A METHOD FOR EMPLOYING SAMPLING TECHNIQUES IN HOUSING SURVEYS

New York (State) Division of Housing
Herman T. Stichman, Commissioner
September 1948

Introduction

The following procedure describes the purposes and techniques of undertaking community housing surveys and contains tables designed to give the non-mathematical reader the sample sizes required to achieve results within a specified level of accuracy in a survey area of known size. This material has been prepared by the Research staff of the Division of Housing to help housing authorities and other civic groups study their housing needs and to answer in adequate detail inquiries from other sources.

The presentation is divided into two sections. The first explains the method of conducting a survey in non-technical language. (Enumeration forms, and detailed instructions as used by the Division of Housing, are available upon request.) The second section gives a technical explanation of the sampling selection procedures employed by the Division. The latter section includes a practical application of the theory to a survey made previously by the Division which indicates that the sampling procedures used are valid and reliable.

Recd 3/24/49

Areal sampling design as outlined herein is of wide interest. Its use is not confined to housing surveys but can be employed whenever population sampling is desired. Persons concerned with market research and with public opinion polls may find this brief guide to sample survey procedures helpful.

This study was undertaken and the report prepared by Arthur Schechter with the collaboration of William Wolman of the Research staff.

Richard W. Hill, Jr., Director
Bureau of Research and Statistics

New York State Division of Housing
270 Broadway
New York 7, New York

The Division of Housing is frequently called upon to provide technical assistance to communities in planning for the redevelopment of blighted and substandard areas. Preliminary to the planning of such a program, it is necessary to know the amount and degree of substandardness, the number of families involved and their characteristics such as size age and income. The federal housing census of 1940 provides block by block housing data for cities of over 50,000 population. This information is helpful in ascertaining the location of blighted areas. However, it is of limited value in determining the facts needed for working out an integrated redevelopment and housing program. Therefore, to obtain the essential area information in a community, a survey should be conducted. Due to time, personnel, and financial considerations, it is generally more practical to undertake a sample survey rather than a complete enumeration since the same conclusions can thus be obtained. The theory and methods of areal sampling design, used extensively by the Bureau of the Census, are of relatively recent origin. Areal sampling methods as used by the Division of Housing in its surveys are special applications of the general sampling theory. The procedure used in conducting such a sample survey consists of the following phases:

a) Selecting the area
b) Block listing
c) Sample selection
d) Household enumeration

Selecting the Area

Blighted areas usually come into existence where the growth of industrial and commercial establishments encroach upon existing residential neighborhoods. These areas are generally found in the older sections of a city. The need for major repairs to residential structures, overcrowded living conditions and the lack of proper sanitary facilities are the general characteristics associated with such blighted or substandard areas. Data from the 1940 housing census can be used to identify such areas in the cities for which block data are available. The exact boundaries of such areas, for survey purposes, can be fixed by visual inspection and in consultation with the local authorities.

Not only do dwelling units in such areas as described above deteriorate into slums but they also hinder necessary commercial and industrial expansion. In the redevelopment of a community, consideration should be given to the expansion needs of business and industry. Emphasis is therefore placed upon surveying contiguous blocks that fall into a redevelopment pattern rather than attempting to cover every block in a city containing either substandard dwelling units or conflicting land uses.

Block Listing

The first phase of the actual survey consists of listing all the land uses in the area under study. From city maps, cards are prepared to identify each block and its street boundaries. These cards are assigned to field workers who are instructed to obtain the following information on forms specially designed for that purpose.

1) The use to which each lot is put (whether residential, commercial or other)
2) The number and type of residential structures by state of repair.
3) The number of dwelling units both occupied and vacant.

Upon completion of block listing, not only is the information that is necessary to the drawing of a random sample available but also a land use picture of the area.

