

行列のスケーリングとその応用

清 智也

2017年6月7日(水)

数理情報工学演習2A

概要

- 複数の指標から1つの総合指標を作りたい、という場面がしばしばある。

例：5科目の成績、大学ランキング、経済指標 etc.

- 目標：「公平」な総合指標を作る。

例：幸福度ランキング

寺島監修「全47都道府県幸福度ランキング2016年度版」, 2016.

- 65指標をもとに総合指標を作り、都道府県をランキングしている。

基本指標 人口増加率、県民所得、選挙投票率、食料自給率、財政健全度

健康分野: 生活習慣病、気分障害、産婦人科、ホームヘルパー、高齢者ボランティア、健康寿命、平均歩数、健康診査受診率、スポーツの活動時間

文化分野: 教養娯楽支出、余暇時間、常設映画館、書籍購入数、NPO認証数、外国人宿泊者、姉妹都市、語学教室、海外渡航者率、留学生数

仕事分野: 若者完全失業率、正規雇用者、高齢者有業率、インターンシップ、大卒者進路、障害者雇用率、製造業労働生産性、事業所新設、特許、本社機能流出流入

生活分野: 持ち家、生活保護受給、待機児童、一人暮らし高齢者、インターネット人口、污水处理人口、道路整備、一般廃棄物リサイクル、エネルギー消費量、地縁団体

教育分野: 学力、不登校児童、司書教諭、大学進学、教員一人当たり児童生徒数、社会教育費、社会教育学級数、学童保育設置率、余裕教室活用率、悩みストレス

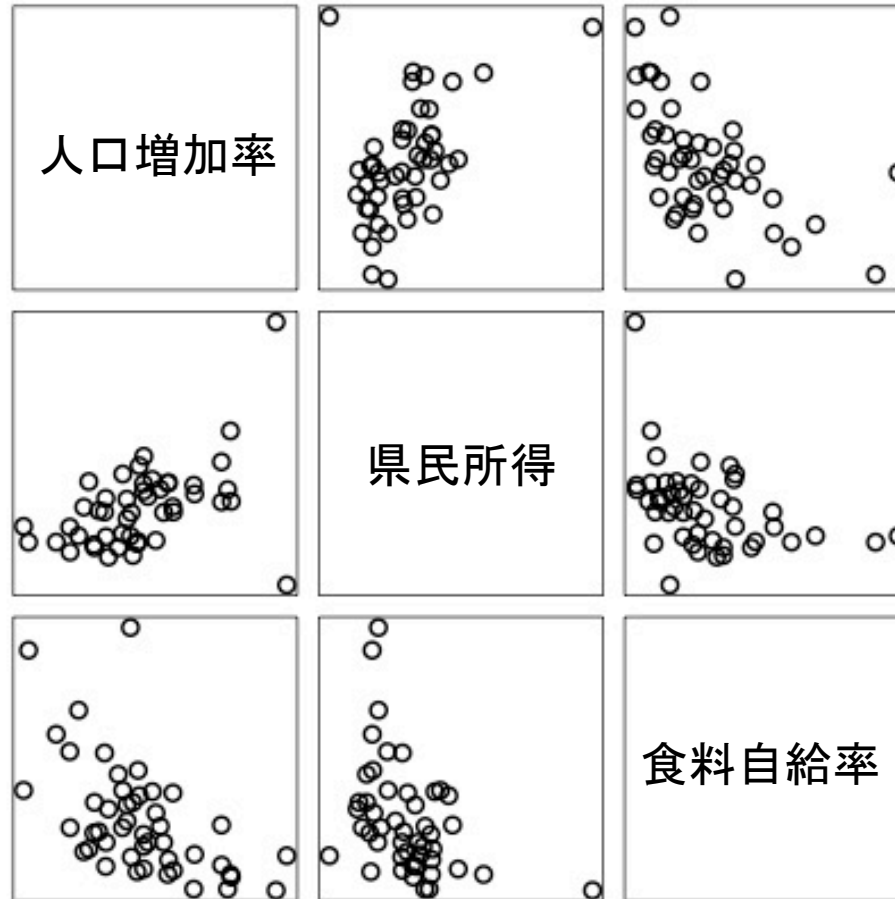
追加指標: 信用金庫利回り、平均寿命、女性労働力、自殺死亡者、子供の運動能力、合計特殊出生率、自主防災組織、刑法犯認知件数、農業の付加価値創出、勤労者世帯可処分所得

