# 2017 New York City
# Housing and Vacancy Survey

# Sample Design, Weighting, and Error Estimation

Updated: August 1, 2018

U.S. Census Bureau, Department of Commerce
New York City Department of Housing Preservation & Development

This page is intentionally blank.

# Table of Contents

Appendixes

This page is intentionally blank.

# 1. Overview

The purpose of this document is to describe the sample design, weighting, and error estimation for the 2017 New York City Housing and Vacancy Survey (NYCHVS). The NYCHVS is sponsored by the New York City Department of Housing Preservation and Development (HPD) and conducted by the US 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 estimate of the vacant available for rent rate for the entire city be no more than one-fourth of 1 percent, if the actual rate was 3 percent.

# 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 2010-2020, the NYCHVS was conducted in 2011, 2014, and 2017. The main sample was selected after the 2010 Decennial Census, and then additional sample units were selected in 2011, 2014, and 2017.

## 2.1. Eligible Universe

The universe of interest for the NYCHVS consists of the 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:

- Transient hotels and motels (those that are less than 75% residential),
- 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.

Residential hotels and motels (those that are at least 75% residential); however, are included in the survey.

## 2.2. Sampling Frames

The 2017 NYCHVS sample consists of housing units selected from the following four sampling frames:

1. Housing Units Included in the 2010 Census

This frame was created from the decennial census from both the Decennial Hundred-Percent File (DHF) and the Census Unedited File (CUF). Both the DHF and CUF are created from the decennial response file, which contains all responses to the 2010 Census. The CUF contains the unedited individual responses to the 2010 Census questionnaire, while the DHF contains the edited responses to the 2010 Census questionnaire. The reason we used these files instead of Master Address File (MAF) is the DHF and CUF had the most current data at the time of our sample selection.

The sample selected from this frame is referred to as the initial and main sample of the 2017 NYCHVS. This sample was included in all three cycles of NYCHVS: 2011, 2014, and 2017. The sample housing units for the 2017 NYCHVS were initially selected from this frame in 2011.

HPD obtained, prepared and provided to the Census Bureau three address list files, which were used as additional sample frames for the 2017 NYCHVS:

2. Housing Units Newly Constructed Since the 2010 Census

This list was based on New Construction data maintained by the NYC Department of City Planning that integrates regular updates from the NYC Dept. of Buildings (DoB). The New Construction list included only unique addresses with Final Certificates of Occupancy issued by the DoB for newly constructed residential units since the last HVS cycle (from December 2013 through November 2016). Addresses were cleaned and prepared by HPD using the City's Geo-Support system to validate house number and street name addresses, to provide a valid Building Identification Number (BIN), BBL, and number of residential units, to eliminate duplicates and invalid or pseudo-BINs. Addresses without valid BINs were excluded.

The resulting frame contained units in buildings constructed after the 2010 Census. This included units constructed prior to the 2011 survey (eligible for the 2011 survey), units constructed since the 2011 survey but prior to the 2014 survey (eligible for the 2014 survey), as well as units constructed since the 2014 survey (eligible for the 2017 survey).

3. Housing Units from Alterations and Conversions

To create this list, HPD obtained from DoB the latest available Alterations File, which was current as of November 30, 2016 inclusive, and extracted listings for addresses where DoB recorded a signed-off Alteration Permit dated between December 1, 2013 and

November 30, 2016.  Only addresses where the records indicated a net gain in number of residential units were extracted, along with the number of newly created units.  By processing the addresses using the City Planning Geo-Support system, each address was linked to its unique BIN.  Only addresses with valid BINs were retained; addresses with duplicate or pseudo-BINs were dropped.  The calendar year of the final sign-off was recorded.

This frame is similar to the newly constructed frame (frame #2 listed above) and contains residential units newly created in existing buildings since the 2010 Census. It includes housing units constructed since the 2010 Census in preexisting buildings altered to create more units (alterations) or converted from nonresidential use (conversions).  This includes units constructed prior to the 2011 survey (eligible for 2011 survey),  units constructed since the 2011 survey but prior to the 2014 survey (eligible for 2014 survey), as well as units constructed since the 2014 survey (eligible for 2017 survey).

### 4. Housing Units in Structures Owned by New York City (*in rem*)

This frame consisted of units in structures owned by New York City as of November 2016.  The City owned these units because the owner failed to pay the real estate tax and/or other charges on the property.  HPD is the City agency responsible for administering the inventory of in rem buildings. There were historically two different administrative groups of in rem buildings, those centrally managed by the Division of Property Management (DPM) and those in programs where responsibility for maintaining and/or upgrading the buildings was delegated to different community organizations or groups.  The stock of in rem buildings is a dwindling universe. Over the years buildings that were in various earlier in rem programs were transferred into other HPD programs for rehabilitation or management.  In order to maintain comparability of the sample frames through the subsequent HVS cycles, subsequent or re-named programs still containing “legacy” in rem buildings were identified and the addresses were consolidated into a “DPM” list and a “community” list.  Only buildings with zero residential units or demolished building addresses were removed.  The remaining “legacy” in rem addresses were processed through the City Planning Geo-Support system, so that each address could be linked to a unique BIN, BBL, etc.  However, addresses that could not be linked to a unique BIN were not removed, because HPD confirmed all addresses on the list as a City-owned residential building.

The domain for this frame changes in every sampling cycle, since some new units are added, and some units get sold and are no longer *in rem*. In the 2017 NYCHVS, the units selected from this frame supplemented the *in rem* sample from previous sample years that are still *in rem*.

The frame size for 2017 was 212 buildings containing roughly 1,700 HUs. The city requires a sample size of approximately 600-700 units in the sample.  Thus, these types of housing units were oversampled to ensure a large enough sample for analysis of this sub-universe.

The HUs of the *in rem* frame may also be part of the 2010 Census frame. We accounted for the overlapping frames by adjusting the probability of selection of units in both frames, and thereby their base weights as suggested by Lohr (2007, 2010).

### 2.3. Sample Selection by Frame

Within each of the four NYCHVS sampling frames, we selected clusters (groups of housing units) of generally four housing units, with the exception of some *in rem* buildings where we selected clusters of five.

Frame 1: Housing Units Included in the 2010 Census

The sample for this frame came from two different Census 2010 files – the DHF and the CUF. The sample from the DHF was selected first and the CUF was second. To ensure no HU was selected in both files, HUs in the CUF were removed if they were already in the DHF.

Within this frame, we sorted housing units by

- Borough,
- Sub-borough,
- Percent renter occupied in the block,
- Tract,
- Block number,
- Basic street address, and
- Unit designation.

We selected a systematic random sample of housing units across all boroughs from the ordered frame. This frame included *in rem* units.

Frame 2: Housing Units Newly Constructed Since the 2010 Census

We selected units in this frame from Certificates of Occupancy (C of Os) issued between April 2010 and November 2016. The list of C of Os was provided by HPD for each survey cycle.

Sample units were selected for the 2011 survey from Certificates issued between April 2010 and November 2010. Additional sample units were selected for the 2014 survey from Certificates issued between December 2010 and November 2013. Additional sample units were selected for the 2017 survey from Certificates issued between December 2013 and November 2016. Units selected from the C of O frame in 2011 and 2014 remained in the sample for 2017.

For the 2017 NYCHVS, we dropped from this frame all housing units that were also on the 2010 Census frame, or previous 2011, 2014 C of O lists. We sorted the housing units by borough, street name, and street number, and then selected a sample of housing units within each borough.  We listed each structure that contained a sample housing unit and then identified the designated sample unit in the order in which the unit appeared on these listings.

### Frame 3: Housing Units from Alterations and Conversions

Housing units added to existing residential buildings (alterations) and housing units in buildings converted from nonresidential use (conversions) were sampled for the 2017 survey. The selection process was conducted for the 2011 and 2014 surveys, and was conducted again, using updated lists from HPD, for the 2017 survey. Addresses were identified by HPD where residential units were created through alterations or conversions with signed off permits between December 2013 and November 2016.

The list of alteration and conversion addresses was matched to the 2017 C of O frame list for newly constructed buildings and to the 2010 Census on basic address.  For matching addresses, the unit counts were compared between the city's alteration and conversion list and the new construction C of O or Census 2010 list.  If the alteration and conversion listing for the address contained more units than the new construction C of O or the Census list, it was considered an alteration and eligible for the alteration sample.  If the alteration and conversion listing for the address contained the same or fewer units than the new construction C of O or the Census list, it was dropped from the alteration and conversion frames because those units should have been accounted for in the C of O or the Census list first.  If the address did not match, the building was considered a conversion and included in the conversion frame.

Within each frame, a sample of buildings was selected.  The sampled buildings went through a listing process from which sample units were identified.  For the alterations sample, a determination was made about which units were not included in the Census or the new construction C of O file.  These units were then eligible for the alterations sample. For the buildings identified as conversions, all units listed were eligible for the conversion sample.

### Frame 4: Housing Units in Structures Owned by New York City (*in rem*)

The *in rem* frame is a special domain identified by HPD. The oversample of *in rem* HUs is selected in each cycle of the survey from a frame that is updated in each cycle of the survey.  The main requirement is that the *in rem* universe is oversampled with a sample size of 600-700 units each sample cycle.

