AMERICAN HOUSING SURVEY

# Quick Guide to Estimating Variance Using Replicate Weights: 2009 to Current

LAST UPDATED: MAY 2021

U.S. DEPARTMENT OF HOUSING AND URBAN DEVELOPMENT
U.S. CENSUS BUREAU
U.S. DEPARTMENT OF COMMERCE

AMERICAN HOUSING SURVEY

## Contents


1. Overview ..... 1
2. AHS-Specific Information ..... 2
   2.1. Replicate Weight Generation Method ..... 2
   2.2. AHS Regular Weight ..... 2
   2.3. AHS Split Sample Weights ..... 2
   2.4. AHS Replicate Weights ..... 2
   2.5. AHS Split Sample Replicate Weights ..... 2
   2.6. AHS Data Sets ..... 3
3. Using Replicate Weights with Built-In SAS Procedures ..... 3
   3.1. Specifying the VARMETHOD Option ..... 3
   3.2. Declaring the REPWEIGHTS Statement ..... 3
   3.3. The Importance of the DOMAIN Statement for Subgroup Analysis ..... 3
4. Using Replicate Weights Manually Compute Variance ..... 4

References ..... 5

Additional Reading ..... 5

Appendix A. ..... 6

Descriptive Statistics with PROC SURVEYMEANS ..... 6

Appendix B. Frequencies with PROC SURVYEFREQ ..... 9

Appendix C. Linear Regression with PROC SURVEYREG ..... 10

Appendix D. Logistic Regression Using PROC SURVEYLOGISTIC ..... 13

## List of Exhibits

Exhibit 3.1. SAS Survey Procedures and Documentation ..... 3
Exhibit A.1. Statistics ..... 6
Exhibit A.2. Statistics for REGION Domains ..... 6
Exhibit A.3. Statistics for TENURE Domains ..... 7
Exhibit A.4. Statistics for REGION*TENURE Domains ..... 8
Exhibit B.1. NUNIT2 by TENURE ..... 9
Exhibit C.1. Parameter Estimates - Estimated Regression Coefficients ..... 11
Exhibit D.1. The SURVEYLOGISTIC Procedure - Parameter Estimates and t Confidence Intervals ..... 14
Exhibit D.2. The SURVEYLOGISTIC Procedure Odds - Ratio Estimates and t Confidence Intervals ..... 14

# 1. Overview

The purpose of this document is to explain how to estimate variances and other measures of variability for the American Housing Survey (AHS) public use files (PUFs) that include replicate weights.[1] Replicate weights are a way to encapsulate the effect of the sampling design on variances. In heuristic terms, the algorithms that generate the replicate weights simulate drawing additional samples using the same design, thus providing a sample of samples used to understand the variability in the data. For a more technical description, see Lewis (2015). Note that statistical software designed to work with a simple random sample and a single survey weight generally computes accurate point estimates, but variances, standard errors, and confidence intervals are too small. Thus, estimates may appear to be statistically significant when they are not. This guide applies to any estimate: mean, median, sum, proportion, count, regression parameter, and so on.

The U.S. Department of Housing and Urban Development (HUD) and the U.S. Census Bureau began supplying replicate weights for the AHS starting with the 2009 national sample PUF. The first year that metropolitan area sample PUFs included replicate weights was 2013. All subsequent national sample and metropolitan area sample PUFs have replicate weights. HUD and the Census Bureau supply the PUFs in SAS and ASCII (CSV) formats.

This document provides some information about the AHS necessary to calculate estimates using replicate weights, describes how to use replicate weights in built-in SAS procedures and provides a general description of how to manually calculate variance using replicate weights, which is for AHS users without statistical software packages that contain built-in procedures.[2]

Appendixes A through D provide examples (exact code and results) of using replicate weights for four different SAS statistical procedures: *PROC SURVEYMEANS*, *PROC SURVEYFREQ*, *PROC SURVEYREG*, and *PROC SURVEYLOGISTIC*. The examples assume that the reader has a general familiarity with SAS, the BASE SAS statistical procedures and the statistical models. Many guides and instructional books are available on both SAS and statistics that explain these subjects at length. Although the models and relationships in the appendix examples are not trivial, they were chosen to illustrate the procedures, and not because of theoretical or statistical validity.

The References section provides additional reading about replicate weights.

[1] This document is based on Vandenbroucke (2017).

[2] Most other statistical packages, such as STATA and R, have built-in procedures to use replicate weights.

# 2. AHS-Specific Information

AHS users must know about the replicate weight generation method and the names of AHS variables containing regular and replicate weights.

## 2.1. Replicate Weight Generation Method

The AHS replicate weights were generated using Fay's method, a variation on the balanced repeated replication method (Dippo, Fay, and Morganstein, 1984).

