

APPENDIX 1: Glossary of statistical terms

Coefficient of variation (CV)

The coefficient of variation is a measure of relative dispersion, calculated by expressing the standard deviation as a percentage of the mean. The greater the coefficient of variation, the greater the variability of income.

Gini Coefficient

The Gini Coefficient is calculated as a ratio of the areas defined by the Lorenz curve. A Gini Coefficient of 1 means that one member of the population earns all the income ("perfect concentration"). A Gini Coefficient of 0 means that every member of the population earns the same income ("perfect equality").

Mean

Commonly known simply as "average", it is defined as the total of a distribution of values divided by the number of values.

Median

The mid-point in a distribution of values which has been arranged in size order, also known as the 50th percentile. In an analysis of incomes, it represents the earnings of a "typical" member of the population (i.e. half the population earns less than the median). In a Lorenz curve diagram, the median can be identified by the 50% mark on the horizontal x-axis.

Lorenz curve

The Lorenz curve was developed by Max O. Lorenz as a graphical representation of income distribution: "Methods for measuring the concentration of wealth", *Publications of the American Statistical Association* 9 (1905): pp. 209-219. A Lorenz curve plots cumulative percentage incomes against cumulative percentage population. It represents a series of statements such as: "the bottom 20% earn 10% of total income"; "the bottom 80% earn 60% of total income" (= "the top 20% earn 40% of total income"). The more "sloped" the curve is, the more unequal is the distribution of wealth in a given population. The Lorenz curve is used to calculate the Gini Coefficient.

Population

The complete set of people (or any collection of items) under consideration.

Sample

A sub-set of the population that is selected for research.

Standard deviation

The standard deviation measures how tightly the various values are clustered around the mean in a set of data. When the data points are “bunched together” the standard deviation is small.

T-statistic

The t-statistic tests for the ratio of a coefficient to its standard error.

***Significant difference at the 1% level

** Significant difference at the 5% level

* Significant difference at the 10% level