

SSM 職歴データを分析するための基礎的な方法論

保田時男
(大阪商業大学)

【要旨】

職業の系列データはSSM調査の最大の特徴であるが、これまで十分に活用されてこなかった。その一因は、分析のための基礎的な方法論が不足していることにある。本稿の目的は、行動観察における系列データの分析方法を職歴データの分析に応用し、職歴データにアプローチする分析方法の選択肢を増やすことにある。

系列データを分析するためには、最初に分析しやすい形式にデータを変換することが重要である。職歴データは状態系列からインターバル系列に変換することが有効なので、第2節ではその実際の変換方法について解説している。適切な変換のために必要となる判断は、コーディング作業と密接に関連しているので、分析段階ではなくコーディング段階で変換を行うことが望ましい。

また、膨大な情報を含む系列データであっても、通常の計量分析と同様に一変量の分布から順に分析を進めていくことが望ましい。そのために、第3節では、職業の各側面について、その発生割合・構成比率・継続期間の3つの指標を利用することを提案している。これらの指標を一貫して用いることで一変量から多変量へと連続的に分析を進めていくことができる。

キーワード：職歴、系列データ、方法論

1. 問題の所在

SSM調査（社会階層と社会移動調査）の最大の特長が、その職歴データにあることは間違いない。回顧的な記憶によるデータとはいえ、すべての職業経歴を余すところなく測定した調査データは珍しく、1955年から継続的に収集されているという点も含めて考えると、世界的にも貴重なデータである。

しかしながら、SSM職歴データに含まれる職業の系列（sequences）を利用した研究は、思いのほか少ない。つまり、ある回答者が、いつからいつまでどのような職業に就き、その後どのような職業へと移動したのかという、在職期間と移動の情報があまり活用されていない。SSM職歴データを利用した多くの研究は、職歴の中から初職や現職、40歳時職といった一部の情報だけを抜き出して利用している。もちろん、職業の系列を利用した分析もいくらか存在し、1975年データの原（1979）、1985年データの盛山（1988）、岩井（1990）、1995年データの佐藤（1998）、岩井（1998, 2000）、渡邊（2004）などが代表的なものとしてあげられる。その他の研究事例については渡邊（2004）に詳しい。このような研究が存在するとはいえ、

SSM 調査データを利用した研究の全体的な数からすれば、やはり職業の系列を利用した研究は少ないと言わざるをえない。

職業の系列が十分に活用されてこなかったのは、方法論的な理由によるものと考えられるであろう。方法論的な問題は、大きく 2 つあると考えられる。1 つは、そのデータの特異性ゆえに、分析の方法が確立されていないことである。もう 1 つは、回顧データであることから、データの信頼性についての疑問である。後者の問題については、回顧的な生活史記録の信頼性を高める 1 つの方策として LHC (ライフ・ヒストリー・カレンダー) 調査が試行され (近藤 2005)、その成果が 2005 年 SSM 調査に一部活用されている。また、前者については、原 (1979) が職歴データを「非定型データ」と呼びその扱い方を模索したことをはじめとして、しばしば議論されている (渡邊 2004 ; 保田 2005 ; 吉田 2005)。本稿で考察するのは、こちらの問題である。

扱いが難しい職歴データの分析方法として、たとえば最適マッチング法 (optimal matching methods) による系列分析 (sequence analysis) が近年注目されている。これは、DNA 配列などの分析のために生物科学で発展してきた最適マッチング法を Andrew Abbott が社会科学に応用したものであり、1990 年代に定着してきたとされる (Abbott and Tsay 2000)。SSM 職歴データに対しても渡邊 (2004) が適用している。また、その他にも職歴データの典型的な分析方法がけっして存在しなかったわけではなく、Arosio (2004) は最適マッチング法を含めて職歴データを分析する 4 つの技法の長所や短所と理論的な活用場面についてまとめている。

職歴データの分析方法についての議論が少なくないなかで、本稿で議論したいのは、より基礎的な分析方法についてである。具体的には、職歴データを分析しやすい形に定式化する方法、そのデータから従業上の地位、仕事内容など、職歴の各側面についての一変量分布から記述を始める方法について考察する。このような基礎的な分析方法についてあえて議論するのは、職歴データを扱う研究者の間でこれらの方法が共有されていないことが、極めて重大な不利益をもたらしていると考えられるためである。たとえば『2005 年 SSM 日本調査基礎集計表』では、「以下、職歴に関しては、初職 (問 7) についてのみ掲載し、問 8 (1~16 段目) の集計は省略する」(p.13) との断りのみで、職業の系列に関する基礎集計は職歴の記入枚数の分布しか掲載されていない。以前の調査でも同様の扱いである。SSM 調査の中心であるはずの職歴データの基礎集計が十分に共有されていないことは問題である。本稿では、このような状況の改善するための 1 つの選択肢を示したい。

2. 職歴データのコーディング形式：状態系列からインターバル系列へ

2.1 行動観察分野における方法の適用

SSM 職歴データが扱いにくいのは、それが膨大な系列データ (sequential data) だからであ

る。つまり、同じ回答者の同じ事項（職業）について、時系列的に連続した測定を行ったデータである。SSM 職歴データは、職業に関する複数の側面（勤務先の産業と規模、回答者の従業上の地位、仕事内容、役職）を測定しており、しかもそれらが互いに関連するので、系列データのなかでもかなり複雑なものといえる。また、2005 年 SSM 日本調査（SSM2005-J）では家族経歴（結婚や離死別、親の死亡、それぞれの子どもの誕生の時期）や教育歴（高校より上の学校の修学時期）も調べられているので、職業だけでなく、家族、教育の面でもある程度の系列データを含んでいる。

社会学者が行う社会調査は、ほとんどの場合が一時点における横断調査なので、一般に、計量社会学者は系列データの扱いには不慣れである。また、近年はパネル調査も増えてきたが、パネルデータはいくつかの時点で非連続の測定を行った結果なので、間断のない連続的な測定である系列データとはやや性格が異なる。系列データの扱い方を社会学者が正面から考察した方法論的な研究は限られており、原（1988, 1992）などがそのまれな例である。

一方で、系列データを取り扱うことが多い学問分野では、当然、その方法論の整備が進んでいる。たとえば、病状を連続的に記録する疫学の分野では、1990 年代の一時期、系列データの効率的な変換プログラムといった限定的な関心で多くの研究論文が発表されている（たとえば、Wood et al. 1997）。また、動物や人間の行動を連続的に記録する行動観察の分野では、1980 年代後半から系列データの変換方法や分析方法の標準化が進み、社会心理学等の研究に応用されている（Bakeman and Quera 1995; Bakeman and Gottman 1997）。

このような他分野の成果を援用することは、系列データの扱いに不慣れな我々にとって有効な手段と考えられる。実際に、Abbott の系列分析も生命科学分野で発展した最適マッチング法を社会科学分野に応用したものである。本稿では、行動観察の分野で発展してきた方法を、SSM 職歴データの基礎分析に応用することを試みる。行動科学と社会科学は、多くの点で統計技法の基礎を共有しており、行動観察のために発展してきた方法は社会学者にとってもなじみやすいものに感じられる。

2.2 系列データのコーディング形式

分析に先立って重要となるのは、分析のためのデータの整備である。Bakeman and Quera (1995) は系列データのコーディング形式を 4 つの水準で整理し、SDIS (Sequential Data Interchange Standard) と呼んでいる。SDIS は最近の文献 (Sayers and McGrath 2004) でも標準的な基準として引用されているので、有効な枠組みとして定着していると考えられる。職歴データについても、SDIS の枠組みから整理することが有効であろう。

