Source of the Data and Accuracy of the Estimates for the Household Pulse Survey – Phase 3.8

Interagency Federal Statistical Rapid Response Survey to Measure Household Experiences during the Coronavirus (COVID-19) Pandemic

SOURCE OF THE DATA

The Household Pulse Survey (HPS), an experimental data product, is an Interagency Federal Statistical Rapid Response Survey developed to Measure Household Experiences during the Coronavirus (COVID-19) Pandemic, conducted by the United States Census Bureau in partnership with 16 other Federal agencies and offices:

- Bureau of Labor Statistics (BLS)
- Bureau of Transportation Statistics (BTS)
- Centers for Disease Control and Prevention (CDC)
- Consumer Financial Protection Bureau (CFPB)
- Department of Defense (DOD)
- Energy Information Administration (EIA)
- Department of Health and Human Services (HHS/ASPE)
- Department of Housing and Urban Development (HUD)
- Food and Drug Administration
- Maternal and Child Health Bureau (MCHB)
- National Center for Education Statistics (NCES)
- National Center for Health Statistics (NCHS)
- National Center for Immunization and Respiratory Diseases (NCIRD)
- National Institute for Occupational Safety and Health (NIOSH)
- USDA Economic Research Service (ERS)
- USDA Food and Nutrition Service (FNS)
- The White House Council of Economic Advisers (CEA)
- The White House Domestic Policy Council (DPC)

These agencies collaborated on the design and provided content for the HPS, which was also reviewed and approved by OMB. (OMB # 0607-1013; expires October 31, 2023.)

The Household Pulse Survey (HPS) ended Phase 3.7 on March 13, 2023. We entered Phase 3.8 to continue collecting information on how the coronavirus pandemic and other emergent issues are impacting households across the country with modifications to the questions. Working with the Office of Management and Budget (OMB), the HPS has been approved to continue collecting the HPS with an expiration date of October 31, 2023. In Phase 3.8, data are planned to be collected for 13 days with the next data collection beginning approximately two weeks later.

The HPS continues asking about core demographic household characteristics (including sexual orientation and gender identity), as well as asking questions about the following topics:

- Access to infant formula
- Childcare arrangements and cost
- COVID-19 vaccinations and long COVID symptoms and impact
- Education, specifically K-12 enrollment
- Employment
- Food sufficiency
- Housing security
- Household spending, including energy expenditures and consumption
- Inflation concerns and changes in behavior due to increasing prices
- Physical and mental health
- Rental assistance from state and local governments
- Transportation, including behavioral changes related to the cost of gas
- Health insurance coverage (including Medicaid)
- Impact of living through natural disasters

The HPS is designed to produce estimates at three different geographical levels. The first level, the lowest geographical area, is for the 15 largest Metropolitan Statistical Areas (MSAs). The second level of geography is state-level estimates for each of the 50 states plus the District of Columbia, and the final level of geography is national-level estimates.

The U.S. Census Bureau conducted Phase 1 of the HPS every week starting April 23, 2020. For details of Phase 1 see the Source and Accuracy Statements at: https://www.census.gov/programs-surveys/household-pulse-survey/technical-documentation.html. Phase 1 of the Household Pulse Survey was collected and disseminated on a weekly basis. Phase 1 collection ended July 21, 2020.

Table 1 provides the beginning and end dates along with the associated reference weeks for Phase 2 through 3.7. Despite going to a two-week collection period (in Phases 2 – 3.7), the Household Pulse Survey continues to call these collection periods "weeks" for continuity with Phase 1.

