補表6　「知性×積極」からの変化と子どもの性格（多項ロジスティック回帰）

N=7,368

	知性×調整に変化				感性×積極に変化				感性×調整に変化			
	B	有意確率		Exp (B)	B	有意確率		Exp (B)	B	有意確率		Exp (B)
切片	-.760	.000	***		-.925	.000	***		-1.239	.000	***	
おとなしい	-.046 0[b]	.637		.955	-.055 0[b]	.570		.947	-.033 0[b]	.770		.967
活発	.094 0[b]	.191		1.099	-.063 0[b]	.392		.939	.179 0[b]	.037	*	1.195
誰にでも愛想がよい	.059 0[b]	.426		1.061	-.040 0[b]	.601		.961	.093 0[b]	.287		1.097
人見知りが激しい	-.360 0[b]	.008	**	.698	.019 0[b]	.884		1.019	.060 0[b]	.683		1.061
お調子者	.055 0[b]	.453		1.056	.005 0[b]	.944		1.005	-.017 0[b]	.843		.983
気が短い	.122 0[b]	.169		1.130	-.036 0[b]	.712		.965	-.109 0[b]	.324		.896
何事にも慎重	-.051 0[b]	.517		.950	-.037 0[b]	.637		.964	-.177 0[b]	.062	+	.838
何事にもマイペース	-.190 0[b]	.006	**	.827	-.051 0[b]	.464		.951	-.059 0[b]	.465		.943
我（気）が強い	.134 0[b]	.079		1.143	.043 0[b]	.592		1.044	.059 0[b]	.524		1.060
気が弱い	-.117 0[b]	.245		.890	.057 0[b]	.566		1.059	.110 0[b]	.330		1.116
勝ち気、負けず嫌い	.116 0[b]	.104		1.123	.102 0[b]	.165		1.108	.033 0[b]	.703		1.033
素直	-.041 0[b]	.544		.960	.317 0[b]	.000	***	1.373	.061 0[b]	.439		1.063
1人でやりたがる	-.063 0[b]	.465		.939	-.018 0[b]	.838		.983	-.053 0[b]	.601		.948
執着心が強い	-.059 0[b]	.533		.943	-.041 0[b]	.675		.960	.030 0[b]	.789		1.030
好奇心が旺盛	-.069 0[b]	.323		.933	.324 0[b]	.000	***	1.383	-.108 0[b]	.197		.898
飽きっぽい	-.060 0[b]	.438		.942	-.163 0[b]	.050	*	.850	.104 0[b]	.251		1.109
落ち着きがない	.122 0[b]	.120		1.130	-.273 0[b]	.002	**	.761	-.087 0[b]	.362		.917
恥ずかしがり屋	.062 0[b]	.380		1.064	-.001 0[b]	.988		.999	-.022 0[b]	.792		.978
甘えん坊	.081 0[b]	.198		1.085	.152 0[b]	.018	*	1.164	.092 0[b]	.220		1.096
のんびり屋	.062 0[b]	.446		1.064	-.089 0[b]	.281		.915	.000 0[b]	.997		1.000
せっかち	-.154 0[b]	.185		.857	-.212 0[b]	.089	+	.809	-.235 0[b]	.107		.791
その他	.030 0[b]	.841		1.030	.007 0[b]	.962		1.007	-.062 0[b]	.731		.940
わからない	.067 0[b]	.957		1.069	1.331 0[b]	.147		3.784	.546 0[b]	.657		1.726
Cox と Snell	.025											
Nagelkerke	.027											
McFadden	.010											

- 「知性×調整」へ　－人見知りが激しい　何事にもマイペース
- 「感性×積極」へ　＋素直　好奇心が旺盛　甘えん坊　せっかち　－飽きっぽい　落ち着きがない
- 「感性×調整」へ　＋活発　－何事にも慎重

