

Title	国立大学教員による科研費採択の政策的意味に関する統計解析
Author(s)	細坪, 護拳; 西井, 龍映
Citation	年次学術大会講演要旨集, 27: 78-83
Issue Date	2012-10-27
Type	Conference Paper
Text version	publisher
URL	http://hdl.handle.net/10119/10979
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

国立大学教員による科研費採択の政策的意味に関する統計解析

○細坪 護孝 (文部科学省 科学技術政策研究所)
西井 龍映 (九州大学 マス・フォア・インダストリ研究所)

1. はじめに

科学技術政策研究所等のアンケート調査等から、科研費制度に対する研究者の評価は極めて高い。本発表では、科研費採択と国立大学教員の属性との関係を統計解析により、客観的に示す。将来的には、国の他の競争的資金制度や科学技術・学術政策に対し、本研究に基づくシミュレーション遂行で政策提言を目指す。

2. 先行研究

Rao (1980), 山崎(1982) による、科学研究生産性に関する統計学的検討の先行研究で、論文数と学者数の関係として、負の二項分布の適合が主張されている。「科研費採択」は「科学研究生産性」自体ではなく、科研費の研究・分析はJSPSやNIIで行っている。しかし、統計解析はほとんど見当たらない。

3. 標本設計

「全国大学職員録」(廣潤社)などの掲載情報から観測年9時点(88,91,94,97,00,03,05,06,08年)8区間での大学教員パネルデータ(高等教育教員異動データベース:HM-DB)を使用して、教員の異動・昇格を統計解析[1-5]した。本データは助教の把握率がやや低いものの、それ以外の職位の把握率は高く、公的統計と比較可能な品質である。

本研究ではHM-DBと科研費採択数を接続して使用する。科研費制度も時代で変化してきた。今回は観測年を1年当たりとし、代表研究者として採択された区間数(0-8回,9水準,以下括弧内数字は水準数を目的変数とし、説明変数として

【観測時間】として下記2変量

【終年】: 観察終了年(8)

【観察時点数】: 観察終了までの接続時点数(8)

【最近の状況】として下記3変量

【所属大学】: 国立大 G1・2・3・4等[1-5, 附録]等(7)

【大学地域】: 関東、中部、近畿、中国・四国等(6)

【職位】: 研究科長等、教授、准教授、講師、助教(5)

【出身属性】として下記5変量

【世代】: 焼跡(35-46年)、団塊(47-49年)、しらけ(50-64年)、バブル(65-69年)、氷河期(70-86年)等(6)

【学歴】: 博士、修士、学士、その他(4)

【出身大学】: 自校出身、外国大学等(9)

【出身大学地域】: 関東、中部、近畿、外国等(7)

【分野】: 数学、総合領域、複合新領域、人文学等(11)

【経緯】として下記3変量

【異動回数】: 大学間の異動回数(5)

【昇格回数】: 観察期間中の昇格回数(5)

【昇格パス】: 観察最初・最後の職位の組み合わせ(20)

の計13共変量を作成した(n=110,056)。

4. 全体および分野ごとの採択数の分布推定

科研費採択件数と教員数を比較すると、代表者採択数0件が多い。そこで8区間の科研費採択数と教員数の観測合計値に対して、PO, NB, ZIP(zero-inflated PO), ZINB(zero-inflated NB)のパラメータを推定すると図1となる。ここで、

$Y \square ZIP$ とは $Y_0 \square PO$ のとき

$$P(Y=0) = \nu + (1-\nu)P(Y_0=0)$$

$$P(Y=k) = (1-\nu)P(Y_0=k) \quad (k=1,2,\dots)$$

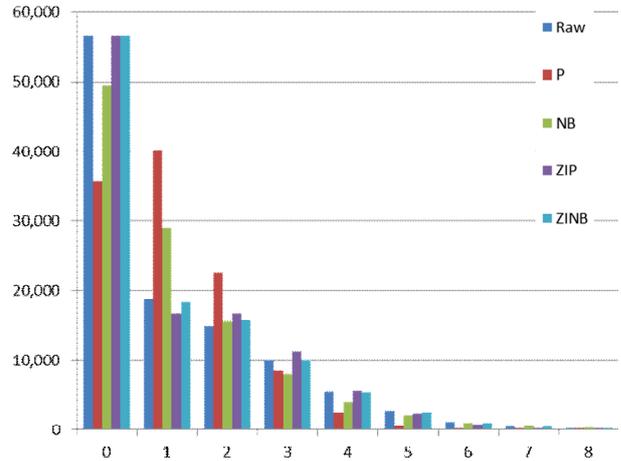


図1 科研費代表者採択数と国立大学教員数の分布

であり、 ν をzero-inflate 確率という。BICの値は、PO: 358089 ($\nu=0$), NB: 323129 ($\nu=0$), ZIP: 318749 ($\nu=0.4384$), ZINB: 318080 ($\nu=0.4058$)となり、zero-inflated models がより重要なモデルであると判明した。POとNBの差ほどZIPとZINBに差は見られない。

分野ごとの分布は図2となり、各分野での最適モデルとそのzero-inflate 確率を調べると、表1となりPOよりNBの方が良いが、前の分布適合に比べると、差は小さい。一部の分野ではZIPの方がZINBより最適になる。また、0件構造は専門分野に大きく依存している。