Table 1. Beginning and End Dates with Associated Reference Weeks for Phases 2 Through 3.7

Phase*	Beginning Date	Reference Week	Ending Date	Reference Week
2	August 19, 2020	13	October 26, 2020	17
3**	October 28, 2020	18	March 31, 2021	27
3.1	April 14, 2021	28	July 5, 2021	33
3.2	July 21, 2021	34	October 11, 2021	39
3.3	December 1, 2021	40	February 7, 2022	42
3.4***	March 1, 2022	43	May 9, 2022	45
3.5	June 1, 2022	46	August 8, 2022	48
3.6	September 14 ,2022	49	November 14, 2022	51
3.7	December 9, 2022	52	February 13, 2023	54

* Despite going to a two-week collection period (in Phases 2 – 3.7), the Household Pulse Survey continues to call these collection periods "weeks" for continuity with Phase 1.
** Phase 3 began with long term approval from OPM.
*** Phase 3.4 introduced new questionnaire content.

Table 2 provides the start and end dates for the current phase of data collection.

Table 2. Data Collection Periods for Phase 3.8 of the Household Pulse Survey

Data Collection Period	Start Date	Finish Date
Phase 3.8 – Week* 55	March 1, 2023	March 13, 2023

* For Phase 3.8 the Household Pulse Survey continues to call these collection periods "weeks" for continuity with Phase 1.

Sample Design

The HPS utilizes the Census Bureau's Master Address File (MAF) as the source of sampled housing units (HUs). Phases 1-3 utilized the January 2020 MAF and Phase 3.1 and 3.2 utilized updates to the MAF as of January 2021. Phase 3.3 takes advantage of the July 2021 MAF updates. Updates from the 2020 Census were incorporated into January 2022 MAF. The source of the sampled HUs for Phase 3.4, Phase 3.5 and Phase 3.6 used the January 2022 MAF. Phase 3.7 introduced updates to the MAF from July 2022 and Phase 3.8 continues to use the July 2022 MAF.

The sample design was a systematic sample of all eligible HUs, with adjustments applied to the sampling intervals to select a large enough sample to create state level estimates[1] and estimates for the top 15 MSAs. Sixty-six independent sample areas were defined. For each data collection period, independent samples were selected, and each sampled HU was interviewed once, unlike the Phase 1 of the HPS.

Sample sizes were determined such that a three percent coefficient of variation (CV) for an estimate of 40 percent of the population would be achieved for all sample areas, the exception being in the 11 smallest states. In these smaller states, the sample size was reduced to produce a 3.5 percent CV. The overall sample sizes within the sampling areas were adjusted for an anticipated response rate of nine percent. For those counties in one of the top MSAs, the sampling interval was adjusted to select the higher of the sampling rate for either the state or MSA.

To enable the use of a rapid deployment internet response system, we added email and mobile telephone numbers from the Census Bureau Contact Frame to the MAF. Since 2013, the Census Bureau has maintained contact frames to allow appended contact information onto sample units within household sample frames to aid in contacting respondents at those households. The primary motivation for creating this contact frame was to support research on potential contact strategies for the 2020 Census.

The Contact Frame information is maintained in two separate files – one containing phone numbers (both landline and cell phones) and the other containing email addresses. Information is obtained primarily from commercial sources, with additions from respondents to the American Community Survey (ACS) and Census tests, as well as participants in some government assistance programs from a few states, as well as from the Alaska Permanent Fund Division. Commercial sources were evaluated against

[1] Including the District of Columbia as a state.

respondent reported phone numbers to determine which sources would be acquired, after determining which vendors provided the best value for the government.

Commercial, survey, and administrative record data providers link phone numbers and email addresses to physical addresses before providing them to the Census Bureau for incorporation into the Contact Frames. Addresses are matched to the MAF. For addresses matched with confidence, the contact information is added to the frame along with the unique identifier from the MAF. In Phase 3.3 we began using the Contact Frame updated with information gathered from the 2020 Census.

