

Sample Design, Weighting, and Error Estimation Survey Cycle 18 2021

1. Overview.....	1
2. Sample Design	1
3. Weighting	7
4. Nonsampling Errors	9
5. Sampling Errors and Replicate Weights.....	13
6. Comparability of Data.....	16
7. References	17
Appendix A. Example of Ratio Adjustments.....	18

1. Overview

This document describes the sample design, weighting, and error estimation for the 2021 New York City Housing and Vacancy Survey (NYCHVS). The Census Bureau has reviewed this data product to ensure appropriate access, use, and disclosure avoidance protection of the confidential source data used to produce this product (Disclosure Review Board (DRB) approval number: **CBDRB-FY23-POP001-0095**). The NYCHVS is sponsored by the New York City Department of Housing Preservation and Development (HPD) and conducted by the U.S. Census Bureau.

The City of New York is required by law to conduct a survey periodically to determine if rent regulations should be continued. A primary tool in this decision is the "*vacant available for rent*" rate, which is defined as the ratio of the vacant available for rent units to the total number of renter-occupied and vacant available for rent units for the entire city. The NYCHVS measures rental and homeowner vacancy rates, as well as various household and person characteristics. The design requires the standard error of the rental vacancy rate for the entire city to be no more than one-fourth of one percent if the actual rate was three percent.

The New York City Housing and Vacancy Survey (NYCHVS) is a citywide, representative survey of New York City's housing stock and population that is collected about every three years by the US Census Bureau on behalf of the City of New York. It has been conducted since 1965, making it the longest running housing survey in the country. The 2021 NYCHVS is the 18th survey cycle.

2. Sample Design

The NYCHVS is a longitudinal survey that is conducted about every three years. The main sample of the survey is selected every decade, and additional new units are selected in each subsequent NYCHVS cycle. For the decade 2020-2030, the NYCHVS was conducted in 2021, with plans for additional survey years of 2023, 2026 and 2029. The main sample was selected using the 2020 July Master Address File (MAF); additional sample units plan to be selected in 2023, 2026, and 2029.

2.1. Eligible Universe

The universe of interest for the NYCHVS consists of all residential housing units (HUs) located within the five boroughs of New York City (Bronx, Brooklyn, Manhattan, Queens, and Staten Island). The principal exclusions are living quarters at locations that are classified as group quarters. These include:

- Correctional facilities,
- Mental health institutions,
- Hospitals,
- Military installations,
- Convents, monasteries, and rectories,
- Shelters, group homes, communes, and halfway houses,
- Home for the aged, disabled, homeless, or needy, and
- Dormitories for students or workers.

2.2. Sampling Frames

The 2021 NYCHVS frame is constructed using several files, including the July 2020 Master Address File (MAF) extract, the American Community Survey (ACS) 2014–2018 5-year file, and several administrative files from HPD. Valid, residential HUs on the July 2020 MAF extract comprise the frame for the 2021 NYCHVS sample selection process, and all other files are used to add variables needed for sorting and strata assignments. MAFIDs (the primary HU identifying variable on the MAF) were appended to the administrative files. The administrative files were merged using MAFID with a block-level version of the ACS file and the July 2020 MAF to create the 2021 NYCVHS sample frame.

The frame was created based on the 2020 July MAF instead of the 2020 Census because the 2020 Census was not available at the time the sample was selected.

2.3. Sample Selection

Housing units on the frame were divided into several strata based on subsidized program participation and Condo or Cooperative (Coop) status. A minimum sample size was determined using 2017 NYCHVS data with a target goal of 30,000 sampled units, which was later reduced to 12,000 (due to constraints caused by COVID-19 pandemic). The sampling design used two variables, Strata 1 and Strata 2, each of which had several categories within them. The sample sizes for each category are presented in Tables 2.1, 2.2, and 2.3 below, first shown by the marginal counts in each variable in Tables 2.1 and 2.2 and then by the cross-sectional counts in Table 2.3.

