

TUCKER2 モデルによる多変量時系列データの解析

— 体格成長データ —

ピーター M. クローネンベルグ, 宮基壽夫

Analysis of Multivariate Longitudinal Data by Tucker2 Model

— Growth Curves Data —

Pieter M. KROONENBERG and Hisao MIYANO

製品科学研究所研究報告 第105号 (昭和61年7月) 別刷
Reprinted from the "Bulletin of Industrial Products
Research Institute" No. 105 (1986) July

TUCKER2 モデルによる多変量時系列データの解析

— 体格成長データ —

ピーター M. クローネンベルグ*1, 宮埜壽夫*2

Analysis of Multivariate Longitudinal Data
by Tucker2 Model

— Growth Curves Data —

Pieter M. KROONENBERG and Hisao MIYANO

Data on the physical growth of Japanese girls between 6 and 14 years old were analysed to illustrate the application of three-mode principal component analysis, in particular the Tucker2 model, to growth curves. The individual differences in growth and growth speed were investigated using the deviation scores with respect to the average growth curves. The results were compared with a similar study of French girls, and with previous analyses of the data by C. Hayashi³⁾ & F. Hayashi⁵⁾ [1982]. In addition, the problem of preprocessing data before a three-mode analysis is discussed.

1. はじめに

本論文では、心理学、人間工学、社会学等の分野においてよく問題とされる多変量時系列データの新しい解析法を提案し、その有効性を示すために女子の体格成長データの解析を行う。

ここで問題とする体格成長データは、一般に人、変量、時間の三つのモード(相)からなる三次元データであり、このようなデータに対しては、これまでいずれかのモードについて平均化し、二相データとして主成分分析等により解析する、あるいは成長曲線を仮定して回帰分析を行うといった方法が通常使われてきた。しかし、このような方法では個人の成長過程が平均的な成長過程とどの程度異なっているのかあるいはどのような成長過程のタイプがあるのかといったことを明らかにすることは出来ない。

これらの事も同時に明らかにするためには、多変量時系列データ(成長データ)が三相であることに留意しな

から解析する必要がある。このための方法としては、PARAFACモデルによる方法¹⁾、動的因子分析法など、これまでに様々な方法が考えられているが、ここではその一つとして三相主成分分析法による解析方法を考える。三相主成分分析法は、Tucker のモデルを基本として、Kroonenberg らにより開発された方法²⁾であり、これまでに様々な実データに対して適用が試みられ、その有効性が確かめられている。また、ここで解析するデータは、林らが Harshman の PARAFAC モデルにより解析を行ったものである。以下では、三相主成分分析法の一つである TUCKALS2 による解析の手順を示すと同時に、林らの解析したデータ^{3), 5)}に適用することにより、この方法によってより明確な解析結果が得られることを示す。

2. 解析方法

2.1 データと前処理

ここで用いた成長データ⁵⁾は、千葉県某女子中学校女学生徒32名の小学校1年から中学校3年までの計9ヶ年における身長、体重、胸囲、座高の計測値である。した

*1 昭和60年度海外招へい研究員

*2 基礎人間工学部心理情報工学課

がって、データ x_{ijk} , $i=1, \dots, 32$, $j=1, \dots, 4$, $k=1, \dots, 9$ は $32 \times 4 \times 9$ の三相データと見ることが出来る。

成長データの解析の目的は、主に平均的な成長パターンの把握と成長パターンの個人差の解析にあると考えられる。そのために、ここではデータを直接解析するかわりに、まずデータの前処理を行い、(1)各変量に関する平均成長曲線(表1、および図1参照)と(2)これらの平均成長曲線からの偏差を求める。ここで、問題となるのは

変量の測定単位が異なるために、偏差がこのままでは互いに比較出来ないことである。そこで、ここでは具体的には解析の目的を考慮して次のようなデータの前処理を行った。

まず、平均的な成長曲線を求めるために、各変量について年毎の平均値 $\bar{x}_{jk}$ と分散 s_{jk}^2 を次式により求めた。

$$\bar{x}_{jk} = (1/I) \sum_i x_{ijk},$$

$$s_{jk}^2 = (1/(IK)) \sum_i \sum_k (x_{ijk} - \bar{x}_{jk})^2, \quad i=1, \dots, I, \\ j=1, \dots, J, \quad k=1, \dots, K.$$

