

身体が高い人は、実際にどちらの成長パターンをとることが多いのだろうかといったことが議論になっている。これらの問題に答えるためには東郷(1998)が主張しているように、同一個人を丹念に調査し記録を残すパネルデータを用いるしかない。東郷の場合、自分の5人の子供の身体計測を28歳ぐらいまで毎月調査したものであるが、『21世紀出生時縦断調査』では46000人を超える子供のデータを集めており、彼らが18歳を超えるまで、毎年必ず、体重・身長を計測し、その他の健康状態に関する調査も一定の期間をおいて繰り返し行えば、その標本サイズからして、人間の成長パターンに関する研究は画期的に進歩するものと考えられる。

現在のところ、最初の5年間の成長データしか利用できない。しかし、このデータは人生の最初にして最大の成長期を多角的に捉えることを可能にしてくれる情報源である。なによりも、この時期の成長が本人の自覚なしに進むということは示唆的である。思春期に入ると、ダイエットやスタイルを気にして自らの発育を抑えようとする内生的な行動をとることがあるが、この最初の成長期では与えられた影響の中でほぼ無意識に成長が進んでいく。人間の成長にとって、そのような時期が必要であるということであろう。

身体発育という観点からは、身体の水準だけでなく、変化(velocity)を見ることも重要であると言われている。そこで体重と身長の2観察時点の差(変化)を取ったものを図示したのが図5-6である。体重・身長の変化は最初の500日ぐらいが最大であとは減速していくことが見て取れる。推定した曲線では体重・身長とも1500日を超えたあたりで再び発育が上昇しているように見えるが、これは一時的な現象で、さらに長い時系列データが蓄積されれば、この傾向は消滅し、次の大きなスパートは思春期に入る11歳から18歳の時期に訪れるはずである。

実際に1日当たりに換算した成長率に関するデータを見てみると、体重で平均0.09%、身長で平均0.04%となっている。体重の成長率の上位1%は0.28%、身長の上位1%は0.11%であるが、最大では体重が1日に25%、身長が1日に15%も伸びる子供がいることになっている。現実的に考えてこれらの数字は誤記入の可能性が高い。また、記録の中にはマイナスの成長率を記録している子供もいる。身長に関しては、この年齢で身長が縮むことはほぼあり得ないと判断して削除してある。一方、体重は子供の体重が過重なので減量させるということはあるが、これにも誤記入があると思われる。

#### 4. パネルデータ推定

『21世紀出生児縦断調査』はパネルデータを構築しようとしているのだが、質問票の設計思想はむしろ、各回毎のクロスセクション調査の発想に縛られており、継続して同じ質問を繰り返し、その変化を分析するというパネルデータ調査の基本的な考え方が必ずしも反映されているとは言えない。ここで継続的に同一項目で調べている数値データは、体重・身長を除けば、親の所得や保育費、および測定日ぐらいであり、あとは質的データが中心

となっている。従って、ここで行ったパネルデータ推定も極めて限られた変数を用いて行わざるを得なかった。具体的には、体重(kg)の対数表示(lnbdywht)、身長(cm)の対数表示(lnbdyhgt)を誕生日からの経過日数(survivalday)、経過日数の2乗(survivalday Sq)、子供の保育料の対数表示(lnkosodate)で説明した。結果は表 9-14 に報告してある。

表 9 は体重の成長に関して男女をプールして推定した結果である。モデルとしては固定効果推定(Fixed)が選ばれており、モデルはかなり高い説明力を持っていることがわかる。誕生日からの経過日数は2次項が負で、1次項が正となっており、成長曲線は上に凸の関数であることが示されている。また保育料は正に有意に効いている。表 10 は同様に身長成長に関して推定したものであるが、結果は体重と同じで固定効果推定が選択されている。ただし、保育料はほぼゼロとなり有意でもなくなっている。このことは身長に関しては自立的成長が強く、外生的な要因が影響を与える余地が少ないことを意味しているのかもしれない。

以下では、男女別に体重と身長成長モデルを推定している。表 11 は男子の体重、表 12 は男子の身長を推定したものである。いずれも固定効果推定が選択されている。保育料の効果は体重では正に有意だが、身長では有意でなくなっている。表 13 は女子の体重、表 14 は女子の身長を推定したものである。ここでも体重に関しては固定効果推定が選ばれるが、身長に関しては全体や男子の場合とは違い、10%の有意水準でランダム効果推定が選択されている。保育料の係数はその有意性も体重の成長に関する方が強く出ていることは全体および男女ともに言えることであるが、女子については身長に関しても正に有意に効いている。この男女の保育料の身長への効果の違いについてはさらに詳細に調べる必要がある。

女子の身長を除いて固定効果推定が選択されるということは重要な発見である。ここで言う固定効果とは何だろうか。これは恐らく、産まれた時の体重・身長などの初期値の違いだけではなく、親から受け継いだ遺伝子情報、あるいはもっと直接的に親の体重・身長の情報、親の経済的状況などが考えられる。ここでは、初期値の高かった子供はどれぐらい長い間この利点を維持できるのだろうかということに関心があるが、現在のデータではそこまでの分析はできない。さらに、この初期値に遺伝子情報が含まれているとすれば、体重・身長だけではなく知能やその他の稼働能力も受け継いでいるはずである。それはどれぐらいその後の環境によって逆転できるものなのだろうか。ゴールトンは英国学術協会人種学部長講演で「進化の過程は2つの相反する行動から成り立っている。一方は収束する動きであり、他方は拡張する動きである。これらは相互に牽制して安定的均衡に達する」と述べている。確かに、ゴールトンが想定したように優良遺伝子を持った人間がその他の人間を席捲しているということではなく、同じ家系の中でも優秀な人間もそれほどでもない人間も生まれてくるというのが現実のように思われる。

ただ、戦後60年間で日本人の体格は大幅に進化したことは事実である。それは一種の進化が起こっているのか、あるいは遺伝子的には潜在力があつたのだが、生後の与えられる環境の貧弱さによって、その遺伝子情報が十分に利用発達できなかったのだが、戦後高度

成長期を通してその制約から解放されたと考えればいいのだろうか。『21世紀出生児縦断調査』はこれらの極めて重要な問題に答えを出すための貴重な資料となることが期待できる。

## 5. おわりに

2001年1月10日・17日と同年7月10日・17日に生まれた日本中の子供を対象にした「21世紀出生児縦断調査」は日本の厚生労働行政のみならず、教育・社会・経済政策にとって役に立つ貴重な資料である。この調査の意義は、(1)個人のリアルタイムの成長の軌道が追えること、(2)パネルデータとして潜在的に多様な利用方法が考えられる。さらに親の情報や祖父母の情報も入れれば超長期の世代間問題にまで分析を及ぼすことができる。(3)これまで、遺伝的形質なのか、後天的形質なのか区別が付きにくかった現象をある程度、識別できる可能性があること、などであろう。

少なくとも体重・身長などの調査はその後の教育履歴のデータとともに継続して調査し、20歳-23歳ぐらいまでの人的資本形成に関わるパネルデータを集めることが望まれる。現在は母親を中心とした保護者がこの調査に答えてくれているが、子供が成長するにつれて、いつの時点かで親から本人が調査に答えることになることが想定される。その時点で、大幅な脱落者が出てくることが予想される。その時、どのような子供が脱落するかは予断を許さないが、特定の属性をもったサンプルが脱落する傾向が強ければ、サンプルに脱落バイアスが出てくる可能性がある。この調査の統計的な代表性を維持したければ、調査当局はかなりの熱意をもって対象者に対して調査への参加の意義を説明し、調査結果の定期的なフィードバックを行う必要がある。それが成功すれば、21世紀初年度に生まれた日本人の代表性を保有した極めて貴重なデータになることは間違いない。

