

# Package ‘FS4’

October 31, 2013

**Type** Package

**Title** First Stage Stratification and Selection in Sampling

**Version** 1.0

**Author** Raffaella Cianchetta

**Maintainer** Raffaella Cianchetta <cianchet@istat.it>

**Description** The R package FS4 has been developed 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

**License** EUPL

**Depends** R (>= 3.0.1), tcltk, tcltk2, svMisc, plyr, sampling

**Imports** utils

**ByteCompile** TRUE

## R topics documented:

allocation . . . . .	1
FS4 . . . . .	2
population . . . . .	4
StratSel . . . . .	5

<b>Index</b>	<b>8</b>
--------------	----------

---

allocation	<i>Allocation data frame</i>
------------	------------------------------

---

## Description

This is an example data frame coming from ISTAT Multipurpose survey data that enables to run R code included in the ‘Examples’ section of the FS4 package help pages.

## Usage

```
data(allocation)
```

**Format**

The allocation data frame contains 105 observations with the following variables:

**dom** Variable identifying domain

**campdom** Variable identifying final sample units

**planned\_final\_sample** Variable identifying planned secondary sample units to observe in each PSU variable between domains

**Examples**

```
data(allocation)
str(allocation)
```

---

FS4

*First Stage Stratification and Selection in Sampling*

---

**Description**

Starts a user friendly GUI (graphical user interface) that allows to access the functionality of the package for first stage stratification and selection in sampling related to two or more stages.

**Usage**

```
FS4()
```

**Details**

The main window of the GUI consists of a menu bar and, from the top to the bottom, of two scrolable 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 section 'Menu Tree').

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 multiple 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 unit size, by the sizes, single or plural (variable between domains), planned in input. All the entries of this form have been supplied with tooltips.

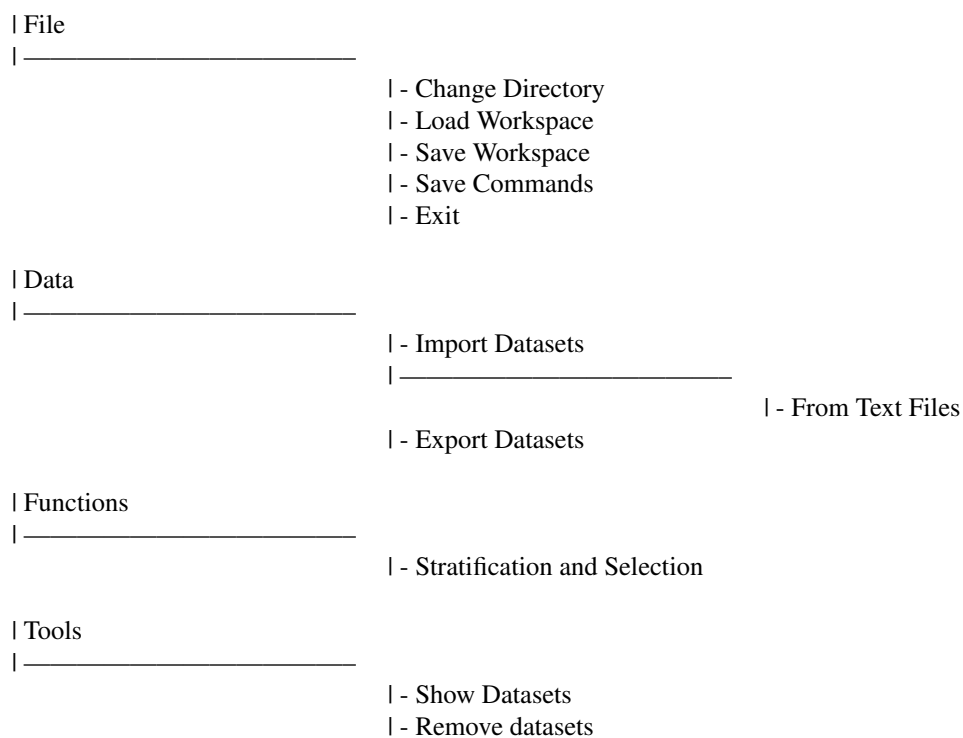
By default, commands generated by the GUI are inserted to the Commands Window (is possible to save commands via *File -> Save Commands* or copy them by pressing *Control-C* or right-clicking the mouse). Warning messages are displayed in the Warnings Window, instead error messages, producing the stop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

In addition to "Help" menu, for each dialog box coming from "Data" and "Functions" menus is available a help button that links to the connected help pages from the corresponding source.

