

主成分分析

— 因子分析との比較 —

2013.7.10.

心理データ解析演習

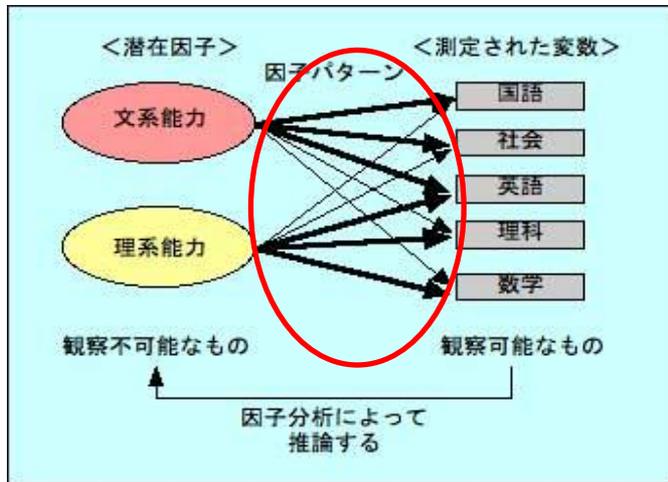
M1 栴田 恵

主成分分析とは

- 主成分分析は、多変量データに共通な成分を探って、一種の合成変数(主成分)を作り出すもの
 - * 主成分はデータを新しい視点でみるための新しい軸
- 主成分分析の目的:「情報を縮約すること」
(データを合成変数(主成分)に総合化)
- ⇔ 因子分析の目的:「共通因子を見つけること」
(データを潜在因子に分解)

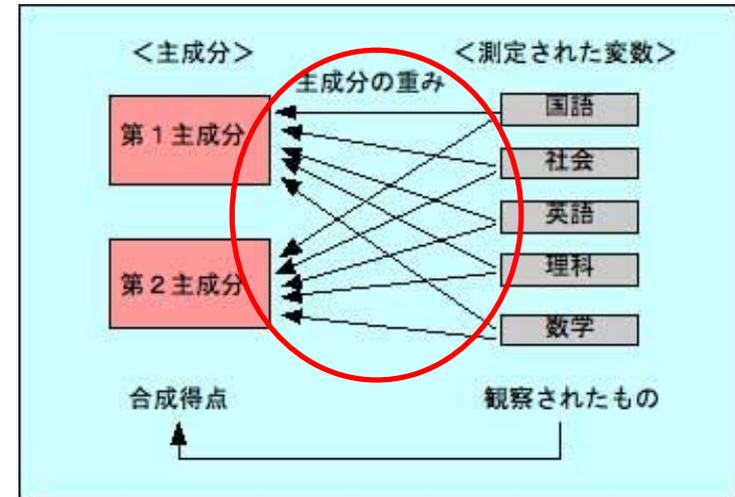
因子分析と比較！

因子分析



- 観測変数に共通な潜在因子を探る
 - 因子は変数を説明する**独立変数**
 - 共通因子の影響を除いたら変数間の偏相関が0になるように因子負荷を求める
- 変数間の相関関係を因子によって説明する

主成分分析



- 観測変数が共有する情報を合成変数として集約する
 - 主成分は変数によってその値が決まる**従属変数**
 - 主成分の分散が最大になるように変数にかかる重みを求める
- もとの変数群の分散をできるだけ取り込むような合成変数を求める

因子分析と比較！

- 因子分析

$$X_j = a_{j1}f_1 + a_{j2}f_2 + e_j$$

- x: 観測変数
- a: 因子負荷
- f: 共通因子
- e: 独自因子

* 共通因子に誤差は含まない
(独自因子に含まれる)