Table 2.1 Strata 1: First Housing Type Allocation

Program Type	Frame Total	30,000 Sample	12,000 Sample
Affordable Owner	21,983	1,144	458
Affordable Renter	92,622	1,205	482
Condo	287,259	2,776	1,110
Coop	405,779	3,650	1,460
Other	2,616,491	17,692	7,077
Public Housing	161,926	2,658	1,063
Rent Control	12,294	880	352

Source: U.S. Census Bureau, July 2020 Master Address File; U.S. Census Bureau, American Community Survey, 2014-2018 5-year file; New York City Department of Housing Preservation and Development administrative files.

Table 2.2 Strata 2: Second Housing Type Allocation

Program Type	Frame Total	30,000 Sample	12,000 Sample
Mitchell Lama Coop	62,669	1,309	524
Mitchell Lama Renter	28,256	1,085	434
Other	2,432,618	19,869	7,947
Rent Stabilized	1,074,811	7,742	3,097

Source: U.S. Census Bureau, July 2020 Master Address File; U.S. Census Bureau, American Community Survey, 2014-2018 5-year file; New York City Department of Housing Preservation and Development administrative files.

Table 2.3 Allocation of Minimum Sample

Strata 1	Strata 2	Frame Total	30,000 Sample	12,000 Sample
Affordable Owner	Mitchell Lama Coop	12,551	562	225
Affordable Owner	Other	7,962	507	203
Affordable Owner	Rent Stabilized	1,470	75	30
Affordable Renter	Mitchell Lama Coop	509	10	4
Affordable Renter	Mitchell Lama Renter	7,486	288	115
Affordable Renter	Other	28,098	288	115
Affordable Renter	Rent Stabilized	56,529	619	248
Condo	Mitchell Lama Renter	219	8	3
Condo	Other	221,661	2,135	854
Condo	Rent Stabilized	65,379	633	253
Coop	Mitchell Lama Coop	49,609	737	295
Coop	Mitchell Lama Renter	1,614	62	25
Coop	Other	290,376	2,290	916
Coop	Rent Stabilized	64,180	561	224
Other	Mitchell Lama Renter	18,937	727	291
Other	Other	1,711,215	11,176	4,470
Other	Rent Stabilized	886,339	5,789	2,316
Public Housing	Other	161,926	2,658	1,063
Rent Control	Other	11,380	815	326
Rent Control	Rent Stabilized	914	65	26

Source: U.S. Census Bureau, July 2020 Master Address File; U.S. Census Bureau, American Community Survey, 2014-2018 5-year file; New York City Department of Housing Preservation and Development administrative files.

A systematic random sample of housing units was selected within each cross-sectional program type in Table 2.3, sorting housing units by

- Borough
- Sub-borough
- Tract
- Median Income by Tract based on 2014–2018 ACS
- Block number
- Binary variable indicating building size as “big” or “small”
- Basic street address
- Unit designation

2.4. Interviews and Response Rates

The total number of sample housing units selected for the 2021 NYCHVS was 12,002. Table 2.4 provides the weighted and unweighted response rates by borough, as well as the distribution of completed interviews and noninterviews.

Table 2.4. Interview Activity for the 2021 New York City Housing and Vacancy Survey

Borough	Unweighted Response Rate	Weighted Response Rate	Selected	Completed Interviews	Type A Non-interviews	Type B & C Non-interviews
Bronx	73%	73%	1,978	1,414	513	51
Brooklyn	74%	74%	3,573	2,501	884	188
Manhattan	74%	74%	3,320	2,354	835	131
Queens	72%	72%	2,607	1,773	683	151
Staten Island	71%	72%	524	352	142	30
Total	73%	73%	12,002	8,394	3,057	551

Source: U.S. Census Bureau, 2021 New York City Housing and Vacancy Survey.

Note: The data are subject to error arising from a variety of sources.

