

統計学I

早稲田大学政治経済学術院

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。

- $x_1, x_2, \dots, x_N \xrightarrow{\text{sort}} x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(N)}$

2. 個体数・変数それぞれについて、下位からの累積和を計算。

- 個体数: $n_{(j)} = 1 + 1 + \dots + 1$

- 変数: $t_{(j)} = x_{(1)} + x_{(2)} + \dots + x_{(j)}$

- とともに j 番目までの個体について。

ローレンツ曲線(2)

3. 累積和を総和で除して相対化する。

- 個体数: $F_{(j)} = n_{(j)} / n_{(N)} = j / N$
- 変数: $T_{(j)} = t_{(j)} / t_{(N)} = \sum_{i=1}^j x_{(i)} / \sum_{i=1}^N x_i$
 - ただし、 $j = 1, 2, \dots, n$.
- $F_{(0)} = 0, T_{(0)} = 0$.

4. 以下の要領で2次元平面に線を描く。

- 点 $(F_{(j-1)}, T_{(j-1)})$ と点 $(F_{(j)}, T_{(j)})$ を直線で結ぶ。
 - ただし、 $j = 1, 2, \dots, n$.
- 点 $(0, 0)$ と点 $(1, 1)$ を直線で結ぶ

ローレンツ曲線(3)

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) $n_{(j)}$	相対累積和 (累積相対度数) $F_{(j)}$	変数	累積和 $t_{(j)}$	相対累積和 $T_{(j)}$
(0)	0	0	$0/N$	0	0	$0/t_{(N)}$
(1)	1	1	$1/N$	$x_{(1)}$	$x_{(1)}$	$t_{(1)}/t_{(N)}$
(2)	1	2	$2/N$	$x_{(2)}$	$x_{(1)} + x_{(2)}$	$t_{(2)}/t_{(N)}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮
(N)	1	N	N/N	$x_{(N)}$	$\sum_{i=1}^N x_{(i)}$	$t_{(N)}/t_{(N)}$
合計	N			$\sum_{i=1}^N x_i$		

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)

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

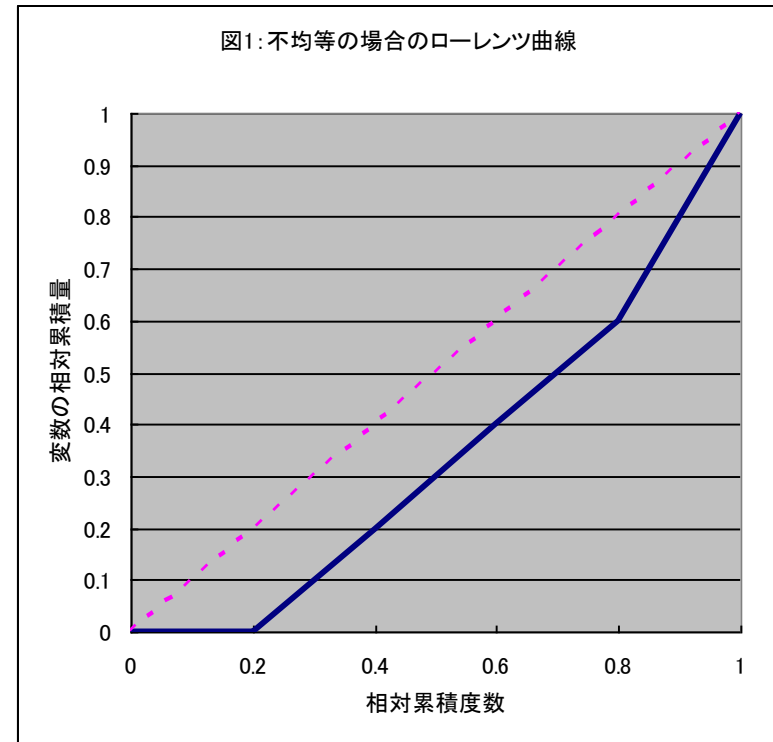
ローレンツ曲線(5)

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) $n_{(j)}$	相対累積和 (累積相対度数) $F_{(j)}$	変数	累積和 $t_{(j)}$	相対累積和 $T_{(j)}$
(0)	0	0	0/5	0	0	0/10
(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

