

2005年SSM調査シリーズ 12

社会調査における測定と分析 をめぐる諸問題

Problems in Measurement and Analysis in Social Surveys

前田 忠彦 編

2005年SSM調査研究会

科学研究費補助金 特別推進研究（16001001）
「現代日本階層システムの構造と変動に関する総合的研究」成果報告書

2005年SSM調査シリーズ 12

社会調査における測定と分析 をめぐる諸問題

前田 忠彦 編

2005年SSM調査研究会

科学研究費補助金 特別推進研究 (16001001)
「現代日本階層システムの構造と変動に関する総合的研究」成果報告書

刊行のことば

本書は、文部科学省科学研究費補助金（特別推進研究）「現代日本階層システムの構造と変動に関する総合的研究」の助成を得て行われた 2005 年社会階層と社会移動調査（SSM 調査）の研究成果報告書『2005 年 SSM 調査シリーズ』（全 15 巻）のうちの一冊である。

SSM 調査は 1955 年以来 10 年毎に行われている全国調査である。このような継続性を持った社会階層と社会移動に関する調査は世界に類を見ない。もちろんそれぞれの年の SSM 調査プロジェクトは独自の研究テーマを持っているが、親や本人の階層などの基本変数は継続的に測定されているので、長期にわたるトレンド分析が可能になる。本シリーズの中にも、このようなトレンド分析を行っている論文が多数収録されている。

この継続性は SSM 調査の貴重な財産である。2005 年 SSM 調査研究プロジェクトでは、このことを踏まえた上で、新たな方向に踏み出した。それは本格的な国際比較と若年層調査である。本プロジェクトの基本的なねらいは、次のような問題群に解答を与えることであった。グローバリゼーションと新自由主義の進行する中で、労働市場の流動性は高まっているのか、それともそうではないのか。また高まっているとすれば、それはどの階層を流動的にしているのか。特定の階層は保護的制度に守られて流動化していないのではないのか。このような「流動化」と「階層の固定化」という一見すると相反する問題にアタックすることが、本プロジェクトの基本的なテーマであった。

このテーマを追求するために、国際比較と若年層調査は不可欠であった。グローバリゼーションと新自由主義はいわば普遍的な変動要因である。ただしこれらは直接的に個々の社会の社会階層・社会移動に影響を及ぼすのではなく、それぞれの社会のローカルな制度との相互作用を通じて、社会階層・社会移動に影響を及ぼしたり、及ぼさなかったりする。また新自由主義や労働市場の流動化に対する人々の評価（これは公共性問題といえよう）も社会によって異なりうる。これらの問題に答えるためには、国際比較が必要になる。しかしあまりに異なる社会と日本を比べることは意味をなさない。そこでわれわれは、同じ儒教文化圏に属し、教育制度も類似しているが、日本よりも早くグローバリゼーションにさらされている韓国と台湾を比較の対象とした。

労働市場の流動化は若年層にもっとも影響を及ぼすと考えられる。フリーターやニートの問題をはじめとして、流動化の矛盾は若年層に集中しているといえよう。この問題に関しては既に多くの研究がなされているが、本プロジェクトでは、SSM 調査の蓄積を活用して、社会階層と社会移動という視点からこの問題にアプローチすることにした。たとえば、誰でもフリーターになるわけではなく、出身階層や本人の学歴によってフリーターになる確率は異なると考えられる。このような社会階層論・社会移動論の道具を用いることで、フリーター・ニート問題に新しい光を当てることができるだろう。

このような理論的関心に基づいて、国際比較と若年層調査を行った。国際比較では、韓

国と台湾の階層研究者 6 名に研究プロジェクトメンバーとなってもらい、彼ら・彼女らの全面的な協力の下に韓国と台湾で SSM 調査を実施した。調査票は日本調査とかなりの部分を共通にして、日本・韓国・台湾で厳密な比較分析が行えるようにした。また産業や職業の国際比較ができるように、それぞれの社会のデータに国際標準産業分類コードと国際標準職業分類コードを割り当てた。日本側メンバーにも東アジアの専門家がいて、膨大な時間を費やしてくれたが、これらの作業は困難を極めた。調査票設計段階の調整から始まり、調査票の翻訳やバックトランスレーション、調査設計の調整、コーディングにおける無数ともいえる細かい確認事項などの作業を経て、調査データが完成した。

若年層調査も多くの困難に直面した。大阪大学の太郎丸博氏をヘッドとする若年層調査タスクグループが実査を担当したが、低い回収率の問題や、郵送調査・ウェブ調査ゆえのデータ・クリーニング、コーディングの難しさがあった。しかし太郎丸氏をはじめとするタスクグループの献身的な努力により、若年層調査データも完成した。

本シリーズに収録されている論文は、このような調査データの分析に基づいたものである。本プロジェクトでは、8つの研究会からなる研究体制をとって、それぞれの研究会でメンバーが論文構想を報告して相互にコメントをしあい、より良い論文を執筆することをめざしてきた。その成果が本シリーズに集められている。これらの論文を通じて、日本のみならず、韓国と台湾の階層状況に対する理解が深まることを期待する。

本プロジェクトを推進するに当たり、実に多くの方々のお世話になった。あえて一人一人のお名前をあげることはしないが、ここに感謝の意を表します。また調査にご協力いただいた対象者の方々にも心より御礼申し上げます。

2008年3月

2005年 SSM 調査研究会

付記1. 本研究会による刊行物のリスト

『2005年SSM日本調査 コード・ブック』 2007年11月

『2005年SSM日本調査 基礎集計表』 2007年11月

2005年SSM調査シリーズ（研究成果報告書集）（2008年3月刊）

第1巻	三輪 哲 小林 大祐	編	『2005年SSM日本調査の基礎分析 —構造・趨勢・方法—』
第2巻	高田 洋	編	『階層・階級構造と地位達成』
第3巻	渡邊 勉	編	『世代間移動と世代内移動』
第4巻	阿形 健司	編	『働き方とキャリア形成』
第5巻	米澤 彰純	編	『教育達成の構造』
第6巻	中村 高康	編	『階層社会の中の教育現象』
第7巻	土場 学	編	『公共性と格差』
第8巻	轟 亮	編	『階層意識の現在』
第9巻	中井 美樹 杉野 勇	編	『ライフコース・ライフスタイルから見た社会階層』
第10巻	菅野 剛	編	『階層と生活格差』
第11巻	太郎丸 博	編	『若年層の社会移動と階層化』
第12巻	前田 忠彦	編	『社会調査における測定と分析をめぐる諸問題』
第13巻	有田 伸	編	『東アジアの階層ダイナミクス』
第14巻	石田 浩	編	『後発産業社会の社会階層と社会移動』
第15巻	佐藤 嘉倫	編	『流動性と格差の階層論』

『2006年SSM若年層郵送調査 コード・ブック』 2008年3月
『2006年SSM若年層郵送調査 基礎集計表』 2008年3月
『2005年SSM韓国調査 コード・ブック』 2008年3月
『2005年SSM韓国調査 基礎集計表』 2008年3月
“Taiwan Social Change Survey, 2005: Social Stratification and Social
Mobility in Three Countries, User Guide and Codebook” February, 2008

付記2. 文部科学省科学研究費補助金研究組織等

研究課題「現代日本階層システムの構造と変動に関する総合的研究」(16001001)

研究種目 特別推進研究

研究組織

研究代表者：佐藤 嘉倫 (東北大学大学院文学研究科教授)
研究分担者：近藤 博之 (大阪大学大学院人間科学研究科教授)
研究分担者：尾嶋 史章 (同志社大学社会学部教授)
研究分担者：斎藤 友里子 (法政大学社会学部教授)
研究分担者：三隅 一百 (九州大学大学院比較社会文化研究院教授)
研究分担者：石田 浩 (東京大学社会科学研究所教授)
研究分担者：中尾 啓子 (首都大学東京都市教養学部教授)
(研究協力者については、全リストを第15巻に掲載した。)

研究経費 (単位 千円)

	直接経費	間接経費	総額
平成16年度	19,700	5,910	25,610
平成17年度	186,600	55,980	242,580
平成18年度	29,400	8,820	38,220
平成19年度	32,700	9,810	42,510
計	268,400	80,520	348,920

研究発表

全リストを第15巻に掲載。

はしがき

本巻は「2005年SSM調査シリーズ」の第12巻にあたり、社会調査における測定と分析をめぐる問題に焦点を当てた9論文を収録した。測定と分析という語はかなり広い意味で用いており、調査の実施にあたって対象者からの回答を得るプロセス全体に対して測定の語を、そこから得られたデータを処理して統計的分析を行い、結果の考察を行う部分、いわば計量分析の全プロセスに対して単に分析という語を当てていると解されたい。調査方法論に関わる論文という側面も持っているが、調査方法そのものが研究テーマのプロジェクトではないので、内容面ではかなり社会科学的な（個別科学としての）テーマ設定を持った論文も多く含まれている。

本巻のタイトルを「SSM調査における」ではなく「社会調査における」としたことには次のような理由がある。およそどのような方法論（測定・分析）も、具体的な適用対象となる素材なしで工夫や新方法の開発を目指すことは困難である。このことは社会の成員である個人を対象とする社会調査において特に顕著に現れる。当然、特定の調査テーマにおいて固有に生じる問題への対処から新しい方法論が生まれるものであろうが、そのような方法論はテーマを超えて、より広い問題群へも適用可能であることが期待される。職業のコーディングや、職歴の変化を扱うことは例えばこのSSM調査に特徴的で、成果を得るためにいろいろな意味で十分なリソースが必要になる。こうしたことへの対処が新しい方法論の開発を動機づけるが、ここで工夫された方法論は、他のテーマ、例えば職業に限らない自由回答のテキストデータの分類や、変化を追う縦断的な調査で現れる問題の解決につながるとか、少なくともヒントを与える、などのことが想定される。要約すれば、方法論の研究は、特定の（固有名詞で語られる）調査において集中的な検討が可能な問題への対応としてなされるという側面は持つが、当該調査だけを視野においたものというわけではないだろう。このことを意識して、本巻もSSM調査という固有名は外した次第である。

刊行のことばに記されたように、2005年SSM調査のプロジェクトは、国際比較と若年層調査を同時期に実施するという新機軸を打ち出した。若年層調査の方法については別巻に譲るとして、本巻においても国際比較の方法を扱った論文が複数収録されている。またこのことに伴い、例えば職業分類の国際標準化への対応なども検討されることになった。また国内の社会調査の環境は悪化を続けており、本プロジェクトの国内調査についても、調査担当者の努力も叶わず、回収率は95年調査と比べて大きく低下した。計量分析に当たり、欠測への対応をいかに考えるかは本調査においても避けて通ることのできない課題である。計量的な社会学研究にあたり、技術としての統計的解析法と社会学の理論の有機的結合は、魅力的な標語でありつつ、実践に結びつけることが難しい。以上思いつくままに本プロジェクトでの調査と「測定と分析をめぐる問題」の例を挙げたが、本巻に収められたのもそのような課題に取り組んだ論文群である。

論文は、著者らにより独立に構想されたものであるが、結果として四つのグループに分けることができよう。編者の判断により、巻内での配列もそのような順に緩やかに揃えたので、その分類を簡単に紹介しておく。各論文の概要は編者の要約が不適切になるよりも各論文の著者自身による要旨に委ねることにしよう。

第一に職歴データの扱いに関するテーマが二題ある（保田論文、田中論文）。共に、SSM調査が職業の経歴を（人によって異なる）数段にわたる系列データそして記録していることへの対応を素材としているが、田中論文はもう少し広くデータ加工のプロセス全般についての提案を含むものになっている。

第二に、国際的職業分類ISCOへのアプローチを述べた二題（高橋論文、鹿又・田辺・竹ノ下論文）。前者は機械学習によるISCOコードの付与自体をテーマとしており、後者はISCOを利用した国際的階層指標の導出を述べている。

以上4題は、豊富な情報を含みながら未だ十分に活用されてこなかったSSM調査の様々な原データから、有用で使いやすい変数をいかに生成するかという点を扱っているという意味で、（データの処理に関する側面を多く含みながら）広い意味の「測定」を扱っている論文と見ることができるかも知れない。

続く5題は測定の問題から次第に分析の要素をより多く含むようになっていく。

特に第三グループの3題（元治論文、Nakao論文、菅野論文）は国際比較の方法に関するもので、前2者は職業威信の測定を通じた日韓比較をテーマとしている。菅野論文は社会関係に関する指標の国際比較を多母集団同時解析のアプローチによって整理した実例である。

第四グループ（高田論文、近藤論文）は統計的分析方法をテーマとする（テーマの一部を含む）二題である。高田論文は社会調査では不可避とも言えるサンプルの欠落への対応を扱い、近藤論文は多重対応分析による計量分析の結果を「社会空間の構築」という形で理論的な枠組みの中に位置づけることに取り組んだ。

各論文の構想は、主に、編者が世話人を務めた「方法論研究会」の中で討論し、それぞれにブラッシュアップを進めたものである。開催回数の制約と世話人の力不足もあり、方法論研究会では、議論を十分に尽くせなかったとの反省も残る。本巻に収められた論文には未だ彫琢を要する部分も残るかも知れないが、これらの点は著者らに今後の研究展開を期待すると共に、読者からも積極的にご意見をお寄せいただきたいと思う。

本巻に収められた論文が、今後、社会階層と社会移動研究のデータをよりよく扱うためのヒントとなり、読者を巻き込んでこの分野の研究の発展に資する材料となるものであることを願っている。

2008年3月

前田忠彦（統計数理研究所）

2005年SSM調査シリーズ 12

社会調査における測定と分析をめぐる諸問題論／目次

刊行のことば

はしがき

1	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	1
2	データ・リダクションのための汎用モジュールの開発— 効率のよい職歴分析のために—	田中重人	21
3	機械学習によるISCO自動コーディング	高橋和子	47
4	SSM職業分類と国際的階層指標：EGP階級分類・ SIOPS・ISEIへの変換	鹿又伸夫・田辺俊介・竹ノ下弘久	69
5	職業評定の日韓比較	元治恵子	95
6	Evaluation of Occupational Prestige Hierarchy in Korea and Japan: A New Approach to Measuring Consensus and Inter-group Variations	Keiko Nakao	113
7	社会関係に関連する指標の日本・台湾・韓国の国際比較 について	菅野 剛	133
8	サンプル選択の影響分析—欠損値・打ち切りデータ—	高田 洋	151
9	多重対応分析を用いた社会空間の構築—階層研究の新た な展開を求めて—	近藤博之	163
	既発表成果一覧		187

Problems in Measurement and Analysis in Social Surveys

Edited by Tadahiko Maeda

CONTENTS

Preface to the 2005 SSM Research Series

Preface to Volume 12

1	A Basic Methodology for Analyzing the Career Data from the SSM Survey	Tokio Yasuda	1
2	Development of Modules for Data Reduction: For Efficient Analysis of Occupational History	Sigeto Tanaka	21
3	Automatic ISCO-88 Coding with Machine	Kazuko Takahashi	47
4	SSM Occupational Code and International Measures of Occupational Status: Conversion into EGP Class Schema, SIOPS and ISEI	Nobuo Kanomata, Shunsuke Tanabe, and Hirohisa Takenoshita	69
5	Occupational Evaluation: A Comparative Study of Japan and Korea	Keiko Genji	95
6	Evaluation of Occupational Prestige Hierarchy in Korea and Japan: A New Approach to Measuring Consensus and Inter-group Variations	Keiko Nakao	113
7	Cross Cultural Comparative Analysis of Measurements on Social Network between Japan,	Tsuyoshi Sugano	133
8	Influential Analysis of Sample Selection: Missing data and Truncated or Censored Data	Hiroshi Takada	151
9	Constructing Social Space by Multiple Correspondence Analysis: A New Orientation in	Hiroyuki Kondo	163
	Appendix.		187

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、中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活用されている（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NTAX」¹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NTAX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	学校所在地区市町村（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	学校所在地区市町村（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歳のうちど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、職業の開始年齢か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年末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が占める割合である。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と思われるので、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、当該のイベントが 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。当該の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ため、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歳時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。

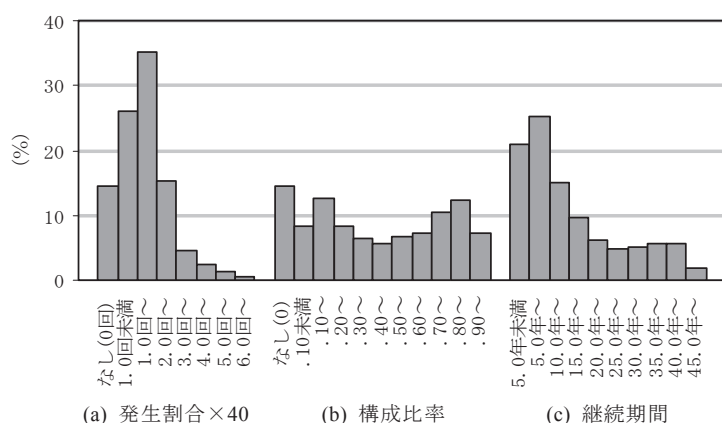


図 1 常時雇用の RPD 指標の分布

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 $156/.078=2.000$ なのに対し、常時雇用の構成比率は $.328/.427=.768$ 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。

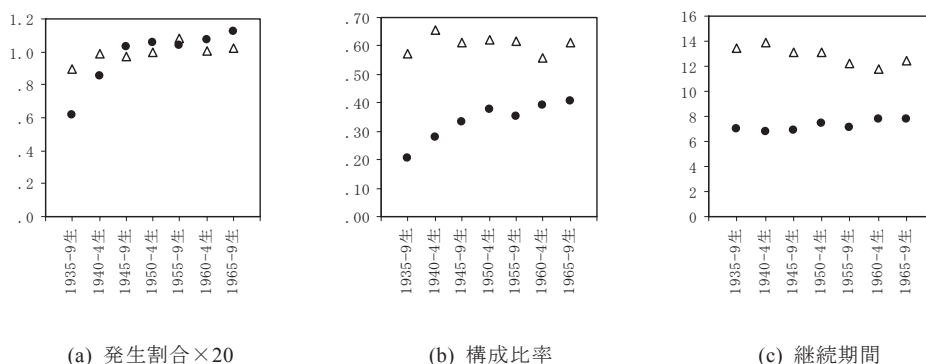


図 2 常時雇用の RPD 指標の平均値 (15～34 歳時、男女・出生コーホート別)

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ー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。

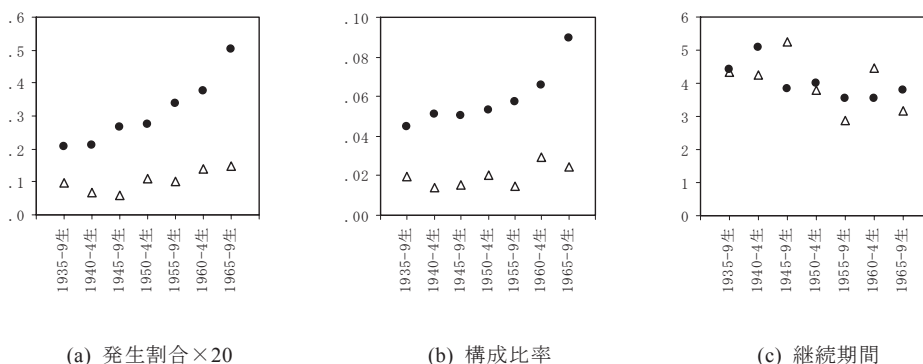


図3 臨時雇用のRPD指標の平均値（15～34歳時、男女・出生コーホート別）

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、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、コーホート別になると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、出生コーホート別に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、どのコー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（新しいコーホートほど、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く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。

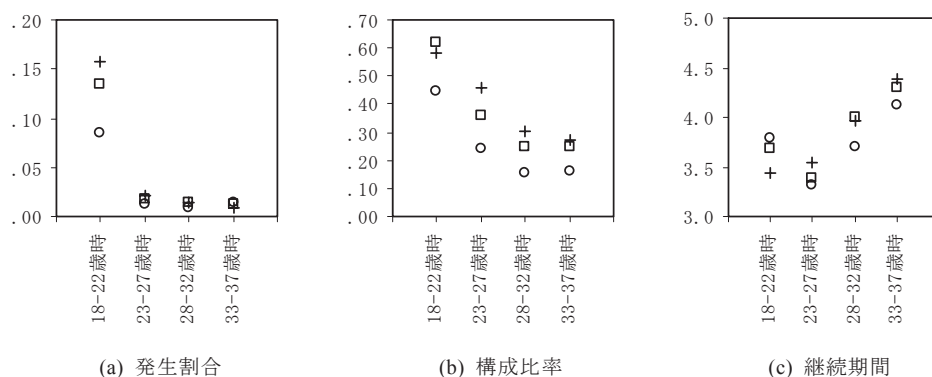


図4 女性の常時雇用についての5歳刻みのRPD指標（18～37歳時、出生コーホート別）

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5歳刻みの推移を、出生コー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歳時に常時雇用に就いていた女性）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ー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。

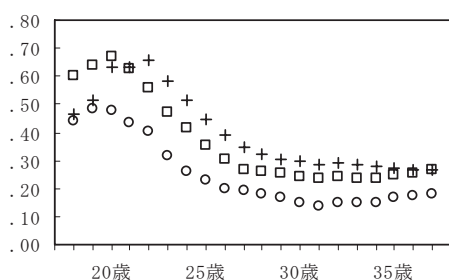


図 5 各年齢時における常時雇用の女性の割合（18～37 歳時、出生コーホート別）

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 に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

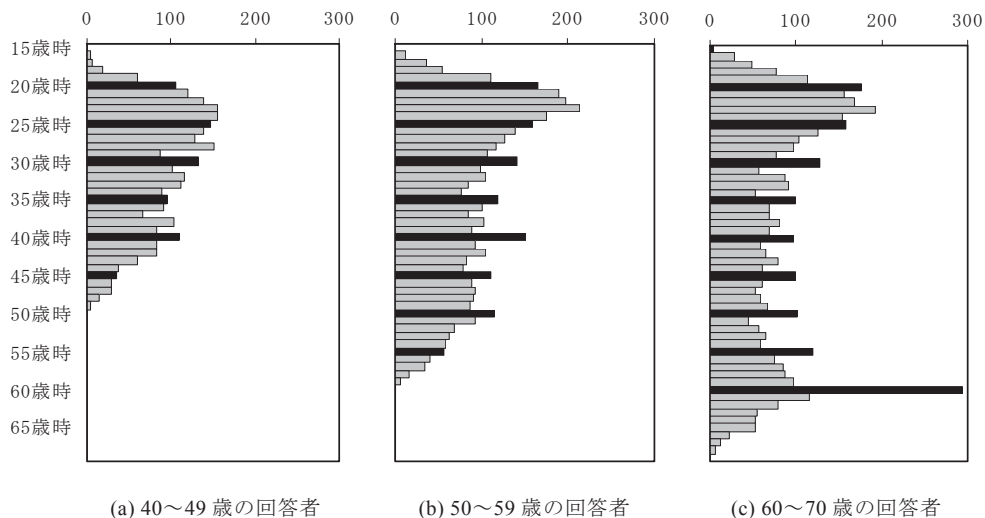


図 6 各年齢時における職業の変化の発生件数（調査時点の年齢層別）

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い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歩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 Vincenzo 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

データ・リダクションのための汎用モジュールの開発

—効率のよい職歴分析のために—

田中重人
(東北大学)

【要旨】

本稿は、複雑な構造を持つデータに対する“toolbox”アプローチを提唱する。このアプローチは、データ加工から統計分析までを統計分析パッケージひとつでおこなう“all-in-one”アプローチに代わるものであり、プログラミング言語を使ってデータ加工をおこなうプロセスを統計分析から切り離そうとするものである。統計分析パッケージの貧弱なデータ加工機能を使うよりも、本格的なプログラミング言語を使うほうが効率がよい。一般的に必要とされる機能をそろえた「モジュール」を用意しておくことで、効率をさらにあげることができる。本稿では、職歴データを加工するためのオブジェクト指向モジュールに特に焦点をあわせる。著者はこれまでに Perl を利用したモジュールを作成してきており、それらのモジュールの概要を紹介する。また、階級分類と“person-year”データを作成するためのプログラム例も紹介する。くわしい情報は <http://www.sal.tohoku.ac.jp/~tsigeto/dredu/> を参照されたい。

キーワード：データ構造, Perl, 職歴分析, オブジェクト指向プログラミング

1. はじめに

1.1. All-in-one アプローチから toolbox アプローチへ

データ分析においては、「統計分析パッケージ」を使って、データの加工から統計分析まですべてをひとつのソフトウェアでこなすのがふつうである。しかし、統計分析パッケージにおいては、主たる機能は統計分析のためのものであり、データ加工のための機能は貧弱である。複雑な構造をもつデータの分析を効率よく進めるためには、統計分析パッケージは統計分析のために使い、データ加工には本格的 (full-fledged) なプログラミング言語を使うのがいい。さまざまな道具の入った工具箱 (toolbox) から必要な道具をその都度えらんで使うように、ソフトウェアも必要に応じて使いわけべきなのである。

この報告では、統計分析の前処理としてのデータ加工 (data reduction) をおこなうためのシステムの設計について、基本的な考えかたを述べる。また、現在作成中の、SSM データの加工をおこなうためのモジュール群について、概要を報告する。

1.2. 歴史

筆者は、1990 年代前半に、1985 年 SSM 調査データ分析用のプログラムを SAS System の

マクロ機能を使ってつくっていた。

その後、1995 年 SSM 調査に参加した際に、このプログラムを Perl (version 4) に移植した。このプログラムは筆者の WWW ページ <http://www.sal.tohoku.ac.jp/~tsigeto/ssm/script.html> から入手できる。

1.3. プラットフォーム

今回のプログラム開発にあたっては、Perl version 5 (Wall, Christiansen, and Orwant 2000) を用いる。Perl は version 5 へのヴァージョン・アップの際にポインタをあつかえるようになり、OOP (オブジェクト指向プログラミング) にも対応した (Conway 2001)。これらの追加機能をふまえて、過去に Perl version 4 で書いたプログラムをもとに、全面的に書きあらためることにした。

プログラムのテストに使ったのは Windows XP (Service Pack 2) 上の ActivePerl (<http://www.activestate.com>) v5.8.7 である。

2. データ・リダクション・システムの要件

2.1. Data reduction という概念

社会調査のデータはケース×変数からなる行列で表現される。統計分析パッケージのデータ・ハンドリングにおいては、通常、先頭から 1 ケースずつ処理を進めることになり、そこでは、データは単純なスカラ値を持つ変数の集合として見えることになる。

この単純な外見は、見かけ上のものである。調査票はもっと複雑な構造を持つ。たとえば、職業のデータは、職業を構成するさまざまな要素を測定した変数が集まったものである。さらに、職業の経歴 (職歴) データの場合、調査対象者が経験してきた職業が順番に並んで職歴の全体を形作るデータとなる。ほかに、家族調査における世帯表 (日本家族社会学会 2005) や妊娠・出産等の履歴 (国立社会保障・人口問題研究所 2007) など、類似の構造を持つデータは多い。

本来複雑なデータであるにもかかわらず、形式上は単純な 1 次元のデータであるかのようにあつかわざるをえないことが、分析上の大きな困難をもたらしている。たとえば SPSS (2005 年 SSM 調査のデータは SPSS のデータ・セットのかたちで配布されている) では、単純なスカラ量以外のデータ型はサポートされていない。職歴データのような複雑なデータ構造は、あつかいようがないのである。

データ・リダクション (data reduction) とは、そのままでは集計できないような複雑なデータから、集計可能なデータを抽出してくるプロセスをいう (Wall and Schwartz 1993)。本稿の文脈では、データを統計分析にかける前に必要な前処理をおこなって、その結果を統計分析

パッケージに渡すまでのプロセスのことである。システムにデータを投入する時点では、調査票の構造を反映したかたちであつかえるようにしておき、1次元のデータとしてあつかえるところまで複雑性を落としてから 統計分析パッケージに渡すのである (図 1)。

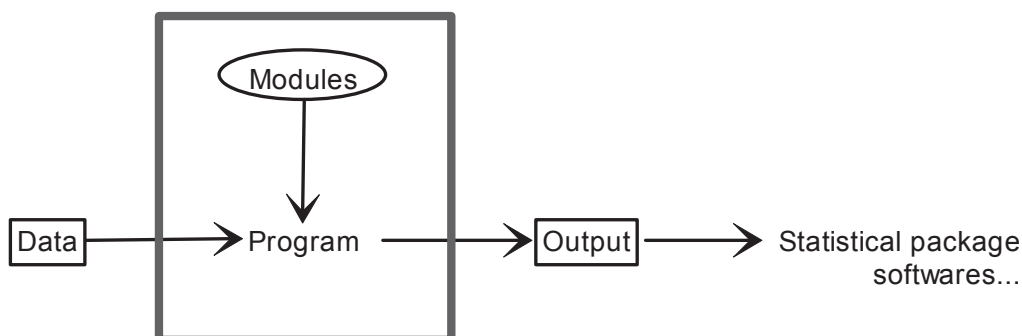


図 1. Data reduction system

2.2. コードの再利用とモジュール

こうしたシステムは複雑なものになるので、ひとつのファイルを最初から最後まで直線的に実行するようなプログラムにはできない。再利用可能なコードはプログラム中で何度も再利用することになるだろう。プログラムの見通しをよくするには、再利用するコードは別のファイルにまとめておき、必要に応じて呼び出すのがいい。こうしたコードのまとまりを「モジュール」(module) と呼ぶ。

モジュールの設計において重要なことは、モジュール同士が不用意に干渉しないように お互いの活動空間を分離しておくことだ。相手呼び出ししたり、呼び出しに応答したりするための「窓口」(interface) を用意しておき、それ以外の部分では、各モジュール内部でなにがおこなわれようと、他のモジュールに影響がおよばないようにしておくのがいい。

2.3. データ型

本格的なプログラミング言語は、さまざまな「型」(type) のデータをサポートする。代表的な例はつぎのようなものである

スカラ (scalar): 単一の数値または文字列

ハッシュ (hash): 「名前」から「値」への写像

関数 (function): コンピュータに実行させる操作を記述したもの

配列 (array): データが一行に並んだもの

構造体 (structure): 一定のテンプレートにしたがって組み立てられたデータ

ポインタ (pointer): 他のデータがおかれている「場所」(address) への参照

データの加工を効率よくおこなうには、これらのデータ型をうまく使いわけ必要がある。たとえば、職業のデータは「従業上の地位」「産業コード」「職業コード」などの名前とそれに対応する値を対にして保持した「ハッシュ」としてあらわすことができる。そして、職歴は、職業を一行にならべたものなので「ハッシュの配列」である。

2.4. カプセル化と前後関係非依存性

いろいろなデータ型をあつかえるのは便利である。しかしそれは同時に面倒なことでもある。さまざまなデータ型が混在していると、いま操作の対象としているデータがどのような型のものであるかを常に意識しながらプログラムを書かなければならなくなる。

たとえば、学歴の変数から教育年数を求める場合を考えてみよう。2005 年 SSM 日本調査のデータでは、学歴はつぎのような値をとる。

中学校: 1

高等学校: 2

高等専門学校: 3

短期大学: 4

大学: 5

大学院: 6

学歴なし: 8

学校教育における標準的な年限にしたがった場合の教育年数を求める標準的な技法は、テーブル参照 (table look-up) だろう。つぎのような変換テーブルを用意しておき、これにしたがって値を変換する。

1 → 9

2 → 12

3 → 14

4 → 14

5 → 16

6 → 18

8 → 0

Perl でこの操作をおこなうためには、つぎのように書けばいい (\$ED_SSM は学歴の変数である)。

```
%edueyear = (
    1, 9,
    2, 12,
    3, 14,
    4, 14,
    5, 16,
    6, 18,
    8, 0,
);

$Eduyear = $edueyear{$ED_SSM};
```

このコード中、\$ はスカラを、% はハッシュを、{ } はハッシュから値を取り出すことをそれぞれあらわしている。こういう妙な記号を使うのは Perl の文法の問題である。他の言語では、別のやりかたが使われるだろう。いずれにせよ、データ型を区別するためになんらかの識別子を使う言語では、これらを書きまちがえるとプログラムはまともに動かない。あるいは、初出時に「宣言」をおこなって変数等の「型」を確定する決まりになっている言語もある。そのような言語では、その変数等の「型」に応じて利用可能な操作が決まっている。その規則に違反するプログラムは実行できない

データ型の区別をつけることは、モジュールを開発しているときなら苦にならないかもしれない。そのときには、データをどのような型で表現してどのように処理すべきかが主要な関心事になっているだろうから。だが、モジュールを使用してデータを加工しているときには、データ型そのものにはあまり注意が向かない。そのような状況で、型をまちがえるたびにエラー・メッセージをみるのは、相当なストレスである。

この問題に対処するためには、モジュールの利用者がデータ型を意識しないですむようにインターフェイスを工夫する。あらゆることをスカラと関数だけで表現できるようにするのである。

上の教育年数変換のコード例は、つぎのようにも書ける

```
package SSM;

# 関数の定義
sub Eduyear {
    ($e) = @_;    # 呼び出し元からわたされた引数
    return 9 if 1==$e;
    return 12 if 2==$e;
    return 14 if 3==$e;
    return 14 if 4==$e;
    return 16 if 5==$e;
    return 18 if 6==$e;
    return 0 if 8==$e;
    return '' ;    # どれにもあてはまらなければ空文字列を返す
}
```

この内容を `SSM.pm` というファイル名で保存する。このコードは、必要な時につぎのようにして呼び出せる。

```
use SSM;      # SSM.pm ファイルを読み込む
.....
$Eduyear = SSM::Eduyear($ED_SSM);
```

呼び出された関数がおなじ結果を返すことを保証できるのであれば、モジュール内部での仕事のこなしかたはどのように実装してもいい。上とおなじように、モジュール内に用意した変換テーブルを参照して変換するには、つぎのようにする。

```
package SSM;

    # ハッシュの定義
    %eduyear = (
        1, 9,
        2, 12,
        3, 14,
        4, 14,
        5, 16,
        6, 18,
        8, 0,
    );

    # 関数の定義
    sub Eduyear {
        ($e) = @_ ;      # 呼び出し元からわたされた引数
        return $eduyear{$e} ;
    }
```

モジュール内部における処理の手続きは全然ちがうものになっている。(たぶんこちらのほうがすこし速い。そのかわり記憶領域を多めに使う。) だがそのことをモジュール利用者が気にかける必要はない。関数 `SSM::Eduyear()` を呼び出した時にえられる出力はかわらないからだ。利用者が知っておかなければならないのは、入力と出力との関係——モジュール `SSM` が提供する関数 `Eduyear()` に教育の変数 `$ED_SSM` をあたえると何を返してもらえるか——だけである。

カプセル化がもたらすもうひとつの望ましい効果は、前後関係からの独立ということである。個々の関数への入力と出力との関係が常に一定であるということは、呼び出す場所の前後でどのような操作がおこなわれていようと、それはその関数の挙動に影響しない、ということだ。プログラムの一部の変更が他の部分に波及してしまうことを防げるから、エラーが起きにくくなる。また、既存のプログラムの一部を切り取って再利用することが簡単にできるようになる。

2.5. 分業の弊害

このような関数を提供するモジュールをたくさんつくっていくと、大きな問題が生じる。

どの関数がどのモジュールで定義されているかをおぼえきれなくなるのである。たとえば、一連の SSM 調査に共通するデータの操作をモジュール SSM で定義し、各々の調査に独自のデータの操作は SSM2005, SSM1995, ... などで定義することにしよう。2005 年 SSM 日本調査の教育年数のデータを操作するには、SSM::Eduyear() を呼び出すべきだろうか? それとも SSM2005::Eduyear() を呼び出すべきだろうか?

