

NAME: S.HARSHAVARTHAN

REG.NO:121012012738

COURSE NAME : DATA VISUALIZATION AND TEXT ANALYSIS

TOPIC : LORENZ CURVE

**Lorenz curve**

In statistics, the Lorenz curve is a graphical representation of the cumulative distribution function (CDF) of a random variable. It is used to represent the distribution of a variable across a population and is commonly used in income inequality studies, risk analysis, and survival analysis.

The Lorenz curve is a graphical tool that is widely used to characterize the concentration of a measure in a population, such as wealth.

The Lorenz curve is typically created by plotting the cumulative proportion of the variable being studied (e.g., income, wealth, failure times) against the cumulative proportion of the population. For example, if we are studying income inequality, we would plot the cumulative proportion of total income earned against the cumulative proportion of the population ranked by income.

To construct the Lorenz curve, we first need to sort the population by the variable being studied (e.g., income) from lowest to highest. We then calculate the cumulative proportion of the variable and the cumulative proportion of the population for each income level. The Lorenz curve is then created by plotting these cumulative proportions on a graph with the horizontal axis representing the cumulative proportion of the population and the vertical axis representing the cumulative proportion of the variable being studied.

The diagonal line on the graph represents perfect equality, where each percentage of the population earns or owns the same percentage of total income or wealth. The more the Lorenz curve bows away from the diagonal line, the greater the inequality in the distribution of income or wealth. For example, if the Lorenz curve bows sharply to the right of the diagonal line, it indicates that a small percentage of the population owns a large proportion of the total income or wealth.

The Gini coefficient can also be calculated from the Lorenz curve to provide a single summary measure of income or wealth inequality. The Gini coefficient is the area between the Lorenz curve and the diagonal line, divided by the total area under the diagonal line. The Gini coefficient ranges from 0, indicating perfect equality, to 1, indicating perfect inequality. A Gini coefficient of 0.5 or higher is generally considered to indicate high inequality.

Consider a case of data analysis where we have household income data i.e. households and their income. We collected a list of all household incomes in the economy in order from low to high

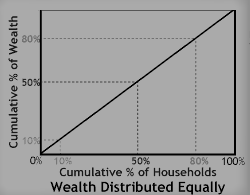
Plotting points on a graph:

Along the x-axis, we measure the[**Cumulative Percent**](https://dictionary.apa.org/cumulative-percentage)**of Households.**

Along the y-axis, we measure the [**Cumulative Percent**](https://dictionary.apa.org/cumulative-percentage)**of Wealth**.

After plotting all the points, you will have the **Lorenz curve**

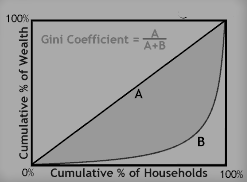
The Lorenz curve for a completely even income distribution (every family has the same income) would lie on the diagonal (as shown below). If all households have the same income, the Lorenz curve is a straight diagonal line, called the line of equality.



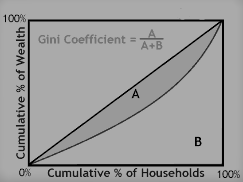
In the above figure, the Lorenz curve is a diagonal line at 45 degrees i.e.

10 % of the households have 10% of the wealth, 50 % of the households have 50% of the wealth, 80 % of the households have 80% of the wealth, and so on.

If there is any inequality in size, then the Lorenz curve falls below the line of equality (shown below).



Consider another case of inequality (shown below), the degree of inequality is lesser here.



The **Gini coefficient** is a numerical measure of inequality based on the Lorenz curve.

Curve B, is Lorenz curve (this is a case of income inequality), and it lies below the line of equality (Curve A).

In the above two figures, you can spot the formula of the Gini-coefficient, A & B denote area as described below:

The Gini Coefficient for a Lorenz curve is defined as the ratio of the area between the Lorenz curve and the diagonal to the triangle below the diagonal line. In the graph above, the area between the Lorenz curve and the diagonal is shown in grey. The area of the triangle is shown as the area of the lower triangle formed by the line of equality.

**The Gini coefficient is then a number that falls in the range from 0 to 1.**

For a completely equal income distribution,the area is zero and the Gini coefficient is zero.

For the completely concentrated distribution, the area is the same as the triangle so the Gini coefficient is one (If all of the income was received by one family the Lorenz would be a zero for every point on the horizontal axis except 100% where it would jump to 100% on the vertical axis. This would be the case of complete concentration of the income distribution and would represent complete income inequality).

**Gini should not be mistaken for an absolute measurement of income or wealth.** A high-income country and a low-income one can have the same Gini coefficient, as long as incomes are distributed similarly within each country.

Overall, the Lorenz curve is a useful tool for visualizing and analysing the distribution of a variable across a population, and it can provide valuable insights into patterns of inequality or risk in a given population.