

Minorities at Risk



MARGene

Minorities at Risk Data Generation and Management Program

Quickstart Guide

Updated January 27, 2003

MARGene allows you to build a data set of Minorities at Risk data, with only the variables, groups, countries, and time points that you want.

Quickstart Guide Contents:

- I. Running the program
- II. Making basic selections
- III. Common questions
 - 1) Want more help? Want details of features?
 - 2) How do I interface with software packages?
 - 3) How can I merge MAR data with other data sets?
 - 4) How do I obtain information on one specific group?
 - 5) How do I obtain lagged values of variables?
- I. Running the program**

To get started with MARGene, double-click the desktop icon for MARGene, or use windows explorer to find the directory where you installed MARGene (typically “C:\program files\MARGene”) and double-click the file “MARGene.exe”.

After MARGene initializes, you should then make several selections to build your dataset. Once your selections are made, press the “Go! Create Data Set Now” button at the bottom of the window, or choose “Create Now” under the “Create Data Set” Menu.

II. Making basic selections

On the first tab, labeled “Main Settings,” you define the underlying characteristics of your data set.

- 1) **Do you want information on individual groups, or do you want information aggregated to the state, region or globe?** Do you want one data point per group? Select “Group” under the “physical aggregation” section of the window. Or, do you want one data point per country or region? Select “Country” or “Region” and data on all of the groups within that country or region will be aggregated to one observation. Want just one observation for the entire world in a given year? Select “Globe.”
- 2) **Do you want only a specific subset of groups, states, or regions, or do you want all cases?** If you want all cases, then leave the button marked for “All Groups” “All States” or “All Regions.” If you want to pick only selected groups, click the button for “Subset of Groups” (or subset of states, or regions) and then press the “Specify Subset” button to pick the individual cases you want.
- 3) **Do you want control groups from other states?** If you only want information on groups that were actually at-risk minorities, or if you want data only on states that actually contained such groups, then leave the “Non-MAR Control Groups” box unchecked. BUT, if you want “dummy” information on other states, for instance so that you can compare countries that have MAR groups to countries without MAR groups, then check this box and “dummy” records will be created in the data set.
- 4) **Do you want the data set to be annual, or at less frequent time intervals?** Check whether you want the observations in the data set to be every year, every 2 years, 5 years, or 10 years. If you select a value greater than 1 year, each observation in your output data set will contain the minimum, maximum, and average of values across that time period.
- 5) **How do you want missing data handled?** If you want MARGene to interpolate missing data to produce a continuous time-series, check “Interpolate Intertemporal Data Points,” and select a method for interpolation (described fully in the main documentation). If you do not want MARGene to do any interpolation, you may decide whether you want MARGene to output the missing cases (where there is no data between existing observations) or omit them (the “Output Missing In Series Cases” box).
- 6) **What time period do you want?** You may specify a starting and ending year for the data set, or simply allow MARGene to output all available years of data.

On the second tab, labeled “Output Options,” you decide where to save your data set and other details of the output.

- 1) **Do you want to save your data set? Where?** You may have data simply appear on the computer screen, send it directly to the printer, or (most likely) save it to a file on the computer. To change the output file, push the “change output file” button.

- 2) **Do you want the data values in the output separated by commas, tabs, or spaces?** Most software will easily read data separated by commas, and so this is the default setting. However, if you want the data to be separated by tabs or spaces, you can change this option.
- 3) **Do you want command files to read the data into other software?** If you are going to be analyzing the Minorities at Risk data in Stata, SPSS, or LIMDEP, MARGene will prepare command files for you. You may use these command files to easily read your data into these software programs, to set appropriate missing values, and as documentation of precisely what settings you chose.
- 4) **Do you want a “header line” in the data output?** If you are going to be reading data into a program like Excel, you may wish to include a header line. With a header line, the names of the variables will appear in the first row of the data set.

On the third tab, labeled “Variables,” you pick the variables you want included in your data set.

- 1) **Choose the variables you want in the data set.** Simply put a check mark next to any variable that you want to have in the output data set. The variables are grouped by the area of the main Minorities at Risk data manual in which they appear.
- 2) **What variables are what?** To determine what displayed variable names correspond to what concepts, you can do two things. First, simply holding the cursor over a variable for approximately 1/2 second will bring up a “pop-up” description/identification of the variable. Second, you may also consult the main Minorities at Risk data manual for detailed variable information.

Once your selections are made, press the “Go! Create Data Set Now” button at the bottom of the window, or choose “Create Now” under the “Create Data Set” Menu.

III. Common questions

- 1) **Want more help? Want details of features?** (like “interpolation” or “including non-MAR states”?) Complete documentation for MARGene is provided with the software. You can load the documentation file automatically by going under MARGene’s HELP menu, and selecting “Help | MARGene Documentation.” The documentation is also available as a separate document in the file “MARGene Documentation.doc” which is located in the “Documentation” subdirectory under the main directory where MARGene was installed. Normally, then, the path to the file will be “c:\program files\MARGene\Documentation\MARGene Documentation.doc”.
- 2) **How do I interface with software packages?** If you check off the option to “create command files” on the “Output Options” tab, a command file will be created for you. You should load this command file into the appropriate software program and then “run” or “execute” the commands therein to load the MAR data into them. The way in which this is done varies slightly by program; you may open the command file in a text editor (like notepad, wordpad, word, or wordperfect) and the exact steps of the procedure for Stata, SPSS, or LIMDEP are listed.

- 3) **How can I merge MAR data with other data sets?** MARGene creates a pooled-time series data set of groups (or states) over time. Each line of data contains critical identifying variables, including “numcode” (identifying the group), “ccode” (identifying the state within which this group lies), and “year” (identifying the year of the observation). With these variables as the “matching” or “key” variables, you can merge the data with data from any other source so long as the other data set has the same key variables. To do this in Stata, you would use the “merge” command. For instance, if you were matching in other annual, country-level data your command would look something like:

```
merge ccode year using otherdata.dta
```

In SPSS, you would use the “match files” command with the “by” option, as (for instance):

```
match files file=* / table='d:\otherdata.sps' / by ccode year
```

- 4) **How do I obtain information on one specific group?** Under the “Main Settings” window, select the group as the physical aggregation. Then, check “Subset of Groups” and press the “Specify Subset” button. Choose the group (or groups) from the available list (the list contains all of the groups identified by the MAR project that have any data in the data set). Choose your other settings (output data set location, variables, etc.), and create the data set. It will contain information only on the specific groups you identified.

- 5) **How do I obtain lagged values of variables?** MARGene does not automatically build lagged values of variables. You should do this in your statistical analysis software. This is easy in many programs. Suppose you had requested annual output on groups, for instance. In Stata you would first issue a command to define the data as panel data:

```
tsset numcode year, yearly
```

This command tells Stata that the dataset 1) contains information on many unique groups which are identified by the “numcode” variable (this is the variable in the data set that has a unique number for each group), and 2) contains annual information, with the date of the year contained in variable “year”. Once the tsset command has been issued, then the 1 year lagged value of a variable may be obtained using the “L.” operator within Stata. For instance, the one year lag of variable “gc2” is obtained by using “L.gc2”. The 2 year lag may be obtained via “L2.gc2”, and so on. Other similar commands (like “lag()” in SPSS) will apply to other software packages, although care should be taken to accommodate the pooled-time-series format of the data.