In past cycles (prior to 2017), the NYCHVS conducted proxy or last resort interviews, where a proxy interview consisted of interviewing a real estate agent, building manager, or someone else knowledgeable about the HU and a last resort interview involved accepting an abbreviated questionnaire as complete for reluctant respondents. In 2021, NYCHVS did not conduct proxy or last resort interviews, which resulted in higher noninterviews.

In 2021, Type A noninterviews included occupied housing units where the occupants were:

- Refusals for interview,
- Unable to locate,
- Not at home after repeated visits,
- Unavailable for some other reason, or
- Absent due to COVID and unable to locate.

Type A noninterviews also include vacant units. In these cases, an interview was not obtained if no informed respondent could be found after repeated visits.

Type B noninterviews indicate an interview cannot be conducted due to structural circumstances related to the sample unit. Type B outcome codes occur rarely. Type B noninterviews indicate sample units not currently eligible for interview but could may be eligible for interview in future years (meaning the sample address should be visited in future survey years).

Type C noninterviews only occur when the sample address cannot be found or the sample unit does not meet the definition of a housing unit. Type C outcome codes occur rarely. Type C noninterviews indicate sample units not currently eligible for interview and will not be eligible for interview in future years (meaning the sample address should be removed from the survey).

The response rate is calculated as the total number of interviews divided by the total eligible sample, which can be written as:

$$\text{Response Rate} = \frac{\text{Total Sample} - \text{Type A noninterviews} - \text{Type B\&C noninterviews}}{\text{Total Sample} - \text{Type B\&C noninterviews}}$$

Note that the weighted response rate just applies the base weight of each HU to the counts.

For calculating response rates, enough of the interview had to be completed for it to be considered a valid interview. For vacant interviews, the entire interview must be completed, including type of vacant unit and reason unit is not available for rent or sale. For non-vacant interviews, all of the following must be answered to be considered as a completed occupied interview:

- Occupancy status,
- Tenure,

AND two of the following five items answered from the household roster for each person:

- Sex,
- Age,
- Relationship to householder,¹
- Hispanic origin, and
- Race.

If these criteria were not met, the sampled unit was classified as a Type A noninterview, following the definitions above.

For evaluation of interviews, a second interview was conducted of all vacant units and five percent of all occupied units. The questions asked during the reinterview included information about the previous field representatives that collected data, the time, date, and length of that interview, tenure, and vacancy status.

3. Weighting

To estimate HU and person characteristics for the 2021 NYCHVS, sample weights are calculated for each sample HU and each sample person. The final weight for each HU is the product of the following weights and adjustments:

3.1. Base Weight

The base weight is the reciprocal of the probability of selecting the unit. This is simply the inverse of the rate at which HUs are selected within the cross-sectional breakdowns in Table 2.3. Note that this sample design resulted in an unequal probability of selection for HUs.

3.2. Nonresponse Adjustment

The base weight of each interviewed HU was adjusted to account for the eligible units that did not respond (Type A noninterviews). This nonresponse adjustment – called a noninterview adjustment factor (NAF) – was calculated within groups of similar responding and nonresponding HUs applied using a which was applied to interviewed HUs to account for Type A noninterviews. The factor was calculated using the weighted sums of interviews and Type A noninterviews divided by the weighted sum of interviews within each of these groups:

$$NAF = \frac{\text{Interviews} + \text{Type A noninterviews}}{\text{Interviews}}$$

¹In very rare situations, having only relationship to householder was considered sufficient if the persons could be verified as real people through other methods.

A new method of calculating the NAF was introduced for the 2021 survey cycle. This involved estimating the probability of response from the responding and nonresponding HUs and grouping HUs with similar response propensities together for this adjustment. Note that some nonresponding HUs were excluded from this modeling because they could not be found within the 2020 Census.

