

FS4

**First Stage Stratification and Selection in
Sampling**

User Manual

Version 2.0

Author:
Raffaella Cianchetta (Istat)

Index

Introduction	3
1 Main features.....	4
2 Installation and loading.....	5
3 FS4 Menu	6
3.1 File menu	7
3.2 Data menu	8
3.3 Functions menu	10
3.4 Tools menu	15
3.5 Results menu	16
3.6 Help menu	17
References.....	18

Introduction

FS4 package is a generalized software for first stage stratification and selection in sampling related to two or more stages with a GUI (Graphical User Interface) based on the tcltk package.

The main window of the GUI consists of a menu bar and, from the top to the bottom, of two scrollable text windows called "Commands Window" and "Warnings Window". The menu bar consists, from left to right, of the following menu items: "File", "Data", "Functions", "Tools", "Results" and "Help" (see FS4 menu "tree" p. 6).

Menus "Data" and "Functions" lead to dialog boxes that need further specifications, while the others ("File", "Tools", "Results" and "Help") allow directly to have the results of commands.

File import and data frame export, in ".txt" and ".csv" formats, are available through "Data" menu. Display and removal of output data frames and of those belonging to user's workspaces, are possible through specific functions of "Tools" menu. With "Results" menu is possible display only all the data frames outputs.

The main form is "Stratification and Selection": it allows user to choose two data frames stored into the user's workspace (GlobalEnv): PSU (Primary Sampling Unit) population and allocation data frames. The choice of each of them, displays the list of variables that forms it, some of which must be selected to populate the form. The input parameters required are represented by the number of sample PSUs to select in each stratum and by the single or multiple¹ size of the planned final sample units to observe in each PSU.

It is possible to launch the procedure in two steps, (in this case the launch is "partial" and radio button is setted on "Yes" by default), so allowing to the user to view the output result and, on the basis of this, modify the input parameters and launch newly the process. With the exception of the "partial launch", the required output fields are four. They refer to the names that the user must assign to the output data frames made available: the first two at domain level, the third and fourth respectively at stratum and PSU level. The first output displays the list of the Self Representative (SR) PSUs and of the Non Self Representative (NSR) PSUs before the stratification of the whole NSR PSUs. In the second output is added the sample size of the final units distinct for SR and NSR PSUs, after the stratification. The third output shows, for each stratum, the number of sample PSUs and the PSUs total number. The fourth output provides, for each PSU, some information such as, for example, the probability of inclusion and the sampling fraction.

FS4 defines, for each PSU, the final sample units size, by the sizes planned in input.

By default, commands generated by the GUI are inserted to the Commands Window (is possible to save commands via File -> Save Commands or copy them by pressing Control-C or right-clicking the mouse). Warning messages are displayed in the Warnings Window, while error messages, producing the stop of computation, are always displayed by means of pop-up dialog boxes.

In addition to "Help" menu, for each dialog box coming from "Data" and "Functions" menus is available a *Function Help*. The end of a GUI session is possible either using the File -> Exit menu or by closing the FS4 main window.

¹ To import by the allocation data frame.

1 Main features

The main features of FS4 are the following:

- it is a flexible software allowing to the user to:
 - choose the measure of size to define stratification and inclusion probability for PPS (Probability Proportional to Size) sampling;
 - insert as input a single or plural (variable between domains) planned size of final sample units to observe in each PSU;
 - launch the procedure in two separate steps, so to observe a first output and on the basis of this modify the input parameters;
- it is a open source package released under EUPL²;
- it has a user friendly GUI that allows also to non-expert R users to use it, although for them is possible to use directly the *StratSel* function from R command line.

*StratSel*³ is the main function of the project: it merges two data frames (whatever PSU population and allocation data frames) and then compute stratification and selection of a fixed number of sample PSUs for stratum using the Sampford's method (unequal probabilities, without replacement, fixed sample size), implemented by the *UPsampford* function of R package *sampling*. About the stratification, the function implements for each estimation domain, the computation of a size threshold for a given size PSU. PSUs with measure of size that exceeds a calculated threshold are identify like SR and each constitutes a stratum by itself, so they come into the sample with probability equal to one. The remaining NSR PSUs are ordered for the measure of size and divided into stratum having size approximately constant to the corrected threshold and with PSUs having sizes as homogeneous as possible.

² European Union Public Licence.

³ For more details see “FS4 2.0 Reference Manual” (pp. 5-7).

2 Installation and loading

FS4 package needs installation of R version 3.0.2 or newer and the following R packages: tcltk2, svMisc, plyr and sampling.

To install this package in R is sufficient, after its download, to launch a R interactive session and use the `install.packages()` function

```
> install.packages(file paths, repos = NULL)
```

where “file paths” is a string of ‘.zip’ or ‘tar.gz’ file’s path.

To load package there are two possible ways:

- a) choice from R “Packages” menu the option “Load package...”;
- b) with the alternative instructions:
 - `require(FS4)`;
 - `library(FS4)`.

FS4(Fig. 1) can properly run under the Rgui in the SDI (single-document interface) mode.