- 主成分分析

$$Z_j = a_{j1}X_1 + a_{j2}X_2$$

- z: 主成分
- a: 主成分負荷量
- x: 観測変数

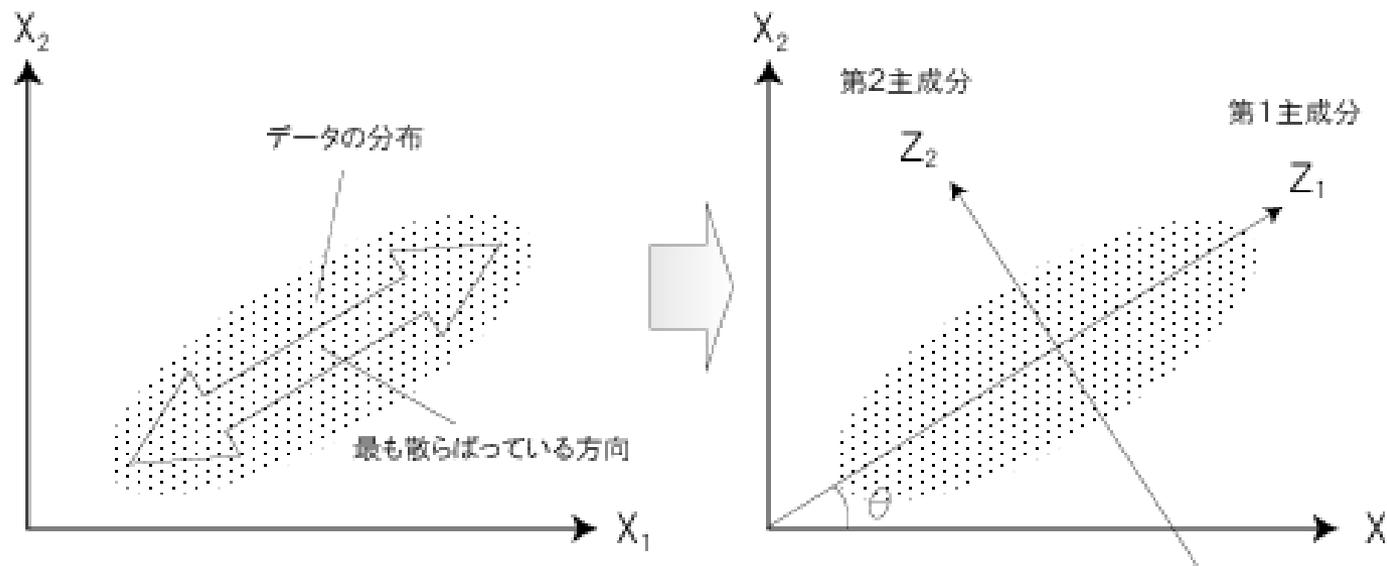
独自因子を考慮しない！

* 主成分は誤差を含む変数をそのまま合成しているため、誤差を含む

$$* a_{j1}^2 + a_{j2}^2 = 1$$

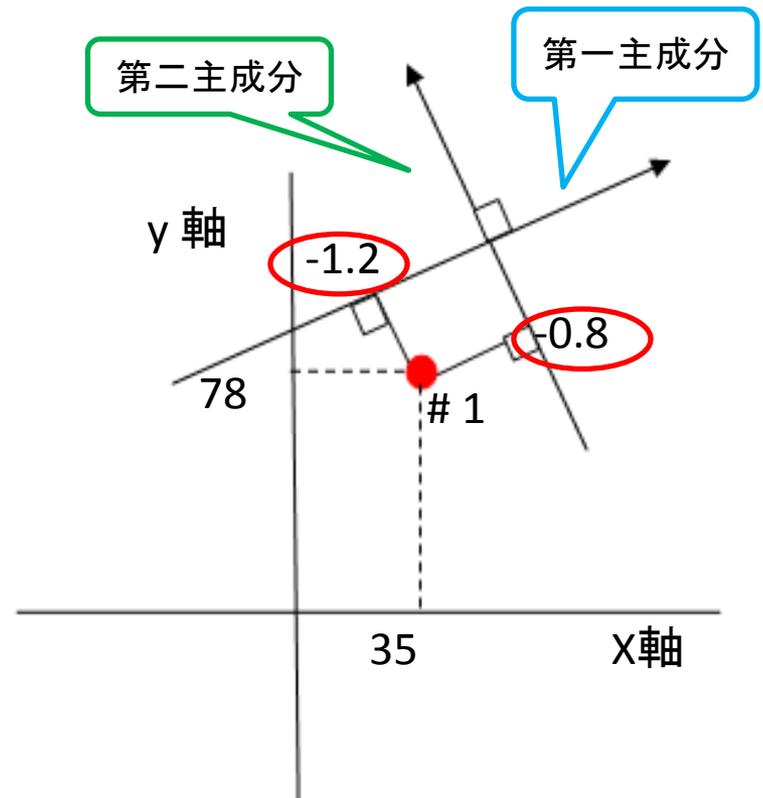
主成分の決め方

- 分散が最大になるように第1主成分を決めたら、次は第1主成分とは相関しないという条件下で分散が最大になるように、第2主成分を決める。第3主成分以降も同様。



主成分得点

- 主成分得点
 - データが新しく得られた主成分の軸上で取る値
 - 主成分の軸にあるデータの点(#1)から下ろした垂線の先の軸上での目盛(座標値)
 - * データの数だけ主成分得点はある
 - * 主成分負荷量: 主成分得点ともとの変数との相関係数(主成分負荷量を見ると、もとの変数が各主成分にどの程度反映されているか分かる)



ここまでを簡単にまとめると...

- 因子分析では「因子」→「項目」というパスが引けるが、主成分分析では「項目」→「合成得点」というパスになる。
- ⇒ 主成分分析は**データの記述**であり、因子分析のような潜在変数を想定したモデルではない。
 - * 主成分分析はデータをまとめる(合成の分析)
- 主成分分析では、**主成分をできる限り少なくすることが目的**
 - 結果において、第一主成分でどの項目も高くなっていることは、主成分分析では喜ばしい結果
 - 究極的には、第一主成分だけ出ればよい
