Webinar: “Calculating Margins of Error the ACS Way”
February 12, 2020
2:00 PM (EST)

Coordinator: Welcome. Thank you for standing by. At this time all participants are on a listen-only mode until the question-and-answer session of today’s conference. At that time you may press Star 1 on your phone to ask a question. I would like to also inform all parties that today’s conference is being recorded. If you have any objections you may disconnect at this time. I would now like to turn the conference over to your host, Mr. Charles Gamble. Thank you, you may begin.

Charles Gamble: Good afternoon and good morning to those of you on the West Coast. My name is Charles Gamble and I’m a Supervisory Survey Statistician in the Outreach and Education Branch in the American Community Survey’s Office. And I’m accompanied by Mr. Sirius Fuller, a Mathematical Statistician in the Decennial Statistical Studies Division, who will be talking about calculating margins of error the ACS way using replicate methodology to calculate uncertainty.

Before we get started we would like to thank you for attending today’s webinar.

Getting started. Here’s what will be covered today. I’ll be giving a brief overview of the American Community Survey or commonly referred to as the ACS. After the ACS overview I will turn the presentation over to Sirius who will then discuss the Variance Replicate Estimates or VRE tables. Due to the size of this section it has been broken into four segments. The first segment being the motivation behind using the VRE tables. This is then followed by a comparison of the VRE tables to ACS detailed tables. Next you will learn
how to obtain the Variance Replicate Estimate’s data. And last you will work through an example how to calculate the margin of error using the VRE data.

After this, Sirius we will also give a brief overview of the public use microdata sample or PUMS files and show a second worked example on how to calculate the margin of error for a percent estimate. And to wrap it up we will open the line for questions. Also as a quick note, today’s slides from this presentation are available in PDF form on our ACS Web site in our event’s section and under our today’s event “Calculating Margins of Error the ACS Way”.

Let’s get started with a brief overview. The ACS is an annual survey that samples over 3.5 million addresses each year. The data collected is used to produce annual social, economic, housing and demographic estimates. These estimates cover more than 40 topics and support many federal uses as well as countless non-federal uses.

As you may know all ACS estimates are published with a measure of uncertainty in the form of margin of error or known as MOE. The MOE reflects the uncertainty due to sampling error. If you wish to learn more a link is provided for the ACS handbook on a later slide toward the end of this presentation.

Note the ACS uses a 90% confidence level margin of error. This will be relevant when we encounter the worked examples later in the webinar.

The ACS sends out surveys to households throughout the year in order to produce our well-known one-year and five-year data products. Twelve-month of survey responses are aggregated together to form the one-year ACS estimates which are available for geographic areas with a population of 65,000
or more. These estimates are published on data.census.gov in the form of detailed tables as well as many other data products such as data profiles.

The five-year ACS estimates are created from 60 months’ worth of responses and are available for geographic areas of all sizes down to the block group level. In addition to the detailed tables and a regular data of products published on data.census.gov, data are also published on the FTP site in the Variance Replicate Estimates tables. The VRE tables are selected ACS detailed tables that include some additional variables.

They allow data users to combine estimates both within a detailed table and between geographies. In addition, data users may use the Variance Replicate Estimates tables to calculate the margin of error for combined estimates using the methodology that is used to calculate the margin of error on data.census.gov.

Finally the Census Bureau publishes Public Use Microdata Sample or PUMS files. These files allow data users to calculate estimates and margins of error for characteristics which are not published in an ACS detailed table or any other data product. The PUMS files are microdata representing a sample of the full ACS records. The method used to calculate margins of error for the VRE tables may also be used with an estimate calculated using the PUMS data which is why there is a worked example using a PUMS file in today’s webinar.

I would now like to turn the webinar over to Mr. Sirius Fuller to discuss Variance Replicate Estimates tables.

Sirius Fuller: Thank you Charles. Once again my name is Sirius Fuller and I’m in DSSD - the Decennial Statistics Studies Division. So the Variance Replicate
Estimates or VRE tables are equivalent to the detailed tables published on data.census.gov. The VRE tables are published on the FTP site in comma separated values, or CSV, formats. This allows you to easily read the data into statistical software. You may then combine estimates and then calculate the margin of error using the same replicate methodology that the ACS uses to calculate margins of error that are seen on data.census.gov.