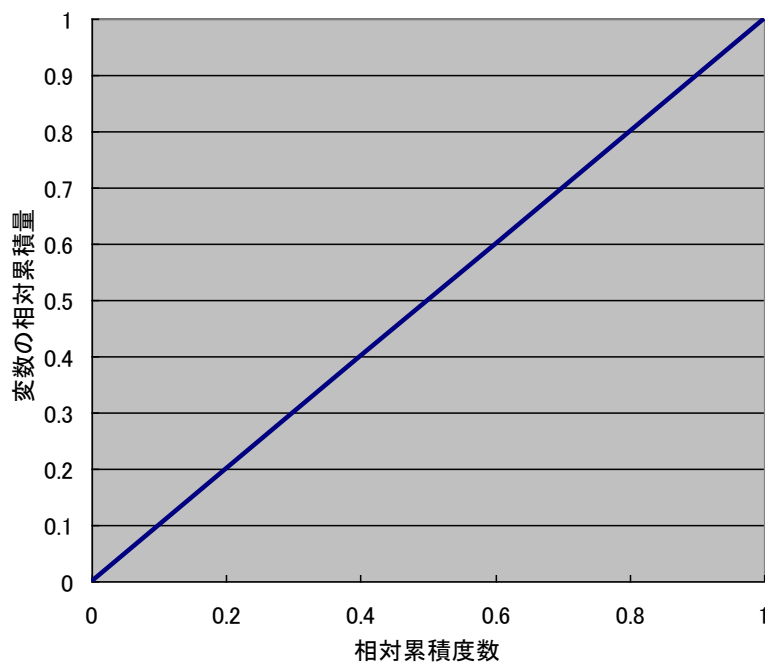
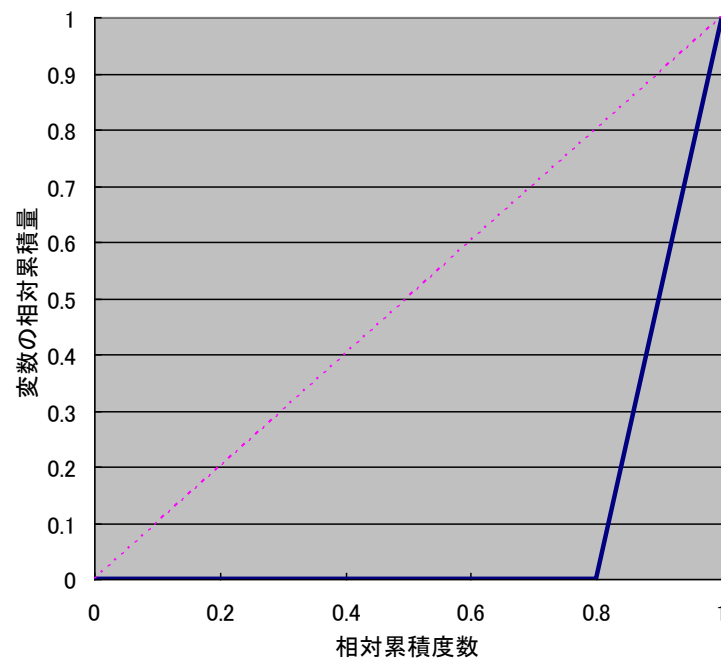