This frame is the most complicated, because the *in rem* universe changes each sample cycle (2011, 2014, and 2017); some units remained in the frame, some new units came in, and some units dropped out.  In 2011, a HU that is *in rem* could be selected into the sample from two different frames: the 2010 Census frame and the 2011 *in rem* frame. For 2017, the third sample cycle of the NYCHVS in the decade, a HU that is *in rem* could be selected into the sample from four different frames:  the 2010 Census frame, and each of the three *in rem* frames (2011 *in rem* universe, 2014 *in rem* universe, and 2017 *in rem* universe).

If the sampled buildings selected in previous surveys, 2011 or 2014, did not drop out of the sample for 2017, the sample units selected within those buildings will continue to remain in the *in rem* sample for 2017. We only selected additional *in rem* units to replace the lost *in rem* sample units from 2014.

We selected a probability-proportional-to-size sample of *in rem* buildings first, then selected sample units within buildings. In this procedure, each building is assigned a probability of selection based on the expected number of housing units in the building. This probability is in direct proportion to this expected number of units. Thus, a building with eight units has twice the probability of selection as a building that has four units.

First, we sorted the buildings by:

- Borough,
- Street name,
- House number

We next selected a systematic random sample of buildings from the ordered frame. Then we listed the individual units in each building, and last we selected a sample of units within each sample building.

### 2.4. Sample Size

The total number of sample housing units selected for the 2017 NYCHVS was 19,020. Table 2.3 provides the total number of selected housing units by borough, as well as the breakdown of completed interviews and non-interviews.

*Table 2.3. Interview Activity for the 2017 NYCHVS*

| Borough | Unweighted response rate | Weighted response rate | Selected | Completed Interviews | Type A Non-interviews | Type C Non-interviews |
|---|---|---|---|---|---|---|
| Bronx | 77% | 76% | 2,863 | 2,168 | 661 | 34 |
| Brooklyn | 83% | 83% | 5,494 | 4,459 | 914 | 121 |
| Manhattan | 83% | 82% | 5,165 | 4,229 | 870 | 66 |
| Queens | 78% | 78% | 4,529 | 3,489 | 975 | 65 |
| Staten Island | 83% | 83% | 969 | 790 | 162 | 17 |
| Total | 81% | 80% | 19,020 | 15,135 | 3,582 | 303 |

Of these 19,020 total sampled housing units, 15,135 interviews were completed. The NYCHVS classifies two types of non-interviews: Type A and Type C.

In 2017, there were 3,582 Type-A non-interviews. These include occupied housing units where the occupants:

- Refused to be interviewed,
- Were not at home after repeated visits, or
- Were unavailable for some other reason.

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

There were an additional 303 Type-C non-interviews, which were not interviewed because they no longer existed or were uninhabitable.

This classification produced an overall unweighted response rate of 81 percent (19,020-3,582-303)/ (19,020-303) = (15,135/18,717). The response rate is calculated as the total number of interviews divided by the total eligible sample, which can written as:

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

Note the response rate using the base weight is 80 percent. For calculating response rates, all of the following must be answered to be considered as a completed interview:

- Occupancy/vacancy status,
- Year moved,
- Coop/condo status,
- Tenure,
- Units in structure,
- Interviewer observations of building condition items,
- Contract rent,
- Type of vacant unit, and
- Asking rent

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

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

If these criteria were not met, the sampled unit was classified as either a Type A or Type C non-interview, following the definitions above.

Interviews started in January of 2017 and continued through May 2017 and survey operation was conducted out of the Census Bureau's New York Regional Office. We hired over 400 field representatives who were tested and trained for this survey. In addition, we filled other positions including field supervisors, automation clerks, administrative clerks, recruiters, and an overall program coordinator. Most of the respondent interviews were personal visits, but sometimes respondents did not agree to a personal interview and in these cases a telephone interview was conducted.

For evaluation of falsification 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 (FR) that collected data, the time, date, and length of that interview, tenure, and vacancy status.

In 2017, we did not conduct proxy or last resort interviews as in past surveys. Last resort interviews in past surveys were conducted for reluctant respondents in which we designated certain questions as essential and accepted an abbreviated questionnaire as complete. Essential items in the past included tenure, rent, vacancy status, year moved, demographic information, among other items. In the past, proxy interviews were conducted when we were not able to conduct an interview, after numerous attempts. Proxy interviews were interviews with a real estate agent, building manager, or someone who knew something about the housing unit.

## 3. Weighting

In order to estimate housing unit and person characteristics based on the data we collected for the 2017 NYCHVS, we calculated sample weights for each sample housing unit, and each sample person. The final weight for each housing unit is the product of the following weights and adjustments:

### 3.1. Base Weight

We determined the base weight as the reciprocal of the probability of selecting the unit. Because *in rem* sample units and some census sample units were eligible for selection from both the 2010 Census and the *in rem* frames, we adjusted the base weights to reflect the fact that housing units had multiple chances of selection given our overlapping frames.

### 3.2. Nonresponse Adjustment

We adjusted the base weight of each interviewed housing unit to account for the 3,582 eligible units that did not respond (Type A non-interviews). We applied this nonresponse adjustment using a non-interview adjustment factor (NIAF). This factor was applied to all interviewed housing units to account for Type-A non-interviews. The factor was calculated using the following ratio:

$$NIAF = \frac{Interviews + \ Type\ A\ noninterviews}{Interviews}$$

We computed the factor separately for old construction (those are sample HUs selected from the 2010 Census) and new construction (those are HUs added to the sample after the 2010 Census) housing units as follows:

Old Construction

For sample housing units selected from the 2010 Census frame, we computed the NIAF separately by borough using the characteristics listed in Table 3.2. We used 2017 NYCHVS response data when available to determine the tenure and characteristics cell of a unit. If the 2017 NYCHVS data were not available, we used 2014 NYCHVS response data. If 2014 NYCHVS response data were not available, we used 2011 NYCHVS response data. If 2011 NYCHVS response data were not available, we used the 2010 Census data. If the 2010 Census data were also not available, we treated it as a vacant housing unit, and assigned it to the "HU without tenure or vacancy status" (unknown) cell of the vacant housing unit table.

The process of determining the tenure and characteristics cell of a unit was different in the 2014 survey cycle or prior. Prior to 2017, we used 2011 NYCHVS response data to determine the tenure status. If 2011 NYCVHS response data were not available, we used 2010 Census data. If 2010 Census data were also not available, we used the current 2014 NYCHVS response data. Starting in 2017, we simplified this process by using the most current values available.

Table 3.2 summarizes the variables used in combination to define cells of the NIAF tables for old construction sample units.

*Table 3.2 Variables Used to Define Nonresponse Adjustment Cells for Old Construction*

| **NIAF Tables by HU Type** | **Variable** | **Values** | | |
|---|---|---|---|---|
| Renter-Occupied Housing Units | Borough | Bronx<br>Brooklyn | Manhattan<br>Queens | Staten Island |
| | Monthly Rent | < $100<br>$100-$199<br>$200-$299 | $300-$399<br>$400-$499<br>$500-$599 | $600-$699<br>$700-$999<br>$1,000+<br>Unknown |
| | Number of Rooms | 1-2, 3, 4-5, 6+, Unknown | | |
| Owner-Occupied Housing Units | Borough | Bronx<br>Brooklyn | Manhattan<br>Queens | Staten Island |
| | Value of the House | < $25,000<br>$25,000-$49,999<br>$50,000-$74,999<br>$75,000-$99,999 | $100,000-$149,999<br>$150,000-$199,999<br>$200,000-$249,999<br>$250,000-$299,999 | $300,000-$399,999<br>$400,000-$499,999<br>$500,000+<br>Unknown |
| | Number of Rooms in the Housing Units | 1-5, 6-7, 8, 9+, Unknown | | |
| Vacant Housing Units | Borough | Bronx<br>Brooklyn | Manhattan<br>Queens | Staten Island |
| | Vacancy Status | Rented/sold/vacant for rent/vacant for sale | Seasonal/Occasional<br>Migrant workers<br>Other | Unknown |
| | Number of Rooms | 1-2, 3, 4-5, 6+, Unknown | | |

New Construction, Alterations, and Conversions

For new construction units, alterations, and conversions, we computed the factor separately using the year the segment was selected (2011, 2014, or 2017) and borough.

### 3.3. Ratio Adjustment Factors for Housing Unit Weights

We adjusted the housing unit weights using three ratio adjustments with the following known totals:

- The 2010 Census frame totals,
- The *in rem* frame totals,
- Housing unit totals produced by demographic analysis

For each ratio estimation procedure, we computed and applied factors separately by cells. The factors were equal to the following ratio:

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

The denominators of the ratios are equal to the sum of the weights of housing units (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 three ratio adjustments are described below:

1. 2010 Census Ratio Adjustment Factor (RAF)

This ratio adjustment accounted for differences between the 2010 Census counts and the Census characteristics of the corresponding weighted sample counts. The factor reduces the variability resulting from sampling the 2010 Census frame. We adjusted the weights of all NYCHVS sample units selected from the 2010 Census frame, and computed the factors separately by borough using the following 2010 Census characteristics. Table 3.3 summarizes the variables used in combination to define cells of the Census ratio adjustment.

