

JGSS 統計分析セミナー

- イベントヒストリー分析の適用例 -

佐々木 尚之

大阪商業大学 JGSS 研究センター

JGSS Statistical Analysis Seminar:
Application Examples of Event History Analysis

Takayuki SASAKI

JGSS Research Center

Osaka University of Commerce

JGSS Research Center hosted a statistical analysis seminar on the theme of event history analysis. Event history analysis is statistical methodology to conduct quantitative investigation of when and how a variety of events in each life stage are likely to happen. Although event history analysis provides many advantages over other analytical methods in the case when life history data is in use, there are few studies that employed this approach. This article introduces three different types of event history analysis with SPSS, namely, discrete-time logit model, piecewise constant rate model, and cox regression model. Characteristics of each model are summarized and application examples of each model in examining factors associated with divorce are demonstrated. It turns out that final conclusions of each model are similar. Most importantly, however, it is necessary to assess which model fits to the data the best.

Key Words: JGSS, event history analysis, divorce

JGSS 研究センターでは、イベントヒストリー分析をテーマに 2008 年度統計分析セミナーを開催した。イベントヒストリー分析とは、人生の過程で起きる様々な出来事が、いつどのような状況で起こる傾向があるのかを、数量的に明らかにする手法である。イベントヒストリー分析は、ライフヒストリー・データを利用する場合に、他の分析方法に比べ多くの利点があるが、日本ではまだ適用頻度が低い。本稿では、SPSS を用いた、離散時間ロジットモデル、時間区分別定率モデル、コックス回帰モデルの 3 つの異なるタイプのモデルを紹介し、それぞれの特徴をまとめて、離婚要因の分析適用例を示した。どのモデルを用いても、最終的な結論にほとんど差異はないが、データに適したモデルを見極める必要性が明らかになった。

キーワード：JGSS，イベントヒストリー分析，離婚

1. 本稿の目的

2008年8月26日と27日の両日に、統計分析のスキルアップを目指す大学院生・研究者を対象とした、JGSS 研究センター主催の統計分析セミナーが開催された⁽¹⁾。セミナーの講師には、2008年に引き続き、山口一男教授（シカゴ大学社会科学部）を招聘した。JGSS 研究センターでは、グローバル化経済の下で急速に進行している働き方やライフスタイルの多様化をとらえることを目的として、28～42歳を対象にライフコース調査を2009年に実施を計画しており、イベントヒストリー分析についての講義をお願いした。イベントヒストリー分析は、ライフヒストリー・データを利用する場合に他の分析方法に比べ多くの利点があるが、日本ではまだ適用頻度が低い。

本稿では、実際のライフヒストリー・データを用いて、JGSS 統計分析セミナーで行われた具体的な適用例の過程をたどる。本稿の目的は、統計分析セミナーで行われた分析演習の詳細な過程を記述することにより、イベントヒストリー分析の理解、さらには適用の進展を目指すことである。セミナーでは、限られた時間の中で、できるだけシンプルな変数を用いていただくようお願いしていたので、紹介するモデルによっては、まだ改善の余地があることをご理解いただきたい。統計分析セミナーの講義の部分では、イベントヒストリー分析の目的と意義、イベントヒストリー分析を行うにあたって必要な、Person-period データファイルの作成方法が解説された。実習部分では、SPSS を用いて、離散時間ロジットモデル、セミ・パラメトリックな等比ハザードモデル、コックス回帰モデルの分析例が紹介され、LEM⁽²⁾を用いて、観察されない異質性を仮定するモデル、競合するイベントへのネステッド・ロジットモデルの分析例が紹介された。本稿では、以上の5つのモデルのうち SPSS を用いた3つのモデルの分析演習に焦点を当てる。

2. イベントヒストリー分析とは

イベントヒストリー分析とは、人生の途上で起きるさまざまな出来事（イベント）が、いつどのような状況で起こる傾向があるのかを、数量的に明らかにする手法である。例えば、結婚、離婚、出産、就職、離職、転居、死亡などは、イベントヒストリー分析に用いられる出来事の代表的な例である。イベント発生時点が明らかであれば、多岐にわたる出来事を対象として分析を行うことができる。言い換えるならば、注目する出来事の有無ではなく、その出来事が人生のどの時点（年齢または年月など）で起きたのかを特定できるデータが必要となる。また同時に、注目する出来事が発生する可能性（リスク）がいつから始まったのかも特定する必要がある⁽³⁾。つまり、注目する出来事が離婚であるならば、そのリスクは結婚時から始まり、注目する出来事が離職であるならば、就職がそのリスクの始まった時点ということになる。イベントヒストリー分析において、この情報は非常に重要な意味を持ち、分析モデルにおいては、リスクが始まってから、いつまで出来事が起こらずに継続しているのかが、分析対象となる。

山口（2001 9月号）はイベントヒストリー分析の利点として4つを挙げている。まず1つ目の利点は、注目する出来事が、ある一定期間に起こらなかった場合の情報を、偏りなく利用できる点である。例えば、離婚の生起に注目する場合、観察期間終了時まで離婚しないケースや結婚数年後に配偶者と死別したため、離婚が起こる可能性がなくなったケースがでてくる。このようなケースを「右センサリング（right censoring）」と呼び、その時点で観察が打ち切られた（右センサーされた）ことを意味する。しかし、個々のケースそれぞれに、離婚のリスクが始まった結婚時から、何年間離婚が起こらなかったという情報を持っており、イベントヒストリー分析を用いることにより、それらのデータをパラメータ推定に生かすことができる。ただし、何年間離婚が起こらなかったかを特定するには、離婚のリスクが始まった結婚時を特定する必要があり、「左センサリング（left censoring）」と呼ばれるリスク開始時が不明な状態を回避する必要がある。