この問題を解決する方法はふたつある。ひとつはモジュールから関数を輸出／輸出 (export/import) する方法、もうひとつはオブジェクト指向プログラミング (object-oriented programming: OOP) である。

「輸入」とは、モジュールを特定せずに関数の名前を指定するだけで関数を呼び出せるようにする、ということの比喩的な表現である。たとえばモジュール SSM から関数 Eduyear() を輸入したとしよう。それ以降、プログラム中で Eduyear() と書けば、それは SSM::Eduyear() と書いたのとおなじことになる。この仕組みをうまく利用すれば、複数のモジュールのどこで何が定義されているかを知らなくてもよくなる。

オブジェクト指向プログラミング (OOP) においては、操作の対象となる「オブジェクト」に指示をあたえると、その指示を実行するためのメソッド (特殊な関数と考えてよい) をそのオブジェクトが探してきて実行する (これも比喩的な表現である)。オブジェクトは、特定の「クラス」(class) に所属していて、そのクラスが定めている規則にしたがってメソッドを探す。たとえば SSM2005::Eduyear() を探してみても、それでみつからなければつぎに SSM::Eduyear() を探す…… というふうに。クラスの設計がきちんとなされていれば、この方法も、問題の解決に役立つ。

輸出／輸入と OOP のどちらがふさわしいかは、状況 (あるいは利用者の思想) による。

操作の対象となるデータの性質が重要であるにもかかわらず、利用者側でそれを把握するのがむずかしい場合がある。たとえばファイルから多数の数値を読み込んで、それを一定の順序で変数に代入していく、というような操作 (後述の Dredu モジュールを参照) を考えよう。この場合、読み込むべき数値がいくつあるとか、どの数値をどの変数に代入すべきとかは、いちいちおぼえておけるようなことではなかろう。そのような情報はオブジェクトに格納しておき、それと読み込んできたデータとを一体のものとしてあつかうのがよい。そうすれば、どのような操作をすべきであるかはオブジェクト自身が決められる。このような場合には、OOP を採用するのが自然である。

一方、上でとりあげた教育年数の変換のような操作は、ふつうは利用者の頭の中に入っているであろう。学歴について分析しようとするのであれば、それぞれの学歴を取得するのにどれだけの年数がかかるかは基礎的な知識として当然知っておくべきことである。そのようなことについて、オブジェクトをいちいち設計するのは労力の無駄である。通常は、関数をどこから呼び出すかという手続きを簡略にするために、「輸入」が適切におこなわれていれば

じゅうぶんであろう。

3. モジュール群の概要

3.1. モジュール間の関係

開発中のモジュール群は、データ・リダクション一般において必要となる汎用の操作を集めたもの (Dredu モジュール群) と SSM 調査に特有の操作を集めたもの (SSM モジュール群) にわかれている。

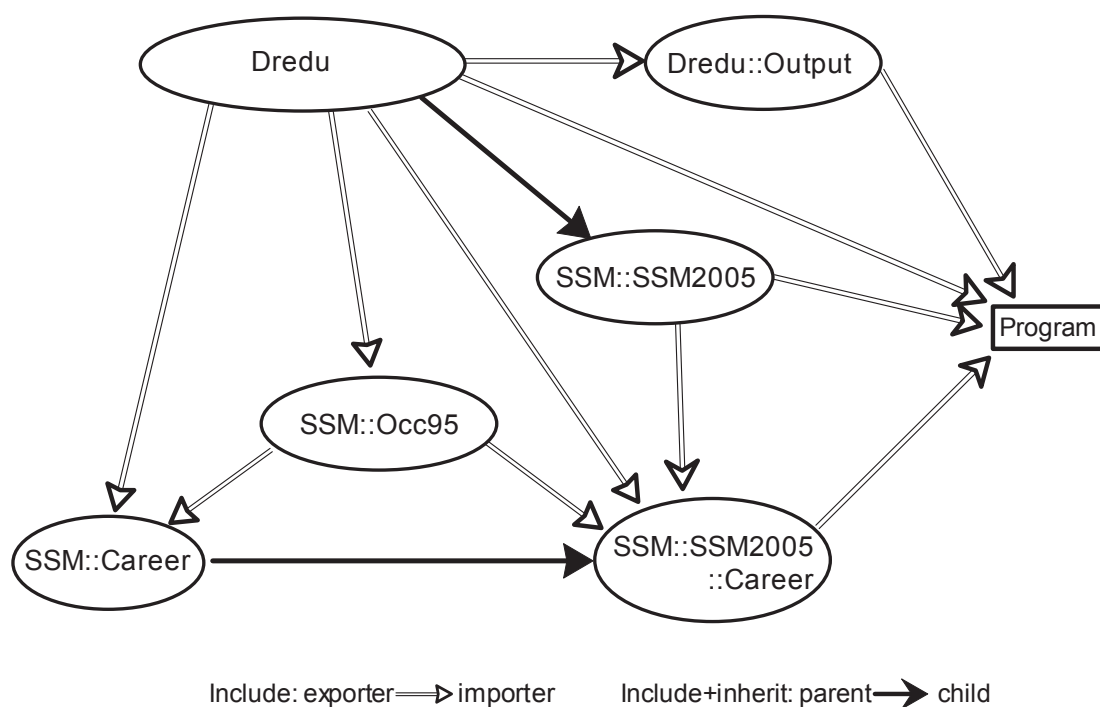


図 2. Dependency among modules

前者については、つぎのモジュール群を用意している

Dredu —— データの読み込み、変数への値の割り当て、欠損値の処理をおこなう。クラス Dredu を定義し、OOP にしたがつたインターフェースを持つ。

Dredu::Output —— データの出力や単純集計をおこなう。クラスは定義せず、関数の輸出のみをおこなう。

後者については、つぎのモジュール群を用意している

SSM::SSM2005, SSM::SSM1995 —— 2005 年, 1995 年 SSM データについて、職業以外の分析をおこなう。Dredu クラスを「継承」(inherit) したうえで必要な機能を追加する。

SSM::Occ95 —— 1995 年版 SSM 職業分類 (1995 年 SSM 調査研究会 2006) 関連の関数を輸出する。SSM::Career などから利用される。

SSM::Career, SSM::SSM2005::Career, SSM::SSM1995::Career —— 職業・職歴の分析をおこなう。クラス SSM::Caer を定義する。OOP にしたがったインターフェースを持つ。

これらのモジュール間の依存・継承関係を図 2 に示す。

3.2. Dredu とその派生クラス

Dredu モジュールは、データ・ファイルから 1 ケースずつデータを読み込んで、それを格納するためのオブジェクトを定義するとともに、オブジェクト操作のための各種メソッドを提供する。

データ・ファイルから読み込んだ各変数の値は、変数名に使われている文字種類にしたがって、つぎのような規則で格納される

- (1) 大文字と数字と下線だけからなる変数名の場合、そのまま Perl の変数として格納される。これにアクセスするには \$ を変数名の先頭につけて \$Q01_1A などとする。
- (2) 上記にくわえて小文字がふくまれる場合、Dredu::Var という名前空間に隔離して格納される。これにアクセスするには \$Dredu::Var:: を変数名の先頭につけて \$Dredu::Var::q01_1a などとする。
- (3) これら以外の場合は、Dredu オブジェクト内部に格納される。

データ・ファイルからあたらしいケースを読み込むたび、これらの変数はあたらしい値に書き換えられる。

いずれの場合も、Dredu オブジェクトは、各変数へのポインタを保持しており、Value メソッド (後述) などを通じてアクセスすることができる。ただし、この記法はずいぶん長いものになる。変数名をすべて大文字 (および下線と数字) だけからなるものにしておくと、通常の Perl の変数として (短い記法で) コードできるので便利である。

SSM::SSM2005 と SSM::SSM1995 は、いずれも Dredu を継承するかたちでつくられている。基本的な部分は Dredu の機能を利用し、各調査に独自の部分をそれぞれのモジュールで

実装している。ほかのさまざまな調査についても、`Dredu` で定義されるメソッドを継承あるいは再定義 (`override`) しながら、それぞれの調査に特化したモジュールをつくっていくことができる。

3.3. `SSM::Career` とその派生クラス

`SSM::Career` は職歴データ分析に特化したモジュールである。このモジュールでは職歴のオブジェクトを定義しており、そのオブジェクトを操作するための共通のメソッドを提供する。各調査の職歴データ分析に特化したメソッドは、各調査のためのモジュール `SSM::SSM2005::Career` や `SSM::SSM1995::Career` で定義されている。詳細は 6 節で述べる。

なお、親職・配偶者職なども、職歴と共通した形式のデータである。このため、一段のみからなる職歴のようなものとみなせる。これらには、職歴用のメソッドをそのまま適用できる。

3.4. 慣習的な約束

上で述べたように、モジュールが提供する機能はすべて関数またはメソッド呼び出しによる。関数やメソッドが返す値のほうも、できるかぎり、通常の (単一の) スカラ値だけにしたいところである。しかし、現実には、通常のスカラ以外の値を返さざるをえないこともある。そこで、つぎのルールにしたがって関数を命名することで、データ型に関する示唆をあたえるようにしてある。

- ・通常の単一のスカラ値 (数値または文字列または真偽) を返す関数の名前は大文字ではじめる
- ・それ以外の値 (配列、ハッシュ、ポインタなど) を返す関数の名前は小文字ではじめる

たとえば、`valid($X, $Y)` は引数のうちで欠損値でないものの配列、`miss($X, $Y)` は欠損値であるものについて欠損値の種類をあらわす文字列の配列を返す。どちらも、単一のスカラ値を返す関数ではないので、小文字からはじまる。

一方、`All_valid($X, $Y)` は、`$X` と `$Y` の少なくとも一方が欠損値であれば 0 (=偽) を、両方が非欠損値であれば 1 (=真) を返す。単一のスカラ値を返すので、大文字ではじまる関数名にしてある。

4. データ分析用の汎用モジュール

社会調査データの分析では、一般に、データを1ケース分ずつ読み込んで加工する方式とられる。このようなデータ加工のプロセスで使う一般的な機能をまとめたものが **Dredu** モジュールである。

4.1. 変数の定義とデータ入力

モジュール **Dredu** を使うには、プログラム冒頭に次のように書いておく。

```
use Dredu;
$data = Dredu -> new; # データ格納用のオブジェクトの作成
```

冒頭の **use** は、モジュールを読み込むことをあらわす。2行目で **Dredu** クラスに属する新しいオブジェクトを作成し、以下 **\$data** として参照する。演算子 **->** はクラスまたはオブジェクトを指定してメソッドを呼び出すことをあらわす。

テキスト形式 (タブ区切り) のファイルからデータを読み込むには、これに続けてつぎのように書く。**DATAFILE** にはデータ・ファイル名が入る。メソッド **fs** は、テキスト・ファイルから入力する際の変数の区切り文字を指定するメソッドである。

```
$data -> fs("\t");      # タブ区切りに
$head = $data -> datafile( 'DATAFILE' , 1 ) ; # ファイルを指定して1行読む
chomp $head;           # 末尾の改行記号を除去
@var = split( $data->fs , $head ); # 配列に分解
$data -> variables(@var); # 変数名を定義
while( $data -> read_text ){ # 1ケースずつ読み込む

    # ここに必要な処理を書く

}
```

変数名等の情報は、定義ファイルであたえることもできる。このファイルには、変数の定義、欠損値、データ・ファイルの名前などを一定の規則にしたがって書いておく。メソッド **configfile()** に「定義ファイル」のファイル名をわたして起動すると、**Dredu** オブジェクトに定義ファイルの内容が格納され、それ以降の処理で利用されるようになる。この機能を使うと、つぎのように書ける。

```
$data -> configfile( 'FILENAME' ); # 読み込むファイルや変数を定義
$data -> read_text;                # 1行読み捨て
while( $data -> read_text ){ .... # 1ケースずつ読み込んで...
```

ただし、巨大なテキスト・ファイルを読み込んで処理するには、かなり時間がかかる。このことを考慮して、固定長のバイナリ・ファイルを読み書きする機能も用意してある。

2005 年 SSM 日本調査データについてこの操作をおこなうためのプログラム `ssm2005.pl` を用意した。カレント・ディレクトリに素データのファイル (`ssm2005.dat`) を置いて `ssm2005.pl` を実行すると、バイナリ・ファイル `ssm2005.bin` ができる。このときに定義ファイルの内容が標準出力に出てくるので、これを適当なファイル名 (たとえば `ssm2005.ini`) にリダイレクトして使うとよい。1995 年 SSM 調査 A 票についても、同様のプログラム `ssm1995a.pl` を作成している。いずれのプログラムにおいても、欠損値の処理 (4.4 節) は自動的におこなわれて定義ファイルに書き出される。変数名にふくまれるアルファベットはすべて大文字に変換される。

バイナリファイルから 1 ケースずつ読み込むには、メソッド `read_binary` を使う。

```
$data -> configfile( 'ssm2005.ini' ); # 読み込むファイルや変数を定義
while( $data -> read_binary ){ .... # 1 ケースずつ読み込んで...
```

4.2. 変数の操作

読み込んだデータは、メソッド `Value( '変数名' )` で読み出せる。ふたつ引数をあてて `Value( '変数名', 値 )` のようにすると、第 1 引数の変数へ第 2 引数の値を代入する。なお、変数名がすべて大文字 (と下線と数字) からなる場合、その変数名の先頭に `$` をつけて参照することができるので、通常はこちらを使うほうが便利だろう。

```
while( $data -> read_binary ){ # 1 ケースずつ読み込む
  print $data -> Value( 'Q01_2A' ); # 年齢を出力
  print $Q01_2A ; # 上におなじ

  $data -> Value( 'Q01_2A' , 30 ); # 年齢を 30 歳に
  $Q01_2A = 30; # 上におなじ
}
```

メソッド `values( 変数名のリスト )` によって、複数の変数の値のリストをえることができる。この場合、変数に値を代入する機能は用意していない。

```
print $data -> values( 'Q01_2YW' , 'Q01_2M' ); # 生年月を出力
```

変数名のリストを指定せずに `values` メソッドを呼び出した場合、全変数の値のリストが返る。変数の順番は、定義ファイル (あるいは `variables` メソッド) で変数を定義したときの順番による。この変数順は、`variables` メソッドを引数なしで呼び出すことで出力できる。

4.3. 別名

変数には、正規の名前以外に「別名」(alias) をあたえることができる。

```
$data -> alias( 'Q01_1' , 'Sex' );
$data -> alias( 'Q01_2A', 'Age' );
$data -> alias( 'Q24_1' , 'MarStat' );
```

```
$data -> alias( 'DANSU' , 'NStage' );
```

メソッド `alias()` の引数には、正規の変数名と別名の組をあたえる。

定義ファイルにつぎのように書いておいても同様のことができる (空白ではなくタブで区切ること)。

```
[alias]
Q01_1    Sex
Q01_2A   Age
Q24_1    MarStat
DANSU    NStage
```

`ssm2005.pl` の出力する定義ファイルには、これらの別名の指定がデフォルトで入っている。`SSM::SSM2005` などの提供するメソッドには、これらの別名が定義されていることが前提になっていることがある。

4.4. 欠損値

欠損値には、適当な値をあたえておく。この「適当な値」はなんでもよいのだが、通常は絶対値の大きい負の整数にしておくのがいいだろう。そうしておく、たいていの演算の結果が変な値になるため、欠損値処理のまちがいをみつけやすくなる。

欠損値の定義には `missing()` メソッドを使う。

```
$data -> missing(
  'NA' ,    -32767,
  'OUT' ,   -32766,
  'BLANK' , -32765,
  'GHOST' , -32764,
  'YET' ,   -32763,
)
```

あるいは、定義ファイル中につぎのように書いておき、`configfile()` メソッドで定義ファイルを読み込んだ際に、欠損値が定義されるようにしてもよい。空白ではなくタブで区切ること。

```
[miss]
NA      -32767
OUT     -32766
BLANK   -32765
GHOST   -32764
YET     -32763
```

これらの欠損値は、つぎのような種類わけによる。

NA: 無回答

OUT: 非該当

BLANK: 素データファイルにおける空白

GHOST: 技術的な理 9 由により存在しない値 (初職の転職理由など)

YET: まだ経験されていない項目

各種メソッド・関数は、これらの欠損値種類が定義されていることを 前提にしてつくられている場合がある。

種々の演算において、欠損値が例外あつかいされる仕様にはしていない。普通に計算すると、ただの数値として処理されてしまう。必要であれば、欠損か否かを前もって判断して条件分岐させること。そのために関数 `miss()`、`valid()`、`All_valid()` が用意されている。たとえば年齢を 10 歳刻みにする場合は、つぎのようにすればよい。

```
if( valid($Q01_2A) ) {           # 欠損値でなければ
    $age10 = int( $Q01_2A / 10 ); # 10 で割って切り捨てる
    $age10 = 7 if 6==$age10 ;     # 70 代は 60 代あつかい
} else {                         # 欠損値なら
    $age10 = $Q01_2A;            # 元の値をそのまま
}
```

欠損値に対応する欠損値種類を調べるには関数 `Missstype()` を使う。この逆の変換を行うのが `Missval()` である。

```
# $FOO が非該当なら $BAR も非該当にする
if( Missstype($FOO) eq 'OUT' ) { $BAR=Missval('OUT') } ;
```

4.5. データ出力

`Dredu::Output` モジュールは、データ出力のための機能を提供する。輸出される関数 `output()` は、変数名と変数値の組を受け取り、テキストを書き出す。区切り記号は Perl の特殊変数 `$,` によって指定する。

```
use Dredu::Output;
$, = "\t" ; # タブ区切り
.....
output(
    'NAME1', $VAR1,
    'NAME2', $VAR2,
    .....
) ;
```

右側の変数 (`$VAR1`, `$VAR2...`) の値がタブ区切りテキストとして書き出される。その際、第 1 行目には、左側の各文字列 (`NAME1`, `NAME2...`) が出力されるので、これを統計パッケージで変数名として読み込むとよい。この 1 行目がいらぬなら、Perl 組み込みの `print` 文をふつうに使えばよい。

```
print $VAR1, $VAR2, $VAR3, .... ;
```

出力先は標準出力 (standard output: STDOUT) である。デフォルトでは、この出力は画面に表示される。ファイルに出力するには、起動時のコマンド・ラインで出力をリダイレクトすること。または、Perl の組み込みの関数 `select()` を使ってもよい。

欠損値はそのまま出力される。統計分析にあたって欠損値処理をおこなえるようにするには、統計分析パッケージの側で欠損値の割り当てをおこなうこと。たとえば SPSS の場合は、「シンタックス」につぎのような一文を入れておけばよい

```
missing values ALL (-32767 thru -32763).
```

関数 `convert_missing_to_alphabet` を実行してから `output()` を実行すると、欠損値はアルファベットに (値のちいさいほうから A, B, C, ... の順に) 変換されてから出力される。

```
convert_missing_to_alphabet;
output( ..... );
```

この関数は、SAS System での分析を念頭においてつくったものである。SAS プログラムの最初につぎのように書いておくと、アルファベット 1 文字からなる値を「特殊欠損値」としてあつかうことができる。

```
missing A B C D E;
```

もし欠損理由を区別しなくてよいなら、欠損値をすべて単一の値に置き換えればよい。関数 `convert_missing_to_null` は欠損値をすべてブランク (空文字) に、関数 `convert_missing_to_dot` は欠損値をすべてピリオド (.) に置き換える指示である (そのあとに出てくる `output()` 関数に有効)。

4.6. 集計

Dreduc::Output モジュールは、単純出力のための関数 `freq()`, `frequencies()`, `class_freqlist()` も輸出する。これらには、引数としてハッシュをあたえる必要がある。前 2 者に対しては、値→出現回数の 1 次元ハッシュ、`class_freqlist()` に対しては、値の組→出現回数の 2 次元ハッシュをあたえる。

```
while( $data->read_binary ){
    .....
    $sex{$Q01_1} ++ ;          # ハッシュ作成 (数え上げ)
    $marstat{$data->Value('MarStat')} ++;
    $sex_marstat{$Q01_1, $data->Value('MarStat')} ++;
}
freq( %sex );                 # 度数分布 (人数のみ)
frequencies( 8, %marstat );   # 度数分布 (人数, %, 累積%)
class_freqlist(8,%sex_marstat); # 男女別度数分布 (人数, %, 累積%)
```

集計のために大した機能を用意する予定はない。output() で出力したテキストファイルを統計分析パッケージで読み込んで分析するのが本来の使いかたである。

4.7. その他のデータ加工関数

Dredu モジュールは、つぎの関数を輸出する。

関数 Common_question() は、調査票のちがいによって同一内容の変数が別の変数名で格納されている場合 (2005 年 SSM 日本調査では留置 A/B 票にそのような質問項目がある)、それらの変数を統合するための簡便な方法を提供する。このメソッドは、引数のリストのなかから、非該当でないものをひとつ抜き出す。

```
$Household_size = Common_question( $A20_1, $B07_1 );
```

関数 Sum() は、引数のリストのうち、欠損値でないものについての合計を求める。

```
$Sum_of_valid = Sum( $A, $B, $C, ... );
```

関数 dummy() は、第 1 引数の値について、第 2 以降の引数と照らし合わせてダミー変数を作成する。

```
@foo = dummy($FOO, 0, 1, 2, 3 );
output(
  'FOO_0', $foo[0], # $FOO==0 なら 1, そうでなければ 0
  'FOO_1', $foo[1], # $FOO==1 なら 1, そうでなければ 0
  'FOO_2', $foo[2], # $FOO==2 なら 1, そうでなければ 0
  'FOO_3', $foo[3], # $FOO==3 なら 1, そうでなければ 0
);
```

5. SSM データ分析のための基本操作モジュール

5.1. オブジェクトの作成

モジュール SSM::SSM2005 と SSM::SSM1995 は Dredu を継承する。つまり Dredu とおなじかたちのオブジェクトを処理対象としており、SSM 調査データの分析に必要な特殊なメソッドをそれに追加するものである。

このモジュールを使う場合、データを読み込むためのオブジェクトは、Dredu クラスではなく、SSM::SSM2005 あるいは SSM::SSM1995 クラスに属するものとして作成する。

```
use Dredu;          # 関数を輸入
use SSM::SSM2005;
$data = SSM::SSM2005 -> new; # データ格納用のオブジェクトの作成
.....
```

5.2. メソッド

メソッド Id は、支局番号 (\$NUMS)、地点番号 (\$NUP), 対象番号 (\$NUMI) を合成して

ID 番号を作成する。

メソッド `Age_marriage`, `Age_first_birth`, `Age_last_birth` は、それぞれ 結婚時、第 1 子誕生、末子誕生の時点での回答者の年齢を求める。子供の誕生年を求めるにあたっては、田中 (1997) とおなじ計算方法を使う。

メソッド `children_age`, `age_births`, `household_member` は、それぞれ 子供年齢、子供誕生時の回答者年齢、世帯構成について配列のかたちにして返す。これらのメソッドは、2005 年 SSM 日本調査についてだけ用意した。子供誕生時年齢の計算は、上と同様、田中 (1997) の方法による。`household_member` は、もとのデータは 1 か 2 の値をとる 2 値データであるが、これを 1 (=該当者あり) と 0 (=該当者なし) に変換する。

世帯構成に関しては、自由回答データの情報も利用して、世帯類型を作成するためのプログラム `SSM/SSM2005/Household.pl` を作成した。このプログラムは関数 `household_type` を定義する。モジュールではないので、つぎのようにして呼び出す。

```
require 'SSM/SSM2005/Household.pl';
.....
household_type;
```

この関数は、呼び出す側と分離されておらず、おなじ「空間」で動作する。世帯類型をあらわすさまざまな変数をつくって呼び出し側の空間を「汚染」するので、注意されたい。

5.3. 関数

SSM データ分析用の関数としては、教育年数を求める `Eduyear()` と、所得を 1 万円単位の値に変換する `Income_10000yen()` を輸出する。

```
$Eduyer = Eduyear ($ED_SSM);
$Pinc_metric = Income_10000yen($Q33A);
```

6. 職歴分析のためのモジュール

`SSM::Career`, `SSM::SSM2005::Career`, `SSM::SSM1995::Career` の各モジュールは、職業の変数処理のためのメソッドを提供する。これらのうち、`SSM::Career` は 1995 年 SSM 調査と 2005 年 SSM 調査の両方に共通のメソッドを提供する「親」のクラスを定義している。あとのふたつは、それぞれ 1995 年調査と 2005 年調査に特殊なメソッドを提供するものであるが、`SSM::Career` から共通処理のためのメソッドを継承しているため、必要があれば共通処理のためのメソッドを「親」クラスから自動的に呼び出す。これらのモジュールは、OOP のインターフェースにしたがって設計されているため、何も輸出しない。

6.1. オブジェクトの設計

SSM 調査における「職歴」(career) のデータは、初職から現職までの職業の変化を全て記録したものである。職業上の移動 (従業先の移動と同一従業先内での異動の両方をふくむ) があるたびにひとつの「段」(stage) として記録される。2005 年データでは最大 16 段、1995 年データでは最大 17 段のデータが記録されている。

個々の「職歴段」は、つぎのような変数の集合体である。

page: 職歴段の番号

number: 従業先の番号

industry: 産業 (SSM 産業分類コード)

size: 企業規模 (従業員数のカテゴリー)

status: 従業上の地位

job: 仕事の内容 (SSM 職業小分類コード)

post: 役職

start: 開始年齢

reason: 転職理由 (2005 年調査のみ; 従業先移動の場合のみ)

reason7: 転職理由の「その他」自由記述 (2005 年調査のみ; 従業先移動の場合のみ)

isic: 国際産業分類 ISIC (2005 年調査のみ)

isco: 国際職業分類 ISCO (2005 年調査のみ)

income: 収入の変化 (2005 年調査のみ; 従業先移動の場合のみ)

これらの変数の実体は、たとえば \$Q08D02 などの名前の変数として格納されている。「職歴オブジェクト」はそれらの変数へのポインタを持つだけである。また、職歴オブジェクトは元データのオブジェクトに関する情報を持っており、職歴以外の変数の情報も呼び出せる。

職歴オブジェクトは、メソッド `new()` によって構築する。引数として、元データの変数へのポインタを渡すと、`new()` は元データから職歴データの格納されている変数名を探して、職歴を組み立てる。その際、職歴の最初に、「第 0 段」を追加する。これは初職より前の職歴にあたるものであり、全員無職、段番号や従業先番号や開始年齢は 0 にしておく。

メソッド `empty` によって、空の (つまり段数 0 の) 職歴オブジェクトをつくることができる。

SSM::Career (およびその派生クラス) が提供するその他のメソッドは、「オブジェクト抽出メソッド」と「メンバ配列メソッド」の 2 種類に分かれる (図 3)。以下、それぞれについて解説する。

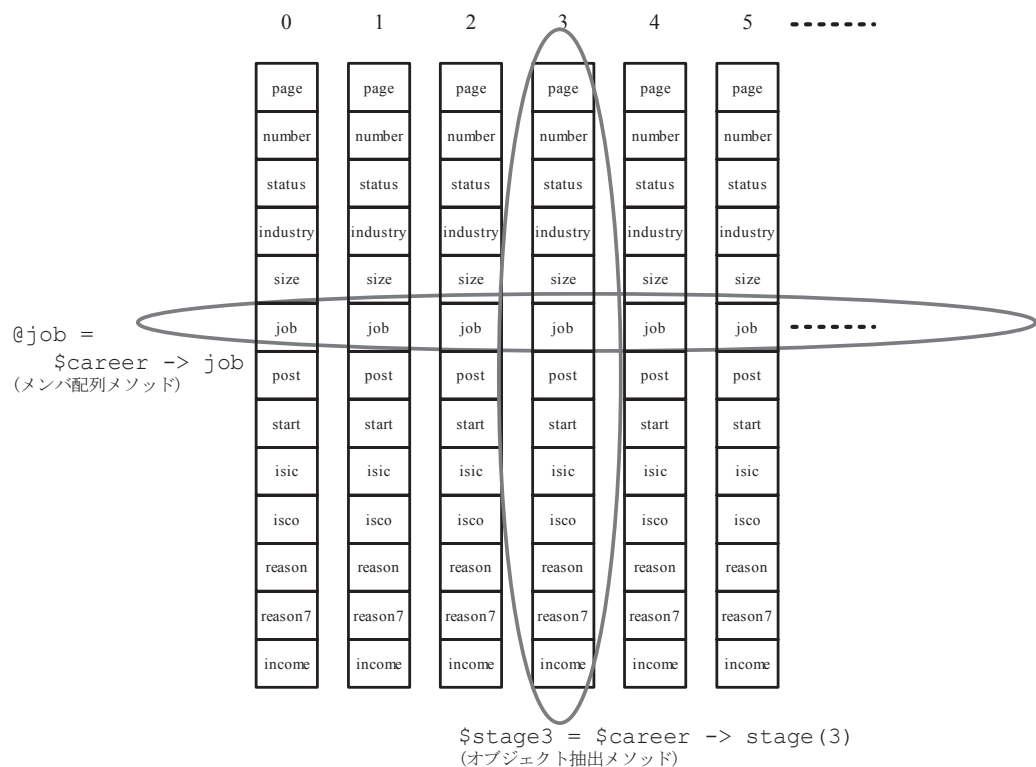


図 3. Manipulation of a career object

6.2. オブジェクト抽出メソッド

「オブジェクト抽出メソッド」は、職歴オブジェクトからいくつかの職歴段を抜き出すものである。抜き出した結果もまた職歴オブジェクトとしてあつかわれるので、メソッドを重ねて使うことができる。

メソッド `valid_range` は、そのケースに対して有効な職歴段の範囲を抜き出す。職歴のデータは、2005 年データでは 16 段、1995 年データでは 17 段存在する。しかしこれは最大の段数をもつケースにあわせたものであり、ほとんどのケースはこれよりもすくない段数だけが記録されていて、現職に到達したあとはすべて欠損値となっている。`valid_range` は、職歴データの末尾についているこのような欠損値部分を切り落とし、「有効」な範囲の職歴段だけを抜き出す。多くの場合、ケースごとにあらかじめこの処理をしてから分析するほうが混乱がすくないだろう。

```
.....
use SSM::SSM2005::Career;
$career0 = SSM::SSM2005::Career -> new($data); # 職歴オブジェクトを定義
.....
while( $data -> read_binary ){
```

```

$career = $career0 -> valid_range;          # 有効部分のみ
# いろいろ分析する
}

```

メソッド `stage()` は、引数であたえた数字に該当する番号の段を抜き出す。引数をふたつ以上あたえてもよい。

```

$third_stage = $career -> stage(3);          # 3 番目の職歴段
$odd_stages = $career -> stage(1,3,5,7);    # 1,3,5,7 番目の職歴段

```

メソッド `first, last` は、それぞれ最初・最後の職歴段を取り出す。

```

$first_stage = $career -> first;             # $career -> stage(0) と等価
$last_stage = $career0->valid_range->last;   # $career0->stage($DANSU) と等価

```

`firstjob` は「初職」を取り出すメソッドである。就業経験がない場合は、職歴の「第 0 段」(上述) を返す。それ以外の場合は、たいていは `stage(1)` とおなじである。ただし、職歴第 1 段の従業上の地位が無職または欠損である場合がありえるため、それらについて例外処理をおこなう。第 1 段が無職であれば、有職の段が見つかるまで 2 段目以降を探索する。従業上の地位が欠損であれば、空の職歴オブジェクト (`empty` メソッドとおなじ) を返す。

結婚前の初職を取り出すのが `firstjob_single` メソッドである。結婚前に就業経験のない者については、第 0 段 (無職) を返す。ただし、結婚前に第 1 子が誕生している場合には、結婚時年齢ではなく第 1 子誕生時年齢 (田中 (1997) と同様の計算方法で調整する) を使って計算する。もし結婚時年齢が無回答なら、空の職歴オブジェクト (`empty` メソッドとおなじ) を返す。これら以外の場合は `firstjob` とおなじ結果を返す。

メソッド `at_age` (下限年齢, 上限年齢) はふたつの引数を取り、下限年齢から上限年齢までの間に経験された職歴段を抽出する。

```

$career_young = $career -> at_age(20,29);   # 20 代の職歴

```

引数をひとつだけ指定することもできる。その場合には、その年齢に該当する職歴段を抽出する。ただし、移動がおこった年齢にあたっている場合、複数の職歴段が抽出されることがある。職歴段をひとつに限定したいなら、`first` (移動前の職歴段を採る) や `last` (移動後の職歴段を採る) などを利用する。

```

$Age_lc = $data->Age_last_birth;            # 末子誕生時年齢
$career_lc = $career -> at_age($Age_lc) -> last;

```

メソッド `each` は、職歴オブジェクトを個々の職歴段に分解し、それぞれを要素に持つ配列を返す。この配列を操作することによって、個々の職歴段を操作することができる。

```

foreach $x ($career->each) {

```

```

    print $x -> status; # 従業上の地位を出力（後述の status メソッドをみよ）
  }

```

6.3. メンバ配列メソッド

「メンバ配列メソッド」は職歴オブジェクトにふくまれる職歴段から各フィールドの値をとりだし、あるいはそれらを操作して一定の値に変換したものの配列を返す関数である。返値はオブジェクトではなく単なるスカラ値の配列なので、メソッドを重ねて使うことはできない。

各職業変数の値を取り出すメソッドには、上述の各フィールドの名前 (page, status, industry, job, size など) がそのままつけてある。これらの名前を指定すれば、その名前に該当する変数の値が配列になって返る。

```

print $career->status; # 従業上の地位の変化を観察する

