USER MANUAL

DASP version 1.4

DASP: Distributive Analysis Stata Package

By

Abdelkrim Araar,
Jean-Yves Duclos

Université Laval
PEP, CIRPÉE and World Bank

December 2007
# Table of contents

Table of contents ........................................................................................................... 2  
List of Figures ................................................................................................................... 4  
1 Introduction .................................................................................................................. 6  
2 DASP and Stata versions .............................................................................................. 6  
3 Installing and updating the DASP package ............................................................... 6  
3.1 Installing DASP modules ......................................................................................... 7  
3.2 Adding the DASP submenu to STATA’s main menu .................................................. 7  
4 DASP and data files ...................................................................................................... 8  
5 Main variables for distributive analysis .................................................................... 8  
6 How can DASP commands be invoked? ................................................................... 9  
7 How can help be accessed for a given DASP module? .............................................. 9  
8 Applications and files in DASP .................................................................................. 10  
9 Basic Notation ............................................................................................................. 11  
10 DASP and poverty indices ......................................................................................... 12  
10.1 FGT and EDE-FGT poverty indices (ifgt) ............................................................... 12  
10.2 Difference between FGT indices (dfgt) .................................................................. 13  
10.3 DASP and multidimensional poverty indices (imd pov) ...................................... 13  
11 Poverty, marginal impacts and elasticities ............................................................... 15  
11.1 FGT-Elasticity with respect to within/between group components of inequality (efgtg) ... 15  
11.2 FGT-Elasticity with respect to within/between income components of inequality (efgtc) ... 16  
12 DASP and inequality indices ..................................................................................... 18  
12.1 Gini and concentration indices (ig ini) ................................................................. 18  
12.2 Difference between Gini/concentration indices (digin i) .................................. 18  
12.3 Generalised entropy index (ientropy) .................................................................... 19  
12.4 Difference between generalised entropy indices (dientropy) ............................. 19  
12.5 Quantile/share ratio indices of inequality (in ineq) .............................................. 20  
12.6 Difference between Quantile/Share indices (din ineq) ....................................... 20  
13 DASP and polarization indices .................................................................................. 21  
13.1 The DER index (ipolar) ......................................................................................... 21  
13.2 Difference between DER polarization indices (dipolar) ..................................... 22  
14 DASP and decompositions ....................................................................................... 22  
14.1 FGT Poverty: decomposition by population subgroups (dfgtg) ................................ 22  
14.2 Decomposition of the variation in FGT indices into growth and redistribution components (dfgtg r). 23  
14.3 Decomposition of the FGT by transient and chronic poverty components (dtc pov) .......................... 24  
14.4 Inequality: decomposition by income sources (digin is) .................................... 26  
14.5 Gini index: decomposition by population subgroups (digin ig) ............................ 27  
14.6 Generalized entropy indices of inequality: decomposition by population subgroups (d en tropyg). 27  
15 DASP and curves ..................................................................................................... 28  
15.1 FGT CURVES (cfg t) ......................................................................................... 28  
15.2 FGT CURVE with confidence interval (cfg ts) ................................................. 30  
15.3 Difference between FGT CURVES with confidence interval (cfg ts2d) ............ 30  
15.4 Lorenz and concentration CURVES (clorenz) ................................................. 30  
15.5 Lorenz/concentration CURVES with confidence intervals (clorenzs) ............. 31  
15.6 Differences between Lorenz/concentration CURVES with confidence interval (clorenzs2d) .......................... 32  
15.7 Poverty curves (cpoverty) .................................................................................... 32  
16 Dominance .................................................................................................................. 33  
16.1 Poverty dominance (dom pov) .......................................................................... 33  
16.2 Inequality dominance (dom ine q) ....................................................................... 33  
16.3 DASP and bi-dimensional poverty dominance (dom b dpov) .............................. 33  
17 Distributive tools ...................................................................................................... 34  
17.1 Quantile curves (c_quantile) .............................................................................. 34
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Output of <code>net describe dasp</code></td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td><code>DASP</code> submenu</td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>Using <code>DASP</code> with a command window</td>
<td>9</td>
</tr>
<tr>
<td>4</td>
<td>Accessing help on <code>DASP</code></td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>Estimating FGT poverty with one distribution</td>
<td>10</td>
</tr>
<tr>
<td>6</td>
<td>Estimating FGT poverty with two distributions</td>
<td>11</td>
</tr>
<tr>
<td>7</td>
<td>Decomposition of the FGT index by groups</td>
<td>23</td>
</tr>
<tr>
<td>8</td>
<td>Decomposition of poverty into transient and chronic components</td>
<td>26</td>
</tr>
<tr>
<td>9</td>
<td>FGT curves</td>
<td>29</td>
</tr>
<tr>
<td>10</td>
<td>Lorenz and concentration curves</td>
<td>31</td>
</tr>
<tr>
<td>11</td>
<td>Survey data settings</td>
<td>47</td>
</tr>
<tr>
<td>12</td>
<td>Setting sampling weights</td>
<td>48</td>
</tr>
<tr>
<td>13</td>
<td>Estimating FGT indices</td>
<td>51</td>
</tr>
<tr>
<td>14</td>
<td>Estimating FGT indices with relative poverty lines</td>
<td>52</td>
</tr>
<tr>
<td>15</td>
<td>FGT indices differentiated by gender</td>
<td>53</td>
</tr>
<tr>
<td>16</td>
<td>Estimating differences between FGT indices</td>
<td>56</td>
</tr>
<tr>
<td>17</td>
<td>Estimating differences in FGT indices</td>
<td>57</td>
</tr>
<tr>
<td>18</td>
<td>FGT differences across years by gender and zone</td>
<td>58</td>
</tr>
<tr>
<td>19</td>
<td>Estimating multidimensional poverty indices (A)</td>
<td>59</td>
</tr>
<tr>
<td>20</td>
<td>Estimating multidimensional poverty indices (B)</td>
<td>61</td>
</tr>
<tr>
<td>21</td>
<td>Drawing FGT curves</td>
<td>63</td>
</tr>
<tr>
<td>22</td>
<td>Editing FGT curves</td>
<td>63</td>
</tr>
<tr>
<td>23</td>
<td>Graph of FGT curves</td>
<td>64</td>
</tr>
<tr>
<td>24</td>
<td>FGT curves by zone</td>
<td>65</td>
</tr>
<tr>
<td>25</td>
<td>Graph of FGT curves by zone</td>
<td>66</td>
</tr>
<tr>
<td>26</td>
<td>Differences of FGT curves</td>
<td>67</td>
</tr>
<tr>
<td>27</td>
<td>Listing coordinates</td>
<td>67</td>
</tr>
<tr>
<td>28</td>
<td>Differences between FGT curves</td>
<td>68</td>
</tr>
<tr>
<td>29</td>
<td>Differences between FGT curves</td>
<td>69</td>
</tr>
<tr>
<td>30</td>
<td>Drawing FGT curves with confidence interval</td>
<td>70</td>
</tr>
<tr>
<td>31</td>
<td>FGT curves with confidence interval</td>
<td>71</td>
</tr>
<tr>
<td>32</td>
<td>Drawing the difference between FGT curves with confidence interval</td>
<td>72</td>
</tr>
<tr>
<td>33</td>
<td>Differences between FGT curves with confidence interval (<code>\(\alpha = 0\)</code>)</td>
<td>72</td>
</tr>
<tr>
<td>34</td>
<td>Difference between FGT curves with confidence interval (<code>\(\alpha = 1\)</code>)</td>
<td>73</td>
</tr>
<tr>
<td>35</td>
<td>Testing for poverty dominance</td>
<td>74</td>
</tr>
<tr>
<td>36</td>
<td>Decomposing FGT indices by groups</td>
<td>75</td>
</tr>
<tr>
<td>37</td>
<td>Lorenz and concentration curves</td>
<td>78</td>
</tr>
<tr>
<td>38</td>
<td>Lorenz curves</td>
<td>78</td>
</tr>
<tr>
<td>39</td>
<td>Drawing concentration curves</td>
<td>79</td>
</tr>
<tr>
<td>40</td>
<td>Lorenz and concentration curves</td>
<td>80</td>
</tr>
<tr>
<td>41</td>
<td>Drawing Lorenz curves</td>
<td>81</td>
</tr>
<tr>
<td>42</td>
<td>Lorenz curves</td>
<td>81</td>
</tr>
<tr>
<td>43</td>
<td>Estimating Gini and concentration indices</td>
<td>82</td>
</tr>
<tr>
<td>44</td>
<td>Estimating concentration indices</td>
<td>83</td>
</tr>
<tr>
<td>45</td>
<td>Estimating differences in Gini and concentration indices</td>
<td>84</td>
</tr>
<tr>
<td>46</td>
<td>Drawing densities</td>
<td>85</td>
</tr>
<tr>
<td>47</td>
<td>Density curves</td>
<td>86</td>
</tr>
<tr>
<td>48</td>
<td>Drawing quantile curves</td>
<td>87</td>
</tr>
<tr>
<td>49</td>
<td>Quantile curves</td>
<td>87</td>
</tr>
<tr>
<td>50</td>
<td>Drawing non-parametric regression curves</td>
<td>88</td>
</tr>
<tr>
<td>51</td>
<td>Non-parametric regression curves</td>
<td>89</td>
</tr>
<tr>
<td>52</td>
<td>Drawing derivatives of non-parametric regression curves</td>
<td>90</td>
</tr>
<tr>
<td>53</td>
<td>Derivatives of non-parametric regression curves</td>
<td>90</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>------------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>54</td>
<td>Plotting joint density function</td>
<td>91</td>
</tr>
<tr>
<td>55</td>
<td>Plotting joint distribution function</td>
<td>93</td>
</tr>
<tr>
<td>56</td>
<td>Testing for bi-dimensional poverty dominance</td>
<td>94</td>
</tr>
<tr>
<td>57</td>
<td>Testing the pro-poor growth (primal approach)</td>
<td>98</td>
</tr>
<tr>
<td>58</td>
<td>Testing the pro-poor growth (dual approach)- A</td>
<td>99</td>
</tr>
<tr>
<td>59</td>
<td>Testing the pro-poor growth (dual approach) – B</td>
<td>100</td>
</tr>
<tr>
<td>60</td>
<td>Benefit incidence analysis</td>
<td>103</td>
</tr>
<tr>
<td>61</td>
<td>Benefit Incidence Analysis (unit cost approach)</td>
<td>105</td>
</tr>
</tbody>
</table>
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 9.2 or higher
- ado files must be updated