As I just said you may wish to combine published estimates to create your own estimates. You can combine estimates either by table, across geographies or both within a table and across geographies.

When you do so you will have to calculate the margin of error. If you use the detailed tables published on data.census.gov you will have to use an approximation formula to calculate the margin of error. The approximate margin of error may become unreasonably large compared to what you would obtain if you were able to calculate the margin of error using internal ACS microdata.

In addition the approximation formula you will have to use will change depending on the type of estimate you wish to calculate. So for example at the bottom of the slide we see two examples of the approximation formulas. The first is for calculating the margin of error when combining two count estimates. And the second is for approximating the margin of error for a percent estimate.

The Variance Replicate Estimates tables allow you to calculate the margin of error using the same methodology that is used to calculate the margin of error seen in ACS detailed tables. It’s not an approximation formula. In addition the same formula may be used for many different types of estimates which makes it a very versatile formula to use.
The formula for calculating the variance is shown in the middle of the slide. After you calculate the variance you can use it to calculate the margin of error. And we will see how to do this in the worked example later on. The variance formula may look complicated, but we will break it down into individual steps and is not as difficult to calculate as it may first appear.

I will now briefly compare the Variance Replicate Estimates tables to the ACS detailed tables published on data.census.gov.

To start off with here’s an example of the ACS detailed tables on data.census.gov. And as a note data.census.gov is the new dissemination platform for publishing ACS detailed tables and other data products such as the data profiles. It is the replacement to American Fact Finder or AFF. And this table shows estimates for sex by age at the national level.

Circled in red in the upper left is the table ID. The table for sex by age has the table ID of B01001. And the table ID is useful to know when working with data from the Variance Replicate Estimates tables. We will see an example in a few slides.

In addition data.census.gov allows you to switch between published years and time periods. This is circled in red at the top of the slide. The time period is next to the word product and says 2017 ACS five-year estimates detailed tables. And this is how it refers to the detailed tables for the 2013-2017 ACS 5-year data. It is also how it’s displayed on American Fact Finder.

As mentioned earlier by Charles, the ACS publishes 1- and 5-year estimates. However only 5-year estimates are published for the Variance Replicate Estimates tables. At the time these slides were created the 2013-2017 ACS 5-
year data were the most currently publicly available data. However the 2014-2018 ACS 5-year data is now available on the FTP site and it was published on January 30 of 2020.

I will now compare the detailed table to the equivalent Variance Replicate Estimates table.

So the variance - excuse me, the ACS detailed table we saw in the previous slide is shown at the bottom left on this slide. The equivalent data from the Variance Replicate Estimates or VRE table is on the top. The VRE tables are published in CSV format and statistics package such as SAS or R or STATA can read in and work with CSV data.

You may also examine a CSV file by opening it with Excel which is how the data are displayed here. Note that the VRE tables are equivalent to the ACS detailed tables with some added features. So let’s compare them.

Both the VRE table and the ACS detailed tables have data for estimates. The values, as well as the variable names, are the same.

Both also displayed the margin of error. In the ACS detailed table on the bottom the variable is simply called margin of error. For the VRE data on top the variable is called CME and CME stands for the character margin of error.

In addition, the VRE data also has a numeric version of the margin of error called MOE. The numeric version allows data users such as yourself to use the margin of error immediately without needing to remove the plus or minus signs from the character version of the variable. The character version allows for special characters such as when an estimate is controlled to the official population estimate. In this case the character version of the margin of error
has five stars or five asterisks. The numeric margin of error is considered to be zero as it matches the independent population estimates which have no sampling variance.

The Variance Replicate Estimates tables also include Variance Replicate Estimates. They are circled in red in the top right of the slide and called Var_Rep followed by a number. There are 80 Variance Replicate Estimates although only the first four are shown on the slide. The variables are unique to the VRE tables and do not appear on the ACS detailed tables or other data products found on data.census.gov.

The Var_Rep naming convention will appear again in the formula for calculating the variance and the margin of error when we do a worked example in a few slides.

Here’s a list of the variables on the VRE table along with a short description. And in the interest of time I will not go through them. This webinar as we said before is available as a PDF and will be published after this presentation as well.

I will now discuss how to obtain the VRE data by downloading it from the Census Bureau’s FTP site.