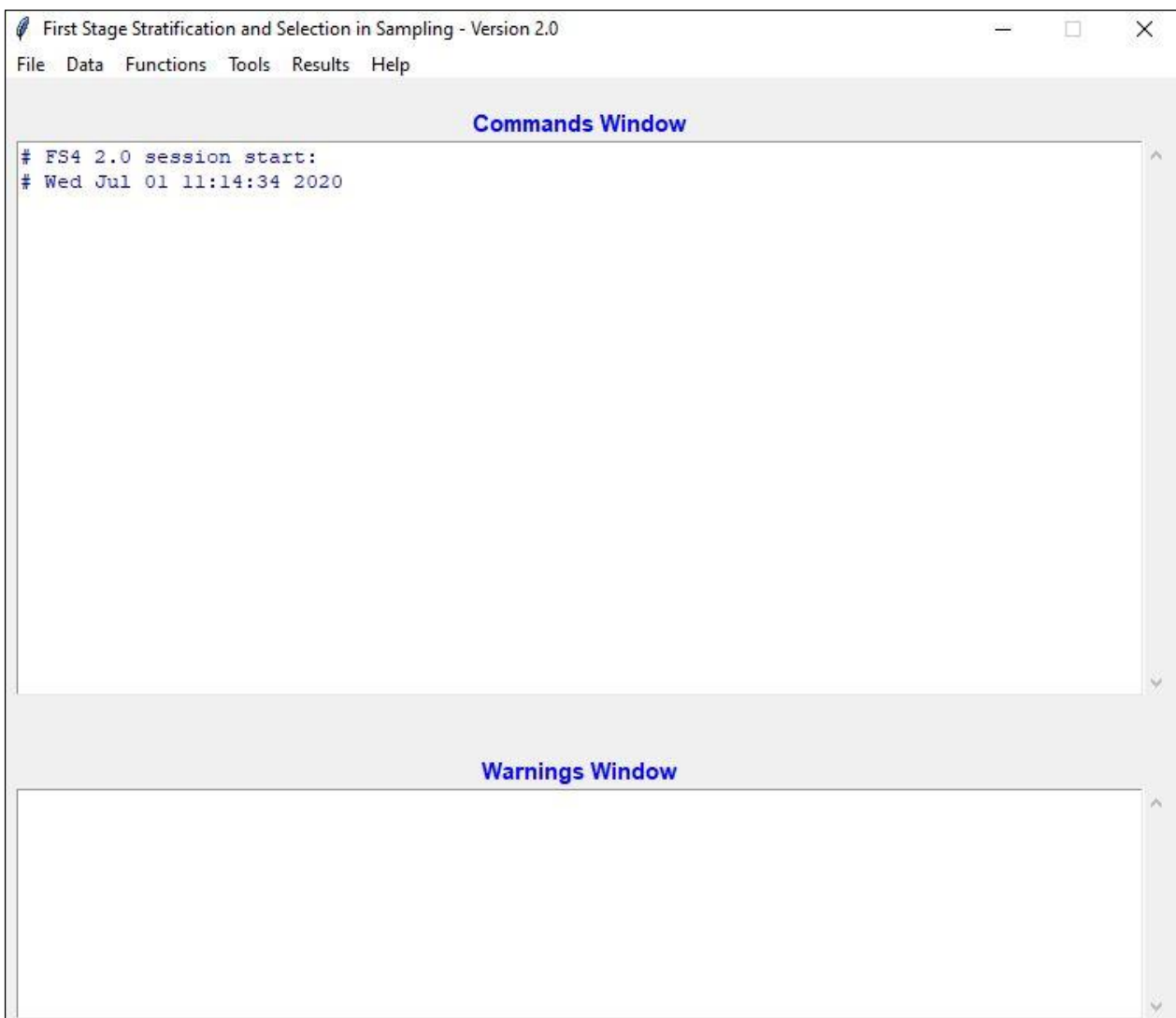


Figure 1: FS4 window at start-up

3 FS4 Menu

The menu "tree" for this FS4 package (version 2.0) is shown below:

| File

-
- | - Change Directory
 - | - Load Workspace
 - | - Save Workspace
 - | - Save Commands
 - | - Exit

| Data

-
- | - Import Datasets
 - | - Export Datasets
- | - From Text Files

| Functions

-
- | - Stratification and Selection

| Tools

-
- | - Show Datasets
 - | - Remove datasets

| Results

-
- | - Show results

| Help

-
- | - R Help on FS4
 - | - FS4 2.0 Reference Manual
 - | - FS4 2.0 User Guide

3.1 File menu

Menu items (Fig. 2) to:

- change directory. It's possible to change the work directory and/or create a new one;
- load workspace. The system can load R workspace files (.RData or .rda extensions).
- save current workspace;
- save commands resulting from “Command Windows”;
- exit. Before exit, there is the possibility to save the current workspace and/or the Commands History.

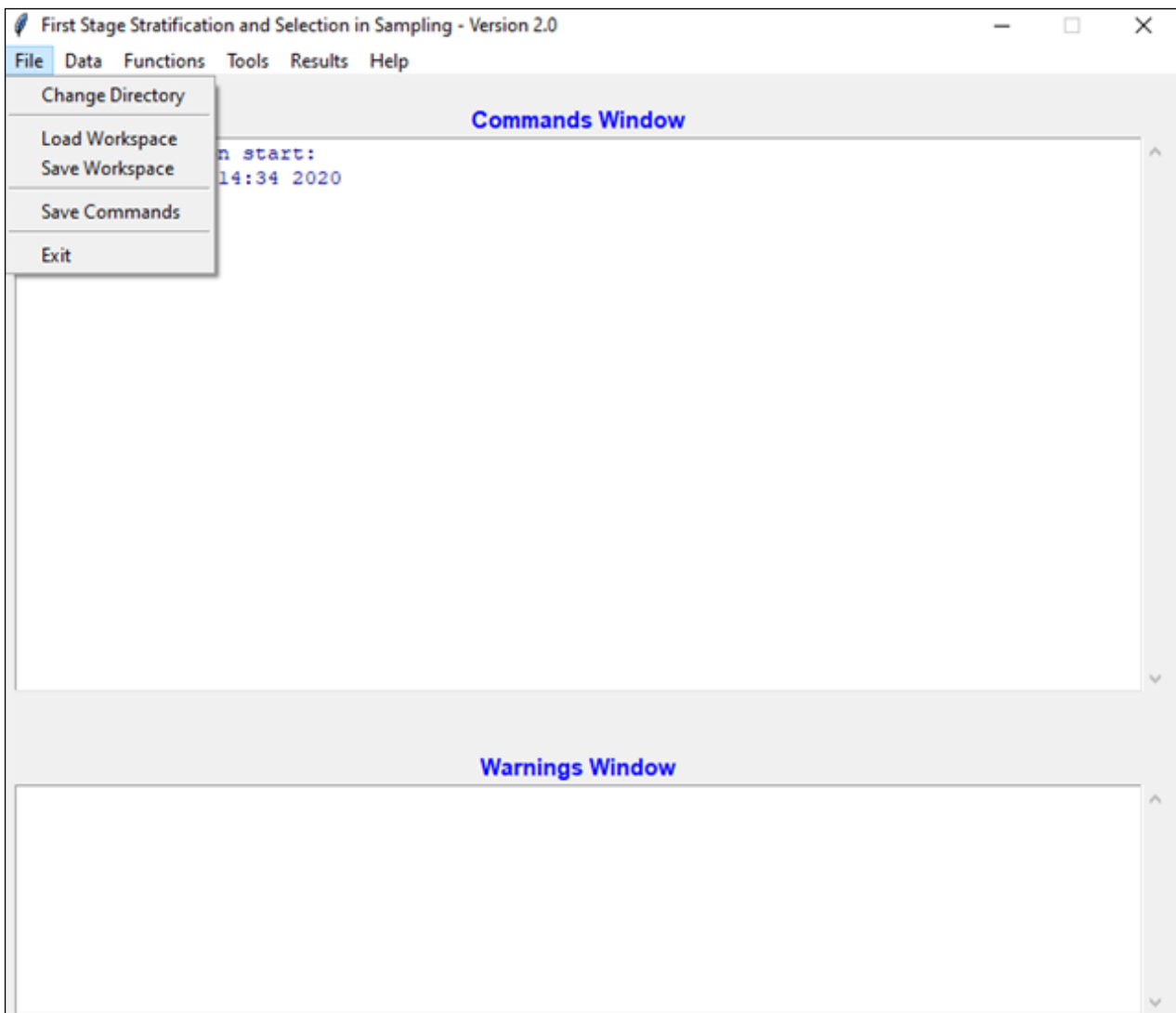


Figure 2: File menu

3.2 Data menu

Menu item to export and to import datasets from Text Files (.txt, .csv and .dat extensions) (Fig. 3).

