



## **Addendum to the WesVar<sup>®</sup> User's Guide (New Features in WesVar<sup>®</sup> 5.1)**

WesVar<sup>®</sup> 5.1 is the latest version of WesVar, Westat's program for the analysis of data collected using complex sampling designs. WesVar 5.1 adds the following features not available in previous versions:

- Descriptive Statistics;
- Batch preparation of WesVar formatted data files, including:
  - Importing files with replicate weights, and
  - Importing, calculating replicate weights, and adjusting replicate weights;
- Batch preparation of table and regression requests; and
- Additional data import and export options

Each of these features is described below. All of the features in WesVar 5.1 are described in this document and the WesVar User's Guide.

### **Descriptive Statistics**

WesVar 5.1 calculates the following descriptive statistics for numerical variables:

- Five largest and five smallest values;
- Five most influential values when calculating the mean and weighted sum;
- Number of nonmissing values;
- Number of missing values;
- The following percentiles: 1, 5, 10, 25, 50, 75, 90, 95, and 99;
- Mean;
- Geometric mean;
- Total;
- Variance;
- CV;
- Skewness; and
- Kurtosis.

When working in a WesVar Workbook, use the following steps to specify the descriptive statistics to be calculated:

- 1) Select the workbook node at the top of the left panel.
- 2) Click the new request "Descriptive Stats" button on the right panel. An additional descriptive request will be added to the bottom of the workbook.

- 3) Specify the variables to be analyzed using the following steps:
  - a) Select the “Analysis Variables” node in the descriptive request.
  - b) In the right panel, specify the variables to be analyzed by moving the names of the variables from the left “Source Variables” box to the right “Selected” box. The variables can be moved between boxes by either 1) dragging the names, 2) double clicking on the names, 3) selecting names and clicking on the “>” or “<” buttons, or clicking on the “>>” or “<<” buttons (to move all names at once). Multiple names can be selected by using a shift-click.
  
- 4) Specify transformed variables to be analyzed. Transformations of the variables in the data file can be specified. The summary statistics for the transformed variable will be presented; however, the transformed variable will not be saved in the data file. Transformations can be defined using the following steps:
  - a) Select the “Transformed Variables” node in the descriptive request, in the left panel.
  - b) Enter an equation into the “Transformed Statistic:” edit box in the right panel. The equation starts with a name for the transformed variable, followed by an equal sign and the formula for calculating the transformed variable. For example, the equation might be:  $x = y + z$ , where y and z are the names of numeric variables in the data file. The equation can be typed into the edit box. Alternatively, the variable names, operators, and functions that WesVar uses can be inserted into the edit box at the cursor position by double clicking on the variable names in the “Source Variables:” list, clicking on the buttons for arithmetic operations, and selecting function names from the function drop-down list.
  - c) Once the equation is entered, click the “Add as New Entry” button to add the equation to the list of transformed variables to be analyzed. If a previously defined transformed variable is selected in the left panel, clicking the “Replace Current Entry” button will replace the selected transformed variable with the variable defined in the edit box.
  
- 5) Specify pairs of variables for which correlations will be calculated, using the following steps:
  - a) Select the “Correlations” node in the descriptive request in the left panel.
  - b) Moving variable names from the “Source Variables” list to the “List 1” and “List 2” lists. WesVar calculates correlations for all combinations of variables from list 1 and variables from list 2. Variable names can be moved between lists in the same ways as used when selecting the analysis variables (item 3 above). The radio button beside “List 1” and “List 2” indicates which list box is used when moving names using a double click or using the buttons.
  
- 6) The statistics are calculated within By groups, if any, defined in the “By” node in the descriptive request, in the left panel. The “By” variables are specified by moving variables from the “Source Variables:” list to the “Selected:” list.

As with other requests, the calculations are not performed until clicking on “Run Workbook Requests” button or the “Run Selected Request” button. These tool bar buttons have green triangles and are labeled when the cursor stops over the button. After the statistics are calculated, the results can be viewed by clicking on the “View Output” tool bar button (it looks like an open book).

The algorithms for calculating the summary statistics are described in Appendix A at the end of this addendum.

