



第6回多変量解析実習(2008.5.26)

主成分分析・因子分析



小野晃典研究会第5期

千葉貴宏

加藤絵美



主成分分析・因子分析とは

- ◎ これまで学んできた回帰分析や分散分析
→独立変数 X (原因)と従属変数 Y (結果)に関するデータを収集した上で、両変数間の因果関係を明らかにする分析手法。
- ◎ 今回学ぶ主成分分析・因子分析
→独立変数や従属変数のいずれか一方に関するデータを収集して、得られた多数の変数から縮約した少数の変数を生成する分析手法。
得られた新たな変数と元々の変数との関係を見ていくことで、その新たな変数を解釈する(命名する)作業があるのが特徴です。

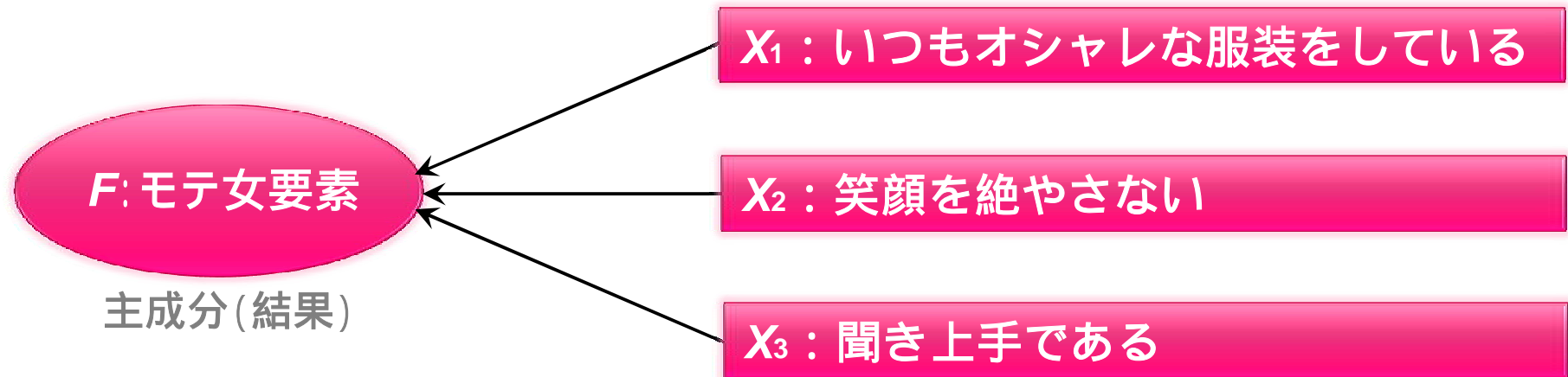
主成分分析と因子分析の違い

< 理念上の違い >

◎ 主成分分析は合成作業

【目的】

多くの生の変数を最もよくまとめる主成分（新たな変数）を生成すること。分析者が自分自身の目的のために「変数群を要約して主成分を新しく生み出す」という感覚。



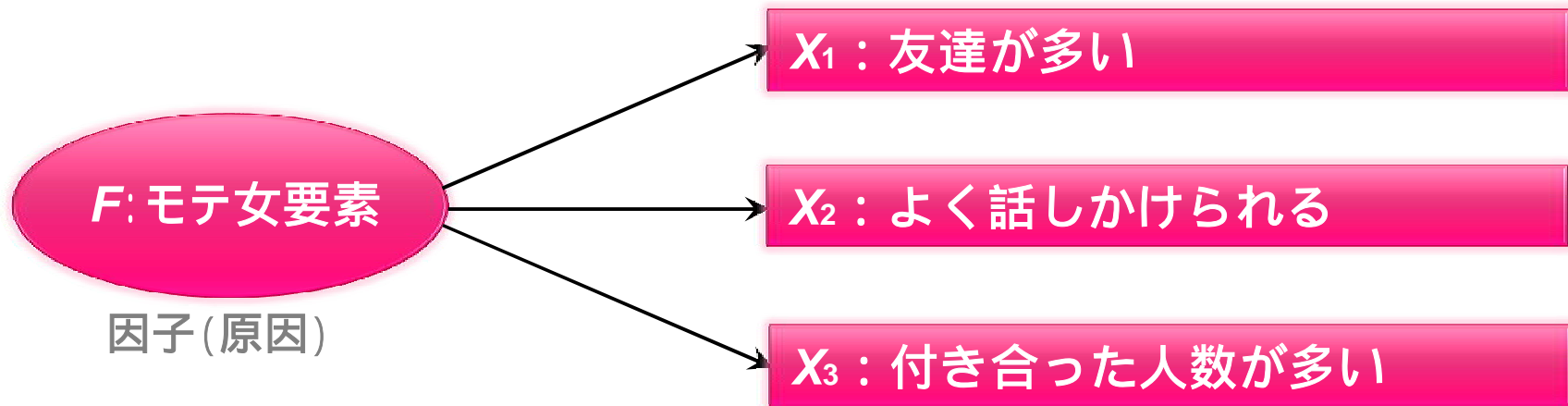
生の変数が原因、主成分が結果

主成分分析と因子分析の違い

○ 因子分析は分解作業

【目的】

多くの生の変数を最もよく説明する因子を抽出すること。
分析者が「変数を縮約してそれらの変数の背後にある潜在的な因子を探り出す」というような感覚。



因子(原因)

因子が原因、生の変数が結果

主成分分析と因子分析の違い

< 実践上の違い >

◎ 主成分分析

主成分分析は、生の変数群と最も相関の高い第1主成分を作り、主成分と無相関であるという条件の下、残りの生の変数群と最も相関の高い第2主成分を作り・・・という作業を繰り返します（これを「主因子法」といいます）。生の変数群と同数の主成分が生成されるため、多数の変数を少数の主成分にまとめるという目的を達成するために、分析後に手動で主成分を取捨選択する作業を行います。

◎ 因子分析

因子分析は、因子を抽出する方法が多様で、主成分分析と同じ「主因子法」ではなく「最尤法」という技法が推奨されています。また、抽出された因子が単純構造になるように「回転」を行う点、回転に伴って、因子の数を分析前に指定する点が主成分分析とは異なります。