補表7　「感性×積極」からの変化と子どもの性格（多項ロジスティック回帰）

N=8,357

	知性×調整に変化				知性×積極に変化				感性×調整に変化			
	B	有意確率		Exp (B)	B	有意確率		Exp (B)	B	有意確率		Exp (B)
切片	-.438	.000	***		.219	.003	**		-.290	.001	***	
おとなしい	.018 0^{b}	.871		1.018	.078 0^{b}	.388		1.081	-.021 0^{b}	.837		.979
活発	-.017 0^{b}	.826		.983	-.122 0^{b}	.059	+	.885	.048 0^{b}	.509		1.049
誰にでも愛想がよい	.059 0^{b}	.457		1.061	-.031 0^{b}	.647		.970	-.041 0^{b}	.585		.960
人見知りが激しい	-.142 0^{b}	.298		.868	-.211 0^{b}	.064	+	.810	-.301 0^{b}	.025	*	.740
お調子者	-.014 0^{b}	.855		.986	.012 0^{b}	.855		1.012	-.084 0^{b}	.262		.919
気が短い	.090 0^{b}	.360		1.094	.113 0^{b}	.173		1.119	.144 0^{b}	.127		1.154
何事にも慎重	-.166 0^{b}	.049	*	.847	-.177 0^{b}	.011	*	.838	-.183 0^{b}	.020	*	.833
何事にもマイペース	.075 0^{b}	.311		1.078	-.025 0^{b}	.685		.975	.021 0^{b}	.761		1.022
我（気）が強い	.230 0^{b}	.006	**	1.258	-.011 0^{b}	.876		.989	-.083 0^{b}	.314		.921
気が弱い	-.040 0^{b}	.712		.961	.124 0^{b}	.151		1.132	-.105 0^{b}	.302		.901
勝ち気、負けず嫌い	-.020 0^{b}	.808		.981	.042 0^{b}	.527		1.043	.007 0^{b}	.924		1.007
素直	-.204 0^{b}	.004	**	.816	-.239 0^{b}	.000	***	.788	-.058 0^{b}	.386		.944
1人でやりたがる	.212 0^{b}	.025	*	1.236	.021 0^{b}	.799		1.021	.119 0^{b}	.186		1.126
執着心が強い	.013 0^{b}	.898		1.014	.128 0^{b}	.136		1.137	.001 0^{b}	.991		1.001
好奇心が旺盛	-.449 0^{b}	.000	***	.638	-.186 0^{b}	.003	**	.830	-.264 0^{b}	.000	***	.768
飽きっぽい	.003 0^{b}	.973		1.003	.004 0^{b}	.951		1.004	-.074 0^{b}	.358		.929
落ち着きがない	.186 0^{b}	.033	*	1.205	.146 0^{b}	.047	*	1.157	.136 0^{b}	.104		1.146
恥ずかしがり屋	-.051 0^{b}	.516		.951	-.063 0^{b}	.329		.939	.076 0^{b}	.293		1.079
甘えん坊	-.107 0^{b}	.121		.899	-.095 0^{b}	.097	+	.909	-.062 0^{b}	.340		.940
のんびり屋	.007 0^{b}	.939		1.007	.089 0^{b}	.218		1.093	.087 0^{b}	.287		1.090
せっかち	.065 0^{b}	.616		1.067	.214 0^{b}	.045	*	1.238	.031 0^{b}	.807		1.031
その他	-.262 0^{b}	.115		.769	-.027 0^{b}	.836		.974	-.169 0^{b}	.262		.844
わからない	-19.672 0^{b}			.000	-19.702 0^{b}	.998		.000	-.809 0^{b}	.485		.445
Cox と Snell	.020											
Nagelkerke	.021											
McFadden	.007											

- 「知性×調整」へ　＋我（気）が強い　1 人でやりたがる　落ち着きがない　－何事にも慎重　素直　好奇心が旺盛
- 「知性×積極」へ　＋落ち着きがない　せっかち　－活発　人見知りが激しい　何事にも慎重　素直　好奇心が旺盛　甘えん坊
- 「感性×調整」　＋好奇心が旺盛　－人見知りが激しい　何事にも慎重

