

# Inequality in Health

## Tutorial 2

Nikolaos Prodromidis, M.Sc.

[nikolaos.prodromidis@uni-due.de](mailto:nikolaos.prodromidis@uni-due.de)



Winter Term 2022/2023

1. Open "*dataset1\_s10.dta*" and merge it with "*deciles1\_s10.dta*".

Set the working directory and open the dataset with

Stata:

```
cd "C:\path\  
use ".\dataset1_s10.dta"
```

and merge it to the deciles dataset with

Stata:

```
merge 1:1 scid using ".\deciles_s10.dta"  
drop if _merge==2  
drop _merge
```

## 2. What is the difference between wage and income?

Wages are the **labor earnings** per time unit (week, month, year, etc.).  
Wages are a part of income which may further consist of **public transfers** (e.g. unemployment benefits) and **capital income** (interest, dividends, etc.).

### 3. Why is it reasonable to consider equivalent incomes in income comparisons?

Household needs grow with household size (e.g. food, rooms) but **not proportionately** (e.g. electricity, internet, lower needs of children in general). Adjusting incomes for the household size makes incomes **more comparable**.

## 4. Plot the income distribution of India using a histogram with 30 bins.

Histograms are plotted by the command

Stata:

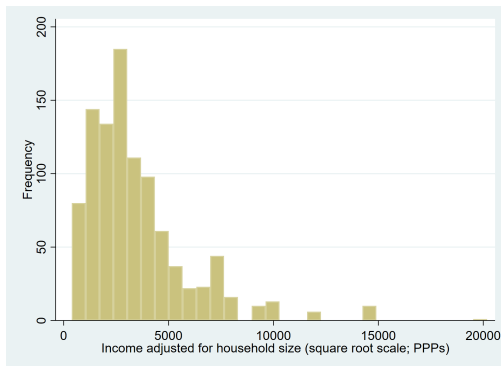
```
histogram var if condition, bin(n_bin) frequency
```

The option `bin(n_bin)` splits the range of the variable *var* from minimum to maximum into *n\_bin* equally sized intervals.

## 4. Plot the income distribution of India using a histogram with 30 bins.

Stata:

```
histogram incompppE if country==8, bin(30) frequency
```



## 5. Plot the same distribution using kernel density estimation.

Histograms have the disadvantage that they do not provide information on the **distribution within a bin**. Further, histograms are not necessarily able to provide a **continuous distribution** of the variable of interest as in our example. Kernel density estimation allows to estimate the probability density function of a variable using the kernel density estimator

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right)$$

where  $h$  is a smoothing parameter (*bandwidth*), and  $K(x)$  is a kernel function (e.g. the Gaussian kernel  $K(x) = \phi(x)$ ).

## 5. Plot the same distribution using kernel density estimation.

In Stata, the command

Stata:

```
kdensity var if condition, bwidth(bw) kernel(gaussian)
```

plots a kernel density graph. Histogram and kernel density graph can be combined by

Stata:

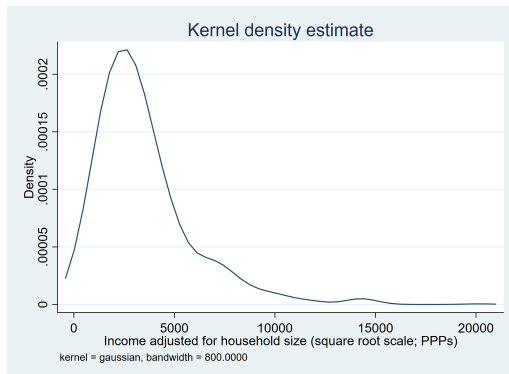
```
twoway (histogram var, bin(n_bin) density) || /// (kdensity  
var, bwidth(bw) kernel(gaussian))
```



## 5. Plot the same distribution using kernel density estimation.

Stata:

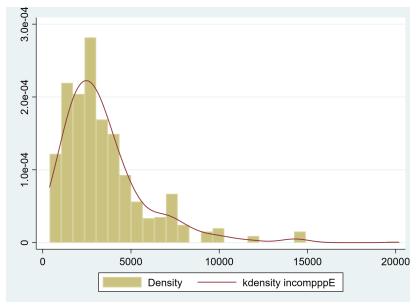
```
kdensity incompppE if country==8, bwidth(800) kernel(gaussian)
```



## 5. Plot the same distribution using kernel density estimation.

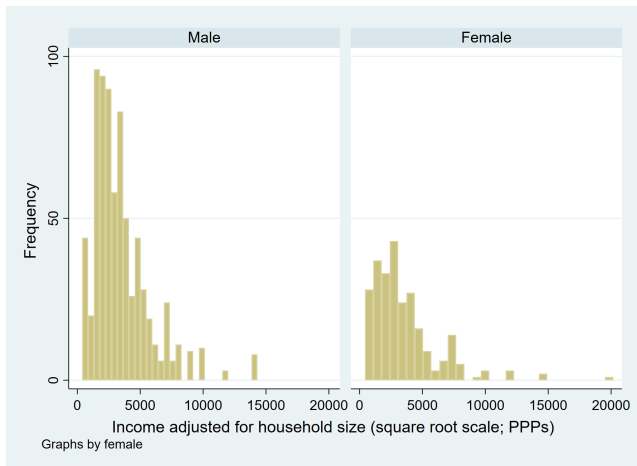
Stata:

```
twoway (histogram incompppE if country==8, bin(30)) || ///
(kdensity incompppE if country==8, ///
bwidth(800) kernel(gaussian))
```



## 6. Plot the income distribution of India by sex.

The option `by(var)` allows to plot separate graphs for each value of `var`.



# 1. Compare the top decile income shares of India and South Africa.

Stata:

```
sum incompppE if country==8 & decile_IN==10  
sum incompppE if country==8
```

shows that the top decile in India consists of **100** individuals with a mean income of **9056.044** whereas the total sample consists of **995** observations with a mean of **3464.451**. Thus, the income share of the top decile in India is more than **26%**.

# 1. Compare the top decile income shares of India and South Africa.

Similarly,

Stata:

```
sum incompppE if country==16 & decile_ZA==10  
sum incompppE if country==16
```

shows that the top decile income share in South Africa is more than **47%**.

2. Plot the Lorenz curves for India and South Africa and add a  $45^\circ$  line. How does the notion of Lorenz dominance apply in this setting?

The Lorenz curve is a graphical representation of the cumulative distribution function of income. The cumulative population share (individuals ordered from lowest to highest income) is shown on the x-axis. The y-axis indicates the share of total income of the poorest  $x$  percent of the population.

2. Plot the Lorenz curves for India and South Africa and add a 45° line. How does the notion of Lorenz dominance apply in this setting?

In Stata, the command

Stata:

```
ssc install glcurve // if not installed  
glcurve var, lorenz pvar(xvar) glvar(yvar)
```

creates the necessary x-axis (*xvar*) and y-axis (*yvar*) coordinates which can be used to create Lorenz curve plots manually.

2. Plot the Lorenz curves for India and South Africa and add a 45° line. How does the notion of Lorenz dominance apply in this setting?

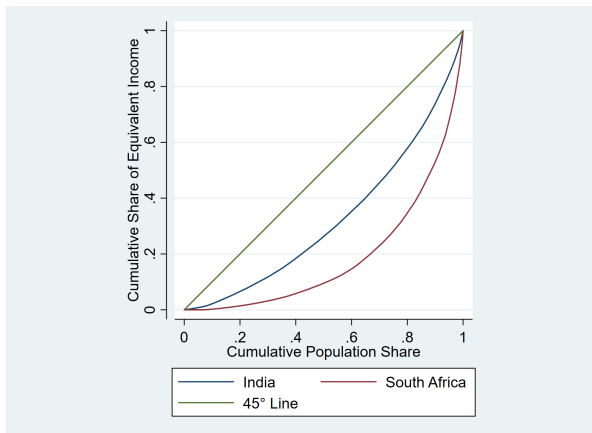
Stata:

```
twoway (line yvar xvar, sort) || ///
(function y = x), aspectratio(1)
```

plots the Lorenz curve (first parentheses) and the 45° line (second parentheses). The plots are combined via "||"; adding further plots is possible.