## **Batch File Preparation**

For calculating parameters and variance estimates, WesVar uses data files in its own format. To analyze data in other formats, the files must first be converted into WesVar format. Converting the files to WesVar format requires reading the data into WesVar, designating the weight variable and any replicate weights on the imported file, specifying the replicate method, and saving the data in WesVar format. WesVar can then be used to create replicate weights and adjust the weights using poststratification or raking. In previous versions of WesVar, these steps were specified manually for each file to be processed. In WesVar 5.1, the processing steps can be specified once and modified or executed multiple times on the same or different files. This can be particularly helpful if the data files need to be modified or additional variables are added to the data.

WesVar provides two methods to import data and create a WesVar data file. The first is used when the replicate weights are on the imported files and do not need to be modified. Multiple files can be processed at a time using the first method. The second method allows creation and modification of replicate weights. Each of these methods of creating a WesVar file is described below.

### **Batch Processing: Importing Files and Replicate Weights**

From the opening WesVar window, select the “Simple Import” item from the “Run Batch” submenu on the Utility menu. A window will open allowing the user to specify the imported file name, the name of the output WesVar data file, the name of the variable with the base (or full sample) weight, the prefix used for the name of the replicate weight variables, and the file name of the file, if any, with JK<sub>n</sub> factors. Each row in the window is used for processing a different data file. Once the specifications have been entered, they can be executed immediately or saved in an ASCII file and executed later. To save the specifications, select “Save” from the file menu. To read in specifications that were previously saved, select “Open” from the file menu, and locate the text file. To execute the specifications and create the WesVar files, click on the “Run” button (the tool bar button with the yellow lightning bolt). To view a summary of the processing and any errors, click on the “View Summary” button.

If there are errors when typing in the file and variable names, the WesVar files may not be created or may not be created correctly. To help minimize errors when entering the file names, place the cursor in the cell in which the file name is to be placed and select the “Browse” menu. Then, navigate to the file. Selecting the file in the navigation window places the corresponding file name in the selected cell.

When specifying a file with JK<sub>n</sub> factors, the order of the factors in the file is important. See the WesVar manual for additional details.

Note that after creating the data file, it may still be necessary to adjust the degrees of freedom associated with the data file.

Imported files using this method must be SAS or SPSS. This method cannot be used to create replicate weights, adjust replicate weights for nonresponse, or adjust weights using poststratification or raking. For these purposes, see the next section.

### **Batch Processing: Importing Files and Creating and Modifying Replicate Weights**

When performing all tasks in the data file window, including:

- Creating replicate weights;
- Nonresponse adjustment;
- Poststratification and raking;
- Subsetting the file;
- Attaching factors;
- Changing the default degrees of freedom associated with the data file;
- Specifying labels; and
- Recoding variables,

WesVar modifies a log file recording the changes to the data file. The log file is located in the same directory as the data file and has the same name but with the extension .log instead of .var.

If the user modifies the original data file from which the WesVar file was created and wishes to recreate the WesVar file using the same steps as before, he or she should follow these steps:

1. From the opening WesVar window, select the “All Functions” item from the “Run Batch” submenu on the Utility menu.
2. A navigation window will open; click on the “Open” button to locate and select the .log file that was previously created when the original WesVar file was created. Any text file can also be used if the contents correspond to those from a WesVar .log file.
3. Click on the “Run” icon. The previous steps will be repeated using the new source data file.

It is also possible to modify an existing .log file to change how the file is processed. If the user is familiar with the layout of the .log file, he/she can easily create/edit one using a text editor. Another way to modify an existing .log file is to edit it directly in the navigation window after reading in the file. Once the specifications have been entered, they can be executed immediately or saved and executed later. To save the specifications, click on the “Save” icon. The modified file should look like a WesVar .log file. The appropriate format can be determined by performing the desired operations on any file and looking at the .log file to see how that operation is described.