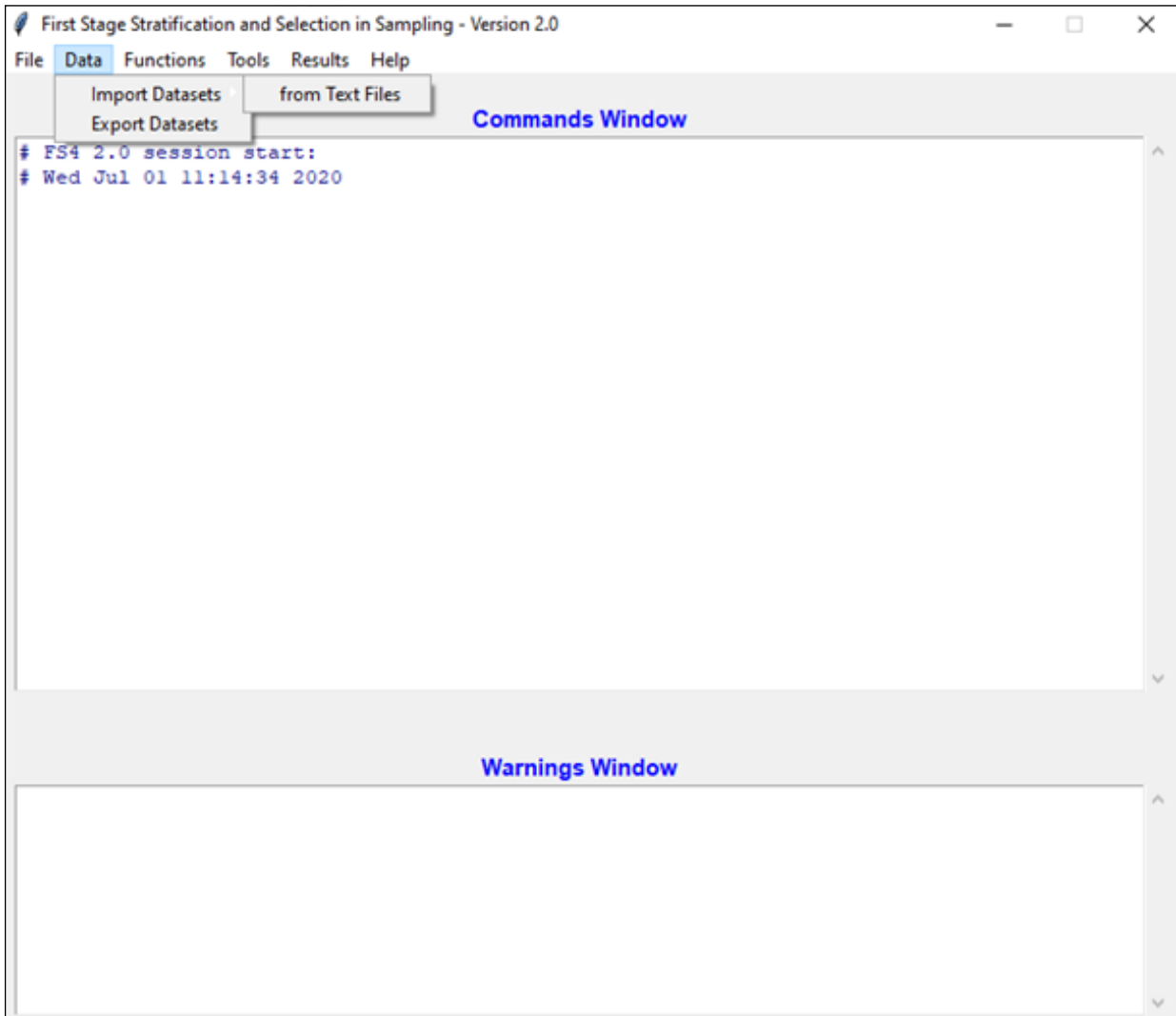


Figure 3: Data menu

Choosing Import Datasets from Text File will be open the window (Fig. 4) for the insertion of dataset name and missing data indicator (NA and “Dataset” are the default respectively), the setting of variables name in the file, the choice of field separator and decimal-point character. Clicking on OK button will display the open-file dialog for reading a text data file which will become an R data frame⁴.

⁴ Note that data sets in FS4 are R data frames.