So as I just said the VRE tables are published on the FTP site. The tables are not currently available through data.census.gov. The FTP site may be accessed through the Census Bureau’s Web site. The URL on the right-hand side of the screen will take you to the page that is shown on the left. Click on the link circled in red that says “5-year Variance Replicate Tables”.
I will be using the 2013-2017 ACS 5-year data in this presentation. It was the most current data available and as I said before a few weeks ago the 2014-2018 5-year data was released. So if you wish you may use that. And as a side note you may also reach this page by going to the main census page, census.gov/acs and then clicking on “data” in the sidebar.

Before I show you how to access the data I would like to point out that you can also access the VRE technical documentation page by clicking on the link circled in red. This includes the technical documentation which goes into how to calculate the margin of error in greater depth. In addition there are a number of other useful things on the technical documentation page. There are parameter files needed to calculate the margin of error for estimates which are zero, a list of detailed tables and geographies that are published for the VRE tables and a link to the ACS table shells.

The table shells are the ACS detailed tables without any data so that you may see the layout of the table and what characteristics are available in each one of the tables. So I will now show you how to access the data.

Once you click on the link for the FTP site you’re taken to the ACS internal FTP site. You will see a list of directories to choose from. They are named using the three-digit geographic summary level codes. You can see them on the left-hand side of the slide. Note the only 11 geographic summary levels are published for the VRE tables. The description of this code is shown in the middle of the slide. And as I mentioned previously the information is also available on the VRE technical documentation page.

Once you choose a geographic summary level you will see a list of files to download. The names of the file is the table ID followed by “.csv.gz”. And the table ID is the same table ID listed in the detailed tables on
data.census.gov. So for example circled in red we have the first table B01001.csv.gz. And this is the same sex by age table that we saw in the previous examples. The “.csv” stands for common separated value and gz indicates the method used to compress the file, that is the gzip method.

On data.census.gov there are over 800 detailed tables published for the ACS 5-year data. Only about 100 of these tables are available as the Variance Replicate Estimates tables. However all ACS detailed tables used to create the ACS data profiles are available through the VRE tables.

Note also that the file names are the same with different geographic summary levels. If you downloaded B01001 file shown here for the national data and then changed to the geographic summary level for the state data, you would see a similar file which is also called B01001. The national file is circled in red on the far left of the slide and the state file is below it which is also circled in red.

If you saved the state file in the same location as the national file, you would write over the national data. So you’ll either have to save the national and state files in separate locations or rename one or both of the files as you download them.

Note that when downloading the data it is recommended to use either the Google Chrome browser or Apple’s Safari browser. If using the Mozilla Firefox browser first save the file before attempting to uncompress it. And finally if you’re using Microsoft Internet Explorer or Edge browsers the file extension may be renamed after you download it. I will show you how to address this in the next slide. If you’re using Chrome or Safari you will not need to worry about this issue which is why those are recommended.
So as I mentioned the files are compressed using the gzip format. If you download the VRE file using Internet Explorer or Edge the file’s extension may be renamed. It’ll change to the ending - from having an ending in .gz to having an ending of .tar. To fix this so you may uncompress the file, select view at the top of the file folder which is circled on the top left of the slide. And then click on the filename extensions to see the file’s extension. This is circled in the top right.

The file for B01001 is circled in the middle of the slide. With the file extension shown you can see that the filename ends in .tar instead of .gz. You’ll have to rename the .tar to .gz and once you do this Windows will pop up a warning. The warning is shown at the bottom right of the slide and click yes to accept the change.

Once you’ve changed the file extension back to .gz you should be able to uncompress the file using programs such as WinZip, 7Zip or Apple’s built in decompression program. If you download the file using IE or Edge and do not change the file’s extension name, you may receive an error message and the file may not uncompress.

Finally at the tract and block group geographic summary levels the names for the VRE files are different than at other geographic summary levels. The format for how the files are named is the table ID followed by an underscore and then the two-digit FIPS state code and then the familiar “.csv.gz”. So for example here we’ve got “B01001_01.csv.gz” which is circled in red. And this is the sex by age table with a FIPS state code “01” for Alabama. As a note the tract and block group geographies are the smallest geographies published for the ACS which is why they’re divided by state.
On the next slide I will show you how to obtain the state FIPS codes and their corresponding state names.

