Title	ペア・集団データにおける階層性の分析
Author(s)	清水, 裕士
Citation	対人社会心理学研究. 6 p.89-p.99
Issue Date	2006
oaire:version	VoR
URL	https://doi.org/10.18910/11478
rights	
Note	

ペア・集団データにおける階層性の分析[1)]

清水　裕士(大阪大学大学院人間科学研究科)

本論文の目的は、ペア・集団データといった階層的構造を持ったデータに対する、有効な方法論を提示し、それらの利点を比較することである。社会心理学で得られる機会の多い、ペア・集団データは従来の重回帰分析や分散分析では適切な分析を行うことができず、また知ることができる範囲も限られる。階層的分析法であるHLMやペアワイズ相関分析を用いることで、従来の分析法におけるエラーを回避できるだけでなく、対人関係や集団のダイナミズムについての重要な知見を得ることが可能になる。また、本論文では特にSASにおけるMixedプロシージャを用いたHLMの適用例と解釈方法について解説した。また、階層的分析の比較とそれぞれの限界点について述べた。

キーワード: ペアデータ、階層データ、階層線形モデル、混合モデル、ペアワイズ相関分析

問題

個人と集団のダイナミズムを研究テーマとする社会心理学、特にグループ・ダイナミクスの領域では、個人に関する変数と集団に関する変数の両方が分析の対象となることがある。このようなデータを本論文では階層的データと呼ぶ。

階層的データを分析するとき、従来のような方法では様々な問題が現れることが指摘されている(e.g., Bryk & Raudenbush, 1992)。

本論文ではこのようなデータを分析するための手法について整理し、どのような場面において必要とされるのか、またどのような分析法を選択するべきなのかについて論ずる。また、特に相互依存的なペアを対象に収集した階層的データの分析法について、いくつか紹介し、その有効性を比較検討する。

社会心理学におけるデータの階層性

社会心理学において、個人の行動や態度が集団によって大きく影響を受けることは古くから主張されてきており、また重要なテーマでもある。

Lewin(1950)は個人の行動が、パーソナリティだけでなく集団の雰囲気や状況といった環境によっても規定されることを有名な式、B=f(P,E)によって表現した。彼の集団力学では、成員間の相互依存性によって生まれる生活空間の誘因と、個人の動因の合力によって個人の行動が現れると考える。

また、成員間の相互依存性が影響力を持つのは集団だけでなく、夫婦や恋愛関係といった親密な 2 者関係、あるいはただ単に会話を行っているペアの場合も同様である。

ペアデータを得る場合、個人に帰属されないペア単位の変数もいくつか測定されることが多い。例えば、夫婦関係や恋愛関係であれば結婚(交際)年月といった変数が、会話場面のペアであれば沈黙時間、パーソナルスペースといった変数が、挙げられる。研究者は個人を対象とした心理変数と、このようなペア単位の変数との関連に興味があるが、これらの変数間の相関係数を素直に算出することは難しい。

さらに、相互依存的なペアから収集したデータは、大抵の場合、ペア内で類似性が見られる。つまり、ある夫が夫婦関係に満足しているとき、その妻も同様に満足しており、ある人が会話場面でジェスチャーの量が多いとき、そのパートナーも同様にジェスチャーが多いのである。このようなペア内に類似性が見られるデータについて、考慮なしに相関係数等の統計指標を算出することは許されない。

このように、ペアデータのような階層的データを扱うことは様々な問題をはらんでいるのである。

階層的データとは

これまで、社会心理学では集団の状況や雰囲気を実験条件として設定し、個人の行動がどのように変化するのかを明らかにしてきた。もし個人が一人一人実験に参加し、行動の測定が行われる場合には、今回述べるような階層的データ構造にはならない。集団に関する変数は実験条件という固定化された効果のみであり、従来の分散分析や回帰分析を行うことで適切な分析が可能である。

問題となるのは複数の個人が集団を形成して実験に参加し、データを収集が行われるときである。このような場合、集団ごとによって固有の雰囲気や状況が形成されるからである。集団ごとによって生まれる固有の状態は実験条件のように固定化されたものではなく、実験参加者によって異なる(実験参加者に依存する)ものである。このように、集団間にばらつきが存在するとき、それ自体を変数として扱わなければならない。それは集団の凝集性であったり、パフォーマンスであったりするような、集団レベルの変数である。

つまり、階層的データとは個人に関する変数(個人レベルの変数と呼ぶ)と、その個人が複数集まって形成されたペアや集団に関する変数(ペアレベル、集団レベルの変数と呼ぶ)が共に存在するデータである。例えば、恋愛関

係において、関係満足感(個人レベル)と、その関係の交際期間(ペアレベル)を測定しているようなデータである。このようなデータは個人ごとに得点が入力される変数(個人レベル)と、ペア内の個人で同じ値が入力される変数(ペアレベル)が混在する(Table1)。

Table1　個人レベルの変数とペアレベルの変数が混在するデータセット例

ID	ペア番号	交際期間	満足感
1	1	1	1.5
2	1	1	5
3	2	5	5
4	2	5	4.5
5	3	15	5
6	3	15	5
7	4	3	5
8	4	3	5
9	5	12	5
10	5	12	4
⋮	⋮	⋮	⋮

※上のデータの場合、交際期間がペアレベル、満足感が個人レベルである。

また、個人が複数回、質問紙に回答する場合も同様に階層的データである可能性がある。各時点における回答は下位レベルの変数、そして回答者の特性といった時系列によって変化しないと仮定される変数が上位レベルの変数となる。見かけは異なるが、個人-集団の階層性と同等の構造を持っている(Figure1)。

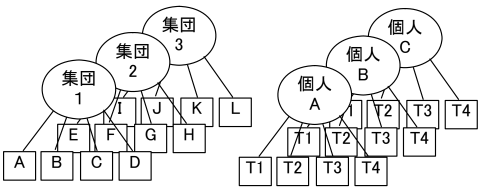

Figure1　階層的データ構造

従来の分析法の問題点

もしわれわれが階層性をもったデータを得た場合、従来の方法を用いることによってどのような問題が生じるのであろうか。

階層的データを扱う上で陥りやすい問題として、1.サンプルの独立性の仮定に違反する、2.ペアや集団内で平均値を算出し、それをペア・集団レベルの変数とする、3.個人レベルの変数間の関連を、ペアレベル・集団レベルの変数を無視して分析・解釈を行う、という三つが挙げられる。