```

上述の each メソッドを使ってつぎのように書いても同様の結果がえられる。

```

print map {$_->status} $career->each; # 職歴段個別に status メソッドを適用

```

オブジェクト抽出メソッドと重ねて使うと、次のような操作が可能である。

```

$Age_lc = $data->Age_last_birth;
$career -> at_age($Age_lc) -> last -> status ;

```

このコードは、つぎのように動作する

- ・末子誕生時の職歴段をとりだす
- ・取り出された職歴段が複数である場合には、最後のものを採る
- ・その職歴段の「従業上の地位」の値を取り出す

メソッド have_job は、status で抜き出した「従業上の地位」の値のうち、有職のものだけについての配列を返す。この値を条件分岐に使うことができる。

```

if( $career->at_age($Age_lc)->last->have_job ){
  # 末子誕生時に有職だった者だけについての処理を何か書く
}

```

メソッド job8598 は、1995年版SSM職業分類（1995年SSM調査研究会 2006）について、1985年以前のSSM調査データとの整合をとるために 管理職に関する変換（2005年社会階層と社会移動調査研究会 2007, pp. 89–94）をおこなった職業小分類コード（3桁）を返す。

メソッド class_job8 は、「SSM職業大分類」（1995年SSM調査研究会 2006, pp. 101–5）の8分類を返す。

```

$j8 = $career -> firstjob -> class_job8;
print '専門' if $j8==1;
print '管理' if $j8==2;
print '事務' if $j8==3;
print '販売' if $j8==4;
print '熟練' if $j8==5;
print '半熟' if $j8==6;
print '非熟' if $j8==7;
print '農林' if $j8==8;
print '無職' if $j8==Missval('OUT');
# これら以外の場合は職業コード欠損

```

職業小分類コードについては、あらかじめメソッド `job8595` によって管理職に関する変換がおこなわれる。もし変換前のコードに基づいた大分類が必要であれば、`SSM::Occ95` で定義されている関数 `occ8()` を呼び出して使うこと。

```

.....
use SSM::Occ95;
.....
while( $data->read_binary ){
    $career = $career0->valid_range;
    print map { occ8($_) } $career->job; # メソッド job の返値を occ8() で変換
}

```

メソッド `class_new8()` は、総合職業分類 (原・盛山 1999) を作成する。引数には、「小企業」の上限を指定する (「企業規模」の変数値を使用する)。6 (300 人未満まで) を指定すると、原・盛山 (1999) の使用している分類とおなじ結果がえられる。値のあたえかたは次のとおりである。

- 1: 専門
- 2: 大企業・官公庁の管理・事務・販売
- 3: 小企業の管理・事務・販売
- 4: 自営業の管理・事務・販売
- 5: 大企業・官公庁の熟練・半熟練・非熟練
- 6: 小企業の熟練・半熟練・非熟練
- 7: 自営業の熟練・半熟練・非熟練
- 8: 農林漁業
- Missval('OUT'): 無職
- その他の欠損値: 職業変数のどれかが欠損

メソッド `class_status4` は、従業上の地位 (および一部は職業分類) に基づいて 4 つの階級に分類したものを返す。

- 1: 経営者・自営・家族従業者
- 2: 常時雇用されている一般従業者
- 3: 臨時雇用・パート・アルバイト・派遣社員・契約社員・内職

9: 無職

これらのメソッドを組み合わせて、さまざまな職業階級分類をつくることができる。たとえば、無職を 0、自営を 9、非正規雇用を 10 として、常時雇用の場合のみ総合職業分類を適用するあたらしいメソッドをつくるには、つぎのようにする。

```
sub class10 {
  my ($self) = @_;                                # 操作の対象になるオブジェクト
  my $N8 = $self -> class_new8(6);                # 総合職業分類
  my $S4 = $self -> class_status4;                # 従業上の地位 4 分類
  my $class;
  if( miss($S4) ) { $class = $S4; }               # 欠損値
  elsif( 1==$S4 ) { $class = 9; }                 # 自営など
  elsif( 2==$S4 ) { $class = $N8; }               # 常時雇用されている一般従業者
  elsif( 3==$S4 ) { $class = 10; }                # 臨時雇用・パート・内職など
  elsif( 9==$S4 ) { $class = 0; }                 # 無職
  else{ warn( "ERROR: Illegal value $S4" ); }     # 以上で全部のはず
  return $class;
}
```

7. 使用例

7.1. 結婚前初職と末子誕生時の職業移動

結婚前の初職と末子誕生時との間の職業移動をとらえるには、つぎのようにするとよい。

```
$Fjs10 =$career -> firstjob_single -> class10;    # 上記の class10 メソッド
$Age_lc = $data->Age_last_birth
$Lc10 = $career -> at_age($Age_lc) -> last -> class10;
output (
  'Id', $data->Id,
  'Sex', $data->Value('Sex'),
  'Age', $data->Value('Age'),
  'MarStat', $data->Value('MarStat'),
  'MarAge', $data->Age_marriage,
  'Age_lc', $Age_lc->Age_last_birth,
  'Fjs10', $Fjs10,
  'Lc10', $Lc10,
);
```

ID, 性別, 年齢, 婚姻状況, 結婚時年齢, 末子誕生時年齢, 結婚前初職分類, 末子誕生時職分類がファイルに出力される。

7.2. 職歴の “parson-year” データ

保田時男氏による SPSS シンタックス ssm05py.sps (2007-07-18) の職歴部分 (およそ 400 行のコード) と同等のことをおこなう。15 歳から 70 歳までの各年齢時の職業に関する変数 (12 個) を書き出したファイルが出力される。出力されるデータは 1 行が回答者ひとり分である。各回答者を特定するための変数 3 個と各年齢時の職業に関する変数 $(70-15+1) \times 12 = 672$ 個

をあわせた 675 個のフィールドが各行に出力される。

```
..... 〔省略〕 .....
# 変数名のリスト
@field = qw(
    page number reason industry isic size
    status job isco post start income
);
@vname = qw(ISNUM ISN ISA ISB ISBI ISC ISD ISE ISEI ISF ISPA ISG);

# 先頭に変数名を出力
@header=();
foreach $a(15..70){
    push @header, map{ $_ . $a }@vname;
}
print 'NUMS' , 'NUMP' , 'NUMI', @header;

# 1 ケースずつ処理
while( $data -> read_binary ){
    $career = $career0 -> valid_range; # 有効な職歴段のみ
    @result=();
    foreach $a(15..70){
        $j = $career->at_age($a)->last; # 各年齢時の職歴段
        foreach $f( @field ){
            push @result, $j -> $f; # 各フィールドの値を配列に格納
        }
    }
    print $NUMS , $NUMP , $NUMI, @result; # 出力
}
```

付録: Perl 語彙

Perl で使われる記号と予約語について、簡単に説明する。本稿を理解するための最小限の説明にとどめているため、不正確な場合がある。

```
# —— ここから行末までコメント
; —— ひとつの「文」の終端
$ —— 単一のスカラ値を持つ変数
% —— ハッシュ全体
@ —— 配列変数
@_ —— 関数・メソッドへわたされた引数の配列
:: —— モジュール名の区切り
-> —— メソッドの呼び出し
'...' —— 文字列
"¥t" —— タブ記号
++ —— 数値をひとつ増やす
= —— 左辺に右辺の値を代入
== —— 両辺が (数値として) 等しい
eq —— 両辺が (文字列として) 等しい
foreach $i(...) {...} —— () 内のリストを順番に $i に代入して {} 内のコードを繰り返す
map {...} —— 配列の全ての要素に対して {} 内のコードを実行
my —— {} で囲まれたブロック内だけで通用する名前を宣言する
package —— モジュール等が使う「名前空間」を定義
print —— 標準出力に出力
```

push — 配列の末尾にあたらしい要素を追加
qw(...) — () 内の文字列を、空白で区切られたリストに変換
require — 他ファイルの Perl プログラムを読み込む
return — 関数・メソッドの処理を終え、「返値」を返す
sub — 関数・メソッドの定義
use — モジュールの読み込み
warn — 警告を表示する
while(...){} — () 内の条件式が真である限り {} 内のコードを繰り返す

文献

- 1995 年 SSM 調査研究会. 2006. 『SSM 産業分類・職業分類 (95 年版)』(修正版).
2005 年社会階層と社会移動調査研究会. 2007. 『2005 年 SSM 日本調査 コード・ブック』.
Conway, Damian; 山根ドキュメンテーション (訳). 2001. 『オブジェクト指向 Perl マスターコース』ピアソン・エデュケーション.
原純輔・盛山和夫. 1999. 『社会階層: 豊かさの中の不平等』東京大学出版会.
国立社会保障・人口問題研究所. 2007. 『わが国夫婦の結婚過程と出生力: 出生動向基本調査 第 3 回』厚生統計協会.
日本家族社会学会 全国家族調査委員会. 2005. 『第 2 回家族についての全国調査 (NFRJ03): 第 1 次報告書』.
田中重人. 1997. 「高学歴化と性別分業」『社会学評論』48: 130–42.
Wall, Larry, and Randal L. Schwartz; 近藤嘉雪 (訳). 1993. 『Perl プログラミング』ソフトバンク.
Wall, Larry, Tom Christiansen, and Jon Orwant. 2000. *Programming Perl* (3rd ed). Sebastopol, CA: O'Reilly.

Development of Modules for Data Reduction: For Efficient Analysis of Occupational History

Sigeto TANAKA

Tohoku University

This paper proposes a “toolbox” approach for analyses of data with a complex structure. In this approach, you use any programming language for preprocessing before statistical analyses. It is an alternative to the “all-in-one” approach in which you use a statistical package to conduct both data processing and statistical analyses. It is efficient to use a full-fledged programming language instead of poor data processing functions of statistical packages. And it is more efficient to utilize “modules” prepared to meet common needs. The paper focuses on the data structure and the object-oriented modules to process occupational history. The author has developed some modules in Perl, and introduces the basic feature of the modules. This paper also includes some programs to make class categories and parson-year data of personal histories. Further details are available from <http://www.sal.tohoku.ac.jp/~tsigeto/dredu/>.

Keywords and phrases: data structure, Perl, career analysis, object-oriented programming

機械学習による ISCO 自動コーディング

高橋和子
(敬愛大学)

【要旨】

国際比較研究の活発化に伴い、社会調査における職業コーディングにおいても、SSM (Social Stratification and Social Mobility) 職業分類コードに加えて、国際標準職業分類である ISCO (International Standard Classification of Occupation) -88 コードの付与が要請されるようになってきた。ISCO-88 コードは SSM 職業分類コードから単純な変換により生成することが困難なため、新たに ISCO コーディングを行って付与する必要がある。しかし、職業コーディングは煩雑な作業でありコードの負担が大きく多大な時間と労力を要するために、2 種類のコードを付与する作業 (ダブル・コーディング) を行うことは、深刻な問題となる。

この問題を解決する一つの方法として、コーディングの自動化が有効である。実は、SSM 職業コーディングに関しては、すでにルールベースや機械学習による自動化のアルゴリズムが提案され、ユーザーインタフェース機能が付加された実用システム (NANACO システム) として利用が進みつつある。これを参考にして、ISCO コーディングにおいても、効果的な自動化の方法を検討することが望まれよう。

本研究では、ISCO 自動コーディングの有効な方法を検討し、SSM 職業分類コードを素性とする機械学習による方法を提案する。提案手法は、「2005 年 SSM 社会階層と社会移動日本調査」データによる実験の結果、有効性を示した。職業コーディングにおいて用いられるコード体系は今後も変更される可能性があるが、提案手法はこのような状況にも容易に対応できる。

キーワード：自動コーディング、ISCO、SSM 職業コード、機械学習、職業コーディング、NANACO システム

1. はじめに

本研究の目的は、国際比較研究の活発化に伴って、最近必要性が高まってきた「ISCO (International Standard Classification of Occupation; 国際標準職業分類) -88 コード」(以下 ISCO コードと略する) を機械学習により自動的に付与する方法を検討し、提案することである。ISCO 自動コーディングの方法を検討する前に、まず、職業コーディングについて、各国の状況も含めて述べておく。

職業コーディングは、自由回答である「仕事の内容」や「従業先事業の種類」と、選択回

答である「従業上の地位」や「役職」など職業に関する複数の回答を総合的に判断し、数百種類の職業コードから該当するコードを1つ選んで付与するという煩雑な作業であるが、職業データを統計処理するためには必須である（原・海野 1984）。職業コーディングにおいて用いられるコード体系は、各国における職業の概念の違いによりさまざまに異なっており¹（岡本 2004, 西澤 2006）、また同じ国であっても年度により異なる場合がある²。我が国においては、ISCOコードの旧版であるISCO-68 コードの流れを組む「日本標準職業分類」（JSCO；Japanese Standard Classification of Occupation）を基本とし、国勢調査ではJSCOに準拠した国勢調査職業分類、社会調査ではJSCOを簡略化した「SSM（Social Stratification and Social Mobility）95 職業分類コード」（以下SSM職業コードと略する）（1995 年SSM調査研究会 1995）が標準的に用いられている。このように各国でさまざまに異なる職業コードが用いられる中で、国際標準の職業コードとして、1988 年に国際労働機関（ILO）で改訂されたISCOコードが存在する（Bureau of Statistics; International Labour Office 2001）。

ISCOコードは、分類の単位を、個人の遂行する課業（task）とそれにとりあう責務（duty）からなる職務（job）に置く。ISCO-68 コードと大きく異なる点は、分類の基準として、新たに「技能度（skill level）」³が採用されたことである（西澤 2006）。ISCOコードは、各桁が分類上の意味を持つ4桁の桁別分類コードで、大分類10個、重大分類28個、中分類116個、小分類390個からなる。一方、SSM職業コードは3桁の順番コードで小分類196個からなるが、ISCOコードのような分類基準をもたないISCO-66 コードを源とするため、ISCOコードとの間に単純な対応関係を見いだすことは困難である⁴。このような事情により、「2005 年SSM社会階層と社会移動日本調査」（以下「2005 年SSM日本調査」と略する）では、ISCOコードをSSM職業コードから変換するのではなく、新たにISCOコーディングを行って付与することにした。

ところで、職業コーディングはテキスト型のデータである自由回答の分類であるために、どのようなコード体系が用いられていてもコードにかかる負担が大きく、多大な時間と労力を要する。また、我が国においては、正確性を期するために同一のデータに対して複数回のコーディングが行われるため、さらに大きな問題となる（高橋 2000, 盛山 2004）。これらの問題は、特に海外で強く認識されており、統計局レベルで自動化に向けての取り組みが行われてきた。その結果、オーストラリア、米国、カナダやフランスなどで自動コーディングの方法が提案されている⁵（Kunz 2003, Creecy et al. 1992, Riviere 1994, 岡本 2004）。し

¹ 例えば、英国ではSOC 2000、オーストラリアではASCO2、米国ではSOC、カナダではNOC-S2001なる職業コード体系が用いられており、小分類はそれぞれ順に353個、340個、449個、520個である。このうち、ASCO2やSOCではさらに細分類があり、それぞれ986個、821個である。

² 例えば、米国センサスでは、1970年と1980年で異なる職業コード体系が用いられた（Rubin 2004）。

³ skill specialization という用語が用いられている。

⁴ 例えば、『職業分類におけるSSM95とISCO-88の対応』（2005年SSM調査研究会職業分類タスクグループ2004）によれば、「2003年仕事と暮らしに関する全国調査」において1対1に対応する事例は約30～40%程度であったとの報告がなされている。

⁵ 例えば、オーストラリアではシソーラスベースのPrecision Data、米国ではAIOCSやPACE、カナダではACTR、

かし、いずれもデータベースにおける検索の域を出ず、単語や文字単位（場合によっては、2-gram⁶や3-gramを単位とする）の完全一致による方法である。これに対して、我が国では、社会調査におけるSSM職業コーディングを対象とし、自然言語処理により、表層的ではあるが「意味」を考慮した方法が提案された（高橋 2000）。すなわち、職業を表現する手がかりとして動作を表す述語に注目し、文中の名詞を述語との関係で捉える「格フレーム」（徳永 1999）の概念を利用して、職業の定義内容を表現するルールを生成した⁷。また、これらのルールと回答の意味的な一致を捉えるために、両者に出現する述語と名詞をそれぞれシソーラスにより拡張した。ルールベースによる方法による自動コーディングの結果は、コードが利用しやすいように表計算ソフトにより表示する機能が追加され、ROCCO (Rule-based OCCupation CODing) システムと名付けられた。ROCCOシステムは、JGSS (Japanese General Social Surveys⁸) を始めとする 5 つの調査で利用され、一定の評価を得た（高橋 2001; 2002; 2003）。

ルールベースによる方法は人間が結果を理解しやすい反面、準備されたルールにマッチした事例しかコードを決定できないことや、ルールセットのメンテナンスに手間がかかるという問題があるために、次に、まったく異なるアプローチとして機械学習による方法が提案された（高橋ほか 2004a; 2004b）。これは、自由回答を非常に短い文書と捉えることで、文書分類と同様の分類方法を適用しようとするもので、自由回答に出現する品質付きの単語を分類器が学習する素性として利用することを想定する。ここで、機械学習にはいくつかの方法があるが、高橋ほか（2004a; 2004b）では、多くの文書分類タスクにおいて精度の高さが評価されているサポートベクターマシン（SVM）（Vapnik 1998, Joachims 1998, Dumais et al. 1998, 平・春野 2000, Sebastiani 2002）を用いた。

機械学習による方法はルールベース手法を上回る性能を示したが、コードと同等の性能を目指すために、ルールベース手法のもつ貴重な「知識」を活かす方法が検討された（高橋ほか 2005a, Takahashi et al. 2005）。その結果、ルールベース手法と機械学習を組み合わせる 4 種類の方法が提案されたが、どの方法も機械学習を単独に用いる方法よりよい結果を示した。中でも、ルールベース手法による自動分類結果を SVM の素性として用いる方法は最もよかったため、ユーザーインタフェース機能が強化されて、コードのためのコーディング支援システム（NANACO システム）として実用化された（高橋ほか 2005b）。NANACO システムは、JGSS や「2003 年仕事と暮らし調査」など 6 つの調査において利用され、コードの負担軽減やコーディング時間の短縮などに役立っている（高橋 2007b）。

フランスでは SICORE と呼ばれる自動コーディングシステムが提案されており、英国、イタリア、アイルランドにおいても検討が行われている（Keogh 1998）。

⁶ 例えば、「work at the office」という回答の場合、単語単位の 2-gram では、「work at」「at the」「the office」が処理の単位となり、文字単位の 2-gram では、「wo」、「or」、・・・、「ic」、「ce」が処理の単位となる。

⁷ この他に、コード・ブック（1995 年 SSM 調査研究会 1996, 西村・石田 2001, 石田・三輪 2006）に記述された内容やコードが判断する際に用いる知識も可能な限りルール化されている。

⁸ <http://jgss.daishodai.ac.jp/japanese/frame/japanesetop.html>.

以上に述べた状況を踏まえ、本研究においては、ISCO 自動コーディングのための有効な方法を検討する。まず、利用される職業コードの体系が今後も変更される可能性を考慮すると、人手により毎回新たなルールを生成することはコストの面で問題があると思われるために、ルールベース手法ではなく機械学習による方法を検討することにする。ただし、機械学習による方法は大量の訓練事例を必要とし、事例数が少ない場合には高い精度を期待できないという問題があるために、これを補う方法を検討する必要がある。そこで、例えば、SSM 職業コーディングにおいて最も性能が高かった「機械学習による方法にルールベース手法を組み合わせた方法」に注目したが、ISCO コーディングにおいてはルールベース手法が存在しないという致命的な問題が存在する。これについては 2 節の関連研究で述べる。

そこで、次に、まったく別のアプローチとして、今回の ISCO コーディングと SSM コーディングを同一の事例に対するダブル・コーディングであると考え、SSM コードしか付与されていない事例は ISCO コードが欠測したデータであると解釈することも可能である。例えば、Rubin (2004) は、1970 年米国センサスのデータに対する 1970 年職業コードと 1980 年職業コードのダブル・コーディングを 1980 年職業コードの欠測値問題として扱った。そこで用いられた変数は、性別、年齢、年収などに限定されたが、もし、今回、自由回答を対象とすると、出現単語のすべてが変数となるために変数空間がスパースになるという問題が生じる。これについても 2 節の関連研究で述べる。

以下、次節の関連研究においては、上述した 2 つのアプローチについて述べる。第 3 節で提案手法を説明し、第 4 節で実験を行って提案手法の有効性を示す。最後に第 5 節でまとめる。

2. 関連研究

ここでは、「ダブル・コーディングにおける欠測値問題」としてのアプローチ (Rubin 2004) および、機械学習による方法にルールベース手法を組み合わせた方法の中で最も性能の高かった「機械学習の素性としてルールベース手法による結果を利用する方法」(高橋ほか 2005, Takahashi et al. 2005) の 2 つの研究を取り上げる。

2.1 ダブル・コーディングにおける欠測値問題としてのアプローチ

Rubin (2004) は、欠損値補充問題の一例として、1970 年の米国センサスのデータに対して、すでに付与されている 1970 年職業コードに加えて、1970 年と全く異なるコード体系の 1980 年職業コードも付与するというダブル・コーディングの問題を取り上げた。前提条件と

して、1970 年のデータの中に両方のコードが付与されている事例の存在⁹が必要であるが、両方のコードが付与されていないデータは 1980 年職業コードが欠測していると考えられることができるために、他の欠測値補充問題と同様のアプローチが可能である。詳細は省略するが、Rubin は、欠測の原因を無視できる (ignorable nonresponse) 場合と無視できない (nonignorable nonresponse) 場合に分けて検討を行った。ダブル・コーディングは無視できる場合であるとして、Hot-deck と呼ばれる方法 (欠測値をもつある事例のいくつかの変数の値と同じ事例を欠測値がない事例から見つけて、欠測値にその事例の値を補充する) および、ロジスティック回帰による方法による補充を提案した。このとき用いられた変数は、年収、性別、年齢などであった。

欠損値を補充するアプローチには、職業コードを決定するために最も重要な情報となる自由回答を変数とすることが困難である。なぜなら、自由回答に出現する単語は多岐にわたるため、これらをすべてダミー変数としてか出現したか否か 1 または 0 で表現すると、変数空間がスパースになることが予想されるからである。したがって、本研究では、ISCO 自動コーディングに対して、欠測値補充問題としてのアプローチは検討しないことにする。

2.2 他の分類器による結果を機械学習の素性として利用する方法

SSM 職業コーディングにおいては、ルールベース手法による自動コーディングの結果 (SSM 職業コード) を機械学習の素性として学習させる方法 (以下 add-code 法と呼ぶ) が最もよい結果を示した (高橋ほか 2005a, Takahashi et al. 2005, 高橋 2007b)¹⁰。add-code 法は、実験時 (訓練事例数が約 13,000 サンプル) に、自動コーディングにおける正解率 (4.1 節式 (1) を参照のこと) の目標値として設定した 75% (JGSS におけるコードの平均正解率) をやや下回る 74.4% であったが、基本素性だけ用いた場合よりは約 3% 高く、また、その後に実際に適用された調査 (JGSS-2003) では、80.7% という高い正解率を示した¹¹ (高橋 2007b)。したがって、機械学習とルールベース手法と組み合わせた add-code 法は、本研究の参考になり得ると考えられる。

add-code 法の適用を検討する際にまず問題となるのは、ISCO コーディングにおいてはルールベース手法が存在しない点である。これに対しては、ルールベース手法は機械学習ではないが分類器の一つであると考えことは可能で、このような立場に立てば、組み合わせる手法がルールベース手法に限定される必要はなく、add-code 法を他の分類器による出力を利用する機械学習法、すなわち一種のスタッキング (Wolpert 1992) であるとみなすことができる。さらにいえば、add-code において必要なものは他のシステムによる分類結果であるた

⁹ 120,000 サンプルに対しては両方のコードが付与されていたが、90%以上の欠測率であった (Rubin 2004)。

¹⁰ ただし、add-code 法による結果は、はずれ値の重みを決めるソフトマージン・パラメータ C の値により、結果に違いがあるために、あらかじめ全訓練事例を用いて C の値をチューニングしておく必要がある。

¹¹ この理由は、訓練事例が約 13,300 サンプルから約 20,000 サンプルに増えたためであると考えられる。

めに、分類器による自動分類結果である必要はなく、例えば、人手による分類の結果でもよいことになる。次に問題となる点は、add-code 法は、機械学習とルールベース手法という 2 つの「分類器」がともに同じ分類クラス（ここでは職業コード体系である。以下クラスと略する）であることを想定しているのに対して、本研究では、両者においては、職業というドメインは同じもののクラスが異なることである。しかし、add-code 法で利用されたものが同じクラス同士の直接的な知識であるとするならば、今回は、何らかの関連がある異なるクラスの結果といういわば間接的な知識であると考えられることができるために、ある程度の有効性が期待できるのではないかとと思われる。

3. 提案手法

ISCO 自動コーディングにおける方法を検討する際にさまざまなアプローチが可能であるが、本研究では、SSM 職業コーディングと同様の機械学習（SVM）を用いた場合¹²に、どのような素性が有効であるかということに焦点を絞って検討する。このとき、まったく新しい素性から検討を始めることはせずに、add-code 法を参考にして、SSM 職業コーディングで有効であった「基本素性」に新しい素性を追加する方針をとる。ここで、本研究における基本素性とは、職業に関するデータにおける「仕事の内容」および「従業先事業の種類」に出現する品詞付き単語（以下単語と略する）、「従業上の地位・役職」の選択肢である。基本素性は、職業に関するデータのすべてではなく、例えば「従業先事業の規模」は用いない¹³。

検討の結果、本研究では、ISCO 自動コーディングのための有効な方法として、SSM 職業コーディングにおける基本素性である「仕事の内容」および「従業先事業の種類」に出現する単語、「従業上の地位・役職」の選択肢に、新たな素性として「SSM 職業コード」を追加した機械学習（SVM）による方法を提案する（図 1）。

- ・ 「仕事の内容」に出現する単語（原形）
 - ・ 「従業先事業の種類」に出現する単語（原形）
 - ・ 「従業上の地位・役職」の選択肢（10 種類）
 - ・ 「SSM 職業コード」（約 200 種類）
- } 基本素性

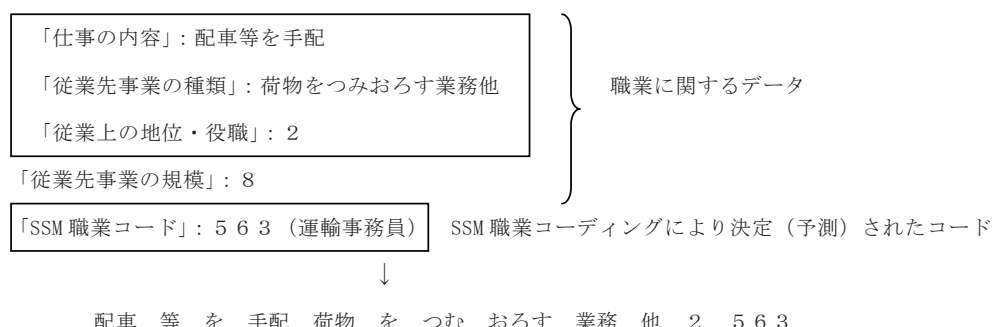
図 1 提案手法で利用される素性

「SSM 職業コード」は、過去の調査などですでに SSM 職業コードが決定されている場合（本研究においては「正解」と呼ぶ）には正解の SSM 職業コード、もしまだ SSM 職業コードが最

¹² 分類器についても、SSM 自動職業コーディングにおいて有効であった分類器を想定する。

¹³ この理由は、SSM 自動職業コーディングの有効な方法を調査する実験の結果、「従業先事業の規模」を素性に加えても正解率の向上が認められなかった（高橋ほか 2005）ことによる。

最終的に決定されていない場合には、人手やコンピュータにより予測された SSM 職業コードを指す。このため、本研究では、SSM 職業コーディングと ISCO コーディングの両方を同時に行う場合には、SSM 職業コーディングを先に行われることを想定している。図 2 に提案手法における素性の例（枠で囲んだ部分）を示す。



注）実際に処理される段階では、素性は単語ではなく素性番号に変換されている（図 3 も同様。後述 STEP 2 を参照のこと）

図 2 提案手法における素性の例

提案手法の手順を示す。

STEP 1

形態素解析¹⁴により、自由回答である「仕事の内容」および「従業先事業の種類」に記述された文を単語に区切る。

STEP 2

STEP1 で得られたすべての単語の原形から素性辞書を作成し、素性辞書に基づいて各単語を素性番号に変換する¹⁵。このとき、同じ単語であっても、品詞が異なれば異なる素性番号にする。また、単語が出現した場所が「仕事の内容」と「従業先事業の種類」で異なる場合には、両者を区別するために異なる素性番号にする必要がある¹⁶。

STEP 3

選択肢「従業上の地位・役職」および、処理対象とする事例に付与された「SSM 職業コード」を素性に追加する。このとき、SSM 職業コードが付与されていない場合は、何らかの「分類器」により予測されたコードを用いる。

STEP 4

すでに ISCO コードの正解が付与されている事例を訓練事例として利用するため、素性の組

¹⁴ 形態素解析用ソフトとしては、京都大学長尾研究室で開発された JUMAN（黒橋・長尾 1998）がある。

¹⁵ 素性を離散値に変換する理由は、<http://chasen.org/~taku/software/TinySVM/> を利用するためである。

¹⁶ 「従業先事業の種類」に出現した単語のすべてに対して、素性辞書中の素性番号にある定数を加えることで機械的に別の素性番号を生成することができる。

と ISCO コードの正解を結合する。例えば、図 2 で用いた例に正解の ISCO コード「4133」（運輸事務員）が付与されている場合には、図 3 に示すような訓練事例が生成される。

4 1 3 3 配車 等 を 手配 荷物 を つむ おろす 業務 他 2 5 6 3

図 3 提案手法における訓練事例の例

STEP 5

訓練事例を用いて分類器を生成する。

STEP 6

分類器により評価事例进行分类する（ISCO コードを付与する）。

4. 実験

実験の目的は、提案手法の有効性を示すこと（実験 1）および、提案手法におけるエラー解析を行う（実験 2）。

4.1 実験設定

分類器

分類器は SVM を用いた。ただし、SVM は本来 2 値分類器であるために、one-versus-rest 法（Kressel 1999）により多値分類器に拡張した。また、はずれ値の重みを決めるソフトマージン・パラメータ C の値は、0.1 から 1.0 まで 10 通りに変化させた。なお、SVM におけるカーネル関数は、線形カーネルを用いた。以上は、SSM 職業コーディングの場合と全く同様である。

データセット

データセットは、「2005 年 SSM 日本調査」における本人現職（有職者）の全 4,133 サンプルを用い、10 分割交差検定を行った。すなわち、サンプル全体の 9/10 を訓練事例、1/10 を評価事例とするセットを 10 通り生成し、各セットにおける訓練事例と評価事例のペアごとに実験を行い（計 10 回）、その結果を平均した値を実験結果とした。

評価尺度

評価尺度としては（1）式により計算される正解率を用いた。正解率が高いほどよい手法であるとした。

$$\text{正解率} = \text{正しく分類できた事例数} / \text{全事例数} \quad (1)$$

4.2 追加する素性確認のための予備実験

「2005 年 SSM 日本調査」による実験に先立ち、「2003 年仕事と暮らし調査」を用いて予備実験を行った。実験の目的は、ISCO 自動コーディングにおける有効な素性の候補を発見することである。追加する素性の候補として、「学歴」および「SSM 職業コード」の 2 つを選んだ。この理由は、学歴については、ISCO コードにおける分類の基準である技能度の情報をもつ変数を素性として組み込むことが有効であると思われるが、これを直接表す変数が存在しないため、最も近い変数として学歴で代用することを考えたためである。また、SSM 職業コードについては、2.2 節で述べたように、コード体系が異なっているにもかかわらず、職業という同じドメインにおいて付与されるコードには有用な情報が含まれるのではないかと考えたためである。なお、「性別」は SSM 自動職業コーディングにおいて有効な素性ではなかったため（高橋 2007a）、今回の予備実験においては追加する素性の候補とはしなかった。

実験は、データセットとして「2003 年仕事と暮らし調査」の本人現職（全 767 サンプル）を用い、10 分割交差検定を行った。分類器は 4.1 節の実験設定とまったく同様に SVM を one-versus-rest 法により多値分類器に拡張し、カーネル関数としては線形カーネルを用いた。また、 C の値を 0.1 から 1.0 まで変化させた。学歴は、選択肢の 7 段階¹⁷をそのまま用いた場合と 5 段階¹⁸にまとめた場合の 2 通りについて実験した。また、今回用いられた予測 SSM 職業コードは、add-code 法により正解率 54.1% で予測された結果である。

実験の結果を図 4 に示す。図 4 における正解率は、 C を変化させて最もよかった場合の値（最大）である。訓練事例数が非常に少ないためにいずれも正解率の値が低かったが、基本素性のみを用いた場合と素性を追加した場合を比較すると、正解率の向上に有効な素性は「正解 SSM 職業コード」および「予測 SSM 職業コード」であった。学歴は 7 段階のままでは正解率が低下し、5 段階にまとめた場合でも正解率に変化がなかった。SSM 職業コードと学歴の両方を追加する方法も有効ではあるが、SSM 職業コードのみを追加する方がより有効であるために、学歴は SSM 職業コードと組み合わせても追加しない方がよいことがわかった。SSM 職業コードは、正解の場合と予測の場合で正解率が大きく異なり、正解の場合は基本素性のみを用いた場合より約 10% 高かったが、予測の場合は約 3% しか高くなかった。

以上の結果から、本研究では基本素性に追加する素性は SSM 職業コードのみとし、実験 1 では、SSM 職業コードの効果的な追加方法についての検討を行うことにした。

