

Title	大学教員の流動性究明に向けた統計学的アプローチ
Author(s)	細坪, 護孝
Citation	年次学術大会講演要旨集, 24: 764-769
Issue Date	2009-10-24
Type	Conference Paper
Text version	publisher
URL	http://hdl.handle.net/10119/8739
Rights	本著作物は研究・技術計画学会の許可のもとに掲載するものです。This material is posted here with permission of the Japan Society for Science Policy and Research Management.
Description	一般講演要旨

大学教員の流動性究明に向けた統計学的アプローチ

○細坪 護孝（文部科学省 科学技術政策研究所）

全国大学職員録（廣潤社）の掲載情報を活用して、大学教員の大学間異動者の特性値を探る試みを実施した^[1]。その後、データベースを拡張し、現在まで 1988 年から 2006 年までの 19 年間 8 時点（私立大学は 7 時点）の国公立大学教員パネルデータを構築した。

今般、そのデータを活用して、異動が及ぼす影響、逆に何が異動に影響を与えるかなどを解明するため、統計学的に分析を行った。具体的には、異動に関連すると思われる各種データを質的・量的データに分類し、以下の分析を実施した。

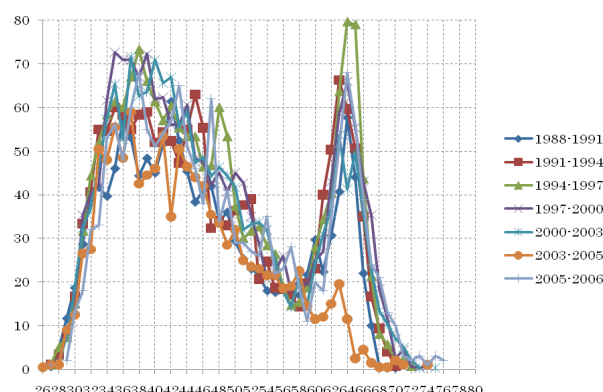
- (1) クロス表やその期待度数などを用いたカテゴリカルデータ分析
- (2) クロス表の度数がどのような因子から影響を受けているのかを調べる対数線形モデル分析
- (3) 異動を2値質的従属変数と見なし、それに影響を及ぼす変数を調べるロジスティック回帰分析
- (4) (3)と同じく、異動を2値の質的従属変数と見なしたパネルデータ分析

0.2006 年度の本学会発表の延長

過去の発表^[1]では、基本的には大学間異動経験教員（以下「異動者」という）と大学間異動未経験教員（以下「不異動者」という）それぞれの属性値の度数分布を描き、それを直接比較するという方法だった。しかし、数々の属性値（時間依存性を持つ変数を含む）が得られると、次はそれらが互いにどのような関係があるのか、例えばどのようなシステムを構築しているのか、を調べない限り、どんな因子が異動に効いているのか、むしろ異動が他の因子に効いているのかなどが分からない。つまり、静的に(static)2 又は 3 次元グラフを数十個組み合わせて並べても、人の認識には限界があり、それから特徴的な動向を科学的かつ実証的に把握することは不可能に近い。逆説的に言えば、観測者の都合のよいデータだけ見れば何でも言えてしまう。しかし、それは事実ではないことは言を待たない。

例えば図表1では年齢別の異動者数の時間変化を図示したものである。図表1からは、大学教員異動数は二山構造となっていることが分かる。また、このデータを分析すると、異動教員の平均年齢が低下していることも分かる。しかし、この他にも教員には様々な属性値があり、その異動者数が年齢だけに依存しているとは言いきれない。

また、この図表1の解釈は比較的分かりやすいが、



図表1 年齢別異動者数の時間変化(2年間変化に補正済)

科研費取得数や昇格回数と異動回数との関係を図示しても解釈が難しくなり、因果関係の問題も発生する。その解決へのアプローチを試みるため、カテゴリカル・データに対するいくつかの統計的手法の導入を試みる。

1. 統計学的手法を用いる前の準備

今回の調査では、職員録の掲載データを全て用いたため、標本数は母集団の規模に近く、非標本誤差はもとと比較的小さいと考えられる。一方、今回は異動を主な対象としていることから、特に異動者と不異動者間の非標本誤差は可能な限り避けたい。異動者・不異動者間で生じる非標本誤差の代表的なものが、観測時間即ち、接続時点数である。もし異動が時間的な確率変数ならば、観測時間が長くなればそれに応じて異動教員の割合は増える。逆に、観測時間を短くすればそれに応じて異動する教員数は減る。極端な話、観測が一時点であれば異動を観測できないため、全員が不異動者となる。

