# Applications in DAD

INTRODUCTION TO APPLICATIONS

Remember that DAD can activate one or two databases. Once a database is activated, the user can then call different applications of DAD. Before you reach those applications, however, you must indicate how many databases are to be used in the application, and which ones. This is done through the following window:

Figure 1: The dialogue box to initiating the number of distributions
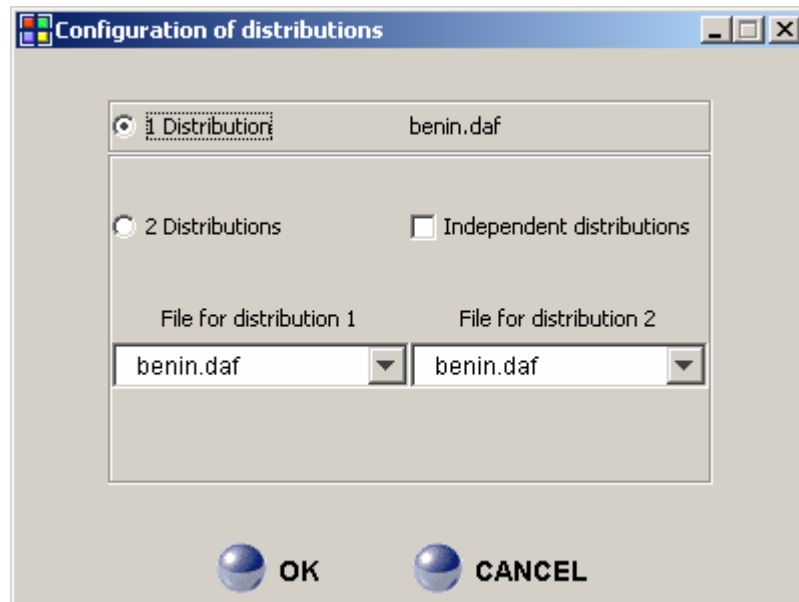
Each database represents one distribution. Generally, you should indicate the following information:

- The number of distributions
- The name of the file representing the first distribution.
- The name of the file representing the second distribution.
- When two distributions are to be used, you should indicate if the two distributions represent dependent or independent samples for the accurate computation of standard errors that use information on the joint distribution.

REMARK: If the number of distributions is one, the activated file is automatically the file specified on the 1st line.

Confirm your choice by clicking on the button **"OK".** Once the choice is confirmed, you can reach the desired application.

Once data have been read into *DAD* and that the sampling design has been specified[1], the field is wide open for the estimation of distributive statistics and for performing distributive tests. For every application programmed in *DAD*, there is a specific application window that facilitates the specification of variables, parameters and options to generate the desired distributive statistics. For example, Figure 2 shows the specific application window for computing the Atkinson index with one distribution. There is a separate specific window for the case of two distributions.