パネルデータは調査を継続することが大前提となる。そのための熱意を維持する努力と、新しいアイデアを持った研究者がこの宝の山のようなデータを利用して様々な興味深い研究成果を生み出すことによって、この調査への関心が高まり、またその価値が認められるという好循環を生み出す努力を怠らないことが望まれる。

## 参考文献

- [1] 北村行伸(2007)「21世紀出生児縦断調査に基づく子供の成長パターンの測定」『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18年度報告書』(厚生労働科学研究費補助金統計情報高度利用総合研究事業：主任研究者金子隆一：課題番号h18-統計-002)平成19年3月, pp.101-123.
- [2] 厚生労働省雇用均等・児童家庭局(2001)『平成12年乳幼児身体発育調査報告書』、厚生労働省
- [3] 産経新聞(2006)「早生まれ損?得?」、産経新聞、2006年12月4日朝刊
- [4] 鈴木隆雄(1996)『日本人のからだ 健康・身体データ集』、朝倉書店

- [5] 東郷正美 (1998) 『身体計測による発育学』、東京大学出版会
- [6] 福井幸男 (1997) 『知の統計学 2』、共立出版
- [7] 増山元三郎(1994) 『成長の個体差』、みすず書房
- [8] Behrman, Jere and Rosenzweig, Mark R. (2001) “The Returns to Increasing Body Weight”, University of Pennsylvania, Penn Institute for Economic Research Working Paper 01-052.
- [9] Boyd, E.(1980) Origin of the Study of Human Growth, University of Oregon Health Science Center Foundations.
- [10] Case, Anne and Paxson, Christina.(2006) “Stature and Status: Height, Ability, and Labor Market Outcomes”, Center for Health and Wellbeing, Princeton University, mimeo.
- [11] Dubner, Stephen, J. and Levitt, Steven D.(2006) “A Stat Is Made”, The New York Times Magazine, Freakonomics, May 7, 2006.
- [12] Duffy, Linda J., Baluch, Bahman. and Ericsson, Andres. (2004) “Dart Performance as a Function of Facets of Practice Among Professional and Amateur Men and Women Players”, International Journal of Sport Psychology, 35, pp.232-245.
- [13] Ericsson, K.Andres., Krampe, Ralf Th., and Tesch-Romer, Clements. (1993) “The Role of Deliberate Practice in the Acquisition of Expert Performance”, Psychological Review, 100(3), pp.363-406.
- [14] Eveleth, Phyllis B. and Tanner, James M.(1990) Worldwide Variation in Human Growth, 2nd ed, Cambridge: Cambridge University Press.
- [15] Hall, Stephen S.(2006a) Size Matters, Houghton Mifflin Company.
- [16] Hall, Stephen S.(2006b) “Essay: Success is Relative, and Height isn’t Everything”, New York Times, November 28, 2006.
- [17] Helsen, Werner F, Winckel, Jan Van. and Williams, A. Mark. (2005) “The Relative Age Effect in Youth Soccer Across Europe”, Journal of Sports Sciences, 23(6), pp.629-636.
- [18] Musch, Jochen and Grondin, Simon. (2001) “Unequal Competition as an Impediment to Personal Development: A Review of the Relative Age Effect in Sport”, Development Review, 21, pp.147-167.
- [19] Presico, Nicola, Postlewaite, Andrew, and Silverman, Dan. (2004) “The Effect of Adolescent Experience on Labor Market Outcomes: The Case of Height”, Journal of Political Economy, 112(5), pp.1019-1053.
- [20] Steckel, Richard H.(1995) “Stature and the Standard of Living”, Journal of Economic Literature, 33(4), pp.1903-1940.
- [21] Steckel, Ricahrd H. and Prince, Joseph M.(2001) “Tallest in the World: Native Americans of the Great Plains in the Nineteenth Century”, American Economic Review, 91(1), pp.287-294.
- [22] Stigler, Stephen M.(1986) The History of Statistics, Harvard University Press.
- [23] Tanner, James M.(1981) A History of the Study of Human Growth, Cambridge University Press.

表1 21世紀出生児縦断調査 調査回数別・男女別体重・身長統計

男子

	第1回		第2回		第3回		第4回		第5回	
	体重	身長	体重	身長	体重	身長	体重	身長	体重	身長
平均	3.07	49.19	10.86	81.03	13.10	89.88	14.91	97.03	16.77	103.86
標準偏差	0.44	2.34	1.14	3.31	1.43	3.85	1.70	3.99	2.04	4.42
最小値	0.50	30.00	5.10	53.00	6.30	57.00	5.90	73.50	7.60	80.00
1%	1.70	42.00	8.40	73.00	10.00	80.00	11.40	88.00	12.80	94.00
中位(50%)	3.10	49.00	10.80	81.00	13.00	90.00	14.90	97.00	16.50	103.80
99%	4.00	54.00	13.90	90.00	16.90	100.00	19.50	108.00	23.00	115.00
最大値	5.50	60.00	20.00	110.20	20.00	110.00	40.00	130.00	37.20	132.00

女子

	第1回		第2回		第3回		第4回		第5回	
	体重	身長	体重	身長	体重	身長	体重	身長	体重	身長
平均	3.00	48.71	10.26	79.73	12.56	88.69	14.47	96.07	16.42	103.05
標準偏差	0.42	2.27	1.07	3.23	1.39	3.76	1.67	3.99	2.02	4.41
最小値	0.50	28.00	5.40	45.00	6.00	68.00	8.10	73.60	7.00	80.50
1%	1.80	42.00	8.00	72.00	9.60	80.00	11.00	87.00	12.50	93.50
中位(50%)	3.00	49.00	10.20	79.90	12.50	89.00	14.40	96.80	16.20	103.00
99%	4.00	53.00	13.00	89.00	16.00	98.00	19.00	106.55	22.30	115.00
最大値	5.50	58.00	20.60	110.00	20.00	105.00	28.00	140.00	32.00	132.00

表2 測定期間別標本分布

年・月・日齢	体 重		身 長	
	男子	女子	男子	女子
0年0-6月末まで	2	-	2	-
0年6-12月末まで	31	30	31	29
1年0-6月末まで	1722	1527	1686	1503
1年6-12月末まで	18896	17618	18435	17210
2年0-6月末まで	3164	2851	2967	2653
2年6-12月末まで	16594	15490	14997	13937
3年0-6月末まで	3514	3274	3498	3243
3年6-12月末まで	18866	17408	18269	16847
4年0-6月末まで	13703	12684	13451	12457
4年6-12月末まで	217	199	204	189

表3 平成12年(2000年)乳幼児身体発育調査による体重の分布

年・月・日齢	男子							女子						
	パーセンタイル値							パーセンタイル値						
	3	10	25	50 中央値	75	90	97	3	10	25	50 中央値	75	90	97
0年0-6月末まで	6.29	6.75	7.22	7.79	8.41	8.98	9.55	5.90	6.30	6.72	7.18	7.74	8.25	8.80
0年6-12月末まで	7.73	8.21	8.72	9.33	9.97	10.57	11.18	7.14	7.59	8.12	8.67	9.28	9.85	10.45
1年0-6月末まで	8.63	9.16	9.70	10.41	11.11	11.83	12.65	8.14	8.65	9.16	9.84	10.51	11.19	11.95
1年6-12月末まで	9.44	10.03	10.64	11.43	12.23	13.05	13.92	9.03	9.60	10.17	10.95	11.72	12.51	13.33
2年0-6月末まで	9.97	10.59	11.26	12.07	12.91	13.81	14.74	9.45	10.07	10.77	11.53	12.38	13.26	14.17
2年6-12月末まで	10.80	11.49	12.18	13.01	13.92	14.97	16.04	10.22	10.95	11.68	12.51	13.46	14.51	15.57
3年0-6月末まで	11.59	12.28	13.06	13.97	14.99	16.14	17.36	11.03	11.78	12.58	13.49	14.54	15.72	16.92
3年6-12月末まで	12.34	13.09	13.93	14.92	16.05	17.33	18.71	11.80	12.62	13.49	14.49	15.65	16.97	18.33
4年0-6月末まで	13.10	13.90	14.82	15.90	17.16	18.60	20.17	12.57	13.46	14.41	15.50	16.79	18.27	19.84
4年6-12月末まで	13.86	14.72	15.72	16.91	18.30	19.93	21.71	13.33	14.29	15.32	16.52	17.96	19.62	21.37
5年0-6月末まで	14.63	15.56	16.65	17.96	19.52	21.38	23.40	14.07	15.10	16.23	17.55	19.31	21.09	23.29
5年6-12月末まで	15.27	16.32	17.48	18.93	20.70	22.85	25.50	14.81	15.93	17.16	18.62	20.66	22.84	25.39