接続 時点 数	合計	異動者数		未調整 不異動者数		調整済 不異動者数		異動者数・ 調整済不 異動者数 の母比率 の差の等 計量 Z	左に対応 する 正規分布 の累積分 布関数
2	1076	429	1.79%	8163	13.84%	647	1.78%	0.052	0.5207
3	3973	1565	6.53%	9770	16.57%	2408	6.64%	-0.536	0.2961
4	7394	2937	12.25%	9345	15.85%	4457	12.28%	-0.134	0.4466
5	9508	3781	15.76%	8097	13.73%	5727	15.78%	-0.057	0.4771
6	11131	4427	18.46%	6944	11.77%	6704	18.47%	-0.05	0.4799
7	6983	2771	11.55%	4525	7.67%	4212	11.61%	-0.201	0.4202
8	20207	8074	33.66%	12133	20.57%	12133	33.44%	0.5824	0.7199
総計	60272	23984	100.00%	58977	100.00%	36288	100.00%		

図表2 接続時点別異動者数と調整済不異動者数

以上から、標本数の大きい不異動者の接続時点別割合を比較的標本数の小さな異動者のそれから大きく離

れないように調整すればよいことが分かる。ここでは、中心極限定理により標本平均が正規分布で近似できることを活用した母比率の差の検定を使用して、不異動者数を調整した(図表 2)。

以下の分析では主にこの異動者数と調整済不異動者数を使用する。

2. クロス表などを用いたカテゴリカルデータ分析

観測される統計量のうち、教員の年齢や世代は量的変量だが、職位や学位、異動の有無などは質的変量といえる。これらには順序は付けられるが、(2 値変数以外)その間隔が等しいとは厳密には言えない。そのため、数値化計算には限界がある。更に、専門分野となれば順序すら付けることは容易ではなく、名義尺度である。

質的変量进行分析するための統計的手法には様々あり、数量化Ⅲ類などが提案されてきた。ここでは、対数線形モデルによる分析の下準備も兼ねて、2×2 のクロス表による分析を行う^[2]。

最も簡単な仮定として、教員の異動の有無はその研究活動の優秀さを表しているとしよう。それと同時に、科研費取得と昇格の有無は教員の研究活動の優秀さを表しているとも仮定する。いずれにしても、「教員の研究活動の優秀さ」を直接計測できないための苦肉の策である。上記の仮定が成り立つとすれば、教員の異動の有無と科研費取得、昇格の有無は密接な関係があるはずである。これら3つの変量をクロス表にまとめたのが図表 3 である。このクロス表では行:異動、列:科研費、層:昇格と置いている。図表3から、行列層を入れ替えたクロス表はあと2セットできる。

	昇格なし			昇格あり		
	科研費なし	科研費あり	総計	科研費なし	科研費あり	総計
異動なし	4251	8907	13158	5088	18042	23130
異動あり	2609	4565	7174	3018	13792	16810
総計	6860	13472	20332	8106	31834	39940

図表 3 異動の有無に対する科研費、昇格の有無

最初に、周辺度数から求めた各セルの期待度数 E_i と実際の度数 O_i との差からカイ 2 乗分布を活用した独立性の検定((1)式)を行った。

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i} \dots\dots\dots(1)$$

その結果、ほとんどの場合では独立性の帰無仮説は棄却された(0.1%有意)が、ただ一つ、科研費なしの場合における異動と昇格の関係では独立性の仮説は棄却されなかった。同じケースにフィッシャーの正確確率検定(Fisher's exact test)を行っても、同様に独立性の仮説は棄却されなかった(有意確率 0.318)。このことから、科研費をとってない教員群では、異動と昇格は独立であるといえる。

このことを念頭に置いて、再び図表 3 のクロス表を見る。昇格の有無即ち層別にクロス表を比較すると、昇格あり

で科研費なしの教員合計数は昇格なしのその 1.182 倍(=8106/6860)である一方、昇格ありで科研費ありの教員合計数は昇格なしのその 2.363 倍(=31834/13472)である。これを科研費の有無で比べるとちょうど2倍の差がある。これは昇格ありの方が昇格なしの方より科研費ありの割合が 2 倍高いということである。しかしこれは昇格したから科研費が採択されるようになったのか、科研費が採択されたから昇格したのかは分からない。いずれにせよ、科研費と昇格には比較的密接な関係がありそうだ。一方、異動に関しては双方にそのような強い関係は見られない。

ここでは、異動と科研費と昇格の 3 つの因子を対象に独立性検定等を行ったが、データがあれば他にも調べられる。平成 20 年度第三期科学技術基本計画フォローアップ調査では国立大学を 3 群に分けてカテゴリー化し、異動元・異動先大学で 3×3 のクロス表を構築し、独立性検定^[3]を行った。

