

Inequality: the Gini Index

Goal

This exercise introduces the Gini index, which is one of the most commonly used income inequality indicators.

Activity

Calculate the Gini index on total disposable income for Finland and the US in 2000, after bottom-coding disposable income at 1 percent of its equivalised mean and top-coding at 10 times the unequivalised median.

Guidelines

- Prepare the data as you did in the previous exercise (drop observations with missing or zero *dpi*).
- In the previous exercise you have seen two different methods of dealing with extreme values, trimming and winsorising (or bottom-/top-coding). The LIS key figures are calculated using a particular type of bottom-/top-coding, which we will replicate in this exercise. The bottom-coding is carried out after the equivalisation of income (on the equivalised income distribution), while the top-coding is carried out before (on the unequivalised distribution) in the following way:
 - at the bottom of the distribution, equivalised income is bottom-coded at 1 percent of its equivalised mean, i.e., all observations for which equivalised income is lower than 1% of the average equivalised income are set to that value.
 - at the top of the distribution, income is top-coded at 10 times the unequivalised median, i.e., all observations for which unequivalised income (or *dpi*) is higher than 10 times the median unequivalised income are set to that value.
 - Use the same frame as the one created in the previous exercise and adapts it to comply with the LIS definition.
- **HELP:** Here is a SAS algorithm to calculate the Gini coefficient. The variable **wt** and **ey** used in this algorithm have been exactly defined the same than for the previous exercises.

```
PROC SORT DATA=gini;
  BY ey;
RUN;

DATA gini (KEEP=gini);
  IF _N_ = 1 THEN
    DO UNTIL (last);
      SET gini END=last;
      swt + wt ;
      swtey + (wt*ey) ;
    END;
  SET gini END=eof;
  IF _N_ = 1 THEN
```

```
DO;
  prewt = 0 ;
  preey = 0 ;
  up     = 0 ;
  sum    = 0 ;
END;
cwt + wt ;
cwtey + (ey*wt) ;
pcwt = cwt / swt * 100 ;
pcwtey = cwtey / swtey * 100 ;
up=(pcwt-prewt)*(pcwtey+preey);
sum + up ;
prewt = pcwt ;
preey = pcwtey ;

RETAIN prewt preey ;
IF eof THEN
DO;
  gini = 1 - (sum / 10000) ;
  OUTPUT;
END;
RUN;
```

Program

```
OPTIONS NOSOURCE NONOTES NOFMterr NODATE NOCENTER LABEL NONUMBER LS=MAX
PS=MAX ;
```

```
%MACRO prep ;
  DATA prep (DROP=hweight);
  SET &&pi.h (KEEP=hweight d4 dpi);
  IF dpi in (. 0) THEN DELETE;
  ey = &var / SQRT(d4);
  wt = hweight * d4;
  RUN;
PROC UNIVARIATE DATA=prep NOPRINT;
  VAR ey dpi ;
  WEIGHT wt ;
  OUTPUT OUT=temp MEAN=aveey MEDIAN=medey medpi ;
RUN ;
DATA _NULL_;
  SET temp;
  CALL SYMPUT("a",aveey);
  CALL SYMPUT("m",medpi);
RUN;
DATA prep (DROP= botlin toplin);
  SET prep ;
  botlin = 0.01 * &a;
  IF ey < botlin THEN ey = botlin ;
  toplin = 10 * &m;
  IF &var > toplin THEN ey = (toplin/(SQRT(d4))) ;
RUN ;
PROC SORT DATA=prep;
  BY ey ;
RUN ;
%MEND prep;
%MACRO Gini ;
  DATA Gini (KEEP=gini) ;
  IF _N_ = 1 THEN
    DO UNTIL (last) ;
      SET prep END=last;
      swt + wt ;
      swtey + (wt*ey) ;
    END ;
  SET prep END=eof;
  IF _N_ = 1 THEN
    DO ;
      prewt = 0 ;
      preey = 0 ;
      up = 0 ;
      sum = 0 ;
    END ;
  cwt + wt ;
  cwtey + (ey*wt);
  pcwt = cwt / swt * 100;
  pcwtey = cwtey / swtey * 100;
  up = (pcwt-prewt) * (pcwtey+preey) ;
  sum + up ;
  prewt = pcwt ;
```

```
preey = pcwtey ;

RETAIN prewt preey ;
IF eof THEN
  DO ;
    gini=1-(sum / 10000) ;
    OUTPUT ;
  END ;
RUN;
TITLE &pi - &var ;
PROC MEANS DATA=Gini MEAN;
RUN;
%MEND Gini ;

%Let var = dpi ; %LET pi = us00 ;
%Prep
%Gini
%LET pi = fi00 ;
%Prep
%Gini
```

Results

	Gini
US 2000	<i>0.36823</i>
Finland 2000	<i>0.24621</i>

Comments

- The Gini index ranges between 0 and 1, with inequality increasing with an increasing index. A value of 0 means there is a completely equal distribution of income; a 1 refers to the extreme situation of one household holding the total population income, and all the rest having no income at all.
- As expected, inequality is much larger in the US than in Finland.
- To see the Ginis for all LIS datasets online, go to <http://www.lisproject.org/key-figures/key-figures.htm>.