統計学入門 第4回

早稲田大学政治経済学部
西郷 浩

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

（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。

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。
     - $x_1, x_2, \ldots, x_n$
       ↓sort
       $x_{(1)} \leqq x_{(2)} \leqq \ldots \leqq x_{(n)}$
  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。

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		

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,…, n については、
     $a_{(i)}$ = {(i) に対応する個体数の相対累積和}
     $b_{(i)}$ = {(i) 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。

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

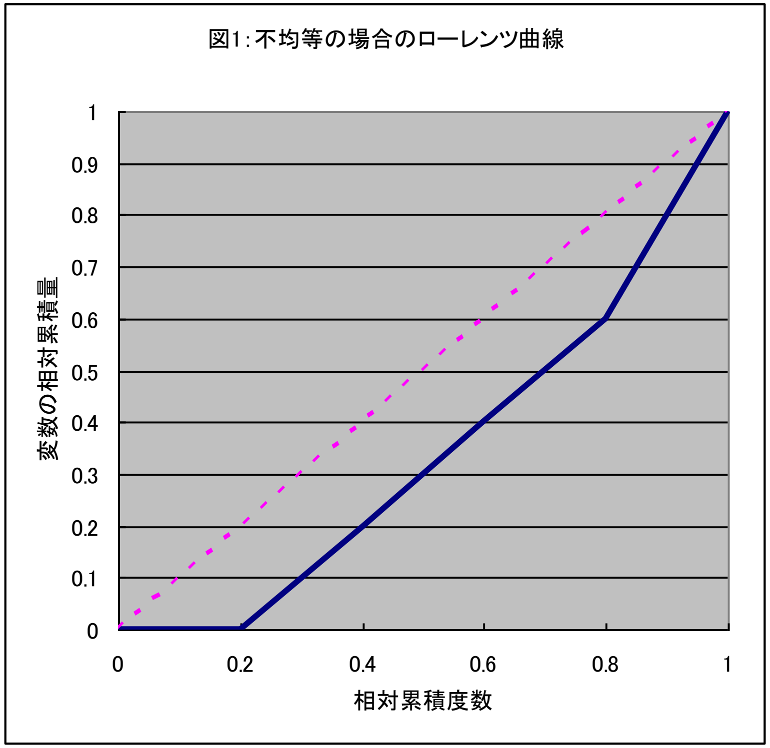
ローレンツ曲線(5)

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		

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



ローレンツ曲線(7)