SDIS における第 1 のコーディング形式は、イベント系列 (event sequences) と呼ばれ、イベントの起こった順序のみを記録するものである。第 2 のコーディング形式は、状態系列 (state sequences) と呼ばれ、順序だけでなくそのイベントがいつから始まったのかという期

間を記録する。状態系列は、どの時点でも排他的に1つのイベントが継続中であるという観察対象について用いられる（そのため、〈状態〉なのである）。その点で、第3のコーディング形式である刻時イベント系列（timed-event sequences）に比べて単純なものである。刻時イベント系列におけるイベントは排他的ではなく、複数のイベントが同時並行で発生する可能性がある。そのため、イベントの開始時点と終了時点の両方を記録する。最後のコーディング形式はインターバル系列（interval sequences）と呼ばれ、一定の時間間隔ごとに（変化の有無にかかわらず）その時の状態を記録する。職歴（特に従業上の地位の経歴）についてこれらの形式を適用した具体例については、保田（2005）がもう少し詳しく記述している。

これらのコーディング形式は、ほとんどの（あるいはすべての）情報を失うことなく互いに置き換え可能なものであり、Bakeman and Quera（1995）はそのための汎用プログラム（GSEQ: Generalized Sequential Querier）を提供している。そのような相互変換が必要なのはそれぞれのコーディング形式に長所と短所があるためであるが、一般的にはインターバル系列がもっとも汎用的で扱いやすい基本となるコーディング形式とされている（Bakeman and Quera 1995）。

SDIS の整理に従うと、SSM 職歴データは副職を除いた主な職業の系列をそれぞれの職業の開始年とともに記録しているので、状態系列のコーディングを行っていることになる。したがって、我々が初めに行うべきことは、これを一般的に扱いやすいインターバル系列に変換することである。この点は、原（1988）が「生活史グリッド」という名称で、「不定長」な職歴データを「定長」なデータに変換することを提案しているのと同じことであり、そのようなデータ形式は person-year data といった呼び名で一般化しつつある。また、吉田（2005）が述べているとおり、このような変換はこれまでの研究でも多く独自に行われてきた作業である。

しかし、ここで強調しておかなければならないことは、このような変換が分析段階で行われるのではなく、〈コーディング〉の段階として行われるべきだということである。それは次の2つの理由による。第1に、SSM 職歴データの変換は必ずしも自動的に行うことができず、いくつかの判断が必要となる。適切な判断は、データ収集により近いコーディングの段階で行われるべきである。第2に、インターバル系列への変換によって発覚する論理エラーが存在するので、変換によってデータクリーニングの必要が発生する。クリーニングの結果は、データ全体に波及するのでやはり早い段階での変換が望ましい。これらの点を明らかにするために、次節では SSM 職歴データのインターバル系列への変換の実際について説明する。

2.3 職歴データのインターバル系列への変換

筆者は、2005年SSM日本調査の職歴データを状態系列からインターバル系列へと変換するために、SPSSシンタックスによるプログラムを作成し、「SSM2005 国内調査 person-year data

変換シンタックス」という名称で公開した。社会学者の間ではGSEQのような汎用変換プログラムが共有されているわけではないので、普及度の高いSPSSを用いてSSM職歴データに特化したプログラムを作成することが最適と考えたからである。以下、このプログラムをPY05と呼ぶことにする。PY05では、表1の37種類の変数について各回答者の15～70歳時点（56歳時点）における状態を表す $37 \times 56 = 2,072$ 個の変数を生成する¹。また、回答者ではなく人年（person-year）をケースの単位とした積み上げ式のデータも別途生成する。

表1 person-year data 変換シンタックスで作成される変数

職歴		教育歴	
isnum	職歴番号	pen	学校種
isn	従業先番号	pea	卒業・中退・在学
isa	離職理由	pepa	在学期間始まり（年齢）
isb	事業内容	pec	国公私立
isbi	事業内容（isic）	ped1a	大学学部
isc	従業員数	ped1	学科系統
isd	従業上の地位	ped2	大学短大名
ise	仕事の内容	pep83	学校所在地都道府県（83年）
isei	仕事の内容（isco）	pec83	学校所在地市町村（83年）
isf	役職名	pep05	学校所在地都道府県（05年）
ispa	在職期間始まり（年齢）	pec05	学校所在地市町村（05年）
isg	収入の変化		
isqlf	取得資格数		
		家族歴	
		pfage	父親の年齢
		pmage	母親の年齢
		psage	婚姻状態・配偶者の年齢
		psacq	配偶者と知り合ってから年数
		pcnum	子ども数
		pclage	第1子の年齢
	
		pc7age	第7子の年齢
		pcyage	末子の年齢

注：各変数名の末尾には、時点年齢を表す2桁の数値が付く。積み上げ式のデータでは人年を表す変数（rage）を含む。また、psacqは留置調査票がB票を割り当てられた回答者のみの変数である。

PY05の変換アルゴリズムは単純なものである。たとえば、30歳時点における職業関連の変数の値は、職歴の初職から現職までの在職期間を順に確認していき、在職期間内に30歳を含む職業からその値を取得する。調査時点で30歳に達していない回答者の値は当然、欠損値（非該当）となる。教育歴や家族歴についても、同様である。

¹ SSM調査のサンプリング時点での対象年齢は20～69歳であるが、調査時点で70歳に達している回答者も存在する。70歳の回答者の現職までのデータをすべて包含するために上限を70歳時点とした。実際の分析では69歳時点までを対象としてもほとんど問題がない。また、下限を15歳時点としたのは、中学校卒業時からの就業を捉えるためである。543ケース（回答者全体の9.5%）が15歳時点で就業しておりその数は無視できない。一方でより若い時点から就業している回答者も存在するが、33ケース（0.6%）とごく少数であるために15歳時点からで十分と考えた。15歳は教育歴の開始年齢とも合致している。

アルゴリズムは単純であるが、変換を確定するためにはいくつかの点で判断の仕方を決定する必要がある。第1に判断が必要な問題は、在職期間に無回答（DK）を含む場合である。PY05では、職業が不明な期間の変数には最大限欠損値（職歴不明）のコードを割り当てる方針を取った。たとえば、30歳から50歳のうちの時点かは不明であるが何年かはパートで働き、あとは無職だったというような場合、この期間の職業はすべて欠損値とし、パートで働いたことはどこにも反映されない。ただし、職業の開始年齢が終了年齢の一方だけは回答されている場合、回答のあった年齢だけにはその職業を反映させている。たとえば、40歳からパートで働いていたが何歳まで働いたかは不明であるという場合、40歳時点の職業だけをパートとし、41歳以降の職業を欠損値としている。

第2に判断が必要となるのは、ちょうど転職した年齢で前後のどちらの職業を採用するかという問題である。PY05では常に後の職業を採用することにした。たとえば、30歳で臨時雇用から常時雇用に転職した場合、30歳の職業は常時雇用となる。同じように、その職業を40歳で辞め無職になった場合には、40歳の職業は無職となる。後の状態を採用するという方針は教育歴にも適用されているので、たとえば18歳から22歳まで大学に在学していた場合、PY05では18～21歳までが学生となり、22歳は卒業後の状態（学生ではない）が割り当てられている²。

