Dra. Linda Llamas / Universidad Estatal de Sonora

Comandos clorenz, cdensity y digini y su aplicación en el análisis de la distribución del ingreso.

Outline

1. Objective
2. Measuring income distribution
3. Standard sintaxis
4. Application with STATA: cdensity, clorenz & digini commands
5. Conclusions

Objective

To illustrate alternative tools for the analysis of income distribution by using DASP module commands (Distributive Analysis Stata Package), such as clorenz, cdensity and digini, which operate in the STATA environment.

Measuring income distribution

Probability density function

Consider any random quantity of X that has probability density function f. Specifying the function f gives a natural description of the distribution of X, and allows probabilities associated with X to be found from the relation

$$P(a > X < b) = \int_a^b f(x)dx \qquad \text{for all } a < b \qquad (1)$$

(Silverman, 1986)

Lorenz curves

Indicates the cumulative percentage of total income held by a cumulative proportion *p* of the population, when individuals are ordered in increasing income values.

$$L(p) = \frac{\int_0^p Q(q)dq}{\int_0^1 Q(q)dq} = \frac{1}{\mu}\int_0^p Q(q)dq \qquad (2)$$

(Duclos & Araar, 2006)

Gini Index

Computes the average distance between cumulated population shares and cumulated income shares.

$$\frac{Gini\ index\ of\ inequality}{2} = \int_0^1 (p - L(p))dp \qquad (3)$$

(Duclos & Araar, 2006)

Standard Sintaxis STATA

Dataset and variables

- Database: **pobreza_22.dta**

  CONEVAL 2022 database for poverty measurement in Mexico (based on National Survey of Household Income and Expenditure, ENIGH 2022).

- Data source:

  https://www.coneval.org.mx/Medicion/Paginas/PobrezaInicio.aspx

- Variables:

  ictpc: Total per capita income (mexican pesos)
  rururb: Rural or urban area
  factor: Survey weight

Density curves

kdensity varname [if] [in] [weight] [, options]

- Main options:
  - kernel(kernel) default is epanechnikov
  - bwidth(#)
  - generate(newvar_x newvar_d)
  - n(#)
  - at(var_x)
  - nograph
- Kernel plot
  - cline_options
- Density plots
  - Normal
  - normopts(cline_options)
  - student(#)
  - stopts(cline_options)
- Add plots
  - addplot(plot)

Example for subgroups:

kdensity ictpc if ictpc<=15000 & rururb==0, kernel(gaussian) bwidth(220.2523) legend(label(1 "Urbano")) addplot(**kdensity** ictpc if ictpc<=15000 & rururb==1, kernel(gaussian) bwidth(220.2523) legend(label(2 "Rural")))