## 2.2. AHS Regular Weight

Every AHS PUF has a general survey weight variable.

2009–2013: *WGT90GEO*. These PUFs also have a variable, *REPWGT0*, which is the same as *WGT90GEO*. Note that these data sets also have a variable called *WEIGHT*. *WEIGHT* is calibrated to produce estimates for 2001 and later using the same methodology used in pre-2001 AHS. For 2001 and later, HUD and the Census Bureau slightly changed the methodology to produce AHS estimates. In short, the revised methodology uses more current information on nonresponse, demographics and housing unit counts to provide what HUD and the Census Bureau believe is a more accurate weight.

HUD and the Census Bureau encourage AHS users to use *WGT90GEO* for producing current-year estimates. For AHS users comparing pre-2001 with post-2001 estimates, HUD and the Census Bureau suggest using *WEIGHT*.

2015 and later: *WEIGHT*.

## 2.3. AHS Split Sample Weights

Beginning in 2013, the AHS began administering some topical modules to one-half of the sample to increase the number of topical modules and limit respondent burden. Variables in these topical modules must be weighted with the split weights corresponding to the half-sample to which the module was administered. See the AHS Codebook entry for the variable in question to see which is appropriate.

2013: *SPLTWGT1* and *SPLTWGT2*.

2015 and later: *SP1WEIGHT* and *SP2WEIGHT*.

## 2.4. AHS Replicate Weights

The AHS PUFs have 160 replicate weight variables. The names vary.

2009–2013: *REPWGT1-REPWGT160*.

2015 and later: *REPWEIGHT1-REPWEIGHT160*.

## 2.5. AHS Split Sample Replicate Weights

Starting in 2015, the AHS PUFs include replicate weights for the split sample variables. The variables are *SP1REPWGT1- SP1REPWGT160* and *SP2REPWGT1- SP2REPWGT160*. Use the set appropriate to the variable of interest.

Note that no split replicate weights exist in the 2013 PUF, and so the variances of the split sample variables may not be estimated using replicate weights for this survey year.

## 2.6. AHS Data Sets

If using the relational structure version of the AHS PUF, the replicate weights are all on the HOUSEHOLD data set[3].

# 3. Using Replicate Weights with Built-In SAS Procedures

SAS/STAT software provides a set of procedures whose names begin with *SURVEY* that are the counterparts of *BASE* SAS procedures. This document concentrates on the basic information needed to make use of replicate weights. SAS procedures have many options and capabilities not discussed in this document. Procedure names with links to additional SAS documentation are in Exhibit 3.1.

**Exhibit 3.1. SAS Survey Procedures and Documentation**

| SAS Procedure | Documentation Link |
|---|---|
| SURVEYMEANS | http://support.sas.com/documentation/onlinedoc/stat/142/surveymeans.pdf. |
| SURVEYFREQ | http://support.sas.com/documentation/onlinedoc/stat/142/surveyfreq.pdf. |
| SURVEYREG | http://support.sas.com/documentation/onlinedoc/stat/142/surveyreg.pdf. |
| SURVEYLOGISTIC | http://support.sas.com/documentation/onlinedoc/stat/142/surveylogistic.pdf. |
| SURVEYPHREG | http://support.sas.com/documentation/onlinedoc/stat/142/surveyphreg.pdf. |

## 3.1. Specifying the VARMETHOD Option

To correctly use the *SURVEY*xxx procedures, the *VARMETHOD*= option must be added to the PROC statement to indicate the method to develop the replicate weights. For the AHS, this method is Fay's variation of the balanced repeated replications method. Thus, the option is always VARMETHOD=BRR(FAY). For example, in PROC SURVEYMEANS, one writes—

PROC SURVEYMEANS DATA=<dataset> VARMETHOD=BRR(FAY) <other options>;

## 3.2. Declaring the REPWEIGHTS Statement

To properly use the *SURVEYxxx* procedures, the *REPWEIGHTS* options must be specified. The *REPWEIGHTS* statement identifies the variables containing replicate weights. When using AHS data, the statement is—

REPWEIGHTS REPWEIGHT1-REPWEIGHT160;

Note the previous section concerning the names of replicate weight variables by year.

## 3.3. The Importance of the DOMAIN Statement for Subgroup Analysis

AHS users frequently want to compute estimates for subgroups within AHS data, such as by tenure, race, and so on. A common way to do so with *BASE* SAS procedures is to use BY group processing or to filter the data using *WHERE*. These methods should be avoided when using the *SURVEY*xxx procedures. For the procedures to use replicate weights correctly, these methods must process the full sample. BY group processing and the *WHERE* subset the sample before computing the statistics. Instead of BY group processing or *WHERE* filtering, users should list the subset variables on the *DOMAIN* statement. For

[3]The replicate weights were kept on separate REPWGT files in the original 2009-2013 release, but were merged into the HOUSEHOLD file for the May 18th, 2021 revisions.

example, include this statement to compute estimates by tenure.

