

統計学(基礎)

第13回 多変量解析

多変量解析とは

- 複数の変数を同時に扱う手法
 - 「多変量解析」という手法があるわけではない
 - 変量とは変数のこと
 - 重回帰分析も多変量解析
- 編集を集約するか、モデルを作るもの
 - 主成分分析、因子分析が有名
 - 看護研究では質問紙検査データでよく利用

多変量解析

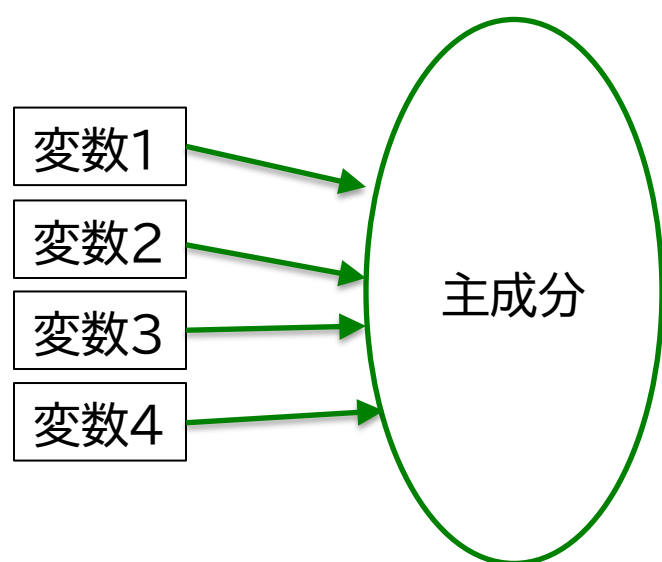
- 絶対的な正解があるわけではない
 - モデルの考え方、手法の選択など方法がいくつかある
 - あまりやり過ぎると、仮説検証ではなくて数理モデルの構築になる
 - 一般性よりそのデータへのフィッティングが目的になってしまう
 - 基本的には、複数の変数でシンプルに説明できるようにすることが目的
 - 結果予測ができるけど、予測精度にとらわれるとオーバーフィッティングの問題が起きる

多変量解析

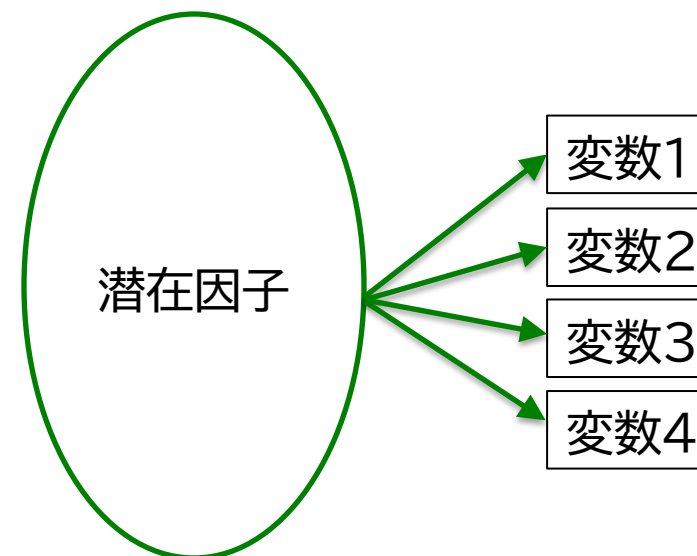
- 複数の変数(観測変数)で何かを説明する
 - 結果として予測もできるけど、あくまで説明が主
 - 投入した変数の全体が変われば、説明も変わるし予測もできない
 - 説明が主なので、シンプルな説明を目指す
 - 「予測精度」を上げようとする、モデルが複雑になる
 - 予測精度よりは変数間の関係性の説明が大事

主成分分析と因子分析の違い

- 変数を集約するか、潜在因子の実態として変数を扱うか



主成分分析



因子分析

主成分分析と因子分析の違い

主成分分析

- 変数を縮約することが目的
- 軸の回転はしない
 - 極力少ない主成分で説明
 - 主成分間の相関は仮定しない
- 正規性はあまり気にしない
 - 方法は主因子法のみ

因子分析

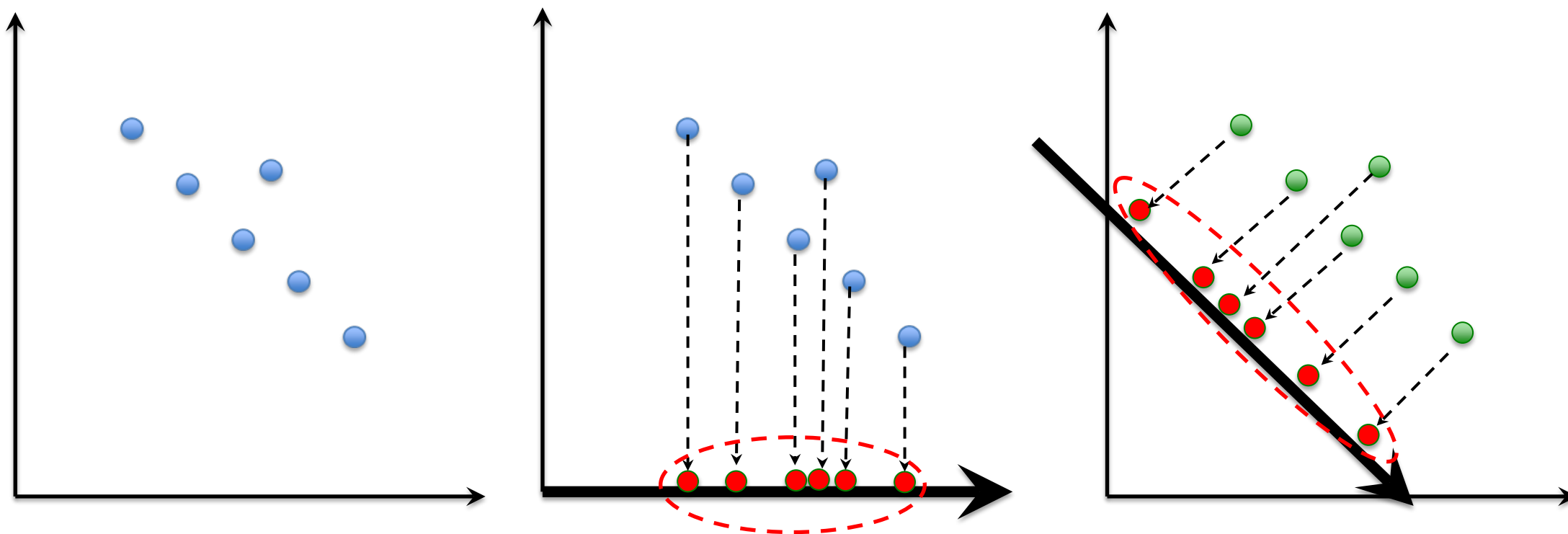
- 潜在因子を仮定し、各変数(観測変数)への影響度を調べる
- 軸を回転して説明しやすくする
 - 因子が複数あってもいい
 - 因子間相関があってもいい
- 原則正規分布(保証できないときは手法を変える)