2. Plot the Lorenz curves for India and South Africa and add a 45° line. How does the notion of Lorenz dominance apply in this setting?



2. Plot the Lorenz curves for India and South Africa and add a  $45^\circ$  line. How does the notion of Lorenz dominance apply in this setting?

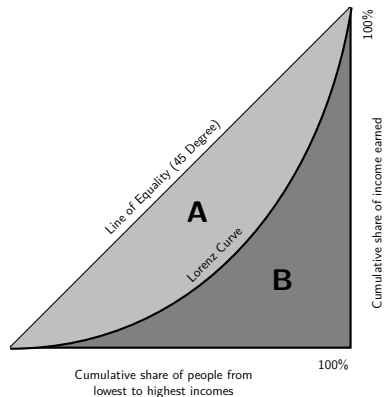
The Lorenz curve of India is closer to the  $45^\circ$  line than the curve of South Africa. Thus, South Africa's income distribution is Lorenz-dominated by India's income distribution.

### 3. What is the Gini coefficient?

The Gini coefficient ( $G$ ) is the area between the 45° line and the Lorenz curve ( $A$ ) compared to the area between the lines of perfect equality and perfect inequality ( $A + B$ ). The higher the Gini coefficient, the more unequal is the distribution.

$$G = \frac{A}{A + B} = 2A$$

### 3. What is the Gini coefficient?



$$G = \frac{A}{A + B} = 2A$$

## 4. What is the Theil index?

The Theil index  $T$  is less intuitive than the Gini coefficient. It is an alternative measure to calculate income inequality. When everybody has the same income, a state of **perfect equality** is reached and the index equals **zero**. The higher the Theil index, the more unequal is the distribution. The maximum value of the index for a population of size  $N$  is  $\ln(N)$  (**perfect inequality**). The index is defined as

$$T = \frac{1}{N} \sum_{i=1}^N \left( \frac{x_i}{\bar{x}} \ln \frac{x_i}{\bar{x}} \right)$$

where  $x_i$  is the individual income and  $\bar{x}$  indicates the population's mean income.

## 5. Calculate the Gini coefficient and the Theil index for India and South Africa.

Gini coefficient and Theil index can be calculated in Stata with

Stata:

```
ssc install inequal2 // if not installed  
inequal2 var
```

## 5. Calculate the Gini coefficient and the Theil index for India and South Africa.

With

Stata:

```
inequal2 incompppE if country==8
inequal2 incompppE if country==16
```

we obtain larger indices for South Africa (**0.63** and **0.74**) than for India (**0.35** and **0.21**) indicating that South Africa has a higher income inequality.

# 1. What is the difference between a Lorenz curve and a concentration curve?

The notions of concentration curve and Lorenz curve are quite similar. In case of the **Lorenz curve**, the variable ordering the population on the x-axis and the (cumulative) variable on the y-axis are the **same**. In case of the **concentration curve**, the two variables are **different**.



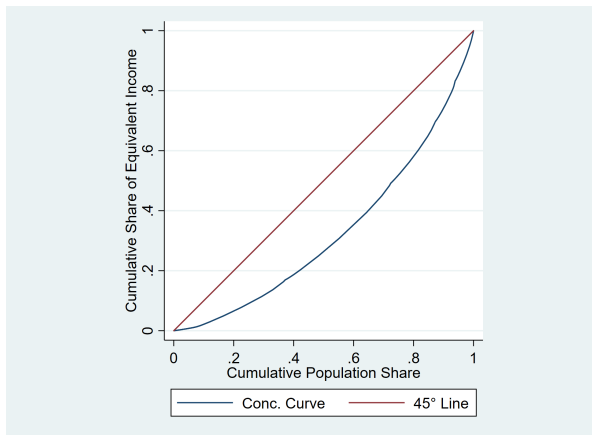
## 2. Plot the concentration curve of equivalent income with respect to household income for India.