Sample Selection

The number of units to be included in a sample is closely related to the reliability of the results. Generally, the larger the sample size the smaller will be the sampling error. However, there is a point beyond which it is not practical to increase the sample size. The most efficient sample design varies with the size of area and the reliability desired. The reliablility of the sample results is measured by the confidence interval. A 95 per cent confidence interval insures that the chances are 19 out of 20 that percentages derived from a sample would not deviate from the true percentages by more than predetermined amounts. Table 1 gives sample sizes to be employed in a survey of a given area depending upon the accuracy desired. While the choice of a desired accuracy would vary with the nature of a particular survey, for our purposes it is sufficient that sample percentages are within five percentage points of true percentages. This represents the absolute maximum error to be expected. It would be preferable to measure the relative rather than the absolute sampling error. However, the larger sample sizes required to achieve a relative maximum error of 5 per cent could not be handled with the resources presently at the disposal of the Division of Housing. The following example is an illustration of the use of Table 1: An area in a city designated for survey was found, after block listing, to contain 5,000 occupied dwelling units. To achieve an accuracy of five percentage points1/ between sample and true percentages it can be seen that the sample should consist of 370 occupied dwelling units.2/ Then, referring to Table 3 which contains the sampling ratios to be employed under the conditions set forth in this survey it can be seen that one out of every fourteen dwelling units should be enumerated. Starting at random, each fourteenth occupied dwelling unit on the block listing is designated for household enumeration.3/ Table 2 contains

---

1/ This five percentage point leeway represents the maximum deviation to be expected in 19 cases out of 20. In 2 cases out of 3 the maximum deviation would be 2 1/2 percentage points.

2/ For a more detailed description of sample size determination see section entitled "Technical Notes".

3/ The results obtained from a sample selected in this manner will not differ appreciably from those obtained from a random sample.

sampling ratios to be employed when an accuracy to within three percentage points is desired.

Household Enumeration

Interviewers are furnished the addresses of the dwelling units comprising the selected sample. The desired information on housing and family characteristics is then obtained from the occupants of these dwelling units. Call backs are required whenever interviews can not be secured with a responsible member of the household. Due to the nature of the sampling method substitutions should not be made without expert guidance.

The completion of this phase concludes the actual survey. The returns are then tabulated and analyzed.

Technical Notes

Samples are ordinarily drawn from very large populations. Under such circumstances, repeated random drawings with or without replacements do not materially change the probability of a given characteristic. When sampling from relatively small populations, population percentages will be affected materially if the sampling units are selected without replacements. The binomial expansion describes the distribution of percentages when sampling from a large population and this, in turn, is approximated by the normal distribution for large samples. When sampling from a moderately large population without replacements, the distribution of a percentage is described by the hypergeometrical distribution which can be approximated by the normal distribution, with the following restrictions:

1. The sample size should not approach the population size.
2. $\frac{1}{\sqrt{n}}$ and $\frac{1}{\sqrt{N-n}}$ must be sufficiently small to be neglected.
3. np_1 and np_2 should be moderately large.

Notation

n = number in sample

n_1 = number in sample possessing the characteristic being measured

N = number in population

$\bar{p}_1 = \frac{n_1}{n}$ proportion of units in sample possessing the characteristic being measured

p_1 = proportion of units in population possessing the characteristic being measured

$\bar{p}_2 = 1-\bar{p}_1$ and $p_2 = 1-p_1$

The maximum likelihood estimate of $p_1 = \bar{p}_1 = \frac{n_1}{n}$

The standard error of a percentage when sampling without replacements is given by the following expression:

$$\sigma_{\%} = \sqrt{\frac{\bar{p}_1 p_2}{n} \left(\frac{N-n}{N-1}\right)} \quad \text{which is approximately} \quad \sqrt{\frac{p_1 p_2}{n} \left(1-\frac{n}{N}\right)}$$

It can be seen from the above expression for the standard error that as N gets larger with respect to n, $\frac{n}{N} \rightarrow 0$ and $\sigma_{\%} \rightarrow \sqrt{\frac{p_1 p_2}{n}}$

which is the standard error of a percent when sampling with replacements.

For all population and sample sizes, the maximum value of the product $p_1 p_2 = (.5)\,(.5) = .25$, and therefore the maximum standard deviation is equal to the following:

$$\text{Max. } \sigma_{\%} = \sqrt{\frac{.25}{n}\left(1-\frac{n}{N}\right)}$$

The confidence interval of $\bar{p}_1 \pm 2\,\sigma_{\%}$ is used to determine a sample size such that the chances are 19 out of 20 that percentages derived from this sample would not deviate from the population percentages by more than predetermined amounts.1/ Therefore, the probability is at least .95[2/] that the population percentage, p_1, of the characteristic being measured falls in the interval of $\bar{p}_1 \pm 2\,\sigma_{\%}$. This is known as a 95 per cent confidence interval. The confidence interval is not a predetermined constant but a chance variable depending on the sample value $\bar{p}_1$; it is the range of the confidence interval that can be of predetermined size.