*Table: 3.3 Variables Used to Define Census Ratio Adjustment Cells*

| **RAF Table by HU Type** | **Variable** | **Values** | |
|---|---|---|---|
| Renter-Occupied Housing Units | Sub-borough | Bronx: 1-10<br>Brooklyn: 1-18<br>Manhattan: 1-10 | Queens: 1-14<br>Staten Island: 1-3 |
| | Number of Persons in the Housing Units | 1, 2, 3-4, 5 or more | |
| | Race and Ethnicity of the Householder | White (non-Hispanic)<br>African American (non-Hispanic)<br>Other (all remaining races) | |
| Owner-Occupied Housing Units | Sub-borough | Bronx: 1-10<br>Brooklyn: 1-18<br>Manhattan: 1-10 | Queens: 1-14<br>Staten Island: 1-3 |
| | Number of Persons in the Housing Units | 1, 2, 3-4, 5 or more | |
| Vacant Housing Units | Type of Vacancy | Vacant for rent<br>Vacant for sale<br>Rented/sold | Seasonal<br>Migrant<br>Other |
| | Borough | Bronx<br>Brooklyn<br>Manhattan | Queens<br>Staten Island |

2. *In rem* Ratio Adjustment

This procedure adjusts for known sampling variability in the *in rem* sample selection. We adjusted the weights of all sample units selected from the *in rem* frame by borough (five cells). We used the total number of units in each borough in the *in rem* frame as control totals.

3. 2017 Housing Unit Ratio Adjustment

This procedure adjusted the 2017 NYCHVS sample estimate for sampling variability and housing unit undercoverage (as described in Section 4.1.) by controlling the sample estimate using independent estimates of 2017 total housing units. The independent estimates were projected to 2017 based on 2010 Census housing unit totals. The independent estimates were derived from the Census Bureau's demographic population estimates program and are used here to correct for the coverage error (for more information, see Census, 2017a). We applied this ratio estimation procedure to all interviewed housing units. We calculated the ratio adjustment factor for each of the boroughs (five cells). The independent estimates were counts of the total number of housing units in each of the boroughs at the time of the 2017 survey.

### 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 within interviewed households. We computed this factor within each borough by age, race, Hispanic Origin, and sex (200 cells).

- The numerator of the ratio equaled the independent estimate of 2017 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 2010 Census person totals (Census, 2017a).

- 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

Since the statistics produced from this survey are estimates derived from a sample, they will differ from the "true values" being estimated. There are two types of errors, which cause estimates based on a sample survey to differ from the true value - sampling error and nonsampling error.

If every housing unit in New York City were interviewed, the estimates of housing unit characteristics could still differ from the true value (for example, the median contract rent). In this instance, the difference is due solely to nonsampling errors. We attribute nonsampling errors in sample surveys to many sources:

- Deficiencies in the sampling frame (i.e., not all housing units are covered),
- Inability to pick up all persons within sample households,
- Inability to obtain information about all cases in the sample,
- Definitional difficulties,
- Differences in the interpretation of questions,
- Inability or unwillingness to provide correct information on the part of the respondents, and
- Mistakes in recording, coding or keying the data obtained.

There are also other errors of collection, response, processing, coverage, and estimation for missing data.

### 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, or 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 housing unit estimate of a group or an area and the independent estimate (or the known totals). The undercoverage rate is calculated as:

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

Based on the Census population estimate for NYC in March 2017, in the 2017 NYCHVS, we missed less than one percent of the housing units in the five boroughs. Overall, we missed about twelve percent of the people in sample households. (See Table 4.1a)

*Table 4.1a Overall Undercoverage for HUs and Persons*

| | **Estimated from NYCHVS** | **Known Total** | **Undercoverage** |
|---|---|---|---|
| Housing Units | 3,469,240 | 3,489,271 | 0.58% |
| Persons | 7,519,528 | 8,418,512 | 11.96% |

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

*Table 4.1b Undercoverage by Race/Ethnicity-Sex Group*

| Race/Ethnicity-Sex Group | Undercoverage |
|---|---|
| White & Other Females | 2% |
| White & Other Males | 3% |
| African American Females | 18% |
| African American Males | 22% |
| Asian Females | 6% |
| Asian Males | 4% |
| Hispanic Females | 18% |
| Hispanic Males | 29% |

We adjusted for this undercoverage through the housing unit and person ratio adjustment factors. These factors adjust the sample weights to population totals provided by the US Census Bureau, the resulting final weight accounts for the undercoverage identified in Table 4.1. 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. We mitigated the error due to nonresponse by applying the non-interview 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 many housing, demographic, and economic questions the Census Bureau imputes missing responses. When imputing, we try 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 we first try to impute based on other household information. Every household must have some demographic information or it would be made a Type C. It is rare that a household is missing all demographic information for one item.

For imputing the housing items, we look for units with similar characteristics. For example, when imputing contract rent, we find a unit with a similar year moved, year built range, units in structure, and input control status (stabilized, public housing, unregulated, etc.) and use that unit's contract rent to impute the recipient's contract rent. If no such unit can be found, we impute contract rent based on the median value for units with the same input control status in the particular borough. In 2017, contract rent was imputed in only four percent of the renter occupied units.

For economic items, we try to achieve the best possible match between donors and recipients through a statistical match with key items. The items used for matching donors and recipients are sex, race, ethnicity, age, relationship, education, worked last week, temporarily absent, looking for work, year last worked, kind of business, type of business, industry, occupation, weeks worked, hours worked, and rent/value. We use all of these criteria to get the best statistical match possible.

All donors and recipients have the same borough, tenure, and either receive public assistance or do not. Appendix B provides the list of variables being imputed.

Variables that can be used to determine imputation rates are in the public use files (PUF) and are defined on the record layout. These variables are shown, beginning on page 23 for occupied units, page 33 for persons, and page 43 for vacant units. For example, using these variables from the PUF, users can see we imputed electricity for 5.6 percent of occupied units, we imputed age for 2.9 percent of all persons, and we imputed stories for 5.4 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 nonresponse error.

### 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 we have an English and Spanish questionnaire, sometimes the respondent does not speak either one of these languages. In these cases, the interview must be rescheduled so that a FR 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

The 2017 NYCHVS was administered using a paper instrument. This requires more processing than most other Census surveys, which are completed using a computerized survey instrument. For each interview, the survey data are keyed and verified by our

National Processing Center, and then transmitted to our programming area while the data are reviewed and edited.

After the data are collected, errors that can be introduced include data capture errors, data coding and classification errors, and data editing and imputation errors. The Census Bureau carefully tests all aspects of the data capture, coding, classification, 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 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 small changes to the text of various questions between surveys. NYCHVS data users comparing estimates with prior year surveys should consult the ‘Overview’ document on the NYCHVS website (See Census 2014 and Census 2017b).

## 5. Sampling Errors

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 housing units been interviewed, under the same conditions, rather than only a sample.

Users of NYCHVS PUF can use replicate weights to estimate errors for any estimate. For further information, see Section 6.

The Generalized Variance Functions (GVFs) are a convenient tool for quick and easy estimation of sampling errors. The text below describes how to calculate sampling errors for counts, percentages, differences, medians, and means using GVFs.

### 5.1. Sampling Error for Counts

Most published estimates from the NYCHVS reflect weighted counts of housing units. The error from sampling for a weighted count is approximated using the following GVF for estimating a 90-percent confidence interval:

$$1.645 * \sqrt{a * \hat{X} + b * \hat{X}^2},$$

Where $\hat{X}$ is the weighted sample estimate from the file, and $a$ and $b$ are GVF parameters that vary depending on the characteristics being estimated.

Sets of GVF parameters have been computed for New York City as a whole, as well as for each of the five boroughs.  For 2017 NYCHVS, the GVF parameters are now calculated using replicate weights, see Section 6 for more information on replicate weights.

Table 5.1a contains GVF parameters for computing standard errors of housing unit characteristics and Tables 5.1b and 5.1c contain the GVF parameters for computing standard errors of person characteristics.

### Housing Unit Characteristics

The parameters provided in Table 5.1a identify three sets of GVF parameters for housing units of NYC and each of the five boroughs.  Use Table 5.1a and Appendix Table C1 and C2 to decipher which GVF parameters to use for a given housing unit characteristic. Table C1 identifies the characteristics to be used with the second set of parameters. Table C2 identifies the characteristics to be used with the third set of parameters.  For a given estimate, consider the geography first and then refer to tables C1 and C2.  If the characteristic can be matched to either table then use the parameters associated with that table. The first column in Tables C1 and C2 lists the characteristics for which the tables are to be applied. The second column lists the applicable subgroup (e.g. total occupied, vacant for rent, etc.)  If the estimate of interest matches to both the first and second column of either table, use the corresponding GVF parameters.  If the characteristic of interest is not identified in either Table C1 or C2 then use the first set of GVF parameters. Exhibit 5.1 illustrates how to correctly select the right set of GVF parameters for calculating sampling errors.

For sub-borough estimates, the sub-borough is treated as a characteristic and can be found on table C1 or C2 depending on the borough.  Match the borough and the characteristic to table C1 and C2 to determine which set of parameters to use.