To obtain the concentration curve coordinates, the `sortvar(svar)` option must be added to the `glcurve` command:

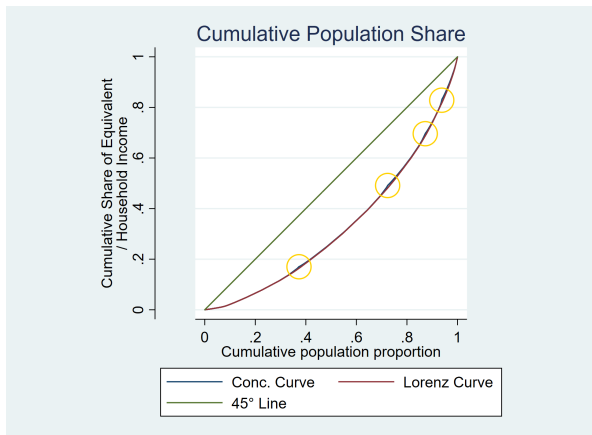
Stata:

```
glcurve var, lorenz p(xvar) gl(yvar) sortvar(svar)
```

## 2. Plot the concentration curve of equivalent income with respect to household income for India.



### 3. Plot the concentration curve and the Lorenz curve for India in one figure. Explain the result.



### 3. Plot the concentration curve and the Lorenz curve for India in one figure. Explain the result.

Although hardly visible (see *yellow circles*), the concentration curve lies above the Lorenz curve at several points. Because e.g. the income share of the poorest 10% is always lower than/equal to the income share of any other 10% of the population, a concentration curve can **never** lie below the Lorenz curve.

# 1. What is the income of the richest person in India? What is the income of the poorest person in India?

Stata:

```
sum incompppE if country==8
```

Variable	Obs	Mean	Std. Dev.	Min	Max
-----+-----					
incompppE	995	3464.451	2432.767	388.0115	20181

The poorest person in India has an equivalent income of **388.0115** whereas the richest person has an equivalent income of **20,181**.

## 2. Transfer an amount of 1,000 per person from the richest decile to the second richest decile in India. How do Gini coefficient and Theil index respond?

To transfer income from one decile to another, we generate a new variable *incompppE\_tr* which is equal to *incompppE*, and then reduce it by 1,000 for the individuals in the 10th decile and increase it by 1,000 for 9th decile individuals.

Stata:

```
gen incompppE_tr = incompppE
replace incompppE_tr = incompppE-1000 if decile_IN==10 & country==8
replace incompppE_tr = incompppE+1000 if decile_IN==9 & country==8
```

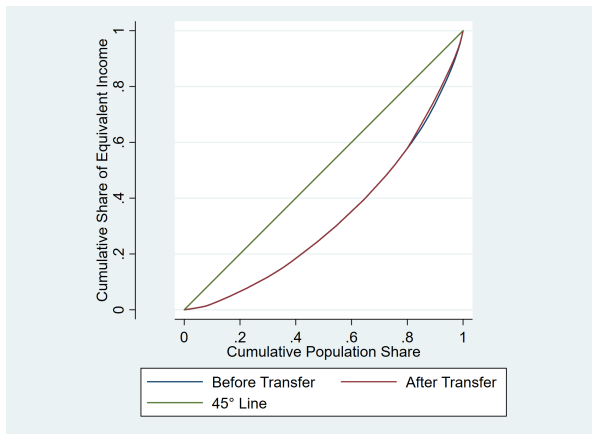
2. Transfer an amount of 1,000 per person from the richest decile to the second richest decile in India. How do Gini coefficient and Theil index respond?

Index	Before the Transfer	After the Transfer
Gini	0.3517	0.3470
Theil	0.2088	0.1994

Both measures indicate that inequality is reduced.

### 3. Explain how the Lorenz curve responds to the transfer.

Generate the new x- and y-axis coordinates with the `g1curve` command and plot the before- and after-transfer Lorenz curves.





### 3. Explain how the Lorenz curve responds to the transfer.

The Lorenz curve shifts up only at its right end and remains unchanged elsewhere.