DOMAIN TENURE;

Cross domain specifications using the asterisk operator in the usual way. To generate statistics by tenure and householder race, including the marginal distributions, AHS users must use the following statement.

DOMAIN TENURE HHRACE TENURE*HHRACE;

## 4. Using Replicate Weights Manually Compute Variance

Many statistical software programs enable users to incorporate replicate weights into survey estimations. If a statistical software program does not support replicate weights, the manual process to estimate variance is straightforward. This section provides the basic steps to make the manual calculation of variances.

Replicate weights are computationally intensive to construct. Once created, however, replicate weights can be used to estimate variances for most estimators, including means, totals, ratios, and regression coefficients.

The general replication procedure consists of the following steps.

1. Estimate the statistic of interest $\hat{\theta}$ using the full sample weights.
2. Estimate the same statistic for each of the 160 replicate weights as $\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_{160}$.
3. Apply the following formula to estimate the variance of this estimator—

$$\hat{v}\left(\hat{\theta}\right) = \frac{4}{160}\sum_{r=1}^{160}\left(\hat{\theta}_r - \hat{\theta}\right)^2 .$$

Thus, AHS users must estimate the statistic 161 times, using whatever method the software requires—once using the full sample weight and 160 times using each replicate weight. Then, AHS users must compute the squared difference between each estimate based on a replicate weight and the estimate based on the full sample weight. Then, AHS users must sum the squared differences and multiply the sum by 4/160 (0.025). This calculation yields the variance of the estimate. With the variance, AHS users can calculate standard errors, confidence intervals, and so on in the usual way. Using loops for the estimation process and to accumulate the sum of squares is helpful.

## References


Dippo, Cathryn S., Robert E. Fay, and David H. Morganstein. 1984. "Computing Variances From Complex Samples With Replicate Weights." In *Proceedings of the Survey Research Methods Section*. Washington, DC: American Statistical Association: 489–494.

Lewis, Taylor. 2015. *Replication Techniques for Variance Approximation*. Paper 2601-2015. College Park, MD: University of Maryland, Joint Program in Survey Methodology. https://support.sas.com/resources/papers/proceedings15/2601-2015.pdf.

SAS Institute Inc. 2016a. *SAS/STAT® 14.2 User's Guide: The SURVEYMEANS Procedure.* Cary, NC: SAS Institute. http://support.sas.com/documentation/onlinedoc/stat/142/surveymeans.pdf.

———. 2016b. *SAS/STAT® 14.2 User's Guide: The SURVEYFREQ Procedure.* Cary, NC: SAS Institute. http://support.sas.com/documentation/onlinedoc/stat/142/surveyfreq.pdf.

———. 2016c. *SAS/STAT® 14.2 User's Guide: The SURVEYREG Procedure.* Cary, NC: SAS Institute. http://support.sas.com/documentation/onlinedoc/stat/142/surveyreg.pdf.

———. 2016d. *SAS/STAT® 14.2 User's Guide: The SURVEYLOGISTIC Procedure.* Cary, NC: SAS Institute. http://support.sas.com/documentation/onlinedoc/stat/142/surveylogistic.pdf.

———. 2016e. *SAS/STAT® 14.2 User's Guide: The SURVEYPHREG Procedure.* Cary, NC: SAS Institute. http://support.sas.com/documentation/onlinedoc/stat/142/surveyphreg.pdf.

Vandenbroucke, David A. 2017. "Not So Simple: Intervals You Can Have Confidence in With Real Survey Data." Paper 970-2017. In *Proceedings of the SAS Global Forum 2017*, April 2–5, Orlando, FL. http://support.sas.com/resources/papers/proceedings17/0970-2017.pdf.


## Additional Reading


Ash, Stephen. 2014. "Using Successive Difference Replication for Estimating Variances," *Survey Methodology* 40 (1): 47–59.

Fay, Robert E. 1989. "Theory and Application of Replicate Weighting for Variance Calculations." In *Proceedings of the Section on Survey Research Methods*. Washington, DC: American Statistical Association: 212–219.

Fay, Robert E., and George F. Train. 1995. "Aspects of Survey and Model-Based Postcensal Estimation of Income and Poverty Characteristics for States and Counties." In *Proceedings of the Section on Government Statistics*. Alexandria, VA: American Statistical Association: 154–159.