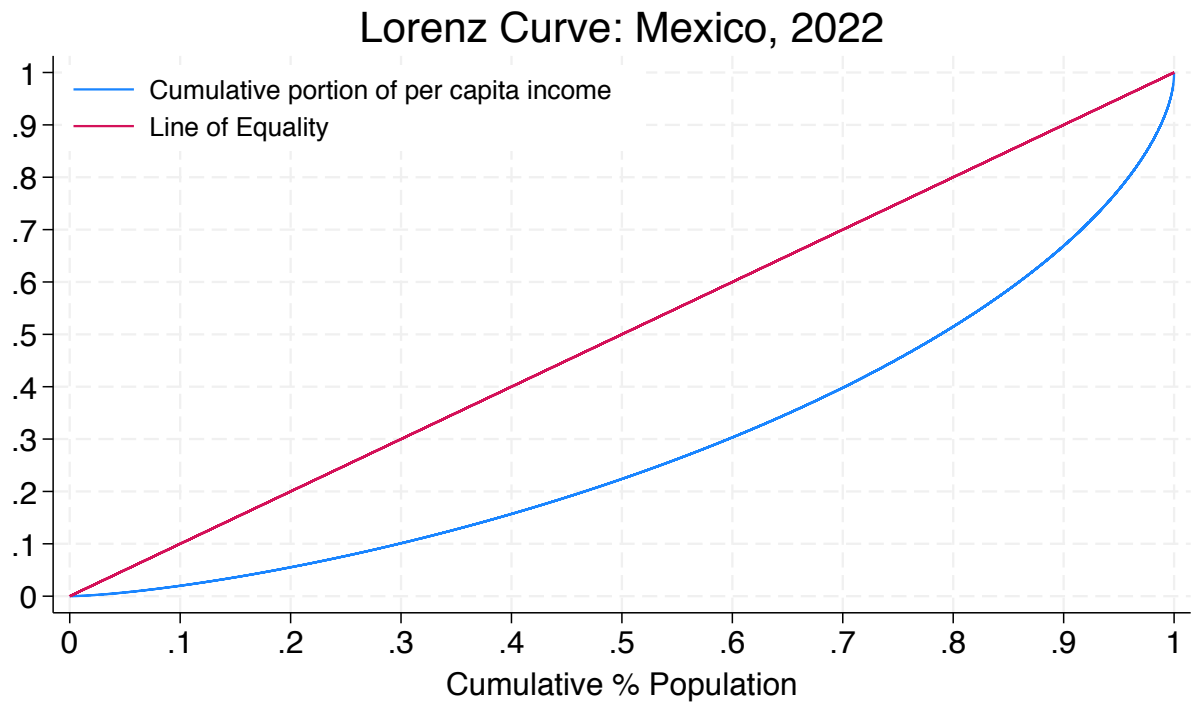
kdensity *varname* [*if*] [*in*] [*weight*] [, *options*]

options	Description
Main	
kernel(*kernel***)**	specify kernel function; default is **kernel(epanechnikov)**
bwidth(#)	half-width of kernel
generate(*newvar_x newvar_d***)**	store the estimation points in *newvar_x* and the density estimate in *newvar_d*
n(#)	estimate density using # points; default is min(N, 50)
at(*var_x***)**	estimate density using the values specified by *var_x*
nograph	suppress graph
Kernel plot	
cline_options	affect rendition of the plotted kernel density estimate
Density plots	
normal	add normal density to the graph
normopts(*cline_options***)**	affect rendition of normal density
student(#)	add Student's *t* density with # degrees of freedom to the graph
stopts(*cline_options***)**	affect rendition of the Student's *t* density
Add plots	
addplot(*plot***)**	add other plots to the generated graph
Y axis, X axis, Titles, Legend, Overall	
twoway_options	any options other than **by()** documented in **[G-3]** *twoway_options*

Lorenz curves

```
* Step 1: Open database
use pobreza_22.dta, clear
* Step 2: Create percentiles
drop if ictpc==.
sort ictpc
ge rnk=_n
egen nt=count(ictpc)
ge p=rnk/nt
* Step 3: Estimation
ge ar=sum(ictpc)
egen str=max(ar)
ge part =ar/str
label var part "Cumulative portion of per capita income"
```

```
* Step 4: Graph
line part p p, xlabel(0(0.1)1) ylabel(0(0.1)1) ti("Lorenz Curve: Mexico, 2022") xti(Cumulative % Population) legend(pos(11) ring(0))
```



Gini index

```
lab var ictpc “Per capita income”
sum ictpc
scalar MU = r(mean)
scalar NOBS = r(N)
sort ictpc
egen gini = sum((_n/NOBS^2)*(2/MU)*(ictpc-MU))
scalar GINI_22 = gini
di "GINI" _col(10) GINI_22

GINI     .42375082
```

Difference between Gini

```
* Step 1: Gini estimate for 1st income distribution
use pobreza_20.dta
sum ictpc
scalar MU = r(mean)
scalar NOBS = r(N)
sort ictpc
egen gini = sum((_n/NOBS^2)*(2/MU)*(ictpc-MU))
scalar GINI_20 = gini
```

```
* Step 2: Gini estimate for 2nd income distribution
use pobreza_22.dta, clear
sum ictpc
scalar MU = r(mean)
scalar NOBS = r(N)
sort ictpc
egen gini = sum((_n/NOBS^2)*(2/MU)*(ictpc-MU))
scalar GINI_22 = gini
* Step 3: Difference between Gini
di "GINI Difference" _col(20) GINI_22 - GINI_20
GINI Difference    -.01717496
```