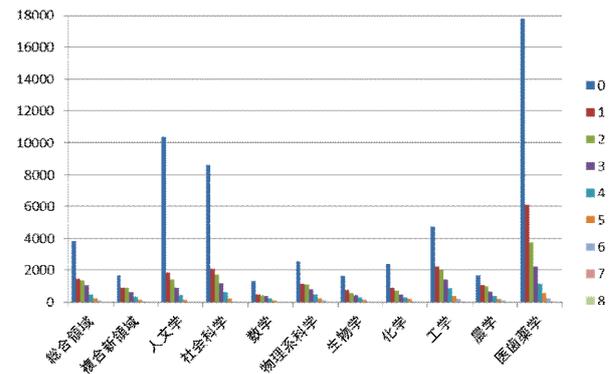


図2 分野ごとの科研費代表者採択数と国立大学教員数の分布

分野	総合領域	複合新領域	人文学	社会科学	数学	物理系
最適モデル	ZINB	ZIP	ZINB	ZINB	ZIP	ZINB
ゼロインフレート確率	0.374	0.282	0.618	0.520	0.384	0.323
分野	生物学	化学	工学	農学	医歯薬学	
最適モデル	ZINB	ZINB	ZINB	ZINB	ZINB	
ゼロインフレート確率	0.349	0.406	0.322	0.243	0.465	

表1 最適モデル(PO, NB, ZIP, ZINB)とゼロインフレート確率

この分野の違いの理由の可能性には、医歯薬学では大学附属病院の医師を含むこと、実験系と理論系科学、学問における科研費の性格の違いが考えられる。

本研究の拠り所となるHM-DB, 科研費データでは、教員の研究特性の理論・実験の別は不明である。

5. PO, NB, ZIP 回帰と多項ロジスティック回帰

ここでは、科研費代表者採択数を PO, NB, ZIP 回帰する。2. で作成した 13 変量の線形項で一般化線形モデル[8-9]を適用すると、水準数が多すぎたり統計学的意義の乏しい変量により、不安定な推定モデルとなる。

そこで、まず、主効果モデルを網羅計算 $2^{13}=8192$ 式の BIC を求め、最適モデルを探索する。すると、

PO 回帰: 276504(-[職位], df=84)

NB 回帰: 273652(-[職位]-[大学地域], df=79)

ZIP 回帰: 271823(-[出身大学地域]-[昇格パス], df=65,

本研究では ZIP 回帰は母数 μ に対してのみ行う)

ただし、-[職位]は説明変数から[職位]を除くことを意味する。

PO より NB の方が適合は良いが、4. の分布適合に比べると差は小さくなっている。この 3 つでは ZIP が最適であり、以降は ZIP を使う。

一方、zero-inflate も含めた時系列的により柔軟なモデル化も考えられる。ここでは、科研費採択回数 0, 1, ..., 8 をカテゴリとした下記の多項ロジスティック回帰(Logit)である。

p 個の説明変量 $x_i = (x_{i1}, \dots, x_{ip})^t$ を持つ i 番目の教員

$$\ln\left(\frac{P([\text{科代}] = k)}{P([\text{科代}] = 0)}\right) = x_i^t \beta_k, \beta_k = (\beta_{k0}, \beta_{k1}, \dots, \beta_{kp})^t, k = 1, 2, \dots, 8$$

多項ロジスティックモデル(Logit)による説明変数の選択では、科研費採択数 0 件で inflate していることから、目的変数を科研費採択 0 件に対する他の採択数とし、13 説明変数から、Logit 推定 [10-11]する。この計算では上記 ZIP と同じく BIC をモデル評価指標とし、網羅計算による最適モデル探索を行う。その結果、採択数と、より強い関連のある共変量を持つ異なるモデルとして、

Logit: 264530 (-[大学地域]-[出身大学]-[出身大学地域]-[昇格回数]-[昇格パス], df=376)

が得られた。この Logit モデルは上記の PO, NB, ZIP より適切である。説明変数と目的変数を書くと、

[科代] ~ [終年]+[観察時点数]+[所属大学]+[世代]+[職位]+[学歴]+[分野]+[異動回数]

となる。

次に、統計学的に意義の乏しい説明変数の水準を統合する。本研究では (1)変量内の水準統合、と (2)モデルの水準統合を分けて考える。構造的な問題を避けるため、[終年]、[観察時点数]の 2 変量は統合の対象外とする。(1)では、

- 1) 前モデルの変量内の各水準間に順序を設定する
- 2) 水準間の効果差(推定値差)が小さい水準を統合する
- 3) BIC が増加に転じるまで、1)-2)を反復

(2)では、BIC の低下が大きな変量から順に水準統合する。前述の Logit モデルでは、[異動回数] → [学歴] → [所属大学] → [世代] → [分野]の順となる。水準統合により BIC は **263693 (df = 288)**にまで低下し、本研究の最適モデルとなった。また、各変量で統合された水準(・)は次のとおり。水準の順番は採択数への正の影響順。