The end of a GUI session is possible either using the *File -> Exit* menu or by closing the FS4 main window.

## Menu Tree

The menu "tree" for this FS4 package (version 1.0) is shown below.



| Results

|-----

| - Show results

| Help

|-----

| - R Help on FS4

| - FS4 Reference Manual

| - FS4 User Guide

### Note

This is a Windows version, FS4() can properly run under the *Rgui* in the SDI (single-document interface) mode.

### Author(s)

Raffaella Cianchetta

### References

R. Cianchetta (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](http://www.istat.it/it/files/2013/06/NewsStat-8_2013-19_giu_2013-pag.20.pdf)

### Examples

```
# Start FS4
## Not run:
FS4()

## End(Not run)
```

---

population

*Population data frame*

---

### Description

This is an example data frame coming from ISTAT Multipurpose survey data that enables to run R code included in the 'Examples' section of the FS4 package help pages.

### Usage

```
data(population)
```

**Format**

The population data frame contains 8094 observations related to Italian municipality with the following variables:

**CAP** Variable identifying Italy's postal code  
**comune\_num** Variable identifying municipalities  
**alfaprov** Variable identifying province of belonging  
**fam** Variable identifying secondary stage units  
**pop** Variable identifying PSU size  
**dom** Variable that refers to the belonging domain

**Examples**

```
data(population)
str(population)
```

---

 StratSel

*Stratification and selection*


---

**Description**

Merge two data frames (whatever PSU population and allocation data frames) and then compute stratification and selection of a fixed number of sample PSUs (Primary Sampling Units) 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 realizes for each domain estimation, 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 (Self Representative) and each constitutes a stratum by itself, so they come into the sample with probability equal to one. The remaining NSR (Non Self Representative) 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.

If the number of NSR PSUs in each stratum is greater than the number of sample PSUs for each stratum, then it is indispensable to define the vector of the inclusion probabilities (`pik`), argument of `UPSampford` function, vice versa NSR PSUs will become SR PSUs. If some `pik` values are greater than 1, `pik` is newly calculated, for those PSUs belonging to the same domain and stratum of the PSUs having `pik > 1`, until `pik` values will be all less than 1, then applies the `UPSampford` function.

**Usage**

```
StratSel(dataPop, idpsu, dom, final_pop, size, PSUsamplestratum, min_sample,
min_sample_index = FALSE, dataAll, domAll, f_sample,
planned_min_sample = NULL, launch = TRUE)
```

**Arguments**

<code>dataPop</code>	PSU population data frame.
<code>idpsu</code>	Formula identifying primary sample units identifier.
<code>dom</code>	Formula identifying the variable domain.
<code>final_pop</code>	Formula identifying secondary stage units.

<code>size</code>	Formula identifying first stage units size.
<code>PSUsamplestratum</code>	Number of sample PSUs to select to each stratum.
<code>min_sample</code>	Number of final sample units to observe in each PSU.
<code>min_sample_index</code>	Identify the absence of <code>planned_min_sample</code> (default = FALSE)
<code>dataAll</code>	Allocation data frame.
<code>domAll</code>	Formula identifying the domain variable for allocation data frame.
<code>f_sample</code>	Formula identifying final sample units.
<code>planned_min_sample</code>	Formula identifying planned final sample units to observe in each PSU, variable between domains; NULL (the default) means the existence of a fixed number of them identify from <code>min_sample</code> .
<code>launch</code>	Identify the parameter related to the launch procedure. If default = TRUE the launch is partial (see 'Details').