表4 21世紀出生児縦断調査による体重の分布

年・月・日齢	男子							女子							
	パーセンタイル値							パーセンタイル値							
	3	10	25	50 中央値	75	90	97	3	10	25	50 中央値	75	90	97	
0年0-6月末まで	-	-	-	-	-	-	-	0年0-6月末まで	-	-	-	-	-	-	-
0年6-12月末まで	7.20	8.00	8.70	9.80	10.40	10.60	12.50	0年6-12月末まで	6.00	7.60	7.90	8.80	9.10	9.60	11.30
1年0-6月末まで	8.50	9.30	9.90	10.50	11.40	12.10	13.00	1年0-6月末まで	8.20	8.80	9.30	10.00	10.70	11.40	12.40
1年6-12月末まで	9.00	9.50	10.00	10.80	11.60	12.40	13.10	1年6-12月末まで	8.50	9.00	9.50	10.20	11.00	11.60	12.50
2年0-6月末まで	10.30	11.00	12.00	13.00	14.00	15.00	15.60	2年0-6月末まで	10.00	10.50	11.50	12.30	13.30	14.00	15.00
2年6-12月末まで	10.70	11.50	12.00	13.00	14.00	15.00	16.00	2年6-12月末まで	10.00	11.00	11.70	12.50	13.50	14.50	15.30
3年0-6月末まで	12.00	12.70	13.50	14.50	15.60	16.70	18.00	3年0-6月末まで	11.50	12.30	13.00	14.00	15.00	16.30	17.70
3年6-12月末まで	12.30	13.00	14.00	15.00	16.40	17.70	19.00	3年6-12月末まで	12.00	12.80	13.60	14.80	16.00	17.20	18.80
4年0-6月末まで	13.60	14.50	15.50	16.60	18.00	19.40	21.00	4年0-6月末まで	13.20	14.00	15.00	16.20	17.60	19.00	20.60
4年6-12月末まで	13.50	14.00	15.00	16.50	18.00	19.40	21.00	4年6-12月末まで	13.40	14.00	15.40	16.90	18.00	20.00	22.00

表5 平成12年(2000年)乳幼児身体発育調査による身長分布

年・月・日齢	(cm)							年・月・日齢	(cm)						
	男子								女子						
	パーセンタイル値								パーセンタイル値						
	3	10	25	50 中央値	75	90	97		3	10	25	50 中央値	75	90	97
0年0-6月末まで	62.6	64.0	65.4	67.0	68.5	69.8	71.4	0年0-6月末まで	61.0	62.4	63.8	65.4	67.0	68.5	69.9
0年6-12月末まで	69.5	71.0	72.6	74.4	76.0	77.4	78.9	0年6-12月末まで	68.5	69.8	71.2	72.7	74.2	75.6	77.0
1年0-6月末まで	74.9	76.6	78.3	80.2	82.0	83.5	85.1	1年0-6月末まで	74.2	75.8	77.3	79.1	80.8	82.3	83.9
1年6-12月末まで	80.1	81.9	83.8	85.8	87.7	89.4	91.0	1年6-12月末まで	79.4	81.0	82.6	84.4	86.2	87.9	89.5
2年0-6月末まで	81.2	83.1	85.0	87.1	89.1	90.9	92.6	2年0-6月末まで	80.7	82.4	84.1	86.0	87.9	89.7	91.4
2年6-12月末まで	85.0	86.9	88.8	91.0	93.2	95.2	97.2	2年6-12月末まで	84.2	86.0	87.8	89.9	92.0	94.0	96.0
3年0-6月末まで	88.3	90.3	92.3	94.6	97.0	99.2	101.4	3年0-6月末まで	87.6	89.5	91.5	93.7	95.9	98.3	100.4
3年6-12月末まで	91.5	93.6	95.8	98.2	100.9	103.3	105.7	3年6-12月末まで	90.9	92.9	95.1	97.4	99.7	102.3	104.6
4年0-6月末まで	94.5	96.8	99.1	101.6	104.5	107.2	109.8	4年0-6月末まで	94.1	96.3	98.5	101.0	103.5	106.1	108.5
4年6-12月末まで	97.4	99.8	102.2	104.9	108.1	110.9	113.7	4年6-12月末まで	96.9	99.3	101.7	104.3	106.9	109.5	111.9
5年0-6月末まで	100.2	102.7	105.3	108.1	111.4	114.4	117.4	5年0-6月末まで	99.8	102.3	104.8	107.6	110.4	112.9	115.4
5年6-12月末まで	103.1	105.8	108.4	111.4	114.9	118.0	121.1	5年6-12月末まで	102.6	105.2	107.9	110.8	113.7	116.4	119.0
6年0-6月末まで	106.2	109.0	111.8	114.9	118.6	121.8	125.1	6年0-6月末まで	105.2	108.0	110.7	113.8	116.9	119.6	122.4

表6 21世紀出生児縦断調査による身長分布

年・月・日齢	(cm)							年・月・日齢	(cm)						
	男子								女子						
	パーセンタイル値								パーセンタイル値						
	3	10	25	50 中央値	75	90	97		3	10	25	50 中央値	75	90	97
0年0-6月末まで	-	-	-	-	-	-	-	0年0-6月末まで	-	-	-	-	-	-	-
0年6-12月末まで	65.80	67.70	72.00	74.20	79.00	81.10	84.40	0年6-12月末まで	60.70	67.60	69.20	70.80	74.20	79.50	81.00
1年0-6月末まで	73.70	76.00	78.00	80.00	82.00	84.00	86.20	1年0-6月末まで	72.20	74.80	76.90	78.80	80.60	82.70	85.00
1年6-12月末まで	75.20	77.40	79.20	81.00	83.00	85.00	87.00	1年6-12月末まで	74.30	76.00	78.00	80.00	81.60	83.50	85.50
2年0-6月末まで	81.50	84.00	86.20	89.00	91.10	94.50	97.00	2年0-6月末まで	80.60	83.00	85.00	87.50	90.00	92.90	95.00
2年6-12月末まで	83.50	85.50	87.90	90.00	92.50	95.00	98.00	2年6-12月末まで	82.00	84.60	86.50	89.00	91.00	94.00	96.00
3年0-6月末まで	89.70	91.60	93.70	96.00	98.50	100.80	104.00	3年0-6月末まで	88.00	90.50	92.80	95.00	97.50	100.00	102.60
3年6-12月末まで	90.50	93.00	95.30	98.00	101.00	105.00	108.50	3年6-12月末まで	90.00	92.00	94.70	97.20	100.00	103.80	107.80
4年0-6月末まで	96.40	98.80	101.00	104.00	107.00	110.00	112.40	4年0-6月末まで	95.60	98.00	100.30	103.00	106.00	108.90	111.30
4年6-12月末まで	96.00	98.00	101.00	105.00	108.50	110.00	112.30	4年6-12月末まで	96.00	99.10	101.40	105.00	108.30	110.00	113.00