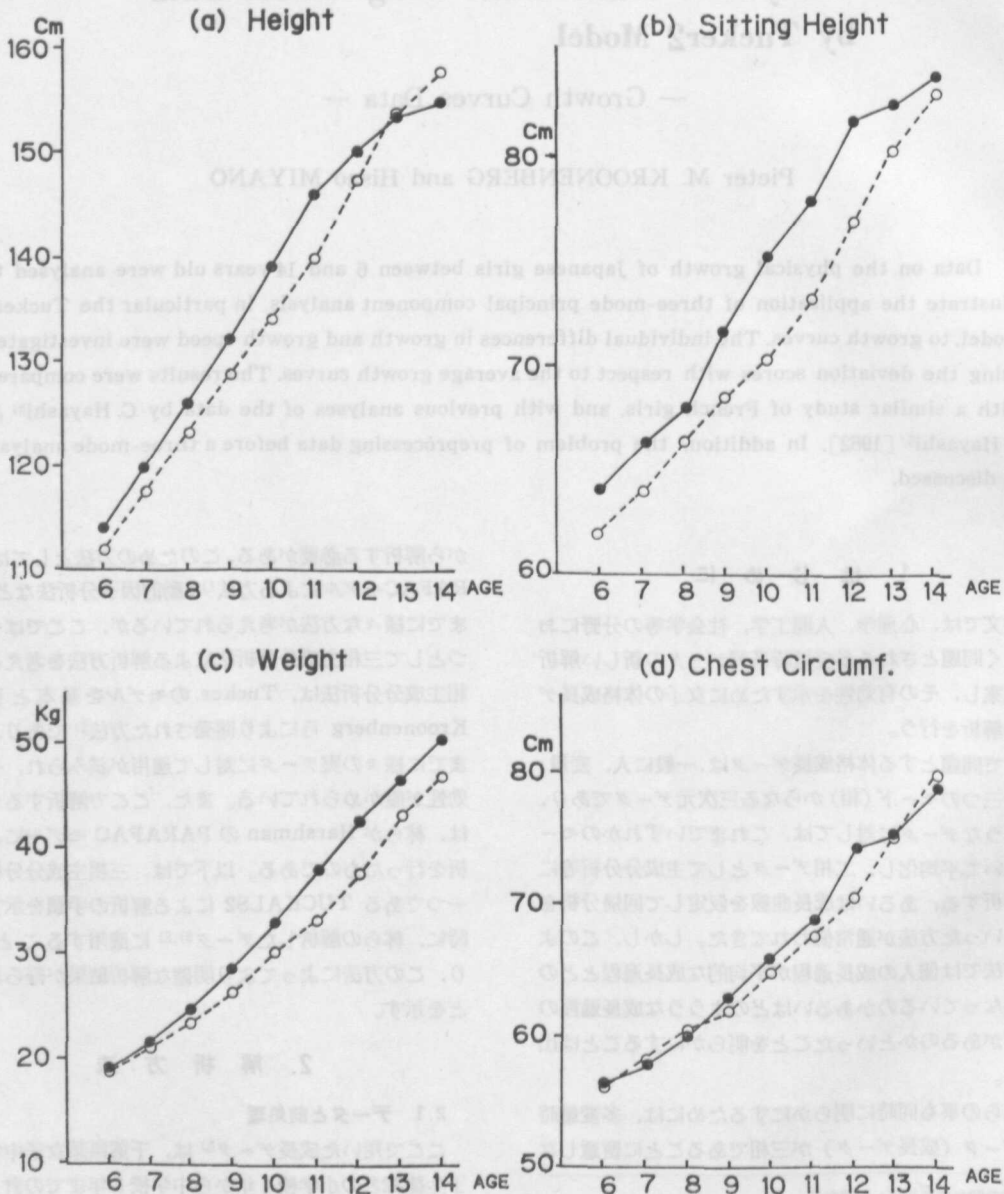


Figure 1. Mean Growth Curves of Japanese (•) and French (◦)