To update the executable file (from 9.0 to 9.2) 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:

<table>
<thead>
<tr>
<th>Priority</th>
<th>Directory</th>
<th>Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>UPDATES:</td>
<td>Official updates of STATA *.ado files</td>
</tr>
<tr>
<td>2</td>
<td>BASE:</td>
<td>*.ado files that come with the installed STATA software</td>
</tr>
<tr>
<td>3</td>
<td>SITE:</td>
<td>*.ado files downloaded from the net</td>
</tr>
<tr>
<td>4</td>
<td>PLUS:</td>
<td>..</td>
</tr>
<tr>
<td>5</td>
<td>PERSONAL:</td>
<td>Personal *.ado files</td>
</tr>
</tbody>
</table>
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

d. `net from c:/dasp`

Figure 1: Ouput of `net describe dasp`

```
Version : Version 1.4
Date     : December 2007
Stata Version : Required 9.2 and higher

Author:
DASP is conceived and programmed by:
Dr. Abdelkrim Araar : aabd@ecn.ulaval.ca

Co-author:
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:

update the ado files:
http://www.stata.com/support/updates/stata9/ado/

The following 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

```
e. Type the syntax
`net install dasp_p1.pkg`, force replace
`net install dasp_p2.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`. 
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

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

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

![Figure 6: Estimating FGT poverty with two distributions](image)

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.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Indication</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y$</td>
<td>variable of interest</td>
</tr>
<tr>
<td>$i$</td>
<td>observation number</td>
</tr>
<tr>
<td>$y_i$</td>
<td>value of the variable of interest for observation $i$</td>
</tr>
<tr>
<td>$hw$</td>
<td>sampling weight</td>
</tr>
<tr>
<td>$hw_i$</td>
<td>sampling weight for observation $i$</td>
</tr>
<tr>
<td>$hs$</td>
<td>size variable</td>
</tr>
<tr>
<td>$hs_i$</td>
<td>size of observation $i$ (for example the size of household $i$)</td>
</tr>
<tr>
<td>$w_i$</td>
<td>$hw_i$ * $hs_i$</td>
</tr>
<tr>
<td>$hg$</td>
<td>group variable</td>
</tr>
<tr>
<td>$hg_i$</td>
<td>group of observation $i$.</td>
</tr>
<tr>
<td>$w_i^*$</td>
<td>$sw_i^*$ if $hg_i = k$, and 0 otherwise.</td>
</tr>
</tbody>
</table>
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

\[
\tilde{P}(z; \alpha) = \frac{\hat{P}(z; \alpha)}{(z)^{\alpha}}
\]

The EDE-FGT index is estimated as:

\[
EDE(P(z; \alpha)) = \left( \frac{\tilde{P}(z; \alpha)}{\tilde{P}(z; \alpha)} \right)^{1/\alpha} \quad \text{for} \quad \alpha > 0
\]

- 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 \( \alpha \).

- 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 21.1
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 $\alpha$ can be chosen for each of the two distributions.

Interested users are encouraged to consider the exercises that appear in Section 21.2.

10.3 DASP and multidimensional poverty indices (imd pov)

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 $X_i = (x_{i,1}, ..., x_{i,J})$ and vector of poverty lines $Z = (z_1, ..., z_J)$), determining $i$’s contribution to total poverty $P(X, Z)$.


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

[2] Extended Watts index
\[ p(X_i, Z) = \sum_{j=1}^{J} a_j \ln \left( \frac{z_j}{\min(z_j; x_{i,j})} \right) \]

[3] Multiplicative extended FGT index

\[ p(X_i, Z) = \prod_{j=1}^{J} \left( \frac{z_j - x_{i,j}}{z_j} \right)^{a_j} \]


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

[5] Intersection headcount index

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

[6] Union headcount index

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


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

where:

\[ C_1 = \left( \frac{z_1 - x_{i,1}}{z_1} \right)^{\gamma} \quad \text{and} \quad C_2 = \left( \frac{z_2 - x_{i,2}}{z_2} \right)^{\gamma} \]

imd pov estimates the above 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 6).
- 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 21.3

11 Poverty, marginal impacts and elasticities

11.1 FGT-Elasticity with respect to within/between group components of inequality (efgtg).

This module estimates the FGT marginal impact and 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):


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

After clicking on SUBMIT, the following should be displayed:
11.2 *FGT-Elasticity with respect to within/between income components of inequality (efgtc).*

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


To open the dialog box of this module, type the command `db efgtc`. 
After clicking on SUBMIT, the following should be displayed:

\[ \eta(k), \lambda, \tau \]

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

---

In case one is interested in changing income-component only among individuals that are effectively active in some economic sectors (schemes \( \eta^* (k), \tau^* \) and \( \lambda^* \) in the paper of reference), the user should select the approach "Truncated income component".
12 DASP and inequality indices

12.1 Gini and concentration indices (igini)

The Gini index is estimated as

\[ \hat{I} = 1 - \hat{\xi} \]

where

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

and

\[ V_i = \sum_{h=i}^{n} w_h \quad \text{and} \quad y_1 \geq y_2 \geq \cdots \geq y_{n-1} \geq y_n. \]

The concentration index for the variable \( T \) when the ranking variable is \( Y \) is estimated as

\[ \hat{IC}_T = 1 - \frac{\hat{\xi}_T}{\hat{\mu}_T} \]

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

\[ V_i = \sum_{h=i}^{n} w_h \quad \text{and} \quad y_1 \geq y_2 \geq \cdots \geq y_{n-1} \geq y_n. \]

- 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 21.9

12.2 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.

