USER MANUAL

DASP version 2.3

DASP: Distributive Analysis Stata Package

By

Abdelkrim Araar,
Jean-Yves Duclos

Université Laval
PEP, CIRPÉE and World Bank

June 2013

Table of contents

List of Figures

1 Introduction

The Stata software has become a very popular tool to transform and process data. It comes with a large number of basic data management modules that are highly efficient for transformation of large datasets. The flexibility of Stata also enables programmers to provide specialized ".ado" routines to add to the power of the software. This is indeed how *DASP* interacts with Stata. *DASP,* which stands for *Distributive Analysis Stata Package*, is mainly designed to assist researchers and policy analysts interested in conducting distributive analysis with Stata. In particular, *DASP* is built to:

- Estimate the most popular statistics (indices, curves) used for the analysis of poverty, inequality, social welfare, and equity;
- Estimate the differences in such statistics;
- Estimate standard errors and confidence intervals by taking full account of survey design;
- Support distributive analysis on more than one data base;
- Perform the most popular poverty and decomposition procedures;
- Check for the ethical robustness of distributive comparisons;
- Unify syntax and parameter use across various estimation procedures for distributive analysis.

For each *DASP* module, three types of files are provided:

***.ado: This file contains the program of the module**
***.hlp: This file contains help material for the given module**
***.dlg : This file allows the user to perform the estimation using the module's dialog box**

The *.dlg files in particular makes the *DASP* package very user friendly and easy to learn. When these dialog boxes are used, the associated program syntax is also generated and showed in the review window. The user can save the contents of this window in a *.do file to be subsequently used in another session.

2 *DASP* and Stata versions

DASP requires

- Stata version 10.0 or higher
- ado files must be updated

To update the executable file (from 10.0 to 10.2) and the ado files, see:
http://www.stata.com/support/updates/

3 Installing and updating the *DASP* package

In general, the *.ado files are saved in the following main directories:

Priority	Directory	Sources
1	UPDATES:	Official updates of Stata *.ado files
2	BASE:	*.ado files that come with the installed Stata software
3	SITE:	*.ado files downloaded from the net
4	PLUS:	..
5	PERSONAL:	Personal *.ado files

3.1 Installing DASP modules

a. Unzip the file ***dasp.zip*** in the directory c:
b. Make sure that you have ***c:/dasp/dasp.pkg*** or ***c:/dasp/stata.toc***
c. In the Stata command windows, type the syntax
   net from c:/dasp

Figure 1: Ouput of *net describe dasp*

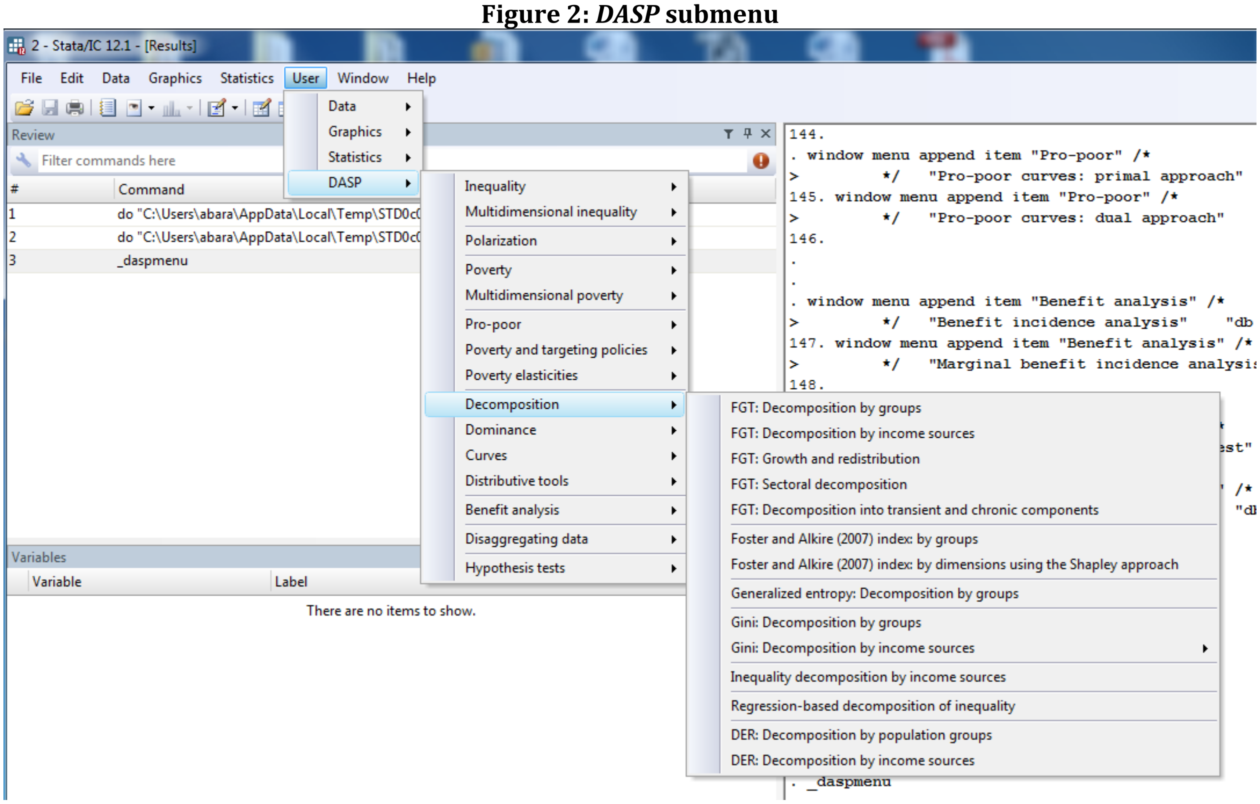
```
Version       : version   2.0
Date          : June 2009
Stata Version : Required 9.2 and higher
==================================================
Author:
DASP is conceived by:

Dr. Abdelkrim Araar  :  aabd@ecn.ulaval.ca
Dr. Jean-Yves Duclos : jyves@ecn.ulaval.ca

==================================================
Before using modules of this package, users have to:

update the executable Stata file  to Stata 9.2   or higher:
http://www.stata.com/support/updates/stata9.html

update the ado files:
http://www.stata.com/support/updates/stata9/ado/
==================================================
The two follwing sub-packages  must be  installed to run DASP.

==================================================

PACKAGES you could -net describe-:
    dasp_p1          Distributive Analysis Stata Package: PART  I
    dasp_p2          Distributive Analysis Stata Package: PART II
```

d. Type the syntax
   net install dasp_p1.pkg, force replace
   net install dasp_p2.pkg, force replace
   net install dasp_p3.pkg, force replace
   net install dasp_p4.pkg, force replace

3.2 *Adding the DASP submenu to Stata's main menu*

With Stata 9, sub menus can be added to the menu item **User**.

Figure 2: *DASP* submenu



To add the *DASP* sub menus, the file ***profile.do*** (which is provided with the *DASP* package) must be copied into the ***PERSONAL*** directory. If the file ***profile.do*** already exists, add the contents of the *DASP* –provided ***profile.do*** file into that existing file and save it. To check if the file ***profile.do*** already exists, type the command: *findfile profile.do .*

4 *DASP* and data files

DASP makes it possible to use simultaneously more than one data file. The user should, however, "initialize" each data file before using it with *DASP*. This initialization is done by:

1. Labeling variables and values for categorical variables;
2. Initializing the sampling design with the command *svyset*;
3. Saving the initialized data file.

Users are recommended to consult appendices A, B and C,

5 Main variables for distributive analysis

VARIABLE OF INTEREST. This is the variable that usually captures living standards. It can represent, for instance, income per capita, expenditures per adult equivalent, calorie intake, normalized height-for-age scores for children, or household wealth.

SIZE VARIABLE. This refers to the "ethical" or physical size of the observation. For the computation of many statistics, we will indeed wish to take into account how many relevant individuals (or statistical units) are found in a given observation.

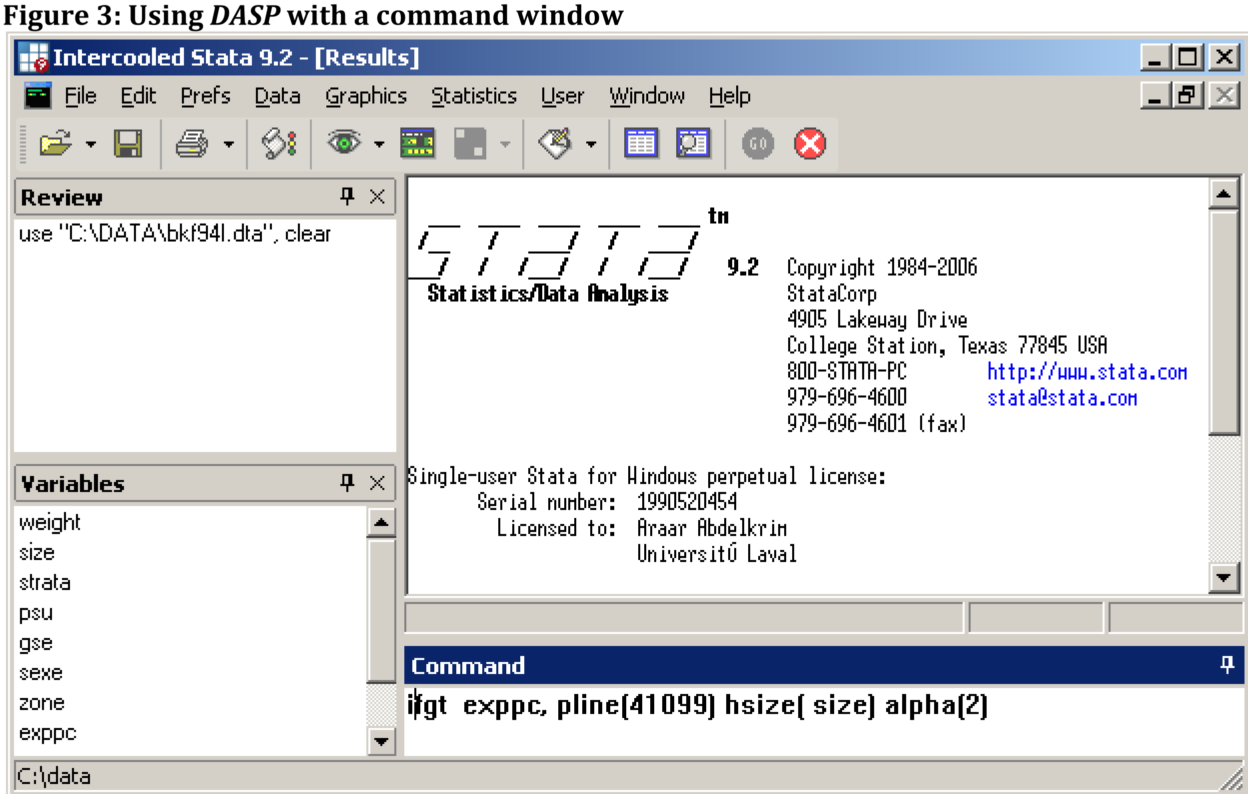
GROUP VARIABLE. (This should be used in combination with GROUP NUMBER.) It is often useful to focus one's analysis on some population subgroup. We might, for example, wish to estimate poverty within a country's rural area or within female-headed families. One way to do this is to force *DASP* to focus on a population subgroup defined as those for whom some GROUP VARIABLE (say, area of residence) equals a given GROUP NUMBER (say 2, for rural area).

SAMPLING WEIGHT. Sampling weights are the inverse of the sampling probability. This variable should be set upon the initialization of the dataset.

6 How can *DASP* commands be invoked?

Stata commands can be entered directly into a command window:

Figure 3: Using *DASP* with a command window



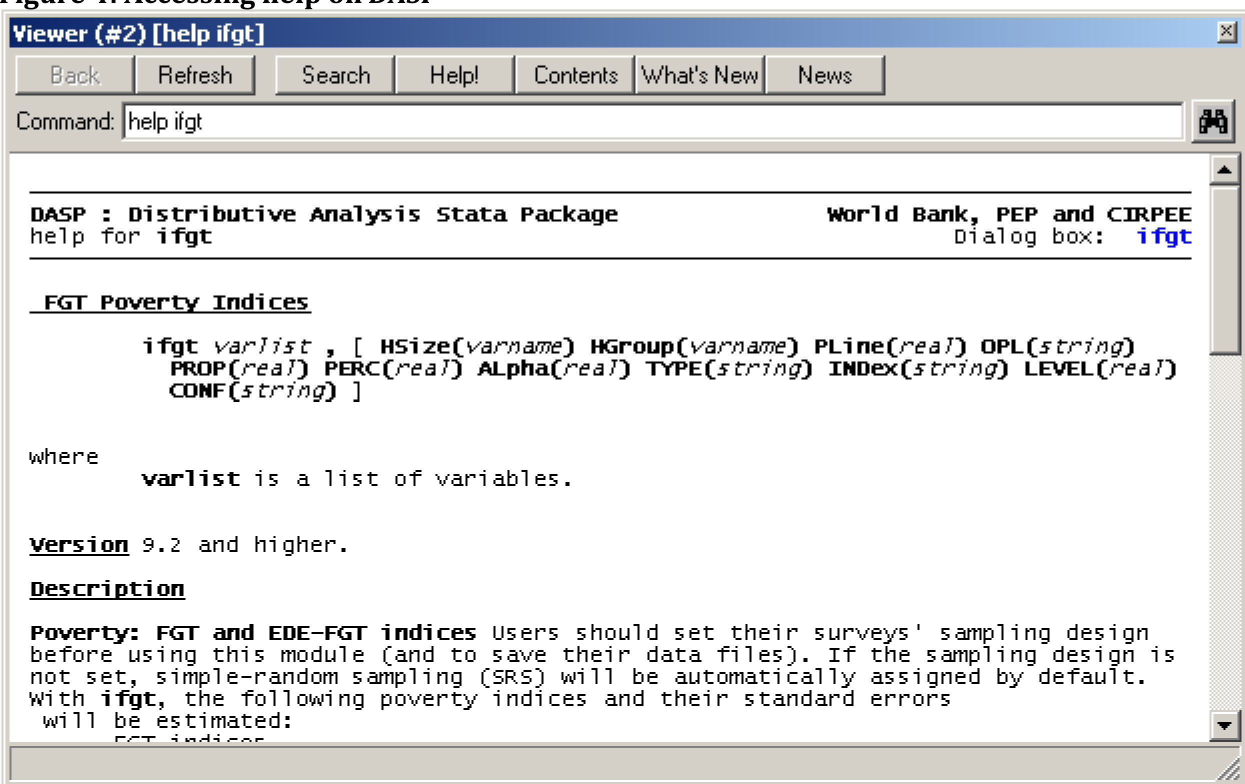
An alternative is to use dialog boxes. For this, the command *db* should be typed and followed by the name of the relevant *DASP* module.
Example:
db ifgt

7 How can help be accessed for a given *DASP* module?

Type the command *help* followed by the name of the relevant *DASP* module.

Example:
help ifgt

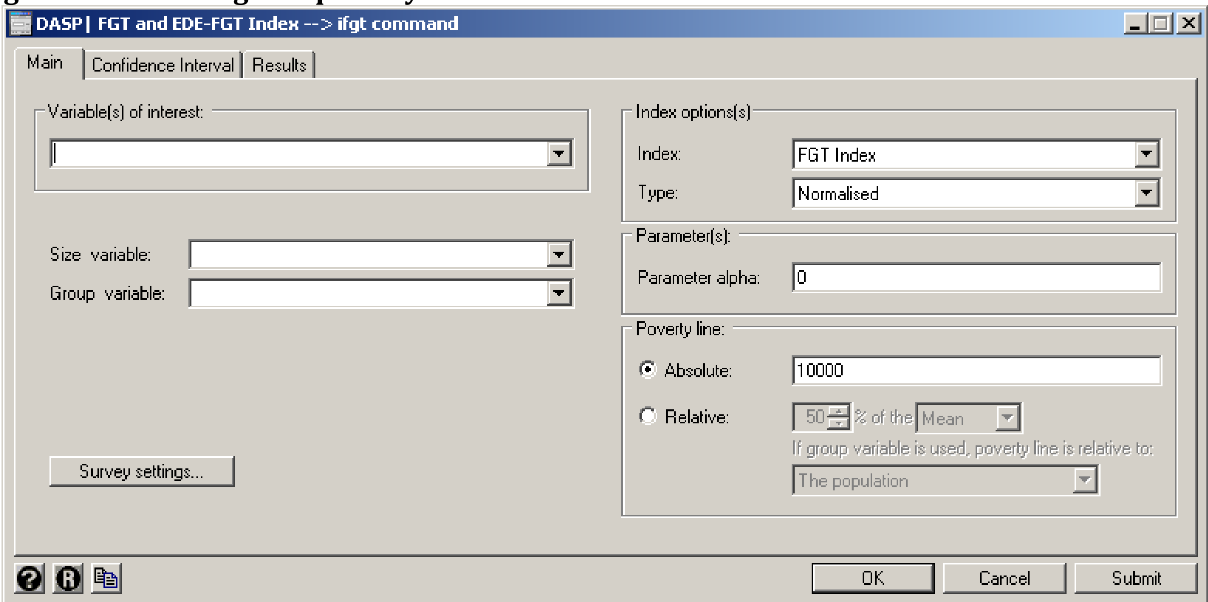
Figure 4: Accessing help on *DASP*



8 Applications and files in *DASP*

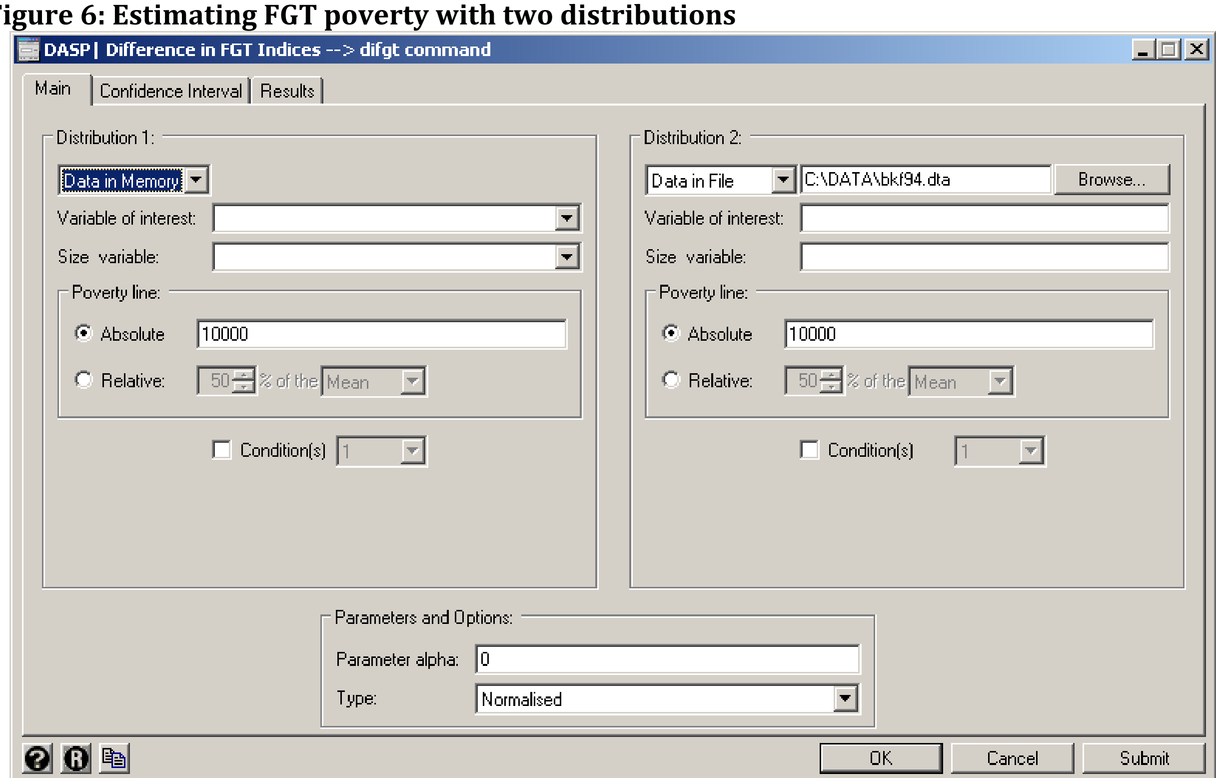
Two main types of applications are provided in *DASP*. For the first one, the estimation procedures require only one data file. In such cases, the data file in memory is the one that is used (or "loaded"); it is from that file that the relevant variables must be specified by the user to perform the required estimation.

Figure 5: Estimating FGT poverty with one distribution



For the second type of applications, two distributions are needed. For each of these two distributions, the user can specify the currently-loaded data file (the one in memory) or one saved on disk.

Figure 6: Estimating FGT poverty with two distributions



Notes:

1. *DASP* considers two distributions to be statistically dependent (for statistical inference purposes) if the same data set is used (the same loaded data or data with the same path and filename) for the two distributions.
2. If the option DATA IN FILE is chosen, the keyboard must be used to type the name of the required variables.

9 Basic Notation

The following table presents the basic notation used in *DASP*'s user manual.

Symbol	Indication
y	variable of interest
i	observation number
y_i	value of the variable of interest for observation i
hw	sampling weight
hw_i	sampling weight for observation i
hs	size variable
hs_i	size of observation i (for example the size of household i)
w_i	hw_i* hs_i
hg	group variable
hg_i	group of observation i.
w_i^k	sw_i^k=sw_i if hg_i = k, and 0 otherwise.
n	sample size

For example, the mean of *y* is estimated by *DASP* as $\hat{\mu}$:

$$\hat{\mu} = \frac{\sum_{i=1}^{n} w_i\, y_i}{\sum_{i=1}^{n} w_i}$$

10 *DASP* and poverty indices

10.1 FGT and EDE-FGT poverty indices (ifgt).

The non-normalised Foster-Greer-Thorbecke or FGT index is estimated as

$$\hat{P}(z;\alpha) = \frac{\sum_{i=1}^{n} w_i\,(z-y_i)_+^{\alpha}}{\sum_{i=1}^{n} w_i}$$

where z is the poverty line and $x_+ = \max(x,0)$. The usual normalised FGT index is estimated as

$$\hat{\bar{P}}(z;\alpha) = \hat{P}(z;\alpha)/(z)^{\alpha}$$

The EDE-FGT index is estimated as:

$$\widehat{EDE}(P(z;\alpha)) = \left(\hat{P}(z;\alpha)\right)^{1/\alpha} \quad for \quad \alpha > 0$$

- There exist three ways of fixing the poverty line:

  1- Setting a deterministic poverty line;
  2- Setting the poverty line to a proportion of the mean;
  3- Setting the poverty line to a proportion of a quantile *Q(p)*.

- The user can choose the value of parameter α.

- The user can select more than one variable of interest simultaneously. For example, one can estimate poverty by using simultaneously *per capita* consumption and *per capita* income.

- A group variable can be used to estimate poverty at the level of a categorical group. If a group variable is selected, only the first variable of interest is then used.

- Standard errors and confidence intervals with a confidence level of 95% are provided. Both the type of confidence intervals provided and the level of confidence used can be changed.

- The results are displayed with 6 decimals; this can be changed.

Interested users are encouraged to consider the exercises that appear in Section **23.1**

10.2 Difference between FGT indices (difgt)

This module estimates differences between the FGT indices of two distributions.

For each of the two distributions:

- There exist three ways of fixing the poverty line:
  1- Setting a deterministic poverty line;
  2- Setting the poverty line to a proportion of the mean;
  3- Setting the poverty line to a proportion of a quantile *Q(p)*
- One variable of interest should be selected.
- Conditions can be specified to focus on specific population subgroups.
- Standard errors and confidence intervals with a confidence level of 95% are provided. Both the type of confidence intervals provided and the level of confidence used can be changed.
- The results are displayed with 6 decimals; this can be changed.
- A level for the parameter α can be chosen for each of the two distributions.

Interested users are encouraged to consider the exercises that appear in Section **23.2.**

10.3 Watts poverty index (iwatts).

The Watts poverty index is estimated as

$$\hat{P}(z) = \frac{\sum_{i=1}^{q} w_i \,(ln(z / y_i))}{\sum_{i=1}^{n} w_i}$$

where z is the poverty line and q the number of poor.

- The user can select more than one variable of interest simultaneously. For example, one can estimate poverty by using simultaneously *per capita* consumption and *per capita* income.

- A group variable can be used to estimate poverty at the level of a categorical group. If a group variable is selected, only the first variable of interest is then used.

- Standard errors and confidence intervals with a confidence level of 95% are provided. Both the type of confidence intervals provided and the level of confidence used can be changed.

- The results are displayed with 6 decimals; this can be changed.

10.4 Difference between Watts indices (diwatts)

This module estimates differences between the Watts indices of two distributions.

For each of the two distributions:

- One variable of interest should be selected.
- Conditions can be specified to focus on specific population subgroups.
- Standard errors and confidence intervals with a confidence level of 95% are provided. Both the type of confidence intervals provided and the level of confidence used can be changed.
- The results are displayed with 6 decimals; this can be changed.
- A level for the parameter α can be chosen for each of the two distributions.

10.5 Sen-Shorrocks-Thon poverty index (isst).

The Sen-Shorroks-Thon poverty index is estimated as:

$$\hat{P}(z) = HP^*(z,\alpha)[1+G_g^*]$$

where z is the poverty line H is the headcount, $P^*(z,\alpha)$ the poverty gap estimated at the level of poor group and G_g^* the Gini index of poverty gaps $(z-y)_+/z$.

- The user can select more than one variable of interest simultaneously. For example, one can estimate poverty by using simultaneously *per capita* consumption and *per capita* income.

- A group variable can be used to estimate poverty at the level of a categorical group. If a group variable is selected, only the first variable of interest is then used.

- Standard errors and confidence intervals with a confidence level of 95% are provided. Both the type of confidence intervals provided and the level of confidence used can be changed.

- The results are displayed with 6 decimals; this can be changed.

10.6 Difference between Sen-Shorrocks-Thon indices (disst)

This module estimates differences between the Watts indices of two distributions.

For each of the two distributions:

- One variable of interest should be selected.
- Conditions can be specified to focus on specific population subgroups.
- Standard errors and confidence intervals with a confidence level of 95% are provided. Both the type of confidence intervals provided and the level of confidence used can be changed.
- The results are displayed with 6 decimals; this can be changed.
- A level for the parameter α can be chosen for each of the two distributions.

10.7 DASP and multidimensional poverty indices

The general form of an additive multidimensional poverty index is:

$$P(X,Z)=\frac{\sum_{i=1}^{n} w_i\, p(X_i,Z)}{\sum_{i=1}^{n} w_i}$$

where $p(X_i,Z)$ is individual *I*'s poverty function (with vector of attributes $X_i=\left(x_{i,1},\ldots,x_{i,J}\right)$ and vector of poverty lines $Z=\left(z_1,\ldots,z_J\right)$), determining *I*'s contribution to total poverty $P(X,Z)$.

[1] Chakravarty et al (1998) index **(imdp_cmr)**

$$p(X_i,Z)=\sum_{j=1}^{J} a_j\left(\frac{z_j-x_{i,j}}{z_j}\right)_+^{\alpha}$$

[2] Extended Watts index **(imdp_ewi)**

$$p(X_i,Z)=\sum_{j=1}^{J} a_j \ln\left(\frac{z_j}{\min(z_j;x_{i,j})}\right)$$

[3] Multiplicative extended FGT index **(imdp_mfi)**

$$p(X_i,Z)=\prod_{j=1}^{J}\left(\frac{z_j-x_{i,j}}{z_j}\right)_+^{\alpha_j}$$

[4] Tsui (2002) index **(imdp_twu)**

$$p(X_i,Z)=\prod_{j=1}^{J}\left(\frac{z_j}{\min(z_j;x_{i,j})}\right)^{b_j}-1$$

[5] Intersection headcount index **(imdp_ihi)**

$$p(X_i,Z)=\prod_{j=1}^{J} I\left(z_j>x_{i,j}\right)$$

[6] Union headcount index **(imdp_uni)**

$$p(X_i,Z)=1-\prod_{j=1}^{J} I\left(z_j<x_{i,j}\right)$$

[7] Bourguignon and Chakravarty bi-dimensional (2003) index **(imdp_bci)**

$$p(X_i,Z)=\left[C_1+\beta^{\gamma/\alpha}C_2\right]^{\alpha/\gamma}$$

where:

$$C_{1=}\left(\frac{z_1-x_{i,1}}{z_1}\right)_+^{\gamma} \text{ and } C_{2=}\left(\frac{z_2-x_{i,2}}{z_2}\right)_+^{\gamma}$$

[8] Alkire and Foster (2011) index **(imdp_afi)**

$$p(\alpha,X_i,Z)=\frac{1}{N}\sum_{i=}^{N}\frac{1}{J}\sum_{j=1}^{J}w_j\left(\frac{z_j-x_{i,j}}{z_j}\right)_+^{\alpha}I(d_i\geq d_c)$$

where $\sum_{j=1}^{J}w_j=J$ *and* d_i denotes the number of dimensions in which the individual i is deprived. d_c denotes the normative dimensional cut-off.

The modules presented above can be used to estimate the multidimensional poverty indices as well as their standard errors.

- The user can select among the seven multidimensional poverty indices.
- The number of dimensions can be selected (1 to 10).
- If applicable, the user can choose parameter values relevant to a chosen index.
- A group variable can be used to estimate the selected index at the level of a categorical group.
- Standard errors and confidence intervals with a confidence level of 95% are provided. Both the type of confidence intervals provided and the level of confidence used can be changed.
- The results are displayed with 3 decimals; this can be also changed.

Users are encouraged to consider the exercises that appear in Section **23.3**

10.8 Multiple overlapping deprivation analysis (MODA) indices

The **imoda** DASP module produces a series of multidimensional poverty indices in order to show the incidence of deprivation in each dimension. Further, this application estimates the incidence of multi-deprivation in the different combinations of dimensions. In this application, the number of dimensions is set to three. Further, the multidimensional poverty is measured by the headcount (union and intersection headcount indices) and the Alkire and Foster (2011) M0 index for different levels of the dimensional cut-off.

- The number of dimensions is three.
- A group variable can be used to estimate the MODA indices at the level of a categorical group.
- Standard errors and confidence intervals with a confidence level of 95% are provided. Both the type of confidence intervals provided and the level of confidence used can be changed.
- The results are displayed with 3 decimals; this can be also changed.

11 *DASP, poverty* and targeting policies

11.1 Poverty and targeting by population groups

The per-capita dollar impact of a marginal addition of a constant amount of income to everyone within a group k – called Lump-Sum Targeting (LST) – on the FGT poverty index $P(k,z;\alpha)$, is as follows:

$$LST = \begin{cases} -\alpha P(k,z;\alpha-1) & \text{if } \alpha \geq 1 \\ -f(k,z) & \text{if } \alpha = 0 \end{cases}$$

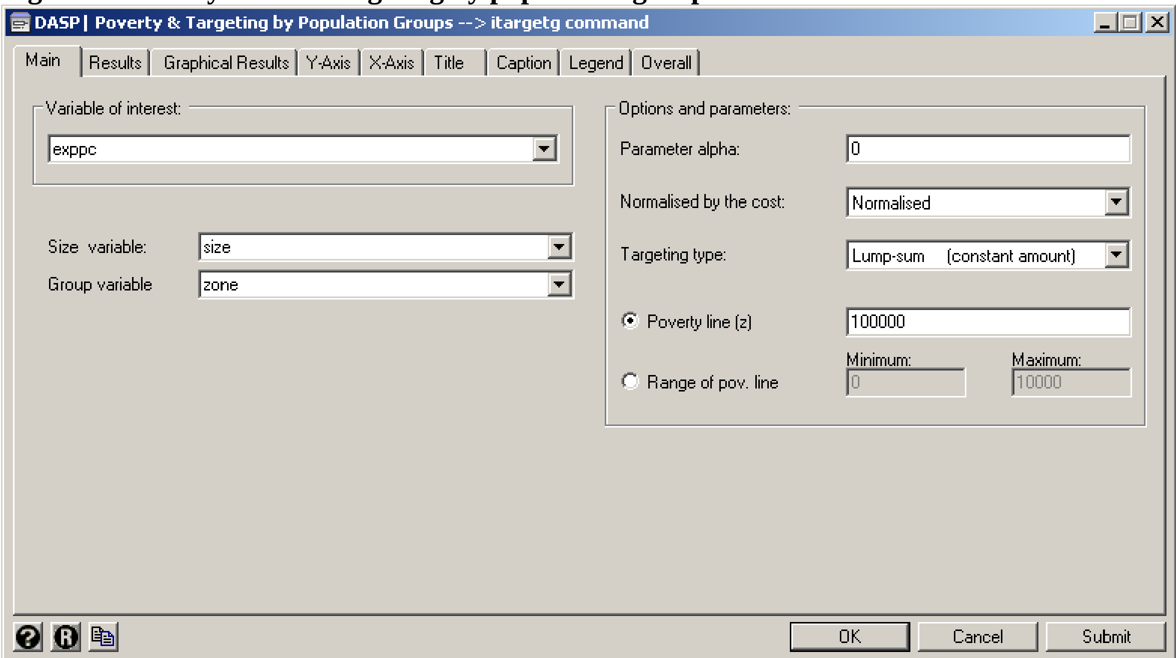
where z is the poverty line, k is the population subgroup for which we wish to assess the impact of the income change, and $f(k,z)$ is the density function of the group k at level of income z. The per-capita dollar impact of a proportional marginal variation of income within a group k, called Inequality Neutral Targeting, on the FGT poverty index $P(k,z;\alpha)$ is as follows:

$$INT = \begin{cases} \alpha \dfrac{P(k,z;\alpha) - zP(k,z;\alpha-1)}{\mu(k)} & \text{if } \alpha \geq 1 \\ -\dfrac{zf(k,z)}{\mu(k)} & \text{if } \alpha = 0 \end{cases}$$

The module **itargetg** allows to:

- Estimate the impact of marginal change in income of the group on poverty of the group and that of the population;
- Select the design of change, constant or proportional to income to keep inequality unchanged;
- Draw curves of impact according for a range of poverty lines;
- Draw the confidence interval of impact curves or the lower or upper bound of confidence interval;
- Etc.

Figure 7: Poverty and the targeting by population groups



Reference:

DUCLOS, J.-Y. AND A. ARAAR (2006): Poverty and Equity Measurement, Policy, and Estimation with DAD, Berlin and Ottawa: Springer and IDRC. (sec. 12.1)

11.2 Poverty and targeting by income components

Proportional change per 100% of component

Assume that total income Y is the sum of J income components, with $Y=\sum_{j=1}^{J}\lambda_j y_j$ and where c is a factor that multiplies income component y_j and that can be subject to growth. The derivative of the normalized FGT index with respect to λ_j is given by

$$\left.\frac{\partial \bar{P}(z,\alpha)}{\partial \lambda_j}\right|_{\lambda_j=1,\, j=1\cdots J} = -CD_j(z,\alpha)$$

where CD_j is the Consumption dominance curve of component j.

Change per $ of component

The per-capita dollar impact of growth in the j^{th} component on the normalized FGT index of the k^{th} group is as follows:

$$\frac{\dfrac{\partial P(z,\alpha)}{\partial y^j}}{\dfrac{\partial \mu_j}{\partial y^j}} = -\overline{CD}^j(z,\alpha)$$

where $\overline{CD}^j$ is the normalized consumption dominance curve of the component j.

Constant change per component

Simply we assume that the change concerns the group with component level greater than zero. Thus, this is similar to targeting by the nonexclusive population groups.

The module **itargetc** allows to:

- Estimate the impact of marginal change in income component on poverty;
- Select the option normalised or non normalised by the average of component;
- Select the design of change, constant (lump sum) or proportional to income to keep inequality unchanged;
- Draw curves of impact according for a range of poverty lines;
- Draw the confidence interval of impact curves or the lower or upper bound of confidence interval;
- Etc.

Reference:

DUCLOS, J.-Y. AND A. ARAAR (2006): Poverty and Equity Measurement, Policy, and Estimation with DAD, Berlin and Ottawa: Springer and IDRC. (sec. 12)

12 Marginal poverty impacts and poverty elasticities

12.1 FGT elasticity's with respect to the average income growth (efgtgr).

The overall growth elasticity (GREL) of poverty, when growth comes exclusively from growth within a group k (namely, within that group, inequality neutral), is given by:

$$GREL = \begin{cases} -\dfrac{zf(k,z)}{F(z)} & if \quad \alpha = 0 \\ \alpha \dfrac{\bar{P}(k,z;\alpha) - \bar{P}(k,z;\alpha-1)}{\bar{P}(z,\alpha)} & if \quad \alpha \geq 1 \end{cases}$$

where z is the poverty line, k is the population subgroup in which growth takes place, $f(k,z)$ is the density function at level of income z of group k, and $F(z)$ is the headcount.

Araar, Abdelkrim and Jean-Yves Duclos, (2007), Poverty and inequality components: a micro framework, **Working Paper: 07-35**. CIRPEE, Department of Economics, Université Laval.
Kakwani, N. (1993) "Poverty and economic growth with application to Côte D'Ivoire", Review of Income and Wealth, 39(2): 121:139.

- To estimate the FGT elasticity's with respect average income growth the group or the whole population;
- The user can select more than one variable of interest simultaneously. For example, one can estimate poverty by using simultaneously *per capita* consumption and *per capita* income.
- A group variable can be used to estimate poverty at the level of a categorical group. If a group variable is selected, only the first variable of interest is then used.
- Standard errors and confidence intervals with a confidence level of 95% are provided. Both the type of confidence intervals provided and the level of confidence used can be changed.
- The results are displayed with 6 decimals; this can be changed.

12.2 FGT elasticities with respect to average income growth with different approaches (efgtgro).

The overall growth elasticity of poverty is estimated using one approach among the following list:

- The counterfactual approach;
- The marginal approach;
- The parameterized approach;
- The numerical approach;

The module ***efgtgro*** allows the estimation of a poverty elasticity or semi-elasticity with respect to growth with the different approaches mentioned above. For more details on these approaches, see:

Abdelkrim Araar, 2012. "**Expected Poverty Changes with Economic Growth and Redistribution**,"Cahiers de recherche 1222, CIRPEE.

- To estimate a FGT elasticity –semi-elasticity- with respect to average income growth in a group or in an entire population;
- A group variable can be used to estimate poverty at the level of a categorical group. If a group variable is selected, only the first variable of interest is then used.
- The results are displayed with 6 decimals; this can be changed.

12.3 FGT elasticity with respect to Gini inequality (efgtineq).

The overall growth elasticity (INEL) of poverty, when growth comes exclusively from change in inequality within a group k, is given by:

$$INEL = \begin{cases} \dfrac{\phi(k)\, f(k,z)\big(\mu(k)-z\big)}{F(z)} / \dfrac{\phi(k)\mu(k)}{\mu}\dfrac{C(k)}{I} & if \quad \alpha = 0 \\ \alpha\dfrac{\bar{P}(k,z;\alpha)-\big(\big(\mu(k)-z\big)/z\big)\bar{P}(k,z;\alpha-1)}{\bar{P}(z,\alpha)} / \dfrac{\phi(k)\mu(k)}{\mu}\dfrac{C(k)}{I} & if \quad \alpha \geq 1 \end{cases}$$

where z is the poverty line, k is the population subgroup in which growth takes place, $f(k,z)$ is the density function at level of income z for group k, and $F(z)$ is the headcount. $C(k)$ is the concentration coefficient of group k when incomes of the complement group are replaced by $\mu(k)$. I denotes the Gini index.

Araar, Abdelkrim and Jean-Yves Duclos, (2007), Poverty and inequality components: a micro framework, **Working Paper: 07-35**. CIRPEE, Department of Economics, Université Laval.

Kakwani, N. (1993) "Poverty and economic growth with application to Côte D'Ivoire", Review of Income and Wealth, 39(2): 121:139.

- To estimate a FGT elasticity with respect to average income growth in a group or in an entire population;
- The user can select more than one variable of interest simultaneously. For example, one can estimate poverty by using simultaneously *per capita* consumption and *per capita* income.
- A group variable can be used to estimate poverty at the level of a categorical group. If a group variable is selected, only the first variable of interest is then used.
- Standard errors and confidence intervals with a confidence level of 95% are provided. Both the type of confidence intervals provided and the level of confidence used can be changed.
- The results are displayed with 6 decimals; this can be changed.

12.4 FGT elasticity with respect to Gini-inequality with different approaches (efgtine).

The overall Gini-inequality elasticity of poverty can be estimated by using one approach among the following list:

- The counterfactual approach;
- The marginal approach;
- The parameterized approach;
- The numerical approach;
-

The module ***efgtine*** allows the estimation of a poverty elasticity or semi-elasticity with respect to inequality with the different approaches mentioned above. For more details on these approaches, see:

Abdelkrim Araar, 2012. "**Expected Poverty Changes with Economic Growth and Redistribution**,"Cahiers de recherche 1222, CIRPEE.

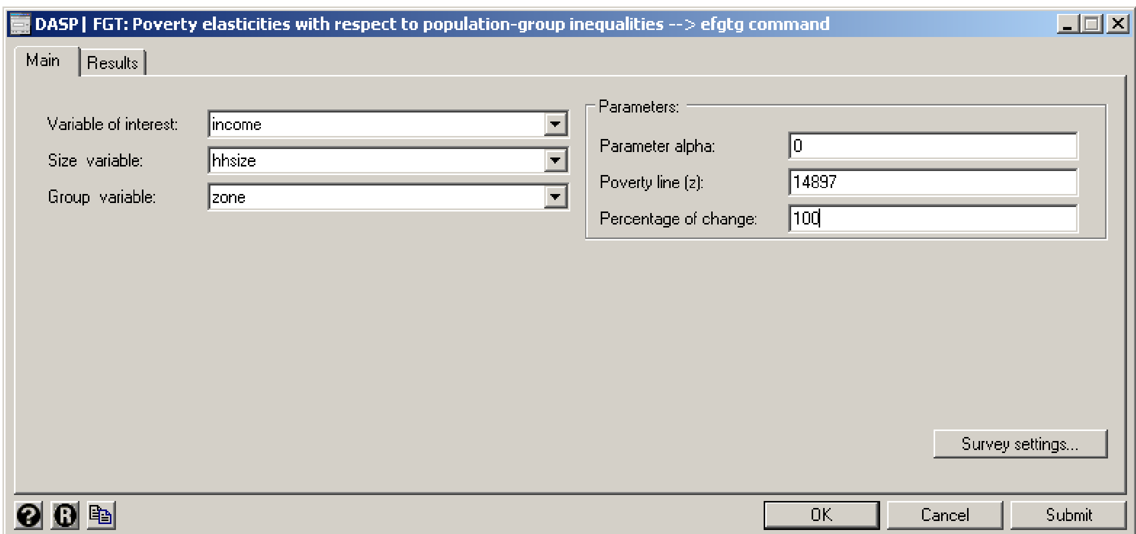
- To estimate a FGT elasticity –semi-elasticity- with respect to inequality;
- A group variable can be used to estimate poverty at the level of a categorical group. If a group variable is selected, only the first variable of interest is then used.
- The results are displayed with 6 decimals; this can be changed.

12.5 FGT elasticities with respect to within/between group components of inequality (efgtg).

This module estimates the marginal FGT impact and FGT elasticity with respect to within/between group components of inequality. A group variable must be provided. This module is mostly based on Araar and Duclos (2007):

> Araar, Abdelkrim and Jean-Yves Duclos, (2007), Poverty and inequality components: a micro framework, **Working Paper: 07-35**. CIRPEE, Department of Economics, Université Laval.

To open the dialog box of this module, type the command **db efgtg.**



After clicking on SUBMIT, the following should be displayed:

```
efgtg income, hgroup(zone) hsize(hhsize) alpha(0) pline(14897) prc(1) dec(3)
```

Poverty and Inequality Indices

Indices	Estimate
FGT	0.585
Gini	0.617

Marginal Impact & Elasticities By Groups

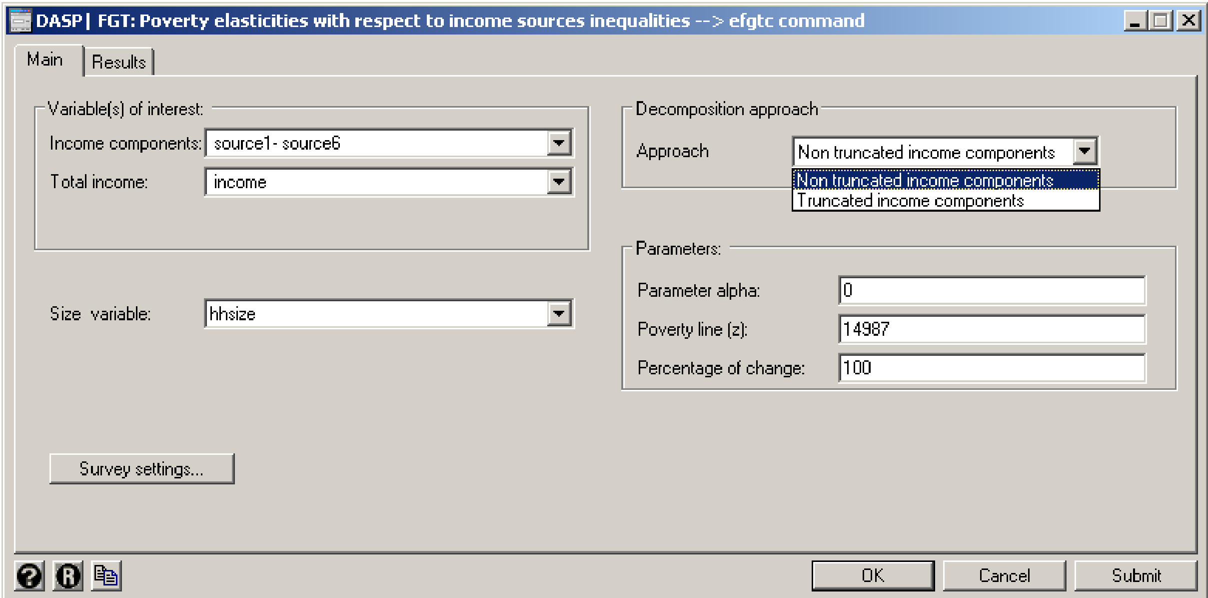
	Group	Population Share	Marginal Impact on Ineq.	Marginal Impact on Pov.	Elasticity
σ(g)	1: South south	0.150	0.110	0.036	0.350
	2: South east	0.119	0.076	0.014	0.196
	3: South west	0.194	0.138	0.077	0.588
	4: North central	0.139	0.065	0.005	0.078
	5: North east	0.135	0.060	0.002	0.029
	6: North west	0.263	0.090	-0.008	-0.090
σ	Within	.	0.539	0.126	0.247
γ	Between	.	0.051	0.051	1.045
λ	Population	1.000	0.617	0.135	0.231

12.6 FGT elasticities with respect to within/between income components of inequality (efgtc).

This module estimates the marginal FGT impact and FGT elasticity with respect to within/between income components of inequality. A list of income components must be provided. This module is mostly based on Araar and Duclos (2007):

> Araar, Abdelkrim and Jean-Yves Duclos, (2007), Poverty and inequality components: a micro framework, **Working Paper: 07-35**. CIRPEE, Department of Economics, Université Laval.