補表8　「感性×調整」からの変化と子どもの性格（多項ロジスティック回帰）

N=8,468

	知性×調整に変化				知性×積極に変化				感性×積極に変化			
	B	有意確率		Exp (B)	B	有意確率		Exp (B)	B	有意確率		Exp (B)
切片	-.237	.002	**		-.310	.000	***		-.550	.000	***	
おとなしい	-.066 0[b]	.489		.936	.135 0[b]	.158		1.144	.100 0[b]	.321		1.105
活発	.007 0[b]	.917		1.007	-.040 0[b]	.565		.961	.039 0[b]	.592		1.040
誰にでも愛想がよい	.001 0[b]	.985		1.001	-.068 0[b]	.340		.934	-.116 0[b]	.121		.890
人見知りが激しい	-.156 0[b]	.191		.855	.003 0[b]	.979		1.003	-.199 0[b]	.131		.820
お調子者	-.042 0[b]	.533		.958	.039 0[b]	.583		1.039	-.123 0[b]	.102		.885
気が短い	.121 0[b]	.151		1.128	.100 0[b]	.256		1.105	-.023 0[b]	.816		.978
何事にも慎重	-.023 0[b]	.751		.977	-.085 0[b]	.268		.919	-.020 0[b]	.800		.980
何事にもマイペース	-.001 0[b]	.991		.999	.026 0[b]	.704		1.026	.046 0[b]	.514		1.047
我（気）が強い	-.010 0[b]	.890		.990	-.099 0[b]	.191		.906	-.160 0[b]	.048	*	.852
気が弱い	.032 0[b]	.731		1.032	.082 0[b]	.380		1.086	.020 0[b]	.844		1.020
勝ち気、負けず嫌い	.126 0[b]	.066	+	1.134	.079 0[b]	.265		1.083	-.010 0[b]	.892		.990
素直	-.085 0[b]	.172		.918	-.118 0[b]	.067	+	.889	.085 0[b]	.205		1.089
1人でやりたがる	-.003 0[b]	.975		.997	.008 0[b]	.928		1.008	.006 0[b]	.945		1.006
執着心が強い	-.028 0[b]	.763		.972	.041 0[b]	.664		1.042	.005 0[b]	.961		1.005
好奇心が旺盛	-.096 0[b]	.144		.908	.020 0[b]	.767		1.020	.143 0[b]	.044	*	1.153
飽きっぽい	-.083 0[b]	.248		.920	-.163 0[b]	.032	*	.850	-.073 0[b]	.359		.930
落ち着きがない	.276 0[b]	.000	***	1.318	.059 0[b]	.457		1.060	-.030 0[b]	.725		.971
恥ずかしがり屋	.015 0[b]	.822		1.015	-.078 0[b]	.257		.925	-.031 0[b]	.664		.969
甘えん坊	-.235 0[b]	.000	***	.790	-.212 0[b]	.001	***	.809	-.105 0[b]	.104		.901
のんびり屋	-.009 0[b]	.903		.991	.045 0[b]	.571		1.046	-.033 0[b]	.690		.968
せっかち	-.149 0[b]	.178		.861	-.003 0[b]	.977		.997	-.123 0[b]	.322		.884
その他	-.002 0[b]	.989		.998	.025 0[b]	.864		1.026	.192 0[b]	.191		1.212
わからない	-17.964 0[b]	.998		.000	.310 0[b]	.827		1.364	-17.926 0[b]			.000
Cox と Snell	.014											
Nagelkerke	.015											
McFadden	.005											

- 「知性×調整」へ　＋勝ち気・負けず嫌い　落ち着きがない　－甘えん坊
- 「知性×積極」へ　＋素直　飽きっぽい　甘えん坊
- 「感性×積極」へ　＋好奇心が旺盛　－我（気）が強い

（3）小括

この簡単な分析ではあまりはっきりしたことは明らかにならなかったが、子供の成長と共に「知性」志向なかでも「知性×積極」志向が強まるという全体的傾向に加えて、親の属性と子どもの性格の双方で、以前よりあった子ども観の分布がより傾向がはっきりする方向に動いている可能性は高いと言えるのではないだろうか。

文献

広田照幸(1999)『日本人のしつけは衰退したか：「教育する家族」のゆくえ』講談社.

本田和子(2007)『子どもが忌避される時代：なぜ子どもは生まれにくくなったのか』新曜社.

柏木恵子(2001)『子どもという価値：少子化時代の女性の心理』中央公論社.

小山静子(2002)『子どもたちの近代：学校教育と家庭教育』吉川弘文館.

元森絵里子(2008)「「出生児縦断調査」による子ども観の分析に向けて：「どのような子に育って欲しいか」の分類および規定要因分析」『パネル調査（縦断調査）に関する総合的分析システムの開発研究（平成 19 年度厚生労働科学研究費補助金統計情報総合研究事業報告書)』, pp.143-164.

元森絵里子(2009)「子ども観と育児方針：第 1 回～第 6 回「出生児縦断調査」の分析から」『パネル調査（縦断調査）に関する統合的高度統計分析システムの開発研究（平成 20 年度厚生労働科学研究費補助金統計情報総合研究事業報告書)』, pp.247-261.

元森絵里子(2010)「子ども観と育児方針２：第 1 回～第 6 回「出生児縦断調査」の分析より―」『パネル調査（縦断調査）に関する統合的高度統計分析システムの開発研究（平成 21 年度厚生労働科学研究費補助金統計情報総合研究事業報告書)』, pp.241-282.

元森絵里子(2011)「子ども観と教育方針：主として「第 7 回出生児縦断調査」の分析より――」『パネル調査（縦断調査）に関する統合的分析システムの応用研究』厚生労働科学研究費補助金統計情報総合研究事業平成 22 年度総括研究報告書, pp.133-165.

沢山美果子(1987)「〈童心〉主義子ども観の展開：都市中間層における教育家族の誕生」『保育幼児教育体系 5 保育の思想』労働旬報社.

沢山美果子(1990)「教育家族の成立」中内敏夫他『教育：誕生と終焉』藤原書店.