表7 昭和35年、45年、55年、平成2年および12年の乳幼児身体発育調査による体重の分布(kg)

年・月・日齢	男 子					
	昭和35年	昭和45年	昭和55年	平成2年	平成12年	21世紀パネ ル調査
0年0-6月末まで	7.40	7.80	7.80	7.75	7.79	8.90
0年6-12月末まで	8.80	9.30	9.49	9.39	9.33	9.56
1年0-6月末まで	10.00	10.40	10.50	10.56	10.37	10.65
1年6-12月末まで	11.00	11.60	11.82	11.70	11.43	10.88
2年0-6月末まで	11.60	12.30	12.18	12.33	12.07	12.88
2年6-12月末まで	12.50	13.20	13.27	13.35	13.12	13.16
3年0-6月末まで	13.30	14.10	14.28	14.32	14.13	14.66
3年6-12月末まで	14.20	15.00	15.22	15.28	15.15	15.32
4年0-6月末まで	15.00	15.80	16.12	16.24	16.15	16.83
4年6-12月末まで	15.80	16.60	17.01	17.22	17.27	16.76
5年0-6月末まで	16.60	17.40	17.91	18.27	18.36	-
5年6-12月末まで	17.40	18.20	18.86	19.38	19.48	-
6年0-6月末まで	-	-	19.88	20.60	20.56	-

年・月・日齢	女 子					
	昭和35年	昭和45年	昭和55年	平成2年	平成12年	21世紀パネ ル調査
0年0-6月末まで	6.90	7.30	7.33	7.23	7.18	-
0年6-12月末まで	8.40	8.90	8.91	8.83	8.67	8.62
1年0-6月末まで	9.50	9.90	10.10	9.95	9.86	10.05
1年6-12月末まで	10.40	11.30	11.34	11.09	10.97	10.28
2年0-6月末まで	11.10	11.70	11.89	11.72	11.55	12.38
2年6-12月末まで	12.00	12.60	12.88	12.79	12.58	12.62
3年0-6月末まで	12.90	13.40	13.86	13.83	13.62	14.21
3年6-12月末まで	13.80	14.30	14.82	14.85	14.63	14.90
4年0-6月末まで	14.60	15.20	15.76	15.88	15.73	16.49
4年6-12月末まで	15.40	16.10	16.67	16.92	16.79	16.89
5年0-6月末まで	16.20	17.00	17.55	17.99	17.92	-
5年6-12月末まで	17.00	18.00	18.38	19.11	18.94	-
6年0-6月末まで	-	-	19.15	20.14	20.04	-



表8 昭和35年、45年、55年、平成2年および12年の乳幼児身体発育調査による身長分布(cm)

年・月・日齢	男子					
	昭和35年	昭和45年	昭和55年	平成2年	平成12年	21世紀パネル調査
0年0-6月末まで	65.5	66.7	66.6	67.1	66.8	67.00
0年6-12月末まで	73.1	74.2	74.3	74.3	74.4	74.60
1年0-6月末まで	78.4	80.1	80.1	80.6	80.2	80.01
1年6-12月末まで	83.0	84.9	85.2	85.3	85.5	81.13
2年0-6月末まで	85.0	87.1	87.2	87.4	87.1	88.94
2年6-12月末まで	88.5	90.8	91.1	91.3	91.0	90.21
3年0-6月末まで	91.9	94.4	94.8	95.0	94.7	96.26
3年6-12月末まで	95.0	97.8	98.2	98.6	98.3	98.60
4年0-6月末まで	98.2	101.2	101.5	102.1	101.6	104.09
4年6-12月末まで	101.4	104.3	104.6	105.4	104.9	104.83
5年0-6月末まで	104.4	107.1	107.6	108.6	108.1	-
5年6-12月末まで	107.4	109.6	110.6	111.6	111.4	-
6年0-6月末まで	-	-	113.6	114.5	114.9	-

年・月・日齢	女子					
	昭和35年	昭和45年	昭和55年	平成2年	平成12年	21世紀パネル調査
0年0-6月末まで	64.0	65.2	65.3	65.4	65.4	-
0年6-12月末まで	71.6	73.0	73.0	73.0	72.7	72.00
1年0-6月末まで	77.1	78.7	79.0	79.4	79.1	78.84
1年6-12月末まで	81.4	83.7	84.1	83.9	84.4	79.83
2年0-6月末まで	83.7	86.1	86.3	86.0	86.0	87.76
2年6-12月末まで	87.2	89.5	90.2	90.1	89.9	89.00
3年0-6月末まで	90.7	93.0	93.9	94.0	93.7	95.24
3年6-12月末まで	94.1	96.4	97.5	97.7	97.4	97.64
4年0-6月末まで	97.3	99.8	100.9	101.3	101.0	103.28
4年6-12月末まで	100.4	103.1	104.1	104.7	104.3	104.77
5年0-6月末まで	103.3	106.2	107.1	107.9	107.6	-
5年6-12月末まで	106.3	109.1	109.8	110.9	110.8	-
6年0-6月末まで	-	-	112.2	113.8	113.8	-

表9 体重の成長に関するパネル推定(全体)

Dependent Variable: lnbdywh	Pooling		Between		Random		Fixed	
	Estimated Coefficient	t	Estimated Coefficient	t	Estimated Coefficient	z	Estimated Coefficient	t
survivalday	0.002	1142.57	0.002	332.03	0.002	1312.04	0.002	1272.91
survivalday Sq	-8.64E-07	-682.65	-8.94E-07	-185.5	-8.65E-07	-780.11	-8.64E-07	-754.53
lnkosodate	0.005	10.16	0.002	2.27	0.004	9.46	0.005	8.56
_cons	1.132	1371.13	1.117	567.76	1.131	1368.35	1.131	1439.73
<b>Diagnostics</b>								
Number of observation	175549		175549		175549		175549	
Number of groups			46526		46526		46526	
R-sq within			0.968		0.968		0.968	
between			0.894		0.894		0.894	
overall	0.947		0.947		0.947		0.947	
F test that all $u_i=0$							F(46525, 129020)=2.40	
sigma_u							0.082	
sigma_e							0.132	
rho							0.277	
Breusch-Pagan Lagrangean multiplier test for random effect							chi2(1) = 19850.17 Prob > chi2 = 0.0000	
Hausman Test							chi2(2) = 9.36 Prob > chi2 = 0.0093	

表10 身長の高成長に関するパネル推定(全体)

Dependent Variable: lnbdyght	Pooling		Between		Random		Fixed	
	Estimated Coefficient	t	Estimated Coefficient	t	Estimated Coefficient	z	Estimated Coefficient	t
survivalday	0.001	1312.41	0.001	410.17	0.001	1480.13	0.001	1430.78
survivalday Sq	-3.15E-07	-718.1	-3.29E-07	-211.7	-3.16E-07	-806.28	-3.15E-07	-776.55
lnkosodate	0.001	3.86	0.000	1.1	0.000	2.55	0.000	1.26
_cons	3.901		3.896	6177.05	3.901		3.901	
<b>Diagnostics</b>								
Number of observation	170336		170336		170336		170336	
Number of groups			46459		46459		46459	
R-sq within			0.979		0.979		0.979	
between			0.937		0.937		0.937	
overall	0.967		0.967		0.967		0.967	
F test that all $u_i=0$							F(46458, 123874)=2.20	
sigma_u							0.026	
sigma_e							0.046	
rho							0.247	
Breusch-Pagan Lagrangean multiplier test for random effect							chi2(1) = 15789.23 Prob > chi2 = 0.0000	
Hausman Test							chi2(2) = 8.68 Prob > chi2 = 0.013	

表 11 体重の成長に関するパネル推定(男子)