2つ目の利点は、時間とともに変化する説明変数をモデルに投入できる点である。人々の状態や環境は、多くの場合、時間に依存している。無職だった人が正規雇用者となり、非正規雇用に転職し、再び無職となるといったような就業上の変化が起きた場合に、ある一時点の状況のみを説明変数とし

て投入すると、結果に大きな歪みをもたらしてしまう。イベントヒストリー分析では、ある時点 t で注目する出来事が起きるリスクが、 t 以前の説明変数の状況と関連しているのかを検証できる。しかしながら、必ずしも因果関係を説明する分析ではない。1年前の健康状態を説明変数に離婚のリスクを検証し、関連性が明らかになったとしても、健康だから結婚を継続しているのか、結婚を継続しているから健康なのかの、どちらかを証明するわけではないので、解釈には注意が必要である。

3つ目の利点は、他の分析方法では、分析から除外されてしまうケースをパラメータ推定に生かし、存在する情報を最大限に利用できる点である。例えば、刑務所から出所後10年間に、再犯を犯すか否かの分析の場合、ロジスティック回帰分析では、出所後10年経っていない人が、分析から除外されてしまうし、出所後1年以内に再犯を犯した人と9年後に犯した人を同質のもとして分析されてしまう。イベントヒストリー分析では、出所後10年以内であったとしても、それまで再犯を犯していない継続時間が重要な情報として利用される。また、再犯が起きた時期の差異も考慮できる。

最後に4つ目の利点は、どの時点で何が起こると、注目する出来事の発生リスクが上がるのかを分析できる点である。再犯率を例にとると、出所直後は人的ネットワークの存在が再犯リスクを低くするかもしれないが、出所数年後には本人の経済状況が再犯リスクとより強い関連をもつかもしれない。これらの決定要因と時間の交互作用効果を検証することにより、各決定要因の影響力が時間とともに変化するのかどうかを分析することができる。

3. データファイルの作成⁽⁴⁾

一般に、イベントヒストリー分析を行う際には、通常とは異なる形式のデータファイルに変換する必要がある。通常の分析で用いられるデータファイルは、表1のように、各個人の情報がそれぞれ各行に記入されている (Person-level データ)。これらのデータを2000年に観察されたライフヒストリー・データと仮定し、「9999」を「非該当」とすると、1番目の分析対象者は、大学卒で、1998年に結婚し、2000年の調査時まで離婚しておらず、配偶者の親との同居歴もない。2番目の分析対象者は、高校卒で、1986年に結婚し、翌年の1987年から配偶者の親と同居を始め、1991年に離婚し、配偶者の親との同居も終えている。この情報をもとに、イベントヒストリー分析では、注目するイベントの生起リスク開始時から終了時までの時間区分をそれぞれの行で表す (Person-period データ)。注目するイベントを離婚として、表1のデータを Person-period データに変換すると、表2ようになる。この場合、時間区分は年単位なので、1番目の分析対象者は離婚のリスクが始まった結婚時の1998年から1999年、そして観察時である2000年までの3行を使って表す。この分析対象者は、観察時まで、離婚が発生しなかったため、右センサーされたケース (観察打ち切り) となるが、結婚後3年間離婚しなかったという情報を持っている。

時間依存しない変数に関しては、各個人に同じ数値が記入されている。この例の場合、最終学歴は時間とともに変わることのない変数である。配偶者の親との同居の有無は時間に依存するため、同一人物であっても、異なる数値が記入されている。時間依存する変数を Person-period データに変換する際には、その変数がイベントの規定要因である前提を満たすように、リスク継続時以前の状態とすることが多い。つまり、この例では、1年のラグをとり (意図的に時間差を作り) 結婚後 n 年目の離婚ハザード率を $n-1$ 年目の配偶者の親との同居の有無によって説明しようと試みている。2番目の分析対象者は、1986年に結婚して、6年目である1991年に離婚している。よって、これ以降のデータは入力されていない。配偶者の親とは結婚2年目に同居を始めたので、2年目の状態が影響を与えうる3年目から同居ありを示す「1」が記入されるよう設定されている。仮説によっては、2年前の状態や3年前の状態を設定することが可能である。SPSSにおける、詳しい変換方法は、Boslaugh (2004)を参照されたい。

表 1 Person-level データ

ID	結婚年	離婚年	最終学歴 (教育年数)	配偶者の親との同居開始年	配偶者の親との同居終了年
1	1998	9999	16	9999	9999
2	1986	1991	12	1987	1991
3	1963	9999	9	1963	9999
4	1959	1974	12	9999	9999
5	1971	9999	14	9999	9999

表 2 Person-period データ

ID	結婚継続年数	離婚生起	最終学歴	配偶者の親との同居
1	1	0	16	0
1	2	0	16	0
1	3	0	16	0
2	1	0	12	0
2	2	0	12	0
2	3	0	12	1
2	4	0	12	1
2	5	0	12	1
2	6	1	12	1
3	1	0	9	0
3	2	0	9	1
3	3	0	9	1
3	4	0	9	1
3	5	0	9	1
3	6	0	9	1
3	∴	∴	∴	∴

4. 分析に用いるデータ

JGSS 統計セミナーでは、日本家族社会学会全国家族調査委員会に、特別利用の申請をした上で、2002 年に実施された全国調査「戦後日本の家族の歩み」(略称 J-S01 調査)のデータが利用された。この調査は、戦後の日本家族の変動を明らかにするために、32~81 歳の女性を対象に遡及法を用いて、結婚歴、親との同居歴、出産歴、教育歴、職歴、介護歴などの多岐にわたる領域のライフヒストリー・データを収集している。この調査の主な概要は表 3 のとおりである。本稿で対象とするイベントは、初婚から 30 年間の観察期間内に生起する離婚である。調査時まで結婚していた 3 272 人の女性のうち、199 人(6 1%)が結婚後 30 年以内に初婚を離婚によって終えていた。平均初婚年齢は 24 7 歳である。