### 12.3 Generalised entropy index (ientropy)

The generalized entropy index is estimated as

$$
\hat{I}(\theta) = \begin{cases} 
\frac{1}{\theta(\theta-1)} \sum_{i=1}^{n} w_i \left( \frac{y_i}{\hat{\mu}} \right)^\theta - 1 & \text{if } \theta \neq 0,1 \\
\frac{1}{n} \sum_{i=1}^{n} w_i \log \left( \frac{\hat{\mu}}{y_i} \right) & \text{if } \theta = 0 \\
\frac{1}{n} \sum_{i=1}^{n} w_i \frac{y_i}{\hat{\mu}} \log \left( \frac{\hat{\mu}}{y_i} \right) & \text{if } \theta = 1 
\end{cases}
$$

- 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.

### 12.4 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.

12.5 Quantile/share ratio indices of inequality (ineiq)

The quantile ratio is estimated as

\[ \hat{QR}(p_1, p_2) = \frac{\hat{Q}(p_1)}{\hat{Q}(p_2)} \]

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

The share ratio is estimated as

\[ \hat{SR}(p_1, p_2, p_3, p_4) = \frac{\hat{GL}(p_2)-\hat{GL}(p_1)}{\hat{GL}(p_4)-\hat{GL}(p_3)} \]

where \( GL(p) \) 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.

12.6 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 **DASP and polarization indices**

### 13.1 The DER index (ipolar)

The Duclos, Esteban and Ray (2004) (DER) polarization index is estimated as Denote the Duclos, Esteban and Ray (DER) index of polarisation for the group k by $\text{DER}(k, \alpha)$. It can be expressed as follows:

$$\text{DER}(\alpha) = \int \int f(x)^{1+\alpha} f(y) |y - x| dy dx$$

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

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

The normalised DER, that the module estimates, is defined as follows:

$$\overline{\text{DER}}(\alpha) = \frac{\text{DER}(\alpha)}{2\mu(1-\alpha)} \in [0,1]$$

Where:

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

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 inequality, for instance 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.
13.2 *Difference between DER polarization indices (dipolar)*

This module estimates differences in DER 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.

14 *DASP and decompositions.*

14.1 *FGT Poverty: decomposition by population subgroups (dfgtg)*

The dfgt 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: \( \left( \frac{\hat{\phi}(g) \hat{P}(z; \alpha; g)}{\hat{P}(z; \alpha)} \right) \)

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

To open the dialog box for module dfgt, type db dfgt in the *command window.*
14.2 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, \( t_1 \) and \( t_2 \), into growth and redistribution components as follows:

\[
\frac{P_2 - P_1}{\text{variation}} = \frac{P(\mu^{t_2}, \pi^{t_1}) - P(\mu^{t_1}, \pi^{t_1})}{C_1} + \frac{P(\mu^{t_1}, \pi^{t_2}) - P(\mu^{t_1}, \pi^{t_1})}{C_2} + R \quad / \text{ref = 1}
\]

\[
\frac{P_2 - P_1}{\text{variation}} = \frac{P(\mu^{t_2}, \pi^{t_2}) - P(\mu^{t_1}, \pi^{t_2})}{C_1} + \frac{P(\mu^{t_2}, \pi^{t_2}) - P(\mu^{t_2}, \pi^{t_1})}{C_2} + R \quad / \text{ref = 2}
\]

where

\begin{align*}
\text{variation} & = \text{difference in poverty between } t_1 \text{ and } t_2; \\
C_1 & = \text{growth component;} \\
C_2 & = \text{redistribution component;} \\
R & = \text{residual;} \\
\text{Ref} & = \text{period of reference.}
\end{align*}
$P(\mu^{t_1}, \pi^{t_1})$ : the FGT index of the first period

$P(\mu^{t_2}, \pi^{t_2})$ : the FGT index of the second period

$P(\mu^{t_2}, \pi^{t_1})$ : the FGT index of the first period when all incomes $y_i^{t_1}$ of the first period are multiplied by $\mu^{t_2} / \mu^{t_1}$

$P(\mu^{t_1}, \pi^{t_2})$ : the FGT index of the second period when all incomes $y_i^{t_2}$ of the second period are multiplied by $\mu^{t_1} / \mu^{t_2}$

The Shapley value decomposes the variation in the FGT Index between two periods, $t_1$ and $t_2$, into growth and redistribution components as follows:

$$\frac{P_{t_2} - P_{t_1}}{\text{Variation}} = C_1 + C_2$$

$$C_1 = \frac{1}{2} \left[ P(\mu^{t_2}, \pi^{t_1}) - P(\mu^{t_1}, \pi^{t_1}) \right] + \left[ P(\mu^{t_2}, \pi^{t_2}) - P(\mu^{t_1}, \pi^{t_2}) \right]$$

$$C_2 = \frac{1}{2} \left[ P(\mu^{t_1}, \pi^{t_2}) - P(\mu^{t_1}, \pi^{t_1}) \right] + \left[ P(\mu^{t_2}, \pi^{t_1}) - P(\mu^{t_2}, \pi^{t_1}) \right]$$

14.3 Decomposition of the FGT by transient and chronic poverty components (dtc pov). This type of decomposition decomposes total poverty, observed over some time periods, into transient and chronic components.

The Jalan and Ravallion (1998) approach

Let $y_i^{t_1}$ be the income of household $i$ in period $t$, and $\mu_i$ be the average income over the $T$ periods for household $i$. Total poverty is defined as follows:

$$TP(\alpha, z) = \frac{\sum_{t=1}^{T} \sum_{i=1}^{N} w_i (z - y_i^{t_1})^\alpha_+}{\sum_{i=1}^{T} \sum_{i=1}^{N} w_i}$$

The chronic poverty component is then defined as:
The transient poverty component is finally defined as:

\[ TPC(\alpha, z) = TP(\alpha, z) - CPC(\alpha, z) \]

**Duclos, Araar and Giles (2006) approach**

Let \( y_t^i \) be the income of household \( i \) in period \( t \) and \( \mu_i \) be the average income over the \( T \) periods for household \( i \). Let \( \Gamma(\alpha, z) \) be the "equally-distributed-equivalent" (EDE) poverty gap such as:

\[ \Gamma(\alpha, z) = \left[ TP(\alpha, z) \right]^{1/\alpha} = \frac{\sum_{t=1}^{T} \sum_{i=1}^{N} w_i (z - y_t^i)^\alpha_+}{T \sum_{i=1}^{N} w_i} \]

The transient poverty component is defined as follows:

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

where \( \theta_i = \gamma_i(\alpha, z) - \gamma_i(1, z) \) and \( \gamma_i(\alpha, z) = \left( \frac{\sum_{t=1}^{T} (z - y_t^i)^\alpha_+}{T} \right)^{1/\alpha} \)

The chronic poverty component is defined as follows:

\[ CPC(\alpha, z) = \Gamma(\alpha, z) - TPC(\alpha, z) \]

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 8: Decomposition of poverty into transient and chronic components**

- The user can select more than one variable of interest simultaneously, where each variable represents the income – standard of living – for one period.
- The user can select one of the two presented approaches above.
- Bias-correction can be done, using either an analytical/asymptotic or 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.