Approximately 147 million HUs are represented in the MAF and considered valid for sampling. After matching to the contact frame and removing previous sampled cases that have opted out of future interviewing, approximately 127 million addresses are eligible HUs for the HPS. Of the 147 million addresses in the MAF, 76 percent of valid addresses are associated with at least one email, and 74 percent of valid addresses with at least one cell-phone number. The updated contact frame has at least one email or one cell-phone number for 86 percent of valid addresses. Unique phone numbers and email addresses are identified using a de-duplication process and assigned to only one HU. Only valid addresses with a phone number and/or email address are included on the Contact Frame as the final, eligible HUs for the HPS. Table 3 shows the number of addresses with updated contact information.

Table 3. Number of Addresses on the Master Address File, as of July 2022, Eligible for the Household Pulse Survey

Total Addresses	146,672,000
Addresses with any contact information	126,583,000
Addresses with cell phone	110,011,000
Addresses with email	111,374,000

Source: U.S. Census Bureau Master Address File Extracts and Contact Frame
Note: The counts for last three cells of this table exclude the addresses that have opted out of future interviewing for HPS.

Sampled households are contacted by both email and SMS if both are available. Only emails from domains with an expected deliverability rate of 90% or more were kept in sample. Emails and SMS invitations were only sent on weekdays and reminders are sent to nonrespondents.

The Census Bureau conducted the HPS online using Qualtrics as the data collection platform. Qualtrics is currently used at the Census Bureau for research and development surveys and provides the necessary agility to deploy the HPS quickly and securely. It operates in the Gov Cloud, is FedRAMP authorized at the moderate level, and has an Authority to Operate from the Census Bureau to collect personally-identifiable and Title-protected data.

Approximately 1,051,000 housing units were selected from the sampling frame for the first collection period of Phase 3.8. Approximately 73,000 respondents answered the online questionnaire. Table 4 shows the sample sizes and the number of responses by collection period for Phase 3.8 of the HPS.

Table 4. Sample Size and Number of Respondents at the National Level

Data Collection Period	Sample Size	Number of Respondents
Phase 3.8 - Week* 55	1,050,882	72,738

Source: U.S. Census Bureau, Household Pulse Survey.
*For Phase 3.8 the Household Pulse Survey continues to call these collection periods "weeks" for continuity with Phase 1.

State-level sample sizes and number of responses can be found in Table A1 on the Appendix A1 tab in the State-level Quality Measures spreadsheet at https://www.census.gov/programs-surveys/household-pulse-survey/technical-documentation.html under the Source and Accuracy Statements section.

Estimation Procedure

The final HPS weights are designed to produce biweekly estimates for the total persons age 18 and older living within HUs. These weights were created by adjusting the household-level sampling base weights by various factors to account for nonresponse, adults per household, and coverage.

The sampling base weights for each incoming sample in each of the 66 sample areas are calculated as the total eligible HUs in the sampling frame divided by the number of eligible HUs selected for interviews each week. Therefore, the base weights for all sampled HUs sum to the total number of HUs for which contact information is known.

The final HPS person weights are created by applying the following adjustments to the sampling base weights:

1. Nonresponse adjustment – the weight of all sample units that did not respond to the HPS are evenly allocated to the units that did respond within the same sample collection period, sample area (MSA or balance of state) and state. After this step, the weights of all respondents sum to the total HUs with contact information in the sampling frame.
2. Occupied HU ratio adjustment – this adjustment corrects for undercoverage in the sampling frame by inflating the HU weights after the nonresponse adjustment to match independent controls for the number of occupied HUs within each state. Each sampled respondent was assigned to the state where they reported their current address, which may be different from the selected state. For this adjustment, the independent controls are the 2021 American Community Survey (ACS) one-year, state-level estimates available at www.census.gov[2]. These controls were updated for Phase 3.7.