主成分分析のプログラム

```
data a;  
input x1-x8;  
cards;  
【ここにデータを入力】  
;  
proc princomp data=a out=b;  
var x1-x8;  
proc print data=b;  
run;
```

主成分分析

(principal component
analysis)

主成分得点の出力

主成分得点の印刷

分析対象変数の指定

主成分分析のアウトプット

< 主成分の取捨選択 >

主成分	固有値	固有値の差異	寄与率	累積寄与率
Prin 1	3.95202	2.31358	0.494003	0.49400
Prin 2	1.63844	0.54508	0.204805	0.69881
Prin 3	1.09336	0.29343	0.136670	0.83548
Prin 4	0.79993	0.48542	0.099992	0.93547
Prin 5	0.31401	0.19314	0.039252	0.97472
Prin 6	0.12087	0.07863	0.015109	0.98983
Prin 7	0.04224	0.00312	0.005280	0.99511
Prin 8	0.03912		0.004890	1.00000

主成分分析のアウトプット

- ◎ 固有値
その主成分がいくつ分の変数を要約しているかを示す変数。
- ◎ 固有値の差異
下位の主成分との固有値の差異
- ◎ 寄与率
固有値をパーセンテージ化した数値（固有値 / 生の変数の数）
- ◎ 累積寄与率
第1主成分から該当する主成分までの寄与率を合計した数値。

主成分分析のアウトプット

< 主成分の取捨選択基準 >

- ◎ 固有値が**1**を下回ったところ
- ◎ 累積寄与率が**70%**または**80%**または**90%**を上回ったところ
- ◎ 固有値や寄与率の値の差異が大きく開いているところ
- ◎ 共通度の低い変数が存在しなくなる場所
(因子分析のみ)