分析に用いた説明変数を表 4 に示す。先行研究では、離婚に影響する要因として、未就学児の有無や性別などの人口学的要因や 15 歳時の定位家族の属性、結婚・離婚に関する価値観(安藏 2003)に加えて、経済成長率(加藤 2005)などが挙げられているが、本稿の目的を考慮し、絞り込んだ変数を用いていることに留意されたい。

表 3 全国調査『戦後日本の家族の歩み』概要

調査時期	2002 年 1 月~3 月
調査地域	日本全国
調査対象	2001 年 12 月末時点で 32~81 歳の女性(1920~1969 年生まれ)
標本規模	5 000 人(有効回収数 3 475 人)
回収率	69.5%
標本抽出	層化 2 段無作為抽出法
地点数	312 地点
調査方法	訪問留置法

表 4 説明変数の記述統計

変数名	度数	割合(%)	変数名	度数	割合(%)
本人の教育年数			結婚1年目の本人の親との同居		
中学校	819	25.0	非同居(別居・死亡)	3093	94.5
高校	1662	50.8	同居	171	5.2
短期大学・高等専門学校	574	17.5	不明	8	0.2
大学・大学院	206	6.3	結婚1年目の配偶者の親との同居		
無回答	11	0.3	非同居(別居・死亡)	2314	70.7
結婚時の配偶者の職業			同居	941	28.8
大企業・団体の正社員	1173	35.8	不明	17	0.5
中小企業・団体の正社員	1374	42.0	結婚1年目の本人の職業		
農林漁業の自営業	219	6.7	無職	1135	34.7
農林漁業以外の自営業	367	11.2	正社員	1567	47.9
臨時雇い	72	2.2	自営・家族従業	216	6.6
仕事に就いていなかった	41	1.3	臨時雇い	246	7.5
無回答	26	0.8	不明	108	3.3

5. 分析

本節では、統計セミナーで行われた分析演習のうち、J-S01 調査のデータを用いて、SPSS のプログラムで、離散時間ロジットモデル、セミ・パラメトリックな等比ハザードモデル、コックス回帰モデルのそれぞれの手法で分析した過程を記す。それぞれの手法で、同一の説明変数を用いたモデル(ベースモデル)を走らせて結果の比較をした後、それぞれのモデルを発展させてみよう。

(ア) 離散時間ロジットモデル⁽⁵⁾

離散時間ロジットモデル(Discrete-time logit model)は、コックス回帰モデルと並び、イベントヒストリー分析の中で最も多く適用されているモデルの一つである。このモデルでは、注目する出来事は、順序のある離散的な時間単位に基づいて観察されており、その時間単位ごとにイベントが生起するという仮定をおいている。つまり、イベントの生起が月単位で記録されている場合は、リスク開始から1カ月目、2カ月目、3カ月目などにイベントが発生する確率(ハザード率)をそれぞれの時間区分ごとに推定する。このハザード率とは、離散時間ロジットモデルにおいて、ある時点 t 以前にイベントが起こらなかったという条件のもとで、時点 t にイベントが起こる確率と定義されており、推定された各変数の回帰係数の exp をとることにより、ある状態の人は何倍のオッズ比で出来事が起こるといような解釈ができる。

分析に入る前に、説明変数をリコードする必要がある。まず、サンプル全体の離婚経験率が約6%であることを考慮すると、調査期間である30年間のそれぞれの年に離婚が発生するケース数は、非常に少ないことが予想されるため、2年ごとの順序カテゴリー変数を作成した。また、最終学歴が未回答のケースは11で、いずれも離婚を経験しておらず、パラメータ推定に整合性がないため、これらのケースは分析から外した。

表5に各説明変数の回帰係数、標準誤差、オッズ比を示す。回帰係数が正の場合は、それ以前に離婚が起こらなかったという条件の下で、その変数の値が大きくなるほど、離婚発生リスクを高めると解釈できる。反対に、その変数の値が小さくなるほど、離婚発生リスクを下げる。今回のベースモデル(Model 1)の結果をひとつひとつ辿ってみよう。まず、リスクの継続年数に注目する。リスクの継続年数をモデルに投入することによって、イベントがどの時点で生じやすいのかが明らかになる。表5の結果では、結婚1~2年目に比べて、結婚10年目までは、離婚の発生するリスクが高く、その後徐々にリスクが低くなる傾向がある。しかし、結婚29~30年後以外は有意でなく、このモデルでは、ピンポイントで結婚何年目の離婚リスクが高い(低い)といった解釈はできない。

次に、時間依存しない説明変数に注目しよう。最終学歴が高等学校の女性に比べて、中学校の女性の離婚するオッズは0.7倍であり、有意に離婚リスクが低い。また、結婚時の夫の職業が大企業・団

体の正社員の女性に比べて、夫の職業が中小企業・団体の正社員、農林漁業以外の自営業、臨時雇いだった女性の離婚するオッズはそれぞれ2.3倍、3.3倍、6.0倍であり、夫の職業により離婚の可能性が異なる。さらに、連続変数として投入された、結婚時の本人の年齢の結果から、結婚した年齢が若いほど、有意に離婚しやすい傾向が見て取れる。つまり、結婚年齢が1年上がるごとに、離婚するオッズが0.9倍となる。