Application with STATA

cdensity, clorenz & digini commands

cdensity command

cdensity varlist, [HSize(varname) HGroup(varname) BAND(real) BCOR(string) BORDER(int) TYPE(string) MIN(real) MAX(real) LRES(int) SRES(string) DGRA(string) SGRA(string) EGRA(string)]

hsize Household size.

hgroup Variable that conditions the analysis on a socio-demographic group.

type option "dde" to draw the derivative of the density.

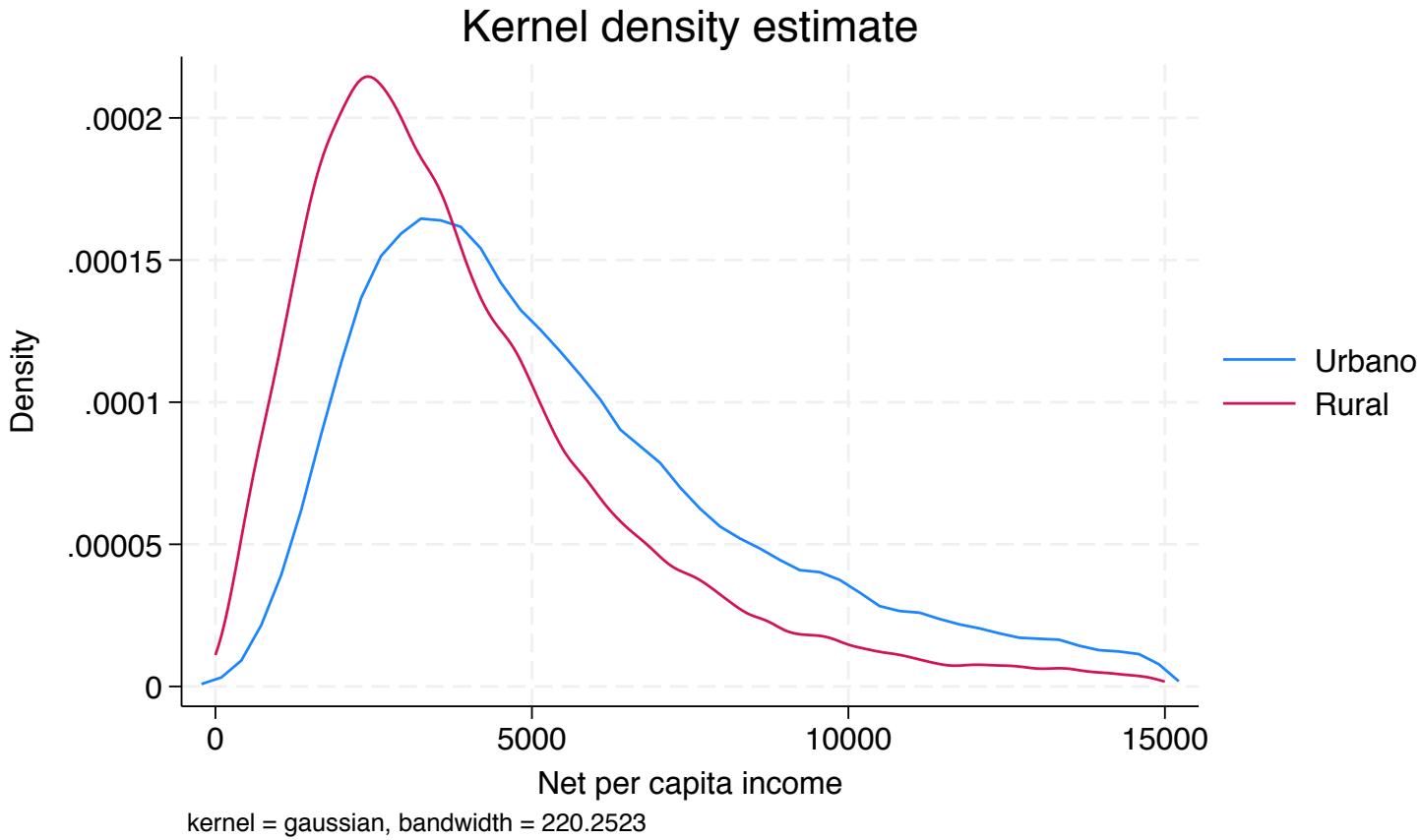
band To indicate the desired bandwidth.