都道府県データ

	2010 → 2014 人口増加率(%)	一人当たり 県民所得(千円)	カロリーベース 食料自給率(%)
北海道	-1.93	2,475	198
青森	-3.81	2,418	118
岩手	-3.47	2,558	105
宮城	-0.86	2,747	74
...			
東京	1.75	4,464	1
...			
沖縄	2.02	2,020	27

➡ 1つの総合指標にまとめたい

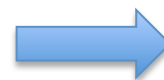
参考：対散布図



総合指標その1: 偏差値の和

- 各変量の偏差値(平均0、分散1)を求め、和をとる。

	人口増加率 偏差値	県民所得 偏差値	食料自給率 偏差値		
北海道	-0.18	-0.68	3.37	総合指標	順位
青森	-1.46	-0.83	1.50	2.51	2
岩手	-1.23	-0.47	1.20	-0.79	34
宮城	0.56	0.02	0.47	-0.50	30
...				1.05	7
東京	2.35	4.44	-1.24	5.54	1
...					
沖縄	2.53	-1.85	-0.63	0.05	23



x_1

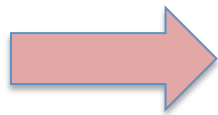
x_2

x_3

$x_1 + x_2 + x_3$

総合指標の公平性

- 「偏差値の和」は公平か？



必ずしも公平とは言えない

都道府県データの相関行列

	人口増加率	県民所得	食料自給率
人口増加率	1	0.47	-0.56
県民所得	0.47	1	-0.40
食料自給率	-0.56	-0.40	1

「偏差値の和」 との共分散
0.91
1.07
0.03

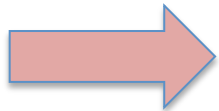
食料自給率と総合指標の
相関がほとんどゼロ。☹
データによっては負にもなり得る

総合指標その2: 主成分得点

- 相関の情報を加味した総合指標として、第1主成分が考えられる。

$$g = 0.60x_1 + 0.54x_2 - 0.58x_3$$

食料自給率に負の重みを掛けてしまう。☹



まったく公平でない

e.g. Baker (1974)

	第1主成分 得点	順位
北海道	-0.24	45
青森	-0.17	41
岩手	-0.18	43
宮城	0.06	17
...		
東京	0.41	1
...		
沖縄	0.10	13

総合指標の数理

- 線形の総合指標 $\mathbf{g} = w_1\mathbf{X}_1 + w_2\mathbf{X}_2 + w_3\mathbf{X}_3$
- 満たしてほしい二つの性質
 - 重みの公平性 $w_i > 0$
 - 共分散の公平性 $\mathbf{x}'_i\mathbf{g} > 0$

	偏差値の和	主成分得点	OGI (次頁)
重みの公平性	○	×	○
共分散の公平性	×	×	○

OGI

変量 $\mathbf{X}_1, \dots, \mathbf{X}_p$ ← 一次独立と仮定。

後で
説明

定理 (Marshall & Olkin, 1968)

次を満たす**正の実数** w_1, \dots, w_p が一意的に存在する:

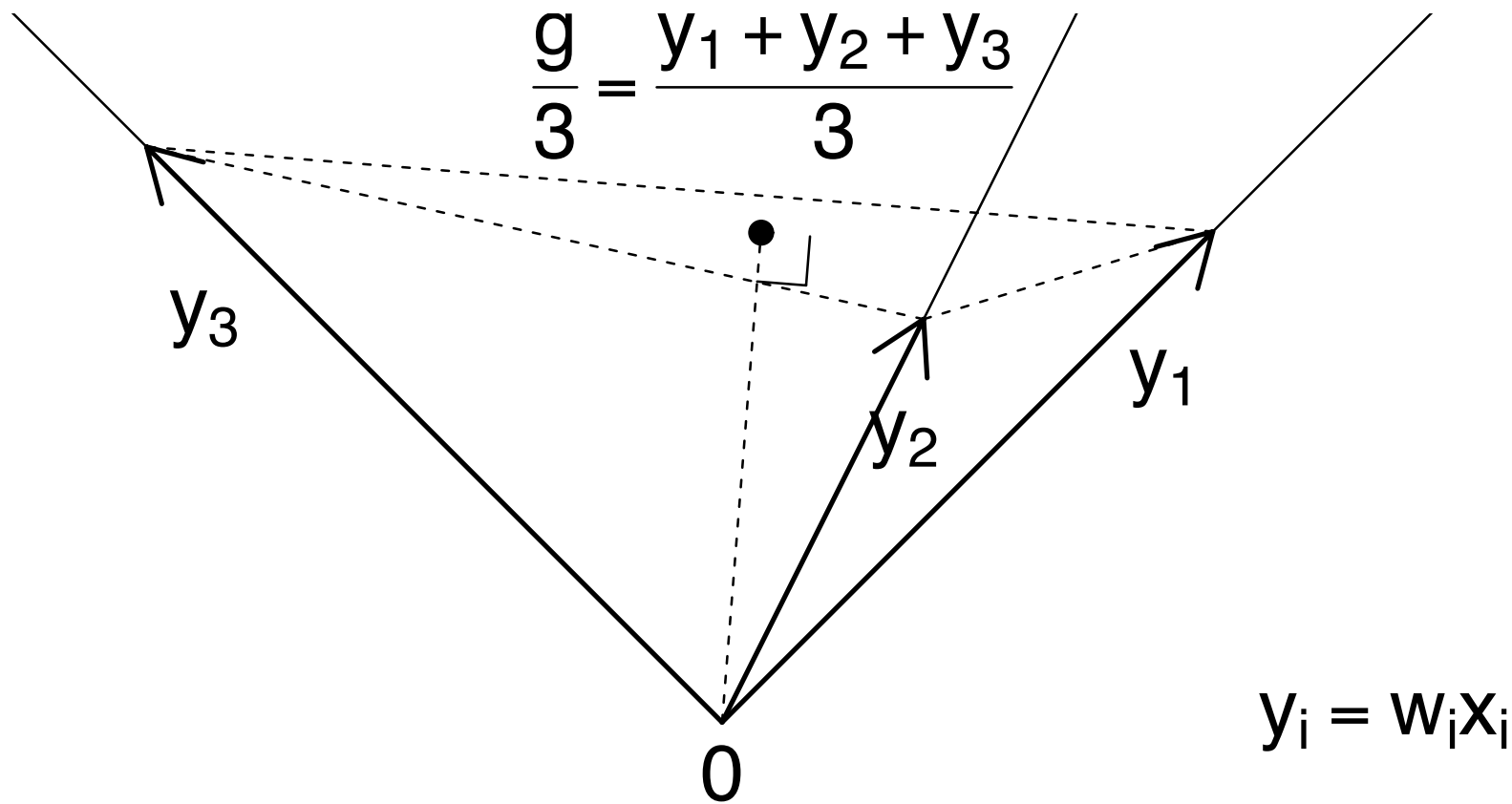
$$\mathbf{g} := \sum_{i=1}^p w_i \mathbf{X}_i, \quad \frac{1}{n} (w_i \mathbf{X}_i)' \mathbf{g} = 1$$

総合指標との共分散が1

この \mathbf{g} を客観的総合指数 **OGI** と呼ぶ。

(OGI = objective general index) Sei (2016)

OGI の幾何学的解釈



都道府県データの OGI

- さっきの例で確認

$$\begin{aligned} \mathbf{g} &= w_1 \mathbf{x}_1 + w_2 \mathbf{x}_2 + w_3 \mathbf{x}_3 \\ &= 1.27 \mathbf{x}_1 + 1.05 \mathbf{x}_2 + 1.72 \mathbf{x}_3 \end{aligned} \quad \text{重みは正！ 😊}$$

$w_i \mathbf{x}_i$ の共分散行列

	人口増加率	県民所得	食料自給率
人口増加率	1.60	0.63	-1.23
県民所得	0.63	1.10	-0.72
食料自給率	-1.23	-0.72	2.95

総合指標 との共分散	
1	
1	
1	

共分散も正！ 😊

得られた総合指標の比較

	人口増加率 偏差値	県民所得 偏差値	食料自給率 偏差値
北海道	-0.18	-0.68	3.37
青森	-1.46	-0.83	1.50
岩手	-1.23	-0.47	1.20
宮城	0.56	0.02	0.47
...			
東京	2.35	4.44	-1.24
...			
沖縄	2.53	-1.85	-0.63

