# 3相データの主成分分析

2004MM025 今泉賢資
指導教員: 田中豊

## 1 はじめに

本研究でいう3相データとは，$N$個体×$P$変量×$T$時点の形で表されるデータのことを指す．つまり，個体×変量の形の2相データを，継続的に測定したデータと言える．たとえば，あるホームセンターの売上データとして，店別，売り上げ品目(部門)別の売り上げが数年分得られている．これは，店舗(個体)×部門(変量)×年度(時点)という構造の3相データである．

本研究は，このような3相データにおける時間による変化をどのように捉えたらよいかを考察するものである．

## 2 アプローチ

3相データを解析する際，各時点ごとに主成分分析を行い，その違いを見ることでも一応の結果は得られるだろう．しかし，それでは各時点ごとに主成分が同じものを表しているとはいえないため，適切とは言いづらい．

１つの考え方として，ある時点の主成分分析の結果に基づいて他の時点での主成分得点のみを計算し，時点間の差を見る方法がある．即ち，ある時点の主成分係数を代表として他の場合にも適用する方法である．このようにすることで，指標となる主成分は同一のものとなり，時点間の差を比べることができる．

また，別の考え方として，各時点での$N$個体×$P$変量の2相データを縦方向や横方向に連結してひとつの大きな2相データとして扱う方法がある．対応分析の分野では，これらLONG Matrix(縦に連結したもの)やBROAD Matrix(横に連結したもの)を対応分析することにより，時点間の差を捉える方法が知られている(Heijden, 1987[1])．この考え方を主成分分析に適用する．

## 3 解析方法

### 3.1 各時点での主成分分析

各時点で別々に主成分分析を行うと，それぞれから得られる主成分は，たとえ同じ第1主成分と呼ばれていようとも係数ベクトルが異なるため，時点が異なれば同じものを表しはしない．

そこで，ある時点での主成分係数を用いて，他の時点でのデータの主成分得点を計算する．即ち，ある時点を基準として，その時点での分析結果に，他の時点でのデータを追加データとして持ち込み主成分得点を計算する．

本研究では，変化前(時点$t=0$)の特徴を基に，主成分得点の変化からデータの変化を捉える．

ただし，このとき基準とする時点(主成分係数を代表値として用いる時点)での平均と標準偏差を用いて追加データを標準化する必要がある．

### 3.2 LONG Matrixの主成分分析

この方法では，各時点の全体的(平均的)な主成分を抽出することができる．その上で，各個体の$t$時点での主成分得点の変化からデータの変化を考察する．

### 3.3 BROAD Matrixの主成分分析

この方法では，$P\times T$変量で主成分分析を行い，各変量の$T$個の時点での主成分係数の変化から，時間の変化とともに主成分の意味の変化を捉える．

## 4 人工データの生成

### 4.1 生成法

本研究では，因子分析モデルを用いて，すでにあるデータの解析結果を元に人工データとして再生成した．

$$X = \mu + Af + e \tag{1}$$

式(1)が因子分析のモデルである．ここで，$\mu$は$P\times 1$の平均ベクトル，$A$は$P\times Q$の因子負荷行列である．$A$は以下の式で与えられる．

$$A = \begin{pmatrix} \sqrt{\lambda_1}V_1 & \sqrt{\lambda_2}V_2 & \cdots & \sqrt{\lambda_Q}V_Q \end{pmatrix} \tag{2}$$

$f$は$Q\times 1$の因子得点ベクトル，$e$は$P\times 1$の独自因子である．

しかし本研究では，各手法を用いることでどのような変化として捉えられるかを目的とする．そのため，シミュレーションに用いる人工データの生成においては，独自因子$e$はすべて0であるとし．

$$X = \mu + Af \tag{3}$$

式(3)のモデルを用いることとする．

式(3)を用いて作成したデータを基準に3相データを生成する．

### 4.2 3相データの生成

4.1節の生成法に手を加えて，$T$個の時点を持つ3相のデータとする．

#### 4.2.1 平均値の変化

$\mu$ や $f$ の値を変化させ平均値に変化を与えた3相のデータを作成する．

$$\mu(t) = \mu_0 + at \tag{4}$$

$$f(t) = \begin{pmatrix} f_1 + at \\ f_2 \end{pmatrix} \tag{5}$$

式(4)や式(5)のように $\mu$ や $f$ を$t$の一次関数として作成した．平均値を変化させることの目的は，主に主成分得点の変化を見ることである．

#### 4.2.2　因子負荷行列の変化

因子負荷行列$A$は、式(2)を用いて定める．固有値$\lambda$の値を変化させたり固有ベクトル$V$を回転させることでその構造に変化を与える．

$$\lambda_1(t) = \lambda_1(0) + at \qquad (6)$$

式(6)のように$\lambda_1$のみを$t$の一次関数として変化させたデータや，固有値$\lambda_1, \lambda_2$を入れ替えるなどの変化を与えたデータなどを作成した．