### **Batch Tables and Regression**

WesVar 5.1 provides the ability to define, using a text file, a list of tables and/or regression models to be created. When the text file is read into WesVar, a workbook is created. The workbook can then be edited, if necessary, and run to obtain the desired statistics. The text file describes the requests and tables or models that go into the workbook. Each line begins with a directive within square brackets. The directive defines the effect of the line and any additional parameters on the line. The following is a simple example to calculate two tables:

```
[workbook]
[request]
[infile]   c:\test\dataset.var
[table]
[var]      age, 4
[var]      race, 3
```

```
[end table]
[table]
[var]    age,4
[var]    sex,2
[end table]
[end request]
[end workbook]
```

A detailed description of the format of the text file can be found in Appendix B at the end of this addendum and in WesVar at the opening screen by selecting “Workbook Batch File Help” on the Help menu.

The text file can be created using any text editor or using another program. WesVar assumes the extension on the text file will be .wvt. However, any other extension, such as .txt, can be used.

The following steps describe how to use the batch file for creating workbooks:

- 1) Prepare the text file.
- 2) In WesVar, with the opening screen active, select the File:Open:WesVar Workbook menu item. If the text file does not have the \*.wvt extension, then select the “All files (\*.\*)” option in the “Files of type” drop-down list.
- 3) Navigate to and select the text file from step 1 and click “Open”. A new workbook will be created with the specified request. If some part of a request is not specified correctly, the output may not be created.
- 4) The requests can be corrected, if necessary, or modified before running.

### **Additional Data Import and Export Options**

WesVar 5.1 can export data to all of the formats that can be used for importing data. After a file is created, select “Export” from the file menu to save the data. Select the format for the data to be exported from the “Save as Type” drop-down list, specify the file name and location, and click on the “Save” button.

## Appendix A: Descriptive Statistics Request

Statistic	Description
Number of nonmissing values (denoted $n$ in formulas below)	If “Exclude all cases with missing values” is checked, $n$ is the number of cases where all analysis variables are not missing (and the weights are positive). If “Exclude all cases with missing values” is <u>not</u> checked, $n$ is the number of cases where the analysis variable <u>associated with the descriptive request</u> is not missing (and the weights are positive). Weighted $n$ is the sum of weights of the cases enumerated by unweighted $n$ .
Number of missing values	If “Exclude all cases with missing values” is <u>not</u> checked, Unweighted Missing is the number of missing values for the analysis variable associate with the descriptive request, and Weighted Missing is the sum of the weights for those cases. If “Exclude all cases with missing values” is checked, then the Missing row will have “N/A” in the Unweighted column.
Five largest and five smallest values	Case identifiers and full sample weights are also displayed.
Five most influential values when calculating the mean and weighted sum	Case identifiers and full sample weights also displayed. A value’s influence for calculating the weighted sum is $w_i y_i$ , with the five most influential values having the five largest values of $ w_i y_i $ . A value’s influence for calculating the mean is $w_i (y_i - \bar{y}_w)$ , where $\bar{y}_w = \sum_{i=1}^n w_i y_i / \sum_{i=1}^n w_i$ , with the five most influential values having the 5 largest values of $ w_i (y_i - \bar{y}_w) $ .
Percentiles: 1, 5, 10, 25, 50, 75, 90, 95, and 99	Weighted estimates of percentiles are calculated by inverse interpolation of the estimated cumulative distribution function. Interpolated values that would be less than the minimum of the sample are set equal to the minimum. Woodruff’s method is the default methods for estimating standard errors of weighted percentiles; however, the user can request that replication-based standard errors be calculated instead. Unweighted estimates of percentiles are calculated by setting $w_i=1$ .
Mean	
Geometric mean	This is not calculated if any of the data values are zero or negative
Total	

Statistic	Description
Variance	<p data-bbox="496 306 646 338"><u>Unweighted:</u></p> $s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$ $= \frac{1}{n-1} \left[ \sum_{i=1}^n x_i^2 - \frac{1}{n} \left( \sum_{i=1}^n x_i \right)^2 \right]$ <p data-bbox="496 653 618 684"><u>Weighted:</u></p> $\hat{\sigma}^2 = \frac{\sum_{i=1}^n w_i (x_i - \bar{x}_w)^2}{\sum_{i=1}^n w_i}$
Coefficient of variation (CV)	<p data-bbox="496 947 646 978"><u>Unweighted:</u></p> $CV = \frac{\sqrt{\text{unweighted population variance}}}{\text{unweighted mean}}$ <p data-bbox="496 1110 618 1142"><u>Weighted:</u></p> $CV = \frac{\sqrt{\text{weighted population variance}}}{\text{weighted mean}}$