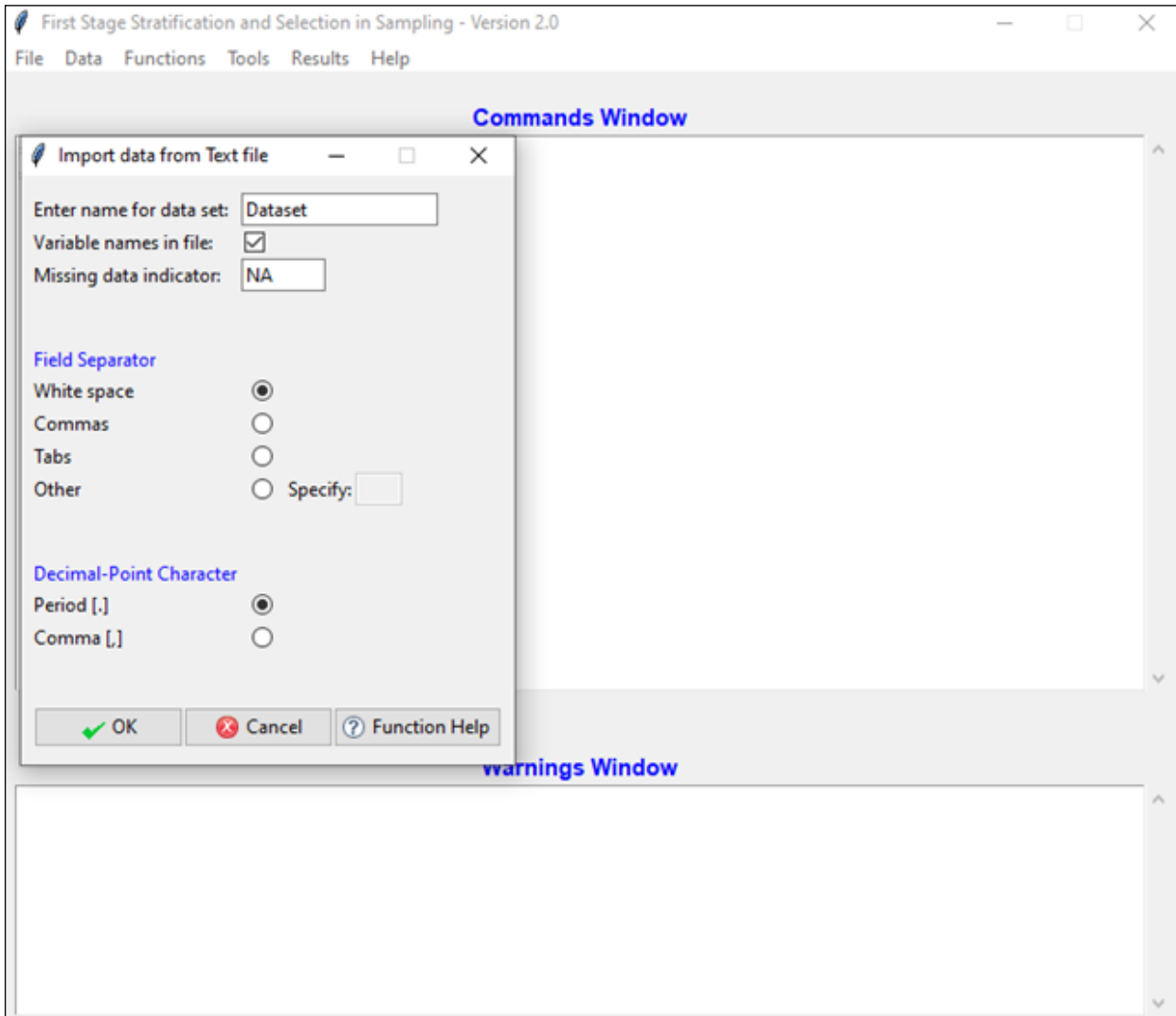
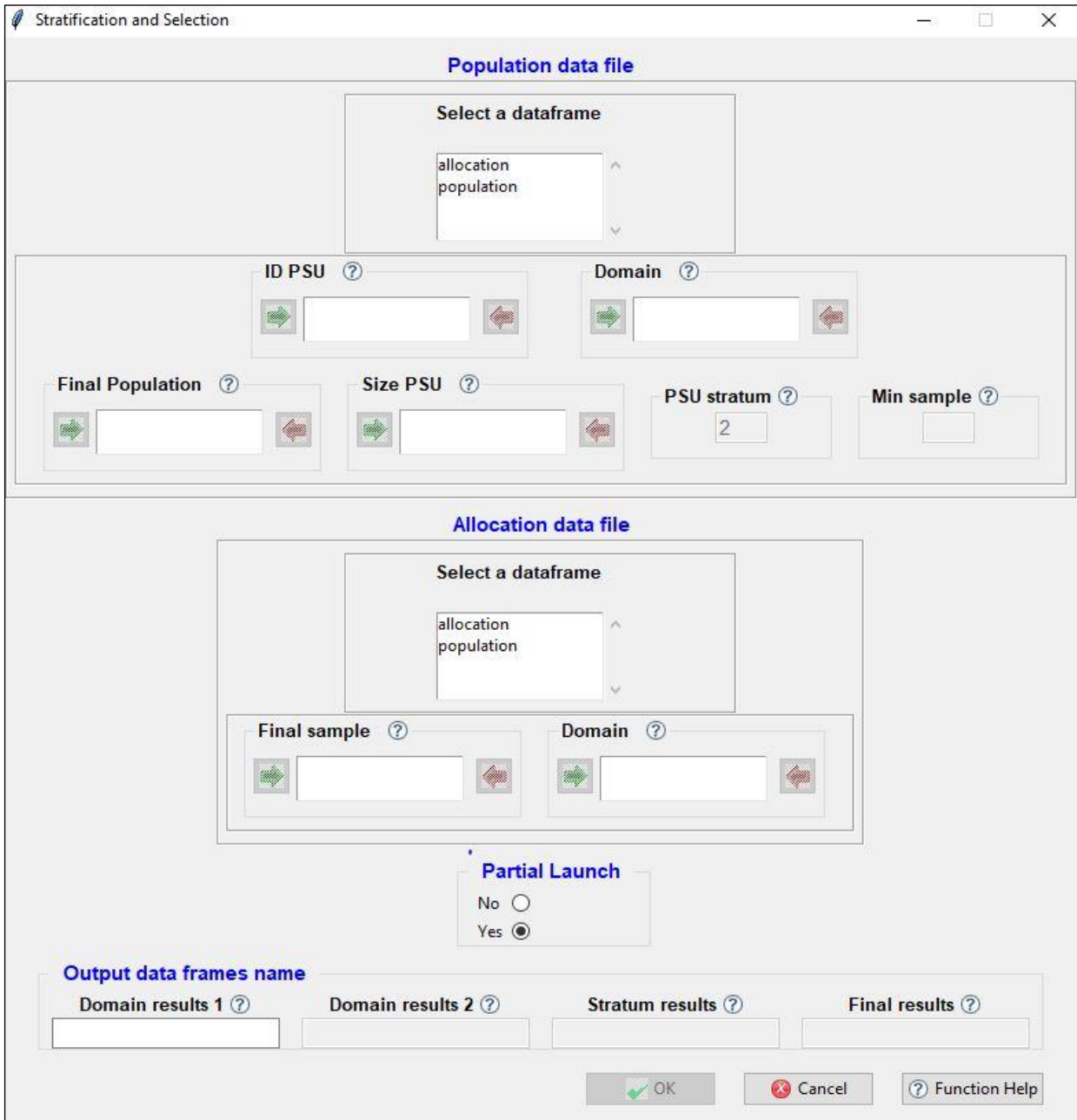


Figure 4: Importing data from a text file

3.3 Functions menu

Menu item for Stratification and Selection. This dialog box consist of four frames: “Population data file”, “Allocation data file”, “Partial Launch” and “Output data frames names”.

At start-up in the snapshot (Fig. 5) are visible only the data frames presents in the user's workspace, the *Cancel* and *Function Help* buttons and is editable only the entry, in the “Output data frames names”, related to the name of the output, at level domain, before the stratification.