Table 1. Comparison of Mean Growth Curves —Japanese (J) and French (F)—

AGE	HEIGHT (cm)		SITTING HEIGHT (cm)		WEIGHT (kg)		CHEST CIRCUMF (cm)	
	J	F	J	F	J	F	J	F
6	113.9	111.9	64.0	62.2	19.7	18.9	56.2	56.0
7	119.9	117.6	66.4	64.1	21.9	21.1	58.0	58.3
8	125.8	123.4	68.0	66.6	25.1	23.6	60.0	60.5
9	132.0	128.8	71.9	68.6	28.9	26.7	62.7	62.0
10	139.2	134.1	75.5	70.5	33.3	30.0	66.1	64.5
11	146.2	140.1	78.2	73.3	38.0	33.4	69.1	67.9
12	150.6	147.5	82.0	77.1	43.2	37.9	74.6	71.4
13	153.6	154.1	82.6	80.6	46.9	43.6	75.5	75.2
14	155.1	158.1	84.0	83.3	50.6	46.7	79.1	80.0

* Number of Japanese Girls=32, Number of French Girls=30.

ここで、I=32 は人数、J=4 は変数の数、K=9 は年数である。平均値 $\bar{x}_{ijk}$ を、表1及び図1に示す。比較

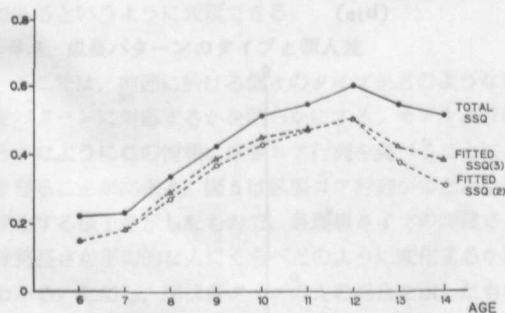


Figure 2. Total Sums of Squares and Fitted Sums of Squares

のために、フランス人の結果⁴⁾も示してある。明らかな違いが身長と座高の発達の割合に見られる。次に、上の二つの量に基づいて、データのいわゆる規格化を次式により行った。

$$\tilde{x}_{ijk} = (x_{ijk} - \bar{x}_{jk}) / s_{jk}, i=1, \dots, I, j=1, \dots, J, k=1, \dots, K$$

この時、平均的な人を平均の成長曲線 $\bar{x}_{jk}$ によりその発達過程が記述される人と定義してよいであろう。このような人が実際にいるか否かが問題であるが、後に示す図4より被験者3および25がほぼ平均的な人であると言ってよいであろう。

2.2 モデル

平均的な人の成長曲線からの“ずれ”（偏差）を表わす $\tilde{x}_{ijk}$ を解析するために、ここではいわゆる Tucker2 モデルをもちいる。このモデルは、次式により記述される。

$$\tilde{x}_{ijk} = \hat{x}_{ijk} + e_{ijk}, \quad e_{ijk}: \text{エラー},$$

$$\hat{x}_{ijk} = \sum_{m=1}^M \sum_{p=1}^P a_{im} b_{jp} g_{mpk}$$

あるいは、行列を用いて次のようにも書ける。

$$\hat{X}_k = A G_k B'$$

ここで、 $A = (a_{im})$ は第1モード（人）に対する因子負荷行列、 $B = (b_{jp})$ は第2モード（変量）の因子負荷行列である。また、行列 G_k は拡張コア行列と呼ばれるもので、第1モードと第2モードの関連の様子を表わしている。特に、行列 A と B が正規直交行列である場合には g_{mpk}^2 は第1モードの m 番目の成分と第2モードの p 番目の成分の両者が解全体に寄与している程度を表わしている。以下では、行列 A 、 B の正規直交性を仮定するが、 A 、 B を正則変換したものもまた解になるので、この仮定は必ずしも必要とはしない。