Dependent Variable: lnbdywh	Pooling		Between		Random		Fixed	
	Estimated Coefficient	t	Estimated Coefficient	t	Estimated Coefficient	z	Estimated Coefficient	t
survivalday	0.002	828.07	0.002	243.95	0.002	937.86	0.002	908.71
survivalday Sq	-8.83E-07	-499.63	-9.17E-07	-137.75	-8.82E-07	-563.28	-8.81E-07	-544.05
lnkosodate	0.004	6.08	0.002	1.41	0.003	4.91	0.003	3.79
_cons	1.145	994	1.130	419.29	1.145	991.44	1.146	1029.7
<b>Diagnostics</b>								
Number of observation	91193		91193		91193		91193	
Number of groups			24185		24185		24185	
R-sq within			0.966		0.967		0.967	
between			0.896		0.896		0.896	
overall	0.947		0.946		0.947		0.947	
F test that all $u_j=0$							F(24184, 67005) = 2.25	
sigma_u					0.079		0.110	
sigma_e					0.135		0.135	
rho					0.255		0.401	
Breusch-Pagan Lagrangean multiplier test for random effect					chi2(1) = 8694.94 Prob > chi2 = 0.0000			
Hausman Test							chi2(2) = 6.29 Prob>chi2 = 0.043	

表 12 身長の高成長に関するパネル推定(男子)

Dependent Variable: lnbdyght	Pooling		Between		Random		Fixed	
	Estimated Coefficient	t	Estimated Coefficient	t	Estimated Coefficient	z	Estimated Coefficient	t
survivalday	0.001	950.71	0.001	300.4	0.001	1064.89	0.001	1028.82
survivalday Sq	-3.19E-07	-523.85	-3.37E-07	-156.59	-3.19E-07	-584.22	-3.19E-07	-562.35
lnkosodate	0.000	2	0.000	0.97	0.000	0.61	0.000	-0.81
_cons	3.906	9926.3	3.901	4511.52	3.906	9919.66	3.907	
<b>Diagnostics</b>								
Number of observation	88536		88536		88536		88536	
Number of groups			24148		24007		24007	
R-sq within			0.979		0.979		0.979	
between			0.938		0.938		0.938	
overall	0.967		0.966		0.967		0.967	
F test that all $u_j=0$							F(24147, 64385) = 2.13	
sigma_u					0.026		0.037	
sigma_e					0.046		0.046	
rho					0.235		0.394	
Breusch-Pagan Lagrangean multiplier test for random effect					chi2(1) = 7570.28 Prob > chi2 = 0.0000			
Hausman Test							chi2(2) = 12.43 Prob>chi2 = 0.0020	

表 13 体重の成長に関するパネル推定(女子)

Dependent Variable: lnbdywh	Pooling		Between		Random		Fixed		
	Estimated Coefficient	t	Estimated Coefficient	t	Estimated Coefficient	z	Estimated Coefficient	t	
survivalday	0.002	799.5	0.002	231.33	0.002	922.15	0.002	894.71	
survivalday Sq	-8.45E-07	-472.44	-8.70E-07	-127.72	-8.45E-07	-542.24	-8.45E-07	-524.46	
lnkosodate	0.006	8.37	0.002	1.84	0.006	8.7	0.007	8.55	
_cons	1.117	959.15	1.102	392.96	1.115	956.05	1.115	1009.89	
<b>Diagnostics</b>									
Number of observation	84356		84356		84356		84356		
Number of groups			22341		22341		22341		
R-sq within			0.969		0.969		0.969		
between			0.896		0.896		0.895		
overall	0.949		0.949		0.949		0.949		
F test that all $u_j=0$							F(22340, 62012) = 2.45		
sigma_u					0.081		0.110		
sigma_e					0.128		0.128		
rho					0.286		0.421		
Breusch-Pagan Lagrangean multiplier test for random effect					chi2(1) = 9888.44 Prob > chi2 = 0.0000				
Hausman Test							chi2(2) = 7.96 Prob > chi2 = 0.019		

表 14 身長の高成長に関するパネル推定(女子)

Dependent Variable: lnbdyht	Pooling		Between		Random		Fixed		
	Estimated Coefficient	t	Estimated Coefficient	t	Estimated Coefficient	z	Estimated Coefficient	t	
survivalday	0.001	916.83	0.001	286.33	0.001	1030.58	0.001	995.45	
survivalday Sq	-3.11E-07	-497.77	-3.22E-07	-146.17	-3.11E-07	-557.12	-3.11E-07	-536.08	
lnkosodate	0.001	3.47	0.000	0.67	0.001	3.12	0.001	2.69	
_cons	3.896	9684.78	3.891	4324.11	3.895	9655.38	3.895	9899.85	
<b>Diagnostics</b>									
Number of observation	81800		81800		81800		81800		
Number of groups			22311		22311		22311		
R-sq within			0.980		0.980		0.980		
between			0.938		0.938		0.938		
overall	0.968		0.968		0.968		0.968		
F test that all $u_j=0$							F(22310, 59486) = 2.15		
sigma_u					0.026		0.037		
sigma_e					0.045		0.045		
rho					0.244		0.399		
Breusch-Pagan Lagrangean multiplier test for random effect					chi2(1) = 7046.58 Prob > chi2 = 0.0000				
Hausman Test							chi2(2) = 0.68 Prob > chi2 = 0.711		