主成分分析

主成分分析(PCA)とは

- 主成分分析(PCA: Principal Component Analysis)
- 分散が最大となる方向を探す
 - 第1主成分→最も広がる方向
- 多くの変数を少数の主成分に要約
 - 基本は第1主成分でほとんどを説明できるように変数を集約する
 - 第3主成分以降はあまり考えない
 - 説明が複雑になるから

分散が最大となる方向を探す



主成分分析で分かること

- 各変数がどの主成分にどれだけ関連するか
- データの全体構造

主成分とは

- 目的は要約なのでデータの分布形より相関構造が重要
 - 極力少ない主成分に縮約することが目的
- 軸の回転は基本しない
 - (第1)主成分の寄与率を最大にすることが目的なので、軸を回転させて複数の主成分の解釈をするようなことは基本的にしない
 - 主成分間は直交

正規性の扱い

- 主成分分析は正規性を前提としない
 - 極端な非正規や外れ値があると主成分方向が歪む可能性はある
 - 外れ値に注意した方がいい
 - 主成分分析も因子分析も相関をもとに構造を理解する手法

主成分数の決定

- 今は平行分析(Parallel Analysis)による決定が一般的
 - 実データから得た固有値とランダムに作ったデータから得た固有値を比べて、「ランダムより大きい主成分だけ残す」方法
 - ノイズによる主成分を排除できる
 - 固有値の過大評価を防ぐ
- 以前は「固有値 > 1 」となる主成分を選択
- スクリーンプロットで確認

実際の手順

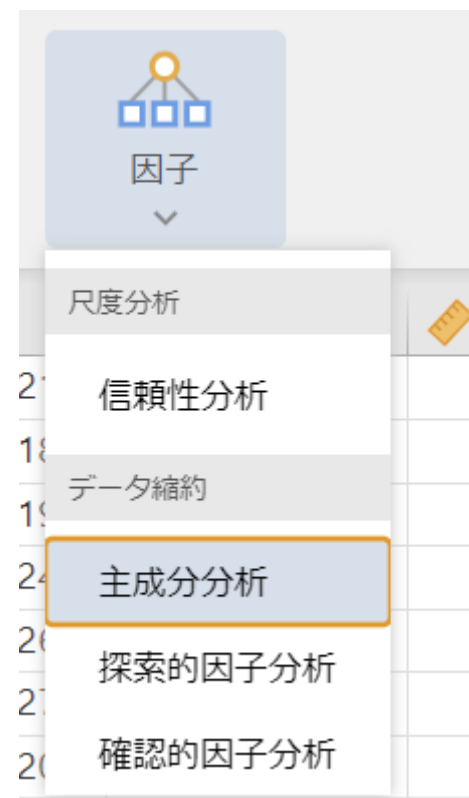
- data13_01
 - BP_sys(収縮期血圧)
 - BP_dia(拡張期血圧)
 - HR_rest(安静時心拍数)
 - BMI
 - Steps(歩数)
 - Sleep_h(睡眠時間)

実際の手順

JASP



jamovi



手順(JASP)

▼ 主成分分析

変数

BP_sys

BP_dia

HR_rest

BMI

Steps

Sleep_h

Data

☒ Raw

☐ Variance-covariance matrix

Sample size

▼ Number of Components

Based on

☒ 平行分析

☒ PCに基づく

☐ FAに基づく

再現性

☐ シードを設定します。

☐ 固有値

以上の固有値

☐ 手動

成分数

何故かデフォルトで斜交回転 (promax) が指定されているので「直交」の「なし」にする

▼ Analysis Options

Rotation Method

☒ 直交

なし

▼

☐ 斜交

promax

▼

Base Decomposition on

☒ 相関行列

☐ 共分散行列

☐ ポリコリック/テトラコリック相関行列

▼ 出力オプション

上記の負荷量を表示する

0

表

☐ 成分間相関

☐ 残差行列

☐ 平行分析

☐ PCに基づく

☐ FAに基づく

仮定のチェック

☐ KMO検定

☐ Bartlettの検定

☐ Mardiaの検定

Order Loadings By

☒ Size

☐ 変数

プロット

☐ バス図

☒ スクリーンプロット

☒ 平行解析結果

欠損値

☒ ペアワイズでケースを除外

☐ リストごとにケースを除外する

手順(jamovi)

何故かデフォルトでバリマックスが指定されているので「なし」にする

主成分分析

→

変数

BP_sys

BP_dia

HR_rest

BMI

Steps

Sleep_h

方法

回転法

なし

主成分数

☒ 平行分析に基づく

☐ 固有値に基づく

☐ 数値指定

固有値の最小基準値

1

1

個の主成分を抽出

前提チェック

☒ バートレットの球面性検定

☐ KMO標本妥当性指標

因子負荷

これ未満の負荷量を隠す

0

☒ 負荷量順に並び替え

追加の出力

☒ 主成分の要約量

☐ 主成分の相関

☐ 回転前の固有値

☒ スクリーンプロット

出力

- 成分負荷量と成分特性

成分負荷量

	PC1	PC2	独自性
BP_sys	0.943	-0.027	0.110
BP_dia	0.922	0.015	0.149
BMI	0.898	-0.046	0.191
HR_rest	0.869	0.083	0.239
Steps	-0.047	0.941	0.113
Sleep_h	0.026	0.946	0.105