 - * 第一主成分は総合指標になることが多い
 - * 上位の主成分ほど個体の散らばりをよく表す

主成分分析の特徴

- (1) 主成分分析は、計算に**繰り返しのない**主因子法を用い、**共通性の初期値を1**として計算(独自性を考慮しない)
 - 主成分分析では、必ず共通性の初期の値はすべて1と出力される！
 - ここで用いる主因子法は因子分析の因子抽出法の1つ。因子分析と違うのは、因子分析の計算では共通性を1とは推定しないということだけ。

主成分分析の特徴

- (2)主成分分析では(基本的に)軸の回転を行わない
 - 主成分分析で用いる主因子法は、第一因子にできるだけ因子寄与を高くしようとするもの
 - 因子分析で回転させるのは、因子寄与を分散させて単純構造にするためであり、因子数を少なくするのではなく、むしろ増えることになる
- 少ない次元で説明することができなくなる
- * 単純構造...特定の因子だけに因子負荷が高い値を示した因子パターンを示す場合

主成分分析の特徴

- (3) 主成分分析では、取り上げなかった下位の主成分を誤差と見なす(上位の主成分を結果として取り上げ、解釈の対象にする)
 - ー 下位の主成分の重みつき合計が各変数の誤差
 - 各変数の誤差は取り上げなかった下位の主成分をその源として共有し、互いに独立なものでない
 - 主成分分析の誤差は各変数に**独自のものではない!**

主成分分析の解釈

- 主成分分析においては、異なる主成分どうしは無相関（軸が直交）

→異なる主成分は互いに無関係な意味をもつ

- 主成分はどちらの方向が正か負かを決められないので、主成分の正負と意味づけにおける正負が逆転してもよい

一例)ある主成分が「対人関係能力の高さ」と解釈した場合...