3.3. Ratio Adjustment Factors for Housing Unit Weights

New methods for implementing ratio adjustment factors (RAFs) within NYCHVS were also introduced. The HU weights were adjusted using two main sources of known totals:

- The July 2021 MAF
- Totals by program type from HPD administrative files

At each step in the ratio estimation procedure, the factors were equal to the following ratio:

$$\frac{\textit{Known Totals}}{\textit{NYCHVS Sample Estimate}}$$

The denominators of the ratios are equal to the sum of the weights of HUs (or persons) with all previous factors applied. Appendix A includes more information on the ratio adjustment factors and examples on how the process works.

The updated process creates three partitions of the sample based on borough membership, subsidized program status, and Condo/ Coop, similar but not identical to the sample design. The weighting program then iterates through these three partitions until the RAF factor at each iteration stabilizes and final estimated totals of the groups within these partitions equal their known totals. The partitions are listed below.

- Partition 1: Affordable Owners, Affordable Renters, Remainder – City-wide
- Partition 2: Mitchell Lama Renter, Mitchell Lama Coop, Remainder – by Borough
- Partition 3: Public Housing, Condo, Coop, Remainder – by Borough

Note that Partition 1 was done city-wide while Partitions 2 and 3 were done by borough. “Remainder” means any HU not fitting into the other categories in the partition. Estimates of total HUs made of these particular program types at the specified level of geography match their known totals.

3.4. Ratio Adjustment Factors for Person Weights

When calculating person weights, the final housing unit weight was used as the base weight for each person, then we added a ratio adjustment to account for sampling variability and known coverage deficiencies (as described in Section 4.1.) for persons other than reference

persons, spouses or married partners within interviewed households. We computed this factor within each borough by age, race, Hispanic Origin, and sex (200 cells), but reduced this number of cells to 89 based on how high the initial factors were and how few cases there were in some of the original 200 cells.

- The numerator of the ratio equaled the independent estimate of 2021 total persons for the cell minus the NYCHVS sample estimate of reference persons and spouses or unmarried partners. The independent estimates were projected based on 2020 Census person totals (Census, 2021) to estimate the population of NYC in April, 2021.
- The denominator of the ratio equaled the NYCHVS sample estimate of persons other than reference persons, spouses, or unmarried partners for the cell. The person ratio estimate factor was applied only to the persons other than reference persons, spouses, or unmarried partners.

The ratio estimation procedures, as well as the overall estimation procedure, reduced the sampling error for most statistics in comparison to what would have been obtained by simply weighting the sample by the base weight.

4. Nonsampling Errors

All numbers from the NYCHVS, except for sample size, are estimates. As in other surveys, two types of general errors occur: sampling errors and nonsampling errors. Sampling errors are discussed in Section 5. The definition of nonsampling errors is—

Nonsampling errors arise mainly due to misleading definitions and concepts, inadequate sampling frames, unsatisfactory questionnaires, defective methods of data collection, tabulation, coding, incomplete coverage of sample units, and so on. These errors are unpredictable and not easily controlled. Unlike sampling error, this error may increase with increases in sample size. If not properly controlled, nonsampling error can be more damaging than sampling error for large-scale household surveys.²

The various types of nonsampling errors are discussed in the following sections.

4.1. Coverage Error

Coverage errors arise from the failure to give some units in the target population any chance of selection into the sample (undercoverage) or giving units more than one chance of selection (overcoverage). To calculate the coverage, we used the sample base weight,

² https://unstats.un.org/unsd/demographic/meetings/egm/Sampling_1203/docs/no_7.pdf

which is the weight prior to any sample adjustments. The sample adjustments, described in Section 3, help to mitigate the undercoverage identified in this section.

The coverage rate is the ratio of the survey population or HU estimate of a group or an area and the independent estimate (or the known totals). The undercoverage rate is calculated as:

$$\text{Undercoverage} = \left(1 - \frac{\text{NYCHVS Sample Estimate}}{\text{Known Totals}} \right) * 100$$

Table 4.1a indicates the estimated undercoverage for both HUs and persons. Based on the July 2021 MAF, we missed about seven and a half percent of the housing units in the five boroughs. Overall, we missed just under fifteen percent of the people in sample households.