Judkins, D.R. 1990. "Fay's Method for Variance Estimation," *Journal of Official Statistics* 6 (3): 223–239.

# Appendix A.

## Descriptive Statistics with PROC SURVEYMEANS

This example generates estimates of mean household income and housing cost by region and tenure, using the flattened version of the 2013 American Housing Survey national file.[3] It assigns formats (defined in an earlier PROC FORMAT step, not shown) to the classification variables.

```
PROC SURVEYMEANS DATA=Flat.AHS2013N PLOTS=NONE
  VARMETHOD=BRR(FAY); /* end of proc statement */
  VAR ZInc2 ZSMHC;
  WEIGHT Wgt90Geo;
  REPWEIGHT RepWgt1-RepWgt160;
  DOMAIN Region Tenure Region*Tenure;
  FORMAT
    Region $Region.
    Tenure $Tenure.
  ; /* end of format statement */
RUN;
```

The statistics and domain exhibits follow.

**Exhibit A.1. Statistics**

| Variable | Label | N | Mean | Standard Error of Mean | 95% CLfor Mean | |
|---|---|---|---|---|---|---|
| ZINC2 | Household income | 60097 | 68089 | 338.555453 | 67420.7590 | 68757.9864 |
| ZSMHC | Monthly housing costs | 60097 | 1136.039415 | 5.481829 | 1125.2133 | 1146.8655 |

**Exhibit A.2. Statistics for REGION Domains**

| Census Region | Variable | Label | N | Mean | Standard Error of Mean | 95% CL for Mean | |
|---|---|---|---|---|---|---|---|
| Northeast | ZINC2 | Household income | 15540 | 77018 | 817.526137 | 75403.6047 | 78632.6720 |
| | ZSMHC | Monthly housing costs | 15540 | 1355.355549 | 9.061006 | 1337.4610 | 1373.2501 |
| Midwest | ZINC2 | Household income | 16218 | 64922 | 671.268133 | 63596.1151 | 66247.4921 |
| | ZSMHC | Monthly housing costs | 16218 | 977.251911 | 8.722198 | 960.0264 | 994.4774 |

[3] Replicate weights are also used with the relational version of the data sets. In some survey years, users must first merge the replicate weights file to the file(s) containing the analysis variable by matching on CONTROL. This document uses the flat file structure in its examples to simplify the presentation.

| Census Region | Variable | Label | N | Mean | Standard Error of Mean | 95% CL for Mean | |
|---|---|---|---|---|---|---|---|
| South | ZINC2 | Household income | 17632 | 62560 | 538.874632 | 61495.5342 | 63623.9829 |
| | ZSMHC | Monthly housing costs | 17632 | 996.466335 | 8.775149 | 979.1363 | 1013.7964 |
| West | ZINC2 | Household income | 10707 | 73155 | 789.894481 | 71595.2277 | 74715.1554 |
| | ZSMHC | Monthly housing costs | 10707 | 1347.782380 | 15.251856 | 1317.6615 | 1377.9033 |

**Exhibit A.3. Statistics for TENURE Domains**

| Owner/Renter Status of Unit | Variable | Label | N | Mean | Standard Error of Mean | 95% CL for Mean | |
|---|---|---|---|---|---|---|---|
| Owned | ZINC2 | Household income | 35852 | 81625 | 478.050393 | 80681.1378 | 82569.3428 |
| | ZSMHC | Monthly housing costs | 35852 | 1241.889028 | 7.361340 | 1227.3511 | 1256.4270 |
| Cash rent | ZINC2 | Household income | 23358 | 42992 | 393.281772 | 42215.7753 | 43769.1608 |
| | ZSMHC | Monthly housing costs | 23358 | 972.605074 | 5.695551 | 961.3569 | 983.8532 |
| No-cash rent | ZINC2 | Household income | 887 | 34582 | 1351.991761 | 31911.8866 | 37251.9879 |
| | ZSMHC | Monthly housing costs | 887 | 170.012093 | 6.158597 | 157.8495 | 182.1747 |
| Not applicable | ZINC2 | Household income | 0 | | | | |
| | ZSMHC | Monthly housing costs | 0 | . | . | . | . |

