

An Introduction to Analyzing the NAWS Public Access Data



July 2020

Release 1.1

Contents

Overview on Conducting Analysis with the NAWSPAD	3
Software Options for Analyzing the NAWS Data	3
Calculate Weighted Means and Proportions	4
Calculate Design-Corrected Standard Errors	4
Additional Documentation to Support NAWSPAD Users	4
How to Calculate Weighted Means and Weighted Proportions using Excel	6
Download the NAWS Excel files	6
Apply Sampling Weights	6
Calculate Weighted Means and Proportions	6
Calculate Weighted Means	7
Calculate Weighted Means for All Fiscal Years	7
Calculate Weighted Means for Specific Fiscal Years	8
Calculate Weighted Proportions	10
Calculate Weighted Proportions for All Fiscal Years	10
Calculate Weighted Proportions for Specific Fiscal Years	16

Overview on Conducting Analysis with the NAWSPAD

This document provides an introduction to the NAWS Public Access Data (NAWSPAD), alerting the user to statistical issues related to the National Agricultural Workers Survey's (NAWS) complex sampling design and outlining available options for addressing them. The NAWS uses a stratified multi-stage sampling design to account for seasonal and regional fluctuations in the level of employment in crop agriculture. The NAWS has seven levels of sampling. The first two levels are to create 36 strata determined by three interviewing cycles and 12 agricultural regions. The next level consists of the primary sampling unit (PSU) within each stratum. The PSU is the county cluster (Farm Labor Area). The next four levels of sampling occur within the county cluster. Within each county cluster, the NAWS draws random samples of counties, ZIP Code regions, and employers. At the final level of sampling, workers are randomly sampled within employers. Further information on the NAWS sampling design is available in the *Statistical Methods of the National Agricultural Workers Survey* available on the NAWS website (<http://www.doleta.gov/agworker/naws.cfm>).

The NAWS's multi-level sampling design adds complexity to data analysis in two ways:

1. To produce accurate statistics, users need to apply sampling and post-sampling weights to account for differences in sampling probabilities and to correct for non-response at the region and cycle levels.
2. Users who wish to calculate the standard error of point estimates need to account for the complex multi-stage sampling design. The NAWSPAD supports using the balanced repeated replication (BRR) method. This method does not require information on the strata or PSU, which are suppressed from the NAWSPAD for privacy reasons.

Software Options for Analyzing the NAWS Data

The NAWSPAD files are available in SAS, Microsoft Excel, and CSV (comma separated values) formats. These files can be read into other data analysis programs, such as SPSS and Stata. NAWS users who do not have access to analysis software such as SAS, SPSS, R, or Stata, or users who only need to quickly calculate the mean or proportion of specific variables, can perform these calculations using the NAWS Excel files. In addition, there are several documents to support users, including codebooks, questionnaires, and explanation of the NAWS methodology. A list of these documents is provided below.

Calculate Weighted Means and Proportions

Users with access to statistical software such as SAS, SPSS, R, or Stata should consult the respective manuals for how to calculate weighted means and proportions using the weighting variable PWTYCRD. MS Excel users can calculate weighted means and proportions using the instructions in Chapter 1 “How to Calculate Weighted Means and Weighted Proportions from NAWS Public Access Data Using Excel.” This chapter provides step-by-step instructions on performing the calculations. The process can be tedious if analyzing many variables but does provide the same means and proportions produced by statistical software.

Calculate Design-Corrected Standard Errors

This document provides specific guidance and detailed explanations on how to calculate weighted means and weighted proportions using excel.

Additional Documentation to Support NAWSPAD Users

The following additional documentation about the NAWS and the NAWSPAD may be useful. These documents are available on the NAWS website (<http://www.doleta.gov/agworker/naws.cfm>).

- NAWSPAD Variables and Labels
- The NAWSPAD codebook
- Supporting statement [Part A of the Paperwork Reduction Act (PRA) Information Collection Request (ICR)]
- Statistical Methods of the National Agricultural Workers Survey (Part B of the PRA ICR)
- Field sampling protocols
- Questionnaires in English and Spanish
- For users interested in regional analysis:
 - Sampling regions
 - Correspondence between NAWS and USDA Farm Labor Survey sampling regions
 - Analysis regions in the NAWSPAD file

How to Calculate Weighted Means and Weighted Proportions using Excel

This chapter provides step-by-step instruction on calculating weighted means and proportions using Excel.

Step 1

Download the NAWS Excel files

From the “Public Data Files” link on the NAWS website (<http://www.doleta.gov/agworker/naws.cfm>), download the two files that are available in Excel format: NAWS_A2E185 (which contains the first half of the variables) and NAWS_F2Y185 (which contains the second half of the variables).

Step 2

Apply Sampling Weights

The NAWS uses a complex sampling design that includes both stratification and clustering; for this reason, users must make use of sampling weights to adjust the relative value of each farmworker so that population estimates may be obtained from the sample. The NAWS sampling weight variable PWTYCRD includes a factor which correctly proportions the data for analysis. The PWTYCRD variable is used for almost all NAWS analysis and allows merging several years of data together. At least two consecutive years of data should be combined to obtain robust. The PWTYCRD variable can be found in both the NAWS_A2E185 and NAWS_F2Y185 files. Further information on how the NAWS sampling weight is calculated is provided in *Statistical Methods of the National Agricultural Workers Survey*, available on the NAWS website (<http://www.doleta.gov/agworker/naws.cfm>).

Step 3

Calculate Weighted Means and Proportions

There are two types of variables in the NAWSPAD: continuous and categorical variables. Users can calculate weighted means for continuous variables using the instructions in Step 3a, and weighted proportions for categorical variables using the instructions in Step 3b. Examples of continuous variables include age of farmworker, wages, number of work days, and numbers of years in the United States. Categorical variables include those for which each farmworker respondent is assigned a value

indicating his/her membership in one of several possible categories. Examples of categorical variables include gender, employment status, and legal status. Users should consult the codebook for more information on the variables contained in the NAWSPAD.