1/ $\frac{\bar{p}_1 - p_1}{\sigma_{\%}}$ is approximately normally distributed with mean 0 and variance unity.

2/ .9546 to be exact.

The following formula was used to determine the sample size (n) needed for a maximum confidence interval of range c and with a 95 per cent level of accuracy:

$$4\,\sigma_{\%} = c$$

$$4\sqrt{\frac{(.5)\,(.5)}{n}\left(1-\frac{n}{N}\right)} = c$$

therefore $$n = \frac{4N}{Nc^2 + 4}$$

Table 1 contains the sample sizes necessary to achieve indicated confidence intervals. Where sample sizes in the table are not given, a complete enumeration is advised because the probability of a percentage deviating from the true percentage cannot be accurately determined by the normal curve approximation. The sample sizes needed to achieve the predetermined ranges of the confidence intervals of Table 1 were based on the maximum value of the standard error, i.e. when $p_1 = p_2 = .5$. Therefore, when $p_1 \neq .5$ and the sample sizes of Table 1 are used the range of the confidence interval is decreased. Table 4 gives the percentage decrease in the range of the confidence interval for different values of p_1. For application see footnotes to Tables 2 and 3.

For samples chosen from large populations (over 100,000) it can be seen that to decrease the size of the sampling error from ±5 per cent to ±1 per cent, the sample size must be increased 25 fold, i.e., from 400 to 10,000, since the sampling error varies inversely with the square root of the sample size. Samples over 500 require a staff that is usually beyond the facilities at the disposal of the Division of Housing. Only well trained enumerators should be utilized for field work because errors in obtaining and recording data cannot be as easily evaluated as sampling errors.

A sampling error of approximately 5 percentage points (in 19 cases out of 20)[1] is not considered excessive for the usual types of housing surveys conducted by the Division. Decreasing the sampling error to ±3 per cent is advisable in some studies where greater accuracy is desired. The sampling ratios required for sampling errors of ± 3 per cent and ± 5 per cent have been computed and are shown in Tables 2 and 3. By employing the recommended sampling ratios corresponding to predetermined confidence intervals and population sizes, greater accuracy than specified will result due to over-sampling since fractional units of a sample cannot be selected.

1/ In 2 cases out of 3 the sampling error is $2\frac{1}{2}$ percentage points.

Conditions Under Which It Is More Efficient to Sample Rather Than Conduct A Complete Enumeration

Although the time and cost of taking a sample is greater per dwelling unit actually enumerated than for a complete enumeration it is the overall cost and time and the precision desired for the whole area that determines whether to sample or not.

To Find the Cost of a Sample Survey

Let P_s = total cost of a sample survey

P_1 = cost of listing a d. u.

P_2 = cost of enumerating and tabulating a d.u.

N = total number of d.u.'s in the area

n = number of d.u.'s in the sample

c = range of confidence interval

then $P_s = NP_1 + n\,P_2$

but $n = \dfrac{4N}{Nc^2 + 4}$

Also assume that $10P_1 = P_2$, i.e. it costs ten times more to enumerate and tabulate a dwelling unit than it does to just list a dwelling unit. The Census Bureau found that actually this ratio was closer to 13:1.

$$\text{therefore } P_s = NP_1 + \left(\frac{4N}{Nc^2 + 4}\right) P_2$$

$$\text{since } P_1 = \frac{P_2}{10} \qquad P_s = \frac{NP_2}{10} + \frac{P_2 4N}{Nc^2 + 4}$$

$$(1) \qquad P_s = NP_2\left(\frac{Nc^2+44}{10\,Nc^2+40}\right)$$

which is the total cost of a sample survey.

To Find the Total Cost of a Complete Enumeration

let P_c = cost of a complete enumeration

P_3 = cost of enumerating and tabulating a d.u.