min The minimum value of the range of the x axis.

max The maximum value of the range of the x axis.

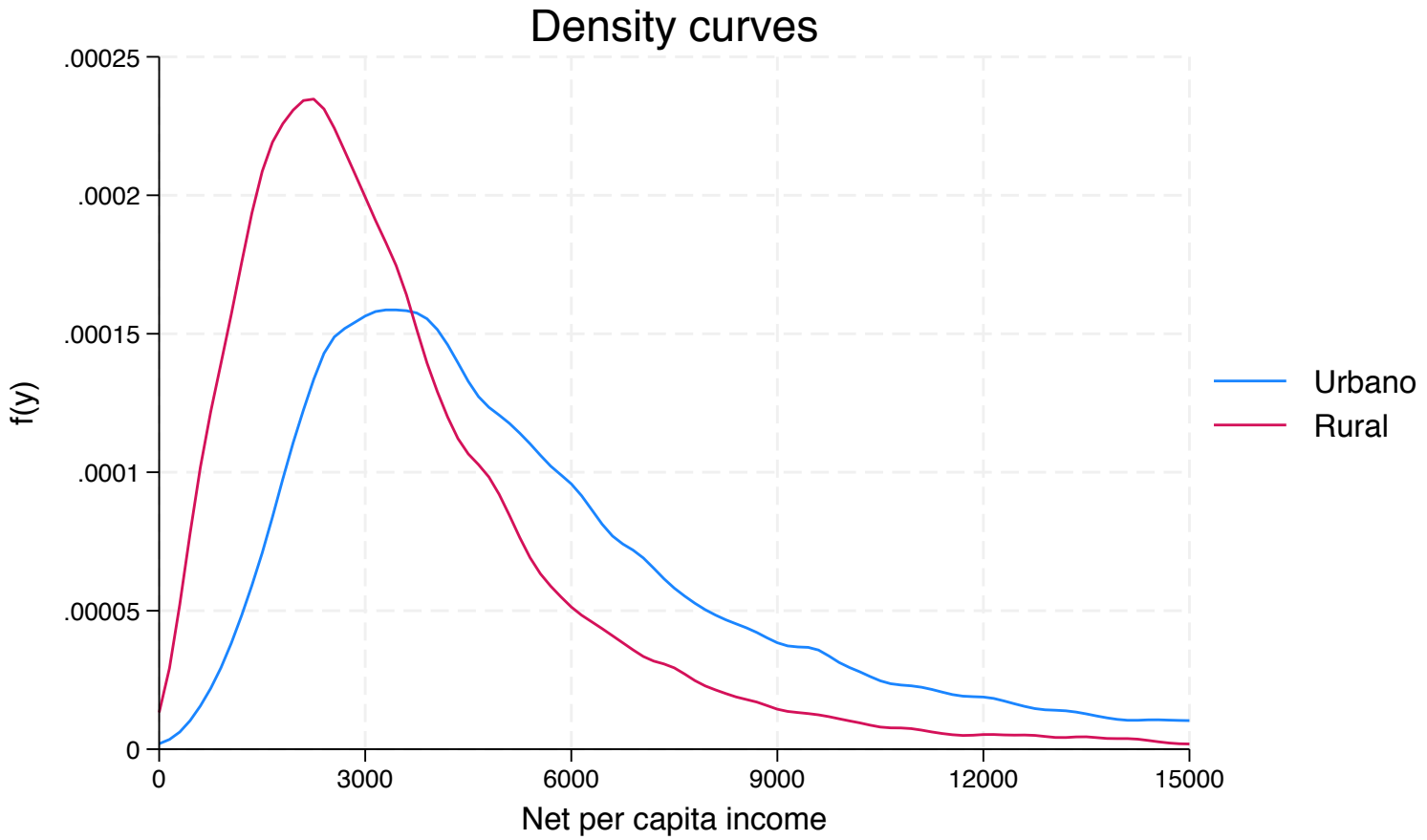
Example 1: kdensity

kdensity ictpc if ictpc<=15000 & rururb==0, kernel(gaussian) bwidth(220.2523) legend(label(1 "Urbano")) addplot(**kdensity** ictpc if ictpc<=15000 & rururb==1, kernel(gaussian) bwidth(220.2523) legend(label(2 "Rural")))



Example 1: cdensity

cdensity ictpc, hsize(factor) hgroup(rururb) min(0) max(15000) band(220.2523) xtitle(Net per capita income)



clorenz command