主成分分析のアウトプット

< 主成分の取捨選択 >

主成分	固有値	固有値の差異	寄与率	累積寄与率
Prin 1	3.95202	2.31358	0.494003	0.49400
Prin 2	1.63844	0.54508	0.204805	0.69881
Prin 3	1.09336	0.29343	0.136670	0.83548
Prin 4	0.79993	0.48542	0.099992	0.93547
Prin 5	0.31401	0.19314	0.039252	0.97472
Prin 6	0.12087	0.07863	0.015109	0.98983
Prin 7	0.04224	0.00312	0.005280	0.99511
Prin 8	0.03912		0.004890	1.00000



主成分分析のアウトプット

< 主成分の解釈 >

◎ 固有ベクトル

注目するは、生の変数と主成分の関係を示す「固有ベクトル」。固有ベクトルは回帰分析の回帰係数に相当する値（主成分係数）から構成される行列です。採用した主成分だけに注目しつつ、各係数の正負や大小をチェックして各主成分の特性を検討し、各主成分に命名します。

表を横に見て、絶対値の一番大きな固有ベクトルをチェックしましょう！！

主成分分析のアウトプット

< 主成分の解釈 >

	prin1	prin2	prin3	prin4	...
x1	0.425000	0.374431	-0.028142	0.125327	
x2	0.374332	0.343946	-0.058029	-0.478774	
x3	0.364637	0.265005	-0.155449	0.559390	
x4	0.356195	-0.282238	0.527531	-0.191157	
x5	0.331749	-0.322025	0.544312	0.196491	
x6	-0.122727	0.598885	0.409822	-0.335708	
x7	-0.274618	0.353920	0.409345	0.502149	
x8	0.467233	-0.011942	-0.248295	0.046348	



主成分分析のアウトプット

- ◎ 主成分得点（ アウトプットの一番最後 ）

各オブザベーションは、生の変数1つ1つに対して1個ずつの値を持つのと同様に、各主成分に対しても、1個ずつの値を持ちます。これが主成分得点です。オブザベーションと各主成分との関係を示します。

主成分分析によって生成された各主成分は無相関なので、主成分得点を用いて回帰分析などの別の分析を行うこともできます。

因子分析のプログラム

```
data a;  
input x1-x8;  
cards;  
【ここにデータを入力】  
;  
proc factor data=a method=ml  
n=4 rotate=hk out=b;  
var x1-x8;  
proc print data=b;  
run;
```

因子分析
(factor analysis)

因子数の指定

回転法の指定

因子抽出法の指定

分析対象変数の指定

因子分析のプログラム

< 因子数の指定 >

因子分析では分析に先立ち因子数を決定しなければならない点が、主成分分析と異なります。因子数を決定するために予備分析が必要です。

主因子法は・・・

主成分分析プロシジャで予備分析を行い、主成分分析の際の主成分の取捨選択と同様の手順で因子数を決定します。

最尤法は・・・

因子数 n を $n=1,2,\dots$ と増やしていく度に1回ずつ因子分析を行い、「因子数は n 個で十分である」という帰無仮説への対立仮説（「 n 個より多くの因子が必要である」）がカイ2乗検定によってはじめて棄却される因子数を本分析に使用します。

因子分析のプログラム

○ 予備分析のアウトプット

	自由度	カイ2乗値	Pr>ChiSq
H ₀ : 4因子で十分である	17	12.7489	0.7528
H _A : より多くの因子が必要である			
Bartlettの修正なしのカイ2乗		12.998644	
赤池の情報量基準		21.001356	
Schwarzのベイズ情報量基準		87.713704	
Tucker-Lewisの信頼度係数		1.007466	

有意確率

この値が、0.10より上回り、非有意であれば、対立仮説である「より多くの因子が必要である」は棄却されたこととなります。0.7528なので棄却ですね！！

因子分析のプログラム

○ 予備分析のアウトプット

	自由度	カイ2乗値	Pr>ChiSq
H ₀ : 4因子で十分である	17	12.7489	0.7528
H _A : より多くの因子が必要である			
Bartlettの修正なしのカイ2乗		12.998644	
赤池の情報量基準 (AIC)		21.001356	
Schwarzのベイズ情報量基準 (SBC)		87.713704	
Tucker-Lewisの信頼度係数 (TLI)		1.007466	