階層性をもったペアデータ構造は Table1 のように、個人レベルの変数にはそれぞれ違う値が、ペアレベルの変数は2人に共有された値が入力される。ここで個人レベルの変数とペアレベルの変数の相関などの関連を見る場合、従来ならば以下に挙げるような二つの方法がとられてきた。一つめは、そのまま(n=個人の数)データ行列を用いて相関係数を算出する、二つめは、ペアごとで個人レベルの変数の平均値を算出し、n=ペア数で相関係数を算出する、である(Table2)。

Table2　個人レベルの変数を平均化し、集団レベルの変数とするデータセット例

ペア番号	交際期間	満足感の平均値
1	1	3.25
2	5	4.75
3	15	5
4	3	5
5	12	4.5
⋮	⋮	⋮

まず一つめの方法、N=個人の数でデータを分析する場合の誤りは、統計学におけるサンプルの独立性の仮定に違反することである。サンプルはそれぞれ相関関係を持つことを許されない。これは分散分析や回帰分析のようなモデルでは、個人の持つ誤差得点が無相関であるという仮定のもとで係数が算出されているからである。もしこの仮定が違反された場合、推定される統計値が歪むと同時に、有効サンプル数を多く見積もっているために有意になりやすい(タイプⅠエラーを犯す)という誤りが生じる。このような誤りについて、われわれはつい過小評価してしまいがちであるが、統計的検定結果を重視する社会心理学的分析では重要な問題であろう。

次に二つめの方法、n =集団の数でデータを分析する誤りは、個人レベルの変数をペアごとに平均化させ、ペアレベルの変数とするという点にある。仮に平均化した値を算出し、ペアレベルの変数との相関係数を求めると、真の値(母集団における真の相関係数)よりも希薄化した値が得られる。これは個人レベルの変数の平均値は、必ず誤差が含まれることによる。この誤差とは個人が独自にもつ部分や、測定誤差などが含まれる。ペアレベル変数の推定値としての得点にこのような誤差が含まれると、相関係数が小さくなるのである。これと同様の問題として、尺度得点を算出するときに信頼性が低いと相関が希薄化されることが挙げられる。このように、ペアや集団内で平均値を算出すると、従来の方法では分析結果に誤差が含まれてしまい、真の値を推定することができなくなるのである。加えて、平均値を用いた方法では、せっかく多くのサンプルを得てもかなりの情報を捨ててしまうことになる。これはデータの効率性という意味で欠点をもつ。

以上のような問題と同時に、個人レベルの変数同士の関

連を見る場合においても、ペアレベルの変数の影響が無視できない場合に、問題が生じるときがある。それは、個人レベルの変数間の相関がペアごとで異なっている場合である。

例えば、多くの恋愛関係のペアからデータを得ることを考えてみよう。ある研究者が、関係安定性の認知(関係が安定していて、長く継続するであろうという認知)が個人の関係満足感によって予測される、という仮説をたてたとき、その研究者は個人レベルの変数として関係安定性の認知と関係満足度を測定するだろう。そして仮説を検証するために関係安定性の認知と関係満足感の相関係数を算出したとき、ペアごとによってこの2変数の相関関係が異なることがある(Figure3)。あるペアでは、関係に満足していれば関係が安定した状態であると認知しているが、別のペアでは関係に満足していても、それほど関係が安定した状態でないと認知しているかもしれない。このようなペアごとの相関関係の違いは、ペアレベルの変数、例えば交際期間などによって説明されるかもしれない。つまり、交際期間が長くなるにつれ、満足感は安定性の認知に影響しなくなる、などが考えられる。

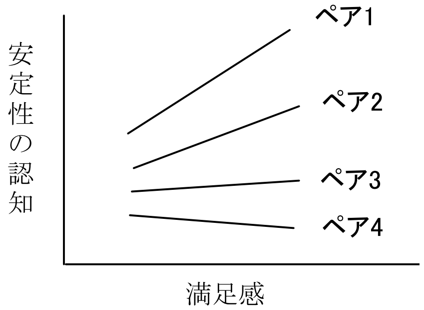

Figure3　ペアごとによる相関関係(回帰直線)のばらつき

階層的分析の発想

それでは階層的データをどのように分析すれば、適切な推定値を得ることができるのであろうか。

階層的分析法の発想は、ペアや集団内の個人の相関を考慮し、一つのモデルで表現することによって、従来の方法によって起こる誤りを防ぐことにある。

ペア内の個人同士の相関を考慮するために、階層的分析では変量効果[2)]を導入し、個人の誤差得点に構造を取り入れている。従来の分散分析や回帰分析では誤差得点はただのあまりもの、という位置づけであったのに対し、階層的分析では説明可能な誤差分散と説明不可能な完全な誤差分散をわけている。このような処理によって有意性の検定が正しくなるだけでなく、さらなる説明の可能性を知ることができるという利点がある。

さらに、階層データを一つのモデルですべて表現することによって、平均値の得点化によって生じる誤差を含まずに推定値を算出することができる。これは SEM において因子得点を算出せずに構造方程式を解くことで、尺度の信頼性の問題を回避できる点と同様である。

また、モデルを一つにすることによって、個人レベル変数同士の相関係数を、新たな潜在変数として導入し、集団レベルの変数との関連を見ることも可能にする。これは先ほどの例でいえば、関係安定性の認知と満足感の相関係数が、ペアの交際期間によってどのように変化するかを明らかにすることである。このような分析が可能であるもの、変量効果の導入による。

階層的分析にはいくつか種類がある。一つは近年社会心理学でもとりあげられている HLM(Hierarchical Linear Model: 階層線形モデル)である。また、分散分析の文脈で発展した、混合モデル(Linear mixed model)がある。この二つのモデルは、数学的には同様の構造であり、結果もほぼ一致する。さらにペアデータに特化した階層的分析としてペアワイズ相関分析(Pair-wise correlation analysis)がある。前者二つと後者の分析法は、同様に階層的データを分析する方法であるが、視点が異なっているため使い分ける必要がある。