3. 対数線形モデル分析

2 節の独立性検定では、一部のクロス表で独立性が(弱い意味で)検出されたものの、ほとんどのケースでは独立性の帰無仮説は棄却された。

つまり、この 3 因子の場合でも、ほとんどのセルの度数は独立、即ち互いの因子の単純積ではなく、「交互作用的因子」が存在すると考えられる。

それを実際に計算するために考え出された方法が対数線形モデルである。まず、あるセルの期待度数を μ_{ij} 、i 番目の行の確率を p_i 、j 番目の列の確率を p_j 、総度数を N とすると、次のように表現できる。

$$\mu_{ij} = N \times p_i \times p_j \dots\dots\dots(2)$$

(2)式の両辺の自然対数をとると、

$$\begin{aligned} \ln \mu_{ij} &= \ln (N \times p_i \times p_j) = \ln N + \ln p_i + \ln p_j \\ &= \lambda + \lambda^{\text{行}}_i + \lambda^{\text{列}}_j \dots\dots\dots(3) \end{aligned}$$

($\lambda = \ln N$ 、 $\lambda^{\text{行}}_i = \ln p_i$ 、 $\lambda^{\text{列}}_j = \ln p_j$)

となる。しかし、実際の多くの場合、観測度数は期待度数と一致しない(≠独立でない)ため、(3)式に

$$\ln \mu_{ij} = \lambda + \lambda^{\text{行}}_i + \lambda^{\text{列}}_j + \lambda^{\text{行列}}_{ij} \dots\dots\dots(4)$$

と交互作用項を入れて期待度数とする。(4)式は飽和モデルであり、飽和モデルの母数の数は観察度数と常に一致する。

対数線形モデルの計算では、2 節で述べたクロス表を使う方法がある。この場合、3 因子なので、飽和モデル(1)、均一連関モデル(1)、条件付独立モデル(3)、一変数独立モデル(3)、独立モデル(1)の計 9 モデルがある。この中から、求められた期待度数が観察度数に当てはまりが最もよいモデル(例:AIC 又は BIC が最も小さなモデル)が採択される。

期待度数からの計算の結果(条件付独立、一変数独立、独立モデルのみ)、昇格に対して異動と科研費の有無が直交する条件付独立モデルが採択された。

これを式で表すと次のとおり。

$$\ln \mu_{ij} = \lambda + \lambda_{i \text{ 異動}} + \lambda_{j \text{ 科研費}} + \lambda_{k \text{ 昇格}} + \lambda_{ik \text{ 異動・昇格}} + \lambda_{jk \text{ 科研費・昇格}} + \dots \quad (5)$$

(5)式のλの値(ijkにより異なる)を手動で求めるのは非常に煩雑な計算になるそうなので、ここから R 言語に頼ることにする。同じデータに対して R で計算すると、周辺度数の手計算から容易に求められない飽和モデル、均一連関モデルに次いで、上記の条件付モデルが最適と判断された。最も適切と判断された飽和モデルの係数(図表 4)を見ると、異動・昇格・科研費のうち、異動に関する項は総じて比較的係数が小さい。飽和モデルが最適となっても、全ての因子が同様に効いている保証にはならない。

0 次項(切片) 8.702812		
一次項(1 の場合。0 のときは逆符号) 異動 昇格 科研費		
-0.24341	0.267093	0.510513
二次項(0×0、1×1 の場合。 0×1、1×0 のときは逆符号) 異動×昇格 異動×科研費 科研費×昇格		
0.045706	0.009171	0.185765
三次項(0×0、1×1 の場合。 0×1、1×0 のときは逆符号) 科研費 = 0 異動×昇格 -0.054234 科研費 = 1 異動×昇格 0.054234		

図表 4 異動×昇格×科研費の 3 因子対数線形モデルの係数(N=60273)

上記の分析結果、即ち、異動は科研費と昇格ほどに強い関係を有していないことは、2節のクロス表分析で得られた結果と類似している。

対数線形モデルでは、説明変数が2値でなくても計算できる。即ち、一定期間中の異動の有無ではなく異動回数、昇格の有無ではなく昇格回数、科研費取得経験の有無ではなく科研費取得数を変量としても計算できる。更に、詳細に分析するため、4 次、5 次、6 次と因子を追加していくこともできる。しかし、網羅的に飽和モデルから独立モデルまで計算する方法では、6 次になると変量間の組み合わせ数が多くなり、結果として最近の計算機でも実行が大変な数になる(図表 5)。7 次元では網羅的な式数は 1,000 億程度と推測される。