主成分 { 正⇒対人関係能力が高い

{ 負⇒対人関係能力が低い

という解釈が妥当なこともある

{ 正⇒対人関係能力が低い

{ 負⇒対人関係能力が高い

という逆転の解釈が妥当なこともある

どんな時に使うのか

- 多変量データ(例えば多数の質問項目)の総合評価をしたい時
 - 各変数の類似性を明確にし、総合的に把握する
 - いわゆる(たったひとつの)総合評価だけでなく、いろいろな観点からの総合評価を抽出する

例) 授業評価のアンケート

- 授業全体の総合評価に加え、内容の総合評価や教員の努力の総合評価を抽出することができる

どんな時に使うのか

- 変数に重みづけをして、合成得点を算出したい時
— 変数に重みづけをすれば、個体の相違がいつそう明瞭になる総得点が得られる
- * 分散をできるだけ大きくする
→ 個体差をできるだけ大きくする

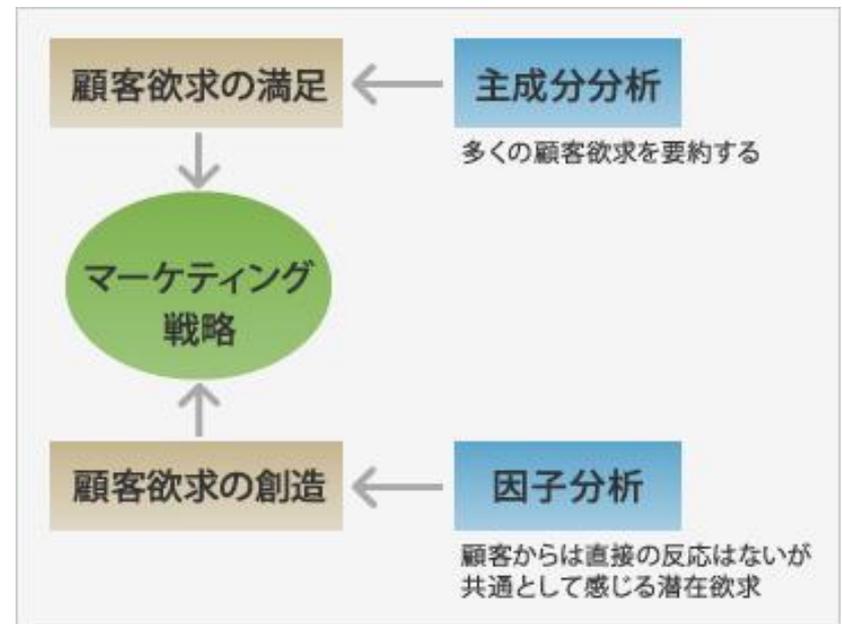
どんな時に使うか

例) テストの合計得点の算出

- 国語の平均が30点(SD 10), 数学の平均点が70点(SD 20)である時
 - 国語が得意なA君は国語が40点, 数学が50点で, 2教科の合計は90点。
 - 数学が得意なB君は国語が20点, 数学が90点で, 2教科の合計は110点。
 - 単に足しあわせた合計得点には, 数学の得点の影響がより大きく反映してしまうのではないか。
 - 数学が得意な学生が上位を占め, 国語が得意な学生の順位が低くなってしまうことになり, あまりフェアなやり方とはいえない。
- 主成分分析を用いると, 各教科の点数に「重みづけをして」, 合成得点を算出することができる

どんな時に使うのか

- 主成分分析は企業経営、マーケティング等でも用いられる
 - 顧客満足度を調査したアンケート結果をまとめるとき（多くの顧客の欲求をまとめ、明らかにする）
 - 消費者のさまざまな商品の購買数量の類似性を明らかにしたいときなど