Step 3a

Calculate Weighted Means

The mean, or average, of a variable is the sum of all values across all the observations divided by the number of observations. This sample mean assumes that each observation in the sample has an equal weight. However, when calculating the mean of a variable in the NAWSPAD, users need to apply the sampling weight variable PWTYCRD to adjust for the relative value of each farmworker in the sample, because each farmworker contributes differently to the final mean depending on his/her sampling weight.

The weighted mean can be calculated in Excel with a formula that uses both the SUMPRODUCT and SUM functions. For example, if users want to calculate the mean age of farmworkers then the SUMPRODUCT function is used to calculate the numerator in the formula, which is the sum of the products of each farmworker's age and his/her sampling weight (PWTYCRD). The SUM function is used to calculate the denominator in the formula, which is the sum of all farmworker sampling weights (PWTYCRD).

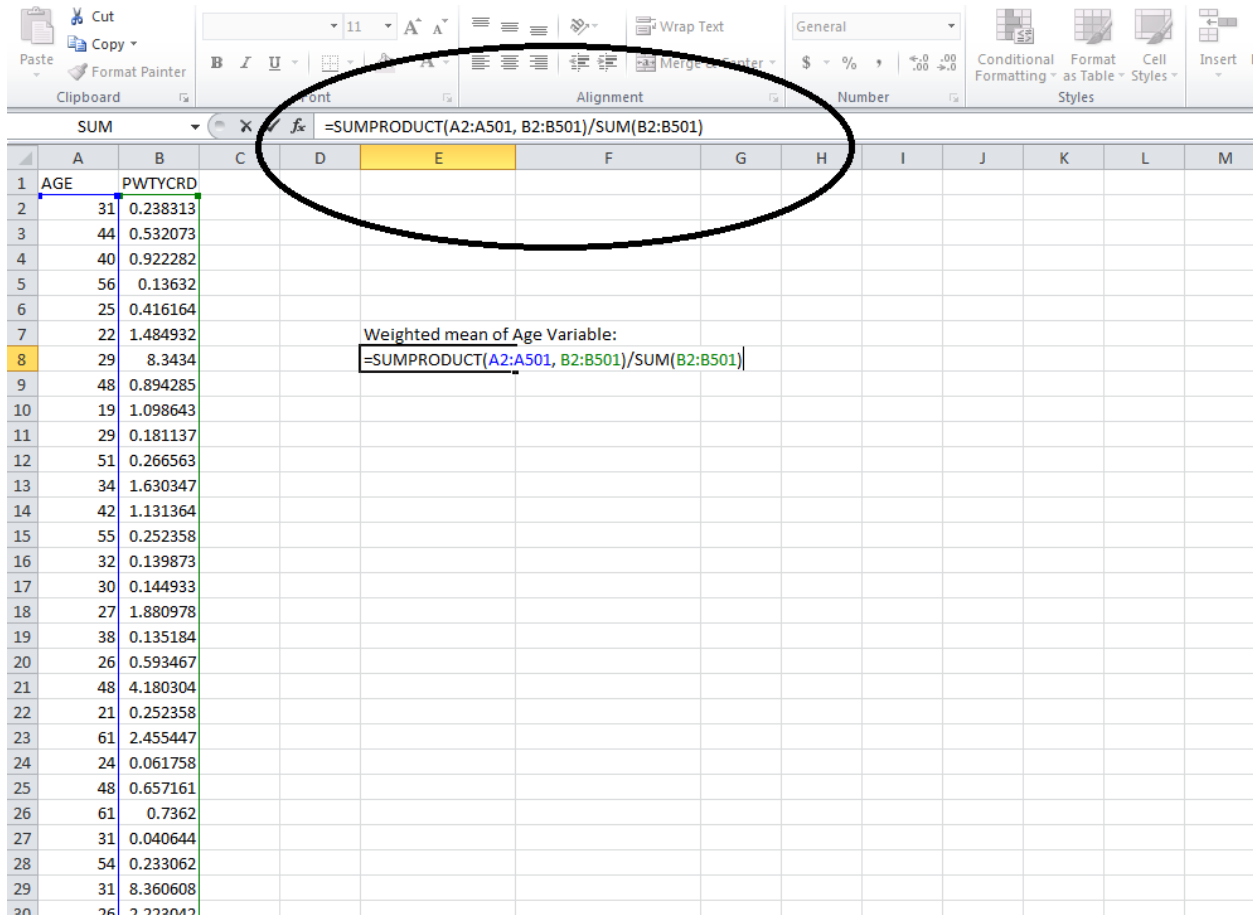
Calculate Weighted Means for All Fiscal Years

The following instructions show how to calculate the weighted mean of a continuous variable for all fiscal years in the dataset combined in Excel. For the purpose of illustration, we will use AGE as the variable of interest and we assume that the Excel data set contains 500 farm workers interviewed in all fiscal years.

1. Determine the variable(s) for which you would like to calculate the weighted mean(s). In our example, we want to calculate mean AGE.
2. Locate the variable of interest and the PWTYCRD variable in the NAWS Excel file. For our example, the variable AGE is located in column A and the variable PWTYCRD is located in column B (see Figure 1). There are 500 rows in our example dataset, one for each farmworker interviewed. The first row contains the variable names.
3. The Excel formula for calculating a weighted mean is `SUMPRODUCT(array1,array2,array3,...)/SUM(number1, [number2], [number3], ...)`. Using our example, the formula to calculate farmworkers' mean age is `SUMPRODUCT(A2:A501, B2:B501)/SUM(B2:B501)`.
 - a. `SUMPRODUCT(A2:A501, B2:B501)` calculates the numerator, which sums the products of each farmworker's age (AGE) and his/her sampling weight (PWTYCRD).