**Exhibit A.4. Statistics for REGION*TENURE Domains**

| Census Region | Owner/Renter Status of Unit | Variable | Label | N | Mean | Standard Error of Mean | 95% CL for Mean | |
|---|---|---|---|---|---|---|---|---|
| Northeast | Owned | ZINC2 | Household income | 8796 | 93388 | 1125.642715 | 91165.3362 | 95611.4033 |
| | | ZSMHC | Monthly housing costs | 8796 | 1524.410306 | 13.388370 | 1497.9696 | 1550.8510 |
| | Cash rent | ZINC2 | Household income | 6560 | 48423 | 1023.451108 | 46401.6171 | 50444.0475 |
| | | ZSMHC | Monthly housing costs | 6560 | 1088.298393 | 12.491286 | 1063.6293 | 1112.9675 |
| | No-cash rent | ZINC2 | Household income | 184 | 40225 | 3734.320098 | 32849.6467 | 47599.4758 |
| | | ZSMHC | Monthly housing costs | 184 | 160.209187 | 16.865054 | 126.9024 | 193.5160 |
| | Not applicable | ZINC2 | Household income | 0 | | | | |
| | | ZSMHC | Monthly housing costs | 0 | | | | |
| Midwest | Owned | ZINC2 | Household income | 10363 | 76939 | 924.885293 | 75112.0206 | 78765.1354 |
| | | ZSMHC | Monthly housing costs | 10363 | 1069.892711 | 10.957067 | 1048.2536 | 1091.5318 |
| | Cash rent | ZINC2 | Household income | 5657 | 37218 | 591.987811 | 36048.8664 | 38387.1017 |
| | | ZSMHC | Monthly housing costs | 5657 | 790.340954 | 8.696447 | 773.1663 | 807.5156 |
| | No-cash rent | ZINC2 | Household income | 198 | 41081 | 3019.834005 | 35117.3709 | 47045.1206 |
| | | ZSMHC | Monthly housing costs | 198 | 166.738999 | 11.234548 | 144.5519 | 188.9261 |
| | Not applicable | ZINC2 | Household income | 0 | | | | |
| | | ZSMHC | Monthly housing costs | 0 | | | | |
| South | Owned | ZINC2 | Household income | 10871 | 74247 | 758.790945 | 72748.4583 | 75745.5331 |
| | | ZSMHC | Monthly housing costs | 10871 | 1070.684401 | 11.582589 | 1047.8099 | 1093.5589 |
| | Cash rent | ZINC2 | Household income | 6392 | 39764 | 572.892731 | 38632.9639 | 40895.7774 |
| | | ZSMHC | Monthly housing costs | 6392 | 890.630361 | 8.178963 | 874.4777 | 906.7830 |
| | No-cash rent | ZINC2 | Household income | 369 | 28527 | 1939.860755 | 24696.1961 | 32358.2641 |
| | | ZSMHC | Monthly housing costs | 369 | 186.400970 | 7.646978 | 171.2989 | 201.5030 |
| | Not applicable | ZINC2 | Household income | 0 | | | | |
| | | ZSMHC | Monthly housing costs | 0 | | | | |
| West | Owned | ZINC2 | Household income | 5822 | 90533 | 1206.538564 | 88150.3677 | 92915.9575 |
| | | ZSMHC | Monthly housing costs | 5822 | 1514.067961 | 22.206519 | 1470.2123 | 1557.9237 |
| | Cash rent | ZINC2 | Household income | 4749 | 47627 | 899.309833 | 45850.9491 | 49403.0458 |
| | | ZSMHC | Monthly housing costs | 4749 | 1133.220443 | 13.903508 | 1105.7624 | 1160.6785 |
| | No-cash rent | ZINC2 | Household income | 136 | 39889 | 2910.336918 | 34140.8916 | 45636.1494 |
| | | ZSMHC | Monthly housing costs | 136 | 137.796207 | 13.991933 | 110.1635 | 165.4289 |
| | Not applicable | ZINC2 | Household income | 0 | | | | |
| | | ZSMHC | Monthly housing costs | 0 | | | | |

# Appendix B. Frequencies with PROC SURVYEFREQ

This program produces weighted frequencies of structure type (NUNIT2) by tenure (TENURE), using the 2013 national PUF. It assigns formats (defined in an earlier PROC FORMAT step, not shown) to the classification variables and includes the options CL and CLWT to get the confidence intervals of the percentages and weighted counts, respectively.

```
PROC SURVEYFREQ DATA=Flat.AHS2013N
  VARMETHOD=BRR(FAY); /* end of proc statement */
  TABLES NUnit2 * Tenure /CL CLWT;
  WEIGHT Wgt90Geo;
  REPWEIGHT RepWgt1-RepWgt160;
  FORMAT
    Nunit2 $nunittwo.
    Tenure $Tenure.
  ; /* end of format statement */
RUN;
```

The resulting cross tabulation Exhibit follows.