一方、対数線形モデルの計算では、変数減少法的な一種の簡便法がある。その方法で解くと一気に式数が減少する(図表 5)上、この場合は※、4、5、6 次元の全て場合で網羅計算による最適解と一致した。

次に、設定したモデルと得られた解の解釈に移る。上記の4、5、6次元のいずれも、最適と導出された解の最高次数は3次である。しかし、図表4の場合とは異なり、実際に係数を見ようとすると、それぞれの変量の取りうる値に応じてλが決まるため、解は長大になる。そこで、解の最高次つまり3次への動向のみに着目して、図示した(図表 6)。この方法では係数の大小は分からないが、高次の対数線形モデルの意味を簡便かつイメージ

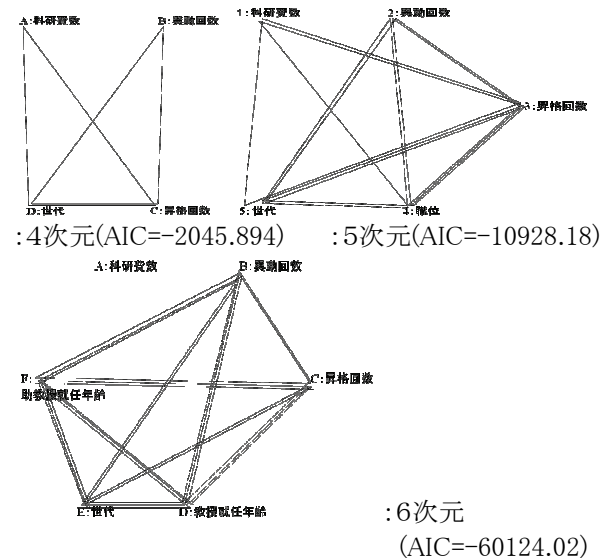
※ この対数線形モデルのデータセットは厳密には1節のものと異なるが、性格はほぼ同じであり、概ね同一のものと見て差し支えない。

	網羅計算	変数減少法 (今回の場合、AICが増加に転じたことを確認する式含む)	グラフィカル対数線形モデリング (今回の場合、AICが増加に転じたことを確認する式含む)
4次元	80	11	12
5次元	2,079	56	28
6次元	1,114,174	316	76

図表 5 対数線形モデルで解を求めるまでの式数

モデル設定

- 4次元: 科研費数(0-8)、異動回数(0-5)、昇格回数(0-5)、世代(1-6)
 5次元: 科研費数(0-8)、異動回数(0-5)、昇格回数(0-5)、職位(1-4)、世代(1-6)
 6次元: 科研費数(0-8)、異動回数(0-5)、昇格回数(2-5)、教授就任年齢(2-6)、世代(2-6)、助教授就任年齢(2-6)

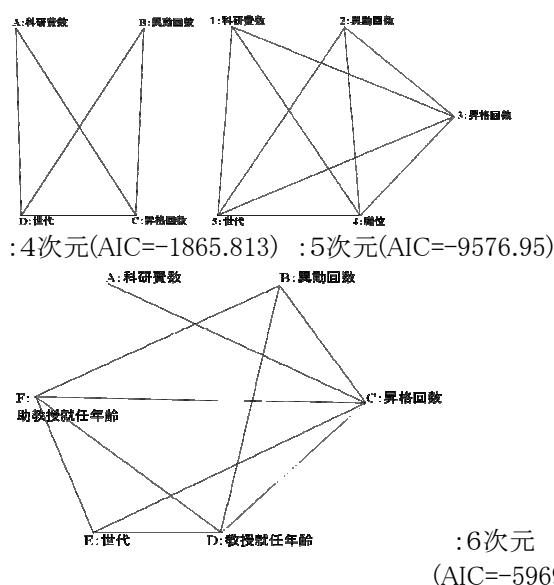


図表 6 4・5・6次元対数線形モデルの簡便近似図示(3 次項を構成変量間の三角形で示す)

を説明する近似的なものとお考えいただきたい。

図表 6 と得られた解から、異動回数と科研費数の連関はほとんどないと言える。これは、「直接的な」関係がないのであって、例えば、科研費が採択された教員が論文等を書いて公表し、外部から高い評価を得たら別のよりよい大学に異動する、といった媒介変量が間にあるため、この2変量間の関係が直接的には明白にならない可能性もある。もともと、両者に全く関係がない可能性も棄却できない。