x_1

x_2

x_3

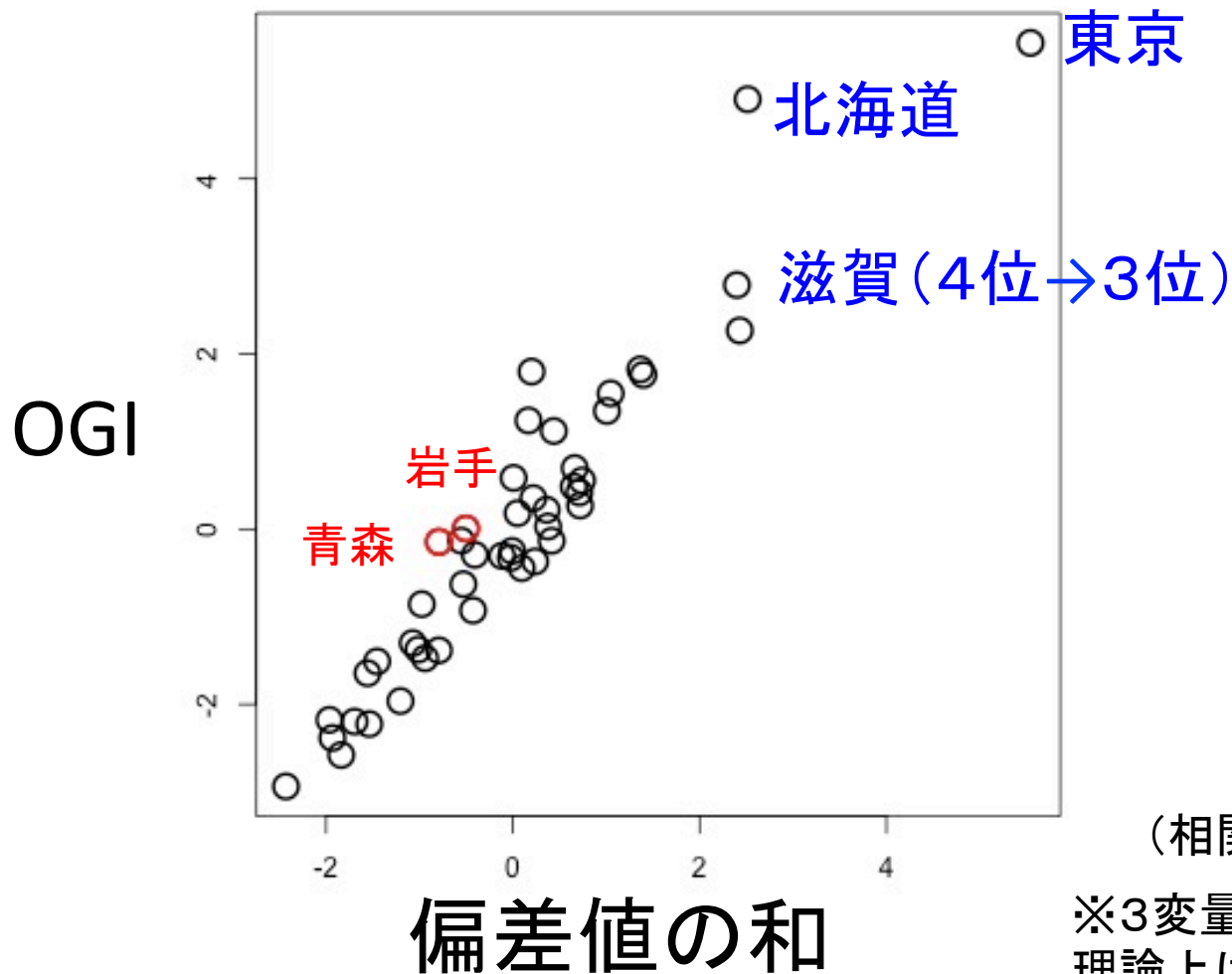
偏差値の和	順位
2.51	2
-0.79	34
-0.50	30
1.05	7
...	
5.54	1
...	
0.05	23

$x_1 + x_2 + x_3$

OGI	順位
4.91	2
-0.15	25
0.01	22
1.55	8
...	
5.55	1
...	
0.18	20

$1.27x_1 + 1.05x_2 + 1.72x_3$

総合指標どうしの散布図

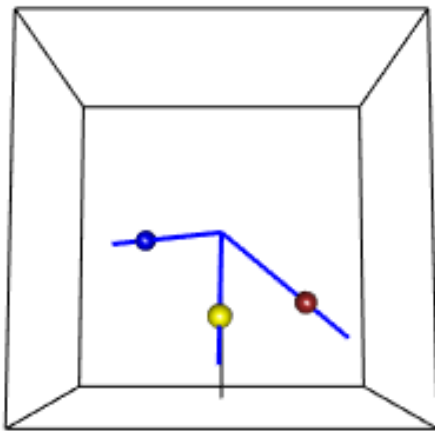


(相関係数 0.95)