So as I said this is how to obtain the state FIPS codes and names. Note that when you go to the URL listed on the right-hand side of the slide you will need to click on the “FIPS Codes for the States and District of Columbia” which is circled in red. The state names, FIPS state codes and state abbreviations will then be displayed. And as a note Puerto Rico is not on the list. The section below it is called “FIPS State Codes for Outlying Areas in the United States and Freely Associated States”. And if you click on that you can see the FIPS state code for Puerto Rico. But to save you some time Puerto Rico has a FIPS state code of 72.

Now I will show you how to calculate the margin of error using the VRE tables.

So first some background. The method we are using is called the Successive Difference Replication method or SDR. An SDR is similar to the Balance Repeated Replication, or BRR, method as well as the Jackknife method. All of these methods use replicate estimates to calculate the variance.

The full method involves a number of steps. However, I will only focus on the formula used to compute the variance. As a note, the Public Use Microdata Sample, or PUMS, files contain replicate weights. And these replicate weights may be used to calculate replicate estimates which then may be used to calculate the margin of error and we’ll see an example of this in the worked example for the PUMS data.

And finally this method may be used to calculate the variance for many different types of estimates such as counts, percentages, means, or ratios.
So as I said the formula is similar to BRR or Jackknife for calculating the variance. The main difference is the presence of the “4/80” factor in the variance formula which is at the top of the screen. The formula may look tricky to use but we are going to break it down into steps in the next few slides. And after we calculate the variance, we will then find the standard error by taking the square root of the variance.

And then finally from the standard error we will multiply it by 1.645 to find the margin of error. Note that the ACS uses a two-sided 90% confidence level margin of error which uses a Z-score of 1.645 to convert from the standard error to the margin of error and so 1.645 in the slide.

So here’s the variance formula again. There are basically four steps to using the formula. The first step, labeled 1a, is to subtract the estimate from each replicate estimates to calculate 80 differences. And this is the portion of the formula within the parentheses that is the Var_Rep with a little sub “i” minus estimate.

Remember that we saw the variables of Var_Rep1, Var_Rep2, et cetera, et cetera when we examine the data from the variance replicate tables in the earlier slide. So this is the same thing. And the little sub “i” in this formula is used to denote that we’re counting from 1 to 80 which is the total number of replicate estimates.

In the second step, labeled 1b, square each difference. And then in 1c sum up all of these square differences. And finally define the variance we multiply this by 4/80. In Step 2 after we calculate the variance we find the standard error by taking the square root and then in Step 3 we multiply the standard
error by 1.645 to obtain the margin of error. And we’ll see these steps as we work through the example as we go along.

The margin of error I will be calculating for the worked example is for the estimate of males under the age of - excuse me ages 15 to 17 for a census tract in Alaska. And I chose this tract because the estimate is small making the calculation easier to follow. Although we will calculate the margin of error for a single estimate, the same method may be used when combining multiple estimates either within the table, across geographies or both within a table and across geographies.

And the data shown on this slide uses the data on data.census.gov, however we will use the data from the variance replicate estimates table.

So here’s the same data that we just saw but only in the variance replicate estimates table. Notice that the replicate estimates are circled in red and only four out of 80 of the variance replicates are shown. The bottom of the slide contains the information to find the data should you wish to work through this example by yourself.

So now we will work through calculating the margin of error. So first we’ll calculate the variance. Recall there were four steps to do this. The first was to create the 80 differences by subtracting the estimate from each replicate estimate. Note that I’m only showing four replicate estimates so as not to overclutter the slides. But you will need to use all 80 replicate estimates when calculating the margin of error. Note also that the variance formula is shown in the middle of the slide. And for each step the relevant portion will be highlighted in red.
So to calculate the first difference we subtract the estimate which is 45 from the first replicate estimate 44. So 44 minus 45 gives you -1. And if you want to do this for the second variance replicate estimate it will be 40 minus 45 which gets you -5 and so forth.

For the second set labeled 1b we will square each difference. So our first difference was -1 and -1 squared just gives you 1. And our second difference was -5. So -5 squared is 25 and so on.

Now for Step 1c we sum up the 80 squared differences. So we add up 1 plus 25 plus 4, et cetera, et cetera, et cetera. And if you do this you’ll get a sum of the 80 square differences equal to 1,458 which is circled in red at the top of the slide.