¹⁷ 1: 中学校 2: 高校 3: 短大・高専 4: 専修学校（高卒後） 5: 大学 6: 大学院 7: その他

¹⁸ 1: 中学校 2: 高校 3: 短大・高専または専修学校（高卒後） 4: 大学または大学院 5: その他

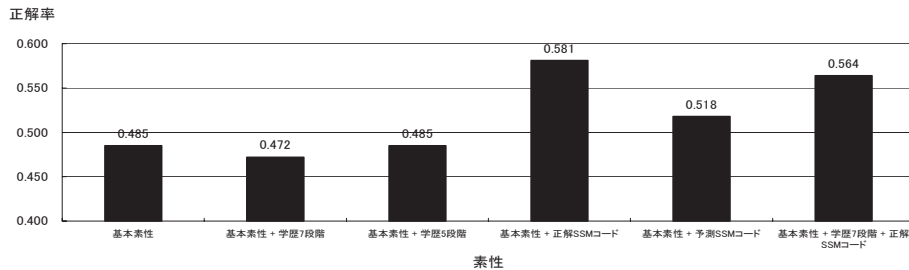


図4 追加した素性別正解率（「2003 年仕事と暮らし調査」データによる）

4.3 実験1（方法の比較）

本研究では人手（一般コード）によるコーディングの結果が残されていないために、提案手法を人手による方法と比較することができない。したがって、実験1では機械学習による方法同士の比較を行った。すなわち、素性にSSMコードを追加せずに基本素性のみを用いる方法をbaselineとし、SSM職業コードを追加した場合と正解率の比較を行った。追加するSSM職業コードとしては、第1位に予測されたもの1個（方法B）、第1位から第5位までに予測されたもの5個（方法C）、正解として決定されているもの1個（方法D）の3通りを考え、baselineを含めた4通りの方法（図5）について実験した。ここで、方法Bや方法Cで用いられた予測SSM職業コードは、add-code法により正解率72.4%（第1位のクラスの場合）¹⁹および、84.6%（第1位から第5位までのクラスの場合）で予測された結果である。

- A) 基本素性のみを用いる方法（baseline）
- B) 基本素性に第1位に予測されたSSM職業コード（1個）を追加する方法
- C) 基本素性に第1位から第5位までに予測されたSSM職業コード（5個）を追加する方法
- D) 基本素性に正解SSM職業コード（1個）を追加する方法

図5 実験1における4つの方法

実験結果

まず、小分類（4桁）のレベルにおける正解率を調査する。表1～表4は、各方法において、「第1位に予測されたクラス」から「第1位から第5位までに予測されたクラス²⁰」までの正解率について、それぞれの最大、最小、平均を示す。表中、カッコ内の数字は最大、最小の場合のCの値を示した。また、各方法とも第1位のクラスにおける最大値を太字で示した。

¹⁹ 訓練事例として、評価事例とやや性質が異なるJGSSデータ（約34,521サンプル）を用いたため、訓練事例数がJGSS-2003より約15,000サンプル多かったにもかかわらず正解率は8.3%低かった（高橋 2007b）。

²⁰ NANACOシステムは、コードのために第1位から第5位までに予測されたクラスを候補として提示する。

表 1 基本素性のみを用いる方法 (baseline) における正解率

	第 1 位のみ	第 2 位まで	第 3 位まで	第 4 位まで	第 5 位まで
最大	0.551 (C=0.6)	0.649 (C=0.5)	0.681 (C=0.5)	0.698 (C=0.5)	0.713 (C=0.6)
最小	0.528 (C=0.1)	0.634 (C=1.0)	0.660 (C=1.0)	0.684 (C=1.0)	0.693 (C=1.0)
平均	0.543	0.641	0.675	0.693	0.707

表 2 第 1 位に予測された SSM 職業コード 1 個を追加する方法 (方法 B) における正解率

	第 1 位のみ	第 2 位まで	第 3 位まで	第 4 位まで	第 5 位まで
最大	0.583 (C=0.5)	0.679 (C=0.5)	0.713 (C=0.6)	0.729 (C=0.5)	0.741 (C=0.6)
最小	0.571 (C=0.1)	0.661 (C=1.0)	0.697 (C=1.0)	0.744 (C=1.0)	0.724 (C=1.0)
平均	0.577	0.674	0.709	0.726	0.735

表 3 第 1 位から第 5 位までに予測された SSM 職業コード 5 個を追加する方法 (方法 C) における正解率

	第 1 位のみ	第 2 位まで	第 3 位まで	第 4 位まで	第 5 位まで
最大	0.546 (C=0.3)	0.648 (C=0.5)	0.686 (C=0.2)	0.706 (C=0.1)	0.720 (C=0.1)
最小	0.532 (C=1.0)	0.629 (C=1.0)	0.670 (C=1.0)	0.691 (C=0.9)	0.699 (C=1.0)
平均	0.542	0.640	0.677	0.698	0.711

表 4 正解 SSM 職業コードを追加する方法 (方法 D) における正解率

	第 1 位のみ	第 2 位まで	第 3 位まで	第 4 位まで	第 5 位まで
最大	0.615 (C=0.2)	0.717 (C=0.2)	0.754 (C=0.5)	0.773 (C=0.2)	0.785 (C=0.1)
最小	0.602 (C=1.0)	0.704 (C=1.0)	0.740 (C=1.0)	0.755 (C=1.0)	0.765 (C=1.0)
平均	0.609	0.713	0.748	0.767	0.779

実験 1 においても訓練事例が少ないために、いずれの方法も正解率が低かったが、SSM 職業コードを素性として追加する方法と基本素性のみを用いる方法を比較した結果、SSM 職業コードを素性として追加する方法の有効性が示された²¹。これは予備実験と同様の結果であった。まず、最も重要な第 1 位に予測されたクラスにおける正解率に注目すると、正解 SSM 職業コードを追加する方法（以下方法 D と略する）が最もよく（約 6% 高い）、次に、予測 SSM 職業コードを 1 個追加する方法（以下方法 B と略する）がよかった（約 3% 高い）。予測 SSM 職業コードを 5 個追加する方法（以下方法 C と略する）はやや悪かった（約 0.5% 低い）。次に、第 1 位から第 5 位までに予測されたクラスにおける正解率も、基本素性のみを用いる方法は約 70% で

²¹ C の値による正解率の違いも、SSM 職業コードを追加した方法の方が小さかった。

あったが、方法Bは約 74%、方法Dは 78%、方法Cも約 72%でいずれも基本素性のみを用いる方法より高い値であった。

ISCO コードは 4 桁であるが、実際には、最後の桁を無視した 3 桁（中分類）までがわかれば有用な場合もある。そこで、正解の基準を 1 桁ずつ緩め、3 桁、2 桁（亜大分類）、1 桁（大分類）のレベルにおける正解率を調査した（ $c=1.0$ の場合）。各方法において第 1 位に予測されたクラスの正解率の変化を示す（図 6）。

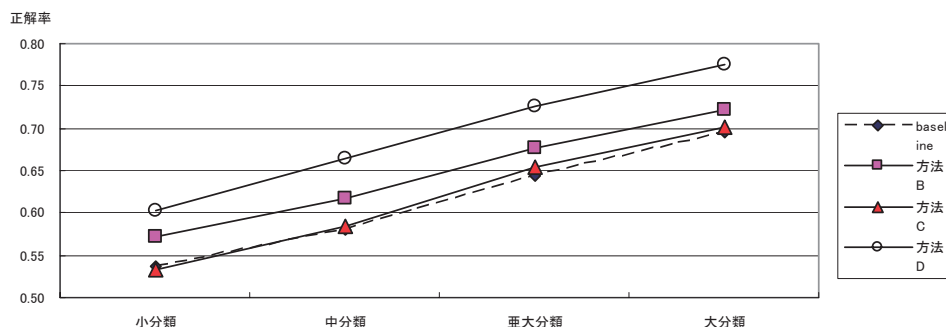


図 6 分類レベルの違いにおける方法別正解率（第 1 位のみ）（ $c=1.0$ の場合）

分類レベルを緩めるにつれて、どの方法においても正解率が上がり、特に方法 D では約 80% 近くになった。上昇の程度は方法間で差がなく、例えば小分類と大分類における正解率の差はいずれの方法においても約 17%であった。

以上より、ISCO コーディングにおいては、どの分類レベルを用いる場合でも提案手法、すなわち SSM 職業コードを素性として追加する方法は有効であった。特に、正解 SSM 職業コードの追加は有効であったため、正解が決定されている場合には正解コードを用いるのがよい。もし SSM 職業コードが決定されていない場合には、第 1 位に予測されたコードを用いてもよい。

最後に、正解率が全般的に低かったことについて、原因の一つとして訓練事例数の少なさが考えられる。単純な比較はできないが、SSM 職業コーディングにおいては訓練事例数の増加に伴う正解率の向上が実験的に示されており（高橋ほか 2000）、実際に訓練事例数が約 6,700 サンプル増加した調査で正解率が約 6.3%向上したこと（高橋 2007b）、また、ISCO 自動コーディングにおいても、予備実験より訓練事例数が約 2,800 サンプル多かった実験 1 で、基本素性のみを用いる方法、方法 B、方法 D においてそれぞれ約 7%、約 3%、約 3% 正解率の向上がみられたことから、今後、正解が決定された事例が増えるにつれて訓練事例数の増大が見込まれるために、正解率の向上が期待できる。もちろん、さらなる改善を検討は必要である。

4.4 実験2（エラー解析）

ここでは、提案手法の中で有効性を示した方法Bと方法Dについて、分類レベル別に出現頻度上位10位までのクラスに注目してエラー解析を行う。

実験結果

まず、最も厳しい基準である小分類レベルについての正解率を調査した。表5に基本素性のみを用いる方法の正解率が高い順に示す（ $C=1.0$ の場合）。表中右2列の「(差)」は、方法Bおよび方法Dと基本素性のみを用いる方法との差（以下差と略する）を示す。アンダーラインのある数値は、それぞれの方法における正解率の平均や差の平均より大きな値であったことを示す（以下表6から表8においても同様である）。小分類レベルにおいて、表5に示した10クラスのカバー率は約35%であった。

表5 小分類における上位10クラスの方法別正解率と差

ISCO コード	出現頻度 (%)	Baseline	方法 B	方法 D	方法 B (差)	方法 D (差)
5141	1.5	<u>0.947</u>	<u>0.965</u>	<u>1.000</u>	0.018	0.053
6111	3.2	<u>0.823</u>	<u>0.858</u>	<u>0.876</u>	0.035	0.053
5122	2.7	<u>0.823</u>	<u>0.865</u>	<u>0.948</u>	<u>0.042</u>	<u>0.125</u>
3415	3.4	<u>0.803</u>	<u>0.773</u>	<u>0.841</u>	-0.030	0.038
4100	8.7	<u>0.795</u>	<u>0.804</u>	<u>0.818</u>	0.009	0.024
5220	6.2	<u>0.794</u>	<u>0.833</u>	<u>0.912</u>	<u>0.040</u>	<u>0.118</u>
4121	3.8	<u>0.791</u>	<u>0.811</u>	<u>0.791</u>	0.020	0.000
9132	1.8	<u>0.731</u>	<u>0.731</u>	<u>0.746</u>	0.000	0.015
8281	1.6	<u>0.561</u>	<u>0.649</u>	<u>0.632</u>	0.088	<u>0.070</u>
4190	1.8	0.403	0.389	0.389	-0.014	-0.014
全クラス平均	—	0.551	0.583	0.615	0.032	0.064

表5に示した10クラスのうち8クラスは正解率が70%を超えた。特に、「5141 Hairdressers, barbers, beauticians and related workers」は基本素性のみを用いる方法で90%を超えたが、方法Bや方法Dによりさらに値が高くなりそれぞれ約97%、100%となった。9位までのクラスにおける正解率は、基本素性のみを用いる方法を含めたどの方法も平均正解率を上回ったことから、出現頻度の高いクラスに対しては機械学習が有効であったことおよび、基本素性のみを用いる方法との差のほとんどがプラスであったことから、SSM職業コードの追加も有効であることがわかった²²。一方で、「4190 Other office clerks」は基本素性のみを用いる方法で正解率50%を下回っていたが、方法Bや方法Dによりいずれもさらに1.4%低下した。「4190」

²² SSM 職業コードの追加が特に有効であったのは、「8281 Mechanical-machinery assemblers」（方法Bにより約8%の上昇）および、「5122 Cooks」と「5220 Shop salespersons and demonstrators」（方法Dによりそれぞれ約13%と12%の上昇）であった。

に対しては、基本素性のみを用いる方法では 22 種類、方法Dでも 12 種類のSSM職業コードが対応した。それぞれ 2 種類のコード（「554」「558」）と 3 種類のコード（「558」「554」「557」）の頻度が特に高かった。比較のために、方法Dが有効であった「8281」を調査すると 7 種類のSSM職業コードしか対応せず、特に頻度の高いコードも 1 種類（「633」）しかなかった。したがって、「4190」は事例数が少ない中で素性が分散したため、分類器の学習がうまくいかなかったものと思われる。他にSSM職業コードの追加により正解率が低下したクラスとして「3415 Technical and commercial sales representatives」があったが、この場合は、方法Bは約 3%低下したが、方法Dは約 4%上昇した。小分類レベルにおいては、調査した事例のカバー率が低かったため、ここでの結果を出現頻度の高くないクラスに対して一般化することはできず、さらなる調査が必要である。

次に中分類レベルについての正解率を示す（表 6）。中分類レベルにおいて表 6 に示した 10 クラスのカバー率は約 41%であった。

表 6 中分類における上位 10 クラスの方法別正解率と差

ISCO コード	出現頻度 (%)	baseline	方法 B	方法 D	方法 B (差)	方法 D (差)
611	4.8	<u>0.906</u>	<u>0.930</u>	<u>0.965</u>	0.023	0.059
233	1.9	<u>0.873</u>	<u>0.873</u>	<u>0.911</u>	0.000	0.038
512	4.3	<u>0.833</u>	<u>0.853</u>	<u>0.968</u>	0.019	<u>0.135</u>
522	6.2	<u>0.794</u>	<u>0.833</u>	<u>0.912</u>	<u>0.040</u>	<u>0.118</u>
410	8.7	<u>0.792</u>	<u>0.804</u>	<u>0.818</u>	0.012	0.026
341	4.5	<u>0.729</u>	<u>0.700</u>	<u>0.777</u>	-0.029	0.047
412	4.2	<u>0.728</u>	<u>0.772</u>	<u>0.747</u>	<u>0.043</u>	0.019
832	3.0	<u>0.726</u>	<u>0.853</u>	<u>0.902</u>	<u>0.128</u>	<u>0.177</u>
828	3.6	<u>0.617</u>	<u>0.680</u>	<u>0.688</u>	<u>0.063</u>	0.070
311	2.1	0.182	0.208	0.286	0.026	<u>0.104</u>
全クラス平均	—	0.580	0.617	0.665	0.037	0.085

中分類レベルにおいても小分類レベルと同様に、出現頻度の高いクラスでは機械学習が有効であり、SSM職業コードの追加も有効であった。特に方法Dはすべてのクラスで有効であった。唯一、「311 Physical and engineering science technicians」は際だって正解率が悪く 20%を下回っており、方法Bや方法Dによっても 30%に達しなかった。「311」における誤り方を調査すると、基本素性のみを用いる方法も方法Bも 24 種類の誤ったISCOコードを付与していた²³。SSM職業コードの追加が有効ではなかったのは、「341 finance and sales associate professionals」における方法Bのみで、約 3%の低下であった。

²³ 特に、基本素性のみを用いる方法では、正解（「311」）より「214 Architects, engineers and related professionals」が付与された事例の方が多かった。

次に亜大分類レベルについての正解率を示す（表 7）。亜大分類において表 7 に示した 10 クラスのカバー率は約 68%で過半数を超えた。

表 7 亜大分類における上位 10 クラスの方法別正解率と差

ISCO コード	出現頻度 (%)	baseline	方法 B	方法 D	方法 B (差)	方法 D (差)
61	5.7	<u>0.872</u>	<u>0.918</u>	<u>0.964</u>	0.046	0.092
23	3.3	<u>0.870</u>	<u>0.878</u>	<u>0.906</u>	0.008	0.038
41	17.3	<u>0.807</u>	<u>0.815</u>	<u>0.862</u>	0.009	0.056
52	6.8	<u>0.802</u>	<u>0.834</u>	<u>0.919</u>	0.032	0.117
51	8.2	<u>0.763</u>	<u>0.811</u>	<u>0.921</u>	0.048	0.158
71	4.0	<u>0.690</u>	<u>0.703</u>	<u>0.736</u>	0.013	0.045
91	5.5	<u>0.649</u>	0.644	0.693	-0.005	0.045
82	7.9	0.627	0.631	0.646	0.004	0.019
34	6.1	0.563	0.567	0.635	0.004	0.071
72	3.7	0.546	0.575	0.603	0.028	0.057
全クラス平均	—	0.643	0.677	0.727	0.034	0.084

亜大分類レベルでは、正解率が特に低いクラスはなく、特に、「61 Skilled agricultural and fishery workers」「51 Personal and protective service workers」「52 Models, salespersons and demonstrators」は、基本素性のみを用いる方法も有効で 80%を上回っており、SSM 職業コードの追加はさらに有効であった。特に方法 D はすべてのクラスで有効で、方法 B も「91 Sales and services elementary occupations」を除くすべてのクラスで有効であった。

最後に大分類レベルについての正解率を示す（表 8）。大分類レベルにおけるクラスは図 7 に示すように全部で 10 個であるため、上位 10 クラスのカバー率が 100%となる。

- | | |
|--|---|
| 1: LEGISLATORS, SENIOR OFFICIALS AND MANAGERS | 2: PROFESSIONALS |
| 3: TECHNICIANS AND ASSOCIATE PROFESSIONALS | 4: CLERKS |
| 5: SERVICE WORKERS AND SHOP AND MARKET SALES WORKERS | 6: SKILLED AGRICULTURAL AND FISHERY WORKERS |
| 7: CRAFT AND RELATED TRADES WORKERS | 8: PLANT AND MACHINE OPERATORS AND ASSEMBLERS |
| 9: ELEMENTARY OCCUPATIONS | 0: ARMED FORCES |

図 7 大分類コードと分類名

表 8 に示したように、大分類レベルでは、「6」「4」「5」のクラスは機械学習が特に有効で、SSM 職業コードの追加も有効であった。この 3 クラスの出現頻度は、それぞれ第 8 位、第 1

位、第2位であり、必ずしも高いものばかりではなかった。SSM 職業コードの追加が有効なクラスは、この他には「9」「2」「3」であった。ただし、このクラスはいずれも基本素性のみを用いる方法による正解率が70%を下回っており、「9」「2」は方法Dにより70%を超えたが、方法Bでは60%台にとどまった。「3」は出現頻度が第4位であったにもかかわらず、基本素性のみを用いる方法による正解率が下位であったことから、分類が困難なクラスであったものと思われる。方法Dは大分類レベルにおいても他と同様にすべてのクラスで有効であったが、方法Bは正解率を「1」で約2%、「7」で約0.2%低下させた。

表8 大分類における上位10クラスの方法別正解率と差

ISCO コード	出現頻度 (%)	baseline	方法 B	方法 D	方法 B (差)	方法 D (差)
6	5.7	<u>0.873</u>	<u>0.919</u>	<u>0.965</u>	<u>0.046</u>	<u>0.091</u>
4	20.2	<u>0.823</u>	<u>0.830</u>	<u>0.882</u>	0.008	0.059
5	15.0	<u>0.797</u>	<u>0.835</u>	<u>0.926</u>	<u>0.037</u>	<u>0.128</u>
8	12.8	0.686	0.711	0.727	0.025	0.041
9	9.0	0.636	0.666	0.728	<u>0.030</u>	<u>0.093</u>
7	12.1	0.636	0.634	0.707	-0.002	0.072
2	8.9	0.629	0.686	0.734	<u>0.057</u>	<u>0.106</u>
3	12.9	0.534	0.585	0.606	<u>0.051</u>	0.072
1	3.4	0.437	0.413	0.508	-0.024	0.071
0	0.0	0.000	0.000	0.000	0.000	0.000
全クラス平均	-	0.696	0.722	0.776	0.026	0.080

大分類レベルにおける誤りは重大であるため、各方法における正誤状況をさらに詳細に調査した。表9は、行に示したクラスに入る事例、が列に示したクラスにどのような割合で分配されたかを示す。すなわち、表の対角部分（斜体）のセルは正解率、それ以外のセルは誤分類の割合を表す。各セルとも、上段は基本素性のみを用いる方法における割合、2段目、3段目はそれぞれ方法B、方法Dにおける割合と基本素性のみを用いる方法の割合との差を示す。したがって、対角部分のセルでは「差」がプラスで大きいほどよく、それ以外のセルでは「差」がマイナスで小さいほどよい。例えば、対角部分である「1」行「1」列のセルは、上段から順に、大分類「1」の基本素性のみを用いる方法による正解率が0.437、方法Bにおいては差が-0.024（方法Bの正解率0.413）、同様に方法Dにおいては差が0.071（方法Dの正解率0.508）であったことを表す²⁴。表9より、全般的な正誤状況としては、「6」が「9」へ誤分類される場合が特に多いことを除き、近隣のクラスへ誤りやすい傾向が見られた。また、表右端列は、各クラスごとに誤分類に対してSSM職業コードの追加が有効であったクラスの数を示す。自分

²⁴ 対角部分の数値は、上段から順にそれぞれ表8における左から3列目と右2列の値に等しい。

以外のクラスは9個であるため、この数が多いほどよく、最低でも5以上が望ましい。この点からは、方法Dはすべてのクラスにおいて有効性が高かったが、方法Bは「2」「3」「5」「8」「9」においては有効性が高かったが、「1」「4」「6」「7」では低かった。

表9 大分類における正誤状況とSSM職業コードの追加による効果（単位：％）

	1	2	3	4	5	6	7	8	9	個
1	<i>0.437</i>	0.079	0.119	0.143	0.095	0.008	0.071	0.016	0.032	3
	<i>-0.024</i>	0.008	0.040	-0.016	0.008	0.000	-0.016	-0.008	0.008	
	<i>0.071</i>	-0.024	0.024	0.000	-0.008	-0.008	-0.032	-0.016	-0.008	
2	0.006	<i>0.629</i>	0.214	0.060	0.020	0.000	0.026	0.031	0.014	6
	0.009	<i>0.057</i>	-0.009	-0.009	-0.011	0.000	-0.014	-0.014	-0.009	
	0.011	<i>0.106</i>	-0.006	-0.046	-0.017	0.000	-0.020	-0.017	-0.011	
3	0.012	0.144	<i>0.534</i>	0.084	0.115	0.004	0.047	0.041	0.018	5
	0.012	-0.016	<i>0.053</i>	-0.012	-0.029	0.000	-0.010	0.008	-0.006	
	0.008	0.027	<i>0.072</i>	-0.014	-0.043	-0.004	-0.018	-0.023	-0.004	
4	0.005	0.030	0.047	<i>0.823</i>	0.029	0.006	0.019	0.024	0.018	3
	0.003	-0.015	0.003	<i>0.008</i>	-0.001	0.000	0.000	-0.003	0.006	
	0.008	-0.019	-0.002	<i>0.060</i>	-0.013	-0.005	-0.013	-0.014	-0.002	
5	0.013	0.017	0.050	0.032	<i>0.797</i>	0.013	0.030	0.022	0.026	6
	0.000	-0.006	-0.004	0.004	<i>0.037</i>	-0.004	-0.013	-0.006	-0.009	
	-0.006	-0.013	-0.032	-0.002	<i>0.128</i>	-0.007	-0.030	-0.019	-0.020	
6	0.005	0.005	0.005	0.010	0.015	<i>0.873</i>	0.010	0.000	0.076	4
	-0.005	-0.005	0.005	0.005	0.000	<i>0.046</i>	-0.010	0.000	-0.036	
	-0.005	-0.005	0.000	-0.005	-0.015	<i>0.091</i>	-0.010	0.000	-0.051	
7	0.017	0.004	0.053	0.036	0.051	0.008	<i>0.636</i>	0.150	0.044	3
	-0.004	0.008	0.004	-0.011	0.000	0.000	<i>-0.002</i>	0.002	0.002	
	-0.008	0.003	-0.017	-0.027	-0.024	-0.004	<i>0.120</i>	-0.046	0.003	
8	0.014	0.007	0.021	0.041	0.023	0.009	0.103	<i>0.686</i>	0.098	5
	-0.009	0.009	-0.005	-0.018	0.002	-0.005	0.009	<i>0.025</i>	-0.009	
	-0.007	-0.002	0.005	-0.023	-0.009	-0.005	0.036	<i>0.041</i>	-0.036	
9	0.018	0.006	0.036	0.054	0.081	0.027	0.030	0.111	<i>0.638</i>	6
	-0.003	-0.003	-0.003	0.003	0.000	-0.003	-0.006	-0.015	<i>0.030</i>	
	-0.012	-0.003	0.000	-0.009	-0.021	-0.009	-0.003	-0.036	<i>0.093</i>	

ここで、大分類における誤り表9における基本素性のみを用いる方法に基づいて、「誤りやすいクラス」を誤りやすさの程度により整理した（表10）。ここで、誤りやすいクラスとは、誤分類の程度が5%より大きいクラスである。例えば、本来「1」に分類されるべき事例

が誤りやすいクラスは、「2」「3」「4」「5」「7」である。表中、誤りやすいクラスの後の（ ）内は、SSM 職業コードを追加する方法（太字）が、有効であるか（「0」）、有効でないか（「X」）を示す。例えば、「2（B0 DX）」は、クラス「2」への誤分類を減らすために、方法 B は有効であったが方法 D は有効でなかったことを表す。表 10 より、最も誤りやすい場合は、誤りやすさの程度からは、「2」が「3」に、「1」が「3」「4」に、「7」が「8」に、「8」が「7」に、「9」が「8」にそれぞれ誤分類される場合であった。誤りやすいクラスの個数からは、「1」が 5 クラスで最も多かった。クラス間の関係からは、「1」「2」「3」間で誤りやすかったが、これらは SSM 職業コードにおける大分類では「専門・技術的職業従事者」や「管理的職業従事者」にほぼ該当する。一方で、「4」「5」や「6」は誤りやすいクラスがまったくないか 1 つだけであったが、これらは SSM 職業コードにおける大分類では「事務従事者」「販売従事者」「保安職業従事者」「農林漁業従事者」にほぼ該当する。最後に、大分類を誤りやすい場合に SSM 職業コードの追加が有効でなかったのは、「8」が「7」に誤分類される場合のみで、他は方法 B と方法 D の少なくともどちらかが有効であった。

表 10 基本素性のみ用いる方法において誤りやすいクラスと方法別の効果

ISCO コード	20%以上	10%以上 20%未満	5%以上 10%未満
1	-	3 (B0 DX) 4 (B0 DX)	2 (B0 DX) 5 (BX D0) 7 (B0 D0)
2	3 (B0 D0)	-	6 (B0 D0)
3	-	2 (B0 DX) 5 (B0 D0)	4 (B0 D0)
4	-	-	-
5	-	-	-
6	-	-	9 (B0 D0)
7	-	8 (BX D0)	3 (B0 D0) 5 (B0 D0)
8	-	7 (BX DX)	9 (B0 D0)
9	-	8 (B0 D0)	4 (B0 D0) 5 (BX D0)

5. おわりに

本研究では、ISCO 自動コーディングの方法として、SSM 職業コードを素性とする機械学習（サポートベクターマシン）による方法を提案した。このとき、SSM 職業コードとしては、すでに決定されたものがあればそのコード、ない場合は予測されたコードを用いる。提案手法は、「2005 年 SSM 日本調査」で収集された本人現職（有職者）データを用いた実験により、有効性が示された。ただし、今回は人手による方法との比較ができなかったために、今後、実験を行って確認する必要がある。また、提案手法は、誤分類が最も重大な大分類レベルに

において、クラスによっては分類精度が不十分であった。今後の課題として、正解率のさらなる向上を目指して検討を行う必要がある。2010 年米国センサスでは、モバイル技術が採用される予定である。今後、社会状況の変化に応じて、調査方法だけでなく用いられる職業コード体系も変化する可能性があるが、提案手法は、現在付与されているコードを有用な情報として容易に利用することができる。

【文献】

- 1995 年 SSM 調査研究会. 1995. 『SSM 産業分類・職業分類 (95 年版)』. 1995 年 SSM 調査研究会.
- 1995 年 SSM 調査研究会. 1996. 『1995 年 SSM 調査コード・ブック』. 1995 年 SSM 調査研究会.
- 2005 年 SSM 調査研究会職業分類タスクグループ. 2004. 『職業分類における SSM95 と ISCO-88 の対応』.
- Bureau of Statistics; International Labour Office. 2001. *Coding Occupation and Industry*.
Bureau of Statistics; International Labour Office.
- Creecy, R. H. , Mas, B. M. Smith, S. J. and D. L. Waltz. 1992. “Trading Mips and Memory for Knowledge Engineering.” *Communication of the ACM* 35(8) : 48-63.
- Dumais, S. , Platt, J. , Hecherman, D. and M. Sahami. 1998. “Inductive Learning Algorithms and Representations for Text Categorization.” *Proceedings of the ACM-CIKM98* : 145-155.
- Gillman, D. W. and M. V. Appel. 1999. “Developing an Automated Industry and Occupation Coding System for CENSUS 2000.” *2000 Proceeding of the American Statistical Association Annual Meeting*, Government Statistics Section.
- 原純輔・海野道郎. 1984. 『社会調査演習』. 東京大学出版会.
- 石田浩・三輪哲. 2006. 『2005 年 SSM 調査 SSM95 職業・産業コーディングマニュアル改訂版』.
東京大学社会科学研究所附属日本社会研究情報センター.
- Joachims, T. 1998. “Text Categorization with Support Vector Machines: Learning with Many Relevant Features.” *Proceedings of the European Conference on Machine Learning* : 137-142.
- Keogh, G. 1998. “Automatically Coding Occupation Description from the 1996 Census of Population of Ireland.” *Technical report in Central Statistic Office(CSO)*.
- 国立国語研究所. 1964. 『分類語彙表』. 秀英出版社.
- Kunz, C. 2003. *CENSUS: OCCUPATION (Census Paper No. 03/06)*. Australian Bureau of Statistics.
- 黒橋禎夫・長尾眞. 1998 『日本語形態素解析システム JUMAN version 3.61』. 京都大学大学情報研究科.
- Kressel, U. 1999. Pairwise classification and Support Vector Machines. Scholkopf, B. ,

- Burgesa, C. J. C. and A. J. Smola (Eds.) *Advances in Kernel Methods Support Vector Learning*. The MIT Press : 255-268.
- 西村幸満・石田浩. 2001. 『SSJ Data Archive Research Paper Series JGSS-2000 調査 職業・産業コーディングインストラクション』. 東京大学社会科学研究所附属日本社会研究情報センター.
- 西澤弘. 2006. 『労働政策研究報告書 No.57 職業紹介における職業分類のあり方を考える「労働省編職業分類」の改訂に向けた論点整理一』. 独立行政法人労働政策研究・研究機構.
- 岡本政人. 2004. 「国内外における統計自動格付法の研究動向」『製表技術参考資料2』独立行政法人統計センター : 46-77.
- Riviere, P. 1997. “SICORE – general automatic coding system.” *Statistical Data Editing Vol.2 United Nations* : 222-231.
- Rubin, D. B. 2004. *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons, Hoboken, New Jersey.
- Sebastiani, F. 2002. “Machine Learning Automated Text Categorization.” *ACM Computing Surveys* 34(1) : 1-47.
- 盛山和夫. 2004. 『社会調査演習』. 東京大学出版会.
- 平博順・春野雅彦. 2000. 「Support Vector Machine によるテキスト分類における属性選択」『情報処理』41 (4) : 1113-1123.
- Takahashi, K. , Takamura, H. and M. Okumura. 2005. “Automatic Occupation Coding with Combination of Machine Learning and Hand-Crafted Rules.” Bao, H. T. , David, C. and L. Huan (Eds.) *Advances in Knowledge Discovery and Data Mining Proceedings Series: Lecture Notes in Computer Science Subseries: Lecture Notes in Artificial Intelligence* Vol.3518. Springer-Verlag, Berlin Heidelberg : 269-279.
- 高橋和子. 2000. 「自由回答のコーディング支援について—格フレームによる SSM 職業コーディングシステム—」『理論と方法』15 (1) : 149-164.
- 高橋和子. 2001. 「自由回答のコーディング自動化システム—「健康と階層」調査における職業コーディング—」(文部省科研費(基礎研究A(2) 福祉社会の価値観に関する実証的研究(研究代表 武川正吾) 研究成果)『敬愛大学国際研究』8 (1) : 31-52.
- 高橋和子. 2002. 「JGSS-2000 における職業・産業コーディング自動化システムの適用」『日本版 General Social Surveys 研究論文集 JGSS で見た日本人の意識と行動東京大学社会科学研究所資料第20集』: 171-183.
- 高橋和子. 2003. 「JGSS-2001 における職業・産業コーディング自動化システムの適用」『日本版 General Social Surveys 研究論文集 [2] JGSS で見た日本人の意識と行動東京大学社会科学研究所資料第22集』: 1791-191.

- 高橋和子. 2004. 「職業コーディングにおける ROCCO システムと SVM の組み合わせ」『日本版 General Social Surveys 研究論文集 [3] JGSS で見た日本人の意識と行動東京大学社会科学研究所資料第 24 集』: 163-174.
- 高橋和子・高村大也・奥村学. 2004. 「ルールベース手法と機械学習による自由回答の分類—職業コーディングの自動化の方法—」『理論と方法』15 (1) : 177-196.
- 高橋和子・高村大也・奥村学. 2005a. 「機械学習とルールベース手法の組み合わせによる自動職業コーディング」『自然言語処理』12 (2) : 3-24.
- 高橋和子・須山敦・村山紀文・高村大也・奥村学. 2005b. 「職業コーディング支援システム (NANACO) の開発と JGSS-2003 における適用」『日本版 General Social Surveys 研究論文集[4] JGSS で見た日本人の意識と行動 <JGSS Research Series No.1>』: 225-242.
- 高橋和子. 2007a. 「機械学習による ISCO 自動コーディングにおける素性選択」2005 年 SSM 調査職業・産業班研究会 (第 4 回) 於: 東大駒場キャンパス.
- 高橋和子. 2007b. 「機械学習を用いた自由回答分類の高度化に関する研究」東京工業大学大学院博士学位論文.
- 徳永健伸. 1999. 『言語と計算 5 情報検索と言語処理』. 東京大学出版会.
- Vapnik, V. 1998. *Statistical Learning Theory*. John Wiley, New York.
- Wolpert, D. H. 1992. “Stacked Generalization.” *Neural Networks* 5 : 241-259.

Automatic ISCO-88 Coding with Machine Learning

Kazuko TAKAHASHI

Keiai University

In recent occupation coding, coders are asked to assign an example to ISCO(International Standard Classification of Occupation)-88 in addition to SSM (Social Stratification and Social Mobility) -95 codes. However, this double-coding needs time consuming and laborious efforts. Therefore, we also propose an automatic coding method with machine learning for ISCO-88 because SSM automatic coding system, which we have already developed, effectively supports coders. In the proposed method, we use either type of SSM codes as features in learning. One is a correct code, and the other is a predicted code. We empirically show that using correct SSM codes is the most effective, and that using predicted codes is more effective.