このように、対数線形モデルは計算も解の解釈も容易ではないが、両者を見越して計算する方法の一つとしてグラフィカル対数線形モデリングがある(図表 7)^[4]。グラフィカル対数線形モデリングでは「計算の最初から」図表 6 のようなグラフを描くことを念頭に置き解を導出する。つまり、計算方法とグラフを対応させて考える。全てのグリッド間(=2変量間)を無向グラフで結んだものを飽和モデルとして(そのため均一連関モデルは考慮しない)、そこから AIC が最も低くなるように次々に辺を取り除いて



図表 7 グラフィカル対数線形モデリングによる無向グラフ

不飽和モデルを計算する。そのため、この方法では、最初から2次以上の項を削除していくため、高次のモデルでは最初から削除対象となる項が非常に多くなり計算が複雑になり、厳密解とは一致しないが、特に高次の場合には計算すべき式数が少なくなる(図表 5)。

興味深いことに、グラフィカル対数線形モデルから得られた結果と、簡便近似図示法とは似ている。但し、6次では異なる部分もある。

4.ロジスティック回帰分析

対数線形モデルの一つのバリエーションでもあるロジスティック回帰分析を活用して、本研究のモデルを調べてみよう。

具体的には、SPSS により、従属変数を異動の有無の2値変量として、説明変数を変数減少法で求めることとする。最初に投入する説明変数は次のとおり。

- (1) 昇格回数(0-5、間隔尺度)→そのまま投入
- (2) 科研費数(0-8、間隔尺度) →そのまま投入
- (3) 助教授就任年齢(比率尺度) →そのまま投入
- (4) 教授就任年齢(比率尺度) →そのまま投入
- (5) 学位(1-2、順序尺度)→ダミー変数化して投入
- (6) 世代(比率尺度) →そのまま投入
- (7) 最終職位(1-4、順序尺度)→ダミー変数化して投入
- (8) 接続時点数(2-8、名義尺度) →ダミー変数化して投入

この変量の中から、有意確率の高いものから順次一つずつ除去していき、最終的に 0.1%有意の変量のみが残れば、そのモデルが妥当であるとする。その計算結果が図表 8 である。

有意な変量として残ったのが、職位 3(助教授、負)、助教授就任年齢(正)、教授就任年齢(負)、世代(正)である。職位以外この結論は妥当である。助教授になるために異動するのであれば、助教授就任年齢と異動の有無とは正の相関があっておかしくない。逆に教授ではそれは成り立たない。若くして教授になった人ほど異動する

と考えると、感覚的には理解できる。つまり、同じ就任年齢でも教授と助教授では異動が(に対して)果たす役割が全く異なる。そして、世代は低い人ほど異動する。これは図表 1 の異動教員の平均年齢が下がっていることとも一致する。

	B	標準誤差	有意確率	Exp(B)	EXP(B) の 99.9% 信頼区間	
					下限	上限
最終職位 3 (助教授)	-.843	.220	.000	.431	.209	.888
助教授就任年齢	.072	.007	.000	1.075	1.050	1.101
教授就任年齢	-.100	.007	.000	.905	.884	.926
世代	.036	.002	.000	1.037	1.030	1.044

図表 8 従属変数を異動の有無としたロジスティック回帰分析の説明変量 I (N=8413、助教授・教授就任年齢含むモデル、0.1%有意)

しかし、8つの変量のうち本当にたった4つの変量しか異動の有無に影響を与えないのか。これは、全体的に年齢記載割合が低いこと、助教授就任年齢を計算するには、観測期間中に助手又は講師などから助教授に昇格していなければならないことから、この情報が判明している割合は全体の 3~4 割に過ぎないことに起因する。教授に関してもほぼ同じである。そしてその2つが同時に判明している割合は全体の 1~2 割程度であり、この割合は図表 8 の標本数に整合する。

そこで、前記の 8 種の変量のうち、標本数を減らす助教授・教授就任年齢を除く変量のモデルから計算する。すると、世代とダミー変量である接続時点数 3 以外の変量は有意になり、図表 9 のようになる。

	B	標準誤差	有意確率	Exp(B)	EXP(B) の 99.9% 信頼区間	
					下限	上限
最終職位 1(助手)	-.364	.056	.000	.695	.577	.836
最終職位 3(助教授)	.637	.043	.000	1.890	1.641	2.178
最終職位 4(教授)	1.130	.042	.000	3.095	2.696	3.553
昇格回数	.300	.013	.000	1.350	1.295	1.407
学位 2(博士)	-.076	.021	.000	.926	.863	.994
科研費数	.030	.004	.000	1.031	1.017	1.046
接続時点数 4	-.179	.039	.000	.836	.735	.951
接続時点数 5	-.356	.038	.000	.700	.618	.793
接続時点数 6	-.534	.038	.000	.586	.518	.664
接続時点数 7	-.702	.042	.000	.496	.431	.570
接続時点数 8	-.829	.038	.000	.437	.385	.495
定数	-1.011	.050	.000	.364		