**Exhibit B.1. NUNIT2 by TENURE**

| NUNIT2 | TENURE | Frequency | Weighted Frequency | Standard Error of Wgt Freq | 95% Confidence Limits for Wgt Freq | | Percent | Standard Error of Percent | 95% Confidence Limits for Percent | |
|---|---|---|---|---|---|---|---|---|---|---|
| One-unit building, detached from any other building | Owned | 29395 | 62691705 | 314079 | 62071429 | 63311980 | 47.1962 | 0.2369 | 46.7283 | 47.6641 |
| | Cash rent | 4904 | 10448469 | 148466 | 10155264 | 10741674 | 7.8659 | 0.1117 | 7.6452 | 8.0866 |
| | No-cash rent | 521 | 1182554 | 59051 | 1065935 | 1299174 | 0.8903 | 0.0445 | 0.8025 | 0.9781 |
| | Not applicable | 4766 | 9069616 | 256275 | 8563498 | 9575734 | 6.8279 | 0.1928 | 6.4471 | 7.2087 |
| | Total | 39586 | 83392343 | 270357 | 82858415 | 83926272 | 62.7802 | 0.2032 | 62.3789 | 63.1816 |
| One-unit building, attached to one or more buildings | Owned | 2608 | 4028501 | 102259 | 3826550 | 4230452 | 3.0328 | 0.0770 | 2.8807 | 3.1849 |
| | Cash rent | 1877 | 2497683 | 78259 | 2343129 | 2652236 | 1.8803 | 0.0589 | 1.7640 | 1.9967 |
| | No-cash rent | 71 | 93272 | 16239 | 61201 | 125343 | 0.0702 | 0.0122 | 0.0461 | 0.0944 |

| NUNIT2 | TENURE | Frequency | Weighted Frequency | Standard Error of Wgt Freq | 95% Confidence Limits for Wgt Freq | | Percent | Standard Error of Percent | 95% Confidence Limits for Percent | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Not applicable | 711 | 961365 | 60265 | 842347 | 1080382 | 0.7237 | 0.0454 | 0.6341 | 0.8133 |
| | Total | 5267 | 7580820 | 153100 | 7278463 | 7883177 | 5.7071 | 0.1153 | 5.4793 | 5.9348 |
| Building with two or more apartments | Owned | 2068 | 3572330 | 91823 | 3390988 | 3753672 | 2.6894 | 0.0691 | 2.5528 | 2.8259 |
| | Cash rent | 16145 | 24105822 | 177021 | 23756223 | 24455420 | 18.1476 | 0.1334 | 17.8841 | 18.4111 |
| | No-cash rent | 226 | 314174 | 25759 | 263302 | 365045 | 0.2365 | 0.0194 | 0.1982 | 0.2748 |
| | Not applicable | 3703 | 5263407 | 195408 | 4877495 | 5649320 | 3.9625 | 0.1471 | 3.6720 | 4.2529 |
| | Total | 22142 | 33255732 | 217877 | 32825447 | 33686017 | 25.0359 | 0.1639 | 24.7121 | 25.3597 |
| Manufactured (mobile) home | Owned | 1781 | 5357738 | 114056 | 5132490 | 5582987 | 4.0335 | 0.0859 | 3.8639 | 4.2030 |
| | Cash rent | 432 | 1358654 | 71688 | 1217078 | 1500230 | 1.0228 | 0.0540 | 0.9163 | 1.1294 |
| | No-cash rent | 69 | 200774 | 24901 | 151598 | 249951 | 0.1511 | 0.0187 | 0.1141 | 0.1882 |
| | Not applicable | 619 | 1686074 | 112992 | 1462926 | 1909222 | 1.2693 | 0.0851 | 1.1013 | 1.4373 |
| | Total | 2901 | 8603240 | 140914 | 8324950 | 8881531 | 6.4768 | 0.1061 | 6.2672 | 6.6863 |
| Total | Owned | 35852 | 75650274 | 307982 | 75042038 | 76258509 | 56.9518 | 0.2326 | 56.4924 | 57.4111 |
| | Cash rent | 23358 | 38410627 | 168557 | 38077744 | 38743510 | 28.9167 | 0.1271 | 28.6657 | 29.1677 |
| | No-cash rent | 887 | 1790774 | 74940 | 1642775 | 1938773 | 1.3481 | 0.0564 | 1.2367 | 1.4596 |
| | Not applicable | 9799 | 16980461 | 439998 | 16111508 | 17849414 | 12.7834 | 0.3311 | 12.1295 | 13.4373 |
| | Total | 69896 | 132832136 | 7750 | 132816831 | 132847441 | 100.0000 | | | |

## Appendix C. Linear Regression with PROC SURVEYREG

This program uses PROC SURVEYREG to estimate a linear regression equation with monthly housing cost (*ZSMHC*) as a function total number of rooms, number of bathrooms, tenure, region, and structure type (*NUNIT2*), using 2013 national American Housing Survey data. The WHERE statement excludes noninterview cases, which have weights of zero. Filtering these out does not cause a problem for the replicate weights, as noninterview cases are not part of the sample. *TENURE*, *REGION*, and *NUNIT2* are listed on a *CLASS* statement, because they are character-format categorical variables. SAS automatically builds dummy variables for the values they take. The /REF=FIRST option means that the first value of the variable should be used as the reference category. The *MODEL* statement uses the *CLPARM* option to display confidence intervals for the parameter estimates and *SOLUTION* to display the parameter estimates. A *FORMAT* statement specifies value labels (defined in a *PROC FORMAT* step, not shown) for the categorical variables.