それでは実際に統計パッケージで実行可能な分析法を見てみよう。

階層的分析のメカニズム

HLM

Bryk & Raudenbush(1992)が提案した方法で、階層線形モデルとも呼ばれる。この分析法は個人レベルの変数を従属変数としたときに、独立変数に個人レベルだけでなくペアレベルの変数も投入し、適切に分析を行うことができる。

HLM ではレベル 1 のモデルとレベル 2 のモデルを分けて表現することが多い。レベル1のモデルとは個人レベルの変数同士の関連を表す回帰式であり、例でいえば、関係安定性の認知を予測する満足感の効果を表現している。レベル 2 のモデルは、レベル 1 のモデルで表現された切片と傾き(大抵は回帰係数を表す)を従属変数とし、ペアレベルの変数(交際期間)を独立変数とした回帰式である。つまり、レベル 2 のモデルは一つの切片と、レベル 1 で投入された変数の数を合わせた分、回帰式が存在することになる。式で表すと式 1～式 3 のようになる。

レベル 1 の式から説明すると、式 1 において最初のY_{ij}とはj番目のペアのiさんの従属変数、つまり安定性の認知の得点を表している。X_{ij}はj番目ペアのiさんの個人レベルの独立変数、つまり満足感を表している。

続いて、βは回帰式で推定される値で、β_{0j}は切片、β

$$\text{レベル1}\quad Y_{ij} = \beta_{0j} + \beta_{1j}\cdot X_{ij} + r_{ij} \qquad \text{式1}$$

$$\text{レベル2}\quad \beta_{0j} = \gamma_{00} + \gamma_{01}\cdot W_j + u_{0j} \qquad \text{式2}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}\cdot W_j + u_{1j} \qquad \text{式3}$$

$$\text{ただし、}\quad \begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{pmatrix}\right] \qquad \text{式4}$$

$_{1j}$は満足感についての回帰係数を表している。βの添え字について、左の数字は係数の通し番号で大抵切片に 0 がつけられる。本論文の例では独立変数が一つしかないので 1 までしかない。右の添え字はペアをあらわしており、ペアごとに切片と回帰係数が求められていることを表している。

r_{ij}はj番目のペアのiさんにおける残差得点である。残差とは「未だ説明されていない」という意味も持ち、残差分散が大きく残っている場合は、まだ説明されていない独立変数がある可能性が示唆される。また、測定誤差などといった完全な誤差もここに含まれる。

このように、レベル 1 のモデルは従来の回帰モデルとまったく同様であり、個人レベル同士の関連を表しているに過ぎない。

HLMの特徴はレベル 2 のモデルを導入することである。レベル 2 のモデルはβ_{0j}とβ_{1j}を従属変数とすることが式2,3 を見ることでわかる。これは、切片や回帰係数がペアによって分散しており、その分散をペアレベルの変数(式ではWで表現)で説明を行おうとしているのである。先ほどの例でいえば、安定性の認知と満足感の関係が、ペアの交際期間によって異なっているのではないか、という仮説を検証することに該当する。

レベル2のモデルにおいて、γとは、レベル2のモデルの係数(切片や回帰係数)を表している。βがレベル 1 の係数であるため区別するために別の文字を使っているのである。つまり、レベル 1 と同様にγ_{00}は切片を、γ_{01}は回帰係数を表している。γの添え字は、左側がβの左側の添え字と対応しており、β_{0j}のモデルの係数ならばγの左の添え字も 0 になっている。右側の添え字はペアレベル変数の独立変数の通し番号になっており、0 が切片を表し、その後1から順番につけられる。今回の例では交際期間という一つの変数しか仮定していないため 1 までしかない。

また、レベル 2 のモデルで見られるu_{0j}とu_{1j}はそれぞれの回帰式における誤差であり、HLMでは式4のように平均 0 の多変量正規分布に従うことが仮定されている。多変量正規分布とはそれぞれの変数が正規分布し、また変数間に共分散が仮定されているような分布である。HLMではu_{0j}とu_{1j}の分散をτ(タウと読む)と呼び、u_{0j}の分散をτ_{00}、u_{1j}の分散をτ_{11}とする。またu_{0j}とu_{1j}の共分散は$\tau_{01}(=\tau_{10})$と呼ぶ。τ_{00}やτ_{11}が有意である場合、β_{0j}やβ_{1j}(つまり切片と回帰係数)はまだペア・集団間でばらつきがあることを示している。つまり、まだ分析に投入されていないペアレベルの変数で説明できる可能性があることを示している。

HLMは、HLM6[3]という専用のパッケージで解析が可能である。これはスチューデント版であれば無料で手に入る 1)。ソフトウェアの使用方法はここでは触れないが、HLM6 を使えば比較的簡単にHLMを実行することができる。またHLM6 は変数のセンタリング(平均値を集団ごとに 0 にすること)がすぐできるので、他のソフトウェアよりも便利である。

混合モデル

混合モデルは一般的には線形混合モデルと呼ばれ、従来の一般線形モデルを拡張し、被験者内要因を含む分析をより適切に扱うことができるモデルである。

先に述べたように、反復測定によって得られたデータ構

$$Y_{ij} = (\gamma_{00} + \gamma_{01}\cdot W_j + u_{0j}) + (\gamma_{10} + \gamma_{11}\cdot W_j + u_{1j})X_{ij} + r_{ij} \qquad \text{式5}$$

$$Y_{ij} = \underbrace{(\gamma_{00} + \gamma_{10}\cdot X_{ij} + \gamma_{01}\cdot W_j + \gamma_{11}\cdot W_j\cdot X_{ij})}_{\text{固定効果}} + \underbrace{(u_{0j} + u_{1j}\cdot X_{ij} + r_{ij})}_{\text{変量効果}} \qquad \text{式6}$$