To open the dialog box of this module, type the command **db efgtc.**



After clicking on SUBMIT, the following should be displayed:

```
efgtc source1- source6, tot(income) hsize(hhsize) alpha(0) pline(14897) prc(1)
```

Poverty and Inequality Indices

Indices	Estimate
FGT	0.584667
Gini	0.616503

Marginal Impacts & Elasticities of poverty with respect to the within/between inequality in income components

	Source	Income Share	Impact on Inequality	Impact on Poverty	Elasticity
	1: source1	0.352966	0.265888	0.097233	0.385605
	2: source2	0.199865	0.063585	-0.032419	-0.537610
η(k)	3: source3	0.023731	0.012489	0.002508	0.211784
	4: source4	0.344093	0.229384	0.067828	0.311798
	5: source5	0.024588	0.013828	0.002247	0.171358
	6: source6	0.054758	0.031356	0.005368	0.180533
λ	Within	.	0.616503	0.134793	0.230546
τ	Between	.	0.049948	0.041846	0.883417

In case one is interested in changing some income component only among those individuals that are effectively active in some economic sectors (schemes $\eta^*(k), \tau^*$ and λ^* in the paper mentioned above), the user should select the approach "Truncated income component".

13 *DASP* and inequality indices

13.1 Gini and concentration indices (igini)

The Gini index is estimated as

$$\hat{I} = 1 - \frac{\hat{\xi}}{\hat{\mu}}$$

where

$$\hat{\xi} = \sum_{i=1}^{n} \left[\frac{(V_i)^2 - (V_{i+1})^2}{[V_1]^2} \right] y_i \quad \text{and} \quad V_i = \sum_{h=i}^{n} w_h \quad \text{and} \quad y_1 \geq y_2 \geq \cdots y_{n-1} \geq y_n \, .$$

The concentration index for the variable *T* when the ranking variable is *Y* is estimated as

$$\widehat{IC}_T = 1 - \frac{\hat{\xi}_T}{\hat{\mu}_T}$$

where $\hat{\mu}_T$ is the average of variable *T*,

$$\hat{\xi}_T = \sum_{i=1}^{n} \left[\frac{(V_i)^2 - (V_{i+1})^2}{[V_1]^2} \right] t_i$$

and where $V_i = \sum_{h=i}^{n} w_h$ and $y_1 \geq y_2 \geq \cdots y_{n-1} \geq y_n$.

- The user can select more than one variable of interest simultaneously. For example, one can estimate inequality, for instance by using simultaneously *per capita* consumption and *per capita* income.
- To estimate a concentration index, the user must select a ranking variable..
- A group variable can be used to estimate inequality at the level of a categorical group. If a group variable is selected, only the first variable of interest is then used.
- Standard errors and confidence intervals with a confidence level of 95% are provided. Both the type of confidence intervals provided and the level of confidence used can be changed.
- The results are displayed with 6 decimals; this can be changed.

Interested users are encouraged to consider the exercises that appear in Section **23.9**

13.2 Difference between Gini/concentration indices (digini)

This module estimates differences between the Gini/concentration indices of two distributions.

For each of the two distributions:

- One variable of interest should be selected;
- To estimate a concentration index, a ranking variable must be selected;
- Conditions can be specified to focus on specific population subgroups;

- Standard errors and confidence intervals with a confidence level of 95% are provided. Both the type of confidence intervals provided and the level of confidence used can be changed.
- The results are displayed with 6 decimals; this can be changed.

13.3 Generalised entropy index (ientropy)

The generalized entropy index is estimated as

$$\hat{I}(\theta)=\begin{cases} \dfrac{1}{\theta(\theta-1)\sum\limits_{i=1}^{n} w_i}\sum\limits_i w_i\left[\left(\dfrac{y_i}{\hat{\mu}}\right)^{\theta}-1\right] & \text{if } \theta\neq 0,1 \\ \dfrac{1}{\sum\limits_{i=1}^{n} w_i}\sum\limits_i w_i\log\left(\dfrac{\hat{\mu}}{y_i}\right) & \text{if } \theta=0 \\ \dfrac{1}{\sum\limits_{i=1}^{n} w_i}\sum\limits_i \dfrac{w_i y_i}{\hat{\mu}}\log\left(\dfrac{y_i}{\hat{\mu}}\right) & \text{if } \theta=1 \end{cases}$$

- The user can select more than one variable of interest simultaneously. For example, one can estimate inequality simultaneously for *per capita* consumption and for *per capita* income.
- A group variable can be used to estimate inequality at the level of a categorical group. If a group variable is selected, only the first variable of interest is then used.
- Standard errors and confidence intervals with a confidence level of 95% are provided. Both the type of confidence intervals provided and the level of confidence used can be changed.
- The results are displayed with 6 decimals; this can be changed.

13.4 Difference between generalized entropy indices (diengtropy)

This module estimates differences between the generalized entropy indices of two distributions.

For each of the two distributions:

- One variable of interest should be selected;
- Conditions can be specified to focus on specific population subgroups;
- Standard errors and confidence intervals with a confidence level of 95% are provided. Both the type of confidence intervals provided and the level of confidence used can be changed.
- The results are displayed with 6 decimals; this can be changed.

13.5 *Atkinson index (iatkinson)*

Denote the Atkinson index of inequality for the group k by $I(\varepsilon)$. It can be expressed as follows:

$$\hat{I}(\varepsilon) = \frac{\hat{\mu} - \hat{\xi}(\varepsilon)}{\hat{\mu}} \quad \text{where} \quad \hat{\mu} = \frac{\sum_{i=1}^{n} w_i\, y_i}{\sum_{i=1}^{n} w_i}$$

The Atkinson index of social welfare is as follows:

$$\hat{\xi}(\varepsilon) = \begin{cases} \left[\dfrac{1}{\sum_{i=1}^{n} w_i} \sum_{i=1}^{n} w_i\, (y_i)^{1-\varepsilon} \right]^{\frac{1}{1-\varepsilon}} & \rightarrow \text{if } \varepsilon \neq 1 \text{ and } \varepsilon \geq 0 \\ \\ \mathrm{Exp}\left[\dfrac{1}{\sum_{i=1}^{n} w_i} \sum_{i=1}^{n} w_i\, \ln(y_i) \right] & \rightarrow \varepsilon = 1 \end{cases}$$

- The user can select more than one variable of interest simultaneously. For example, one can estimate inequality simultaneously for *per capita* consumption and for *per capita* income.
- A group variable can be used to estimate inequality at the level of a categorical group. If a group variable is selected, only the first variable of interest is then used.
- Standard errors and confidence intervals with a confidence level of 95% are provided. Both the type of confidence intervals provided and the level of confidence used can be changed.
- The results are displayed with 6 decimals; this can be changed.

13.6 *Difference between Atkinson indices (diatkinson)*

This module estimates differences between the Atkinson indices of two distributions.
For each of the two distributions:

- One variable of interest should be selected;
- Conditions can be specified to focus on specific population subgroups;
- Standard errors and confidence intervals with a confidence level of 95% are provided. Both the type of confidence intervals provided and the level of confidence used can be changed.
- The results are displayed with 6 decimals; this can be changed.

13.7 Coefficient of variation index (icvar)

Denote the coefficient of variation index of inequality for the group k by CV. It can be expressed as follows:

$$\widehat{CV} = \left[\frac{\sum_{i=1}^{n} w_i y_i^2 / \sum_{i=1}^{n} w_i - \hat{\mu}^2}{\hat{\mu}^2} \right]^{\frac{1}{2}}$$

- The user can select more than one variable of interest simultaneously. For example, one can estimate inequality simultaneously for *per capita* consumption and for *per capita* income.
- A group variable can be used to estimate inequality at the level of a categorical group. If a group variable is selected, only the first variable of interest is then used.
- Standard errors and confidence intervals with a confidence level of 95% are provided. Both the type of confidence intervals provided and the level of confidence used can be changed.
- The results are displayed with 6 decimals; this can be changed.

13.8 Difference between coefficient of variation (dicvar)

This module estimates differences between coefficient of variation indices of two distributions. For each of the two distributions:

- One variable of interest should be selected;
- Conditions can be specified to focus on specific population subgroups;
- Standard errors and confidence intervals with a confidence level of 95% are provided. Both the type of confidence intervals provided and the level of confidence used can be changed.
- The results are displayed with 6 decimals; this can be changed.

13.9 Quantile/share ratio indices of inequality (inineq)

The quantile ratio is estimated as

$$\widehat{QR}(p_1, p_2) = \frac{\hat{Q}(p_1)}{\hat{Q}(p_2)}$$

where $Q(p)$ denotes a p-quantile and p_1 and p_2 are percentiles.

The share ratio is estimated as

$$\widehat{SR}(p1,p2,p3,p4) = \frac{\widehat{GL}(p2)-\widehat{GL}(p1)}{\widehat{GL}(p4)-\widehat{GL}(p3)}$$

where $\mathrm{GL(p)}$ is the Generalised Lorenz curve and p_1, p_2, p_3 and p_4 are percentiles.

- The user can select more than one variable of interest simultaneously. For example, one can estimate inequality simultaneously for *per capita* consumption and for *per capita* income.
- A group variable can be used to estimate inequality at the level of a categorical group. If a group variable is selected, only the first variable of interest is then used.
- Standard errors and confidence intervals with a confidence level of 95% are provided. Both the type of confidence intervals provided and the level of confidence used can be changed.
- The results are displayed with 6 decimals; this can be changed.

13.10 Difference between Quantile/Share indices (dinineq)

This module estimates differences between the Quantile/Share indices of two distributions.

For each of the two distributions:

- One variable of interest should be selected;
- Conditions can be specified to focus on specific population subgroups;
- Standard errors and confidence intervals with a confidence level of 95% are provided. Both the type of confidence intervals provided and the level of confidence used can be changed;
- The results are displayed with 6 decimals; this can be changed.

13.11 The Araar (2009) multidimensional inequality index

The Araar (2009) the multidimensional inequality index for the K dimensions of wellbeing takes the following form:

$$I = \sum_{i=1}^{K} \phi_k \left[\lambda_k I_k + (1-\lambda_k) C_k \right]$$

where ϕ_k is the weight attributed to the dimension k (may take the same value across the dimensions or can depend on the averages of the wellbeing dimensions). I_k, C_k are respectively the relative –absolute- Gini and concentration indices of component k. The normative parameter λ_k controls the sensitivity of the index to the inter-correlation between dimensions. For more details, see:

Abdelkrim Araar, 2009. "The Hybrid Multidimensional Index of Inequality," Cahiers de recherche 0945, CIRPEE: http://ideas.repec.org/p/lvl/lacicr/0945.html

14 *DASP* and polarization indices

14.1 The DER index (ipolder)

The Duclos, Esteban and Ray (2004) (DER) polarization index can be expressed as:

$$\mathrm{DER}(\alpha) = \iint f(x)^{1+\alpha} f(y)\left|y - x\right| dydx$$

where f(x) denotes the density function at x. The discrete formula that is used to estimate this index is as follows:

$$\mathrm{DER}(\alpha) = \frac{\sum_{i=1}^{n} w_i f(y_i)^{\alpha} a(y_i)}{\sum_{i=1}^{n} w_i}$$

The normalized DER estimated by this module is defined as:

$$\overline{\mathrm{DER}}(\alpha) = \frac{\mathrm{DER}(\alpha)}{2\mu^{(1-\alpha)}}$$

where:

$$a(y_i) = \mu + y_i\left(\left(\frac{2\sum_{j=1}^{i} w_j - w_i}{\sum_{i=1}^{N} w_i}\right) - 1\right) - \left(\frac{2\sum_{j=1}^{i-1} w_j y_j + w_i y_i}{\sum_{i=1}^{N} w_i}\right)$$

The Gaussian kernel estimator is used to estimate the density function.

- The user can select more than one variable of interest simultaneously. For example, one can estimate polarization by using simultaneously *per capita* consumption and *per capita* income.
- A group variable can be used to estimate polarization at the level of a categorical group. If a group variable is selected, only the first variable of interest is then used.
- Standard errors and confidence intervals with a confidence level of 95% are provided. Both the type of confidence intervals provided and the level of confidence used can be changed.
- The results are displayed with 6 decimals; this can be changed.

Main reference

DUCLOS, J.-Y., J. ESTEBAN, AND D. RAY (2004): "Polarization: Concepts, Measurement, Estimation," *Econometrica*, 72, 1737–1772.

14.2 Difference between DER polarization indices (dipolder)

This module estimates differences between the DER indices of two distributions.

For each of the two distributions:

- One variable of interest should be selected;
- Conditions can be specified such as to focus on specific population subgroups;
- Standard errors and confidence intervals with a confidence level of 95% are provided. Both the type of confidence intervals provided and the level of confidence used can be changed.

- The results are displayed with 6 decimals; this can be changed.

14.3 The Foster and Wolfson (1992) polarization index (ipolfw)

The Foster and Wolfson (1992) polarization index can be expressed as:

$$FW = 2\left[2\left[0.5 - \text{Lorenz}(p = 0.5)\right] - \text{Gini}\right]\frac{\mu}{\text{median}}$$

- The user can select more than one variable of interest simultaneously. For example, one can estimate polarization by using simultaneously *per capita* consumption and *per capita* income.
- A group variable can be used to estimate polarization at the level of a categorical group. If a group variable is selected, only the first variable of interest is then used.
- Standard errors and confidence intervals with a confidence level of 95% are provided. Both the type of confidence intervals provided and the level of confidence used can be changed.
- The results are displayed with 6 decimals; this can be changed.

Main reference

FOSTER, J. AND M. WOLFSON (1992): "Polarization and the Decline of the Middle Class: Canada and the U.S." mimeo, Vanderbilt University.

14.4 Difference between Foster and Wolfson (1992) polarization indices (dipolfw)

This module estimates differences between the FW indices of two distributions.

For each of the two distributions:

- One variable of interest should be selected;
- Conditions can be specified such as to focus on specific population subgroups;
- Standard errors and confidence intervals with a confidence level of 95% are provided. Both the type of confidence intervals provided and the level of confidence used can be changed.
- The results are displayed with 6 decimals; this can be changed.

14.5 The Generalised Esteban, Gradin and Ray (1999) polarisation index (ipoger)

The proposed measurement of polarisation by Esteban and Ray (1994) is defined as follows:

$$P^{ER}(f,\alpha) = \sum_{j=1}^{m}\sum_{k=1}^{m} p_j^{1+\alpha} p_k \left|\mu_j - \mu_k\right|$$

where μ_j and p_j denote respectively the average income and the population share of group j. The parameter $\alpha \in [1,1.6]$ reflects sensitivity of society to polarisation. The first step for the estimation requires to define exhaustive and mutually exclusive groups, ρ. This will involve some degree of error. Taking into account this idea, the measure of polarisation proposed by Esteban et al. (1999) is obtained after correcting the $P^{ER}(\alpha)$ index applied to the simplified representation of the original distribution with a measure of the grouping error. Nonetheless, when dealing with personal or spatial income distributions, there are no unanimous criteria for establishing the precise demarcation between different groups. To address this problem, Esteban et al. (1999) follow the methodology proposed by Aghevli and Mehran (1981) and Davies and Shorrocks (1989) in order to find an optimal partition of the distribution for a given number of groups, ρ^*. This means selecting the partition that minimises the Gini index value of within-group inequality, $\text{Error} = G(f) - G(\rho^*)$ (see Esteban et al., 1999). The measure of polarisation proposed by Esteban et al. (1999) is therefore given by:

$$P^{EGR}(f,\alpha,\rho^*,\beta) = \sum_{j=1}^{m}\sum_{k=1}^{m} p_j^{1+\alpha} p_k \left|\mu_j - \mu_k\right| - \beta\left(G(f) - G(\rho^*)\right)$$

where, $\beta \geq 0$ is a parameter that informs about the weight assigned to the error term. (In the study of Esteban et al. (1999), the value used is $\beta = 1$).

The Stata module ipoegr.ado estimates the generalised form of the Esteban et al. (1999) polarisation index. In addition to the usual variables, this routine offers the three following options:

1. The number of groups. Empirical studies use two or three groups. The user can select the number of groups. According to this number, the program searches for the optimal income interval for each group and displays them. It also displays the error in percentage, ie: $\frac{G(f) - G(\rho^*)}{G(f)} * 100$;
2. The parameter α;
3. The parameter β.

To respect the scale invariance principle, all incomes are divided by average income i.e. $\left(\dot{\mu}_j = \mu_j / \mu\right)$. In addition, we divide the index by the scalar 2 to make its interval lie between 0 and 1 when $\alpha = 1$.

$$\dot{P}^{EGR}(f,\alpha,\rho^*,\beta) = 0.5\left(\sum_{j=1}^{m}\sum_{k=1}^{m} p_j^{1+\alpha} p_k \left|\dot{\mu}_j - \dot{\mu}_k\right| - \beta\left(G(f) - G\left(\rho^*\right)\right)\right)$$

14.6 *The Inaki (2008) polarisation index (ipoger)*

Let a population be split into N groups, each one of size $n_i > 0$. The density function, the mean and the population share of group i are denoted by $f_i(x)$, μ_i and π_i respectively. μ is the overall mean. We therefore have that $\int f_i(x) = 1$, $\sum_{i=1}^{N} \pi_i \mu_i = \mu$ and $\sum_{i=1}^{N} \pi_i = 1$. Using Inaki (2008), a social polarisation index can be defined as:

$$P(F) = \sum_{i=1}^{N} \left[P_W(i,F) + P_B(i,F) \right]$$

where

$$P_W(i,F) = \mu^{\alpha-1} \pi_i^{2+\alpha} \int\int f_i^{\;1+\alpha}(x) f_i\;(y) \,|\, x - y \,|\, dydx$$

and

$$P_B(i,F) = \mu^{\alpha-1} \pi_i^{1+\alpha} \left[\left(\mu - \pi_i \mu_i\right) \int f_i^{\;1+\alpha}(x)dx + \left(1 - \pi_i\right) \int f_i^{\;1+\alpha}(x) x dx \right]$$

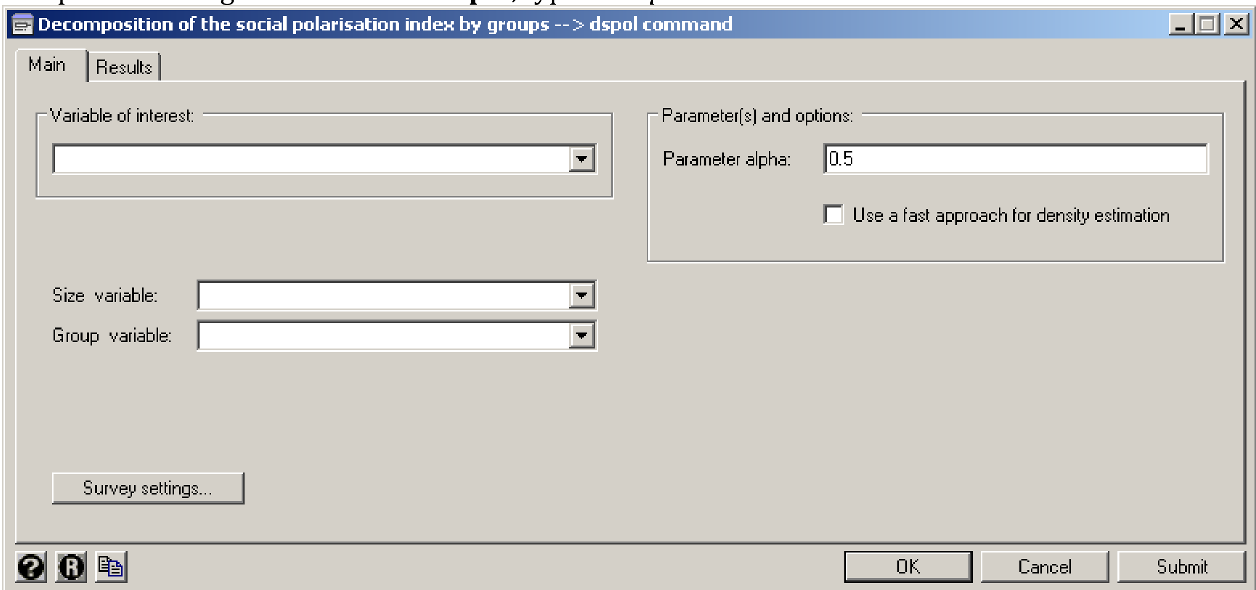
The module Stata **dspol** allows performing the decomposition of the social polarisation index $P(F)$ into group components.

- The user can select the parameter alpha;
- The user can select the use of a faster approach for the estimation of the density function;
- Standard errors are provided for all estimated indices. They take into account the full sampling design;
- The results are displayed with 6 decimals by default; this can be changed;
- The user can save results in Excel format.

The results show:

- The estimated population share of subgroup i : π_i;
- The estimated income share of subgroup i: $\pi_i \mu_i / \mu$;
- The estimated $P_W(i,F)$ index of subgroup i;
- The estimated $P_B(i,F)$ index of subgroup i;
- The estimated $P_W = \sum_i P_W(i,F)$ index;
- The estimated $P_B = \sum_i P_B(i,F)$ index;
- The estimated total index P_F

To open the dialog box for module **dspol**, type *db dspol* in the command window.



Example:

For illustrative purposes, we use a 1996 Cameroonian household survey, which is made of approximately 1700 households. The variables used are:

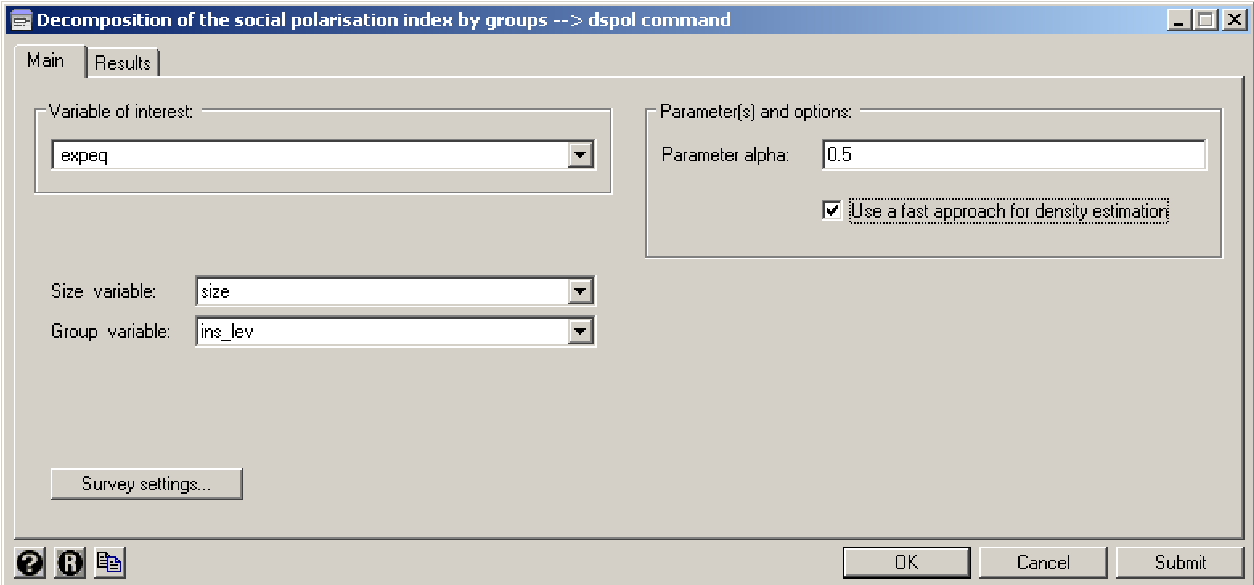
Variables:

STRATA	Stratum in which a household lives
PSU	Primary sampling unit of the household
WEIGHT	Sampling weight
SIZE	Household size
INS_LEV	Education level of the head of the household

1. Primary;
2. Professional Training, secondary and superior;
3. Not responding.

We decompose the above social polarization index using the module **dspol** by splitting the Cameroonian population into three exclusive groups, according to the education level of the household head. We first initialize the sampling design of the survey with the dialog box **svyset** as shown in what follows**:**

After that, open the dialog box by typing *db dspol*, and choose variables and parameters as in:



After clicking SUBMIT, the following results appear:

```
Decomposition of the social polarisation index by population groups
Household size  :  size
Sampling weight :  weight
Group variable  :  ins_lev
Parameter alpha :  0.50
Social polarisation index : 0.65413204 (0.00854258)
```

Group	Population Share	Income Share	Within-Group Component	Between-Group Component
Group: 1	**0.350447**	**0.286228**	**0.028080**	**0.211355**
	0.021118	0.019042	0.004381	0.015226
Group: 2	**0.282791**	**0.451516**	**0.027714**	**0.119772**
	0.021539	0.032054	0.005099	0.009773
Group: 3	**0.366762**	**0.262256**	**0.029005**	**0.238206**
	0.028543	0.027982	0.005985	0.020541
Total	**1.000000**	**1.000000**	**0.084798**	**0.569334**
	0.000000	0.000000	0.002360	0.010026

Main references

1. DUCLOS, J.-Y., J. ESTEBAN, AND D. RAY (2004): “Polarization: Concepts, Measurement, Estimation,” *Econometrica*, 72, 1737–1772.
2. Tian Z. & all (1999) "Fast Density Estimation Using CF-kernel for Very Large Databases". http://portal.acm.org/citation.cfm?id=312266
3. Iñaki Permanyer, 2008. "**The Measurement of Social Polarization in a Multi-group Context**," UFAE and IAE Working Papers 736.08, Unitat de Fonaments de l'Anàlisi Econòmica (UAB) and Institut d'Anàlisi Econòmica (CSIC).

15 *DASP* and decompositions

15.1 FGT Poverty: decomposition by population subgroups (dfgtg)

The **dgfgt** module decomposes the FGT poverty index by population subgroups. This decomposition takes the form

$$\hat{P}(z;\alpha) = \sum_{g=1}^{G} \hat{\phi}(g)\hat{P}(z;\alpha;g)$$

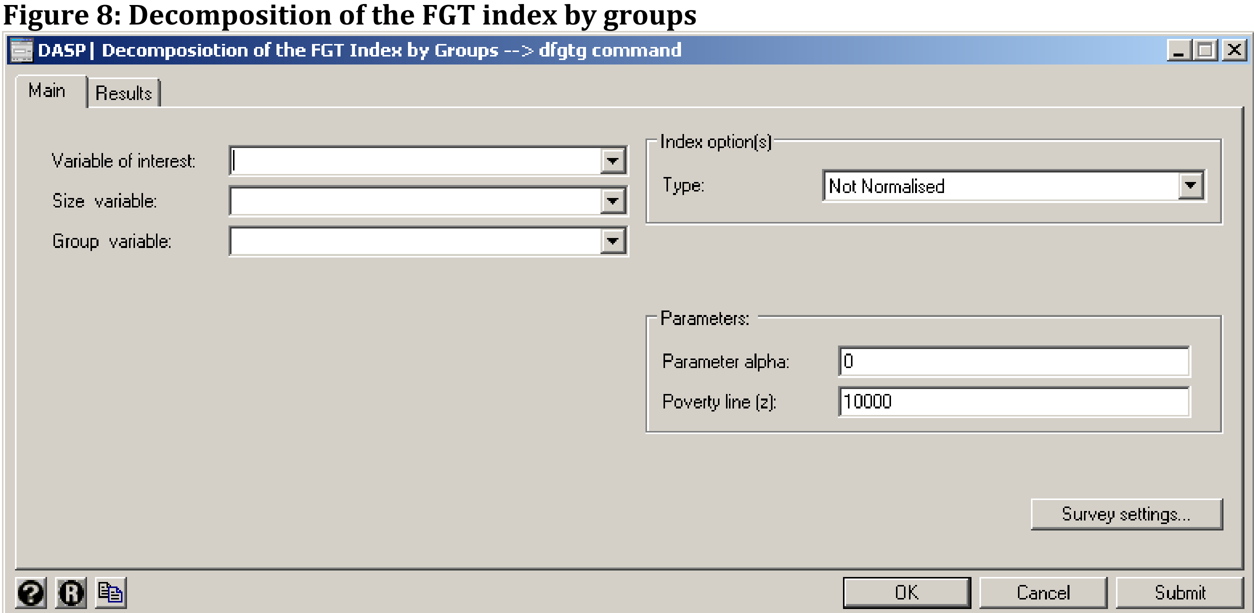
where G is the number of population subgroups. The results show:

- The estimated FGT index of subgroup g : $\hat{P}(z;\alpha;g)$
- The estimated population share of subgroup g : $\hat{\phi}(g)$
- The estimated absolute contribution of subgroup g to total poverty: $\hat{\phi}(g)\hat{P}(z;\alpha;g)$
- The estimated relative contribution of subgroup g to total poverty: $\left(\hat{\phi}(g)\hat{P}(z;\alpha;g)\right)/\hat{P}(z;\alpha)$

An asymptotic standard error is provided for each of these statistics.

To open the dialog box for module **dfgtg**, type *db dfgtg* in the command window.

Figure 8: Decomposition of the FGT index by groups



Note that the user can save results in Excel format.

Interested users are encouraged to consider the exercises that appear in Section **23.7**

15.2 FGT Poverty: decomposition by income components using the Shapley value (dfgts)

The **dfgts** module decomposes the total alleviation of FGT poverty into a sum of the contributions generated by separate income components. Total alleviation is maximal when all individuals have an income greater than or equal to the poverty line. A negative sign on a decomposition term indicates that an income component reduces poverty. Assume that there exist K income sources and that s_k denotes income source k. The FGT index is defined as:

$$\hat{P}\left(z;\alpha;y=\sum_{k=1}^{K} s_k\right)=\frac{\sum_{i=1}^{n} w_i\left(1-{}^{y}\!/\!_{z}\right)_+^{\alpha}}{\sum_{i=1}^{n} w_i}$$

where w_i is the weight assigned to individual i and n is sample size. The **dfgts** Stata module estimates:

- The share in total income of each income source k;
- The absolute contribution of each source k to the value of ($\hat{P}-1$);
- The relative contribution of each source k to the value of ($\hat{P}-1$);

Note that the **dfgts** ado file requires the module **shapar.ado**, which is programmed to perform decompositions using the Shapley value algorithm developed by Araar and Duclos (2008).

- *Araar A and Duclos J-Y (2008), "An algorithm for computing the Shapley Value", PEP and CIRPEE. Tech.-Note: Novembre-2008: http://dad.ecn.ulaval.ca/pdf_files/shap_dec_aj.pdf*

Empirical illustration with a Nigerian household survey

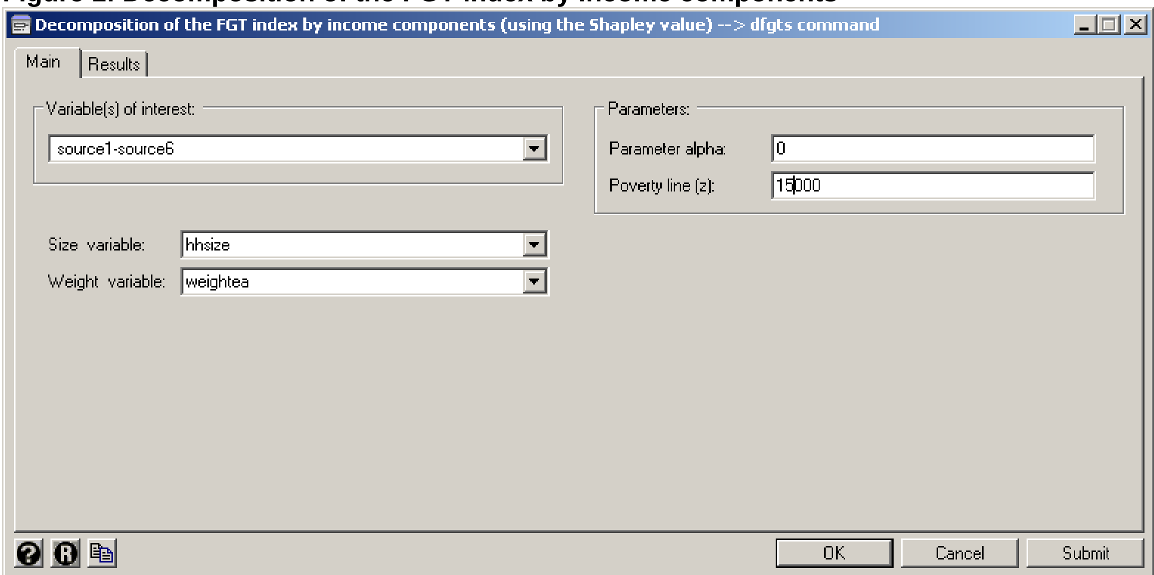
We use a survey of Nigerian households (NLSS, using 17764 observations) carried out between September 2003 and August 2004 to illustrate the use of the **dfgts** module. We use *per capita* total household income as a measure of individual living standards. Household observations are weighted by household size and sampling weights to assess poverty over all individuals. The six main income components are:

- *source_1*: Employment income;
- *source_2:* Agricultural income;
- *source_3*: Fish-processing income;
- *source_4*: Non-farm business income;
- *source_5*: Remittances received;
- *source_6*: All other income;

The Stata data file is saved after initializing its sampling design with the command **svyset**.

- To open the dialog box for module **dfgts**, type *db dfgts* in the command window.

Figure 1: Decomposition of the FGT index by income components



- Indicate the *varlist* of the six income sources.
- Indicate that the poverty line is set to 15 000 $N.
- Set the variable HOUSEHOLD SIZE.
- Set the variable HOUSEHOLD WEIGHT.
- Click on the button SUBMIT. The following results appear:

```
dfgts  source1- source6, pline(15000) hsize(hhsize) hweight( weightea)

  Decomposition of the FGT index by income components (using the Shapley value).
  Execution  time :         5.03 second(s)
  Parameter alpha :         0.00
  Poverty line    :     15000.00
  FGT index       :     0.584910
  Household size  :  hhsize
  Sampling weight :  weightea
```

Sources	Income Share	Absolute Contribution	Relative Contribution
1: source1	0.352966	-0.131599	0.317036
2: source2	0.199865	-0.091798	0.221151
3: source3	0.023731	-0.011535	0.027789
4: source4	0.344093	-0.150365	0.362247
5: source5	0.024588	-0.009299	0.022403
6: source6	0.054758	-0.020494	0.049373
Total	1.000000	-0.415090	1.000000

Marginal contributions:

Source	level_1	level_2	level_3	level_4	level_5	level_6
1: source1	-0.023228	-0.022837	-0.022328	-0.021734	-0.021085	-0.020387
2: source2	-0.014041	-0.014698	-0.015221	-0.015637	-0.015971	-0.016230
3: source3	-0.001848	-0.001953	-0.001988	-0.001971	-0.001925	-0.001850
4: source4	-0.025759	-0.025661	-0.025404	-0.025021	-0.024543	-0.023976
5: source5	-0.001297	-0.001494	-0.001597	-0.001636	-0.001643	-0.001631
6: source6	-0.003113	-0.003378	-0.003506	-0.003536	-0.003510	-0.003451

15.3 Alkire and Foster (2011) MD index of poverty: decomposition by population subgroups (dmdafg)

The dmdafg module decomposes the MD Alkire and Foster index of poverty index by population subgroups. This decomposition takes the form. The results show:

- The estimated Alkire and Foster index of each subgroup:
- The estimated population share of subgroup;
- The estimated absolute contribution of subgroup g to total poverty;
- The estimated relative contribution of subgroup g to total poverty;

An asymptotic standard error is provided for each of these statistics.

15.4 Alkire and Foster (2011: decomposition by dimensions using the Shapley value (dmdafs)

The **dmdafs** module decomposes the Alkire and Foster (2011) multidimensional poverty indices into a sum of the contributions generated by each of the poverty dimensions. It uses the Shapley characteristic function. The non-presence of a given factor –dimension- is obtained by setting the level of that dimension to its specific poverty line, thus ensuring the non-contribution of this dimension to the AF (2011) indices. Note that the **dmdafs** ado file requires the module **shapar.ado**, which is programmed to perform decompositions using the Shapley value algorithm developed by Araar and Duclos (2008).

- *Araar A and Duclos J-Y (2008), "An algorithm for computing the Shapley Value", PEP and CIRPEE. Tech.-Note: Novembre-2008: http://dad.ecn.ulaval.ca/pdf_files/shap_dec_aj.pdf*

15.5 FGT Poverty: decomposition by income components using the Shapley value (dfgts)

The **dfgts** module decomposes the total alleviation of FGT poverty into a sum of the contributions generated by separate income components. Total alleviation is maximal when all individuals have an income greater than or equal to the poverty line. A negative sign on a decomposition term indicates that an income component reduces poverty. Assume that there exist K income sources and that s_k denotes income source k. The FGT index is defined as:

$$\hat{P}\left(z;\alpha;y=\sum_{k=1}^{K} s_k\right)=\frac{\sum_{i=1}^{n} w_i\left(1-y/z\right)_+^{\alpha}}{\sum_{i=1}^{n} w_i}$$

where w_i is the weight assigned to individual i and n is sample size. The **dfgts** Stata module estimates:

- The share in total income of each income source k;
- The absolute contribution of each source k to the value of ($\hat{P}-1$);
- The relative contribution of each source k to the value of ($\hat{P}-1$);

Note that the **dfgts** ado file requires the module **shapar.ado**, which is programmed to perform decompositions using the Shapley value algorithm developed by Araar and Duclos (2008).

- *Araar A and Duclos J-Y (2008), "An algorithm for computing the Shapley Value", PEP and CIRPEE. Tech.-Note: Novembre-2008: http://dad.ecn.ulaval.ca/pdf_files/shap_dec_aj.pdf*

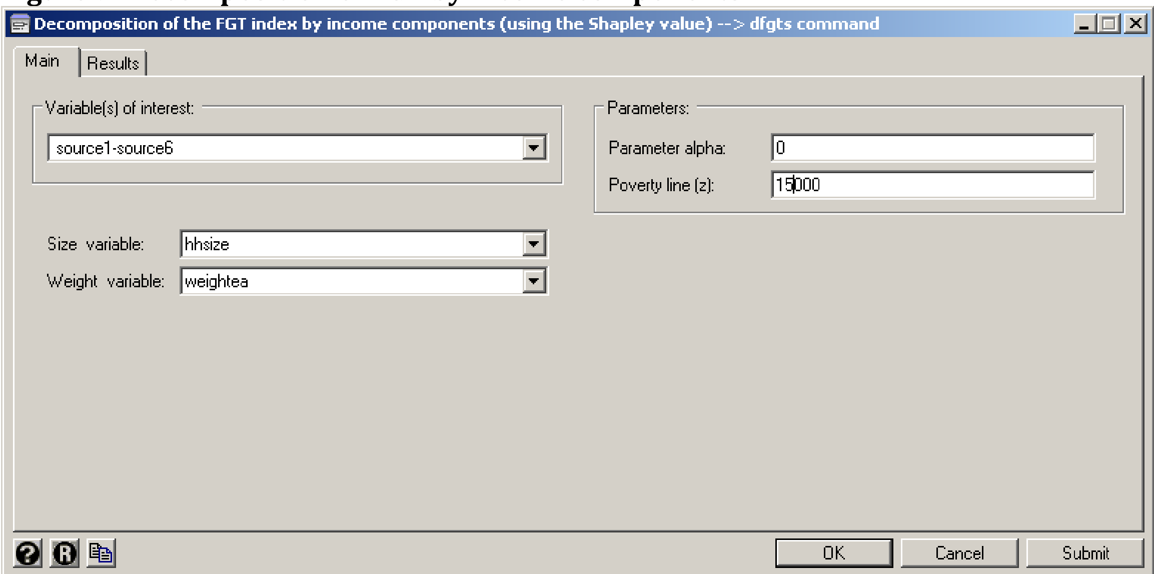
Empirical illustration with a Nigerian household survey

We use a survey of Nigerian households (NLSS, using 17764 observations) carried out between September 2003 and August 2004 to illustrate the use of the **dfgts** module. We use *per capita* total household income as a measure of individual living standards. Household observations are weighted by household size and sampling weights to assess poverty over all individuals. The six main income components are:

- *source_1*: Employment income;
- *source_2:* Agricultural income;
- *source_3*: Fish-processing income;
- *source_4*: Non-farm business income;
- *source_5*: Remittances received;
- *source_6*: All other income;

The Stata data file is saved after initializing its sampling design with the command **svyset**. To open the dialog box for module **dfgts**, type *db dfgts* in the command window.

Figure 9: Decomposition of FGT by income components



- Indicate the *varlist* of the six income sources.
- Indicate that the poverty line is set to 15 000 $N.
- Set the variable HOUSEHOLD SIZE.
- Set the variable HOUSEHOLD WEIGHT.
- Click on the button SUBMIT. The following results appear:

```
dfgts  source1- source6, pline(15000) hsize(hhsize) hweight( weightea)

  Decomposition of the FGT index by income components (using the Shapley value).
  Execution  time :        5.03 second(s)
  Parameter alpha :        0.00
  Poverty line    :    15000.00
  FGT index       :    0.584910
  Household size  :  hhsize
  Sampling weight :  weightea
```

Sources	Income Share	Absolute Contribution	Relative Contribution
1: source1	0.352966	-0.131599	0.317036
2: source2	0.199865	-0.091798	0.221151
3: source3	0.023731	-0.011535	0.027789
4: source4	0.344093	-0.150365	0.362247
5: source5	0.024588	-0.009299	0.022403
6: source6	0.054758	-0.020494	0.049373
Total	1.000000	-0.415090	1.000000

Marginal contributions:

Source	level_1	level_2	level_3	level_4	level_5
1: source1	-0.023228	-0.022837	-0.022328	-0.021734	-0.021085
2: source2	-0.014041	-0.014698	-0.015221	-0.015637	-0.015971
3: source3	-0.001848	-0.001953	-0.001988	-0.001971	-0.001925
4: source4	-0.025759	-0.025661	-0.025404	-0.025021	-0.024543
5: source5	-0.001297	-0.001494	-0.001597	-0.001636	-0.001643
6: source6	-0.003113	-0.003378	-0.003506	-0.003536	-0.003510

Source	level_6
1: source1	-0.020387
2: source2	-0.016230
3: source3	-0.001850
4: source4	-0.023976
5: source5	-0.001631
6: source6	-0.003451

15.6 Decomposition of the variation in FGT indices into growth and redistribution components (dfgtgr)

Datt and Ravallion (1992) decompose the change in the FGT index between two periods, t1 and t2, into growth and redistribution components as follows:

$$\underbrace{P_2 - P_1}_{\text{variation}} = \underbrace{\left[P(\mu^{t2}, \pi^{t1}) - P(\mu^{t1}, \pi^{t1}) \right]}_{C1} + \underbrace{\left[P(\mu^{t1}, \pi^{t2}) - P(\mu^{t1}, \pi^{t1}) \right]}_{C2} + R \quad / \text{ref} = 1$$

$$\underbrace{P_2 - P_1}_{\text{variation}} = \underbrace{\left[P(\mu^{t2}, \pi^{t2}) - P(\mu^{t1}, \pi^{t2}) \right]}_{C1} + \underbrace{\left[P(\mu^{t2}, \pi^{t2}) - P(\mu^{t2}, \pi^{t1}) \right]}_{C2} + R \quad / \text{ref} = 2$$

where

variation	= difference in poverty between t1 and t2;
C1	= growth component;
C2	= redistribution component;
R	= residual;
Ref	= period of reference.

$P(\mu^{t1}, \pi^{t1})$: the FGT index of the first period
$P(\mu^{t1}, \pi^{t1})$: the FGT index of the second period

$P(\mu^{t2},\pi^{t1})$: the FGT index of the first period when all incomes y_i^{t1} of the first period are multiplied by μ^{t2}/μ^{t1}

$P(\mu^{t1},\pi^{t2})$: the FGT index of the second period when all incomes y_i^{t2} of the second period are multiplied by μ^{t1}/μ^{t2}

The Shapley value decomposes the variation in the FGT Index between two periods, t1 and t2, into growth and redistribution components as follows:

$$\underbrace{P_2 - P_1}_{\text{Variation}} = C_1 + C_2$$

$$C_1 = \frac{1}{2}\left(\left[P(\mu^{t2},\pi^{t1}) - P(\mu^{t1},\pi^{t1})\right] + \left[P(\mu^{t2},\pi^{t2}) - P(\mu^{t1},\pi^{t2})\right]\right)$$

$$C_2 = \frac{1}{2}\left(\left[P(\mu^{t1},\pi^{t2}) - P(\mu^{t1},\pi^{t1})\right] + \left[P(\mu^{t2},\pi^{t2}) - P(\mu^{t2},\pi^{t1})\right]\right)$$

15.7 Decomposition of change in FGT poverty by poverty and population group components –sectoral decomposition- (dfgtg2d).