- b. $\text{SUM}(B2:B501)$ calculates the denominator, which is the sum of all 500 farmworker sampling weights.
- c. The full formula $\text{SUMPRODUCT}(A2:A501, B2:B501)/\text{SUM}(B2:B501)$ calculates the weighted mean age for the 500 farmworkers in the data set. Figure 1 shows an illustration of the Excel data set with the formula written in it.

Figure 1: Formula for Weighted Mean



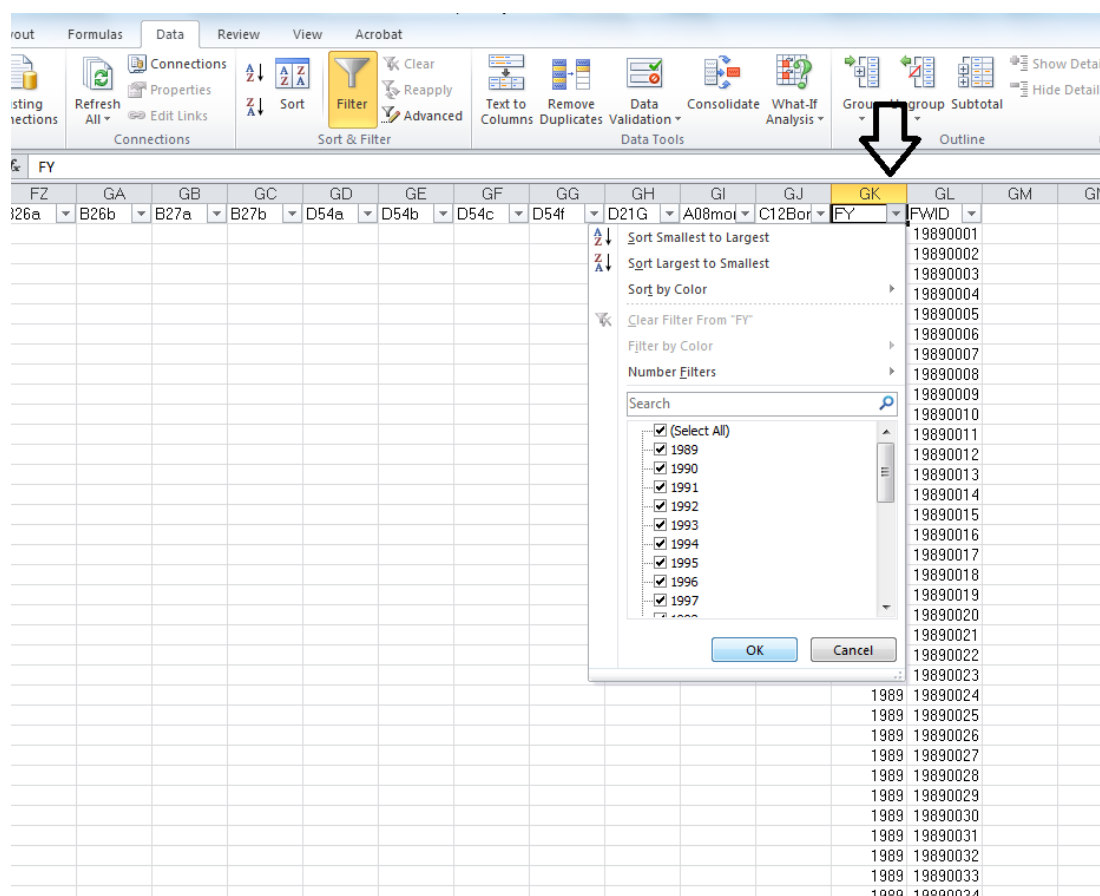
Calculate Weighted Means for Specific Fiscal Years

The previous step described how to calculate weighted means for a continuous variable using all fiscal years in the dataset. It is also possible to calculate weighted means for specific years using the Fiscal Year (FY) variable. To ensure that your results are as accurate as possible, you need to combine at least two consecutive fiscal years of data (e.g., 2015 and 2016). For more information about combining fiscal years, please consult the *Statistical Methods of the National Agricultural Workers Survey*, available on the NAWS website (<http://www.doleta.gov/agworker/naws.cfm>), which describes the statistical methods of the NAWS.

There are two options to create a worksheet containing data for specific fiscal years.

Option 1: Use the Filter option in Excel Before entering any formulas into your worksheet, select all cells in the worksheet, go to the “Data” tab on the menu bar at the top of the worksheet (see Figure 2), and click the “Filter” button. A small drop-down arrow will appear next to each variable name. Click on the drop-down arrow next to the “FY” variable name (see Figure 2). This will open a dialogue box and at the bottom you will see a list of all the fiscal years available for selection (see Figure 2). Select the fiscal years you are interested in and click “OK”. You can pick as many fiscal years as you want, but you should pick at least two consecutive years. Now that you have applied the filter, you can calculate weighted means using the formula described in the “Calculate Weighted Means for All Fiscal Years” section above, steps 1 through 3.

Figure 2: Apply a Filter to "FY"



Option 2: Copy and Paste Before you do any calculations, copy the column headings for all the variables you need (i.e., the variable of interest, PWTYCRD, and FY) and the rows corresponding to the fiscal years you are interested in. Paste them into a new worksheet. Make sure you copy all the rows included in the fiscal years you have chosen; missing even one row can change your results. In the new worksheet, calculate

weighted means using the formula described in the “Calculate Weighted Means for All Fiscal Years” section above, steps 1 through 3.

For more help with calculating weighted means in Excel, please consult the “How to calculate weighted averages in Excel” page on the Microsoft website (<http://support.microsoft.com/kb/214049>).