表5 離散時間ロジットモデル (Model1: ベースモデル) の結果

	B	標準誤差	Exp(B)
リスクの継続年数			
(対: 結婚1~2年目)			
3~4年目	571	371	1.769
5~6年目	570	377	1.769
7~8年目	478	384	1.612
9~10年目	533	382	1.704
11~12年目	-113	427	0.893
13~14年目	-355	454	0.701
15~16年目	-333	454	0.717
17~18年目	-522	480	0.593
19~20年目	-944	553	0.389
21~22年目	-321	466	0.726
23~24年目	-1084	597	0.338
25~26年目	-797	553	0.451
27~28年目	-1222	663	0.295
29~30年目	-2235	1052	0.107
最終学歴			
(対: 高等学校)			
中学校	-387	191	0.679
短期大学・高等専門学校	311	195	1.365
大学・大学院	-439	426	0.645
結婚時の配偶者の職業			
(対: 大企業・団体の正社員)			
中小企業・団体の正社員	839	197	2.314
農林漁業の自営業	087	459	1.091
農林漁業以外の自営業	1198	249	3.314
臨時雇い	1800	338	6.049
無職	941	534	2.563
無回答	1061	732	2.890
本人の親との同居			
(対: 非同居)			
同居	606	231	1.834
不明	1082	526	2.952
非該当	-135	534	0.873
配偶者の親との同居			
(対: 非同居)			
同居	-1003	238	0.367
不明	464	372	1.590
本人の職業			
(対: 無職)			
正社員	785	186	2.192
自営・家族従業	318	295	1.375
臨時雇い	562	231	1.755
不明	696	344	2.006
結婚時の本人の年齢			
	-087	024	0.916
定数	-4665	686	0.09

p 10 p 05 p 01 p 001

各変数のカッコ内の変数を参照カテゴリーとして分析した。以下の分析結果の表も同様である。

今度は、時間依存する説明変数の結果を見てみよう。今回の分析では、1年間のラグをとっており、1年前の状況が、どのように離婚リスクに関連するかを検証している。親との同居の影響については、妻の親と同居するのか、夫の親と同居するのかで、結果は正反対である。妻の親と同居している人は、別居している人に比べ、1年後に離婚するオッズが1.8倍となる。一方、夫の親と同居している人は、別居している人に比べ、1年後に離婚するオッズは0.4倍である。さらに、妻の職業も有意に離婚リスクに関連していた。無職であった人に比べて、正社員であった人、臨時雇いであった人は、それぞれ1年後の離婚オッズが2.2倍、1.8倍である。

Model 1のベースモデルでは、結婚1～2年目を参照カテゴリーと指定しており、リスク継続年数のうち、結婚後29～30年目のみ有意であったが、リスク継続年数のカテゴリー変数全体としては有意であった (ald 52.89 df 14 p 0.01)。そこで、回帰係数の一番大きい値(離婚リスクが一番高い時点)である結婚3～4年目を参照カテゴリーに変更して、分析し直した(Model 2: 結婚3～4年目参照カテゴリーモデル)。その結果、結婚3～4年目に比べて11年目以降30年目までのすべてにおいて、離婚リスクが有意に低いことが明らかになった。つまり、離婚リスクは、結婚3～4年目を頂点とする曲線である可能性が示唆された。そこで、Model 3(離婚ハザード曲線モデル)では、2年ごとのリスク継続年数カテゴリーを連続変数とし、その二乗項をモデルに投入した上で、結婚年数と離婚リスクの間に曲線的な関係があるのかどうかを検定した(表6)。

表6 離散時間ロジットモデル(Model3: 離婚ハザード曲線)の結果

	B	標準誤差	E p(B)
リスクの継続年数(2年毎)	-155	025	856
最終学歴			
(対: 高等学校)			
中学校	-383	191	682
短期大学・高等専門学校	320	195	1378
大学・大学院	-422	426	656
結婚時の配偶者の職業			
(対: 大企業・団体の正社員)			
中小企業・団体の正社員	842	197	2321
農林漁業の自営業	083	459	1086
農林漁業以外の自営業	1198	249	3313
臨時雇い	1801	338	6058
無職	954	535	2596
無回答	1060	732	2886
本人の親との同居			
(対: 非同居)			
同居	611	231	1841
不明	1101	525	3007
非該当	-687	452	503
配偶者の親との同居			
(対: 非同居)			
同居	-997	238	369
不明	497	371	1643
本人の職業			
(対: 無職)			
正社員	708	184	2031
自営・家族従業	287	294	1333
臨時雇い	516	230	1675
不明	669	344	1952
結婚時の本人の年齢	-087	024	917
リスクの継続年数2乗項	-009	006	991
定数	-3583	643	028

p 10 p 05 p 01 p 001

二乗項をモデルに加える際の注意点として、共線性の問題が挙げられる。ある変数とその二乗項には高い相関性があり、同時にモデルに投入するには問題がある。この2つの変数を独立させるさまざまな方法が提案されているが、今回は、リスク継続年数の二乗項をリスク継続年数で単回帰し、得られた回帰係数を2で割り、その数値を、あらかじめリスク継続年数から引いた上で、新たなリスク継続年数の二乗項を算出した。この一連の過程により、2つの変数間の独立性を維持しつつ、パラメータ推定に偏りを引き起こすことを防ぐことができる。離婚ハザード曲線モデル (Model 3) の結果、リスクの継続年数の二乗項は有意ではなく、リスクの継続年数は有意であった。このことから、結婚年数と離婚リスクの関係は、曲線ではなく線形であることが示唆された。つまり、結婚年数が経てば経つほど、一律に離婚のリスクが下がるという結果が出た。

離散時間ロジットモデルの最後に、リスク継続年数カテゴリーを2年区切りにしたことが妥当であったのかどうかを、結婚1年目ダミーを投入することによって確認した (Model 4 : 結婚1年目ダミーモデル)。その結果、結婚1年目の離婚リスクは有意に低く、結婚2～3年目にかけて離婚リスクが上昇し、徐々に下がっていく曲線であることが推測された (表7)。つまり、今回の分析では、30年のリスク継続年数を、一律2年ごとの15個の順序カテゴリーを作成したが、最初の数年は1年ごととし、その後、2年またはそれ以上の括りでカテゴリーを作成することによって、うまく曲線をとらえ、モデルの改善につながる可能性が示唆された。