Ⅲ.　パネルデータ分析法ガイド

パネルデータ分析法ガイド

平成 24～25 年度厚生労働科学研究費補助金政策科学推進研究事業
「縦断および横断調査によるライフコース事象の経時変化分析と
施策への応用に関する研究」（研究代表者：金子隆一）編

目次

はじめに

このパネルデータ分析法ガイドは、厚生労働省大臣官房統計情報部で実施されている３つの縦断調査の分析を行う上で有用と考えられる、パネルデータに関するいくつかの分析手法の理論とその応用例を示したものである。

第1章

パネルデータの利点と課題

パネル調査 (縦断調査) では、同一調査対象を継続的に調査し、その実態や意識の変化を時系列で捉えることによって、対象に生ずる事象のタイミングや因果関係に対する強力な推論が行える。しかし、その有効性を十分に引き出すためには横断調査とは異なる統計手法が必要となる。『パネルデータ分析ガイド』は、そうしたパネルデータ特有の分析手法を概説したもので、入門者から本格的な分析研究を目指す者までを対象に実践的なガイドとなることを目指している。わが国では従来パネルデータの蓄積が遅れていたが、近年に至って多くのパネル調査が創設され、分析への関心は高まっている。とりわけ 21 世紀縦断調査は国が行う初の公的パネル調査であり、国民生活の多様な側面を大規模な標本と経時的な調査で捕捉しようとする画期的なものである。本書ではこの 21 世紀縦断調査を中心的な題材としている。ここではまずパネルデータ分析法理解への第一歩として、パネルデータというものの利点と課題について整理をしておきたい。

1.1 調査法と分析デザイン

調査法の種別は大きく分けて、横断調査 cross-sectional survey と、縦断調査 longitudinal survey に分けられる。横断調査は 1 時点における多数の客体に対する調査である。一方、縦断調査は同一または比較可能な客体について、経時的比較を目的に、複数時点で繰り返し実施される調査である (Menard 1991)。縦断調査のうち同一の客体に対して実施する調査はパネル調査と呼ばれる*[1]。調査法分類の一例を図 1 に示した。

横断調査においても、対象の過去の履歴を調べ、これを時系列データと見なしてパネル調査同様に経時分析を行うことが出来る。これは回顧的パネル設計と呼ばれる。しかし、対象者の記憶に頼るため、遠い過去ほど不正確となるなど時間に依存した誤差が生じやすい点や、過去の意識、意欲といった心理的項目を捉えることが困難な点で、真のパネル調査には及ばない。

繰り返し横断調査とは、同一母集団の変化を捉えるために、異なる時点において異なる標本を抽出して実施するタイプの調査である*[2]。官庁などが同一テーマについて定期的に実施する調査の多

*[1] 継続的に保持される対象者一覧表をパネルと呼ぶことからこのように呼ばれる。社会科学における実地調査の開発・発展に寄与した社会学者 Paul H. Lazarsfeld (1901-76) の命名とされる。

*[2] 繰り返し横断調査はその経時性に着目して縦断調査の一形態として分類されることもある (Menard 1991)。

図1　調査の体系

- **横断調査**　cross-sectional survey
  - **回顧的パネル設計**　retrospective panel design
  - **繰り返し横断調査**　repeated cross-sectional survey
- **縦断調査**　longitudinal survey
  - **時系列調査**　time series survey
  - **パネル調査**　panel survey
    - **フォローアップ調査**　follow-up survey
    - **コーホート調査**　cohort survey

くはこれに属す。

一方の縦断調査の中で、単一の対象を経時的に捉えるケースは時系列調査に分類されるが、社会科学における統計調査では多数の対象を標本として分析することが普通なので、縦断調査の語を狭義に用いて、パネル調査と同義に用いることも提案されている（Baltes and Nesselroade 1979, Wall and Williams 1970 など）。ただしこれらの語の用法については専門家間でも必ずしも一致を見ない（Menard 1991）。本稿では、複数時点の変数の比較・分析に共通する手法やデザインを指して縦断と呼ぶことにし、縦断調査をパネル調査と同義に用いることとする。

パネル調査の中でも、1回の調査で捉えた標本について、後に追加的情報の取得を目的に行われる調査はフォローアップ調査と呼ばれる。また特定の事象を同時に経験した集団（コーホート）を定期的、継続的に調査する場合はコーホート調査と呼ばれている。ちなみに、厚生労働省の行う21世紀縦断調査はパネル調査の中のコーホート調査に相当し、出生児調査は出生コーホート調査 birth-cohort survey、成年者調査と中高年者調査は、年齢コーホート調査 age-cohort survey ということになる[*3]。

パネル調査データの統計分析に際しては、上述の特徴を反映して横断調査で用いられる統計手法（回帰分析に代表される多変量一般線形モデル）に時系列分析手法を複合して適用することが必要となる。実際、Frees (2004) は、縦断データ分析は時系列分析と回帰分析の結婚であると表現している。したがって、パネル調査は横断調査における母集団の代表性と時系列調査における経時性の両面を同時に備えた調査と言える。ただし、パネル調査では、調査回を重ねるごとに標本の一部脱落が繰り返され、しだいに標本の代表が損なわれて行くという性質がある。この脱落への対処こそがパネルデータの分析法の最大の課題と言える。これについては後述する。