[2] The one-year estimates are at this URL: B25002: OCCUPANCY STATUS - Census Bureau Table

3. Person adjustment – this adjustment converts the HU weights into person weights by multiplying them by the number of persons age 18 and older that were reported to live within the household. The number of adults is based on subtracting the number of children under 18 in the household from the number of total persons in the household. This number was capped at 10 adults. If the number of total persons and number of children was not reported, then it is imputed.
4. Iterative Raking Ratio to Population Estimates – this procedure controls the person weights to independent population controls by various demographics within each state. The ratio adjustment is done through an iterative raking procedure to simultaneously control the sample estimates to two sets of population controls (both updated for Phase 3.8) -- Educational attainment estimates from the 2021 1-year ACS estimates (Table B15001)[3] by age and sex, and the July 1, 2023 Hispanic origin/race by age and sex estimates from the Census Bureau's Population Estimates Program (PEP). PEP provided July 1, 2023 household population estimates by single year of age (0-84, 85+), sex, race (31 groups), and Hispanic origin for states from the Vintage 2023 estimates series[4]. The ACS 2021 estimates were adjusted to match the 2023 pop controls within states by sex, and the five age categories in the ACS educational attainment estimates. Tables 5 and 6 show the demographic groups formed.

   Before the raking procedure was applied, cells containing too few responses were collapsed to ensure all cells met the minimum response count requirement. The cells after collapsing remained the same throughout the raking. These collapsed cells were used in the calculation of replicate weights.

Table 5: Educational Attainment Population Adjustment Cells within State

Age	No HS diploma Male	No HS diploma Female	HS diploma Male	HS diploma Female	Some college or Associate's degree Male	Some college or Associate's degree Female	Bachelor's degree or higher Male	Bachelor's degree or higher Female
18-24								
25-34								
35-44								
45-64								
65+								

[3] The1-year state-level detailed table B15001 is located at this URL: B15001 - Census Bureau Tables.
[4] The Vintage 2022 estimates methodology statement is available at this URL: methods-statement-v2022.pdf (census.gov). Note: The Vintage 2023 methodology has not yet been released – The Vintage 2022 methodology has been provided for reference.
The Modified Race Summary File methodology statement is available at this URL: https://www2.census.gov/programs-surveys/popest/technical-documentation/methodology/modified-race-summary-file-method/mrsf2010.pdf

Table 6: Race/Ethnicity Population Adjustment Cells within State

Age	Hispanic Any Race Male	Hispanic Any Race Female	Non-Hispanic White-Alone Male	Non-Hispanic White-Alone Female	Non-Hispanic Black-Alone Male	Non-Hispanic Black-Alone Female	Non-Hispanic Other Races Male	Non-Hispanic Other Races Female
18-24								
25-29								
30-34								
35-39								
40-44								
45-49								
50-54								
55-64								
65+								

Starting in week 13, the microdata file also contains a household weight for creating estimates of household-level characteristics. The final HPS household weights are created by applying the following adjustments to the final HPS person weights:

1. Housing Unit adjustment – this adjustment converts the person level weight back into a HU weight by dividing the person level weight by the number of persons age 18 and older that were reported to live within the household. The number of adults is the same value used to create the person adjustment.
2. Occupied HU ratio adjustment – this adjustment ensures that the final HPS household weights will sum to the 2021 American Community Survey (ACS) one-year, state-level estimates available at www.census.gov[2]. This ratio adjustment is the same adjustment applied to the person weights but is needed again because state totals may have changed as a result of the iterative raking adjustment in the final step of the person weight creation.

The detailed tables released for this experimental Household Pulse Survey show frequency counts rather than percentages. Showing the frequency counts allows data users to see the count of cases for each topic and variable that are in each response category and in the 'Did Not Report' category. This 'Did Not Report' category is not a commonly used data category in U.S. Census Bureau tables. Most survey programs review these missing data and statistically assign them to one of the other response categories based on numerous characteristics.

In these tables, the Census Bureau recommends choosing the numerators and denominators for percentages carefully, so that missing data are deliberately included or excluded in these counts. In the absence of external information, the percentage based on only the responding cases will most closely match a percentage that would result from statistical imputation. Including the missing data in the denominator for percentages will lower the percentages that are calculated.

