因子分析

因子分析

- 複数の変数に共通して影響を与えるもの(因子)と変数の関係を分析する
  - 探索的因子分析の目的は得られたデータをもとに因子を探すこと
  - 確認的因子分析の目的は想定した因子構造にデータが適合するかを検討すること
- 探索的因子分析(Exploratory Factor Analysis; EFA)
  - 複数の変数に共通して影響を与えるもの(＝因子)を探す分析
  - 因子構造の仮説がない時や，確認的因子分析のあてはまりが悪かった場合に使われる
    - 違う言語で作られた尺度を翻訳して使う場合など
    - 因子構造(factor structure)：どのような因子があって，どの変数がどの因子から影響を受けているか
- 確認的因子分析(Confirmatory Factor Analysis; CFA)
  - あらかじめ想定した因子構造にデータがあてはまるかを確認する分析
  - 因子構造に関する仮説がある場合はこちらを使う事が多い
    - 新しく尺度を作る時など

因子分析のイメージ 相関が高い変数同士でグループを作る

- A，B，Cの間には高い相関関係
- D，E，Fの間にも高い相関関係
- ABCとDEFの間は高い相関関係はない
- ABCとDEFはそれぞれ似たような質問をしている可能性
→ABC，DEFの背後にそれぞれ共通する概念(＝因子)があるのではないか？
- 因子分析では変数間の相関関係の情報から因子を抽出する

	A	B	C	D	E	F
A	1	.608	.575	.211	.200	.184
B	.608	1	.615	.146	.141	.124
C	.575	.615	1	.123	.130	.102
D	.211	.146	.123	1	.636	.632
E	.200	.141	.130	.636	1	.646
F	.184	.124	.102	.632	.646	1

A	私が受けてきた教育は、私の能力を高めてくれている
B	私が受けてきた教育は、私の人間関係を豊かにしてくれている
C	私が受けてきた教育は、私を幸福にしてくれている
D	教育の力がなくても人間のできることはたくさんある
E	教育の力がなくても豊かな人間関係を作ることはできる
F	教育の力がなくても人は幸せになれる

探索的因子分析のモデル

$$ZA_i = a_{a1}f_{1i} + a_{a2}f_{2i} + u_{ai}$$
$$Z\mathrm{B}_i = a_{b1}f_{1i} + a_{b2}f_{2i} + u_{bi}$$
$$ZC_i = a_{c1}f_{1i} + a_{c2}f_{2i} + u_{ci}$$
$$\vdots$$
$$ZF_i = a_{f1}f_{1i} + a_{f2}f_{2i} + u_{fi}$$

i:対象者番号， ZA_i～ ZF_i:観測変数を平均0，標準偏差1に標準化したもの

a: 因子負荷量(因子から各項目への影響の大きさ，－1から1の間の値をとる)
f: 因子得点，u: 独自因子

この例では観測変数は6項目(ZAからZF)で因子は2つ(f1,f2)だが，項目数も因子数も変えて良い

- ZA-ZFが観測変数(得られたデータ)を標準化(元のデータから平均を引いて，標準偏差で割ったもの)したもの
→因子分析で扱える観測変数は**量的変数**
- f1，f2が因子
  - データを元に因子負荷量を推定する→因子抽出(factor extraction)
- 3つの仮定
  - 因子得点と独自因子はどれも平均0，分散1の正規分布に従う
  - 因子得点と独自因子は互いに無相関
  - 独自因子は互いに無相関

探索的因子分析のステップ

1. 分析する項目を選ぶ
2. 因子の数を決める
3. 因子の抽出方法を決める
4. 因子の回転を行うか，行うならどの方法を使うか決める
5. 分析を実行する
6. 結果を確認する
   - 共通性
   - 因子抽出の妥当性
   - 因子負荷量
7. 因子に名前をつける
8. 結果に納得がいけば終了．納得がいかなければ項目の組み合わせや因子数などを変えてやり直し

