Inequality in Health

Tutorial 3

Nikolaos Prodromidis, M.Sc.

nikolaos.prodromidis@uni-due.de

Winter Term 2022/2023

1. Open "*dataset1_s10.dta*" in Stata. It contains data on women and their children in India and Mali

Set the working directory and open the dataset with

Stata:

```
cd "C:\path\"
use ".\dataset2_s10.dta"
```

2. Plot the Lorenz curve for malnutrition and the concentration curve with respect to wealth for India, adding the 45° line

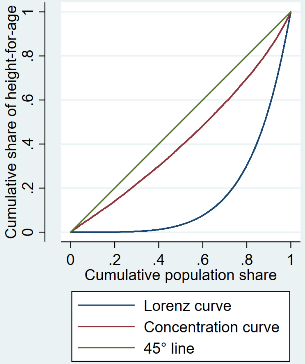
We determine the Lorenz and concentration curve coordinates with

Stata:

```
* Lorenz curve
glcurve var, lorenz p(xvar) gl(yvar)

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

2. Plot the Lorenz curve for malnutrition and the concentration curve with respect to wealth for India, adding the 45° line

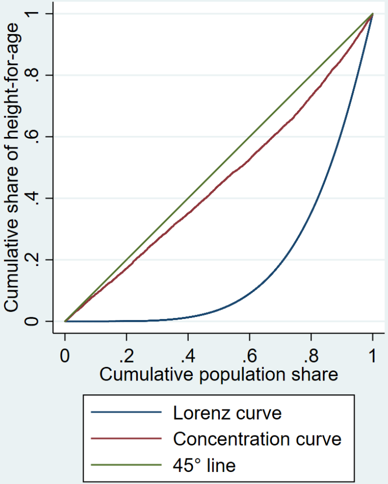


Lorenz and concentration curves for child malnutrition in India. The population is ordered by height-for-age for the Lorenz curve and by wealth level for the concentration curve, respectively, on the x-axis.

2. Plot the Lorenz curve for malnutrition and the concentration curve with respect to wealth for India, adding the 45° line

The shape of the **Lorenz curve** is very extreme, indicating a **very high inequality** in child malnutrition in India. For example, the 60% of children with the lowest values of height-for-age (i.e. the highest malnutrition), have less than 10% of children's total sum of height-for-age values.
The **concentration curve** with respect to wealth is **less extreme**, indicating that although poorer children have lower height-for-age values, differences in the wealth level can by far not explain the high inequality in child malnutrition shown by the Lorenz curve.

3. Plot the same curves for Mali.

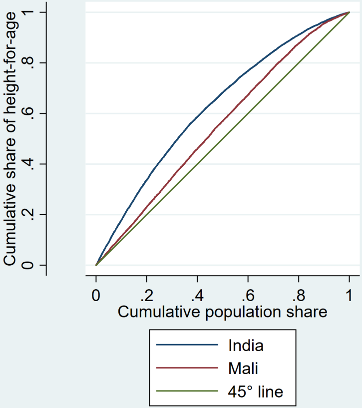


Lorenz and concentration curves for child malnutrition in Mali.

3. Plot the same curves for Mali.

The Lorenz curve for Mali has a similar shape as the curve for India. However, the concentration curve indicates **less inequality** in the lower quantiles.

4. Plot the concentration curves for under-five deaths with respect to wealth for India and Mali.



Concentration curves for India and Mali. The population is ordered from lowest to highest wealth level on the x-axis.

5. If we want to test for the statistical significance of dominance, why do we need a decision rule? What are the possible outcomes?

By comparing standard errors we can only test whether **two points** are statistically different. The concentration curve, however, is a **line** estimate. Therefore, we must decide **at how many points** the two curves must be statistically different to be overall considered different from each other.

5. If we want to test for the statistical significance of dominance, why do we need a decision rule? What are the possible outcomes?

Possible outcomes:

1. Concentration curve dominance: at least x significant differences (x according to decision rule) in one direction, none in the other
2. Non-dominance: no significant differences in any direction
3. Curves cross: significant differences in both directions

1. How is the concentration index related to the Gini coefficient?

The Gini coefficient measures the inequality determined by some Lorenz curve. The concentration index measures the inequality determined by some concentration curve, i.e. it equals twice the area between the curve and the 45° line.

2. Calculate the concentration index of child malnutrition with respect to wealth for India using the formula from the question sheet.

We are interested in estimating the concentration index as the coefficient β of the regression

$$2\sigma_r^2 \left(\frac{h_i}{\mu}\right) = \alpha + \beta r_i + \varepsilon_i$$

where

- h_i Height-to-age of child i
- r_i Rank of child i with respect to wealth
- σ_r^2 Variance of the rank variable
- μ Mean of the malnutrition variable

2. Calculate the concentration index of child malnutrition with respect to wealth for India using the formula from the question sheet.

The concentration index can be expressed as $C = \frac{2}{\mu} Cov\left(h, r\right)$. The coefficient β of a regression $y = \alpha + \beta x + u$ is defined as $\beta^{OLS} = \frac{Cov(x,y)}{Var(x)}$. Thus, our regression estimates

$$\begin{aligned}
\beta &= \frac{Cov\left(r, 2\sigma_r^2\left(\frac{h}{\mu}\right)\right)}{Var\left(r\right)} \\
&= \frac{2}{\mu}\frac{\sigma_r^2}{Var\left(r\right)} Cov\left(h, r\right) \\
&= \frac{2}{\mu} Cov\left(h, r\right) = C
\end{aligned}$$

2. Calculate the concentration index of child malnutrition with respect to wealth for India using the formula from the question sheet.

First, we drop Mali, as it is no longer needed:

Stata:

```
drop if country==2
```

Then, we generate our rank variable with

Stata:

```
glcurve wealth, p(rank) nograph
```

Note that we are only interested in the rank so we do not need to generate the y-axis coordinates with the `gl` command.

2. Calculate the concentration index of child malnutrition with respect to wealth for India using the formula from the question sheet.

Next, we need to determine the mean of the malnutrition via

Stata:

```
sum hfa
scalar mean_hfa=r(mean)
```

and the variance of the rank variable with

Stata:

```
sum rank
scalar var_rank=r(Var)
```

2. Calculate the concentration index of child malnutrition with respect to wealth for India using the formula from the question sheet.

```
      Source |       SS           df       MS      Number of obs   =    40,984
-------------+----------------------------------   F(1, 40982)     =   1731.78
       Model |  90.4132318         1  90.4132318   Prob > F        =    0.0000
    Residual |  2139.59468    40,982  .052208157   R-squared       =    0.0405
-------------+----------------------------------   Adj R-squared   =    0.0405
       Total |  2230.00791    40,983  .054412998   Root MSE        =    .22849

------------------------------------------------------------------------------
         lhs |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        rank |   .1645037    .003953    41.61   0.000     .1567557    .1722517
       _cons |   .0897268   .0021662    41.42   0.000     .0854811    .0939725
------------------------------------------------------------------------------
```

2. Is the concentration index significantly different from zero?

The estimated concentration index is 0.1645. Given that the corresponding t-value is larger than 1.96, the concentration index is **statistically different from zero** and **inequality** in child malnutrition by wealth level is indeed present.

3. Standardize the concentration index by *educ*, *age*, *watertime*, and *electricity*.

```
      Source |       SS           df       MS      Number of obs   =    38,376
-------------+----------------------------------   F(5, 38370)     =    382.51
       Model |  98.055479         5  19.6110958   Prob > F        =    0.0000
    Residual | 1967.22016    38,370  .051269746   R-squared       =    0.0475
-------------+----------------------------------   Adj R-squared   =    0.0474
       Total | 2065.27564    38,375  .053818258   Root MSE        =    .22643

------------------------------------------------------------------------------
         lhs |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        rank |   .1134621   .0066264    17.12   0.000     .1004743      .12645
        educ |   .0049058   .0003067    16.00   0.000     .0043047     .005507
         age |   .0005914   .0002144     2.76   0.006     .0001711    .0010117
   watertime |   .0001108     .00007     1.58   0.114    -.0000265     .000248
 electricity |  -.0062344   .0033297    -1.87   0.061    -.0127608     .000292
       _cons |     .07402   .0063521    11.65   0.000     .0615698    .0864703
------------------------------------------------------------------------------
```

3. Do these covariates explain the estimated inequality in malnutrition by wealth level?.

After controlling for covariates, the concentration index **shrinks** to 0.1135, but it is still **statistically different from zero**. Hence, differences in mother's education, mother's age, time to get to water source and access to electricity can explain only a portion of the estimated inequality in child malnutrition by wealth level.

1. What is an Oaxaca-Blinder decomposition?

Generally speaking, decompositions reveal to what extent differences in the outcome variable can be explained by inequalities in its determinants. The Oaxaca-Blinder decomposition explains the gap in the means of an outcome variable between two groups.

1. What is an Oaxaca-Blinder decomposition?

The gap is decomposed in two parts (and an interaction):

1. Group differences in the levels of the determinants (ΔX, explained component)
2. Group differences in the effects of the determinants ($\Delta\beta$, unexplained component)

$$\mathbb{E}\left(Y_1 - Y_2\right) = \underbrace{\mathbb{E}\left(\Delta X\right)'\beta_2}_{E} + \underbrace{\mathbb{E}\left(X_2\right)'\Delta\beta}_{C} + \underbrace{\mathbb{E}\left(\Delta X\right)'\Delta\beta}_{CE} \tag{1}$$

- E Gap in **endowments** (explained)
- C Gap in **coefficients** (unexplained)
- CE **Interaction** of these two gaps.

2. Compare the mean differences in child malnutrition between poor and non-poor. Are they different?