固有ベクトル$V$の回転に関しては，

$$V(t) = V(0)\begin{pmatrix} \cos(\theta t) & \sin(\theta t) \\ -\sin(\theta t) & \cos(\theta t) \end{pmatrix} \qquad (7)$$

式(7)のように，時点ごとに定数$\theta$だけ回転させたデータを作成した．これらのデータは，因子負荷行列$A$に変化があった場合にどのようにその変化が表れるかを見ることが目的である．

## 5　解析結果

本節では，実際に作成したデータと解析結果を示す．

多変量統計解析法[2]にある身体測定データ（p.84 表3）を元に以下の4通りの操作で3時点，30個体，4変量の3相データを作成した．

$$u(t) = u_0 + u_0/10 \times t \qquad (8)$$

$$f(t) = \begin{pmatrix} f_1 - 2t \\ f_2 \end{pmatrix} \qquad (9)$$

$$\sqrt{\lambda_1}(t) = 1.881781 - 0.5t, \sqrt{\lambda_2} = 0.5598064 \qquad (10)$$

$$V(t) = V(0)\begin{pmatrix} \cos(30t) & \sin(30t) \\ -\sin(30t) & \cos(30t) \end{pmatrix} \qquad (11)$$

なお，ここでは式(8)から式(11)までの1例ずつの例示に留めるが，本節に示す結果は他の場合についても同様の結果が得られている．ここでは，元データが身体測定データであることを考慮し，すべて標準化して主成分分析を行った．

### 5.1　主成分得点の変化にみる差異

LONG Matrixを用いた主成分分析の結果を主成分得点プロットとして図1から図4に示す．図を見れば明らかであるが，この方法で変化のタイプを分類することができた．すなわち，主成分得点が第1，第2主成分を用いたプロット図上において，次のように捉えられる．

- 平均値（因子得点）の変化は，平行移動
- 固有値の変化は，原点への集中・拡散
- 固有ベクトルの回転は，回転移動

### 5.2　主成分係数の変化にみる差異

BROAD Matrixを用いた主成分分析を行った場合，同一変数の主成分係数が時点ごとに変化しているかどうかで分類できる．結果として因子負荷量$A$に変化がある場合，主成分係数に変化が見られる．

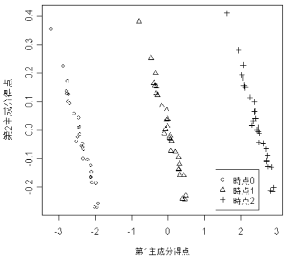


図 1: $\mu$を変化させた場合

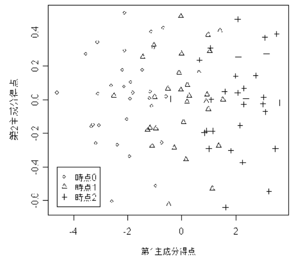


図 2: $f$を変化させた場合

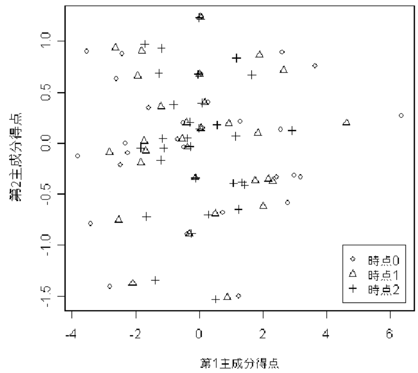


図 3: $\lambda$を変化させた場合

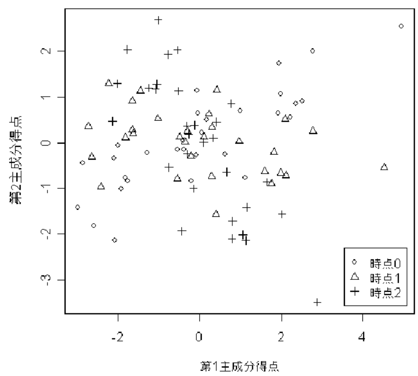


図 4: $V$を回転させた場合

## 6　おわりに

本研究では，3相のデータ解析方法について3種の解析法を用いて考察した．結果，LONG Matrixを用いることですべてのタイプを分けることができた．

本研究は現時点で人工データを用いた解析までであり，主に相関行列を用いた分析を行ってきた．今後は分散共分散行列を用いた分析や，実データを用いてその検証するなどより発展的な研究を行っていきたい．

## 参考文献

[1] VAN DER HEIJDEN P. G. M.: Correspondence analysis of longitudinal categorical data, DSWO Press (1987).

[2] 田中 豊, 脇本 和昌：多変量統計解析法, 現代数学社 (1983/01).