N = total number of d.u.'s in area

then $P_c = NP_3$

Now the cost of enumerating and tabulating a d.u. in a sample survey, P_2, will be larger than the cost of enumerating and tabulating a d.u. for a complete enumeration, P_3. Assume that the cost of P_3 is 3/4 the cost of P_2. The Census Bureau assumes that it costs twice as much to enumerate a sample d.u. than a d.u. for a complete enumeration, but since the Census is dealing with much larger universes the relationship $P_3 = 3/4\ P_2$ seems more applicable in the type of study outlined previously.

$$(2) \qquad \text{therefore} \quad P_c = \frac{3NP_2}{4}$$

As can be seen from equation (1) and (2), $P_s = P_c$ (cost of sample survey equal to cost of a complete enumeration).

$$\text{if} \quad \frac{Nc^2 + 44}{10Nc^2 + 40} = \frac{3}{4}$$

$$N = \frac{56}{26\ c^2} = \frac{2.15}{c^2}$$

Therefore for a precision of within five percentage points

$$(c = 10\%) \qquad N = \frac{2.15}{.01} = 215$$

and for a precision of within three percentage points

$$(c = 6\%) \qquad N = \frac{2.15}{.0036} = 600$$

therefore for $N > 215$ $\quad P_s < P_c$ for $c = 10\%$

and for $N > 600$ $\quad P_s < P_c$ for $c = 6\%$

In conclusion it can be said that in order for it to be more economical to sample there must be more than 215 d.u.'s in the total area if a precision of within 5 percentage points is tolerated and there must be more than 600 d.u.'s in the total area if a 3 percentage point precision is to be acceptable.

Table 1

Sample Sizes Required for Specified Areas
and Confidence Intervals

95% Confidence Interval
(p_1 = .5)

Number of Occupied D.U.'s in Area	$\bar{p}_1 \pm 1\%$ a/	$\bar{p}_1 \pm 2\%$	$\bar{p}_1 \pm 3\%$	$\bar{p}_1 \pm 4\%$	$\bar{p}_1 \pm 5\%$	$\bar{p}_1 \pm 10\%$
500					222	83
1,000				385	286	91
1,500			638	441	316	94
2,000			714	476	333	95
2,500		1,250	769	500	345	96
3,000		1,364	811	517	353	97
3,500		1,458	843	530	359	97
4,000		1,538	870	541	364	98
4,500		1,607	891	549	367	98
5,000		1,667	909	556	370	98
6,000		1,765	938	566	375	98
7,000		1,842	959	574	378	99
8,000		1,905	976	580	381	99
9,000		1,957	989	584	383	99
10,000	5,000	2,000	1,000	588	385	99
15,000	6,000	2,143	1,034	600	390	99
20,000	6,667	2,222	1,053	606	392	100
25,000	7,143	2,273	1,064	610	394	100
50,000	8,333	2,381	1,087	617	397	100
100,000	9,091	2,439	1,099	621	398	100
$\rightarrow \infty$	10,000	2,500	1,111	625	400	100

NOTE: Where sample size is not indicated a complete enumeration is advised

a/ $\bar{p}_1$ = Proportion of units in sample possessing characteristic being measured

Table 2

Sampling Ratios Required for Specified Areas for a Maximum 95% Confidence Interval of $\bar{p}_1 \pm 3\%$[a/]

Number of Occupied D.U.'s in Area	Sample Size	Sampling Ratio	Per Cent of Population Sampled
400		1:1	100%
800		1:1	100%
1,200	577	1:2	50%
1,600	656	1:2	50%
2,000	714	1:2	50%
2,400	759	1:3	33%
2,800	795	1:3	33%
3,200	825	1:3	33%
3,600	849	1:4	25%
4,000	870	1:4	25%
4,400	887	1:5	20%
4,800	902	1:5	20%
5,200	915	1:5	20%
5,600	927	1:6	17%
6,000	938	1:6	17%
6,400	947	1:6	17%
6,800	955	1:7	14%
7,200	963	1:7	14%
7,600	969	1:7	14%
8,000	976	1:8	12%
10,000	1,000	1:10	10%
50,000	1,087	1:45	2%
$\to \infty$	1,111		

a/ $\bar{p}_1$ = Proportion of sample units possessing characteristic being measured.