### Person Characteristics

The parameters provided in Table 5.1b and Table 5.1c identify seven sets of GVF parameters for person estimates of NYC and each of the five boroughs.  To help determine which parameter set to use for a given person estimate, first consider the geography then identify matching characteristics. If no characteristics can be matched to the ones listed then use the parameters identified for “person characteristics not listed above”.  If multiple sets of characteristics can be matched then use the set of parameters yielding the higher standard error.

For sub-borough estimates, find the parameters given for the borough and use the parameter for the person characteristics “Borough and Sub-borough”.

*Exhibit 5.1: Decision Tree on How to Determine Which Set of GVF Parameters to Use*

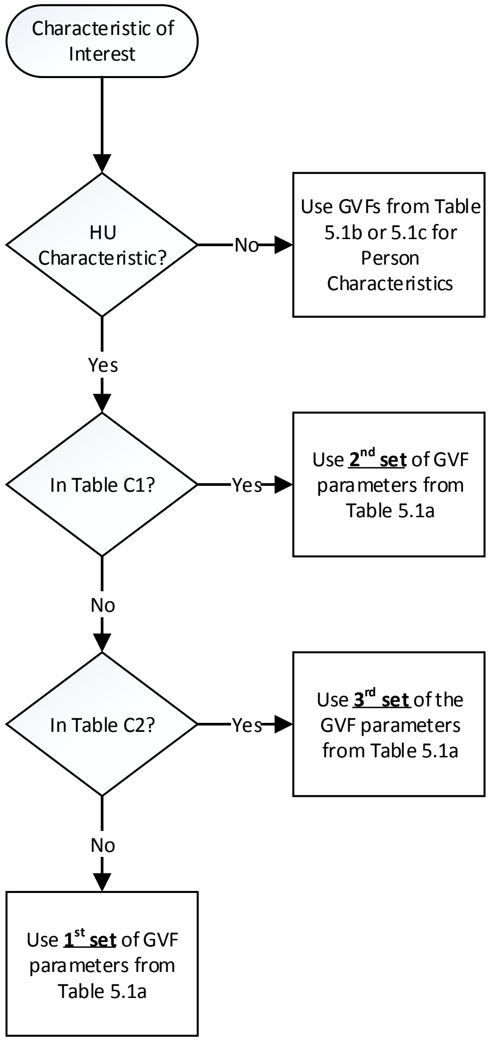
Characteristic of Interest
HU Characteristic?
No
Use GVFs from Table 5.1b or 5.1c for Person Characteristics
Yes
In Table C1?
Yes
Use **2nd set** of GVF parameters from Table 5.1a
No
In Table C2?
Yes
Use **3rd set** of the GVF parameters from Table 5.1a
No
Use **1st set** of GVF parameters from Table 5.1a

*Table 5.1a The Value of a and b Parameters for Housing Unit Characteristics by Borough*

| Borough | HU Characteristics … | *a* | *b* |
|---|---|---|---|
| City Wide | … not listed in C1, C2 | 284.23 | -0.000082 |
| | … listed in table C1 | 368.42 | -0.000106 |
| | … listed in table C2 | 304.08 | -0.000087 |
| Bronx | … not listed in C1, C2 | 322.97 | -0.000613 |
| | … listed in table C1 | 404.18 | -0.000767 |
| | … listed in table C2 | 380.36 | -0.000722 |
| Brooklyn | … not listed in C1, C2 | 296.17 | -0.000286 |
| | … listed in table C1 | 378.88 | -0.000366 |
| | … listed in table C2 | 312.72 | -0.000302 |
| Manhattan | … not listed in C1, C2 | 312.90 | -0.000354 |
| | … listed in table C1 | 396.23 | -0.000449 |
| | … listed in table C2 | 385.20 | -0.000437 |
| Queens | … not listed in C1, C2 | 297.82 | -0.000348 |
| | … listed in table C1 | 378.94 | -0.000443 |
| | … listed in table C2 | 334.69 | -0.000391 |
| Staten Island | … not listed in C1, C2 | 315.61 | -0.001746 |
| | … listed in table C1 | 373.75 | -0.002071 |
| | … listed in table C2 | 469.29 | -0.002604 |

*Table 5.1b The Value of a and b Parameters for Person Characteristics for City Wide*

| Borough | Person Characteristics | *a* | *b* |
|---|---|---|---|
| City Wide | White and Other Race Ethnicity | 544.38 | -0.000034 |
| | Males | 441.75 | -0.000030 |
| | Female | 459.51 | -0.000050 |
| | Under 25 years old and other special characteristics[1] | 396.74 | 0.000157 |
| | African Americans, American Indians or Native Alaskans | 607.77 | 0.000127 |
| | Borough and Sub-borough[2] | 604.47 | -0.000054 |
| | Person characteristics not listed above | 518.85 | -0.000021 |

[1] Special characteristics include: retired, income less than $20,000, highest education level is H.S. diploma and not enrolled in any other education, and self-employed.

[2] Exclude total population in households. Use the set of GVF parameters for “characteristics of persons not listed above” for these person characteristics.

*Table 5.1c The Value of a and b Parameters for Person Characteristics by Borough*

| Borough | Person Characteristics | *a* | *b* |
|---|---|---|---|
| Bronx | White and Other Race Ethnicity | 576.83 | -0.000168 |
| | Males | 452.41 | -0.000195 |
| | Female | 478.43 | -0.000139 |
| | Under 25 years old and other special characteristics[3] | 393.93 | 0.000705 |
| | African Americans, American Indians or Native Alaskans | 564.87 | 0.000848 |
| | Borough and Sub-borough[4] | 643.66 | -0.000138 |
| | Person characteristics not listed above | 512.98 | -0.000046 |
| Brooklyn | White and Other Race Ethnicity | 511.72 | 0.00002 |
| | Males | 432.17 | 0.000022 |
| | Female | 426.81 | -0.000109 |
| | Under 25 years old and other special characteristics[3] | 372.57 | 0.000429 |
| | African Americans, American Indians or Native Alaskans | 583.45 | 0.000556 |
| | Borough and Sub-borough[4] | 596.27 | -0.000048 |
| | Person characteristics not listed above | 470.45 | 0.000000 |
| Manhattan | White and Other Race Ethnicity | 525.09 | -0.000076 |
| | Males | 427.18 | -0.000225 |
| | Female | 406.88 | -0.000119 |
| | Under 25 years old and other special characteristics[3] | 342.23 | 0.001138 |
| | African Americans, American Indians or Native Alaskans | 401.88 | 0.000832 |
| | Borough and Sub-borough[4] | 570.34 | -0.000141 |
| | Person characteristics not listed above | 445.07 | -0.000062 |
| Queens | White and Other Race Ethnicity | 547.76 | 0.000001 |
| | Males | 444.81 | -0.000113 |
| | Female | 479.12 | -0.000194 |
| | Under 25 years old and other special characteristics[3] | 440.31 | 0.000473 |
| | African Americans, American Indians or Native Alaskans | 574.96 | 0.000290 |
| | Borough and Sub-borough[4] | 640.30 | -0.000117 |
| | Person characteristics not listed above | 515.26 | -0.000063 |
| Staten Island | White and Other Race Ethnicity | 492.36 | -0.000106 |
| | Males | 380.83 | -0.000526 |
| | Female | 436.90 | -0.000571 |
| | Under 25 years old and other special characteristics[3] | 356.57 | 0.001552 |
| | African Americans, American Indians or Native Alaskans | 521.66 | 0.010359 |
| | Borough and Sub-borough[4] | 568.62 | -0.000398 |
| | Person characteristics not listed above | 455.15 | -0.000157 |

[3] Special characteristics include: retired, income less than $20,000, highest education level is H.S. diploma and not enrolled in any other education, and self-employed.

[4] Exclude total population in households. Use the set of GVF parameters for "characteristics of persons not listed above" for these person characteristics.

The parameters in Table 5.1a, 5.1b, and 5.1c, citywide and for each borough, allow you to compute a range of error such that there is a known probability of being correct if you say the actual value is within the range. The error formulas are approximations to the errors. They indicate the order of magnitude of the errors rather than the actual errors for any specific characteristic. To construct the range, add and subtract the error computed from the formulas to the estimate.

We will continue with an example using the equation we provided for estimating the sample error for counts. In the 2017 NYCHVS, there are 22,537 vacant-for-rent units in Brooklyn, that is *A*=22,537. To compute a 90-percent confidence interval, you would use the first set of GVF parameters in Table 5.1a and you would compute the margin of error as follows:

$$1.645 * \sqrt{(296.17 * 22{,}537) + (-0.000286 * 22{,}537^2)} = 4{,}203$$

Thus, there is a 90-percent chance you will be correct if you conclude the actual number of vacant-for-rent units in Brooklyn is 22,537 plus or minus 4,203 or in the range 18,334 to 26,740.

If the estimate involves two characteristics from Tables 5.1a, 5.1b or 5.1c, use the set of GVF parameters with the larger *a* parameter.