**Main references**


### 14.4 Inequality: decomposition by income sources (diginis)

The diginis module decomposes the Gini index 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)


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

The diginig module decomposes the Gini index 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:

\[
I = \sum_{g=1}^{G} \phi_g \phi g + \overline{I}_{\text{Within}} + R_{\text{Overlap}}
\]

where
- \( \phi_g \): the population share of group g;
- \( \phi g \): the income share of group g.
- \( \overline{I} \): the between group inequality (when each individual has the average income of its group).
- \( R \): The residue implied by group income overlap


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

\[
\hat{I}(\theta) = \sum_{k=1}^{K} \hat{\phi}(k) \left( \frac{\hat{\mu}(k)}{\hat{\mu}} \right)^\theta \hat{I}(k; \theta) + \hat{I}(\theta)
\]

where:
- \( \phi(k) \) is the proportion of the population found in subgroup k.
- \( \mu(k) \) is the mean income of group k.
- \( I(k; \theta) \) is inequality within group k.
- \( I(\theta) \) is population inequality if each individual in subgroup k is given the mean income of subgroup k, \( \mu(k) \).
15 DASP and curves.

15.1 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**.
Interested users are encouraged to consider the exercises that appear in Section 21.4.
15.2 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, lowerbounded 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 21.5.

15.3 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, lowerbounded 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 21.5.

15.4 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:
  o *.gph : STATA format;
  o *.wmf : typically recommended to insert graphs in Word documents;
  o *.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 10: Lorenz and concentration curves**

Interested users are encouraged to consider the exercises that appear in Section 21.8.

### 15.5 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:
  o *.gph : STATA format;
  o *.wmf : typically recommended to insert graphs in Word documents;
  o *.eps : typically recommended to insert graphs in Tex/Latex documents.
• Many graphical options are available to change the appearance of the graphs.

15.6 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:
  o *.gph : STATA format;
  o *.wmf : typically recommended to insert graphs in Word documents;
  o *.eps : typically recommended to insert graphs in Tex/Latex documents.
• Many graphical options are available to change the appearance of the graphs.

15.7 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:

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

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:
  o *.gph : STATA format;
  o *.wmf : typically recommended to insert graphs in Word documents;
  o *.eps : typically recommended to insert graphs in Tex/Latex documents.
• Many graphical options are available to change the appearance of the graphs.
16 Dominance

16.1 Poverty dominance (dompov)

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

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

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 $\zeta$ and $P_1(\zeta; \alpha)$ for which $P_1(\zeta; \alpha) = P_2(\zeta; \alpha)$ when:

$$\text{sign}(P_1(\zeta-\eta; \alpha) - P_2(\zeta-\eta; \alpha)) = \text{sign}(P_2(\zeta+\eta; \alpha) - P_1(\zeta+\eta; \alpha))$$

for a small $\eta$. The crossing points $\zeta$ 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):


Interested users are encouraged to consider the exercises that appear in Section 21.6.

16.2 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):


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

16.3 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
\[ \hat{P}_D(Z; A) = \sum_{i=1}^{n} \left[ \prod_{k=1}^{2} \left( z^k_i - y^k_i \right)^{\alpha_k} \right] \]

where \( Z = (z_1, z_2) \) and \( A = (\alpha_1, \alpha_2) \) are vectors of poverty lines and parameters \( \alpha \) respectively, and \( x_i \) = max(x, 0).

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

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

The DASP `domdpov` 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 21.12.

## 17 Distributive tools

### 17.1 Quantile curves (c_quantile)

The quantile at a percentile \( p \) of a continuous population is given by:

\[ Q(p) = F^{-1}(p) \quad \text{where} \quad p = F(y) \quad \text{is the cumulative distribution function at} \ y. \]

For a discrete distribution, let \( n \) observations of living standards be ordered such that \( y_1 \leq y_2 \leq \cdots \leq y_i \leq y_{i+1} \leq \cdots \leq y_n \). If \( F(y_i) < p \leq F(y_{i+1}) \), 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 21.10.

### 17.2 Density curves (cdensity)

The Gaussian kernel estimator of a density function \( f(x) \) is defined by
\[
\hat{f}(x) = \frac{\sum_{i=1}^{n} w_i K_i(x)}{\sum_{i=1}^{n} w_i} \quad \text{and} \quad K_i(x) = \frac{1}{h\sqrt{2\pi}} \exp\left(-\frac{0.5}{\lambda_i(x)^2}\right) \quad \text{and} \quad \lambda_i(x) = \frac{x-x_i}{h}
\]

where h is a bandwidth that acts as a “smoothing” parameter.

Interested users are encouraged to consider the exercises that appear in Section 21.10.

### 17.3 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:

#### 17.3.1 Nadaraya-Watson approach

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

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

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

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

#### 17.3.2 Local linear approach

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

\[
K_i(x)^{\frac{1}{2}} y_i = \mu(x)K_i(x)^{\frac{1}{2}} + \mu'(x)K_i(x)^{\frac{1}{2}}(x_i - x) + v
\]

or, alternatively, of:

\[
K_i(x)^{\frac{1}{2}} y_i = \alpha K_i(x)^{\frac{1}{2}} + \beta K_i(x)^{\frac{1}{2}}(x_i - x) + v_i
\]

Estimates are then given by:

\[
E\left( y \mid x \right) = \alpha, \quad E\left( \frac{dy}{dx} \mid x \right) = \beta
\]

Interested users are encouraged to consider the exercises that appear in Section 21.10.
17.4 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:

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

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;
- coordinates can be saved in Stata or GnuPlot-ASCII format.

Interested users are encouraged to consider the exercises that appear in Section 21.11.

17.5 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:

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

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 21.11.

18 DASP and pro-poor growth
18.1 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):**

   \[
   \text{Index} = \frac{W_1(z) - W_2(z)}{F_1(z)}
   \]

   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):**

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

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

   \[
   \text{Index}_1 = \frac{P_1(z,\alpha) - P_2(z\alpha)}{P_1(z,\alpha) - P_1(z(\mu_1/\mu_2),\alpha)}
   \]

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

   \[
   \text{Index}_2 = \text{Index}_1 - g
   \]

- 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 \( \alpha \) can be chosen for each of the two distributions.

18.2 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.

18.2.1 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:

$$\Delta(z, s) = (P_{2}(z + cons, \alpha = s - 1) - P_{1}(z, \alpha = s - 1)) < 0 \forall z \in [0, z^+]$$

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

$$\Delta(z, s) = z \left( P_{2}(z \frac{\mu_{2}}{\mu_{1}}, \alpha = s - 1) - P_{1}(z, \alpha = s - 1) \right) < 0 \forall z \in [0, z^+]$$

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 21.13.