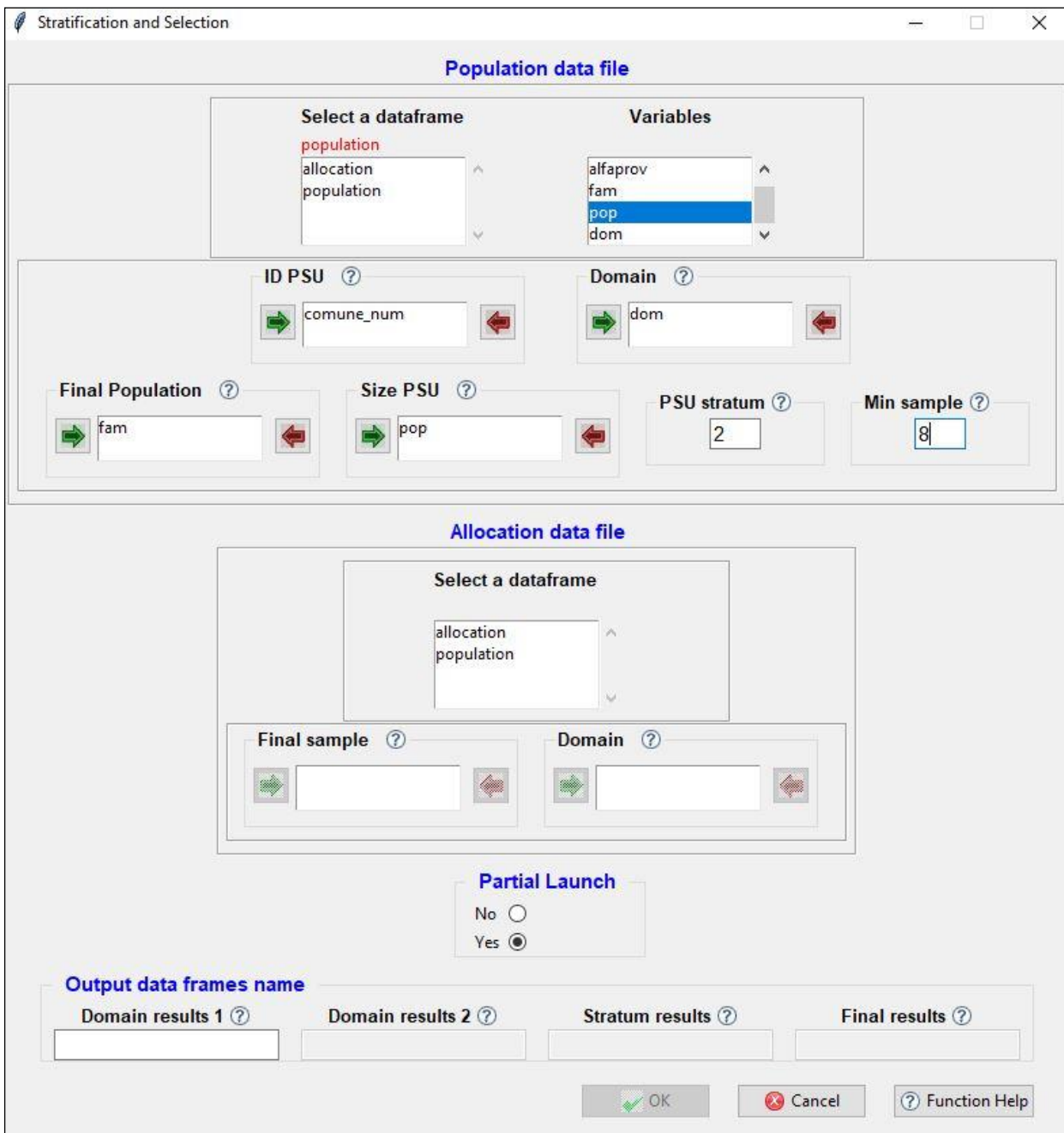
The dialog box is titled "Stratification and Selection" and contains the following sections:

- Population data file:** A "Select a dataframe" list box containing "allocation" and "population". Below it are input fields for "ID PSU" and "Domain", each with a green arrow on the left and a red arrow on the right.
- Allocation data file:** A "Select a dataframe" list box containing "allocation" and "population". Below it are input fields for "Final sample" and "Domain", each with a green arrow on the left and a red arrow on the right.
- Partial Launch:** Radio buttons for "No" and "Yes" (selected).
- Output data frames name:** Four input fields labeled "Domain results 1", "Domain results 2", "Stratum results", and "Final results".

At the bottom, there are three buttons: "OK" (with a green checkmark), "Cancel" (with a red X), and "Function Help" (with a question mark).

Figure 5: Stratification and Selection window at start-up

“Population data file” frame, as well as “Allocation data file” frame, is the part of the form in which you must choose a data frame (a red label indicates the selected data frame) and its related variables from two distinct list boxes. With the first selection (Fig. 6) a list box of the corresponding variables is shown and become clearly visible the arrow buttons (green and red for transfer and remove variables, respectively) related to the “Population data file” frame, as well as the entries concerning the number of sample PSUs to select for each stratum (*PSU stratum*, default value is 2) and the number of final sample units to observe in each PSU (*Min sample*).

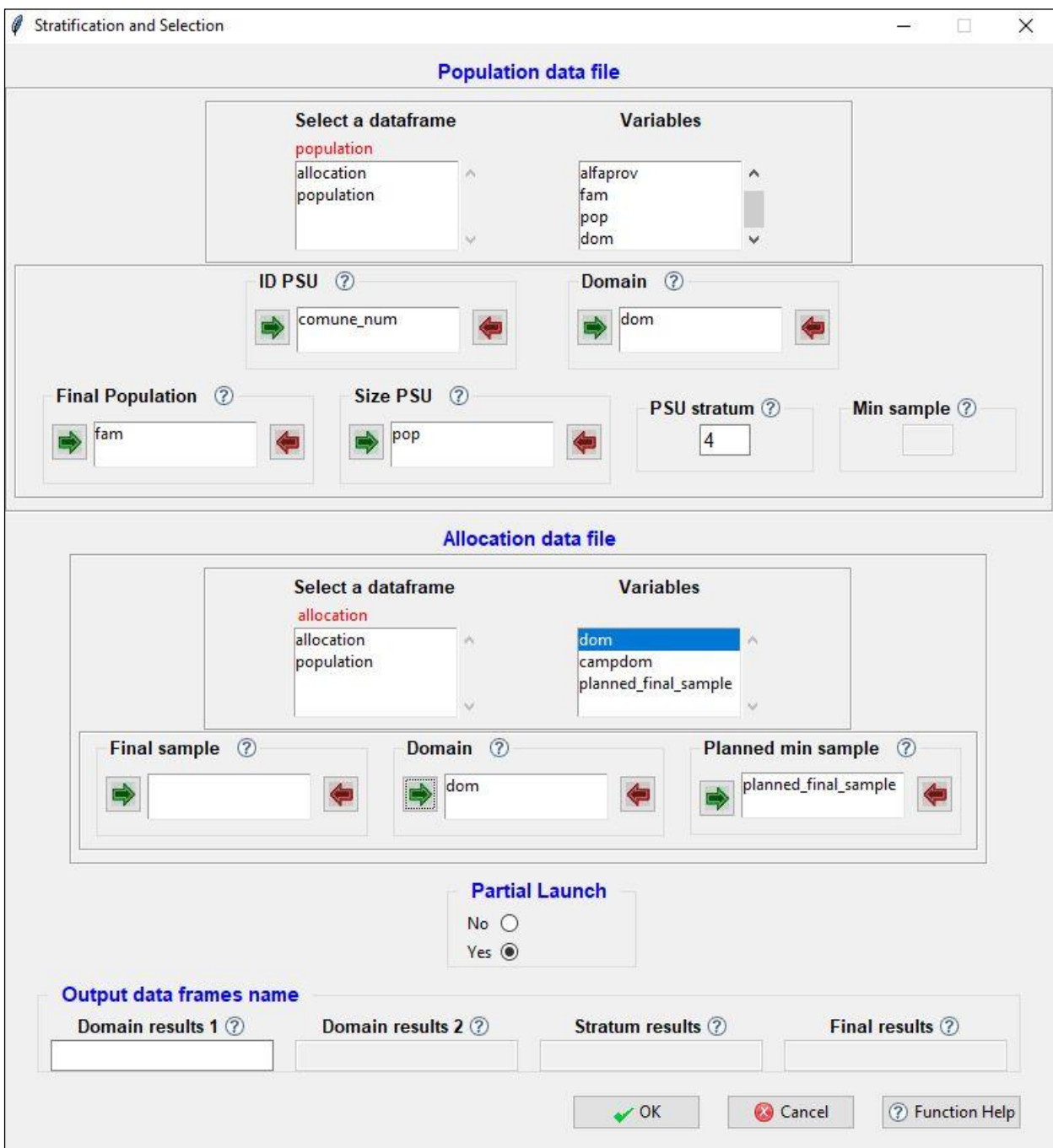


The screenshot shows the "Stratification and Selection" window with the following sections:

- Population data file**
 - Select a dataframe:** A list box containing "population" (highlighted in red), "allocation", and "population".
 - Variables:** A list box containing "alfaprov", "fam", "pop" (highlighted in blue), and "dom".
 - ID PSU:** A text box containing "comune_num" with green and red arrow buttons.
 - Domain:** A text box containing "dom" with green and red arrow buttons.
 - Final Population:** A text box containing "fam" with green and red arrow buttons.
 - Size PSU:** A text box containing "pop" with green and red arrow buttons.
 - PSU stratum:** A text box containing the value "2".
 - Min sample:** A text box containing the value "8".
- Allocation data file**
 - Select a dataframe:** A list box containing "allocation" and "population".
 - Final sample:** An empty text box with green and red arrow buttons.
 - Domain:** An empty text box with green and red arrow buttons.
- Partial Launch:** Radio buttons for "No" and "Yes" (selected).
- Output data frames name:** Four text boxes labeled "Domain results 1", "Domain results 2", "Stratum results", and "Final results".
- Buttons:** "OK", "Cancel", and "Function Help" buttons at the bottom.

Figure 6: Stratification and Selection window after the variables choice in the population data frame

It is indispensable the transfer of each variable in the corresponding field, the insert the *PSU stratum* and the *Min sample* fields. About this, is possible omit it if the number of final sample units isn't constant. Only in this this way, choosing the data frame related to the “Allocation data file” (Fig. 7), the *Min sample* field will not be more editable and will be completely visible the field *Planned min sample* in which is compulsory transfer the data frame variable related to the number of final sample units between domains. In addition to this field, in the “Allocation data file” frame are always compulsory the transfers of the *Final sample* and *Domain* variables. OK button is visible only after the choice of the allocation data frame.



The screenshot shows the 'Stratification and Selection' window with the 'Allocation data file' section active. The 'Population data file' section is visible but disabled. The 'Allocation data file' section contains the following fields:

- Select a dataframe:** A list box with 'allocation' (highlighted in red) and 'population'.
- Variables:** A list box with 'dom' (highlighted in blue), 'campdom', and 'planned_final_sample'.
- ID PSU:** A text box containing 'comune_num'.
- Domain:** A text box containing 'dom'.
- Final Population:** A text box containing 'fam'.
- Size PSU:** A text box containing 'pop'.
- PSU stratum:** A text box containing '4'.
- Min sample:** An empty text box.
- Final sample:** An empty text box.
- Domain:** A text box containing 'dom'.
- Planned min sample:** A text box containing 'planned_final_sample'.

Below the 'Allocation data file' section is the 'Partial Launch' section with radio buttons for 'No' and 'Yes' (selected). At the bottom, there are four text boxes for 'Output data frames name': 'Domain results 1', 'Domain results 2', 'Stratum results', and 'Final results'. At the very bottom are three buttons: 'OK', 'Cancel', and 'Function Help'.

Figure 7: Stratification and Selection window after the variables choice in all the data frames

“Partial Launch” frame is represented by two radio buttons (Yes/No, default is “Yes”) identifying the procedure launch. If launch is partial (radio button setted on “Yes”) means that procedure will be launched in two steps. In this case only one output field of the “Output data frames names” is editable. Once the process is finished, the user, on the basis of output results, can repeat newly the procedure with new input parameters or set radio button on “No” to complete it (Fig. 8). In this way, it’s compulsory the insert (in the meantime editable) of the other fields related to output object names.

Output object/s is/are also stored in the user’s workspace and are available by the “Results” menu and by the “Tools” menu (together with all the other data frames).

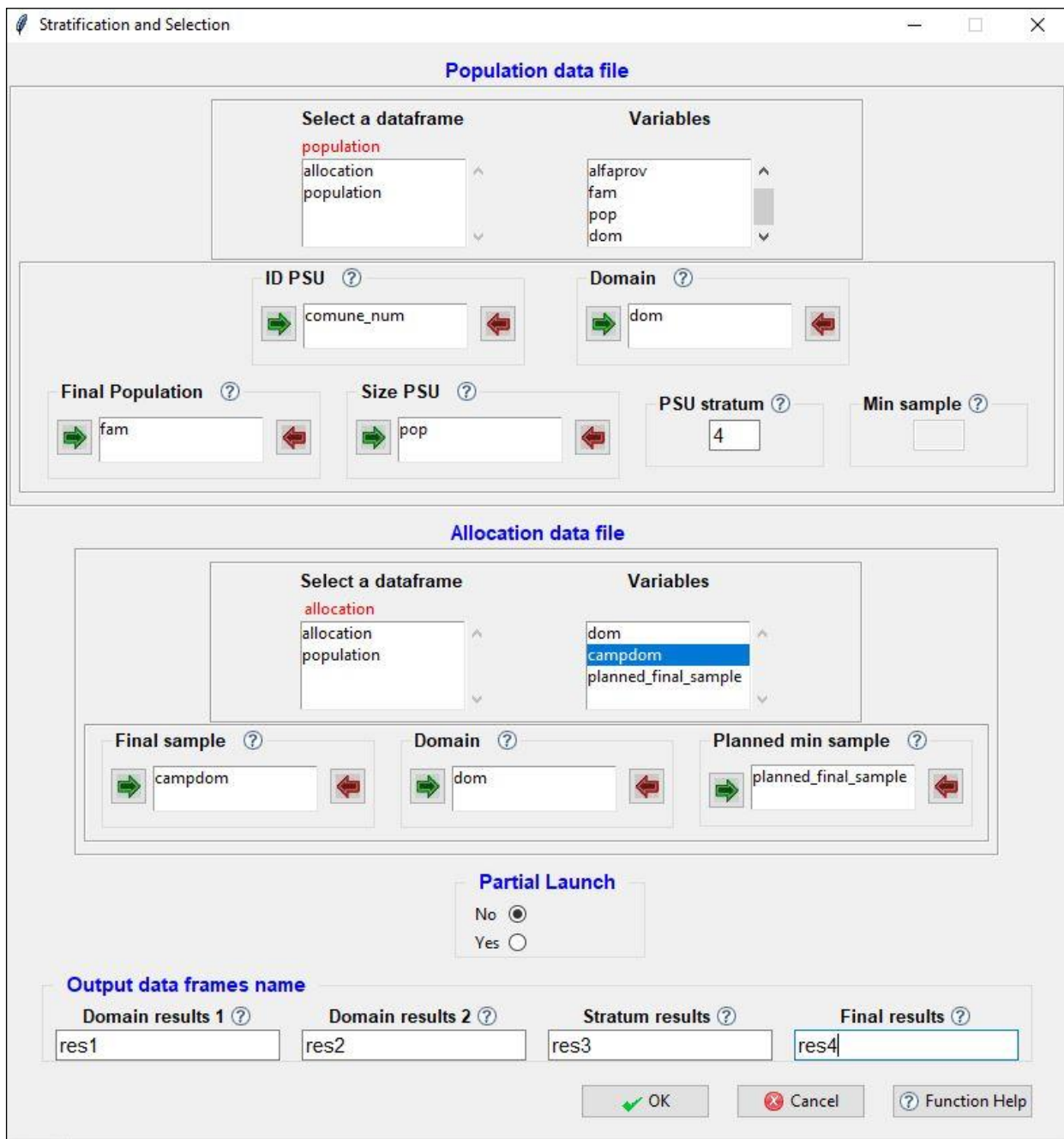


Figure 8: Stratification and Selection window completed

All these mandatory elements are indispensable for the function processing. If some elements aren't selected or inserted, pressing the OK button is displayed an error pop-up specifying the field in which is occurred. If the operation is executed, a pop-up is displayed specifying the success of the transaction). The output dataframe/s is/are stored in the workspace. If launch is partial the command generated by the function with its output object is inserted in the "Commands Window". If launch total on the "Commands Window" will be displayed a command assigned to a list called "Output_list" (see Fig. 9) composed by the four output data frames.