### 5.2. Sampling Error for Percentages

Any subgroup can be shown as a percentage of a larger group. The error from sampling for a 90 percent confidence interval for this percentage is computed as:

$$1.645 * \sqrt{\frac{a * P * (100 - P)}{A}}$$

where:

$a$: the parameter *a* from Table 5.1a, 5.1b or 5.1c,<br>
$P$: is the percent you calculate, and<br>
$A$: is the weighted denominator of the percent.

For example, there are 580,484 occupied home owner conventional housing units in New York City and 130,487, or 22.5 percent, were built between 1947 and 1973. Use Table 5.1a for City Wide, together with Table C1 and C2 in Appendix C. Since the characteristic (year building built) is listed in Table C2, the applicable subgroups for this characteristic do not include occupied home owner conventional housing units, you must use the first set of the parameters from Table 5.1a. To compute a 90-percent confidence interval you would plug the following numbers into the above formula:

$$1.645\sqrt{\frac{284.23 * 22.5 * 77.5}{580{,}484}} = 1.52$$

Thus, if you say that the actual percentage of occupied home owner conventional housing units in New York City built between 1947 and 1973 is between 20.98 percent and 24.02 percent, there is a 90-percent chance you will be correct.

### 5.3. Sampling Error for Differences

People often ask whether two numbers are actually different. Two numbers from the NYCHVS, for example, 21 and 34, or 34 percent and 55 percent, have a statistically significant difference if their 90-percent confidence intervals do not overlap.
When 90-percent confidence intervals do overlap, numbers are still statistically different if the result of subtracting one from the other is more than:

$$1.645 * \sqrt{\sigma_1^2 + \sigma_2^2}$$

Where:

$\sigma_1$: the standard error for the first number
$\sigma_2$: the standard error for the second number

This formula is quite accurate for (a) the difference between estimates of the same item in two different areas or (b) the difference between separate and uncorrelated items in the same area. If there is a high positive correlation between the two items, the formula will overestimate the error. If there is a high negative correlation, the formula will underestimate the error. The following illustration shows how to compute the error of a difference.

There are 5,603 condominium housing units in Queens with 20 to 49 units in the building and 7,605 condominium housing units in Queens with 50 to 99 units in the building. Follow the steps in Table 5.3a to compute the 90-percent confidence interval for the difference between those two numbers.

*Table 5.3a Steps to Compute the 90% Confidence Interval for a Difference*

| Steps for Calculations | The Formula | An Example |
|---|---|---|
| Which set of GVF parameters should we use? (since the characteristic of interest is units in Condo building in Queens, and this matches to both columns in Table C2 of Appendix C, use the third set of the parameters for Queens from Table 5.1a) | $a$<br>$b$ | 334.69<br>-0.000391 |
| How many total units in Queens with 20-49 units in the building? | $\widehat{X_1}$ | 5,603 |
| What's the estimated standard error of total units in Queens with 20-49 units in the building? | $\sigma_1 = \sqrt{a \times \hat{X}_1 + b \times \hat{X}_1^2}$ | $\sqrt{334.69 \times 5{,}603 - 0.000391 \times 5{,}603^2}$<br>$= 1{,}364$ |
| How many total units in Queens with 50-99 units in the building? | $\widehat{X_2}$ | 7,605 |
| What's the estimated standard error of total units in Queens with 50-99 units in the building? | $\sigma_2 = \sqrt{a \times \hat{X}_2 + b \times \hat{X}_2^2}$ | $\sqrt{334.69 \times 7{,}605 - 0.000391 \times 7{,}605^2}$<br>$= 1{,}588$ |
| What's the difference of the two numbers you are interested in? | $Diff = \widehat{X_2} - \widehat{X_1}$ | 7,605-5,603 = 2,002 |
| What is the margin of error for a 90-percent confidence interval for the difference? | $ME = 1.645 * \sqrt{\sigma_1^2 + \sigma_2^2}$ | $1.645 * \sqrt{1{,}364^2 + 1{,}588^2} = 2{,}612$ |
| The 90-percent confidence interval for the difference is: | $Diff \pm SE$ | $2{,}002 \pm 2{,}612$ |

Thus, a 90-percent confidence interval of (-610, 4,614) includes zero. Therefore, the difference between condominium housing units in Queens with 20 to 49 units and 50 to 99 units is not statistically significant.

Here, we demonstrate how to compare the same estimate of two boroughs. For example, we want to know whether the estimated number of rent stabilized housing units in the Bronx is different from the Manhattan estimate. Table 5.3b provides the steps that compute the 90-percent confidence interval for the difference between those two numbers.

*Table 5.3b Steps to Compute the 90% Confidence Interval for a Difference*

| Steps for Calculations | The Formula | An Example |
|---|---|---|
| Which set of GVF parameters should we use? (since the characteristic of interest is units with stabilized rent in Bronx and Manhattan, and this matches neither tables of Appendix C, use the first set of the parameters for Bronx and Manhattan from Table 5.1a) | Bronx:<br>$a_1$<br>$b_1$<br>Manhattan:<br>$a_2$<br>$b_2$ | Bronx:<br>322.97<br>-0.000613<br>Manhattan:<br>312.90<br>-0.000354 |
| How many total units in Bronx are rent stabilized? | $\widehat{X_1}$ | 233,502 |
| What's the estimated standard error of total units in Bronx with stabilized rent? | $\sigma_1 = \sqrt{a_1 \times \hat{X}_1 + b_1 \times \hat{X}_1^2}$ | $\sqrt{322.97 \times 233{,}502 - 0.000613 \times 233{,}502^2}$ $= 6{,}480$ |
| How many total units in Manhattan are rent stabilized? | $\widehat{X_2}$ | 249,000 |
| What's the estimated standard error of total units in Manhattan with stabilized rent? | $\sigma_2 = \sqrt{a_2 \times \hat{X}_2 + b_2 \times \hat{X}_2^2}$ | $\sqrt{312.9 \times 249{,}000 - 0.000354 \times 249{,}000^2}$ $= 7{,}481$ |
| What's the difference of the two numbers you are interested in? | $Diff = \widehat{X_2} - \widehat{X_1}$ | 249,000 - 233,502 = 15,498 |
| What is the margin of error for a 90-percent confidence interval for the difference? | $ME = 1.645 * \sqrt{\sigma_1^2 + \sigma_2^2}$ | $1.645 * \sqrt{6{,}480^2 + 7{,}481^2} = 16{,}281$ |
| The 90-percent confidence interval for the difference is: | $Diff \pm SE$ | $15{,}498 \pm 16{,}281$ |

Thus, a 90-percent confidence interval of (-783, 31,779) includes zero, so we conclude that the difference between rent stabilized housing units in the Bronx and Manhattan is not statistically significant.

### 5.4. Sampling Error for Medians

The median is the value 50-percent of the way through the distribution. Thus, 50-percent of the total falls below and 50-percent falls above the median. Note that the median presented in this example is the true median (i.e., computed by statistical package) not an approximation. You can construct a confidence interval around the median by computing the standard error on a 50-percent characteristic and then translating that into an interval for the characteristic.

Steps to compute the sampling errors for medians:

1) First, get the estimated standard error of a 50-percent characteristic, using the same formula for errors for percent (section 5.3), but substitute 50 for the *P*:

$$\sqrt{\frac{a * 50 * (100 - 50)}{A}} = \sigma$$

2) Then, calculate the standard error from sampling for the median as:

$$(U - L) * \frac{\sigma}{p} = SE_{median}$$

where:

*a:* is parameter *a* from Table 5.1a, 5.1b or 5.1c.
*A*: is the total number of housing units from the distribution.
*U-L*: is the width of the interval that contains the median. U is the upper bound of the interval, and L is the lower bound of the interval.
σ: is the error for a 90-percent confidence interval for the 50-percent characteristic.
*p*: is the percent of cases that fall in the interval containing the median.

3) Last, calculate a 90 percent confidence interval for the true median by adding and subtracting to the median:

$$Median \ \pm 1.645 * SE_{median}$$

For example, the median household income for all occupied housing units in New York City is $57,500. The number of occupied housing units in the distribution of household income is presented in the Table 5.4a.

*Table 5.4a: Distribution of Household Income*

| Household Income | Number of HUs | Percent | Cumulative Percent |
|---|---|---|---|
| Less Than $5,000/no income/loss | 166,827 | 5.36 | 5.36 |
| $5,000-$9,999 | 139,239 | 4.48 | 9.84 |
| $10,000-$14,999 | 161,977 | 5.21 | 15.05 |
| $15,000-$19,999 | 155,689 | 5.01 | 20.06 |
| $20,000-$24,999 | 156,732 | 5.04 | 25.10 |
| $25,000-$29,999 | 131,104 | 4.22 | 29.31 |
| $30,000-$34,999 | 128,896 | 4.14 | 33.46 |
| $35,000-$39,999 | 122,756 | 3.95 | 37.40 |
| $40,000-$49,999 | 216,410 | 6.96 | 44.36 |
| $50,000-$59,999 | 201,798 | 6.49 | 50.85 |
| $60,000-$69,999 | 185,237 | 5.96 | 56.81 |
| $70,000-$79,999 | 148,787 | 4.78 | 61.59 |
| $80,000-$89,999 | 138,940 | 4.47 | 66.06 |
| $90,000-$99,999 | 120,038 | 3.86 | 69.92 |
| $100,000-$124,999 | 262,477 | 8.44 | 78.36 |
| $125,000-$149,999 | 164,428 | 5.29 | 83.65 |
| $150,000 or more | 508,620 | 16.35 | 100.00 |
| TOTAL | 3,109,955 | 100.00 | |

The error on a 50-percent characteristic based on 3,109,955 units is calculated as illustrated in the Table 5.4b.