第3に、同じ年齢のうちに複数の転職を経験した場合の判断が必要になる。PY05では先に示した「後の職業を採用する」という方針を踏襲し、その年齢の最後の職業を採用している。たとえば、それまで無職で40歳のときに職についたが、40歳のうちにまた無職に戻ったという場合、PY05ではずっと無職となり、職に就いたという情報は職歴番号（isnum）の変化にしか反映されない。同じ年齢で複数の転職を経験している回答者は473ケース（8.2%）あり、けっして無視できない。ただし、その大部分は短い無職期間を回答しているものであり（2005年SSM日本調査では3カ月以上の無職期間については回答を求めている）、その点にこだわる分析目的でない限りは、大きな問題とはならないであろう。

第4に、インターバルの基準として各時点での回答者の年齢を採用するのか、暦年を採用するのかという判断が必要になる。すでに自明なとおり、PY05では回答者の年齢を採用し

² 在学期間の判断の仕方は、分析関心によっては重要となる。たとえば、大学の在学期間が18～22歳までで、23歳から初職に就いたと回答した場合、PY05では22歳に学生でも就労者でもない無職の期間が存在したことになるが、これが事実と合致しているかどうかには疑問の余地がある。1年留年した後に就職したという意味で回答している可能性もあるし、実際には4年間で卒業し22歳で就職したが、卒業後すぐという意味で1歳上の年齢で就職したと回答している可能性もある。クリーニングにあたっては、香山めい氏との協力で、卒業・中退の情報（面接票 問19(2)a）と学卒後すぐに就職したかという情報（面接票 問7e）をもとに、可能な限り補正を行ったが、不明な点が多い。職歴や教育歴の記録は、調査票では暦年を中心に設計されていたが、実際の回答は年齢を中心になされていたことが問題の一因と考えられる。特に在学期間の回答には多く混乱が見られたので、改善の余地がある。何歳で入学し、何年間で学校を出たか（1年未満の端数は切り捨て）を尋ねる方が信頼性の高いデータが得られるように思える。

また、高校の在学期間は尋ねられていないので、便宜的に一律15～17歳としている。ただし、中退の場合、あるいは中退か卒業か不明の場合は、やはり便宜的に15歳のみを在学とし、16～17歳の教育歴を不明としている。

ている。暦年を採用することの利点も皆無ではないが、幅広い出生コーホートを対象とする調査データで暦年を採用することは煩雑であろう³。

もとなる状態系列のデータに不整合がなければ、インターバル系列への変換において問題となる主な判断は、以上のようにそう多くはない。しかし、データに不整合があった場合の対応を含めると個別に膨大な判断が必要となる。PY05 の作成においては、その途中で判明したデータの不整合については、クリーニング担当者と連絡を取り合い修正することができた。そのため、最終的には不整合のないデータを前提としてプログラムを単純化することができた。先に述べたとおり、コーディングの過程としてインターバル系列への変換を行うことが重要である⁴。

3. 職歴データの基本指標：発生割合・構成比率・継続期間

3.1 行動観察における系列データの RPD 指標

では次に、インターバル系列に変換された職歴データの分析方法へと視点を移そう。我々が調査データを計量的に分析する際には、分析の目的にかかわらず、まず関心のある変数の一変量分布を確認する。つまり、度数分布表を作成したり、平均値や標準偏差といった要約統計量を算出したりする。ところが、系列データで単純に同じことを行うことは難しい。状態系列のデータ（つまり変換されていない職歴データ）の場合には、1 つ目の職業、2 つ目の職業、3 つ目の……と集計を繰り返すことができるが、その結果得られるものはあまり有効ではない。一方で、インターバル系列に変換したデータであれば、各年齢時点における職業について集計を行うことができる。これはいくらか有効な情報であるが、変数の数（観察年齢の数）は膨大なので、要約的でない。

系列データである職歴については、度数分布表の作成や要約統計量の算出に該当する段階としてどのような分析を行うべきなのであろうか。その方法論的な共通認識がないために、『基礎集計表』には職歴の集計結果が掲載されておらず、それぞれの分析者がこの段階から分析を行う（あるいはその分析方法を模索する）ことになってしまう。

³ 期間情報には、もともとは暦年で回答されているものも含まれることに、注意が必要である。在職期間をはじめとして 2005 年 SSM 日本調査では多くの期間や時点を回答者の年齢か暦年のいずれか回答しやすい方で尋ねる方法を採用している。多くの場合、その両方または年齢のみで回答が得られているが、一部は暦年のみで回答されている。プロジェクト全体の方針として暦年の回答は年齢に置き換えてコーディングがなされているが、当然正確な年齢からは 1 歳の誤差が発生することがある。回答者の思い違いも含めて考えると、1、2 歳程度の誤差は頻繁に発生していると考えるべきであろう。PY05 では回答者の年齢を単位として 1 歳刻みのインターバル系列を生成しているが、1 歳刻みの厳密さを要求する分析を実際に行うことは難しい。

⁴ ただし、データの確定前であることから思いもよらない問題が発生することもある。PY05 の作成にあたっては、クリーニングの過程でプログラムに利用していた変数の定義が変わったり、利用していた変数そのものがなくなったりすることがあった。また、積み上げ式のデータの生成においてはケース数（有効回答者数）が重要となるが、有効回答票数にも動きがあった。このため、7 月 18 日時点で公開したプログラムでは、ver.4.2 以降のデータに対して積み上げ式データが正しく生成されない。

ここでもやはり、行動観察における系列データの分析方法を参考にすることは有効である。系列データにおいてあるイベントの発生状況を記述する（つまり、系列データにおけるあるイベントの一変量分布を把握する）場合、そのイベントの発生割合・構成比率・継続期間（rates, probabilities, durations）の3つを指標として、まずその指標の分布を集計することが一般的である（Bakeman and Gottman 1997; Sayers and McGrath 2004）。

たとえば、職歴データにおいて、常時雇用の仕事をするというイベントを例に考えてみよう。各回答者がどのくらい常時雇用職に就いているかを測るもっとも単純な方法は、常時雇用の仕事に就いた回数を数えることである。しかし、系列データでは、その観察期間がケースごとにまちまちであることが多いので、合計回数を単位時間あたりの発生割合に置き換えることが普通である。1年あたりの〈発生割合〉に置き換えるならば、たとえば、20年間の職歴のなかで3回常時雇用の仕事に就いた回答者は、発生割合が0.15回／年となる。

もう1つの指標は、イベントの〈構成比率〉に着目するものである。構成比率はイベント単位と時間単位の2種類が考えられる。イベント単位の構成比率は、相互に排他的で包括的なイベント群のなかで当該イベントの発生回数占める割合である。つまり、ある回答者が「常時雇用→無職→常時雇用→臨時雇用→常時雇用」と移動していれば、イベント単位での常時雇用の構成比率は $3/5=.6$ （60%）となる。一方、時間単位の構成比率は全観察期間に占める、そのイベントが発生していた期間が占める割合である。つまり、ある回答者が「常時雇用1年→無職3年→常時雇用4年→臨時雇用2年→常時雇用10年」と移動していれば、時間単位での常時雇用の構成比率は $(1+4+10)/(1+3+4+2+10)=.75$ （75%）となる。構成比率は同時にある観察時点でその回答者が常時雇用である確率も表していることになるので、通常は観察確率（probabilities）と表現される。しかし、職歴の分析について考える場合には、むしろ構成比率と表現した方が発生割合との違いがわかりやすいように思われるので、本稿では観察確率ではなく構成比率と表現することにする。また、職歴の場合に限れば、イベント単位ではなく時間単位の構成比率の方がほとんどの分析目的にとって適切であろうから、以下では構成比率と書けば時間単位の構成比率を指すものとする。