Eventual warning messages are displayed with green colour, in the "Warnings Window", while error messages, producing the stop of computation, are always displayed by means of pop-up dialog boxes.

Cancel button allows the exit from the function, without saving any data.

Function Help button links directly to the FS4 package - StratSel documentation. All the entries of this form have been supplied with tooltips.

All FS4 dialog boxes in general are modal, (except those that trigger directly the corresponding command) so that isn't possible to begin a new operation if the current one isn't finished.

3.4 Tools menu

Menu item (Fig. 9) to display and remove datasets (including results coming from Selection and Stratification process) existing in the current workspace. If no objects are found, a pop-up message is displayed.

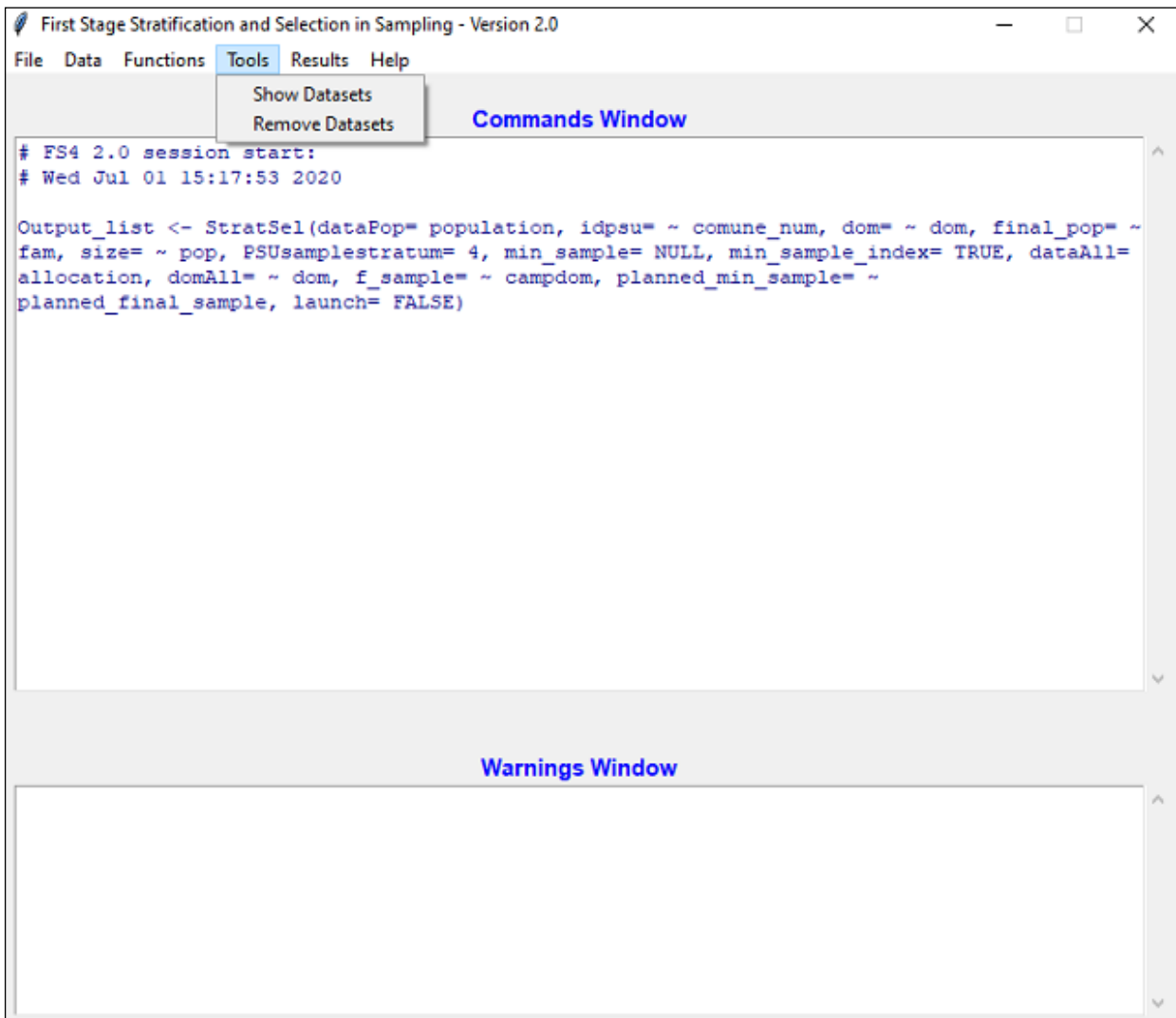


Figure 9: Tools menu

3.5 Results menu

Menu item (Fig. 10) to display only the available output data frames coming from the selection and stratification process. If no results are found, a pop-up message is displayed.