Table3 HLM・混合モデルのデータセット例

ID	ペア番号	性別	交際期間	安定性	満足感	満足感センタリング
1	1	1	1	3.9875	1.5	−1.75
2	1	2	1	4.4	5	1.75
3	2	1	5	3.275	5	0.25
4	2	2	5	4.275	4.5	−0.25
5	3	1	15	4.1125	5	0
6	3	2	15	4.6	5	0
7	4	1	3	4.4125	5	0
8	4	2	3	4.525	5	0
9	5	1	12	4.4375	5	0.5
10	5	2	12	3.975	4	−0.5
：	：	：	：	：	：	：

造は階層的データの一種であるので、HLM と同様の問題が生じる。よって反復測定を含む被験者内要因計画では、混合モデルという階層性を扱うモデルが提案されているのである。

原理的には HLM も混合モデルも同様の分析法である。HLM の式を展開すれば混合モデルの式と一致する。式1に式2と式3を代入すると、式5となる。そして式5を展開すると式6となり、混合モデルでいうところの固定効果と変量効果に分離することができる。

最初のγ_{00}は式全体の切片を表す。そしてγ_{10}、γ_{01}、γ_{11}はそれぞれ効果の推定値であり、回帰係数と同様である。特にγ_{11}はW_jとX_{ij}の交互作用という形で表現されているが、これは個人レベル変数であるYとXの関連に及ぼす、集団レベルの変数Wの効果であると読み取ることができる。

階層的分析の実際

混合モデルは、SAS や SPSS といった社会心理学者が比較的慣れ親しんだ統計ソフトウェアで実行が可能なので、HLM6 と比べとっつきやすいかもしれない。しかし、HLM6 のようにセンタリング(後述)を自動的に行ってくれる機能がない上に、自分で式を展開して入力する変数を確認しなければならない分、面倒な作業が必要となる。

ここでは、SAS による混合モデルの使い方を簡単に説明する。SAS では PROC Mixed を用いて混合モデルを実行することができる。

今回は、関係安定性の認知が、関係の満足感によって予測されることを例に、実際に恋愛関係にあるペア 194 組のデータを用いて分析を行った。調査回答者は近畿地方および中部地方の大学生とその恋人である。関係安定性の認知として清水・大坊(2004)の関係性認知尺度(18 項目の SD 尺度)を、満足感として「私は関係に満足している」などの 2 項目を、用いて測定した。また、交際期間は月単位で測定した。データセットは Table3 のように並べ SAS に入力した。なお、標準化係数を出力したい場合、変数をすべて標準化しておく必要がある。SAS には Standard プロシージャがあるので、それを用いて mean =0 std =1 を指定すれば標準化されたデータを得ることができる。

Figure3 は SAS のプログラム例である。

```
PROC Mixed  covtest noitprint;
class       pair;
model       stability =  satisfaction duration / solution  ddfm = bw;
random      intercept  satisfaction / sub = pair type = un;
run;
```

Figure3 SASの混合モデルのプログラム例[4)]

上の例では、従属変数として関係安定性の認知(stability)を、個人レベル変数の独立変数として満足感(satisfaction)を、ペアレベル変数として交際期間(duration)を投入している。なお、満足感の得点はペア内でセンタリングしたものを用いている。

まず、class ステートメントでは pair という変数を投入している。この変数はペアを弁別するためのもので、ペア番号が入力されている。

次の model ステートメントには、左側に従属変数、右側にはレベル 1,2 両方の独立変数をすべて投入する。この式に指定されるのは固定効果の独立変数である。スラッシュの後にある solution オプションは回帰係数を出力するためのものである。

その下の random ステートメントは変量効果を指定する。具体的にはまず切片を表す intercept、そしてレベル 1 の独立変数(この場合は satisfaction)を投入する。切片の変量効果とは関係安定性認知のペアの平均値についての、ペア間分散を表している。つまり関係安定性がペア間でどれほどばらついているかを表す。満足感の変量効果とは関係安定性認知と満足感の相関係数についての、ペア間分散を表している。つまり、関係安定性認知と満足感の関連がペア間でどれほどばらついているかを表す。スラッシ

ュの後にある sub オプションは変量効果をどの変数で設定するかを指定する。この例ではペア間の分散を変量効果として設定するので pair を投入している。

また、階層的分析を行う場合、変数をセンタリングしておくことが勧められている(Bryk & Raudenbush, 1992)。センタリングとは、それぞれのペアの平均値を個人レベル変数の得点から減ずることである。こうすることで、レベル 1 のモデルではペア間の違いを統制し、純粋な個人レベルの関係を解釈することができるのである。またレベル 2 の変数をセンタリングするときは全体の平均値との差を求める必要がある。レベル 1 の変数についてセンタリングを行うには、ペアごとの平均値との差を求める必要があり、SAS では行うことができない。よって、元のデータセットにもどりセンタリングした変数を作成しておく必要がある。レベル2はmeansプロシージャで得た平均値を減じればセンタリングすることができる。

SAS による混合モデルの結果は 2 種類の出力に注目する必要がある Covariance parameter estimates と、Solution for Fixed Effects である。

前者の出力は変量効果についてのもので、切片やレベル1のモデルに投入した独立変数についての分散がどれほど残っているかを示している。この値が有意であるならば、まだ説明されていない分散が残っている可能性があることを示している。Table4 の場合、切片も満足感も有意に分散が存在していることが示されている(UN(1,1)が切片の分散 τ_{00}を、UN(2,2)が満足感の傾きの分散 τ_{11}を表している。またUN(2,1)が両者の共分散 τ_{10}である)。

Table4 SAS による混合モデルの変量効果出力

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
UN(1,1)	pair	.584	.078	7.450	<.0001
UN(2,1)	pair	−.016	.036	−.450	.655
UN(2,2)	pair	.091	.037	2.430	.008
Residual		.278	.047	5.940	<.0001

また、投入した独立変数の説明力を知りたいとき、これらの分散減少量から計算することができる。レベル 1 の変数の説明力は Residual の減少量から知ることができる。例えば満足感の説明力は、まず何も独立変数を投入しないときの Residual の分散を算出する。このデータの場合、Residual の分散の estimate は 0.533 であった。つまり 0.533 の分だけ安定認知には分散が存在しているのである。そしてTable3のように満足感をレベル1に投入することで 0.278 まで減少していることから、(0.533-0.278) / 0.533 によって 48%の分散が説明されていることを知ることができる。これは0.533のうち、減少した分散の割合を計算することで分散説明力を計算しているのである。