最後の指標である〈継続期間〉は、当該イベントが1回あたり平均どれだけ持続したかを表す。先ほどの移動の例の場合、 $(1+4+10)/3=5$ 年/回となる。当該イベントが発生していなければ、当然、継続期間は算出できない。以下では簡略のために、発生割合・構成比率・持続期間の3つの指標をまとめてRPD指標（rate, probability, durationの頭文字）と呼ぶことにする。

3つのRPD指標は、互いに関連し合っている。ある回答者が職歴全体の中で常時雇用になった回数を C 、常時雇用であった年数を Y 、職歴全体の観察年数を T と表すならば、3つの指標は次のように算出される。

$$\text{発生割合} = \frac{C}{T}, \quad \text{構成比率} = \frac{Y}{T}, \quad \text{継続期間} = \frac{Y}{C}$$

したがって、3つの指標の間には、構成比率＝発生割合×継続期間という単純な等式が成り立っている。このことから、3つの指標のうち2つの集計結果を示せば残る1つの集計結果もほぼ推測できるので、3つすべてを示すことは冗長である⁵。たとえば、あるイベントについて発生割合は低い構成比率が高いならば、1回の継続期間が長いものと推測できる。

3.2 RPD 指標による一変量分析

具体的に、職歴データについて RPD 指標（発生割合・構成比率・継続期間）を適用した基礎分析を行ってみよう。インターバル系列のデータが整ってさえいれば、RPD 指標の算出にはほとんど手間がかからない。前節で示したとおり、これらの指標は、イベントの起こった回数（C）、イベントが占める年数（Y）、集計対象とした全体年数（T）という3つの値から構成されているので、各回答者についてまずこれら3つの値を導き、それぞれ組み合わせればよい。

たとえば、職業の各側面のなかでも従業上の地位に注目して考えてみよう。従業上の地位は常時雇用、臨時雇用などの選択肢をもつ1つのカテゴリ変数であるが、RPD 指標の算出においては、それぞれのカテゴリを1つのイベントとして別々に考える必要がある。かりに常時雇用に焦点を絞ることにしよう。前節で示したように3つの指標をすべて示すことは冗長であるが、ここでは例示のためにすべての指標を算出する。

インターバル系列のデータを前提にすれば、イベントの起こった回数（C）は、15歳時から順に従業上の地位を確認し、値が常時雇用になった回数（前の年齢時から値が2に変化した回数）をカウントすればよい。イベントが占める年数（Y）は、やはり15歳時から順に従業上の地位の値を確認し、常時雇用であった年数をカウントする。集計対象とした全体年数（T）は、15歳時から順に回答者が該当年齢に達していた年数をカウントする。これらの値をもとに、常時雇用の発生割合（ C/T ）、構成比率（ Y/T ）、継続期間（ Y/C ）が算出できる。SPSS シンタックスであれば、たとえば Appendix A のようなプログラムで常時雇用の RPD 指標が算出できる。

図1のヒストグラムは、常時雇用のRPD指標の度数分布を表している。単純に全体サンプルの全職歴（15～70歳時）を集計対象とした結果であり、その分布には雑多な攪乱要因が影響しているが⁶、この分布には職歴データから得られる常時雇用職の全体的な特徴が反映さ

⁵ ただし、この等式はそれぞれの指標の集計データ（平均値）の間では成り立たない。なぜならば、当該のイベントが1回も発生していない場合（ $C \neq 0$ の場合）には、継続期間が算出されないで、継続期間の平均値は他の指標とは対象が異なるからである。「ほぼ推測できる」と述べているのはそのためである。

⁶ たとえば、初職に就く前の学生の期間も集計対象に含まれているので、入職後に限った場合と比べれば、その発生割合や構成比率は低く見積もられている。また、回答者の年齢には幅があるので、70歳までのほとんどの期間をまっとうしている回答者もいれば、まだ数年の職歴しかもっていない回答者もいる。その

れている。1 年あたりの発生割合は、さすがに頻度が少なく読み取りにくいので、ここでは値を 40 倍し 40 年あたりの発生割合に換算している。40 年はおおまかに一生の職歴期間を想定している。図 1 から発生割合を確かめると、生涯で常時雇用就くことはせいぜい 1～2 回程度であり、あらためて 3 回も 4 回も就くことはまれであることがわかる。構成比率については、職歴の 1～2 割程度を占める低比率の場合と、7 割以上を占める高比率の 2 つの山があることがわかる。1 回の常時雇用の継続期間は 15 年程度までの場合が大半を占めるが、40 年以上の継続もまれではない。継続期間が短い方に偏っているのは主に若い回答者の職歴データが打ち切られているためであろう。

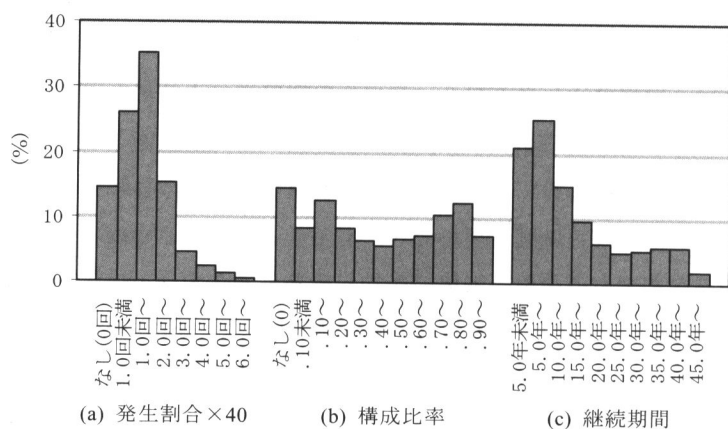


図 1 常時雇用の RPD 指標の分布

注：SSM2005-J (ver.4.2) の 15～70 歳時の職歴について、常時雇用の発生割合・構成比率・継続期間の分布を示している (n=5742)。1 年あたりの発生割合は値が小さく読み取りにくいので、40 倍して 40 年の職歴における平均発生回数に換算している。継続期間は常時雇用職に一度でも就いたことがある回答者だけが集計対象である (n=4903)。

他の従業上の地位の RPD 指標と比較すれば、常時雇用の特徴はよりわかりやすくなる。1 つ 1 つの従業上の地位について分布を細かく観察することは煩雑なので、平均値と標準偏差で要約し、表 2 にまとめた。無職を別にすれば、常時雇用の平均発生割合はすべての従業上の地位のなかで最大であり、ついで値の高い臨時雇用と比べても 2.5 倍程度も頻繁に発生するポピュラーな職歴であることがわかる。また、その平均構成比率から職歴全体の 4 割以上を常時雇用が占めており、この比率は臨時雇用の 5 倍以上に及ぶ。継続期間は、経営者や自営業、家族従業者と並んで長い部類である。それぞれのばらつき具合について確認すると、常時雇用に限らず、いずれも平均値の規模に比して標準偏差がかなり大きい。つまり、0 や 0 に近い値をもつ回答者（その地位にまったく、あるいはほとんど就かない者）が相当に多いことを示唆している。そのなかでは、常時雇用の各指標の標準偏差は相対的に小さく、やは