注 回転方法は適用されませんでした。

成分特性

	固有値	分散説明率	累積
Component 1	3.303	0.551	0.551
Component 2	1.789	0.298	0.849

分散説明率
= 寄与率
累積
= 累積寄与率

主成分負荷量

	主成分		独自性
	1	2	
BP_sys	0.9428	-0.0268	0.110
BP_dia	0.9223	0.0149	0.149
BMI	0.8980	-0.0457	0.191
HR_rest	0.8686	0.0829	0.239
Sleep_h	0.0262	0.9457	0.105
Steps	-0.0470	0.9407	0.113

ノート、「回転なし」の結果です

[3]

主成分統計量

要約

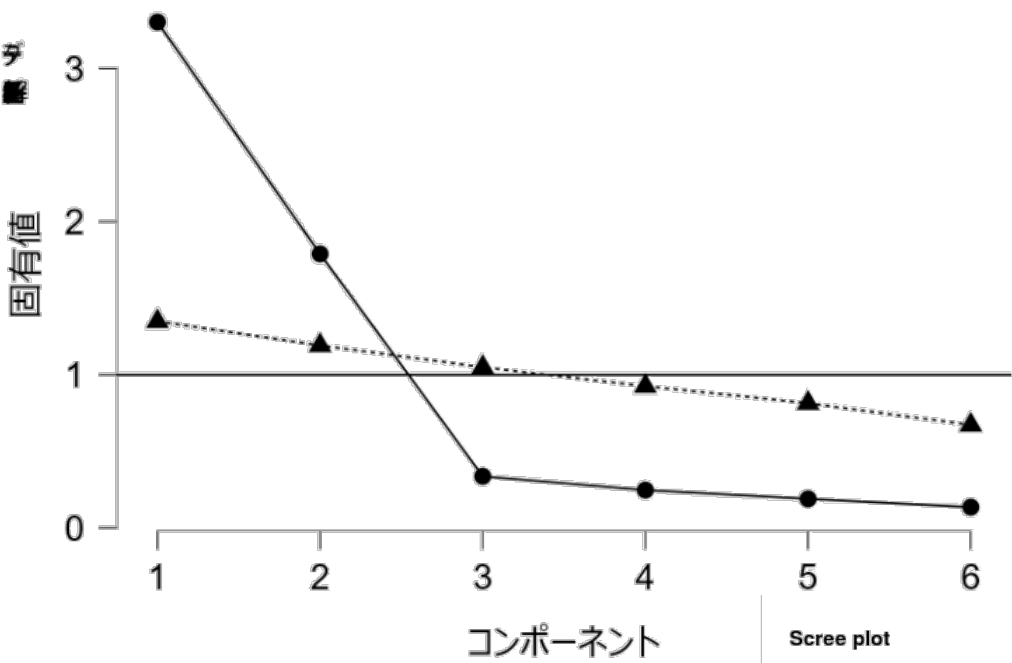
主成分	負荷量2乗和	分散説明率 (%)	累積%
1	3.30	55.1	55.1
2	1.79	29.8	84.9

出力

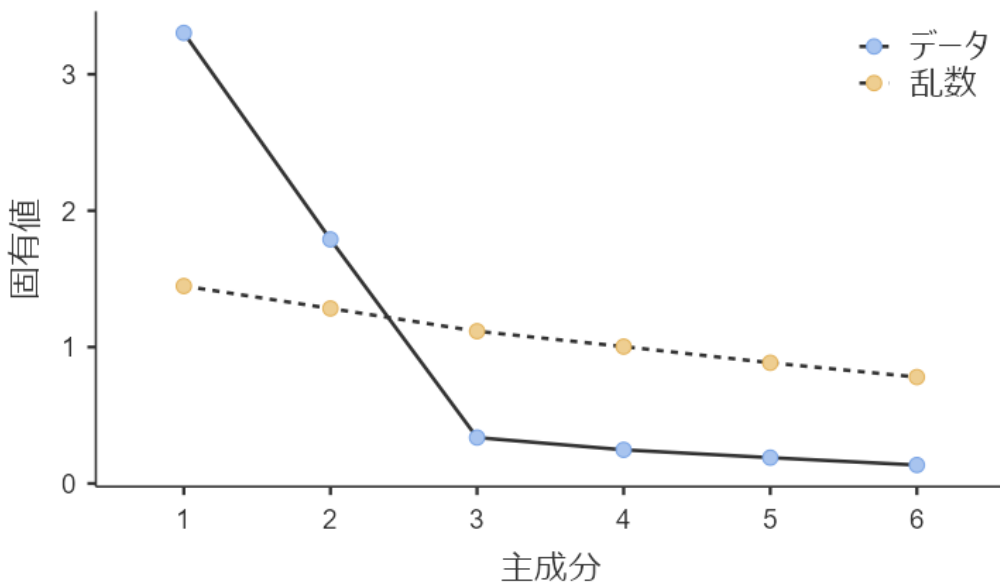
- 主成分負荷量
 - 主成分に対するその変数の重み付け
- 固有値
 - その主成分がどれだけ分散を説明しているか
- 分散説明率(寄与率)
 - その固有値を全体の分散で割った割合
 - 主成分分析ではデータを標準化しているので、全体の分散は「変数の数」になるから寄与率は $\text{固有値} \div \text{変数の数}$

