

SBS to ARGUS Translator

Anco Hundepool
9 November 2009

Introduction

τ -ARGUS is a program for the Statistical Disclosure Control of tabular data. τ -ARGUS can read both microdata as well as tabular data. When using microdata τ -ARGUS offers all its flexibility, while using tabular data the flexibility is more restricted. Although τ -ARGUS can handle different varieties of tabular input, it was till now never requested to read the traditional SBS-format into τ -ARGUS. Nevertheless the SBS-format is a well-known format in Europe. So an additional program (SBS2ARGUS) has been written to translate the specific SBS-format into a τ -ARGUS-readable input file. If this program proves to be useful we could eventually include the functionality in τ -ARGUS.

Program description

SBS2ARGUS reads a SBS-file and will create a tab-file, containing the data, a RDA-file containing the metadata and a a-priory file containing the status of the unsafe cells.

The input file

The input file is a free-format file (separator to be specified) and will contain the following columns

- Up to 3 codes of the spanning variable
- The cell value
- The cell frequency
- The status flag (A=freq-unsafe, B= (1,k) unsafe, C = (2,k) unsafe). All other values are treated as safe
- The dominance-percentage

Optionally the first line of the datafile may contain some metadata. i.e. the name of the columns. If available SBS2ARGUS will use the names, otherwise the variable names have to be entered manually.

```

Region;Size;Value;Freq;Flag;Dom
Total;Total;16847646.84;42723;V;0
Total;2;20;9;D;0
Total;4;25;5;D;0
Total;5;2711808;20002;V;0
Total;6;2320534;8831;V;0
Total;7;2505042.58;5498;V;0
Total;8;2799074.26;4594;V;0
Total;9;6510758;3779;V;0
Total;99;385;5;V;0
Nr;Total;4373664;11395;V;0
Nr;2;5;6;D;0
Nr;4;5;1;A;0
Nr;5;719049;5137;V;0
Nr;6;659680;2471;D;0
Nr;7;688962;1487;V;0
Nr;8;756529;1426;V;0
Nr;9;1549049;862;V;0
Nr;99;385;5;V;0
" 1";Total;1986129;6112;V;0
" 1";2;5;1;A;0
" 1";4;5;1;A;0
" 1";5;398062;2856;V;0
" 1";6;348039;1334;D;0