ため、多くの回答者が集計対象となる低年齢時の特徴ほど、指標の値を強く規定している。

り他の従業上の地位に比べれば多くの人々が同様に経験するポピュラーな職歴であることがわかる⁷。

表 2 各従業上の地位についての RPD 指標の要約統計量（平均値・標準偏差）

	発生割合×40	構成比率	継続期間			
経営者・役員	.070 (.280)	.025 (.114)	15.7	(12.1)	n=374	
常時雇用	1.447 (1.160)	.427 (.328)	15.3	(12.8)	n=4903	
臨時雇用	.589 (1.160)	.078 (.156)	6.8	(6.9)	n=1894	
派遣社員	.072 (.480)	.007 (.046)	4.4	(3.9)	n=188	
契約社員・嘱託	.077 (.400)	.010 (.058)	6.3	(6.4)	n=272	
自営業主・自由業者	.122 (.360)	.051 (.174)	19.1	(14.1)	n=650	
家族従業者	.147 (.480)	.053 (.172)	16.6	(13.0)	n=718	
内職	.024 (.200)	.005 (.048)	10.3	(10.2)	n=109	
無職（学生含む）	2.175 (1.520)	.334 (.270)	6.8	(6.7)	n=5509	
DK	.027 (.200)	.011 (.087)	19.8	(15.6)	n=118	

注 1：SSM2005-J（ver.4.2）の 15～70 歳時の従業上の地位について、各指標の平均値と標準偏差（括弧内）を示している（n=5742）。継続期間は、当該の従業上の地位になったことがある回答者のみから算出されるので、別途分析ケース数を示している。

注 2：在学期間も分析対象に含むので、義務教育終了後に進学した場合には無職期間が 1 回発生する。

従業上の地位以外の職業の各側面についても、RPD 指標の平均値を算出し Appendix B にまとめた。そこに示される特徴はおよそ常識的なものであり、即座にはあまり貴重な情報とはならないが、職歴データに RPD 指標を用いることは少なくとも次の 3 つの点で有益である。第 1 に、これらの指標は職歴データ全体の様子を要約している。繰り返し述べるが、『基礎集計表』には職歴を要約した集計結果がまったく掲載されていない。職歴データ全体の RPD 指標は、いままさに現代日本に生活している人々がもつ職業経験の布置を表している。このような基礎資料を共有することは重要である。第 2 に、PRD 指標は職歴の特徴を数量的に表している。職歴のカテゴリカルな分類をもとにした分析方法も有益であるが、属性による違いや時代による変化の方向性や程度を簡便に表すには数量的な指標の方が向いている。数量的な RPD 指標は、分析方法の選択の幅を広げてくれる。第 3 に、上の 2 つから当然導き出される点として、RPD 指標は数量的な多変量解析へと連続的に繋げることができる。図 1 や表 2 では全体サンプルの全職歴について RPD 指標の単純集計（一変量分布）を示したが、当然、回答者の属性別に指標の平均値を比較することや、職歴全体ではなくいくつかの年齢幅に区切って指標の値を比較することもできる。また、他の変数との相関を分析することや、複数の RPD 指標の間での相関を分析することもでき、回帰分析等によるモデル化も可能である。つまり、我々が通常慣れ親しんでいる変数志向の分析発想がそのまま適用できる。

⁷ 相対的なばらつきの規模を変動係数で比較すると、臨時雇用の変動係数は $1.156/.078=2.000$ なのに対し、常時雇用の構成比率は $1.328/.427=.768$ と比較的小さい。

最後の点は特に重要である。変数志向の発想では明らかにできない現象が存在することは事実であるが、それは変数志向の分析を行わない理由にはならない。現在、職歴データの分析に不足しているのはオーソドックスな変数志向の分析に思える。Pollock (2007) は Abbott の系列分析をベースにして、複数の変数（職歴、居住歴、家族歴）からなる系列を扱う multiple-sequence analysis を提唱しているが、それは系列分析を変数志向の多変量解析へと引き戻す試みであると主張している (pp.178-179)。RPD 指標の利用は、その端緒は異なるもののこの主張に呼応するものである。

3.3 属性グループによる RPD 指標の比較

変数志向の発想で RPD 指標を用いる例を具体的に示そう。図 2 は職歴全体ではなく 15～34 歳時に範囲を絞った上で、男女別、出生コーホート別に常時雇用の RPD 指標を算出し、それらの平均値をプロットしたものである。職歴の初期に常時雇用に就くことについて、男女別によどのような時代的变化があったかを表している。男性の指標には大きな変化は認められないが、女性の指標はこの 30 年ほどの間の変化を的確に示している。はじめの 10 年ほどの間に女性の常時雇用の発生割合は上昇し、男性と変わらない程度になった。その結果、構成比率も当初の 2 倍の 40% 程度に跳ね上がったが、それでも男性の 3 分の 2 程度に留まっている。なぜならば、その継続期間にほとんど変化が見られず、男性に比べて 4～5 年ほど短いままだからである。つまり、女性が常時雇用に就く機会は増えたが、短い期間で職を離れるという特徴にはまったく変化が見られないということである。

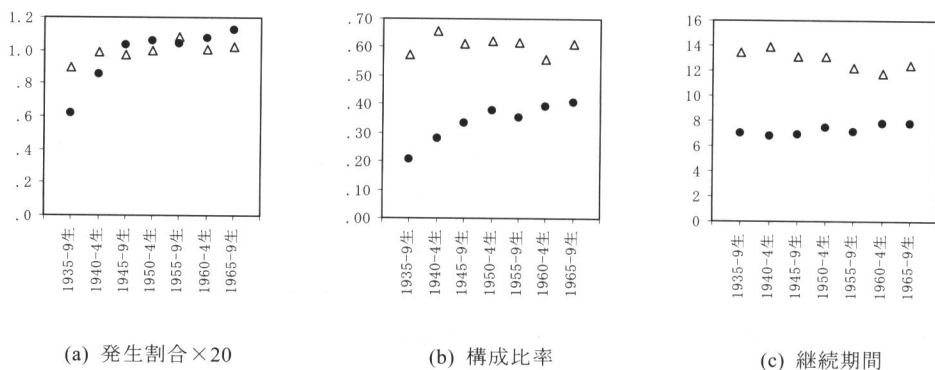


図 2 常時雇用の RPD 指標の平均値 (15～34 歳時、男女・出生コーホート別)

注 1 : △ 男性 ● 女性

注 2 : SSM2005-J (ver.4.2) から 15～34 歳時の職歴について、常時雇用の発生割合 (20 年あたりの回数)・構成比率・継続期間の平均値を、男女・出生コーホート別に示している。集計対象は、調査時点で 34 歳までの職歴を終えている 1935～1969 年生まれの回答者 (n=4452) で、継続期間については常時雇用職に一度でも就いたことがある回答者だけが集計対象である (n=3826)。

図3は、同様の分析を臨時雇用についても行った結果である。やはり、女性の変化が大きく、常時雇用よりも遅れてその発生割合が急激に上昇している。その結果、構成比率も最近のコホートで急増しており、男性との差は広がっている。常時雇用と同様に継続期間には大きな変化が見られないが、常時雇用の場合と異なり男性も同じ程度の継続期間である。つまり、臨時雇用に就いた場合の継続は、男性も女性も同様に短いという特徴に変化は見られないが、女性が臨時雇用に就く機会だけが後半15年ほどの間に急増している。