図3: 独占の場合



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

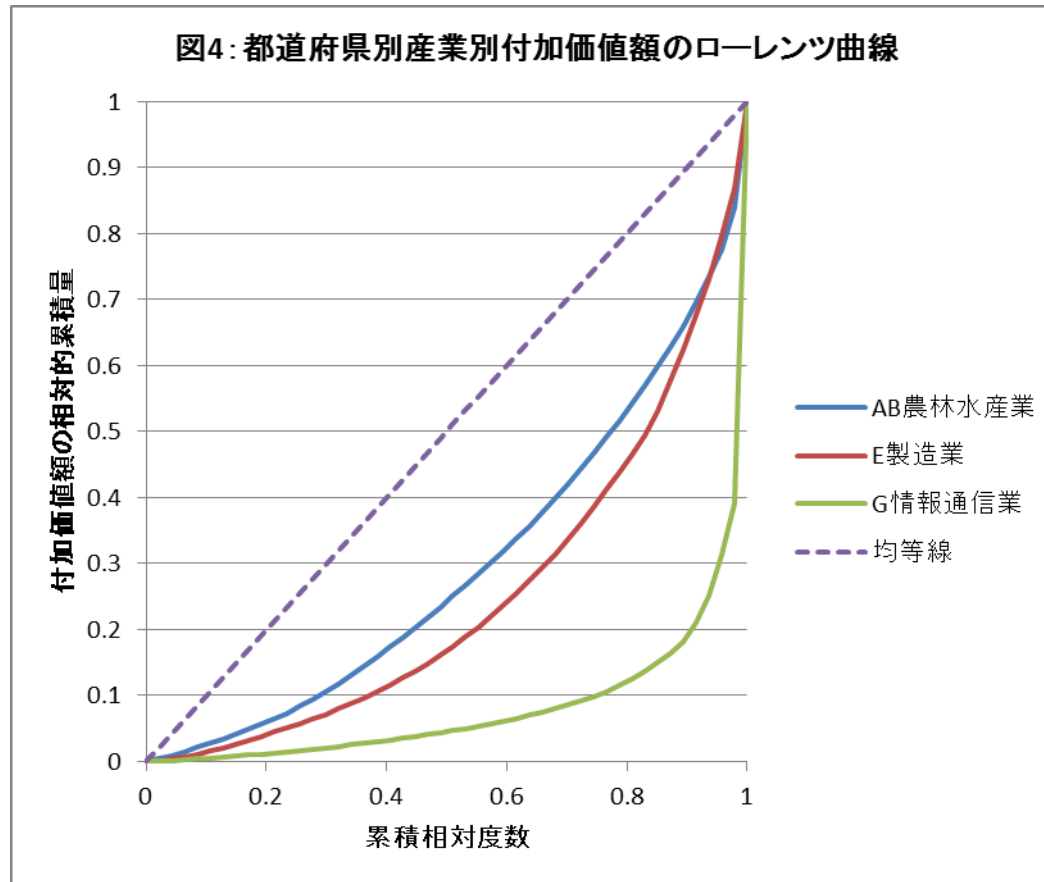
ローレンツ曲線(9)

- 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)

- 都道府県別付加価値額(2016年)
 - AB農林水産業
 - E製造業
 - G情報通信業
- 都道府県の間の不平等の程度は、どの産業においてもっとも顕著であるか。

ローレンツ曲線の応用(2)



資料: 総務省・経済産業省「平成28年経済センサス-活動調査」

不均等度の尺度：ジニ係数(1)

- ジニ係数 $G/$

- $2 \times$ (ローレンツ曲線と均等線の間の弓型面積)