[*3] 21世紀出生児縦断調査は2001年1月10日から17日の間及び7月10日から17日の間に出生した子, 21世紀成年者縦断調査は2002年月末時点で20~34歳であった全国の男女及びその配偶者, 中高年者縦断調査は2005年10月末現在で50~59歳である全国の男女をそれぞれ母集団としている。

1.2 パネル調査の利点と因果分析

パネル調査の主要な利点として個々の対象（本稿では個人と呼ぶことにする）に起こる変化を経時的に追うことで、この変化の原因に関する統計的な推論ができることが挙げられる。この要因間の因果関係の特定は、一般に科学的研究の近接的目標であり、これをもとにして事象のモデル化や科学的理論の構築がなされ、ひいては科学的予測 scientific prediction を行うことが可能となる。因果関係の特定は、厳密には科学実験によってのみ可能である。しかし、容易に実験の行えない社会科学の分野では、これに準ずる因果特定の方途を与えるパネル調査とその分析法は重要な位置づけを持つ。とりわけ政策的観点からは、有効な施策の立案・実施は因果モデルによってのみ実現できるものであり、これを目指すための統計調査はパネル型が基礎になるといっても過言ではない。このようにパネル調査は社会科学的な実証分析や科学的根拠に基づいた政策形成において中心的な役割を担うものである。

つぎに因果関係の特定法について簡単に考えよう。一般に一つの変数Xが他の変数Yの変化（変異）の原因であるためには、次の三つの条件を満たす必要がある。(1) XとYに相関の存在すること（関連性）、(2) XがYに時間的に先行すること（先行性）、(3) 相関が見かけ上の関係 spurious relationship ではないこと（竹内 1989, Menard 1991）[*4]。見かけ上の関係とは、第3の変数（潜在的独立変数）の因果的介在による相関関係のことであり、要件 (3) はXとYが他の変数を介さない直接の関係を持っていること意味する。横断調査では、(1)（相関性）を見い出すことはできる。しかし、 (2)（先行性）は一般に正確に捉えることは難しい。回顧的 retrospective に記述された変数を用いて先行関係を特定することもできるが、短期的な記憶等に依存する事柄については不正確であり、科学的分析としては不十分となるケースが多い。このように因果関係の要件に時間的要素があることから、純粋な横断調査ではその科学的特定が困難であり、縦断的デザインが必要となる。また、その場合に横断調査とは異なった因果モデル（因果関係を前提または想定したモデル）をベースとした統計分析手法が用いられる。では、パネル調査データが横断調査データに比べて因果分析に強いということは、統計モデルから見るとどのように説明できるのであろうか。

1.3 変数変化のモデル

統計的分析の対象として、パネル調査データが横断調査データと最も異なる点は、前者では同一対象を繰り返し調べることによって、関心のある変数の「変化」を明示的に分析の対象とすることができる点であろう。すなわち、変化をモデル上の一つの変数として扱うことができる。

まず、横断調査において二つの変数（X、Y）の因果関係をモデル化する場合を考えよう。線形回帰モデルによって、X の値が Y の値に対して影響を与えていることを表現すれば、以下のようになる。

[*4] これに加えて、異なる対象や時間にわたる普遍性を意味する (4) 関連の普遍性または一致性 (consistency of association)、理論的な整合性を意味する (5) 関連の整合性 (coherence of association) も要件とされることがある (竹内 1989)。

$$Y_{i,t} = \beta_{0,t} + \beta_{x,t} X_{i,t} + \varepsilon_{i,t} \tag{1.1}$$

ここで、$Y_{i,t}$、$X_{i,t}$ は、時刻 t における個人 i の変数値であり、$\beta_{0,t}$,$\beta_{x,t}$ は切片および回帰係数、また $\varepsilon_{i,t}$ は $X_{i,t}$ と独立に分布する誤差項である。

しかし、横断調査データにおいて、Y の値が X の値にともなって変化していたとしても、それは必ずしも真の「変化」ではなく、時間 t における個人間の「差異」を変化と見なしていることになる。この差異の中には、X では説明できないもともと個人間に存在する違い（いわゆる個人差）が含まれている。すなわち、個人 i の変数 Y における個人差を f_i とすると、

$$Y_{i,t} = \beta_{0,t} + \beta_{x,t} X_{i,t} + f_i + \varepsilon'_{i,t} \tag{1.2}$$

となる（ここでは f_i は時間によらないとし、$\sum f_i = 0$ とする）[*5]。横断調査、すなわち 1 時点 t のみの観察においては、個人差 f_i は誤差項 $\varepsilon_{i,t}$ に含まれ区別することはできないので、もし X が個人差 f_i と相関を持つなら、モデル (1) による X の効果 β_x の推定値はバイアス（unobserved heterogeneity bias）を受けることになる[*6]。