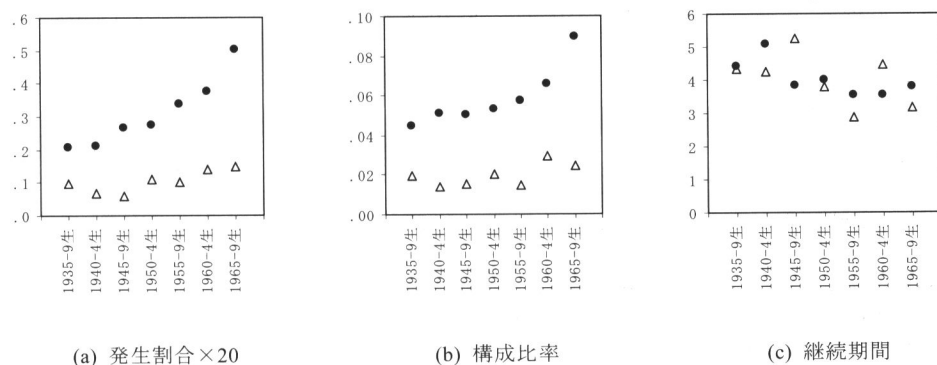


図3 臨時雇用のRPD指標の平均値（15～34歳時、男女・出生コホート別）

注1：△ 男性 ● 女性

注2：SSM2005-J (ver.4.2) から15～34歳時の職歴について、臨時雇用の発生割合（20年あたりの回数）・構成比率・継続期間の平均値を、男女・出生コホート別に示している。集計対象は、調査時点で34歳までの職歴を終えている1935～1969年生まれの回答者（ $n=4452$ ）で、継続期間については臨時雇用職に一度でも就いたことがある回答者だけが集計対象である（ $n=809$ ）。男性は継続期間の集計対象が少なく、コホート別によると21～31人しかいないため、標本誤差が大きいことに注意が必要である。

3.4 RPD指標による世代内移動の分析

職業の系列データを利用することの重要な利点は、各回答者の水準で世代内移動を捉えることができることであるが、RPD指標で直接的に世代内移動の分析を行うことは難しい。RPD指標は一定の期間の特徴を要約しているため、期間内の変化（移動）については記述してくれないからである。しかしながら、期間を細かく区切り指標の値の推移を観察すれば、間接的にはあるが世代内移動の様子を記述することができる。

図4はそのような分析の例である。女性の常時雇用に焦点を絞り、出生コホート別にRPD指標の5歳刻みの推移を表している。発生割合のグラフからわかるように、どのコホートでも常時雇用の発生は最初の5年間にほぼ限定されている（新しいコホートほど、その発生割合は高くなっている）。したがって、年齢層が高くなるにつれて構成比率が低くなっているのは主に流出傾向が続いているためである。5年間に占める常時雇用の継続期間からは、

より明確に常時雇用の安定性の変化が読み取れる。平均継続期間が短いほど、5 年間の途中で流出した（あるいは流入した）女性が多いことを示すので、加齢に伴って安定性は増している。

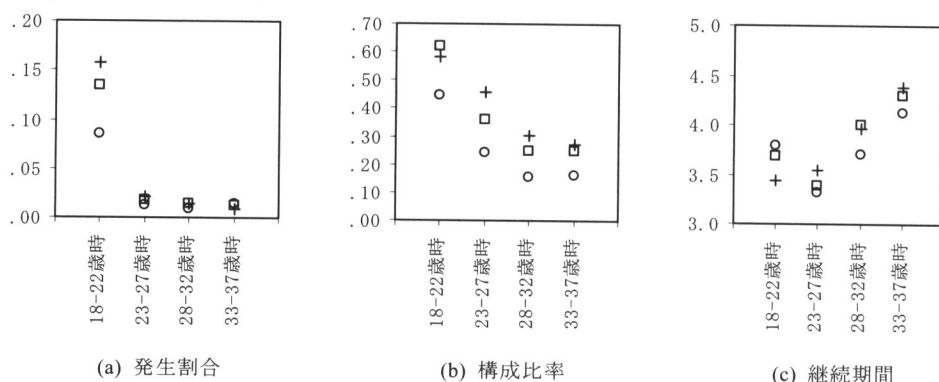


図4 女性の常時雇用についての5歳刻みのRPD指標（18～37歳時、出生コホート別）

注1：○ 1935～44年生まれ □ 1945～54年生まれ + 1955～64年生まれ

注2：SSM2005-J (ver.4.2) から女性の18～37歳の職歴について、常時雇用の発生割合（1年あたり）、構成比率、継続期間の平均値の5歳刻みの推移を、出生コホート別に示している。発生割合・構成比率の集計対象者数は、1935～44年生まれから順に、1190、781、614である。継続期間の集計対象者は、グループごとに異なるが、145（1935～44年生まれで33～37歳時に常時雇用就いていた女性）が最小である。

注3：ここでの発生割合は、観察前の年齢時点を含めた新規の発生回数だけをカウントして算出した。つまり、17歳のときにも常時雇用だった女性が18歳時に常時雇用であったとしても、それは回数にはカウントされない。一方で、継続期間の算出においては観察対象の5年間の中で常時雇用を経験した回数を分母としているので、ここでは発生割合と継続期間における発生回数（C）は別のものである。

以上のようにRPD指標からは、間接的にいつどのような世代内移動が起こっているのかを読み取ることができる。ここでは常時雇用をイベントとして集計を行ったが、イベントの定義を工夫すれば幅広い移動をRPD指標で捉えることができるであろう。

3.5 類似の方法との関係

RPD指標を用いた世代内移動の分析方法は、従来の方法とどのような関係にあるのか、最後に若干の考察を加えたい。RPD指標のなかでも構成比率は従来から比較的良好に用いられてきた指標である。Arosio (2004) がclass time budgetの研究としてまとめている分析方法は、まさに職業の構成比率を出身階層やコホート別に比較するものである。SSM職歴データに関しては、岩井が各年齢時点における構成比率を属性別に集計することで平均的なライフコースを記述している（岩井 1990, 1998, 2000）。また、発生割合という指標もしばしば用いられている。たとえば、吉田 (2005) は各年齢時点における各種の移動の発生率を属性別に集計することで、それぞれの移動のプロフィールを記述している。これは、ある種の移動をイ

ベントと定義してその発生割合を算出しているものとみなすことができる。

このようにすでに類似の方法があるなかで、本稿で RPD 指標を提案することの意義を 2 点強調しておきたい。第 1 に、RPD 指標はそれぞれの年齢時点ではなく一定の幅をもつ期間について記述することを前提としている。この点が系列データとしての職歴の特徴を表すためには重要である。幅を最小（1 歳）にした場合、発生割合と構成比率は同義となり、また継続期間は必ず 1 年なので意味をなさない。観察期間に幅をもたせることで職歴を 3 つの方向から多角的に捉えることができるようになる。

たとえば図 4 と同じく、1935～64 年生まれの女性を 3 つの出生コーホートにわけ、各年齢時において常時雇用に就いていた回答者の割合を算出すると、図 5 のようになる。これは図 4（b）の構成比率よりも詳細で、特に若年時の分布には、高学歴化の影響による変化がよく表れている。しかし、一方で 23 歳以降では新規の常時雇用の発生が非常に少ないこと（発生割合の激減）や、30 代に入ってから常時雇用の安定性が増し続けていること（継続期間の上昇）は、この図からは読み取れない。