Keywords and phrases: automatic coding, ISCO-88, SSM occupation coding system, machine learning, SVMs, NANACO system, multiclass classification

SSM 職業分類と国際的階層指標： EGP 階級分類・SIOPS・ISEI への変換

鹿又伸夫・田辺俊介・竹ノ下弘久
(慶應義塾大学・東京大学・静岡大学)

【要旨】

2005 年 SSM 日本調査のデータには、職業（仕事の内容）について、SSM 職業分類のコードとともに ISCO-88 のコードが記録されているため、後者をもちいて国際的に利用頻度の高い EGP 階級分類、SIOPS そして ISEI などを作成できる。ところが 1995 年までの SSM 調査データには ISCO コードが含まれていないので、これらの分類や指標を容易に作成できない。そこで、国際的に利用頻度の高い階層分類と階層指標を利用可能にするため、SSM 職業小分類と ISCO-88 との対応関係を提示する。この対応関係から直接に SIOPS と ISEI を作成できる。しかし、EGP 階級分類は ISCO コードだけでは作成できないので、Ganzeboom and Treiman (1996)の手続きに準拠しながら、従業上地位、職位（役職）、企業規模をもちいた変換方法を提示する。ただし、第 1 回 1955 年調査では職位情報が欠落しているため、ここでの EGP 階級分類への変換方法は第 2 回 1965 年調査以降についてしか適用できない。

キーワード：SSM 職業分類、ISCO-88、EGP 階級分類、SIOPS、ISEI

1. はじめに

日本の「社会階層と社会移動全国調査 (SSM 調査)」は、1955 年から 10 年間隔で実施されてきており、国際的にみても早い段階から活発に研究され多くの成果が蓄積されてきた。しかし、日本を含む国際比較研究という点では、それほど活発でなかったといわざるをえない。それは、日本独自の分類方式である SSM 職業分類を、各国で使用されている職業分類と対応させて、比較可能な階級・階層分類あるいは階層的地位指標に変換する、という煩雑な作業が必要だったことにも一因があると思われる。

第 6 回 2005 年 SSM 日本調査のデータには、職業（仕事の内容）について、SSM 職業小分類のコードとともに、国際標準職業分類 ISCO-88 (International Standard Classification of Occupations 1988 年改訂版：以下では ISCO と略記する) のコードが記録されている。そのため 2005 年調査については ISCO を利用して、国際的に利用頻度の高い EGP 階級分類 (EGP class schema)、SIOPS (Standard International Occupational Prestige Scale)そして ISEI (International Socio-Economic Index of Occupational Status)などに変換できる。Ganzeboom and Treiman (1996) は、ISCO-88 と、EGP 階級分類、SIOPS そして ISEI との対応関係をすでに明示しているので、これを利用して上記の分類や指標を作成できるのである。

ところが 1995 年以前の調査データには ISCO コードが含まれていないので、1955 年から 1995 年までの 5 回分の調査データをもちいた国際比較研究は容易ではない。そこで本稿では、国際的に利用頻度の高い階層分類と階層指標を利用可能にするため、SSM 職業小分類と ISCO-88 との対応関係を提示する。この対応関係から直接に SOPS と ISEI を作成できる。しかし、EGP 階級分類は ISCO コードとの対応関係をつけることだけでは作成できないので、Ganzeboom and Treiman (1996)の手続きに準拠しながら、従業上地位、職位（役職）、企業規模をもちいた変換方法を提示する。ただし、1955 年調査では職位情報が欠落しているため、ここでの EGP 階級分類への変換方法は 1965 年調査以降についてしか適用できない。

2. EGP 階級分類と国際比較研究

社会階層と移動の分野の中でも、中心的なテーマとされてきたのは、世代間移動の研究である。そしてその国際比較研究では、親から子どもへの世代間移動における機会格差が諸国間で同様なパターンを示すのか、世代間地位継承の程度に諸国間で相違があるのか、などが検討されてきた。こうした研究では、職業階層分類または階級分類を作成して、それらにもとづいて世代間移動を分析する。そして国際的に近年の移動研究をみれば、EGP 階級分類を使用する研究が多くなっている。この分類は、Erikson, Goldthorpe and Portocarero (1979)や Erikson and Goldthorpe (1992)の階級図式にしたがって作成されるので、一般的に EGP 階級分類または EG 階級分類と呼ばれる。

EGP 分類は、1970 年代にヨーロッパ 7 カ国で実施された調査プロジェクト CASMIN (Comparative Analysis of Social Mobility in Industrial Nations)によって、国際的に標準的な階級分類の 1 つとして頻繁に利用されるようになった。だが他方では、各国で異なる職業分類をどのようにもちいて EGP 分類に変換・作成するかについて、その詳細な手続きが明示化されていないとも指摘されている(Ganzeboom and Treiman 1996: 204)。

この点を解消するために、Ganzeboom and Treiman (1996)は、ISCO-88 をもちいて EGP 分類を作成する変換手続きを明示化した。各国調査の職業データが独自の分類を使用しているも、その分類を ISCO コードに変換し、さらにかねらの変換手続きにしたがえば、同等の EGP 分類を作成できる。その変換は、第 1 段階での ISCO unit（小分類）コードから EGP 分類への暫定的割り当て、第 2 段階での補正処理から構成される。ただし、第 1 段階は上記論文に明示されているが、第 2 段階の補正処理は SPSS シンタックスとして記載されているだけで、その処理の論理について説明されていない。

2005 年 SSM 調査データには、SSM 職業小分類とともに ISCO コードが記録されている。このため、ISCO を利用して EGP 分類を作成すれば、ISCO をもちいた他の国の調査データと一貫し

た分類を作成できる。しかし、1995 年以前のSSM調査データにはISCOコードは含まれていない。このため 1995 年以前のデータを活用するには、まず第 1 段階の処理をおこなうためにSSM職業小分類とISCOコードとの対応関係をしめさなければならない。また第 2 段階の補正処理をおこなうためには、SSM調査における従業上地位、職位、企業規模の適用方法をしめさなければならない¹。

以降では、SSM 職業小分類から直接に EGP 階級分類を作成する方法を提示する。ここで直接に作成するとは、SSM 職業小分類に ISCO コードとの対応関係をもたせながらも、変換作業ではその ISCO コードを使用せずに SSM 職業コードから EGP 分類を作成する、という意味である。次の 3 節では、まず SSM 職業小分類と ISCO コードとの対応関係をしめす。この対応関係によって、SSM 職業小分類から EGP 分類（変換の第 1 段階）、SIOPS、ISEI への割り当てが可能となる。4 節では Ganzeboom and Treiman (1996)による第 2 段階の補正処理の論理を整理し、5 節では SSM 職業小分類から EGP 分類への変換方法をしめす。

3. SSM 職業小分類と ISCO コードの対応関係

表 1 は、SSM 職業小分類と ISCO の対応関係をしめしている。「SSM95」の欄は 1995 年版の SSM 職業小分類のコード番号で、「SSM75」の欄は前者に対応する 1975 年版の SSM 職業小分類である。これらの小分類は、1995 年版で 188、1975 年版で 288 の職業項目から構成されている。

「ISCO」と「ISCO Title」はSSM小分類コードに対応するISCO unit（小分類）コードとその職業名が記されている。「EGP9」はEGP分類を作成するための第 1 段階の暫定的割り当てであり、この分類を完成するには、次節以降で説明する第 2 段階の補正処理が必要である。ここでわれわれがおこなったのは、SSM分類にたいするISCO分類の割り当てである。そのISCOコードに対応する「EGP9」「SIOPS」「ISEI」の割り当ては、Ganzeboom and Treiman (1996)が提示したものをほぼそのまま採用している²。

SSM項目（コード）にたいするISCO項目（コード）の割り当てにおいては、これらの項目indexおよび説明を参照して、なるべく近いものを対応させた。ISCO小分類は 4 桁で表示される 390 項目なので³、SSM小分類の項目数よりも多い。そのため 1 つのSSM職業項目に複数のISCO項目が該当することもある。表の冒頭には、そうした例として 1995 年版SSM

¹ 本稿で提示する SSM 職業小分類から作成する場合と同じく、2005 年調査データの ISCO コードから作成する場合も、第 2 段階では従業上地位、職位、企業規模の各情報の適用方法を作成する必要がある。

² 農林漁業従事者についての EGP コードを 11 としたことをのぞけば、Ganzeboom and Treiman (1996)が ISCO コードに割りふった SIOPS、ISEI、EGP のコードを採用した。ただし、かれらが ISCO コードに割りあてた EGP コードは、Ganzeboom が <http://homes.fsw.vu.nl/%7Eganzeboom/pisa/index.htm> で公開しているプログラムと一部異なっている。本稿では、上記論文の ISCO コードと EGP コードの対応表にもとづいた。

³ 実際には、各国独自の職業コードを加えて使用することもあるので 390 以上になることがある。

表 1 SSM 職業小分類・ISCO-88・EGP 階級 9 分類・SIOPS・ISEI の対応関係

SSM 95	SSM95 職業小分類	SSM 75	SSM75 職業小分類	EGP9	ISCO	ISCO Title (n.e.c. not elsewhere classified, n.f.s.: not further specified)	SIOPS	ISEI
501*	自然科学系研究者	1	自然科学系研究者	1	2100	Physical, mathematical and engineering science professionals	63	69
501*	自然科学系研究者	1	自然科学系研究者	1	2120	Mathematicians, statisticians and related professionals	56	71
501*	自然科学系研究者	1	自然科学系研究者	1	2200	Life science and health professionals	70	80
502	人文科学系研究者	2	人文科学系研究者	1	2440	Social science and related professionals	58	65
503	機械・電気・化学技術者							
				1	2140	Architects, engineers and related professionals	63	73
		4	金属精錬技術者	1	2147	Mining engineers, metallurgists and related professionals	61	67
		5	機械技術者	1	2145	Mechanical engineers	66	67
		6	電気技術者	1	2143	Electrical engineers	65	68
		7	化学技術者	1	2146	Chemical engineers	66	71
504	建築・土木技術者			1	2140	Architects, engineers and related professionals	63	73
		8	建築技術者	1	2141	Architects, town and traffic planners	72	69
		9	土木技術者	1	2142	Civil engineers	70	69
505	農林技術者	10	農林技術者	2	3212	Agronomy and forestry technicians	47	50
506	情報処理技術者	11	情報処理技術者	1	2130	Computing professionals	51	71
507	その他の技師・技術者							
		3	鉱山技術者	1	2149	Architects, engineers and related professionals (n.e.c.)	56	69
		12	その他の技術者	1	2147	Mining engineers, metallurgists and related professionals	61	67
				1	2149	Architects, engineers and related professionals (n.e.c.)	56	69
508	医師	13	医師	1	2221	Medical doctors	78	88
509	歯科医師	14	歯科医師	1	2222	Dentists	70	85
510	薬剤師	15	薬剤師	1	2224	Pharmacists	64	74
511	助産婦	16	助産婦	2	2230	Nursing and midwifery professionals	54	43
512	保健婦	17	保健婦	2	3229	Modern health associate professionals (except nursing) (n.e.c.)	45	51
513	栄養士	18	栄養士	2	3223	Dieticians and nutritionists	52	51
514	看護婦、看護士	19	看護婦、看護士	2	2230	Nursing and midwifery professionals	54	43
515	あん摩・はり・きゅう師、柔道整復師	20	あん摩・はり・きゅう師、柔道整復師	2	3241	Traditional medicine practitioners	29	51
516	その他の保健・医療従事者	21	その他の保健・医療従事者	2	3229	Modern health associate professionals (except nursing) (n.e.c.)	45	51
517	裁判官、検察官、弁護士	22	裁判官、検察官、弁護士	1	2420	Legal professionals	73	85
518	その他の法務従事者	23	その他の法務従事者	1	2429	Legal professionals (n.e.c.)	71	82
519	公認会計士、税理士	24	公認会計士、税理士	1	2411	Accountants	62	69

520	幼稚園教員	25	幼稚園教員	2	2332	Pre-primary education teaching professionals	49	43
521	小学校教員	26	小学校教員	2	2331	Primary education teaching professionals	57	66
522	中学校教員	27	中学校教員	2	2320	Secondary education teaching professionals	60	69
523	高等学校教員	28	高等学校教員	2	2320	Secondary education teaching professionals	60	69
524	大学教員	29	大学教員	1	2310	College, university and higher education teaching professionals	78	77
525	盲・ろう・養護学校教員	30	盲・ろう・養護学校教員	2	2340	Special education teaching professionals	62	66
526	その他の教員	31	その他の教員	2	2300	Teaching professionals (n.f.s.)	61	69
527	宗教家	32	宗教家	2	2460	Religious professionals	60	53
528	文芸家、著述家	33	文芸家、著述家	2	2451	Authors, journalists and other writers	58	65
529	記者、編集者	34	記者、編集者	2	2451	Authors, journalists and other writers	58	65
530	彫刻家、画家、工芸美術家	35	彫刻家、画家、工芸美術家	2	2452	Sculptors, painters and related artists	57	54
531	デザイナー	36	デザイナー	2	3471	Decorators and commercial designers	49	53
532	写真師、カメラマン	37	写真師、カメラマン	2	3131	Photographers and image and sound recording equipment operators	46	48
533	音楽家(個人に教授するものを除く)	38	音楽家(個人に教授するものを除く)	2	2453	Composers, musicians and singers	45	64
534	俳優、舞踏家、演芸家(個人に教授するものを除く)】	39	俳優、舞踏家、演芸家(個人に教授するものを除く)	2	2450	Writers and creative or performing artists	57	61
535	職業スポーツ家(個人に教授するものを除く)	40	職業スポーツ家(個人に教授するものを除く)	2	3475	Athletes, sports persons and related associate professionals	49	54
536	獣医師	41	獣医師	1	2223	Veterinarians	61	83
537	保母、保父	42	保母、保父	2	2332	Pre-primary education teaching professionals	49	43
538	社会福祉事業専門職員	43	社会福祉事業専門職員	2	2446	Social work professionals	52	51
539	個人教師	44	個人教師	3	3340	Other teaching associate professionals	50	38
540	不動産鑑定士	45	他に分類されない、専門的・技術的職業従事者	2	3000	Technicians and associate professionals	48	54
541	経営コンサルタント			2	2419	Business professionals (n.e.c)	57	69
542	アナウンサー(ラジオ・テレビ)			2	2419	Business professionals (n.e.c)	57	69
543	図書館司書			2	3472	Radio, television and other announcers	50	64
544	その他の専門的・技術的職業従事者			2	2432	Librarians and related information professionals	54	65
545	管理的公務員	46	管理的公務員	2	3000	Technicians and associate professionals	48	54
546	国会議員	47	国会議員	1	1120	Senior government officials	71	77
547	地方議員	48	地方議員	1	1110	Legislators	64	77
548	会社役員	49	会社役員	1	1110	Legislators	64	77
549	その他の法人・団体の役員			1	1210	Directors and chief executives	70	70
				2	1140	Senior officials of special-interest organizations	63	58

50	公社・公団の役員	2	1140	Senior officials of special-interest organizations	63	58
51	その他の法人・団体の役員	2	1140	Senior officials of special-interest organizations	63	58
550	会社・団体の管理職員	2	1240	Office managers	55	58
551	駅長・区長	1	1226	Production and operations department managers in transport, storage and communication	60	59
552	郵便局長、電報・電話局長	1	1226	Production and operations department managers in transport, storage and communication	60	59
553	その他の管理的職業従事者	2	1300	General managers	50	51
55	一般事務員	3	4190	Other office clerks	37	39
554	総務・企画事務員	3	4190	Other office clerks	37	39
555	受付・案内事務員	3	4222	Receptionists and information clerks	38	52
556	出荷・受荷事務員	3	4131	Stock clerks	30	32
557	営業・販売事務員	3	4190	Other office clerks	37	39
558	その他の一般事務員	3	4190	Other office clerks	37	39
559	会計事務員	3	4121	Accounting and book-keeping clerks	45	51
560	郵便・通信事務員	3	4190	Other office clerks	37	39
561	集金人	3	4212	Tellers and other counter clerks	42	46
562	その他の外務事務従事者	3	4190	Other office clerks	37	39
563	運輸事務員	3	4133	Transport clerks	37	45
564	速記者、タピスト、キーハンチャー	3	4111	Stenographers and typists	42	51
61	速記者、タピスト	3	4111	Stenographers and typists	42	51
62	せん孔機等操作員	3	4113	Data entry operators	45	50
63	電子計算機等操作員	3	4114	Calculating-machine operators	45	51
566	小売店主	5	1314	General managers in wholesale and retail trade	46	49
567	卸売店主	5	1314	General managers in wholesale and retail trade	46	49
568	飲食店主	5	1315	General managers of restaurants and hotels	38	44
569	販売店員	3	5220	Shop, stall and market salespersons and demonstrators	32	43
570	行商人、呼売人、露天商	9	9000	Elementary occupations (n.f.s.)	21	20
571	再生資源卸売人・回収人	9	9161	Garbage collectors	13	23
572	商品仲立人	3	5230	Stall and market salespersons	24	37
573	外交員(保険・不動産を除く)	3	4190	Other office clerks	37	39
574	保険代理人・外交員	3	3412	Insurance representatives	44	54
575	不動産仲買人・売買人	2	3413	Estate agents	49	59
576	質屋店主・店員	3	4214	Pawnbrokers and money-lenders	15	40
577	その他の販売類似職業従事者	3	5200	Models, salespersons and demonstrators	31	43

578	女中、家政婦、家事サービス職業従事者	9	9131	Domestic helpers and cleaners	22	16
269	家事女中（住込みのもの）	9	9131	Domestic helpers and cleaners	22	16
270	家政婦	9	9131	Domestic helpers and cleaners	22	16
271	その他の家事サービス職業従事者	9	9131	Domestic helpers and cleaners	22	16
579	理容師、美容師	8	5141	Hairdressers, barbers, beauticians and related workers	32	29
272	理容師	8	5141	Hairdressers, barbers, beauticians and related workers	32	29
273	美容師	8	5141	Hairdressers, barbers, beauticians and related workers	32	29
580	クリーニング職、洗張職	9	9133	Hand-launders and pressers	22	16
581	料理人	8	5122	Cooks	31	30
582	バーテンパー	9	5123	Waiters, waitresses and bartenders	21	34
278	給仕人	9	5123	Waiters, waitresses and bartenders	21	34
583	給仕係	9	5123	Waiters, waitresses and bartenders	21	34
584	スチュワーデス、スチュワード'	3	5111	Travel attendants and travel stewards	50	34
585	接客社交係	9	5149	Other personal services workers (n.e.c.)	29	19
586	娯楽場等の接客員	3	4211	Cashiers and ticket clerks	34	53
283	その他の個人サービス職業従事者	9	5149	Other personal services workers (n.e.c.)	29	19
587	旅行・観光案内人	3	5113	Travel guides	29	34
588	その他の個人サービス職業従事者	9	5149	Other personal services workers (n.e.c.)	29	19
589	旅館主人・番頭、ホテル支配人	2	1315	General managers of restaurants and hotels	38	44
590	下宿・アパートの管理人、舎監、寮母	3	5121	Housekeepers and related workers	37	30
591	ファッションモデル	3	5210	Fashion and other models	28	43
592	その他のサービス職業従事者	9	5149	Other personal services workers (n.e.c.)	29	19
274	浴場従事者	9	5149	Other personal services workers (n.e.c.)	29	19
280	芸者、ダンサー	9	5149	Other personal services workers (n.e.c.)	29	19
282	一時預かり人、下足番、賃借人	9	5149	Other personal services workers (n.e.c.)	29	19
286	ファッションモデル等の広告宣伝人	9	5149	Other personal services workers (n.e.c.)	29	19
288	他に分類されないサービス職業従事者	9	5149	Other personal services workers (n.e.c.)	29	19
593	自衛官	8	5164	Armed forces, soldiers	39	40
594	警察官・海上保安官・鉄道公安官	8	5162	Police officers	40	50

595	消防員	265	消防員	8	5161	Fire-fighters	35	42
596	看守、守衛、監視人	266	看守、守衛、監視人	9	5163	Prison guards	39	40
597	その他の保安職従事者	267	その他の保安職従事者	9	5169	Protective services workers (n.e.c)	30	40
598	旧職業軍人	268	旧職業軍人	8	5164	Armed forces, soldiers	39	40
599	農耕・養蚕作業者	76	農耕・養蚕作業者	11	6110	Market gardeners and crop growers (n.f.s.)	40	23
600	植木職、造園師	82	植木職、造園師	11	6113	Gardeners, horticultural and nursery growers	40	23
601	畜産作業者	77	畜産作業者	11	6120	Market-oriented animal producers and related workers	40	23
602	林業作業者			11	6140	Forestry and related workers	24	22
78	育林作業者			11	6141	Forestry workers and loggers	24	22
79	伐木・造材作業者			11	6141	Forestry workers and loggers	24	22
80	集材・運材作業者			11	6141	Forestry workers and loggers	24	22
81	製炭・製薪作業者			11	6142	Charcoal burners and related workers	16	22
83	その他の農林作業者			11	6100	Skilled agricultural and fishery workers	38	23
603	その他の農林作業者			11	6150	Fishery workers, hunters and trappers	28	28
604	漁業作業者			11	6150	Fishery workers, hunters and trappers	28	28
84	漁ろう作業者			11	6150	Fishery workers, hunters and trappers	28	28
86	海草・貝採取人			11	6152	Inland and coastal waters fishery workers	23	28
87	水産養殖作業者			11	6151	Aquatic life cultivation workers	23	28
88	その他の漁業作業者			11	6150	Fishery workers, hunters and trappers	28	28
605	漁船の船長・航海士・機関長・機関士	85	漁船の船長・航海士・機関長・機関士	11	6150	Fishery workers, hunters and trappers	28	28
606	電車・機関車運転士			8	8311	Locomotive-engine drivers	43	41
97	電気機関士、蒸気機関士			8	8311	Locomotive-engine drivers	43	41
98	電車・気動車運転士			8	8311	Locomotive-engine drivers	43	41
607	自動車運転者	99	自動車運転者	9	8320	Motor-vehicle drivers	32	34
608	船長・航海士(漁船を除く)、水夫	100	船長・航海士、水夫人(漁船を除く)	2	3142	Ships' deck officers and pilots	55	52
609	船舶機関長・機関士(漁船を除く)	101	船舶機関長・機関士(漁船を除く)	2	3141	Ships' engineers	60	52
610	航空機械操縦士、航空士、航空機関士	102	航空機械操縦士、航空士、航空機関士	1	3143	Aircraft pilots and related associate professionals	60	69
611	車掌	103	車掌	3	5112	Transport conductors	32	34
612	鉄道員			9	8312	Railway brakemen, signalers and shunters	29	32
104	操車掛、信号掛、転てつ手・連結手			9	8312	Railway brakemen, signalers and shunters	29	32
613	船員			9	8340	Ships' deck crews and related workers	29	32
105	甲板員			9	8340	Ships' deck crews and related workers	29	32
106	船舶機関員			8	8162	Steam-engine and boiler operators	35	27

614	その他の運輸従事者	107	他に分類されない運輸従事者	9	9330	Transport laborers and freight handlers	20	29
615	無線通信士、無線技術士	108	無線通信士、無線技術士	2	3132	Broadcasting and telecommunications equipment operators	49	57
616	有線通信士	109	有線通信士	3	4190	Other office clerks	37	39
617	電話交換手	110	電話交換手	3	4223	Telephone switchboard operators	38	52
618	郵便・電報外務員	111	郵便・電報外務員	9	4142	Mail carriers and sorting clerks	33	39
619	その他の通信従事者	112	その他の通信従事者	2	3145	Air traffic safety technicians	46	50
620	採鉱員、採炭員			9	7111	Miners and quarry workers	34	30
		89	採鉱員	9	7111	Miners and quarry workers	34	30
		90	採炭員	9	7111	Miners and quarry workers	34	30
		93	支柱員	9	7111	Miners and quarry workers	34	30
		94	坑内運搬員	9	7111	Miners and quarry workers	34	30
		95	選鉱員、連炭員	9	7111	Miners and quarry workers	34	30
621	石切出作業者	91	石切出作業者	9	7111	Miners and quarry workers	34	30
622	その他の採掘作業者			9	7110	Miners, shofifers, stone cutters and carvers	34	30
		92	砂利・砂・粘土採取人	9	7110	Miners, shofifers, stone cutters and carvers	34	30
		96	その他の採鉱・採石作業者	9	7110	Miners, shofifers, stone cutters and carvers	34	30
623	陶磁器工、絵付作業者			9	7320	Potters, glass-makers and related trades workers	28	28
		196	陶磁器工	9	7321	Abrasive wheel formers, potters and related workers	25	27
		197	窯業絵付工	8	7324	Glass, ceramics and related decorative painters	31	29
624	石工	201	石工	9	7122	Bricklayers and stonemasons	34	29
625	ガラス・セメント製品製造作業者			9	8130	Glass, ceramics and related plant operators	31	22
		195	ガラス製品成形工	9	8131	Glass and ceramics kiln and related machine operators	31	22
		199	セメント製造工	8	8152	Chemical-heat-treating-plant operators	43	35
		200	セメント製品製造工	9	8212	Cement and other mineral products machine operators	30	30
626	その他の窯業・土石製品製造作業者			9	8139	Glass, ceramics and related plant operators (n.e.c.)	31	22
		194	窯業原料工	9	8139	Glass, ceramics and related plant operators (n.e.c.)	31	22
		198	れんが・かわら・土管製造工	9	8139	Glass, ceramics and related plant operators (n.e.c.)	31	22
		202	その他の窯業・土石製品製造作業者	9	8139	Glass, ceramics and related plant operators (n.e.c.)	31	22
627	製鉄工、製鋼工、精錬工			9	8120	Metal-processing plant operators	40	30
		113	製鉄工、製鋼工	9	8120	Metal-processing plant operators	40	30
		114	非鉄金属精錬工	9	8121	Ore and metal furnace operators	45	31
628	鋳物工、鍛造工、金属材料製造作業者			8	7200	Metal, machinery and related trades workers	40	34
		115	鋳物工	8	7211	Metal molders and coremakers	38	29

116	鍛造工	8	7221	Blacksmiths, hammer-smiths and forging-press workers	35	33
117	圧延工	9	8122	Metal melters, casters and rolling-mill operators	36	30
118	伸線工	8	7215	Riggers and cable splicers	32	30
119	金属熱処理工	9	8123	Metal heat-treating-plant operators	38	28
120	その他の金属材料製造作業者	8	7200	Metal, machinery and related trades workers	40	34
629	化学製品製造作業者	8	8150	Chemical-processing-plant operators	42	35
217	化学工	8	8151	Crushing-, grinding-, and chemical-mixing-machinery operators	43	35
218	油脂加工工	8	8153	Chemical-filtering-, and separating-equipment operators	43	35
219	その他の化学製品製造作業者	8	8159	Chemical-processing-plant operators (n.e.c.)	43	35
630	金属工作機械工、めっき工、金 属加工作業者	8	7200	Metal, machinery and related trades workers	40	34
121	金属工作機械工	9	8211	Machine-tool operators	38	36
122	金属プレス工	8	7221	Blacksmiths, hammer-smiths and forging-press workers	35	33
127	めっき工	9	8223	Metal finishing-, plating- and coating-machine operators	28	30
128	手仕上工	8	7224	Metal wheel-grinders, polishers and tool sharpeners	27	24
129	その他の金属加工作業者	8	7220	Blacksmiths, tool-makers and related trades workers	37	35
631	鉄工、板金工	8	7213	Sheet-metal workers	34	33
124	鉄工、びょう打工、製かん工	8	7213	Sheet-metal workers	34	33
125	板金工	8	7213	Sheet-metal workers	34	33
126	金属彫刻工	8	7224	Metal wheel-grinders, polishers and tool sharpeners	27	24
123	金属溶接工	8	7212	Welders and flame cutters	39	30
632	金属溶接工	9	8281	Mechanical-machinery assemblers	30	30
130	一般機械組立工・修理工	9	8281	Mechanical-machinery assemblers	30	30
131	一般機械修理工	9	8281	Mechanical-machinery assemblers	30	30
137	自動車組立工	9	8281	Mechanical-machinery assemblers	30	30
634	電気機械器具組立工・修理工	9	8282	Electrical-equipment assemblers	48	34
132	電気機械器具組立工・修理工	9	8282	Electrical-equipment assemblers	48	34
133	半導体製品製造工	9	8283	Electronic-equipment assemblers	48	34
134	電球・真空管組立工	9	8282	Electrical-equipment assemblers	48	34
135	被覆電線製造工	9	8282	Electrical-equipment assemblers	48	34
136	その他の電気機械器具組立・修理作業者	9	8282	Electrical-equipment assemblers	48	34
635	自動車組立工・整備工	8	7231	Motor vehicle mechanics and fitters	43	34
137	自動車組立工	9	8281	Mechanical-machinery assemblers	30	30
138	自動車修理工	8	7231	Motor vehicle mechanics and fitters	43	34
139	鉄道車両組立工・修理工	8	7231	Motor vehicle mechanics and fitters	43	34

637	船舶ぎ装工(他に分類されない)	140	船舶ぎ装工(他に分類されない)	8	7214	Structural-metal preparers and erectors	44	30
638	航空機組立工・修理工	141	航空機組立工・修理工	8	7232	Aircraft engine mechanics and fitters	50	42
639	自転車組立工・修理工	142	自転車組立工・修理工	9	8281	Mechanical-machinery assemblers	30	30
640	その他の輸送機械組立・修理工	143	その他の輸送機械組立・修理工	8	7233	Agricultural- or industrial-machinery mechanics and fitters	42	33
641	時計組立工・修理工	144	時計組立工・修理工	8	7311	Precision-instrument makers and repairers	47	38
642	光学機械・精密機械器具組立 作業者			8	7310	Precision workers in metal and related materials	45	38
		145	レンズ研ま工・調整工	9	7322	Glass-makers, cutters, grinders and finishers	37	29
		146	光学機械器具組立・修理作業 者	8	7311	Precision-instrument makers and repairers	47	38
		147	計器組立工・調整工	8	7311	Precision-instrument makers and repairers	47	38
		148	その他の精密機械器具組立・修 理作業者	8	7310	Precision workers in metal and related materials	45	38
643	精穀工・製粉工	203	精穀工・製粉工	9	8273	Grain- and spice-milling-machine operators	33	29
644	パン・菓子・めん類・豆腐製造工			8	7410	Food processing & related trade workers (n.f.s.)	28	30
		204	パン・菓子製造工	8	7412	Bakers, pastry-cooks and confectionery makers	33	31
		205	めん類製造工	8	7410	Food processing & related trade workers (n.f.s.)	28	30
		206	豆腐・こんにやく・ふ製造工	8	7410	Food processing & related trade workers (n.f.s.)	28	30
645	味噌・醤油・缶詰食品・乳製品製 造工、飲食料品製造作業者			9	8270	Food and related products machine operators (n.f.s.)	33	29
		207	精糖工	9	8276	Sugar production machine operators	45	29
		208	製茶工	9	8277	Tea-, coffee- and cocoa-processing-machine operators	34	29
		209	味そ・しょう油製造工	8	7410	Food processing & related trade workers (n.f.s.)	28	30
		210	かん詰・びん詰食品製造工	9	8270	Food and related products machine operators (n.f.s.)	33	29
		212	清涼飲料製造工	9	8270	Food and related products machine operators (n.f.s.)	33	29
		213	乳製品製造工	9	8272	Dairy-products machine operators	34	29
		214	水産物加工工	9	8271	Meat- and fish-processing-machine operators	31	29
		215	動植物油脂製造工	9	8270	Food and related products machine operators (n.f.s.)	33	29
		216	その他の飲食料品製造作業者	9	8278	Brewers, wine and other beverage machine operators	34	29
646	たばこ製造工	240	たばこ製造工	9	8279	Tobacco production machine operators	39	29
647	酒類製造工	211	酒類製造工	8	7410	Food processing & related trade workers (n.f.s.)	28	30
648	製糸作業者			9	8261	Fiber-preparing-, spinning- and winding-machine operators	29	29
		149	繰糸工	9	8261	Fiber-preparing-, spinning- and winding-machine operators	29	29
		150	粗紡工、精紡工	9	8261	Fiber-preparing-, spinning- and winding-machine operators	29	29
		151	合糸工、ねん糸工、加工糸工	9	8261	Fiber-preparing-, spinning- and winding-machine operators	29	29
		152	繰返工、かせ取工	9	8261	Fiber-preparing-, spinning- and winding-machine operators	29	29
		159	その他の製糸・紡織作業者	9	8261	Fiber-preparing-, spinning- and winding-machine operators	29	29

658	印刷・製本作業者	紙製品製造工 その他のハルプ・紙・紙製品製造 作業者	9	8253	Paper-products machine operators	28	28
181			9	8140	Wood-processing- and papermaking-plant operators	28	27
182		製版工	9	8250	Printing-, binding- and paper-products machine operators	41	38
183		文選工、植字工	8	7342	Stereotypers and electrotypers	41	40
184		印刷工	8	7341	Compositors, typesetters and related workers	42	40
185		製本工	9	8251	Printing-machine operators	41	38
186		その他の印刷・製本作業者	9	8252	Book-binding-machine operators	32	38
659	ゴム・プラスチック製品製造作業者		9	8250	Printing-, binding- and paper-products machine operators	41	38
187		ゴム工	9	8230	Rubber- and plastic-products machine operators	30	30
188		ゴム製品成形工	9	8231	Rubber-products machine operators	30	30
189		プラスチック製品成形工・加工工・ 仕上工	9	8232	Rubber-products machine operators	30	30
190		その他のゴム・プラスチック製品製 造作業者	9	8230	Rubber- and plastic-products machine operators	30	30
660	くつ製造工・修理工、かわ・かわ 製品製造作業者		8	7440	Pelt, leather and shoemaking trades workers	27	31
191		製革工	8	7441	Pelt dressers, tanners and fellmongers	22	31
192		くつ製造工・修理工	8	7442	Shoe-makers and related workers	27	31
193		その他のかわ・かわ製品製造作 業者	8	7440	Pelt, leather and shoemaking trades workers	27	31
241		塗装工、画工、看板工	8	7141	Painters and related workers	31	29
242		漆塗工、まき絵師	8	7324	Glass, ceramics and related decorative painters	31	29
663		表具工、内張工	8	7124	Carpenters and joiners	37	29
243		内張工	8	7124	Carpenters and joiners	37	29
244		表具師	8	7124	Carpenters and joiners	37	29
245		和がさ・ちようちん・うちわ職	9	7331	Handicraft workers in wood and related materials	31	29
246		貴金属・宝石・甲・角等細工	8	7313	Jewelry and precious-metal workers	43	38
247		印判師	8	7343	Printing engravers and etchers	41	42
248		洋傘組立工	9	8280	Assemblers	33	31
249		かばん・袋物製造工	8	7440	Pelt, leather and shoemaking trades workers	27	31
250		がん具製造工	9	8280	Assemblers	33	31
670		製図工、現図工	2	3118	Draftpersons	55	51
251		製図工、写図工	2	3118	Draftpersons	55	51
252		現図工	8	7214	Structural-metal preparers and erectors	44	30
254		映写技士	8	7344	Photographic and related workers	42	40