図表 9 従属変数を異動の有無としたロジスティック回帰分析の説明変量 II (N=60272、助教授・教授就任年齢除外モデル、0.1%有意)

図表 9 から、異動に対して正の効果を有するのは、高い最終職位、昇格回数、そして僅かながら科研費数である。一方、負の効果を有するのは接続時点数と学位である。一見、接続時点数の増加につれて、異動しないモデルになっていることは奇妙である。おそらく、接続時点の増加よりも、職位の向上、つまり昇格の効果が上回っているのだろう。

また、図表 8 のケースと異なるようにも思われるが、早々に教授に昇格した研究者が異動するとも考えられる。しかし、そうであれば異動教員は比較的若いはずであり、学位が負に効くことの説明は難しい。

本研究で最も厄介なのは、異動には年齢依存性の高い2種類があることであり(図表 1)、これを一貫して解釈するモデルを求めることは非常に大変である。

そこで、この異動システムの変曲点と思われる 58 歳以下と 59 歳以上の年齢で標本全体を二分割し、母集団自体がそもそも2つであると考えことにする。この2つの母集団の標本は時間に応じて一方向に流れる特性を有するが、それぞれの母集団に属している期間中は、その属する母集団の特性に従うものとする。

	B	標準誤差	有意確率	Exp(B)	EXP(B) の 99.9% 信頼区間	
					下限	上限
助教授 就任年齢	.059	.007	.000	1.060	1.036	1.086
教授 就任年齢	-.093	.007	.000	.911	.890	.933
世代	.040	.003	.000	1.041	1.031	1.050

図表 10 従属変数を異動の有無としたロジスティック回帰分析の説明変量Ⅲ (N=7539、助教授・教授就任年齢含むモデル、58 歳以下、0.1%有意)

そのように仮定して、改めて 58 歳以下の教員の異動をロジスティック回帰で調べたのが図表 10,11 である。全体的に係数の絶対値が大きくなり、傾向の把握がしやすくなった。図表 10 から図表 8 とほぼ同じことが言えるが、58 歳以下では特に教授の異動が効く。

	B	標準誤差	有意確率	Exp(B)	EXP(B) の 99.9% 信頼区間	
					下限	上限
職位 1 (助手)	-.458	.060	.000	.633	.519	.771
職位 3 (助教授)	.668	.045	.000	1.951	1.682	2.264
職位 4 (教授)	1.252	.050	.000	3.498	2.968	4.124
昇格回数	.456	.018	.000	1.578	1.486	1.676
世代	.056	.002	.000	1.058	1.051	1.065
科研費数	.031	.005	.000	1.032	1.015	1.049
接続時点 数 3	-.574	.090	.000	.563	.419	.758
接続時点 数 4	-.743	.086	.000	.476	.358	.631
接続時点 数 5	-.922	.086	.000	.398	.300	.528
接続時点 数 6	-.992	.086	.000	.371	.279	.493
接続時点 数 7	-.976	.089	.000	.377	.281	.504
接続時点 数 8	-1.068	.087	.000	.344	.258	.458
定数	-4.204	.152	.000	.015		

図表 11 従属変数を異動の有無としたロジスティック回帰分析の説明変量Ⅳ (N=41206、助教授・教授就任年齢除外モデル、58 歳以下、0.1%有意)

また、図表 11 から、58 歳以下では異動と昇格には密接な関係がある。一方、ここでも接続時点数が大きいほど異動しないモデルになっており、多重共線性の疑いがある。

一方、59 歳以上ではどうか。まず、助教授・教授就任年齢を含むモデル(図表 12。この計算でダミー項を含めるとデータ数に比して求める項が多すぎて解が求めら

れないため、このみダミー項を使用しない。)では、接続時点数以外同じ項が残るが、世代効果の正負が逆転する(図表 10、図表 12)。つまり、58 歳以下では低い世代ほど異動するが、59 歳以上では高い世代ほど異動する。これが母集団を2つに分けた本質である。

助教授・教授就任年齢効果を除いた場合でも、この2つの集団の特性が反映される。58 歳以下では昇格回数の影響力が大きい、59 歳以上では最終職位の影響力が非常に大きい(図表 13)。