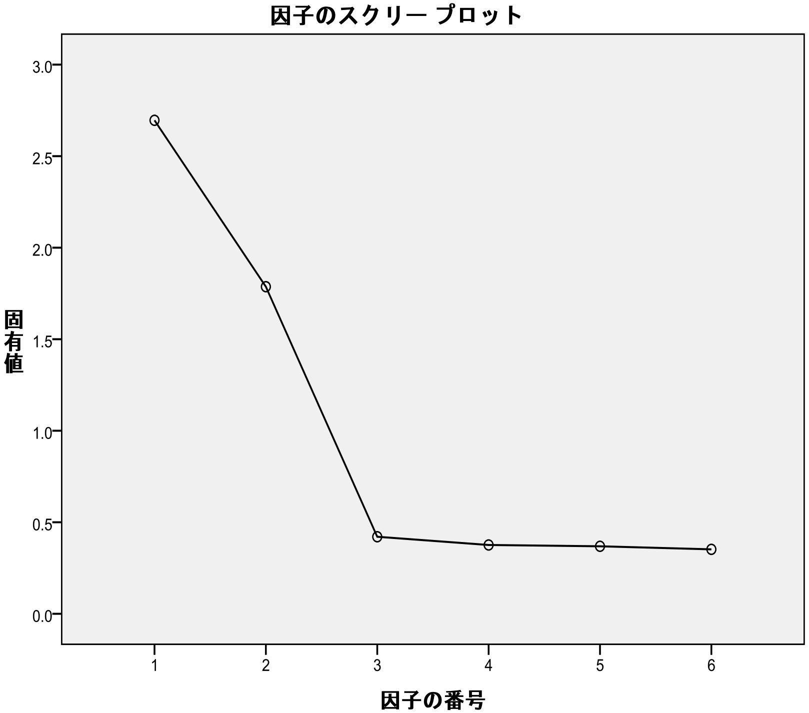
因子の数を決める

- 因子の数を決めないと分析できない
  - モデルの方程式が定まらない
  - 統計ソフトの中には決めなくても結果が出るものがあるが(SPSSなど), それはソフトで基準を作って勝手に因子数を決めて分析しているだけ
- どうやって決める?
  - 理論や仮説に基づいて
    - 特に尺度開発をする場合は測定する概念やその下位概念を定義しているはずなので，それに従うのが原則
  - 因子が解釈できるかどうか
  - 統計的な基準を参考にして
    - 固有値1以上(カイザー基準)，スクリーテスト，適合度検定，最小偏相関平均，平行テスト，寄与率など

因子数を決めるときの参考指標(1)

- カイザー基準(Kaiser criterion)
  - 固有値が1よりも大きい因子を採用
- スクリーテスト(scree test)
  - 固有値の大きさが急に小さくなる直前の因子
  - スクリープロット(固有値をプロットしたもの)を見て判断
- カイ二乗検定(Chi-squared test)
  - 因子抽出法が最尤法の場合にできる
  - カイ二乗検定が有意にならない最小の因子数を採用する
  - サンプルサイズが大きいと因子数を過大評価する傾向

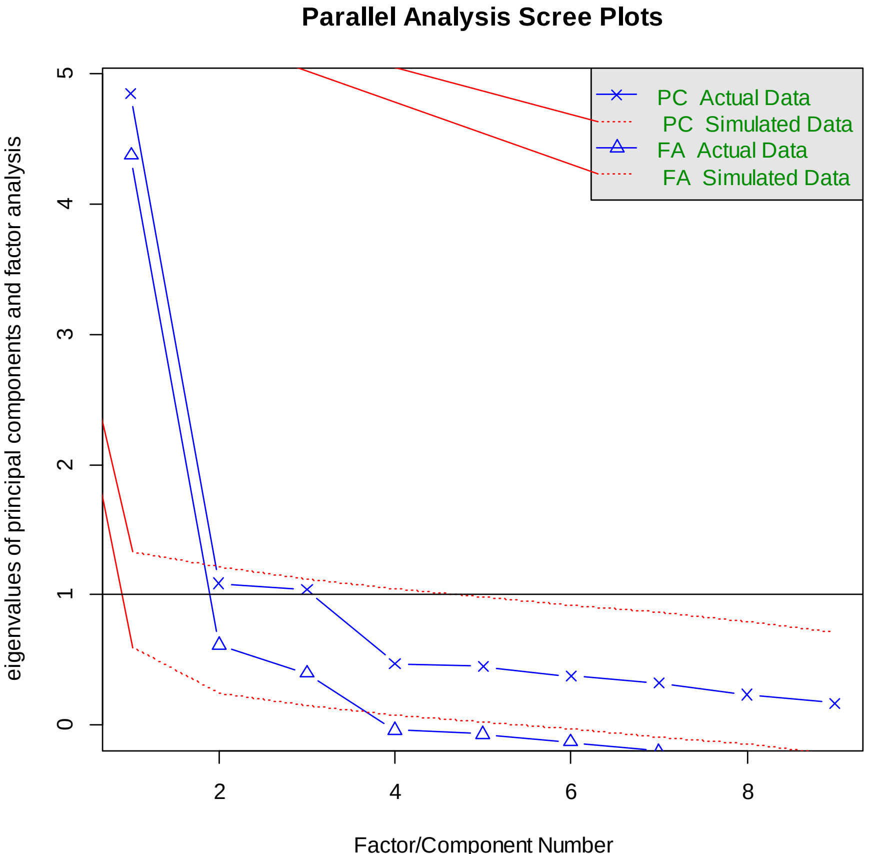
SPSSでできるのはここまで



因子数を決めるときの参考指標(2)

