

統計学入門 第4回

早稲田大学政治経済学部
西郷 浩

1

本日の目標

- 不均等度の測定
 - 必要性
 - ローレンツ曲線
 - ローレンツ曲線の応用
 - ジニ係数
 - 付録: ジニ係数の数理

2

(不)均等度の測定の必要性

- 均等性
 - 社会の要求
 - 法の下での平等
 - 一票の格差(不均等)
 - 機会の均等
 - 出発点は同一条件であるべき。
 - 結果の均等との関連
- 均等の測定
 - 不均等である=バラツキが大きい
 - 均等であることに焦点をあてた特別の測定方法がある。

3

ローレンツ曲線(1)

- ローレンツ曲線作成の手順
 1. データを昇順に並べ替える。
 - $x_1, x_2, \dots, x_n$
↓ sort
 $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$
 2. 個体数・変数それぞれの累積和を計算する。
 3. 累積和を総和で除して相対化する。

4

ローレンツ曲線(2)

表1: ローレンツ曲線作成のための計算表

昇順	個体数	累積和 (累積度数)	相対累積和 (累積相対度数)	変数	累積和	相対累積和
(1)	1	1	$1/n$	$x_{(1)}$	$x_{(1)}$	$x_{(1)}/T$
(2)	1	2	$2/n$	$x_{(2)}$	$x_{(1)} + x_{(2)}$	$(x_{(1)} + x_{(2)})/T$
...	...	...	...	...	...	...
(n)	1	n	n/n	$x_{(n)}$	T	T/T
合計	n			T		

5

ローレンツ曲線(3)

4. 以下のように $a_{(i)}$ と $b_{(i)}$ を定める。

- $a_{(0)} = 0, b_{(0)} = 0$
- $i=1, 2, \dots, n$ については、
 $a_{(i)} = \{(i) \text{ に対応する個体数の相対累積和}\}$
 $b_{(i)} = \{(i) \text{ に対応する変数の相対累積和}\}$

5. 座標平面に以下の要領で線を描く。

- $(a_{(i-1)}, b_{(i-1)})$ と $(a_{(i)}, b_{(i)})$ とを直線で結ぶ。
- $(0, 0)$ と $(1, 1)$ とを直線で結ぶ。

6

ローレンツ曲線(4)

■ 数値例

■ ケースA(完全均等)

- $x_1=2, x_2=2, x_3=2, x_4=2, x_5=2$

■ ケースB(不均等)

- $x_1=0, x_2=2, x_3=4, x_4=2, x_5=2$

■ ケースC(独占)

- $x_1=0, x_2=10, x_3=0, x_4=0, x_5=0$

7

ローレンツ曲線(5)

表2: ケースBについてのローレンツ曲線作成のための計算表

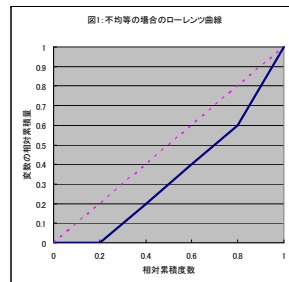
昇順	個体数	累積和 (累積度数)	相対累積和 (累積相対度数)	変数	累積和	相対累積和
(1)	1	1	$1/5$	0	0	$0/10$
(2)	1	2	$2/5$	2	2	$2/10$
(3)	1	3	$3/5$	2	4	$4/10$
(4)	1	4	$4/5$	2	6	$6/10$
(5)	1	5	$5/5$	4	10	$10/10$
合計	5			10		

8

ローレンツ曲線(6)

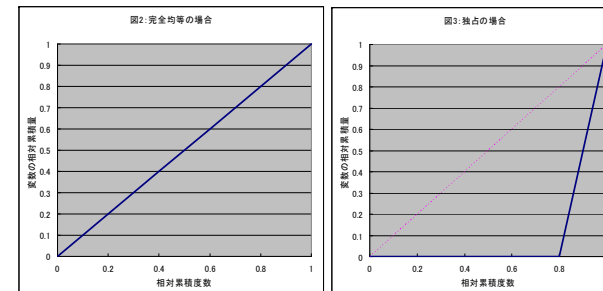
■ 線の名称

- 青い線:
ローレンツ曲線
- ピンクの点線:
均等線



9

ローレンツ曲線(7)



10

ローレンツ曲線(8)

■ 数値例からの観察

- ローレンツ曲線の位置は、均等線よりも下になる。
- 変数 x の状態: 均等 $\longleftrightarrow$ 不均等
ローレンツ曲線: 上 $\longleftrightarrow$ 下
- したがって、ローレンツ曲線が下位にある分布ほど不均等の程度が大きい。
 - 不均等の程度が大きい $\Leftrightarrow$ 散らばりが大きい

11

ローレンツ曲線(9)