[所属大学]: 国立大 G1, G2, G3, 私立大, G4・公立大, 大学校等

[世代]: 氷河期, バブル, しらけ, 焼跡・団塊, 大正昭和一桁

[学歴]: 博士, 修士・その他, 学士

[分野]: 医歯薬学, 農学, 生物学・化学・工学, 総合領域・

物理系科学, 複合新領域・数学, 人文学・社会科学,

[異動回数]: 1 以上, 0

この係数推定値を表 2 に示す。この水準統合から、科研費代表者となる教員の異動「回数」は関係なく、有無しか意味がないと分かる。

係数推定値	切片	終年								観察時点数							
		88年	91年	94年	97年	00年	03年	05年	06年	r.11	r.12	r.13	r.14	r.15	r.16	r.17	r.18
1	0.84	0.00	-0.56	-0.45	-0.46	-0.40	-0.44	-0.22	-0.26	-0.89	-0.20	-0.25	-0.10	-0.08	-0.05	-0.03	0.00
2	-1.11	0.00	-0.76	-0.47	0.05	-0.72	-0.61	0.36	0.77	-0.35	-0.57	-0.12	-0.19	-0.18	0.10	0.14	0.00
3	-1.76	0.00	-1.66	0.16	0.31	-0.70	0.58	0.64	1.17	0.37	-1.25	-0.25	-0.38	-0.02	-0.15	0.11	0.00
4	-0.80	0.00	0.39	0.14	-0.28	-0.51	-0.60	-0.29	0.87	0.13	-0.26	-4.25	-0.77	-0.40	-0.69	0.15	0.00
5	1.23	0.00	0.24	0.19	0.11	0.49	0.47	0.23	0.96	0.18	0.53	2.68	0.89	0.79	0.30	0.00	0.00
6	-0.85	0.00	0.21	-0.29	-0.24	-0.58	0.22	0.59	0.16	0.18	-0.20	0.05	-0.61	-0.84	-1.22	-0.21	0.00
7	-1.64	0.00	-0.02	-0.77	-0.74	-0.52	-0.69	0.18	0.88	-0.19	-0.45	-0.53	0.62	0.05	-0.98	-0.68	0.00
8	-1.22	0.00	0.05	-0.61	-1.40	-0.70	-0.45	-0.09	0.62	0.24	0.13	-0.28	0.92	0.47	0.06	-1.38	0.00

係数推定値	所属大学				世代				職位							
	G1	G2	G3	公立大	私立大	大学校等	大正昭和	焼跡・しらけ	バブル	氷河期	助教	講師	准教授	教授	研究科長等	
1	0.00	-0.22	-0.92	-0.81	-0.58	-1.72	-0.28	-0.83	0.00	0.48	0.78	-0.61	-0.25	-0.32	0.00	0.94
2	0.00	-0.22	-0.96	-1.15	-1.09	-1.89	-0.15	-0.47	0.00	0.64	1.28	-1.25	-1.06	-0.80	0.00	0.62
3	0.00	-0.41	-1.22	-1.44	-1.31	-2.00	-0.18	-0.89	0.00	0.89	1.93	-1.56	-1.46	-0.78	0.00	1.13
4	0.00	-0.81	-1.30	-1.72	-1.42	-2.30	-0.68	-0.58	0.05	1.02	1.74	-0.40	-1.06	-1.31	0.00	1.07
5	0.00	-0.78	-1.56	-2.04	-1.58	-2.43	-0.22	-0.99	0.00	0.89	1.88	-2.44	-2.14	-1.81	0.00	1.02
6	0.00	-0.29	-1.50	-2.01	-1.79	-2.25	-1.43	-0.53	0.43	1.14	1.66	-3.27	-2.84	-1.86	0.00	1.28
7	0.00	-0.67	-1.81	-2.09	-1.85	-1.98	-1.28	-0.63	0.00	1.52	2.20	-2.79	-2.19	-1.84	0.00	1.28
8	0.00	-1.19	-1.58	-2.18	-1.99	-2.43	-0.47	0.17	0.00	1.01	1.23	-2.32	-1.65	-1.92	0.00	1.32

係数推定値	学歴		分野											
	修士・その他	博士	工学・人文系	数学・複合領域	学・生物	社会科学	総合領域	物理系科学	農学	化学	工学	医歯薬学	0	1 以上
1	0.00	0.20	0.94	-0.21	-0.82	-0.77	-0.44	-0.07	0.03	0.00	0.00	0.00	0.14	0.14
2	0.00	0.69	1.85	-0.33	-0.92	-0.46	-0.70	0.02	0.00	0.00	0.00	0.00	0.36	0.36
3	0.00	0.42	1.53	-0.08	-1.17	-0.59	-0.61	0.00	0.00	0.00	0.00	0.00	0.56	0.56
4	0.00	0.66	1.93	-0.10	-1.24	-0.60	-0.77	-0.06	0.00	0.00	0.00	0.00	0.84	0.84
5	0.00	0.59	1.84	-0.43	-1.78	-0.85	-0.78	-0.14	0.00	0.00	0.00	0.00	0.98	0.98
6	0.00	-0.02	1.40	-0.41	-1.95	-1.29	-1.08	0.18	0.00	0.00	0.00	0.00	1.24	1.24
7	0.00	-0.27	1.24	-0.71	-2.31	-1.10	-1.50	-0.69	0.00	0.00	0.00	0.00	0.43	0.43
8	0.00	0.24	1.62	-0.54	-2.07	-1.64	-0.73	0.03	0.00	0.00	0.00	0.00	1.23	1.23

表 2 Logit 最適モデルの係数推定値

また、上記 Logit の最適モデルで ZIP と ZINB 回帰を行うと、ZINB (272701, df = 37), ZIP (272712, df = 38)となる。同様に ZIP の最適モデルの水準を統合し、Logit に勝るかを調べた。ZIP 最適モデルの水準統合は、[分野] → [出身大学] → [大学地域] → [所属大学] → [世代] → [昇格回数] → [異動回数]の順となる。水準統合により BIC は 271671(df = 48)にまで低下するが、上記の Logit モデルには及ばない。また、各変量で統合された水準(・)は次のとおり。水準の順番は採択数への正の影響順(以下同じ)。
[所属大学]: 国立大 G1, G2, G3・私立大, G4・公立大, 大学校等
[大学地域]: 関東・中部, 近畿, 北海道/東北・中国/四国・九州
[世代]: 氷河期, バブル, しらけ, 大正昭和一桁・焼跡・団塊
[出身大学]: 国立大 G1, 自校出身, G2・G3, 公立大・外国大学, G4・私立大, 大学校等
[分野]: 医歯薬学, 生物学・農学, 化学・工学, 総合領域・複合新領域・物理系科学, 数学, 社会科学, 人文学
[異動回数]: 2・3, 1, 4 以上, 0
[昇格回数]: 0, 1, 2・3, 4 以上