- 最小偏相関平均(Minimum Average Partial(correlation); MAP)
  - 主成分を制御変数とした観測変数間の偏相関係数を求め，偏相関の2乗平均が最小となる因子数を採用する方法

- 平行分析(Parallel analysis)
  - 元のデータと同じ変数数，サンプル数の正規乱数行列の相関行列の固有値と因子分析をしたときの固有値を比較
  - 正規乱数行列の相関行列の固有値よりも大きい因子数を採用する方法



- MAPと平行分析を併用して挟み込みで決める方法が有効(堀，2005)
MAPの計算や平行分析はRでできる

(参考)MAPなどの算出

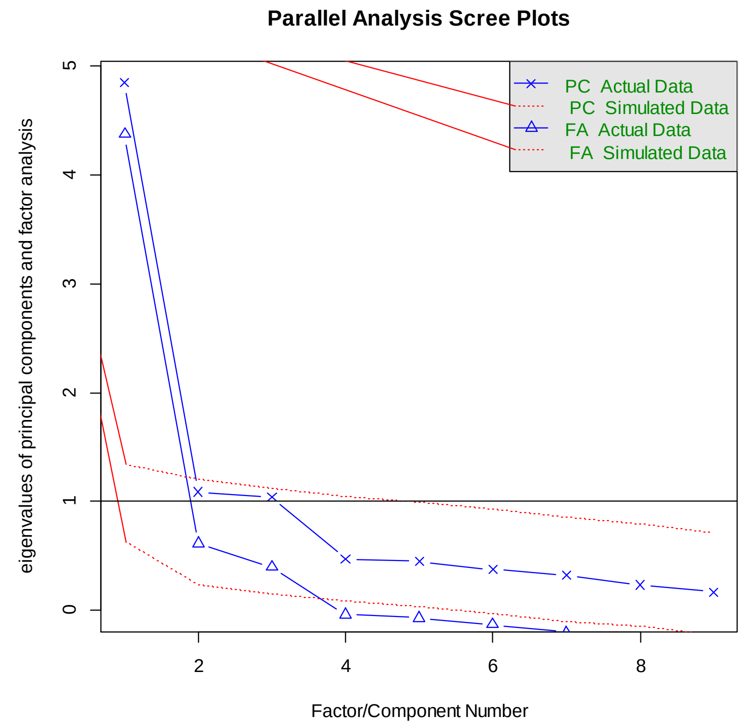
- nfactors(x,n,rotate=“varimax”,fm=“minres”,...)
  - psychパッケージ
  - MAP等の指標を算出する関数
  - 主な引数
    - x: 因子分析をするデータフレームまたは相関行列
    - n: 抽出する因子数
    - rotate:因子の回転方法．none(回転なし)，varimax(バリマックス回転),oblimin(オブリミン回転),promax(プロマックス回転)
    - fm: 因子抽出方法．minres(最小二乗法)，ml(最尤法),wls(重み付け最小二乗法),gls(一般化最小二乗法),pa(主因子法)などが選べる

Statistics by number of factors

	vss1	vss2	map	dof	chisq	prob	sqresid	fit	RMSEA	BIC	SABIC	complex	eChisq	SRMR	eCRMS	eBIC
1	0.46	0.00	0.0493	252	1927	2.3e-255	34.1	0.46	0.117	361	1161	1.0	8947	0.180	0.188	7381
2	0.84	0.85	0.0061	229	247	2.0e-01	9.5	0.85	0.014	-1176	-449	1.0	191	0.026	0.029	-1232
3	0.84	0.85	0.0083	207	202	5.8e-01	9.0	0.86	0.000	-1084	-427	1.1	149	0.023	0.027	-1137
4	0.84	0.86	0.0109	186	168	8.3e-01	8.6	0.86	0.000	-988	-398	1.1	122	0.021	0.026	-1034
5	0.84	0.86	0.0139	166	142	9.1e-01	8.3	0.87	0.000	-890	-363	1.2	99	0.019	0.024	-933
6	0.84	0.86	0.0169	147	115	9.8e-01	7.9	0.88	0.000	-799	-332	1.3	76	0.017	0.023	-838
7	0.75	0.87	0.0210	129	90	1.0e+00	7.4	0.88	0.000	-712	-302	1.3	62	0.015	0.022	-740
8	0.79	0.87	0.0249	112	69	1.0e+00	7.1	0.89	0.000	-627	-271	1.4	45	0.013	0.020	-651