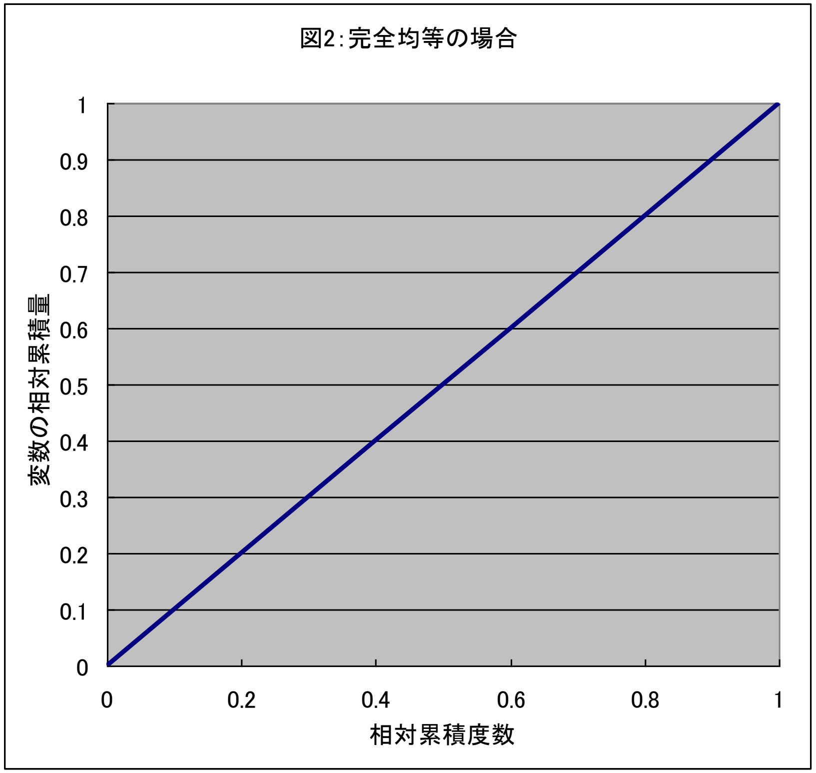


図2:完全均等の場合
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
変数の相対累積量
0
0.2
0.4
0.6
0.8
1
相対累積度数

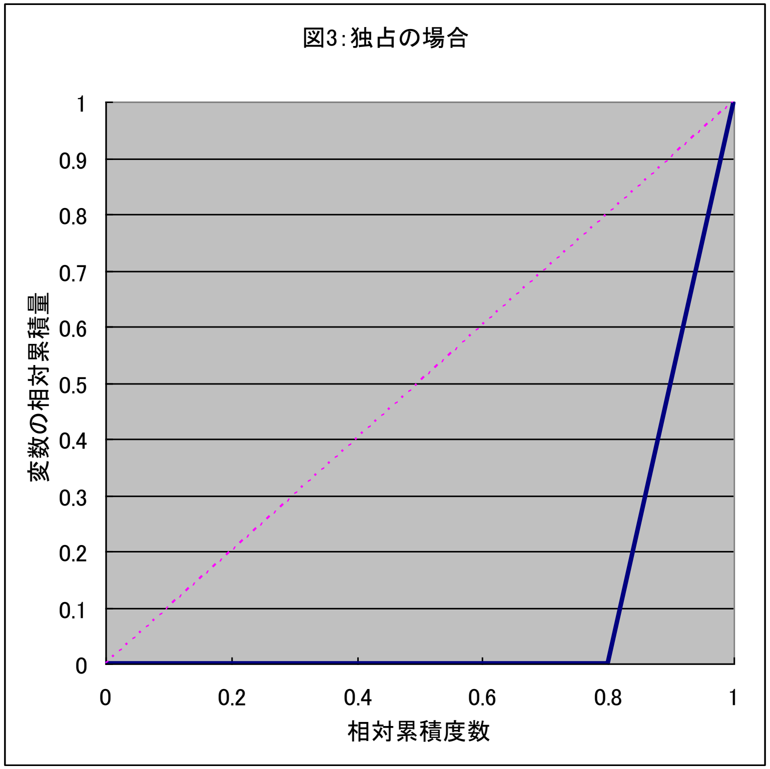


図3:独占の場合
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
変数の相対累積量
0
0.2
0.4
0.6
0.8
1
相対累積度数

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 ⟷ 不均等
    ローレンツ曲線：　上 ⟷ 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⇔散らばりが大きい

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

ローレンツ曲線の応用(1)