■ 使い方

- 異なる集団間の均等度の比較
- 同一集団内の異なる変数(異なる時点間を含む)の均等度の比較

12

ローレンツ曲線の応用(1)

■ 参議院議員選挙(選挙区) H22, H10

■ 一票の格差

■ 都道府県別

$$x = (\text{議員定数} / \text{有権者数})$$

- 新聞等では (有権者数/議員定数) が使われる。
- 有権者を対象集団とするので、 x を使う。

13

ローレンツ曲線の応用(2)

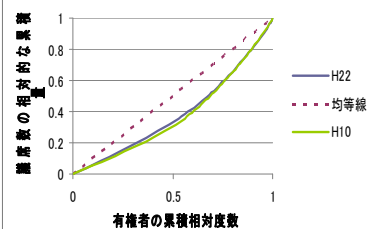
■ 注意点

- 対象集団: 全国の有権者
 - 都道府県ではない。
 - x の (47個の) 合計を47で除しても、 x の全国平均にならない。
 - x の全国平均は以下のように計算する。
 - $\Sigma (\text{ある県の有権者数} \times \text{その県の} x) / \text{全有権者数}$
 - $= \Sigma (\text{ある県の有権者数構成比} \times \text{その県の} x)$
 - これは加重平均と呼ばれる。
- 横軸の累積相対度数は、有権者数で計算する。

14

ローレンツ曲線の応用(3)

図4: 有権者1000人当たり議席数のローレンツ曲線



15

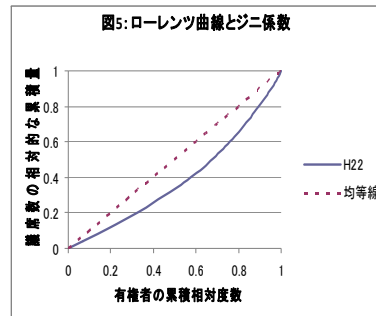
不均等度の尺度: ジニ係数(1)

■ ジニ係数 $G/$

- $2 \times (\text{ローレンツ曲線と均等線の面積})$
- 均等 不均等
 $0 \leftarrow G/ \rightarrow 1$
- 所得分布などでは、0.5より大きいと不均等度が大きいとみなされる。
 - 対象などによって一概にいけない。

16

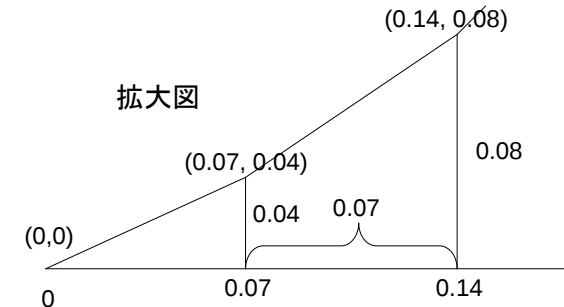
不均等度の尺度:ジニ係数(2)



17

不均等度の尺度:ジニ係数(3)

図5: ローレンツ曲線の下側の面積の計算



18

不均等度の尺度:ジニ係数(4)

■ G/I の変化

- 平成10年: 0.25 → 平成22年: 0.23

■ G/I は以下のようにも計算できる。

- $$GI = \frac{1}{2\bar{x}} \left(\frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n |x_i - x_j| \right)$$

個体の変数間の差の絶対値の平均値

- G/I が散らばりの尺度のひとつであることが分かる。

19

G/I の数理(1)

$$y_{(i)} \equiv \frac{x_{(i)}}{\sum_{i=1}^n x_{(i)}} = \frac{x_{(i)}}{\sum_{i=1}^n x_i} = \frac{x_{(i)}}{n\bar{x}}$$

とすれば、

$$a_{(j)} = j/n$$

$$b_{(j)} = \sum_{i=1}^j x_{(i)} / \sum_{i=1}^n x_i = \sum_{i=1}^j y_{(i)}$$

20

GI の数理(2)

$a_{(0)} = b_{(0)} = y_{(0)} = 0$ とすれば、

$$\begin{aligned} U &= \frac{1}{2} \sum_{j=1}^n (b_{(j)} + b_{(j-1)}) (a_{(j)} - a_{(j-1)}) \\ &= \frac{1}{2n} \sum_{j=1}^n (b_{(j)} + b_{(j-1)}) \\ &= \frac{1}{2n} \sum_{j=1}^n \left(\sum_{i=0}^j y_{(i)} + \sum_{i=0}^{j-1} y_{(i)} \right) \end{aligned}$$

21

GI の数理(3)

$$\begin{aligned} &= \frac{1}{2n} \sum_{j=1}^n \left(2 \sum_{i=1}^j y_{(i)} - y_{(j)} \right) \\ &= \frac{1}{n} \sum_{j=1}^n \sum_{i=1}^j y_{(i)} - \frac{1}{2n} \sum_{j=1}^n y_{(j)} \end{aligned}$$