The range of the 95% confidence interval depends on p_1, the proportion of units in the total area possessing the characteristic being measured: e.g.

p_1	95 Per Cent Confidence Interval
.1 or .9	$\bar{p}_1 \pm 1.8\%$
.2 or .8	$\bar{p}_1 \pm 2.4\%$
.3 or .7	$\bar{p}_1 \pm 2.8\%$
.4 or .6	$\bar{p}_1 \pm 2.9\%$
.5	$\bar{p}_1 \pm 3\%$

Table 3

Sampling Ratios Required for Specified Sized Areas for a Maximum 95% Confidence Interval of $\bar{p}_1 \pm 5\%$[a]

Number of Occupied D.U.'s in Area	Sample Size	Sampling Ratio	Per Cent of Population Sampled
400	200	1:2	50%
800	267	1:3	33%
1,200	300	1:4	25%
1,600	320	1:5	20%
2,000	333	1:6	17%
2,400	343	1:7	14%
2,800	350	1:8	12%
3,200	356	1:9	11%
3,600	360	1:10	10%
4,000	364	1:11	9%
4,400	367	1:12	8%
4,800	369	1:13	8%
5,200	371	1:14	7%
5,600	373	1:15	7%
6,000	375	1:16	6%
6,400	376	1:17	6%
6,800	378	1:18	6%
7,200	379	1:19	5%
7,600	380	1:20	5%
8,000	381	1:21	5%
10,000	385	1:26	4%
50,000	397	1:126	1%
$\rightarrow \infty$	400		

a/ $\bar{p}_1$ = Proportion of units in sample possessing the characteristic being measured.

The range of the 95 per cent confidence interval depends on p_1, the proportion of units in the area possessing the characteristic being measured: e.g.

p_1	95% Confidence Interval
.1 or .9	$\bar{p}_1 \pm 3\%$
.2 or .8	$\bar{p}_1 \pm 4\%$
.3 or .7	$\bar{p}_1 \pm 4.6\%$
.4 or .6	$\bar{p}_1 \pm 4.9\%$
.5	$\bar{p}_1 \pm 5\%$

Table 4

Effect Of Varying Proportions on Confidence Intervals

Proportion p_1	Percentage decrease in range of 95% confidence intervals for sample sizes of Table 1[a] (per cent)
.1	40
.2	20
.3	8
.4	2
.5	None
.6	2
.7	8
.8	20
.9	40

a/ Table 1 gives the necessary sample sizes for a 95 per cent confidence interval when p_1 = .5.

A practical test was made to establish whether the sampling methods described would produce the desired results. A housing and family census had been conducted by the Division of Housing in a substandard section of Schenectady, New York. To apply the sampling techniques used a sample was selected from this complete enumeration by utilizing the procedures outlined above. By comparing the results derived from the sample with those of the complete area, the accuracy of the Division's sampling technique could then be measured.

The area consisted of 942 occupied d.u.'s. The sample size for this study was selected such that the percentages derived from the sample would, in 19 cases out of 20, be within approximately 5 percentage points of the percentages obtained from the complete enumeration. This level of accuracy required a sampling ratio of 1:4.

Using this ratio, it was found that $\bar{p}_1 \pm 5.7$ per cent would be the actual maximum confidence interval to be expected. None of the sample results were found to deviate from the population results by an amount greater than was expected (see Tables 5 to 10). The largest difference between percentages calculated from the complete enumeration and from the corresponding sample percentages was 2.8 percentage points which was well within the 5.7 percentage point leeway permitted.

In some cases, however, percentages based on sub-groups of the sample were found to differ from their corresponding population percentages by more than 5.7 percentage points (see Table 11). This was expected since the predetermined confidence interval was valid only for percentages derived from the entire sample.

The confidence interval for percentages based on a sub-group of a chosen sample will be larger than the confidence interval predetermined for percentages based on the entire sample as can be seen from

Table 1. Therefore if a maximum error of 5 percentage points is to be tolerated, then the size of the smallest sub-group of the sample must not be less than the sample sizes indicated in Table 3.

The Chi Square test was employed to determine whether there was a significant difference between population and sample results. None of the Chi Square values were found to be significant (See Tables 12 to 17). This indicated that the differences between sample and population results were due to chance factors inherent in sampling and not due to improper sampling design. In addition, sample means were found not to differ significantly from the corresponding population means (see Tables 18 and 19).