<出典：清水功次『マーケティングのための多変量解析』>

<http://www.macromill.com/landing/works/b001.html>

主成分分析の利点

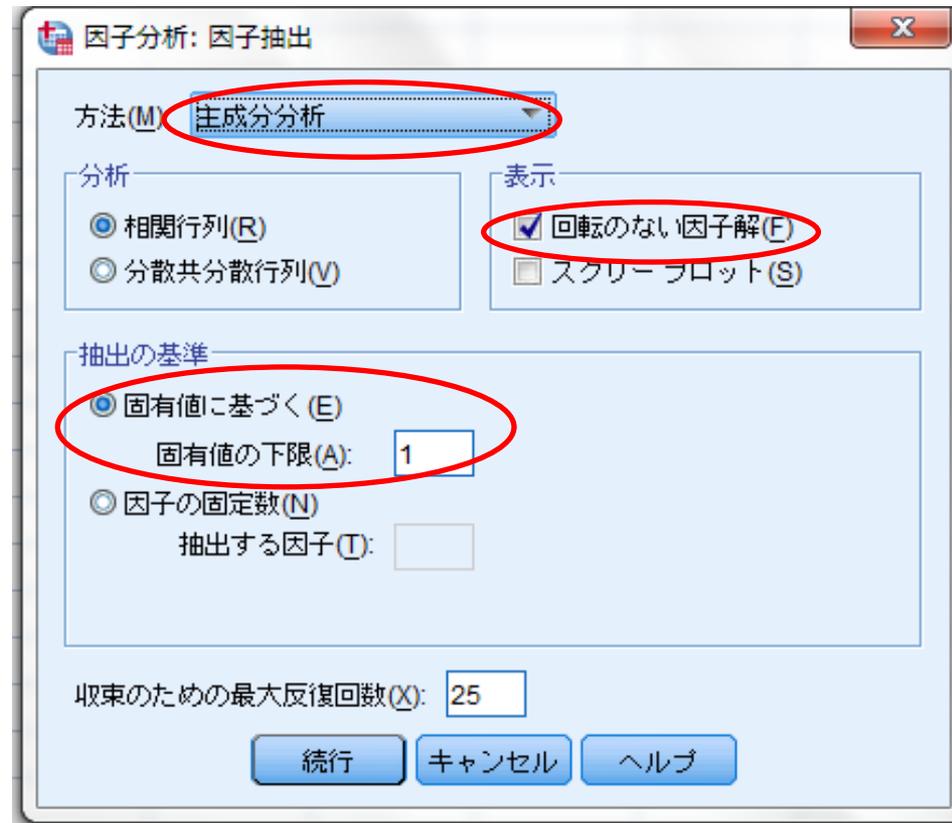
- 数学的な原理や主成分を求める方法が単純明快
← 変数間の相関行列さえあれば計算できる
- 因子分析のように複数の推定法があって、それぞれで結果が違ふということはない
- データ数より変数の数が多いデータにも適用できる
(得られる主成分の数は「データ数 - 1」)
- 因子分析では、データ数が少ないと不適解が生じやすいが、不適解が生じないため、常に結果が求められる
- 1回で計算が終わる

注意！

- みなさんはそんなことしないと思いますが...
- SPSSでは、因子分析のデフォルトが「主成分分析」になっているために、因子分析をするつもりが、誤って主成分分析を行ってしまうケースがある
 - 計算上の違いは共通性の初期値の違いだけであるので、大きな違いは生じない場合もあるが、ふつうの因子分析よりいい結果に見えてしまうことがあるので注意！
 - 因子分析と主成分分析では得点の算出法が大きく異なるので注意！
- * 因子得点...各因子と各個体(各データ)の相関の程度

デモ

- 因子抽出
 - 「方法」で「主成分分析」(デフォルト)を選択
 - 「抽出の基準」の「最小の固有値」は「1」でよい
 - 「回転のない因子解」にチェックをいれておく
- 「続行」
- * 主成分分析では軸を回転させない



デモ

- 主成分得点を算出したい時には「得点」をクリック
 - 「変数として保存」にチェックを入れる
 - 「回帰法」を選択
 - 「続行」 → 「OK」



デモ

共通性

- 因子分析の時と同様に，共通性が出力される。
- ただし，初期の固有値はすべて「1」になる。

共通性

	初期	因子抽出後
外向性	1.000	.689
社交性	1.000	.656
積極性	1.000	.765
知性	1.000	.661
信頼性	1.000	.648
素直さ	1.000	.792