同様に、レベル 2 の変数の説明力を知りたいとき、切片の分散の減少量を計算する必要がある。このデータの例では、まず満足感のみを投入したモデルの切片の分散のestimateを算出し、その後レベル2の変数、つまり交際期間を投入したモデルの切片の分散の estimate を算出する。満足感のみを投入したモデルの切片の分散は 0.592 であった。そして Table4 のように交際期間を投入したあとのモデルでは0.584であった。これから1.4%の説明力を持っていたことがわかる。

このように、レベル 1 とレベル 2 の変数では説明している分散が異なっている。それは従属変数の分散の中で、個人レベルで説明できる部分とペアレベルで説明できる部分を切り分けているからである。

Table5 SAS による混合モデルの固定効果の出力

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	.009	.062	191.000	.150	.881
sat	.247	.044	183.000	5.620	<.0001
dur	.129	.061	183.000	2.140	.034

後者の出力(Table5)は各変数の回帰係数を示したものである。今回の例では標準化した変数を投入しているので推定値は標準化係数である。

満足感は.247 と中程度の効果を持っており、交際期間はやや小さめであるが有意な効果が見られた。つまり、関係安定性の認知は、個人の満足感が高いほど大きく、また交際期間が長くなるにつれやや大きくなることがわかる。

先ほどの例では、ただ単にレベル1の変数とレベル2の変数を別に投入したにすぎないが、HLM や混合モデルでは従属変数とレベル 1 の変数との回帰係数の分散もレベル 2 の変数で説明することができる。

実際の SAS プログラムでは以下のように、満足感と交際期間の交互作用という形で表現する。

```
PROC Mixed  covtest noitprint;
class    pair;
model    stability =  satisfaction  duration  satisfaction*duration / solution  ddfm = bw;
random  intercept  satisfaction / sub = pair type = un;
run;
```

Figure4 レベル 1 とレベル 2 の交互作用を追加した SAS のプログラム例

このモデルを実行すると、変量効果の出力はTable6のようになる。ここで Table4 と見比べると満足感の分散が減少していることがわかる。つまり、満足感の回帰係数の分散が一部説明されたことを示している。その割合は 3%であった。

Table6 SAS による混合モデルの変量効果の出力 2

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
UN(1,1)	pair	.584	.078	7.470	<.0001
UN(2,1)	pair	−.020	.036	−.550	.579
UN(2,2)	pair	.088	.036	2.440	.007
Residual		.278	.046	6.050	<.0001

またそれぞれの標準化係数を算出すると Table7 のようになる。有意傾向ではあるが、関係安定性認知と満足感の関連は、交際期間が長くなるにつれ小さくなる(あるいは負の係数になる)ことがわかる。このことから、交際期間が長くなると関係の状態は個人の感情によって変動しなくなる、つまり関係独自のダイナミクスを形成し始める、ということが示唆される。

Table7 交互作用項の標準化係数

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	.009	.062	191.000	.150	.884
satisfaction	.232	.044	182.000	5.240	<.0001
duration	.133	.061	182.000	2.190	.030
satisfaction*duration	−.083	.047	182.000	−1.760	.080

HLM・混合モデルの限界と解決案

さて、これまでの分析例において、実は見落としている部分がある。それはペア・集団内で個人レベルの変数の得点が類似している場合、ペア間の分散の効果を説明していない点である。HLM や混合モデルのレベル 1 のモデルで説明されているのは、ペア・集団間の分散をセンタリングによって取り除いた効果である。よって、ペア・集団内で得点が類似している場合、レベル 1 の分析結果からでは、ペア間の分散の効果を知ることができないのである。

ペア・集団間の効果とは、いわばペア・集団の平均値の効果である。個人レベルの変数は個人に帰属するものであるが、変数によってはペアや集団の影響を受けている可能性もある。具体的にいうならば、もし安定性の認知と満足感がペア内で類似している場合、個人の安定性の認知と満足感との関連は、個人が満足すればその個人が関係安定性を高く評価するということ(個人レベルの関連)と、2 人ともが満足している関係は安定性が高い(2 人とも安定性を高く認知している)ということ(ペアレベルの関連)を同時に含んでいる、ということである。実際、安定性の認知と満足感のペア内の類似度を表す級内相関係数は両変数とも有意であった(安定性の認知; r=.47, p < .01, 満足感; r=.25, p<.01)。

HLM ではレベル 1 のモデルに、ペア・集団内で類似した変数を投入する場合、上記のようなペア・集団間の効果、つまりペアレベルの関連を分析することができず、説明されないまま残ってしまうのである。

上記の問題についての解決法として思い浮かぶのは、妥当でないものを含めて三つある。それは、1.満足感をセンタリングせずに分析に投入する、2.ペア内の満足感の平均値を算出し、それをレベル 2 の変数として分析に投入する、3.別の階層的分析を用いる、である。

まず、一つめの方法は、妥当でない方法であり、いくつかの問題点がある。それは分析によって得られた係数が個人レベルの効果なのか、ペアレベルの効果なのか判別がつかないこと、もう一つはレベル 2 の変数との交互作用を見ることができない、というものである。まず一つめの問題は、センタリングしないことによって得られた係数が、「個人が満足すれば、個人は関係を安定していると認知している傾向にある」ことを表しているのか、「ペアが両方とも満足していると、関係が安定している(2 人とも安定していると認知している)」ことを表しているのか、が判別できないことである。これは個人レベルの効果とペアレベルの効果の混在の問題である(Griffin & Gonzalez, 2000)。もしセンタリングした変数を投入していると、ペアの平均値の分散が取り除かれていることから、前者の効果であることが特定できるのである。二つめの問題は、先ほど例に示したような、安定性の認知と満足感の関連自体に交際期間が影響する、というモデルを検討できないことである。