(参考)平行分析

- fa.parallel(x,n.obs,fm="minres",...)
  - psychパッケージに収録されている
  - 平行分析を実行する関数
  - 主な引数
    - x: 因子分析をするデータフレームまたは相関行列
    - n.obs: サンプル数．xを相関行列にした場合は必須
    - fm: 因子抽出方法．minres(最小二乗法)，ml(最尤法),wls(重み付け最小二乗法),gls(一般化最小二乗法),pa(主因子法)などが選べる



因子抽出の方法

- よく使われるのは以下の方法
  - 最近はコンピュータの性能が上がって計算できるようになったので，最尤法が好まれる
  - 最尤法(maximum likelihood estimation)
    - 観測データが得られる確率(尤度)が最大になるように推定
    - 適合度を出せる
  - 最小二乗法(least squares estimation)
    - データから得られる相関行列とモデルから得られる相関行列のズレの二乗が最小になるように推定
  - (反復)主因子法(principal axis factoring)
    - 共通性の初期値から因子負荷を推定し，推定した因子負荷から共通性を推定というのを解が収束するまで繰り返す

結果のチェックー共通性

- 各項目が受ける因子負荷量の2乗の総和を各項目の共通性h_j^2（communality）と呼ぶ（0～1の間をとる）
- 共通性が低い項目は他の項目と異質であることを表す
→尺度を構成する項目として不適切な可能性を示す
→尺度から削除する候補になる
- 確認するポイント
  - 共通性が1を超えていないか
    - 共通性の範囲は0から1なので，1を超えている場合は不適解
  - 共通性の大きさ
    - 0に近い(どの因子からの影響力も小さい)項目は削除の候補になる

共通性

	初期	因子抽出後
Q1D_32 教育の力がなくても豊かな人間関係を作ることはできる	.506	.649
Q1D_33 私が受けてきた教育は、私の能力を高めてくれている	.450	.578
Q1D_34 私が受けてきた教育は、私の人間関係を豊かにしてくれている	.475	.649
Q1D_35 私が受けてきた教育は、私を幸福にしてくれている	.443	.584
Q1D_36 教育の力がなくても人間のできることはたくさんある	.493	.624
Q1D_37 教育の力がなくても人は幸せになれる	.501	.643

因子抽出法: 最尤法

固有値と因子寄与率

- 固有値
  - 固有値とは:
    正方行列Aに対して
    $$Ax = \lambda x$$
    となるようなベクトルxを固有ベクトル，λを固有値いう
  - SPSSでは分析する変数の相関行列の固有値を算出したものが「初期の固有値」の出力となる
  - カイザー基準やスクリー基準で因子数を決定する場合はこの結果を使う

説明された分散の合計

因子	初期の固有値			抽出後の負荷量平方和			回転後の負荷量平方和[a]
	合計	分散の %	累積 %	合計	分散の %	累積 %	合計
1	2.696	44.930	44.930	2.313	38.548	38.548	2.028
2	1.787	29.779	74.709	1.413	23.553	62.101	1.917
3	.421	7.014	81.722				
4	.376	6.265	87.987				
5	.369	6.146	94.133				
6	.352	5.867	100.000				

因子抽出法: 最尤法

a. 因子が相関する場合は、負荷量平方和を加算しても総分散を得ることはできません。

負荷量平方和と因子寄与率

- 負荷量平方和
  - 因子分析によって推定した因子負荷量を2乗して合計したもの
- 因子寄与率
  - 負荷量平方和を質問項目の数で割った値が因子寄与率
  - 各因子で全体の分散のどのくらいを説明しているかがわかる
  - 後述の斜交回転を行った場合は因子寄与率は計算できない
    - 因子間に相関があると仮定するので個々の因子独自の説明力をだすことができないため

説明された分散の合計

因子	初期の固有値			抽出後の負荷量平方和			回転後の負荷量平方和[a]
	合計	分散の %	累積 %	合計	分散の %	累積 %	合計
1	2.696	44.930	44.930	2.313	38.548	38.548	2.028
2	1.787	29.779	74.709	1.413	23.553	62.101	1.917
3	.421	7.014	81.722				
4	.376	6.265	87.987				
5	.369	6.146	94.133				
6	.352	5.867	100.000				

因子抽出法: 最尤法

a. 因子が相関する場合は、負荷量平方和を加算しても総分散を得ることはできません。