Microdata will be available by FTP in the future. Users may develop statistical imputations for the missing data but should ensure that they continue to be deliberate and transparent with their handling of these data.

ACCURACY OF THE ESTIMATES

A sample survey estimate has two types of error: sampling and nonsampling. The accuracy of an estimate depends on both types of error. The nature of the sampling error is known given the survey design; the full extent of the nonsampling error is unknown.

Sampling Error

Since the HPS estimates come from a sample, they may differ from figures from an enumeration of the entire population using the same questionnaires, instructions, and enumeration methods. For a given estimator, the difference between an estimate based on a sample and the estimate that would result if the sample were to include the entire population is known as sampling error. Standard errors, as calculated by methods described below in "Standard Errors and Their Use," are primarily measures of the magnitude of sampling error. However, the estimation of standard errors may include some nonsampling error.

Nonsampling Error

For a given estimator, the difference between the estimate that would result if the sample were to include the entire population and the true population value being estimated is known as nonsampling error. There are several sources of nonsampling error that may occur during the development or execution of the survey. It can occur because of circumstances created by the respondent, the survey instrument, or the way the data are collected and processed. Some nonsampling errors, and examples of each, include:

- Measurement error: The respondent provides incorrect information, the respondent estimates the requested information, or an unclear survey question is misunderstood by the respondent.
- Coverage error: Some individuals who should have been included in the survey frame were missed.
- Nonresponse error: Responses are not collected from all those in the sample or the respondent is unwilling to provide information.
- Imputation error: Values are estimated imprecisely for missing data.

To minimize these errors, the Census Bureau applies quality control procedures during all stages of the production process including the design of the survey, the wording of questions, and the statistical review of reports.

Two types of nonsampling error that can be examined to a limited extent are nonresponse and undercoverage.

Nonresponse

The effect of nonresponse cannot be measured directly, but one indication of its potential effect is the nonresponse rate. Table 7 shows the unit response rates by collection period.

Table 7. National Level Weighted Response Rates by Collection Period for the Household Pulse Survey

Data Collection Period	Response Rate (Percent)
Phase 3.8 - Week* 55	6.7

Source: U.S. Census Bureau, Household Pulse Survey
* For Phase 3.8 the Household Pulse Survey continues to call these collection periods "weeks" for continuity with Phase 1.

State-level response rates can be found in Table A1 on the Appendix A1 tab in the State-level Quality Measures spreadsheet at https://www.census.gov/programs-surveys/household-pulse-survey/technical-documentation.html under the Source and Accuracy Statements section.

In accordance with Census Bureau and Office of Management and Budget Quality Standards, the Census Bureau will conduct a nonresponse bias analysis to assess nonresponse bias in the HPS.

Responses are made up of complete interviews and sufficient partial interviews. A sufficient partial interview is an incomplete interview in which the household or person answered enough of the questionnaire to be considered a complete interview. Some remaining questions may have been edited or imputed to fill in missing values. Insufficient partial interviews are considered to be nonrespondents.

Undercoverage

The concept of coverage with a survey sampling process is defined as the extent to which the total population that could be selected for sample "covers" the survey's target population. Missed housing units and missed people within sample households create undercoverage in the HPS. A common measure of survey coverage is the coverage ratio, calculated as the estimated population before poststratification divided by the independent population control. The national household-level coverage ratio is 0.96. State household-level coverage ratios can be found in Table A1 on the Appendix A1 tab in the State-level Quality Measures spreadsheet at https://www.census.gov/programs-surveys/household-pulse-survey/technical-documentation.html under the Source and Accuracy Statements section.

HPS person coverage varies with age, sex, Hispanic origin/race, and educational attainment. Generally, coverage is higher for females than for males and higher for non-Blacks than for Blacks. This differential coverage is a general problem for most household-based surveys. The HPS weighting procedure tries to mitigate the bias from undercoverage within the raking procedure. However, due to small sample sizes, some demographic cells needed collapsing to increase sample counts within the raking cells. In this case convergence to both sets of the population controls was not attained. Therefore, the final

coverage ratios are not perfect for some demographic groups. Table 8 shows the coverage ratios for the person demographics of age, sex, Hispanic origin/race, and educational attainment before and after the raking procedure is run.