6. 分野ごとの説明変数の選択と水準の統合

5. で、全体を説明するには ZIP より Logit (多項ロジット回帰) モデルの方が適切と判明したが、4. に示したとおり、専門分野に応じて zero-inflate 確率は大きく異なり、これは、0 件採択者の事情が分野間で大きく異なることを示唆する。そのため、分野別の解析も必要となる。全体の結果と異なり、分野に分けると、全分野で Logit より ZIP の方がよいモデルとなる。また、Logit では、各分野最適モデルの BIC 合計値(277009) > 全分野の BIC (264530, df = 376)となる一方、ZIP では、各分野最適モデルの BIC 合計値(271430) < 全分野の BIC (272664, df = 48)となり、分野別解析で ZIP モデルを使用することが適切と分かる。しかし、**水準統合で逆転**することもある。そこで次に、ZIP, Logit モデルに対して分野別に水準統合と交互作用項の検証を行う。

こうして、分野毎の網羅計算による最適モデル探索を行う。以下に、採択数とより強い関連のある共変量を持つ異なるモデルを示す。

6.1 総合領域(n= 8490):

[Logit][科代][観察時点数]+[学歴](BIC=24193,df=88)

[ZIP][科代][観察時点数]+[所属大学]+[世代]+[職位]+[学歴](BIC=23911,df=27)

総合領域の ZIP モデルの水準統合は[所属大学] → [世代] → [職位]の順であり、各変量で統合された水準(・)は次のとおり。

[所属大学]: 国立大学 G1, G2, G3・私立大学・大学校等, G4・公立大学

[世代]:氷河期, バブル, しらけ・団塊, 焼跡・大正昭和一桁
 [職位]:研究科長等, 教授, 准教授・助教・講師
 この結果BICは23866(df=20)まで低下する。2次交互作用項は存在しない。また、Logit で水準統合するとBICは24140(df=80)となりZIPに及ばず、2次交互作用項も存在しない。

6.2 複合新領域(n=4628):

【Logit】[科代]~[観察時点数]+[学歴](BIC=14802,df=88)
 【ZIP】[科代]~[観察時点数]+[所属大学]+[世代]+[職位]+[学歴]+[異動回数](BIC=14231,df=31)
 複合新領域のZIPモデルの水準統合は[所属大学]→[世代]→[職位]→[異動回数]→[学歴]の順であり、各変量で統合された水準(・)は次のとおり。
 [所属大学]:国立大学 G1, G2, G3, G4・公立大学・私立大学・大
 学校等
 [世代]:氷河期, バブル, しらけ・大正昭和一桁, 焼跡・団塊
 [職位]:研究科長等, 教授, 准教授, 助教・講師
 [異動回数]:3, 1・2, 0, 4以上
 [学歴]:博士, 修士・その他, 学士
 この結果BICは14171(df=23)まで低下する。2次交互作用項は存在しない。また、Logit で水準統合するとBICは14738(df=80)となりZIPに及ばず、2次交互作用項も存在しない。

6.3 人文学(n=15165):

【Logit】[科代]~[観察時点数]+[所属大学]+[世代]+[学歴](BIC=29403,df=176)
 【ZIP】[科代]~[終年]+[観察時点数]+[所属大学]+[世代]+[職位]+[学歴]+[異動回数](BIC=29112,df=38)
 人文学のZIPモデルの水準統合は[所属大学]→[異動回数]→[世代]→[職位]→[学歴]の順であり、各変量で統合された水準(・)は次のとおり。
 [所属大学]:国立大 G1・G2・私立大, G3・G4・大学校等, 公立大
 [世代]:氷河期, バブル, しらけ, 大正昭和一桁・焼跡・団塊
 [職位]:研究科長等, 教授, 准教授・助教, 講師
 [学歴]:博士, 修士, 学士・その他
 [異動回数]:2・3, 1, 0・4以上
 この結果BICは29027(df=28)まで低下する。2次交互作用項は存在する。モデルが大きいため、3項までは網羅計算で求め、4項以上はBIC変数増加法とする(以下同じ)。すると、
主効果項 + [大学]: [世代] + [世代]: [学歴] + [大学]: [学歴] (BIC = 28936, df = 44)
 が最適となる。交互作用項の推定値は表3となる。

		大学		
		G12私	G34大	公立大
世代	大正昭一 焼跡 団塊	0.000	-0.362	-0.613
	しらけ	0.000	0.000	0.000
	バブル	0.000	0.175	0.215
	氷河期	0.000	0.176	0.192

		学歴		
		学士その他	修士	博士
世代	大正昭一 焼跡 団塊	0.000	0.188	0.581
	しらけ	0.000	0.000	0.000
	バブル	0.000	0.974	0.912
	氷河期	0.000	0.215	0.154

		大学		
		G12私	G34大	公立大
学歴	学士その他	0.000	0.000	0.000
	修士	0.000	0.360	-0.376
	博士	0.000	0.493	0.136