```
. sum hfa if poor==0

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
         hfa |     26,273    2169.693    2747.957          0       9979

. sum hfa if poor==1

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
         hfa |     14,711    1405.179    2396.336          0       9978
```

Non-poor children have a mean height-for-age score of **2169.693** whereas the mean among poor children is **1405.179** ($\approx$65%).

2. Does it seem reasonable to use the Oaxaca-Blinder decomposition in this setting?

It makes sense to apply the Oaxaca-Blinder decomposition if there is a difference in outcomes among two groups, if there are systematic differences in the explaining factors, and if the effects of these factors vary. The previous tables indicate that there actually is a difference in malnutrition between poor and non-poor children.

2. Does it seem reasonable to use the Oaxaca-Blinder decomposition in this setting?

```
. sum hfa if poor==0 & educ==0

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
         hfa |      6,236     1647.21    2539.646          0       9978

. sum hfa if poor==0 & educ>0

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
         hfa |     20,037    2332.302    2789.803          0       9979
```

If we look at malnutrition of non-poor children for mothers with at least some and mothers without any education, we find **education** to be **positively associated with height-for-age**. Less than **24%** of the non-poor children have mothers **without any education**.

2. Does it seem reasonable to use the Oaxaca-Blinder decomposition in this setting?

```
. sum hfa if poor==1 & educ==0

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
         hfa |     10,214    1360.552    2395.295          0       9978

. sum hfa if poor==1 & educ>0

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
         hfa |      4,497    1506.539     2395.88          0       9978
```

In case of poor children, almost **70%** have **uneducated mothers**. Here we also find a positive **correlation** between education and height-for-age but it is much **weaker**. This indicates that there are systematic differences in mother's education of poor and non-poor and that the effect of education on malnutrition differs as well.

3. Run a regression of malnutrition on the poverty dummy and its interactions with the following covariates: *childage*, *son*, *educ*, *bmi*, *watertime*.

Source	SS	df	MS
Model	2.0397e+10	6	3.3995e+09
Residual	2.4507e+11	38,157	6422746.18
Total	2.6547e+11	38,163	6956204.29

Number of obs	=	38,164
F(6, 38157)	=	529.29
Prob > F	=	0.0000
R-squared	=	0.0768
Adj R-squared	=	0.0767
Root MSE	=	2534.3

hfa	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
poor	-222.0798	31.7869	-6.99	0.000	-284.383	-159.7766
childage	-319.8329	9.348174	-34.21	0.000	-338.1555	-301.5102
son	-29.05925	25.98032	-1.12	0.263	-79.98136	21.86287
educ	70.42887	3.017454	23.34	0.000	64.51458	76.34316
bmi	80.66869	3.990225	20.22	0.000	72.84774	88.48963
watertime	-.3081186	.7740032	-0.40	0.691	-1.825185	1.208948
_cons	635.3963	87.09713	7.30	0.000	464.6836	806.109

3. Run a regression of malnutrition on the poverty dummy and its interactions with the following covariates: *childage*, *son*, *educ*, *bmi*, *watertime*.

hfa	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
poor	414.0104	203.4453	2.03	0.042	15.25232	812.7685
childage	-331.2526	11.72489	-28.25	0.000	-354.2337	-308.2715
poor#c.childage	28.28952	19.39938	1.46	0.145	-9.733766	66.31281
son	-24.85768	32.48739	-0.77	0.444	-88.53381	38.81845
poor#son	-12.13402	54.02316	-0.22	0.822	-118.0208	93.75279
educ	78.941	3.350674	23.56	0.000	72.37359	85.50841
poor#c.educ	-57.35259	7.916095	-7.25	0.000	-72.86834	-41.83684
bmi	84.00223	4.465983	18.81	0.000	75.24878	92.75567
poor#c.bmi	-30.27286	10.07227	-3.01	0.003	-50.01477	-10.53096
watertime	-2.441207	1.14734	-2.13	0.033	-4.690024	-.19239
poor#c.watertime	3.941614	1.553127	2.54	0.011	.8974437	6.985783
_cons	540.7939	97.31352	5.56	0.000	350.0569	731.531

3. Check whether the differences in the effects on malnutrition between poor and non-poor are systematic.

hfa	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
poor	414.0104	203.4453	2.03	0.042	15.25232	812.7685
childage	-331.2526	11.72489	-28.25	0.000	-354.2337	-308.2715
poor#c.childage	28.28952	19.39938	1.46	0.145	-9.733766	66.31281
son	-24.85768	32.48739	-0.77	0.444	-88.53381	38.81845
poor#son	-12.13402	54.02316	-0.22	0.822	-118.0208	93.75279
educ	78.941	3.350674	23.56	0.000	72.37359	85.50841
poor#c.educ	-57.35259	7.916095	-7.25	0.000	-72.86834	-41.83684
bmi	84.00223	4.465983	18.81	0.000	75.24878	92.75567
poor#c.bmi	-30.27286	10.07227	-3.01	0.003	-50.01477	-10.53096
watertime	-2.441207	1.14734	-2.13	0.033	-4.690024	-.19239
poor#c.watertime	3.941614	1.553127	2.54	0.011	.8974437	6.985783
_cons	540.7939	97.31352	5.56	0.000	350.0569	731.531

4. Test whether a Oaxaca-Blinder decomposition would help in explaining the gap in child malnutrition between poor and non-poor in India, using the set of covariates from the previous question.

The **t-statistic** is useful if we are interested in testing whether **one coefficient** is significantly different from zero. However, if we want to test **multiple hypotheses** at the same time, we must use a **Wald test**. In Stata, the command `testparm` followed by the variables for the coefficients we want to test, provides the F-statistic and p-value for the hypothesis that all coefficients are jointly equal to zero.

4. Test whether a Oaxaca-Blinder decomposition would help in explaining the gap in child malnutrition between poor and non-poor in India, using the set of covariates from the previous question.

Stata:

```
testparm poor 1.poor#c.childage 1.poor#1.son 1.poor#c.educ 1.poor#c.bmi ///
1.poor#c.watertime
```

```
( 1)  poor = 0
( 2)  1.poor#c.childage = 0
( 3)  1.poor#1.son = 0
( 4)  1.poor#c.educ = 0
( 5)  1.poor#c.bmi = 0
( 6)  1.poor#c.watertime = 0

      F(  6, 38152) =   21.16
           Prob > F =    0.0000
```

4. Test whether a Oaxaca-Blinder decomposition would help in explaining the gap in child malnutrition between poor and non-poor in India, using the set of covariates from the previous question.

As the p-value is below 0.05, the null hypothesis of **homogeneous effects** for poor and non-poor can be **rejected**. Hence, the regression coefficient vector beta differs systematically between the poor and the non-poor. Therefore, an Oaxaca-Blinder decomposition is reasonable in this context.

5. Conduct the Oaxaca-Blinder decomposition and interpret the results.

In Stata, an Oaxaca-Blinder decomposition is conducted via the command `decompose`:

Stata:

```
ssc install decompose // if not installed
decompost decomp_var covariates, by(group_var) detail estimates
```

5. Conduct the Oaxaca-Blinder decomposition and interpret the results.

```
-----------------------------------
  Mean prediction high (H):2142.298
   Mean prediction low (L):1386.943
Raw differential (R) H-L: 755.355
   - due to endowments (E): 208.635
 - due to coefficients (C): 152.640
 - due to interaction (CE): 394.080
-----------------------------------
```

Differences in endowments are more important than differences in the coefficients (**E=208.635** vs **C=152.640**). However, the main part of the gap in malnutrition between poor and non-poor remains unexplained (**CE=394.080**).

5. Conduct the Oaxaca-Blinder decomposition and interpret the results.

Decomposition results for variables:

				explained: D =			
Variables	E(D=0)	C	CE	1	0.5	0.639	*
childage	2.775	-59.220	0.259	3.034	2.905	2.941	2.922
son	-0.805	6.183	0.264	-0.541	-0.673	-0.636	-0.559
educ	112.993	100.426	300.181	413.174	263.083	304.842	415.759
bmi	106.397	577.622	59.947	166.344	136.370	144.710	167.608
watertime	-12.725	-58.360	33.428	20.703	3.989	8.640	9.253
_cons	0.000	-414.010	0.000	0.000	0.000	0.000	0.000
Total	208.635	152.640	394.080	602.715	405.675	460.496	594.983

5. Conduct the Oaxaca-Blinder decomposition and interpret the results.

Overall, especially differences in **mother's education and BMI** between poor and non-poor contribute to the gap in child malnutrition. However, we cannot conclude that these variables cause malnutrition!

6. What is the advantage of decomposing the concentration index over an Oaxaca-Blinder decomposition?

A decomposition of the concentration index allows decomposing inequalities in health across the **full distribution** of wealth (rather than simply between the poor and the non-poor).
On the other hand, it does not make distinctions between the contributions of differences in endowments and differences in effects.