次に、二つめの方法は、あまり妥当ではないが、しかし有効な方法である。まず妥当でない理由は、平均値を算出してレベル 2 の変数とすることは、相関係数の希薄化を招くからである。しかし、もし変数のペア内類似度が十分大きいなら、その希薄化の影響は最低限にとどめることができる。そしてペアの平均値を投入することによって、関係安定性の認知に及ぼす満足感の影響について、個人レベル・ペアレベル両方の視点から解釈を行うことができるのである。実際に満足感の平均値を投入したモデルを解析すると、平均値の効果は β=.53, p < .01 であり、また切片の分散に対する説明力は 79%であった。これはかなりの影響力があり、個人レベルの効果よりも大きい。つまり、個人が満足していることは個人の安定性の認知を高めるが、それ以上に、ペアが満足していることが関係安定性を予測していることを知ることができる。

なお、ペア・集団内の類似度を知る方法として、内的一貫性の指標である α 係数を算出することが挙げられる。α 係数はペアの平均値の分散のうち、真のペアレベルの変数の値の分散が占める割合を表している。この値が十分高いならば平均値を算出しても、相関の希薄化はそれほど問題にはならないだろう。

このペア・集団内の内的一貫性を表す α 係数はペアであっても集団であっても同様の方法で算出できる。まず α

係数を算出したい変数を従属変数、そしてペア・集団を弁別するカテゴリー変数(上記の例では pair の変数がそれにあたる)を独立変数とした 1 元配置の分散分析を行い、水準間変動と水準内変動を得る。そして 1-(水準内変動／水準間変動)によって α 係数を得ることができる(Kenney & La voie, 1985; 村上, 1999)。なお、先の例であげた満足感のペア内 α 係数は.43 と非常に低い値であるため、平均値は投入するべきではなかったことがわかる。

もし、ペアでなく一つの集団に多くの個人が属しているようなデータであれば、α 係数は比較的高くなるため、平均値を用いることはそれほど問題ではなくなる。ペアよりもっと規模が大きい集団の階層的データを扱う場合は HLM や混合モデルで個人レベルと集団レベルの効果を比較することは有効であると考えられる。

ただし、このようなペア・集団間の分散を問題とする場合、ペア・集団内で類似性が認められることが前提とされる。つまりペア内で類似性がない変数は平均値を算出する意味がそもそもないのである。それは真に個人に帰属される変数であり、ペア間の分散は存在しないからである。もし、先述の α 係数算出の際、1 元配置分散分析の水準間効果が有意でない場合(これはペア・集団内の級内相関係数が有意でないことを示す)、ペア・集団内の類似性がないと考えてよい(Kenney & La voie, 1985)。なお、級内相関係数は式 7 で得ることができる(ただし、MS_Bは水準間変動、MS_Wは水準内変動、nはペア・集団内にいる人数を表す)。

$$\frac{MS_B - MS_W}{MS_B + MS_W(n-1)} \qquad \text{式 7}^{5)}$$

なお、級内相関とはペア・集団内の個人の得点間の相関係数の平均値である。つまり、どれほどペア・集団内で得点が類似しているかの指標である。

ペアワイズ相関分析

三つめの解決方法は、Griffin & Gonzalez(1995)が提案したペアワイズ相関分析を用いることであり、上記の疑問を解決する上で最も妥当で有効な方法である。この方法は、ペア内で類似している変数について、個人レベルとペアレベルに分けて分析を行うことができる。

ペアワイズ相関分析は Figure5 のようなモデルを仮定している。Figure5 を見ると各観測変数は二つ用意されていることがわかる。これは Table8 のように、ペアで入れ替えられたもう一つのデータセットを用いて分析に投入されているからである。なお、詳しい分析方法については

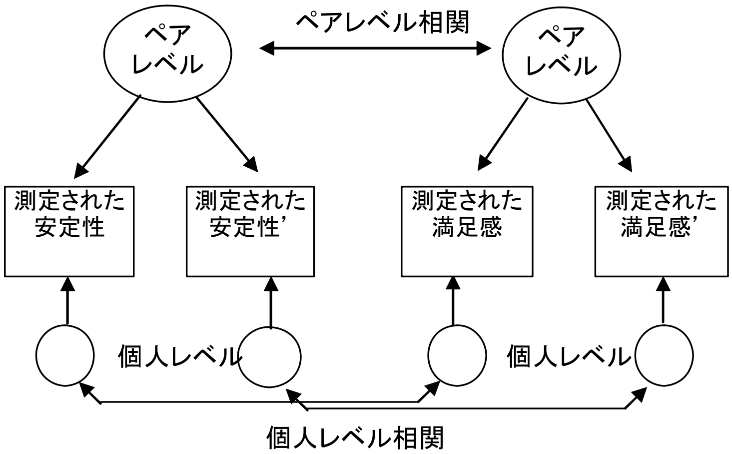

Figure5　ペアワイズ相関分析のモデル図
Gonzalez & Griffin (2000)を基に作成

Griffin & Gonzalez(1995)、石盛・清水(2004)に説明がある。

また Figure5 では、潜在変数としてペアレベルの変数と個人レベルの変数が推定されている。これはペアレベルにおいて、ペアで共有されている安定性や満足感を推定しているのである。このように、潜在変数を用いてモデルを構築することにより、平均値を算出する必要がないため相関係数が希薄化されるという問題を避けることができる。

Figure5 で示される、個人レベルの相関とは、ペアの平均値の効果を取り除いた効果という意味で、センタリングしたレベル 1 の HLM や混合モデルと同様の結果を得ることができる。またペアレベルの相関は、HLM や混合モデルでは正確に推定を行うことができないが、上記の平均値をレベル 2 のモデルに投入する方法で代用することができる。

ペアワイズ相関分析は、個人レベルの変数を用いて、さらに平均値を算出することなくペアレベルの関連を見ることができる点で、HLM や混合モデルよりも秀でている。また、ペアワイズ相関分析は SEM(構造方程式モデル)を用いて行うことができる(Gonzalez & Griffin, 1999)。よって適合度を算出することができると同時に、他の分析に応用することができる点で、汎用性も優れている。