因子抽出法: 主成分分析

デモ

説明された分散の合計

- 一番左上の部分が、主成分分析では「成分」となっている。
- 回転を行っていないので、「回転後の負荷量平方和」は出力されない。
- 全分散のうち2つの主成分で説明される部分は70.21%となっている。

説明された分散の合計

成分	初期の固有値			抽出後の負荷量平方和		
	合計	分散の %	累積 %	合計	分散の %	累積 %
1	2.691	44.853	44.853	2.691	44.853	44.853
2	1.521	25.358	70.211	1.521	25.358	70.211
3	.715	11.909	82.119			
4	.482	8.036	90.156			
5	.334	5.567	95.723			
6	.257	4.277	100.000			

因子抽出法: 主成分分析

デモ

成分行列

- 因子分析では「因子行列」であったが、主成分分析では「成分行列」となる。
- ここで表示される数値は「重み(負荷量)」と呼ばれる。
- 第1主成分には6つの変数いずれも正の重みを示している→第1主成分:「総合評価」
- 第2主成分は「外向性」「社交性」「積極性」が正の重み,「知性」「信頼性」「素直さ」が負の重みを示している
→第2主成分:「対人関係能力と知的能力のいずれが優位か」

成分行列^a

	成分	
	1	2
外向性	.706	.436
社交性	.656	.475
積極性	.635	.601
知性	.667	-.466
信頼性	.662	-.458
素直さ	.689	-.563

因子抽出法: 主成分分析

a. 2 個の成分が抽出されました

因子分析と比較

- 因子分析 (主因子法・バリマックス回転)
- 主成分分析

共通性

	初期	因子抽出後
外向性	.432	.536
社交性	.434	.456
積極性	.532	.695
知性	.440	.459
信頼性	.471	.443
素直さ	.513	.816

因子抽出法: 主因子法

共通性

	初期	因子抽出後
外向性	1.000	.689
社交性	1.000	.656
積極性	1.000	.765
知性	1.000	.661
信頼性	1.000	.648
素直さ	1.000	.792

因子抽出法: 主成分分析

回転後の因子行列^a

	因子	
	1	2
外向性	.204	.703
社交性	.156	.657
積極性	.051	.832
知性	.658	.163
信頼性	.648	.151
素直さ	.900	.082