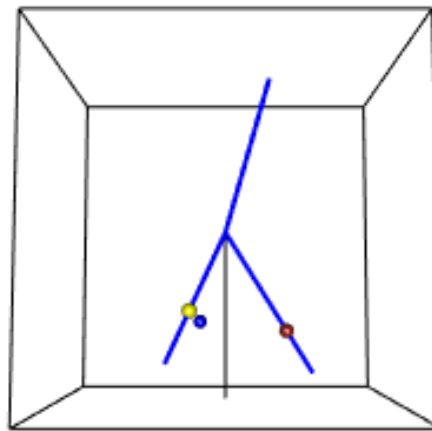
※3変量以上の場合、
理論上は任意の正の相関
係数を取り得る。

不公平性の可視化

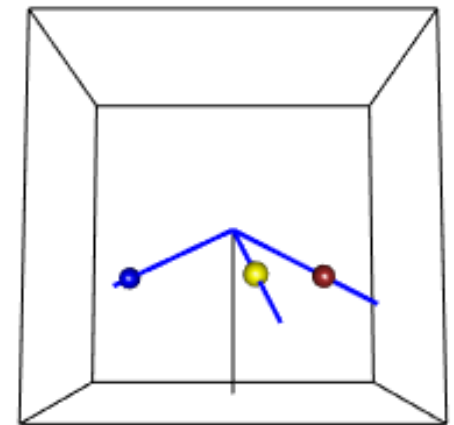
- 不公平性は「やじろべえ」で可視化できる。



偏差値の和



主成分



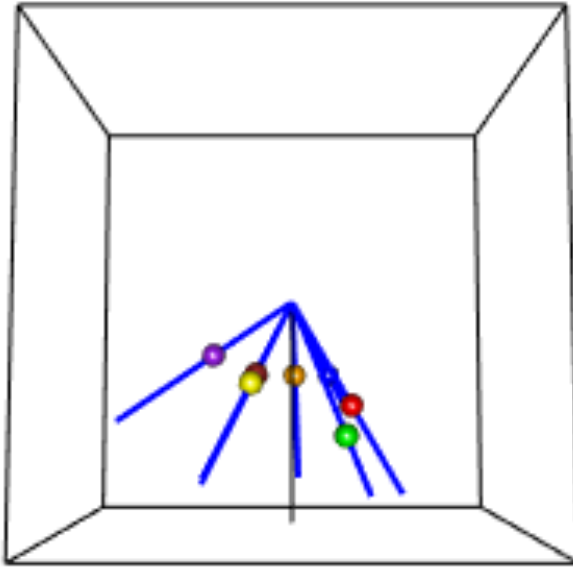
OGI

水平につりあう！

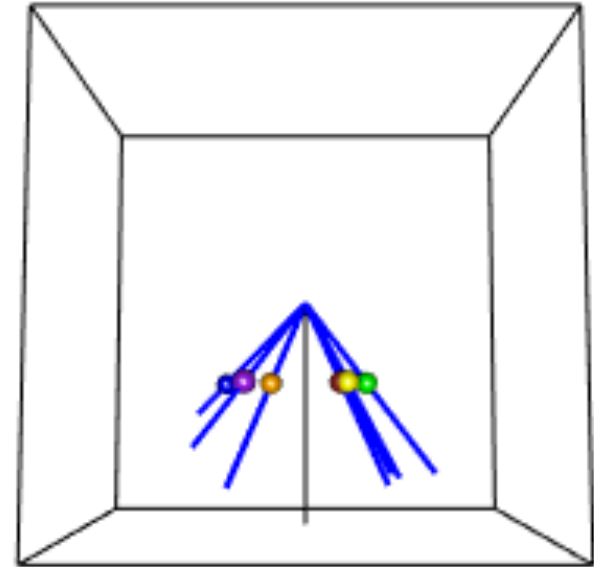
- 人口増加率
- 県民所得
- 食料自給率

- ・総合指標を鉛直下方向に描いた図
- ・内積が共分散に対応

女子七種競技の例



IAAFルール



OGI

- 100mH
- HJ
- SP
- 200m
- LJ
- Javelin
- 800m

3次元に射影して描いた図。
本当は7次元必要。

この例は選択バイアスがあるので注意

まとめ

- 偏差値の和や主成分は、総合指標としての公平性に欠ける。
- OGI は公平性を持った総合指標である。
- ただし、「そもそもどの変数を使うか？」という設計の方が大事であり、難しい。
- 以下、数学的な部分を黒板で説明します。

参考文献

- 1) R.J. Baker (1974). Selection indexes without economic weights for animal breeding, *Can. J. Animal Sci.*, 54 (1), 1–8.
- 2) A.W. Marshall, I. Olkin (1968). Scaling of matrices to achieve specified row and column sums, *Numer. Math.*, 12, 83–90.
- 3) T. Sei (2016). An objective general index for multivariate ordered data, *J. Multivariate Anal.*, 147, 247–264.
- 4) 寺島 実郎監修, (一財)日本総合研究所編／日本ユニシス総研編「全47都道府県幸福度ランキング2016年度版」, 東洋経済新報社, 2016.