```

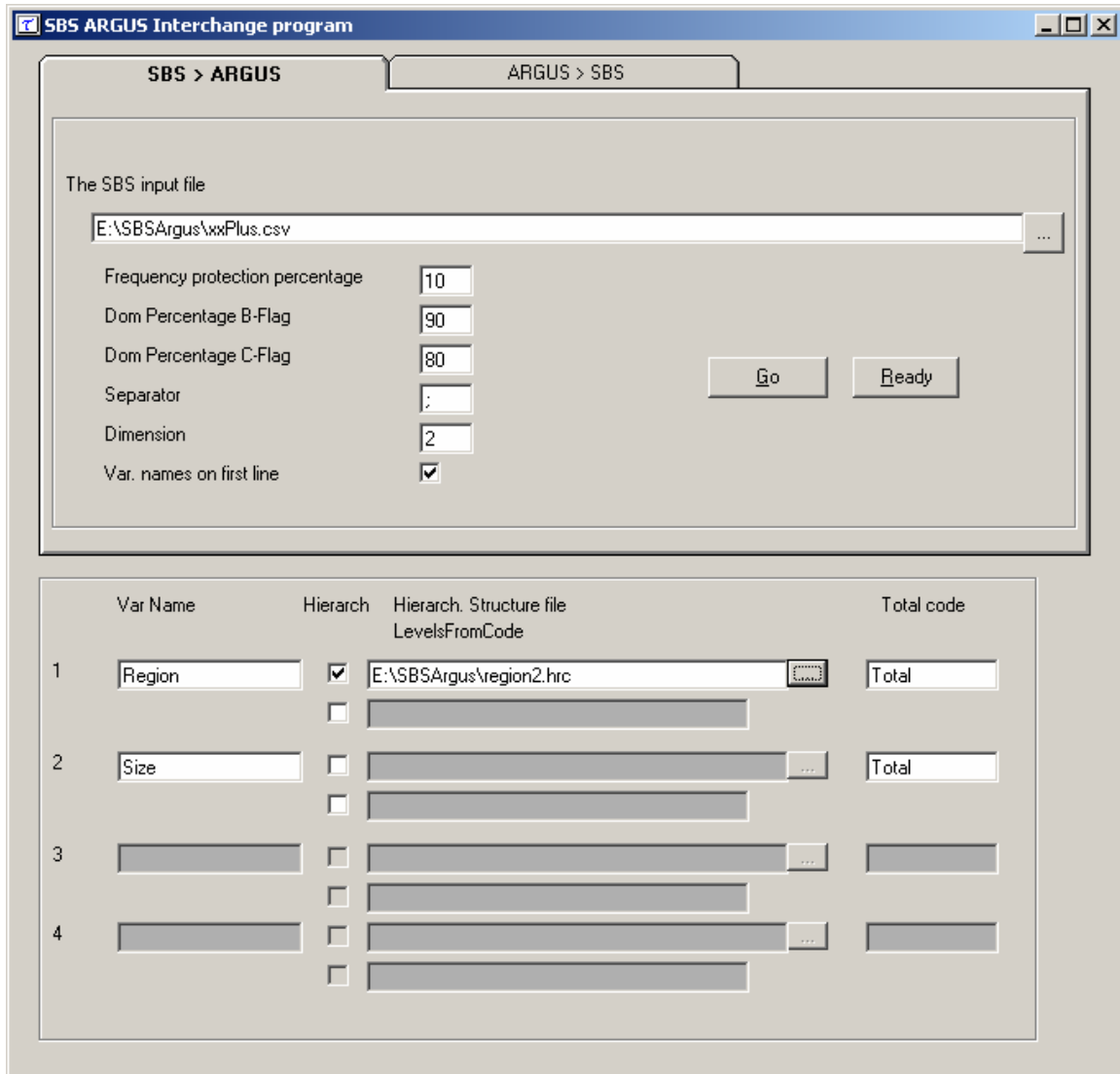
Example of a SBS-inputfile (with column names)

The program interface

On the top half you specify the file name of the SBS-file, the safety range for the freq-unsafe cells (flag A) ,the dom-percentage (K) for the 1,K- rule (B-flag), the dom-percentage of the 2,K-rule (C-Flag) and the separator used in the free format data files.

All statuses other that A, B, or C will treated as safe!

Also the dimension of the table must be specified and whether or not the first line contains variable names.



On the lower part you specify the names of the explanatory variables, whether they are hierarchical, and the code for the total.

There are two options for specifying the hierarchical structure:

1. The first option is to be checked if the hier. structure comes from a .HRC file.
2. The second one is to be checked if the hier. structure comes from the codes itself. Just enter a string like "1 2 1 1"

The output files will bear the name of the inputfile plus ".tab" for the data and ".rda" for the metadata, ".hst" for the apriory file and ".log" for the log file.

Note1: The reason for creating the apriory file is the following. Sometimes it happens that the table given to τ -ARGUS is not strictly additive. Additivity is a strong condition in τ -ARGUS. If a table is not additive, τ -ARGUS cannot protect it. Of course we advise to supply a strictly additive table to τ -ARGUS, but

τ -ARGUS has an option to make a table additive. The drawback of this option is that the status of all (sub-) totals will be lost. By applying the apriory file generated by SBS2ARGUS, the statuses are again set to the required values.

Note2: Only the first tab-page works. I do not yet know whether the second will ever be used. Maybe for future use.

Note3: The logfile will report a.o. about problematic cells. E.g. if the dom. percentage specified in the datafile is smaller than the global dom. percentage used for the rule. This is not consistent and the protection levels cannot be computed correctly. In that case the cell is assumed to be unsafe by the frequency rule and the corresponding protection level is computed.

Note4: All remarks, suggestions for improvement and questions are to be send to Anco Hundepool (aj.hundepool@cbs.nl)