Additive poverty measures, like the FGT indices, can be expressed as a sum of the poverty contributions of the various subgroups of population. Each subgroup contributes by its population share and poverty level. Thus, the change in poverty across time depends on the change in these two components. Denoting the population share of group k in period t by $\phi^t(k)$, the change in poverty between two periods can be expressed as (see Huppi (1991) and Duclos and Araar (2006)):

$$\begin{aligned} & P^2(z,\alpha) - P^1(z,\alpha) \\ & = \sum_k^K \underbrace{\phi^1(k)\left(P^2(k,z,\alpha) - P^1(k,z,\alpha)\right)}_{\text{within-group poverty effects}} \\ & + \sum_k^K \underbrace{P^1(k,z,\alpha)\left(\phi^2(k) - \phi^1(k)\right)}_{\text{demographic or sectoral effects}} \\ & + \underbrace{\sum_k^K \left(\left(P^2(k,z,\alpha) - P^1(k,z,\alpha)\right)\left(\phi^2(k) - \phi^1(k)\right)\right)}_{\text{interaction term}} \end{aligned} \qquad (06)$$

This decomposition use the initial period as the one. If the reference period is the final, the decomposition takes the form:

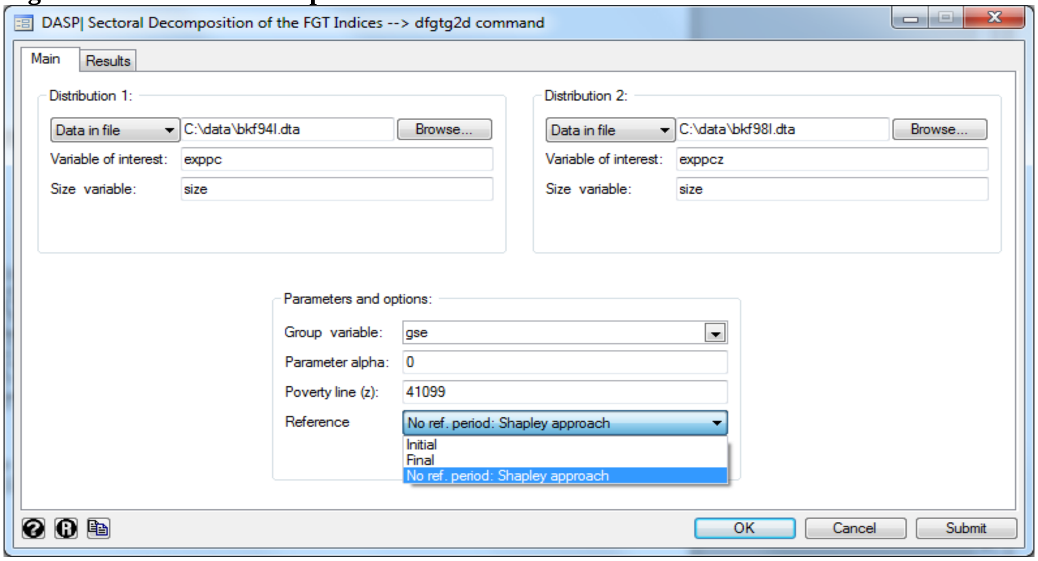
$$
\begin{aligned}
P^2(z;\alpha) - P^1(z;\alpha)
&= \sum_{k}^{K} \underbrace{\phi^2(k)\left(P^2(k;z;\alpha) - P^1(k;z;\alpha)\right)}_{within-group\ poverty\ effects} \\
&+ \sum_{k}^{K} \underbrace{P^2(k;z;\alpha)\left(\phi^2(k) - \phi^1(k)\right)}_{demographic\ or\ sectoral\ effects} \\
&+ \underbrace{\sum_{k}^{K} \left(\left(P^2(k;z;\alpha) - P^1(k;z;\alpha)\right)\left(\phi^1(k) - \phi^2(k)\right)\right)}_{interaction\ term}
\end{aligned}
\qquad (06)
$$

To remove the arbitrarness in selecting the reference period, we can use the Shapley decomposition approach, finding:

$$
\begin{aligned}
P^2(z;\alpha) - P^1(z;\alpha)
&= \sum_{k}^{K} \underbrace{\bar{\phi}(k)\left(P^2(k;z;\alpha) - P^1(k;z;\alpha)\right)}_{within-group\ poverty\ effects} \\
&+ \sum_{k}^{K} \underbrace{\bar{P}(k;z;\alpha)\left(\phi^2(k) - \phi^1(k)\right)}_{demographic\ or\ sectoral\ effects}
\end{aligned}
\qquad (07)
$$

where $\bar{\phi}(k)$ is the average population share $= 0.5(\phi^1(k) + \phi^2(k))$ and $\bar{P}(k;z;\alpha) = 0.5(P^1(k;z;\alpha) + P^2(k;z;\alpha))$. The DASP module dfgtg2d performs this sectoral decomposition, and this by selecting the reference period of the Shapley approach (see the following dialog box):

Figure 10: Sectoral decomposition of FGT




```
. dfgtg2d exppc exppcz, alpha(0) hgroup(gse) pline(41099) file1(C:\data\bkf94I.dta) hsize1(size) file2(C:\data\bkf98I.dta) hsize2(size) ref(0)
```

Decomposition of the FGT index by groups
Group variable : gse
Parameter alpha : 0.00

Population shares and FGT indices

Group	Initial Pop. share	Initial FGT index	Final Pop. share	Final FGT index	Difference in FGT index
Wage-earner (public sector)	0.042971	0.022406	0.041403	0.059094	0.036688
	0.003790	0.012599	0.003927	0.023396	0.026573
Wage-earner (private sector)	0.026598	0.067271	0.029035	0.111283	0.044012
	0.002164	0.024093	0.002624	0.023087	0.033369
Artisan or trader	0.062640	0.097548	0.055795	0.126776	0.029228
	0.004288	0.014712	0.004666	0.018202	0.023404
Other type of earner	0.006650	0.194481	0.005689	0.293404	0.098923
	0.001308	0.060817	0.000923	0.089680	0.108357
Crop farmer	0.104402	0.500707	0.167806	0.424391	-0.076316
	0.014896	0.034911	0.014125	0.024457	0.042625
Subsistence farmer	0.680885	0.514999	0.653552	0.533956	0.018957
	0.016403	0.021132	0.015083	0.011572	0.024093
Inactive	0.075856	0.414986	0.046719	0.386852	-0.028134
	0.004839	0.035336	0.003354	0.032340	0.047901
Population	1.000000	0.444565	1.000000	0.452677	0.008113
	0.000000	0.016124	0.000000	0.010927	0.019477

Decomposition components

Group	Poverty Component	Population Component	Interaction Component
Wage-earner (public sector)	0.001548	-0.000064	0.000000
	0.001117	0.001931	0.000000
Wage-earner (private sector)	0.001224	0.000218	0.000000
	0.000930	0.001222	0.000000
Artisan or trader	0.001731	-0.000768	0.000000
	0.001380	0.002417	0.000000
Other type of earner	0.000610	-0.000234	0.000000
	0.000700	0.000930	0.000000
Crop farmer	-0.010387	0.029328	0.000000
	0.005992	0.013963	0.000000
Subsistence farmer	0.012648	-0.014336	0.000000
	0.016127	0.018726	0.000000
Inactive	-0.001724	-0.011681	0.000000
	0.002932	0.004317	0.000000
Population	0.005650	0.002463	0.000000
	===	===	===

15.8 Decomposition of FGT poverty by transient and chronic poverty components (dtcpov)

This decomposes total poverty across time into transient and chronic components.

The Jalan and Ravallion (1998) approach

Let y_i^t be the income of individual i in period t and μ_i be average income over the T periods for that same individual i, i=1,…,N. Total poverty is defined as:

$$TP(\alpha, z) = \frac{\sum_{t=1}^{T} \sum_{i=1}^{N} w_i (z - y_i^t)_+^{\alpha}}{T \sum_{i=1}^{N} w_i}$$

The chronic poverty component is then defined as:

$$CPC(\alpha, z) = \frac{\sum_{i=1}^{N} w_i (z - \mu_i)_+^{\alpha}}{\sum_{i=1}^{N} w_i}$$

Transient poverty equals:

$$TPC(\alpha, z) = TP(\alpha, z) - CPC(\alpha, z)$$

Duclos, Araar and Giles (2006) approach

Let y_i^t be the income of individual i in period t and μ_i be average income over the T periods for individual i. Let $\Gamma(\alpha, z)$ be the ”equally-distributed-equivalent” (EDE) poverty gap such that:

$$\Gamma(\alpha, z) = \left[TP(\alpha, z)\right]^{1/\alpha}$$

Transient poverty is then defined as

$$TPC(\alpha, z) = \frac{\sum_{i=1}^{N} w_i \theta_i(\alpha, z)}{\sum_{i=1}^{N} w_i}$$

where $\theta_i = \gamma_i(\alpha, z) - \gamma_i(1, z)$ and B $\gamma_i(\alpha, z) = \left(\left(\sum_{i=t}^{T} (z - y_i^t)_+^{\alpha}\right) / T\right)^{1/\alpha}$

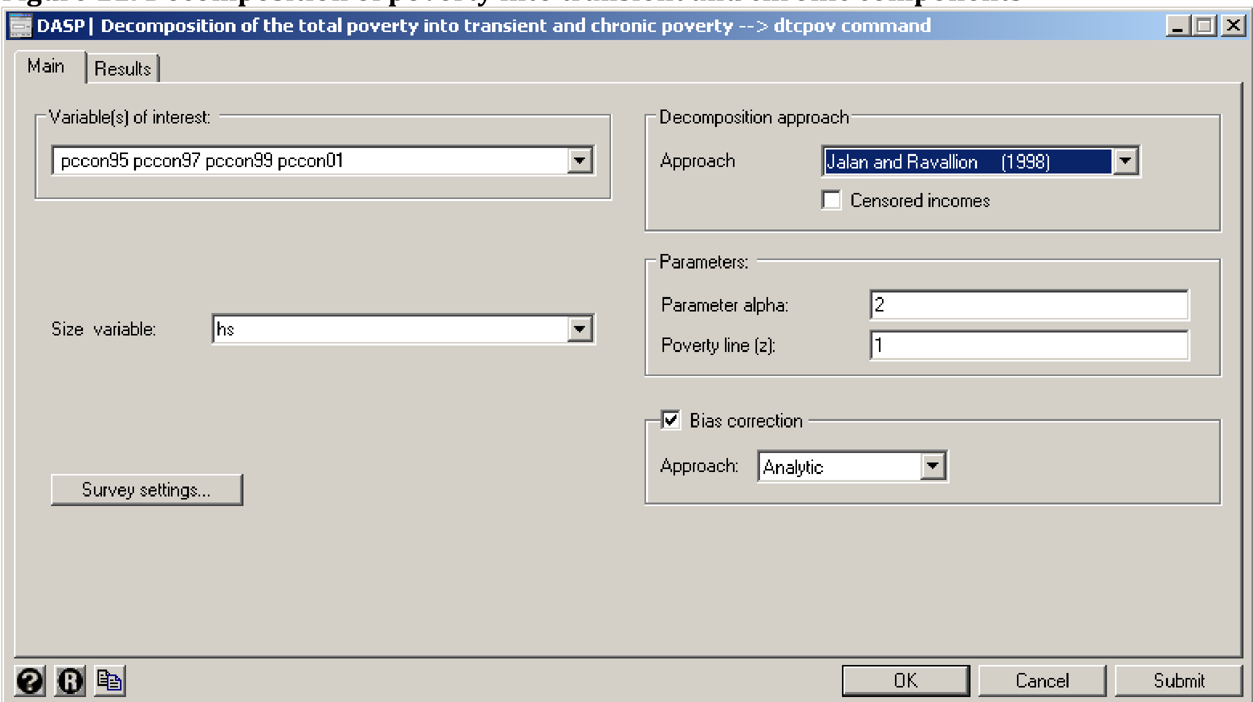
and chronic poverty is given by

$$\mathrm{CPC}(\alpha, z) = \Gamma(\alpha, z) - \mathrm{TPC}(\alpha, z)$$

Note that the number of periods available for this type of exercise is generally small. Because of this, a bias-correction is typically useful, using either an analytical/asymptotic or bootstrap approach.

To open the dialog box for module **dtcpov**, type *db dtcpov* in the command window.

Figure 11: Decomposition of poverty into transient and chronic components



- The user can select more than one variable of interest simultaneously, where each variable represents income for one period.
- The user can select one of the two approaches presented above.
- Small-T-bias-corrections can be applied, using either an analytical/asymptotic or a bootstrap approach.
- Standard errors and confidence intervals with a confidence level of 95% are provided. Both the type of confidence intervals provided and the level of confidence used can be changed.
- The results are displayed with 6 decimals; this can be changed.

References

Jalan Jyotsna, and Martin Ravallion. (1998) "Transient Poverty in Postreform Rural China" Journal of Comparative Economics, 26(2), pp. 338:57.

Jean-Yves Duclos & Abdelkrim Araar & John Giles, 2006. "Chronic and Transient Poverty: Measurement and Estimation, with Evidence from China," Working Paper 0611, CIRPEE.

15.9 Inequality: decomposition by income sources (diginis)

Analytical approach

The **diginis** module decomposes the (usual) relative or the absolute Gini index by income sources. The three available approaches are:

- Rao's approach (1969)
- Lerman and Yitzhaki's approach (1985)
- Araar's approach (2006)

Reference(s)

- Lerman, R. I., and S. Yitzhaki. "Income Inequality Effects by Income Source: A New Approach and Applications to the United States." Review of Economics and Statistics 67 (1985): 151-56.
- Araar Abdelkrim (2006). On the Decomposition of the Gini Coefficient: an Exact Approach, with an Illustration Using Cameroonian Data, Working paper 02-06, CIRPEE.

Shapley approach

The **dsineqs** module decomposes inequality indices into a sum of the contributions generated by separate income components. The **dsineqs** Stata module estimates:

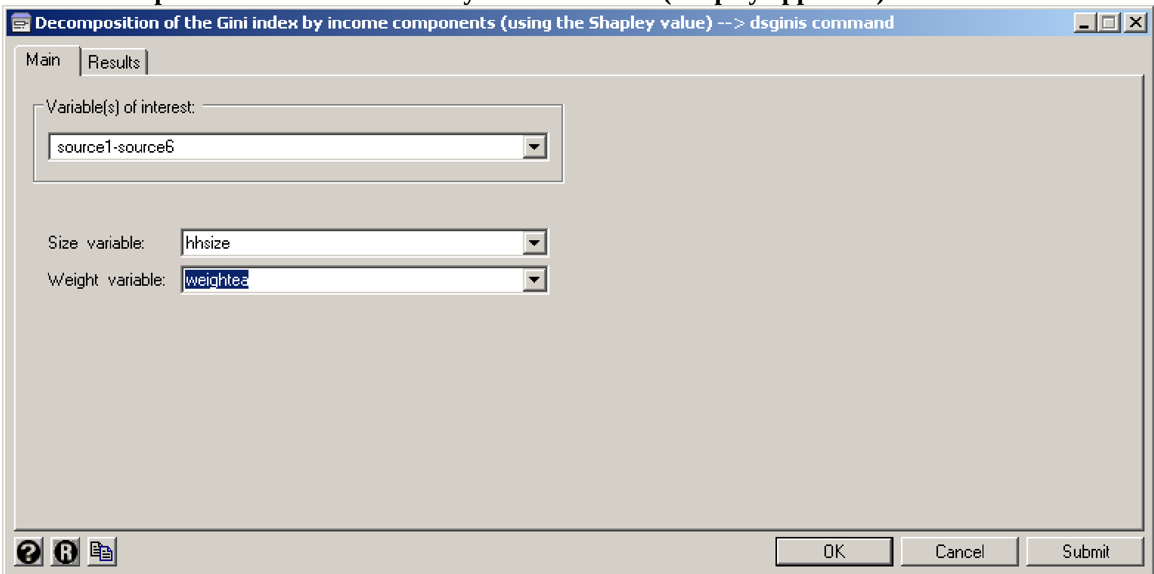
- The share in total income of each income source k;
- The absolute contribution of each source k to the Gini index;
- The relative contribution of each source k to the Gini index;

For the Shapley decomposition, the rule that is used to estimate the inequality index for a subset of components is by suppressing the inequality generated by the complement subset of components. For this, we generate a counterfactual vector of income that equals the sum of the components of the subset plus the average of the complement subset. Note that the **dsineqs** ado file requires the module **shapar.ado**, which is programmed to perform decompositions using the Shapley value algorithm developed by Araar and Duclos (2008).

- *Araar A and Duclos J-Y (2008), "An algorithm for computing the Shapley Value", PEP and CIRPEE. Tech.-Note: Novembre-2008: : http://dad.ecn.ulaval.ca/pdf files/shap dec aj.pdf*

To open the dialog box for module **dsginis**, type *db dsginis* in the command window.

Figure 12: Decomposition of the Gini index by income sources (Shapley approach)



15.10 *Regression-based decomposition of inequality by income sources*

A useful approach to show the contribution of income covariates to total inequality is by decomposing the latter by the predicted contributions of covariates. Formally, denote total income by y and the set of covariates by $X = \{x_{1,} x_2, \cdots, x_K\}$. Using a linear model specification, we have:

$$y = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_k x_k + \cdots + \hat{\beta}_K x_K + \hat{\varepsilon}$$

where $\hat{\beta}_0$ and $\hat{\varepsilon}$ denote respectively the estimated constant term and the residual.

Two approaches for the decomposition of total inequality by income sources are used:

1- *The Shapley approach:* This approach is based on the expected marginal contribution of income sources to total inequality.

2- *The Analytical approach:* This approach is based on algebraic developments that express total inequality as a sum of inequality contributions of income sources.

With the Shapley approach:

- The user can select among the following relative inequality indices;
  - Gini index
  - Atkinson index
  - Generalized entropy index
  - Coefficient variation index
- The user can select among the following model specifications;
  - Linear : $y = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_K x_K + \hat{\varepsilon}$
  - Semi Log Linear : $\log(y) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_K x_K + \hat{\varepsilon}$

With the Analytical approach:

- The user can select among the following relative inequality indices;
  - Gini index
  - Squared coefficient variation index
- The model specification is linear.

<u>Decomposing total inequality with the analytical approach:</u>

Total income equals $y = s_0 + s_1 + s_2 + \cdots + s_K + s_R$ where s_0 is the estimated constant, $s_k = \hat{\beta}_k X_k$ and s_R is the estimated residual. As reported by Wang 2004, relative inequality indices are not defined when the average of the variable of interest equals zero (the case of the residual). Also, inequality indices equal zero when the variable of interest is a constant (the case of the estimated constant). To deal with these two problems, Wang (2004) proposes the following basic rules:

Let $\hat{y} = s_0 + s_1 + s_2 + \cdots + s_K$ and $\tilde{y} = s_1 + s_2 + \cdots + s_K$, then: $I(y) = cs_0 + I(\tilde{y}) + cs_r$

The contribution of the constant: $cs_0 = I(y) - I(\hat{y})$

The contribution of the residual: $cs_R = I(\hat{y}) - I(\tilde{y})$

<u>The Gini index:</u>

Using Rao 1969's approach, the relative Gini index can be decomposed as follows:

$$I(\tilde{y}) = \frac{\mu_k}{\mu_{\tilde{y}}} \tilde{C}_k$$

where $\mu_{\tilde{y}}$ is the average of $\tilde{y}$ and $\tilde{C}_k$ is the coefficient of concentration of s_k when $\tilde{y}$ is the ranking variable.

<u>The Squared coefficient of variation index:</u>

As shown by Shorrocks 1982, the squared coefficient of variation index can be decomposed as:

$$I(\tilde{y}) = \sum_{k=1}^{K} \frac{C\,\mathrm{ov}(\tilde{y}, s_k)}{\mu_{\tilde{y}}^2}$$

<u>Shapley decompositions:</u>

The Shapley approach is built around the expected marginal contribution of a component. The user can select among two methods to define the impact of missing a given component.

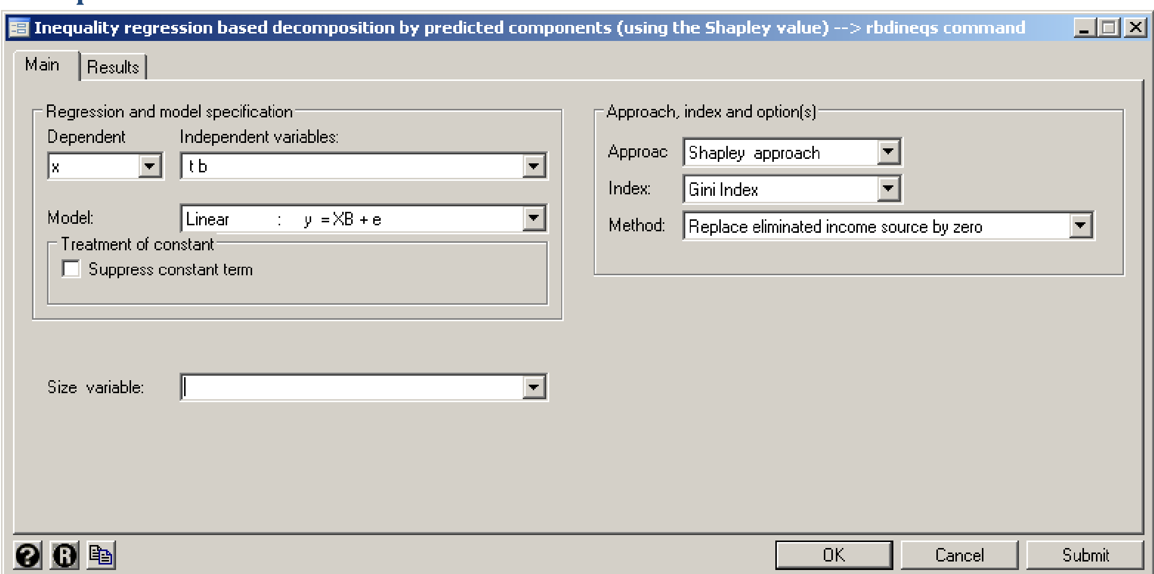
- With option: method(mean), when a component is missing from a given set of components, it is replaced by its mean.
- With option: method(zero), when a component is missing from a given set of components, it is replaced by zero.

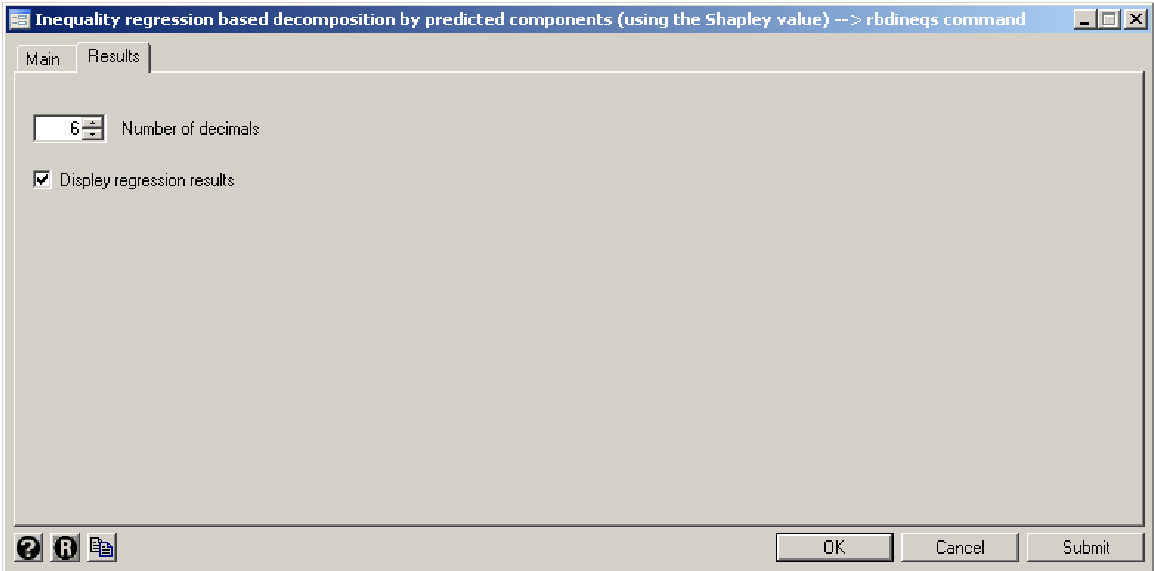
As indicated above, we cannot estimate relative inequality for the residual component.

- For the linear model, the decomposition takes the following form: $I(y) = I(\hat{y}) + cs_r$, where the contribution of the residual is $cs_r = I(y) - I(\hat{y})$.
- For the Semi-log linear model, the Shapley decomposition is applied to all components including the constant and the residual.

With the Shapley approach, the user can use the log linear specification. However, the user must indicate the income variable and not the log of that variable (DASP automatically runs the regression with log(y) as the dependent variable).

Example 1




Inequality regression based decomposition by predicted components (using the Shapley value) --> rbdineqs command
Main
Results
6
Number of decimals
Displey regression results
OK
Cancel
Submit

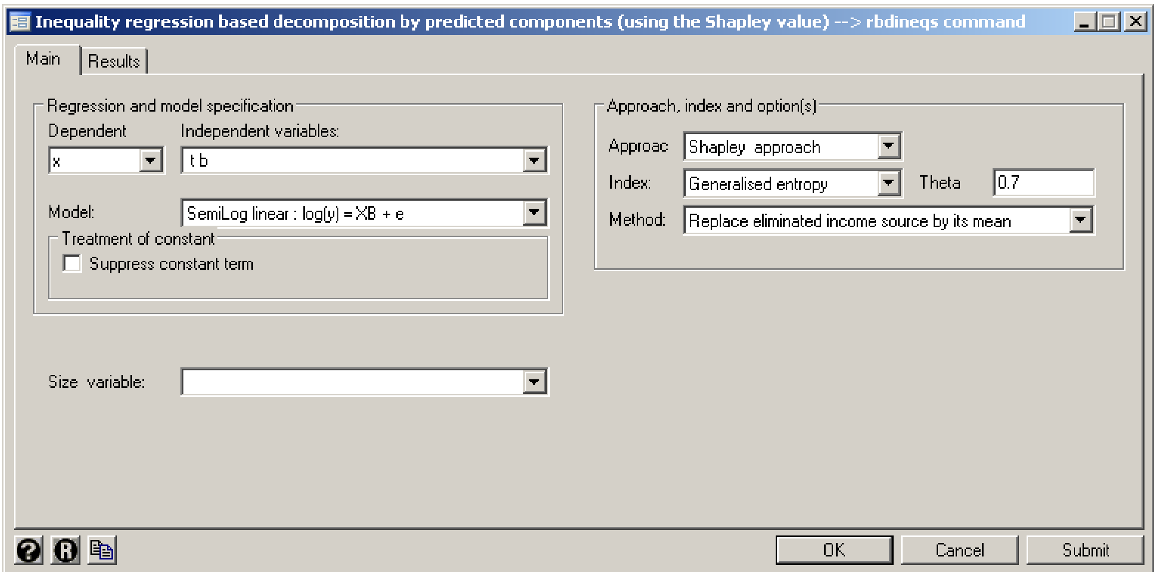
```
. rbdineqs t b, dep(x) method(zero) dregres(1)
(sum of wgt is   1.0000e+03)

      Source |       SS       df       MS              Number of obs =    1000
-------------+------------------------------           F(  2,   997) = 6409.95
       Model |  5.0832e+11     2  2.5416e+11           Prob > F      =  0.0000
    Residual |  3.9532e+10   997  39650534.2           R-squared     =  0.9278
-------------+------------------------------           Adj R-squared =  0.9277
       Total |  5.4785e+11   999   548395658           Root MSE      =  6296.9

------------------------------------------------------------------------------
           x |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           t |    2.53851   .0241632   105.06   0.000     2.491093    2.585926
           b |  -.8753363   .0464088   -18.86   0.000    -.9664065    -.784266
       _cons |   11054.16   291.9053    37.87   0.000     10481.34    11626.98
------------------------------------------------------------------------------

    Inequality regression based decomposition by predicted income components (using the Shapley value).
    Execution  time       :  0.86      second(s)
    Inequality index      :  Gini index
    Estimated inequality  :  0.508456
```

Sources	Income Share	Absolute Contribution	Relative Contribution
1: _p_cons	0.599814	-0.118428	-0.232918
2: _p_t	0.572300	0.596276	1.172719
3: _p_b	-0.172114	-0.051300	-0.100894
4: _p_residual	0.000000	0.081909	0.161093
Total	0.000000	0.508456	1.000000

Marginal contributions:

Source	level_1	level_2	level_3
1: _p_cons	0.000000	0.094680	-0.213108
2: _p_t	0.205723	0.332915	0.057638
3: _p_b	-0.210092	0.117056	0.041735

Example 2




```
. rbdineqs t b, dep(x) index(ge) theta(0.7) model(semilog) dregres(0)
Warning-> 115  OBS are omitted:
 - Dependant variable should not be =< 0 with the semilog specification.

    Inequality regression based decomposition by predicted income components (using the Shapley value).
    Execution  time        :  0.17      second(s)
    Inequality index       :  Generalised entropy index
    Estimated inequality   :  0.368465
```

Sources	Income Share	Absolute Contribution	Relative Contribution
1: _p_cons	.	-0.000000	-0.000000
2: _p_t	.	3.946080	10.709499
3: _p_b	.	0.057117	0.155012
4: _p_residual	.	-3.634731	-9.864511
Total	.	0.368465	1.000000

Marginal contributions:

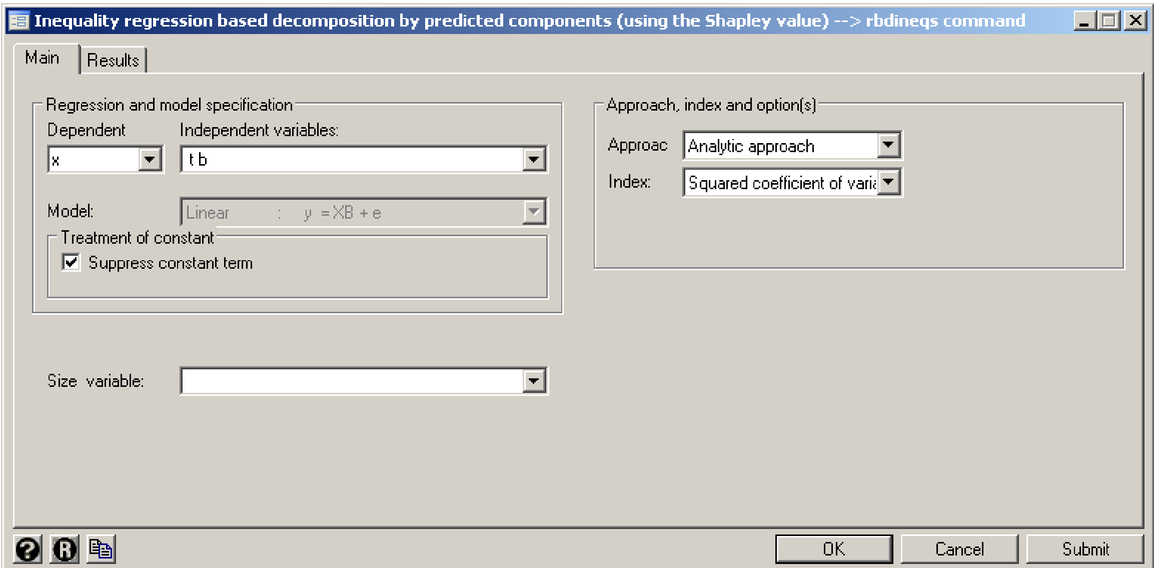
Source	level_1	level_2	level_3
1: _p_cons	0.000000	-0.000000	0.000000
2: _p_t	1.324649	1.315360	1.306071
3: _p_b	0.028328	0.019039	0.009750

With this specification, we have $y = E\,\text{xp}(s_0 + s_1 + s_2 + \cdots + s_K + s_R)$. Then:

- We cannot estimate the income share (no linear form);

- The contribution of the constant is nil. $y = E\,\mathrm{xp}(s_0).\prod_{k=1}^{K} E\,\mathrm{xp}(s_k).E\,\mathrm{xp}(s_E)$. Adding a constant will have not any impact.

Example 3




```
. rbdineqs t b, dep(x) index(scvar) appr(analytic) noconstant dregres(0)

    Inequality regression based decomposition by predicted income components.
    Execution time       :  0.16      second(s)
    Inequality index     :  Squared coefficient of variation index
    Estimated inequality :  1.613027
```

Sources	Income Share	Absolute Contribution	Relative Contribution
1: _p_t	0.668583	1.637250	1.015018
2: _p_b	0.052303	-0.028919	-0.017928
3: _p_residual	0.000000	-1.869025	-1.158707
Total	0.000000	1.613026	1.000000

15.11 Gini index: decomposition by population subgroups (diginig).

The **diginig** module decomposes the (usual) relative or the absolute Gini index by population subgroups. Let there be G population subgroups. We wish to determine the contribution of every one of those subgroups to total population inequality. The Gini index can be decomposed as follows:

$$I = \underbrace{\sum_{g=1}^{G} \phi_g \varphi_g I_g}_{\text{Between}} + \underbrace{\bar{I}}_{\text{Within}} + \underbrace{R}_{\text{Overlap}}$$

where

ϕ_g	the population share of group g;
φ_g	the income share of group g.
$\overline{\mathrm{I}}$	between-group inequality (when each individual is assigned the average income of his group).
R	The residue implied by group income overlap

15.12 Generalized entropy indices of inequality: decomposition by population subgroups (dentropyg).

The Generalised Entropy indices of inequality can be decomposed as follows:

$$\hat{\mathrm{I}}(\theta)=\sum_{k=1}^{K}\hat{\phi}(k)\left(\frac{\hat{\mu}(k)}{\hat{\mu}}\right)^{\theta}.\hat{\mathrm{I}}(k;\theta)+\hat{\bar{\mathrm{I}}}(\theta)$$

where:

B $\phi(k)$	is the proportion of the population found in subgroup k.
B $\mu(k)$	is the mean income of group k.
B $\mathrm{I}(k;\theta)$	is inequality within group k.
B $\bar{\mathrm{I}}(\theta)$	is population inequality if each individual in subgroup k is given the mean income of subgroup k, $\mu(k)$.

15.13 Polarization: decomposition of the DER index by population groups (dpolag)

As proposed by Araar (2008), the Duclos, Esteban and Ray index can be decomposed as follows:

$$P=\underbrace{\sum_{g}\varphi_g^{1+\alpha}\psi_g^{1-\alpha}R_gP_g}_{\text{Within}}+\underbrace{\tilde{P}}_{\text{Between}},$$

where

$$R_g=\frac{\int a_g(x)\pi_g(x)f(x)^{1+\alpha}dx}{\varphi_g\int a_g(x)f_g(x)^{1+\alpha}dx},$$

- φ_g and ψ_g are respectively the population and income shares of group g.

- $\pi_g(x)$ denotes the local proportion of individuals belonging to group g and having income x;
- $\tilde{P}$ is the DER polarization index when the within-group polarization or inequality is ignored;
-

The **dpolas** module decomposes the DER index by population subgroups.

Reference(s)

Abdelkrim Araar, 2008. "**On the Decomposition of Polarization Indices: Illustrations with Chinese and Nigerian Household Surveys**," Cahiers de recherche 0806, CIRPEE.

15.14 Polarization: decomposition of the DER index by income sources (dpolas)

As proposed by Araar (2008), the Duclos, Esteban and Ray index can be decomposed as follows:

$$P = \sum_k \psi_k CP_k$$

where $CP_k = \frac{\int f(x)^{1+\alpha} a_k(x)dx}{\psi_k^{\alpha} \mu_k^{\alpha-1}}$ and ψ_k are respectively the pseudo concentration index and income share of income source k. The **dpolas** module decomposes the DER index by income sources.

Reference(s)

Abdelkrim Araar, 2008. "**On the Decomposition of Polarization Indices: Illustrations with Chinese and Nigerian Household Surveys**," Cahiers de recherche 0806, CIRPEE.

16 *DASP* and curves.

16.1 FGT CURVES (cfgt).

FGT curves are useful distributive tools that can *inter alia* be used to:

1. Show how the level of poverty varies with different poverty lines;
2. Test for poverty dominance between two distributions;

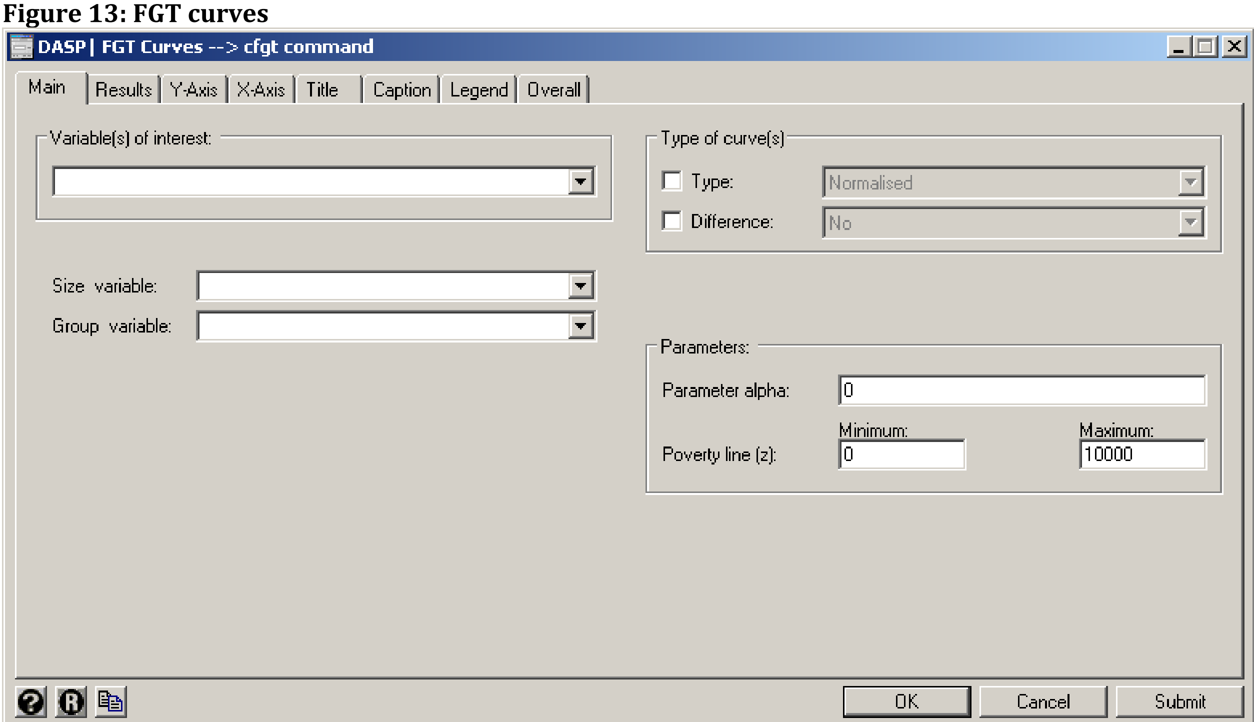
3. Test pro-poor growth conditions.

FGT curves are also called primal dominance curves. The **cfgt** module draws such curves easily. The module can:

- draw more than one FGT curve simultaneously whenever more than one variable of interest is selected;
- draw FGT curves for different population subgroups whenever a group variable is selected;
- draw FGT curves that are not normalized by the poverty lines;
- draw differences between FGT curves;
- list or save the coordinates of the curves;
- save the graphs in different formats:
  - *.gph : Stata format;
  - *.wmf : typically recommended to insert graphs in Word documents;
  - *.eps : typically recommended to insert graphs in Tex/Latex documents.
- Many graphical options are available to change the appearance of the graphs.

To open the dialog box of the module **cfgt**, type the command *db dfgt* in the command window.

Figure 13: FGT curves



Interested users are encouraged to consider the exercises that appear in Section **23.4**.

FGT CURVE with confidence interval (cfgts).

The **cfgts** module draws an FGT curve and its confidence interval by taking into account sampling design. The module can:

- draw an FGT curve and two-sided, lower-bounded or upper-bounded confidence intervals around that curve;
- condition the estimation on a population subgroup;
- draw a FGT curve that is not normalized by the poverty lines;
- list or save the coordinates of the curve and of its confidence interval;
- save the graphs in different formats:
  - *.gph : Stata format;
  - *.wmf : typically recommended to insert graphs in Word documents;
  - *.eps : typically recommended to insert graphs in Tex/Latex documents.
- Many graphical options are available to change the appearance of the graphs.

Interested users are encouraged to consider the exercises that appear in Section **23.5**.

16.3 Difference between FGT CURVES with confidence interval (cfgts2d).

The **cfgts2d** module draws differences between FGT curves and their associated confidence interval by taking into account sampling design. The module can:

- draw differences between FGT curves and two-sided, lower-bounded or upper-bounded confidence intervals around these differences;
- normalize or not the FGT curves by the poverty lines;
- list or save the coordinates of the differences between the curves as well as the confidence intervals;
- save the graphs in different formats:
  - *.gph : Stata format;
  - *.wmf : typically recommended to insert graphs in Word documents;
  - *.eps : typically recommended to insert graphs in Tex/Latex documents.
- Many graphical options are available to change the appearance of the graphs.

Interested users are encouraged to consider the exercises that appear in Section **23.5**.

Lorenz and concentration CURVES (clorenz).

Lorenz and concentration curves are useful distributive tools that can *inter alia* be used to:

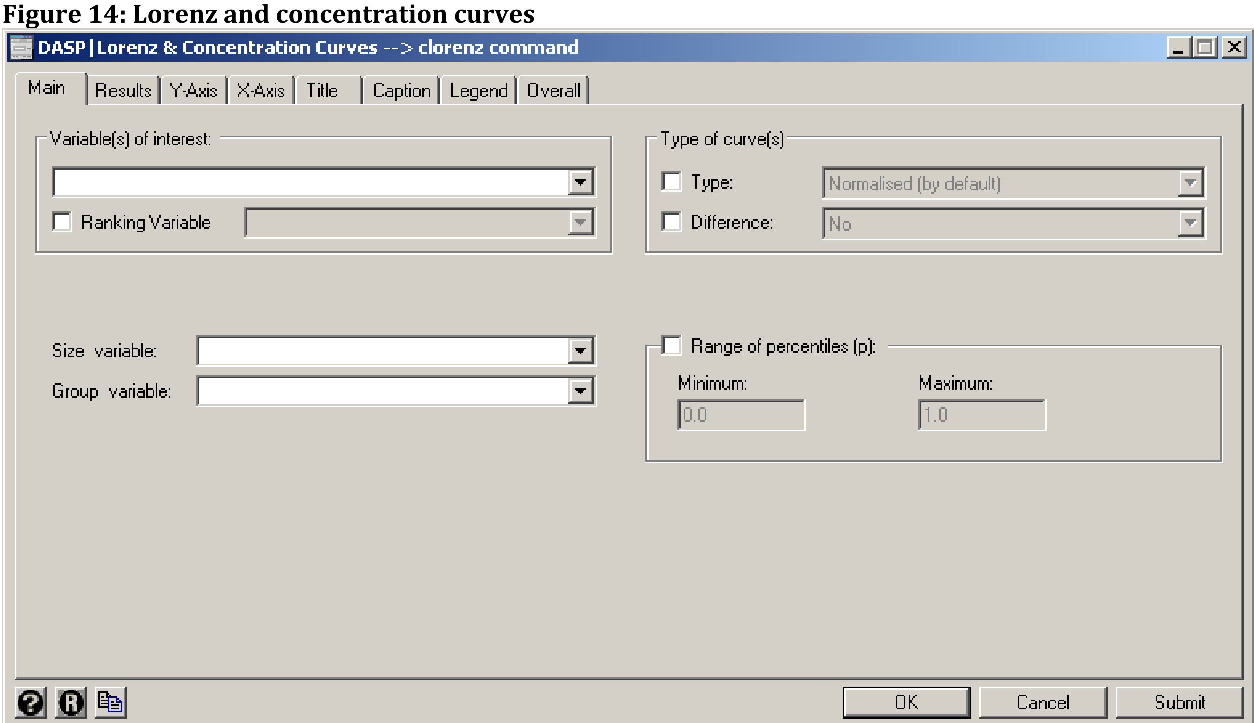
1. show the level of inequality;
2. test for inequality dominance between two distributions;
3. test for welfare dominance between two distributions;
4. test for progressivity.

The **clorenz** module draws Lorenz and concentration curves simultaneously. The module can:

- draw more than one Lorenz or concentration curve simultaneously whenever more than one variable of interest is selected;
- draw more than one generalized or absolute Lorenz or concentration curve simultaneously whenever more than one variable of interest is selected;
- draw more than one deficit share curve;
- draw Lorenz and concentration curves for different population subgroups whenever a group variable is selected;
- draw differences between Lorenz and concentration curves;
- list or save the coordinates of the curves;
- save the graphs in different formats:
  - *.gph : Stata format;
  - *.wmf : typically recommended to insert graphs in Word documents;
  - *.eps : typically recommended to insert graphs in Tex/Latex documents.
- Many graphical options are available to change the appearance of the graphs.

To open the dialog box of the module **clorenz**, type the command *db clorenz* in the command window.

Figure 14: Lorenz and concentration curves



Interested users are encouraged to consider the exercises that appear in Section **23.8**.

16.5 Lorenz/concentration curves with confidence intervals (clorenzs).

The **clorenzs** module draws a Lorenz/concentration curve and its confidence interval by taking sampling design into account. The module can:

- draw a Lorenz/concentration curve and two-sided, lower-bounded or upper-bounded confidence intervals;
- condition the estimation on a population subgroup;
- draw Lorenz/concentration curves and generalized Lorenz/concentration curves;
- list or save the coordinates of the curves and their confidence interval;
- save the graphs in different formats:
  - *.gph : Stata format;
  - *.wmf : typically recommended to insert graphs in Word documents;
  - *.eps : typically recommended to insert graphs in Tex/Latex documents.
- Many graphical options are available to change the appearance of the graphs.

16.6 Differences between Lorenz/concentration curves with confidence interval (clorenzs2d)

The **clorenz2d** module draws differences between Lorenz/concentration curves and their associated confidence intervals by taking sampling design into account. The module can:

- draw differences between Lorenz/concentration curves and associated two-sided, lower-bounded or upper-bounded confidence intervals;
- list or save the coordinates of the differences and their confidence intervals;
- save the graphs in different formats:
  - *.gph : Stata format;
  - *.wmf : typically recommended to insert graphs in Word documents;
  - *.eps : typically recommended to insert graphs in Tex/Latex documents.
- Many graphical options are available to change the appearance of the graphs.

16.7 Poverty curves (cpoverty)

The **cpoverty** module draws the poverty gap or the cumulative poverty gap curves.

- The poverty gap at a percentile p is: $G(p;z) = (z - Q(p))_+$
- The cumulative poverty gap at a percentile p, noted by $CPG(p;z)$, is given by:

$$CPG(p;z) = \frac{\sum_{i=1}^{n} w_i \, (z - y_i)_+ I(y_i \le Q(p))}{\sum_{i=1}^{n} w_i}$$

The module can thus:

- draw more than one poverty gap or cumulative poverty gap curves simultaneously whenever more than one variable of interest is selected;
- draw poverty gap or cumulative poverty gap curves for different population subgroups whenever a group variable is selected;

- draw differences between poverty gap or cumulative poverty gap curves;
- list or save the coordinates of the curves;
- save the graphs in different formats:
  - *.gph : Stata format;
  - *.wmf : typically recommended to insert graphs in Word documents;
  - *.eps : typically recommended to insert graphs in Tex/Latex documents.
- Many graphical options are available to change the appearance of the graphs.