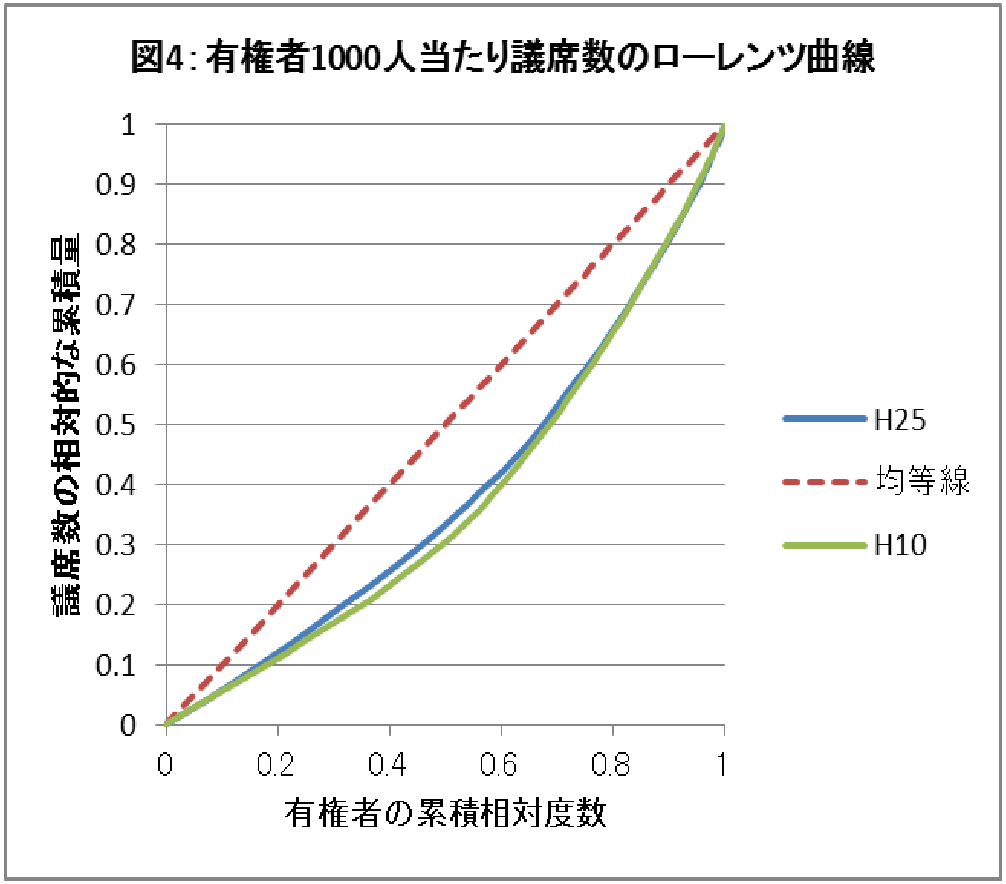
- 参議院議員選挙（選挙区）H25, H10
  - 一票の格差
    - 都道府県別

      x = (議員定数/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。

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。
      - Σ(ある県の有権者数×その県のx)/全有権者数
        = Σ(ある県の有権者数構成比×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。

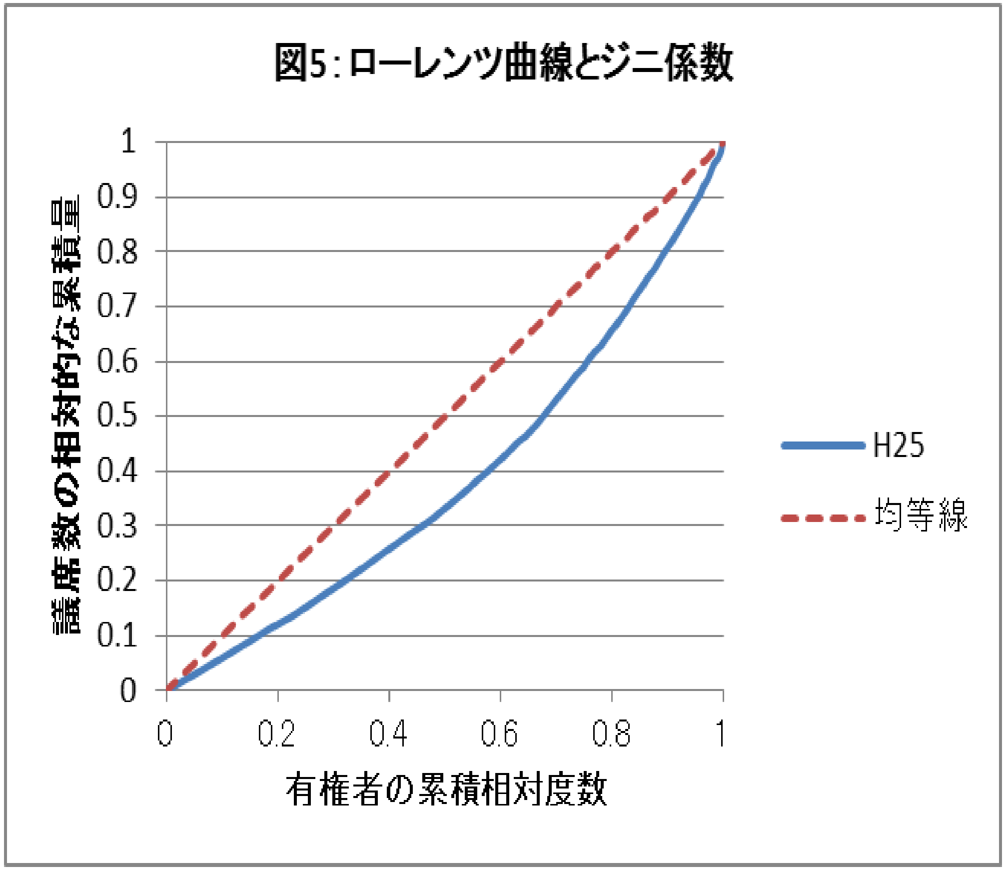
ローレンツ曲線の応用(3)