672	その他の技能工・生産工程作業者			9	8400	Semiskilled workers n.f.s.	33	24
		253	包装工	9	8400	Semiskilled workers n.f.s.	33	24
		255	他に分類されない技能工・生産工程作業者	9	8400	Semiskilled workers n.f.s.	33	24
673	汽かん士、気かん火夫	231	汽かん士、気かん火夫	8	8162	Steam-engine and boiler operators	35	27
674	起重機・建設機械運転作業者			8	8332	Earth-moving and related plant operators	32	26
		232	起重機・巻上機運転工	8	8333	Crane, hoist and related plant operators	33	28
		233	建設機械運転工	8	8332	Earth-moving and related plant operators	32	26
675	その他の定置機関運転作業者	234	その他の定置機関運転作業者	8	8160	Power-production and related plant operators	38	32
676	発電員、変電員	235	発電工、変電工	8	8161	Power-production plant operators	42	33
677	電気工事・電話工事作業者			8	7240	Electrical and electronic equipment mechanics and fitters	38	40
		236	電線架線工	8	7245	Electrical line installers, repairers and cable joiners	36	38
		237	電気工事人	8	7240	Electrical and electronic equipment mechanics and fitters	38	40
		238	電信電話機械据付工・保守工	8	7244	Telegraph and telephone installers and servicers	35	40
		239	その他の電気作業者	8	7240	Electrical and electronic equipment mechanics and fitters	38	40
678	土木・建築請負師	220	土木・建築請負師	8	7129	Building frame and related trades workers (n.e.c.)	28	30
679	大工、左官、とび職			8	7120	Building frame and related trades workers	34	30
		221	大工	8	7124	Carpenters and joiners	37	29
		222	屋根職	9	7131	Roofers	31	19
		223	左官	8	7133	Plasterers	31	31
		224	とび職	8	7129	Building frame and related trades workers (n.e.c.)	28	30
680	れんが積工、配管工			8	7136	Plumbers and pipe fitters	34	33
		225	れんが積工・タイル張工	9	7122	Bricklayers and stonemasons	34	29
		226	配管工・鉛工	8	7136	Plumbers and pipe fitters	34	33
681	量職	227	量職	9	7331	Handicraft workers in wood and related materials	31	29
682	土工、道路工夫	228	土工、道路工夫	9	9312	Construction and maintenance laborers: roads, dams and similar constructions	15	21
683	鉄道線路工夫	229	鉄道線路工夫	9	9312	Construction and maintenance laborers: roads, dams and similar constructions	15	21
684	現場監督、その他の建設作業者			8	7120	Building frame and related trades workers	34	30
685	倉庫夫、仲仕	230	その他の建設作業者	8	7120	Building frame and related trades workers	34	30
			倉庫夫	9	9333	Freight handlers	20	30
		257	倉庫夫	9	9333	Freight handlers	20	30
		258	沖仲仕、沿岸仲仕	9	9333	Freight handlers	20	30

686	運搬作業者	259	陸仲仕、運搬夫	9	9333	Freight handlers	20	30
				9	9330	Transport laborers and freight handlers	20	29
		256	荷造工	9	9330	Transport laborers and freight handlers	20	29
		261	配達人	9	9330	Transport laborers and freight handlers	20	29
687	清掃員	287	清掃員	9	9162	Sweepers and related laborers	13	23
688	その他の労務作業者			9	9300	Laborers in mining, construction, manufacturing and transport	18	23
		260	駅手	9	9300	Laborers in mining, construction, manufacturing and transport	18	23
		262	その他の単純労働者	9	9300	Laborers in mining, construction, manufacturing and transport	18	23

コード番号 501、1975 年版コード番号 1 の「自然科学系研究者」に該当する ISCO の 3 つの項目 (2100, 2120, 2200) をしめた。ただし、これ以降ではこのような複数該当の場合は記載していない。この複数該当のままでは、SSM コードを ISCO、EGP、SOPS、ISEI に変換できない。1 つの SSM 項目に ISCO の複数コードを割りふることになってしまうからである。そこでこのような例の場合、SIOPS と ISEI を参照し、原則としてそれらの中間的なスコアをもつ ISCO 項目を割り当てた。上の例では、2100 を割り当てることになる。

また SSM 職業項目じたいが 1995 年版と 1975 年版とで違うので、それによる齟齬が生じる場合もある。たとえば、95 年版の「583 給仕係、584 スチュワーデス/スチュワード」と 75 年版の「278 給仕人」、前者の「623 陶磁器工、絵付作業員」と後者の「196 陶磁器工、197 窯業絵付工」のように、一方が他方を包含する職業項目とされた場合である。こうした場合、より包括的な SSM 項目にたいして、ISCO のより包括的なコード (下 1 桁目が 0)、または従事者数がより多いと考えられる職業項目を割り当てた。

4. Ganzeboom and Treiman による EGP 階級分類作成 (第 2 段階)

EGP 階級分類は 11 分類として提示されたが、データ分析で実際に使用されてきたほとんどのものは、その一部を合併した 6 分類か 7 分類である。しかし、Ganzeboom and Treiman(1966)が変換方法を提示したのは 10 分類である。それは、11 分類の中の IIIa (Higher grade routine nonmanual) と IIIb (Lower grade routine nonmanual) が実際には分離して利用されてこなかったからである。かれらはこれらを合併して、表 2 のように III (Higher and lower grade routine nonmanual) としている。

かれらによる ISCO から EGP 階級分類への変換は、2 段階によっておこなわれる。第 1 段階では、ISCO コードにたいして EGP コードを暫定的に割り当てる変換である。このとき割り当てられる EGP コードは表 2 の中の 1, 2, 3, 7, 8, 9, 10, 11 のみである。6 のコードがないのは、上述したように IIIa と IIIb を合併したためである。他の 4 (Small employer), 5 (Independent) の自営業層は、次の第 2 段階で割り当てられる。また第 1 段階で 7 (Manual foreman) が割り当てられるのは、ISCO コード 3452 の Armed forces non commissioned officers だけなので、7「監督的マニュアル」も実質的にはほとんど第 2 段階で割り当てられる。

Ganzeboom and Treiman (1996) では、第 2 段階の処理が SPSS シンタックスとして記載されているが、その処理の論理について説明をしていない。ここでは、その処理の論理を整理しておく。第 2 段階では次の(1)～(8)のような補正がおこなわれる。それらの処理では、第 1 段階で割り当てた EGP コードにくわえて、ISCO コード、自営、監督的地位が使用される。自営は、自営業であるか否かの 2 分類である。監督的地位は、管理職者については部下の数、経営者については従業員数によって分類され、0 人、1～9 人、10 人以上の 3 分類である。

表2 EGP階級10分類

1	I	Higher service 上層サービス階級 Includes mostly professionals, large enterprise employers and higher managers (>10 subordinates)
2	II	Lower service 下層サービス階級 Includes mostly associate professionals, lower managers (1-10 subordinates), higher sales
3	III	Routine clerical/sales 事務販売 Includes routine clerical and sales workers
4	IVa	Small employer 自営業主 Includes small entrepreneurs (1-10 subordinates)
5	IVb	Independent 独立自営 Own account workers, no employees
7	V	Manual foreman 監督的マニュアル Manual workers with supervisory status (>1 subordinate)
8	VI	Skilled manual 熟練マニュアル Mostly craft workers, some skilled service, and skilled machine operators
9	VIIa	Semi-Unskilled manual 半熟練・非熟練マニュアル Mostly machine operators, elementary laborers, elementary sales and services
10	VIIb	Farm workers 農業労働者 Employed farm workers, irrespective of skill level; also family farm workers
11	IVc	Farmers/Farm managers 農業経営 Self-employed and supervisory farm workers, irrespective of skill level

(1)事務・販売から下層サービス階級への変更

第1段階で「3 (Routine clerical/sales)」に割り当てられ、かつ「監督的地位」(部下または雇用者が1人以上)である場合に、「2 (Lower service)」に変更する。

(2)自営業主の分離

この直前段階までにおいて「2 (Lower service)または3 (Routine clerical/sales)」に割り当てられ、かつ「自営」であり、かつISCOが下記(a)(b)(c)に該当する場合に「4 (Small employer)」へ変更する。

(a)1310 番台(1310 から 1319 まで)の[Small enterprise] General managers

(b)3400 番台(3400 から 3439 まで)の Other associate professionals

(c)4000 番台と 5000 番台(4000 から 5230 まで)の Clerks と Service workers and shop & market sales workers

ここでは、(a)小企業(10人未満)の管理職、(b)経理・販売・法務・税務など事業にたずさわる準専門職、(c)事務職・販売職・サービス職の従事者の中から自営業者を分離している。いいかえれば、ここで自営業者に変更されない者は、ISCOでの職業が優先され、第1段階で割り当てられた Lower service または Routine clerical/sales のままとされる。

(3)独立自営の分離

第1段階で「7 (Manual foreman), 8 (Skilled manual), 9 (Semi-Unskilled manual)のいずれか」に割り当てられ、かつ「自営」であり、かつ(d)「ISCOが9300番台以外(9300から9333ま

での Laborers in mining, construction, manufacturing and transport 以外)である」場合に、「5 (Independent)」に変更する。ここでは、マニュアル職で自営の場合に、独立自営に変更する。ただし鉱業、建設業、製造業、運輸業の単純作業労働者の場合は、自営であっても第1段階で割りふった EGP コードのままとする。

(4) 熟練労働者から監督的マニュアルへの変更

第1段階で「8 (Skilled manual)」に割り当てられ、かつ「監督的地位」(部下または雇用者が1人以上)である場合に、「7 (Manual foreman)」に変更する。

(5) 農業労働者から農業経営への変更

第1段階で「10 (Farm workers)」に割り当てられ、かつ「自営」である場合、「11 (Farmers/Farm managers)」に変更する。

(6) 自営業主から独立自営への変更

この直前段階までにおいて「4 (Small employer)」に割り当てられ、かつ「監督的地位」でない(部下または雇用者がいない)場合に、「5 (Independent)」に変更する。

(7) 独立自営から自営業主への変更

この直前段階までにおいて「5 (Independent)」に割り当てられ、かつ「監督的地位」(部下または雇用者が1人以上)である場合に、「4 (Small employer)」に変更する。

(8) 上層サービス階級への変更

この直前段階までにおいて「2 (Lower service), 3 (Routine clerical/sales), 4 (Small employer) のいずれか」に割り当てられ、かつ「監督的地位」で部下または雇用者が10人以上である場合に、「1 (Higher service)」に変更する。

表3 Ganzeboom and Treiman (1996)の変換補正処理

補正前 EGP	自営	部下または 雇用者の数	該当 職業	補正後 EGP
(1) 事務販売		1人以上		下層サービス
(2) 下層サービス/事務販売	自営		DG	自営業主
(3) 監督的/熟練/半・非熟練マニュアル	自営		PR	独立自営
(4) 熟練マニュアル		1人以上		監督的マニュアル
(5) 農業労働者	自営			農業経営
(6) 自営業主		0人		独立自営
(7) 独立自営		1人以上		自営業主
(8) 下層サービス/事務販売/自営業主		10人以上		上層サービス

表3は、Ganzeboom and Treiman (1996)による第2段階の補正処理の論理を整理したものである。ここでは、「自営」、「部下または雇用者の人数」、「該当職業(小分類)をあらわす変数 DG と PR」がすべて AND 条件とされている。DG は上記(2)の(a)(b)(c)の該当職業(ISCO)をあらわす。また PR は同じく(3)の(d)の該当職業をあらわす。

5. SSM 職業分類から EGP 階級分類への変換

この節では、SSM 職業分類から EGP 階級分類を作成する方法を説明する。Ganzeboom and Treiman (1996)による EGP 階級は 10 分類であったが、ここでは「10 Farm workers」を「11 Farmers/Farm managers」に包含させた 9 分類に変換する方法を提示する。そのため表 1 では、農林漁業にかんする SSM 職業項目に対応する「EGP9」コードはすべて 11 としてあり、10 はない。9 分類としたのは、日本では、被雇用の農林漁業従事者数がきわめて少なく、自営が大多数だからである。

表 4 1995 年版 SSM 職業小分類から EGP9 分類への変換(第 1 段階): 2005 年用

```

COMPUTE EG= SSM95.
RECODE EG
      (701=559)(702=679)(703=521)(704=633)(706=607).
RECODE EG
      (501 thru 504=1)(505=2)(506 thru 510=1)(511 thru 516=2)
      (517 thru 519=1)(520 thru 523=2)(524=1)(525 thru 535=2)(536=1)
      (537, 538=2)(539=3)(540 thru 544=2)(545 thru 548=1)(549, 550=2)
      (551, 552=1)(553=2)(554 thru 565=3)(566, 567=5)(568=5)
      (569=3)(570 thru 572=9)(573, 574=3)(575=2)(576, 577=3)
      (578=9)(579=8)(580=9)(581=8)(582, 583=9)(584=3)(585=9)(586, 587=3)
      (588=9)(589=2)(590, 591=3)(592=9)(593 thru 595=8)(596, 597=9)(598=8)
      (599 thru 605=11)(606=8)(607=9)(608, 609=2)(610=1)(611=3)
      (612 thru 614=9)(615=2)(616, 617=3)(618=9)(619=2)(620 thru 622=9)
      (623 thru 627=9)(628 thru 632=8)(633, 634=9)
      (635 thru 638=8)(639=9)(640 thru 642=8)(643=9)(644=8)
      (645, 646=9)(647=8)(648 thru 650=9)(651=8)(652,653=9)
      (654, 655=8)(656,657=9)(658=9)(659=9)(660 thru 663=8)
      (664=9)(665, 666=8)(667=9)(668=8)(669=9)(670=2)(671=8)(672=9)
      (673 thru 680=8)(681 thru 683=9)(684=8)(685 thru 688=9)
      (ELSE=999).

```

表 4 と表 5 は、表 1 に整理された<SSMコードーISCOコードーEGP9 分類>の対応関係にしたがって、SSMコードからEGPコードを暫定的に割り当てる第 1 段階のシンタックスである。それぞれ第 1 行はSSM職業小分類コードを変数EGにコピーし、第 2 行以下のRECODE文で EGP9 分類コードに変換する。表 4 は 1995 年版用だが 2005 年調査データを変換するための RECODE文を追加してある。追加したのは、1 番目のRECODE文であり、2005 年調査で新たに作られたコードを既存のSSM職業分類コードに変換させている⁴。1995 年調査データについては、このRECODEは必要ない。変数SSM95 は、1995 年版SSM職業小分類コードをその値

⁴ 2005 年調査で新たに新設されたのは、701 レジ・キャッシャー、702 大工、703 教員、704 製品製造作業
者、705 会社員、706 宅配便の配達、707 自営業である。これらのうち 707 自営業は、職業（仕事内容）に
かんする情報がえられないので既存コードへの変換をしていない。2005 年の新設コードについては、2005
年社会階層と社会移動調査研究会『2005 年 SSM 日本調査 コードブック』（pp.85-91）を参照されたい。

とする変数である。表 5 は 1975 年版SSM職業分類用で、1965 年から 1985 年までの各調査データに適用できる。変数SSM75 は 1975 年版SSM職業小分類コードである。これらの表のシタックスは、記載された部分だけならば父職、初職、現職などのいずれか 1 つを変換するものになっているが、DO REPEATなどをもちいれば複数を同時に変換できる。

表 6 は、第 2 段階の補正処理をおこなう論理を整理したものである。この表の(1)～(4)、(6)～(8)は、表 3 の同じ番号の補正処理に対応している。(5)は、農林漁業従事職を 1 つのカテゴリに合併したために必要ないので設定していない。「自営」には、経営者・役員を含むが、家族従業者は含まない。記載されている「自営」「職位」「企業規模」「該当職業 DG と PR」はすべて AND 条件である。

表 5 1975 年版 SSM 職業小分類から EGP9 分類への変換(第 1 段階)

COMPUTE EG=SSM75.	
RECODE EG	
(1 thru 9=1)(10=2)(11 thru 15=1)(16 thru 21=2)(22 thru 24=1)	
(25 thru 28=2)(29=1)(30 thru 40=2)(41=1)(42, 43=2)(44=3)(45=2)	
(46 thru 49=1)(50, 51=2)(52, 53=1)(54=2)(55 thru 63=3)(64, 65=5)	
(66=5)(67=3)(68, 69=9)(70 thru 72=3)(73=2)(74, 75=3)	
(76 thru 88=11)(89 thru 96=9)(97, 98=8)(99=9)(100, 101=2)(102=1)	
(103=3)(104, 105=9)(106=8)(107=9)(108=2)(109, 110=3)(111=9)(112=2)	
(113, 114=9)(115, 116=8)(117=9)(118=8)(119=9)(120=8)(121=9)(122=8)	
(123 thru 126=8)(127=9)(128, 129=8)(130 thru 137=9)(138 thru 141=8)	
(142=9)(143, 144=8)(145=9)(146 thru 148=8)(149 thru 159=9)	
(160 thru 162=8)(163 thru 170=9)(171, 172=8)(173 thru 181=9)	
(182, 183=8)(184 thru 190=9)(191 thru 193=8)(194 thru 196=9)	
(197=8)(198=9)(199=8)(200 thru 203=9)(204 thru 206=8)(207, 208=9)	
(209=8)(210=9)(211=8)(212 thru 216=9)(217 thru 221=8)(222=9)	
(223, 224=8)(225=9)(226=8)(227 thru 229=9)(230=8)(231 thru 239=8)	
(240=9)(241 thru 244=8)(245=9)(246, 247=8)(248=9)(249=8)(250=9)	
(251=2)(252=8)(253=9)(254=8)(255 thru 262=9)(263 thru 265=8)	
(266, 267=9)(268=8)(269 thru 271=9)(272, 273=8)(274, 275=9)(276=8)	
(277 thru 280=9)(281=3)(282, 283=9)(284=2)(285=3)(286 thru 288=9)	
(ELSE=999).	

表 6 SSM 調査データから EGP 分類への変換補正処理

補正前 EGP	自営 ^a	職位	企業規模 ^b	該当職業	補正後 EGP
(1) 事務販売		係長以上			下層サービス
(2) 下層サービス/事務販売	自営			DG	自営業主
(3) 監督的/熟練/半・非熟練マニュアル	自営			PR	独立自営
(4) 熟練/半・非熟練マニュアル		職・班長/ 係・課長/部長			監督的マニュアル
(6) 自営業主			1 人		独立自営
(7) 独立自営			2-29 人		自営業主
(8) 下層サービス/事務販売/自営業主		部長以上	30 人以上		上層サービス

a 自営には経営者・役員を含む

b 本人を含む人数

表 7 は、1995 年版 SSM 職業小分類から EGP9 分類への変換における第 2 段階の補正をおこなう SPSS シンタックスで、2005 年調査データに適用するものである。左側の行番号は説明のために便宜的につけたものである。また変数 EG は、すでに表 4 または表 5 にしたがって暫定的な EGP9 コードが与えられていることを前提としている。

表 7 1995 年版 SSM 職業小分類から EGP9 分類への変換(第 2 段階): 2005 年調査用

1	COMPUTE	PR=SSM95.	
2	COMPUTE	DG=SSM95	
3	RECODE	PR (501 thru 613=1)(615 thru 681=1)(684, 687=1)(ELSE=0).	
4	RECODE	DG (553 thru 565=1)(569=1)(572 thru 577=1)(579=1) (581 thru 592=1)(597, 611=1)(616 thru 618=1)(ELSE=0).	
5	RECODE	SV (2=1)(3 thru 4=2)(5=3)(6=4)(ELSE=0).	
6	RECODE	SE (6, 7=1)(1=2)(ELSE=0).	
7	RECODE	SZ (1=1)(2 thru 4=2)(5 thru 10=3)(ELSE=0).	
8	IF (EG=3 AND SV>=2)		EG=2.
9	IF ((EG=2 OR EG=3) AND SE>=1 AND DG=1)		EG=4.
10	IF ((EG>=7 AND EG<=9) AND SE>=1 AND PR=1)		EG=5.
11	IF ((EG=8 OR EG=9) AND (SV>=1 AND SV<=3))		EG=7.
12	IF ((EG=4) AND SZ=1)		EG=5.
13	IF ((EG=5) AND SZ=2)		EG=4.
14	IF ((EG>=2 AND EG<=4) AND SV>=3 AND SZ=3)		EG=1.

表 8 1975 年版 SSM 職業小分類から EGP9 分類への変換(第 2 段階): 1975 年調査用

1	COMPUTE	PR=SSM75.	
2	COMPUTE	DG=SSM75	
3	RECODE	PR (1 thru 106=1)(108 thru 227=1)(230 thru 255=1)(266 thru 288=1) (ELSE=0).	
4	RECODE	DG (54 thru 63=1)(67=1)(70 thru 75=1)(103=1)(109 thru 111=1) (267=1)(272 thru 274=1)(276 thru 286=1)(288=1)(ELSE=0).	
5	RECODE	SV (25=1)(22 thru 24=2)(27, 28=2)(21=3)(10=4)(ELSE=0).	
6	RECODE	SE (3=1)(4=2)(1=3)(ELSE=0).	
7	RECODE	SZ (1=1)(2, 3=2)(4 thru 8=3)(ELSE=0).	
8	IF (EG=3 AND SV>=2)		EG=2.
9	IF ((EG=2 OR EG=3) AND SE>=1 AND DG=1)		EG=4.
10	IF ((EG>=7 AND EG<=9) AND SE>=1 AND PR=1)		EG=5.
11	IF ((EG=8 OR EG=9) AND (SV>=1 AND SV<=3))		EG=7.
12	IF (EG=4 AND SZ=1)		EG=5.
13	IF (EG=5 AND SZ=2)		EG=4.
14	IF ((EG>=2 AND EG<=4) AND SV>=3 AND SZ=3)		EG=1.

行 1 と 2 では SSM 職業小分類コードを変数 PR と変数 DG にコピーしている。行 3-4 は前節(3)の(d)の条件を、行 5-6 は同様に(2)の(a)(b)(c)の条件を設定している。変数 SV は職位(役職)変数で、行 5 ではそれを 1=職長・班長、2=係長・課長、3=部長、4=社長・役員等、0=その他(役職なし)に再コードしている。変数 SE は従業上地位変数で、行 6 ではそれを 1=自営業者、2=経営者・役員、0=その他に再コードしている。変数 SZ は従業企業規模で、行 7 ではそれを 1=1 人(本人のみ)、2=2~29 人、3=30 人以上と官公庁、0=その他に再コード

している。IF 文の 8 行目は表 6 の(1)、9 行目は(2)、10 行目は(3)、11 行目は(4)、12 行目は(6)、13 行目は(7)、14 行目は(8)の処理をおこなう。

表 8 は、1975 年版 SSM 職業小分類を EGP9 分類に変換補正するシンタックスで、1975 年調査データ用である。表 7 の 1995 年版 SSM 職業小分類からの変換との違いは、1-2 行で SSM 職業分類の変数名を SSM75 にしていること、3-7 行目の変数 PR、DG、SV、SE、SZ にかんする RECODE 文にある。他方で、8-14 行の補正処理じたいは表 7 と同一である。

1975 年版 SSM 職業小分類は 1965 年、1975 年、1985 年の調査データで、1995 年版は 1995 年と 2005 年の調査データでもちいられている。ここまで表 5 と表 8 で 1975 年用、表 4 と表 7 で 2005 年用の変換プログラムをしめた。つまり、1965 年と 1985 年調査については前者の 2 表、1995 年調査については後者の 2 表を使えば EGP9 分類を作成できる。ただし、SSM 調査での職位、従業上地位、企業規模にかんするコード番号は各調査年で変更されている。そのため、各調査年の変数 SV、SE、SZ について、表 9 にしたがって 5-7 行の RECODE 文の内容を各年について書き換えなければならない。とくに 1965 年調査では、従業上地位が職位とともに「役職コード」としてデータ化されていることに注意されたい。

表 9 各調査年の原コード名と RECODE 文の内容

変数	調査年	原コード名	RECODE
職位 SV	1965年	役職	(25=1) (22 thru 24=2) (27, 28=2) (21=3) (10=4)
	1975年	役職	(25=1) (22 thru 24=2) (27, 28=2) (21=3) (10=4)
	1985年	役職	(2=1) (3, 4=2) (5=3) (ELSE=0)
	1995年	役職	(2=1) (3, 4=2) (5=3) (6=4) (ELSE=0)
	2005年	役職	(2=1) (3, 4=2) (5=3) (6=4) (ELSE=0)
従業上 地位SE	1965年	役職	(30=1) (40=2) (10=3) (ELSE=0)
	1975年	従業上地位	(3=1) (4=2) (1=3) (ELSE=0)
	1985年	従業上地位	(4=1) (5=2) (1=3) (ELSE=0)
	1995年	従業上地位	(5=1) (1=3) (ELSE=0)
	2005年	従業上地位	(6=1) (1=3) (ELSE=0)
企業規模 SZ	1965年	従業企業規模	(1=1) (2, 3=2) (4 thru 7=3) (ELSE=0)
	1975年	従業企業規模	(1=1) (2, 3=2) (4 thru 8=3) (ELSE=0)
	1985年	従業企業規模	(1=1) (2 thru 4=2) (5 thru 10=3) (ELSE=0)
	1995年	従業企業規模	(1=1) (2 thru 4=2) (5 thru 10=3) (ELSE=0)
	2005年	従業企業規模	(1=1) (2 thru 4=2) (5 thru 10=3) (ELSE=0)

6. おわりに

SSM 職業小分類を国際的階層指標に変換するため、SSM 職業小分類と ISCO 職業項目との対応関係を整理し、さらに Ganzeboom and Treiman (1996)にしたがった EGP 階級分類の作成方法を提示した。表 10 には、1965 年から 2005 年調査まで、男性回答者の調査時点現職につ

いての EGP9 分類の集計結果をしめした。

ただしこのEGP階級分類について、注意すべき点がある。その第 1 は、管理職の分類にかんするものである。Ganzeboom らは、ISCO1200 番台のCorporate managersが企業規模 10 人以上、1300 番台のGeneral managersが 10 人未満と区別されることを前提としている。しかし、SSM職業分類における管理職の企業規模基準はそれと異なっていて、1. Higher serviceと 2. Lower serviceは 10 人を基準にして区別されていない⁵。企業規模をもちいてこの区別をおこなう補正も可能である。しかし実際的には、ほとんどの研究でこれらを合併してI+II. Service classとして扱っているの、あえてそのような補正を加えなかった。

表 10 各調査年男性回答者現職の EGP9 階級分類の構成比率(%)

		1965年	1975年	1985年	1995年	2005年
1	I Higher service	6.2	8.4	10.3	14.5	14.0
2	II Lower service	12.2	13.4	16.0	17.0	15.5
3	III Routine clerical/sales	13.3	13.9	13.8	11.7	13.0
4	IVa Small employer	11.7	8.3	7.0	13.8	10.3
5	IVb Independent	5.5	7.7	10.1	3.8	3.5
7	V Manual foreman	4.8	5.2	6.7	8.0	8.1
8	VI Skilled manual	10.2	12.0	10.4	10.6	11.4
9	VIIa Semi-Unskilled manual	16.3	16.0	18.2	14.7	18.1
11	IVc Farmers/Farm workers	19.9	15.2	7.5	5.8	6.0
	計	100.0	100.0	100.0	100.0	100.0
	N	1965	2532	2214	2182	2202

第 2 は、SSM職業分類と、ISCO分類・EGP階級分類との間では、ノンマニュアルとマニュアルの区別において、そして熟練マニュアルと半・非熟練マニュアルの区別において、一部に矛盾が生じることである。たとえば、警察官や消防員はSSM分類では事務職とされるが、EGP分類では熟練マニュアルになる。また、SSM分類で熟練とされるがEGP分類で半・非熟練になる職業も表 1 に含まれている。これらの相違の根底には、EGP分類そしてそれが依拠するISCOの分類体系や技能・熟練のとらえ方と、日本のSSM分類のそれらとの相違がある⁶。

第 3 に、4. Small employerと 5. Independentの区別が一貫しないことである。SSM調査では従業上の地位にかんして、1985 年まで従業員のいない単独自営と従業員のいる自営業者を区別していた。しかし、1995 年と 2005 年の調査では、この区別をしていない。表 10 で 1985 年から 1995 年にかけて、自営業主の比率が急増し、独立自営の比率が急減している大きな原

⁵ SSM 調査では、1985 年まで規模 5 人以上、1995 年で 30 人以上をめやすとして管理職か否かを判断してコーディングされた。しかし、実際のコーディングでは他の情報も考慮して総合的に判断されたので、この基準に完全に統一されていたとはいえないだろう。日本だけでなく各国でのこうしたローカルなコーディングルールの存在は、分類基準の不一致をもたらす原因になっている。ただし 2005 年 SSM 調査では、日本、韓国、台湾でのコーディングルール統一のために、企業規模の基準を 30 人としている。

⁶ SSM 分類と ISCO 分類の分類枠組・体系の相違については、田辺(2006)が詳しく論じているので参照されたい。

因の1つは、この自営にかんする区別の変更によるものと考えられる⁷。このため、自営業主と独立自営の区分は信頼性がなく、これらをそのまま分析にもちいるべきではない。ただし実際上は、ほとんどの研究においてこれらを合併したIVab. *Petty bourgeoisie*として分析している。したがって、合併して扱えば問題はないだろう。

第4点も自営にかんすることである。5節で述べたSSM職業小分類からEGP分類への変換補正処理における「自営」（つまり表6の「自営」、そして表7, 8, 9の変数SE）は、従業上地位について自営業者（従業員のいない単独自営および従業員のいる自営業者）と回答した者とともに、経営者・役員と回答した者も含んでいる。しかし、家族従業者と回答した者については含んでいない。家族従業者を自営業者とみなすか否かについては2つの立場が考えられる。1つの立場は、事業経営を家族が一体として行っており、事業の世代間継承が前提されているケースが多いとして、自営業者とみなすべきだというものである。他方の立場は、家族経営であっても職責の分離や所得・家計の分離があり、世代間継承が未確定のケースもあることを重視して、自営業者とみなさないというものである。本稿で採用したのは後者の立場であり、それは後者がより慎重だと考えたからである。

最後に、ここでの変換方法によって作成された2005年調査のEGP分類は、ISCOコードをもちいて変換したそれと完全には一致しない。SSM職業分類から作成したEGP分類を、既存研究で利用されてきた6分類（I+II. *Service class*, III. *Routine clerical/sales*, IVa+IVb. *Petty bourgeoisie*, V+VI. *Skilled manual*, VIIa. *Semi-Unskilled manual*, IVc+VIIb. *Farmers and farm workers*）として、これとISCOコードをもちいて変換したEGP6分類⁸の一致度を検討した結果、その一致率は比較的に高かった。それぞれから作成した階級カテゴリーが一致したのは、男性について86.1%（双方からそれぞれEGP階級分類を作成できた2197人のうち1892人が一致）、女性について85.1%（同じく1937人のうち1648人が一致）であった。V+VI. *Skilled manual*とVIIa. *Semi-Unskilled manual*を合併した5分類の場合の一致率は、男性で90.0%、女性で90.8%だった。さらにI+II. *Service*とIII. *Routine clerical/sales*を合併した4分類では、男性で94.3%、女性で95.7%だった。このように、SSM職業分類から作成する場合と、ISCOコードから作成する場合では、その間に矛盾が生じる。だがその齟齬は致命的というほどではないだろう。

上記の諸点がしめすように、本稿で提示した変換方法によって作成されるEGP分類、SIOPS、ISEIなどの国際的な比較可能性は完全だとはいいいがたい。そもそも完全な比較可能性は理想ではあっても、それを到達するには実際的にさまざまな困難をとまなう。本稿では、むしろそうした問題点も明示しながら、SSM職業小分類を国際的に標準的に使用されている階層分

⁷ 1965年から1985年までの男性回答者で、従業上地位が従業員のいない単独自営だった比率は20.1, 19.2, 15.7%、従業員のいる自営業者の比率は8.6, 6.6, 6.1%であった。1995年と2005年は、従業員の有無を区別しない自営業者の比率は16.9, 13.6%であった。

⁸ 第2段階の補正処理は、SSM職業小分類コードからする場合とまったく同一の処理をおこなっている。注1も参照されたい。

類と地位指標になるべく近づけることによって、比較可能性をより担保しようとする試みをおこなった。この試みによって SSM 職業小分類から EGP 分類、SIOPS、ISEI への変換が簡便になり、日本を含む国際比較研究が活発になることが望まれる。

文献

- Erikson, R., J. H. Goldthorpe and L. Portocarero. 1979. "Intergenerational Class Mobility in Three Western European Societies: England, France and Sweden." *British Journal of Sociology* 30:415-41.
- Erikson, Robert and John H. Goldthorpe. 1992. *The Constant Flux: A Study of Class Mobility in Industrial Societies*. Oxford: Clarendon Press.
- Ganzeboom, Harry B.G. and Donald Treiman. 1996. "Internationally Comparative Measures of Occupational Status for 1988 International Standard Classification of Occupations." *Social Science Research* 25:201-39.
- Ganzeboom, Harry B.G. 2007 年 12 月現在. "Tools for deriving status measures from ISKO-88 and ISCO-68." <http://homes.fsw.vu.nl/%7Eganzeboom/pisa/index.htm>.
- 田辺俊介. 2006. 「ISCO と SSM 職業分類の相違点の検討ー国際比較調査における職業データに関する研究ノートー」『社会学論考』第 27 号: 53-78.
- 2005 年社会階層と社会移動調査研究会. 2007. 『2005 年 SSM 日本調査 コードブック』

**SSM Occupational Code and International Measures of Occupational Status:
Conversion into EGP Class Schema, SIOPS and ISEI**

Nobuo KANOMATA

Keio University

Shunsuke TANABE

The University of Tokyo

Hirohisa TAKENOSHITA

Shizuoka University

The data of Social Stratification and Social Mobility (SSM) Survey in Japan 2005 contains SSM occupational code as well as ISCO-88 code. The latter would be easily transformed into EGP class schema, SIOPS and ISEI which have been employed frequently in cross-national studies. However, the Japanese data of SSM Survey collected before 2005 did not include ISCO-88 but only SSM code which has brought about difficulty with regard to constructing internationally popular classification and index of occupations because of its peculiar classification. This paper shows correspondence between SSM code and ISCO-88 and presents how to construct EGP class schema, SIOPS and ISEI from SSM code. Since the data collected in 1955 lacks the information on job-ranking, the method of conversion to EGP class schema to be presented can be applicable from the second (in 1965) to the sixth (in 2005) survey data.