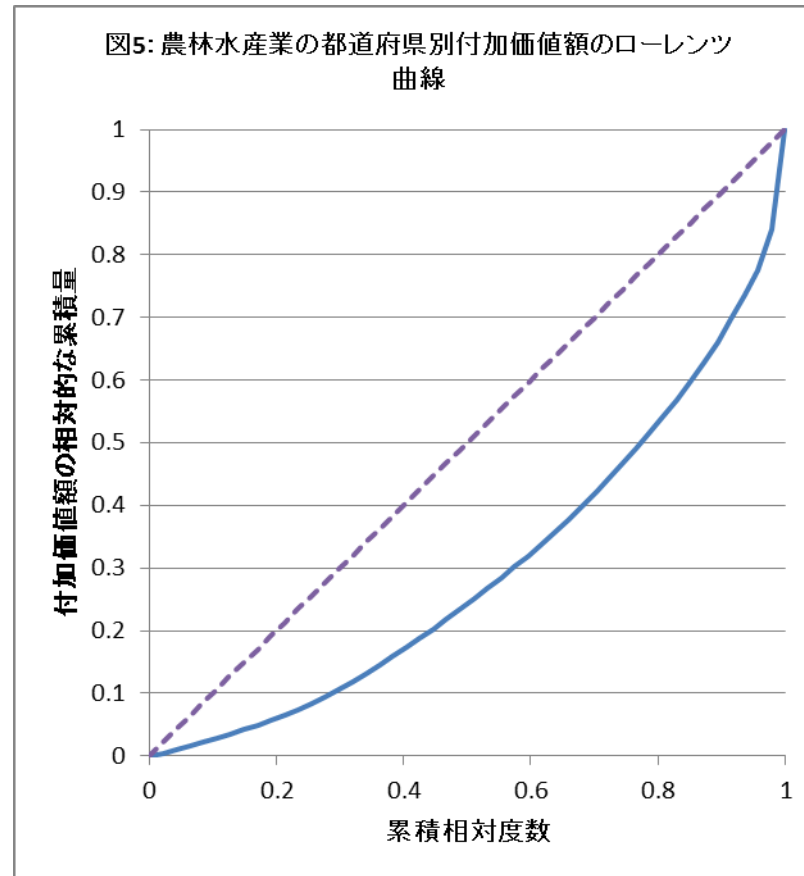
- 均等 不均等

- $0 \longleftarrow G/ \longrightarrow 1$

- 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。

- おおよその目安にすぎない。

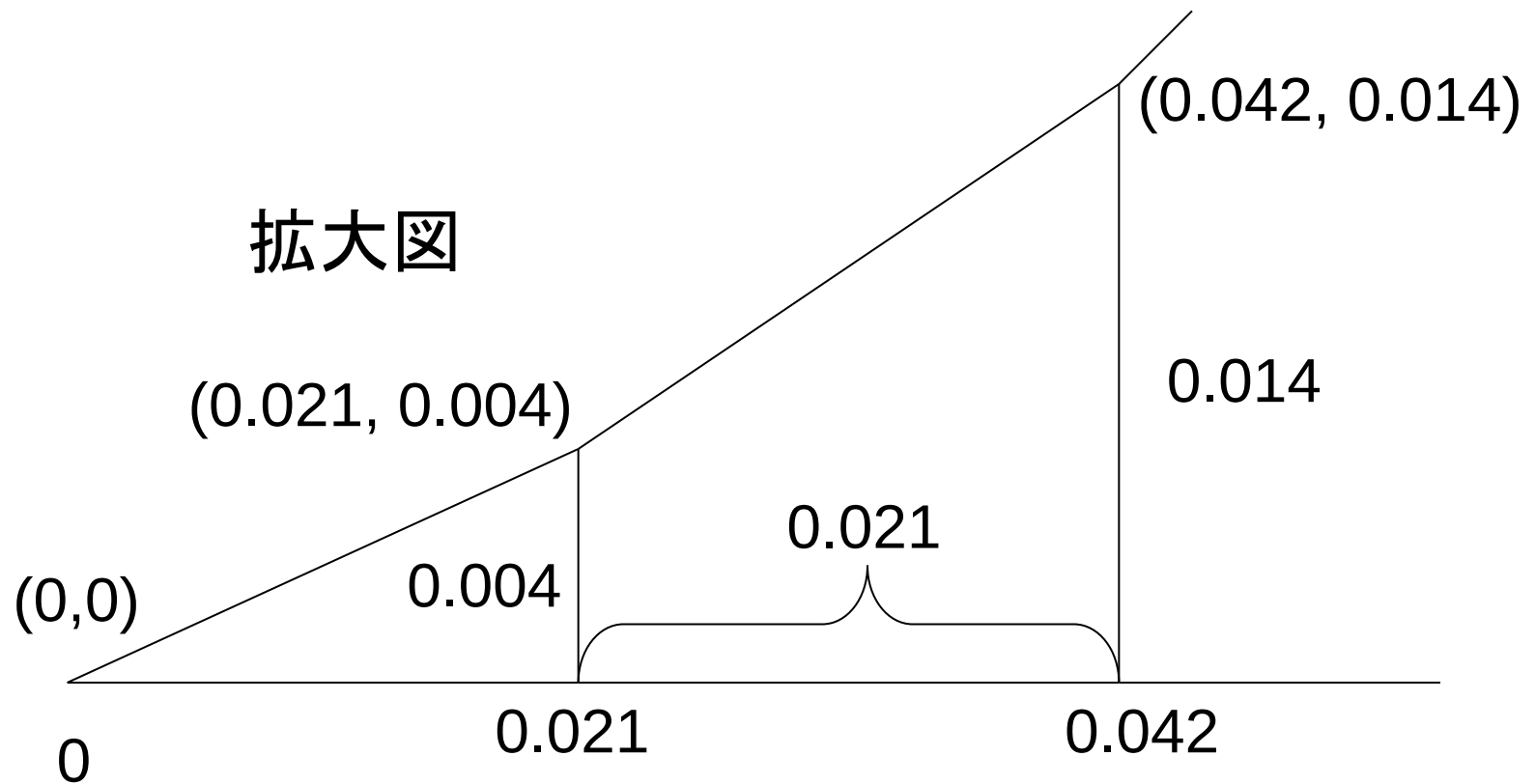
不均等度の尺度：ジニ係数(2)