16.8 Consumption dominance curves (cdomc)

Consumption dominance curves are useful tools for studying the impact of indirect tax fiscal reforms on poverty. The j^{th} **C**ommodity or **C**omponent dominance (C-Dominance for short) curve is defined as follows:

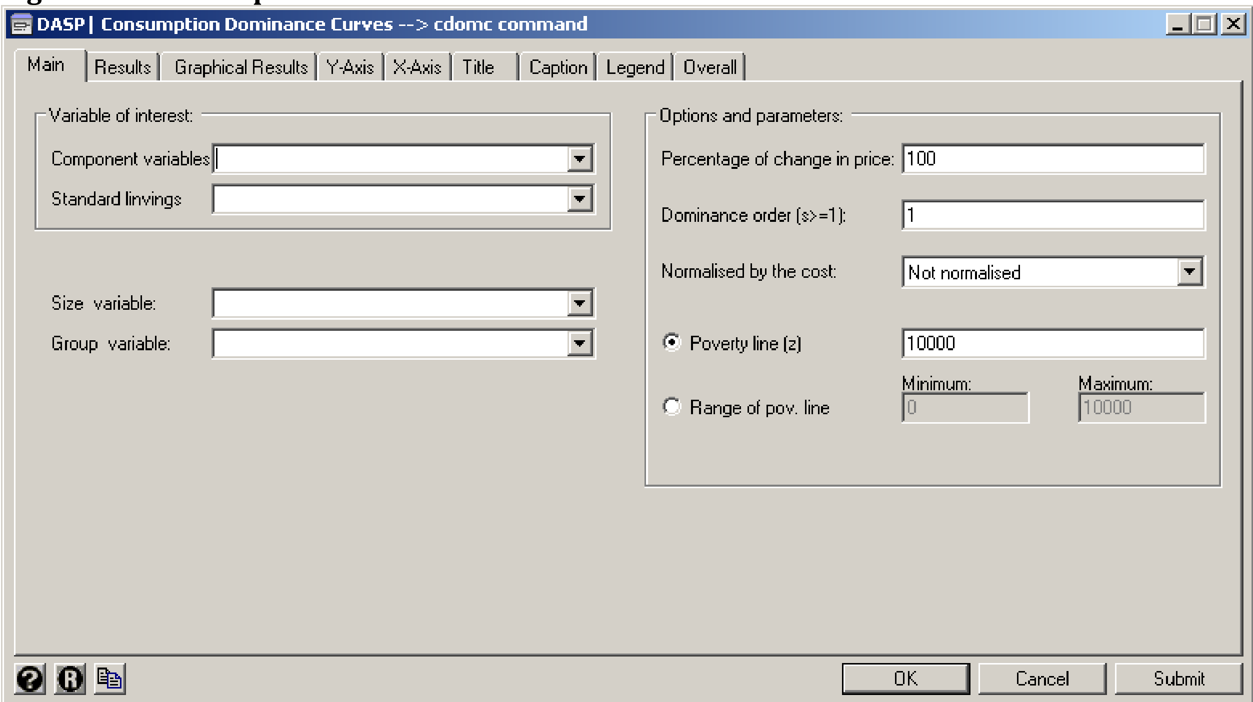
$$CD^j(z,s) = \begin{cases} \dfrac{\sum_{i=1}^{n} w_i\,(z-y_i)_+^{s-2}\,y_i^j}{\sum_{i=1}^{n} w_i} & if \quad s \geq 2 \\ E\left[y^j \mid y=z\right] f(z) = \dfrac{\sum_{i=1}^{n} w_i\,K(z-y_i)\,y_i^j}{\sum_{i=1}^{n} w_i} & if \quad s=1 \end{cases}$$

where K( ) is a kernel function and y^j is the j^{th} commodity. Dominance of order s is checked by setting $\alpha = s-1$. The **cdomc** module draws such curves easily. The module can:

- draw more than one CD curve simultaneously whenever more than one component is selected;
- draw the CD curves with confidence intervals;
- estimate the impact of change in price of a given component on FGT index (CD curve) for a specified poverty line;
- draw the normalized CD curves by the average of the component;
- list or save the coordinates of the curves;
- save the graphs in different formats:
  - *.gph : Stata format;
  - *.wmf : typically recommended to insert graphs in Word documents;
  - *.eps : typically recommended to insert graphs in Tex/Latex documents.
- Many graphical options are available to change the appearance of the graphs.

To open the dialog box of the module **cdomc**, type the command *db cdomc* in the command window.

Figure 15: Consumption dominance curves



16.9 Difference/Ratio between consumption dominance curves (cdomc2d)

The **cdomc2d** module draws difference or ratio between consumption dominance curves and their associated confidence intervals by taking sampling design into account. The module can:

- draw differences between consumption dominance curves and associated two-sided, lower-bounded or upper-bounded confidence intervals;
- list or save the coordinates of the differences and their confidence intervals;
- save the graphs in different formats:
  - *.gph : Stata format;
  - *.wmf : typically recommended to insert graphs in Word documents;
  - *.eps : typically recommended to insert graphs in Tex/Latex documents.
- Many graphical options are available to change the appearance of the graphs.

16.10 DASP and the progressivity curves

16.10.1 *Checking the progressivity of taxes or transfers*

The module cprog allows checking whether taxes or transfers are progressive.
Let X be a gross income, T be a given tax and B be a given transfer.
The tax T is Tax Redistribution (TR) progressive if :

$$\mathrm{PR}(p) = \mathrm{L}_X(p) - \mathrm{C}_T(p) > 0 \ \forall p \in]0,1[$$

The transfer B is Tax Redistribution (TR) progressive if :

$$\mathrm{PR}(\mathrm{p}) \;=\; \mathrm{C}_B(\mathrm{p})-\mathrm{L}_X(\mathrm{p})>0\ \forall p\in\left]0,1\right[$$

The tax T is Income Redistribution (IR) progressive if :

$$\mathrm{PR}(\mathrm{p}) \;=\; \mathrm{C}_{X-T}(\mathrm{p})-\mathrm{L}_X(\mathrm{p})>0\ \forall p\in\left]0,1\right[$$

The transfer B is Income Redistribution (IR) progressive if :

$$\mathrm{PR}(\mathrm{p}) \;=\; \mathrm{C}_{X+B}(\mathrm{p})-\mathrm{L}_X(\mathrm{p})>0\ \forall p\in\left]0,1\right[$$

16.10.2 *Checking the progressivity of transfer vs tax*

The module cprogbt allows checking whether a given transfer is more progressive than a given tax.

The transfer B is more Tax Redistribution (TR) progressive than a tax T if :

$$\mathrm{PR}(\mathrm{p}) \;=\; \mathrm{C}_B(\mathrm{p})+\mathrm{C}_T(\mathrm{p})-2\mathrm{L}_X(\mathrm{p})>0\ \forall p\in\left]0,1\right[$$

The transfer B is more Income Redistribution (TR) progressive than a tax T if :

$$\mathrm{PR}(\mathrm{p}) \;=\; \mathrm{C}_{X+B}(\mathrm{p})\text{-}\mathrm{C}_{X-T}(\mathrm{p})>0\ \forall p\in\left]0,1\right[$$

17 Dominance

17.1 Poverty dominance (dompov)

Distribution 1 dominates distribution 2 at order s over the range $\left[z^-,z^+\right]$ if only if:

$$P_1\ (\zeta;\alpha)<P_2\ (\zeta;\alpha)\quad\forall\quad\zeta\in\left[z^-,z^+\right]\ \text{for}\ \alpha=\mathrm{s}-1.$$

This involves comparing stochastic dominance curves at order *s* or FGT curves with $\alpha=s-1$. This application estimates the points at which there is a reversal of the ranking of the curves. Said differently, it provides the crossing points of the dominance curves, that is, the values of ζ and $P_1\ (\zeta;\alpha)$ for which $P_1\ (\zeta;\alpha)=P_2\ (\zeta;\alpha)$ when:
$sign(P_1\ (\zeta-\eta;\alpha)-P_2\ (\zeta-\eta;\alpha))=sign(P_2\ (\zeta+\eta;\alpha)-P_1\ (\zeta+\eta;\alpha))$ for a small η. The crossing points ζ can also be referred to as "critical poverty lines".

The **dompov** module can be used to check for poverty dominance and to compute critical values. This module is mostly based on Araar (2006):

Araar, Abdelkrim, (2006), Poverty, Inequality and Stochastic Dominance, Theory and Practice: Illustration with Burkina Faso Surveys, **Working Paper: 06-34**. CIRPEE, Department of Economics, Université Laval.

Interested users are encouraged to consider the exercises that appear in Section **23.6.**

17.2 Inequality dominance (domineq)

Distribution 1 inequality-dominates distribution 2 at the second order if and only if:

$$L_1(p) \leq L_2(p) \quad \forall \quad p \in [0,1]$$

The module **domineq** can be used to check for such inequality dominance. It is based mainly on Araar (2006):

Araar, Abdelkrim, (2006), Poverty, Inequality and Stochastic Dominance, Theory and Practice: Illustration with Burkina Faso Surveys, **Working Paper: 06-34**. CIRPEE, Department of Economics, Université Laval.

Intersections between curves can be estimated with this module. It can also used to check for tax and transfer progressivity by comparing Lorenz and concentration curves.

17.3 DASP and bi-dimensional poverty dominance (dombdpov)

Let two dimensions of well-being be denoted by $k = 1,2$. The intersection bi-dimensional FGT index for distribution D is estimated as

$$\hat{P}_D(Z;A) = \frac{\sum_{i=1}^{n} w_i \left[\prod_{k=1}^{2} (z^k - y_i^k)_+^{\alpha_k} \right]}{\sum_{i=1}^{n} w_i}$$

where $Z = (z_1, z_2)$ *and* $A = (\alpha_1, \alpha_2)$ *are* vectors of poverty lines and parameters α respectively, and $x_+ = \max(x,0)$.

Distribution 1 dominates distribution 2 at orders (s_1, s_2) over the range $\left[0, Z^+\right]$ if and only if:

$$P_1(Z; A = s-1) < P_2(Z; A = s-1) \quad \forall \quad Z \in \left[0, z_1^+\right] \times \left[0, z_2^+\right] \text{ and for } \alpha_1 = s_1 - 1, \quad \alpha_2 = s_2 - 1.$$

The DASP **dombdpov** module can be used to check for such dominance.

For each of the two distributions:

- The two variables of interest (dimensions) should be selected;
- Conditions can be specified to focus on specific population subgroups;
- Standard errors and confidence intervals with a confidence level of 95% are provided. Both the type of confidence intervals provided and the level of confidence used can be changed.

- Surfaces showing the difference, the lower bound and the upper bound of the confidence surfaces are plotted interactively with the GnuPlot tool.
- Coordinates can be listed.
- Coordinates can be saved in Stata or GnuPlot-ASCII format.

Interested users are encouraged to consider the exercises that appear in Section **23.12.**

18 Distributive tools

18.1 Quantile curves (c_quantile)

The quantile at a percentile *p* of a continuous population is given by:

$Q(p) = F^{-1}(p)$ where $p = F(y)$ is the cumulative distribution function at *y*.

For a discrete distribution, let *n* observations of living standards be ordered such that $y_1 \le y_2 \le \cdots \le y_i \le y_{i+1} \le \cdots \le y_n$. If $F(y_i) < p \le F(y_{i+1})$, we define $Q(p) = y_{i+1}$. The normalised quantile is defined as $\overline{Q}(p) = Q(p)/\mu$.

Interested users are encouraged to consider the exercises that appear in Section **23.10.**

18.2 Income share and cumulative income share by group quantiles (quinsh)

This module can be used to estimate the income shares, as well as, the cumulative income shares by quantile groups. The user can indicate the number of group partition. For instance, if the number is five, the quintile income shares are provided. We can also plot the graph bar of the estimated income shares.

18.3 Density curves (cdensity)

The Gaussian kernel estimator of a density function $f(x)$ is defined by

$$\hat{f}(x) = \frac{\sum_i w_i K_i(x)}{\sum_{i=1}^{n} w_i} \quad \textit{and} \quad K_i(x) = \frac{1}{h\sqrt{2\pi}} \exp\left(-0.5\, \lambda_i(x)^2\right) \quad \textit{and} \quad \lambda_i(x) = \frac{x - x_i}{h}$$

where h is a bandwidth that acts as a "smoothing" parameter.

Interested users are encouraged to consider the exercises that appear in Section **23.10.**

Boundary bias correction:

A problem occurs with kernel estimation when a variable of interest is bounded. It may be for instance that consumption is bounded between two bounds, a minimum and a maximum, and that we wish to estimate its density "close" to these two bounds. If the true value of the density at these two bounds is positive, usual kernel estimation of the density close to these two bounds will be biased. A similar problem occurs with non-parametric regressions.

Renormalisation approach:
One way to alleviate these problems is to use a smooth "corrected" Kernel estimator, following a paper by Peter Bearse, Jose Canals and Paul Rilstone. A boundary-corrected Kernel density estimator can then be written as

$$\hat{f}(x) = \frac{\sum_i w_i K_i^*(x) K_i(x)}{\sum_{i=1}^{n} w_i}$$

where

$$K_i(x) = \frac{1}{h\sqrt{2\pi}} \exp\left(-0.5\, \lambda_i(x)^2\right) \quad \text{and} \quad \lambda_i(x) = \frac{x - x_i}{h}$$

and where the scalar $K_i^*(x)$ is defined as

$$K_i^*(x) = \psi(x)'\, P(\lambda_i(x))$$

$$P(\lambda) = \begin{pmatrix} 1 & \lambda & \frac{\lambda^2}{2!} & \cdots & \frac{\lambda^{s-1}}{(s-1)!} \end{pmatrix}$$

$$\psi(x) = M^{-1} 1_s{}' = \left(\int_A^B K(\lambda) P(\lambda) P(\lambda)' d\lambda \right)^{-1} 1_s \quad : A = \frac{x - \max}{h}, \quad B = \frac{x - \min}{h}, \quad 1_s{}' = (1 \quad 0 \quad 0 \cdots 0)$$

min is the minimum bound, and *max* is the maximum one. *h* is the usual bandwidth. This correction removes bias to order h^s.

DASP offers four options, without correction, and with correction of order 1, 2 and 3.
Refs:

- Jones, M. C. 1993, simply boundary correction for Kernel density estimation. Statistics and Computing 3: 135-146.
- Bearse, P., Canals, J. and Rilstone, P. Efficient Semi parametric Estimation of Duration Models With Unobserved Heterogeneity, *Econometric Theory*, **23**, 2007, 281–308

Reflection approach:
The reflection estimator approaches the boundary estimator by "reflecting" the data at the boundaries:

$$\hat{f}(x) = \frac{\sum_i w_i K_i^r(x)}{\sum_{i=1}^{n} w_i}$$

$$K^r(x) = K\left(\frac{x - X}{h}\right) + K\left(\frac{x + X - 2\min}{h}\right) K\left(\frac{x + X - 2\max}{h}\right)$$

Refs:

- Cwik and Mielniczuk (1993), *Data-dependent Bandwidth Choice for a Grade Density Kernel Estimate*. Statistics and probability Letters 16: 397-405

- Silverman, B. W. (1986), *Density for Statistics and Data Analysis.* London Chapman and Hall (p 30).

18.4 Non-parametric regression curves (cnpe)

Non-parametric regression is useful to show the link between two variables without specifying beforehand a functional form. It can also be used to estimate the local derivative of the first variable with respect to the second without having to specify the functional form linking them. Regressions with the **cnpe** module can be performed with one of the following two approaches:

18.4.1 Nadaraya-Watson approach

A Gaussian kernel regression of y on x is given by:

$$E\left(y|x\right) = \Phi(y \mid x) = \frac{\sum_i w_i K_i(x) y_i}{\sum_i w_i K_i(x)}$$

From this, the derivative of $\Phi(y \mid x)$ with respect to *x* is given by

$$E\left(\frac{dy}{dx}|x\right) = \frac{\partial \Phi(y \mid x)}{\partial x}$$

18.4.2 Local linear approach

The local linear approach is based on a local OLS estimation of the following functional form:

$$K_i(x)^{1/2} y_i = \mu(x) K_i(x)^{1/2} + \mu'(x) K_i(x)^{1/2} (x_i - x) + v$$

or, alternatively, of:

$$K_i(x)^{1/2} y_i = \alpha K_i(x)^{1/2} + \beta K_i(x)^{1/2} (x_i - x) + v_i$$

Estimates are then given by:

$$E\left(y|x\right) = \alpha \,,\; E\left(\frac{dy}{dx}|x\right) = \beta$$

Interested users are encouraged to consider the exercises that appear in Section **23.10.**

18.5 DASP and joint density functions.

The module **sjdensity** can be used to draw a joint density surface. The Gaussian kernel estimator of the joint density function $f(x, y)$ is defined as:

$$\hat{f}(\bar{x},\bar{y}) = \frac{1}{2\pi h_x h_y \sum_{i=1}^{n} w_i} \sum_{i=1}^{n} w_i \ \exp\left(-\left(\frac{1}{2}\right)\left(\left(\frac{\bar{x}-x_i}{h_x}\right)^2 + \left(\frac{\bar{y}-y_i}{h_y}\right)^2\right)\right)$$

With this module:

- The two variables of interest (dimensions) should be selected;
- specific population subgroup can be selected;
- surfaces showing the joint density function are plotted interactively with the GnuPlot tool;
- coordinates can be listed;c
- coordinates can be saved in Stata or GnuPlot-ASCII format.

Interested users are encouraged to consider the exercises that appear in Section **23.11???**

18.6 DASP and joint distribution functions

The module **sjdistrub** can be used to draw joint distribution surfaces. The joint distribution function $F(\bar{x},\bar{y})$ is defined as:

$$\hat{F}(\bar{x},\bar{y}) = \frac{\sum_{i=1}^{n} w_i \ \ I(x_i \le \bar{x})I(y_i \le \bar{y})}{\sum_{i=1}^{n} w_i}$$

With this module:

- The two variables of interest (dimensions) should be selected;
- specific population subgroups can be selected;
- surfaces showing the joint distribution function are plotted interactively with the GnuPlot tool;
- coordinates can be listed;
- coordinates can be saved in Stata or GnuPlot-ASCII format.

Interested users are encouraged to consider the exercises that appear in Section **23.11**

19 *DASP* and pro-poor growth

19.1 DASP and pro-poor indices

The module **ipropoor** estimates simultaneously the three following pro-poor indices:

1. ***The Chen and Ravallion pro-poor index (2003):***

$$Index = \frac{W_1(z) - W_2(z)}{F_1(z)}$$

where $W_D(z)$ is the Watts index for distribution $D \in [1,2]$ and $F_1(z)$ is the headcount for index for the first distribution, both with poverty lines *z*.

2. ***The Kakwani and Pernia pro-poor index (2000):***

$$Index = \frac{P_1(z,\alpha) - P_2(z\alpha)}{P_1(z,\alpha) - P_1(z(\mu_1/\mu_2),\alpha)}$$

3. ***The Kakwani, Khandker and Son pro-poor index (2003):***

$$Index\_1 = g\frac{P_1(z,\alpha) - P_2(z\alpha)}{P_1(z,\alpha) - P_1(z(\mu_1/\mu_2),\alpha)}$$

where the average growth is $g(\mu_2 - \mu_1)/\mu_1$ and where a second index is given by:

$$Index\_2 = Index\_1 - g$$

- One variable of interest should be selected for each distribution.
- Conditions can be specified to focus on specific population subgroups.
- Standard errors and confidence intervals with a confidence level of 95% are provided. Both the type of confidence intervals provided and the level of confidence used can be changed.
- The results are displayed with 6 decimals; this can be changed.
- A level for the parameter α can be chosen for each of the two distributions.

19.2 DASP and pro-poor curves

Pro-poor curves can be drawn using either the primal or the dual approach. The former uses income levels. The latter is based on percentiles.

19.2.1 Primal pro-poor curves

The change in the distribution from state 1 to state 2 is s-order absolutely pro-poor with standard $cons$ if:

$$\Delta(z,s) = \left(P_2(z+cons,\alpha=s-1) - P_1(z,\alpha=s-1)\right) < 0 \ \forall\, z \in \left[0,z^+\right]$$

The change in the distribution from state 1 to state 2 is s-order relatively pro-poor if:

$$\Delta(z,s) = z\left(P_2(z\frac{\mu_2}{\mu_1},\alpha=s-1) - P_1\left(z,\alpha=s-1\right)\right) < 0 \ \forall\, z \in \left[0,z^+\right]$$

The module **cpropoorp** can be used to draw these primal pro-poor curves and their associated confidence interval by taking into account sampling design. The module can:

- draw pro-poor curves and their two-sided, lower-bounded or upper-bounded confidence intervals;
- list or save the coordinates of the differences between the curves as well as those of the confidence intervals;
- save the graphs in different formats:
    - *.gph : Stata format;
    - *.wmf : typically recommended to insert graphs in Word documents;
    - *.eps : typically recommended to insert graphs in Tex/Latex documents.

Many graphical options are available to change the appearance of the graphs.

Interested users are encouraged to consider the exercises that appear in Section **23.13.**

19.2.2 Dual pro-poor curves

Let:

$Q(p)$: quantile at percentile p.
$GL(p)$: Generalised Lorenz curve at percentile p.
μ : average living standards.
The change in the distribution from state 1 to state 2 is first-order absolutely pro-poor with standard *cons=0* if:

$$\Delta(z,s) = Q_2(p) - Q_1(p) > 0 \ \forall \ p \in \left[0, \mathrm{p}^+ = F(z^+)\right]$$

or equivalently if:

$$\Delta(z,s) = \frac{Q_2(p) - Q_1(p)}{Q_1(p)} > 0 \ \forall \ p \in \left[0, \mathrm{p}^+ = F(z^+)\right]$$

The change in the distribution from state 1 to state 2 is first-order relatively pro-poor if:

$$\Delta(z,s) = \frac{Q_2(p)}{Q_1(p)} - \frac{\mu_2}{\mu_1} > 0 \ \forall \ p \in \left[0, \mathrm{p}^+ = F(z^+)\right]$$

The change in the distribution from state 1 to state 2 is second-order absolutely pro-poor if:

$$\Delta(z,s) = GL_2(p) - GL_1(p) > 0 \ \forall \ p \in \left[0, \mathrm{p}^+ = F(z^+)\right]$$

or equivalently if:

$$\Delta(z,s) = \frac{GL_2(p) - GL_1(p)}{GL_1(p)} > 0 \ \forall \ p \in \left[0, \mathrm{p}^+ = F(z^+)\right]$$

The change in the distribution from state 1 to state 2 is first-order relatively pro-poor if:

$$\Delta(z,s) = \frac{GL_2(p)}{GL_1(p)} - \frac{\mu_2}{\mu_1} > 0 \ \forall\ p \in \left[0, \mathrm{p}^+ = F(z^+)\right]$$

The module **cpropoord** can be used to draw these dual pro-poor curves and their associated confidence interval by taking into account sampling design. The module can:

- draw pro-poor curves and their two-sided, lower-bounded or upper-bounded confidence intervals;
- list or save the coordinates of the differences between the curves as well as those of the confidence intervals;
- save the graphs in different formats:
  - *.gph : Stata format;
  - *.wmf : typically recommended to insert graphs in Word documents;
  - *.eps : typically recommended to insert graphs in Tex/Latex documents.

Many graphical options are available to change the appearance of the graphs.

Interested users are encouraged to consider the exercises that appear in Section **23.13**

20 *DASP* and Benefit Incidence Analysis

20.1 Benefit incidence analysis

The main objective of benefit incidence is to analyse the distribution of benefits from the use of public services according to the distribution of living standards.

Two main sources of information are used. The first informs on access of household members to public services. This information can be found in usual household surveys. The second deals with the amount of total public expenditures on each public service. This information is usually available at the national level and sometimes in a more disaggregated format, such as at the regional level. The benefit incidence approach combines the use of these two sources of information to analyse the distribution of public benefits and its progressivity.

Formally, let

$\mathrm{w_i}$ be the sampling weight of observation i;

$\mathrm{y_i}$ be the living standard of members belonging to observation i (i.e., *per capita* income);

$\mathrm{e_i^s}$ be the number of "eligible" members of observation *i*, i.e., members that "need" the public service provided by sector *s*. There are *S* sectors;

$\mathrm{f_i^s}$ be the number of members of observation *i* that effectively use the public service provided by sector *s*;

$\mathrm{g_i}$ be the socio-economic group of eligible members of observation *i* (typically classified by income percentiles);

c_i be a subgroup indicator for observation *i* (e.g., 1 for a rural resident, and 2 for an urban resident). Eligible members can thus be grouped into population exclusive subgroups;

E_r^s be total public expenditures on sector s in area r. There are R areas (the area here refers to the geographical division which one can have reliable information on total public expenditures on the studied public service);

E^s be total public expenditures on sector s $\left(E^s = \sum_{r=1}^{R} E_r^s\right)$.

Here are some of the statistics that can be computed.

1. The share of a g in sector s is defined as follows:

$$SH_g^s = \frac{\sum_{i=1}^{n} w_i f_i^s I(i \in g)}{\sum_{i=1}^{n} w_i f_i^s}$$

Note that: $\sum_{g=1}^{G} SH_g^s = 1$.

2. The rate of participation of a group g in sector s is defined as follows:

$$CR_g^s = \frac{\sum_{i=1}^{n} w_i f_i^s I(i \in g)}{\sum_{i=1}^{n} w_i e_i^s I(i \in g)}$$

This rate cannot exceed 100% since $f_i^s \leq e_i^s \ \forall i$.

3. The unit cost of a benefit in sector s for observation j, which refers to the household members that live in area r :

$$UC_j^s = \frac{E_r^s}{\sum_{j=1}^{n_r} w_j f_j^s}$$

where n_r is the number of sampled households in area r.

4. The benefit of observation i from the use of public sector s is:

$$B_i^s = f_i^s UC_i^s$$

5. The benefit of observation i from the use of the S public sectors is:

$$B_i = \sum_{s=1}^{S} B_i^s$$

6. The average benefit at the level of those eligible to a service from sector s and for those observations that belong to a group g, is defined as:

$$ABE_g^s = \frac{\sum_{i=1}^{n} w_i B_i^s I(i \in g)}{\sum_{i=1}^{n} w_i e_i^s I(i \in g)}$$

7. The average benefit for those that use the service s and belong to a group g is defined as:

$$ABF_g^s = \frac{\sum_{i=1}^{n} w_i B_i^s I(i \in g)}{\sum_{i=1}^{n} w_i f_i^s I(i \in g)}$$

8. The proportion of benefits from the service from sector s that accrues to observations that belong to a group g is defined as:

$$PB_g^s = \frac{B_g^s}{E^s}$$

where $B_g^s = \sum_{i=1}^{n} w_i B_i^s I(i \in g)$.

These statistics can be restricted to specific socio-demographic groups (e.g.,. rural/urban) by replacing $I(i \in g)$ by $I(i \in c)$.

.

The **bian.ado** module allows the computation of these different statistics.

Some characteristics of the module:

- Possibility of selecting between one and six sectors.
- Possibility of using frequency data approach when information about the level of total public expenditures is not available.
- Generation of benefit variables by the type of public services (ex: primary, secondary and tertiary education levels) and by sector.
- Generation of unit cost variables for each sector.
- Possibility of computing statistics according to groups of observations.
- Generation of statistics according to social-demographic groups, such as quartiles, quintiles or deciles.

Public expenditures on a given service often vary from one geographical or administrative area to another. When information about public expenditures is available at the level of areas, this information can be used with the ***bian*** module to estimate unit cost more accurately.

<u>Example 1</u>

Observation i	HH size	Eligible HH members	Frequency	Area indicator	Total level of regional public expenditures
1	**7**	**3**	**2**	**1**	**14000**
2	**4**	**2**	**2**	**1**	**14000**
3	**5**	**5**	**3**	**1**	**14000**
4	**6**	**3**	**2**	**2**	**12000**
5	**4**	**2**	**1**	**2**	**12000**

In this example, the first observation contains information on household 1.

- This household contains 7 individuals;
- Three individuals in this household are eligible to the public service;
- Only 2 among the 3 eligible individuals benefit from the public service;
- This household lives in area 1. In this area, the government spends a total of 14000 to provide the public service for the 7 users of this area (2+2+3).

The unit cost in area 1 equals: 14000/7=2000
The unit cost in area 2 equals: 12000/3=4000

By default, the area indicator is set to 1 for all households. When this default is used, the variable Regional public expenditures (the fifth column that appears in the dialog box) should be set to total public expenditures at the national level. This would occur when the information on public expenditures is only available at the national level.

<u>Example 2</u>

Observation i	HH size	Eligible members	Frequency	Area indicator	Regional public expenditures
1	**7**	**3**	**2**	**1**	**28000**
2	**4**	**2**	**2**	**1**	**28000**
3	**5**	**5**	**3**	**1**	**28000**
4	**6**	**3**	**2**	**1**	**28000**
5	**4**	**2**	**1**	**1**	**28000**

The unit cost benefit (at the national level) equals: 28000/10=2800

Interested users are encouraged to consider the exercises that appear in Section **23.14**

21 Disaggregating grouped data

The **ungroup** *DASP* module generates disaggregated data from aggregate distributive information. Aggregate information is obtained from cumulative income shares (or Lorenz curve ordinates) at some percentiles. For instance:

Percentile (p)	0.10	0.30	0.50	0.60	0.90	1.00
Lorenz values: L(p)	0.02	0.10	0.13	0.30	0.70	1.00

The user must specify the total number of observations to be generated. The user can also indicate the number of observations to be generated specifically at the top and/or at the bottom of the distribution, in which case the proportion (in %) of the population found at the top or at the bottom must also be specified.

Remarks:

- If only the total number of observations is set, the generated data are self weighted (or uniformly distributed over percentiles).
- If a number of observations is set for the bottom and/or top tails, the generated data are not self weighted and a weight variable is provided in addition to the generated income variable.
  - Example: Assume that the total number of observations to be generated is set to 1900, but that we would like the bottom 10% of the population to be represented by 1000 observations. In this case, weights will equal 1/1000 for the bottom 1000 observations and 1/100 for the remaining observations (the sum of weights being normalized to one).
- The generated income vector takes the name of _y and the vector weight, _w.
- The number of observations to be generated does not have to equal the number of observations of the sample that was originally used to generate the aggregated data. The **ungroup** module cannot in itself serve to estimate the sampling errors that would have occurred had the original sample data been used to estimate poverty and/or inequality estimates.
- The user can select any sample size that exceeds (*number_of_classes*+1), but it may be more appropriate for statistical bias-reduction purposes to select relatively large sizes.

STAGE I ***Generating an initial distribution of incomes and percentiles***

S.1.1: Generating a vector of percentiles

Starting from information on the importance of bottom and top groups and on the number of observations to be generated, we first generate a vector of percentiles.

Examples:
Notations:
NOBS: number of total observations
F: vector of percentiles
B_NOBS: number of observations for the bottom group
T_NOBS: number of observations for the top group.

- For NOBS=1000 spread equally across all percentiles, F=0.001, 0.002... 0.999, 1. To avoid the value F=1 for the last generated observation, we can simply replace F by F-(0.5/NOBS).
- For NOBS=2800, B_NOBS=1000 and T_NOBS=1000, with the bottom and top groups being the first and last deciles:
    a. F=0.0001, 0.0002,..., 0.0999, 0.1000 in 0001/1000
    b. F=0.1010, 0.1020,…, 0.8990, 0.9000 in 1001/1800
    c. F=0.9001, 0.9002,..., 0.9999, 1.0000 in 1801/2800

  Adjustments can also be made to avoid the case of F(1)=1.

The weight vector can easily be generated.

S.1.2: Generating an initial distribution of incomes

The user must indicate the form of distribution of the disaggregated data.

***- Normal and log normal distributions*:**
Assume that x follows a lognormal distribution with mean μ and variance σ^2. The Lorenz curve is defined as follows:

$$L(p) = \Phi\left(\frac{Ln(x)-(\mu-\sigma^2)}{\sigma}\right) \text{ and } p = \Phi\left(\frac{Ln(x)-\mu}{\sigma}\right)$$

We assume that $\mu = 1$ and we estimate the variance using the procedure suggested by Shorrrocks and Wan (2008): a value for the standard deviation of log incomes, σ, is obtained by averaging the $m-1$ estimates of $\sigma^k = \Phi^{-1}\left(p_k\right) - \Phi^{-1}\left(L(p_k)\right) \quad k = 1, \ \ldots, \ m-1$
where m is the number of classes and Φ is the standard normal distribution function (Aitchison and Brown 1957; Kolenikov and Shorrocks 2005, Appendix).

***- Generalized Quadratic Lorenz Curve*:**
It is assumed that:

$$L(1-L) = a(p^2 - L) + bL(p-1) + c(p-L)$$

We can regress $L(1-L)$ on $(p^2 - L)$, $L(p-1)$ and $(p-L)$ without an intercept, dropping the last observation since the chosen functional form forces the curve to go through (1,1).

We have $Q(p) = -\frac{b}{2} - \frac{\left(2mp+n\right)\left(mp^2+np+e^2\right)^{-0.5}}{4}$

$$
\begin{aligned}
e &= a+b+c+1 \\
m &= b^2-4a \\
n &= -2be-4c
\end{aligned}
$$

***- Beta Lorenz Curve*:**

It is assumed that:

$$\log\left(p-L\right)=\log(\theta)+\gamma\log(p)+\delta\log(1-p)$$

After estimating the parameters, we can generate quantiles as follows

$$Q\left(p\right)=\theta+p^{\gamma}\left(1-p\right)^{\delta}\left[\frac{\gamma}{p}-\frac{\delta}{\left(1-p\right)}\right]$$

See also Datt (1998).

- The Singh-Maddala distribution

The distribution function proposed by Singh and Maddala (1976) takes the following form:

$$F(x)=1-\left[\frac{1}{1+(x/b)^{a}}\right]^{q}$$

where $a\geq 0, b\geq 0, q\geq 1/a$ are parameters to be estimated. The income (x) is assumed to be equal to or greater than zero. The density function is defined as follows:

$$f(x)=\left(aq/b\right)\left(1+\left(x/b\right)^{a}\right)^{-(q+1)}\left(x/b\right)^{(a-1)}$$

Quantiles are defined as follows:

$$Q(p)=b\left(\left(1-p\right)^{-1/q}-1\right)^{1/a}$$

We follow Jenkins (2008)'s approach for the estimation of parameters. For this, we maximize the likelihood function, which is simply the product of density functions evaluated at the average income of each class:

http://stata-press.com/journals/stbcontents/stb48.pdf

<u>STAGE II</u> *Adjusting the initial distribution to match the aggregated data (optional).*

This stage adjusts the initial vector of incomes using the Shorrocks and Wan (2008) procedure. This procedure proceeds with two successive adjustments:

- *<u>Adjustment 1:</u>* Correcting the initial income vector to ensure that each income group has its original mean income.
- *<u>Adjustment 2:</u>* Smoothing the inter–class distributions.

The generated sample is saved automatically in a new Stata data file (called by default *ungroup_data.dta*; names and directories can be changed). The user can also plot the Lorenz curves of the aggregated (when we assume that each individual has the average income of his group) and generated data.

Dialog box of the **ungroup** module

Figure 16: ungroup dialog box

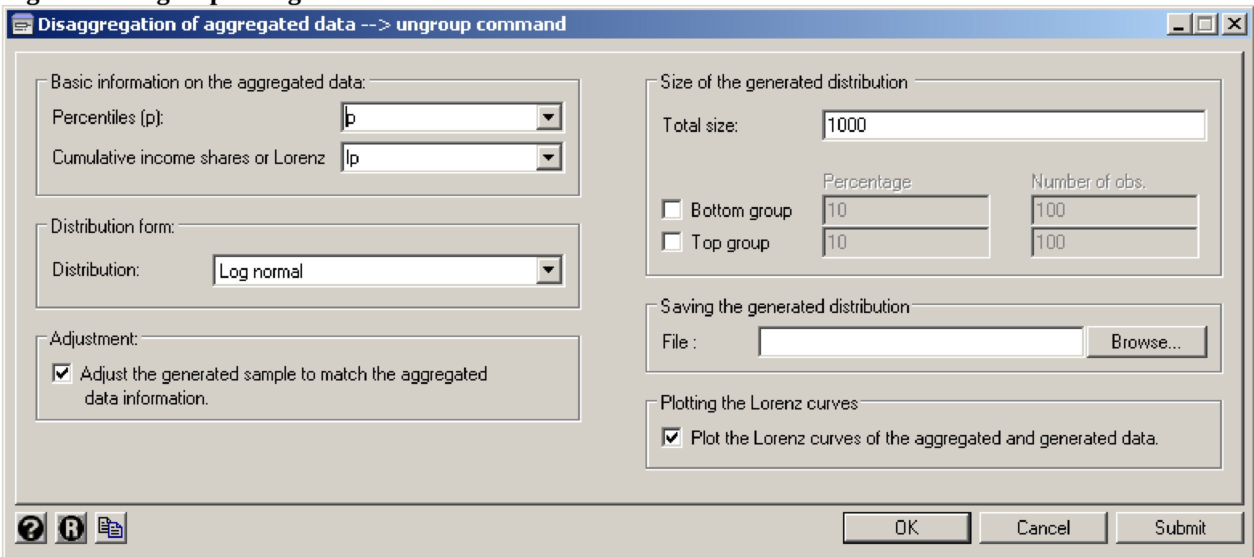


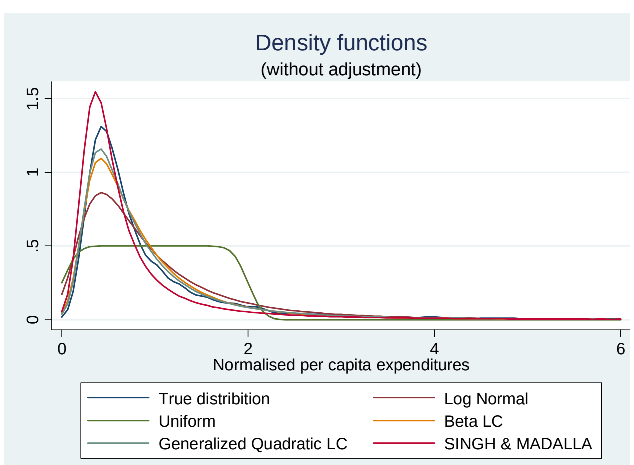
Illustration with Burkina Faso household survey data

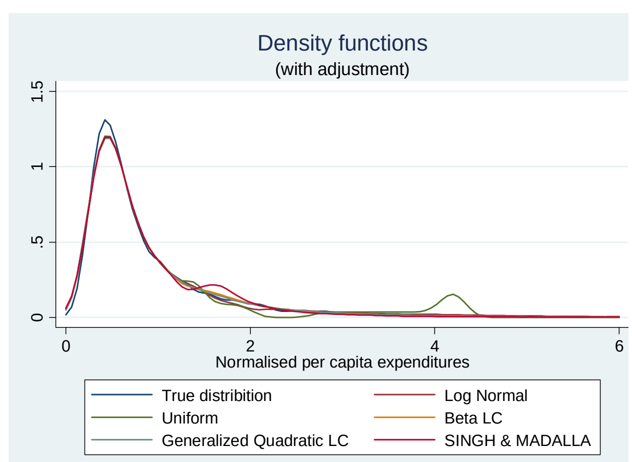
In this example, we use disaggregated data to generate aggregated information. Then, we compare the density curve of the true data with those of the data generated through disaggregation of the previously aggregated data.

```
gen fw=size*weight
gen y=exppc/r(mean)
clorenz y, hs(size) lres(1)
```

Aggregated information:

p	L(p)
.1	.0233349
.2	.0576717
.3	.0991386
.4	.1480407
.5	.2051758
.6	.2729623
.7	.3565971
.8	.4657389
.9	.6213571
1	1.00000





22 Appendices

22.1 Appendix A: illustrative household surveys

22.1.1 The 1994 Burkina Faso survey of household expenditures (bkf94I.dta)

This is a nationally representative survey, with sample selection using two-stage stratified random sampling. Seven strata were formed. Five of these strata were rural and two were urban. Primary sampling units were sampled from a list drawn from the 1985 census. The last sampling units were households.

List of variables

strata Stratum in which a household lives

psu Primary sampling unit

weight Sampling weight

size Household size

exp Total household expenditures

expeq Total household expenditures per adult equivalent

expcp Total household expenditures per capita

gse Socio-economic group of the household head
- 1 wage-earner (public sector)
- 2 wage-earner (private sector)
- 3 Artisan or trader
- 4 Other type of earner
- 5 Crop farmer
- 6 Subsistence farmer
- 7 Inactive

sex Sex of household head
- 1 Male
- 2 Female

zone Residential area
- 1 Rural
- 2 Urban

22.1.2 The 1998 Burkina Faso survey of household expenditures (bkf98I.dta)

This survey is similar to the 1994 one, although ten strata were used instead of seven for 1994. To express 1998 data in 1994 prices, two alternative procedures have been used. First, 1998 expenditure data were multiplied by the ratio of the 1994 official poverty line to the 1998 official poverty line: z_1994/z_1998. Second, 1998 expenditure data were multiplied by the ratio of the 1994 consumer price index to the 1998 consumer price index: ipc_1994/ipc_1998.

List of new variables

expcpz Total household expenditures per capita deflated by (z_1994/z_1998)

expcpi Total expenditures per capita deflated by (ipc_1994/ipc_1998)

22.1.3 Canadian Survey of Consumer Finance (a sub sample of 1000 observations – can6.dta)

List of variables

X Yearly gross income per adult equivalent.

T Income taxes per adult equivalent.

B1 Transfer 1 per adult equivalent.

B2 Transfer 2 per adult equivalent.

B3 Transfer 3 per adult equivalent.

B Sum of transfers *B1*, *B2* and *B3*

N Yearly net income per adult equivalent (*X* minus *T* plus *B*)

22.1.4 Peru LSMS survey 1994 (A sample of 3623 household observations - PEREDE94I.dta)

List of variables

exppc Total expenditures, per capita (constant June 1994 soles per year).

weight Sampling weight

size Household size

npubprim Number of household members in public primary school

npubsec Number of household members in public secondary school

npubuniv Number of household members in public post-secondary school

22.1.5 Peru LSMS survey 1994 (A sample of 3623 household observations – PERU_A_I.dta)

List of variables

hhid Household Id.

exppc Total expenditures, per capita (constant June 1994 soles per year).

size Household size

literate Number of literate household members

pliterate literate/size

22.1.6 The 1995 Colombia DHS survey (columbiaI.dta)

This sample is a part of the Data from the Demographic and Health Surveys (Colombia_1995) witch contains the following information for children aged 0-59 months

List of variables

hid Household id
haz height-for-age
waz weight-for-age
whz weight-for-height
sprob survival probability
wght sampling weight
Asset asset index

22.1.7 The 1996 Dominican Republic DHS survey (Dominican_republic1996I.dta)

This sample is a part of the Data from the Demographic and Health Surveys (Republic Dominican_1996) witch contains the following information for children aged 0-59 months

List of variables

hid	Household id
haz	height-for-age
waz	weight-for-age
whz	weight-for-height
sprob	survival probability
wght	sampling weight
Asset	asset index

22.2 *Appendix B: labelling variables and values*

- The following .do file can be used to set labels for the variables in ***bkf94.dta.***
- For more details on the use of *label* command, type *help label* in the command window.

```
================================lab_bkf94.do =================================

# delim ;

/*   To drop all label values   */

label drop _all;

/* To assign labels */
label var strata   "Stratum in which a household lives";

label var psu      "Primary sampling unit";
label var weight   "Sampling weight";
label var size     "Household size";
label var totexp   "Total household expenditures";
label var exppc    "Total household expenditures per capita";
label var expeq    "Total household expenditures per adult equivalent";
label var gse      "Socio-economic group of the household head";

/* To define the label values that will be assigned to the categorical variable gse */

label define lvgse

1 "wage-earner (public sector)"
2 "wage-earner (private sector)"
3 "Artisan or trader"
4 "Other type of earner"
5 "Crop farmer"
6 "Subsistence farmer"
7 "Inactive"
;

/*To assign the label values "lvgse" to the variable gse */
label val gse lvgse;

label var sex      "Sex of household head";
```

label def lvsex
1 Male
2 Female
;

label val sex lvsex;

label var zone   "Residential area";
label def lvzone
1 Rural
2 Urban
;

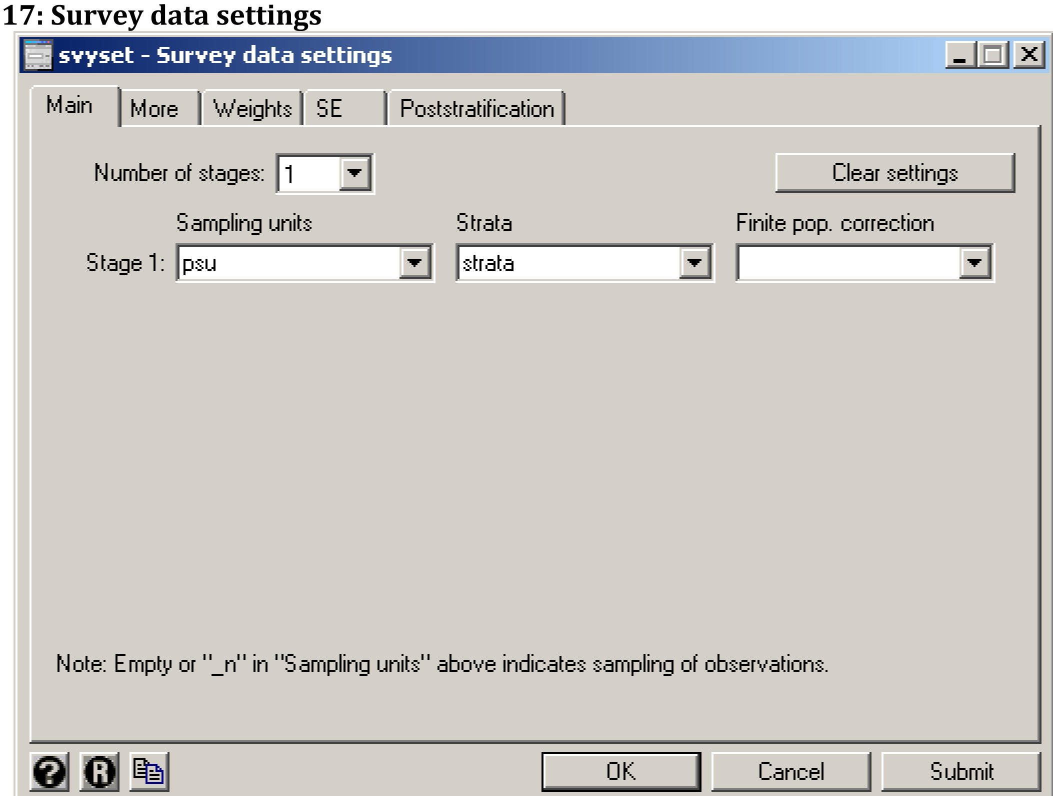
label val zone lvzone;

===================================End=====================================

22.3 Appendix C: setting the sampling design

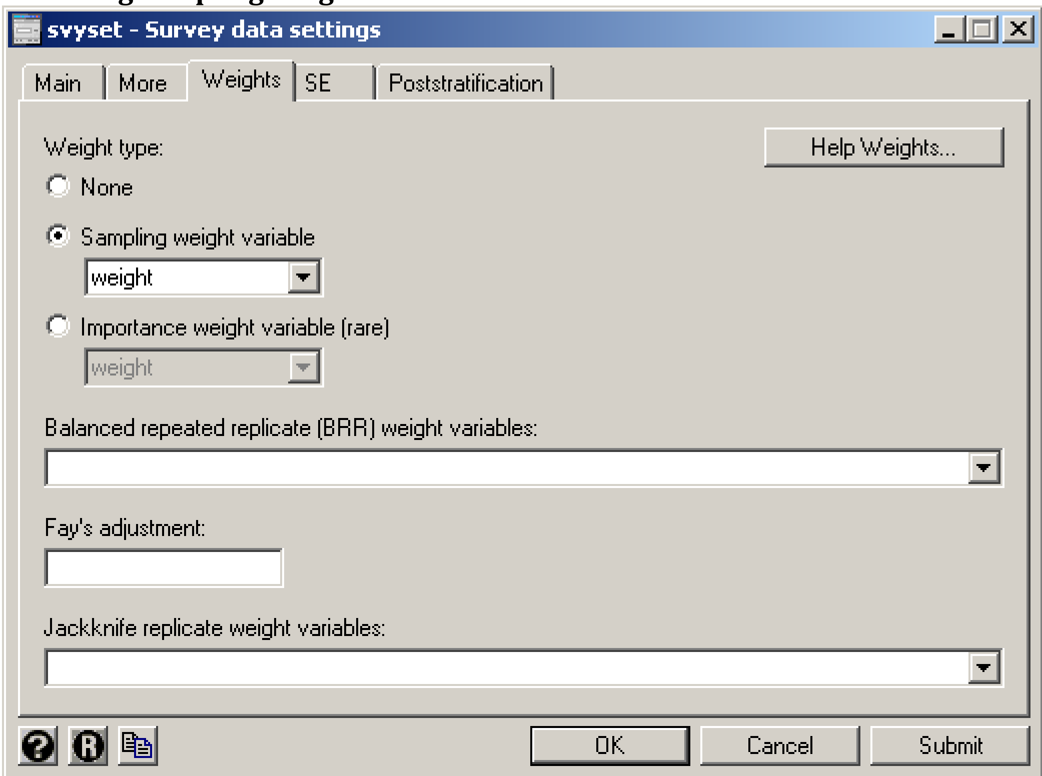
To set the sampling design for the data file ***bkf94.dta***, open the dialog box for the command *svyset* by typing the syntax *db svyset* in the command window. In the Main panel, set STRATA and SAMPLING UNITS as follows:

Figure 17: Survey data settings



In the Weights panel, set SAMPLING WEIGHT VARIABLE as follows:

Figure 18: Setting sampling weights



Click on OK and save the data file.