Step 3b

Calculate Weighted Proportions

When calculating weighted proportions for categorical variables, users must apply the sampling weight variable PWTYCRD to adjust for the relative value of each farmworker in the sample (see Step 2 Apply Sampling Weights at the beginning of this document). The most efficient way to calculate weighted proportions in Excel is to first organize the data using PivotTables. The following instructions use PivotTables to help you easily calculate weighted proportions for all fiscal years combined or for specific fiscal years.

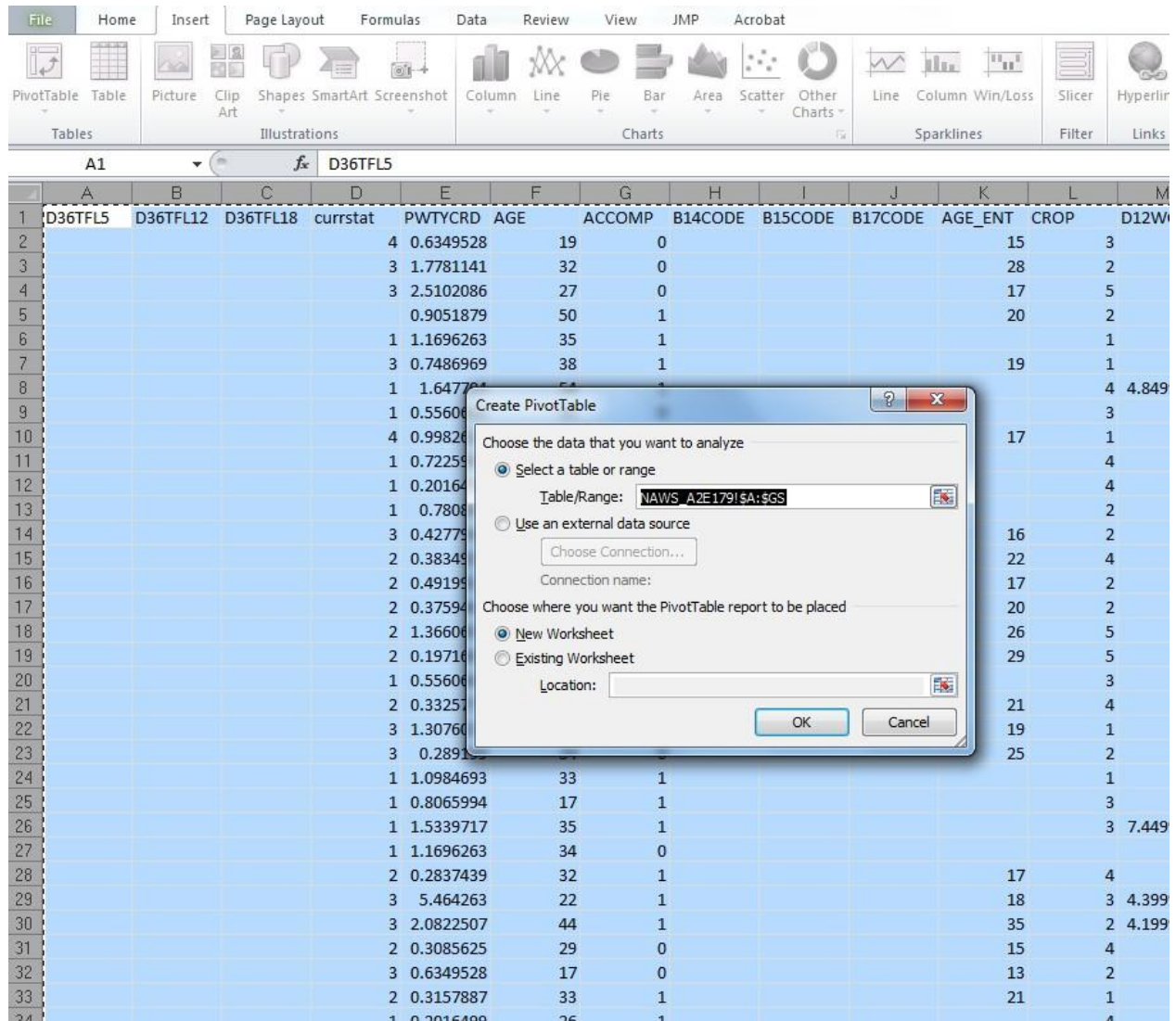
Calculate Weighted Proportions for All Fiscal Years

The instructions below describe the method for calculating weighted proportions for all fiscal years in the dataset. For the purpose of illustration, we will use the variable CURRSTAT (the farmworker’s legal status at the time of interview), which has four categories coded with the numeric values 1 through 4. The calculation described here will yield the weighted proportion for each of the four categories of CURRSTAT.

To calculate weighted proportions for all fiscal years combined:

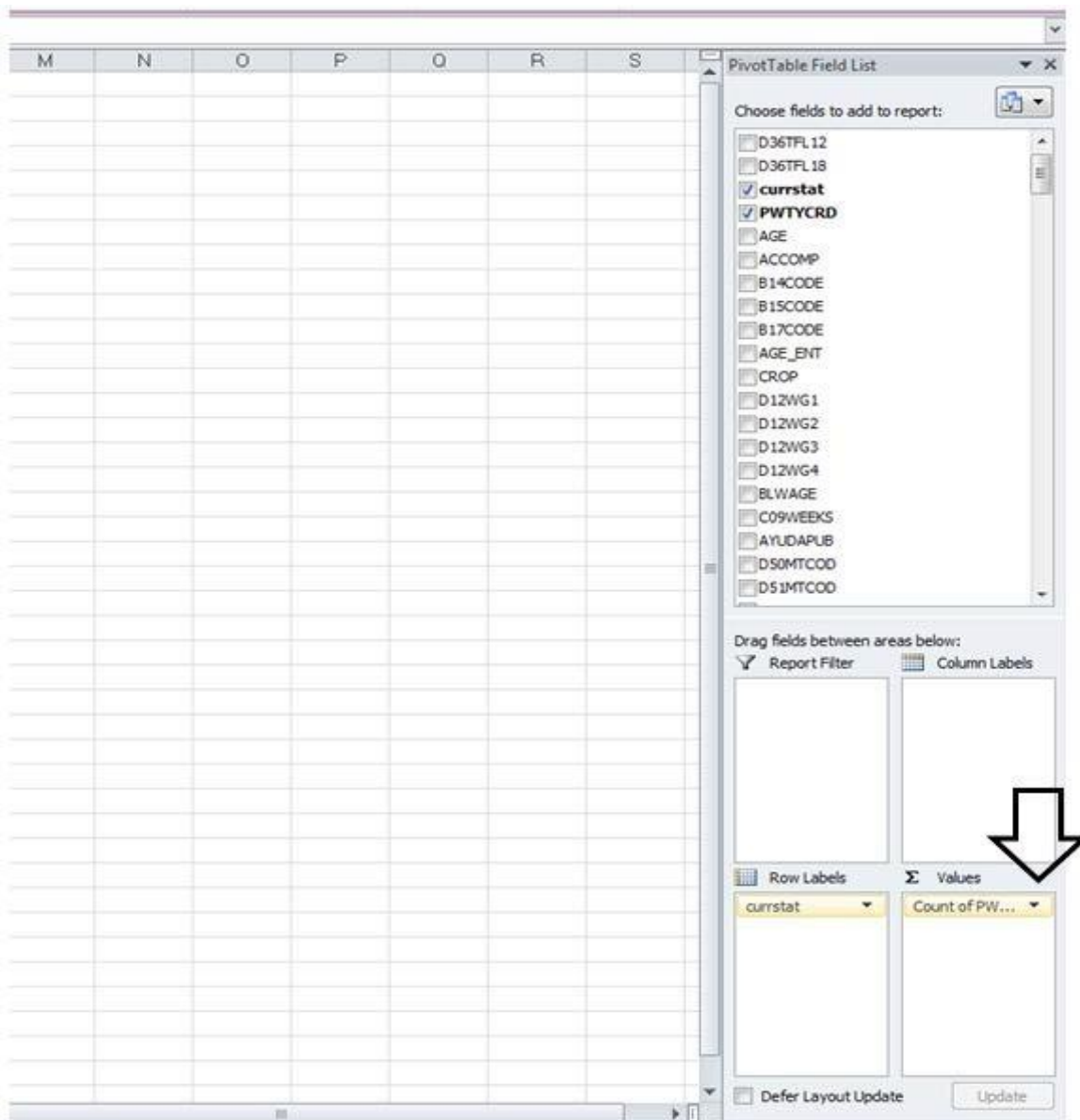
1. Determine the variable(s) for which you would like to calculate the weighted proportions. In our example, we want to calculate the weighted proportions of CURRSTAT.
2. Highlight all the cells in your spreadsheet. Go to the “Insert” tab on the menu bar at the top of the worksheet and select “PivotTable” (see Figure 3). The “Create PivotTable” dialogue box will appear, and two options will automatically be selected: “Select a table or range” (which shows the range of all the cells selected in our spreadsheet) and “New Worksheet” (see Figure 3). Click “OK”.

Figure 3: Create a PivotTable



3. After clicking “OK”, a new worksheet will open. On the left side of the new worksheet you will see an empty PivotTable. On the right side you will see a PivotTable Field List from which you can select the variables that you want to add to your PivotTable (see Figure 4). Using our example of CURRSTAT as the variable of interest:
 - a. Drag CURRSTAT and drop it into the “Row Labels” area.
 - b. Drag PWTYCRD and drop it into the “Values” area; you will see “Count of PWTYCRD” with a drop-down arrow next to it (see Figure 4).

Figure 4: PivotTable Field List



- c. Click on the drop-down arrow next to "Count of PWTYCRD" and click "Value Field Settings". In the Value Field Settings dialogue box, select the "Sum" option under "Summarize value field by" (see Figure 5).

Figure 5: Value Field Settings for PWTYCRD

The screenshot shows the 'Value Field Settings' dialog box in Microsoft Excel. The dialog box has the following fields and options:

- Source Name: PWTYCRD
- Custom Name: Sum of PWTYCRD
- Summarize Values By: Show Values As
- Summarize value field by: Choose the type of calculation that you want to use to summarize data from the selected field. The list includes Sum, Count, Average, Max, Min, and Product. 'Sum' is selected.
- Number Format: (empty)
- Buttons: OK, Cancel

The background shows a PivotTable with the following data:

currstat	Total
1	14978
2	14707
3	3882
4	26786
(blank)	858
Grand Total	61211

- d. Click “OK” to create your PivotTable. The left column in the table contains the four categories of CURRSTAT. The right column contains the sum of the PWTYCRD values for each CURRSTAT category (see Figure 6); this is the column that you will use to calculate your weighted proportions.