カイ2乗値 / 自由度

3以下が好ましいです。

AICとSBC

他の因子数の場合より小さい方が好ましいです。

TLI

0.95以上が好ましいです。

因子分析のプログラム

◎ 予備分析の注意点

因子数 n を1から順に増やしていき、はじめて棄却された n を因子数とします。

(n を3、4、5としたときに仮説が棄却された場合は迷わず、3を因子数としましょう。)

対立仮説が棄却される前に共通性が1.0より大きくなるという「超Heywoodケース」に陥ってしまった場合、変数の数を調整して、もう1度予備分析を行います。

(この場合、「共通度が1.0より大きいです。」というエラーメッセージがアウトプットに表示されます。

正常終了した場合は、「収束基準が満たされました。」というメッセージが表示されます。)

因子分析のプログラム

< 回転法の指定 >

◎ 直交回転

(代表例：バリマックス回転 (v))

因子間に相関がないという前提の下*、因子軸と因子軸とを直角を保って回転します。

* 抽出された因子に対して相関分析を行って、相関が見つかった場合、斜交回転に切り替えて再分析すべきです。

◎ 斜交回転

(代表例：プロマックス回転 (p)、ハリス カイザー回転 (hk))

因子間に相関があるという前提の下で、各因子軸同士が直角を保つことなくそれぞれ回転します。「直交」という制約がないため、より変数に近い軸を描けることが多いです。

因子分析のプログラム

< 回転法の指定 >

現在の流れでは、直交回転より斜交回転が推奨されています。直交回転が斜交回転より優れている点は、因子アウトプットの解釈が容易である点のみであり、直交回転の様子が許されるためには、回転後の因子に相関があってはならないという厳しい条件があります。また、直交回転後の分析結果から得られる有用な知見のほとんどは、斜交回転後の分析結果でも得られることが、直交回転より斜交回転が推奨されている理由でしょう。

今回の課題で用いるのは・・・

ハリス カイザー回転（小野ゼミ初登場）！！！！

プロマックス回転では、バリマックス回転解（直交回転解）を利用して回転を行うため、正確なモデルを再現することができませんが、ハリス カイザー回転は、回転するだけでなく、尺度変換がなされるのでより正確な回転解を求めることができます。

因子分析のプログラム

< 因子抽出法の指定 >

最尤法 (ml)

観測変数が多変量正規分布に従うと仮定して、観測変数の同時確率をもとにパラメータ（分析で求めたいもの）に関する情報を伝達するという尤度という指標が算出されま
す。そして、この尤度が最大となるような因子負荷量と残差分散が導かれます。

最尤法は、不適解が出ることが多く、初心者には使いこなすことが難しいという短所を持つ一方で、適合度検定が可能である、望ましい漸近特性を持っている、大標本のものでは主因子法よりも良い推定量が求まるといった長所を持っているため、現在は、以前主流であった主因子法より推奨されています。

その他の因子抽出法は・・・

主因子法 (prin)、反復主因子法 (prinit)、最小2乗法 (uls)

因子分析のアウトプット

< 最尤法 & ハリス カイザー回転 >

◎ 因子の分散（固有値）

3種類のアウトプットが出力されます。

因子の分散

因子	重み付け	重み付けなし
Factor1	22.4501470	3.99675530
Factor2	6.8460687	1.48325951
Factor3	2.5332254	0.63737133
Factor4	2.1250383	0.68733873

因子分析のアウトプット

他の因子の影響を因子から取り除いた後の因子の分散 ←

因子	重み付け	重み付けなし
Factor1	2.78562488	1.02791200
Factor2	2.89322897	0.80749208
Factor3	6.03469867	1.30226867
Factor4	9.71369192	1.08994198

当該因子の全変数に対する独自の説明分散の割合
(他の因子の存在は排除されている)