SEM を用いてペアワイズ相関分析を行う手順は、1.分析用のデータセットを用意する、2.Amos などの SEM 用ソフトにデータを投入し、構造方程式を書く、3.得られた係数をもとに有意性検定を行う、の 3 ステップである。

まずステップ 1 に、Table8 のようなペアでデータを入れ替えたデータセットを用意する。ここで注意すべきことは、もしペア内で個人を識別できるカテゴリー変数がある場合、それも同時にデータセットに含めておくべきことだ。なぜなら、識別可能なデータと識別不能なデータでは分析手法が異なるからである(Gonzalez & Griffin, 1999)。上の

Tabel8 ペアワイズ相関分析のデータセット例

ID	ペア番号	性別	交際期間	安定性	安定性’	満足感	満足感’
1	1	1	1	3.9875	4.4	1.5	5
2	1	2	1	4.4	3.9875	5	1.5
3	2	1	5	3.275	4.275	5	4.5
4	2	2	5	4.275	3.275	4.5	5
5	3	1	15	4.1125	4.6	5	5
6	3	2	15	4.6	4.1125	5	5
7	4	1	3	4.4125	4.525	5	5
8	4	2	3	4.525	4.4125	5	5
9	5	1	12	4.4375	3.975	5	4
10	5	2	12	3.975	4.4375	4	5
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

例では性別というカテゴリー変数を投入している。しかし、今回は解説が目的のため、性別の効果は見ない。

次にステップ2に、構造方程式をAmosなどのSEM用のソフトウェアで記述する。今回はAmosによる例を示す。具体的にはFigure6のようにモデルを描くことで推定値を得ることができる。解の推定のときは、1.個人レベルの潜在変数(e1~e4)から観測変数へのパスはすべて1に固定する、2.ペアレベルの潜在変数(F1～F2)のうち、どれか一つの潜在変数は、観測変数へのパスを二つとも1に固定し、その他の潜在変数は一つだけ1に固定する、という二つの制約を加えることで識別可能となる。また、結果は標準化された係数しか意味を持たないので、出力のオプションで「標準化推定値」にチェックをしておこう。以上の準備をしてから分析を実行するとFigure6のような結果が得られる。F1とF2の間の係数がペアレベルの相関係数、e1
とe3(あるいはe2とe4)の間の係数が個人レベルの相関係数である。

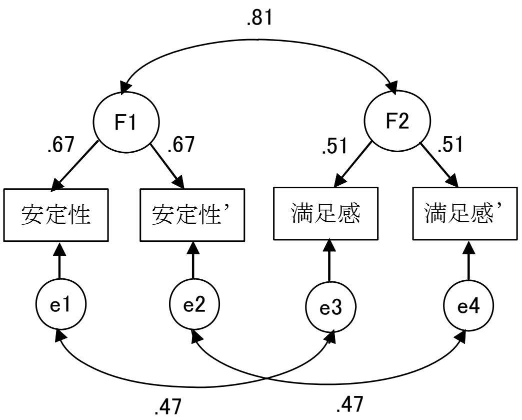

Figure6 Amosによるペアワイズ相関分析

最後に、有意性の検定を行う必要がある。それはAmosで出力されるp値はn=個人の数で計算されているが、ペアワイズ相関分析では、それぞれの係数は調整されたnを用いてp値が計算されるためである。ペアレベルの相関であるr_Dの有意性検定は式8で得られるZ値に基づいて行われる(ただし、nはペアの数)。Z値が1.96を超えると5%水準で有意となる。

$$Z = r_D \sqrt{N_D} \qquad \text{ただし、} N_D = \frac{2n(r_{xx'} \cdot r_{yy'})}{1 + r_{xx'} \cdot r_{yy'} + r^2_{xy}}$$

式8

また、個人レベルの相関であるr_Iの有意性検定は式9で得られる、自由度n-1のt値に基づいて行われる(ただし、nはペアの数)。

$$t_{n-1} = \frac{r_I \sqrt{n-1}}{\sqrt{1-r^2_I}}$$

式9

識別可能なときの有意性検定についてはGonzalez & Griffin(1999)に詳しい説明がある。

以上のように、HLMや混合モデルの欠点を補うことができるペアワイズ相関分析にも当然、限界点がある。それは、ペア内類似性が極端に高い、あるいは極端に低い場合に推定値が不安定になる点である。これはペアワイズ相関分析がペア内類似度を仮定したモデルだからである。しかし、もし極端に類似度が高い場合はペアレベルの変数として分析に含めればよいし、低い場合は従来の方法を用いて分析すれば問題ない。ペアワイズ相関分析は従来の方法では十分に扱えなかったデータ(ペア内の類似性が有意である個人レベルの変数を含んだデータ)に特化した方法であるといえる。

階層的分析の比較

ここで、これまでとりあげた階層的分析の簡単な比較を行い、どのようなデータのときにどの分析法を用いるのがよいのかについてまとめておこう。

まず、階層性のないデータのときはHLMや混合モデル、ペアワイズ相関分析を用いる必要がない。従来のよう

な分散分析、回帰分析を用いて分析することは問題がない。

階層性のあるデータとは、個人に帰属される変数とペアや集団に帰属される変数の両方が含まれているデータであった。このようなデータを用いて分析する場合は HLM や混合モデルを使う必要がある。もしこれらを使わない場合は、有意性検定の正確さを失い、得られる係数も希薄化されたものになってしまう可能性がある。

HLM と混合モデルの違いはほとんどない。しかし、混合モデル(Mixed プロシージャ)は一般線形モデルの文脈で発展している点から、カテゴリーデータを扱う場合には利点がある。要因間の交互作用や単純主効果の下位検定も簡単に行うことができるので、便利である。HLM(HLM6)の利点としてはセンタリングがすぐに行える点、あと解の推定に関する様々な設定(収束基準や推定方法など)を変更することができる点にある。

また、個人レベルの変数に類似性が見られる(級内相関が有意である)場合には、HLM や混合モデルでも適切な解を得ることができないこともある。このようなときはペアワイズ相関分析を用いて個人レベルとペアレベルを分けて分析する必要がある。ただペアワイズ相関分析はペアデータである必要があるため、制約が大きい。もし一つの集団の規模が十分大きいデータのときは、信頼性係数を算出して内的整合性を評価し、十分大きい値が得られたなら平均値を算出して HLM や混合モデルに投入しても大きな問題はないと考えられる。