To check if the sampling design has been well set, type the command *svydes*. The following will be displayed:

```
Survey: Describing stage 1 sampling units

      pweight: weight
          VCE: linearized
     Strata 1: strata
         SU 1: psu
        FPC 1: <zero>

                                          #Obs per Unit
                                 ----------------------------
Stratum    #Units     #Obs       min       mean       max
--------   --------   --------   --------   --------   --------
       1         42        838         19       20.0         20
       2         37        733         17       19.8         20
       3         98       1959         19       20.0         20
       4         55       1093         19       19.9         20
       5         66       1286         13       19.5         20
       6         41        779          1       19.0         20
       7         97       1937         19       20.0         20
--------   --------   --------   --------   --------   --------
       7        436       8625          1       19.8         20
```

23 Examples and exercises

23.1 Estimation of FGT poverty indices

"How poor was Burkina Faso in 1994?"

1. Open the ***bkf94.dta*** file and label variables and values using the information of Section 22.1.1. Type the *describe* command and then *label list* to list labels.
2. Use the information of Section 22.1.1. to set the sampling design and then save the file.
3. Estimate the headcount index using variables of interest *expcc* and *expeq*.
    a. You should set SIZE to household size in order to estimate poverty over the population of individuals.
    b. Use the so-called 1994 official poverty line of 41099 Francs CFA per year.
4. Estimate the headcount index using the same procedure as above except that the poverty line is now set to 60% of the median.
5. Using the official poverty line, how does the headcount index for male- and female-headed households compare?
6. Can you draw a 99% confidence interval around the previous comparison? Also, set the number of decimals to 4.

Answer
Q.1
If ***bkf94.dta*** is saved in the directory ***c:/data***, type the following command to open it:
use "C:\data\bkf94.dta", clear

If ***lab_bkf94.do*** is saved in the directory ***c:/do_files***, type the following command to label variables and labels:

do "C:\do_files\lab_bkf94.do"

Typing the command *describe*, we obtain:

```
obs:      8,625
vars:     9                   31 Oct 2006 13:48
size:     285,087 (99.6% of          memory free)
storage  display    value
variable  name      type      format          label           variable label

weight              float     %9.0g                           Sampling weight
size                byte      %8.0g                           Household size
strata              byte      %8.0g                           Stratum in which a household lives
psu                 byte      %8.0g                           Primary sampling unit
gse                 byte      %29.0g          gse             Socio-economic group of the household head
sex                 byte      %8.0g           sex             Sex of household head
zone                byte      %8.0g           zone            Residential area
exp                 double    %10.0g                          Total household expenditures
expeq               double    %10.0g                          Total household expenditures per adult equivalent
exppc               float     %9.0g                           Total household expenditures per capita
```

Typing *label list*, we find:

```
zone:
          1         Rural
          2         Urban
```

sex:

1	Male
2	Female

gse:

1	wage-earner (public sector)
2	wage-earner (private sector)
3	Artisan or trader
4	Other type of earner
5	Crop farmer
6	Food farmer
7	Inactive

Q.2

You can set the sampling design with a dialog box, as indicated in Section 22.3, or simply by typing

svyset psu [pweight=weight], strata(strata) vce(linearized)

Typing *svydes*, we obtain

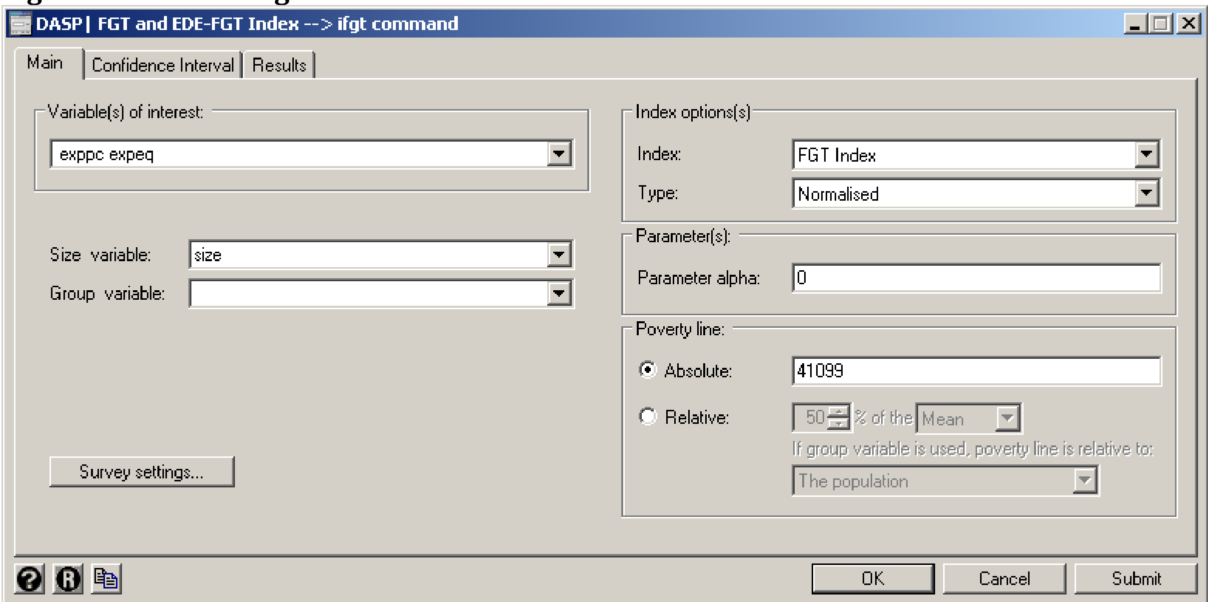
```
Survey: Describing stage 1 sampling units

      pweight: weight
          VCE: linearized
     Strata 1: strata
         SU 1: psu
        FPC 1: <zero>
```

			#Obs per Unit		
Stratum	#Units	#Obs	min	mean	max
1	42	838	19	20.0	20
2	37	733	17	19.8	20
3	98	1959	19	20.0	20
4	55	1093	19	19.9	20
5	66	1286	13	19.5	20
6	41	779	1	19.0	20
7	97	1937	19	20.0	20
7	436	8625	1	19.8	20

Q.3

Type *bd ifgt* to open the dialog box for the FGT poverty index and choose variables and parameters as indicated in the following window. Click on SUBMIT**.**

Figure 19: Estimating FGT indices



The following results should then be displayed:

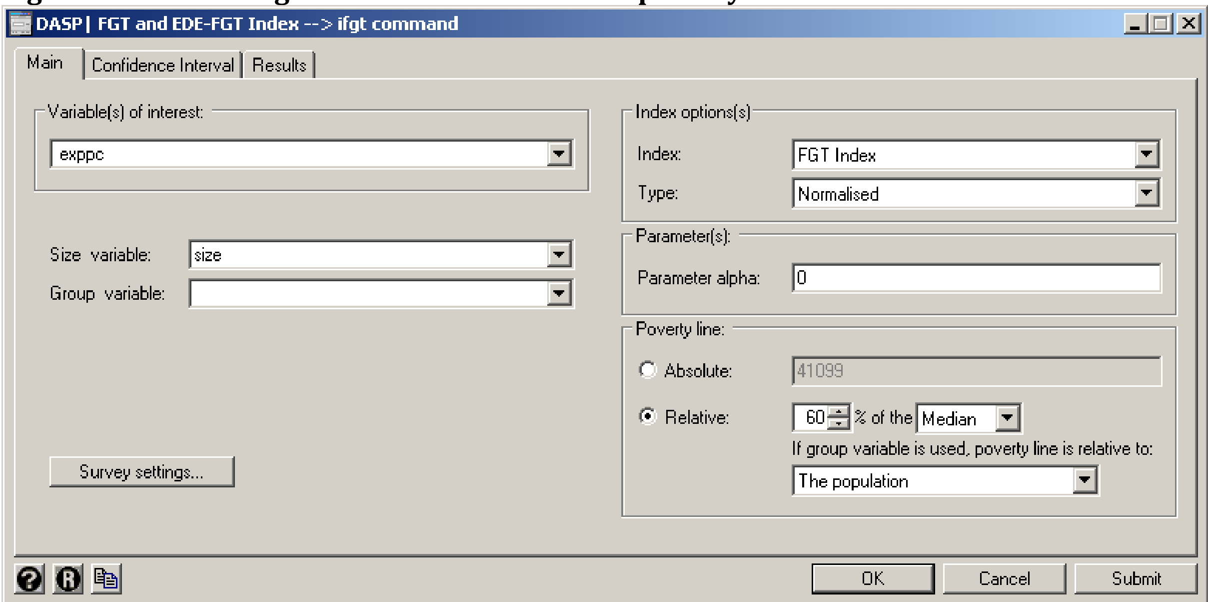
```
. ifgt exppc expeq, alpha(0) hsize(size) pline(41099)

    Poverty Index   :  FGT Index
    Household size  :  size
    Sampling weight :  weight
    Parameter alpha :  0.00
```

Variable	Estimate	STD	LB	UB	P. Line
exppc	**0.444565**	**0.016124**	**0.412873**	**0.476256**	41099.00
expeq	**0.255400**	**0.013326**	**0.229208**	**0.281592**	41099.00

Q.4

Select RELATIVE for the poverty line and set the other parameters as above.

Figure 20: Estimating FGT indices with relative poverty lines



After clicking on SUBMIT, the following results should be displayed:

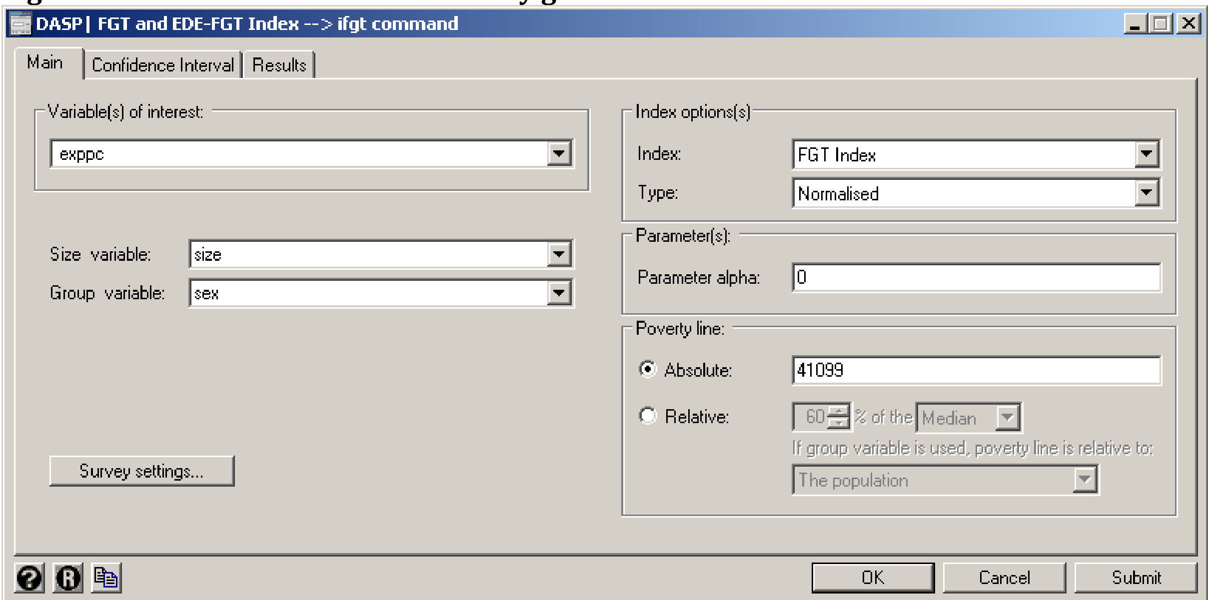
```
. ifgt exppc, alpha(0) hsize(size) opl(median) prop(60)

   Poverty Index   :  FGT Index
   Household size  :  size
   Sampling weight :  weight
   Parameter alpha :  0.00
```

Variable	Estimate	STD	LB	UB	P. Line
exppc	**0.185243**	**0.008576**	**0.168386**	**0.202099**	27046.71

Q.5

Set the group variable to *sex*.

Figure 21: FGT indices differentiated by gender



Clicking on SUBMIT, the following should appear:

```
. ifgt exppc, alpha(0) hsize(size) hgroup(sex) pline(41099)

    Poverty Index   :  FGT Index
    Household size  :  size
    Sampling weight :  weight
    Group variable  :  sex
    Parameter alpha :  0.00
```

Group	Estimate	STD	LB	UB	P. Line
1: Male	**0.452176**	**0.016633**	**0.419484**	**0.484867**	41099.00
2: Female	**0.281850**	**0.028206**	**0.226411**	**0.337290**	41099.00
POPULATION	0.444565	0.016124	0.412873	0.476256	41099.00

Q.6

Using the panel CONFIDENCE INTERVAL, set the confidence level to 99 % and set the number of decimals to 4 in the RESULTS panel.

```
. ifgt exppc, alpha(0) hsize(size) hgroup(sex) dec(4) level(99) pline(41099)

    Poverty Index   :  FGT Index
    Household size  :  size
    Sampling weight :  weight
    Group variable  :  sex
    Parameter alpha :  0.00
```

Group	Estimate	STD	LB	UB	P. Line
1: Male	**0.4522**	**0.0166**	**0.4091**	**0.4952**	41099.00
2: Female	**0.2819**	**0.0282**	**0.2089**	**0.3548**	41099.00
POPULATION	0.4446	0.0161	0.4028	0.4863	41099.00

23.2 Estimating differences between FGT indices.

"Has poverty Burkina Faso decreased between 1994 and 1998?"

1. Open the dialog box for the difference between FGT indices.
2. Estimate the difference between headcount indices when
    a. Distribution 1 is year 1998 and distribution 2 is year 1994;
    b. The variable of interest is **exppc** for 1994 and **exppcz** for 1998.
    c. You should set size to household size in order to estimate poverty over the population of individuals.
    d. Use 41099 Francs CFA per year as the poverty line for both distributions.
3. Estimate the difference between headcount indices when
    a. Distribution 1 is rural residents in year 1998 and distribution 2 is rural residents in year 1994;
    b. The variable of interest is **exppc** for 1994 and **exppcz** for 1998.
    c. You should set size to household size in order to estimate poverty over the population of individuals.
    d. Use 41099 Francs CFA per year as the poverty line for both distributions.
4. Redo the last exercise for urban residents.
5. Redo the last exercise only for members of male-headed households.
6. Test if the estimated difference in the last exercise is significantly different from zero. Thus, test:

    $$H_0 : \Delta P(z = 41099, \alpha = 0) = 0 \quad \textit{against} \quad H_1 : \Delta P(z = 41099, \alpha = 0) \neq 0$$

    Set the significance level to 5% and assume that the test statistics follows a normal distribution.

Answers

Q.1

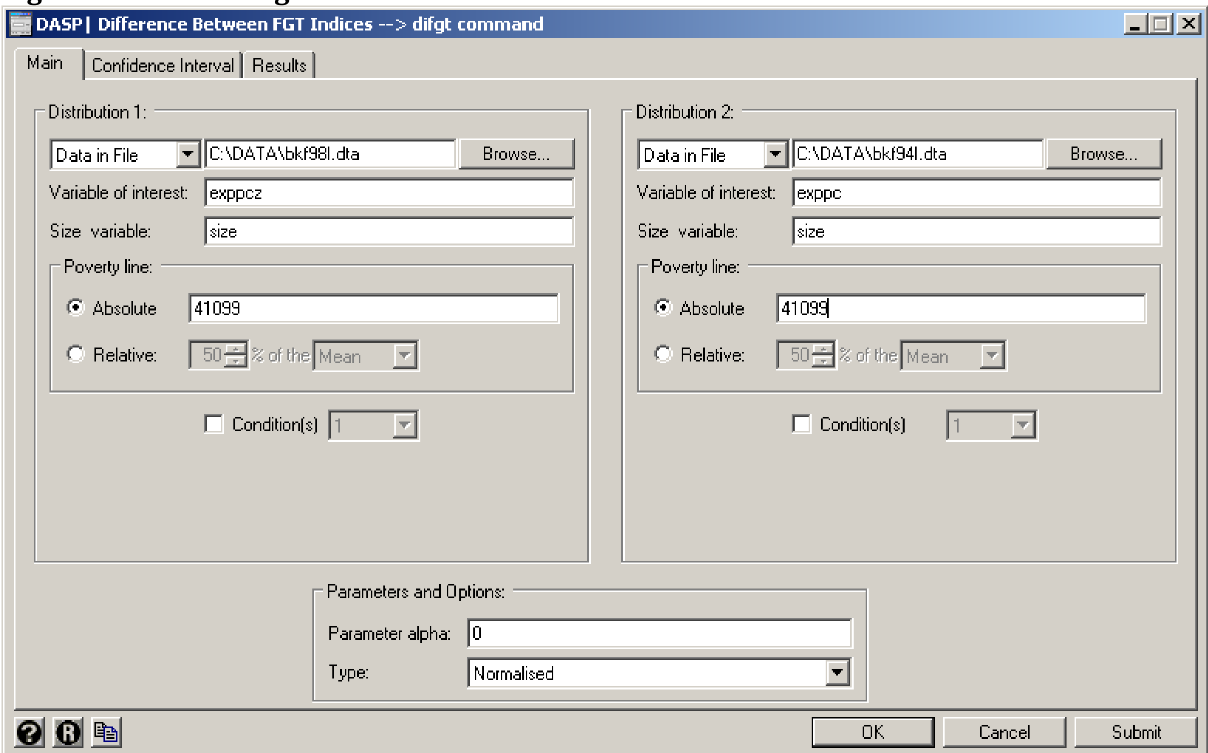
Open the dialog box by typing

db difgt

Q.2

- For distribution 1, choose the option DATA IN FILE instead of DATA IN MEMORY and click on BROWSE to specify the location of the file ***bkf98I.dta***.
- Follow the same procedure for distribution 2 to specify the location of ***bkf94I.dta***.
- Choose variables and parameters as follows:

Figure 22: Estimating differences between FGT indices



After clicking on SUBMIT, the following should be displayed:

```
. difgt exppcz exppc, alpha(0) file1(C:\DATA\bkf98I.dta) hsize1(size) file2(C:\DATA\bkf94I.dta) hsize2(size) pline1(41099) pline2(41099)
```

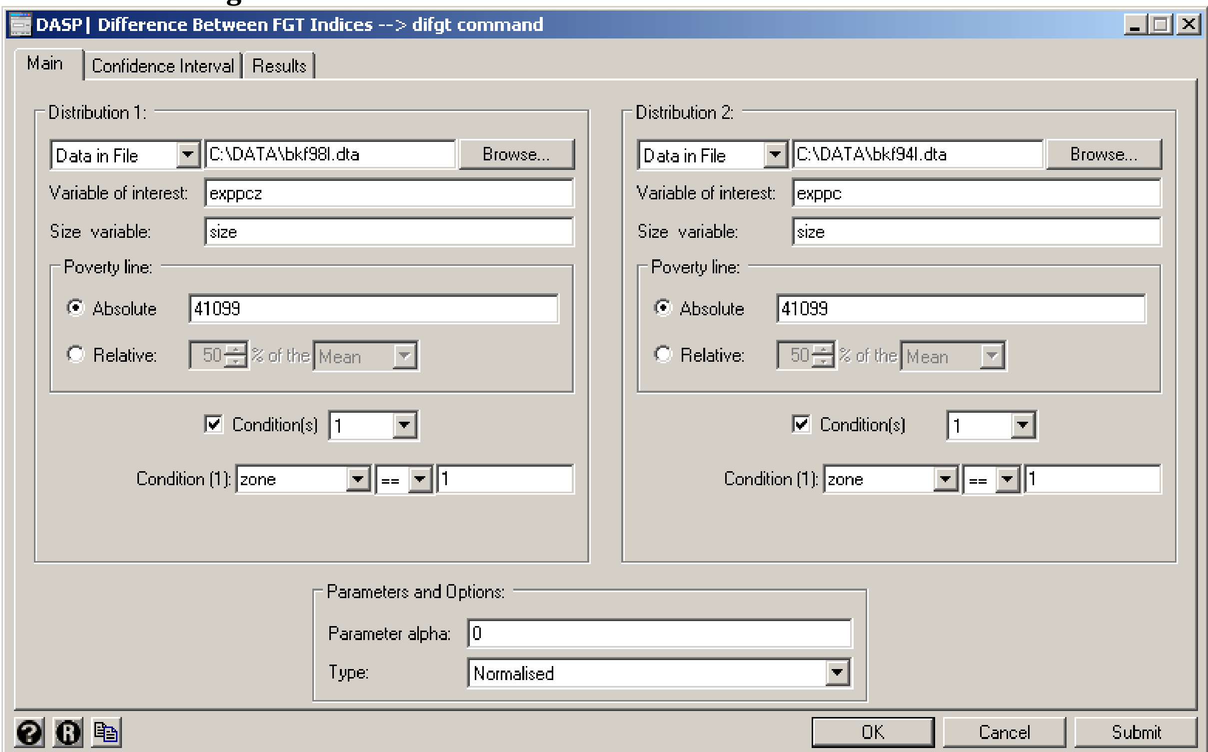
Poverty Index : FGT Index
Parameter alpha : 0.00

	Estimate	STD	LB	UB	P. Line
Distribution_1	**0.452677**	**0.010927**	**0.431199**	**0.474156**	41099.00
Distribution_2	**0.444565**	**0.016124**	**0.412873**	**0.476256**	41099.00
Difference	0.008113	0.019477	-0.030062	0.046288	---

Q.3

- Restrict the estimation to rural residents as follows:
    - o Select the option Condition(s)
    - o Write ZONE in the field next to CONDITION (1) and type 1 in the next field.

Figure 23: Estimating differences in FGT indices



After clicking on SUBMIT, we should see:

```
Poverty Index   :   FGT Index
Parameter alpha :        0.00
```

	Estimate	STD	LB	UB	P. Line
Distribution_1	**0.510344**	**0.011601**	**0.487539**	**0.533149**	41099.00
Distribution_2	**0.510497**	**0.019975**	**0.471236**	**0.549758**	41099.00
Difference	-0.000153	0.023100	-0.045427	0.045121	---

Q.4

Poverty Index : FGT Index
Parameter alpha : 0.00

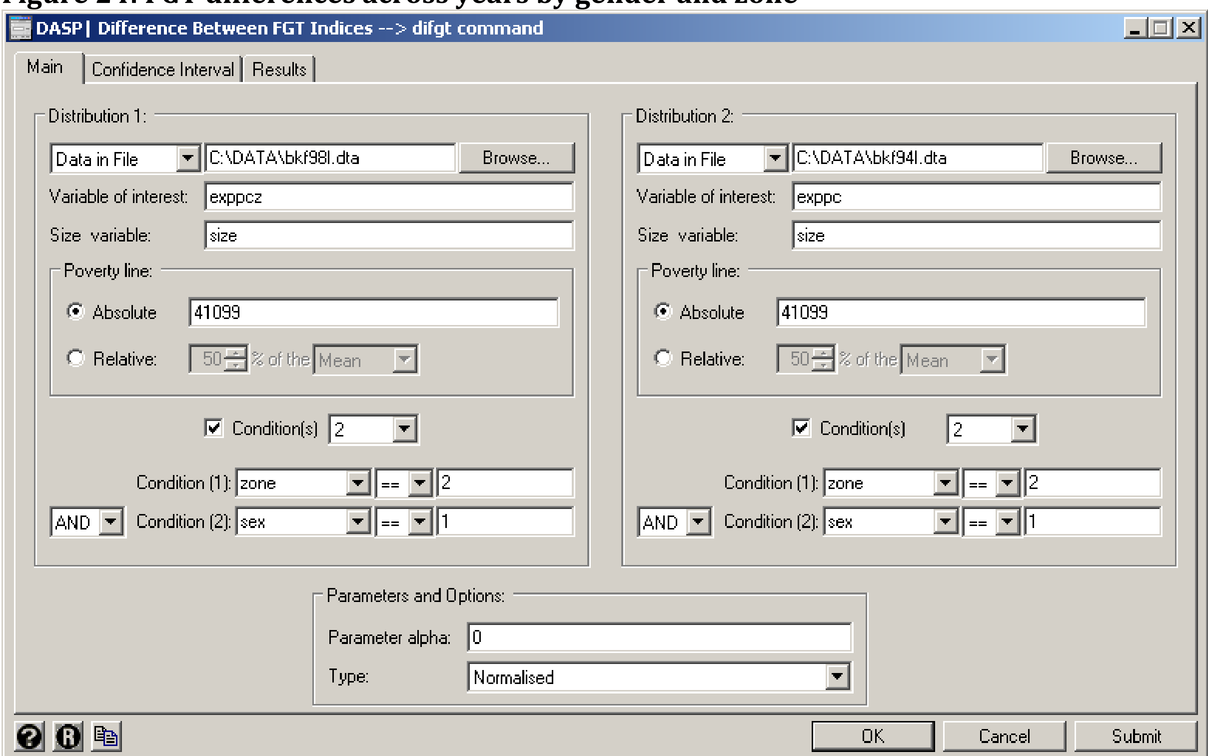
	Estimate	STD	LB	UB	P. Line
Distribution_1	**0.164573**	**0.016297**	**0.132538**	**0.196608**	41099.00
Distribution_2	**0.103684**	**0.013419**	**0.077309**	**0.130059**	41099.00
Difference	0.060889	0.021111	0.019513	0.102265	---

One can see that the change in poverty was significant only for urban residents. **Q.5**

Restrict the estimation to male-headed urban residents as follows:

- Set the number of Condition(s) to 2;
- Set **sex** in the field next to **Condition (2)** and type **1** in the next field.

Figure 24: FGT differences across years by gender and zone



After clicking on SUBMIT, the following should be displayed:

Poverty Index : FGT Index
Parameter alpha : 0.00

	Estimate	STD	LB	UB	P. Line
Distribution_1	**0.172384**	**0.017701**	**0.137590**	**0.207179**	41099.00
Distribution_2	**0.105997**	**0.013945**	**0.078588**	**0.133405**	41099.00
Difference	0.066388	0.022534	0.022222	0.110553	---

Q.6

We have that:

Lower Bound: = 0.0222
Upper Bound: = 0.1105

The null hypothesis is rejected since the lower bound of the 95% confidence interval is above zero.

23.3 Estimating multidimensional poverty indices

"How much is bi-dimensional poverty (total expenditures and literacy) in Peru in 1994?"

Using the ***peru94I.dta*** file,

1. Estimate the *Chakravarty et al (1998)* index with parameter alpha = 1 and

	Var. of interest	Pov. line	a_j
Dimension 1	exppc	400	1
Dimension 2	pliterate	0.90	1

2. Estimate the *Bourguignon and Chakravarty (2003)* index with parameters alpha=beta=gamma= 1 and

	Var. of interest	Pov. line
Dimension 1	exppc	400
Dimension 2	literate	0.90

Q.1

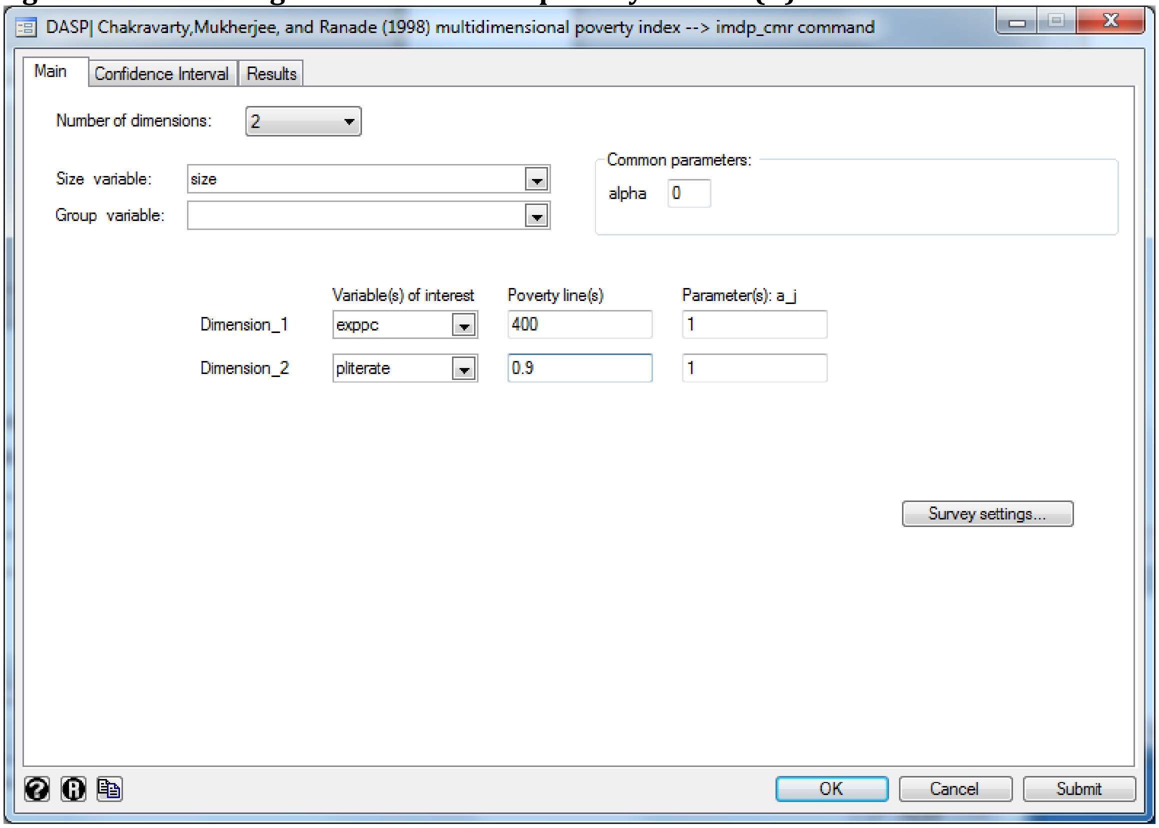
Steps:

- Type
  use "C:\data\ peru94_A_I.dta ", clear

- To open the relevant dialog box, type
  db imdp_bci

- Choose variables and parameters as in

Figure 25: Estimating multidimensional poverty indices (A)



After clicking SUBMIT, the following results appear.

```
imdpov exppc pliterate, hsize(size) index(1) alpha(0) a1(1) pl1(400) a2(1) pl2(0.9)

 M.D. Poverty index :  Chakravarty et al (1998)
 Household size     :  size
```

	Estimate	STD	LB	UB
'opulation	**0.418**	**0.009**	**0.403**	**0.433**

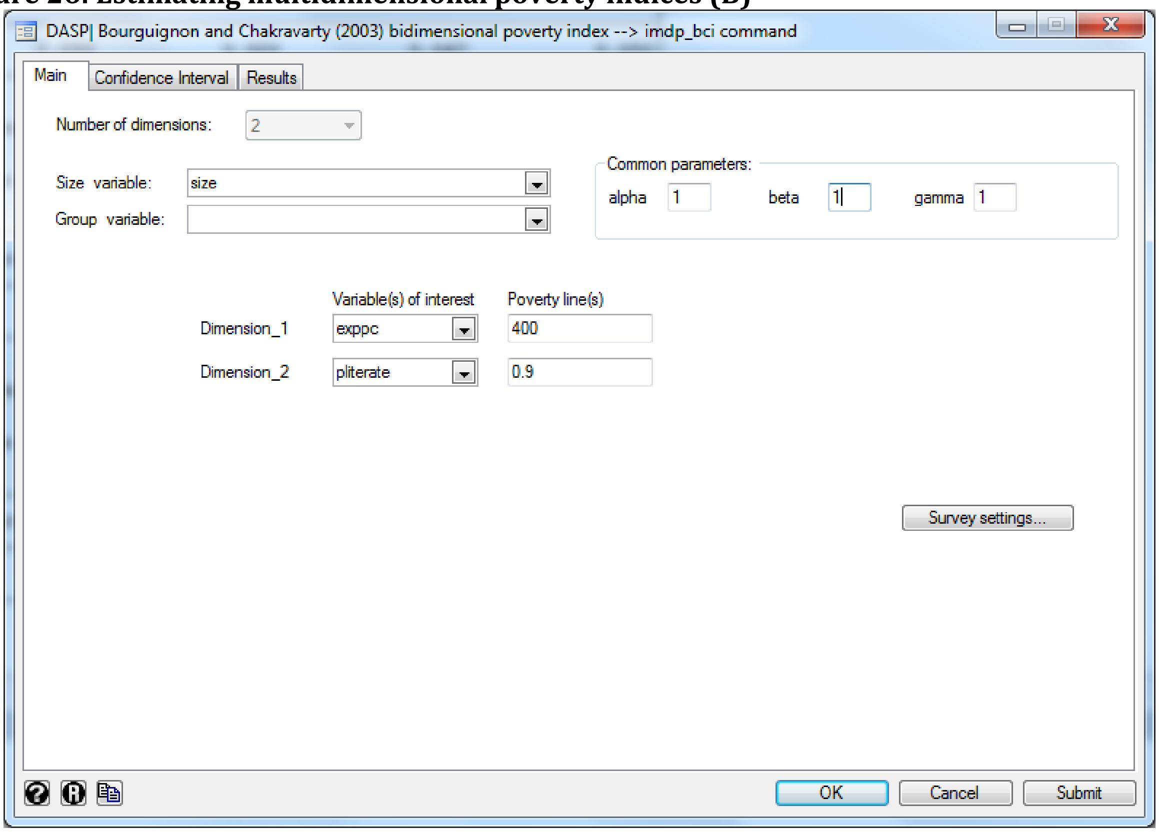
Q.2

- To open the relevant dialog box, type
  db imdp_cmr

Steps:

- Choose variables and parameters as in

Figure 26: Estimating multidimensional poverty indices (B)



After clicking SUBMIT, the following results appear.

```
. imdpov exppc pliterate, hsize(size) index(7) alpha(1) beta(1) gamma(1) pl1(400) pl2(0.9)

   M.D. Poverty index :  Bourguignon and Chakravarty (2003)
   Household size     :  size
```

	Estimate	STD	LB	UB
Population	**0.098**	**0.003**	**0.093**	**0.103**

23.4 Estimating FGT curves.

"How sensitive to the choice of a poverty line is the rural-urban difference in poverty?"

1. Open ***bkf94I.dta***
2. Open the FGT curves dialog box.
3. Draw FGT curves for variables of interest *exppc* and *expeq* with
   a. parameter $\alpha = 0$;
   b. poverty line between 0 and 100,000 Franc CFA;
   c. size variable set to *size*;
   d. subtitle of the figure set to "Burkina 1994".
4. Draw FGT curves for urban and rural residents with
   a. variable of interest set to *expcap*;
   b. parameter $\alpha = 0$;
   c. poverty line between 0 and 100,000 Franc CFA;
   d. size variable set to *size*.
5. Draw the difference between these two curves and
   a. save the graph in *.gph format to be plotted in Stata and in *.wmf format to be inserted in a Word document.
   b. List the coordinates of the graph.
6. Redo the last graph with $\alpha = 1$.

Answers

Q.1

Open the file with

use "C:\data\bkf94I.dta", clear

Q.2

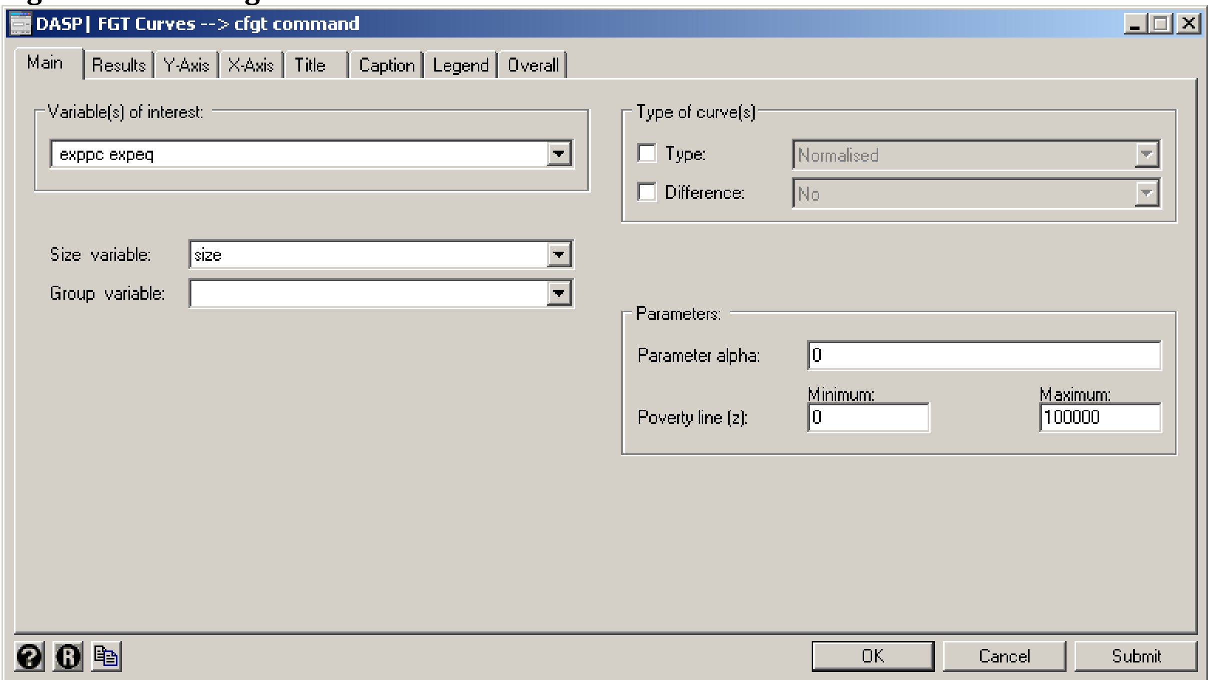
Open the dialog box by typing

db difgt

Q.3

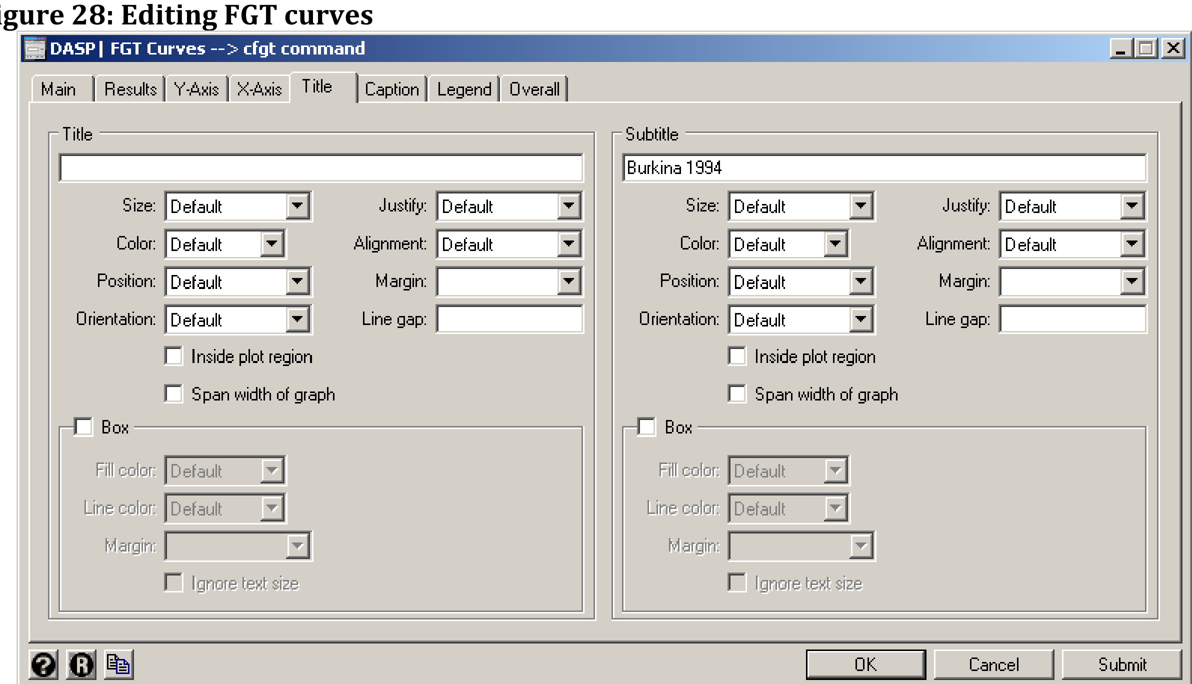
Choose variables and parameters as follows:

Figure 27: Drawing FGT curves



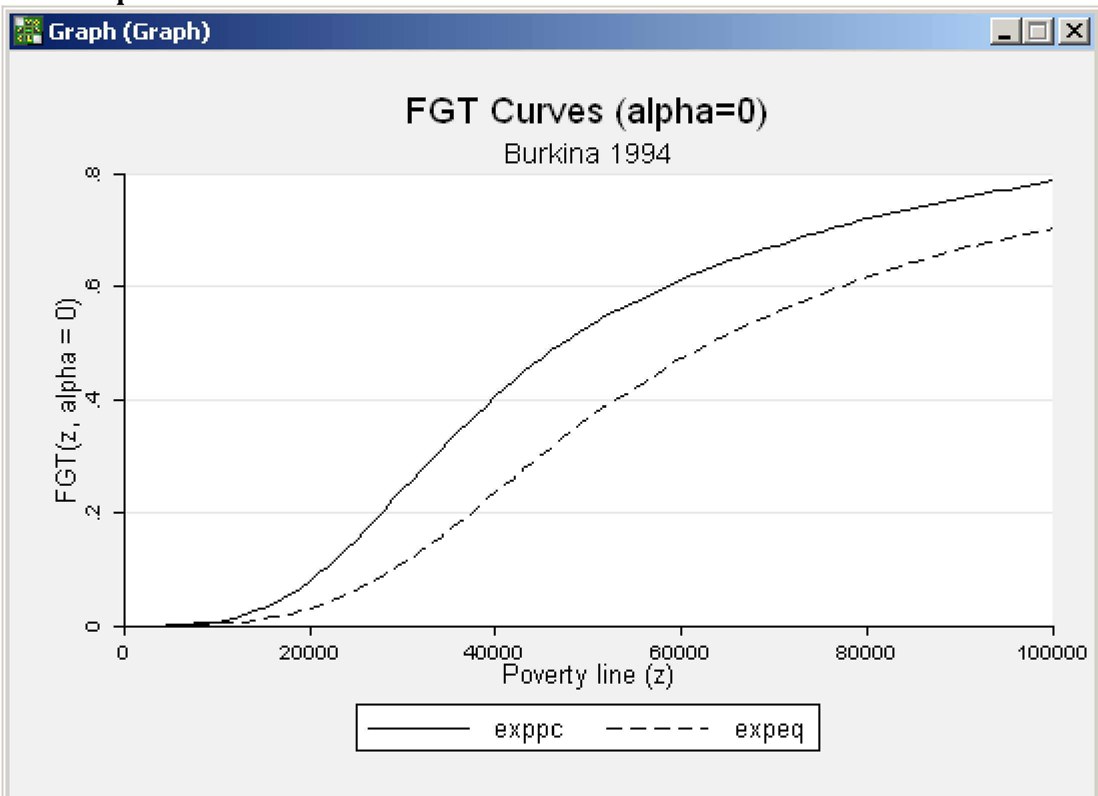
To change the subtitle, select the Title panel and write the subtitle.