Table 5

Dwelling Units by Facility

Facility	Total Area	Sample	Difference
Number of Dwelling Units	942	235	
Per Cent	100.0	100.0	
Heating			
Central	71.1	71.5	- .4
Other	28.9	28.5	+ .4
Total	100.0	100.0	
Cooking			
Electric or Gas	96.5	97.0	- .5
Other Types	3.3	3.0	+ .3
None	.2	-	+ .2
Total	100.0	100.0	
Refrigerator			
Mechanical	64.4	66.4	-2.0
Ice Box	30.5	29.8	+ .7
None	5.1	3.8	+1.3
Total	100.0	100.0	
Water Supply			
Running Hot	88.4	90.2	-1.8
Cold only	10.5	9.4	+1.1
None	1.1	.4	+ .7
Total	100.0	100.0	
Flush Toilet			
Private inside unit	76.2	74.5	+1.7
Shared inside unit	8.4	8.9	- .5
Private outside unit	1.3	1.7	- .4
Shared outside unit	14.1	14.9	- .8
Total	100.0	100.0	
Bath			
Private inside unit	69.5	68.5	+1.0
Shared inside unit	8.0	9.4	-1.4
Private outside unit	1.0	1.7	- .7
Shared outside unit	14.3	15.3	-1.0
None	7.2	5.1	+2.1
Total	10010	100.0	

Table 6

Dwelling Units by Tenure

Tenure	Total Area	Sample	Difference
Number of Dwelling Units	942	235	
Percent	100.00	100.00	
Owner occupied	20.0	17.9	+ 2.1
Tenant occupied	80.0	82.1	- 2.1

Table 7

Families by Color

Race	Total Area	Sample	Difference
Number of Families	926	240	
Percent	100.0	100.0	
White	94.8	93.3	+ 1.5
Non-White	5.2	6.7	- 1.5

Table 8

Dwelling Units by Monthly Contract or Estimated Rent

Monthly Contract or Estimated Rent	Total Area	Sample	Difference
Number of Dwelling Units Reporting	939	234	
Per Cent	100.0	100.0	
Under $15	6.2	4.7	+ 1.5
$15 - 19	13.6	13.2	+ .4
$20 - 24	11.0	10.3	+ .7
$25 - 29	11.1	10.7	+ .4
$30 - 39	26.6	26.9	- .3
$40 - 49	19.5	20.1	- .6
$50 - 74	11.5	13.3	- 1.8
$75 and over	.5	.8	- .3
Median Rent	$33.05	$34.13	- $1.08

Table 9

Families by Family Income

Weekly Family Income	Total Area	Sample	Difference
Number of Families Reporting	939	235	
Per Cent	100.0	100.0	
No income	1.7	1.3	+ .4
Under $20	6.0	5.5	+ .5
$20 - 29	9.1	9.4	- .3
$30 - 39	18.2	17.5	+ .7
$40 - 49	23.5	23.8	- .3
$50 -59	16.7	15.7	+ 1.0
$60 - 74	12.1	13.6	- 1.5
$75 - 99	7.8	7.7	+ .1
$100 and over	4.9	5.5	- .6
Median Income	$46.38	$46.85	- $ .47

Table 10

Families by Number of Persons per Family

Number of Persons per Family	Total Area	Sample	Difference
Number of Families	962	240	
Per Cent	100.0	100.0	
One Person	23.7	22.1	+ 1.6
Two Persons	40.4	42.5	- 2.1
Three Persons	17.5	19.6	- 2.1
Four Persons	9.7	10.0	- .3
Five Persons	5.3	2.5	+ 2.8
Six Persons	1.9	2.5	- .6
Seven Persons	.6	.4	+ .2
Eight Persons	.5	.4	+ .1
Nine Persons	.4	-	+ .4
Median family	2.6 persons	2.7 persons	.1 person

Table 11

Families by Income and Number of Persons per Family

Number of Persons per Family

Weekly Income	One Person		Two Persons		Three Persons		Four Persons		Five or more persons	
	Total Area	Sample	Total Area	Sample	Total Area	Sample	Total Area	Sample	Total Area	Sample
Number Reporting	225	53	378	99	165	46	87	23	84	14
Per Cent	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
No Income	5.3	5.7	.5	-	.6	-	1.2	-	-	-
Under $20	16.9	17.0	3.4	4.0	.6	-	2.3	-	2.4	-
$20 - 29	20.5	24.5	6.4	7.1	5.4	2.2	5.7	4.4	1.2	-
$30 - 39	24.4	22.6	16.7	16.2	18.2	19.5	16.1	13.0	10.7	7.1
$40 - 49	19.1	17.0	24.6	28.3	27.9	26.1	23.0	26.1	22.6	7.1
$50 - 59	9.3	11.3	22.5	17.1	18.2	21.7	12.6	13.0	11.9	7.1
$60 - 74	3.2	1.9	16.4	19.2	12.1	10.9	9.2	13.0	20.2	28.6
$75 - 99	.4	-	5.3	7.1	10.3	8.7	21.8	17.4	19.1	21.5
$100 and over	.9	-	4.2	1.0	6.7	10.9	8.1	13.0	11.9	28.6
Median Income	$32.99	$31.24	$49.35	$48.02	$49.03	$51.09	$51.35	$55.00	$60.59	$75.12