	B	標準誤差	有意確率	Exp(B)	EXP(B) の 99.9% 信頼区間	
					下限	上限
助教授 就任年齢	.107	.025	.000	1.112	1.024	1.208
教授 就任年齢	-.088	.023	.000	.916	.848	.989
世代	-.084	.023	.000	.920	.853	.991
接続時点 数	.404	.095	.000	1.498	1.094	2.050

図表 12 従属変数を異動の有無としたロジスティック回帰分析の説明変量Ⅴ (N=874、助教授・教授就任年齢含むモデル、59 歳以上、0.1%有意)

	B	標準誤差	有意確率	Exp(B)	EXP(B) の 99.9% 信頼区間	
					下限	上限
職位 4 (教授)	1.898	.104	.000	6.671	4.732	9.404
昇格 回数	.235	.033	.000	1.265	1.136	1.409
世代	-.390	.007	.000	.677	.661	.693
科研費数	.028	.008	.001	1.028	1.001	1.056
接続時点 数 3	1.049	.151	.000	2.856	1.736	4.697
接続時点 数 4	1.920	.147	.000	6.818	4.198	11.074
接続時点 数 5	2.903	.150	.000	18.221	11.110	29.884
接続時点 数 6	3.825	.158	.000	45.841	27.289	77.006
接続時点 数 7	4.130	.173	.000	62.162	35.225	109.70
接続時点 数 8	5.103	.169	.000	164.463	94.159	287.26
定数	9.449	.270	.000	12693.8		

図表 13 従属変数を異動の有無としたロジスティック回帰分析の説明変量Ⅵ (N=19066、助教授・教授就任年齢除外モデル、59 歳以上、0.1%有意)

5. パネルデータ分析

本研究対象となる職員録の掲載データは、異時点間で接続させており、教員×年のパネルデータを構成している。

0 節から 4 節までは各教員が有する変量をいわば静的(static)な教員固有の属性値として入力データとしてきたが、さらに異動の因果関係などを明らかにするためには、異動に関する変量の変量間の関係も考慮しつつ、その時間的変化なども考慮する必要がある。

以上から、本研究ではパネルデータ分析を実施する。パネルデータ分析の基本として、次の3つのモデルが考えられている。

$$(1) y_{it} = \alpha + x_{it} \beta + \varepsilon_{it}$$

pooling 推計: 一般最小二乗法

個体ごとに平均値も、誤差項の分散も同一である。基本的に pooling 推計は OLS になる。

$$(2) y_{it} = \alpha_i + x_{it}\beta + \varepsilon_{it}$$

within 推計: 固定効果モデル

個体ごとに切片項(α_i)が異なる。個体ごとの平均値を算出し、説明・被説明変数をその平均値との乖離に変換した後、OLSで推計を行う。

$$(3) y_{it} = \alpha + x_{it}\beta + \nu_i + \varepsilon_{it}$$

ランダム推計: 変量効果モデル

個体ごとに誤差項の分散が異なる。pooling 推計後に、個体ごとの誤差分散(ν_i)を計測し、それを除数として GLS を用いて計測する。

それぞれにデータを入力して統計量の推定を行った後、検定を実施して最適なモデルを決める。

以上から、異動の有無を2値従属変量とし、職位、学位、年齢、世代、科研費取得を説明変量としたモデルで分析を試みた^{*}。昇格は職位で説明されるものとし、このモデルでは省いている。説明変量のうち質的変量である職位、学位、科研費についてはダミー変数化した。具体的には、職位(1=助手、2=講師、3=助教授、4=教授)、学位(1=修士以下、2=博士)、科研費(0=なし、1=分担研究者、2=代表研究者)とした。

Rによる推定の結果、poolingとwithinは2時点以上の不完備パネルデータで計算できるが、ランダム推計では一定以上の時点がないと計算できない。そして、検定結果ではランダム推計が残るため、長期間(具体的には6時点以上)の不完備パネルデータを対象とする。

その結果、得られたランダム推計の結果が図表14である。

	推定量	標準誤差	t 値	Pr(> t)
切片	0.206	0.00838	24.5951	< 2.2E-16***
職位 1	-0.0576	0.00217	-26.5458	< 2.2E-16***
職位 3	0.0137	0.00191	7.1536	8.45E-13***
職位 4	0.0497	0.00211	23.6066	< 2.2E-16***
学位 2	0.00766	0.00108	7.1056	1.20E-12***
年齢	-0.00301	0.0000971	-30.9931	< 2.2E-16***
世代	-0.000273	0.0000938	-2.9072	0.003646**
科研費 0	-0.0123	0.00129	-9.5076	< 2.2E-16***
科研費 2	-0.0133	0.00146	-9.11	< 2.2E-16***