表3 人文学のZIP2次最適モデルの交互作用項の推定値

一方、Logit で水準統合するとBICは29183(df=136)となりZIPに及ばず、2次交互作用項も存在しない。
 人文学の特徴として、採択数への私立大学の果たす役割が大きく、講師の科研費取得数が少ない。これは語学教員の影響によるものと考えられる。また、基本的に大学教員は世代が最近になるにつれ、博士号取得者が多いが、人文学で科研費取得と博士号取得、世代間には簡単な線形関係はない。人文学における博士号取得システムが単純ではないことを示唆する。

6.4 社会科学(n=14567):

【Logit】[科代]~[観察時点数]+[所属大学]+[世代]+[学歴](BIC=34713,df=176)
 【ZIP】[科代]~[終年]+[観察時点数]+[所属大学]+[世代]+[職位]+[学歴](BIC=34597,df=34)
 社会科学のZIPモデルの水準統合は[世代]→[所属大学]→[職位]の順であり、各変量で統合された水準(・)は次のとおり。
 [所属大学]:国立大学 G1, G2・私立大学, G3・G4, 公立大学, 大学校等
 [世代]:氷河期, バブル, 大正昭和一桁・焼跡・団塊・しらけ
 [職位]:研究科長等, 教授, 准教授, 助教・講師
 この結果BICは34554(df=28)まで低下する。2次交互作用項は存在しない。
 一方、Logit で水準統合するとBICは34393(df=136)となりZIPに勝る。Logitモデルの水準統合は[所属大学]→[学歴]→[世代]の順であり、各変量で統合された水準(・)は次のとおり。
 [所属大学]:国立大学 G1, G2, G3・G4・私立大学, 公立大学・大学校等
 [世代]:氷河期, バブル, 大正昭和一桁, しらけ, 焼跡・団塊
 [学歴]:博士, 修士, 学士・その他
 各分野のLogitでは2次交互作用項は存在しない。これは分野を問わない。なぜならば、Logitの交互作用モデルでは科研費採択数と時間に依存するゼロ制約項(0と置くべき項)が爆発的に増加し、無制約下の現行モデルでは不要な推定母数が増え、ZIPより不利になるためと考えられる。理論上、これらを推定モデルから省くことは可能である。当該母数は推定モデルに最初から存在しないと見なせばよい。しかし、実際の計算では実現できていない。

6.5 数学(n=2953):

【Logit】[科代]~[世代]+[職位](BIC=8733,df=80)
 【ZIP】[科代]~[終年]+[観察時点数]+[所属大学]+[世代]+[職位]+[学歴](BIC=8013,df=34)
 数学のZIPモデルの水準統合は[所属大学]→[世代]→[職位]→[学歴]の順であり、各変量で統合された水準(・)は次のとおり。
 [所属大学]:国立大学 G1・G2・私立大学, G3・G4・公立大学, 大学校等
 [世代]:氷河期, バブル, しらけ, 焼跡・団塊, 大正昭和一桁
 [職位]:研究科長等, 教授, 准教授, 助教・講師
 [学歴]:博士, 学士・修士, その他

この結果 BIC は 7966(df=27)まで低下する。2 次交互作用項は存在しない。また、Logit で水準統合すると BIC は 8653 (df=64)となり ZIP に及ばない。

6.6 物理系科学(n=6557):

【Logit】[科代]~[観察時点数]+[所属大学]+[世代]+[職位]
(BIC=19950,df=184)

【ZIP】[科代]~[終年]+[観察時点数]+[所属大学]+[世代]+[職位]
+[学歴](BIC=19186,df=34)

物理系科学の ZIP モデルの水準統合は[所属大学] → [世代] → [職位] → [学歴]の順であり、各変量で統合された水準(・)は次のとおり。

[所属大学]: 国立大学 G1, G2・私立大学, G3・G4・公立大学, 大学校等

[世代]: 氷河期, バブル, しらけ, 大正昭和一桁・焼跡・団塊

[職位]: 研究科長等, 教授, 准教授, 助教・講師

[学歴]: 博士, 修士・その他, 学士

この結果 BIC は 19121(df=26)まで低下する。2 次交互作用項を入れた最適モデルは次のとおり。

主効果項 + [世代]: [学歴] (BIC = 19119, df = 29)

交互作用項の推定値は表 4 となり、物理系科学では世代を遡るほど、高学歴者の採択数が増える。

一方、Logit で水準統合すると BIC は 19649(df=144)となり ZIP に及ばない。

		世代		
		大正昭和一桁・焼跡・団塊・しらけ	バブル	氷河期
学歴	学士	0.000	0.000	0.000
	修士その他	1.616	0.000	-0.264
	博士	2.422	0.000	-

表 4 物理系科学の ZIP2 次最適モデルの交互作用項の推定値

6.7 生物学(n=3890):

【Logit】[科代]~[観察時点数](BIC=11833,df=64)

【ZIP】[科代]~[終年]+[観察時点数]+[所属大学]+[世代]+[職位]
+[学歴]+[異動回数]+[出身地](BIC=11148,df=44)

生物学の ZIP モデルの水準統合は[所属大学] → [異動回数] → [世代] → [出身地] → [職位] → [学歴] の順であり、各変量で統合された水準(・)は次のとおり。

[所属大学]: 国立大学 G1・G2, 公立大学・私立大学・大学校等, G3・G4

[世代]: 氷河期, バブル・大正昭和一桁, 焼跡・団塊・しらけ

[職位]: 研究科長等, 教授, 准教授, 助教・講師

[学歴]: 博士, 修士・その他, 学士

[出身地]: 関東, 外国, 中部・近畿, 北海道/東北・九州, 中国/四国

[異動回数]: 1・2・3・4 以上, 0

この結果 BIC は 11045(df=30)まで低下する。2 次交互作用項を入れた最適モデルは次のとおり。

主効果項 + [大学]: [異動回数] + [学歴]: [異動回数]