ところが、これがもしパネル調査によるデータであり、同じ変数に対する調査が以前に（時間 $t-1$ とする）行われていたとすると、その 2 時点間の変化自体をモデル化することができる。すなわち、それぞれの調査時における式 1.2 を用いて、

$$Y_{i,t} - Y_{i,t-1} = (\beta_{0,t} - \beta_{0,t-1}) + (\beta_{x,t} - \beta_{x,t-1})(X_{i,t} - X_{i,t-1}) + (\varepsilon'_{i,t} - \varepsilon'_{i,t-1}).$$

ここで 2 時点間の各個人の Y の変化を、$\Delta Y_i = Y_{i,t} - Y_{i,t-1}$ などと表し、X の Y に対する効果 $\beta_{x,t}$ が、時間によらない（β_x）と考えると、

$$\Delta Y_i = \Delta\beta_0 + \beta_x \Delta X_i + \Delta\varepsilon'_i \tag{1.3}$$

と表され、β_x に対する正しい推定が期待出来る。すなわち、Y の分散のうち個人差に由来する部分を取り除き、変化を正しく評価することができる[*7]。 この β_x はモデル 1.1 に対する係数 β_x と同じものであり (ただし時間によらないと仮定)、式 1.3 の回帰推定によってモデル 1.1 が正しく推定できたことになる。

このことは個人差 f_i を何らかの個人属性に帰着させたり、あるいは部分的に個人属性によると考えても同じように扱うことができる。すなわち、式 1.2 における f_i の項が、個人属性 U による

[*5] 式 1.2 は個人 i の効果を切片に含め、$Y_{i,t} = \beta_{i,t} + \beta_{x,t} X_{i,t} + \varepsilon'_{i,t}$ と表すこともできる。

[*6] 実験などで行われるように X の値が個人に対して無作為に与えられるような場合には、f_i は X との独立性が正当化され β_x は不偏推定量となる。しかし、社会調査においては一般にこれが成り立つことは少ない。その場合には、β_x 不偏推定量を得るためには、X と相関を持つ f_i 自身か、あるいはこれを表現する観測変数すべて明示的にモデルに入れる必要がある。

[*7] モデル 1.3 は、unconditional change-score model、または method of first differences などと呼ばれている。パラメータの標準誤差、検定量等も通常の回帰推定と同様に正しく推定される。

効果 $\beta_u U_i$ に置き換えられるか、あるいは追加されるだけで、2時点間の差を取ると、それらは相殺消去され、結局式 1.3 に帰結する。つまり、時間変化がないか、あるいは変化の小さい個人属性 U はモデルに取り入れなくとも X の効果の推定には影響を与えない。このような統計モデルを固定効果モデル（fixed effects model）（解説は第 5 章を参照のこと）という。

横断的データに対するモデル 1.1 では、上述のように f_i を表現しうるような X と相関を持つ変数をすべて明示的にモデルに入れなければ β_x の推定値はバイアスを持つため、X の Y に対する因果的関係を統計的に正当化される形で把握することは諦めざるを得ない場合がほとんどである。この点について、縦断データでは、変数の「変化」を明示的に分析の対象とすることができることから、この問題（unobserved heterogeneity、または omitted variables の問題）を回避することができるのである（Frees 2004, Menard 1997 など）。

ここで取り上げたモデルは最も単純な形式のものであり、実際の分析では後の章で紹介されるように、より複雑なものを扱わなくてはならない。しかし、パネルデータの利点を活かすための機構についての基本的な考え方は同一であると考えてよい。

1.4 欠損値に対する統計的対処

統計調査、とりわけ回答者自身が記入する形式の調査では、回答がなされなかったり、不適切であったりして、データに欠損値が生ずることは避けられない。ところが一般の統計モデルや理論においては、変数値はすべて揃っていることが前提である。もし欠損が特定の値に偏っている場合には、これらモデルや理論の前提が整わないため結論を誤ってしまわないとも限らない。したがって欠損値の生じ方のパターンや偏りの程度を把握して、統計分析上の適切な対処をする必要がある。とりわけパネル調査においては、調査回を重ねるごとに標本には脱落が生じるため、もし脱落が分析対象の変数値に相関して生ずる場合には分析に深刻な影響を与えることになる。したがって、パネル調査分析においては、常に脱落について注意を払っておく必要がある。以下では欠損値に関する課題を簡単に見ておこう。

統計的な観点から欠損が問題となるのは、欠損に偏りが有る場合、すなわち、その変数あるいは他の変数の値に依存して欠損の生じ方（確率）が異なる場合である。逆に、ある変数の欠損値がその変数の値、または他のいかなる変数の値とも独立に生じている場合は、「完全にランダムな欠損 missing completely at random (MCAR)」と呼ばれ、この欠損を含む標本を除いたデータセットは、もとの標本からの無作為標本となることから、通常の統計手法がそのまま適用できることになる (Allison 2001 など)。また、2つの変数 X と Y を考えたとき、X をコントロールすると Y の欠損確率が Y に依存しない場合には、「ランダムな欠損 missing at random (MAR)」と呼ばれる。これは Y の欠損が X の値に依存していても、X を固定したときに Y の欠損が自身の値にランダムに生じている状況を表している。原則として、通常の多変量解析を行う際、MAR の条件が満たされているとき（したがって、MCAR も含まれる）、欠損値を除いた標本を通常の無作為標本と見なしてよい[*8]。しかし、逆に言うとそうでない場合には、欠損値の統計分析結果に対する影響は無視