Table 4.1a Overall Undercoverage for Housing Units and Persons

	Estimated from 2021 NYCHVS Base Weights	Known Total	Undercoverage
Housing Units	3,369,717*	3,644,065	7.53%
Persons	7,124,692	8,354,047	14.72%

Source: U.S. Census Bureau, 2021 New York City Housing and Vacancy Survey.

Note: The data are subject to error arising from a variety of sources.

*Sum of base weights for interviews and Type A noninterviews.

Table 4.1b provides the various sources of undercoverage for HUs.

Table 4.1b Undercoverage by Source for Housing Units

Source of Undercoverage	Undercoverage
Growth from July 2020 MAF to July 2021 MAF	1.26%
Type B&C Drops	5.26%
Nonrespondents not matched to 2020 Census	1.01%
Total	7.53%

Source: U.S. Census Bureau, 2021 New York City Housing and Vacancy Survey internal data files.

The within-household undercoverage varied by age, race, sex, and borough. Table 4.1c gives the undercoverage of the various race-sex groups for the city as a whole.

Table 4.1c Undercoverage by Race/Ethnicity-Sex Group

Race/Ethnicity-Sex Group	Undercoverage
White & Other Females	12.97%
White & Other Males	16.28%
African American Females	17.35%
African American Males	28.94%
Asian Females	-1.00%
Asian Males	-2.84%
Hispanic Females	14.08%
Hispanic Males	21.55%

Source: U.S. Census Bureau, 2021 New York City Housing and Vacancy Survey.

Note: The data are subject to error arising from a variety of sources.

We adjusted for this undercoverage through the HU and person ratio adjustment factors. These factors adjust the sample weights to population totals provided by the Census Bureau, so the resulting final weight accounts for the undercoverage identified in Tables 4.1a, 4.1b, and 4.1c. For more information on the sample adjustment process, see Sections 3.3 and 3.4. NYCHVS data users do not have to take any additional steps to account for coverage error.

4.2. Nonresponse Error

Some respondents refuse the interview or cannot be located. The Census Bureau mitigated the error due to nonresponse by applying the noninterview adjustment factors into the weighting process, as discussed in Section 3.2. NYCHVS data users do not have to take any additional steps to account for nonresponse error.

4.3. Measurement Error from Missing Responses to Questions

Some respondents participate in an interview but refuse to answer questions or do not know a particular answer. For certain questions, the Census Bureau imputes missing responses. When imputing, the Census Bureau tries to find households or persons with similar characteristics to fill in missing data. For each imputation, records are divided into 'donors' and 'recipients'.

For the demographic items, the Census Bureau first tries to impute based on other household information or household members. Every household must have some

demographic information for each person in the household or it would be made a Type A. It is rare that a household is missing all demographic information for one item.

For imputing the housing items (including housing quality questions), units with similar characteristics are grouped. For example, when imputing contract rent, a unit with a similar year moved, year-built range, number of bedrooms, and input control status (stabilized, public housing, unregulated, etc.) is found and unit's contract rent is used to impute the recipient's contract rent. If no such unit can be found, contract rent is imputed based on the median value for units in 2017 (adjusted for inflation) with the same input control status in the particular borough. In 2021, contract rent was imputed in 11.6 percent of the renter-occupied units. For some characteristics (like mortgage information), where we didn't collect similar data in the 2017 NYCHVS, we used NYC data from the 2019 American Housing Survey with similar housing characteristics, for starting median values when, in rare cases no donor could be found. For example, we based median starting values for first mortgage payments by grouping AHS 2019 data into purchase price and condo/Coop/other owner status in NYC.

For economic items, such as income and employment status, the best possible match between donors and recipients is achieved through a statistical match with key items. The items used for matching donors and recipients are public assistance/non-public assistance, borough, tenure, gender, race, ethnicity, age, relationship, education, worked last week, hours worked, type of work, government/nongovernment, and rent/value. All of these criteria are used to get the best statistical match possible. There are 33 income variables in 2021; in rare cases where a suitable donor was not found, income amount is imputed based on the median value of that income category. All donors and recipients have the same borough, tenure, and either receive public assistance or do not.