Table 12

Chi-Square Test of Goodness of Fit Applied
To Distribution of Monthly Contract or Estimated Rent

Monthly Contract or Estimated Rent	Sample frequencies f_s	Theoretical frequencies f_t	$(f_s - f_t)^2$	$\frac{(f_t - f_s)^2}{f_t}$
Under $15	11	14	9	.64
$15 - 19	31	32	1	.03
$20 - 24	24	26	4	.15
$25 - 29	25	26	1	.04
$30 - 39	63	62	1	.02
$40 - 49	47	46	1	.02
$50 - 59	24	21	9	.43
$60 - 74	7	6		
	9	7	4	.57
$75 and over	2	1		
Totals	234	234		$1.90 = X^2$

A chi-square value of 1.90 for 7 degrees of freedom is not statistically significant. Therefore, the sample may be said to be representative of the area.

Table 13

Chi-Square Test of Goodness of Fit Applied to Distribution of Family Income

Weekly Family Income	Sample frequencies f_s	Theoretical frequencies f_t	$(f_t - f_s)^2$	$\frac{(f_t - f_s)^2}{f_t}$
Under $20	13	14	1	.07
$20 - 29	22	21	1	.05
$30 - 39	41	43	4	.09
$40 - 49	56	56	0	.0
$50 - 59	37	39	4	.01
$60 - 74	32	29	9	.03
$75 - 99	18	18	0	.0
$100 and over	13	12	1	.08
Totals	232	232		.60 = X^2

A chi-square value of .60 for 7 degrees of freedom is not statistically significant. Therefore, the sample may be said to be representative of the area.

Table 14

Chi-Square Test of Goodness of Fit Applied to Distribution of Families by Family Size

Size of Family	Theoretical frequencies f_t	Sample frequencies f_s	$(f_t - f_s)$	$(f_t - f_s)^2$	$\frac{(f_t - f_s)^2}{f_t}$
One Person	56.9	53	3.9	15.2	.27
Two Persons	97.0	102	5.0	25.0	.26
Three Persons	41.9	47	5.1	26.0	.62
Four Persons	23,2	24	.8	.64	.03
Five Persons	12.7	6	6.7	44.9	3.54
Six Persons	4.5	6			
Seven Persons	1.6 } 8.3	1 } 8	.3	.1	.01
Eight Persons	1.2	1			
Nine Persons	1.0	0			
Total	240	240			$4.73 = \chi^2$

A chi-square value of 4.73 for 5 degrees of freedom is not statistically significant. Therefore, the sample may be said to be representative of the area.

Table 15

Chi-Square Test Applied to Distribution of Families by Family Income (One person families only)

Weekly Family Income	Sample frequencies f_s	Theoretical Frequencies f_t	$(f_s - f_t)^2$	$\frac{(f_s - f_t)^2}{f_t}$
No income	3	2.8	.04	.14
Under $20	9	9.0	.0	.0
$20 - 29	13	10.8	4.84	.45
$30 - 39	12	13.0	1.00	.08
$40 - 49	9	10.1	1.21	.12
$50 - 59	6	4.9		
$60 - 74	1	1.6	.09	.12
$75 - 99	0 (7)	.3 (7.3)		
$100 and over	0	.5		
Total	53	53.0		$1.31 = X^2$

A chi-square value of 1.31 for 5 degrees of freedom is not statistically significant. Therefore, the sample may be said to be representative of the area.