Figure 2: Application window for estimating the FGT poverty index – one distribution.
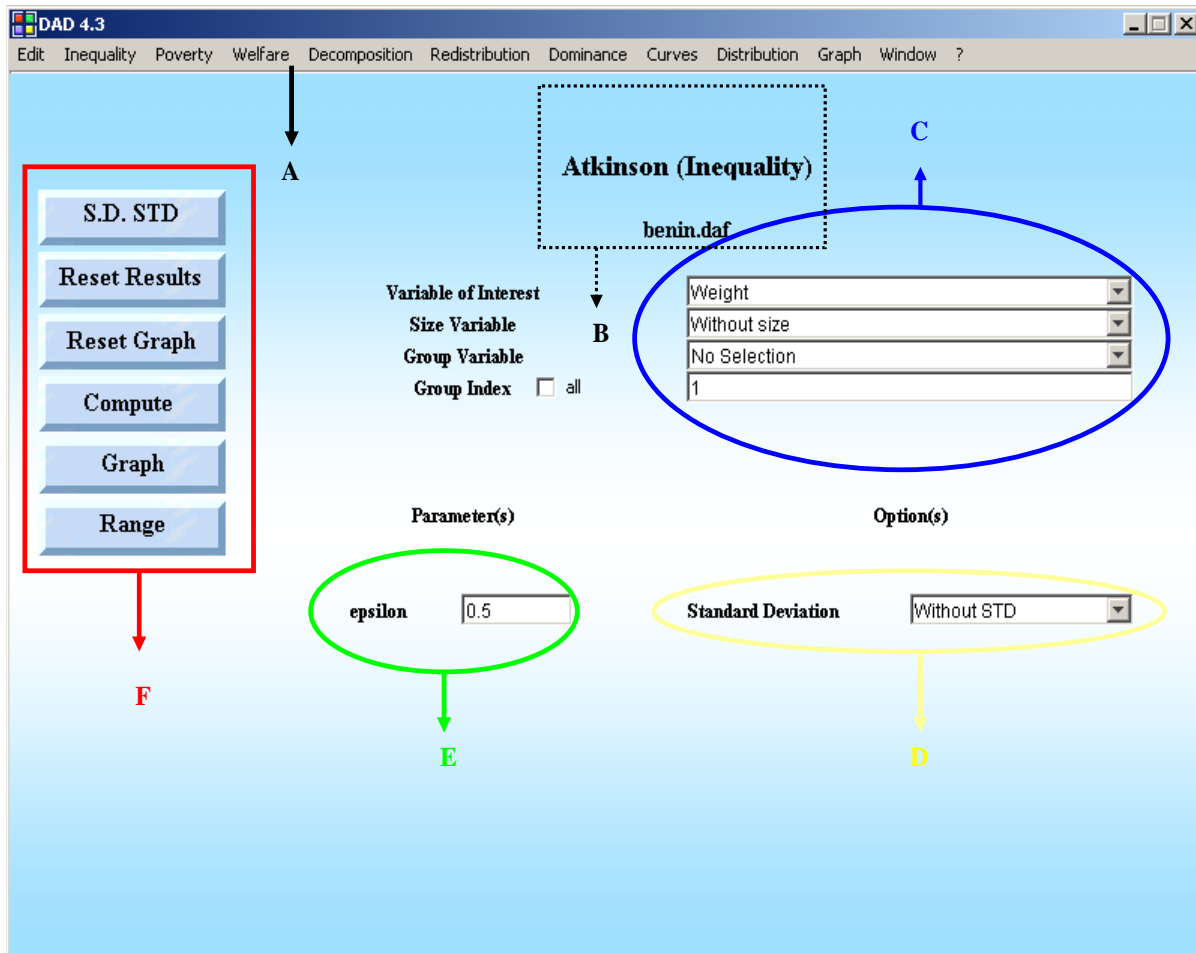
The Panels for this form are:

A: Main menu
B: The name of the application and the name of the file used
C: Set of variables and parameters to be chosen as:
  ➢ Choice of variable of interest.
  ➢ Choice of size variable.
  ➢ Choice of group variable.
  ➢ Choice of group number.
D: Option to compute with or without standard deviation.
E: Parameters to be specified.
F: Set of Commands for this application.

---

[1] See section : **Sampling Design.**

Most application windows, including that of Figure 2, are divided into three main panels. The first panel is used to specify the relevant database variables needed for the estimation. The second panel (generally at the bottom of the application window) specifies the parameter values and options to be used by the estimator – examples include the level of inequality aversion, the value of the poverty line and the percentile to be considered as well as whether indices should be normalized and whether statistical inference should be performed. The third panel activates buttons in order that various types of results may be generated. Some application windows can also generate popping-up dialogue boxes.

Variables:

The following basic variables are typically required for carrying out *DAD*'s computations.

Variable of interest: This is the variable that usually captures living standards. It can represent, for instance, expenditures per adult equivalent, calorie intake, normalized height-for-age scores for children, *etc*.

Size variable: This refers to the "ethical" of physical size of the observation. For the computation of many distributive statistics, we will indeed wish to take into account how many relevant individuals (or statistical units) are found in a given observation. We might, for instance, wish to estimate inequality across individuals or the proportion of children who are poor. Individuals and children will then be respectively the statistical units of interest. Households do differ, however, in their size or in the number of children they contain. *DAD* takes this into account through the use of the Size variable. When an observation represents a household, computing inequality across individuals requires specifying household size as the Size variable, whereas computing poverty among children requires putting the number of children in the household as the Size variable. If the statistics of interest were the proportion of households in poverty, then no Size variable would be needed.