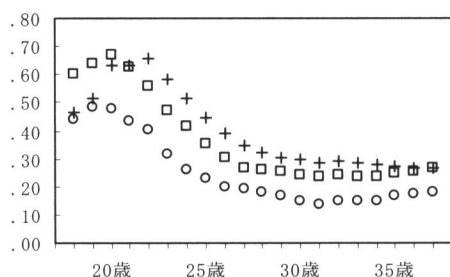


図 5 各年齢時における常時雇用の女性の割合（18～37 歳時、出生コーホート別）

注 1：○ 1935～44 年生まれ □ 1945～54 年生まれ + 1955～64 年生まれ

注 2：SSM2005-J（ver.4.2）から女性の 18～37 歳の職歴について、常時雇用の発生割合（1 年あたり）、構成比率、継続期間の平均値の 1 歳刻みの推移を、出生コーホート別に示している。集計対象者は図 4 と同様である。

また、1 歳刻みで指標を算出することには、若干危険な面もあることを指摘しておきたい。図 6 は、1 歳刻みで何らかの職業の変化があった（調査票で新しいページが使用された）件数を集計したものである。黒く塗りつぶしている棒は、切りのよい年齢（30 歳、35 歳など）での発生件数を示しているが、一見してわかるように特に高齢の回答者では切りのよい年齢に回答が集まる傾向が読み取れる。定年退職を別にすれば、これはおそらく事実在即したものではないであろう。記憶があいまいなため「35 歳のころ」というように切りのよい年齢に回答が偏っていることを意味していると考えられる。1 歳刻みの集計には一定のバイアスがかかっている可能性には注意すべきであろう⁸。

⁸ 吉田（2005）は、前後 2 年ずつの移動平均をとることでグラフを平滑化し、傾向を読み取りやすくしてい

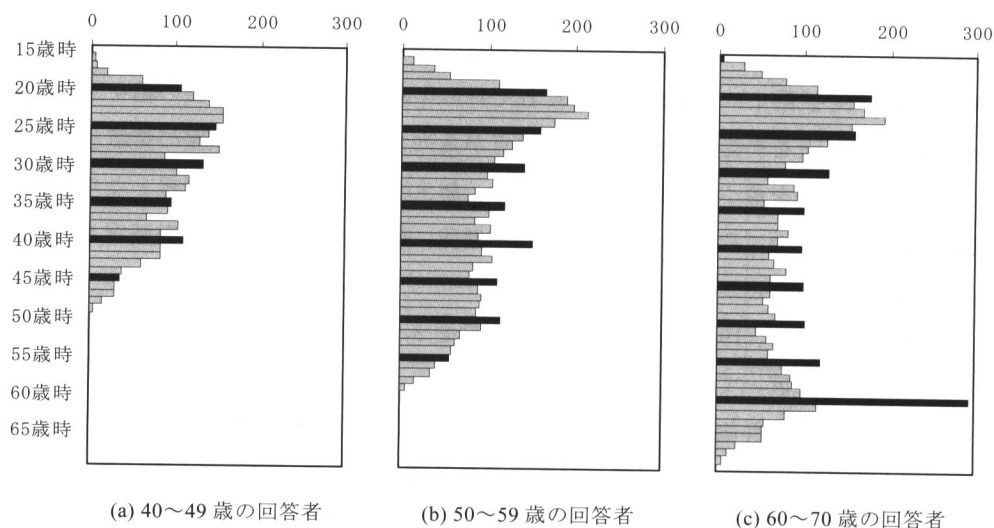


図 6 各年齢時における職業の変化の発生件数（調査時点の年齢層別）

注：SSM2005-J（ver.4.2）の40代・50代・60代の回答者について、新しい職業（2つ目以降の職業）が発生した件数を発生年齢別にカウントした（順に、 $n=1109$ 、 1424 、 1523 ）。

RPD 指標を提案することのもう 1 つの意義は、3 つの指標が明確な形で互に関連し合っているという点である。それぞれの指標には、それぞれに目的があるので構成比率だけでは表せない側面があるのは当然のことであり、不足な面は他の指標で補えばよい。しかし、その際にばらばらの出所の指標を活用するよりは、ひとまとまりの関連性をもつ一群の指標を用いる方が分析結果の解釈や発展が容易になるであろう。RPD 指標の統一性は、職歴データの分析でも有効に働くと考えられる。

4. まとめ

本稿では、職歴データの基礎的な分析を行うための方法論として 2 つの点について考察した。第 1 の点は、系列データである職歴を状態系列から分析しやすいインターバル系列に変換する方法についてである。変換は自動的に行うことはできず、いくつかの判断が必要となること、また変換に伴ってクリーニングの必要が発生することを考慮すると、コーディング段階での変換が望ましいことを強調した。

第 2 に、一変量分布の確認から多変量解析までを一貫して行うために、基礎的な分析から利用することができる指標について考察した。行動観察で用いられている RPD 指標（発生割合・構成比率・継続期間）は、汎用的に利用することができる有効な指標であることを示し

るが、この操作は切りのよい年齢へのバイアスを緩和する上で有効と考えられる。

た。

当然、インターバル系列にしても RPD 指標にしても万能ではない。しかし、職歴データの分析に利用できる選択肢を増やすことは重要である。特に、本稿で示した方法は、基礎的な分析から活用できるものであり、職歴データ分析の入り口を広げる上で重要な一步であると考えている。

【文献】

- Abbott, Andrew and Angela Tsay. 2000. "Sequence Analysis and Optimal Matching Methods in Sociology: Review and Prospect." *Sociological Methods & Research* 29(1): 3-33.
- Arosio, Laura. 2004. "Occupational Careers and Longitudinal Data: Tools and Perspectives of Research." *Quality & Quantity* 38: 435-56.
- Bakeman, Roger and John M. Gottman. 1997. *Observing Interaction: An Introduction to Sequential Analysis, Second Edition*. Cambridge University Press.
- Bakeman, Roger and Vincenc Quera. 1995. *Analyzing Interaction: Sequential Analysis with SDIS and GSEQ*. Cambridge University Press.
- 原純輔. 1979. 「職業経歴の分析」 富永健一編 『日本の階層構造』 東京大学出版会: 198-231.
- . 1988. 「非定形データの処理・分析」 海野道郎・原純輔・和田修一編 『数理社会学の展開』 数理社会学研究会: 461-471.
- . 1992. 「定型データと非定型データ」 原純輔編 『非定型データの処理・分析法に関する基礎的研究（文部科学省科学研究費報告書）』 :1-13.
- 岩井八郎. 1990. 「高度成長期以降の学歴とライフコース」 『教育社会学研究』 46: 71-95.
- . 1998. 「女性のライフコースの動態：日米比較研究」 岩井八郎編 『1995 年 SSM 調査シリーズ 13 ジェンダーとライフコース』 1995 年 SSM 調査研究会: 1-29.
- . 2000. 「近代階層理論の浸透：高度成長期以降のライフコースと教育」 近藤博之編 『日本の階層システム 3 戦後日本の教育社会』 : 199-220.
- 近藤博之編. 2005. 『ライフヒストリーの計量社会学的研究（文部科学省科学研究費報告書）』
- Pollock, Gary. 2007. "Holistic trajectories: a study of combined employment, housing and family careers by using multiple-sequence analysis." *Journal of Royal Statistical Society: Series A* 170(1): 167-183.
- Sayers, Steven L. and Kathleen, McGrath. 2004. "Data Analytic Strategies for Couple Observational Coding Systems." Pp. 43-63 in *Couple Observational Coding Systems*, edited by Patricia K. Kerig and Donald H. Baucom. Lawrence Erlbaum Associates.
- 佐藤嘉倫. 1998. 「戦後日本における職業移動パターンの変遷」 佐藤嘉倫編 『1995 年 SSM 調査シリーズ 3 社会移動とキャリア分析』 1995 年 SSM 調査研究会: 45-64.
- 盛山和夫. 1988. 「職歴移動の分析」 今田高俊・盛山和夫編 『1985 年社会階層と社会移動全国調査報告書 1 社会階層の構造と過程』 1985 年社会階層と社会移動全国調査委員会: 251-305.
- 渡邊勉. 2004. 「職歴パターンの分析：最適マッチング分析による分析」 『理論と方法』 19(2): 213-234.
- Wood, Joy, David Richardson, and Steve Wing. 1997. "A Simple Program to Create Exact Person-Time Data in Cohort Analyses." *International Journal of Epidemiology* 26(2): 395-99.
- 保田時男. 2005. 「ライフ・ヒストリー系列データへのラグ系列分析の適用」 近藤博之編 『ライフヒストリーの計量社会学的研究（文部科学省科学研究費報告書）』 : 23-33.
- 吉田崇. 2005. 「学歴移動プロフィールの記述：LHC 調査をもちいた定型分析」 近藤博之編 『ライフヒストリーの計量社会学的研究（文部科学省科学研究費報告書）』 : 51-61.