Table 8. Person-Level Coverage Ratios at the National Level for Household Pulse Survey Before and After Raking for Collection Week* 55: March 1, 2023 – March 13, 2023

Demographic Characteristic	Before Raking	After Raking
Total Population	1.06	1.01
Male	0.94	1.01
Female	1.17	1.01
Age 18-24	0.31	0.69
Age 25-29	0.68	1.14
Age 30-34	0.96	1.09
Age 35-39	1.22	1.07
Age 40-44	1.29	1.09
Age 45-49	1.34	1.04
Age 50-54	1.33	1.03
Age 55-64	1.25	1.00
Age 65+	1.10	1.01
Hispanic	0.76	1.03
Non-Hispanic white-only	1.19	1.00
Non-Hispanic black-only	0.76	0.96
Non-Hispanic other races	1.12	1.10
No high-school diploma	0.27	0.74
High-school diploma	0.49	1.10
Some college or associate's degree	1.15	1.00
Bachelor's degree or higher	1.69	1.00

Source: U.S. Census Bureau, Household Pulse Survey
* For Phase 3.8 the Household Pulse Survey continues to call these collection periods "weeks" for continuity with Phase 1.

The previous data collection's national person-level coverage ratios and state person-level coverage ratios can be found in Table A2 on the Appendix A2 tab in the State-level Quality Measures spreadsheet at https://www.census.gov/programs-surveys/household-pulse-survey/technical-documentation.html under the Source and Accuracy Statements section.

Biases may also be present when people who are missed by the survey differ from those interviewed in ways other than age, sex, Hispanic origin/race, educational attainment, and state of residence. How this weighting procedure affects other variables in the survey is not precisely known. All of these considerations affect comparisons across different surveys or data sources.

Comparability of Data

Data obtained from the HPS and other sources are not entirely comparable. This is due to differences in data collection processes, as well as different editing procedures of the data, within this survey and others. These differences are examples of nonsampling variability not reflected in the standard errors. Therefore, caution should be used when comparing results from different sources.

A Nonsampling Error Warning

Since the full extent of the nonsampling error is unknown, one should be particularly careful when interpreting results based on small differences between estimates. The Census Bureau recommends that data users incorporate information about nonsampling errors into their analyses, as nonsampling error could impact the conclusions drawn from the results. Caution should also be used when interpreting results based on a relatively small number of cases.

Standard Errors and Their Use

A sample estimate and its standard error enable one to construct a confidence interval. A confidence interval is a range about a given estimate that has a specified probability of containing the average result of all possible samples. For example, if all possible samples were surveyed under essentially the same general conditions and using the same sample design, and if an estimate and its standard error were calculated from each sample, then approximately 90 percent of the intervals from 1.645 standard errors below the estimate to 1.645 standard errors above the estimate would include the average result of all possible samples.

A particular confidence interval may or may not contain the average estimate derived from all possible samples, but one can say with the specified confidence that the interval includes the average estimate calculated from all possible samples.

The context and meaning of the estimate must be kept in mind when creating the confidence intervals. Users should be aware of any "natural" limits on the bounds of the confidence interval for a characteristic of the population when the estimate is near zero – the calculated value of the lower bound of the confidence interval may be negative. For some estimates, a negative lower bound for the confidence interval does not make sense, for example, an estimate of the number of people with a certain characteristic. In this case, the lower confidence bound should be reported as zero. For other estimates such as income, negative confidence bounds can make sense; in these cases, the lower confidence interval should not be adjusted. Another example of a natural limit is 100 percent as the upper bound of a percent estimate.