図1 調査回別の身長の高さのヒストグラム

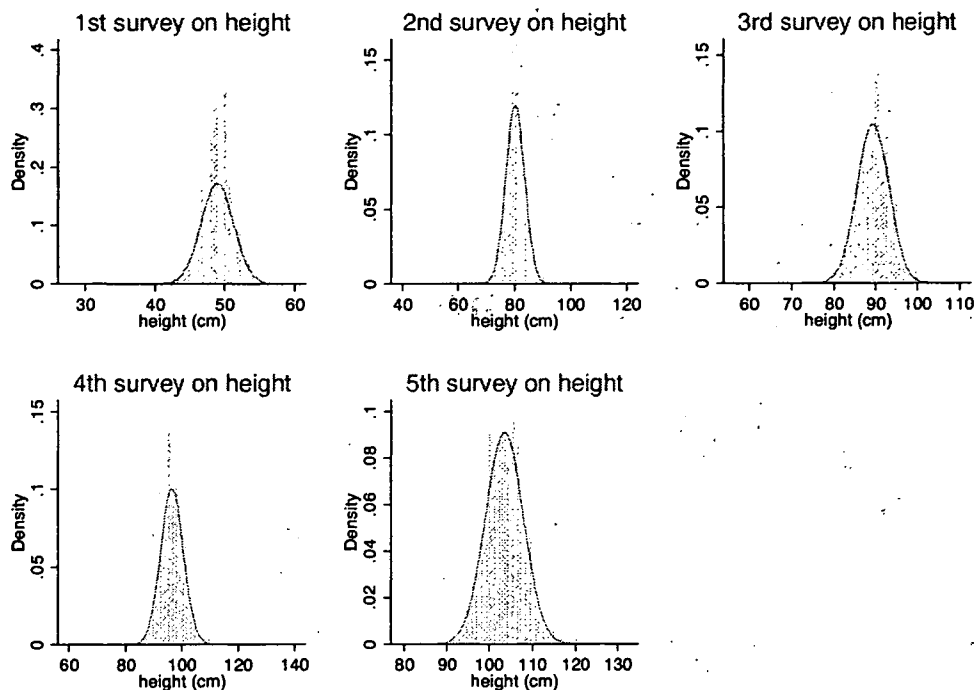


図2 調査回別の体重のヒストグラム

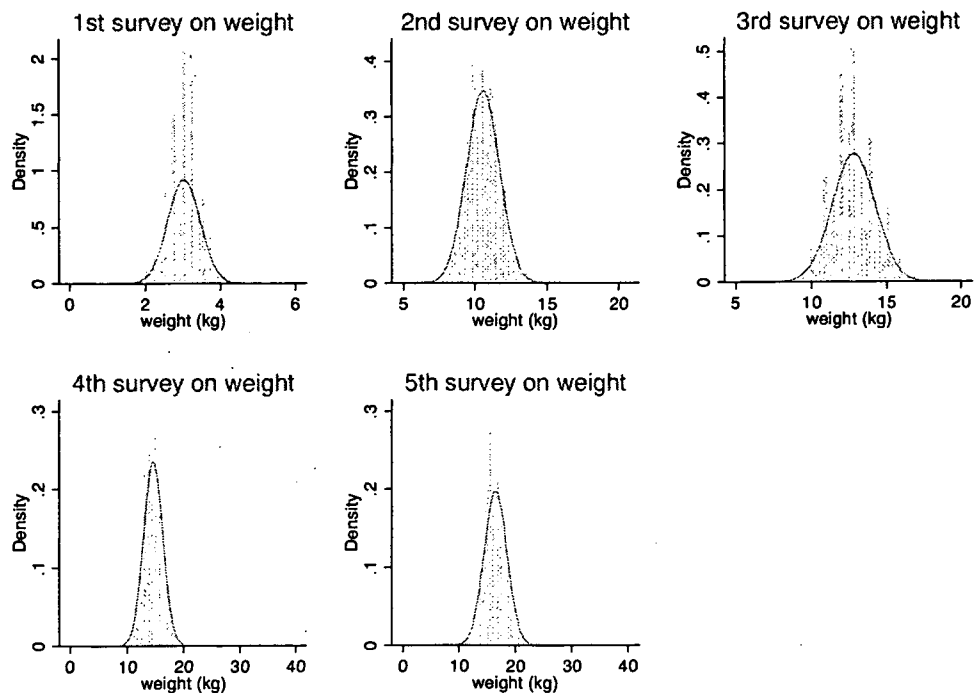


図 3 身長と体重のプロットと統計的推定図

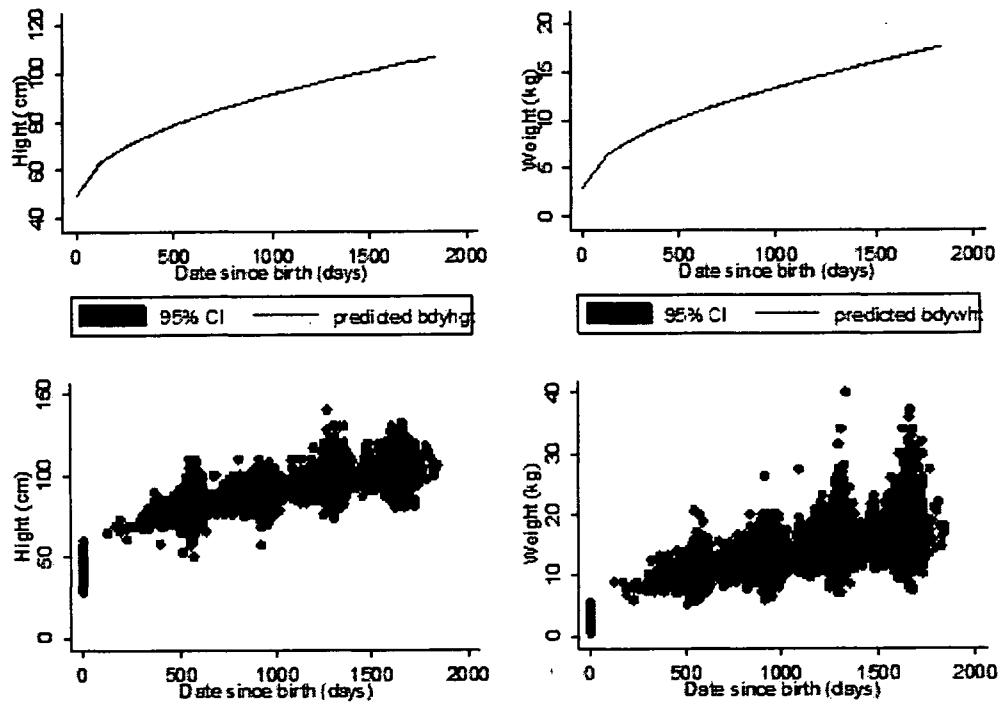


図 4 身長と体重のクロスプロット図

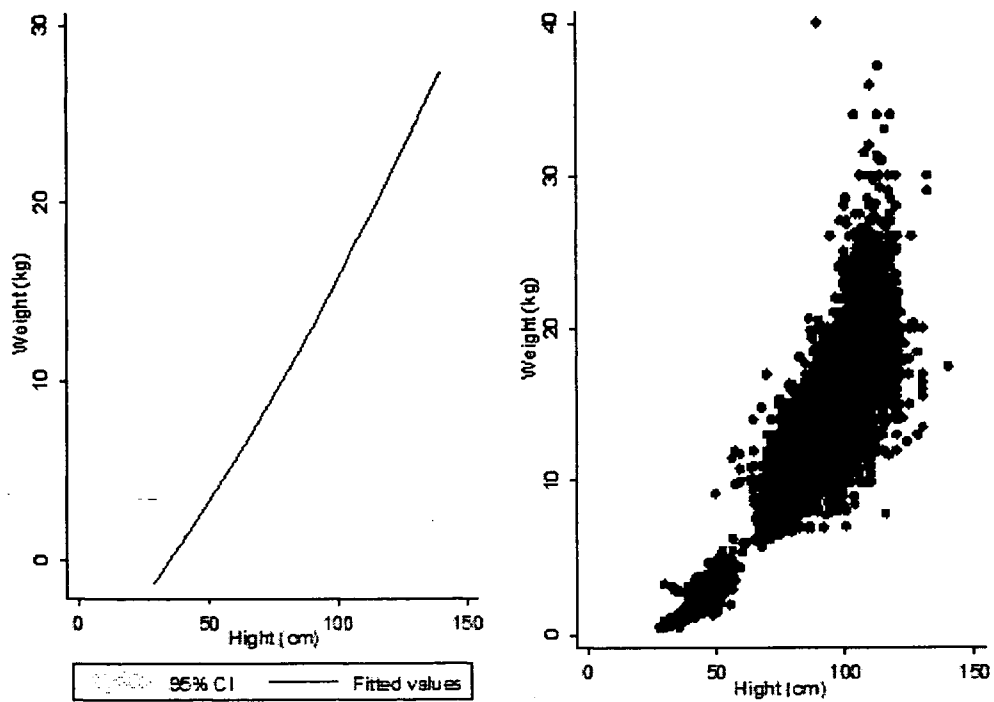


図5 身長と体重の変化のプロット図と統計的推定図

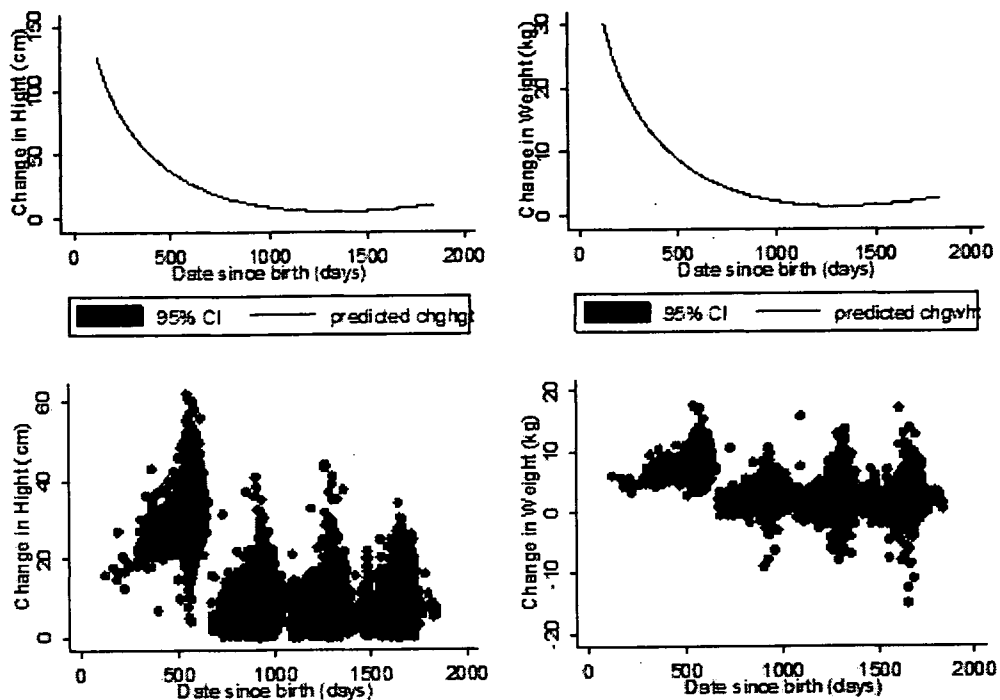
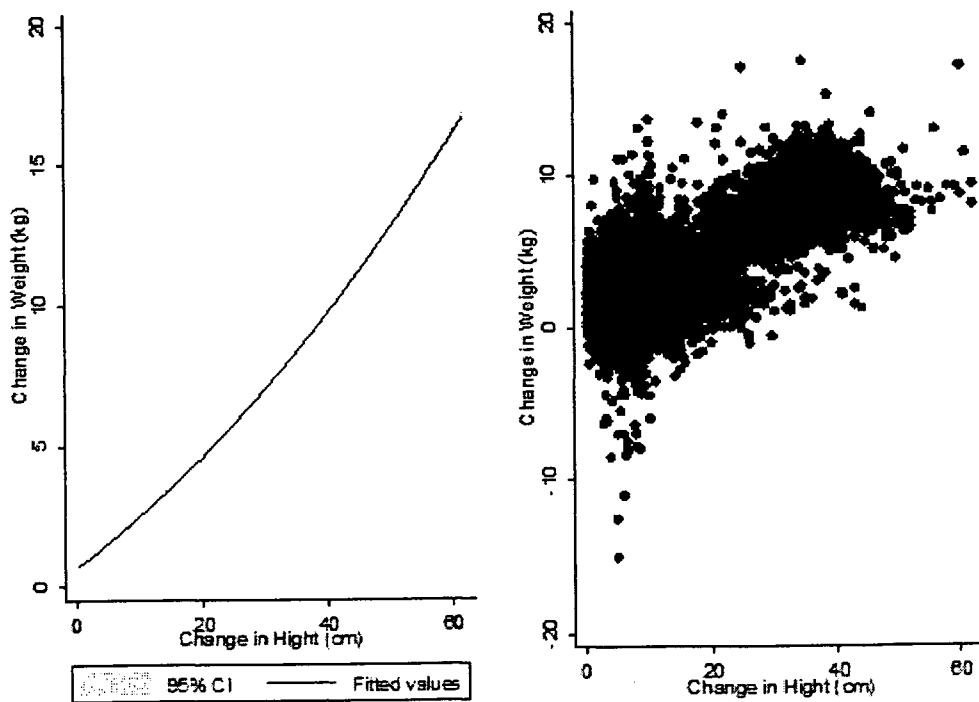


図6 身長と体重の変化のクロスプロット図



## 6 結婚と出生に関する分析



# (1) 成年者縦断調査を用いた初婚の要因分析：ネステッド・ロジットモデルによる初婚と脱落の競合ハザード分析

福田 節也

## 1. はじめに

パネル調査において脱落は不可避の問題である。脱落はサンプル数の減少やサンプル属性の偏向をもたらす。そのため、記述統計やクロス集計といった単純統計の算出においては、脱落の取り扱いは分析者にとって喫緊の課題である<sup>1</sup>。一方、多変量解析によるパラメータ推定においては、脱落による影響はそれほど深刻ではないことが報告されている (Fitzgerald et. Al. 1998, Lillard and Panis 1998, Alderman et. Al. 2001, Falaris 2003)。いずれの研究においても、モデルが正しく設定されている限り、脱落は多変量解析による回帰係数の推定には有意な影響を与えないとの結論を得ている。しかし、この結論は多変量解析において脱落による影響を無視してよいということを意味するのではない。使用する分析手法や従属変数によっては、脱落は分析の妥当性を左右する重要な問題となりうる。

近年、わが国でも盛んに用いられるようになった分析手法の 1 つにイベントヒストリー分析 (event-history analysis) がある。パネルデータを用いたイベントヒストリー分析では、最も一般的な方法として、脱落はセンサリング (censoring) (観察打ち切り例) として扱われてきた。しかし、脱落と対象とするイベントとが独立に生起しない場合、このような処置はパラメーターの推定にバイアスをもたらす。本稿では、イベントヒストリー分析の一種である離散時間ロジットモデル (discrete-time logit model) を用いた初婚の要因分析を例として、脱落がパラメーター推定に与える影響について明らかにする。具体的には、1) 脱落をセンサリングとして扱い、通常の離散時間ロジットモデルを行った場合と、2) 脱落を初婚に競合するイベントとして扱い、双方のパラメーターを同時推定 (simultaneous estimation) によって求める場合で初婚要因のパラメーターを比較する。なお、後者のモデルにおいては、Hill ら (1993) によって提案された離散時間ネステッド・ロジットモデル (discrete-time nested logit model) を適用する<sup>2</sup>。分析によって、1) 初婚と脱落が独立に生起しているのかを統計的に検定し、両者が非独立であった場合、2) 脱落をセンサリングとして扱くと、初婚要因のパラメーター推定にどの程度バイアスが生じるのかを提示する。