スクリープロット

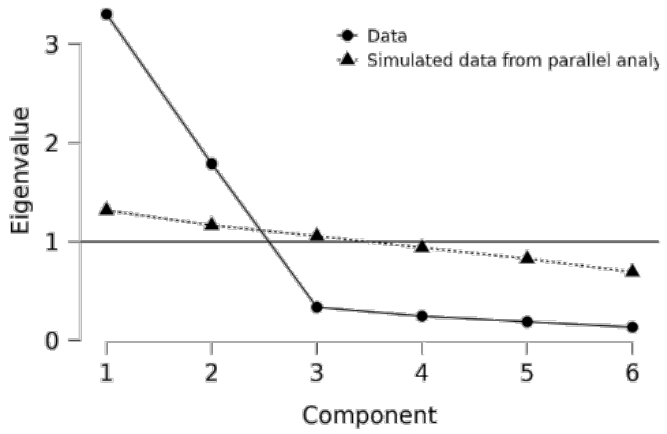
スクリープロット



スクリープロット



Scree plot



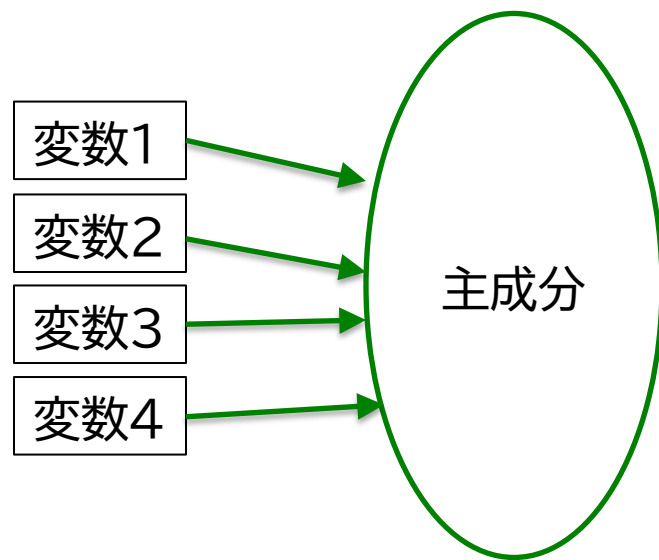
因子分析

因子分析(探索的因子分析:EFA)とは

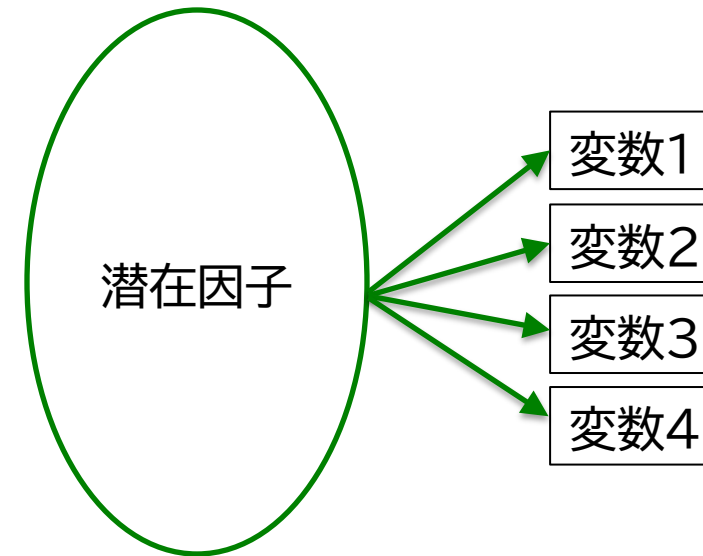
- 探索的因子分析(EFA: Exploratory Factor Analysis)
- 背後の潜在因子を推定する分析
 - どの変数がどの因子に属するかは、データに決めさせる
- 主成分分析とは目的が異なる
 - 確認的因子分析(CFA: Confirmatory Factor Analysis)
 - 確認的因子分析は構造方程式モデリング(共分散構造分析)で行う潜在変数を用いた分析。すでに変数と潜在変数の関係が明らかな状態で、係数を求める
 - どの変数がどの因子に属するかは、分析者が決めて分析する

主成分分析と因子分析の違い(再)

- 変数を集約するか、潜在因子の実態として変数を扱うか



主成分分析



因子分析

因子モデルのイメージ

- 潜在因子→観測項目に影響
- 因子負荷量＝関連の強さ

因子分析:方法の違い

- 因子分析は「潜在因子が観測変数を生む」というモデルに基づく
 - 最小残差法(MINRES: MINimum RESidual)
 - 正規性を前提としない。最尤法に似ている
 - 最尤法(maximum likelihood)
 - 解釈がしやすい 多変量正規を前提とする手法
 - 主因子法(Principal axis factoring)
 - 正規性の仮定が弱く、以前はよく使われていた

因子分析の正規性の扱い

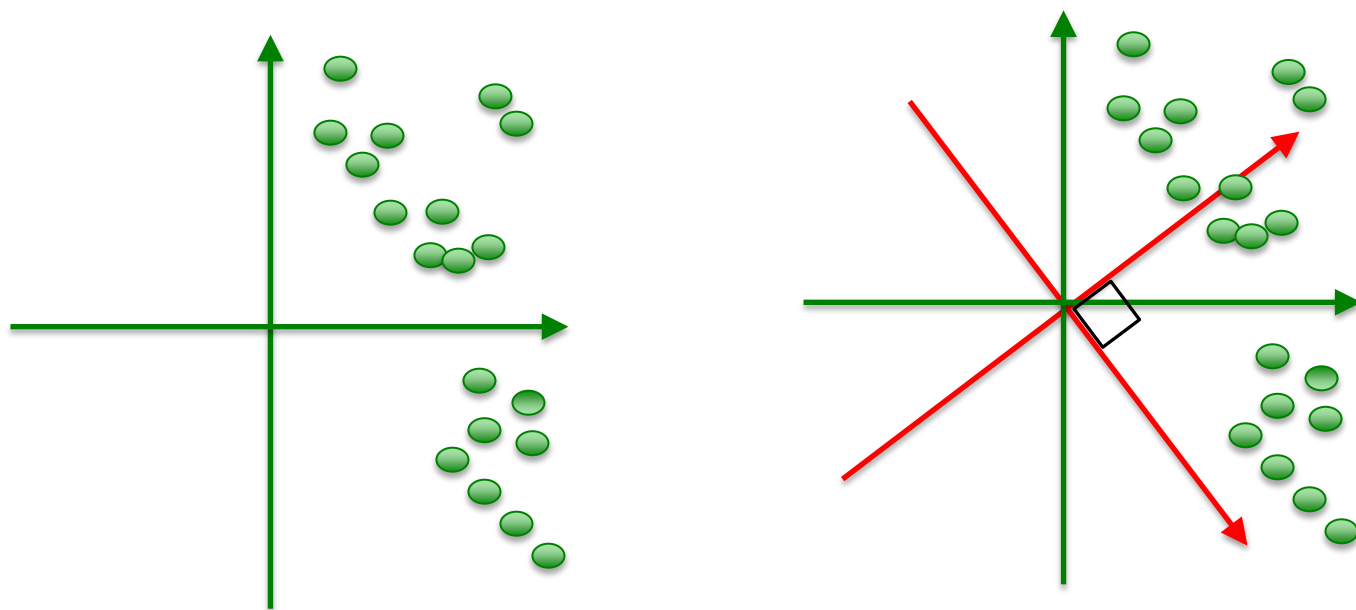
- 極端に偏った分布でなければ大きな問題は生じにくい
 - 探索的因子分析は比較的ロバスト
 - 正規分布の仮定ができないなら最尤法以外の方法で行う
 - 外れ値や極端なゆがみがある場合や、順序尺度では最尤法を選ばない