Standard errors may also be used to perform hypothesis testing, a procedure for distinguishing between population parameters using sample estimates. The most common type of hypothesis is that the population parameters are different. An example of this would be comparing the household distributions in spending sources in the last seven days between weeks 51 and 52.

Tests may be performed at various levels of significance. A significance level is the probability of concluding that the characteristics are different when, in fact, they are the same. For example, to conclude that two characteristics are different at the 0.10 level of significance, the absolute value of the estimated difference between characteristics must be greater than or equal to 1.645 times the standard error of the difference.

The Census Bureau uses 90-percent confidence intervals and 0.10 levels of significance to determine statistical validity. Consult standard statistical textbooks for alternative criteria.

Estimating Standard Errors

The Census Bureau uses successive difference replication to estimate the standard errors of HPS estimates. These methods primarily measure the magnitude of sampling error. However, they do measure some effects of nonsampling error as well. They do not measure systematic biases in the data associated with nonsampling error. Bias is the average over all possible samples of the differences between the sample estimates and the true value.

Eighty replicate weights were created for the HPS. Using these replicate weights, the variance of an estimate (the standard error is the square root of the variance) can be calculated as follows:

$$Var(\hat{\theta}) = \frac{4}{80}\sum_{i=1}^{80}(\theta_i - \hat{\theta})^2 \quad (1)$$

where $\hat{\theta}$ is the estimate of the statistic of interest, such as a point estimate, ratio of domain means, regression coefficient, or log-odds ratio, using the weight for the full sample and θ_i are the replicate estimates of the same statistic using the replicate weights. See reference Judkins (1990).

Creating Replicate Estimates

Replicate estimates are created using each of the 80 weights independently to create 80 replicate estimates. For point estimates, multiply the replicate weights by the item of interest to create the 80 replicate estimates. You will use these replicate estimates in the formula (1) to calculate the total variance for the item of interest. For example, say that the item you are interested in is the difference in the number of people with a loss in employment income in week 1 compared to the number of people with a loss in employment income in week 2. You would create the difference of the two estimates using the sample weight, $\hat{x}_0$, and the 80 replicate differences, x_i, using the 80 replicate weights. You would then use these estimates in the formula to calculate the total variance for the difference in the number of people with a loss in employment income from week 1 to week 2.

$$Var(\hat{x}_0) = \frac{4}{80}\sum_{i=1}^{80}(x_i - \hat{x}_0)^2$$

Where x_i is the ith replicate estimate of the difference and $\hat{x}_0$ is the full estimate of the difference using the sample weight.

Users may want to pool estimates over multiple weeks by creating averages for estimates with small sample sizes. For pooled estimates, where two or more weeks of data are combined to make one estimate for a longer time period, one would divide the unit-level weights that formed $\hat{x}_0$ and x_i (for each of the 80 replicate weights) for each week by the number of weeks that are combined. Then, form 80 replicate pooled estimates, $\hat{x}_{i,pooled}$ and the estimate, $\hat{x}_{0,pooled}$. Then use the pooled estimates in formula (1) to calculate the pooled variance for the item of interest.

Example for Variance of Regression Coefficients

Variances for regression coefficients β_0 can be calculated using formula (1) as well. By calculating the 80 replicate regression coefficients $\beta_i's$ for each replicate and plugging in the replicate β_i estimates and the β_0 estimate into the above formula,

$$Var(\hat{\beta}_0) = \frac{4}{80}\sum_{i=1}^{80}(\beta_i - \hat{\beta}_0)^2$$

gives the variance estimate for the regression coefficient β_0.

TECHNICAL ASSISTANCE

If you require assistance or additional information, please contact the Demographic Statistical Methods Division via e-mail at dsmd.source.and.accuracy@census.gov.

REFERENCES

Judkins, D. (1990) "Fay's Method for Variance Estimation," Journal of Official Statistics, Vol. 6, No. 3, 1990, pp.223-239.