表7 離散時間ロジットモデル (Model4 : 結婚1年目ダミー) の結果

	B	標準誤差	E p (B)
リスクの継続年数	- 140	021	869
結婚1年目ダミー	- 875	435	417
最終学歴 (対: 高等学校)			
中学校	- 387	191	679
短期大学・高等専門学校	325	195	1 384
大学・大学院	- 415	426	660
結婚時の配偶者の職業 (対: 大企業・団体の正社員)			
中小企業・団体の正社員	847	197	2 333
農林漁業の自営業	080	459	1 084
農林漁業以外の自営業	1 200	249	3 320
臨時雇い	1 802	338	6 063
無職	961	535	2 613
無回答	1 058	732	2 879
本人の親との同居 (対: 非同居)			
同居	628	230	1 874
不明	1 109	525	3 033
配偶者の親との同居 (対: 非同居)			
同居	- 988	238	372
不明	511	371	1 668
本人の職業 (対: 無職)			
正社員	660	182	1 934
自営・家族従業	273	294	1 314
臨時雇い	503	230	1 654
不明	659	344	1 933
結婚時の本人の年齢	- 086	024	918
定数	-3 827	622	022

p 10 p 05 p 01 p 001

(イ) セミパラメトリックな等比ハザードモデル (時間区分別定率モデル)⁽⁶⁾

次に、セミパラメトリックな等比ハザードモデルである時間区分別定率モデル (Piece wise constant rate model) に移る。離散時間ロジットモデルに対して、時間区分別定率モデルは、連続する時間単位に基づいて観察されていると仮定する。しかしながら、時間単位が細くなればなるほど、ある時点 t でイベントが起こる確率はほぼ0となってしまう。そこで、注目する出来事の起りやすさが時間とともに変化するとし、そのイベントの生起確率密度としてハザード率を表す。離散時間ロジットモデルと同様に、このハザード率には、ある時点 t 以前にイベントが起こらなかったという条件が付いているが、離散時間ロジットモデルとは異なり、推定された各変数の回帰係数の e_p をとることにより、ある状態の人の出来事が起こる確率が何倍であるというような解釈ができる。このことは、ハザード率と説明変数の関連を式で表すと分かりやすい。離散時間ロジットモデルの式 (1) では、説明変数が0である基底状態に比べて、 X_k が1単位増えるごとに、ハザード率のオッズが e_p (回帰係数) 倍となっているが、等比ハザードモデルの式 (2) では、基底状態に比べて、 X_{ij} が1単位増えるごとに、ハザード率が e_p (回帰係数) 倍と読むことができる。

$$\frac{h(t_j|X)}{\frac{1-h(t_j|X)}{h_0(t_j)}} = \exp\left(\sum_k b_k X_k\right) \quad (1)$$

$$h(t|X_i) = h_0(t) \exp\left(\sum_j b_j X_{ij}(t)\right) \quad (2)$$

分析結果を報告する際に、「オッズ比」に比べて、よりなじみのある「率」で説明できるという点において、等比ハザードモデルの方が、より説得力があると言えよう。では、離散時間ロジットモデル (Model 1 : ベースモデル) とまったく同じ説明変数を使って、時間区分別定率モデル (Model 5 : ベースモデル) を分析してみよう。このモデルは、SPSS において一般化線形モデル (Generalized Linear Model) を使用し、観測時間が連続していると仮定してパラメータ推定する。一般化線形モデルでは、分析者が説明変数と従属変数の期待値の関連性 (連結関数: link function) や誤差のばらつき方 (誤差構造: error structure) を指定できる。例えば、重回帰分析であれば、説明変数と従属変数は線形 (identity link) で、誤差は正規分布 (normal) すると設定する。ロジスティック回帰であれば、連結関数は logit link で、誤差は二項分布 (binomial) と設定する。つまり、一般化線形モデルを用いて、連結関数と誤差の構造をロジスティック回帰のように設定すると、Model 1 とまったく同じパラメータが推定される。ただし、時間区分別定率モデルでは説明変数と従属変数の期待値の関連性は complementary log-log link であると仮定しているので、連結関数の設定を変更する。このようにして、得られた結果は表 8 のとおりであり、パラメータ推定値は離散時間ロジットモデルのそれとほとんど変わらない。しかしこの場合、オッズ比ではなくハザード率として解釈することができる。

統計セミナー後、今回の分析で、離散時間ロジットモデルと時間区分別定率モデルのパラメータ推定値になぜあまり差異がないのか調べを進めたところ、logit link と complementary log-log link の2つの連結関数の分布に関係していることが明らかになった。logit link はハザード率を $\log(\text{ハザード率} / (1 - \text{ハザード率}))$ と上限・下限のない数値に変換してパラメータ推定を行うが、complementary log-log link は $\log(-\log(1 - \text{ハザード率}))$ として上限・下限のない数値に変換する。軸をハザード率、縦軸をハザード率の変換値としてグラフにすると、図 1 のようになる。2つの連結関数は、ハザード率 25% あたりまでの低い値の場合は、ほとんど同じ値をとるが、ハザード率が高くなると差異が顕著となる。今回分析に用いたサンプル全体での離婚率は 6% 程度だったため、各時点における離婚発生率はかなり低いことが予想される。そのために、どちらの連結関数を用いたとしても、非常に近い値になったと結

論できる。また、解釈上の注意すべき点として、logit lin の場合は曲線が対称であるため、回帰係数の正と負を逆転させることによって、イベントが起こらない見込みとして解釈できるが、complementary log-log lin の場合は曲線が非対称であるため、回帰係数の正と負を単に逆転しても、イベントが起こらない確率とはならない。