2.3 パラメータ推定のアルゴリズム

Tucker2 モデルに含まれるパラメータは、交互最小二乗 (ALS) 法により推定することが出来る。すなわち、まず B を固定して A の推定を行い、次に逆に A を固定して B の推定を行う。この過程を繰り返すことにより与えられた二乗評価関数を最小にする A および B を求める。この時、拡張コア行列 G_k は交互推定が終了した時点において、次式により計算する。

$$G_k = A' X_k B, \quad X_k = (\tilde{x}_{ijk})$$

以上の方法の妥当性は、Kroonenberg²⁾ らにより明らかにされており、計算プログラム⁸⁾ (FORTRAN77) も既に公開されている。

3. 解析結果

ここでは、解析結果をいくつかに分けて示す。まず、解全体の特徴について議論を行い、次に変量と人に関する

る結果を示すと共に、拡張コア行列を用いてこれらの成分間の関係が年とともにどのように変化して行くかを示す。また、最後に平均的な人にくらべ、特徴的な成長パターンをとる何人かの人の変化パターンを示す。

3.1 あてはまりの良さ

モデルの全体的なあてはまりの様子を考慮して、変量については2次元($P=2$)、人については3次元($M=3$)の解を求めた。その結果、全体としてはデータのバラツキの82.7%が説明された。また、Varimax 回転をほどこした変量に関する解の説明率は、1次元目が43.3%、2次元目が39.5%であり、人に関する解の説明率は、それぞれ64.1%、13.6%、5.0%であった。

図2に各年毎のあてはまりの程度 (Fitted Sums of Squares) と前処理したデータに含まれているバラツキの程度 (Sums of Squares) を示す。この図より、体格のバラツキは年とともに増大し、12才頃にピークを迎え、その後減少することが分かる。この傾向は、フランス人についても見られるものであり、バラツキの増大は成長の速度および成長の開始時期の違いを反映していると考えられる。なお、図には人についても2次元解を求めた時のあてはまりの様子も示されている。

3.2 平均成長曲線

図1より平均的にはこの期間では体格は成長しつづけることが分かるが、身長などの高さに関する成長の様子と重さに関する成長の様子には明らかな違いが見られる。すなわち、12才頃からは高さの成長が鈍化するのに対して、体重と胸囲は期間中、常に増加している。

一方、通常よく認められる急激な成長点はこの平均成長曲線には見られない。これは、異なる位置に急激な成長点を持つS字形の成長曲線を平均化した場合によく見られる現象であり、そのためと考えられる。実際、後で見るようになりに多くの人が急激な成長点を有している。

フランス人と日本人を比較した場合、全体としてはあまり大きな違いはないようである。しかし、体のプロポーションには明らかな違いがあり、幼年時にはあまり違わなかったものが、中学卒業時には前者が身長のわりに足が長いのにに対して、後者は身長のわりに胴が長いという違いが見られる。

3.3 変量と人に関する解

図3に変量に関する解に Varimax 回転を加えた解を示す。図より第1成分は、体重と胸囲、第2成分は身長と座高に対応することが分かる。ここでは、第1成分を頑健さ、第2成分を骨格長さと呼ぶことにする。

図4は人に関する解を示したものである。ただし、三

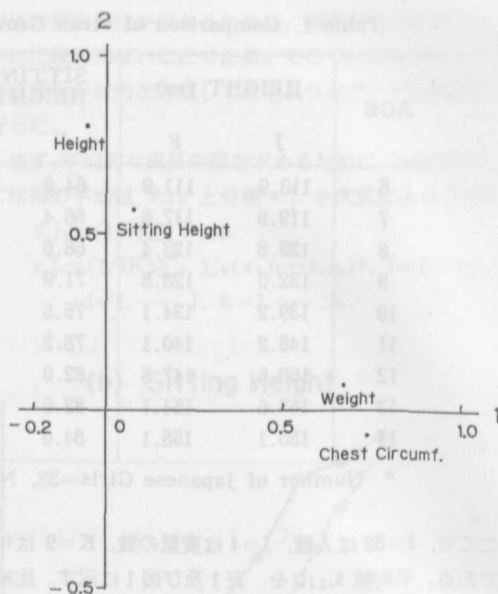


Figure 3. Compromise Solution for Variables (b_{1p})