因子分析のアウトプット

他の因子の存在を因子から無視した後の分散 ←

因子	重み付け	重み付けなし
Factor1	12.3768643	2.97251814
Factor2	10.0437078	2.36378904
Factor3	8.52966439	1.95895368
Factor4	20.9634827	3.2800825

当該因子の全変数に対する説明分散の割合
(他の因子は存在している)

因子分析のアウトプット

◎ 寄与率 & 累積寄与率

因子分析プロシジャにおいては、算出されないので固有値を用い、手計算で算出します（「因子の分散」と書かれた固有値を参照します）。

寄与率 = 固有値 / 元の変数の数

累積寄与率 = 第1因子から該当する因子までの寄与率の合計

因子分析のアウトプット

- ◎ 因子負荷量 3種類のアウトプットが出力されます。

回転後の因子パターン (標準化回帰係数)

	Factor1	Factor2	Factor3	Factor4
x1	0.76282	0.30314	0.02021	0.53289
x2	0.46736	0.21247	-0.00679	0.85153
x3	0.75467	0.11985	-0.04592	0.30974
x4	0.14831	0.81256	-0.15130	0.35631
x5	0.23221	0.89620	-0.18901	-0.05089
x6	-0.01237	-0.11189	0.79821	0.30714
x7	-0.03340	-0.15462	0.71597	-0.35190
x8	0.57643	0.30971	-0.55675	0.39124
x9	0.51339	0.02259	0.04132	0.22596
x10	0.02369	0.00023	0.01015	0.75126
x11	0.04680	0.05013	-0.02065	0.67119

因子分析のアウトプット

参照構造（半偏相関係数）

	Factor1	Factor2	Factor3	Factor4
x1	0.78322	0.32039	0.02388	0.50327
x2	0.48232	0.19825	-0.00882	0.84324
x3	0.78242	0.12894	-0.06742	0.31124
x4	0.07862	0.84422	-0.18704	0.34284
x5	0.25534	0.90013	-0.17823	-0.06722
x6	-0.03599	-0.14128	0.78903	0.27367
x7	-0.04460	-0.13577	0.70379	-0.33463
x8	0.59939	0.26359	-0.53045	0.37958
x9	0.50226	0.03063	0.03613	0.19333
x10	0.02565	0.00023	0.01373	0.72927
x11	0.06747	0.04922	-0.01965	0.64229

因子分析のアウトプット

因子構造（相関係数）

	Factor1	Factor2	Factor3	Factor4
x1	0.77890	0.20349	0.12778	0.58843
x2	0.53722	0.22189	-0.00182	0.89994
x3	0.99750	0.12215	0.11182	0.32328
x4	0.18547	0.90115	0.19901	0.41002
x5	0.34981	0.86317	0.11290	0.09139
x6	0.12849	0.14291	0.81190	0.13284
x7	-0.04380	0.00027	0.93019	-0.23018
x8	0.60489	0.34965	-0.42238	0.35821
x9	0.48926	0.03063	0.02339	0.20001
x10	0.13435	0.00234	0.11193	0.80143
x11	0.12059	0.13062	0.11389	0.60114

因子分析のアウトプット

- ◎ 基準構造 (sasでは「参照構造」と誤訳されています。)

基準構造とは、他の因子の影響を取り除いた後の当該因子と変数との半偏相関行列です。因子間が無相関であれば、基準構造は因子負荷量に等しくなります。因子負荷量と大きな差異があれば、再度分析を行う必要があります。

- ◎ 因子構造

因子負荷量が因子得点から各変数への回帰係数であるのに対し、因子構造とは、因子得点と各変数との単純相関係数です。因子間が無相関であれば、因子構造は因子負荷量に等しくなります。因子負荷量と大きな差異があれば再度分析を行う必要があります。

因子分析のアウトプット

○ 因子間相関

因子間相関行列

	Factor1	Factor2	Factor3	Factor4
Factor1	1.00000	0.33178	0.19367	0.56974
Factor2	0.33178	1.00000	0.33555	0.44769
Factor3	0.19367	0.33555	1.00000	0.16974
Factor4	0.56974	0.44769	0.16974	1.00000