22

GI の数理(4)

$$\begin{aligned} &\sum_{j=1}^n \sum_{i=1}^j y_{(i)} \\ &= y_{(1)} \\ &\quad + y_{(1)} + y_{(2)} \\ &\quad + y_{(1)} + y_{(2)} + y_{(3)} \\ &\quad + \cdots \\ &\quad + y_{(1)} + y_{(2)} + y_{(3)} + \cdots + y_{(n)} \\ &= n y_{(1)} + (n-1) y_{(2)} + \cdots + 2 y_{(n-1)} + y_{(n)} \\ &= \sum_{j=1}^n (n-j+1) y_{(j)} \end{aligned}$$

23

GI の数理(5)

$$\begin{aligned} U &= \frac{1}{n} \sum_{j=1}^n \sum_{i=1}^j y_{(i)} - \frac{1}{2n} \sum_{j=1}^n y_{(j)} \\ &= \frac{1}{n} \sum_{j=1}^n (n-j+1) y_{(j)} - \frac{1}{2n} \sum_{j=1}^n y_{(j)} \\ &= \frac{1}{n} \sum_{j=1}^n \left(n + \frac{1}{2} - j \right) y_{(j)} \end{aligned}$$

24

GI の数理(6)

$$GI = 1 - 2U$$

$$= 1 - \frac{2}{n} \sum_{j=1}^n \left(n + \frac{1}{2} - j \right) y_{(j)}$$

$$= 1 - \frac{2}{n} \sum_{j=1}^n \left(n + \frac{1}{2} - j \right) \frac{x_{(j)}}{n\bar{x}}$$

$$= \frac{1}{n\bar{x}} \left\{ n\bar{x} - \sum_{j=1}^n \frac{2n+1-2j}{n} x_{(j)} \right\}$$

25

GI の数理(7)

$$= \frac{1}{n\bar{x}} \left\{ \sum_{j=1}^n x_{(j)} - \sum_{j=1}^n \left(\frac{2n+1-2j}{n} \right) x_{(j)} \right\}$$

$$= \frac{1}{n\bar{x}} \sum_{j=1}^n \left\{ 1 - \left(\frac{2n+1-2j}{n} \right) \right\} x_{(j)}$$

$$= \frac{1}{n\bar{x}} \sum_{j=1}^n \frac{2j-n-1}{n} x_{(j)}$$

$$= \frac{1}{n^2\bar{x}} \sum_{j=1}^n (2j-n-1) x_{(j)}$$

26

GI の数理(8)

ところで、

$$\sum_{i=1}^n \sum_{j=1}^n |x_i - x_j| = \sum_{i=1}^n \sum_{j=1}^n |x_{(i)} - x_{(j)}|$$

$$\begin{aligned} &= |x_{(1)} - x_{(1)}| + |x_{(1)} - x_{(2)}| + |x_{(1)} - x_{(3)}| + \cdots + |x_{(1)} - x_{(n)}| \\ &\quad + |x_{(2)} - x_{(1)}| + |x_{(2)} - x_{(2)}| + |x_{(2)} - x_{(3)}| + \cdots + |x_{(2)} - x_{(n)}| \\ &\quad + \cdots \\ &\quad + |x_{(n)} - x_{(1)}| + |x_{(n)} - x_{(2)}| + |x_{(n)} - x_{(3)}| + \cdots + |x_{(n)} - x_{(n)}| \end{aligned}$$

27

GI の数理(9)

$$\begin{aligned} &= 0 - (x_{(1)} - x_{(2)}) - (x_{(1)} - x_{(3)}) - \cdots - (x_{(1)} - x_{(n)}) \\ &\quad + (x_{(2)} - x_{(1)}) + 0 - (x_{(2)} - x_{(3)}) - \cdots - (x_{(2)} - x_{(n)}) \\ &\quad + \cdots \\ &\quad + (x_{(n)} - x_{(1)}) + (x_{(n)} - x_{(2)}) + (x_{(n)} - x_{(3)}) + \cdots + 0 \\ &= 2(1-n)x_{(1)} + 2(3-n)x_{(2)} + 2(5-n)x_{(3)} + \cdots + 2(n-1)x_{(n)} \\ &= 2 \sum_{j=1}^n (2j-n-1)x_{(j)} \end{aligned}$$

28



GI の数理(10)

したがって、

$$\begin{aligned} GI &= \frac{1}{2n^2\bar{x}} \sum_{i=1}^n \sum_{j=1}^n |x_i - x_j| \\ &= \frac{1}{2\bar{x}} \left(\frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n |x_i - x_j| \right) \end{aligned}$$