clorenz varlist, [HSize(varname) HGroup(varname) RANK(varname) TYPE(string) MIN(real) MAX(real) DIF(string) LRES(int) SRES(string) DGRA(string) SGRA(string) EGRA(string)]

hsize Household size.

hgroup Variable that conditions the analysis on a socio-demographic group.

rank Option to set the ranking variable (concentration curves).

type Option "dde" to draw the derivative of the density.

band To indicate the desired bandwidth.

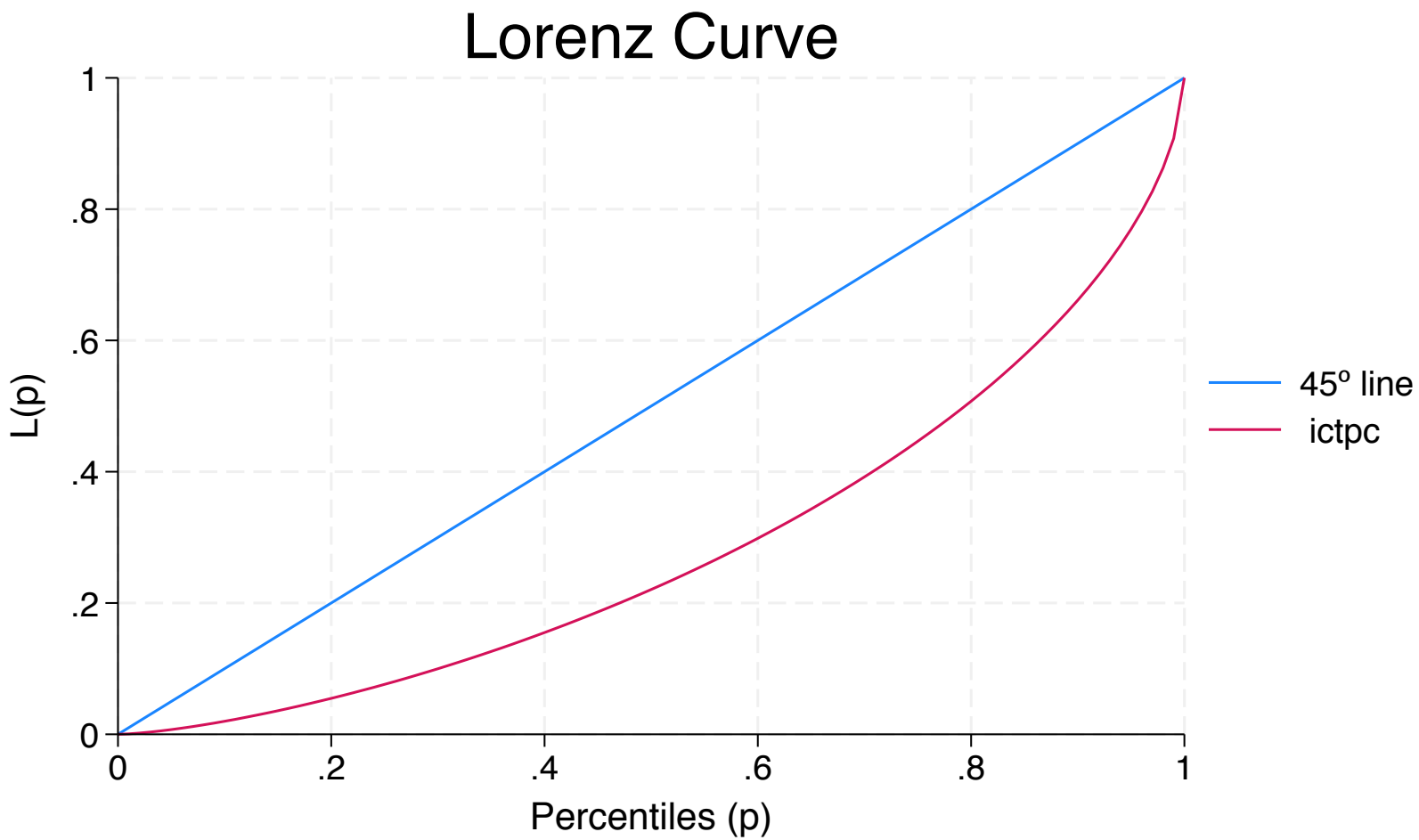
min The minimum value of the range of the x axis.