因子軸の回転

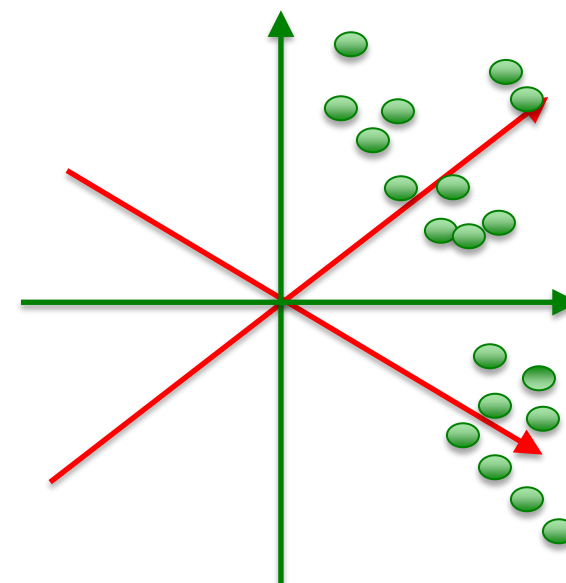
- 因子負荷量が単純構造になることを目指して回転
- 直交回転
 - 因子同士が垂直に交わる(直交)
 - 因子間相関が0 ($\cos 90^\circ = 0$)
- 斜交回転
 - 因子間の角度は任意に設定
 - 因子間の相関を仮定
- 因子間相関が仮定できない場合を除き、基本斜交回転

因子軸の回転のイメージ

直交回転



斜交回転



因子軸の回転

- 直交回転
 - バリマックス(varimax)回転
 - 分散(variance)を最大(max)にする方法。計算が簡単で以前はよく使われていた
- 斜交回転
 - プロマックス(promax)回転
 - バリマックス回転の結果からさらにプロクラテス回転を行う。計算が速いが、今は積極的に使う理由がない
 - オブリミン(oblimin)回転
 - 因子負荷行列の共分散を最小にして、単純構造を目指す方法
 - シンプリマックス(simplimax)回転
 - プロマックスの改良版。1990年台に登場。解釈が少し難しい

因子負荷量の推定方法と回転の選択

- 絶対的正解があるわけではない(特に回転)
- 正規性が仮定できる(orデータが300以上ある)
 - 最尤法でオブリミン回転
- 正規性が仮定できない
 - ミンレス(最小残差法)でオブリミン回転
- 絶対的に因子間相関が仮定できない場合はバリマックス回転でも可(もしくは先行研究との対比)

因子数の決め方

- 主成分分析と考え方は一緒
 - 平行分析(Parallel Analysis)による決定が一般的
 - ただし、主成分分析では相関行列(共分散行列)の固有値を使用しているが、因子分析は共通性を考慮した因子モデル由来の固有値を使用しており手法が違う
 - 以前は「固有値 >1 」となる主成分を選択
 - スクリーンプロットで確認

実際の手順

- data13_02 150名分のデータ

- 1～5 の Likert 尺度(「全くそう思わない」～「とてもそう思う」)

ストレス因子(ST)

ST1_寝つきが悪い

ST2_気分が張りつめている

ST3_疲れが取れにくい

ST4_イライラしやすい

ST5_最近いろいろなことが気になる

自己効力感因子(SE)

SE1_新しいことにも挑戦できる

SE2_問題があっても自分で解決できると思う

SE3_多少の失敗は乗り越えられる

SE4_自分の力を信じて行動できる

SE5_やる気があれば何でもできると思う

ソーシャルサポート因子(SS)