(BIC = 11026, df = 34)

交互作用項の推定値は表 5 となり、生物学では科研費取得数の少ない大学や、学歴の低く、1 回以上異動する教員は科研費取得しやすい。

一方、生物学の Logit モデルは水準統合できない。

		大学		
		G12	公私大学校等	G34
異動回数	1,2,3,4以上	0.000	0.241	0.281
	0	0.000	0.000	0.000

		学歴		
		学士	修士その他	博士
異動回数	1,2,3,4以上	0.000	-0.873	-1.761
	0	0.000	0.000	0.000

表 5 生物学の ZIP2 次最適モデルの交互作用項の推定値

6.8 化学(n=4986):

【Logit】[科代]~[観察時点数]+[所属大学]+[世代]+[職位]
(BIC=13877,df=184)

【ZIP】[科代]~[終年]+[観察時点数]+[所属大学]+[世代]+[職位]
+[学歴]+[異動回数]+[出身地]+[昇格回数]
(BIC=13075,df=48)

化学の ZIP モデルの水準統合は[出身地] → [所属大学] → [異動回数] → [世代] → [昇格回数] → [職位] → [学歴] の順であり、各変量で統合された水準(・)は次のとおり。

[所属大学]: 国立大学 G1, G2・私立大学, G3・G4・公立大学・大学校等

[世代]: 氷河期, バブル, 大正昭和一桁・しらけ, 焼跡・団塊

[職位]: 研究科長等, 教授, 准教授, 助教・講師

[学歴]: 博士, 修士・その他, 学士

[出身地]: 関東・近畿・外国, 北海道/東北・中部・九州, 中国/四国

[異動回数]: 1・2・3・4 以上, 0

[昇格回数]: 0, 1, 2・3・4 以上

この結果 BIC は 12945(df=31)まで低下する。2 次交互作用項を入れた最適モデルは次のとおり。

主効果項 + [大学]: [異動回数] + [世代]: [職位]

(BIC = 12865, df = 40)

交互作用項の推定値は表 6 となる。

		大学			
		G1	G2私立大学	G34公大学校等	
異動回数	1,2,3,4以上	0.000	0.090	0.476	
	0	0.000	0.000	0.000	

		職位			
		助教講師	准教授	教授	研究科長等
世代	大正昭和一桁	-0.650	-0.343	0.000	-
	焼跡・団塊	-1.251	-1.050	0.000	-0.038
	バブル	0.000	0.000	0.000	0.000
	氷河期	0.393	0.227	0.000	-

表 6 化学の ZIP2 次最適モデルの交互作用項の推定値

一方、Logit で水準統合すると BIC は 13568(df=144)となり ZIP に及ばない。

6.9 工学(n=11879):

【Logit】[科代]~[観察時点数]+[所属大学]+[世代]+[職位]+[学歴]
(BIC=34315,df=208)

【ZIP】[科代]~[終年]+[観察時点数]+[所属大学]+[世代]+[職位]
+[学歴]+[異動回数]+[昇格回数]
(BIC=33927,df=42)

工学の ZIP モデルの水準統合は[異動回数] → [昇格回数] → [世代] → [所属大学] → [学歴] → [職位] の順であり、各変量で統合された水準(・)は次のとおり。

[所属大学]: 国立大 G1, G2, 私立大, G3・G4・公立大, 大学校等

[世代]: 氷河期, バブル, しらけ, 大正昭和一桁・焼跡・団塊

[職位]: 研究科長等, 教授, 准教授, 助教・講師

[学歴]: 博士, 修士, 学士・その他

[異動回数]: 0・3・4 以上, 1・2

[昇格回数]: 0, 1, 2・3・4 以上

この結果 BIC は 33839(df=31)まで低下する。2 次交互作用項を入れた最適モデルは次のとおり。

**主効果項 + [大学]:[異動回数] + [世代]:[職位]
+ [職位]:[学歴] + [異動回数]:[昇格回数]
(BIC = 33565, df = 48)**

交互作用項の推定値は表 7 となり、工学では昇格すると科研費採択数が増える模様。基本的に職位が高くなるほど科研費の取得が多くなるが、工学では若い世代が科研費を取得する傾向が強く、高い職位になると科研費以外のグラントに申請するのではないかと考えられる。

異動回数		大学				
		G1	G2	G34公	私立	大学校等
0,3,4 以上	1,2	0.000	0.058	0.225	-0.005	1.334
1,2		0.000	0.000	0.000	0.000	0.000

世代	職位			
	助教講師	准教授	教授	研究科長等
大正昭一 焼跡団塊 しらけ	-1.094	-0.841	0.000	0.019
バブル	0.000	0.000	0.000	0.000
氷河期	0.525	0.205	0.000	-
	0.858	0.414	0.000	-

異動回数	昇格回数		
	0	1	2,3,4以上
0,3,4以上	0.000	-0.205	-0.189
1,2	0.000	0.000	0.000

学歴	職位			
	助教講師	准教授	教授	研究科長等
学士その他	0.000	0.000	0.000	0.000
修士	-0.396	-0.242	0.000	-
博士	-1.309	-1.148	0.000	-

表 7 工学の ZIP2 次最適モデルの交互作用項の推定値

一方、Logit で水準統合すると BIC は 33813(df=152)となり ZIP に勝る。Logit モデルの水準統合は[所属大学] → [世代] → [職位] → [学歴]の順であり、各変量で統合された水準(・)は次のとおり。