<sup>1</sup> この問題への対処法としては、脱落サンプルの属性を加味したウェイトの使用などが提案されている (Lepkowski and Couper 2002, Wooldridge 2002)。

<sup>2</sup> Hill ら (1993) は、SURF (Shared Unmeasured Risk Factors) モデルの呼称を用いている。

## 2. パネルデータを用いた離散時間ロジットモデル

### 2-1. イベントヒストリー分析の概要

パネルデータに対する主要な分析手法の1つとして、イベントヒストリー分析がある。イベントヒストリー分析とは、あるイベントの発生パターンとその要因に関する分析手法の総称である。別名、生存分析 (survival analysis) ともいわれる。

イベントヒストリー分析では、リスク人口 (population at risk) におけるイベント発生確率である「ハザード率 (hazard rate)」を分析の対象とする。リスク人口とは、イベントを経験する可能性がある人口を指す。例えば、離婚をイベントとして分析を行う場合、離婚のリスク人口は有配偶の男女であり、未婚者や死別者、離別者はリスク人口に含まれない。ハザード率は、より正確には「時間  $t$  に至るまでの期間に、当該イベントが起こらなかったという条件のもとでの、時間  $t$  におけるイベント発生の瞬間確率 (instantaneous rate)」(津谷 2002, p. 429 右段, ll. 49-52) を指し、以下のように表わされる。

$$h(t) = \lim_{\Delta t \rightarrow 0} [P(t + \Delta t > T \geq t | T \geq t) / \Delta t] \quad \dots (1)$$

ハザードとは、英語で「危険」を意味する言葉であるが、これはハザード率の概念が死亡を分析対象とすることの多い生物統計において発展したことに由来している。通常、リスク人口におけるイベント発生確率は、イベント発生のリスク開始時点からの「時間」によって異なる。また、イベント発生確率が時間の経過とともにどのようなパターンを示すのかも、対象となる集団・人口によって異なる。イベントヒストリー分析は、このハザード率を時間の関数として特定し、それが単数あるいは複数の説明要因によってどのように変化するのかを明らかにする多変量回帰分析である。モデルのパラメータは、最尤法 (maximum likelihood method) もしくは部分尤度法 (partial likelihood method) によって推定される。時間の関数として表わされるハザード率は、ベースライン・ハザード (baseline hazard) と呼ばれ、モデルの他の要因を統制した場合におけるイベント発生確率の基本的なパターンを表わす。