*Table 5.4b. Steps to Compute the 90% Confidence Interval for a Median Household Income*

| **Steps for Calculations** | **The Formula** | **An Example** |
|---|---|---|
| How many total units is the median based on (in thousands, exclude “not reported” and “don’t know”)? | $A$ | 3,109,955 |
| What’s the parameter *a*? (since household income is not a characteristic listed on either Table C1 and C2 of the Appendix C, use the first set of parameters for citywide from Table 5.1) | $a$ | 284.23 |
| What is the estimated standard error of a 50-percent characteristic with a base equaling the total units? | $\sigma = \sqrt{\frac{a(0.5)(1-0.5)}{\mathrm{A}}}$ | $\sqrt{\frac{284.23(0.5)(1-0.5)}{3,109,955}}$ $= 0.0048$ |
| What are the end points of the category the median is in? | $U - L$ | $59,999.5 – $49,999.5 |
| What is the width of this category (in dollars, rooms, or whatever the item measures)? | $W$ | $10,000 |
| How many housing units are in this median category? | $B$ | 201,798 |
| What is the estimated proportion of the total units falling in the category containing the sample median? | $P = \frac{B}{A}$ | $\frac{201,798}{3,109,955} = 0.0649$ |
| Then the standard error from sampling for the median is approximately: | $se_{median} = \frac{\sigma \times W}{P}$ | $\frac{0.0048 \times \$10,000}{0.0649} = \$739.60$ |
| The 90-percent confidence interval for the median is: | $Median \pm 1.645 \times se_{median}$ | $\$57,500 \pm \$1,217$ |

Thus, there is a 90-percent chance that you will be correct if you conclude that the actual median household income for all occupied housing units in New York City is between $56,283 and $58,717.

### 5.5. Sampling Error for Means

The mean and the median usually differ. The mean is usually higher because it is influenced more heavily than the median by very large values. Use the following equation to calculate a 90-percent confidence interval of the mean:

$$1.645 * \sqrt{\frac{\left(\sum_{i=1}^{n} p_i x_i^2 - \left(\sum_{i=1}^{n} p_i x_i\right)^2\right)}{c} * a}$$

where:

$p_i$: is the proportion of total households or persons from a distribution in the $i^{th}$ interval.
$x_i$: is the midpoint of the $i^{th}$ interval (The midpoint of the open-ended interval is 2.5 times the lower limit).
$c$: is the total number of households or persons in the distribution (Subtract the number of "not applicable" from the total to get $c$).
$n$: is the total number of intervals in the distribution.
$a$: is the parameter a from Table 5.1a, 5.1b or 5.1c.

For example, the mean (or average) household income of all occupied housing units in New York City was $97,132 (compared to a median of $57,500). The distribution from which the mean was computed is given in Table 5.5.

*Table 5.5: Distribution of Household Income from the Mean*

| **Household Income** | **Number of HUs** | $p_i$ | $x_i$ |
|---|---|---|---|
| Less Than $5,000/no income/loss | 166,827 | 0.0536 | $2,500 |
| $5,000-$9,999 | 139,239 | 0.0448 | $7,500 |
| $10,000-$14,999 | 161,977 | 0.0521 | $12,500 |
| $15,000-$19,999 | 155,689 | 0.0501 | $17,500 |
| $20,000-$24,999 | 156,732 | 0.0504 | $22,500 |
| $25,000-$29,999 | 131,104 | 0.0422 | $27,500 |
| $30,000-$34,999 | 128,896 | 0.0414 | $32,500 |
| $35,000-$39,999 | 122,756 | 0.0395 | $37,500 |
| $40,000-$49,999 | 216,410 | 0.0696 | $45,000 |
| $50,000-$59,999 | 201,798 | 0.0649 | $55,000 |
| $60,000-$69,999 | 185,237 | 0.0596 | $65,000 |
| $70,000-$79,999 | 148,787 | 0.0478 | $75,000 |
| $80,000-$89,999 | 138,940 | 0.0447 | $85,000 |
| $90,000-$99,999 | 120,038 | 0.0386 | $95,000 |
| $100,000-$124,999 | 262,477 | 0.0844 | $112,500 |
| $125,000-$149,999 | 164,428 | 0.0529 | $137,500 |
| $150,000 or more | 508,620 | 0.1635 | $375,000 |
| Total | 3,109,955 | 1.000 | |

The error for a 90-percent confidence interval on the mean value is computed as follows:

$$1.645 * \sqrt{\frac{26{,}772{,}828{,}750-(106{,}827.50^2)}{3{,}109{,}955} * 284.23} = \$1{,}949$$

Thus, there is a 90-percent chance of being correct if you say the mean household income of all occupied housing units in New York City is between $95,183 and $99,081.

## 6. Replicate Weights

New to the NYCHVS 2017 data files are replicate weights. These replicate weights provide the data user an alternative method of producing variance estimates. Both the GVFs provided in this document and replicate weights on the data file can be used to produce estimates of variance. The GVFs provide a convenient way of producing a variance estimate while the use of replicate weights would be a more specialized and technical way to estimate variance. Both are acceptable approaches to variance estimation, and the method chosen would depend on the data user's familiarity with each method, access to statistical software, and data user preferences.

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. So to estimate the variance we need multiple samples but we only observe one. Replication uses the single observed sample to generate several replicate samples. These replicate samples can then be used to measure the variation of the estimates. Replication allows small changes to a single probability sample to create a set of replicate samples. This is done through subsets selected from the original sample in a process that mimics the original sample design. Each sample replication (*r*) is then fully weighted, using the same process as the original sample, to ensure each replicate sample, *r*, represents the population of interest. This process forms the set of final replicate weights {$w_r$ | *r* = 1, … ,R}, this is similar to what is provided in the 2017 NYCHVS data file. 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 {$\hat{\theta}_r$ | *r* = 1, … , R} represents the variability, or dispersion, of the estimate of interest under multiple samples of the population. Using the replicated estimates together with the 2017 NYCHVS equation of replicate variance, data users can calculate an estimated variance of an estimate of interest.

The 2017 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). To calculate the variance of an estimate, use the replication variance estimator:

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

where $\hat{\theta}$ is the weighted 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. This is similar to our GVF parameters, we have two different sets of GVF parameters one set for housing characteristics and another set for 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.  We are going to estimate the variance using the 80 replicate weights provided for the NYCHVS.  Please note that for 2017 NYCHVS, the weights in Replicate 1 equals full sample weights, or the weight used to derive sample estimates.  The Hadamard matrix was used to derive replicate factors to apply to individual full sample weights in creating replicate weights.

The goal of this example is to estimate the total number of renter-occupied housing units in Queens for 2017 and its corresponding estimate of variance.

For example, we have 1,810 completed interviews that are renter-occupied housing units in Queens. Table 6.1 displays the first four and last one renter-occupied sample units in Queens. The result show below are the sample cases in Queens with responses to tenure status question as renters.

*Table 6.1: Example of Estimating Variances with Replication*

| **Sample HU** | **Full Sample Weight** | **Replicate Weights** | | | | |
|---|---|---|---|---|---|---|
| | | **Replicate 1** | **Replicate 2** | **Replicate 3** | | **Replicate 80** |
| 1 | 250.430 | 250.430 | 234.225 | 75.769 | … | 272.506 |
| 2 | 241.448 | 241.448 | 224.532 | 254.145 | … | 263.398 |
| 3 | 240.695 | 240.695 | 416.378 | 225.885 | …. | 74.076 |
| 4 | 178.260 | 178.260 | 303.175 | 184.240 | … | 52.920 |
| … | … | … | … | … | | … |
| *1810* | 11.598 | 11.598 | 3.566 | 10.865 | … | 11.525 |

In NYCHVS, the full sample estimate and the full sample weight are referred to as the replicate estimate 0 and replicate weight 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 2017. 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 in Queens Estimate:

$\hat{N}$ = 250.430 + 241.448 + … + 11.598 = 439,257.02

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

The replicate survey estimates are as follows:

Replicate 1 Renter-Occupied Estimate $\hat{N}_{r=1}$= 250.430 + 241.448 + … + 11.598 = 439,257.02
Replicate 2 Renter-Occupied Estimate $\hat{N}_{r=2}$= 234.225 + 224.532 + … + 3.566 = 440,785.37
Replicate 3 Renter-Occupied Estimate $\hat{N}_{r=3}$= 75.769 + 254.145 + … + 10.865 = 435,992.59
$\vdots$ $\vdots$ $\vdots$
Replicate 80 Renter-Occupied Estimate $\widehat{N}_{r=80}$= 272.506 + 263.398 + … + 11.525 = 436,801.68

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

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

$$= 0.05 \times [(439{,}257.02 - 439{,}257.02)^2 + (440{,}785.37 - 439{,}257.02)^2 + (435{,}992.59 - 439{,}257.02)^2 + \cdots + (436{,}801.68 - 439{,}257.02)^2]$$

$$= 0.05 \times [0 + 2{,}335{,}853.72 + 1{,}0656{,}503.22 + \cdots + 6{,}028{,}694.52]$$

$$= 29{,}362{,}077.47$$

The estimate of the variance of total renter-occupied HUs in Queens is $\hat{v}(\hat{N})$ =29,362,077.47.