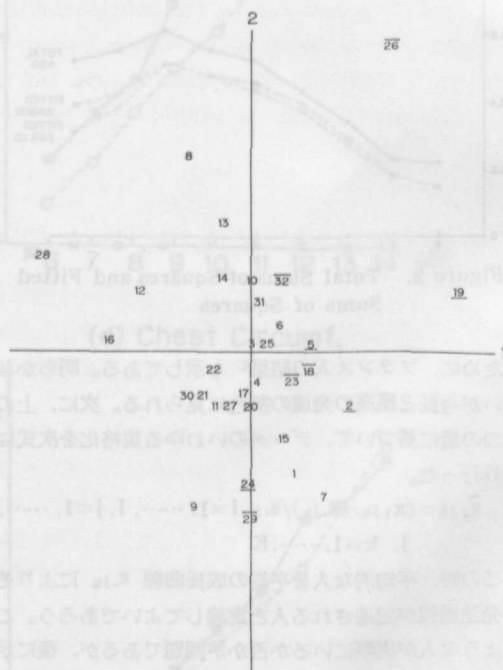


Figure 4. Compromise Solution for Girls (a_{1m})

次元目の値は正の大きな値をとる時は上に線をつけ、負の大きな値をとる時は下に線をつけることにより大難把に示した。しかし、人の場合には、変量の場合と異なりあまり明確な解釈をすることは出来ないように思われ

る。すなわち、図を見ると人は平均的な人の周りに分布しており、人をなんらかの方法でグルーピングすることは不可能であるように思われる。しかし、極端な特徴を有する人に着目すれば、どのような成長パターンのタイプの人がいるかをある程度指示できると考えられる。そこで、ある一つの軸について大きな値をとり、その他の二つの軸についてはほとんどゼロをとるような人を Tucker にしたがって“理想タイプ”と呼ぶことにするとこの場合は3次元解であるから、Type 1+, Type 1-, …, Type 3- の六つのタイプ（実際には、例えば Type 1+ は 1- の逆の関係を表わすから、三つのタイプ）が考えられる。問題は、このようなタイプの人が実際に存在するかであるが、図より被験者16は Type 1-, 被験者13は Type 2+, 被験者1は Type 2-, 被験者18.5はそれぞれ Type 3+ および 3- にほぼ対応していると考えられる。このようなタイプ分けの利点は、各人の特徴が簡潔に捉えられることにある。例えば被験者29は Type 2- と Type 3+ の特徴を有する人であるというように表現できる。

3.4 成長パターンのタイプと個人差

ここでは、前節に於ける個々のタイプがどのような成長パターンに対応するかを明らかにする。モデルより明らかにようにこの情報は拡張コア行列を調べることにより得ることが出来る。図5は拡張コア行列が年とともに変化する様子を示したもので、各理想タイプの頭健さと骨格長さが平均的な人にくらべどのように変化するかがわかる。ただし、図は各タイプの+の場合を示したものである。タイプの一の曲線は、対応する曲線の符号を変えたものになる。

大雑把に言うならば、Type 1+ は頭健さと骨格長さの両方が早い時期に成長を始めるものの12才頃には成長が鈍化するタイプを表わし、Type 2+ は頭健さはこの期間中平均を上回る成長をする一方、骨格長さの成長が

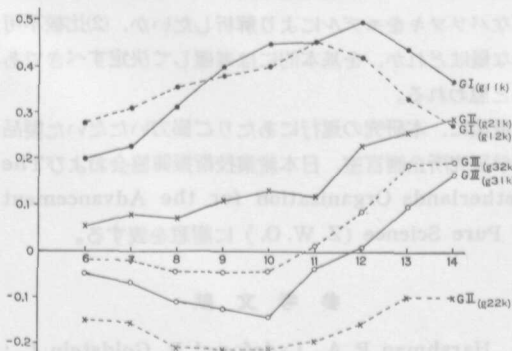


Figure 5. Time Trend of Extended Core Matrix
— : Stoutness, ... : Skeltal Length.