また、モデルにおける説明変数は共変量 (covariate) と呼ばれる。共変量には時間によって値が変化する変数と、そうでないものがある。前者を「時間依存性共変量 (time-varying covariate)」といい、年齢や配偶関係、職業、あるいは学歴といった変数がこれにあたる。一方、後者を「時間独立共変量 (time constant covariate)」と呼ぶ。性別や生年月日、出身地などがこれにあたる。時間依存性共変量を用いることができるのは、時間の概念をもつイベントヒストリー分析ならでの利点である。

イベントヒストリー分析において重要な概念にセンサリングがある。観察対象となるイベントのリスク期間について、その終了時点が明らかではない場合をセンサリングという。このうち、観察期間中にイベントが生起しないケースを右センサリング (right-censoring)

といい、観察期間前にイベントが生起しているケースを左センサリング (left-censoring) という<sup>3</sup> (Guo 1993, Allison 1995)。左センサリングについては、イベントヒストリー分析をはじめ、多くの分析において対処することができない。しかし、右センサリングについては、イベントヒストリー分析では、イベントが生起しなかった時点までの情報を分析に反映して、リスク人口全体を対象とした分析を行うことができる。また、パラメーター推定についても、右センサリングがイベントの生起ハザード率と独立に発生していると仮定できる場合、バイアスのない値を算出することができる (Allison 1995)。

## 2-2. 離散時間ロジットモデル

イベントヒストリー分析にはいくつかのモデルがある。本論文において用いるのは、イベントヒストリー分析のうち、時間の測定単位が連続的 (際限なく細かい) とは仮定できず、離散的 (序数的) である場合に利用される分析手法である離散時間ロジットモデル (Allison 1982) である。離散時間ロジットモデルのモデル式は以下によって表される。

$$\ln[P_t/(1-P_t)] = a_t + b_1 X_1(t) + b_2 X_2(t) + \dots + b_k X_k(t) \quad \dots (2)$$

$P_t$ : ハザード確率、 $a_t$ : 時間変数、 $b_k$ : 共変量  $X_k$  の回帰係数、 $X_k$ : 共変量  $k$

(2) 式より分かるように、離散時間ロジットモデルは、各リスク時点でのハザード確率  $P_t$  のロジット<sup>4</sup>を被説明変数とする回帰モデルである。ここでいうハザード確率とは、時間  $t$  までにイベントが発生していないという条件の下で、時間  $t$  においてイベントが発生する確率を意味する。前述のハザード率は、時間の区切りが無視できるほど小さい場合に定義される確率密度 (probability density) であり、ここでいうハザード確率とは異なるものであることに留意されたい。(2) 式はロジットモデル (ロジスティック回帰分析) と類似しており、係数を指数化してハザード確率のオッズ比として解釈することができる。ただし、ロジットモデルでは確率  $P$  を扱うのに対して、離散時間ロジットモデルでは、ハザード確率  $P_t$  を用いる。また、離散時間ロジットモデルでは、定数  $a$  や共変量  $X$  がリスク期間中に変化することを許容している点も通常のロジットモデルとは異なる。係数  $b_k$  は、共変量  $X_k$  がハザード確率のロジットに与える効果を意味している。ただし、離散時間ロジットモデルでは、係数  $b_k$  は共変量  $X_k$  のリスク期間を通じた平均的な効果を表していることに留意

<sup>3</sup> 社会科学において、左センサリングの定義は曖昧であり、リスク期間の開始時点が不明な場合を左センサリングという場合もある (Guo 1993, Allison 1995)。本稿では Guo (1993) に倣い、そのようなケースは左打ち切り (left-truncation) と呼ぶことにする。

<sup>4</sup> ロジットとはオッズを自然対数化した値をいう。オッズとは、イベントが生起しない確率 ( $1-P$ ) に対するイベント生起確率 ( $P$ ) の比ことを指し、 $P/(1-P)$  として表される。

する必要がある<sup>5</sup>。また、時間変数  $a_t$  は、ベースライン・ログオッズ (baseline log odds) である。ベースライン・ログオッズは、すべての共変量  $X$  が 0 であった場合におけるハザード確率のロジットの時間推移を表しており、時間経過にともなうイベントの基本的な発生パターンを表す。

離散時間ロジットモデルの適用においては、はじめに、リスク開始からイベントが発生するか、もしくはセンサリングとなった時点までの人・期間別データ (person-period data) を作成する<sup>6</sup>。次に、この人・期間別データに対して、イベントが生起するか否かのダミー変数を従属変数とする通常のロジット分析を行う。なお、パラメーターの推定は最尤法によって行われる。

離散時間ロジットモデルは、パネルデータと最も親和性が高いイベントヒストリー分析であるといえる。なぜならば、通常個人を対象としたパネル調査では、調査が行われるのは年に 1 回であり、各年における結婚や出産、就業状態等の変化は、調査時点の状態の変化によって測定されることが多いためである。例えば結婚であれば、ある個人が結婚したか否かは、前年の調査で未婚であった人が当年の調査で有配偶であることによって把握されることが多い<sup>7</sup>。そのため、結婚の生起は  $t-1$  年から  $t$  年の間に起きたことは明らかであっても、具体的にいつ、例えば何月に起きたのかまでは不明である場合が多々ある。このような場合には、イベントの生起時点に関する情報は年単位でしか把握することができず、連続時間を仮定することができない。したがって、イベント発生月が不明である場合には、ハザード率の近似として、リスク期間別のハザード確率を用いた離散時間ロジットモデルを利用することが最も簡便かつ実際的である<sup>8</sup>。

<sup>5</sup> 共変量  $X_k$  と時間変数  $a_t$  の交互作用項をモデルに組み入れることで、係数  $\beta_k$  がリスク期間を通じて変化することを許容するモデルを構築することが可能である (山口 2002c, 津谷 2002)。

<sup>6</sup> 人・期間別データの詳細については、山口 (2001b) を参照されたい。

<sup>7</sup> なかにはイベント発生時点に関する質問を追加して、結婚や出産などのイベントについて、月単位での生起時点を把握しているパネル調査もある。本稿で用いる「21 世紀成年者縦断調査」もそうした調査の 1 つである。イベント発生月に関する情報があるパネルデータでは、連続時間モデルの適用が可能である。しかし、イベント生起の有無については明らかであっても、発生月に関する情報が欠損している場合には、なんらかの仮定によって、発生月を代入する必要がある。また、後に述べるように本稿では脱落を初婚に競合するイベントとして扱う。脱落をイベントとして扱う場合、調査と調査の間でいつ脱落が発生したのかを論理的に定義することは不可能である。そのため、本稿における分析では、離散時間モデルを適用することが最も合理的な選択となる。

<sup>8</sup> Allison (1982) は、ロジットモデルの代わりに complementary log-log モデルを使用して同様の分析が可能であることを示している。complementary log-log モデルを使用した離散時間ハザードモデルによって算出される係数  $\beta$  は、連続時間を仮定する等比ハザードモデルから算出された係数  $\beta$  の近似値となる。そのため、離散時間ロジットモデルの  $\exp(\beta)$  はハザード確率のオッズ比を表すが、離散時間 complementary log-log モデルの  $\exp(\beta)$  は、連続時間を仮定する等比ハザードモデルと同様にハザード比として直接的に解釈できるという利点がある (Allison 1995)。近年の統計パッケージには、complementary log-log モデルをサポートするものも多いため、併せてその使用を検討すべきである。