The survey estimate for total renter-occupied population in Queens is 439,257.02 housing units. This survey estimate has an estimated variance of 29,362,077.47, or a standard error of 5,418.68 housing units.

## 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.

Lohr, S.L. (2007). Recent developments in multiple frame surveys. Proceedings of the Survey Research Methods Section, American Statistical Association, 3257-3264. Accessed online at http://www.amstat.org/sections/srms/Proceedings/ on September 1, 2015.

Lohr, S. (2010). "Dual Frame Surveys: Recent Developments and Challenges," paper presented at the Scientific Meeting of the 45th Italian Statistical Society, Padua, Italy, June 16-18.

U.S. Census Bureau (2014). Overview. https://www2.census.gov/programs-surveys/nychvs/about/overview/overview-2014. Date retrieved April 23, 2018.

U.S. Census Bureau (2017a). Methodology for United States Population Estimates: Vintage 2017. https://www2.census.gov/programs-surveys/popest/technical-documentation/methodology/2010-2017/2017-natstcopr-meth.pdf Date retrieved April 23, 2018.

U.S. Census Bureau (2017b). Overview. https://www2.census.gov/programs-surveys/nychvs/about/overview/overview-2017. Date retrieved April 23, 2018.

# Appendix A. Example of Ratio Adjustments

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

For this example, assume weights were calculated for a sample and the weights included all weighting adjustments up to a nonresponse adjustment. With these weights, totals by two categories of tenure status (owner or renter) and two categories of type of construction (old or new) were created. Table A1 summarizes the estimated totals resulting from the hypothetical sample and weights.

*Table A1: Estimated Totals*

| | **Owners** | **Renters** | **Total** |
|---|---|---|---|
| **New** | 110 | 91 | 201 |
| **Old** | 97 | 107 | 204 |
| **Total** | 207 | 198 | 405 |

Suppose the control totals were as shown in table A2.

*Table A2: Example 1 Control Totals*

| | **Owners** | **Renters** | **Total** |
|---|---|---|---|
| **New** | 115 | 105 | 220 |
| **Old** | 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 1 Ratio Adjustment Factors*

| | **Owners** | **Renters** |
|---|---|---|
| **New** | 115/110 = 1.0455 | 105/91 = 1.1583 |
| **Old** | 95/97 = 0.9794 | 105/107 = 0.9813 |

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

Note that ratio-adjusted weights for the combination of owners and new construction is the product of the weight before the RAF, that is,

*Ratio-adjusted weight = original weight* ´ 1.0455.

The ratio-adjusted weights for the other three cells are defined similarly.

# Appendix B: List of Variables Imputed for 2017 NYCHVS

*Table B1: List of Variables Imputed for Occupied Units*

| Occupied Units | | |
|---|---|---|
| **Item Name** | **Variable Name** | **Imputed?** |
| Additional Source(s) of Heat | SC187 | No |
| Any Buildings with Broken or Boarded-Up Windows (Observation) | SC24 | No |
| Broken Plaster or Peeling Paint on Ceiling or Inside Walls | SC192 | No |
| Broken Plaster or Peeling Paint on Ceiling or Inside Walls Larger than 8 1/2 x 11 | SC193 | No |
| Borough | BORO | No |
| Combined Gas/Electricity Cost | UF14 | Yes |
| Complete Kitchen Facilities | UF83 | Yes |
| Complete Plumbing Facilities | UF81 | Yes |
| Condition of Building (Observation) | SC23 | No |
| Condition of External Walls | UF1_1, UF1_3 - UF1_6 | No |
| Condition of Floors | UF1_17, UF1_19 - UF1_22 | No |
| Condition of Stairways (Exterior and Interior) | UF1_12 - UF1-16, UF1_35 | No |
| Condition of Building Recode | REC21 | No |
| Condition of Windows | UF1_7 - UF1_11 | No |
| Condo/Coop Before Acquisition | SC121 | No |
| Condo/Coop Conversion Done Through a Non-Eviction Plan | SC118 | No |
| Condo/Coop Status | SC114 | No |
| Control Status Recode | UF_CSR | No |
| Cracks or Holes in Interior Walls or Ceiling | SC190 | No |
| Down Payment | UF5 | No |
| Electricity Paid Separately | SC159 | Yes |
| Electricity Monthly Cost | UF12 | Yes |
| Exterminator Service | SC189 | No |
| Federal, State, or Local Government Payments for Rent | SC184, SC541-SC544 | No |
| Fire and Liability Insurance Paid Separately | SC141 | No |
| First Occupants of Unit | SC54 | No |
| Floor the Unit is On | UF50 | No |
| Functioning Air Conditioning | SC197 | No |
| Gas Paid Separately | SC161 | Yes |
| Gas Monthly Cost | UF13 | Yes |
| General Health of Respondent | SC574 | No |
| Heating Equipment Breakdown | SC185 | No |
| Holes in Floors | SC191 | No |
| Household Below Specified Income Level Recode | REC39 | Yes |
| Household Composition Recode | REC1 | Yes |

| Occupied Units | | |
|---|---|---|
| **Item Name** | **Variable Name** | **Imputed?** |
| Household Income from Farm or Nonfarm Business, Proprietorship, or Partnership Recode | UF35 | Yes |
| Household Income from Retirement, Survivor, or Disability Pensions Recode | UF38 | Yes |
| Household Income from Social Security or Railroad Retirement Payments Recode | UF37 | Yes |
| Household Income from SSI, TANF/Family Assistance, Safety Net, or Other Public Assistance or Public Welfare Payments, Including Shelter Allowance Recode | UF39 | Yes |
| Household Income from VA Payments, Unemployment Compensation, Child Support, Alimony, or Other Source of Income Recode | UF40 | Yes |
| Household Income from Wages, Salaries, Commissions, Bonuses, or Tips Recode | UF34 | Yes |
| Household Member Under Age of 6 | UF75 | Yes |
| Household Member Under Age of 18 | REC7 | Yes |
| Householder of Spanish/Hispanic Origin | HHR5 | Yes |
| Householder Moved to the United States as Immigrant | SC560 | No |
| Householder's Age Recode | UF43 | Yes |
| Householder's Race | UF61 | Yes |
| Householder's Sex | HHR2 | Yes |
| Kitchen Facilities Functioning | SC157 | Yes |
| Length of Lease | SC181 | No |
| Householder Lived in Unit at Time of Coop/Condo Conversion | SC117 | No |
| Medical Device in Home | SC198 | No |
| Monthly Contract Rent | UF17 | Yes |
| Monthly Contract Rent as a Percent of Household Income Recode | UF29 | Yes |
| Monthly Contract Rent per Room Recode | UF27 | Yes |
| Monthly Gross Rent | UF26 | Yes |
| Monthly Gross Rent as a Percent of Household Income Recode | UF30 | Yes |
| Monthly Gross Rent per Room Recode | UF28 | Yes |
| Monthly Owner Cost Recode | UF105 | Yes |
| Mortgage Interest Rate (Current) | UF7A | No |
| Mortgage Origination Year | UF68 | Yes |
| Mortgage Status | SC127 | No |
| Most Recent Place Householder Lived for 6 Months or More | UF79 | No |
| Number of 1987 Maintenance Deficiencies Recode | REC54 | No |
| Number of 2017 Maintenance Deficiencies Recode | REC53 | No |
| Number of Bedrooms | UF78 | Yes |

| Occupied Units | | |
|---|---|---|
| **Item Name** | **Variable Name** | **Imputed?** |
| Number of Cockroaches | SC571 | No |
| Number of Heating Equipment Breakdowns | SC186 | No |
| Number of Persons from Temporary Residence | UF2A_1 | No |
| Number of Persons per Room Recode | CPPR | Yes |
| Number of Persons Recode | UF73 | No |
| Number of Rooms | UF77 | Yes |
| Number of Units in Building | UF47 | Yes |
| Number of Stories in Building | UF11 | Yes |
| Occupancy Status Before Acquisition | SC120 | No |
| Other Fuels Paid Separately | SC166 | Yes |
| Other Fuels Annual Cost | UF16 | Yes |
| Out of Pocket Rent | UF17A | Yes |
| Owner Lives in Building | SC147 | No |
| Passenger Elevator in Building | SC149 | No |
| Place of Householder's Birth | SC111 | No |
| Place of Householder's Father's Birth | SC112 | No |
| Place of Householder's Mother's Birth | SC113 | No |
| Postponement of Health Care for Financial Reasons | SC647-SC651 | No |
| Presence of Mice and Rats | SC188 | No |
| Presence of Nonrelatives Recode | UF46 | No |
| Receipt of Public Assistance or Welfare Payments | SC548-SC551, SC175, SC199 | No |
| Race and Ethnicity of Householder Recode | UF69 | Yes |
| Race Recode 1 (Householder) | UF60 | Yes |
| Real Estate Taxes Paid Separately | SC144 | No |
| Respondent Line Number | UF71 | No |
| Respondent Rating of Residential Structures in Neighborhood | SC196 | No |
| Respondent Opinions of their Housing Unit's Affordability | SC168, SC169, SC183 | No |
| Senior Citizen Carrying Charge Increase Exemption (SCRIE) | SC184 | No |
| Service Interruptions for Financial Reasons | SC131, SC132, SC136, SC137, SC138 | No |
| Sidewalk to Elevator Without Using Steps or Stairs | SC173 | No |
| Sidewalk to Unit Without Using Steps or Stairs | SC171 | No |
| Structure Class Recode | UF74 | No |
| Sub-borough Area | CD | No |
| Telephone (Landline) in Apartment (House) | SC575 | No |
| Tenure (1) Owner/Renter | SC115 | No |
| Tenure (2) Cash Rent/Rent Free | SC116 | No |
| Toilet Breakdowns | UF82 | No |
| Total Household Income Recode | UF42 | Yes |
| Type of Heating Fuel | SC158 | Yes |
| Type of Schedule Code | UF76 | No |