不均等度の尺度：ジニ係数(1)

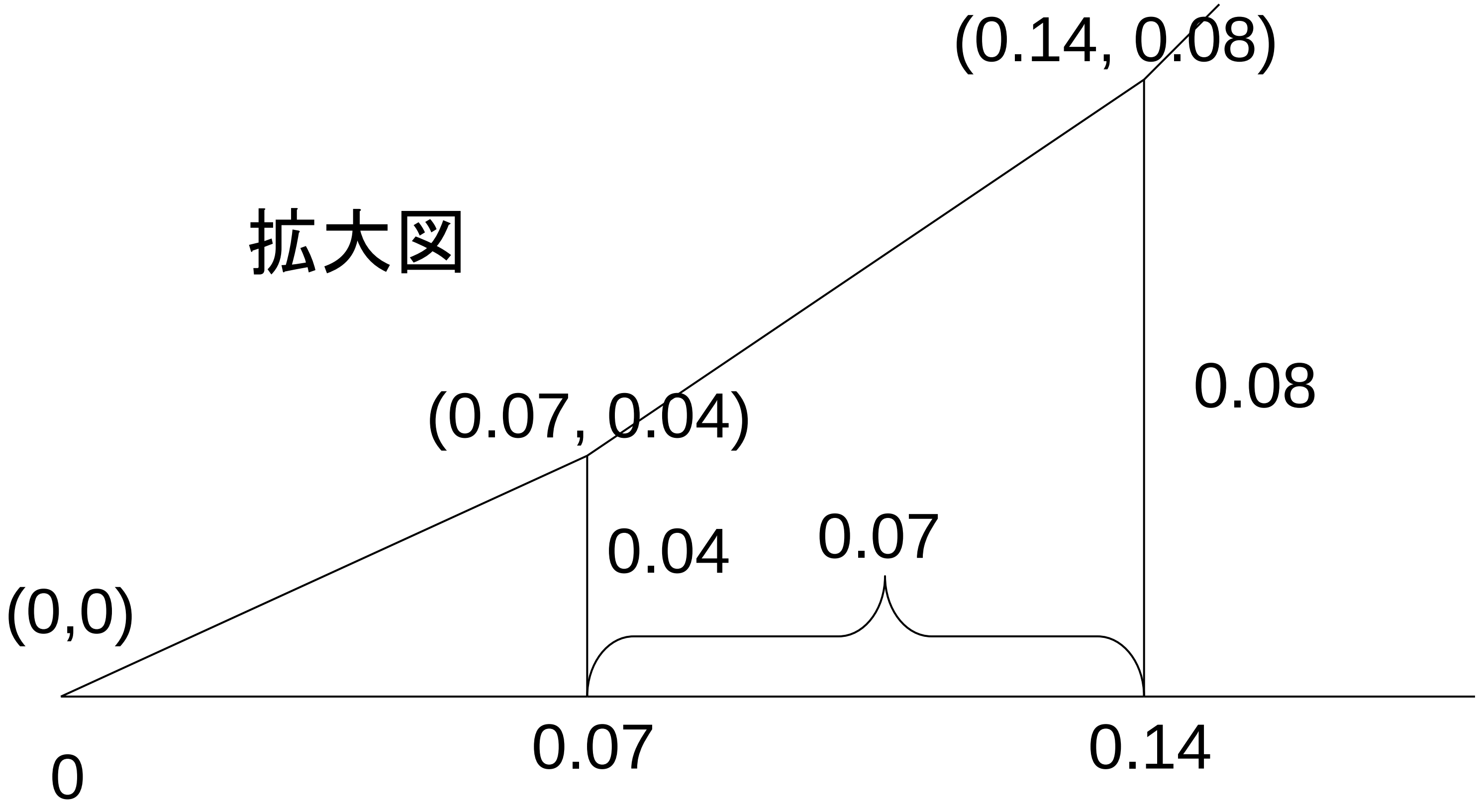
- ジニ係数 *GI*
  - ２×（ローレンツ曲線と均等線の面積）
  - 均等　　　　不均等
    $0 \longleftarrow GI \longrightarrow 1$
  - 所得分布などでは、0.5より大きいと不均等度が大きいとみなされる。
    - 対象などによって一概にいえない。