因子抽出法: 主因子法
回転法: Kaiserの正規化を伴うバリマックス法

a. 3回の反復で回転が収束しました。

成分行列^a

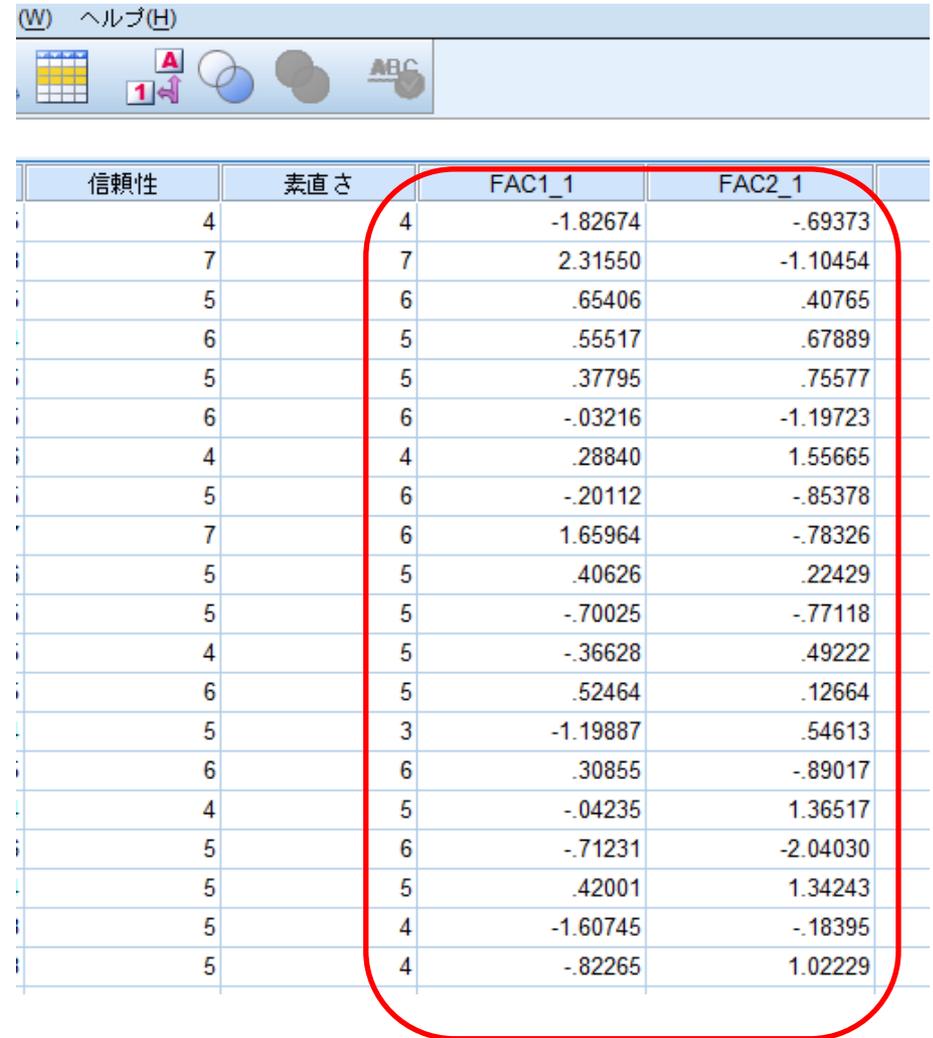
	成分	
	1	2
外向性	.706	.436
社交性	.656	.475
積極性	.635	.601
知性	.667	-.466
信頼性	.662	-.458
素直さ	.689	-.563

因子抽出法: 主成分分析

a. 2個の成分が抽出されました

デモ

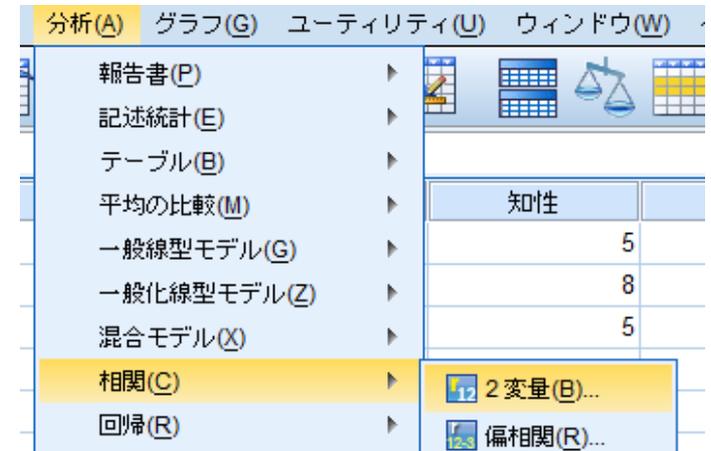
- 「主成分得点」を算出するよう指定したので、2つの主成分に相当する得点が各ケースについて算出される。
- 主成分得点は平均が「0」、分散が「1」になる。