### 18.2.2 Dual pro-poor curves

Let:

- $Q(p)$ : quantile at percentile $p$.
- $GL(p)$ : Generalised Lorenz curve at percentile $p$.
- $\mu$ : 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:

$$\Delta(z, s) = Q_{2}(p) - Q_{1}(p) > 0 \forall p \in [0, p^+ = F(z^+)]$$

or equivalently if:

$$\Delta(z, s) = \frac{Q_{2}(p) - Q_{1}(p)}{Q_{1}(p)} > 0 \forall p \in [0, p^+ = F(z^+)]$$

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

$$\Delta(z, s) = \frac{Q_{2}(p)}{Q_{1}(p)} \cdot \frac{\mu_{2}}{\mu_{1}} > 0 \forall p \in [0, p^+ = F(z^+)]$$
The change in the distribution from state 1 to state 2 is second-order absolutely pro-poor if:

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

or equivalently if:

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

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

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

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 21.13

19  
**DASP and Benefit Incidence Analysis**

19.1 **Benefit incidence analysis**

The main objective of using a benefit incidence approach 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 the access of household members to public services. This information can be found in the 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

$$w_i$$  
be the sampling weight of observation $$i$$;
be the living standard of members belonging to observation \( i \) (i.e., \textit{per capita} income);

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;

\( f_i^s \)
be the number of members of observation \( i \) that effectively use the public service provided by sector \( s \);

\( g_i \)
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 group 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 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:

\[
ABE_g^s = \frac{\sum_{i=1}^{n} w_i B_i^s I(i \in g)}{\sum_{i=1}^{n} w_i c_i^s I(i \in g)}
\]

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 \texttt{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 (e.g., 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.

Generally, public expenditures on a given service can vary from one geographical or administrative area to another. When the 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.

**Example 1**

<table>
<thead>
<tr>
<th>Observation</th>
<th>HH size</th>
<th>Eligible HH members</th>
<th>Frequency</th>
<th>Area indicator</th>
<th>Total level of regional public expenditures</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>14000</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>14000</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>5</td>
<td>3</td>
<td>1</td>
<td>14000</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>12000</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>12000</td>
</tr>
</tbody>
</table>

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.

**Example 2**

<table>
<thead>
<tr>
<th>Observation</th>
<th>HH size</th>
<th>Eligible HH members</th>
<th>Frequency</th>
<th>Area indicator</th>
<th>Regional public expenditures</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>28000</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>28000</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>5</td>
<td>3</td>
<td>1</td>
<td>28000</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>28000</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>28000</td>
</tr>
</tbody>
</table>

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

Interested users are encouraged to consider the exercises that appear in Section 21.14.
20 Appendices

20.1 Appendix A: illustrative household surveys

20.1.1 The 1994 Burkina Faso survey of household expenditures (bkf94l.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
20.1.2 The 1998 Burkina Faso survey of household expenditures (bkf98l.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 \( ipc_{1994}/ipc_{1998} \)

20.1.3 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 \( B \))

20.1.4 Peru LSMS survey 1994 (A sample of 3623 household observations - PEREDE94l.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

---

### 20.1.5  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

---

### 20.1.6  The 1995 Colombia DHS survey (columbian.dta)

This sample is a part of the [Data from the Demographic and Health Surveys](http://example.com) (Colombia_1995) which 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

---

### 20.1.7  The 1996 Dominican Republic DHS survey (Dominican_republic1996l.dta)

This sample is a part of the [Data from the Demographic and Health Surveys](http://example.com) (Republic Dominican_1996) which contains the following information for children aged 0-59 months

List of variables

- **hid**: Household id
- **haz**: height-for-age
- **waz**: weight-for-age
20.2 Appendix B: labelling variables and values

- The following .do file can be used to set labels for the variables in `b kf94.dta`.
- For more details on the use of `label` command, type `help label` in the command window.

```bash
# 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 lgse
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 "lgse" to the variable gse */
label val gse lgse;
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;

20.3 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:

![Survey data settings](image.png)

Figure 11: Survey data settings

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

Note: Empty or ",_n" in "Sampling units" above indicates sampling of observations.
Click on OK and save the data file.

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

\begin{verbatim}
Survey: Describing stage 1 sampling units

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

\hline
Stratum & #Units & #Us & min & mean & max
\hline
 1   & 42  & 838 & 19 & 21.0 & 20
 2   & 37  & 733 & 17 & 19.8 & 20
 3   & 98  & 1959 & 19 & 21.0 & 20
 4   & 55  & 1893 & 19 & 19.3 & 20
 5   & 66  & 1206 & 13 & 19.5 & 20
 6   & 41  & 779 & 1 & 19.0 & 20
 7   & 97  & 1957 & 19 & 20.0 & 20
\hline
 7   & 436 & 8625 & 1 & 19.8 & 20
\end{verbatim}
21 Examples and exercises

21.1 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 20.1.1. Type the *describe* command and then *label list* to list labels.

2. Use the information of Section 20.1.1. 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:

```
zeone:
1  Rural
2  Urban

sex:
1  Male
2  Female
```

49
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 20.3, or simply by typing

```
svys set psu [pweight=weight], strata(strata) vce(linearized)
```

Typing `svydes`, we obtain

```
Survey: Describing stage 1 sampling units

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

<table>
<thead>
<tr>
<th>Stratum</th>
<th>#Units</th>
<th>#Obs</th>
<th>min</th>
<th>mean</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>42</td>
<td>838</td>
<td>19</td>
<td>20.0</td>
<td>20</td>
</tr>
<tr>
<td>2</td>
<td>37</td>
<td>733</td>
<td>17</td>
<td>19.8</td>
<td>20</td>
</tr>
<tr>
<td>3</td>
<td>98</td>
<td>1959</td>
<td>19</td>
<td>20.0</td>
<td>20</td>
</tr>
<tr>
<td>4</td>
<td>55</td>
<td>1093</td>
<td>19</td>
<td>19.9</td>
<td>20</td>
</tr>
<tr>
<td>5</td>
<td>66</td>
<td>1286</td>
<td>13</td>
<td>19.5</td>
<td>20</td>
</tr>
<tr>
<td>6</td>
<td>41</td>
<td>779</td>
<td>1</td>
<td>19.0</td>
<td>20</td>
</tr>
<tr>
<td>7</td>
<td>97</td>
<td>1937</td>
<td>19</td>
<td>20.0</td>
<td>20</td>
</tr>
</tbody>
</table>

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.
The following results should then be displayed:

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

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>STD</th>
<th>LB</th>
<th>UB</th>
<th>P. Line</th>
</tr>
</thead>
<tbody>
<tr>
<td>epopc</td>
<td>-0.44565</td>
<td>0.016124</td>
<td>-0.412973</td>
<td>0.472956</td>
<td>41099.00</td>
</tr>
<tr>
<td>expeq</td>
<td>0.255400</td>
<td>0.013326</td>
<td>0.229908</td>
<td>0.280592</td>
<td>41099.00</td>
</tr>
</tbody>
</table>

**Q.4**

Select RELATIVE for the poverty line and set the other parameters as above.
Figure 14: Estimating FGT indices with relative poverty lines

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

```
. fgt expcc, alpha(0) hsize(size) opl(median) prop(60)

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

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>STD</th>
<th>LB</th>
<th>UB</th>
<th>P. Line</th>
</tr>
</thead>
<tbody>
<tr>
<td>expcc</td>
<td>0.195243</td>
<td>0.008576</td>
<td>0.168386</td>
<td>0.202099</td>
<td>27046.71</td>
</tr>
</tbody>
</table>
```

**Q.5**

Set the group variable to sex.
Clicking on SUBMIT, the following should appear:

```
*.fgt wppc, alpha(0) hs(size) hgroup(sex) pline(41099)
```