max The maximum value of the range of the x axis.

dif Difference between percentiles and lorenz (concentration) curves.

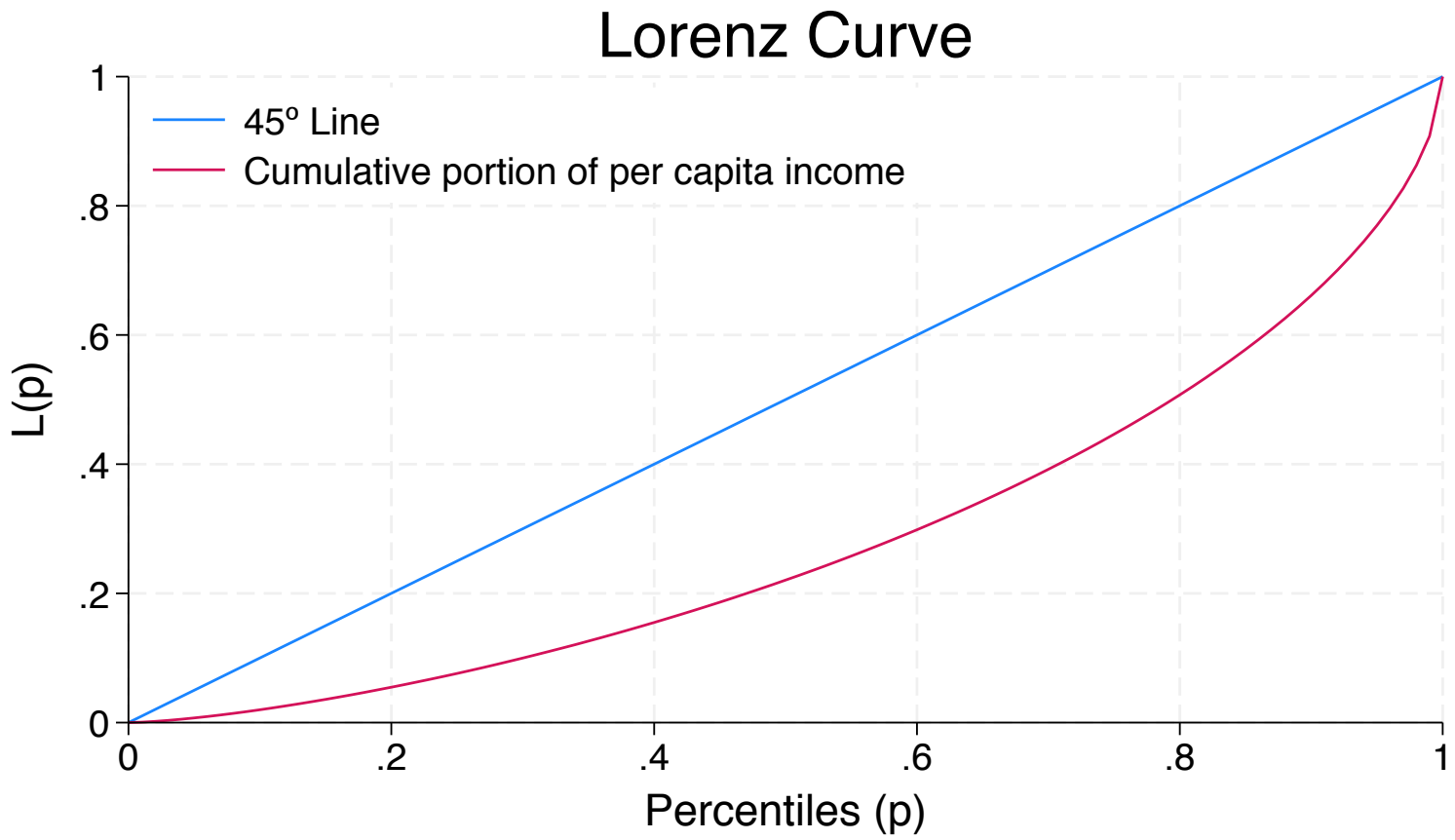
Example 2: clorenz

clorenz ictpc, hsize(factor)