| Occupied Units | | |
|---|---|---|
| **Item Name** | **Variable Name** | **Imputed?** |
| Value | UF6 | Yes |
| Water and Sewer Paid Separately | SC164 | Yes |
| Water and Sewer Annual Cost | UF15 | Yes |
| Water Leakage Inside Apartment | SC194 | No |
| Wheelchair Accessibility | SC36, SC37, SC38 | No |
| Year Built Recode | UF23 | Yes |
| Year Householder Moved Into Unit | UF66 | Yes |
| Year Householder Moved to the United States | UF53 | No |
| Year Householder Moved to New York City | UF54 | No |

*Table B2: List of Variables Imputed for Persons*

| **Persons** | | |
|---|---|---|
| **Item Name** | **Variable Name** | **Imputed?** |
| Age Recode | UF43 | Yes |
| Average Hours Worked per Week in 2016 | UF96 | Yes |
| Borough | BORO | No |
| Check Item G | CHK_G | Yes |
| Current Enrollment in Job Training/Education | ITEM50A | No |
| Educational Attainment | EDUCTN | Yes |
| Hours Worked Last Week | UF95 | Yes |
| Income from Own Farm or Nonfarm Business, Proprietorship, or Partnership | UF18A | Yes |
| Income from Interest, Dividends, Net Rental or Royalty Income, or Income from Estates and Trusts | UF18B | Yes |
| Income from Retirement, Survivor, or Disability Pensions | UF18E | Yes |
| Income from Social Security or Railroad Retirement Payments | UF18C | Yes |
| Income from SSI, TANF/Family Assistance, Safety Net, or Other Public Assistance or Public Welfare Payments, Including Shelter Allowance | UF18D | Yes |
| Income from VA Payments, Unemployment Compensation, Child Support, Alimony, or Other Source of Income | UF18F | Yes |
| Income from Wages, Salaries, Commissions, Bonuses, or Tips | UF18 | Yes |
| Labor Force Status Recode | UF59 | Yes |
| Last Time Worked | ITEM44 | Yes |
| Looking for Work | ITEM42 | Yes |
| Major Industry Type | ITEM45C | Yes |
| Number of Weeks Worked in 2016 | ITEM48A | Yes |
| Person from Temporary Residence | UF3 | No |
| Person Number of 1st Parent | UF87 | No |
| Person Number of 2nd Parent | UF88 | No |
| Person Number of Spouse/Partner | UF86 | No |
| Race | UF62 | Yes |
| Race and Ethnicity of Householder Recode | UF70 | Yes |
| Race Recode 1 | UF60 | Yes |
| Relationship | UF92 | No |
| Sex | SEX | Yes |
| Spanish/Hispanic Origin | HSPANIC | Yes |
| Sub-borough Area | CD | No |
| Temporarily Absent or on Layoff from a Job Last Week | ITEM41 | Yes |
| Total Person Income Recode | UF41 | Yes |
| Type of Worker | UF90 | Yes |
| Workers' Industry Code | UF94 | Yes |

| Persons | | |
|---|---|---|
| **Item Name** | **Variable Name** | **Imputed?** |
| Worked Last Week | ITEM40A | Yes |
| Workers' Occupation Code | UF93 | Yes |
| Year Non-Householder Moved Into Unit | UF55 | No |

*Table B3: List of Variables Imputed for Vacant Units*

| Vacant Units | | |
|---|---|---|
| **Item Name** | **Variable Name** | **Imputed?** |
| Any Buildings with Broken or Boarded-Up Windows (Observation) | SC24 | No |
| Borough | BORO | No |
| Complete Kitchen Facilities | UF84 | Yes |
| Complete Plumbing Facilities | UF91 | Yes |
| Condition of Building (Observation) | SC23 | No |
| Condition of Building Recode | REC21 | No |
| Condition of External Walls | UF1_1, UF1_3 - UF1_6 | No |
| Condition of Floors | UF1_17, UF1_19 - UF1_22 | No |
| Condition of Stairways (Exterior and Interior) | UF1_12 - UF1-16, UF1_35 | No |
| Condition of Windows | UF1_7 - UF1_11 | No |
| Condo/Coop Status | SC530 | No |
| Condo/Coop Status Before Vacancy | SC533 | No |
| Control Status Recode | UF_CSR | No |
| Duration of Vacancy | SC531 | No |
| First Occupancy | SC518 | No |
| Floor the Unit is On | UF51 | No |
| Monthly Asking Rent | UF31 | Yes |
| Monthly Asking Rent per Room Recode | UF32 | Yes |
| Number of Bedrooms | UF78 | Yes |
| Number of Rooms | UF77 | Yes |
| Number of Units in Building | UF47 | Yes |
| Owner in Building | SC520 | No |
| Passenger Elevator in Building | SC522 | No |
| Reason Unit Not Available for Rent or for Sale | UF80 | No |
| Sidewalk to Elevator without Using Steps or Stairs | SC553 | No |
| Sidewalk to Unit without Using Steps or Stairs | SC555 | No |
| Status of Vacant Unit | SC534 | No |
| Status Prior to Vacancy | SC532 | No |
| Stories in Building | UF33 | Yes |
| Structure Class Recode | UF74 | No |
| Sub-borough Area | CD | No |
| Type of Heating Fuel | SC529 | Yes |
| Type of Schedule | UF76 | No |
| Vacant Unit Respondent | SC30 | No |
| Wheelchair Accessibility | SC36, SC37, SC38 | No |
| Year Built Recode | UF23 | No |

## Appendix C: Housing Unit Characteristics Associated With GVF Parameters

For characteristics and subgroups matching to this table (Table C1), use the second of the three sets of parameters from the housing unit GVF parameters (Table 5.1a).

***Table C1****: Housing Unit Characteristics Associated with the Second of Three Sets of Parameters on Table 5.1a*

| Characteristics | Applicable Subgroups |
|---|---|
| Race and Ethnicity of Householder<br>(White, non-Hispanic and African American, non-Hispanic) | Total Housing Units |
| Borough Totals | Renter Occupied (Stabilized, Mitchell Lama, Public Housing) and<br>Owner Occupied (Condominiums and Total Cooperatives) |
| Sub-borough of Staten Island Totals | Total Housing Units, Total Occupied Housing Units, Total Rental Housing Units and Total Occupied Rental Housing Units |
| Contract Rent < $300 | Total Housing Units and Total Occupied Housing Units |
| Wheel Chair Accessibility | All subgroups except<br>Renter Occupied - Controlled and<br>Owner Occupied - Conventional |
| Floor Unit is on (except basement) | |
| Access from Sidewalk to Elevator/Unit without using Stairs | |
| Households Not Receiving Part of Monthly Rent from Government Programs | |
| Condition of Building External Walls, Windows, Stairways, and Floors of Building | Total Occupied and Total Renter Occupied |
| Number of Building Condition Problems 1-4 | |

For characteristics and subgroups matching to this table (Table C2), use the third of the three sets of parameters from the housing unit GVF parameters (Table 5.1a).

***Table C2**: Housing Unit Characteristics Associated with the Third of Three Sets of Parameters on Table 5.1a*

| **Characteristics** | **Applicable Subgroups** |
|---|---|
| Sub-borough Totals (All Boroughs Except Staten Island) | Total Housing Units, Total Occupied Housing Units, Total Rental Housing Units and Total Occupied Rental Housing Units |
| Structure Classification - Multiple dwelling units | Total Housing Units and<br>Total Occupied Housing Units |
| Structure Classification - One or 2 family house | Total Housing Units |
| Rent Control Status | Total Rental Housing Units and Total Occupied Rental Housing Units |
| Year Building Built | Total Occupied and Total Renter Occupied |
| Number of Stories in Building | |
| Number of Units in Building | |
| Presence of Owner in Building | |
| Elevator in Building with 2 or more stories | |
| State/City Assisted Cooperatives | Total Owner Housing Units and Total Occupied Owner Housing Units |
| Private Cooperatives | |
| Private Condominiums | |