SS1_相談できる人がいる

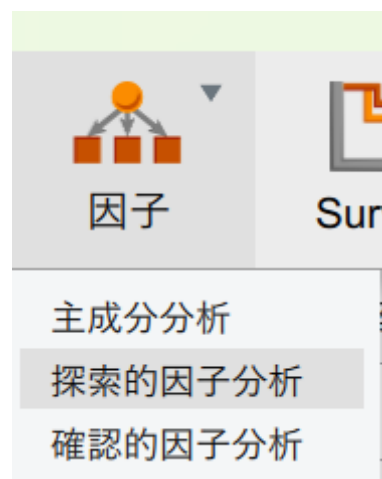
SS2_職場や学校で支えてくれる人がいる

SS3_家族に支えられていると感じる

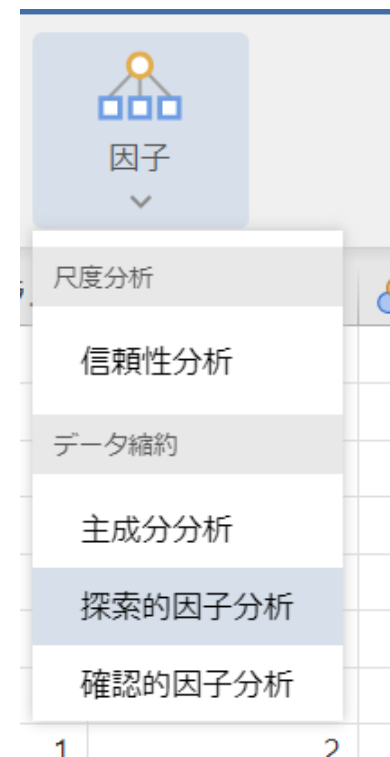
SS4_困ったとき助けてくれる人が多い

実際の手順

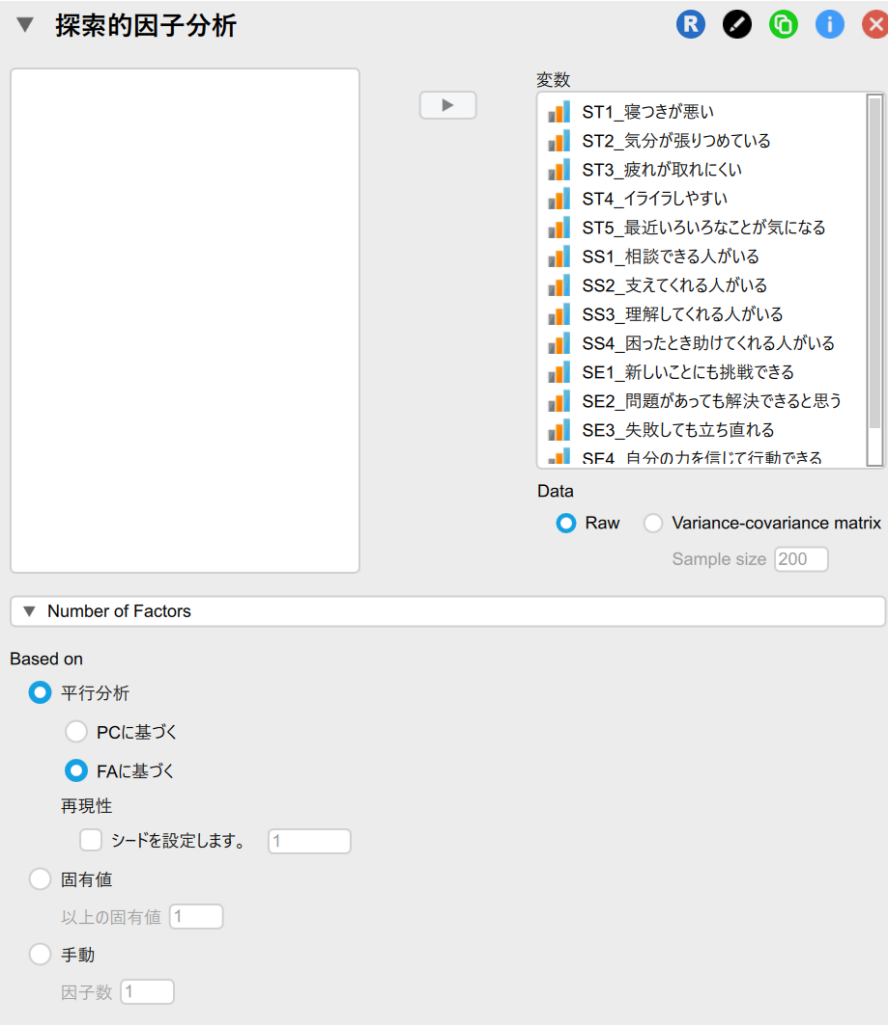
JASP



jamovi



手順(JASP)



ポリコリック相関・テトラコリック相関

- ポリコリック相関
 - 対象: 順序尺度(リッカート尺度など)
 - 前提: 背後に 連続・正規分布する潜在変数があり、それをカテゴリ化した結果が観測変数
- テトラコリック相関
 - 2値版(Yes/No, 0/1)
- ピアソンの積率相関がダメという訳ではない
 - 5件法以上なら、実務的には大きな問題にならないことも多い

手順(JASP)

▼ 出力オプション

上記の負荷量を表示する

0.4

表

☐ 構造行列

☒ 因子間相関

☐ 追加の適合指数

☐ 残差行列

☒ 平行分析

☐ PCに基づく

☒ FAに基づく

仮定のチェック

☒ KMO検定

☒ Bartlettの検定

☐ Mardiaの検定

☐ Anti-image correlation matrix

☐ Add FA scores to data

Prefix

Order Loadings By

☒ Size

☐ 変数

プロット

☐ パス図

☒ スクリーンプロット

☒ 平行解析結果

欠損値

☒ ペアワイズでケースを除外

☐ リストごとにケースを除外する

手順(jamovi)

探索的因子分析

変数

←

SS4_困ったとき助けてくれる人が...
SS3_理解してくれる人がいる
SS2_支えてくれる人がいる
SS1_相談できる人がいる
ST5_最近いろいろなことが気にな...
ST4_イライラしやすい
ST3_疲れが取れにくい

方法

抽出法 最小残差法 ▾

回転法 オブリミン ▾

因子数

☒ 平行分析に基づく

☐ 固有値に基づく

☐ 数値指定

固有値の最小基準値 0

1 個の因子を抽出

前提チェック

☒ バートレットの球面性検定

☒ KMO標本妥当性指標

因子負荷

これ未満の負荷量を隠す 0.3

☒ 負荷量順に並び替え

追加の出力

☒ 因子の要約量

☒ 因子の相関

☐ モデル適合度指標

☒ 回転前の固有値

☒ スクリーンプロット

探索的因子分析 ▼

Kaiser-Meyer-Olkin 検定 ▼	
	MSA
全体的な MSA	0.828
SE1_新しいことにも挑戦できる	0.857
SE2_問題があっても解決できると思う	0.834
SE3_失敗しても立ち直れる	0.839
SE4_自分の力を信じて行動できる	0.813
SE5_やる気があれば何でもできると思う	0.301
SS1_相談できる人がいる	0.832
SS2_支えてくれる人がいる	0.846
SS3_理解してくれる人がいる	0.830
SS4_困ったとき助けてくれる人がいる	0.828
ST1_寝つきが悪い	0.854
ST2_気分が張りつめている	0.814
ST3_疲れが取れにくい	0.856
ST4_イライラしやすい	0.801
ST5_最近いろいろなことが気になる	0.655