Variables that can be used to determine imputation rates are in the public use files (PUF) and are defined on the record layout and PUF User Guide and Codebook. For example, using these variables from the PUF, users can see that summer gas and electricity costs were imputed for 11.4 percent of occupied units, age was imputed for 6.3 percent of all persons, and stories was imputed for 0.1 percent of vacant units.

The Census Bureau does not know how close the imputed values are to the actual values. For other items, "not reported" is used as an answer category. NYCHVS data users do not have to take any additional steps to account for measurement error from missing responses to questions.

4.4. Quality Validity Error

In order to design a survey question that accurately measures the constructs of interest, the Census Bureau carefully tests each new survey question to ensure it is measuring the construct of interest. While the questionnaire is provided in multiple languages, sometimes the respondent does not speak those languages. In these cases, the interview must be rescheduled so that a field representative that speaks the same language as the

respondent can administer the interview. Although some respondents might misinterpret questions, the Census Bureau does not have any additional information to estimate validity error rates. NYCHVS data users do not have to take any additional steps to account for validity error.

4.5. Processing Error

After the data are collected, errors that can be introduced include data capture errors and data editing and imputation errors. The Census Bureau carefully tests all aspects of the data capture and the editing and imputation procedures. Although mistakes are possible, the Census Bureau believes they are minimal. If a processing error is discovered, the Census Bureau will let NYCHVS data users know and, in some cases, will publish revised estimates. NYCHVS data users do not have to take any additional steps to account for processing error.

4.6. Additional Considerations

The NYCHVS is a longitudinal survey conducted about every three years. Many NYCHVS users compare current year NYCHVS with prior year estimates. Users should be aware that HPD and the Census Bureau often make changes to the text of various questions between surveys and sometimes to the underlying weighting methodology or sample design. NYCHVS data users comparing estimates with prior year surveys should consult the 'Overview' document on the NYCHVS website (<https://www.census.gov/programs-surveys/nychvs/about/overview.html>).

5. Sampling Errors and Replicate Weights

Sampling error is a measure of how estimates from a sample vary from the actual value. By the term "actual value," we mean the value we would have gotten had all HUs been interviewed, under the same conditions, rather than only a sample.

Users of NYCHVS PUF should use replicate weights to estimate errors for any estimate. This is different from prior survey cycles in 2010, where the Census Bureau provided replicate weights as well as generalized variance function (GVF) parameters as an alternative method to estimate variance.³ GVFs are not available for this release and will not be made available for future releases of NYCHVS.

Variance estimation for surveys refers to the variation of an estimate due to selecting a sample from the set of all possible samples for a given sample design. To estimate the variance, multiple samples are needed but only one is observed. Replication allows small

³ 2017 was the first NYCHVS survey cycle to include replicate weights upon release. 2011 and 2014 had replicate weights released after their initial release.

changes to a single probability sample to create a set of replicate samples, which can then be used to measure the variation of the estimates. Replication is done through subsets selected from the original sample in a process that mimics the original sample design. Each replicate sample, r , is then fully weighted, using the same process as the original sample, to ensure it represents the population of interest. This process forms the set of final replicate weights $\{w_r \mid r = 1, \dots, R\}$. Considering a particular estimate of interest, each replicate weight, w_r , can be used to create a replicate estimate $\hat{\theta}_r$. The set of replicated estimates $\{\hat{\theta}_r \mid r = 1, \dots, R\}$ represents the variability, or dispersion, of the estimate of interest under multiple samples of the population.

The Hadamard matrix was used to derive replicate factors to apply to individual full sample weights in creating replicate weights. Please note that for 2021 NYCHVS, the weights in Replicate 1 equal full sample weights, the weight used to derive sample estimates.