Keywords and phrases: SSM occupational code, ISCO-88, EGP class schema, SIOPS, ISEI

職業評定の日韓比較

元治恵子
(立教大学)

【要旨】

本稿では、日本で実施された「職業に関する全国調査（職業威信）」と韓国で実施された「2004年韓国職業威信調査」のデータを使い、日本と韓国における職業評定構造の違いを検討した。分析の結果、職業威信スコアにより、両国の職業評定構造を比較すると、相対的な順位が異なる職業もあるものの、相関係数レベルで見れば、非常に類似した構造であることが明らかになった。また、INDSCAL やクラスター分析の結果からも、両国には類似した職業評定構造があることが明らかになった。一方、評定基準では、相対的に重視されている項目とされていない項目には、類似した傾向が見られたが、基準間の関連の仕方が両国では異なることが確認された。また、「人々にとって重要なこと」との関連を見ると、関連の仕方や程度の違いはあるものの、日韓ともに、人々が重要だと考えていることと職業評定の際に重視する基準は、密接に関連していることが明らかになった。

キーワード：職業評定の構造、職業威信スコア、職業評定基準

1. 問題の所在

職業（的地位）は、社会階層や社会移動の研究において、人々の社会的（階層的）地位を表わす指標として利用されてきた。この背景には、職業が、職種や従業上の地位に現れる技能的水準、権限としての勢力水準、就業資格としての学歴水準、そして、職業の格付けという社会的評価を含む認定的指標であると考えられていることがある（Reiss, Duncan, Hatt, & North 1977）。人々の職業的地位を、量的に扱う場合には、職業威信スコアを各々の職業に割り当て、分析に用いる¹。このスコアは、人々の職業に対する評定（評価）から、職業ごとに平均的な評定スコアを割り振ることにより得られる²。これまでの研究では、時代（元治・都築 1998; Hauser 1982; Hodge, Siegel, and Rossi 1964; Nakao and Treas 1994）、評定者の社会的属性（Bose and Rossi 1983; 元治・都築 1998; Siegel 1970）、産業化や国民総生産の異なる国々（Treiman 1975, 1977）の差異によらず、職業の相対的な序列構造に関し

¹ 職業の社会経済的地位指標（SEI）（Blau and Duncan 1967）が利用される場合も多い。この指標は、当該職業従事者の所得や学歴から推定されるものである。社会移動の研究においては、この指標の方が職業威信スコアよりも有効な指標であるという議論もある（Featherman and Hauser 1976; Nakao 1992）。

² 人々の職業に対する5段階の評定に対し、カテゴリーごとに等間隔の評定値を与え、職業ごとに平均評定値を算出し、その値を職業威信スコアとする。わが国では、1955年以来伝統的に、順位づけられた5つのカテゴリーに、最上位100点、最低位0点として、25点間隔の評定値を与えてきている（直井・鈴木, 1977）。

て、かなり高い相関が見られ、安定した有効な尺度であることが、おおかたの見方となっている (Bian 1996; Blau and Duncan 1967; Hauser and Featherman 1977; Hodge, Treiman, and Rossi 1966; 直井・鈴木 1977; Treiman and Yip 1990)。

日本では、職業威信に関する調査は、1955 年、1975 年、1995 年の 3 回実施されているが、それぞれの調査データに基づく分析、および、時点間の比較分析にとどまり、他の国の調査結果との比較分析は十分に行われてこなかった (直井・鈴木 1977; 都築編 1998 など)。また、韓国においても、有田 (2005) によれば、職業威信に関する研究は少なく (李 1957; 李・金 1966; 金 1970, 1979; 石 1992 など)、国際比較の面でも、いまだに多くの検討課題が残されたままになっているという。

これまでの、日韓の職業評定および職業威信スコアに関する研究から、職業威信スコアのヒエラルキーにおいては、非常に高い相関が見られる (有田 2005; 中尾 2005) ことが明らかになっている。しかし、各職業の威信スコアの絶対的水準では、大きな相違がある (有田 2005) という。また、Nakao (2006) は、職業威信評定において、韓国と日本の評定者に共有されている部分は大きい、それぞれの国に固有の違いも見られ、性、年齢、教育歴、職業などのサブグループでの違いに比べ大きいと指摘している。そして、Yoo (2006) による中国を含めた職業構造や職業威信序列の 3 国間の比較研究においては、3 つの国の職業威信序列は類似しており、時点の違いによる変動も大きくないが、韓国の職業威信スコアの幅が大きい (韓国 > 日本 > 中国) ことが明らかになっている。これらを踏まえると、日韓の職業評定や職業威信スコアにおいては、かなり共通する部分が多いと言える。しかし、これまでの研究では、比較できる職業の数も少なく、その影響を受けた可能性も否定できない³。

本稿では、「1995 年社会階層と社会移動調査」の「職業に関する全国調査 (職業威信)」と「2004 年韓国職業威信調査」のデータを用いて、職業評定構造の日韓比較分析を行い、それぞれの国に固有の構造が浮かび上がってくるのか、それとも、共通の構造が見出せるのかを検討する。また、職業評定構造の背景にある要因についても検討も行う。

2. データ

日本の職業評定に関する分析には、「1995 年社会階層と社会移動調査」の「職業に関する全国調査 (職業威信)」のデータを使用した。この調査は、1994 年 12 月末現在満 20 歳から 69 歳の男女を対象として、個別面接調査により、1995 年 10 月から 11 月に実施された。有効サンプル数は 1214 サンプル (回収率 72.5%) であった⁴。一方、韓国の職業評定に関する分析には、「2004 年韓国職業威信調査」のデータを使用した。この調査は、2003 年 12 月末現在

³ ただし、Nakao(2006)では、本稿と同じデータを使用している。

⁴ 詳細は、『1995 年 SSM 調査 コードブック』1995 年 SSM 調査研究会 (編) を参照のこと。

満 20 歳から 69 歳の男女を対象として、個別面接調査により、2004 年 2 月から 3 月に実施された。有効サンプル数は 903 サンプル（回収率 45.2%）であった⁵。分析の中心となる変数は、両国とも職業威信スコアを算出するもととなる共通する 55 の職業の社会的地位に関する質問（注 2 参照）と職業を評定する際に使用した基準についての質問への回答である。

ここで、留意しなければならないのは、使用するデータは、日本が 1995 年、韓国が 2004 年に実施された調査から得られたものであり、およそ 10 年の開きがあることだ。職業を評定する際に、調査時点の産業化レベルや産業構造が異なることが影響を与える可能性もある⁶。職業分布は、産業化レベルや産業構造を示す指標の一つと考えられる。両国の調査時点の職業分布を比較して見ると、非常に類似した分布構造が見られ（図 1）、同じ調査年のデータを比較するよりも、むしろ、両国間の職業評定構造の純粋な差異を明らかにすることができるのではないだろうか⁷。

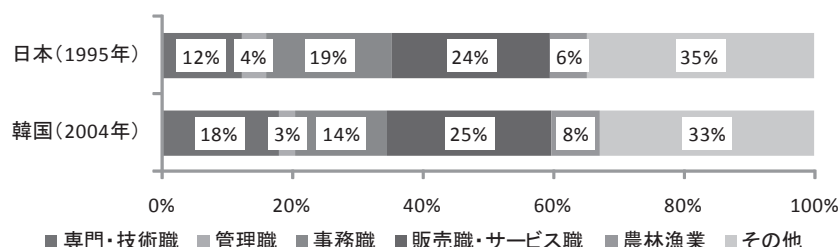


図 1 職業分布の日韓比較⁸

出所：ILO “LABORSTA” (Online Database)

3. 分析結果

3.1 職業威信スコアの序列構造の日韓比較

⁵ 詳細は、『2004 年韓国・職業に関する全国調査（職業威信）コードブック・基礎集計表（修正版）』2003 年階層調査研究会（編）を参照のこと。

⁶ 前述したように、これまでの研究（Treiman 1975, 1977）では、産業化の程度による違いは、大きくないことが明らかになっているが、その可能性は否定できない。

⁷ 有田（2006）も、対象職業が異なるため、厳密な比較はできないという限定つきながら、韓国での 1990 年衡平調査の結果は、日本における 1995 年 SSM 調査の結果よりも 1975 年 SSM 調査の結果の方が類似している可能性を指摘している。相関係数レベルで見れば、前者とは 0.954、後者とは 0.979 であり、わずかな差はあるが、1975 年 SSM 調査との関連が強い傾向が見られている。

⁸ 日本は ISCO-68、韓国は ISCO-88 による分類のため、カテゴリーを結合するなど再集計したものである。厳密な意味では、若干ずれがあるが、大枠を捉えるうえでは、支障がないと考える。

表1 職業威信スコア

SSM 大分類	職業	韓国			日本			順位差 (韓国 -日本)	スコア差 (韓国 -日本)
		順位	威信スコア	SD	順位	威信スコア	SD		
専門	判事(裁判官)	1	86.1	19.5	3	86.9	16.9	-2	-0.9
専門	医師	2	83.9	19.2	1	90.1	14.8	1	-6.2 **
管理	大企業の社長	3	83.4	22.9	2	87.3	16.7	1	-4.0 **
専門	大学教授	4	80.4	19.0	4	84.3	17.1	0	-3.9 **
専門	飛行機の操縦士(パイロット)	5	75.7	19.4	5	82.5	17.9	0	-6.7 **
管理	官公庁の局長(高級官僚)	6	74.2	18.1	6	77.5	21.2	0	-3.2 **
管理	中小企業の経営者	7	73.5	19.1	12	68.9	15.8	-5	4.6 **
専門	薬剤師	8	69.7	17.6	15	65.7	15.1	-7	4.0 **
専門	建築士	9	67.8	19.6	8	72.0	16.2	1	-4.1 **
管理	国会議員	10	67.4	31.0	7	74.9	22.7	3	-7.5 **
専門	プロスポーツの選手	11	67.4	19.9	11	69.0	19.8	0	-1.6
専門	自動車設計技術者	12	66.3	19.4	14	66.3	15.7	-2	0.0
専門	公認会計士	13	65.8	20.6	9	70.8	16.1	4	-5.0 **
販売	スチュワーデス	14	64.4	19.6	10	70.0	17.2	4	-5.6 **
専門	音楽家	15	64.1	19.2	13	66.6	16.0	2	-2.5 **
管理	大企業(大会社)の課長	16	62.9	15.7	18	63.2	14.6	-2	-0.3
専門	小学校の教師(教諭)	17	62.7	17.3	17	63.6	15.6	0	-0.9
管理	市役所の課長	18	62.6	17.6	23	56.9	13.3	-5	5.7 **
専門	衣装(服飾)デザイナー	19	62.0	17.3	16	64.6	15.3	3	-2.5 **
熟練	伝統工芸の職人(家具職人)	20	60.6	22.9	30	52.1	13.5	-10	8.5 **
事務(販売)	大企業の事務職(営業社員)	21	59.6	16.7	22	57.4	14.2	-1	2.2 **
管理	中小企業の課長	22	57.0	15.5	26	56.1	12.9	-4	0.9
事務	銀行員	23	55.6	15.6	25	56.4	12.9	-2	-0.8
事務	警察官	24	55.5	18.9	21	57.9	15.3	3	-2.4 **
熟練	土木・建築の現場監督	25	54.6	17.1	24	56.7	13.7	1	-2.1 **
専門	看護師	26	54.0	18.1	20	59.7	17.0	6	-5.7 **
専門	宗教人(寺の住職)	27	51.4	23.0	19	60.3	18.2	8	-8.9 **
半熟練	地下鉄の運転技師(電車運転士)	28	51.4	17.6	32	51.3	10.7	-4	0.1
専門	幼稚園の先生(保母)	29	50.8	17.1	28	52.9	13.0	1	-2.1 **
事務	中小企業の事務職(事務員)	30	50.4	13.8	41	47.0	10.6	-11	3.5 **
熟練	レストランの料理人(コック)	31	49.9	18.6	31	51.6	13.3	0	-1.7 **
販売	卸売商の主人(卸売店主)	32	49.8	17.1	29	52.9	11.1	3	-3.1 **
事務	銀行の窓口係	33	49.2	15.6	37	49.4	10.0	-4	-0.1
熟練	大企業の機械組立工	34	48.8	20.1	34	51.1	14.3	0	-2.3 **
事務	鉄道の駅務員(鉄道の駅員)	35	46.2	18.5	39	47.8	11.0	-4	-1.6 **
熟練	電気工事者(電気工事人)	36	46.0	19.7	35	50.4	13.1	1	-4.4 **
熟練	美容師(理容師・理髪師)	37	45.9	19.2	36	49.7	11.5	1	-3.8 **
熟練	中小企業の機械組立工	38	45.2	19.4	43	46.7	14.5	-5	-1.5 *
販売	小売商の主人(小売店主)	39	43.0	17.8	33	51.3	11.2	6	-8.3 **
販売	自動車のセールスマン	40	42.7	17.6	40	47.2	11.9	0	-4.6 **
事務	郵便集配員(郵便配達人)	41	41.1	19.5	45	46.2	12.0	-4	-5.1 **
熟練	自動車修理工	42	40.7	20.4	42	46.8	11.9	0	-6.1 **
販売	保険設計士(外販員)(保険の勧誘員)	43	38.4	19.7	47	44.3	13.8	-4	-5.9 **
熟練	大工	44	37.2	23.1	27	53.1	15.4	17	-15.9 **
半熟練	かんづめの製造工(食品かんづめ工)	45	36.7	18.9	50	42.2	13.7	-5	-5.5 **
半熟練	バスの運転手(バス運転手)	46	35.6	20.5	38	48.9	11.5	8	-13.3 **
販売	商店の店員	47	30.8	19.8	49	42.4	14.0	-2	-11.6 **
農林	農業	48	29.7	24.1	46	45.6	17.6	2	-15.9 **
非熟練	道路工事の人夫(道路工夫)	49	29.0	21.9	53	39.0	17.6	-4	-10.0 **
熟練	印刷工	50	28.8	20.8	48	44.0	13.7	2	-15.3 **
半熟練	紡績工	51	28.1	20.3	51	42.0	15.1	0	-13.9 **
農林	漁師(漁業者)	52	26.2	21.7	44	46.5	16.2	8	-20.3 **
販売	ウェイトレス	53	25.7	20.9	54	38.1	15.5	-1	-12.3 **
事務	守衛	54	23.8	21.2	52	39.9	16.1	2	-16.1 **
非熟練	炭鉱夫	55	21.8	21.8	55	36.7	19.0	0	-14.8 **

まず、両国の職業威信スコアの全体的傾向と基本的な特徴を見てみよう（表1）。表には示していないが、これまでの知見と同様（有田 2005；中尾 2005；Nakao 2006）、両国の職業威信スコアの相関係数は0.940と高く、全体的な職業の序列構造としては、非常に類似した

構造であると言える。個々の国別の特徴を見ると、日本における職業威信スコアは、最高値は、医師の90.1（韓国では2位）、最低値は、炭鉱夫の36.7であり、53.4ポイントの範囲に職業が分布している。また、職業威信スコアの平均値は57.1、標準偏差は13.6であった。一方、韓国における職業威信スコアは、最高値が、判事の86.1（日本では3位）、最低値は、日本と同じく、炭鉱夫の21.8であり、64.3ポイントの範囲に職業が分布している。また、職業威信スコアの平均値は52.4、標準偏差は16.7であった。これらのことから両国における職業威信スコアの全体的傾向を概観すれば、相関係数レベルでは、非常に類似した傾向が見られるが、最上位の職業では違いが見られ、最下位の職業は同じであった。また、韓国の方が広範囲に分布しており、平均値は低く、ばらつきが大きいという違いが見られた。

続いて、個々の職業に見られる違いに注目してみよう。職業威信スコアの比較、とく

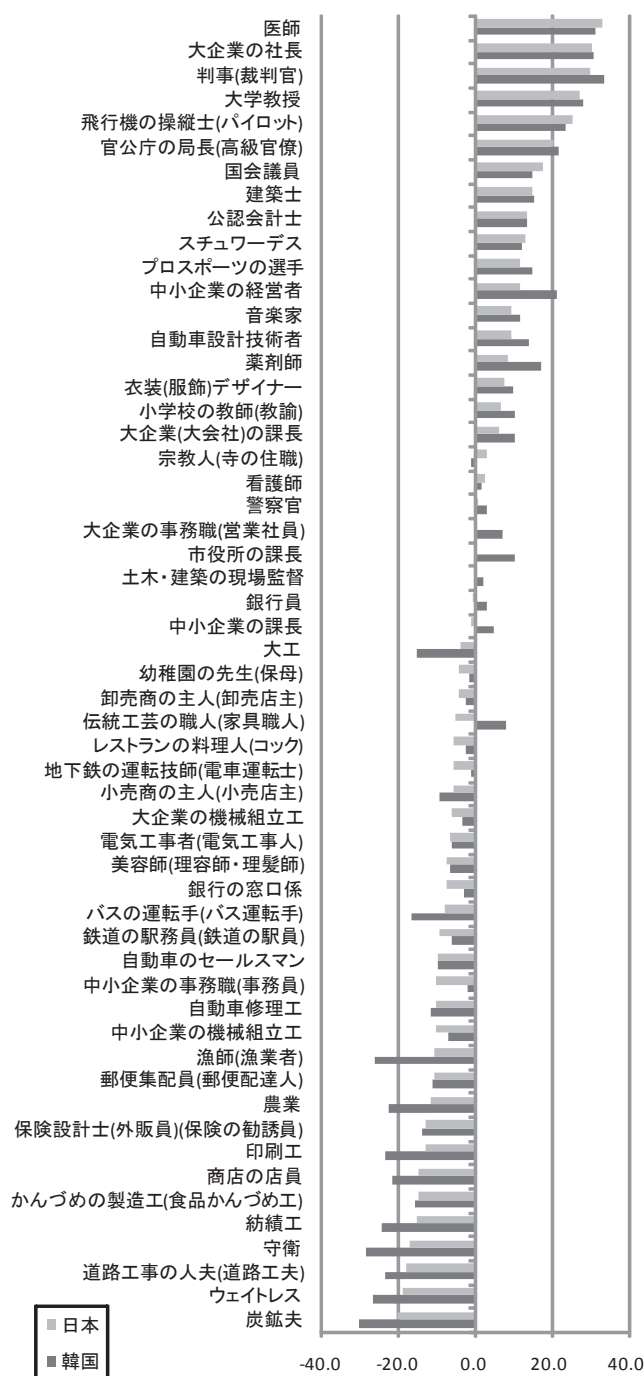


図2 補正スコアの日韓比較

に国際比較を行う場合には、調査における質問に対する回答の国民性による影響を考慮する必要がある。それゆえ、構造の類似性を相関係数から判断したり、個々の職業のスコアを比較する場合でも、当該国のスコアの平均値により補正した補正スコアを用いて比較するなどが行われてきた（園田 2005）。日本では中間回答の選択率が高い傾向や極端な表現を好まない傾向が見られるという（林 1995, 2001）。このことが、職業威信スコアの範囲や標準偏差に影響を与えた可能性も考えられる。そこで、回答の国民性による影響をコントロールするという観点から、それぞれの国の職業威信スコアの平均値を、個々の職業威信スコアから引いた補正スコアに注目して見ていく（図2）。韓国の方が高く評定されているのは22職業、低く評定されているのは職業23である。両国の評定で10ポイント以上の開きがある職業のうち、韓国の方が高く評定されているものは、市役所の課長（13.2ポイント差：以下、職業の後の数値はポイント差）と伝統工芸の職人⁹（10.5）の2つのみであるのに対し、低く評定されているものは、漁師（15.5）、守衛（11.4）、大工（11.2）、農業（11.2）、印刷工（10.5）、炭鉱夫（10.1）の6職業ある。韓国の方が低い職業のうち、日本の大工¹⁰を除いては、両国において職業評定構造の下位に位置づけられる職業が多いが、この下位層での差が大きいことがわかる。つまり、下位に位置づけられる職業は、日本よりも韓国において、より低く評定される傾向が見られる。

3.2 職業威信スコアに対する評定者の属性効果

それでは、評定者の属性は、職業評定に影響を及ぼすのであろうか。職業威信スコアの属性効果を相関係数から検討してみよう。まず、男女間で相関係数を見ると、両国とも0.996と極めて高い。両国の男性間の相関係数は0.936、女性間の相関係数は0.941と、女性間の方が高いものの、それぞれの国における男女間の相関係数よりも、両国の同性間の相関係数の方が低い傾向が見られる。次に年齢を10歳ごとにカテゴリー化して、そのカテゴリー間の相関係数を見ると、両国ともに30代と40代の間の相関が最も高い（韓国 $r=0.997$ 、日本 $r=0.996$ ）。一方、最も低いのは、やはり両国ともに20代と60代の間の相関係数である（韓国 $r=0.960$ 、日本 $r=0.965$ ）。また、両国の同世代間¹¹について見ると、日韓それぞれの国における世代間の相関係数よりも低い相関係数となることが確認された。さらに、職業¹²による効果を見たところ、日本では、専門職と事務職の間の相関係数が最も高く（ $r=0.995$ ）、管理と非熟練の間の相関係数が最も低い（ $r=0.957$ ）。韓国では、販売職と熟練の間の相関係数が最も高く（ $r=0.989$ ）、最も低いのは、日本と同様、管理と非熟練の間の相関係数である（ r

⁹ 1995SSM調査では「家具職人」

¹⁰ 韓国における大工の相対的低さについて、有田（2005）は、韓国における熟練技能に対する評価の低さに起因するとしつつも、日本における相対的位置の高さに原因がある可能性を指摘している。

¹¹ 調査時点を無視した同世代と10年ずらした世代間の両方の相関係数を算出し、比較検討した。

¹² SSM8分類により、評定者を分類し、それぞれのカテゴリー間の相関係数を算出した。

=0.929)。両国の同じ職業カテゴリー間で見ると、熟練間の 0.897 から専門職間の 0.943 の間であり、同国内異職業カテゴリー間よりも、全般的に低い傾向が確認された。以上のように、相関係数レベルで見れば、どのカテゴリー間も相関が高く、評定者の属性効果があるとは言いつ切れない。全般的には、同国内における異なるカテゴリー間の方が、両国の同じカテゴリー間の相関よりも高く、評定者の属性効果よりも、国による違いの方が職業威信評定に影響を及ぼしている可能性が示唆された。

3.3 職業の評定構造の日韓比較

これまでの分析から、両国の職業評定構造は、相関係数レベルで見れば、非常に類似した傾向があることが確認された。しかし、相対的な職業の序列構造には違いも見られ、個々の職業では、大きく評定の異なる職業もあった。また、評定者の属性効果よりも国による違いの大きいことが示唆された。そこで、両国に共通の構造を仮定し、その上での違いを検討する。国ごとに 55 の職業に対する評定から職業間のユークリッド距離行列を算出し、INDSCAL（個人差ユークリッド距離モデル）により分析した。データに対する布置の適合度を示す VAF 比およびその変化や共通対象布置、重み布置などの解釈により、解の次元数は 2 次元を採用することにした¹³。図 3 は、2 次元の共通対象布置を示したものである¹⁴。

共通対象布置および職業威信スコアと布置における職業の各座標との相関係数などから次元を解釈していこう。両次元ともに正（右上）から負（左下）の方向へと高職業威信スコアの職業から低職業威信スコアの職業が並んでいる。次元 1 は、韓国の職業威信スコアとの相関が高く（ $r=0.998$ ）、『韓国の職業評定における職業序列』と解釈できる。一方、次元 2 は、日本の職業威信スコアとの相関が高く（ $r=0.997$ ）、『日本の職業評定における職業序列』と解釈できる。韓国における職業評定では、次元 1 方向で正の座標をもつ職業の評定は高くなり、負の座標をもつ職業の評定は低くなり、座標間の距離の大きい職業間の違いは強調され、職業間の評定の違いがより大きくなると考えられる。逆に次元 2 方向で、座標間の距離の小さい職業間の違いは強調され、より類似した評定になると考えられる。一方、日本では、この逆の解釈ができる。つまり、次元 1 が正、次元 2 が負の象限に位置するような職業は、韓国では、より高い評定、日本ではその逆となり、両国における職業評定に差が見られることになる。前節の補正スコアの分析から、市役所の課長と伝統工芸の職人は、韓国の方が 10 ポイント以上高く評定されていたが、この 2 つの職業などは、その典型と言えよう。

しかし、次元 1 と次元 2 の間の相関も高く（ $r=0.915$ ）、INDSCAL の結果からも両国の職業

¹³ VAF 比の変化は、3 次元解から 1 次元解で順に 0.988→0.985→0.936 であった。

¹⁴ 図の左上および右下の数値は、各国の次元に対する重みを示している。また、矢印は次元の方向を示す。つまり、それぞれの国の個人布置は、日本では、共通対象布置を次元 2 の方向が長い縦長の布置になり、韓国では、共通対象布置を次元 1 の方向が長い横長の布置になる。

評定の構造が非常に類似したものであることに変わりはない¹⁵。

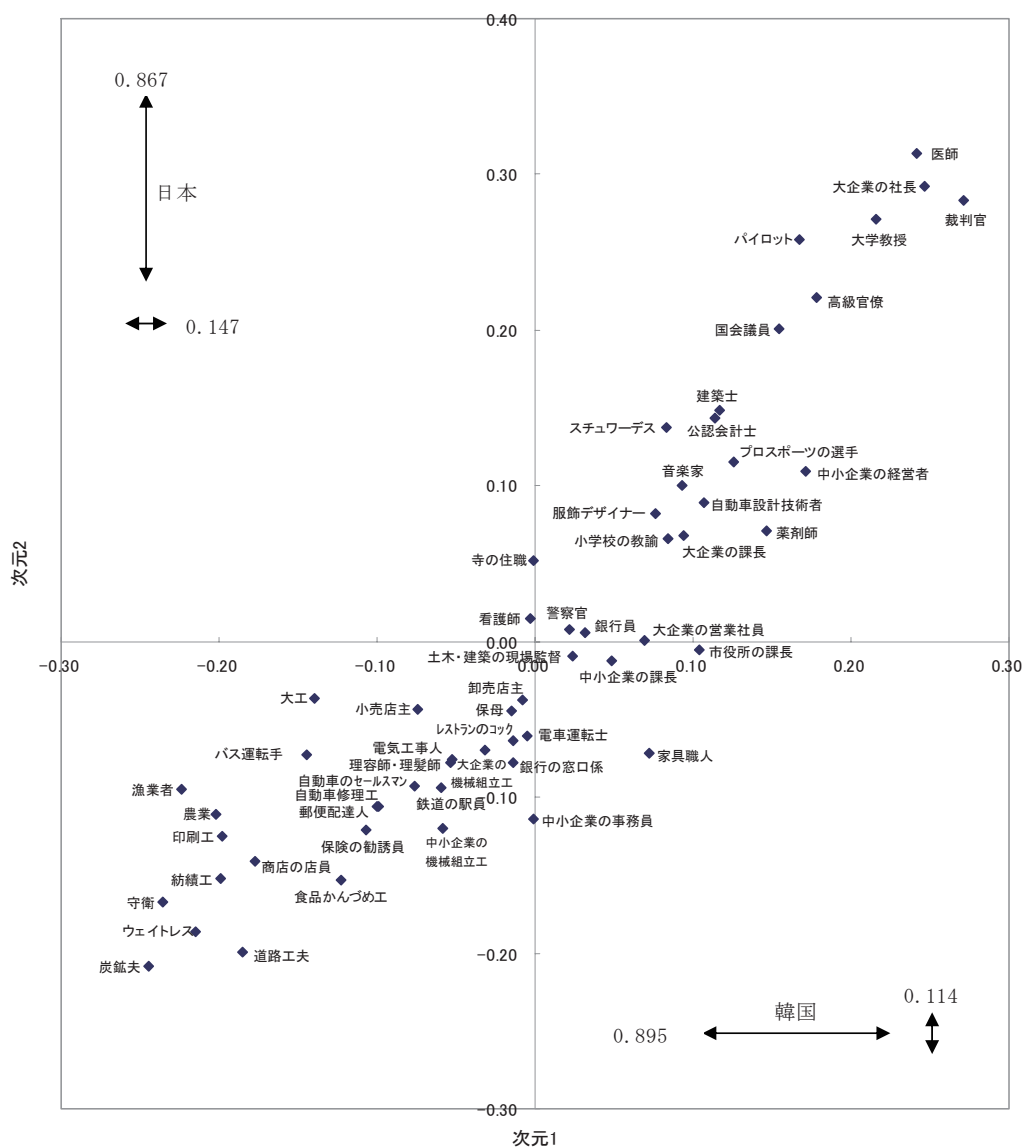


図3 INDSCAL 分析結果：共通対象布置（2次元）および両国の各次元に対する重み

¹⁵ 分析結果は示さないが、ウォード法によるクラスター分析の結果、両国とも大きくは高職業威信クラスターと低職業威信クラスターに分類され、それぞれのクラスターに所属する職業も4つの職業を除いて同じであった。

表2 職業威信スコア、INDSCAL分析の2次元布置職業座標間のピアソンの積率相関係数

	韓国	次元1	次元2
日本	0.940	0.925	0.997
韓国		0.998	0.929
次元1			0.915

3.4 職業評定基準の基本的特性の日韓比較

それでは、人々は、職業を評定する際にどのような基準を重視しているのだろうか。評定基準に対する回答を、「非常に重視した」を4点、「やや重視した」を3点、「あまり重視しなかった」を2点、「全く重視しなかった」を1点とし得点化して平均値を見ると、どの基準も韓国が高く、有意な差が見られる（図4）。先ほどの職業威信スコアと同様、国の違いによる回答バイアスを取り除くため、補正スコアを算出し比較すると、両国とも「責任」、「技能」の2基準は相対的に重視され、「権力」、「格好よさ」の2基準は相対的に重視されていない。相対的に重視されている項目とされていない項目は、類似した傾向が見られる。しかし、「創造力」、「自律性」、「教育」の3基準は、職業評定の際に重視する基準としては、日韓で異なった評価のされ方をしており、特に「教育」での重視度が異なっていることがわかる（図5）。また、韓国よりも日本の方が、評定基準の重視度にばらつきが見られる。

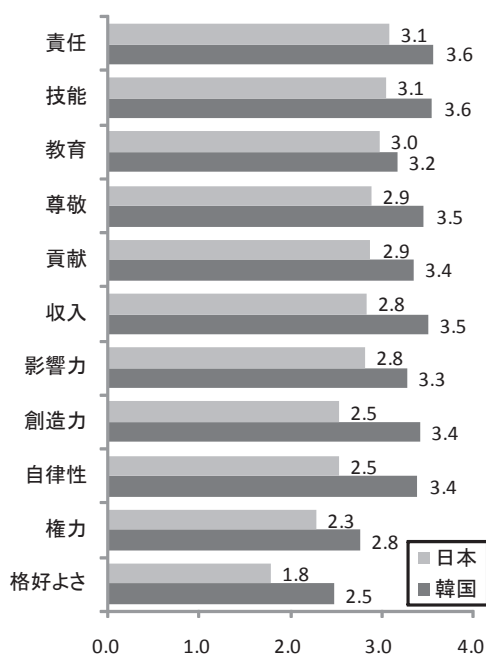


図4 評定基準平均値

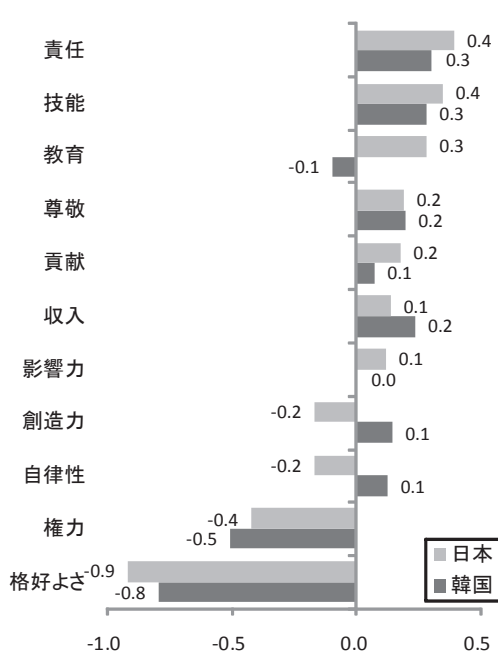


図5 評定基準補正スコア

3.5 評価基準の構造の日韓比較

人々は、職業評価をする際に、何らかの基準に従って職業を評価していると考えられる。ある1つの評価基準のみを使って評価している場合は少ない。人によって、重視する評価基準の組み合わせや各々の基準の重視度が違ったり、同じ人であっても評価する職業によっても異なるだろう。ここでは、データの制約上および、本稿の目的とはずれるので、個人に注目することはしないが、国ごとに職業評価基準間の関係がどのようなになっているのかについて見ていく。3.3節と同様に INDSCAL によりその構造を探った。分析の結果2次元解が適当と考えられ、共通対象布置と重み布置を得た。共通対象布置の各次元に対する重みは、韓国では、次元1が0.980、次元2が0.141、日本では、次元1が0.819、次元2が0.547であった。次元2への重みが、韓国では、0.141と小さく、ほぼ1次元でとらえられていると考えられる。共通対象布置に重みをかけ、国別の布置を示したものが図6と図7である。

次元1は、「技能」、「責任」、「自律性」、「創造性」が正の方向で大きな値を示し、「権力」と「格好よさ」が負の方向で大きな値を示していることから、『職業特性（仕事内容）』と解釈できるだろう。一方、次元2は、「教育」と「収入」が正の方向で大きな値を示し、「格好よさ」が負の方向で大きな値を示しており、『社会的資源』と解釈できるだろう。それなりの職業（的地位）に就くためには、学歴は重要な役割を果たし、また、高学歴によって就いた職業は、高収入を得ることが保証されるという、学歴・職業・収入の3つの関係が人々のなかに強く共有されていると考えることができるだろう。