Group variable: (This should be used in combination with Group number.) It is often useful to limit some distributive analysis to some population subgroup. We might for example wish to estimate poverty within a country's rural area or within the group of public workers. One way to do this is to set Size variable to zero for all of the observations that fall outside these groups of interest. Another way is by defining a Group variable whose values will allow DAD to identify which are the observations of interest.

Group number: Group number tells *DAD* on which value of the Group variable to condition the computation of some distributive statistics. The value for Group number should have been an integer. For example, rural households might be assigned a value of 1 for some variable denoted as $region$. Setting Group variable to $region$ and Group number to 1 makes *DAD* know that we wish the distributive statistics to be computed only within the group the rural households.

REMARK: For some applications, a checkbox appears to allow the user to choose the option "**All**" to compute an index for all groups.

Parameters:

Many distributives indices and curves depend on the level of a specific parameter. For example, the FGT index of poverty depends on the level of a parameter usually denoted as α. Another example is the Lorenz curve with its height depending on the level of a parameter p (a percentile). We can change the default level of such parameters and indicate any desired level.

Options:

- DAD allows the computation of the sampling distribution, via the estimation of the standard deviation (STD), of most distributive indices and curves. To compute an estimator with its STD, choose the option "**With STD**".
- For some applications, other options are also available. For example, for the poverty FGT application, the option "Normalised" allows one to choose if the FGT index should be normalised or not.

Commands:

S.D-STD: Starting with version 4.4 of DAD, one can, for some of the applications, compute confidence intervals and perform statistical tests by using standard or pivotal bootstrap approaches. For more information, see the section: **Standard deviation, confidence intervals and hypothesis testing.**

Reset Results: By clicking on this command, the earlier results computed by DAD are deleted.

**Reset Graphs:** By clicking on this command, the earlier curves drawn by DAD are deleted.

Compute: To perform the computation of the distributive index, press on the button **"Compute".**

Graph: This command draws a curve over a range of parameter values. For example, for the Lorenz application, a Lorenz curve can be drawn over a range of percentiles.

Range: By clicking on the button, a bar menu appears at the bottom of the window to allow the user to specify a range for the horizontal axis of the curve to be drawn. If you like to make this frame disappear, click again on this button.

DAD applications for two distributions can be launched after having loaded one or two databases. Each time an application that can support two distributions is launched, a dialog box, shown on figure 1, opens to make it possible for the user to indicate the desired number of distributions and also the name of the data bases if two distributions are to be used. The application windows for two distributions are very similar to that for only one, the only difference being the addition of a second panel to specify the relevant variables to be used for the second distribution. Generally