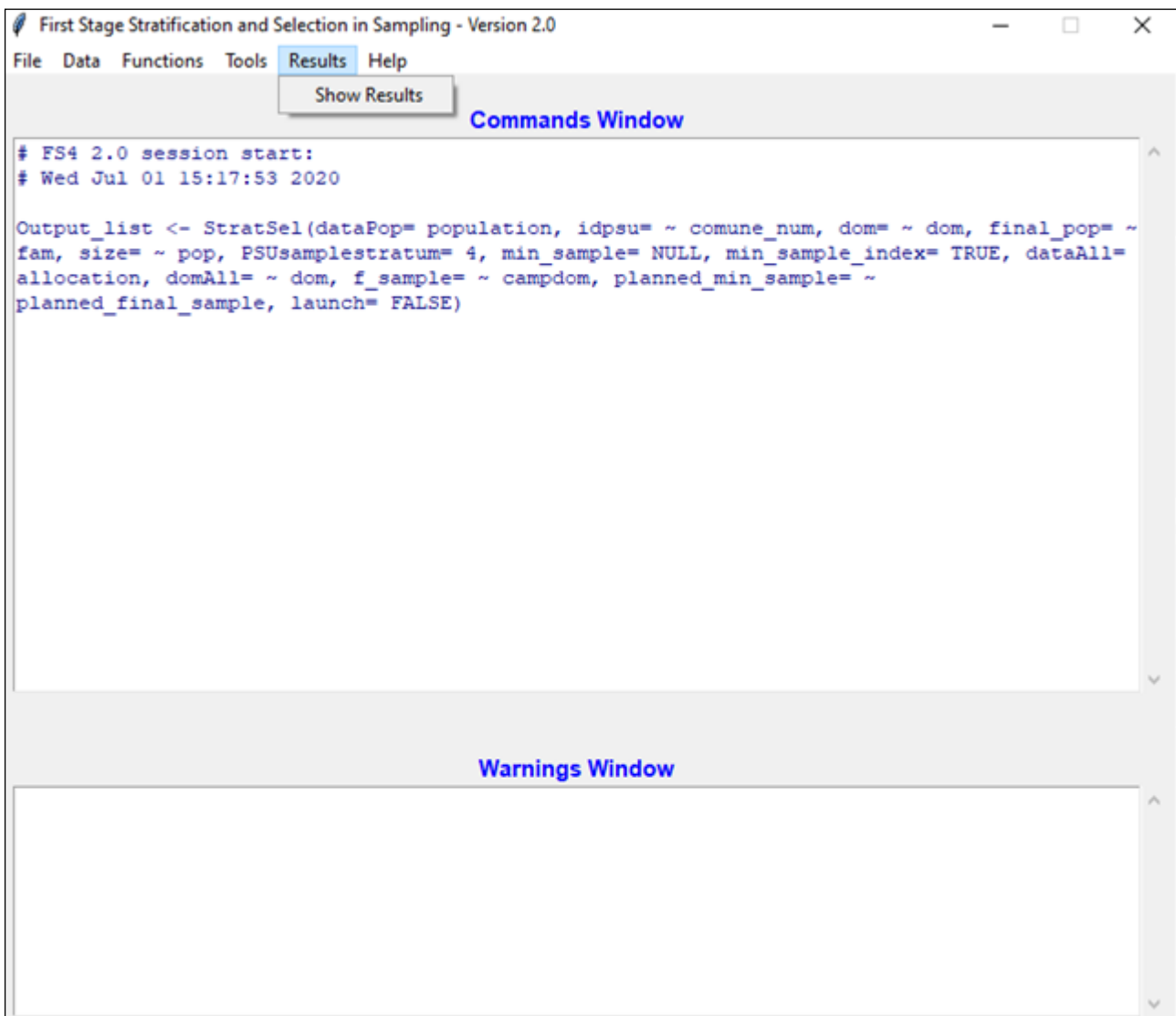


Figure 10: Results menu

3.6 Help menu

Menu item (Fig. 11) to:

- display R documentation about FS4 package;
- FS4 Reference Manual including two functions: FS4() and StratSel() and some examples;
- the current user guide.

Note that each FS4 dialog box has an Help button.

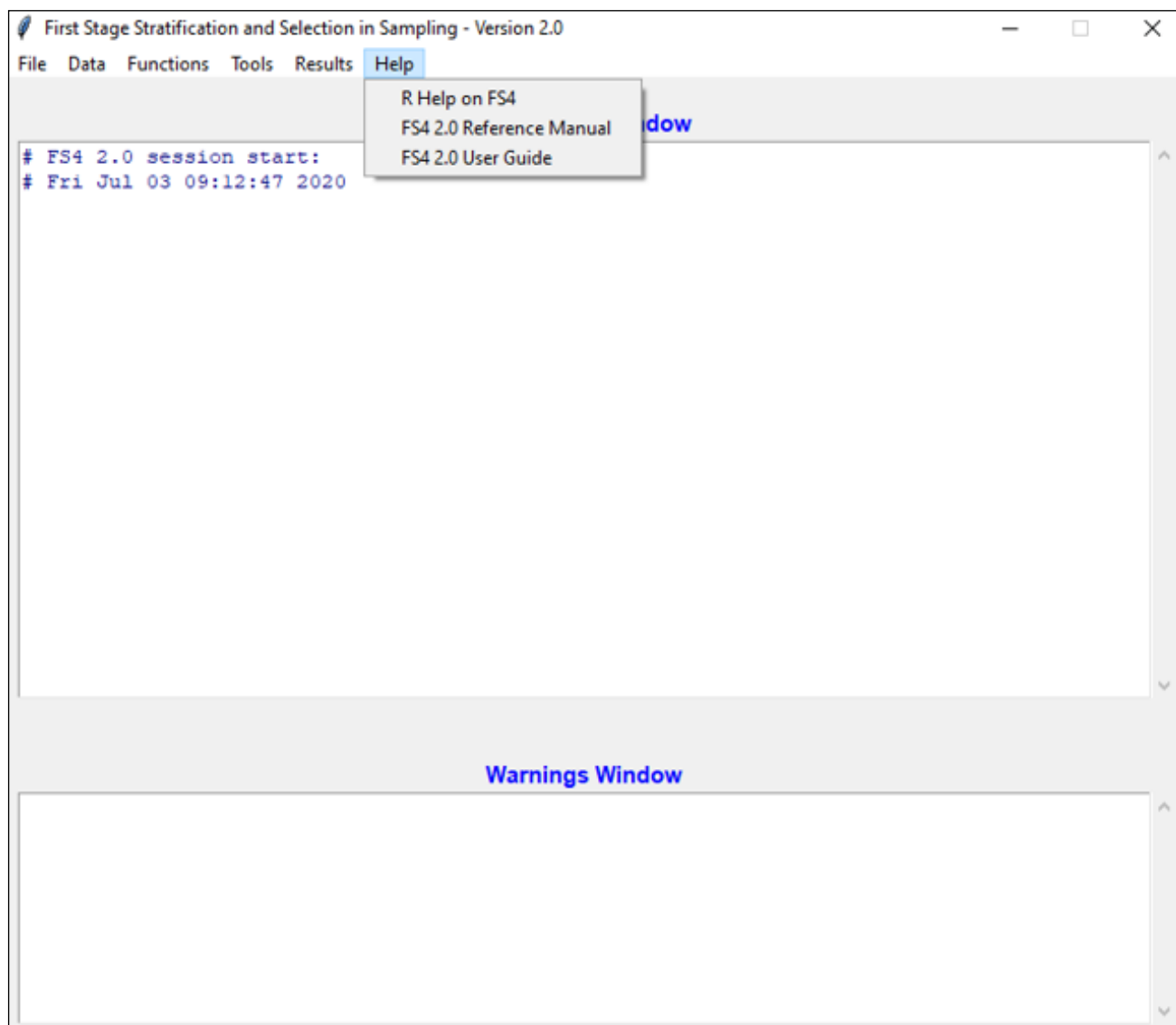


Figure 11: Help menu

References

Cianchetta, R. (2013), In arrivo FS4, nuovo software Open Source per il campionamento, *NewsStat* n. 8 p. 20, http://www.istat.it/it/files/2013/06/NewsStat-8_2013-19_giu_2013-pag.20.pdf.

Sampford, M. (1967), On sampling without replacement with unequal probabilities of selection, *Biometrika*, 54:499-513.