Finally for the fourth step called 1d we multiply the sum by 4/80 to obtain a variance. So we obtain 72.9.

Now that we have the variance we’ll take two more steps to find the margin of error. So first we calculate the standard error by taking the square root of the variance. So the square root is 72.9 is roughly 8.538.

And then in Step 3 we multiply the standard error by 1.645 to obtain a margin of error. And so if we do this we obtain 14.045.

Examining the original data from the VRE table we see that if we round our calculated standard error and margin of error to whole numbers they match the values for the standard error, called SE, and the margin of error, called MOE. And these are circled in red on the VRE table.
So for example for the standard error if we rounded 8.538 we would get 9 which is down below.

The worked examples we just did was for a count estimate. However remember that this method may be used for other types of estimates, such as means, ratios and percentages. When we do the worked example using the PUMS data we will see how to calculate a margin of error for a percent.

In addition there are several special cases that you may encounter. So for example you may want to know how to calculate the margin of error for an estimate which is zero. The special cases are covered in the technical documentation which may be found in the URL on the slide.

As a note the method we just used in the worked example may also be used when combining estimates. So for example, perhaps you wish to find the total number of males under the age of 18 in the tract of Alaska from our example. You would have to add up the four estimates circled in red in the middle of the slide to create a new estimate. And in addition you will need to combine the replicate estimates to create new replicate estimates for the males under 18.

So here’s the same data that I just showed you. So to find the estimates of males under the age of 18 you’d add up the estimates circled in red. So 42 plus 74, et cetera, et cetera. And you would get a total of 249.

And then you do the same thing for each replicate estimate. So for example for the first replicate estimate you’d add up the four estimates circled in red in the middle of the slide and obtain the new variance replicate estimate, first variance replicate estimate of 264.
Once you have the estimate and the 80 replicate estimates you may then use the variance formula and the steps from the previously worked examples to calculate the margin of error. In the interest of time I will not go through all the steps. However I’ve put the results of the variance and margin of error on the bottom of the slide should you wish to work through this yourself.

You may also combine estimates across geographies. So here’s an example for males age 15 to 17 for three census tracts in Alaska. And the first is for the tract - is the worked example that we just saw. So to obtain a combined estimate simply add up the estimates from the three geographies. So you would add up the 45 plus 3 plus 62 to create a new estimate. And you would also add up the replicate estimates in a similar fashion. Once you have your estimates and your 80 replicate estimates then you would use the steps from the worked example to calculate the variance and then the margin of error. And as a side note if you wish to combine estimates both within a table and between geographies you would do that as well using the same method.

So now I will switch gears and talk about the Public Use Microdata Sample, or PUMS, files.

The Public Use Microdata Sample, or PUMS, is a publicly available subsample of the ACS records. Additional restrictions are added to protect data confidentiality. The PUMS files allow data users to calculate their own estimates and margin of errors that may not be available on data.census.gov. Statistical software is recommended when working with the PUMS data.

And as a note I am including this section on PUMS for two reasons. The first is to show that the variance formula seen in the previous example may be used to calculate the margin of error with PUMS data. And the second reason is to
show how to calculate a margin of error for a different type of estimate, namely a percent estimate.

On the screen is the screenshot for accessing the data using the webpage. And as I said in the previous slide PUMS data provides individual records for the data users … the data users must aggregate to form estimates. Unlike on data.census.gov or in the VRE tables there are no pre-tabulated data. Weights are included on the PUMS files so that the data user may create weighted population estimates.

Data users will also need to calculate replicate estimates. And they may do so by summing the replicate weights and we’ll see what the replicate weights look like in the worked example in a few slides.

These slides are just a brief overview of the PUMS data. Data users may wish to examine the introduction to PUMS training for more information. In addition there is the PUMS technical documentation page which provides useful references such as the PUMS Data Dictionary and the PUMS Estimates for User Verification. And the PUMS Estimates for User Verification provides estimates, standard errors and margin of errors so that data users may check that they are using the PUMS files correctly.

As a note there will be a presentation of the Introduction to PUMS on March 11. It’s from 2 to 3 pm. The URL circled in red will take you to the announcement page which will provide a link and dial in information on the day of the training similar to this training.