Figure 28: Editing FGT curves



After clicking SUBMIT, the following graph appears:

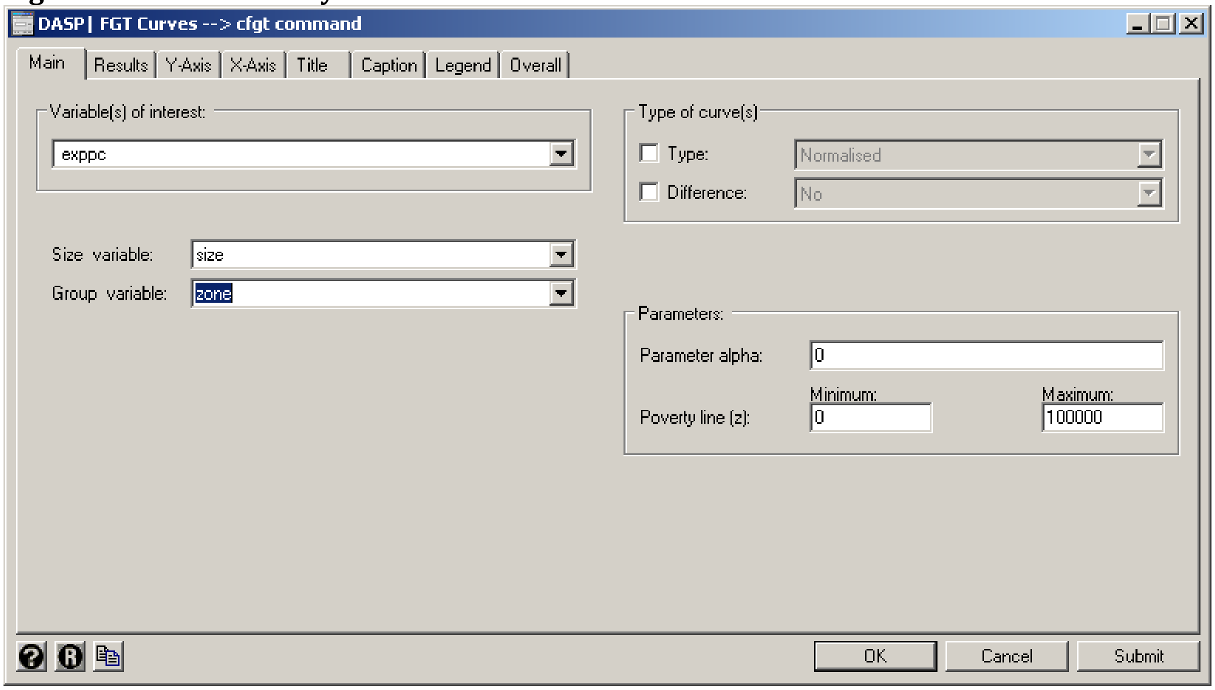
Figure 29: Graph of FGT curves



Q.4

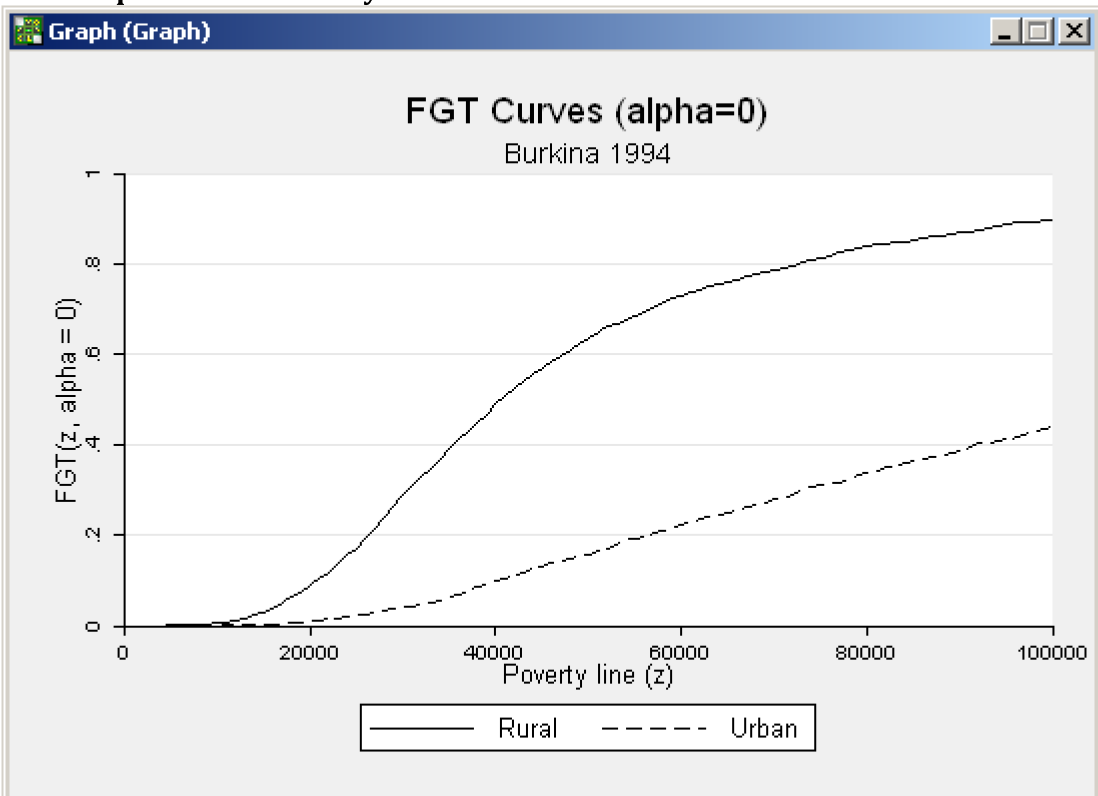
Choose variables and parameters as in the following window:

Figure 30: FGT curves by zone



After clicking SUBMIT, the following graph appears:

Figure 31: Graph of FGT curves by zone



Q.5

- Choose the option DIFFERENCE and select: WITH THE FIRST CURVE;
- Indicate that the group variable is *zone*;
- Select the Results panel and choose the option LIST in the COORDINATES quadrant.
- In the GRAPH quadrant, select the directory in which to save the graph in gph format and to export the graph in wmf format.

Figure 32: Differences of FGT curves

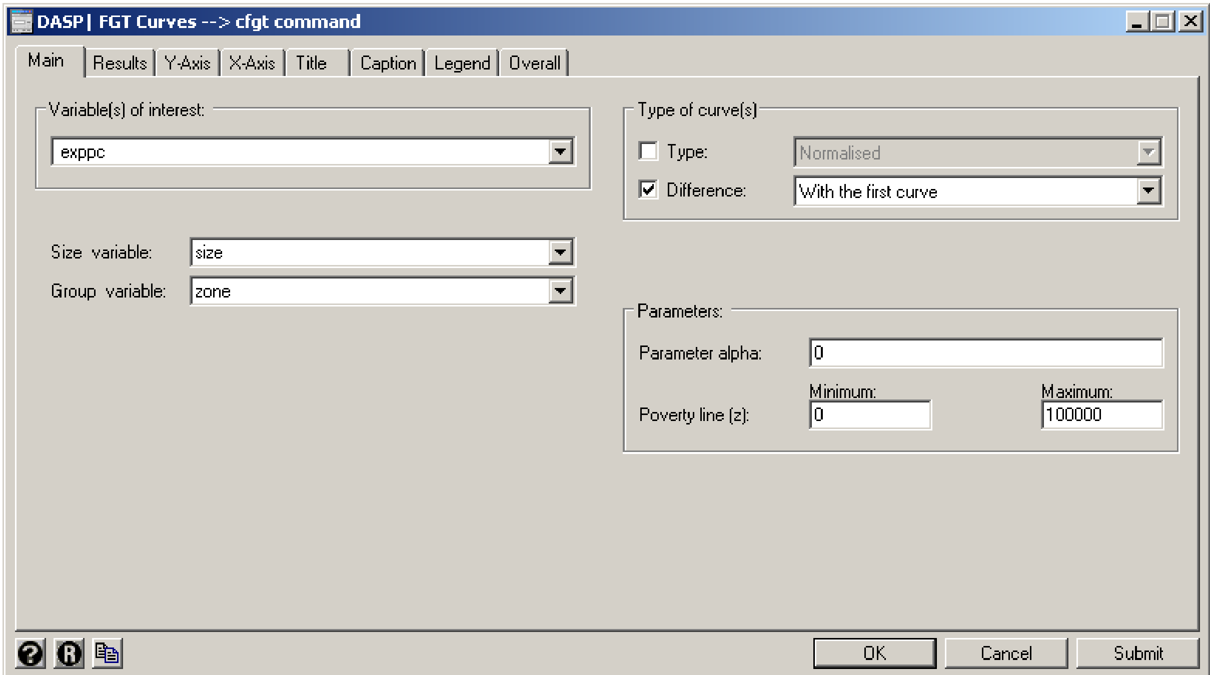
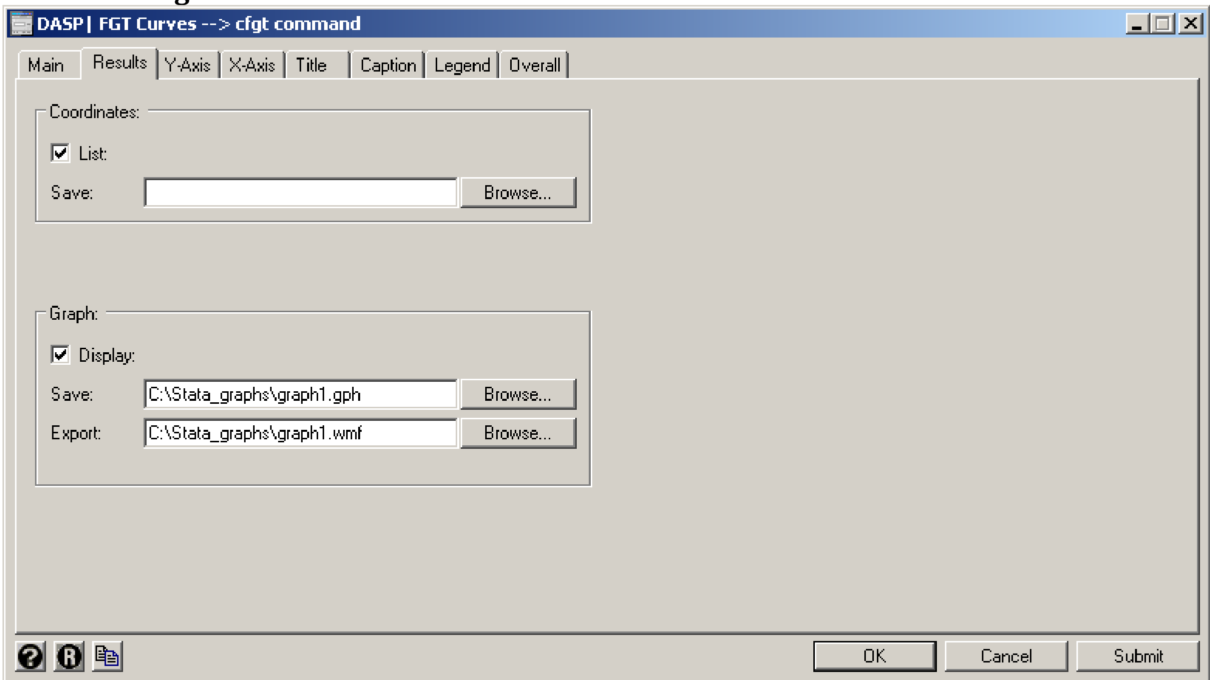
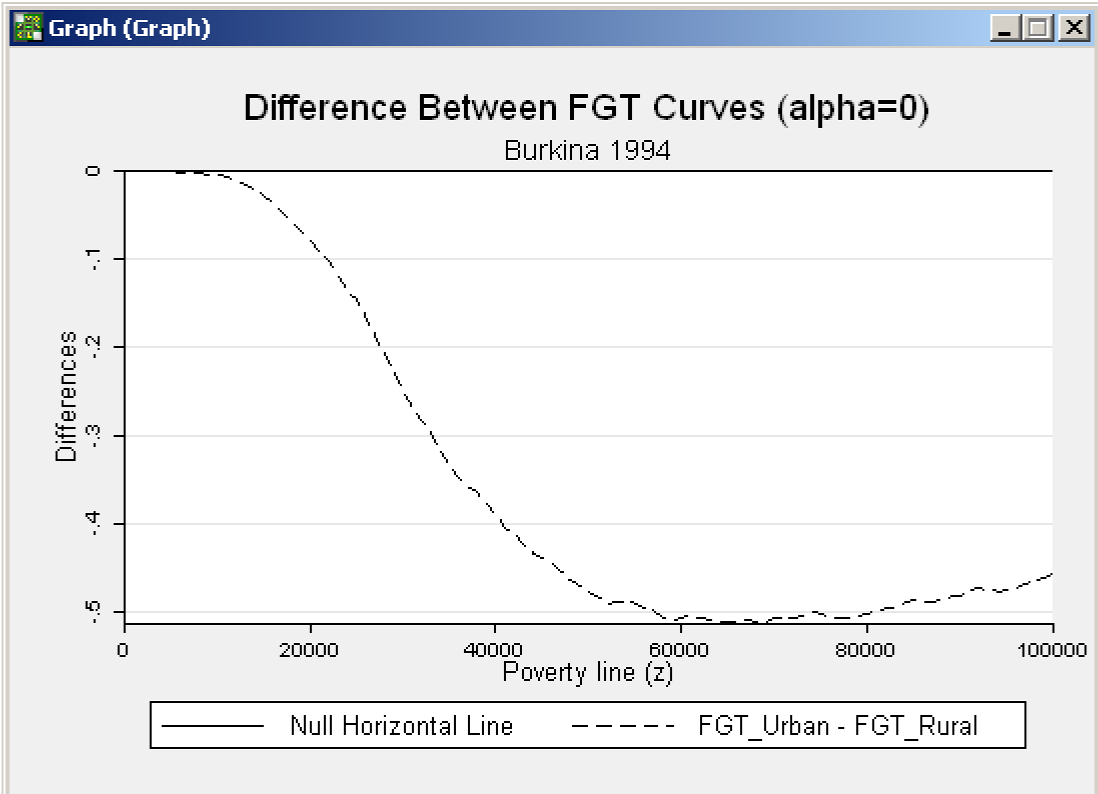


Figure 33: Listing coordinates



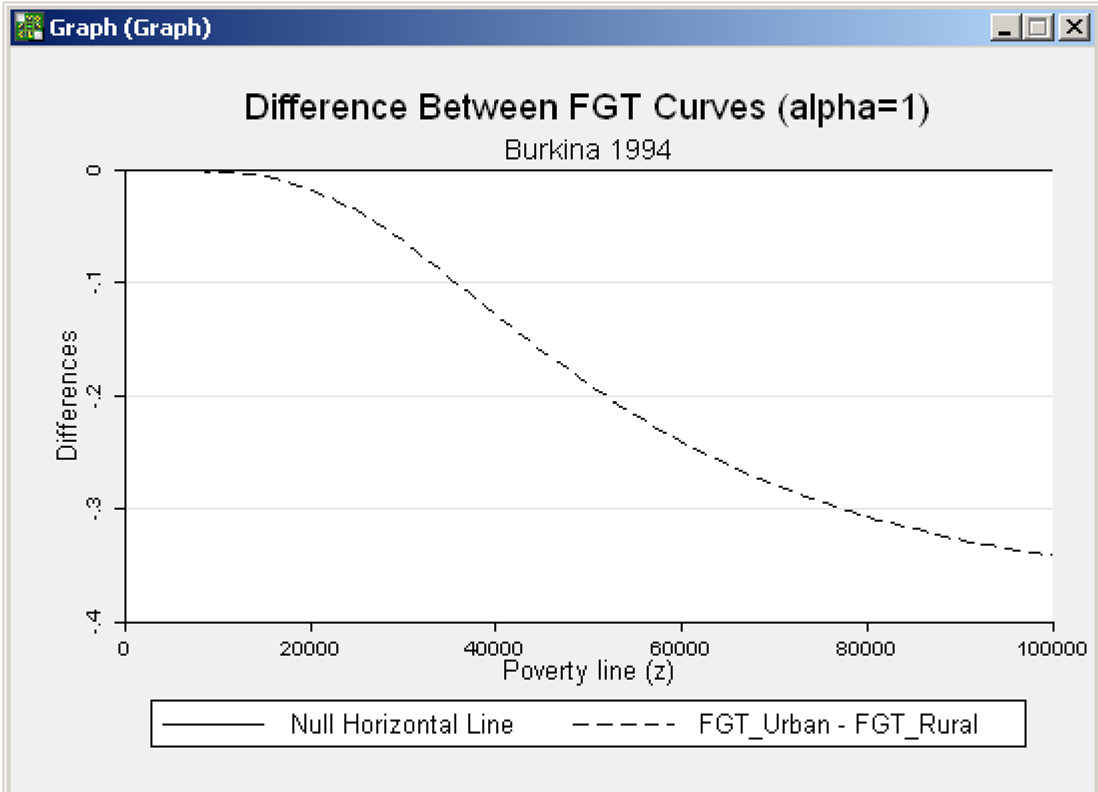
After clicking SUBMIT, the following appears:

Figure 34: Differences between FGT curves



Q.6

Figure 35: Differences between FGT curves



23.5 Estimating FGT curves and differences between FGT curves with confidence intervals

"Is the poverty increase between 1994 and 1998 in Burkina Faso statistically significant?"

1) Using the file ***bkf94I.dta***, draw the FGT curve and its confidence interval for the variable of interest *exppc* with:
   a) parameter $\alpha = 0$;
   b) poverty line between 0 and 100,000 Franc CFA;
   c) size variable set to *size*.

2) Using simultaneously the files ***bkf94I.dta*** and ***bkf98I.dta***, draw the difference between FGT curves and associated confidence intervals with:
   a) The variable of interest *exppc* for 1994 and *exppcz* for 1998.
   b) parameter $\alpha = 0$;
   c) poverty line between 0 and 100,000 Franc CFA;
   d) size variable set to *size*.

3) Redo 2) with parameter $\alpha = 1$.

Answers

Q.1

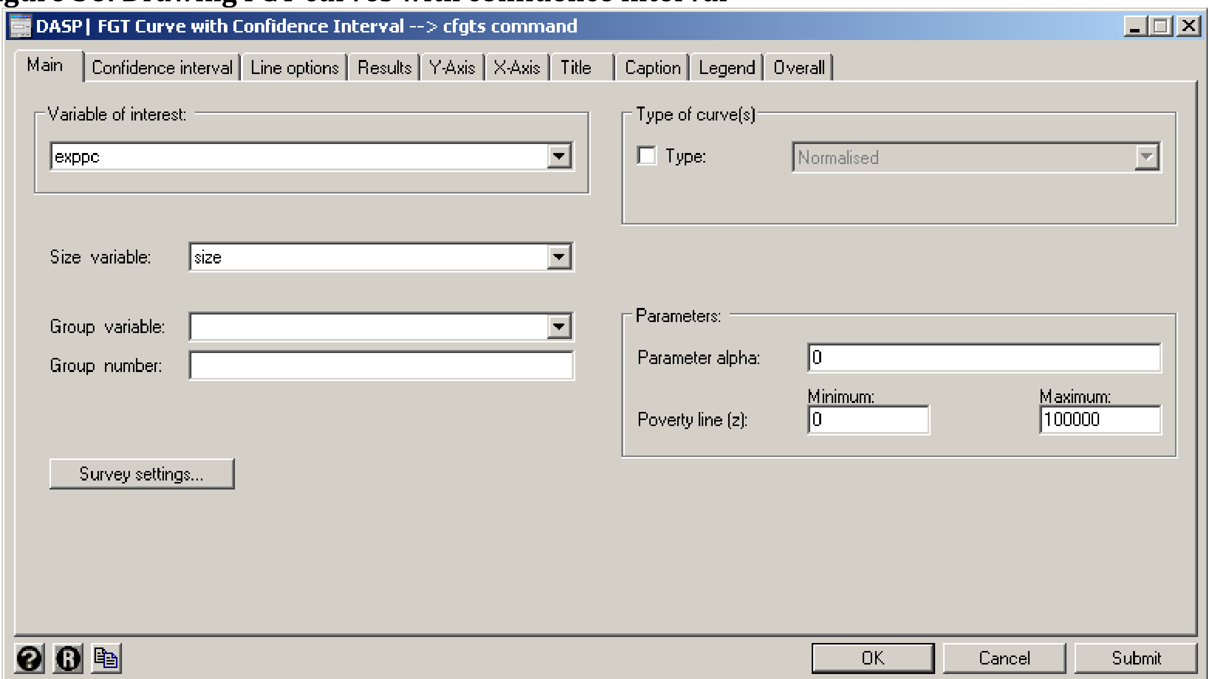
Steps:

- Type
  use "C:\data\bkf94I.dta", clear

- To open the relevant dialog box, type
  db cfgts

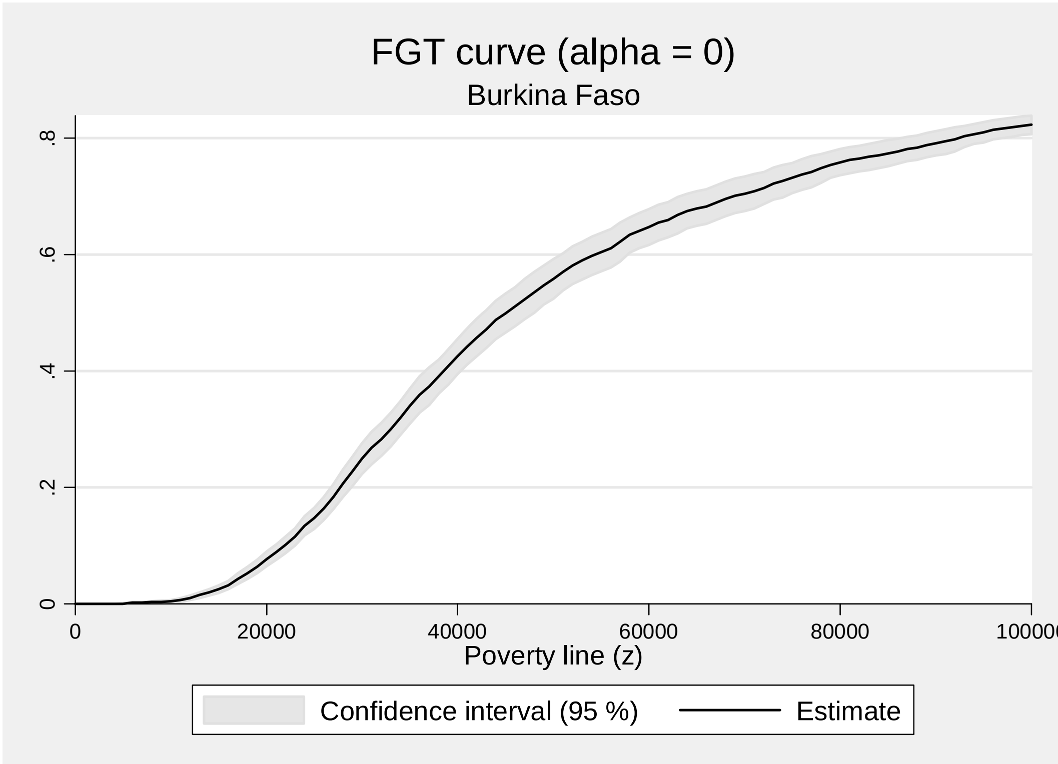
- Choose variables and parameters as in

Figure 36: Drawing FGT curves with confidence interval



After clicking SUBMIT, the following appears:

Figure 37: FGT curves with confidence interval



Q.2

Steps:

- To open the relevant dialog box, type
  db cfgtsd2

- Choose variables and parameters as in

Figure 38: Drawing the difference between FGT curves with confidence interval

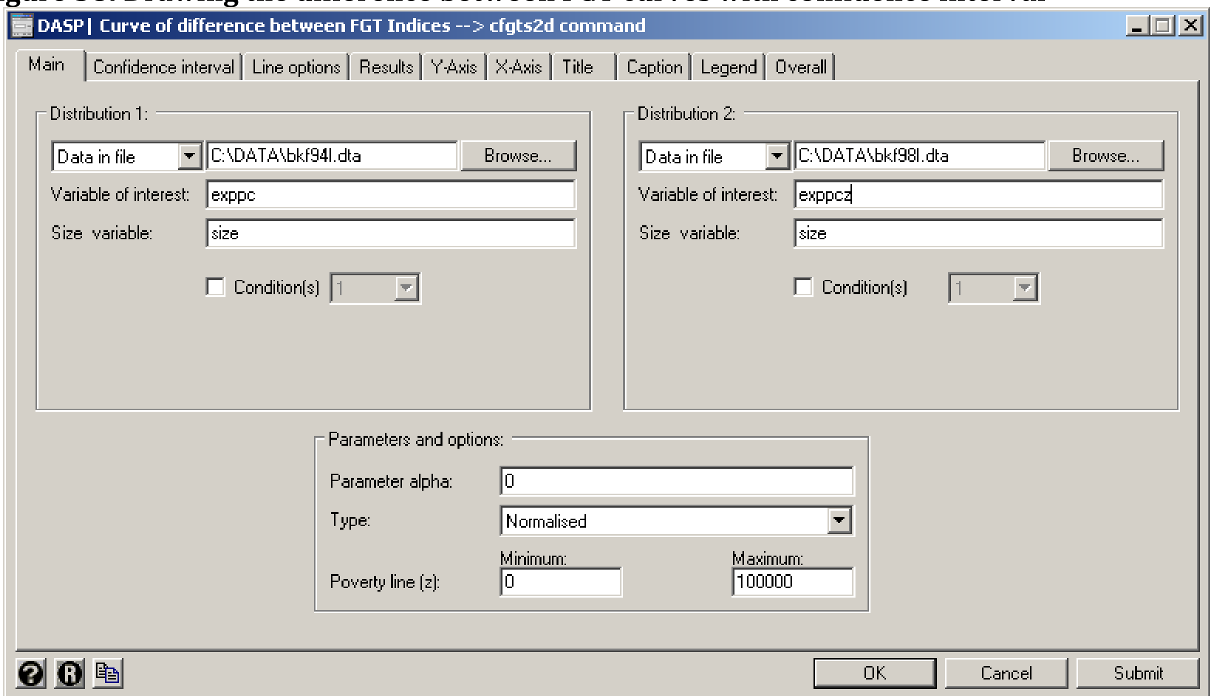


Figure 39: Difference between FGT curves with confidence interval $(\alpha = 0)$

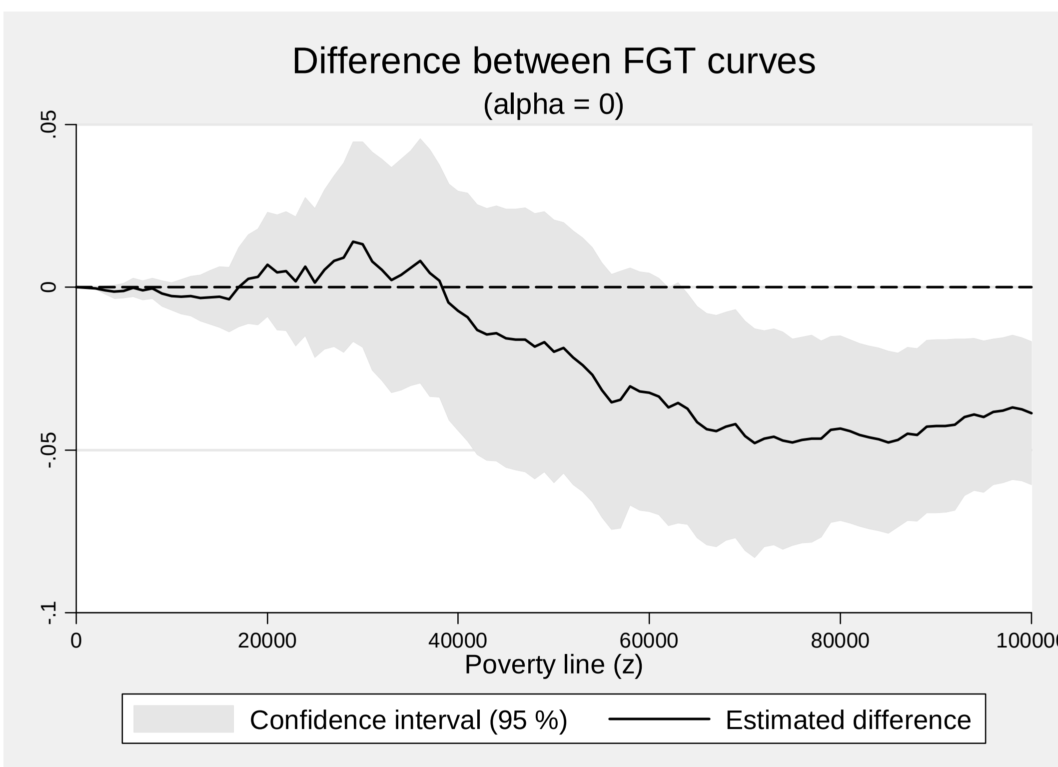
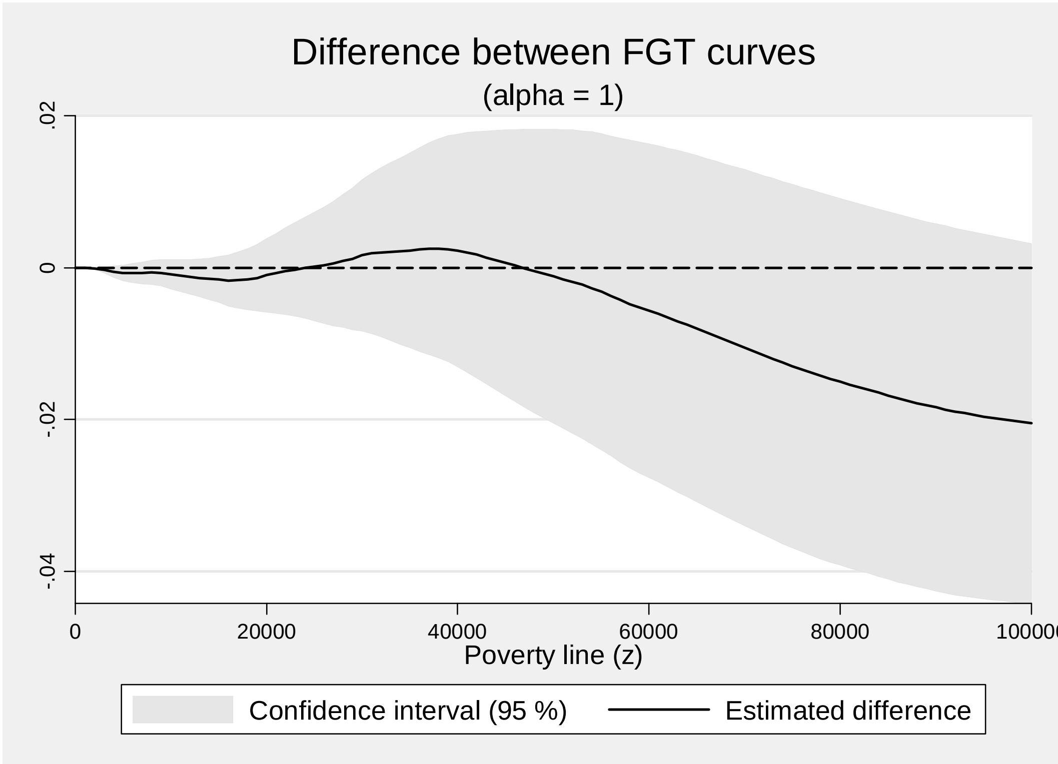


Figure 40: Difference between FGT curves with confidence interval $(\alpha = 1)$



23.6 Testing poverty dominance and estimating critical values.

"Has the poverty increase in Burkina Faso between 1994 and 1998 been statistically significant?"

1) Using simultaneously files *bkf94I.dta* and *bkf98I.dta*, check for second-order poverty dominance and estimate the values of the poverty line at which the two FGT curves cross.
   a) The variable of interest is *exppc* for 1994 and *exppcz* for 1998;
   b) The poverty line should vary between 0 and 100,000 Franc CFA;
   c) The size variable should be set to *size*.

Answers

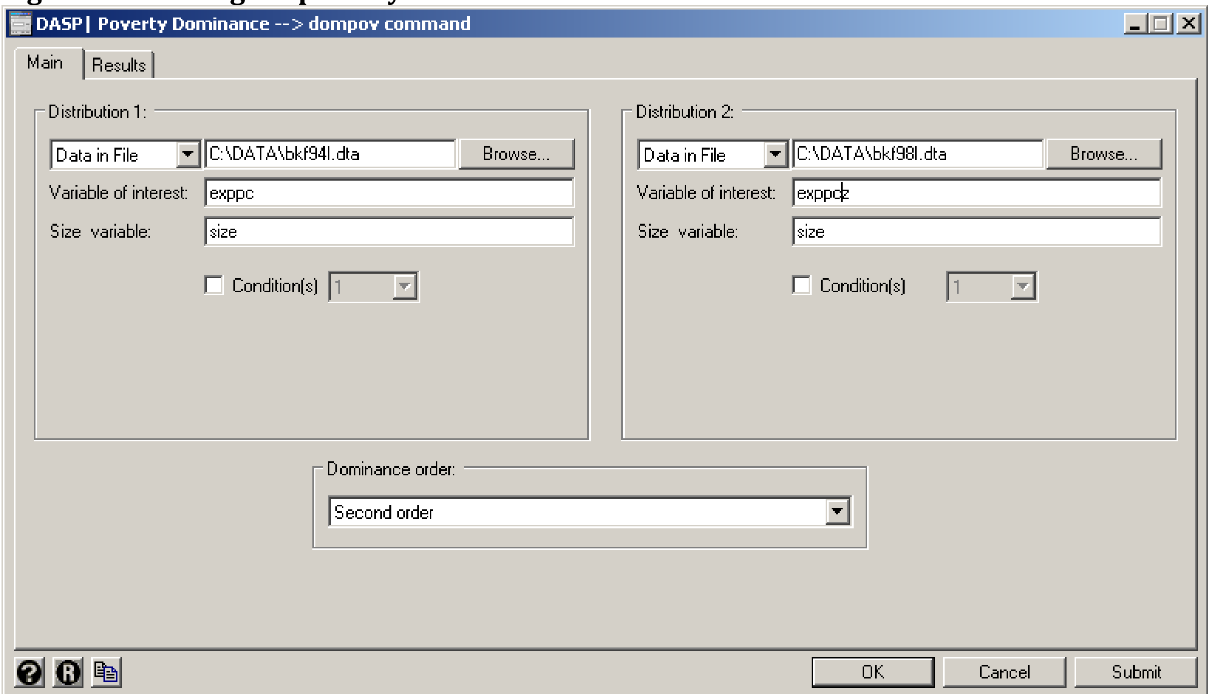
Q.1

Steps:

- To open the relevant dialog box, type
  db dompov

- Choose variables and parameters as in

Figure 41: Testing for poverty dominance



After clicking SUBMIT, the following results appear:

Number of intersection	Critical pov. line	Min. range of pov. lines	Max. range of pov. line	Case
1	24262.871	.	.	A
2	46775.652	.	.	B

Notes :
_case A: Before this intersection, distribution 2 dominates distribution 1.
_case B: Before this intersection, distribution 1 dominates distribution 2.
_case C: No dominance before this intersection.

23.7 Decomposing FGT indices.

"What is the contribution of different types of earners to total poverty in Burkina Faso?"

1. Open ***bkf94I.dta*** and decompose the average poverty gap
    a. with variable of interest *exppc*;
    b. with size variable set to *size*;
    c. at the official poverty line of 41099 Francs CFA;
    d. and using the group variable *gse* (Socio-economic groups).

2. Do the above exercise without standard errors and with the number of decimals set to 4.

Answers

Q.1

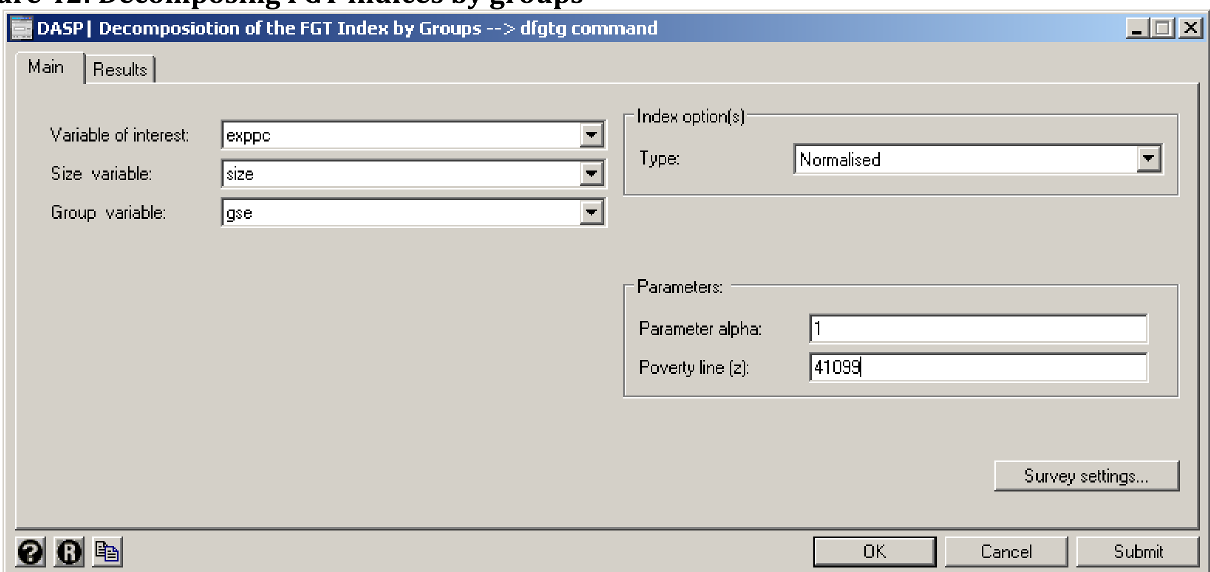
Steps:

- Type
  use "C:\data\bkf94I.dta", clear

- To open the relevant dialog box, type
  db dfgtg

- Choose variables and parameters as in

Figure 42: Decomposing FGT indices by groups



After clicking SUBMIT, the following information is provided:

```
dfgtg exppc, hgroup(gse) hsize(size) alpha(1) pline(41099) type(nor)
```

FGT Index: Decomposition by Groups

Group	FGT Index	Population Share	Absolute Contribution	Relative Contribution
1: wage-earning (public sector)	**0.004237**	**0.042971**	**0.000182**	**0.001308**
	0.002571	0.003790	0.000117	0.000840
2: wage-earning (private sector)	**0.022176**	**0.026598**	**0.000590**	**0.004237**
	0.010678	0.002164	0.000291	0.002083
3: Artisan or trading	**0.027741**	**0.062640**	**0.001738**	**0.012484**
	0.004653	0.004288	0.000325	0.002371
4: Others activities	**0.063853**	**0.006650**	**0.000425**	**0.003050**
	0.025805	0.001308	0.000170	0.001203
5: Farmers (crop)	**0.137525**	**0.104402**	**0.014358**	**0.103148**
	0.011808	0.014896	0.002459	0.016980
6: Farmers (food)	**0.162894**	**0.680885**	**0.110912**	**0.796800**
	0.008643	0.016403	0.005823	0.019015
7: Inactive	**0.144916**	**0.075856**	**0.010993**	**0.078973**
	0.014994	0.004839	0.001332	0.008520
POPULATION	**0.139197**	**1.000000**	**0.139197**	**1.000000**
	0.006553	0.000000	0.006553	0.000000

Q.2

Using the RESULTS panel, change the number of decimals and unselect the option DISPLAY STANDARD ERRORS.

After clicking SUBMIT, the following information is obtained:

```
. dfgtg exppc, hgroup(gse) hsize(size) alpha(1) pline(41099) dstd(0) type(nor) dec(4)
```

FGT Index: Decomposition by Groups

Group	FGT Index	Population Share	Absolute Contribution	Relative Contribution
1: wage-earning (public sector)	**0.0042**	**0.0430**	**0.0002**	**0.0013**
2: wage-earning (private sector)	**0.0222**	**0.0266**	**0.0006**	**0.0042**
3: Artisan or trading	**0.0277**	**0.0626**	**0.0017**	**0.0125**
4: Others activities	**0.0639**	**0.0066**	**0.0004**	**0.0031**
5: Farmers (crop)	**0.1375**	**0.1044**	**0.0144**	**0.1031**
6: Farmers (food)	**0.1629**	**0.6809**	**0.1109**	**0.7968**
7: Inactive	**0.1449**	**0.0759**	**0.0110**	**0.0790**
POPULATION	**0.1392**	**1.0000**	**0.1392**	**1.0000**

23.8 Estimating Lorenz and concentration curves.

"How much do taxes and transfers affect inequality in Canada?"

By using the ***can6.dta*** file,

1. Draw the Lorenz curves for gross income *X* and net income *N*. How can you see the redistribution of income?
2. Draw Lorenz curves for gross income *X* and concentration curves for each of the three transfers *B1*, *B2* and *B3* and the tax *T*. What can you say about the progressivity of these elements of the tax and transfer system?

"What is the extent of inequality among Burkina Faso rural and urban households in 1994?"

By using the ***bkf94I.dta*** file,

3. Draw Lorenz curves for rural and urban households

    a. with variable of interest *exppc*;
    b. with size variable set to *size*;
    c. and using the group variable *zone* (as residential area).

Q.1

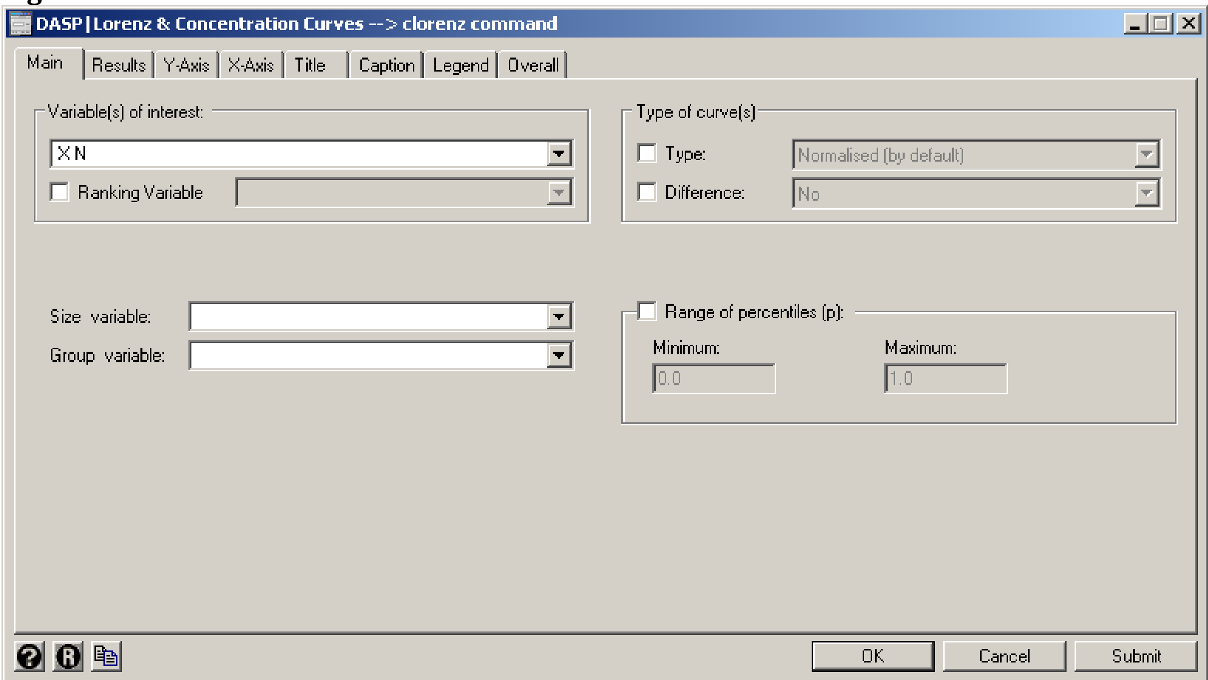
Steps:

- Type
  use "C:\data\can6.dta", clear

- To open the relevant dialog box, type
  db clorenz

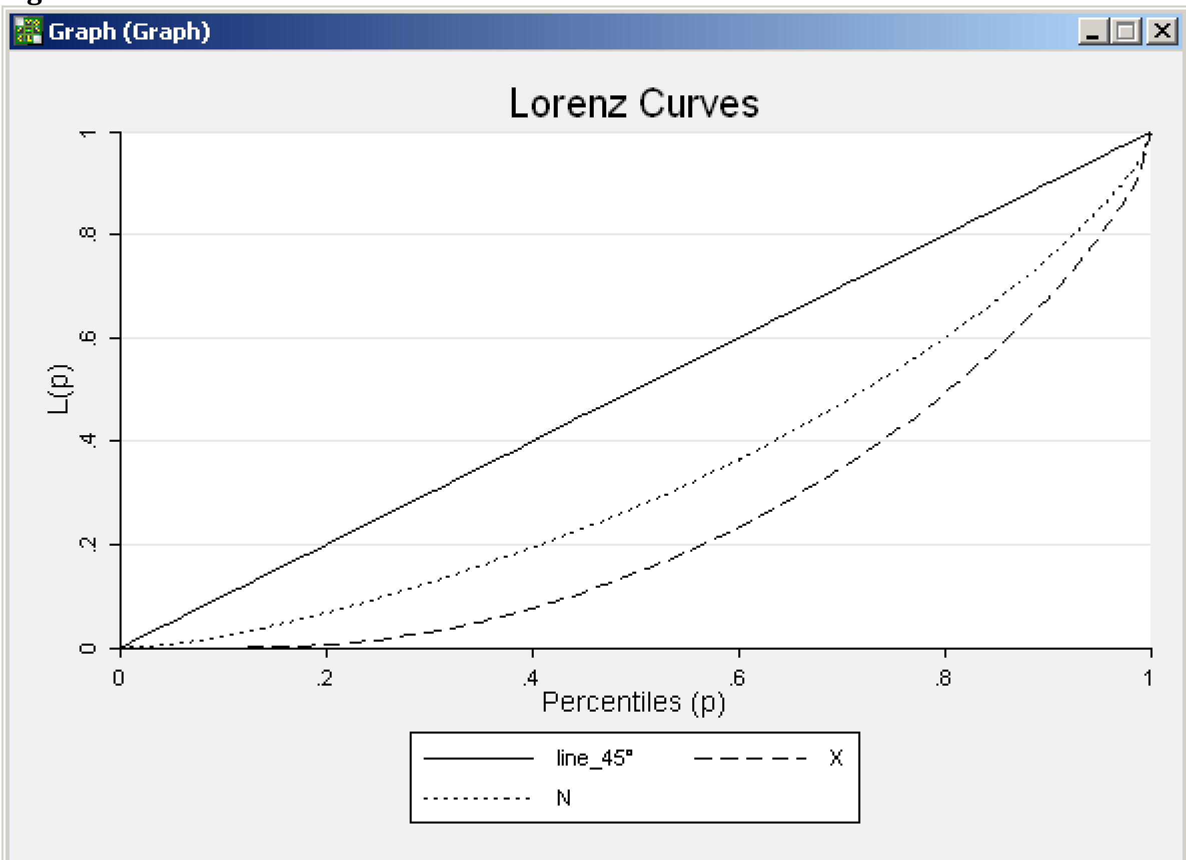
- Choose variables and parameters as in

Figure 43: Lorenz and concentration curves



After clicking SUBMIT, the following appears:

Figure 44: Lorenz curves

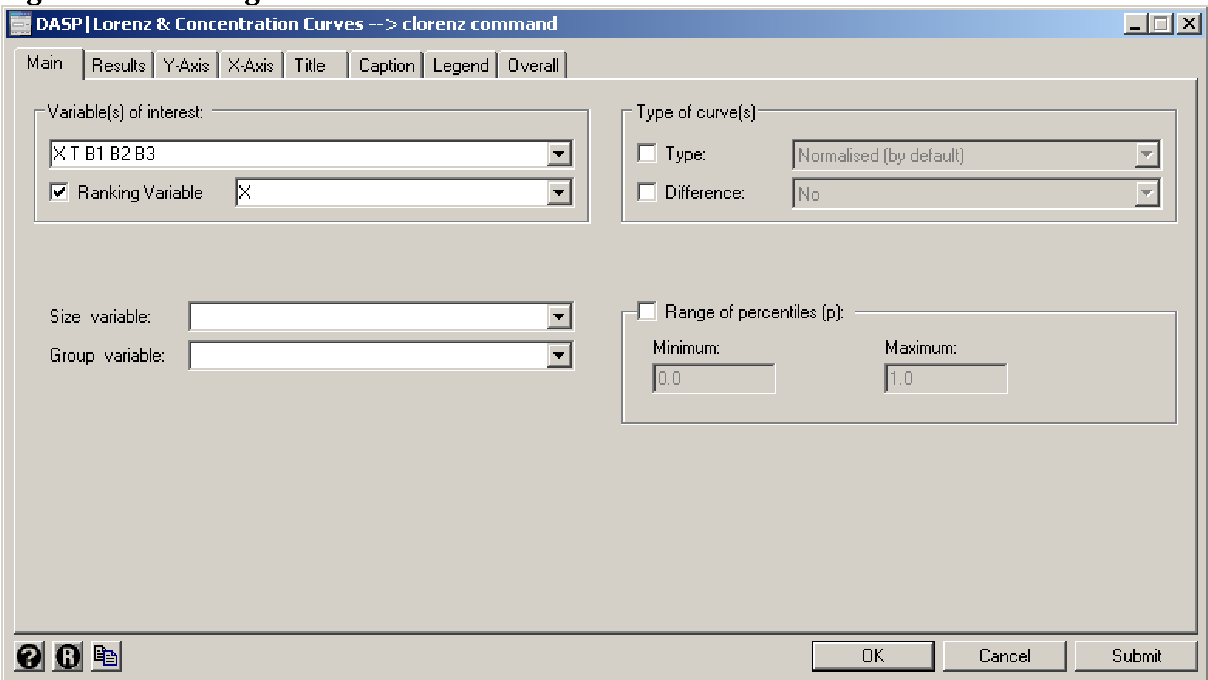


Q.2

Steps:

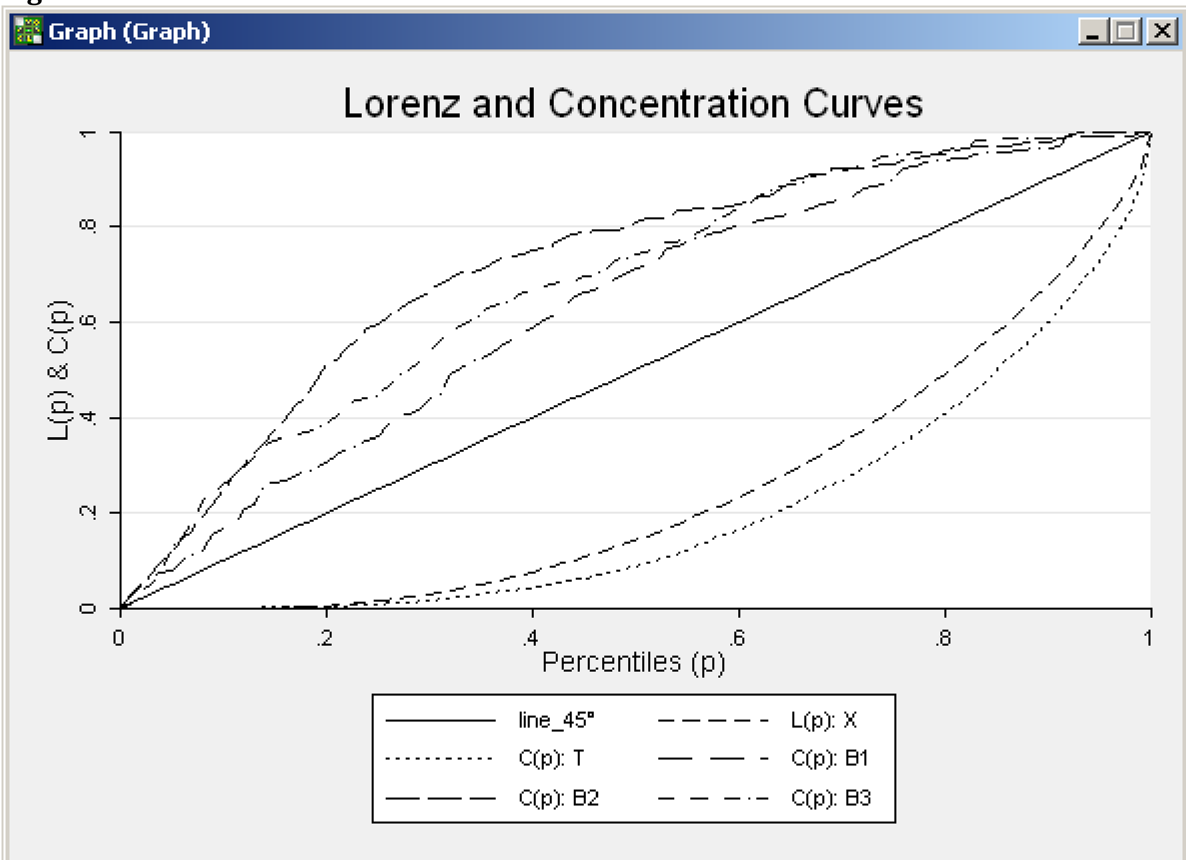
- Choose variables and parameters as in

Figure 45: Drawing concentration curves



After clicking on SUBMIT, the following appears:

Figure 46: Lorenz and concentration curves



Q.3

Steps:

- Type
  use "C:\data\bkf94I.dta", clear

- Choose variables and parameters as in

Figure 47: Drawing Lorenz curves

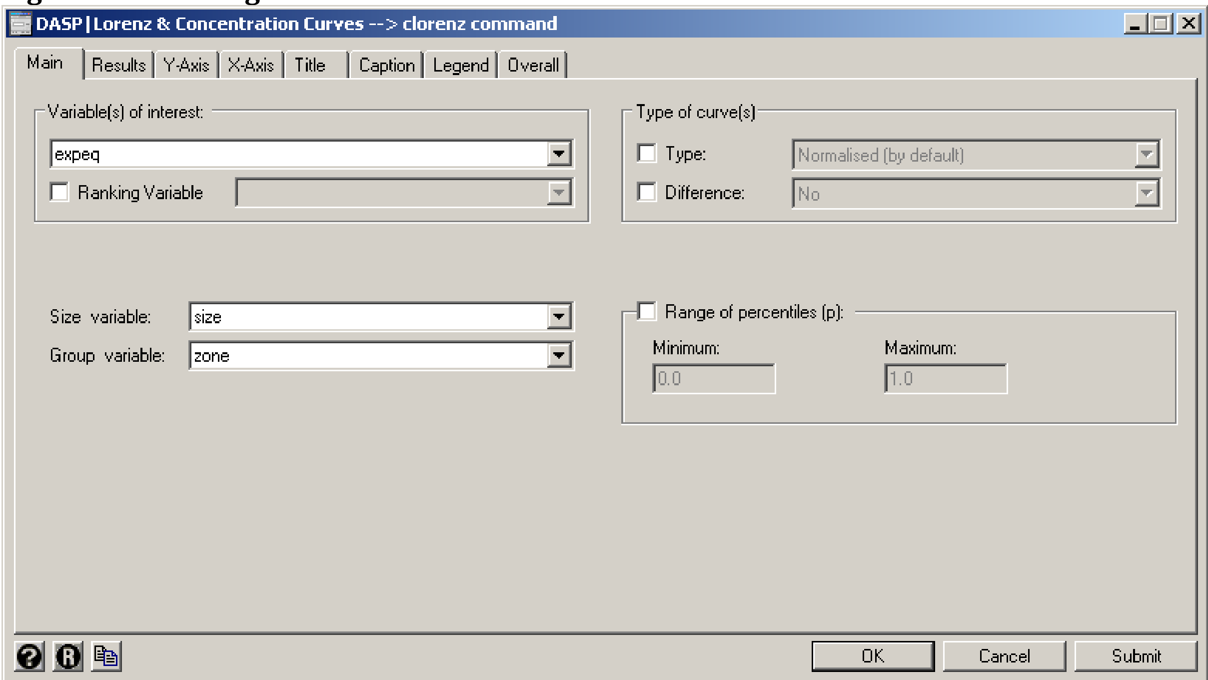
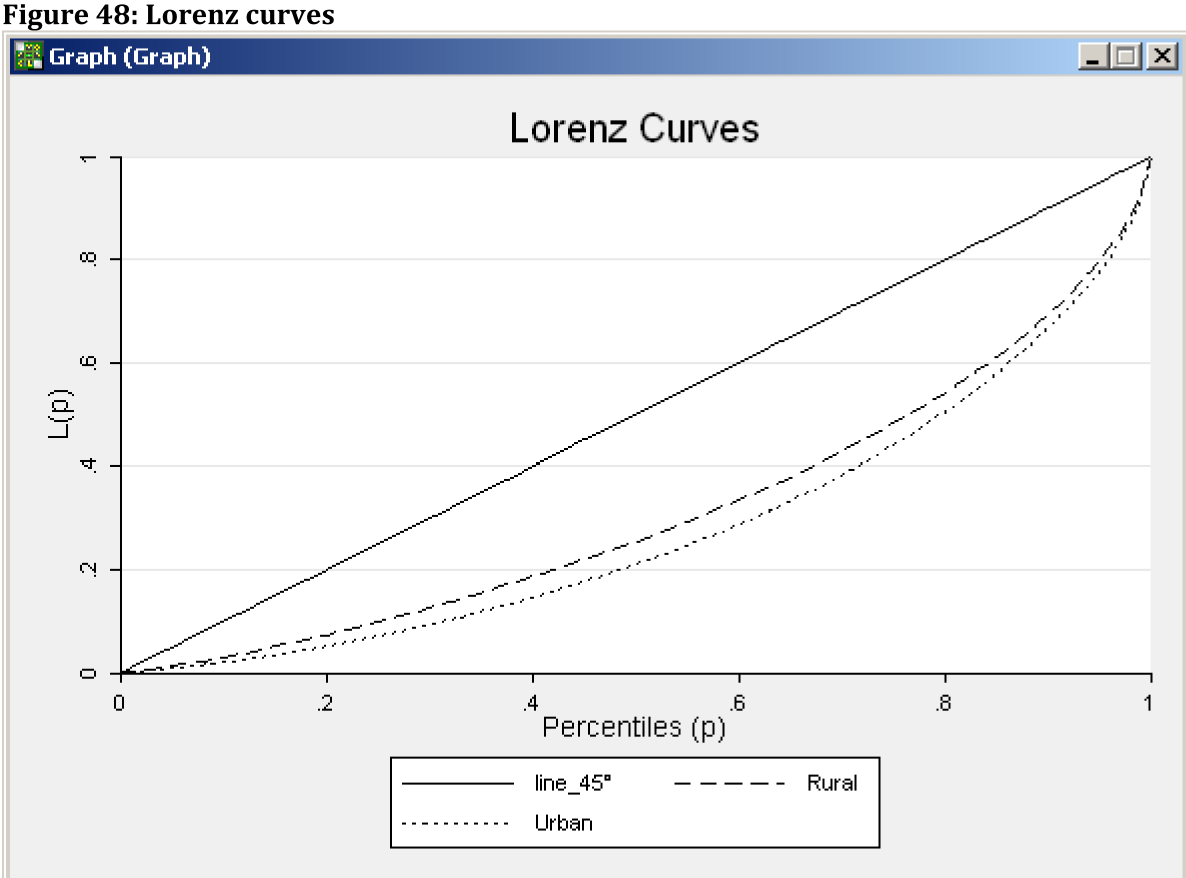


Figure 48: Lorenz curves



23.9 *Estimating Gini and concentration curves*

"By how much do taxes and transfers affect inequality in Canada?"

Using the ***can6.dta*** file,

1. Estimate the Gini indices for gross income *X* and net income *N*.
2. Estimate the concentration indices for variables *T* and *N* when the ranking variable is gross income *X*.

"By how much has inequality changed in Burkina Faso between 1994 and 1998?"

Using the ***bkf94I.dta*** file,

3. Estimate the difference in Burkina Faso's Gini index between 1998 and 1994

    a. with variable of interest *expeqz* for 1998 and *expeq* for 1994;
    b. with size variable set to *size*.

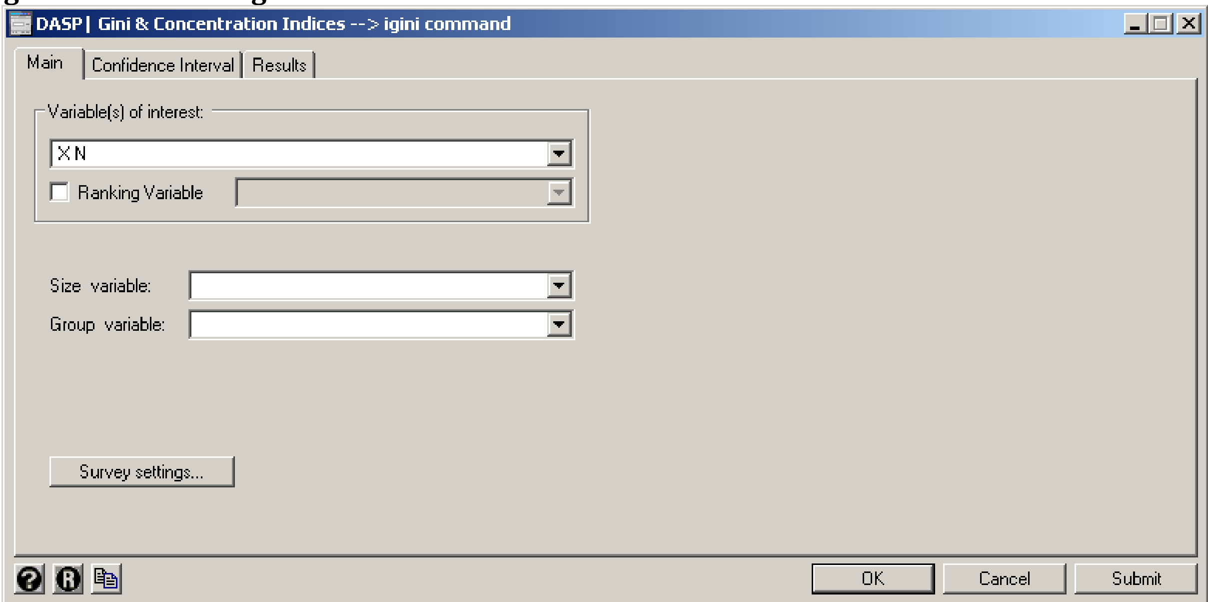
Q.1
Steps:

- Type
  use "C:\data\can6.dta", clear

- To open the relevant dialog box, type
  db igini

- Choose variables and parameters as in

Figure 49: Estimating Gini and concentration indices



After clicking SUBMIT, the following results are obtained:

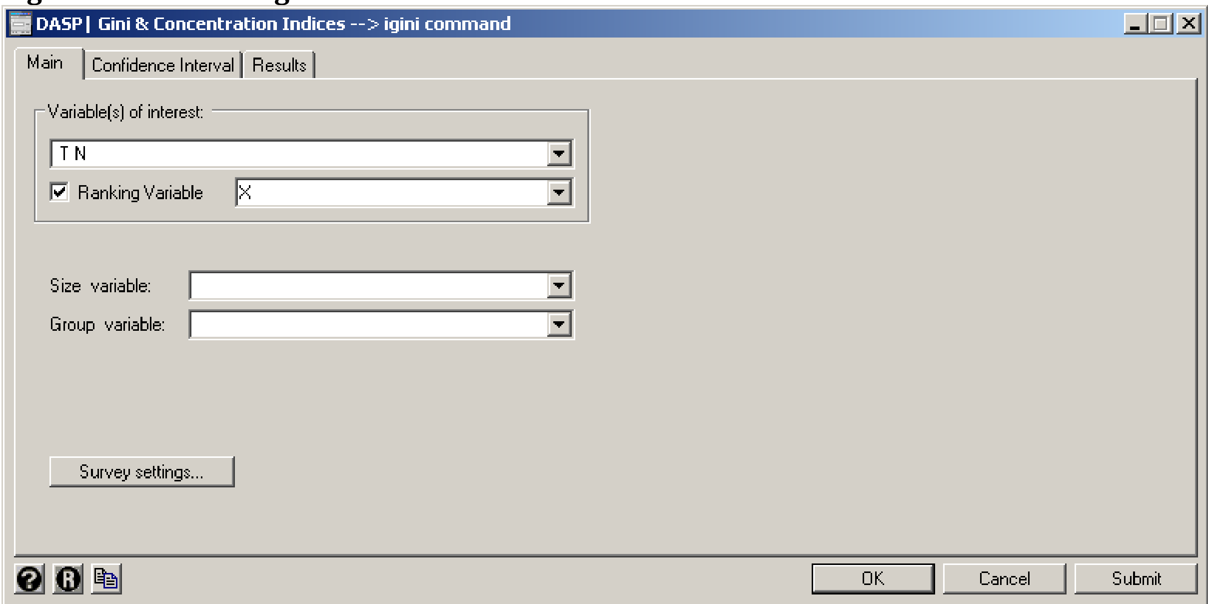
Variable	Estimate	STD	LB	UB
1: GINI_X	0.508456	0.016234	0.476599	0.540313
2: GINI_N	0.332355	0.012758	0.307318	0.357391

Q.2

Steps:

- Choose variables and parameters as in

Figure 50: Estimating concentration indices



After clicking SUBMIT, the following results are obtained:

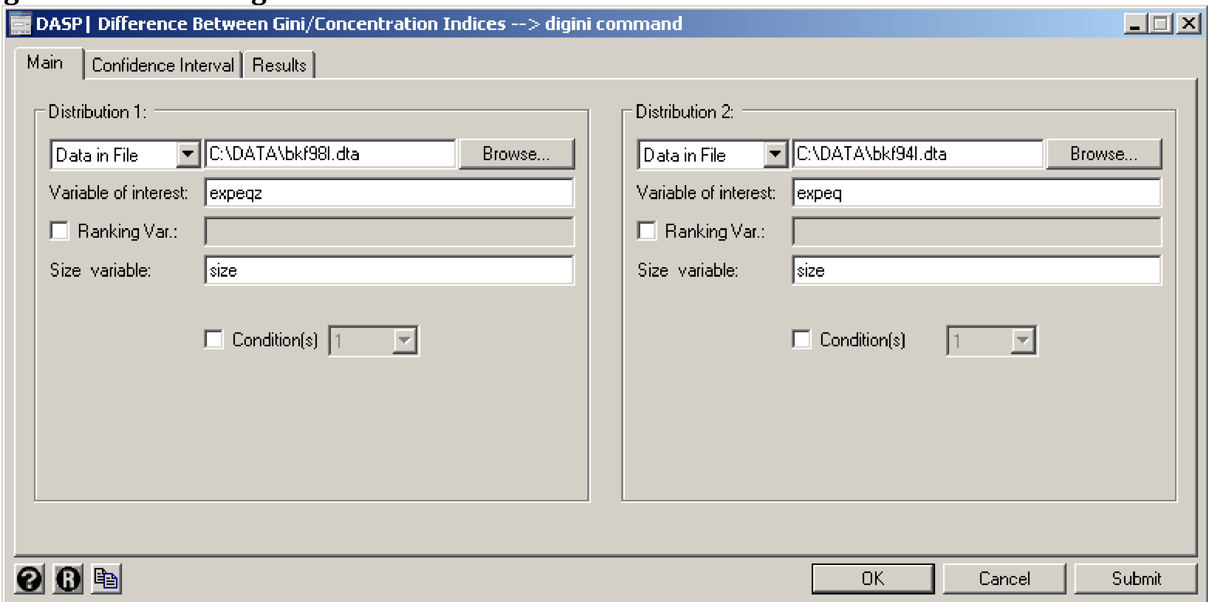
Variable	Estimate	STD	LB	UB
1: CONC_T	0.595339	0.022931	0.550340	0.640338
2: CONC_N	0.306050	0.013268	0.280014	0.332087

Q.3

Steps:

- To open the relevant dialog box, type
  db digini
- Choose variables and parameters as in

Figure 51: Estimating differences in Gini and concentration indices



After clicking SUBMIT, the following information is obtained:

```
. digini expeqz expeq, file1(C:\data\bkf98I.dta) hsize1(size) file2(C:\data\bkf94I.dta) hsize2(size)
```

	Estimate	STD	LB	UB
Distribution_1:(GINI)	**0.444563**	**0.012816**	**0.419371**	**0.469755**
Distribution_2:(GINI)	**0.450055**	**0.008618**	**0.433116**	**0.466994**
Difference	-0.005492	0.015444	-0.035762	0.024778

23.10 *Using basic distributive tools*

"What does the distribution of gross and net incomes look like in Canada?"

Using the ***can6.dta*** file,

1. Draw the density for gross income *X* and net income *N*.
   - The range for the x axis should be [0, 60 000].
2. Draw the quantile curves for gross income *X* and net income *N*.
   - The range of percentiles should be [0, 0.8]
3. Draw the expected tax/benefit according to gross income *X.*
   - The range for the x axis should be [0, 60 000]
   - Use a *local linear estimation* approach.
4. Estimate marginal rates for taxes and benefits according to gross income *X.*
   - The range for the x axis should be [0, 60 000]
   - Use a *local linear estimation* approach.

Q.1

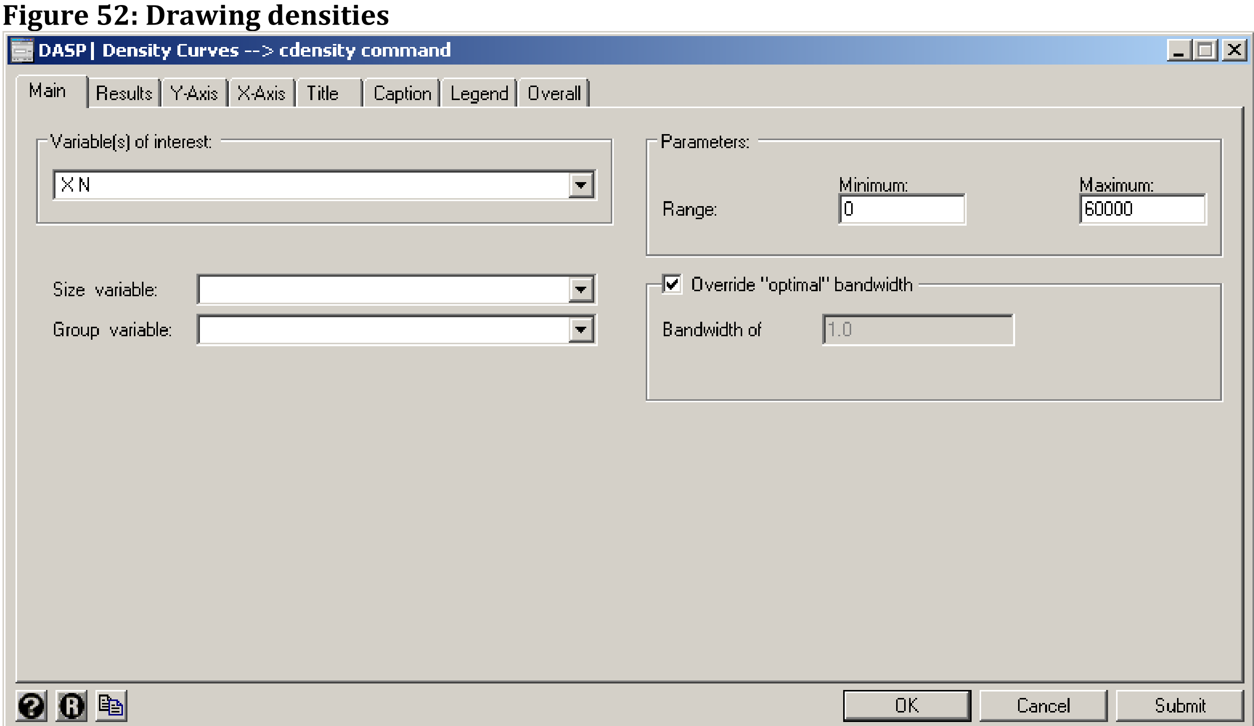
Steps:

- Type
  use "C:\data\can6.dta", clear

- To open the relevant dialog box, type
  db cdensity

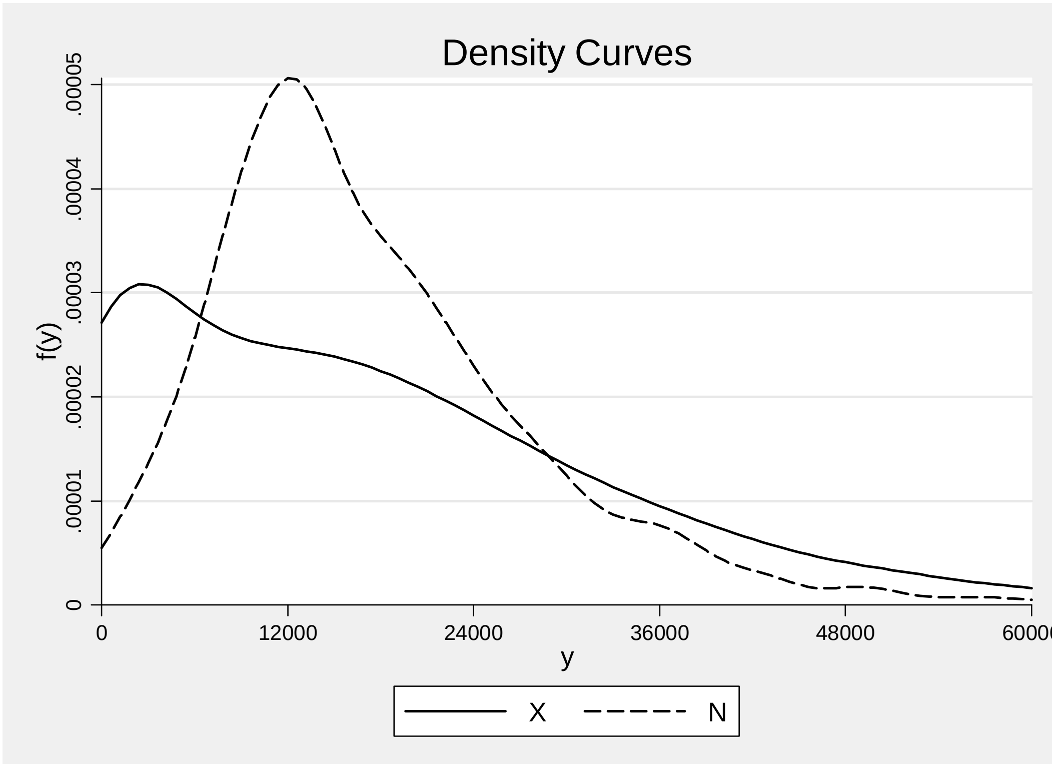
- Choose variables and parameters as in

Figure 52: Drawing densities



After clicking SUBMIT, the following appears:

Figure 53: Density curves

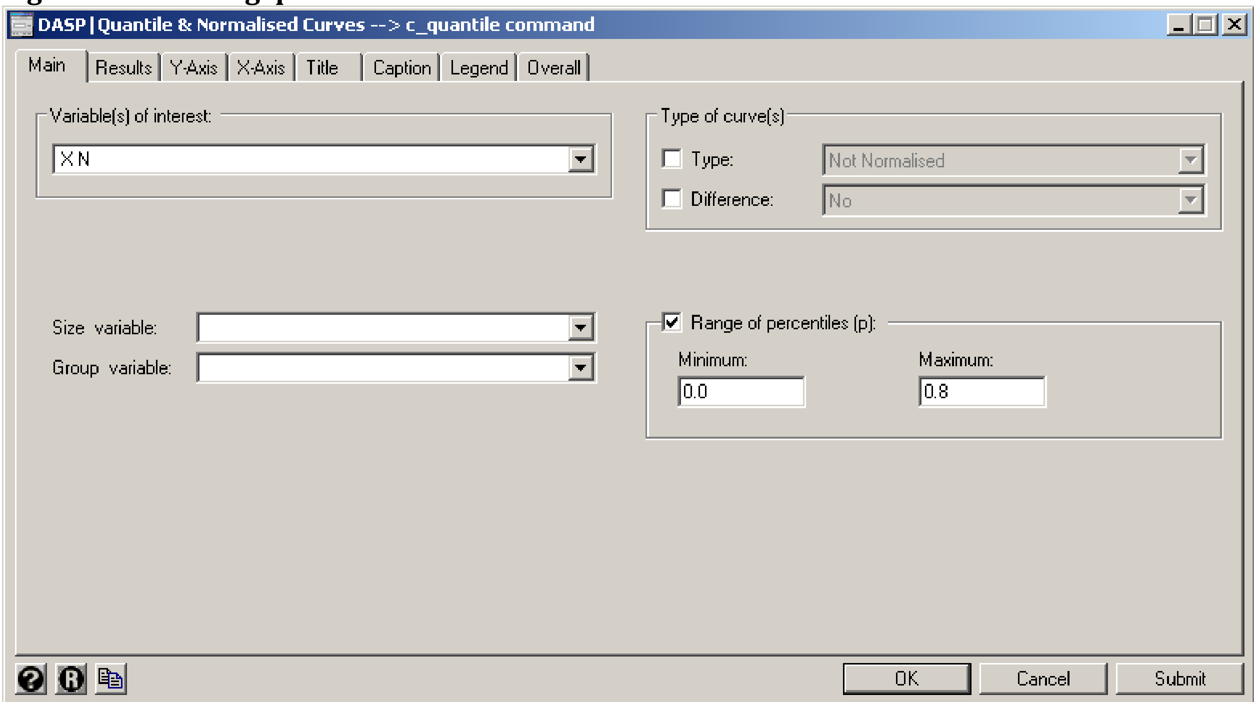


Q.2
Steps:

- To open the relevant dialog box, type
  db c_quantile

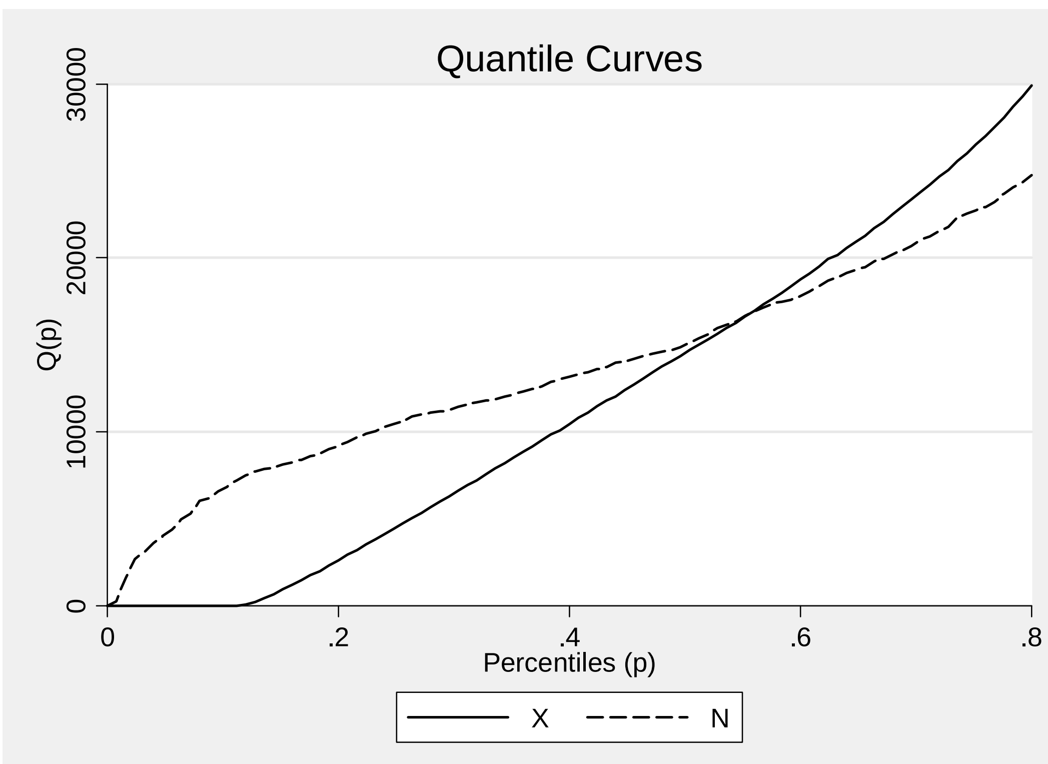
- Choose variables and parameters as in

Figure 54: Drawing quantile curves



After clicking SUBMIT, the following appears:

Figure 55: Quantile curves

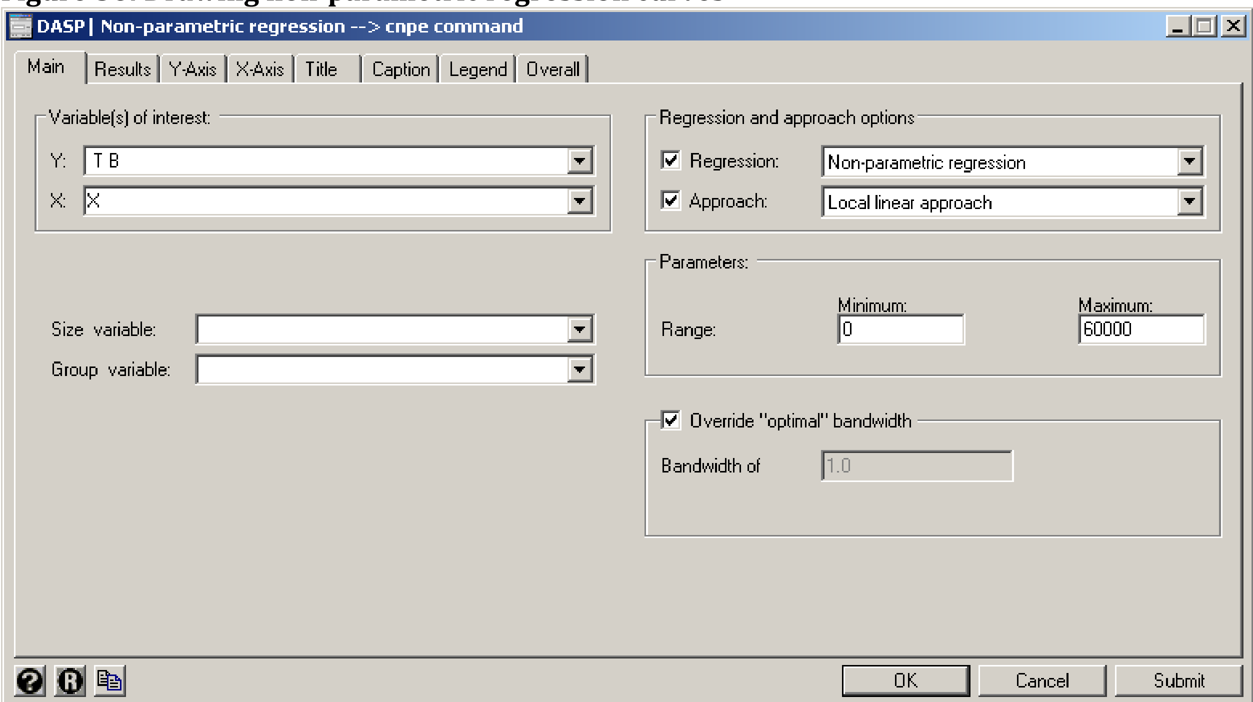


Q.3
Steps:

- To open the relevant dialog box, type
  db cnpe

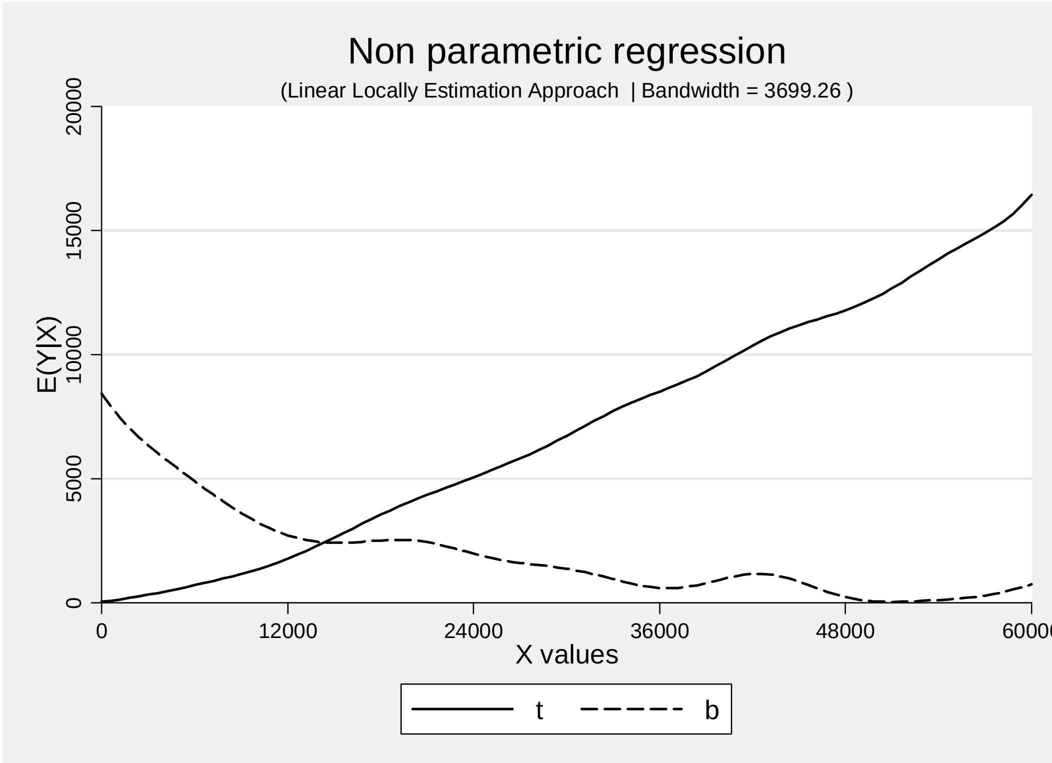
- Choose variables and parameters as in

Figure 56: Drawing non-parametric regression curves



After clicking SUBMIT, the following appears:

Figure 57: Non-parametric regression curves

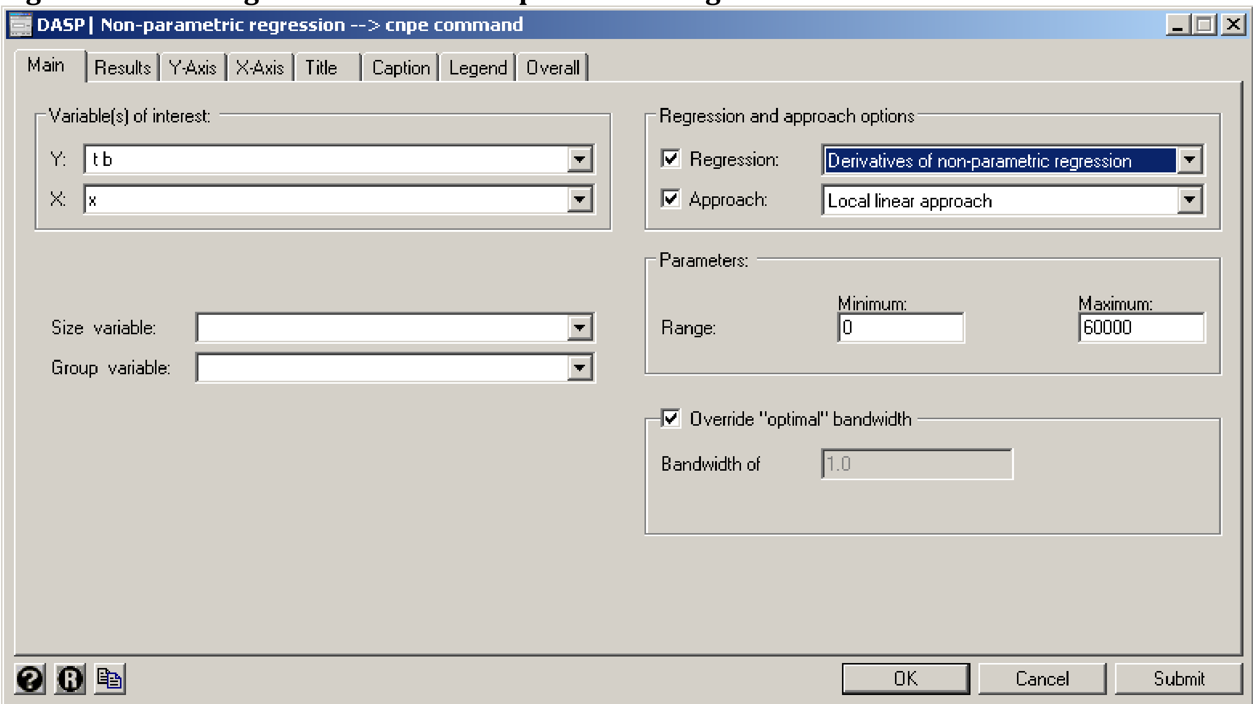


Q.4

Steps:

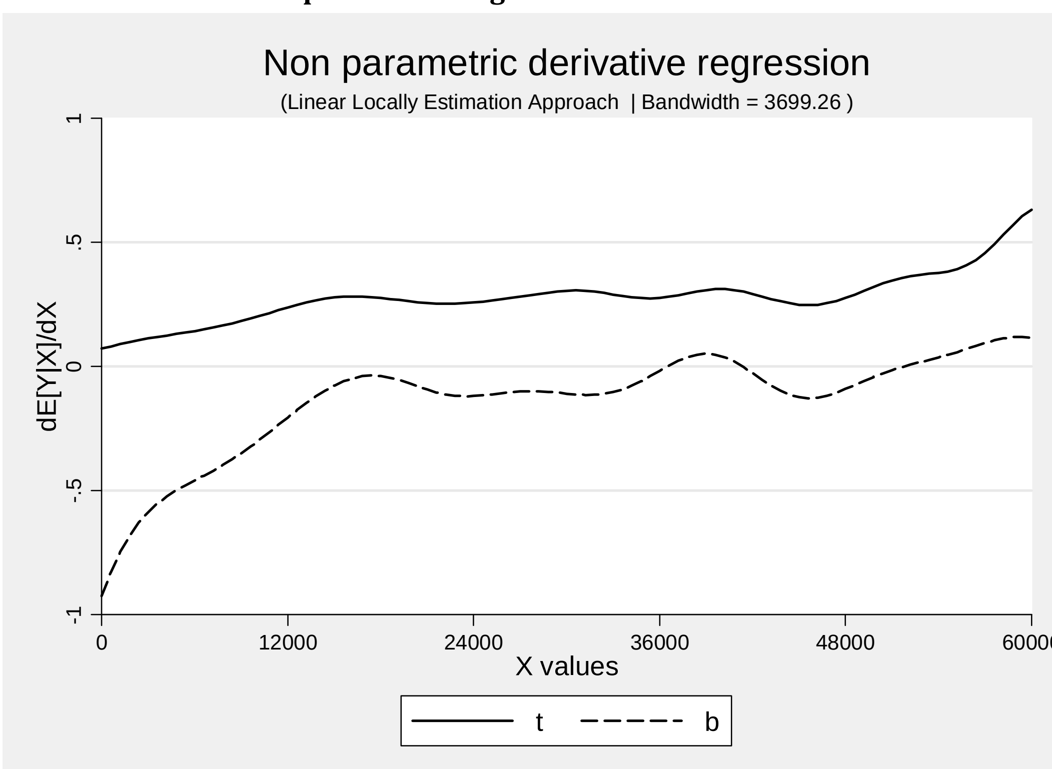
- Choose variables and parameters as in

Figure 58: Drawing derivatives of non-parametric regression curves



After clicking SUBMIT, the following appears:

Figure 59: Derivatives of non-parametric regression curves



23.11 *Plotting the joint density and joint distribution function*

"What does the joint distribution of gross and net incomes look like in Canada?"