表 8 時間区分別定率モデル (Model5 : ベースモデル) の結果

	B	標準誤差	Exp(B)
リスクの継続年数			
(対: 結婚1~2年目)			
3~4年目	570	370	1.768
5~6年目	569	376	1.766
7~8年目	477	383	1.611
9~10年目	531	381	1.701
11~12年目	-113	426	0.893
13~14年目	-354	453	0.702
15~16年目	-331	453	0.718
17~18年目	-520	479	0.594
19~20年目	-941	552	0.390
21~22年目	-320	465	0.726
23~24年目	-1.081	596	0.339
25~26年目	-795	552	0.452
27~28年目	-1.219	662	0.295
29~30年目	-2.232	1.051	0.107
最終学歴			
(対: 高等学校)			
中学校	-387	190	0.679
短期大学・高等専門学校	311	195	1.365
大学・大学院	-438	425	0.645
結婚時の配偶者の職業			
(対: 大企業・団体の正社員)			
中小企業・団体の正社員	837	197	2.310
農林漁業の自営業	088	458	1.091
農林漁業以外の自営業	1.195	248	3.304
臨時雇い	1.793	336	6.009
無職	938	532	2.555
無回答	1.061	730	2.889
本人の親との同居			
(対: 非同居)			
同居	604	230	1.829
不明	1.069	521	2.912
非該当	-135	533	0.874
配偶者の親との同居			
(対: 非同居)			
同居	-1.001	237	0.367
不明	466	370	1.594
本人の職業			
(対: 無職)			
正社員	783	185	2.187
自営・家族従業	318	294	1.375
臨時雇い	562	231	1.754
不明	689	343	1.992
結婚時の本人の年齢	-0.87	0.24	0.917
定数	-4.671	685	0.09

p 10 p 05 p 01 p 001

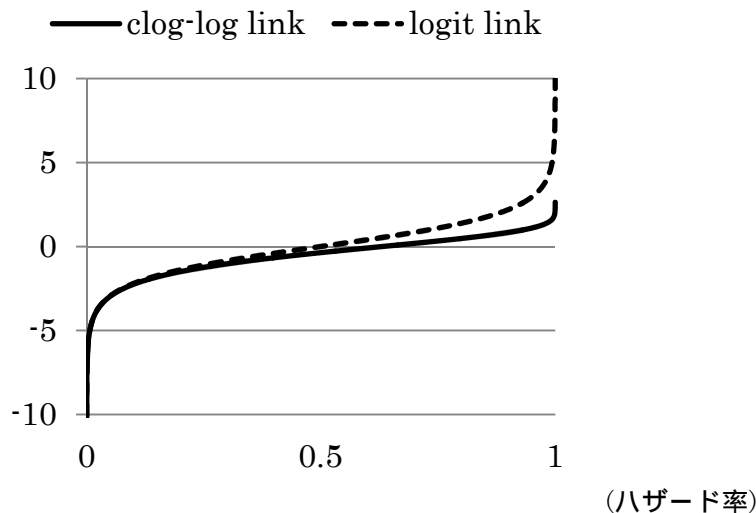


図1 ハザード率とハザード率変換値の曲線

時間区分別定率モデルの応用例として、交互作用項を加えるモデルを検討してみよう。最初に注目する交互作用は、夫婦の職業の違いである。この差が離婚率に関連すると予測して、結婚時の夫の職業と本人の職業の交互作用項をモデルに投入した (Model 6)。カテゴリ変数の交互作用項を作成するにあたっては、一方の変数をダミー変数に変換し、もう片方の変数の参照カテゴリーを 0 と変換し、2 つの変数を掛け合わせる方法を取った。このとき、複数のカテゴリーの値は連続している必要はない。この分析では、妻が有職であったかどうかのダミー変数を作成し、夫の職業が大企業・団体の正社員だった場合を参照カテゴリー (0) に変換し、2 つの変数を掛け合わせた。この作業により、夫の職業カテゴリーと同じ数の交互作用項カテゴリーが作成され、夫の職業が大企業・団体の正社員だった場合と本人が無職であった場合がそれぞれ参照カテゴリーとなる。つまり、交互作用項のカテゴリーの 1 つが有意であったとすると、妻が無職のケースにおいて、夫が大企業・団体の正社員だった場合に比べて、夫が当該の職業に就いていた場合は、1 年後離婚する確率が $e^{p(\cdot)}$ 倍となる。しかし、このモデルでは交互作用項は有意ではなかった。

次に注目した交互作用は、大学卒であることと結婚年齢の関係である。ベースとなる Model 5 (表 8) では、結婚年齢が上がるほど、離婚リスクが下がるという結果が出たが、結婚年齢の効果は、最終学歴によって異なるのではないかという仮説を検証するために、大学卒ダミーと結婚年齢の交互作用項を投入した。この交互作用項は有意であった (表 9)。結婚年齢の回帰係数から、最終学歴が大学未満のものは、結婚年齢が 1 年上がることによって離婚率が $e^{p(-0.79)} = 0.92$ 倍となる。最終学歴が大学以上のものは、結婚が 1 年遅くなることによって離婚率が $e^{p(-0.79)} = (-0.507) = 0.56$ 倍となる。つまり、遅く結婚するメリットは大学卒業者にとってより大きいという解釈が成り立つ。最終学歴が大学以上のカテゴリーの回帰係数が非常に大きく (11.883)、不自然に感じるかもしれないが、これは、大学卒ダミーと結婚年齢の交互作用項をモデルに投入したため、結婚年齢が 0 歳の場合の大学卒者の回帰係数ということになる。つまり、大学卒者の結婚年齢から回帰係数を計算するには、 $11.883 + \text{結婚年齢} \times (-0.507)$ の計算式が成り立ち、30 歳で結婚する大学卒者は同年齢で結婚する高卒者に比べて離婚率が $e^{p(-3.327)} = 0.04$ 倍であると解釈できる。