Figure 6: Created PivotTable

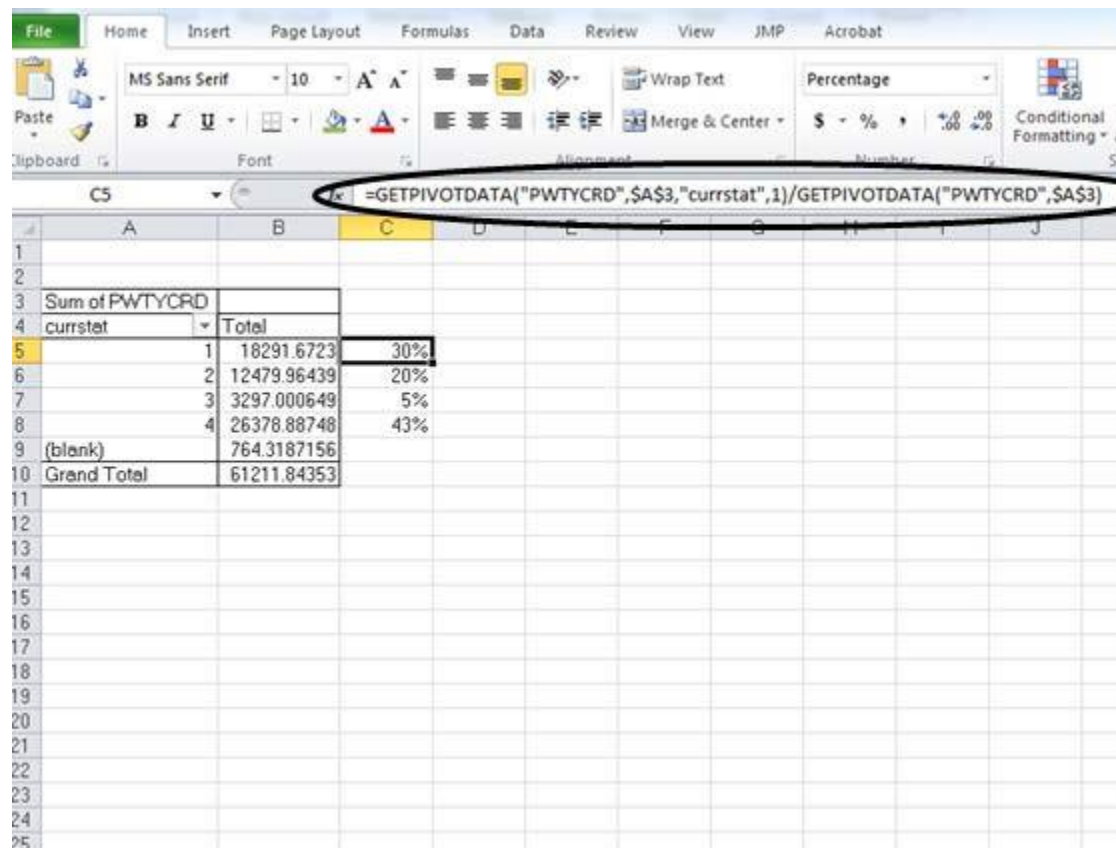
Drop Report Filter Fields Here							
Sum of PWTYCRD							
currstat	Total						
1	18291.6723						
2	12479.96439						
3	3297.000649						
4	26378.88748						
(blank)	764.3187156						
Grand Total	61211.84353						

Note: The columns in your PivotTable may have headings different from those pictured based on the version of Excel you are using. The values in the right column will likely differ from those pictured, depending on the fiscal years of data you are analyzing.

4. Now you are ready to calculate the weighted proportions of the variable (i.e., how often each category of the variable occurs). For each category of the variable (i.e., row in the PivotTable):
 - a. Create a formula that divides the value in the right column of the PivotTable by the “Grand Total” (see Figure 7).
 - b. Convert the resulting values from decimals to percentages (see Figure 7).
 Note: Because you are calculating weighted proportions based on cells that were created using a PivotTable, you cannot use the fill handle to apply the formula for the first category of the variable to the other

categories of the variable. If you want to be able to use the fill handle option, copy the PivotTable and paste it elsewhere in your worksheet using “Paste Special” with the “Values” option. This will remove all embedded formulas and provide you with a static table that you can use for your calculations.

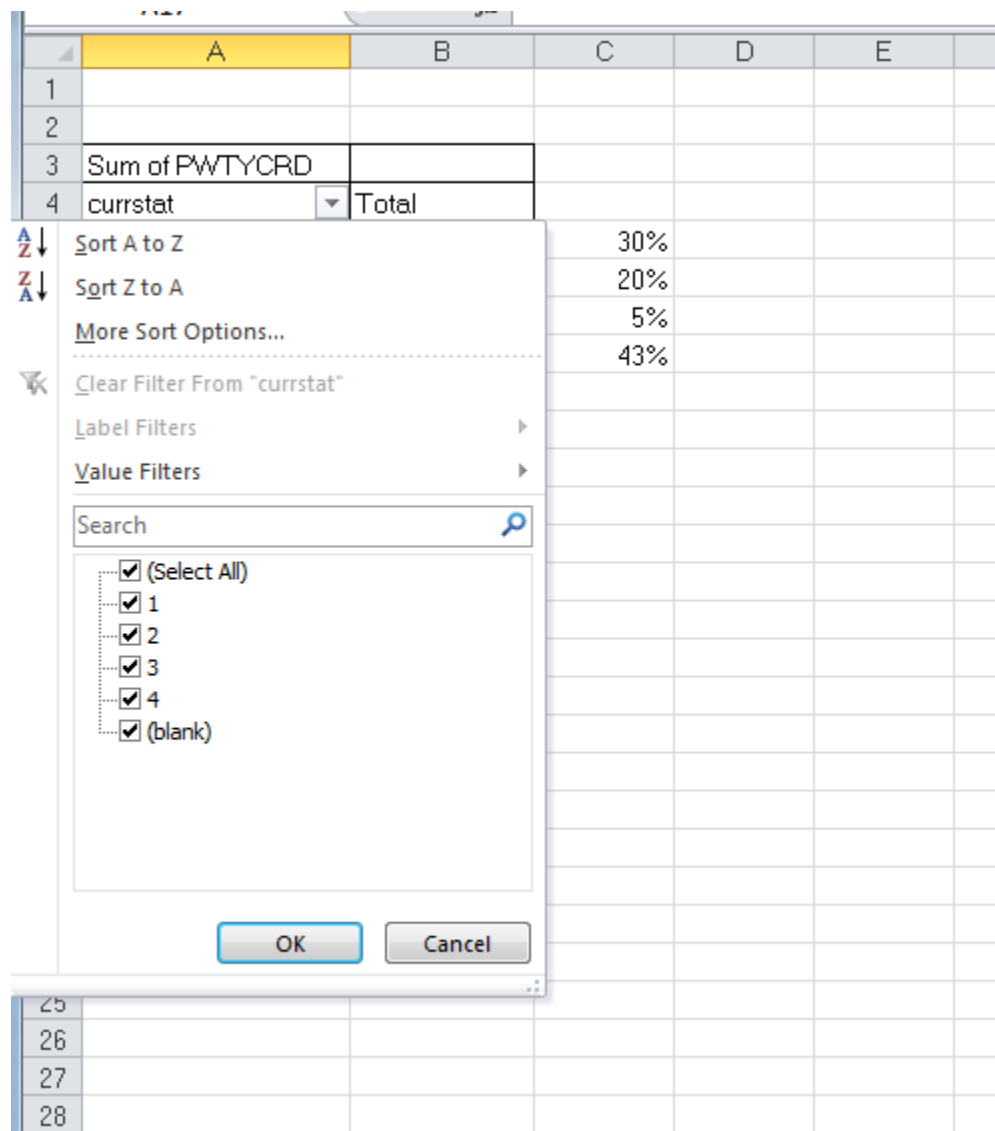
Figure 7: Formula for Weighted Proportion