Using the ***can6.dta*** file,

4. Estimate the joint density function for gross income *X* and net income *N*.
    - X range : [0,60000]
    - N range : [0,60000]

5. Estimate the joint distribution function for gross income *X* and net income *N*.
    - X range : [0,60000]
    - N range : [0,60000]

Q.1

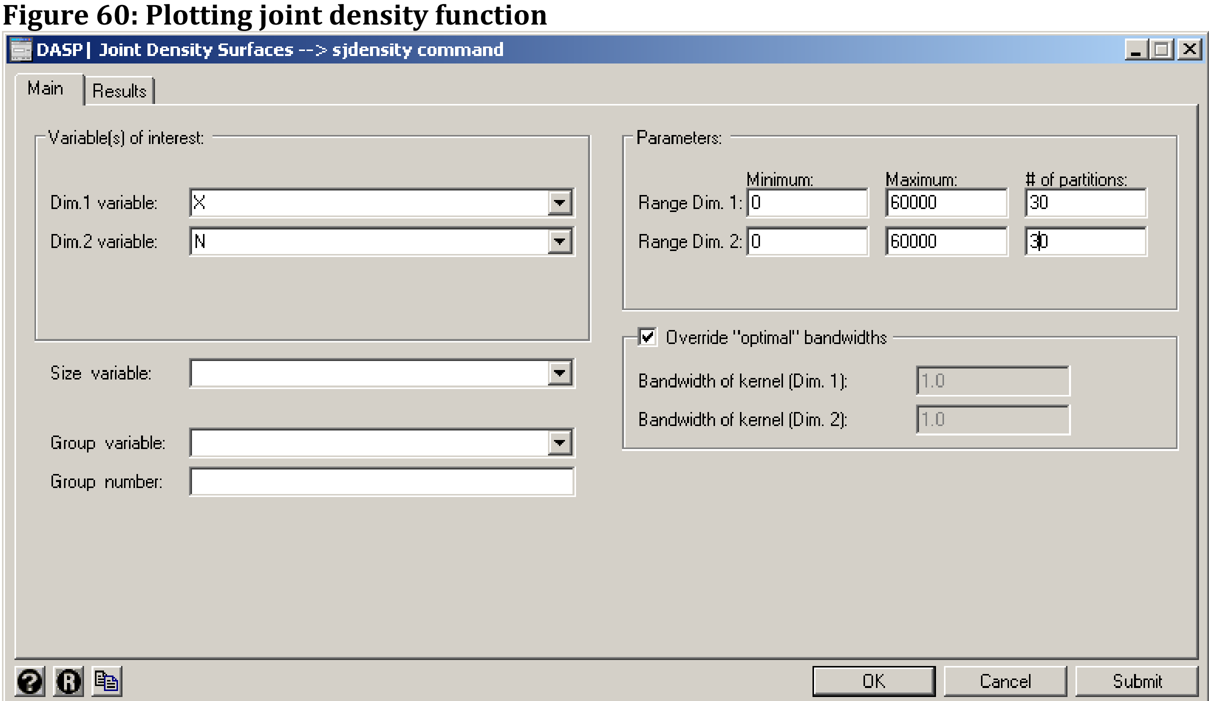
Steps:

- Type
  use "C:\data\can6.dta", clear

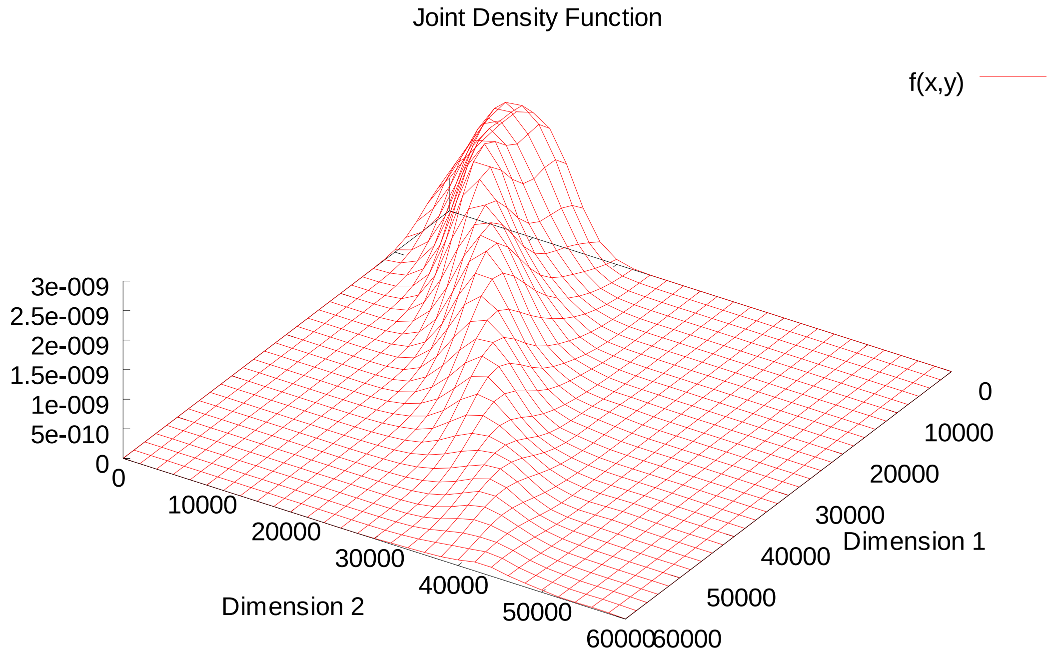
- To open the relevant dialog box, type
  db sjdensity

- Choose variables and parameters as in

Figure 60: Plotting joint density function



After clicking SUBMIT, the following graph is plotted interactively with Gnu Plot 4.2:



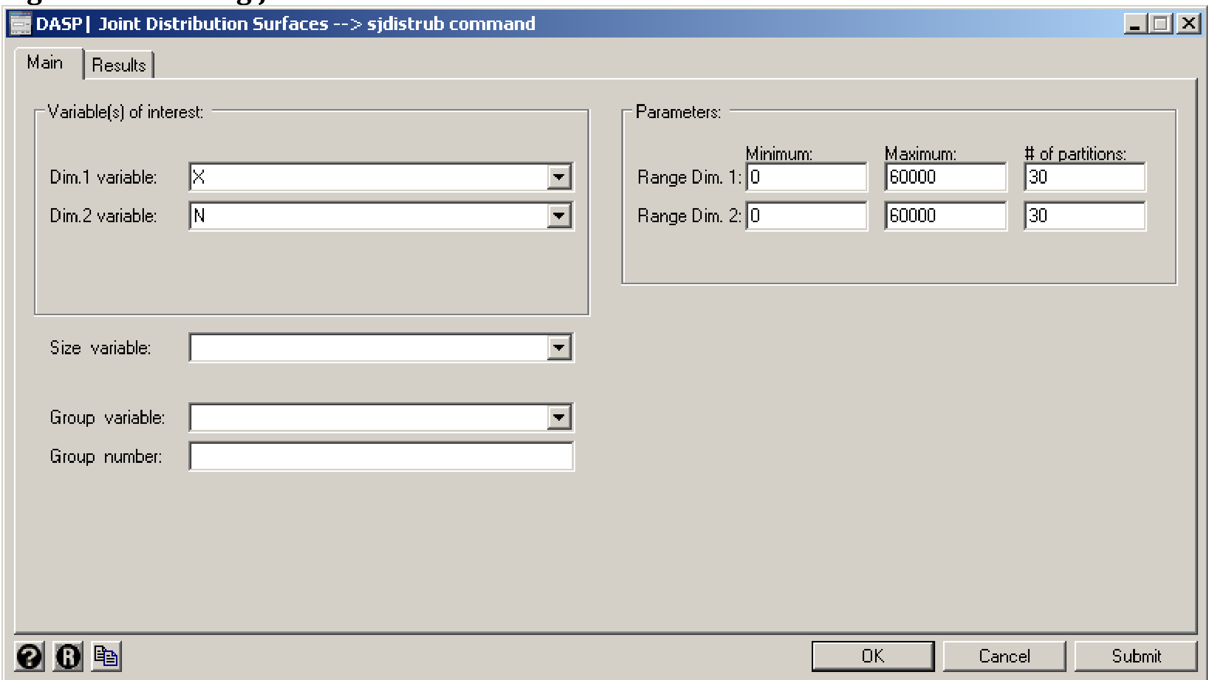
Q.2

Steps:

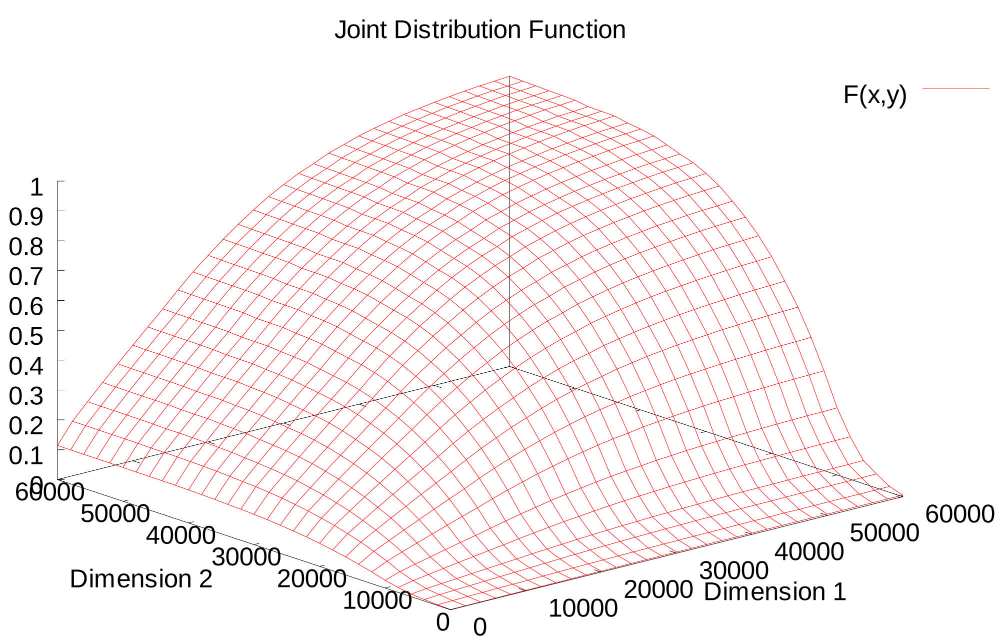
- To open the relevant dialog box, type
  db sjdistrub

- Choose variables and parameters as in

Figure 61: Plotting joint distribution function



After clicking SUBMIT, the following graph is plotted interactively with Gnu Plot 4.2:



23.12 Testing the bi-dimensional poverty dominance

Using the ***columbia95I.dta*** (distribution_1) and the ***dominican_republic95I.dta*** (distribution_2) files,

1. Draw the difference between the bi-dimensional multiplicative FGT surfaces and the confidence interval of that difference when

	Var. of interest	Range	alpha_j
Dimension 1	*haz* : height-for-age	-3.0 / 6.0	0
Dimension 2	*sprob* : survival probability	0.7 / 1.0	0

2. Test for bi-dimensional poverty using the information above.

Answer:

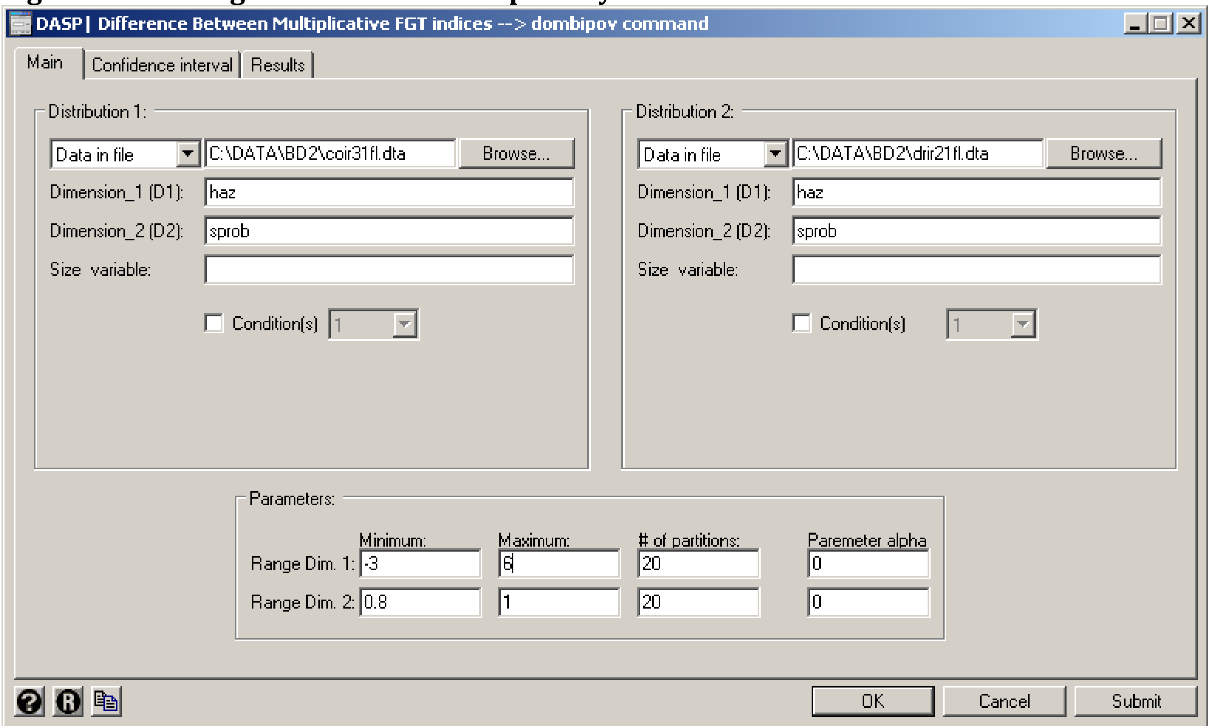
Q.1

Steps:

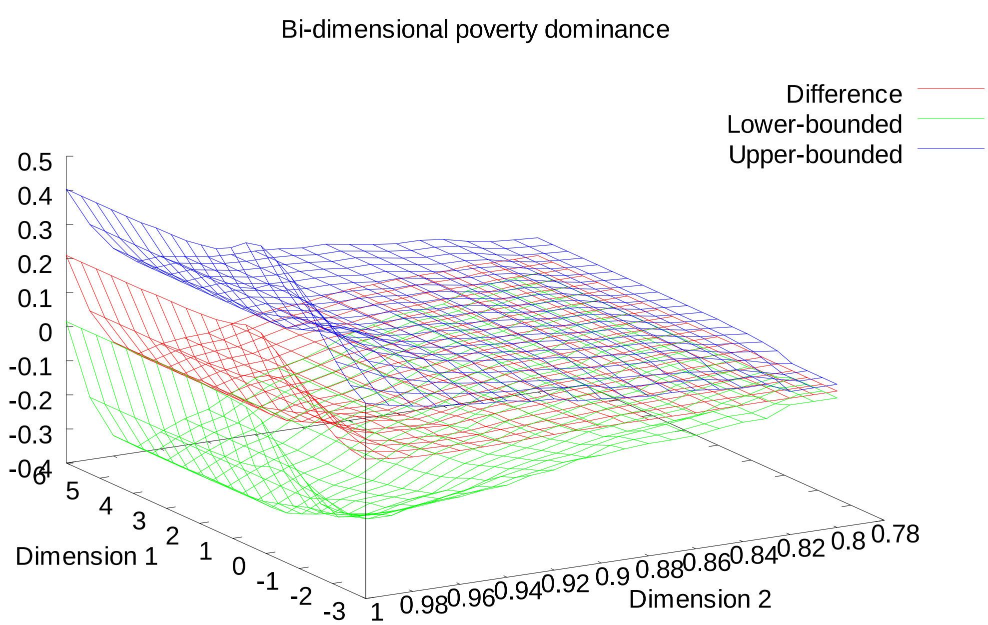
- To open the relevant dialog box, type
  db dombdpov

- Choose variables and parameters as in

Figure 62: Testing for bi-dimensional poverty dominance



After clicking SUBMIT, the following graph is plotted interactively with Gnu Plot 4.2:

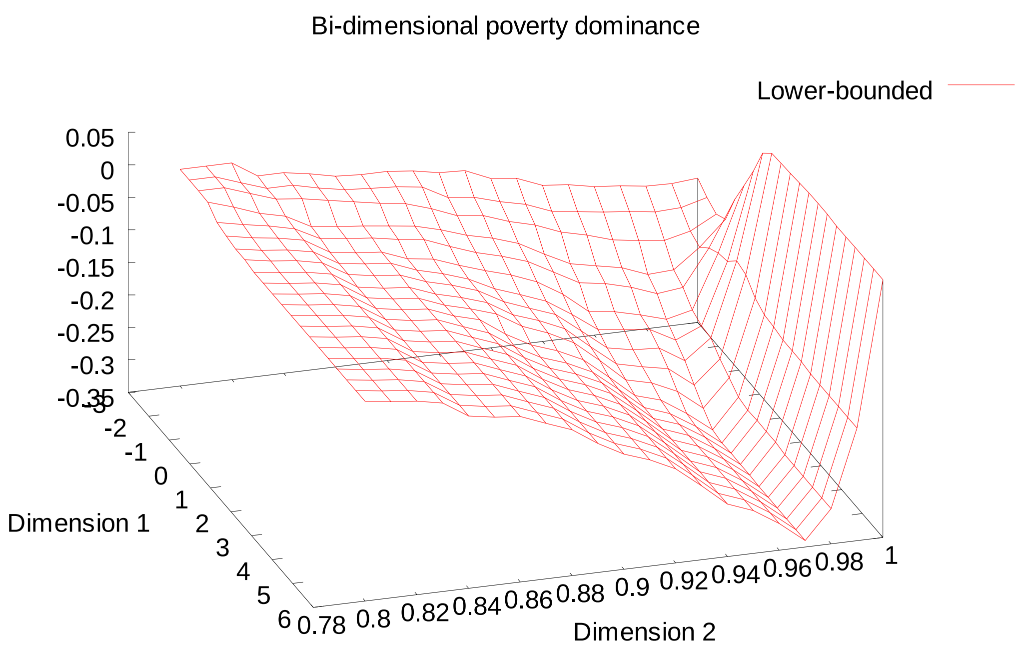


Q.2

To make a simple test of multidimensional dominance, one should check if the lower-bounded confidence interval surface is always above zero for all combinations of relevant poverty lines – or conversely.

- For this, click on the panel "Confidence interval" and select the option lower-bounded.
- Click again on the button Submit.

After clicking SUBMIT, the following graph is plotted interactively with Gnu Plot 4.2:



23.13 Testing for pro-poorness of growth in Mexico

The three sub-samples used in these exercises are sub-samples of 2000 observations drawn randomly from the three ENIGH Mexican household surveys for 1992, 1998 and 2004. Each of these three sub-samples contains the following variables:

strata	The stratum
psu	The primary sampling unit
weight	Sampling weight
inc	Income
hhsz	Household size

1. Using the files mex_92_2mI.dta and mex_98_2mI.dta, test for first-order relative pro-poorness of growth when:
   - The primal approach is used.
   - The range of poverty lines is [0, 3000].

2. Repeat with the dual approach.

3. By using the files mex_98_2mI.dta and mex_04_2mI.dta, test for absolute second-order pro-poorness with the dual approach.

4. Using mex_98_2mI.dta and mex_04_2mI.dta, estimate the pro-poor indices of module ipropoor.
   - Parameter alpha set to 1.
   - Poverty line equal to 600.

Answer:

Q.1

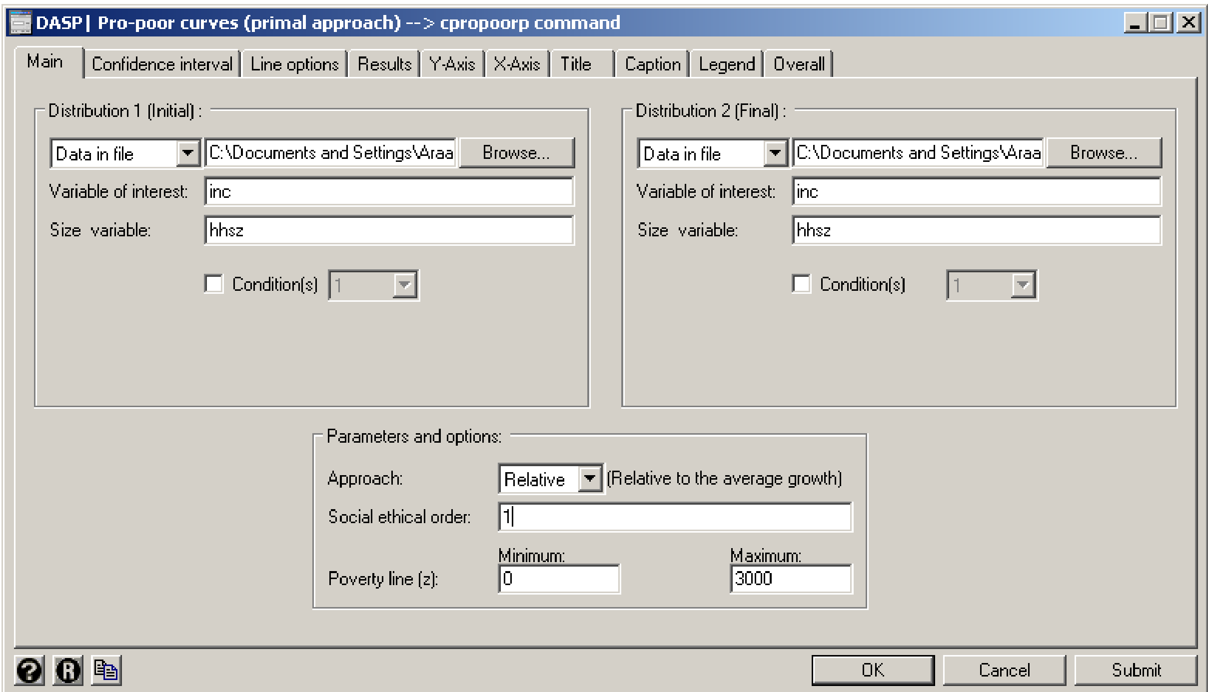
Steps:

- To open the relevant dialog box, type

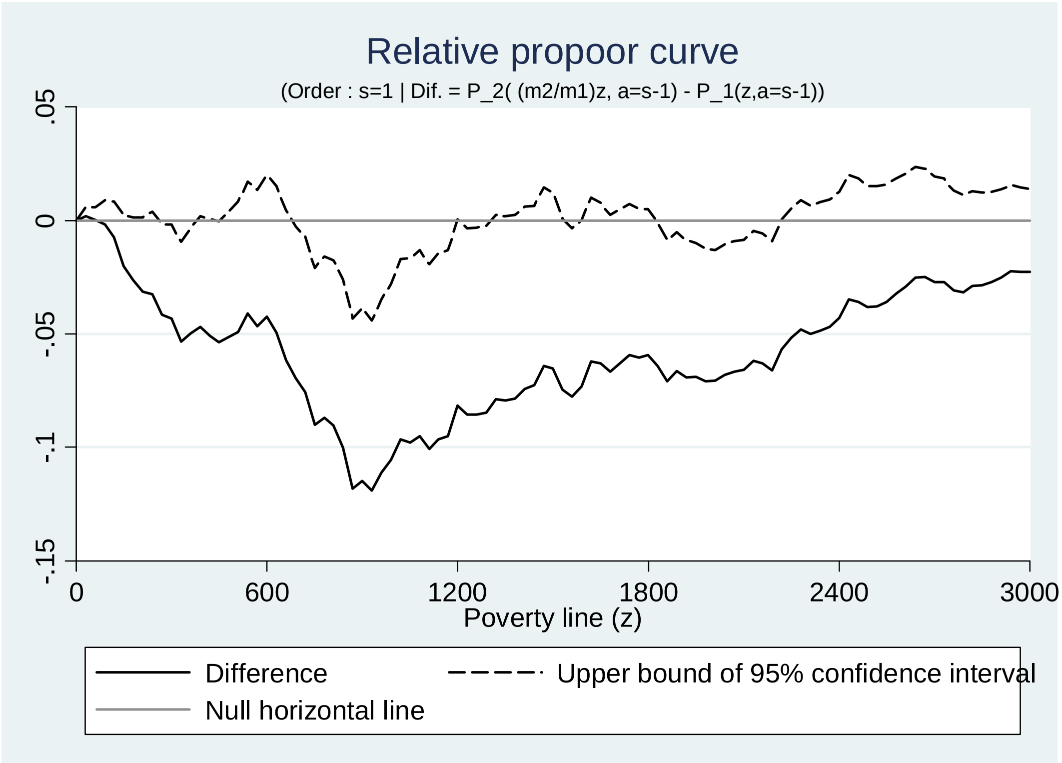
db cpropoorp

- Choose variables and parameters as in (select the upper-bounded option for the confidence interval):

Figure 63: Testing the pro-poor growth (primal approach)



After clicking SUBMIT, the following graph appears



Q.2

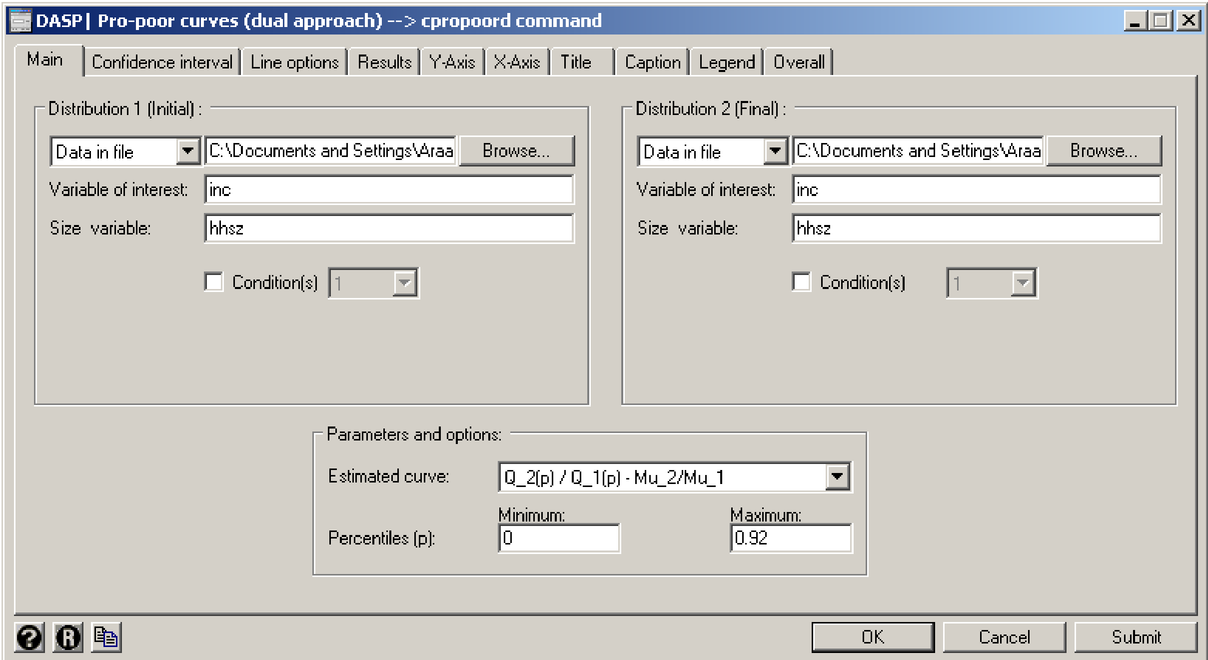
Steps:

- To open the relevant dialog box, type

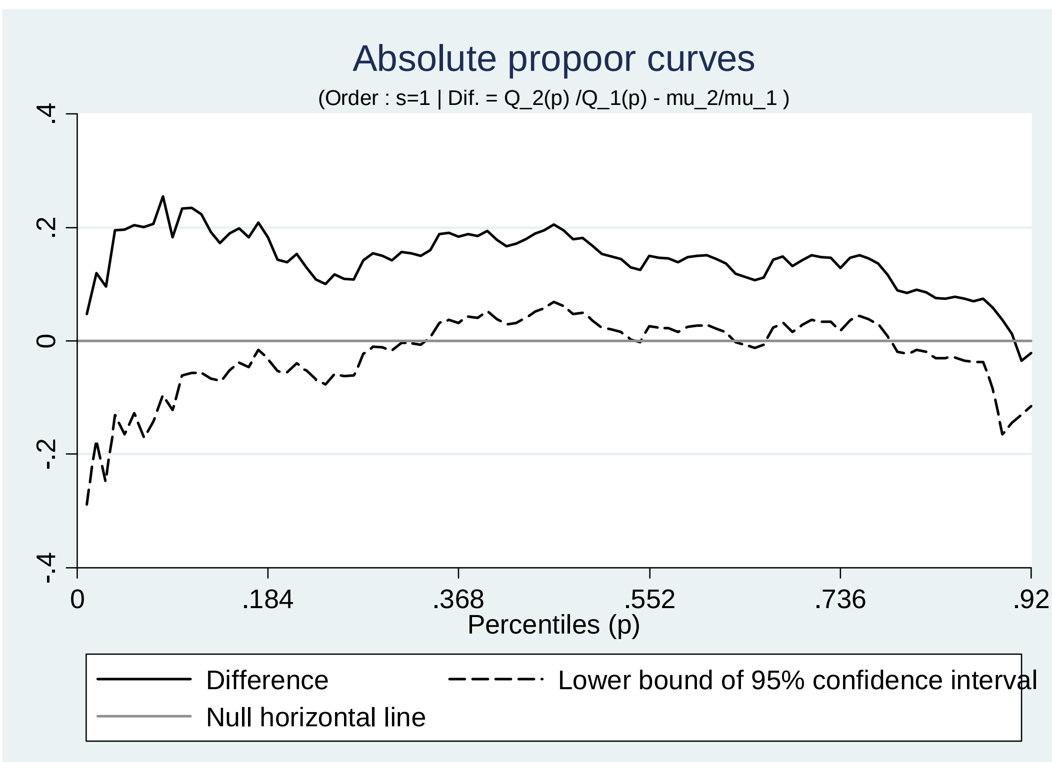
db cpropoord

- Choose variables and parameters as in (with the lower-bounded option for the confidence interval):

Figure 64: Testing the pro-poor growth (dual approach)- A



After clicking SUBMIT, the following graph appears



Q.2

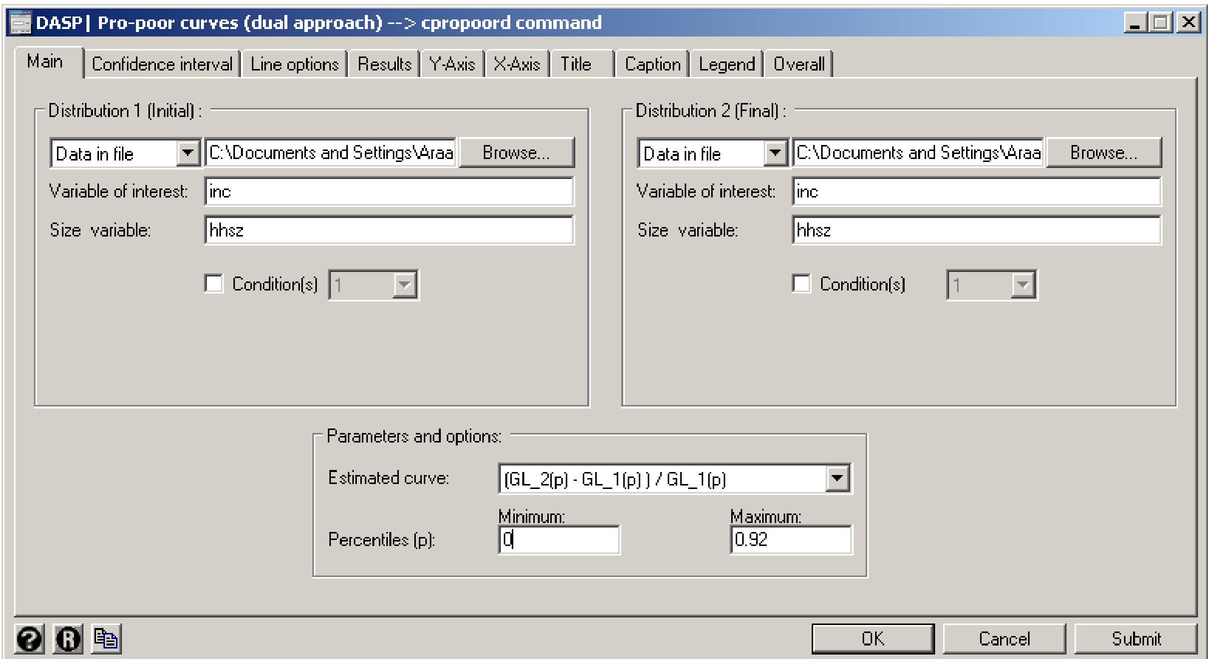
Steps:

- To open the relevant dialog box, type

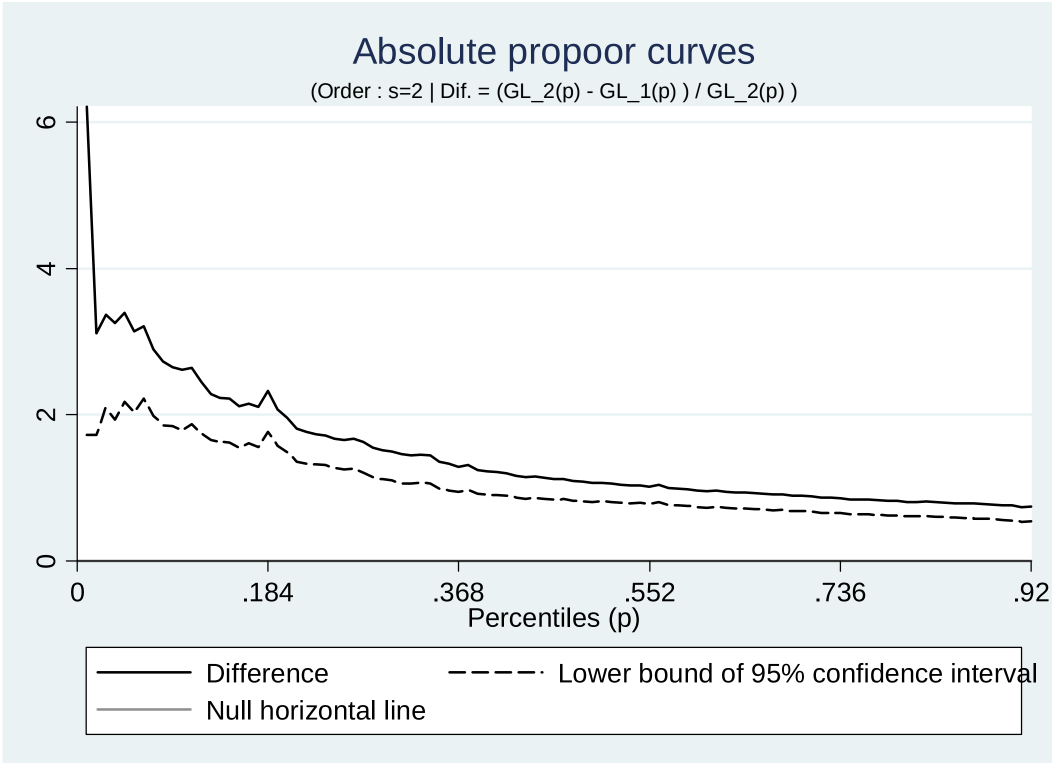
db cpropoord

- Choose variables and parameters as in (with the lower-bounded option for the confidence interval):

Figure 65: Testing the pro-poor growth (dual approach) – B



After clicking SUBMIT, the following graph appears



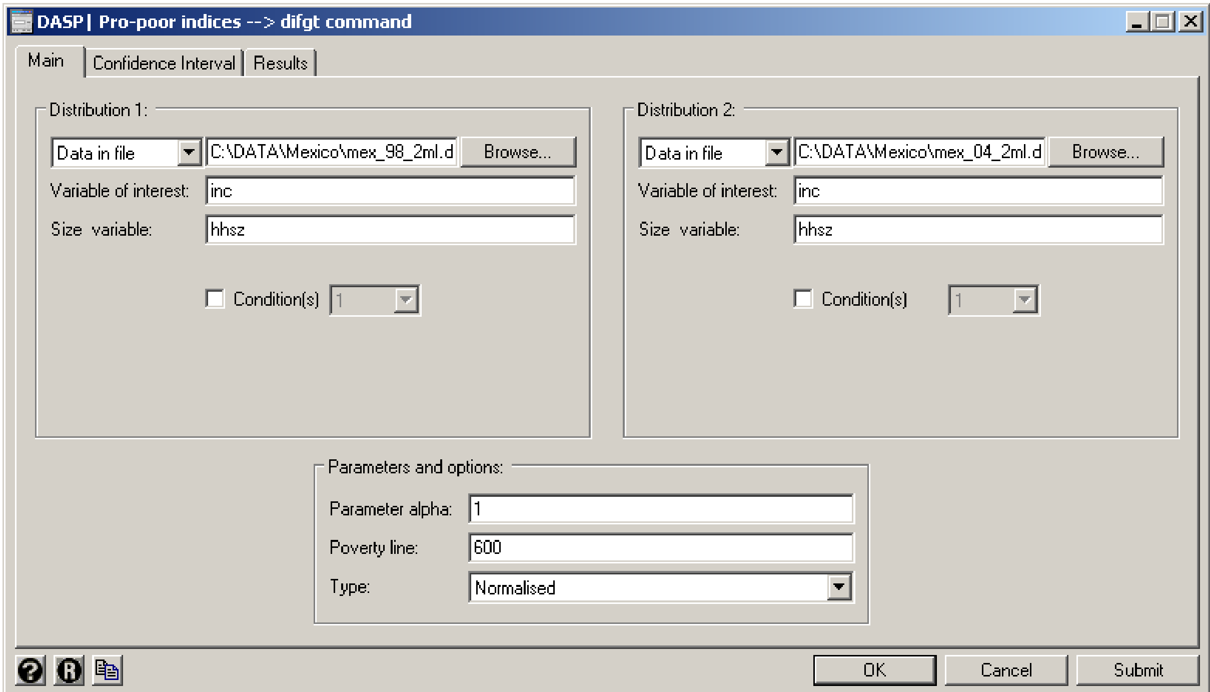
Q.4

Steps:

- To open the relevant dialog box, type

db ipropoor

- Choose variables and parameters as.



After clicking SUBMIT, the following results appear:

Poverty line : 600.00
Parameter alpha : 1.00

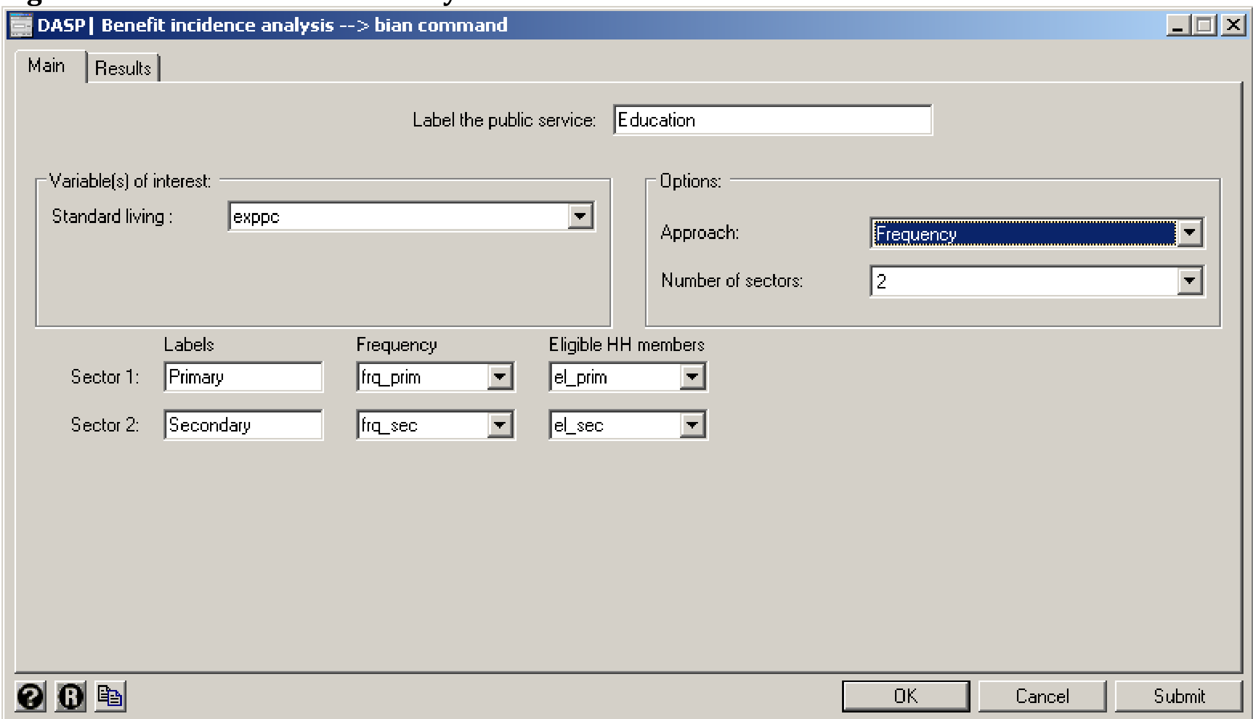
Pro-poor indices	Estimate	STD	LB	UB
Growth rate(g)	**0.582359**	**0.125512**	**0.336361**	**0.828357**
Chen & Ravallion (2003) index	**0.712285**	**1.009337**	**-1.265979**	**2.690549**
Kakwani & Pernia (2000) index	**1.325436**	**0.107047**	**1.115627**	**1.535244**
PEGR index	**0.771879**	**0.137331**	**0.502716**	**1.041042**
PEGR - g	**0.189520**	**0.049357**	**0.092783**	**0.286257**

23.14 Benefit incidence analysis of public spending on education in Peru (1994).

1. Using the peredu94I.dta file, estimate participation and coverage rates of two types of public spending on education when:
   - The standard of living is **exppc**
   - The number of household members that benefit from education is **fr_prim** for the primary sector and **fr_sec** for the secondary one.
   - The number of eligible household members is **el_prim** for the primary sector and **el_sec** for the secondary one.
   - Social groups are quintiles.

Answer:
Type **db bian** in the windows command and set variables and options as follows:

Figure 66: Benefit incidence analysis



After clicking on **Submit,** the following appears:

Benefit Incidence Analysis: Education

Share by Quintile Groups.

Groups	Primary	Secondary
Quintile 1	**0.218**	**0.226**
Quintile 2	**0.231**	**0.252**
Quintile 3	**0.221**	**0.225**
Quintile 4	**0.195**	**0.194**
Quintile 5	**0.135**	**0.102**
All	**1.000**	**1.000**

Rate of Participation by Quintile Groups.

Groups	Primary	Secondary
Quintile 1	**0.762**	**0.325**
Quintile 2	**0.808**	**0.363**
Quintile 3	**0.772**	**0.323**
Quintile 4	**0.683**	**0.279**
Quintile 5	**0.472**	**0.147**
All	**0.699**	**0.287**

2. To estimate total public expenditures on education by sector at the national level, the following macro information was used:
   - Pre-primary and primary public education expenditure (as % of all levels), 1995: 35.2%
   - Secondary public education expenditure (as % of all levels), 1995: 21.2%
   - Tertiary public education expenditure (as % of all levels), 1995: 16%
   - Public education expenditure (as % of GNP), 1995 = 3%
   - GDP per capita: about 3 800.

   Using this information, the following variables are generated

```
cap drop _var1;
gen _var1 = size*weight*3800;
qui sum _var1;
qui gen pri_pub_exp=0.03*0.352*`r(sum)';
qui gen sec_pub_exp=0.03*0.212*`r(sum)';
qui gen uni_pub_exp=0.03*0.160*`r(sum)';
cap drop _var1;
```

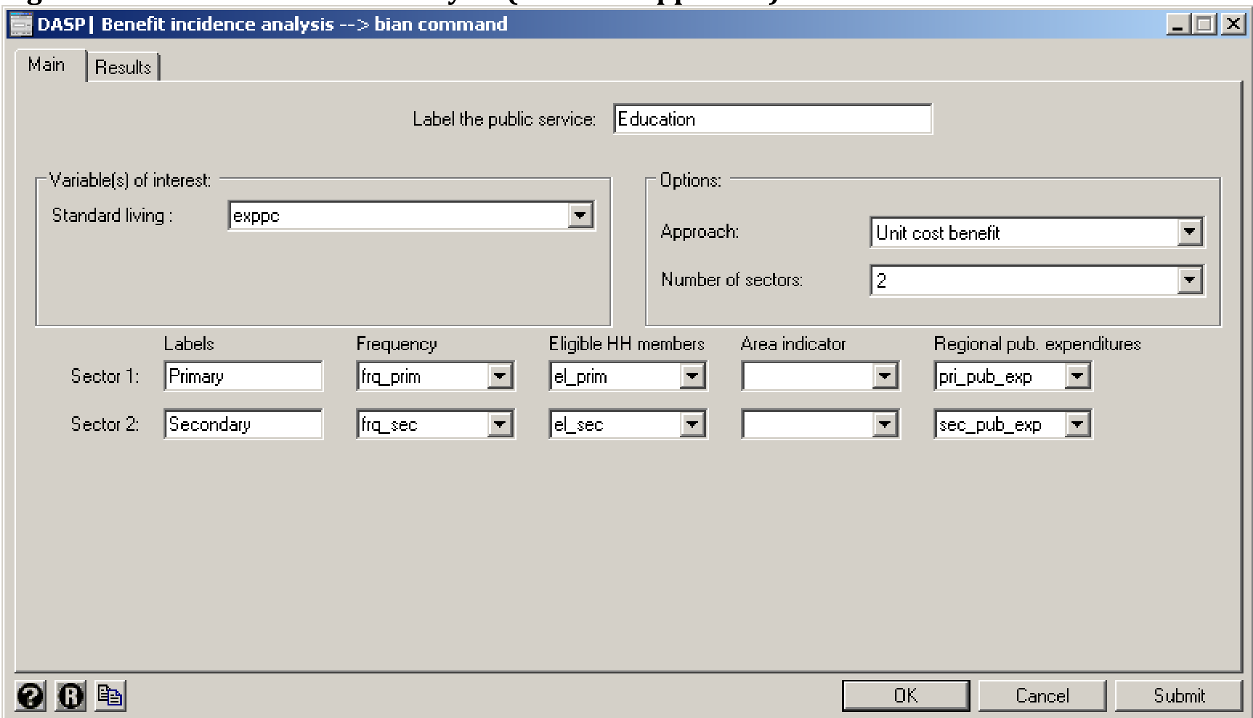
   - Total public expenditures on primary sector : **pri_pub_exp**
   - Total public expenditures on secondary sector : **sec_sec_exp**
   - Total public expenditures on university sector : **uni_pub_exp**

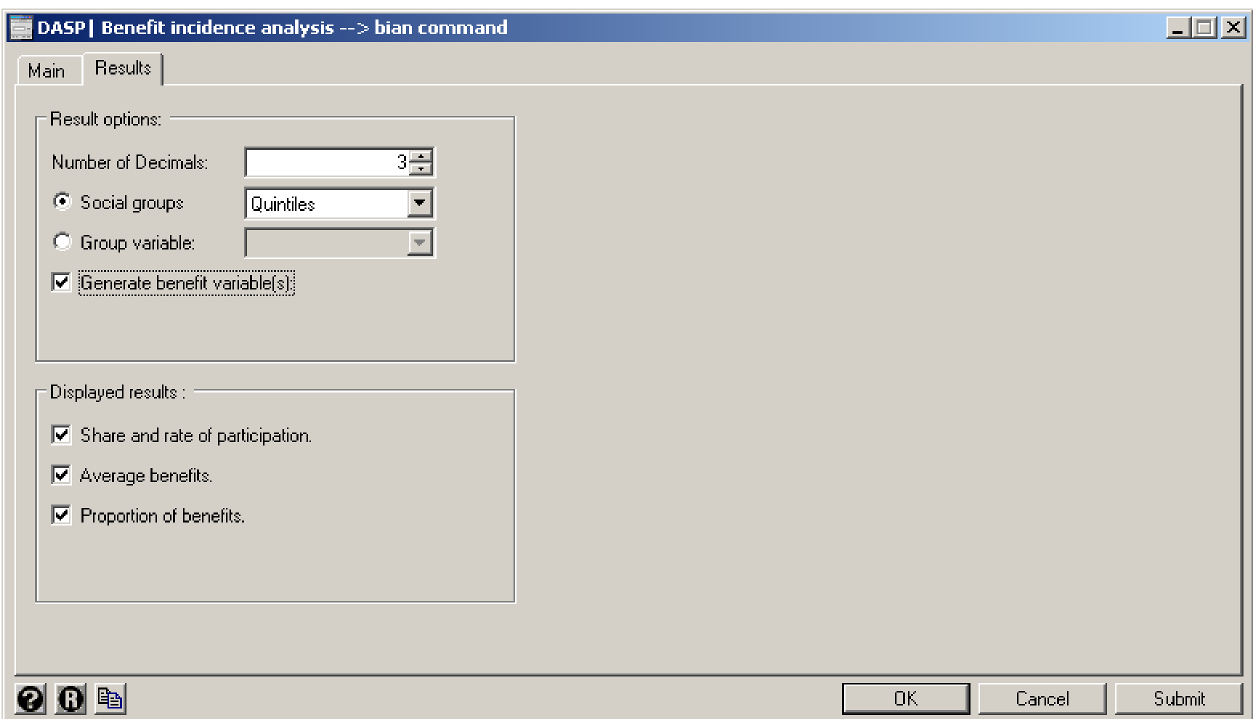
   Estimate the average benefits per quintile and generate the benefit variables.

Answer:

Set variables and options as follows:

Figure 67: Benefit Incidence Analysis (unit cost approach)





After clicking on **Submit,** the following appears:

Average Benefits by Quintile Groups: (at the level of eligible members)

Groups	Primary	Secondary
Quintile 1	276.936	213.366
Quintile 2	293.796	238.235
Quintile 3	280.692	211.999
Quintile 4	248.121	183.199
Quintile 5	171.520	96.228
All	254.201	188.585

Average Benefits by Quintile Groups: (at the level of members that use the public service)

Groups	Primary	Secondary
Quintile 1	363.446	656.563
Quintile 2	363.446	656.563
Quintile 3	363.446	656.563
Quintile 4	363.446	656.563
Quintile 5	363.446	656.563
All	363.446	656.563

Proportion of Benefits by Quintile Groups and by Sectors.

Groups	Primary	Secondary
Quintile 1	0.136	0.085
Quintile 2	0.144	0.095
Quintile 3	0.138	0.085
Quintile 4	0.122	0.073
Quintile 5	0.084	0.038
All	0.624	0.376