So now we will work through the example. As I mentioned in the previous slides, PUMS provides individual records. To obtain an estimate sum the PUMS weights for the records which are for your characteristic. The estimate
we will calculate for this worked example is the percent of people age 18 and older. We will need the enumerator and the denominator to calculate the percent. The numerator is all records where the age is greater than or equal to 18. And the denominator will be all people. That is, all records.

The column called AGEP is for age. And we can see this by looking at the variable description on the right-hand side of the slide. The PUMS variables and their descriptions are available in the PUMS Data Dictionary found in the PUMS Technical Documentation page. The URL is provided at the end of this presentation in the reference section.

Note that there are many more PUMS records and variables but for this example we will only use the records shown here.

So to calculate the numerator we sum up the PUMS weight for the records where AGEP is greater or equal to 18. The PUMS weight variable is called PWGTP. So we would add those circled in red, the 5 plus 25 plus 79, et cetera, et cetera, and obtain a numerator of 399. Notice that in the last column there’s a variable called PWGTP1 and this is a first replicate weight and we’ll come back to that in a moment.

Before we do that we’ll calculate the denominator. The denominator has no restrictions so we sum up all of the weights. That is we add the column in red. And when we do that we would obtain a weighted denominator of 467. Notice that if our universe for the denominator were different we would not sum all the records. So for example if our denominator were for people age 10 and over then the weight for the third record where AGEP is 9 would not be included.
Now we will calculate the percent in the usual way. We divide the numerator by the denominator and then multiply by 100. And we obtain the percent of 85.4%. Notice that the rest of the table is blank. There are placeholders for the replicate percent estimates. To calculate the margin of error we will calculate 80 replicate percent estimates and then use the variance formula in the steps that we saw on the previous example.

So to calculate the first replicate percent we will have to calculate the replicate numerator and replicate denominator. So for the first replicate numerator we will sum up the same records we did for our numerator. However we will use the first replicate weight which is called PWGTP1 and circled in red. And if you do this you obtain first replicate numerator of 561.

Similarly we’ll do the same thing for the first replicate denominator. And there are no restrictions for a denominator so we add up all of the weights for our records. And when we do this we obtain a first replicate denominator of 628. If we wish to find a second replicate enumerator and denominator we’d use the second replicate weight called PWGTP2 which is not shown here. And recall there are 80 replicate weights and we’ll need to find 80 replicate numerators and 80 replicate denominators to find our 80 replicate percents.

So here’s the table we saw before. I put in the first replicate numerator and denominator circled in red. And then to find the first replicate percent you would divide the numerator by denominator, so 561/628 and obtain the first replicate percent estimate of 89.3.

As I said we’ll need to do this 80 times, once for each replicate weight. The second and 80th replicate numerators, denominators and percents are circled in red. Note that you should not round the percent or replicate percents until
the very end. And if you must do so, once you obtain the margin of error then round.

Once you have the percent and 80 replicate percent estimates use the steps at the bottom of the slides to calculate the variance and then the margin of error. And these are the same steps that we saw before.

So here are the steps we worked out. In the interest of time I will not walk through them. However if you do the steps we did before you would obtain a percent estimate of 85.4% plus or minus 26.1. So the margin of error is 26.1.

And as I said before the power of this method is that you can use this to calculate the margin of error for many different types of estimates. For example, for a ratio calculate your ratio estimate and then calculate 80 replicate ratios. And then use the steps before to calculate your variance and then obtain the margin of error.

So this concludes my presentation. I’m now going to hand it over to Charles to go over some reference slides and then we’ll open it to questions. So Charles.

Charles Gamble: Thank you Sirius. We’d like to provide you some references to help with the information provided to you today. On this slide you will find a link to the ACS Design and Methodology document. Please see Chapter 12 of this document for more information about the variance formula. Also below is the Fay and Train paper that discusses the replicate method that we used to calculate the margin of error. As a reminder the slides for this presentation are available in PDF form on our ACS website and our event section and under today’s event Calculating Margins of Error the ACS way. There you can find copies of the links.
On this slide is a link to the ACS handbook called Understanding and Using the ACS Data. This handbook goes more in depth about the margin of error. You may also be interested in other Handbooks. At the bottom of this slide is a link to the list of our handbooks.

Last, here are the links to the PUMS Training and PUMS Technical Documentation’s page.