Bartlett の検定		
X ²	df	p
1032.558	91.000	< .001

Chi-Squared Test			
	値	df	p
Model	54.136	52	.393

出力

前提チェック

バートレットの球面性検定		
χ^2	自由度	p
846	91	< .001

KMO 標本妥当性指標 (MSA)	
	MSA
全体	0.829
SE5_やる気があれば何でもできると思う	0.302
SE4_自分の力を信じて行動できる	0.821
SE3_失敗しても立ち直れる	0.841
SE2_問題があっても解決できると思う	0.821
SE1_新しいことにも挑戦できる	0.863
SS4_困ったとき助けてくれる人がいる	0.828
SS3_理解してくれる人がいる	0.835
SS2_支えてくれる人がいる	0.837
SS1_相談できる人がいる	0.832
ST5_最近いろいろなことが気になる	0.675
ST4_イライラしやすい	0.802
ST3_疲れが取れにくい	0.853
ST2_気分が張りつめている	0.825
ST1_寝つきが悪い	0.842

Bartlett の球面性検定と KMO

- 因子分析を「やってよいか」を確認する指標
 - Bartlett: 相関があるか？
 - KMO: その相関を因子でまとめられるか？

Bartlett の球面性検定と KMO

Bartlett の球面性検定

- 変数同士に相関があるか
- 相関行列が単位行列でないか
- 有意 → まとめる意味がある

KMO(Kaiser-Meyer-Olkin)

- その相関は因子で説明できそうか
- 共通因子っぽさの指標
- 高いほど因子分析に向く
 - 少なくとも0.5以上、0.7あればいい、0.8以上ならとてもいい

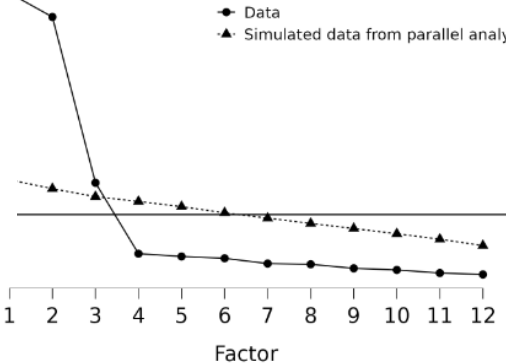
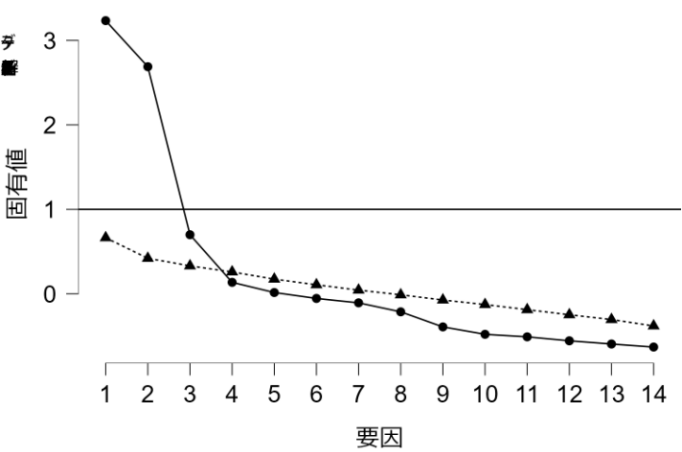
- どっちも因子分析を「やってよいか」を確認する指標
 1. Bartlett が有意 → 相関あり
 2. KMO が十分 → 因子分析に向く
- 両方確認してから因子分析

平行分析

	実データ因子固有値	シミュレーションデータ平均固有値
要因 1*	3.234	0.732
要因 2*	2.688	0.426
要因 3*	0.699	0.338
要因 4	0.136	0.247
要因 5	0.015	0.177
要因 6	-0.055	0.115
要因 7	-0.107	0.047
要因 8	-0.214	-0.008
要因 9	-0.392	-0.068
要因 10	-0.480	-0.130
要因 11	-0.510	-0.186
要因 12	-0.557	-0.248
要因 13	-0.594	-0.318
要因 14	-0.630	-0.390

注 "*" = Factor を保持すべきである。FAに基づく平行分析の結果。

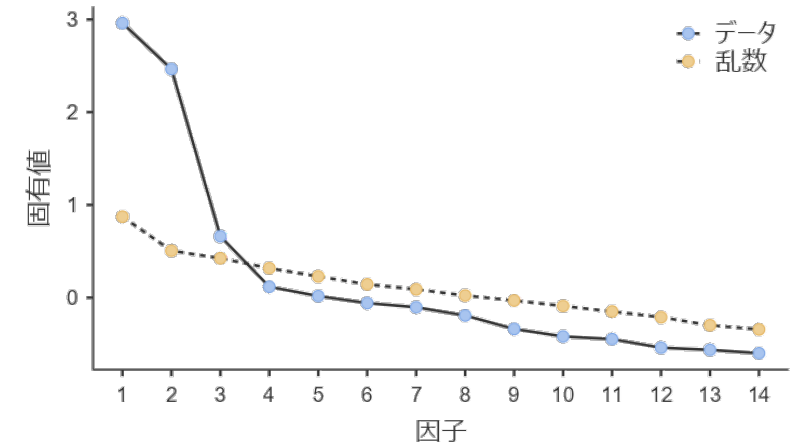
Scree plot



回転前の固有値

因子	固有値
1	2.9602
2	2.4651
3	0.6617
4	0.1176
5	0.0169
6	-0.0587
7	-0.1034
8	-0.1925
9	-0.3378
10	-0.4185
11	-0.4476
12	-0.5399
13	-0.5633
14	-0.5995

スクリープロット



出力

出力

- 平行分析と固有値
 - 固有値: 各因子が説明する共通分散量
 - 固有値 > 1 という基準で因子を採用
 - 平行分析の結果
 - 乱数から作成される固有値より大きい因子までを採用
- スクリーンプロット
 - 固有値をプロットしたもの
 - ガクンと落ちてるところに注目