Note: The replicate weight statement keyword is *REPWEIGHTS* (plural) in this procedure, not *REPWEIGHT*, as it is in *SURVEYMEANS* and *SURVEYFREQ*.[4]

```
PROC SURVEYREG DATA=RegExample
  VARMETHOD=BRR(FAY); /* end of proc statement */
  WHERE Wgt90Geo > 0;
  CLASS Tenure Region NUNit2 /REF=FIRST;
  MODEL ZSMHC = Rooms Baths Tenure Region NUNIT2
   /CLPARM SOLUTION;
  WEIGHT Wgt90Geo;
  FORMAT
   TENURE $TENURE.
   Region $Region.
   NUnit2 $Nunittwo.
  ; /* end of format statement */
  REPWEIGHTS RepWgt1-RepWgt160;
RUN;
```

The parameter estimates table follows. The parameters of the reference categories are 0.

**Exhibit C.1. Parameter Estimates - Estimated Regression Coefficients**

| Parameter | Estimate | Standard Error | T Value | Pr > \|t\| | 95% Confidence Interval | |
|---|---|---|---|---|---|---|
| Intercept | -104.66166 | 18.1479232 | -5.77 | <.0001 | -140.50202 | -68.82130 |
| ROOMS | 116.61517 | 4.2212006 | 27.63 | <.0001 | 108.27871 | 124.95162 |
| BATHS | 365.59410 | 9.4725956 | 38.59 | <.0001 | 346.88666 | 384.30155 |
| TENURE Occupied without payment of rent | -795.83079 | 20.3353932 | -39.14 | <.0001 | -835.99119 | -755.67039 |
| TENURE Owned or being bought by someone in your household | -13.15237 | 10.3528143 | -1.27 | 0.2058 | -33.59816 | 7.29342 |
| TENURE Rented | 0.00000 | 0.0000000 | . | . | 0.00000 | 0.00000 |
| REGION Northeast | 403.19262 | 10.7208018 | 37.61 | <.0001 | 382.02009 | 424.36515 |
| REGION South | -31.97085 | 10.0564427 | -3.18 | 0.0018 | -51.83134 | -12.11037 |
| REGION West | 318.00654 | 17.7286402 | 17.94 | <.0001 | 282.99423 | 353.01886 |

[4] In the Display Manger editor (SAS version 9.4), this keyword appears in red as if it were a syntax error, but it is correct.

| Parameter | Estimate | Standard Error | T Value | Pr > \|t\| | 95% Confidence Interval | |
|---|---|---|---|---|---|---|
| REGION Midwest | 0.00000 | 0.0000000 | . | . | 0.00000 | 0.00000 |
| NUNIT2 Manufactured (mobile) home | -571.74937 | 15.9383553 | -35.87 | <.0001 | -603.22606 | -540.27269 |
| NUNIT2 One-unit building, attached to one or more buildings | -71.49472 | 17.5224792 | -4.08 | <.0001 | -106.09989 | -36.88955 |
| NUNIT2 One-unit building, detached from any other building | -183.40940 | 12.7733021 | -14.36 | <.0001 | -208.63542 | -158.18339 |
| NUNIT2 Building with two or more apartments | 0.00000 | 0.0000000 | | | 0.00000 | 0.00000 |

## Appendix D. Logistic Regression Using PROC SURVEYLOGISTIC

This example uses *PROC SURVEYLOGISTIC* to model ownership as a function of region; metropolitan status; household income; persons in household; and the householder's race, ethnicity, age, and education.

```
PROC SURVEYLOGISTIC DATA=LogExample
  VARMETHOD=BRR(FAY); /* end of proc statement */
  WHERE Wgt90GEO > 0;
  CLASS Region(REF='South')
   Metro3(REF='Metro Urban')
   HHRace(REF='White')
   HHSPAN(REF='Not Hispanic')
   HHAGE(REF='30-64')
   HHGRAD(REF='More than High School')
   ;  /* end of class */
  MODEL TENURE(DESC) = ZInc2 Per Region Metro3 HHRace HHSPAN HHAGE HHGRAD
   /CLPARM CLODDS; /* end of model */
  WEIGHT Wgt90Geo;
  REPWEIGHTS RepWgt1-RepWgt160;
  FORMAT
   Tenure $owner.
   HHRace $race.
   HHSpan $hispanic.
   HHGrad edu.
   HHAGE age.
   Region $region.
   Metro3 $metro.
  ;  /* end of format */
RUN;
```

