

What's New in Wcalc5?

Update: March 14, 2010

This document assumes that readers are familiar with the older version of Wcalc (4) and the web document, Wcalc.pdf.

1 New Features

Easier Input Formats Older versions of Wcalc supported only text input and required a considerable amount of file preparation. Version 5 will now read Stata .dta files and worksheet (.wk1) files (default). The principal advantages of the new binary formats is that they considerably reduce the possibility of errors that often prevented older versions from running, and (2) they eliminate the file preparation step by reading archive datasets in their native formats.

Whereas the .txt format required exactly four items in a specified order, the binary formats can read any number of columns, while using only those four that are logically required. A new selection logic allows users to either designate the required columns by naming them at the file preparation stage or by designating each of them at run time if standard names are not used. Because columns are selected by name, they can be in any order.

The older .txt format is still supported. The one minor change is that a header line is now tolerated (but not required).

Topic Query If you have numeric codes for topics or subtopics in your dataset, you may query to select cases for analysis at run time. This again eliminates file preparation by allowing datasets to contain items not used in a particular analysis.

Automatic Sorting The logic of Wcalc requires that input data be sorted by variables and then by dates (ascending) within variables. Version 5 automatically checks the input sort order and does an auto resort if required.

Bug Fixes Known bugs in version 4 have been fixed. These include printing issues in output files and a failure, under some circumstances, to read the last record in a file.

The documentation files, formerly named .rpt is now .log, a more standard usage.

2 Binary Input

ASCII text input is an extremely limiting format. To read correctly it requires that the user prepare a file containing exactly four items per line in the required order. This process is both tedious and error prone.

The logic of binary input is to allow any number of columns (.wk1) or variables (.dta) to be read, from which the four to be used are selected at run time. Rather than specify the order of input variables, as in the text version, for binary files the header (in worksheets) or variable name (in dta files) defines contents. The program looks for “variable” or “varname” to define the column of variable names. A variable (or column) named “date” defines the date item. “Value” or “index” define the column of numeric values, and “n” or “ncases” defines the column with sample size information. Header names are not case sensitive, so varname = VARNAME = Varname all point to the same column.

If any of the four standard headers is not found at run time, the program provides a numbered list of headers in the file and prompts the user to select the appropriate ones.

3 Topic Query

The query facility permits selecting a subset of a data file defined by content at run time. (This operates independently from selection by date, which has always been part of Wcalc. Thus one may select by topic, by date, by both, or by neither.) This permits selection from data files containing materials

not wanted for a particular analysis.

Query assumes that a file contains a column of content codes. And so the task of the user is specify which column that is and then what the selection criterion is.

Example: mycode=13 (where mycode is a column header in the file). The query prompt is below:

Case Selection:

Enter a selection criterion of the form VARNAME OPERATOR VALUE,
where (1) VARNAME is one of the column names above,
(2) OPERATOR is one these 6: <, >, <=, >=, =, or <>
(3) VALUE is a number.

Example: topic<10

Conditions may also be joined with the boolean operators AND and OR as in:
topic=1 and subtopic<3

Specifications are not case sensitive.

Specify a condition for selection

Query is not available for .txt file input.

4 Auto Sorting

The logic of Wcalc requires that records be sorted by variable name, and within that, by date (ascending). (Variables can be in any order.) In older versions of Wcalc that required that the user perform the sort as part of file preparation. Failure to sort correctly produced nonsense analyses. On input Wcalc5 detects the actual sort ordering of the file. If it does not comply with the dates within variables requisite, the file is automatic resorted to be in

the correct order. Variables are sorted in alphabetical order and dates within variables are ascending.

Beware that variable names are case sensitive in Wcalc. Thus two names which are identical except for capitalization are treated as two different variables for analysis.