少ない比較的背の低い、太ったタイプを表わしている。また、Type 3+ はこの期間中の頃から急激に頭健さ骨格長さの両方が成長するタイプを表わしている。

これと同様に、各人の成長のパターンも平均的な人に対する頭健さと骨格長さの変動の様子から捉えることができる。すなわち、各年における各人の頭健さと骨格長さの二つの成分に関する得点 d_{1Pk} を次式により計算し、上と同様にプロットすればよい。

$$d_{1Pk} = \sum_{m=1}^M a_{1m} g_{mPk}$$

図6、7は、特徴的な成長パターンを示す何人かの人の結果である。

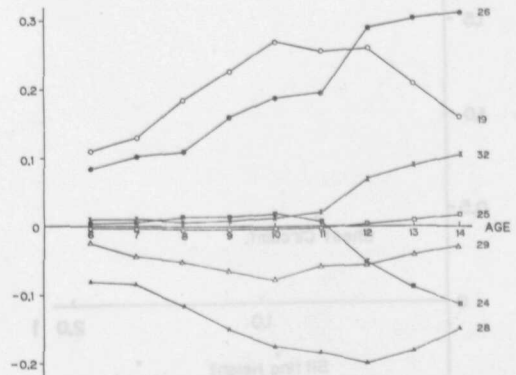


Figure 6. Typical Growth Patterns (Stoutness)

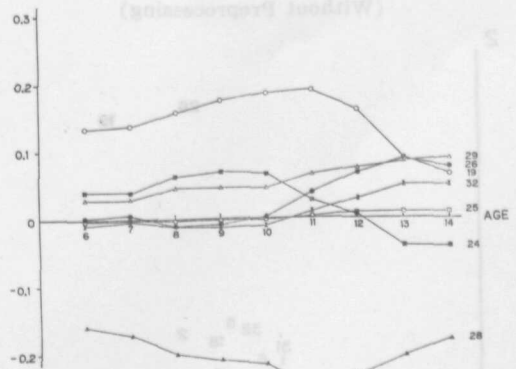


Figure 7. Typical Growth Patterns (Skeltal Length)

4. 考 察

本論では、多変量時系列データを TUCKALS2 により解析する方法について議論し、その有効性を成長データに適用することにより示してきた。ここでは、方法の適用において特に注意しなければならないデータの前処理について林らの解析結果と比較しながら議論する。

林らは、ここで解析したデータを Harshman の PARAFACモデルにより解析した。この時、彼らはデータの前処理を行うと解釈可能な結果が得られないとして、前処理をしない場合の解析結果を示した。ところで、PARAFAC モデル¹⁾は

$$\hat{x}_{ijk} = \sum_q a_{iq} b_{jq} c_{kq}$$

により与えられる。従って、データをまずTUCKALS2モデルにより解析し、次にその拡張コア行列をPARAFACモデルにより解析すれば、全体としてPARAFAC

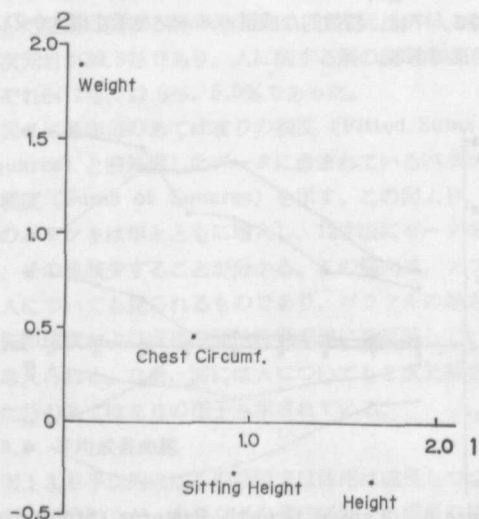


Figure 8. PARAFAC Solution for Variables (Without Preprocessing)