Note that because your formula for calculating weighted proportion uses cells in a PivotTable, you cannot use the fill handle to apply the formula for the first category of the variable to the other categories of the variable. If you want to be able to use the fill handle option, copy the PivotTable and paste it elsewhere in your worksheet using “Paste Special” with the “Values” option. This will remove all embedded formulas and provide you with a static table that you can use for your calculations.

The category labeled “(blank)” represents the respondents who are missing a value for the variable. If left in the PivotTable, the category of missing values will comprise a share of respondents. If you want to exclude missing values from your analysis, you can filter them out by clicking the down arrow next to CURRSTAT, unchecking the box next to “(blank)” in the filter menu, then clicking “OK” (see Figure 8).

Figure 8: Filter “(blank)” from Created PivotTable



For additional help creating PivotTables in Excel, please consult the “Create or delete a PivotTable or PivotChart report” page on the Microsoft website (<https://support.office.com/en-us/article/Create-or-delete-a-PivotTable-or-PivotChart-report-d09e4d07-8cd6-4b60-afad-8fb67418800f?CorrelationId=213dd195-0deb-4f22-8ec8-e80eca1c2fd0&ui=en-US&rs=en-US&ad=US>).

Calculate Weighted Proportions for Specific Fiscal Years

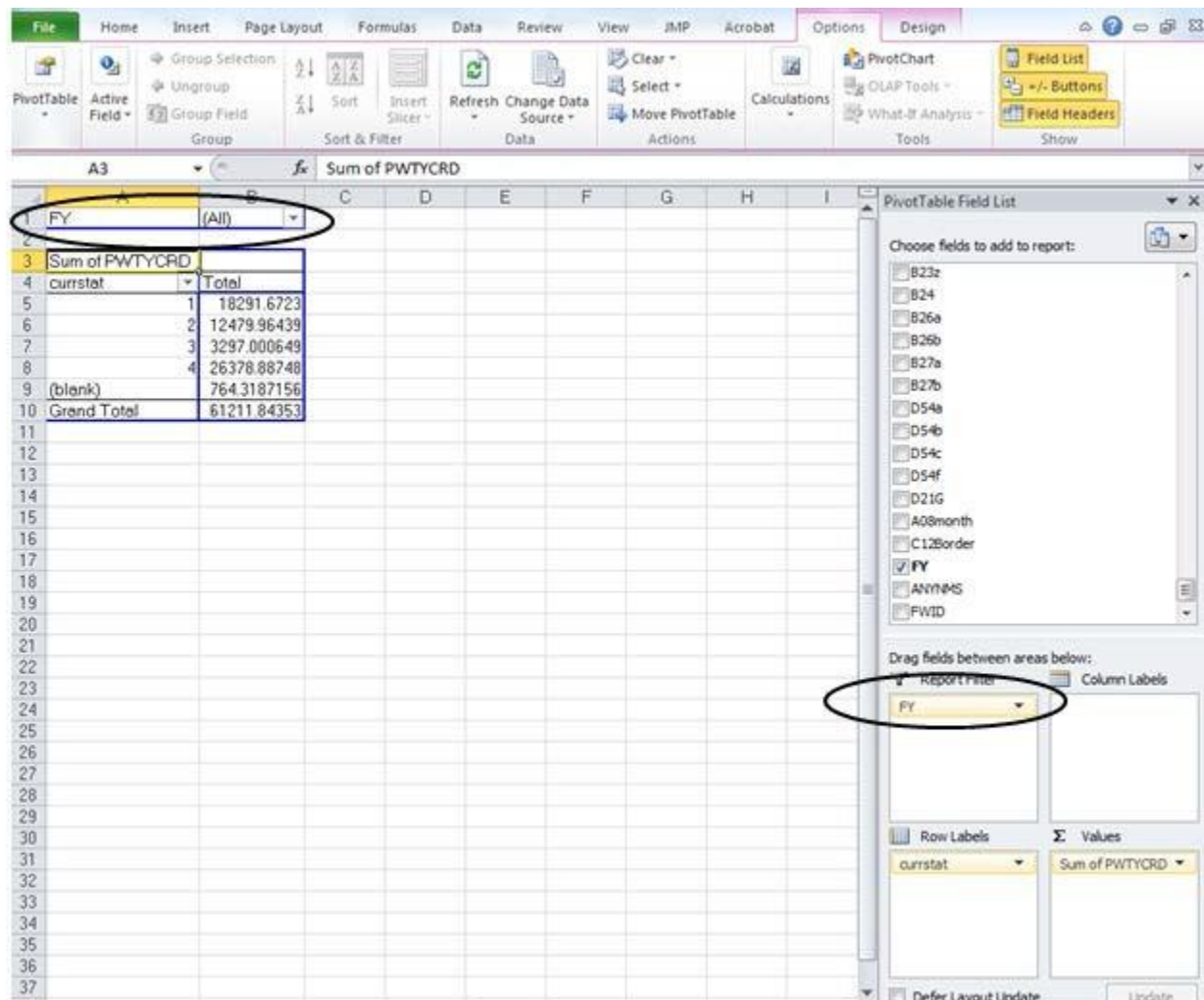
The process for calculating weighted proportions for specific fiscal years is very similar to the process for calculating weighted proportions for all fiscal years. To ensure that the data results are as accurate as possible, you need to combine at least two consecutive fiscal years of data (e.g., 2015 and 2016). For more information about combining fiscal