Appendix A インターバル系列から RPD 指標を算出するシンタックス

```
comp c = 0. /* 常時雇用になった回数.
comp y = 0. /* 常時雇用であった年数.
comp t = 0. /* 総観察年数.

do repeat x = isd15 to isd70 /*15歳～70歳時の状態を順に参照*/
  / lg = 777777 isd15 to isd69. /*1つ前の年の状態を参照.
  if (x = 2) y = y + 1. /*常時雇用であれば年数を加算.
  if (x = 2 & lg ^= 2) c = c + 1. /*常時雇用になった年であれば回数を加算.
  if (x ^= 888888) t = t + 1. /*該当年齢に達していれば観察年数を加算（注）.
end repeat.

do if (t = 0). /* 観察年数が0の場合、発生割合と構成比率は非該当.
  comp rt = 88.
  comp pr = 88.
else.
  comp rt = c/t.
  comp pr = y/t.
end if.
do if (c = 0). /* 常時雇用になった回数が0の場合、継続期間は非該当.
  comp dr = 88.
else.
  comp dr = y/c.
end if.
missing values rt pr dr (88).
exe.

freq rt pr dr /statistics = mean stdev.
```

注：観察年数の数え方は分析目的によって異なってくる。たとえば、無職や学生の期間を除いて集計するのであれば、if(x ^= 9 & x ^= 10 & x ^= 777777 & x ^= 888888) t = t + 1. となる。

Appendix B 職歴全体の PRD 指標の平均値

a) 従業上の地位	発生割合 × 40	構成 比率	継続 期間	継続期間 の n
経営者・役員	.07	.025	15.7	374
常時雇用	1.45	.427	15.3	4903
臨時雇用	.59	.078	6.8	1894
派遣社員	.07	.007	4.4	188
契約社員・嘱託	.08	.010	6.3	272
自営業主・自由業者	.12	.051	19.1	650
家族従業者	.15	.053	16.6	718
内職	.02	.005	10.3	109
無職（学生含む）	2.17	.334	6.8	5509
DK	.03	.011	19.8	118

b) 仕事内容	発生割合 × 40	構成 比率	継続 期間	継続期間 の n
専門	.63	.136	14.1	1003
管理	.12	.034	13.4	472
事務	1.66	.266	10.8	2385
販売	.99	.148	10.5	1620
熟練	.74	.166	15.1	1584
半熟練	.73	.139	12.0	1529
非熟練	.38	.051	7.9	807
農業	.13	.044	19.0	449
DK	.06	.016	18.3	138

c) 役職	発生割合 × 40	構成 比率	継続 期間	継続期間 の n
役職なし	4.14	.849	16.1	5491
監督等	.18	.035	10.0	535
係長等	.16	.031	8.7	572
課長等	.12	.027	10.0	484
部長等	.05	.011	9.6	219
社長等	.09	.030	16.2	347
DK	.07	.017	16.9	172

d) 産業	発生割合 × 40	構成 比率	継続 期間	継続期間 の n
農業	.12	.038	19.3	389
林業	.01	.002	15.5	28
漁業	.02	.005	16.3	59
鉱業	.01	.001	11.4	24
建設業	.35	.081	16.2	762
製造業	1.17	.235	13.3	2200
電気・ガス・ 熱供給・水道業	.03	.006	14.7	66
運輸業・旅行業	.19	.042	12.7	507
卸売・小売業、 飲食店	1.40	.208	11.0	2065
金融・保険業	.21	.033	9.4	367
不動産業	.04	.007	9.4	88
新聞・放送・出版業、 広告業、映画制作業	.06	.009	8.7	99
情報・通信サービス業	.12	.023	13.0	209
医療・福祉サービス業	.49	.077	9.6	697
教育・研究サービス業	.21	.042	13.6	367
法律・会計サービス業	.04	.006	9.0	77
その他サービス業	.84	.127	9.9	1423
公務	.17	.042	15.4	392
DK	.07	.016	19.1	134

e) 企業規模	発生割合 × 40	構成 比率	継続 期間	継続期間 の n
1 人	.09	.028	15.1	348
2-4 人	.56	.135	15.7	1421
5-9 人	.63	.091	9.1	1243
10-29 人	.94	.137	9.1	1727
30-99 人	.84	.131	9.6	1568
100-299 人	.62	.096	9.3	1193
300-499 人	.30	.044	9.4	518
500-999 人	.27	.049	11.4	510
1000 人以上	.80	.144	12.6	1395
官公庁	.32	.082	16.9	677
DK	.53	.064	8.4	812

注 1：SSM2005-J における 15～70 歳の職歴について、3 つの指標の平均値を示している。a) 従業上の地位については、無職も集計対象に含む（サンプル全体なので、n=5742）b) 仕事内容では、無職を集計対象外としている（該当年齢期間中に一度も職に就いたことがない回答者が 186 人いるので、n=5556）。それ以外の変数については、さらに内職を集計対象外としている（内職にしか就いたことがない回答者が 2 人いるので、n=5554）。

注 2：一番右の列は、継続期間の集計対象者数（つまり、その職に一度でも就いたことのある回答者数）を示している。

注 3：仕事内容の 8 分類には、三輪哲氏作成の「SSM 職業 8 分類（旧 8 分類）への変換」シンタックス（2007 年 10 月 1 日作成）を加工して利用した。

A Basic Methodology for Analyzing the Career Data from the SSM Survey

Tokio YASUDA

Osaka University of Commerce

The sequential data of occupational career is the most outstanding feature in the SSM survey, but the data has not been used enough. One of the causes of it is the shortage of basic methodology to analyze the career data. The purpose of this paper is to apply the methods used for the analysis of observed behaviors, to the analysis of the career data. It would add a new option to the methods for approaching the career data.

For making procedures of analysis not difficult, it is important to convert the sequential data into the appropriate form. It is effective to convert the data from the state sequences into the interval sequences. Therefore, the second section explains the actual conversion procedures. Because some judgments needed to complete the appropriate conversion are closely related to the coding works, it is preferable to conduct the conversion not at the analysis phase but at the coding phase.

Even though sequential data include huge information, it is preferable that researchers analyzing the data take the same order as for normal analyses. It means that you should do univariate analyses first, and then advance to multivariate analyses. The third section proposes to use three indices for each aspect of occupation. Three indices are rates, probabilities, and duration. The analyses of career data can be advanced continuously from the univariate to the multivariate by consistently using these indices.

Keywords and phrases: occupational career, sequential data, methodology