### Details

It is possible to launch the procedure in two separate steps, so that user can see a first output (`launch = TRUE`) and decide if modify the input parameters or continue the procedure setting `launch = FALSE`.

### Value

An object of class `list` or `data.frame` depending from argument `launch`. If `launch = TRUE` the only output is, at domain level, a data frame of the Self Representative (SR) PSUs and of the Non Self Representative (NSR) PSUs before stratification. If `launch = FALSE` the output is a `list` composed by four members of `data.frame` class. The first component, at domain level, is the same data frame obtained if `launch = TRUE`. The second element is, at domain level, a data frame of SR and NSR PSUs after stratification, with their totals like the first output, but with additional information such as the final units sample size distinctly for SR and NSR PSUs and their mean. The third component is a data frame that supply, for each stratum, the number of sample PSUs selected and the total number of PSUs. The fourth element provides, for each PSU, some information like the inclusion probability and the sampling fraction.

### Author(s)

Raffaella Cianchetta

### References

- S. Falorsi A. Russo (2001), Il disegno di rilevazione per indagini Panel sulle famiglie, *Rivista di Statistica Ufficiale*, N. 3, p. 55-90.
- Sampford, M. (1967), On sampling without replacement with unequal probabilities of selection, *Biometrika*, 54:499-513.
- Yves Tille' and Alina Matei (2012). `sampling: Survey Sampling`. R package version 2.5.  
<http://CRAN.R-project.org/package=sampling>.

**Examples**

```
# Start StratSel
## Not run:
load(FS4)
data(population)
data(allocation)

# The function with a fixed number of final sample units (min_sample= 8)
# to observe in each PSU and partial launch of the procedure with
# only one output data frame(see Value)
Output_list <- StratSel(dataPop= population, idpsu= ~ comune_num, dom= ~ dom,
  final_pop= ~ fam, size= ~ pop, PSUsamplestratum= 6, min_sample= 8,
  min_sample_index= FALSE, dataAll= allocation, domAll= ~ dom,
  f_sample= ~ campdom, planned_min_sample= NULL, launch= TRUE)

# It can also be written as below due to default values:
Output_list <- StratSel(dataPop= population, idpsu= ~ comune_num, dom= ~ dom,
  final_pop= ~ fam, size= ~ pop, PSUsamplestratum= 6, min_sample= 8,
  dataAll= allocation, domAll= ~ dom, f_sample= ~ campdom)

# The function with a fixed number of final sample units (min_sample= 8)
# to observe in each PSU and full launch of the procedure with
# the output list composed of four data frames(see Value)
Output_list <- StratSel(dataPop= population, idpsu= ~ comune_num, dom= ~ dom,
  final_pop= ~ fam, size= ~ pop, PSUsamplestratum= 6, min_sample= 8,
  dataAll= allocation, domAll= ~ dom, f_sample= ~ campdom, launch= FALSE)

# The function with a variable number of final sample units (planned_min_sample=
# ~ planned_final_sample) and partial launch of the procedure
Output_list <- StratSel(dataPop= population, idpsu= ~ comune_num, dom= ~ dom,
  final_pop= ~ fam, size= ~ pop, PSUsamplestratum= 6, min_sample= NULL,
  min_sample_index= TRUE, dataAll= allocation, domAll= ~ dom,
  f_sample= ~ campdom, planned_min_sample= ~ planned_final_sample)

# The function with a variable number of final sample units (planned_min_sample=
# ~ planned_final_sample) and full launch of the procedure with the output list
#composed of four data frames
Output_list <- StratSel(dataPop= population, idpsu= ~ comune_num, dom= ~ dom,
  final_pop= ~ fam, size= ~ pop, PSUsamplestratum= 6, min_sample= NULL,
  min_sample_index= TRUE, dataAll= allocation, domAll= ~ dom,
  f_sample= ~ campdom, planned_min_sample= ~ planned_final_sample,
  launch= FALSE)

## End(Not run)
```

# Index

\*Topic **datasets**  
    allocation, 1  
    population, 4  
\*Topic **survey**  
    StratSel, 5

allocation, 1

FS4, 2

population, 4

StratSel, 5