資料: 総務省・経済産業省「平成28年経済センサス-活動調査」

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)

- 産業別の GI
 - 第1次産業：0.40；第2次産業：0.50；
第3次産業：0.83
- 補足
 - GI は以下のようにも計算できる。

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

個体の変数間の差
の絶対値の平均値

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、

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

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

G の数理(1)

- 以下の通り変数を定義する。

$$- y_{(j)} = \frac{x_{(j)}}{\sum_{i=1}^N x_{(i)}} = \frac{x_{(j)}}{N\bar{x}} \quad (j = 1, 2, \dots, N)$$

- これを使って以下を得る。

$$- F_{(j)} = j/N$$

$$- T_{(j)} = \sum_{i=1}^j x_{(i)} / \sum_{i=1}^N x_i = \sum_{i=1}^j y_{(i)}$$

- 以下の通り定義する。

$$- y_{(0)} = 0$$

GI の数理(2)

- ローレンツ曲線の下側の面積を U とする。

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

GI の数理(3)

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

GI の数理(4)

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

$$= y_{(1)}$$

$$+ y_{(1)} + y_{(2)}$$

$$+ y_{(1)} + y_{(2)} + y_{(3)}$$

$$+ \cdots$$

$$+ y_{(1)} + y_{(2)} + y_{(3)} + \cdots + y_{(N)}$$

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

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

GI の数理(5)

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

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)

$$\begin{aligned} &= \frac{1}{N \bar{X}} \left\{ \sum_{j=1}^N x_{(j)} - \sum_{j=1}^N \left(\frac{2N+1-2j}{N} \right) x_{(j)} \right\} \\ &= \frac{1}{N \bar{X}} \sum_{j=1}^N \left\{ 1 - \left(\frac{2N+1-2j}{N} \right) \right\} x_{(j)} \\ &= \frac{1}{N \bar{X}} \sum_{j=1}^N \frac{2j-N-1}{N} x_{(j)} \\ &= \frac{1}{N^2 \bar{X}} \sum_{j=1}^N (2j-N-1) x_{(j)} \end{aligned}$$

GI の数理(8)

ところで、

$$\begin{aligned} \sum_{i=1}^N \sum_{j=1}^N |x_i - x_j| &= \sum_{i=1}^N \sum_{j=1}^N |x_{(i)} - x_{(j)}| \\ &= |x_{(1)} - x_{(1)}| + |x_{(1)} - x_{(2)}| + |x_{(1)} - x_{(3)}| + \cdots + |x_{(1)} - x_{(N)}| \\ &\quad + |x_{(2)} - x_{(1)}| + |x_{(2)} - x_{(2)}| + |x_{(2)} - x_{(3)}| + \cdots + |x_{(2)} - x_{(N)}| \\ &\quad + \cdots \\ &\quad + |x_{(N)} - x_{(1)}| + |x_{(N)} - x_{(2)}| + |x_{(N)} - x_{(3)}| + \cdots + |x_{(N)} - x_{(N)}| \end{aligned}$$

GI の数理(9)

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

GI の数理(10)

したがって、

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