Table 16

Chi-Square Test of Goodness of Fit Applied to Distribution of Family Income by Size of Family

(two person families only)

Weekly Family Income	Sample frequencies f_s		Theoretical frequencies f_t		$(f_s - f_t)^2$	$\frac{(f_s - f_t)^2}{f_t}$
No Income	0		.5			
Under $20	4	11	3.4	10.2	.64	.06
$20 - 29	7		6.3			
$30 - 39	16		16.5		.25	.02
$40 - 49	28		24.4		12.96	.53
$50 - 59	17		22.3		28.09	1.26
$60 - 74	19		16.2		7.84	.48
$75 - 99	7	8	5.2	9.4	1.96	.21
$100 and over	1		4.2			
Total	99		99			$2.56 = X^2$

A chi-square value of 2.56 for 5 degrees of freedom is not statistically significant. Therefore, the sample may be said to be representative of the area.

Table 17

Chi-Square Test of Goodness of Fit Applied to Distribution of Family Income (Three person families only)

Weekly Income	Sample frequencies f_s		Theoretical frequencies f_t		$(f_s - f_t)^2$	$\frac{(f_s - f_t)^2}{f_t}$
No Income	0	10	.3	11.5	2.25	.20
Under $20	0		.3			
$20 - 29	1		2.5			
$30 - 39	9		8.4			
$40 - 49	12		12.8		.64	.05
$50 - 59	10		8.4		2.56	.30
$60 - 74	5		5.5		.25	.05
$75 - 99	4	9	4.7	7.8	1.44	.18
$100 and over	5		3.1			
Totals	46		46			$.78 = X^2$

A chi-square value of .78 for four degrees of freedom is not statistically significant. Therefore, the sample may be said to be representative of the area.

Table 18

To Determine a Confidence Interval for the Mean Monthly Contract or Estimated Rent

Monthly Contract or Estimated Rent	Midpt	f_s	(f_s) (Midpt)	d	d^2	$f_s d^2$
$10 - 14	12.50	11	137.50	22.48	505.35	5558.85
$15 - 19	17.50	31	542.50	17.48	305.55	3472.05
$20 - 24	22.50	24	540.00	12.48	155.75	3738.00
$25 - 29	27.50	25	687.50	7.48	55.95	1398.75
$30 - 39	35.00	63	2205.00	.02	.01	.63
$40 - 49	45.00	47	2115.00	10.02	100.40	4718.80
$50 - 59	55.00	24	1320.00	20.02	400.80	9619.20
$60 - 74	67.50	7	472.50	32.52	1057.55	7402.85
$75 - 90	83.00	2	166.00	48.02	2305.92	4611.84
		234	8186.00			40520.97

$\bar{X}_s$ = $34.98 = Average monthly rent

$$\sigma_s = \sqrt{\frac{40520.97}{234}} = \sqrt{173.17} = \$13.16$$

$$\text{Standard Deviation } \bar{X} = \frac{13.16}{\sqrt{234}} = .86$$

95% Confidence interval of mean = $34.98 ± 2 (.86) = $33.26 to $36.70
Population Mean = $33.83

Table 19

To Determine a Confidence Interval for the Mean Family Income

Income	f_s	Midpt	(f_s) (Midpt)	d	d^2	$f_s d^2$
\$10 - 19	13	15.00	195	36.29	1316.96	17120.48
\$20 - 29	22	25.00	5500	26.29	691.16	15205.52
\$30 - 39	41	35.00	1435	16.29	265.36	10879.76
\$40 - 49	56	45.00	2520	6.29	39.56	2215.36
\$50 - 59	37	55.00	2305	3.71	13.76	509.12
\$60 - 74	32	67.50	2160	16.21	262.76	8408.32
\$75 - 99	18	87.50	1575	36.21	1311.16	23600.88
\$100 - 119	13	110.00	1430	58.71	3446.86	44809.18
	232		11900			122748.62

$\bar{X}_s = \frac{11,900}{232} = \$51.29$ = Average family income

$\sigma_s = \sqrt{\frac{122,748.62}{232}} = \sqrt{529.09} = \$23.00$

Standard deviation of $\bar{\bar{X}} = \frac{23}{\sqrt{232}} = \$1.50$

95% Confidence interval of mean = $51.29 \pm 2\ (1.5) = \$48.29$ to \$54.29.

Population mean = \$50.57

New York(State) Division of
Housing.
A Method for Employing Sampling Techniques in...

DATE	ISSUED TO
1-5-73	GENE GRAY 4112 755 53 56
	1-12-73
GAYLORD 40	

DATE	ISSUED TO
	PRINTED IN U.S.A.