Statistic	Description
Skewness	<p data-bbox="483 306 638 338"><u>Unweighted:</u></p> $k_3 = \frac{n}{(n-1)(n-2)} \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{(s^2)^{3/2}}$ <p data-bbox="483 625 610 657"><u>Weighted:</u></p> $\kappa_3 = \frac{1}{\hat{\sigma}^3} \frac{\sum_{i=1}^n w_i (x_i - \bar{x}_w)^3}{\sum_{i=1}^n w_i}$
Kurtosis	<p data-bbox="483 942 638 974"><u>Unweighted:</u></p> $k_4 = \frac{(n+1)n}{(n-1)(n-2)(n-3)} \frac{\sum_{i=1}^n (x_i - \bar{x})^4}{(s^2)^2} - 3 \frac{(n-1)^2}{(n-2)(n-3)}$ <p data-bbox="483 1289 610 1320"><u>Weighted:</u></p> $\kappa_4 = \frac{1}{\hat{\sigma}^4} \frac{\sum_{i=1}^n w_i (x_i - \bar{x}_w)^4}{\sum_{i=1}^n w_i} - 3$

## Appendix B: File Format for the Batch Workbook Request

### General rules

- Spacing does not matter.
- Capitalization does not matter.
- Lines where the first non-blank character is not “[“ will be ignored. So, the line “[table]” will be ignored.
- Syntax errors will not be reported but may cause a feature to be omitted. For example, cell functions that use illegal cell definitions (see the cell definition explanation below) will cause the functions to be omitted.

### Directives

WesVar can create a workbook from one or more text files containing text strings called *directives*. Directives control the processing of the text files, or they specify parameters a user would provide via an interactive screen. Parameters are documented in the WesVar User’s Manual. Parameters not specified by directives are set to default values, which may not be the same values set in Preferences. Be sure to review the workbook in WesVar after opening a batch file to avoid unexpected results.

### Shell

Five directives have corresponding “end” directives: [workbook], [request], [table], [regression], and [model]. For example, the directive [table] must be followed at some point by the directive [end table]. All directives between [table] and [end table] apply to that table only. Here is a shell for the text file:

```
[workbook]
    general directives that apply to the entire workbook

    [request]
        directives that apply to the table request

            [table]
                directives that apply to the table
            [end table]

        [end request]

    [regression]
        directives that apply to the regression request

            [model]
                directives that apply to the regression model
            [end model]

        [end regression]

[end workbook]
```

Note that tables and regression models can be included in the same workbook.

### [include] Directive

The [include] directive includes text from a specified text file into the current text file. For example, assume the text file REQFILE.TXT specifies two tables and associated parameters and the file FUNCFILE.TXT defines cell functions that will be used in multiple requests.

REQFILE.TXT might look something like:

```
[workbook]
  [request]
  [infile]  c:\test\dataset.var
  [alpha]   0.05

      [table]
      [var]   age,4
      [var]   race,3
      [include] funcfile.txt
  [end table]

      [table]
      [var]   age,4
      [var]   sex,2
      [include] funcfile.txt
  [end table]

  [end request]
[end workbook]
```

FUNCFILE.TXT might look something like:

```
[function] F1X=(R01C01-R02C01) / R03C03 * 100.0
[function] F2X=(R02C02-R03C02) / R03C03 * 100.0
. . .
[function] F9X=(R04C04-R05C05) / R99C99 * 100.0
```

### Conventions for specifying table cells and table variables

For a one-way table, cells are named C00, C01,...,C99 where C00 designates a missing cell and C99 designates a total cell. For a two-way table, cells are named R00C00, R00C01,...,R99C99 where the same numbering scheme applies. Cells in a three-way table are named as PxxRxxCxx. The specification for a table variable consists of the variable name followed by a comma followed by the number of nonmissing, nontotal levels.