The 2021 NYCHVS uses a replicate variance estimator derived from a variance equation called the successive differences estimator. This estimator was first introduced by Fay and Train (1995) and then expanded for replication by Ash (2014). Using the replicated estimates, data users can calculate an estimated variance of an estimate of interest using the replication variance estimator:

$$\hat{v}(\hat{\theta}) = \frac{4}{80} \sum_{r=1}^{80} (\hat{\theta}_r - \hat{\theta}_0)^2$$

where $\hat{\theta}$ is the weighted point estimate of the statistic of interest, such as a total, median, mean, proportion, regression coefficient, or log-odds ratio, using the weight for the full sample and $\hat{\theta}_r$ is the replicate estimate for replicate r of the same statistic using the replicate weights. $\hat{\theta}_0$ is the full sample estimate. The value of 80 in $\hat{v}(\hat{\theta})$ is the number of replicates used (NYCHVS uses 80 replicates).

There are two sets of replicate weights. One set of replicate weights is used for computing standard errors of housing unit characteristics and the second set is used for computing standard errors of person characteristics.

To calculate a standard error, the measure of dispersion when parameter estimates are calculated through repeated sampling from the population, obtain the square root of the variance estimate.

The following example illustrates how a statistic would be estimated, replicated, and combined to form a variance estimate. The goal of this example is to estimate the total number of renter-occupied HUs in Queens for 2021 and its corresponding estimate of variance.

For 2021, there are 893 completed interviews that are renter-occupied HUs in Queens (sample cases in Queens with responses to tenure status question as renters). Table 6.1 displays the first four and last one renter-occupied sample units in Queens. Note that the ordering in Table 6.1 is based on the variable *CONTROL*.

Table 6.1: Example of Estimating Variances with Replication of Renters in Queens

Sample HU	Full Sample Weight	Replicate Weights				
		Replicate 1	Replicate 2	Replicate 3	...	Replicate 80
1	432.920	432.920	430.180	724.691	...	128.270
2	1,223.797	1,223.797	2,071.712	351.266	...	365.787
3	36.056	36.056	35.119	11.083	61.436
4	476.777	476.777	470.671	793.062	...	485.903
...
893	432.899	432.899	125.991	724.655	...	747.576

Source: U.S. Census Bureau, 2021 New York City Housing and Vacancy Survey.

Note: The data are subject to error arising from a variety of sources.

In NYCHVS, the full sample weight and the full sample estimate are referred to as replicate weight 0 (w_0) and replicate estimate 0 ($\hat{\theta}_0$), respectively.

Step 1: Calculate the full sample weighted survey estimate.

The statistic of interest is the total number of renter-occupied housing units in Queens for 2021. Add the full sample weights of the sample cases that meet your criteria of interest. Therefore, the total number of renter-occupied housing units in Queens is calculated as follows:

Full Sample Renter-Occupied HUs in Queens Estimate:

$$\hat{\theta}_0 = \sum_{i=1}^{893} w_{0,i} = 432.920 + 1,223.797 + \dots + 432.899 = 467,730$$

Step 2: Calculate the weighted survey estimate for each of the replicate samples.

The replicate survey estimates are as follows:

$$\begin{aligned}
 \text{Rep 1 Rent-Occ Estimate } \hat{\theta}_1 &= \sum_{i=1}^{893} w_{1,i} = 432.920 + 1,223.797 + \dots + 432.899 = 467,730 \\
 \text{Rep 2 Rent-Occ Estimate } \hat{\theta}_2 &= \sum_{i=1}^{893} w_{2,i} = 430.180 + 2,071.712 + \dots + 125.991 = 468,933 \\
 \text{Rep 3 Rent-Occ Estimate } \hat{\theta}_3 &= \sum_{i=1}^{893} w_{3,i} = 724.691 + 351.266 + \dots + 724.655 = 476,408 \\
 &\vdots \\
 &\vdots \\
 &\vdots \\
 \text{Rep 80 Rent-Occ Estimate } \hat{\theta}_{80} &= \sum_{i=1}^{893} w_{80,i} = 128.270 + 365.787 + \dots + 747.576 = 469,330
 \end{aligned}$$