結論

階層的分析の方法は、比較的新しくメカニズムが複雑なため、とっつきにくく敬遠されがちである。しかし、ある程度の原理と使用法を知れば十分活用できる範囲のものであると思われる。特に階層的データを入手する機会が多い社会心理学者はぜひ習得しておくべき方法論であると考える。

「階層的分析は、従来の方法が統計的に間違えているから使うべきである」という消極的な主張は、十分ではないだろう。階層的分析を用いることによって、これまででは知ることができなかったことを知ることができる、という積極的な意味を本論文では強調したい。これは具体的には個人レベルとペア・集団レベルを一つのモデルで同時に分析することによって、個人レベルの推定値のペア・集団間の分散を説明したり、ペア・集団レベルにおける関連を個人レベルの変数から推定したりすることができることを指している。このような知見は階層的分析によって解析的に捏造されたものではなく、集団や対人関係のダイナミクスに関心がある研究者がこれまで知ろうとしていたが知ることができなかったものである。ただ単に階層的分析をやってみて、結果が出たから解釈するというのではなく、個人と集団や対人関係のダイナミクスについての仮説を持っていれば十分に解釈可能なものであると考える。

しかし、階層的分析も統計的分析法の一種であることには変わりはない。よって、それによって示された結果が真実である保証はない。階層的分析において得られた結果がどの範囲で有効であるのか、何を意味しているのかをしっかり理解した上で使用する必要がある。

引用文献

Bryk, A. S. & Raudenbush, S. W. 1992 *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage.

Gonzalez, R. & Grffin, D. 1999 The correlational analysis of dyad-level data in the distinguishable case. *Personal Relationships*, 6, 449-469.

Gonzalez, R. & Grffin, D. 2000 On the Statistics of Interdependence: Treating Dyadic Data with Respect. In W. Ickes & S. Duck (Ed.) *The Social Psychology of Personal Relationships*. New York: Wiley & Sons, Ltd.(石盛真徳 2004 相互依存性についての統計学 二者間データの慎重な取り扱い 大坊郁夫・和田実 監訳 パーソナルな関係の社会心理学 (pp.221-256) 北大路書房)

Griffin, D., & Gonzalez, R. 1995 Correlational Analysis of Dyad-Level Data in the Exchange Case. *Psychological Bulletin*, 118, 430-439.

石盛真徳・清水裕士 2004 二者間データ分析へのペアワイズ・アプローチ 対人社会心理学研究, 4, 127-133.

Kenney, D. A. & La Voie, L. 1985 Separating individual and group effects. *Journal of Personality and Social Psychology*, 48, 339-348.

Lewin, K. 1950 *Field theory in Social Science*. New York,Harper. (猪股佐登留訳 社会科学における場の理論 誠信書房)

村上隆 1999 信頼性係数と級内相関係数(繁桝・柳井・森 Q&Aで知る統計データ解析 Dos and Don'ts(pp.224-226) サイエンス社)

清水裕士・大坊郁夫 2005 恋愛関係における関係性認知が精神的健康に及ぼす影響 対人社会心理学研究, 5, 59-65.

註

1)本論文執筆にあたり、大阪大学大学院人間科学研究科行動データ科学分野の兼清道雄氏に、多大な助言をいただきました。感謝申し上げます。

2)変量効果とは、研究によって水準数が一定でなく、ランダムである効果を指す。変量効果は固定効果のように定数(平均)で得られるのではなく、平均 0 の正規分布が仮定された確率変数として推定する。なお、固定効果とは研究によって水準数が一定のもので、性別や実験

条件などを指し、主に研究で興味がある要因の効果である。変量効果はそれに対して積極的にモデルに組み込まれないものが多く、主に検出力の向上のために導入される。

変量効果の最も代表的な例は被験者内要因における被験者効果である。実験参加者は母集団からランダムに抽出されたものであるため、その効果は確率変数として扱われる。確率変数として扱うことのメリットは、分析結果を母集団に一般化することができることが挙げられる。

HLM では研究のモデルに組み込まれる要因について平均だけでなく分散も考慮に入れているため確率変数として変量効果を導入する必要がある。つまり HLM の特徴は固定効果に指定する独立変数に、同時に変量効果も指定するところにある。

3)HLM6 は以下の HP でダウンロードすることができる。

http://www.ssicentral.com/

ただし、スチューデント版は独立変数が 4 つまでという制約がある。

4)ddfm オプションは各固定効果の検定に用いられる自由度のタイプを指定するものである。bw は被験者間と被験者内を区別して自由度を指定する。このほかに、誤差共分散の構造を考慮する kenwardroger などがある。

5)被験者効果を変量効果として考慮した級内相関を算出する場合、Mixed プロシージャで独立変数を指定しない（変量効果も指定しない）混合モデルを実行することで得られる、切片の分散と誤差の分散を用いて以下の式で得ることができる(ただし、この場合も SUB オプションはペア・集団を識別する変数を指定すること)。この級内相関係数は切片の分散が有意である場合、$\rho=0$ の帰無仮説を棄却する。

級内相関係数= 切片の分散/切片の分散+誤差の分散

The hierarchical analysis on pair or group data.

Hiroshi SHIMIZU (*Graduate School of Human Sciences, Osaka University*)

The purpose of this article is to show the effective methodologies for analyzing data which have hierarchical structures such as pairs and groups, and compare advantages of those analyses. It is difficult to perform sufficient analysis to pair or group data by using conventional analysis, despite plenty opportunities to obtain such data in the field of social psychology. HLM and pair-wise correlation analysis not only avoid errors made the conventional analysis but also become possible to acquire the important information about dynamics of interpersonal relationships or groups. This article explained examples of application and interpretative method of HLM, using a MIXED procedure of SAS. Also comparison and limitation of each hierarchical analysis were described.

Keywords: Pair data, hierarchical data, Hierarchical linear model, Mixed model, pair-wise correlation analysis