[所属大学]: 国立大 G1, G2・私立大, G3・G4・公立大, 大学校等
[世代]: 氷河期, バブル, しらけ, 大正昭和と一桁・焼跡・団塊
[職位]: 研究科長等, 教授, 准教授, 助教・講師

[学歴]: 博士, 修士・その他, 学士
結果、科研費採択数への私立大学の果たす役割が大きいことが分かる。

6.10 農学(n=5087):

[Logit][科代][観察時点数]+[学歴](BIC=16289,df=88)
[ZIP][科代][観察時点数]+[所属大学]+[世代]+[職位]+[学歴]
(BIC=15646,df=27)

農学の ZIP モデルの水準統合は[所属大学] → [世代] → [職位] → [学歴]の順であり、各変量で統合された水準(・)は次のとおり。

[所属大学]: 国立大学 G1, G2・私立大学, G3・公立大学, G4, 大学校等

[世代]: 氷河期, バブル, 団塊・しらけ, 大正昭和と一桁・焼跡
[職位]: 研究科長等, 教授, 准教授, 助教・講師

[学歴]: 博士, 修士・その他, 学士

この結果 BIC は 15604(df=21)まで低下する。2 次交互作用項を入れた最適モデルは次のとおり。

主効果項 + [世代]:[職位] (BIC = 15599, df = 28)

交互作用項の推定値は表 8 となり、世代が若くて職位が高い程、科研費取得数が多い。

世代	職位			
	助教講師	准教授	教授	研究科長等
大正昭一 焼跡	-1.183	-1.089	0.000	0.038
団塊しらけ	-0.302	-0.296	0.000	-
バブル	0.000	0.000	0.000	0.000
氷河期	-0.031	0.011	0.000	-

表 8 農学の ZIP2 次最適モデルの交互作用項の推定値

一方、Logit で水準統合すると BIC は 16229(df=80)となり ZIP に及ばない。

6.11 医歯薬学(n=31854):

[Logit][科代][観察時点数]+[所属大学]+[世代]+[職位]+[学歴]
(BIC=68901,df=208)

[ZIP][科代][終年]+[観察時点数]+[所属大学]+[世代]+[職位]
+[学歴]+[出身大学](BIC=68584,df=42)

医歯薬学の ZIP モデルの水準統合は[出身大学] → [所属大学] → [学歴]の順であり、各変量で統合された水準(・)は次のとおり。

[所属大学]: 国立大 G1, G2, G3, 公立大・私立大, G4・大学校等
[学歴]: 博士, 学士・修士・その他
[出身大学]: 国立大 G1, G2・G3・公立大・外国大学・自校,
G4・私立大学, 大学校等

この結果 BIC は 68498(df=33)まで低下する。2 次交互作用項を入れた最適モデルは次のとおり。

**主効果項 + [大学]:[学歴] + [世代]:[学歴]
(BIC = 68471, df = 42)**

交互作用項の推定値は表 9 となり、G4 と大学校等、公立大学と私立大学が統合されることから、地域の大学総合病院との関係が示唆される。また、出身大学はかなり大括りになる。更に、科研費取得と博士号取得、世代間には簡単な線形関係はないことが分かる。

学歴	大学				
	G1	G2	G3	公私	G4大学校等
修士 学士その他	0.000	0.000	0.000	0.000	0.000
博士	0.000	0.148	0.281	0.314	0.518

学歴	世代					
	大正昭一	焼跡	団塊	しらけ	バブル	氷河期
修士 学士その他	0.000	0.000	0.000	0.000	0.000	0.000
博士	0.424	0.543	0.556	0.000	0.032	0.120

表 9 医歯薬学の ZIP2 次最適モデルの交互作用項の推定値

一方、Logit で水準統合すると BIC は 68449(df=144)となり ZIP に勝る。**医歯薬学の Logit モデルは上記の ZIP 交互作用モデルにも勝る。**Logit モデルの水準統合は[所属大学] → [世代] → [学歴] → [職位]の順であり、各変量で統合された水準(・)は次のとおり。

[所属大学]: 国立大 G1, G2, G3, G4・公立大・私立大・大学校等

[世代]: 氷河期, バブル, しらけ, 大正昭和と一桁・焼跡・団塊
[学歴]: 博士, 学士・その他, 修士

[職位]: 研究科長等, 教授, 助教・講師・准教授

結果、G4、公私立大学等統合され、地域大学総合病院との関係の示唆とともに、科研費採択件数上、准教授以下の職位は同水準と見なされることが分かる。

7. まとめ

本研究では、公表されている科研費代表者の情報からその取得件数と教員との関係を調べた。無論、科研費には審査があり、申請しても採択されないこともある。しかし、申請件数は文科省やJSPSから公表された分野別等の大括りの集計数しか存在せず、教員ベースのデータはない。この推定は今後の課題である。

本研究から科研費代表者取得件数と教員数には、0件取得者数が極めて多いというzero-inflateが見られる。そのため、一般化線形モデル(PO, NB, ZIP, ZINB)でモデルを評価すると、ゼロに特別の重みを付けるZIPやZINBが最適と推定された。