The syntax is like *PROC SURVEYREG*. Note that several variables are given formats. *HHGRAD* and *HHAGE* are numeric variables, but the formats essentially recoded them into categories. The *CLASS* statement specifies the reference category for each variable individually. The *MODEL* statement includes the *CLPARM* and *CLODDS* options to show the confidence intervals of the parameter and odds ratio estimates, respectively. The (DESC) option after *TENURE* indicates that the model is estimated for *TENURE='1'* (owner).

The parameter estimates and odd ratio exhibits follow.

**Exhibit D.1. The SURVEYLOGISTIC Procedure - Parameter Estimates and t Confidence Intervals**

| | | Estimate | 95% Confidence Limits | |
|---|---|---|---|---|
| Intercept | | -1.0132 | -1.0762 | -0.9501 |
| ZINC2 | Household income | 0.000013 | 0.000013 | 0.000014 |
| PER | Persons in household | 0.1071 | 0.0912 | 0.1231 |
| REGION | Midwest | 0.2194 | 0.2006 | 0.2382 |
| REGION | Northeast | -0.1874 | -0.2100 | -0.1648 |
| REGION | West | -0.2241 | -0.2482 | -0.2000 |
| METRO3 | Central city | -0.6626 | -0.7045 | -0.6207 |
| METRO3 | Metro rural | 0.5514 | 0.4821 | 0.6207 |
| METRO3 | Nonmetro rural | 0.5421 | 0.4609 | 0.6233 |
| METRO3 | Nonmetro urban | -0.2608 | -0.3387 | -0.1828 |
| HHRACE | Black | -0.3884 | -0.4219 | -0.3549 |
| HHRACE | Other | -0.0656 | -0.1255 | -0.00568 |
| HHSPAN | Hispanic | -0.2574 | -0.2793 | -0.2355 |
| HHAGE | 65 plus | 1.2836 | 1.2547 | 1.3125 |
| HHAGE | Less than 30 | -1.4043 | -1.4466 | -1.3620 |
| HHGRAD | High school | 0.0680 | 0.0274 | 0.1086 |
| HHGRAD | Not high school | -0.2573 | -0.3044 | -0.2102 |

Note: The degrees of freedom in computing the confidence limits is 160.

**Exhibit D.2. The SURVEYLOGISTIC Procedure Odds - Ratio Estimates and t Confidence Intervals**

| Effect | Unit | Estimate | 95% Confidence Limits | |
|---|---|---|---|---|
| ZINC2 | 1.0000 | 1.000 | 1.000 | 1.000 |
| PER | 1.0000 | 1.113 | 1.095 | 1.131 |
| REGION Midwest versus South | 1.0000 | 1.028 | 0.998 | 1.059 |
| REGION Northeast versus South | 1.0000 | 0.684 | 0.661 | 0.708 |
| REGION West versus South | 1.0000 | 0.660 | 0.635 | 0.685 |
| METRO3 central city versus metro urban | 1.0000 | 0.611 | 0.577 | 0.647 |
| METRO3 metro rural versus metro urban | 1.0000 | 2.057 | 1.868 | 2.266 |
| METRO3 nonmetro rural versus metro urban | 1.0000 | 2.038 | 1.847 | 2.249 |
| METRO3 nonmetro urban versus metro urban | 1.0000 | 0.913 | 0.828 | 1.007 |

| | | | | |
|---|---|---|---|---|
| HHRACE Black versus White | 1.0000 | 0.431 | 0.416 | 0.446 |
| HHRACE other versus White | 1.0000 | 0.595 | 0.542 | 0.652 |
| HHSPAN Hispanic versus non-Hispanic | 1.0000 | 0.598 | 0.572 | 0.624 |
| HHAGE 65 plus versus 30–64 | 1.0000 | 3.199 | 3.087 | 3.316 |
| HHAGE less than 30 versus 30–64 | 1.0000 | 0.218 | 0.204 | 0.232 |
| HHGRAD high school versus more than high school | 1.0000 | 0.886 | 0.836 | 0.939 |
| HHGRAD not high school versus more than high school | 1.0000 | 0.640 | 0.596 | 0.687 |

Note: The degrees of freedom in computing the confidence limits is 160.

U.S. Department of Housing and Urban Development

Office of Policy Development and Research

Washington, DC 20410-6000