speaking, applications for two distributions serve to compute differences (as opposed to levels) between indices and curves.
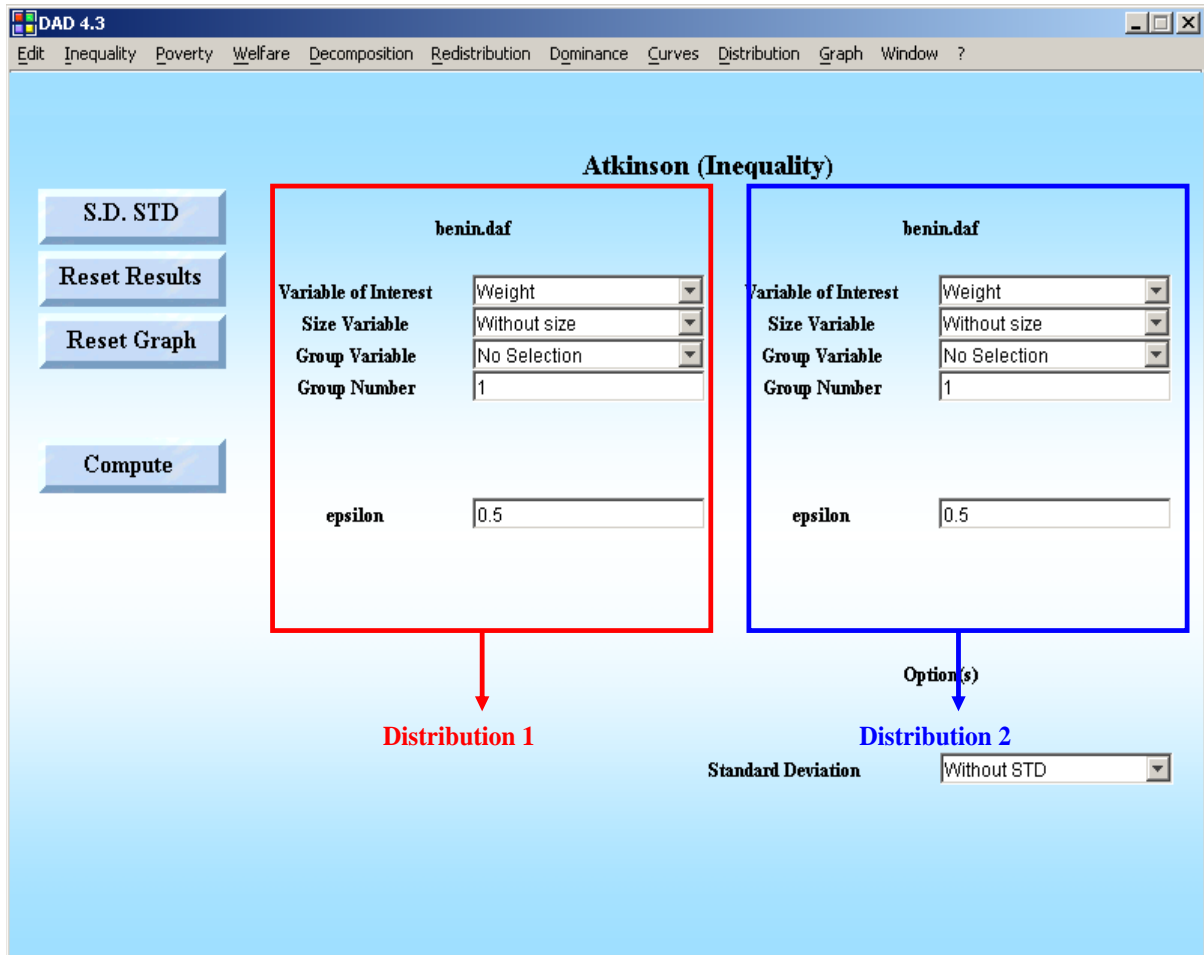
Using applications with two distributions can be useful when we have to compare between two distributions, such as when we wish
- to compute the difference between inequality or poverty across two periods (here, every period represents one distribution).
- to compute the difference between inequality or poverty across two groups (in this case, we use the same database for the two distributions, but we indicate a different group number for every distribution)

Example:

Suppose that you wish to compute the mean of a variable *y*, with y_i denoting the ith household observation. We call the vector *y* to be used the "Variable of Interest". The following table displays the observations of *y* for a sample of ten households. The vector of w_i (**"Sampling Weight variable"**) is the sampling weight to be applied to these observations and s_i is the size of household i – namely, the number of statistical units (typically individuals) whom we want to consider in our computations. We can also assign to each of these observations a code k_i that indicates the subgroup of the population to which the ith observation belongs. For example, code 1 may indicate that households live in town "V1" and code 2 that they live in town "V2":

Observation	y_i	k_i	w_i	s_i
i	Variable of interest	Group Variable	Sampling Weight variable	Size Variable
1	500	1	3	2
2	200	2	1	1
3	300	1	1	4
4	1000	1	2	5
5	700	2	3	5
6	450	1	1	7
7	300	1	1	3
8	200	2	3	3
9	300	2	2	4
10	400	1	1	8

The user has then no less than six possibilities for computing the mean, as shown in the following table. Note that we would normally always wish to ensure that sampling weights are incorporated in the computations – the estimators are weighted ones – and this is best done by initializing sampling weights in the Edit Sample Design window.

	Estimating the mean	Variable of Interest	Size Variable	Group Variable	Index of group
1	For the 10 households **Without size (mean across households)**	y_i	Without Size	No selection	1 (*)
2	For the 10 households **With size (mean across individuals)**	y_i	s_i	No selection	1 (*)
3	For households living in town V1 **Without size (mean across households)**	y_i	Without Size	k_i	1
4	For households living in town V1 **With size (mean across individuals)**	y_i	s_i	k_i	1
5	For households living in town V2 **Without size (mean across households)**	y_i	Without Size	k_i	2
6	For households living in town V2 **With size (mean across individuals)**	y_i	s_i	k_i	2

1- (*): This choice does not affect the results since no group variable was selected.
2- Consult the Sampling design section to know how sampling weights can be initialized.