<table>
<thead>
<tr>
<th>Group</th>
<th>Estimate</th>
<th>STD</th>
<th>LB</th>
<th>UB</th>
<th>P. Line</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>0.42176</td>
<td>0.016633</td>
<td>0.41944</td>
<td>0.424067</td>
<td>41099.00</td>
</tr>
<tr>
<td>Female</td>
<td>0.280850</td>
<td>0.028206</td>
<td>0.27641</td>
<td>0.285290</td>
<td>41099.00</td>
</tr>
<tr>
<td>POPULATION</td>
<td>0.444565</td>
<td>0.016124</td>
<td>0.442073</td>
<td>0.447056</td>
<td>41099.00</td>
</tr>
</tbody>
</table>

**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 evppc, alpha(0) hsize(size) hgroup(sex) dec(4) level(99) pline(41099)

<table>
<thead>
<tr>
<th>Group</th>
<th>Estimate</th>
<th>STD</th>
<th>LB</th>
<th>UB</th>
<th>P. Line</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: Male</td>
<td>0.4522</td>
<td>0.0166</td>
<td>0.4091</td>
<td>0.4952</td>
<td>41099.00</td>
</tr>
<tr>
<td>2: Female</td>
<td>0.2819</td>
<td>0.0282</td>
<td>0.2309</td>
<td>0.3349</td>
<td>41099.00</td>
</tr>
<tr>
<td>POPULATION</td>
<td>0.4446</td>
<td>0.0161</td>
<td>0.4028</td>
<td>0.4863</td>
<td>41099.00</td>
</tr>
</tbody>
</table>
**21.2 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:
   
   \[ H_0 : \Delta P(z = 41099, \alpha = 0) = 0 \quad \text{against} \quad H_1 : \Delta P(z = 41099, \alpha = 0) \neq 0 \]

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

```
.digt expcpc expcpc, alpha(0) file2(C:\DATA\bail981.dta) hsize2(size) file2(C:\DATA\bail941.dta) hsize2(size) pline1(4109) pline2(4109)
```

<table>
<thead>
<tr>
<th>Poverty Index</th>
<th>FGT Index</th>
<th>Parameter alpha</th>
<th>Estimate</th>
<th>STD</th>
<th>LB</th>
<th>UB</th>
<th>P. Line</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distribution_1</td>
<td>0.452677</td>
<td>0.010927</td>
<td>0.431199</td>
<td>0.474156</td>
<td>4109.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distribution_2</td>
<td>0.444565</td>
<td>0.016124</td>
<td>0.412873</td>
<td>0.478256</td>
<td>4109.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Difference</td>
<td>0.008113</td>
<td>0.019477</td>
<td>-0.030062</td>
<td>0.046288</td>
<td>---</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Q.3

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

Figure 17: Estimating differences in FGT indices

After clicking on SUBMIT, we should see:

<table>
<thead>
<tr>
<th>Poverty Index : FGT Index</th>
<th>Parameter alpha : 0.00</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
</tr>
<tr>
<td>Distribution_1</td>
<td>0.500344</td>
</tr>
<tr>
<td>Distribution_2</td>
<td>0.505977</td>
</tr>
<tr>
<td>Difference</td>
<td>-0.000153</td>
</tr>
</tbody>
</table>

Q.4

<table>
<thead>
<tr>
<th>Poverty Index : FGT Index</th>
<th>Parameter alpha : 0.00</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
</tr>
<tr>
<td>Distribution_1</td>
<td>0.164573</td>
</tr>
<tr>
<td>Distribution_2</td>
<td>0.103864</td>
</tr>
<tr>
<td>Difference</td>
<td>0.060809</td>
</tr>
</tbody>
</table>
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 18: FGT differences across years by gender and zone

![Graph showing FGT differences across years by gender and zone]

After clicking on SUBMIT, the following should be displayed:

<table>
<thead>
<tr>
<th>Poverty Index</th>
<th>FGT Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter alpha</td>
<td>0.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>STD</th>
<th>LB</th>
<th>UB</th>
<th>P. Line</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distribution 1</td>
<td>(0.172304)</td>
<td>(0.017701)</td>
<td>(0.137591)</td>
<td>(0.207279)</td>
<td>41099.00</td>
</tr>
<tr>
<td>Distribution 2</td>
<td>(0.105997)</td>
<td>(0.013945)</td>
<td>(0.078531)</td>
<td>(0.132405)</td>
<td>41099.00</td>
</tr>
<tr>
<td>Difference</td>
<td>(0.066307)</td>
<td>(0.022534)</td>
<td>(0.022222)</td>
<td>(0.110553)</td>
<td>---</td>
</tr>
</tbody>
</table>

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.
21.3 Estimating multidimensional poverty indices

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

Using the peru94l.dta file,

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

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Var. of interest</th>
<th>Pov. line</th>
<th>a_j</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>exppc</td>
<td>400</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>literate</td>
<td>0.90</td>
<td>1</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Var. of interest</th>
<th>Pov. line</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>exppc</td>
<td>400</td>
</tr>
<tr>
<td>2</td>
<td>literate</td>
<td>0.90</td>
</tr>
</tbody>
</table>

Q.1

Steps:
- Type
  `use "C:\data\peru94l.dta", clear`
- To open the relevant dialog box, type `db imdpov`
- Choose variables and parameters as in

Figure 19: Estimating multidimensional poverty indices (A)
After clicking SUBMIT, the following results appear.

```
indpov expcc pliterate, hsize(size) index(1) alpha(0) a1(1) p1(1.0) a2(1) p2(0.9)
```

Household size: size

<table>
<thead>
<tr>
<th>Population</th>
<th>Estimate</th>
<th>STD</th>
<th>LB</th>
<th>UB</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.418</td>
<td>0.009</td>
<td>0.403</td>
<td>0.433</td>
</tr>
</tbody>
</table>

Q.2

Steps:

- Choose variables and parameters as in
Figure 20: Estimating multidimensional poverty indices (B)

After clicking SUBMIT, the following results appear.

```
. indpov exppc pliterate, hsize(size) index(7) alpha(1) beta(1) gamma(1) p11(400) p12(0.9)

  Household size     : size

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>STD</th>
<th>LB</th>
<th>UB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population</td>
<td>0.098</td>
<td>0.003</td>
<td>0.093</td>
<td>0.103</td>
</tr>
</tbody>
</table>
```
21.4 Estimating FGT curves.

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

1. Open `bkf94l.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\bkf94l.dta", clear`

Q.2

Open the dialog box by typing

`db difgt`

Q.3

Choose variables and parameters as follows:
Figure 21: Drawing FGT curves

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

Figure 22: Editing FGT curves

After clicking SUBMIT, the following graph appears:
Figure 23: Graph of FGT curves

FGT Curves (alpha=0)
Burkina 1994

- FGT(z, alpha = 0)
- Poverty line (z)

Graph (Graph)
Q.4

Choose variables and parameters as in the following window:

**Figure 24: FGT curves by zone**

After clicking SUBMIT, the following graph appears:
Figure 25: Graph of FGT curves by zone

FGT Curves (alpha=0)
Burkina 1994

Poverty line (z)

FGT(z, alpha = 0)

Rural  Urban
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 26: Differences of FGT curves

Figure 27: Listing coordinates
After clicking SUBMIT, the following appears:

**Figure 28: Differences between FGT curves**

![Graph](image)

**Difference Between FGT Curves (alpha=0)**

Burkina 1994

---

**Q.6**
21.5 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:
   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\bkf941.dta", clear

- To open the relevant dialog box, type 
  db cfgts

- Choose variables and parameters as in

**Figure 30: Drawing FGT curves with confidence interval**

![FGT Curves Dialog Box](image-url)

After clicking SUBMIT, the following appears:
Figure 31: FGT curves with confidence interval

FGT curve (alpha = 0)
Burkina Faso

Q.2

Steps:

- To open the relevant dialog box, type `db cfgtsd2`
- Choose variables and parameters as in
Figure 32: Drawing the difference between FGT curves with confidence interval

Figure 33: Difference between FGT curves with confidence interval \((\alpha = 0)\)
21.6 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
After clicking SUBMIT, the following results appear:

<table>
<thead>
<tr>
<th>Number of intersection</th>
<th>Critical pov. line</th>
<th>Min. range of pov. lines</th>
<th>Max. range of pov. line</th>
<th>Case</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>24262.571</td>
<td>.</td>
<td>.</td>
<td>A</td>
</tr>
<tr>
<td>2</td>
<td>46775.652</td>
<td>.</td>
<td>.</td>
<td>B</td>
</tr>
</tbody>
</table>

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.