[*8] この状況は ignorable と呼ばれる (Allison 2001)。

することが出来ない*9。その際には、欠損値の発生パターンに対する統計モデルを特定または想定することによって、一般の統計モデルによる分析法を修正する必要が生ずる。以下ではその対処として欠損値を扱う主な統計手法の種別を挙げておこう。

(1) **欠損値標本の削除 listwise deletion、complete-case analysis**
分析対象となる変数に欠損値を含む標本をすべて分析対象から外す方法であり、一般に最も広く行われている方法となる。様々なタイプの欠損に対して意外に頑健 robust な方法であることが知られている。

(2) **欠損値変数の削除 pairwise deletion、avairable-case analysis**
欠損値を含む変数を分析対象から外す方法であり、具体的には対象とする変数の（平均）分散共分散行列を用いてパラメータの推定を行う。

(3) **ダミー変数法 dummy variable adjustment**
欠損値をカテゴリー変数における一つのカテゴリーと同様に扱い、ダミー変数を立てる方法である。

(4) **代入による方法 imputation**
欠損値に何らかの統計的方法による推定値を代入する方法の総称。具体的な推定方法により様々な方法が考えられる。平均値や多変量回帰モデルの予測値などの単一の値を代入する方法は単一値代入法（single imputation）と呼ばれる。

(5) **最尤推定法 maximum likelihood method**
統計モデルのパラメータの最尤推定の際に、欠損値の発生確率をもとにした尤度を組み込み、欠損値の発生を考慮した推定を行う方法。反復法 iteration method や EM 法 expectation-maximization algorithm などの有効な方法が知られている。

上記の他にも、欠損値の予測そのものが目的ではないものの、多重代入法 (multiple imputation) といった方法も知られている。多重代入法では、欠損値にランダムな誤差をもつ値を代入したデータを複数作成し、それらのデータを用いて目的となる統計的分析を行い、最後に複数の分析結果を統合することで、欠損値によるバイアスのないパラメーター（回帰係数）の推定を行うことを目的とする。

1.5 おわりに

パネル調査 (縦断調査) は、実験の困難な人間相手の科学、すなわち医科学や社会科学において、科学的分析の根幹である因果関係の特定に有効な調査デザインであるが、特有の手法の適用を以てはじめてその真価を発揮すると考えられる。『パネルデータ分析ガイド』はまさにそうしたパネルデータ特有の分析手法について実例を付して紹介したものである。もちろん、取り上げていない手法も数多くあるが、パネル調査分析手法一般の基礎について理解を得るように構成されている。本書が 21 世紀縦断調査とともに、わが国のパネル調査の発展と利活用に寄与することを期待する。

*9 この状況は nonignorable missing と呼ばれる。

参考文献

Allison, Paul D. (2001) *Missing Data*, Sage University Paper series on Quantitative Applications in the Social Sciences, 07-136. Sage, Newbury Park, CA.

Frees, Edward W. (2004) Longitudinal and Panel Data: Analysis and Applications in the Social Sciences, Cambridge Univ. Press.

Menard, Scott W. (1991) *Longitudinal Research*, Sage University Paper series on Quantitative Applications in the Social Sciences, 07-076. Sage, Newbury Park, CA.

竹内 啓 編 (1989)「18 因果分析法」『統計学事典』pp.501、東洋経済新報社.

第2章

生存時間分析

2.1 生存時間分析の基本量

パネルデータでは経時観察がされることから、特定のイベント発生までの長さを分析することに利用できるが、このような際に用いられるのが生存時間分析 (survival analysis) である。ここでは生存時間分析に使われる基本的な関数などについて簡単に説明する。[*1]

X をあるイベントが起きるまでの時間を表す確率変数であるとする。生存時間分析では、イベントの生起を死亡にみなして、イベントの起きるまでの時間を生存時間と呼ぶ。このとき、対象がある時間 x を越えて生存する確率は、

$$S(x) = Pr(X > x)$$

で定義されるが、これを生存関数という。生存関数は累積分布関数 $F(x) = Pr(X \leq x)$ と $S(x) = 1 - F(x)$ との関係にある。x が連続で確率密度関数 $f(x)$ が存在するとすれば、

$$S(x) = Pr(X > x) = \int_x^\infty f(t)dt$$

となるので、

$$f(x) = -\frac{dS(x)}{dx}$$

が成立する。さらに、

$$h(x) = \lim_{\Delta x \to 0} \frac{Pr(x \leq X < x + \Delta x | X \geq x)}{\Delta x}$$