Example 3: clorenz

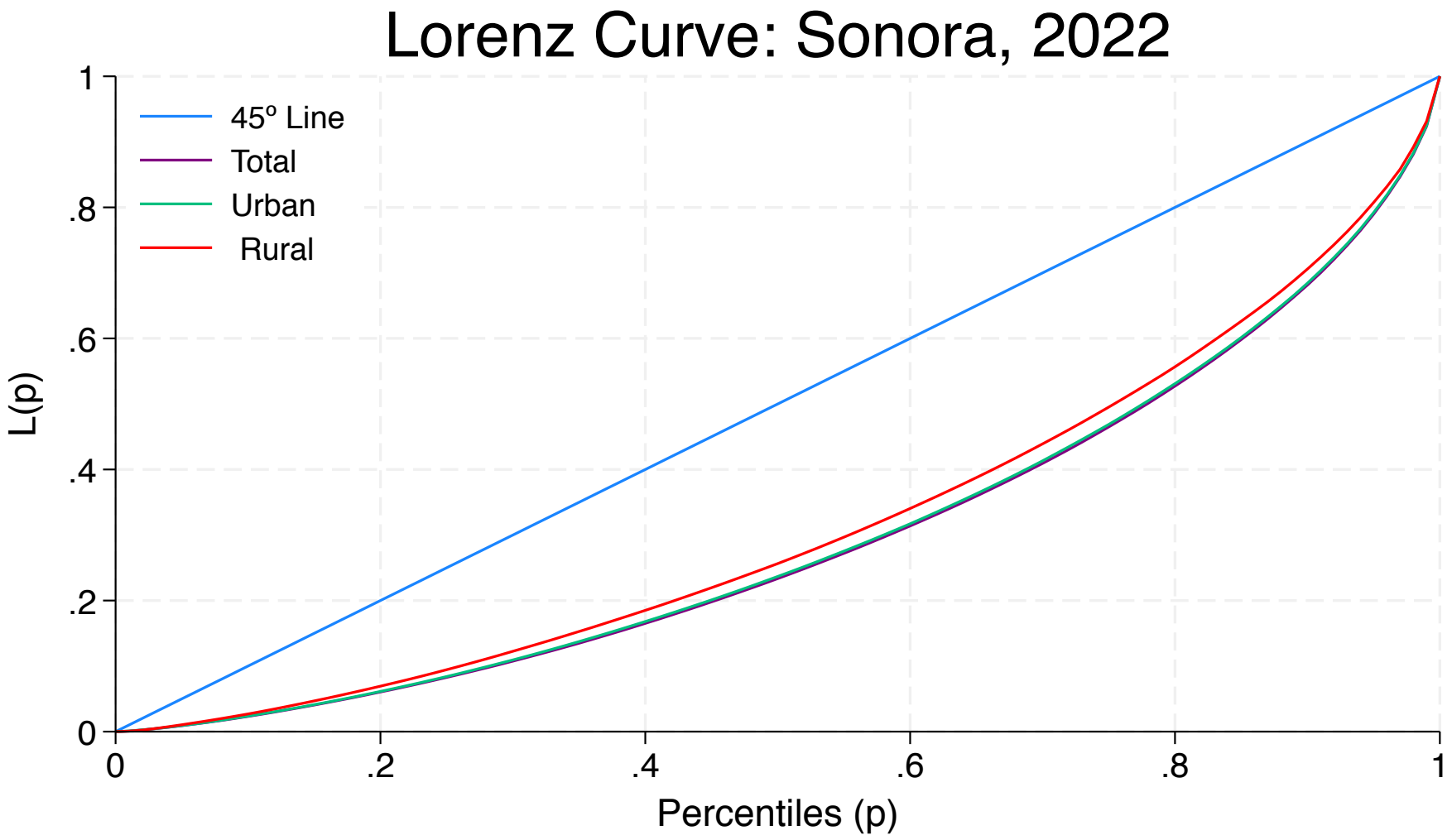
clorenz ictpc, hsize(factor) legend(label(1 "45º Line") label(2 "Cumulative portion of per capita income") pos(11) ring(0))