表9 時間区分別定率モデル (Model7 : 本人学歴と結婚年齢交互作用) の結果

	B	標準誤差	E p (B)
リスクの継続年数			
(対 : 結婚 1 ~ 2 年目)			
3 ~ 4 年目	571	370	1 770
5 ~ 6 年目	575	376	1 777
7 ~ 8 年目	484	383	1 623
9 ~ 10 年目	538	381	1 713
11 ~ 12 年目	- 107	426	899
13 ~ 14 年目	- 350	453	705
15 ~ 16 年目	- 329	452	720
17 ~ 18 年目	- 516	479	597
19 ~ 20 年目	- 938	552	391
21 ~ 22 年目	- 318	465	728
23 ~ 24 年目	-1 080	596	340
25 ~ 26 年目	- 802	552	448
27 ~ 28 年目	-1 219	662	296
29 ~ 30 年目	-2 242	1 051	106
最終学歴			
(対 : 高等学校)			
中学校	- 366	190	694
短期大学・高等専門学校	296	195	1 344
大学・大学院	11 883	4 349	144784 253
結婚時の配偶者の職業			
(対 : 大企業・団体の正社員)			
中小企業・団体の正社員	822	197	2 275
農林漁業の自営業	069	458	1 071
農林漁業以外の自営業	1 182	248	3 261
臨時雇い	1 654	347	5 228
無職	932	533	2 540
無回答	1 041	729	2 832
本人の親との同居			
(対 : 非同居)			
同居	617	230	1 853
不明	1 111	520	3 037
非該当	- 124	533	883
配偶者の親との同居			
(対 : 非同居)			
同居	- 998	237	369
不明	476	371	1 610
本人の職業			
(対 : 無職)			
正社員	801	185	2 228
自営・家族従業	341	294	1 406
臨時雇い	546	231	1 726
不明	697	342	2 008
結婚時の本人の年齢	- 079	024	924
大卒 x 結婚年齢	- 507	187	602
定数	-4 866	685	008

p 10 p 05 p 01 p 001

(ウ) コックス回帰モデル⁽⁷⁾

つづいて、コックス回帰モデルに移る。コックス回帰モデルの最大の特徴は、これまでに紹介したモデルと異なる推定方法を用いてパラメータ推定を行う点である。離散時間ロジットモデルと時間区分別定率モデルでは、手持ちの観測データから最も尤もらしいパラメータ値を推定する最尤推定法 (Maximum likelihood estimation) が用いられているが、コックス回帰モデルでは、部分最尤推定法 (Partial likelihood estimation) が採用されている。最尤推定法を用いたイベントヒストリー分析では、ある状態の人はイベントの起こる確率が高いか低いかという推定と、イベントがいつ起こりやすいの

かという情報を利用している。一方、部分最尤推定法では、相対的なリスク比は考慮するが、いつイベントが起こりやすいのかという情報は厳密でなくてもよい。つまり、注目する出来事が起こるリスクがいつから始まったのかという時間の特定ができない場合でも、ある状態の人が先にイベントを経験したかどうかという相対的な順位が特定できるのであれば、分析に用いることができる。

このように、コックス回帰モデルは、より少ない制約で、これまでに紹介したイベントヒストリーモデルと同じように、時間により変化する説明変数やセンサーされた観察値を偏りなく利用できるという画期的な手法である。ただし、サンプルサイズが小さい場合、注目する出来事が同時に起こり、生起順位が決められないケースが多い場合、または説明変数の影響力が時間依存する場合などは、パラメータ推定値の信頼性に問題のあることが指摘されている。コックス回帰モデルのもうひとつの特徴は、少なくとも SPSS においては、Person-period データではなく Person-level を用いることができるため、データファイルの変換作業を必要としないことである。ただし、時間とともに変化する説明変数を投入する場合は、その設定を行うために、データファイルの変換と同程度の作業が必要となる。

これまでに紹介した、離散時間ロジットのベースモデル (Model 1) ならびに時間区分別定率のベースモデル (Model 5) の場合と対応する説明変数を用いて行ったコックス回帰モデル (Model 8: ベースモデル) の結果が、表 10 である。ここで、「同一」ではなく「対応する」と述べたのは、コックス回帰モデルにおいては、上記のようにリスクの時間的变化には関心がなく特定できないために、リスクの継続時間の主効果をモデルに投入できないからである。しかしながら、有意となる説明変数はほぼ同じで、関連の方向性も一致しており、基本的に結論は同じである。しかし、ハザード率の推定値は、Model 5 と比べると、少し開きがある。コックス回帰モデルでは、本人の親と同居している者は別居している者に比べて、1年後の離婚ハザード率は 2.20 倍であるが、時間区分別定率モデルでは 1.83 倍である。また、配偶者の親と同居している者は別居している者に比べて、1年後の離婚ハザード率はコックス回帰モデルでは 0.05 倍、時間区分別定率モデルでは 0.37 倍である。カテゴリーに少し差異があるとは言え、推定値には明らかな相違が見て取れる。

表 10 コックス回帰モデル (Model 8: ベースモデル) の結果

	B	標準誤差	E p (B)
最終学歴 (対: 高等学校)			
中学校	- 313	180	731
短期大学・高等専門学校	262	191	1 300
大学・大学院	- 348	396	706
結婚時の配偶者の職業 (対: 大企業・団体の正社員)			
中小企業・団体の正社員	707	189	2 027
農林漁業の自営業	560	386	1 751
農林漁業以外の自営業	1 208	238	3 348
臨時雇い	1 585	327	4 877
無職	737	528	2 089
無回答	871	727	2 389
本人の親との同居 (対: 非同居)	790	187	2 203
配偶者の親との同居 (対: 非同居)	-2 922	508	054
本人の職業 (対: 無職)			
正社員	1 255	169	3 507
自営・家族従業	421	286	1 523
臨時雇い	980	212	2 665
不明	2 585	516	13 259
結婚時の本人の年齢	- 095	023	909