### List of Directives

#### General Workbook Directives

```
[workbook] . . .
[end workbook]
```

```
[include] c:\file.txt
```

Include lines from another file at this point (can be placed anywhere in text file)

[infile] c:\file.var	Specify name of WesVar data file (can also be placed within a request)
<u>Table Directives</u>	
[request] . . . [end request]	Designates a table request. The parameter is ignored by WesVar; it can be used as a comment.
[subset] x < 3 [alpha] n [fpc] n [df] n [complete] on/off	Subset criteria Alpha level Finite population correction Degrees of freedom (can be infinite) Complete option (exclude all cases with missing value), see WesVar documentation
[estimate] on/off [stderror] on/off [standard] on/off [wilson] on/off [cv] on/off [n] on/off [deff] on/off [eff_samp] on/off [prob>t] on/off	Output estimate Output standard error Output standard confidence interval Output Wilson confidence interval Output coefficient of variation Output unweighted sample size Output design effect Output effective sample size Output t value and prob >  t
[summary] on/off [flog] on/off [width] n [dec1] n [dec2] n [dec3] n [outest] on/off	Used to show/suppress summary info at top of request Show cells and cell function definitions in summary info How wide numeric fields will be on output Decimal places in the mantissa, scientific notation Decimal places in the estimate, fixed notation Decimal places in the standard error, fixed notation Create an auxiliary output file that contains the replicate estimates
[avar] x [compute] x = y + z	Analysis variable (request level) Compute statistic (request level)
[table] . . . [end table]	Designates a table specification. The parameter is ignored by WesVar; it can be used as a comment.
[var] var, levels [missing] on/off [rs2] on/off [rs3] on/off [f_value] on/off [f_pcnt] on/off [f_rowp] on/off [f_colp] on/off [v_value] on/off [v_pcnt] on/off	Variable to be included in table and number of levels Include missing cells RS2 option for Chi-squared statistic RS3 option for Chi-squared statistic Sum of weights, value Sum of weights, percent Sum of weights, row percent Sum of weights, column percent Analysis variables, value Analysis variables, percent

[v_rowp] on/off	Analysis variables, row percent
[v_colp] on/off	Analysis variables, column percent
[function] x = y + z	Cell function specification
<b><u>Regression Directives</u></b>	
[regression] . . .	Designates a regression request. The parameter is ignored by WesVar; it can be used as a comment.
[end regression]	
[type] linear/logistic/multinomial	Type of model
[alpha] n	Alpha level
[maxiter] n	Maximum number of iterations (logistic models)
[ci] on/off	Confidence interval in output
[correlation matrix] on/off	Correlation matrix in output
[covariance matrix] on/off	Covariance matrix in output
[ftest] on/off	
[summary] on/off	Used to show/suppress summary info at top of request
[width] n	How wide numeric fields will be on output
[dec1] n	Decimal places in the mantissa, scientific notation
[dec2] n	Decimal places in the estimate, fixed notation
[dec3] n	Decimal places in the standard error, fixed notation
[model] ...	Designates a regression model specification. The parameter is ignored by WesVar; it can be used as a comment.
[end model]	
[definition] x = y[2] z	Specifies regression model (class variables must have number of levels in brackets)
[intercept] on/off	Intercept in model
[standardized coefficients] on/off	Standardized coefficients in model
[df] n	Degrees of freedom (can be infinite)
[by] x, y	By variables for model
[test] a: x = 0	Hypothesis test (label: test)

### Text-file example

```
[workbook]
  [request]
  [infile]      c:\abc\yourfile.var
  [avar]       x
  [avar]       y
  [compute]    c1=q*r
  [compute]    c2=q/r
  [subset]     panel < 4
  [summary]    on
  [outest]     off
  [complete]   on
  [width]     12
  [df]        0
```

```
[fpc]          1.0
[alpha]        0.05
  [table]      THIS IS SEX BY AGE TABLE
  [var]        sex,2
  [var]        age,4
  [f_value]    on
  [f_pcnt]     on
  [f_rowp]     on
  [f_colp]     on
  [rs2]        off
  [rs3]        off
  [v_value]    on
  [v_pcnt]     off
  [v_rowp]     on
  [v_colp]     off
  [missing]    on
  [function]   F1=R01C01*R01C01
  [include]    c:\abc\funcs.txt
  [end table]
  [end request]
[end workbook]
```