Also, I want to remind you that there is a user’s group and online community specifically for users of ACS data. The purpose of the ACS data user’s group is to improve understanding of the value and utility of ACS data, as well as promote information sharing among data users about key ACS issues and applications. The user’s group website contains PowerPoint presentations and some video recordings from past data user conferences as well. Go to acsdatacommunity.prb.org to learn more including how to sign up to be one of the over 2,400 users in the ACS online community.

If you’re looking for further assistance with how to obtain or understand ACS data, our data dissemination specialists or DDS who are located within your region can provide you with assistance about Census Bureau data. These specialists usually provide help in English but sometimes in other languages as well as depending on the needs of their communities. Whether conducting one-on-one webinars with business start-ups or conducting large scale presentations at universities these specialists strive to put the public in touch with the data they need.

DDS provide a wide variety of assistance for free. So if you were interested in a specific type of training or presentation, please reach out to a specialist in your area using the contact information on the slide.
In closing I encourage you to connect with us directly. You can sign up for and manage alerts on the ACS via gov delivery. Also add yourselves to gov delivery if you want the slide from this presentation or any other presentations we provide. We will send out an email alert when materials are available. You can visit our website census.gov/acs or connect on the various social media platforms using the hashtag #acsdata. We also have an email to help support data users who may have questions - acso.users.support@census.gov.

And one last thing before we open the line for questions. If you are using ACS estimates make sure to source the Census Bureau’s American Community Survey as to where you receive the data. It helps people know that information they’re using is powered by the American Community Survey.

I will now open the line for questions. Operator at this time do we have any questions? Operator.

Coordinator: Hello yes thank you. We will now begin the question and answer session. If you would like to ask a question please press Star 1, unmute your phone and record your name. Your name is required to introduce your question. If you need to withdraw your question press Star 2. Again to ask a question please press Star 1. It will take a few moments for the questions to come through. Please standby.

(Question 1): Yes I wanted to know if the particular detailed table we’re looking for in the VRE documentation - the variant estimate data set, is there another way to get those estimates or is that information just not available.
Sirius Fuller: So currently only the ones that are listed in about a hundred tables are only publicly available. If you are interested in adding certain detailed tables you can send a request. But there’s no guarantee that we’ll add the tables.

(Question 1): Okay.

Sirius Fuller: But we’d love to hear from data users who are using the data what they would like to have.

(Question 1): Okay great, thank you.

Sirius Fuller: Thank you.

(Question 2): Hi. I have a question about the adjustment factor you mentioned that we multiply the 4 divided by 80. I understand the denominator being 80 but I think I missed the explanation why we use 4 in the numerator.

Sirius Fuller: Yes I did not explain it. The 4/80. So for anyone who didn’t - missed that, the 80 is because there are 80 replicate weights. And I believe the American Housing Survey uses 160 replicate weights. The 4 is there - in the hand waving method the answer is because of math. In order to create the replicate estimates you use replicate weights which we saw in the PUMS data.

And in order to create the replicate weights you use these things called replicate factors. And the replicate factors are created - you need to use this thing called a Hadamard matrix and it’s this orthogonal matrix. And one of the properties you end up having a 4 fall out of the math. And so the four propagates through and ends up in your variance formula. And if you look at the American Housing Survey documentation they have a 4/160 because they have the 160 replicate weights. I hope that answers your question.
(Question 2): Thank you.

Sirius Fuller: Thank you.

(Question 3): Hi yes I was wondering if you had any suggestions for programs that you prefer to use to unzip those gz files?

Sirius Fuller: We don’t have any recommendations. I know gzip should work or 7zip or if you’re using Unix you can use the gunzip command or I believe it’s just built in, currently, the current MAC products, you know, if you click on the file it should unzip. But yes there’s no recommended method to unzip the files.

(Question 3): Okay thanks.

Sirius Fuller: Thank you.

(Caller 4): Yes I’m looking for some published example of how from the row estimates you collapse or you compare a closed geography and adjust margin of error. Do you have any published documentation?

Sirius Fuller: So for - so these slides are really the first worked examples that are publicly available. There are worked examples in the Variance Replicate Estimate’s Technical Documentation that are similar to these slides. So if you want to see something, like, a percent estimate to calculate or some other things there are more examples in the Variance Replicate Estimate Technical Documentation.