更に、教員の13共変量を用いた回帰分析でもPOやNBよりZIPモデルの優位が確認された。ただし、本研究におけるZIP回帰では、モデル平均母数のみのモデルを構築しており、zero-inflate確率に関しては、定数と仮定して推定している。本来、ZIPでは母数が2つあり、回帰式を2つ立てて推定する。本研究では、2つの理由から後者の母数を定数と置いた。1)計算量の縮小。13変量の主効果モデルの網羅計算だけでも8,192式である。これに加えて、別の母数を網羅的に求めることは不可能である。しかも、一般化線形モデルのモデル評価の近似計算手法としてステップワイズ(BICの変数増減法等)があるが、この場合には、ステップワイズによる評価は使えず、2)zero-inflate確率を回帰式でモデル定式化しても、実は本研究の目的とはあまり関係がない。具体的には、科研費0件のinflateとは、そもそも研究を実質本務としない教員の特性との差の問題であり、そういった教員の存在自体はおかしなことではない。この現象を学術的・科学的に追究する意味はあるが、政策的意義には乏しい。なぜならば、本研究の目的は、科研費などのグラントが研究者(大学教員)にどのような影響を及ぼすのか、及ぼされるのかを究明し、よりパフォーマンスの高い政策への貢献が第一であって、各分野の御家事情の暴露・解明ではないからである。

前記のとおりZIPモデルの優位性とともに、0件構造分析も含めた柔軟なカテゴリカルモデルとして、多項ロジスティック(多項ロジット)回帰モデル解析も行った。後者の特徴は、目的変数、即ち科研費取得件数がZIPなどの分布を仮定せず、より制約の緩いモデルであることである。しかし、例えば、観測終年が91年(2番目)とすると、観測時点数は1か2、氷河期世代(70-86年)の教員もありえない(日本では大幅な飛び級は存在しない)。また、目的変数の科研費取得件数は0, 1, 2しかありえない。この制約をモデルに入れないと意味ない母数も推定しているが、現時点でこの問題は解決されていない。変量間の関係による制約まで厳密に規定すると難しく、単なるゼロ制約だけならば、理論的には最初から未知母数を設定しなければよいだけである。しかし、多項ロジスティック回帰に関するゼロ制約計算の例は見当たらず、本研究での計算では実現されていない。

全体と各分野に対して、ZIPと多項ロジット(Logit)の計算を行ったが、結果的にどちらが必ず優性とはならなかった。これは、必ずしも科研費採択数がZIP分布に沿わない一方、現時点のLogitモデルでは情報量にムダが発生し、結果両者が均衡する結果である。

また、いずれにしても13変量や各水準設定の妥当性は検証されていない。そのため、ステップワイズ(BIC変数増減法)を使うと最適モデルと違うモデルを引っ張ってしまう。そのため、LogitとZIPそれぞれで、8,192式の網羅計算、水準統合により、BICによる最適モデルを導出し、2次交互作用項を3項までは網羅計算、4項以降はBIC変数増加法で計算した。

結果の解釈は本文で述べたので割愛するが、Logitモデルでも、全体と各分野それぞれで、若い世代で、職位が上位の教員ほど、科研費代表採択と正の関係があると判明した。この相反傾向は相場観としては尤もで、本モデルで統計的に把握できた意義は大きい。

また、教員の異動回数は1以上であれば概ね回数に関係なく、採択数に正の関係があることも判明した。このことは政策的にも重要である。更に、G1大学や博士号(満期修了も含む)が正の関係が大きいと判明した。

加えて、特に分野別解析に関しては、大学教員は高度な学術的背景を要する業務であり、容易に専門分野を変えることはできない。即ち、分野の独立性が高いという背景がある。

以上から11分野に対して解析したところ、人文学や社会科学、歯歯薬学のゼロインフレーション率が高い一方、農学や複合新領域、工学では低いと判明した。

今後、更にモデル改善を図り、他制度への適用等を考えたい。

参考文献

- [1] 細坪護拳, 科学技術政策研究所 Discussion Paper No.60 (2010年2月)
- [2] M. Hosotsubo, Scientometrics (2011), Vol. 86 (2), pp.405-430
- [3] 細坪護拳, 研究・技術計画学会第25回年次学術大会講演要旨 (2010年10月)
- [4] 細坪護拳, 第15回情報・統計科学シンポジウム (2010年12月)
- [5] 細坪護拳, 応用統計学会2011年度年会 (2011年6月)
- [6] I. K. R. Rao, Journal of the American Society for Information Science (1980), Vol. 31, pp.111-122.
- [7] 山崎博敏, 広島大学 大学教育研究センター 大学論集 第11集 (1982), pp.1-21.
- [8] 藤井良宜, Rで学ぶデータサイエンス1 カテゴリカルデータ解析, 共立出版 (2010)
- [9] 辻谷奨明・竹澤邦夫, Rで学ぶデータサイエンス6 マシニング, 共立出版 (2009)

附録

本研究では、当所の論文分析[6]で国立大を論文数シェア※で4群分類

- G1: 5%以上 (論文数の概ね上位25%相当)、4大学
- G2: 1~5% (G1以外の上位50%まで)、10大学
- G3: 0.5~1% (G1・G2以外の上位75%まで)、17大学
- G4: 0.5%以下 (G1・G2・G3以外)、55大学

G1: 東北大、東京大、大阪大、京都大 (4大学)

G2: 名古屋大、北海道大、九州大、東工大、筑波大、千葉大、神戸大、広島大、金沢大、岡山大 (10大学)

G3: 愛媛大、名工大、東京農工大、静岡大、三重大、信州大、富山大、岐阜大、新潟大、東京医科歯科大、横浜国大、山形大、山口大、群馬大、長崎大、熊本大、鹿児島大 (17大学)

G4: その他 55 国立大

【群内大学順不同】

※ 論文数シェア

論文著者の所属機関所在国に対して件数を分数カウントで積算した国内シェア

例) 東大理学部、東大工学部、京大、カリフォルニア工科大の所属研究者による共著論文 ⇒ 日本: 3/4 件, 米国: 1/4 件とカウント

対象 DB: トムソンロイター社 Web of Science (Science Citation Index Expanded)