Step 3: Use the replicate estimates $\hat{\theta}_r$ in the formula below to calculate the variance estimate for the total renter-occupied HUs in Queens.

$$\begin{aligned}
 \hat{v}(\hat{\theta}) &= \frac{4}{80} \sum_{r=1}^{80} (\hat{\theta}_r - \hat{\theta}_0)^2 \\
 &= 0.05 \times [(467,730 - 467,730)^2 + (468,933 - 467,730)^2 + (476,408 - 467,730)^2 \\
 &\quad + \dots + (469,330 - 467,730)^2] \\
 &= 0.05 \times [0 + 1,449,144 + 75,322,223 + \dots + 2,560,800] \\
 &= 117,126,870
 \end{aligned}$$

The survey estimate for total renter-occupied population in Queens is 467,730 housing units, with an estimated variance of 117,126,870 or a standard error of 10,823 housing units, rounded to the nearest whole number.

6. Comparability of Data

Data obtained from the NYCHVS are not necessarily comparable to other surveys or data sources due to differences in survey processes, question text, how the information were obtained, mode of data collection, or timing of data collection, among other reasons. These differences are examples of nonsampling variability and are not reflected in the standard errors. Caution should be used when NYCHVS comparing results to other data sources.

The NYCHVS was redesigned for the longitudinal panel that begins with 2021. The redesign included the selection of housing units for the new longitudinal panel, improved training on field operations, and revised questionnaire content. Notably, beginning in 2021 the survey was conducted via computer-assisted personal interviewing (CAPI), whereas previously it was a paper survey questionnaire completed by personal interview. The transition to CAPI allowed for response category edit checks and consistency checks among questionnaire items within the laptop instrument.

7. References

- Fay, R. E. and Train, G. F. (1995). Aspects and Survey and Model-based Postcensal Estimation of Income and Poverty Characteristics for States and Counties. *Proceeding of the Sections on Government Statistics*, American Statistical Association, 154-159.
- Ash, S. E. (2014) Using Successive Difference Replication for Estimating Variances. *Survey Methodology*, Statistics Canada, Catalogue no.12-001-X Business Survey Method Division, Vol. 40, No.1, pp.47-59.

Appendix A. Example of Ratio Adjustments

This appendix provides one hypothetical example that demonstrates how the sample weights are ratio-adjusted so that they are consistent with a set of control totals.

For this example, assume weights were calculated for a sample, including all weighting adjustments up to a nonresponse adjustment. With these weights, totals by two categories – simply identified as A or B for Category 1 and C or D for Category 2 – were created. Table A1 summarizes the estimated totals resulting from the hypothetical sample and weights, and Table A2 shows the hypothetical control totals.

Table A1: Example Estimated Totals

Cat1\Cat2	C	D	Total
A	110	91	201
B	97	107	204
Total	207	198	405

Table A2: Example Control Totals

Cat1\Cat2	C	D	Total
A	115	105	220
B	95	105	200
Total	210	210	420

The control totals of Table A2 are used to improve the weights by making the estimates from the weights consistent with the control totals. Table A3 shows the Ratio Adjustment Factor (RAF) that will make the estimated totals consistent with the control totals.

Table A3: Example Ratio Adjustment Factors

Cat1\Cat2	C	D
A	$115/110 = 1.0455$	$105/91 = 1.1583$
B	$95/97 = 0.9794$	$105/107 = 0.9813$

If the factors from Table A3 are multiplied to the weights of the sample units, then the estimates from the revised weights will be consistent with the totals of Table A2.

For example, the ratio-adjusted weights for the combination of Category 1 = A and Category 2 = C is the product of the original weight and the RAF for the A/C combination:

$$\text{Ratio-adjusted weight} = \text{original weight} \times 1.0455$$

The ratio-adjusted weights for the other combinations of Categories 1 and 2 are calculated in the same way, using the corresponding RAF for each combination.