因子負荷量

	因子 1	因子 2	因子 3	独自性
SE3_失敗しても立ち直れる	0.836			0.307
SE2_問題があっても解決できると思う	0.834			0.319
SE4_自分の力を信じて行動できる	0.833			0.300
SE1_新しいことにも挑戦できる	0.749			0.361
SS4_困ったとき助けてくれる人がある		0.837		0.330
SS2_支えてくれる人がある		0.823		0.344
SS1_相談できる人がある		0.736		0.334
SS3_理解してくれる人がある		0.729		0.431
ST4_イライラしやすい			0.802	0.382
ST1_寝つきが悪い			0.770	0.411
ST3_疲れが取れにくい			0.756	0.316
ST2_気分が張りつめている			0.730	0.476
SE5_やる気があれば何でもできると思う				0.993
ST5_最近いろいろなことが気になる				0.967

注 Applied rotation method is oblimin.

因子特性

	非回転解			回転解		
	固有値	分散説明率	累積	負荷量の平方和	分散説明率	累積
Factor 1	3.487	0.249	0.249	2.767	0.198	0.198
Factor 2	3.203	0.229	0.478	2.549	0.182	0.380
Factor 3	1.039	0.074	0.552	2.413	0.172	0.552

因子間相関 ▼

	Factor 1	Factor 2	Factor 3
Factor 1	1.000	0.182	-0.297
Factor 2	0.182	1.000	0.350
Factor 3	-0.297	0.350	1.000

出力

因子負荷

	因子			独自性
	1	2	3	
SE2_問題があっても解決できると思う	0.815			0.360
SE3_失敗しても立ち直れる	0.800			0.359
SE4_自分の力を信じて行動できる	0.787			0.366
SE1_新しいことにも挑戦できる	0.716			0.413
ST5_最近いろいろなことが気になる				0.969
SE5_やる気があれば何でもできると思う				0.996
SS4_困ったとき助けてくれる人がある		0.796		0.395
SS2_支えてくれる人がある		0.787		0.402
SS1_相談できる人がある		0.708		0.381
SS3_理解してくれる人がある		0.697		0.482
ST4_イライラしやすい			0.778	0.421
ST1_寝つきが悪い			0.746	0.448
ST3_疲れが取れにくい			0.724	0.366
ST2_気分が張りつめている			0.691	0.531

ノート: 「最小残差法」で抽出した因子に「オブリミン回転」を適用した結果です

[3]

因子統計量

要約

因子	負荷量2乗和	分散説明率 (%)	累積%
1	2.55	18.2	18.2
2	2.33	16.6	34.8
3	2.23	15.9	50.8

因子間の相関

	1	2	3
1	—	0.184	-0.281
2		—	0.349
3			—

出力

- 因子負荷量(の推定値)
 - 潜在因子からの影響の強さの推定値
- 因子特性表(要約表)
 - 固有値(因子負荷量平方和): 回転前の共通分散量(説明量)
 - 負荷量平方和: 回転後の因子の説明量
 - 分散説明率: 寄与率
 - 累積: 累積寄与率
- 因子間相関

解釈

- KMOの低い観測変数
- どの因子にも高負荷量にならない観測変数
 - 役に立たないかも
- 因子間相関の有無
 - 無かったら直交回転の方がいいかも
- 出てきた因子が思ったように潜在因子として説明できるか
→条件を変えてやり直す

なんか変だぞ

少し説明

JASP と jamovi で結果が少し違う

- 同じ因子負荷量設定、回転設定(MINRES・oblimin)
- でも、結果が完全一致しないことはよくある

理由① MINRES は反復最適化法

- 最小残差法(MINRES) は一意解を与える方法ではない
 - 残差を最小化する方向に少しずつ解を探す
 - 初期値・更新順・停止条件が違うと、微差が生じる
 - JASPとjamoviで初期共通性の与え方、収束判定、数値丸めが微妙に違う
 - 小数2～3位で差が生じる

理由② オブリミン回転は解が一意でない

- 斜交回転は因子間相関を許す
 - 回転解が一意に定まらない
 - 回転の初期位置最適化アルゴリズムの差で、因子負荷量や因子相関が少し違う

理由③ 相関行列の設定

- ポリコリック相関とピアソン相関の違い
 - jamoviはポリコリック相関を採用していないので結果に影響することがある

重要なのは数値の一致ではない

- 高負荷項目の配置が同じか
 - どの観測変数の項目がどの因子に載るか
 - 因子の意味づけ(解釈)が一致しているか
- これが一致していれば、同じ因子構造
 - 小数点以下の違いを追う意味は無い

重要なのは数値一致ではない

- 問題なし
 - 因子1と因子2が入れ替わる
 - 負荷量の符号が全体に反転する($+\leftrightarrow-$)
 - 要注意
 - 因子数が違う
 - 高負荷項目(観測変数)が全然違う因子に載る
 - 因子間相関が極端に違う
- 負荷量の推定方法、回転方法、相関行列の種類、欠損値処理などが違っている

まとめ

- JASP と jamovi の微差は正常
 - MINRES+oblimin では特に起きやすい
 - 因子構造が同じなら結果は同等と考えてよい
 - わざわざ両方のソフトで実施する必要は無い

ほんとのまとめ

まとめ

まとめ

主成分分析

- データの要約
- 変数の情報を主成分が「受け取って要約」する
- 主成分は変数の線形結合で作られる
- 矢印の向き: 変数 \rightarrow 主成分

因子分析

- 潜在因子の探索
- 潜在因子が観測変数を「生み出す」モデル
- 因子が原因、観測変数は結果
- 矢印の向き: 因子 \rightarrow 変数

まとめ

- 主成分分析・因子分析の本質は、「データに合わせる」ものではなく、理論的仮説を数式化し検証する方法
 - 何を説明したいのかを明確にする
 - 投入する変数は仮説から吟味する
 - オーバーフィッティングに注意
- 分析の目的はモデルを説明すること
 - シンプルな方が理解がしやすい
 - 結果的にモデルができるので予測ができるが、予測を重視しすぎると、オーバーフィッティングの問題が発生