一般的に高い相関であるといわれる0.60を上回る相関係数をもつ因子群が存在した場合には、変数の数を調整し、再度分析を行う必要があります。

因子分析のアウトプット

◎ 共通度

最終的な共通性の推定値と変数の重み

共通性の合計：重み付け = 33.954479 重み付けなし = 6.804725

変数	共通性	重み
x1	0.24720834	1.3283853
x2	0.84132902	6.3023300
x3	0.60658544	2.5418534
x4	0.53117217	2.1329822
x5	0.76916326	4.3320539
x6	0.28970838	1.4078736
x7	0.42791906	1.7480084
x8	0.69272926	3.2544549
x9	0.65743449	2.9191403
x10	0.92635349	13.5784044
x11	0.81512196	5.4089927

4つの因子で変数11個のうち約6.80個分を縮約したという意味です。

因子分析のアウトプット

◎ 因子得点(各因子について各オブザベーションが持つ値)

(観測No. (元の変数)

(因子)

OBS	x1	...	x11	factor1	factor2	factor4	factor4
1	4	...	2	0.55526	0.62048	0.22552	0.73171
2	1	...	5	-0.97232	-1.36175	0.84394	-1.63624
3	2	...	4	0.53532	-0.98575	-0.40422	-1.17240
...
162	2	...	3	-0.49159	-0.86566	0.70563	-0.28450

例えば、第1因子を因子負荷量より「デレデレ度」とした場合、1番目のオブザベーションの「デレデレ度」は0.55526ということです。

探索的因子分析と確認的因子分析

- ◎ 探索的因子分析（exploratory factor analysis）

全ての変数を対象としてそこから複数の因子を探索的に算出します。

例えば、国語・社会・数学・理科の得点データから、背景にある能力が何であるかを探る場合に用います。

- ◎ 確認的因子分析（confirmatory factor analysis）

特定の変数のみを対象としてそこから各1つずつの因子を算出します。

例えば、国語・社会・数学・理科のうち、国語と社会の得点データは「文系能力」、数学と理科のデータは「理系能力」とそれぞれ呼びうる別個の2つの能力を背景に有していることを確認したい場合に用います。

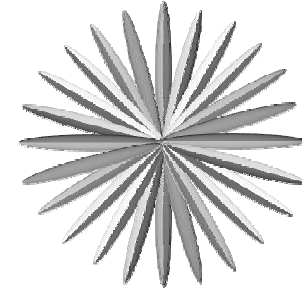
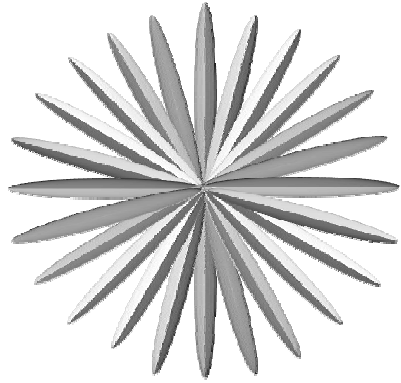
課題

先日、慶應義塾大学の学生がどのような恋愛上の特徴を持っているのかを知るための調査をしました（被験者500人）。

この恋愛調査に用いた質問項目群（20項目）からいくらか任意に取捨選択を行った上で、それらの質問項目に対して因子分析を行い、それぞれの質問項目がいかなる因子に縮約可能であるかについてのコメントをしてください。

コメントに添えて、因子負荷量・固有値・寄与率・累積寄与率・共通度をひとまとめに示す表を作ってください。また、因子負荷量に基づいて生成された因子に命名をしてください。そのうえで、多数の被験者からゼミ生を選択し、どの被験者がどのような恋愛性向を持っているのかを示すプロダクトマップならぬ「恋愛性向マップ」を作成し、最後に考察を加えてください。

さあ、6期生は一体どんな恋愛上の特徴をもっているのでしょうか？



ご清聴ありがとうございました