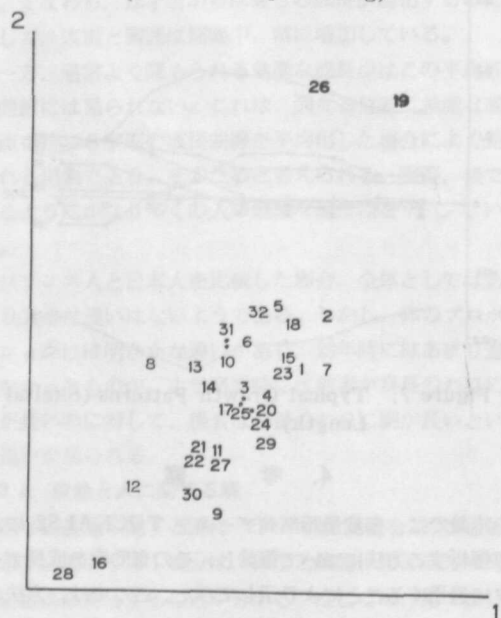


Figure 9. PARAFAC Solution for Girls (Without Preprocessing)

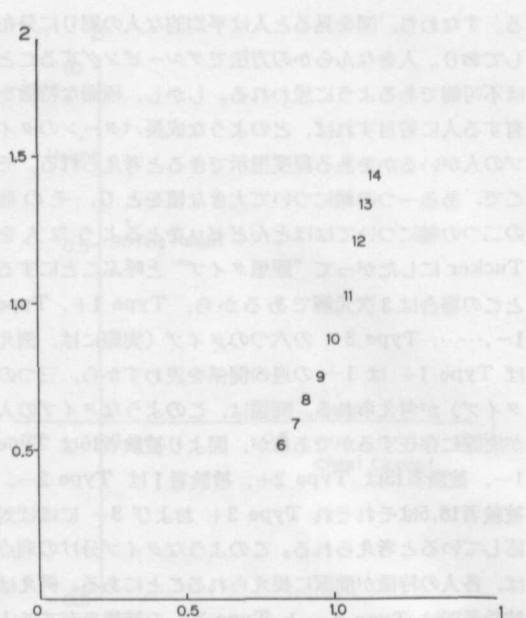


Figure 10. PARAFAC Solution for Ages (Without Preprocessing)

モデルによる解析とほぼ同じ結果が得られると考えられる。実際、データの前処理を行わずに上の方法により成長データを解析すると図8, 9, 10の結果が得られるがこれは林らの得た結果⁵⁾とほとんど同じである。図8は変量についての結果であり、図9は人についての結果である。図より人については、原点の位置と布置の回転を除いてはTUCKALS2の結果とほぼ一致していることが分かる。異なるのは変量についての結果であり、特に“胸囲”の位置が異なっている。これは、データの前処理を行わなかったためと考えられる。すなわち、彼らの結果は変量の測定単位が異なることによる影響を含んでいるものと考えられる。

従って、データの前処理は(1)データに含まれるどのようなパラツキをモデルにより解析したいか、(2)比較可能な量はどれか、を基本的には考慮して決定すべきであると思われる。

最後に、本研究の遂行にあたりご協力いただいた製品科学研究所企画官室、日本産業技術振興協会およびThe Netherlands Organization for the Advancement of Pure Science (Z. W. O.) に謝意を表する。

参考文献

- 1) Harshman, R. A., Ladefoged, P., Goldstein, L.: Journal of Acoust. Soc. Am. Vol. 62, No. 3, 693(1977)

- 2) Kroonenberg, P. M. : Three-mode principal component analysis, Leiden, DSWO Press (1983)
- 3) Hayashi, C. : Behaviormetrika, 11, 49(1982)
- 4) Kroonenberg, P. M. : WEP-reeks WR 85—14—EX, Leiden. Dept. of Education (1985)
- 5) 林 文, 山岡和枝, 寺尾浩明 : 日本行動計量学会誌, 9巻1号, 47(1982)
- 6) Kroonenberg, P. M., Brouwer, P. : WEP-reeks WR 85—12—RP, Leiden, Dept. of Education (1985)