不均等度の尺度：ジニ係数(2)



不均等度の尺度：ジニ係数(3)

図6：ローレンツ曲線の下側の面積 の計算



不均等度の尺度：ジニ係数(4)

- *GI*の変化
  - 平成10年：0.25 → 平成25年：0.23
- *GI* は以下のようにも計算できる。
  - $$GI = \frac{1}{2\bar{x}}\left(\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left|x_i - x_j\right|\right)$$

    個体の変数間の差の絶対値の平均値
  - *GI* が散らばりの尺度のひとつであることが分かる。

GI 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、

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

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

GI の数理(2)

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

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

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

$$= \frac{1}{2n}\sum_{j=1}^{n}\left(\sum_{i=0}^{j} y_{(i)} + \sum_{i=0}^{j-1} y_{(i)}\right)$$

GI の数理(3)

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

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

GI の数理(4)

$$\sum_{j=1}^{n}\sum_{i=1}^{j} y_{(i)}$$

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

$$= n\,y_{(1)} + (n-1)y_{(2)} + \cdots 2y_{(n-1)} + y_{(n)}$$

$$= \sum_{j=1}^{n}(n-j+1)y_{(j)}$$

GI の数理(5)

$$U=\frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{j}y_{(i)}-\frac{1}{2n}\sum_{j=1}^{n}y_{(j)}$$

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

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

GI の数理(6)

$$GI = 1 - 2U$$

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

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

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

GI の数理(7)

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

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

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

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

GI の数理(8)

ところで、

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

$$= \left|x_{(1)} - x_{(1)}\right| + \left|x_{(1)} - x_{(2)}\right| + \left|x_{(1)} - x_{(3)}\right| + \cdots + \left|x_{(1)} - x_{(n)}\right|$$

$$+ \left|x_{(2)} - x_{(1)}\right| + \left|x_{(2)} - x_{(2)}\right| + \left|x_{(2)} - x_{(3)}\right| + \cdots + \left|x_{(2)} - x_{(n)}\right|$$

$$+ \cdots$$

$$+ \left|x_{(n)} - x_{(1)}\right| + \left|x_{(n)} - x_{(2)}\right| + \left|x_{(n)} - x_{(3)}\right| + \cdots + \left|x_{(n)} - x_{(n)}\right|$$

GI の数理(9)

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

GI の数理(10)

したがって、

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

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