p 10 p 05 p 01 p 001

6. まとめ

本稿では、2008年夏のJGSS統計分析セミナーにおいて取り上げられた分析手法のうち、イベントヒストリー分析について、詳述している。セミナーでは、日本家族社会学会全国家族調査委員会が実施した全国調査「戦後日本の家族の歩み」のデータを基に、離婚の生起に影響を与える要因について、3つのタイプのイベントヒストリー分析法が紹介された。統計セミナーでは、セミナーの参加者が講師とともに実践するという演習形式を取っており、モデルを精緻化することよりも、シンプルなモデルを用いた学習に主眼が置かれていた。したがって、離婚の生起に影響を与える要因としては、セミナー(本稿)で組み込まなかった要因が多数存在しうる。そのため、分析結果そのものについて、詳細に議論することはあまり意味を持たないが、興味深い結果も得られた。

例えば、妻自身の親と同居すると離婚率が高くなるが、夫の親との同居は離婚率を下げる。額面どおりに解釈すると、離婚を防ぐには夫の親と同居するのが望ましいとなるが、実際は、そういった家族形態の女性は不幸せな結婚であっても離婚できないのかもしれない。また、中卒者の離婚率が高卒者に比べて有意に低かったが、中卒は中高齢者に多く、中高齢者の離婚率の低さを鑑みると、出生コーホートの効果があると予想される。

離散時間ロジットモデル、時間区分別定率モデル、そして、コックス回帰モデルのそれぞれのベースモデルの結果からも見て取れるように、最終的な結論はほとんど変わらない。しかし、解釈の仕方に違いが生じる。離散時間ロジットモデルでは、オッズを比較するのに対して、他の2つのモデルでは、ハザード率を比較する。また、コックス回帰モデルは、他の2つのモデルと異なる推定法を用いており、同時に注目するイベントが生起した場合は、近似式でパラメータ推定することになる。本稿のように観察時間の単位が年である場合は、離婚の同時発生のケースが増えたことによって、同じ説明変数を投入したにもかかわらず、時間区分別定率モデルのハザード率推定と差が生じたと考えられる。それぞれのモデルには特徴があり、分析に用いるデータに適したモデルを見極める必要がある。

JGSS研究センターでは、2009年に28~42歳を対象としてライフコース調査を実施する。この調査では、教育歴、職歴、結婚歴、同居歴など多岐にわたるライフヒストリー・データを収集するため、イベントヒストリー分析を行うのに適したデータと言える。本稿がきっかけとなり、イベントヒストリー分析の利用が増え、さらに新たな利用法に繋がれば幸いである。

[謝辞]

データは日本家族社会学会全国家族調査委員会によって行われた全国調査「戦後日本の家族の歩み」(J-S01)を許可を得て使用した。

[注]

- (1) 統計セミナーでは、山口一男先生が2001年から2002年にかけて、15回にわたり『統計』に連載した「イベントヒストリー分析」を資料とした。以下は、それぞれの文献情報である。

山口一男 2001 「イベントヒストリー分析(1)」『統計』9月号:74-79
 山口一男 2001 「イベントヒストリー分析(2)」『統計』10月号:70-75
 山口一男 2001 「イベントヒストリー分析(3)」『統計』11月号:73-78
 山口一男 2001 「イベントヒストリー分析(4)」『統計』12月号:73-78
 山口一男 2002 「イベントヒストリー分析(5)」『統計』1月号:68-73
 山口一男 2002 「イベントヒストリー分析(6)」『統計』2月号:65-70
 山口一男 2002 「イベントヒストリー分析(7)」『統計』3月号:69-74
 山口一男 2002 「イベントヒストリー分析(8)」『統計』4月号:73-78
 山口一男 2002 「イベントヒストリー分析(9)」『統計』5月号:55-60
 山口一男 2002 「イベントヒストリー分析(10)」『統計』6月号:69-74
 山口一男 2002 「イベントヒストリー分析(11)」『統計』7月号:58-63

山口一男 2002 「イベントヒストリー分析(12)」『統計』8月号:63-68

山口一男 2002 「イベントヒストリー分析(13)」『統計』9月号:60-65

山口一男 2002 「イベントヒストリー分析(14)」『統計』10月号:66-71

山口一男 2002 「イベントヒストリー分析(最終回)」『統計』11月号:55-60

- (2) Jeroen Janssen が開発した無料統計解析ソフトで、下記のウェブサイトからダウンロードできる。LEM の説明は都村ほか(2008)を参照されたい。
- (3) コックス回帰モデルにおいては、リスク開始時間の特定は必要ではない。
- (4) 統計セミナーでは、大阪商業大学の保田時男(総合経営学科専任講師)が SPSS を用いた Person-period データの作成方法を解説した。
- (5) この節の記述は、山口(2002 1月号)に詳しい。
- (6) この節の記述は、山口(2002 2月号)に詳しい。
- (7) この節の記述は、山口(2002 3月号)に詳しい。

[参考文献]

- 安藏伸治 2003 「離婚とその要因 - わが国における離婚に関する要因分析 - 」『JGSS 研究論文集』2 25-45
- Boslaugh Sarah 2004 An Intermediate Guide to SPSS Programming Using Syntax for Data Management thousand Oaks Sage
- 加藤彰彦 2005 「離婚の要因 - 家族構造・社会構造・経済成長」熊谷苑子・大久保孝治編『コーホート比較による戦後日本の家族変動の研究』日本家族社会学会全国家族調査委員会 pp77-90
- 都村聞人・岩井紀子・保田時男・穴戸邦章 2008 「JGSS-2005 を用いた通信機器利用の潜在クラスモデル - 統計セミナーにおける適用例 - 」『JGSS 研究論文集』7 233-249
- Maguire John 1991 Event Historical Analysis for Parallels Sage