図表 14 従属変数を異動の有無としたパネルデータ分析の説明変量 I ($R^2=0.98845$, $n=38874$, $T=6-8$, $N=280954$, ***0.1%有意, **1%有意。以下同じ)

図表 14 から、異動に対して大きな正の影響を及ぼすのは職位や学位であり、科研費を取ると負の影響を及ぼす。また、年齢は若いほど異動するが、世代は高いほど異動する。この全体像の把握は難しい。

3節でも述べたが、これは異なる異動現象を持つ標本群を無理に一つのモデルに押し込んだために起きたものと考えられる。そこで教員を58歳以下と59歳以上に分割し、改めてそれぞれパネルデータ分析を実施した。

58歳以下の場合において、3つのモデルのうち再びランダム推計が採択されたが、時点数が7以上でないと推計できない。その推定結果が図表15である。

図表15と図表16を比べると、世代と科研費の効果が逆である点が興味深い。58歳以下では、世代が低いほど異動し、科研費の取得は異動に対して正の効果を及ぼす。一方、59歳以上の教員では、教員数が少なく、時点数が少ないためか、ランダム推計が完備パネルデ

ータでも不可能になる。そのため、検定の結果から pooling 推計よりはよい within 推計の結果を図表16に示す。

図表16から59歳以上の教員の異動は、職位が教授や、博士であれば多い一方、科研費の取得は負の効果を及ぼす。また世代と年齢は若ければ多い。

以上から、58歳以下と59歳以上の教員異動のシステムの違いがあり、その内容は4節のロジスティック回帰分析の結果と整合する。

	推定量	標準誤差	t 値	Pr(> t)
切片	0.278646	0.012511	22.273	< 2.2E-16***
職位 1	-0.0698	0.002735	-25.517	< 2.2E-16***
職位 3	0.031054	0.002412	12.8736	< 2.2E-16***
職位 4	0.074121	0.002755	26.8999	< 2.2E-16***
学位 2	0.014936	0.001425	10.4779	< 2.2E-16***
年齢	-0.00623	0.000136	-45.6842	< 2.2E-16***
世代	0.000588	0.00016	3.6806	0.0002327***
科研費 0	-0.01673	0.001687	-9.9199	< 2.2E-16***
科研費 2	-0.00902	0.001893	-4.7623	< 1.9E-06***

図表 15 従属変数を異動の有無としたパネルデータ分析の説明変量 II (58歳以下, $R^2=0.97273$, $n=20095$, $T=7-8$, $N=153962$)

	推定量	標準誤差	t 値	Pr(> t)
職位 1	-0.07474	0.079094	-0.9449	0.344704
職位 3	0.029925	0.05768	0.5188	0.603897
職位 4	0.259126	0.057563	4.5016	6.7E-06***
学位 2	0.061788	0.027128	2.2776	0.022749*
年齢	-0.002	0.000763	-2.6211	0.008764**
世代	-0.12102	0.059468	-2.035	0.041853*
科研費 0	0.049719	0.00841	5.9121	3.4E-09***
科研費 2	-0.05493	0.009105	-6.0331	1.6E-09***

図表 16 従属変数を異動の有無としたパネルデータ分析の説明変量 III (59歳以上, $R^2=0.98979$, $n=24739$, $T=1-8$, $N=54608$, *5%有意)

※このパネルデータ分析のデータセットは厳密には1節のものと異なるが、性格はほぼ同じであり、概ね同一のものと見て差し支えない。

6. 結論

本稿では職員録データベースに対して統計学的アプローチから、大学教員の異動メカニズムを解明することを目指した。手法にはまだ改善の余地が多いと思われるものの、現象を理解するためのアプローチとして一定の成果は得られた。しかし、様々な手法を試したものの、今後の課題は因果関係の究明である。

7. 謝辞

科学技術政策研究所の西井龍映客員研究官(九州大学教授)には数多くの貴重な助言を戴いた。また、桑原輝隆総務研究官と永田晃也総括主任研究官には本研究を実施する機会を提供していただいた。ここに謝意を表する。

【参考文献】

- [1] 2006年度研究・技術計画学会第21回年次学術大会、2G25「大学教員人材の流動性に関する予備的調査」、細坪護学
- [2] 人文・社会科学のためのカテゴリカル・データ入門、ナカシニヤ出版、太郎丸博
- [3] NISTEP REPORT No.122、「日本の大学に関するシステム分析」、2009年3月、付属資料4「国立大学教員の流動性計測のためのデータベースの構築とこれによる予備的調査分析」、細坪護学
- [4] グラフィカルモデリングの実践、日本品質管理学会テクノメトリックス研究会編、日科技連