Example 4: clorenz

```
preserve
keep if ent==26
clorenz ictpc, hsize(factor) hgroup(rururb) legend(label(1 "45º Line") label(2 "Cumulative portion of per capita income") pos(11) ring(0)) title(Lorenz Curve: Sonora, 2022)
restore
```

Example 4: clorenz



digini command

digini varlist , [FILE1(string) FILE2(string) RANK1(varname) RANK2(varname) HSIZE1(string) HSIZE2(string) COND1(string) COND2(string) LEVEL(real) CONF(string) TEST(real)]

varlist is a list of two variables

file1 (file 2) the file used for the first (second) distribution.

hsize1 (hsize2) Household size for the first (second) distribution.

rank1 (rank2) To set the ranking variable for the first (second) distribution.

cond1 (cond2) To set a logical expression that will capture a socio-demographic group for the first (second) distribution.

level To indicate the confidence level.

test To test if the difference equals a given value.

Example 5: digini

digini ictpc ictpc, file1(pobreza_20.dta) hsize1(factor) file2(pobreza_22.dta) hsize2(factor)

Index	Estimate	Std. Err.	t	P>|t|	[95% Conf.	Interval]
GINI_Dis1	.4496415	.0019724	227.967	0.0000	.4457757	.4535073
GINI_Dis2	.4306915	.0016087	267.726	0.0000	.4275385	.4338445
diff.	-.01895	.0025453	-7.44509	0.0000	-.0239387	-.0139613

Conclusions

- DASP: Distributive Analysis Stata Package
  - Freely available (updated on 2021)
  - http://dasp.ecn.ulaval.ca
- Dialog box for all commands
- Sintaxis for commands:
  - Easy to use
  - Use help command for more information
- Twoway options are included
- Efficient performance
- Other commands to explore: lorenz (Jann, 2016)

References

- Araar, A. and Duclos, J. Y. (2007), *DASP: Distributive Analysis Stata Package*, PEP, World Bank, UNDP and Université Laval.
- Duclos, J. Y., & Araar, A. (2006). *Poverty and equity: measurement, policy and estimation with DAD* (Vol. 2). IDRC.
- Jann, B. (2016). Estimating Lorenz and concentration curves. *The Stata Journal*, *16*(4), 837-866.
- Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. Chapman & Hall.

Dra. Linda Llamas / Universidad Estatal de Sonora

Comandos clorenz, cdensity y digini y su aplicación en el análisis de la distribución del ingreso.