And then if you’re using PUMS data there are - you can use the same method that I showed here. And then you can check if you get your standard error and
margin of error. If you’re using it correctly you should be able to match up to the published characteristics which are the PUMS Estimates for User Verification. And there’s about 30 or so, 40 characteristics. I hope that answers your question.

(Caller 4): Can I ask one more question to that?

Sirius Fuller: Sure.

(Caller 4): Yes because I read the Pew research and Urban Institute type of people, when they use ACS estimates they do not give what exactly their margin of error with regard to aggregates or comparison across geographies. What’s actually happening there?

Sirius Fuller: So I have not seen the data you’re talking about. Although I know about Pew. And it might be that they’re using the Variance Replicate Tables. But if not - if they’re using the tables that are published on data.census.gov then they would be approximating the margin of error. And if you go to the ACS Technical Documentation there are formulas there and a few worked examples. There’s some documentation there that shows you how to approximate the margin of error.

And the approximate margin of error for example for counts does not take into account the covariance. So as I mentioned if you have a lot of estimates your approximate margin of error may diverge from the actual margin of error. And the Variance Replicate Estimates Tables, you know, allows you to take the covariance into effect.

(Caller 4): Okay thank you very much, thank you, great.
Sirius Fuller: Thank you.

(Caller 5): Thank you. My question is about whether public disclosure procedures are going to be applied to the ACS data, yes or no?

Sirius Fuller: So yes for everybody else is wondering about this. There are - there’s a little bit of background information. The 2020 census data will have disclosure avoidance methods applied to it and it’s called Differential Privacy. The ACS is moving towards Differential Privacy and the last I heard I think it’s I think in the 2025 would be the earliest. But don’t take my word for it. I believe there’s some official I don’t know Charles do you know of anything besides this? So don’t take my word for it. We’re still working on it. The answer is yes we’re moving that way but the next year my understanding is it will not be applied to the ACS.

(Caller 5): Okay, thank you.

Sirius Fuller: Thank you.

(Caller 5): Can I have a follow up question?

Sirius Fuller: Sure.

(Caller 5): Never mind.

Sirius Fuller: Okay, well thank you for asking.

(Caller 5): Yes.
(Caller 6): Thank you. At the beginning of the presentation you mentioned you were going to talk about calculating MOE when the estimate is a zero. I was just wondering whether you wanted to talk about that.

Sirius Fuller: Yes thank you. So I didn’t mention it for time purposes but you will need to use these parameter files that are available on the VRE Technical Documentation page. And if you’re doing count estimates there’s a parameter called the k-value as well as another parameter file for the average weights. And in the technical documentation page it’ll show you the formula to use. I believe the average weights are also used to calculate estimates which are zero - percent estimates which are zero. And that’s also covered in the Technical Documentation.

(Caller 6): Okay thank you. I have a follow up question here too if I’m allowed.

Sirius Fuller: Yes.

(Caller 6): The ACS also publishes the design factors for people to use. Do you recommend the VRE method over the design factor method?

Sirius Fuller: Right. So for the PUMS data there’s also – there’s design factor methodology as well as using the replicate estimates in the PUMS would be the replicate weights. And so the design factors are model-based and they’re modeled onto the SDR variance, you know, the variance shown with the VREs. And so we would recommend that you use the variance replicate formula that were shown in these slides - the SDR variance is the official name over the design factor method which is a generalized variance function.

(Caller 6): Thank you.
Sirius Fuller: Thank you.

(Caller 7): Hello can you guys hear me all right?

Sirius Fuller: Yes go ahead.

(Caller 7): All right. I’ve been trying to combine data across geographies using R. And I’m wondering if you guys have any resources to help us do that and write the code for it?

Sirius Fuller: Unfortunately no. But I believe there are people who have done this in R and I think they might be on the data user group webpage. You could ask there. I know resources have been - are available and do exist in R but the Census Bureau doesn’t have anything unfortunately.

(Caller 7): Okay thank you.

Sirius Fuller: Thank you.

Coordinator: We show no further questions at this time. Again press 1 to ask a question. <PAUSE>. We show no further questions at this time.

Charles Gamble: Okay well thank you everybody for attending today. We look forward to working with you in the future and keep using ACS data. Thank you.

END