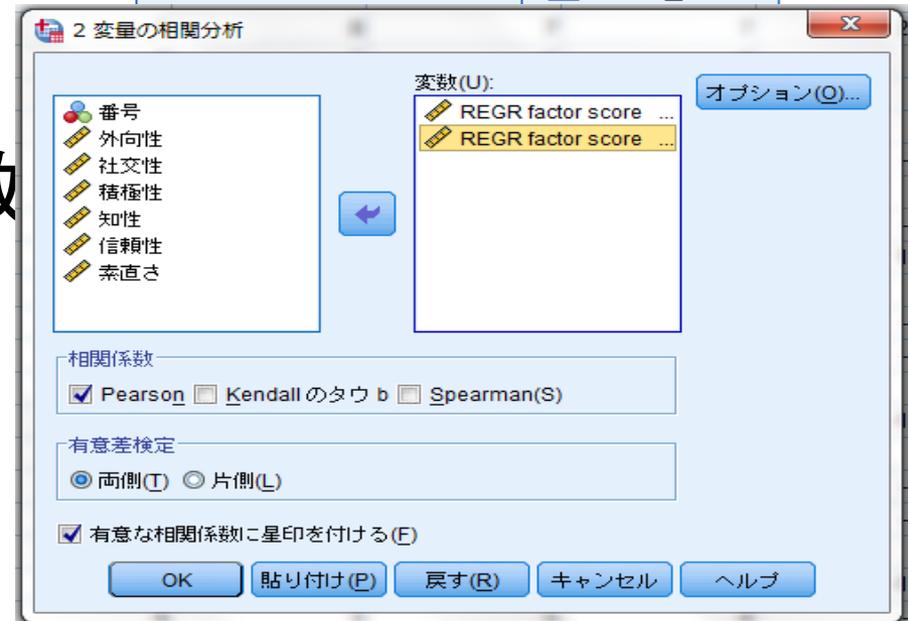
信頼性	素直さ	FAC1_1	FAC2_1
4	4	-1.82674	-.69373
7	7	2.31550	-1.10454
5	6	.65406	.40765
6	5	.55517	.67889
5	5	.37795	.75577
6	6	-.03216	-1.19723
4	4	.28840	1.55665
5	6	-.20112	-.85378
7	6	1.65964	-.78326
5	5	.40626	.22429
5	5	-.70025	-.77118
4	5	-.36628	.49222
6	5	.52464	.12664
5	3	-1.19887	.54613
6	6	.30855	-.89017
4	5	-.04235	1.36517
5	6	-.71231	-2.04030
5	5	.42001	1.34243
5	4	-1.60745	-.18395
5	4	-.82265	1.02229

デモ

- 主成分得点間の相関を調べる
「分析」→「相関」→「2変量」



- 新たに算出された変数
2つを右図のように選択



デモ

- 出力されたデータを見ると主成分得点間の相関係数は「 $r = 0$ 」、無相関になっている
→主成分同士が無相関

相関係数

		REGR factor score 1 for analysis 1	REGR factor score 2 for analysis 1
REGR factor score 1 for analysis 1	Pearsonの相関係数	1	.000
	有意確率 (両側)		1.000
	N	20	20
REGR factor score 2 for analysis 1	Pearsonの相関係数	.000	1
	有意確率 (両側)	1.000	
	N	20	20

参考文献

- 足立浩平(2006)多変量データ解析法ー心理・教育・社会系のための入門ー ナカニシヤ出版
- 加藤剛(2013) 本当に使えるようになる多変量解析超入門 技術評論社
- 小塩真司HP:心理データ解析Basic
(http://www.f.waseda.jp/oshio.at/edu/data_b/top.html)
- 南風原朝和・平井洋子・杉澤武俊(2009)心理統計学ワークブックー理解の確認と深化のために 有斐閣
- 市場調査・アンケート調査のマクロミルHP
ー主成分分析(<http://www.macromill.com/landing/words/b007.html>)
ー因子分析(<http://www.macromill.com/landing/words/b001.html>)
- 松尾太加志・中村知靖(2002) 誰も教えてくれなかった因子分析ー数式が絶対に出てこない因子分析入門ー 北大路書房
- 村山航さん(Department of Psychology, University of California, Los Angeles)HP
(<http://www4.ocn.ne.jp/~murakou/factor.htm>)