years, please consult the *Statistical Methods of the National Agricultural Workers Survey* which describes the statistical methods of the NAWS.

There are two options for calculating weighted proportions for specific fiscal years.

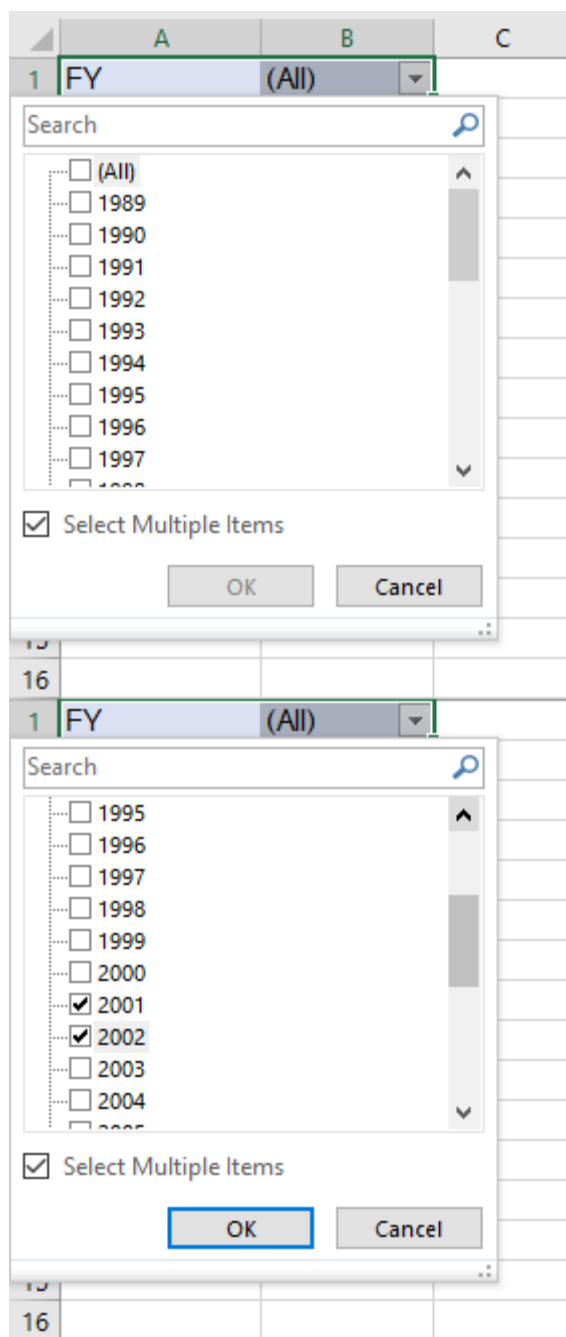
Option 1: Use the Report Filter option in a PivotTable Create a PivotTable by following steps 1 through 3 in the “Calculate Weighted Proportions for All Fiscal Years” section above. After you have changed the “Value Field Settings” for PWTYCRD to “Sum of PWTYCRD” as described in step 3, return to the PivotTable Field List on the right side of the screen, locate the variable “FY”, and drag and drop it into the “Report Filter” area below the list. Once you have done this, the variable “FY” will appear in the “Report Filter” section above your created PivotTable (see Figure 9).

Figure 9: Apply “FY” to the Report Filter



The next step is to select the fiscal years you are interested in. Click the down arrow next to (All) and check the “Select Multiple Items” box. A check box will appear to the left of each Fiscal Year in the list; click in the check box to the left of (All) to deselect all of them. Locate the years that you are interested in, check the box for each (we use 2001-2002, for the purpose of illustration), then click “OK” (see Figure 10).

Figure 10: Select Specific Fiscal Years



Once the report filter is applied, the sum of the PWTYCRD values for each category of CURRSTAT and the Grand Total for the table will recalculate to reflect the weighted frequencies for only those fiscal years which you have selected. The final step is to calculate the weighted proportions according to the procedure described in step 4 of the “Calculate Weighted Proportions for All Fiscal Years” section above.

Option 2: Copy and Paste Before you do any of the steps involved in creating a PivotTable, copy the column headings for all the variables you need (i.e., the variable of

interest, PWTYCRD, and FY) and the rows corresponding to the fiscal years you are interested in. Paste them into a new worksheet. Make sure you copy all the rows included in the fiscal years you have chosen; missing even one row can change your results. In the new worksheet, create a PivotTable and calculate weighted proportions according to the process described in steps 1 through 4 of the “Calculate Weighted Proportions for All Fiscal Years” section above.

For additional help creating PivotTables in Excel, please consult the “Create or delete a PivotTable or PivotChart report” page on the Microsoft website

(<https://support.office.com/en-us/article/Create-or-delete-a-PivotTable-or-PivotChart-report-d09e4d07-8cd6-4b60-afad-8fb67418800f?CorrelationId=213dd195-0deb-4f22-8ec8-e80eca1c2fd0&ui=en-US&rs=en-US&ad=US>)