をハザード関数と呼ぶ。連続な場合には、

$$h(x) = \frac{f(x)}{S(x)} = -\frac{d}{dx} \log S(x)$$

が成立する。

なお、生命表では、$S(x)$ を l_x、$f(x)$ を d_x、$h(x)$ を μ_x と表している。

[*1] 本章の内容については、Klein and Moeschberger (2009)、中澤 (2007) を参考にしている。

2.2 カプラン・マイヤー推定量

Freinreich et al による白血病治療データ (Gehan データと呼ぶ) は、急性白血病にかかり、寛解*2の状態にある 42 人の子供について、プラセボを投与する群（対照群）と抗がん剤の 6-MP（6-メルカプトプリン）を投与した群（処置群）に分けて、再発までの時間（月数）を観測したデータである。データでは、各対象者の id（pair)、再発あるいは観察打ち切り（censoring）までの時間 (time)、その時間がイベント（再発）の生起か打ち切りかの別（cens)、そして対照群か処置群かの別（treat）が示されている。まず、データを見てみよう。

Gehan データの読み込みと表示

```
library(survival)
library(MASS)
data(gehan)
gehan
```

出力結果

```
   pair time cens   treat
1     1    1    1 control
2     1   10    1    6-MP
3     2   22    1 control
4     2    7    1    6-MP
5     3    3    1 control
6     3   32    0    6-MP
7     4   12    1 control
8     4   23    1    6-MP
9     5    8    1 control
10    5   22    1    6-MP
11    6   17    1 control
12    6    6    1    6-MP
13    7    2    1 control
14    7   16    1    6-MP
15    8   11    1 control
16    8   34    0    6-MP
17    9    8    1 control
18    9   32    0    6-MP
19   10   12    1 control
20   10   25    0    6-MP
21   11    2    1 control
22   11   11    0    6-MP
23   12    5    1 control
24   12   20    0    6-MP
25   13    4    1 control
26   13   19    0    6-MP
27   14   15    1 control
28   14    6    1    6-MP
29   15    8    1 control
30   15   17    0    6-MP
31   16   23    1 control
32   16   35    0    6-MP
33   17    5    1 control
34   17    6    1    6-MP
35   18   11    1 control
36   18   13    1    6-MP
37   19    4    1 control
38   19    9    0    6-MP
39   20    1    1 control
40   20    6    0    6-MP
41   21    8    1 control
42   21   10    0    6-MP
```

*2 症状が一時的に軽くなったり、消えたりした状態のこと

このうち、6-MP を投与したグループについて、時間でソートしたデータを表示してみる。

コードと出力結果

```
gehanT <- subset(gehan, treat == "6-MP")
gehanT[order(gehanT$time),]


   pair time cens treat
12    6    6    1  6-MP
28   14    6    1  6-MP
34   17    6    1  6-MP
40   20    6    0  6-MP
4     2    7    1  6-MP
38   19    9    0  6-MP
2     1   10    1  6-MP
42   21   10    0  6-MP
22   11   11    0  6-MP
36   18   13    1  6-MP
14    7   16    1  6-MP
30   15   17    0  6-MP
26   13   19    0  6-MP
24   12   20    0  6-MP
10    5   22    1  6-MP
8     4   23    1  6-MP
20   10   25    0  6-MP
6     3   32    0  6-MP
18    9   32    0  6-MP
16    8   34    0  6-MP
32   16   35    0  6-MP
```

ここで、イベントが起きた時刻 t_i において、イベントを体験する可能性のある個体数を Y_i、発生したイベントの数を d_i、打ち切りの数を c_i とする。このとき、以下のような表が作成できる。

t_i	Y_i	d_i	c_i
0	21		
6	21	3	1
7	17	1	0
9	16	0	1
10	15	1	1
11	13	0	1
13	12	1	0
16	11	1	0
17	10	0	1
19	9	0	1
20	8	0	1
22	7	1	0
23	6	1	0

そこで、d_i のある時刻だけに改めてインデックス i を振り直し、以下のような生存関数の推定量を考える。

$$\hat{S}(t) = \begin{cases} 1 & (t < t_1) \\ \prod_{i(t_i<t)} \left(1 - \frac{d_i}{Y_i}\right) & (t \geq t_1) \end{cases}$$

例えば、$0 \leq t \leq 6$ では、$\hat{S}(t) = 1$、$6 \leq t \leq 7$ では、$\hat{S}(t) = 1 - \frac{3}{21}$、$7 \leq t \leq 10$ では、$\hat{S}(t) = (1 - \frac{3}{21})(1 - \frac{1}{17})$ などとなる。これをカプラン・マイヤー推定量 (Kaplan-Meier estimator) と呼ぶ。