**21.7 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\bkf941.dta", clear*

- To open the relevant dialog box, type
  
  *db dfgtg*

- Choose variables and parameters as in

Figure 36: Decomposing FGT indices by groups

![FIGURE_36](image)

After clicking SUBMIT, the following information is provided:

dfgtg expc, bgroup(gs) hsize(size alpha(l) pline(41099) type(nor))

**FGT Index: Decomposition by Groups**

<table>
<thead>
<tr>
<th>Group</th>
<th>FGT Index</th>
<th>Population Share</th>
<th>Absolute Contribution</th>
<th>Relative Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: wage-earning (public sector)</td>
<td>0.004237</td>
<td>0.042971</td>
<td>0.00182</td>
<td>0.001308</td>
</tr>
<tr>
<td>2: wage-earning (private sector)</td>
<td>0.002571</td>
<td>0.003790</td>
<td>0.00117</td>
<td>0.000840</td>
</tr>
<tr>
<td>3: Artisan or trading</td>
<td>0.022176</td>
<td>0.026598</td>
<td>0.00550</td>
<td>0.004237</td>
</tr>
<tr>
<td>4: Others activities</td>
<td>0.010678</td>
<td>0.002164</td>
<td>0.000291</td>
<td>0.000203</td>
</tr>
<tr>
<td>5: Farmers (crop)</td>
<td>0.013755</td>
<td>0.119402</td>
<td>0.014358</td>
<td>0.010110</td>
</tr>
<tr>
<td>6: Farmers (food)</td>
<td>0.010004</td>
<td>0.600995</td>
<td>0.110932</td>
<td>0.096880</td>
</tr>
<tr>
<td>7: Inactive</td>
<td>0.014916</td>
<td>0.075056</td>
<td>0.00993</td>
<td>0.007873</td>
</tr>
<tr>
<td>POPULATION</td>
<td>0.139197</td>
<td>1.000000</td>
<td>0.139197</td>
<td>1.000000</td>
</tr>
</tbody>
</table>

0.006553  0.003320  0.008520
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(qse) hsiz(size1 alpha(1}) plms(41099) dstd(0) type(mor) dec(4)

FGT Index: Decomposition by Groups

<table>
<thead>
<tr>
<th>Group</th>
<th>FGT Index</th>
<th>Population Share</th>
<th>Absolute Contribution</th>
<th>Relative Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: wage-earning (public sector)</td>
<td>0.0042</td>
<td>0.0430</td>
<td>0.0002</td>
<td>0.0013</td>
</tr>
<tr>
<td>2: wage-earning (private sector)</td>
<td>0.0222</td>
<td>0.0266</td>
<td>0.0006</td>
<td>0.0042</td>
</tr>
<tr>
<td>3: Artisan or trading</td>
<td>0.0277</td>
<td>0.0266</td>
<td>0.0017</td>
<td>0.0125</td>
</tr>
<tr>
<td>4: Others activities</td>
<td>0.0039</td>
<td>0.0066</td>
<td>0.0004</td>
<td>0.0031</td>
</tr>
<tr>
<td>5: Farmers (crop)</td>
<td>0.1375</td>
<td>0.3944</td>
<td>0.0244</td>
<td>0.1031</td>
</tr>
<tr>
<td>6: Farmers (food)</td>
<td>0.1829</td>
<td>0.6059</td>
<td>0.1109</td>
<td>0.7068</td>
</tr>
<tr>
<td>7: Inactive</td>
<td>0.1449</td>
<td>0.0759</td>
<td>0.0110</td>
<td>0.0790</td>
</tr>
<tr>
<td>POPULATION</td>
<td>0.1392</td>
<td>1.0000</td>
<td>0.1392</td>
<td>1.0000</td>
</tr>
</tbody>
</table>
```
21.8 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 **b kf94l.dta** file,

3. Draw Lorenz curves for rural and urban households
   a. with variable of interest \( \text{exppc} \);
   b. with size variable set to \( \text{size} \);
   c. and using the group variable \( \text{zone} \) (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 37: Lorenz and concentration curves

After clicking SUBMIT, the following appears:

Figure 38: Lorenz curves
Q.2
Steps:

- Choose variables and parameters as in Figure 39: Drawing concentration curves

After clicking on SUBMIT, the following appears:
Figure 40: Lorenz and concentration curves

Steps:

- Type
  
  `use "C:\data\bkf94l.dta", clear`

- Choose variables and parameters as in...
Figure 41: Drawing Lorenz curves

Figure 42: Lorenz curves
21.9 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 igin`

- Choose variables and parameters as in

**Figure 43: Estimating Gini and concentration indices**

![Dialog box for estimating Gini and concentration indices](image)

After clicking **Submit**, the following results are obtained:
Q.2
Steps:

- Choose variables and parameters as in Figure 44: Estimating concentration indices

After clicking SUBMIT, the following results are obtained:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>STD</th>
<th>LB</th>
<th>UB</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: $GMI_X$</td>
<td>0.500456</td>
<td>0.016234</td>
<td>0.470399</td>
<td>0.530513</td>
</tr>
<tr>
<td>2: $GMI_M$</td>
<td>0.323558</td>
<td>0.012758</td>
<td>0.307318</td>
<td>0.357391</td>
</tr>
</tbody>
</table>

Q.3
Steps:

- To open the relevant dialog box, type `db digini`

- Choose variables and parameters as in
After clicking SUBMIT, the following information is obtained:

\[ \text{. digini expnz expnz, file1(c:\data\bkgf91.dta) hsize1(size) file2(c:\data\bkgf94.dta) hsize2(size)} \]

<table>
<thead>
<tr>
<th>Distribution 1: (GINI)</th>
<th>Distribution 2: (GINI)</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \bar{G} )</td>
<td>( \bar{G} )</td>
<td>( \bar{G} )</td>
</tr>
<tr>
<td>0.44563</td>
<td>0.45055</td>
<td>-0.00542</td>
</tr>
<tr>
<td>0.012816</td>
<td>0.008618</td>
<td>0.004544</td>
</tr>
<tr>
<td>0.419371</td>
<td>0.433116</td>
<td>-0.013745</td>
</tr>
<tr>
<td>0.465755</td>
<td>0.466696</td>
<td>-0.001941</td>
</tr>
<tr>
<td>( \text{STD} )</td>
<td>( \text{STD} )</td>
<td>( \text{STD} )</td>
</tr>
<tr>
<td>0.012816</td>
<td>0.008618</td>
<td>0.004544</td>
</tr>
<tr>
<td>0.419371</td>
<td>0.433116</td>
<td>-0.013745</td>
</tr>
<tr>
<td>0.465755</td>
<td>0.466696</td>
<td>-0.001941</td>
</tr>
<tr>
<td>( \text{LB} )</td>
<td>( \text{LB} )</td>
<td>( \text{UB} )</td>
</tr>
<tr>
<td>0.012816</td>
<td>0.008618</td>
<td>0.004544</td>
</tr>
<tr>
<td>0.419371</td>
<td>0.433116</td>
<td>-0.013745</td>
</tr>
<tr>
<td>0.465755</td>
<td>0.466696</td>
<td>-0.001941</td>
</tr>
<tr>
<td>( \text{UB} )</td>
<td>( \text{UB} )</td>
<td>( \text{UB} )</td>
</tr>
<tr>
<td>0.012816</td>
<td>0.008618</td>
<td>0.004544</td>
</tr>
<tr>
<td>0.419371</td>
<td>0.433116</td>
<td>-0.013745</td>
</tr>
<tr>
<td>0.465755</td>
<td>0.466696</td>
<td>-0.001941</td>
</tr>
</tbody>
</table>
21.10 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 46: Drawing densities

![Figure 46: Drawing densities](image)

After clicking SUBMIT, the following appears:
To open the relevant dialog box, type
\textit{db c\_quantile}

Choose variables and parameters as in
After clicking SUBMIT, the following appears:

**Figure 49: Quantile curves**
Q.3
Steps:

- To open the relevant dialog box, type `db cnpe`

- Choose variables and parameters as in

**Figure 50: Drawing non-parametric regression curves**

![Dialog box for non-parametric regression](image)

After clicking SUBMIT, the following appears:
Figure 51: Non-parametric regression curves

Non parametric regression
(Linear Locally Estimation Approach | Bandwidth = 3699.26 )

Q.4
Steps:

- Choose variables and parameters as in
Figure 52: Drawing derivatives of non-parametric regression curves

After clicking SUBMIT, the following appears:

Figure 53: Derivatives of non-parametric regression curves

Non parametric derivative regression
(Linear Locally Estimation Approach | Bandwidth = 3699.26 )
21.11  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 54: 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
After clicking SUBMIT, the following graph is plotted interactively with **Gnu Plot 4.2**:

![Joint Distribution Function](image)
21.12 Testing the bi-dimensional poverty dominance

Using the `columbia95l.dta` (distribution_1) and the `dominicar_republic95l.dta` (distribution_2) files,

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

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Var. of interest</th>
<th>Range</th>
<th>alpha_j</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dimension 1</td>
<td>haz: height-for-age</td>
<td>-3.0 / 6.0</td>
<td>0</td>
</tr>
<tr>
<td>Dimension 2</td>
<td>sprob: survival probability</td>
<td>0.7 / 1.0</td>
<td>0</td>
</tr>
</tbody>
</table>

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 56: 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:
21.13 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:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>strata</td>
<td>The stratum</td>
</tr>
<tr>
<td>psu</td>
<td>The primary sampling unit</td>
</tr>
<tr>
<td>weight</td>
<td>Sampling weight</td>
</tr>
<tr>
<td>inc</td>
<td>Income</td>
</tr>
<tr>
<td>hhsz</td>
<td>Household size</td>
</tr>
</tbody>
</table>

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

Figure 57: Testing the pro-poor growth (primal approach)

After clicking SUBMIT, the following graph appears

Relative propoor curve
(Order: s=1 \mid \text{Dif.} = P_2(\frac{m_2}{m_1}z, a=s-1) - P_1(z, a=s-1))

- Difference
- Upper bound of 95% confidence interval
- Null horizontal line
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 58: 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 59: Testing the pro-poor growth (dual approach) – B*
After clicking SUBMIT, the following graph appears:

![Graph showing absolute propoor curves](image)

**Q.4**

Steps:
- To open the relevant dialog box, type `db ipropoor`
Choose variables and parameters as.

After clicking submit, the following results appear:

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Estimation</th>
<th>SE</th>
<th>LB</th>
<th>UB</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Growth rate</strong></td>
<td>0.598959</td>
<td>0.125832</td>
<td>0.366361</td>
<td>0.829357</td>
</tr>
<tr>
<td><strong>Chan &amp; Ravallion (2003) index</strong></td>
<td>0.712265</td>
<td>1.099537</td>
<td>-1.265079</td>
<td>2.695490</td>
</tr>
<tr>
<td><strong>Kakwani &amp; Pernia (2000) index</strong></td>
<td>1.325436</td>
<td>0.107847</td>
<td>1.156287</td>
<td>1.535244</td>
</tr>
<tr>
<td><strong>PEG index</strong></td>
<td>0.721879</td>
<td>0.137331</td>
<td>0.582716</td>
<td>1.041042</td>
</tr>
<tr>
<td><strong>PEG - g</strong></td>
<td>0.169520</td>
<td>0.049357</td>
<td>0.092783</td>
<td>0.286257</td>
</tr>
</tbody>
</table>

**21.14 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 \texttt{exppc}
   - The number of household members that benefit from education is \texttt{fr\_prim} for the primary sector and \texttt{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 60: Benefit incidence analysis**

![Figure 60: Benefit incidence analysis](image)

After clicking on **Submit**, the following appears:
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_pub_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 61: Benefit Incidence Analysis (unit cost approach)

After clicking on Submit, the following appears:
Average Benefits by Quintile Groups: (at the level of eligible members)

<table>
<thead>
<tr>
<th>Groups</th>
<th>Sector 1</th>
<th>Sector 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quintile 1</td>
<td>240.602</td>
<td>128.548</td>
</tr>
<tr>
<td>Quintile 2</td>
<td>257.403</td>
<td>179.816</td>
</tr>
<tr>
<td>Quintile 3</td>
<td>250.395</td>
<td>186.119</td>
</tr>
<tr>
<td>Quintile 4</td>
<td>225.527</td>
<td>192.940</td>
</tr>
<tr>
<td>Quintile 5</td>
<td>157.962</td>
<td>143.327</td>
</tr>
<tr>
<td><strong>All</strong></td>
<td>227.961</td>
<td>166.095</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Groups</th>
<th>Sector 1</th>
<th>Sector 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quintile 1</td>
<td>312.004</td>
<td>200.540</td>
</tr>
<tr>
<td>Quintile 2</td>
<td>312.004</td>
<td>200.540</td>
</tr>
<tr>
<td>Quintile 3</td>
<td>312.004</td>
<td>200.540</td>
</tr>
<tr>
<td>Quintile 4</td>
<td>312.004</td>
<td>200.540</td>
</tr>
<tr>
<td>Quintile 5</td>
<td>312.004</td>
<td>200.540</td>
</tr>
<tr>
<td><strong>All</strong></td>
<td>312.004</td>
<td>200.540</td>
</tr>
</tbody>
</table>

Proportion of Benefits by Quintile Groups and by Sectors.

<table>
<thead>
<tr>
<th>Groups</th>
<th>Sector 1</th>
<th>Sector 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quintile 1</td>
<td>0.136</td>
<td>0.058</td>
</tr>
<tr>
<td>Quintile 2</td>
<td>0.141</td>
<td>0.081</td>
</tr>
<tr>
<td>Quintile 3</td>
<td>0.137</td>
<td>0.084</td>
</tr>
<tr>
<td>Quintile 4</td>
<td>0.123</td>
<td>0.087</td>
</tr>
<tr>
<td>Quintile 5</td>
<td>0.087</td>
<td>0.065</td>
</tr>
<tr>
<td><strong>All</strong></td>
<td>0.624</td>
<td>0.376</td>
</tr>
</tbody>
</table>