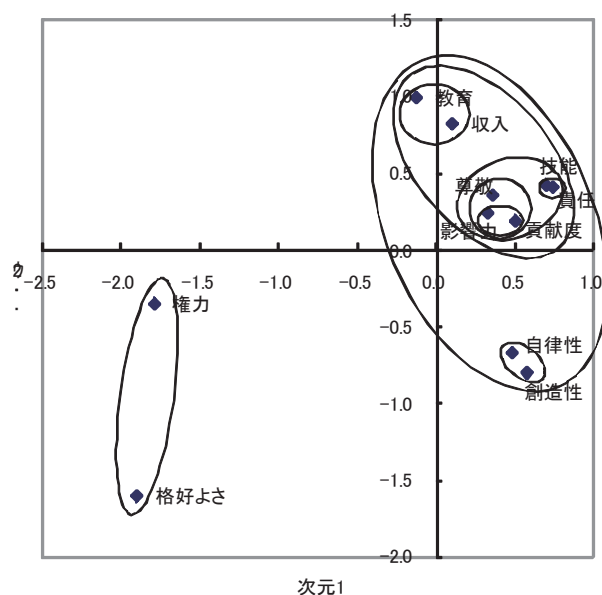


図6 INDSCAL 分析の布置とクラスター凝集過程（日本）

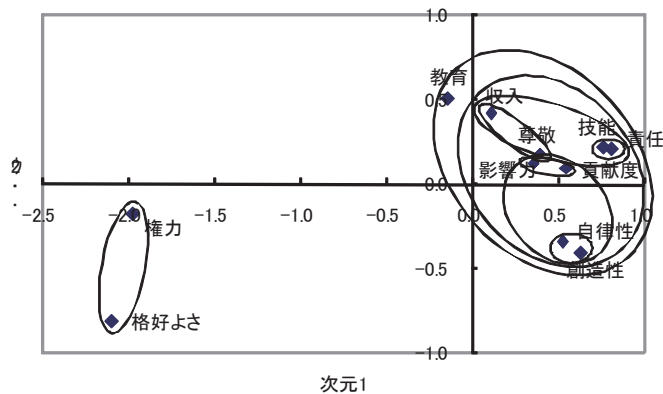


図7 INDSCAL 分析の布置とクラスター凝集過程（韓国）

上記の分析と並行してウォード法によるクラスター分析も行った。図に示してある円は、クラスター凝集過程を示したものである。クラスター分析の結果から、両国とも2つのクラスターが抽出された。これらの2つのクラスターに分類される評価基準は共通しており、「格好よさ」と「権力」のクラスターとそれ以外の基準のクラスターである。しかし、その凝集過程に注目すると、違いも見られる。「格好よさ」と「権力」以外の基準のクラスターの中で、日本では、「創造性」と「自律性」は、最終段階で他の基準と結合されるのに対し、韓国では、早い段階で、「社会に対する貢献」や「社会に対する影響力」と結合している。これは、両国では、仕事に対する考え方が異なることを示している可能性がある。つまり、韓国では、人々は、仕事の自律性が高く、その仕事に対し創造力を発揮することにより、社会に対し貢献し、影響力をもつことができると考えている。一方、日本では、自律性や創造力は、社会に対する貢献や影響力とは直接結びにくく、むしろ、社会に対する貢献や影響力は、社会からの尊敬の大きさと関連が強いと考えられている。また、「教育（学歴）」に注目すると、日本は、割と早い段階で「収入」と結合されるのに対し、韓国では、抽出された2つのクラスターのうちの1つのクラスター内において、最終段階で結合している。このことは、INDSCALの結果において学歴・職業・収入の3つの強い関連が示唆されたものの、両国では、職業に対しての「教育」のもつ意味が若干異なる可能性があることを示唆している。両国の大学進学率¹⁶を見ると、韓国¹⁷では、2004年時点で98.8%（男子98.9%、女子98.8%）であるのに対し、日本¹⁸では、1995年時点で45.2%（男子42.9%、女子47.6%）（文部科学省 2006）と大き

¹⁶ 進学率＝高等教育機関入学人数／該当年齢人口

¹⁷ 元の資料出所：「教育統計年報」韓国教育人的資源部・韓国教育開発院。大学、教育大学、専門大学、産業大学、技術大学への進学率。

¹⁸ 元の資料出所：「学校基本調査報告書」文部科学省。大学（学部）・短期大学（本科）への進学率（浪人を含む）。

な差がある。国によって高等教育卒業資格のもつ意味が異なるが、このような大学進学率の差が、職業を評定する際の「教育」基準の違いに影響を与えた可能性が考えられる。

3.6 職業評定基準と「重要なこと」

これまでの分析から、職業の序列構造においては、非常に類似している構造をもつことが明らかになったが、職業を評定する際に重視する評定基準に関しては、両国では違いも見られた。このような違いはどのような要因が背景にあるのだろうか。人々は、ある職業について何らかのイメージをもっていることも少なくない。そして、ある職業よりも別の職業に対し、良いイメージをもち、無意識のうちに、序列づけている可能性もある。調査において、職業評定の際に重視する（した）評定基準を問われれば、人々が日ごろ無意識のうちに序列づけしている際に重視している基準が顕在化する（あらためて認識する）可能性はある。しかし、それは、世間一般では、人々が職業に対してこんなイメージをもっているという考えのもと、評定基準の重視度を回答している可能性も否定できない。そこで、人々が日ごろの生活で重要だと認識している項目と職業評定基準との関連を見ることにより、職業評定の背景にある要因、言わば本音のようなものを正準相関分析を用いて探ることにする。

分析の結果、日本では第3正準相関、韓国では第6正準相関までが有意（ $p < .01$ ）であったが、正準相関係数が0.3以上であり、単独説明率が10%以上である正準相関を採択し、日本は第2正準相関まで、韓国は第3正準相関までを検討する。表3は各正準相関と元変数群との構造係数を示したものである。

日本では、第1正準相関に対する構造係数は、両変数群ともすべて正である。職業評定基準変数群では「教育」と「権力」、重要項目変数群では「高学歴」と「社会的評価」が特に大きな値を示している。高い学歴をもつことや社会的評価の高い職業に就くことを重要と考えている人は、職業を評定する際にも教育や権力を重視している。第2正準相関に対する構造係数は、職業評定基準変数群では「格好よさ」と「権力」が負で、それ以外が正、重要項目変数群では「高収入」、「高地位」、「多くの財産」が負でそれ以外が正の値を示している。特に大きな値を示しているのは、職業評定基準変数群では「責任」、重要項目変数群では「社会活動に参加」である。社会活動に参加することや家族との関係を重視し、収入、財産、地位などを重要とは考えていない人は、仕事の特性を基準に職業を評定する傾向が強いと言える。

一方、韓国では、第1正準相関に対する構造係数は、両変数群ともすべて正である。職業評定基準変数群では「教育」と「権力」、重要項目変数群では「高学歴」と「高地位」が特に大きな値を示している。高い学歴や高い地位に就くことを重要と考えている人は、職業を評定する際にも、教育や権力を重視している。また、第2正準相関に対する構造係数で大きな値を示しているのは、職業評定基準変数群では「貢献」、「影響力」、「創造性」、「自律性」、重要項目変数群では「社会活動に参加」と「家族からの信頼と尊敬」である。この正準相関は、

日本の第2正準相関と非常に類似しており、同じように解釈できるだろう。第3正準相関に対する構造係数では、特に大きな構造係数はないが、職業評定基準変数群では「技能」、「責任」、「収入」、重要項目変数群では「高収入」、「多くの財産」などがやや大きな値を示している。この正準相関に対する構造係数で特徴的なのは、重要項目変数群の学歴と職業評定基準変数群の高学歴が負の値であることである。日ごろ学歴を重要と考えていない人は、職業評定においても、学歴基準は重視していないと言える。

表3 正準相関分析の結果

	日本		韓国		
	第1正準相関	第2正準相関	第1正準相関	第2正準相関	第3正準相関
職業評定基準					
教育	0.803	0.203	0.763	0.366	-0.179
技能	0.337	0.453	0.102	0.362	0.516
責任	0.186	0.657	-0.087	0.374	0.561
収入	0.555	0.022	0.370	-0.044	0.527
尊敬	0.616	0.309	0.392	0.274	0.307
貢献	0.207	0.590	0.211	0.656	0.115
影響力	0.294	0.592	0.358	0.565	0.129
創造性	0.179	0.591	-0.080	0.616	0.387
自律性	0.258	0.621	-0.144	0.596	0.418
格好よさ	0.552	-0.347	0.417	0.087	0.457
権力	0.717	-0.221	0.768	-0.236	0.343
(冗長性)	0.292	0.020	0.064	0.037	0.017
重要項目					
社会的評価が高い職業に就く	0.738	0.216	0.477	0.486	0.069
高い収入を得る	0.616	-0.230	0.511	-0.029	0.553
高い学歴をもつ	0.888	0.217	0.846	0.350	-0.145
家族からの信頼と尊敬	0.029	0.455	0.030	0.669	0.405
社会活動に参加	0.036	0.775	-0.103	0.757	0.251
趣味や余暇活動など	0.316	0.385	0.138	0.228	0.503
多くの財産を所有	0.507	-0.338	0.533	-0.287	0.517
高い地位につく	0.668	-0.303	0.734	-0.171	0.310
(冗長性)	0.040	0.015	0.097	0.039	0.017
正準相関係数	0.357	0.300	0.615	0.447	0.335
単独説明率(%)	42.8	29.2	50.2	20.6	10.4
ラオの近似F	4.040	2.930	10.550	6.970	5.360
自由度	(88, 6999.7)	(70, 6228.4)	(88, 5419.3)	(70, 4823.2)	(54, 4221.5)
p	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001

前節におけるクラスター分析の結果から、「教育」は、他の変数との関係において、両国では違いが見られることが示唆された。このことを正準相関分析の結果から再度検討しておこう。前述したように、どちらの国においても、それぞれの第1正準相関に対し、「教育」は大きな構造係数を示していた。しかし、職業評定基準変数群の中で、この正準相関に大きな構造係数を示す他の変数を見ると両国では違いが見られる。日本では、「権力」に加え、「収入」、「尊敬」、「格好よさ」なども0.5以上の構造係数を示しているのに対し、韓国では、これらの変数の構造係数は低い。また、重要項目変数群を見ても、日本では、「高学歴」と「社会的

評価の高い職業に就く」が高い構造係数を示すのに対し、韓国では、「高学歴」と「高地位」の構造係数が大きい。韓国では、人々の意識の中で、学歴と地位がストレートに結びついてある可能性がある。職業評定構造の分析結果（3.1 節参照）から、韓国の方が職業的地位の序列構造に関して格差が大きいことが明らかになったが、格差が大きいがゆえに、高い地位につくことを重要と考え、そのためには高い学歴をもつことが必要ということになり、その関連が明確な形で示されたと言えよう。

以上の分析結果から、学歴、収入、地位など社会的資源の側面を重要と考える人は、職業評定においても、そのような側面を重視していること、一方、社会や家族とのつながりを重要と考える人は、仕事の特性側面を重視していることなどが明らかになった。関連の仕方や程度の違いはあるものの、日韓ともに人々が重要だと考えていることと、職業評定の際に重視する基準とは整合的であり、密接に関連していると言えよう。

4. まとめと今後の課題

本稿では、日本における「職業に関する全国調査（職業威信）」と韓国における「2004 年韓国職業威信調査」のデータを用いて、職業評定構造の日韓比較分析を行った。分析の結果明らかになったことを要約しておこう。

まず、職業評定構造と職業威信スコアについては、以下のことが明らかになった。

第1に、職業威信スコアにより、両国の職業評定構造を比較すると、相対的な順位が異なる職業もあるものの、相関係数レベルで見れば、非常に類似した構造であった。しかし、韓国の方が広範囲に分布しており、平均値は低く、ばらつきは大きいという違いも見られた。

第2に、個々の職業に注目すると、日韓では違いも見られた。それぞれの国の職業威信スコアの平均値を、個々の職業威信スコアから引いた補正スコアで見れば、職業威信スコア下位層での違いが大きく、下位に位置づけられる職業は、日本よりも韓国において、より低く評定される傾向が見られた。

以上の2点は、韓国が、日本以上に、伝統的職業観（＝伝統社会の儒教的職業観：精神労働を重視し、商工業に関連する職業を低く見る）の影響を受け、職業間での威信格差がかなり大きい社会（有田 2006：75）であることの証左と言えるだろう。

第3に、同国内における異なるカテゴリー間の方が、両国の同じカテゴリー間の相関よりも高く、評定者の属性効果よりも、国による違いの方が職業威信評定に影響を及ぼしていた。

第4に、INDSCAL 分析結果における2次元解では、それぞれの次元が、各国の職業評定の序列構造を表わすものであると解釈できたが、次元における職業座標間の相関は高く、両国の職業評定構造は非常に類似したものであった。

次に、職業評定基準については、以下のことが明らかになった。

第1に、職業評定基準の補正スコアで見ると、相対的に重視されている項目とされていない項目には、類似した傾向が見られたが、「創造力」、「自律性」、「教育」の3基準は、異なった評価のされ方をしていた。

第2に、INDSCAL 分析結果における2次元解の各次元は、職業特性（仕事内容）を表わす次元と社会的資源を表わす次元であると解釈された。両国では、次元に対する重み（重要度）に違いが見られ、職業評定基準の重視のされ方が異なっていた。また、クラスター分析の結果、両国では、職業に対しての「教育」のもつ意味が異なる可能性が示唆された。

第3に、「人々にとって重要なこと」との関連を見ると、関連の仕方や程度の違いはあるものの、日韓ともに、人々が重要だと考えていることと職業評定の際に重視する基準は、密接に関連していた。

以上のことが明らかになったが、課題も残されている。職業評定およびその結果である職業威信スコアとその際に重視される評定基準の関係である。両国では、類似する評定構造が見られ、職業威信スコアにおいても高い相関が見られたが、職業評定基準について見れば、個々の基準の重視度や用いられ方に違いも見られた。結果として表面的に現れてくる職業評定構造の類似性も、その生成過程には違いがある可能性も考えられる。職業の序列構造において見られた相対的な順位の違いや同一の職業における職業威信スコアや補正スコアにおける絶対的なスコア値の差異などを生じさせる要因となっているのかもしれない。元治（2000）は、本稿と同じ「職業に関する全国調査（職業威信）」のデータを用いて、職業評定と評定基準の関連について分析し、評定基準において社会的資源と職業特性の2要因が見出され、この2要因が職業評定において大きな規定力をもつこと、評定される職業により重視される要因が異なり、各要因の影響度も異なることを明らかにした。韓国においても、同様の結果が見られるのだろうか。この点については、今後の課題としたい。

【文献】

- 有田伸, 2005, 「韓国における職業評定の分析—日本との比較を中心に—」 園田茂人編著『東アジアの階層比較』中央大学出版部, 25-56.
- 有田伸, 2006, 『韓国の教育と社会階層 「学歴社会」 への実証的アプローチ』東京大学出版会.
- Bian, Y., 1996, "Chinese Occupational Prestige A comparative analysis," *International Sociology* 11(2): 161-186.
- Blau, P. M. and O. D. Duncan, 1967, *The American Occupational Structure*. Wiley.
- Bose, C. E. and P. H. Rossi, 1983, "Prestige Standings of Occupations as Affected by Gender," *American Sociological Review* 48: 316-330.
- Featherman, D. L. and R. M. Hauser, 1976, "Prestige or Socioeconomic Scales in the Study of Occupational Achievement," *Sociological Methods & Research* 4: 403-433.
- 元治恵子, 2000, 「職業評定の構造—正準相関分析を用いて—」『応用社会学研究』42: 35-45.
- 元治恵子・都築一治, 1998, 「職業評定の比較分析—威信スコアの性差と調査時点間の差異」都築一治編『1995年SSM調査シリーズ5・職業評価の構造と職業威信スコア』1995年SSM調査研究会, 45-68.
- Hauser, R. M., 1982, "Occupational Status in the nineteenth and twentieth Centuries," *Historical*

Methods 15: 111-126.

Hauser, R. M. and D. L. Featherman, 1977, *The Process of Stratification: Trends and Analysis*. Academic Press.

林知己夫, 1995, 『数字からみた日本人のこころ』 徳間書店.

林知己夫, 2001, 『データの科学』 朝倉書店.

Hodge, R. W., P. M. Siegel, and P. H. Rossi, 1964, "Occupational Prestige in the United States, 1925-63," *American Journal of Sociology* 70: 286-302.

Hodge, R. W., D. J. Treiman, and P. H. Rossi, 1966, "A Comparative Study of Occupational Prestige," in Reinhard Bendix & Seymour Lipset (eds.), *Class, Status, and Power*. The Free Press, pp. 309-321.

ILO "LABORSTA" (Online Database) (<http://laborsta.ilo.org/>)

中尾啓子, 2005, 「韓国の職業威信スコア—2004年威信調査の概要と威信スコアの作成—」 尾島史章編『現代日本におけるジェンダーと社会階層に関する総合的研究』 183-205.

Nakao, K. 1992, "Occupation and Stratification: Issues of Measurement," *Contemporary Sociology* 21(5): 658-662.

Nakao, K. 2006, "Evaluation of Occupational Prestige Hierarchy in Korea and Japan," The proceeding of KGSS Symposium 2006 in Korea.

Nakao, K. and J. Treas, 1994, "Updating Occupational Prestige and Socioeconomic Scores: How the New Measures Measure Up," *Sociological Methodology* 24: 1-72.

直井優・鈴木達三, 1977, 「職業の社会評価の分析—職業威信スコアの検討」『現代社会学』 4 (2) : 115-156.

文部科学省 (生涯学習政策局調査企画課), 2006, 「教育指標の国際比較」 (平成 18 年版) (http://www.mext.go.jp/b_menu/houdou/18/03/06032718/001.htm).

文部科学省, 「学校基本調査」.

Reiss, A. J., O. D. Duncan, P. K. Hatt, & C. C. North, (eds.), 1977, *Occupations and Social Status*. Arno Press.

Siegel, P. M., 1970, "Occupational Prestige in the Negro Subculture," *Sociological Inquiry* 40: 156-171.

園田茂人, 2005, 「職業評価の社会力学—日中比較からの知見—」 園田茂人編著『東アジアの階層比較』 3-23. 中央大学出版部.

Treiman, D. J., 1975, "Problems of Concept and Measurement in the Comparative Study of Occupational Mobility," *Social Science Research* 4: 183-230.

Treiman, D. J., 1977, *Occupational Prestige in Comparative Perspective*. Academic Press.

Treiman, D. J. and K.-B. Yip, 1990, "Educational and Occupational Attainment in 21 Countries," in M. L. Kohn (eds.), *Cross-National Research in Sociology*. Sage, pp. 373-394.

都築一治編, 1998, 『1995年SSM調査シリーズ5・職業評価の構造と職業威信スコア』 1995年SSM調査研究会.

Yoo, H. J., 2006, "The Occupational Structure and Prestige Rankings in China: A Comparative Study with Korea and Japan", Paper to be presented at the seminar at Tohoku University.

【謝辞】統計数理研究所の前田忠彦氏に 3.6 節の正準相関分析についてのご協力をいただきました。また、東北大学の佐藤嘉倫氏に Yoo(2006)についての情報と資料をいただきました。ここに記して感謝いたします。ただし、分析結果の解釈などについては、すべて筆者の責任である。

Occupational Evaluation: A Comparative Study of Japan and Korea

Keiko GENJI

Rikkyo (St. Paul's) University

The present paper investigates the structure of occupational evaluations in Japan and Korea by analyzing the data collected from a nationwide survey. The occupational prestige hierarchy between Japan and Korea is very similar (Person's $r = 0.940$). But the relative rankings of some occupations are different. By the INDSCAL analysis and the method of cluster analysis, a high level of agreement in the evaluation of occupations between two countries was also noted. The means of each evaluative criterion in Korea are Japan show some differences. The way of evaluating the occupational status between two countries, i.e. the weights of each evaluative criterion, is different.

Keywords and phrases: structure of occupational evaluations, occupational prestige score, evaluative criteria

Evaluation of Occupational Prestige Hierarchy in Korea and Japan: A New Approach to Measuring Consensus and Inter-group Variations

Keiko Nakao
(Tokyo Metropolitan University)

Abstract

Numerous empirical studies of occupational prestige in the past several decades have accumulated a great deal of knowledge about the ways in which people evaluate the social standing of various occupations. The general empirical findings point to that of stability. It is understood not only that different measuring instruments produce very similar results, but also that the prestige hierarchy is remarkably stable over time (Hodge, Seigel, and Rossi, 1964; Nakao and Treas, 1994) and consistent across societies (Treiman, 1977). Recent studies that compared the prestige hierarchy among East Asian societies also found a high degree of consistency between Korea and Japan (Arita, 2005) and between China and Japan (Sonoda, 2005). Such high degree of consistency may easily steer us to conclude that the status hierarchy of occupations are quite similar among East Asian countries. These statistical results were, however, obtained in analyses based on the prestige scales already constructed by aggregating individuals' ratings. Since the prestige score of an occupational title is equivalent to the average score over all respondents' ratings on that title, analyses that are based solely on the computed prestige scores inevitably ignore variability among individual raters.

The objective of this paper, therefore, is to re-examine the degree of agreement and disagreement in individuals' evaluations. Here I propose a new approach to measure the degree of consensus in occupational evaluation, based on the model called a theory of cultural consensus. The application of this approach enables us to assess whether there is a collective consensus. It also allows us to measure the portion of prestige perception that is commonly shared by all individuals in both countries relative to the amount of variation not shared by the respondents in two countries. The analysis showed an existence of strong consensus within each country as well between two countries, confirming prior findings; however, there still remains a portion of variance yet to be explained by various subgroups of individuals.

Key words: Occupational Prestige, Consensus Model, and Korea-Japan Comparisons

1. Introduction

Numerous empirical studies of occupational prestige in the past several decades have accumulated a great deal of knowledge about the ways in which people evaluate the social standing of various occupations. Prior studies not only produced widely used scales of prestige and the socioeconomic index (SEI) for various occupations, researchers have made efforts to investigate systematic properties of prestige scales as well as how

people evaluate social standing of occupations. The general empirical findings point to that of stability. It is understood not only that different measuring instruments produce very similar results, but also that the prestige hierarchy is remarkably stable over time (Hodge, Seigel, and Rossi, 1964; Nakao and Treas, 1994) and consistent across societies (Treiman, 1977). Recent studies that compared the prestige hierarchy among East Asian societies also found a high degree of consistency. Between Korea and Japan, for example, although Arita (2005) found some differences in the prestige scores on some particular occupational titles (e.g., “Pharmacist” and “Carpenter”), the overall relative rankings produced a very strong association, i.e., 0.954 (Pearson’s r), between the data from two countries. Similarly, comparing China and Japan, 0.88 (Spearman’s ρ) was reported between the prestige scales from the two countries (Sonoda, 2005).

Such high values of correlation may easily steer us to conclude that the status hierarchy of occupations are quite similar among East Asian countries, leaving little room for investigating any meaningful structural differences.

These results were, however, obtained in analyses based on the prestige scales already constructed by aggregating individuals’ ratings. Since the prestige score of an occupational title is equivalent to the average score over all respondents’ ratings on that title, analyses that are based solely on the computed prestige scores inevitably ignore variability among individual raters. Such analysis would be adequate if we assumed that a societal consensus existed in occupational evaluation, and indeed, much prior research did attest to a strong agreement among individuals’ ratings. At the same time, however, there are others asserting that the variation among individuals’ ratings is more than just random error (e.g., Wegener, 1992). The objective of this paper, therefore, is to re-examine the degree of agreement and disagreement in individuals’ evaluation. This was considered to be a necessary step prior to making substantive comparisons between Korea and Japan about the structure of occupational hierarchy. Here I propose a new approach to measure the degree of consensus in occupational evaluation, based on the model called a theory of cultural consensus. The application of this approach enables us to assess whether there is a consensus, as well as to measure the portion of prestige perception that is commonly shared by all individuals relative to the amount of variation observed between subgroups and individuals.

2. Occupational Prestige as a Collective Conscience

The discussions of consensus in prestige judgments took off when strong inter-rater correlations were found in several studies in the 1970's. For example, Balkwell et al (1980, 1982) reported .745 as the average correlation of individuals' judgments. While other researchers reported lower values of inter-rater correlations, from .42 to .48 (Jencks et al, 1972; Goldthorpe and Hope, 1974; Kraus et al, 1978), they were still regarded as indicating substantial agreement among the raters. Challenging such view of the prestige consensus, Guppy (1982) argued that the degree of consensus varied according to the social strata, thus no collective conscience can be claimed. He pointed out that the level of agreement is greater among higher status individuals than among those who are in lower strata. Guppy and Goyder (1984) further reported the differences in the level of agreement based on education, occupation, and race of individuals. Acknowledging such variations in agreement, however, Hodge et al (1982) contended that characteristics of evaluators never explain more than 25% of the variance in prestige ratings and that they would be far less than variations within subgroups. More recently, Wegener (1992) reiterated the issue of prestige consensus and the importance of investigating individual differences. Using the psychological scaling techniques, Wegener found a polarization of judgments among higher status individuals while a lack of discriminating responses was found in lower social strata. He suggested that this difference of variance in responses might attribute to the difference in the level of agreement between various social groups.

In order to clarify unresolved issues about consensus in occupational prestige evaluation, I adopted the conceptual basis of the cultural consensus model and its related statistical techniques recently developed in psychometrics and often applied in anthropology (Romney, Weller, and Batchelder, 1986; Batchelder and Romney, 1988; Romney, Moore, Batchelder, and Hsia, 2000).

3. Theory of Cultural Consensus and Analytical Approach

The approach presented in this paper follows the theory of cultural consensus, the model first developed in the field of anthropology (Romney, Weller, and Batchelder, 1986; Batchelder and Romney, 1988). In the effort to objectively elicit an aspect of culture that is not directly observable, the theory of cultural consensus views culture (or

an aspect of culture) as information shared and stored in the minds of the society's members. It assumes that the correspondence between any two informants in their responses is a function of the extent to which each shares the knowledge about the aspect of culture being investigated, an idea that traces back to Spearman's landmark 1904 article (Spearman, 1904). The principle idea of this model, therefore, is to examine the pattern of agreement among informants' responses. It allows us to make inferences about how much each informant shares the knowledge of the information constituting culture, from which the aspect of the culture can be estimated.

I see occupational prestige similar to what anthropologists view an aspect of culture as above. The respondents' ratings of occupational status are subjective judgments based on their perceptions about how occupations are hierarchically located in the society. Like culture, occupational prestige ranking is information shared in the minds of the members in a society, which is not directly observable. If all respondents rated the same way, i.e., in the case of a total consensus, we may say that their responses constitute the social reality about the society's occupational prestige hierarchy, and that every member of the society shares such knowledge and responds without error. In practice, of course, individuals' ratings are not all the same. Variation in individuals' responses could result from, aside from unavoidable response error, either one of the following situations: (1) there is no single prestige hierarchy on which the society's members agree; (2) there is a single hierarchy of prestige as a social reality that is commonly shared by the members of the society, but all members are not equal in the degree to which they share the knowledge about the common pattern. Disagreement between any two individuals' ratings could be attributed to the difference in their understanding or knowledge about the social reality. Likewise, the agreement between any two people's responses is a function of how much each shares the information about the society's prestige hierarchy. Based on this model of shared social reality and individuals' social knowledge about the social reality, the approach employed in this paper is to examine the patterns of individuals' agreement with each other. Such information would lead us to find out (1) whether there is a single occupational prestige ranking that people agree on, and if so, (2) how much consensus there is.

4. Data and Preliminary Analyses

In the analyses that follow, I used three sets of data from nation-wide surveys in

Korea and Japan, all of which requested the respondents to evaluate the social standing of various occupations on a scale of five levels (i.e., from “very high” to “very low”). KGSS-2005 questionnaire contained 30 occupational titles to be rated, while Korean SSM survey in 2004 included 66 titles. The Japanese SSM in 1995 had 56 titles, 55 of which overlapped with the 2004 Korean SSM survey. I first constructed the prestige scale via a conventional manner for each data set and Table 1 presents the summary statistics and correlations among the three scales. (See Appendix for prestige scores computed for all occupational rated in each data set.)

Two Korean data sets show a similar mean prestige score (51.28 for KGSS and 52.42 for Korean SSM), while Japanese SSM gives a higher mean (56.92) than Korean surveys. Disparity among various prestige scores is also different between the two countries. For example, the values for standard deviation are greater for Korean data (17.64 for KGSS and 19.16 for Korean SSM) than that of the Japanese (14.79). Correlation values among the three scales shown in Table 2, however, are remarkably high, i.e., all above 0.9.

Table 1: Summary Statistics

	KGSS 2005	Korea SSM 2004	Japan SSM 1995
# of occ. Titles	30	66	56
mean prestige score	51.28	52.42	56.92
sd of prestige scores	17.64	19.16	14.79
min score	15.04	21.84	36.68
max score	90.71	86.06	90.12
range	75.67	64.22	53.44
Sample size	1613	903	1214
mean % non-response	2.98	0.78	3.38
mean sd	20.61	16.23	13.57

Table 2: Correlations among prestige scales

Correlation	KGSS	KoreaSSM	JapanSSM
KGSS		0.967	0.956
KoreaSSM	15		0.932
JapanSSM	12	55	

(lower diagonals: # of overlapping titles)

Correlation	KGSS	KoreaSSM	JapanSSM
KGSS		0.967	0.956
KoreaSSM	15		0.932
JapanSSM	12	55	

(lower diagonals: # of overlapping titles)

The above analyses demonstrate overall similarities and differences between the two countries in the way the occupational inequality is perceived. Compared to the Japanese, Korean respondents evaluate occupations lower on the average, and perceive a greater inequality among various occupations in their social standing. At the same time, however, the ranking of various occupations in terms of their status is shown quite similar to each other, both within and across societies. These observations are consistent with the findings in the previous literature mentioned earlier.

5. Measuring Consensus

Now I proceed to examine the differences in rating by individual respondents. I applied the statistical techniques developed for the model of cultural consensus to the prestige ratings. The basis for the analysis is the inter-rater correlations, which contain the information about the patterns of individual agreement in the way respondents rated various occupations. In order to elucidate the structure underlying inter-rater correlations, singular value decomposition was applied as follows:

$$\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^T,$$

$\mathbf{X}$ is an $n \times m$ matrix, where n is the number of respondents and m is the number of occupational titles rated, containing each individual's response rating of m occupations. $\mathbf{S}$ contains the singular values on the diagonals in a decreasing order. Performing singular value decomposition required a non-zero variance for each row of $\mathbf{X}$, therefore, I used only the respondents whose responses showed non-zero standard deviations (i.e.,

1583 cases for KGSS, 903 cases for Korea SSM, and 1146 cases for Japan SSM, eliminating respondents who did not differentiate status among any occupations). Table 3 presents the first five singular values and their corresponding percentages.

Table 3: Singular value decomposition
(Singular values and corresponding %)

	KGSS		Korea SSM		SSM95		Korea SSM		SSM95	
	Singular value	%	Singular value	%	Singular value	%	Singular value	%	Singular value	%
1	1018.0	64.3	438.6	48.6	603.8	52.7	450.7	49.9	602.7	52.6
2	41.9	2.6	32.9	3.6	42.7	3.7	36.0	4.0	43.1	3.8
3	36.2	2.3	21.8	2.4	34.7	3.0	22.8	2.5	34.7	3.0
4	34.5	2.2	18.6	2.1	26.7	2.3	21.9	2.4	26.9	2.4
5	32.7	2.1	17.1	1.9	22.8	2.0	19.0	2.1	23.0	2.0
n	1583		903		1147		903		1147	
# of titles	30		66		56		55		55	

As we observe in Table 3, the first singular value is overwhelmingly greater than the rest of the values for all data sets analyzed. This is a strong indication of the existence of a single underlying factor in the matrix of subject-by-subject correlations. In other words, the respondents' judgments show a strong agreement such that a single prestige order of occupations can be elicited. In all data sets, the first factor alone accounts for almost half the variance in the agreement among respondents' ratings, while the other factors each account for only 4% or less. To validate this dominant factor, I examined its correspondence with the prestige scores computed in the conventional manner, i.e., taking the mean of the ratings. The values of correlations were .99 or above in all cases.

The above analysis confirmed the existence of strong consensus on occupational evaluation, and this collective view corresponds to the conventional prestige scores. Is this consensus shared equally by all individuals? If so, the portion of variance not accounted for by the dominant factor is likely to be error variance. In the next section, I will examine whether there is any systematic pattern in which respondents share the collective view.

6. Conformity to the collective evaluation

The model of cultural consensus allows us to examine how each individual

deviates from (or conforms to) the single factor extracted. Index of such conformity is the values of the left singular vector. A high value of this index indicates that a respondent's evaluation on all occupations is similar to the collective view. Using this index, therefore, we can find out the attributes of respondents having a similar perception about the occupational hierarchy to the societal view. Table 4 below summarizes the results.

Table 4: Conformity to the collective view by respondents' attributes

	KGSS			Korea SSM			Japan SSM		
	mean	n	sig.	mean	n	sig.	mean	n	sig.
Sex			*			*			*
Male	0.739	724		0.635	368		0.703	529	
Female	0.700	859		0.670	535		0.697	617	
Education			**			**			*
< High school	0.662	871		0.632	570		0.693	860	
> High school	0.787	704		0.697	332		0.725	280	
Employment Status			**			ns			ns
Not working	0.746	891		0.667	382		0.701	312	
Working	0.682	692		0.648	521		0.701	830	
	Corr.	n	sig.	Corr.	n	sig.	Corr.	n	sig.
Age	-0.282	1583	**	-0.062	903	ns	0.051	1146	ns
Occ. Prestige	0.138	1325	**	0.152	509	**	0.111	834	**

**: $p < .01$, *: $p < .05$

As illustrated in Table 4, the index of conformity is associated with the educational level and occupational prestige of the respondents in all surveys. The higher the educational level and occupational status, the stronger conformity one has with the collective evaluation. Sex and age also show some association with the degree of conformity; however, such association is not consistent in all surveys, implying that it may depend on the occupational titles presented to the respondents (i.e., KGSS has a different set of titles from the other two surveys). The above results suggest that the view of the respondents with higher socioeconomic status is more similar to the societal consensus. Two different explanations might be possible. (1) Based on the theory of consensus, individuals with higher social status have better knowledge about the social reality than those with lower social status. If this were the case, higher status individuals would agree among themselves, while those in lower social strata would show little agreement among themselves. (2) Subgroups of individuals (i.e., higher status and lower status) each agree on a different view of occupational hierarchy but the perception of the dominant subgroup emerged as the collective consensus.

In order to pursue this further, inter-group variations are more closely examined. Using the subject-by-subject correlation matrices based on their ratings, I observed how much agreement there is within each subgroup and across subgroups of individuals. To allow the data more comparable between the two countries, analyses in this section are based on two data sets, Korea SSM and Japan SSM, which shared a common set of 55 occupational titles rated in their survey design. Table 5 presents the mean subject-by-subject correlations for within subgroups of individuals and between subgroups.

Table 5: Mean inter-subject correlation

	Korea SSM			Japan SSM		
	n	within-group corr.	between-group corr.	n	within-group corr.	between-group corr.
Male	368	0.428	0.457	529	0.521	0.525
Female	535	0.471		617	0.527	
20–39 yrs. Old	399	0.462	0.462	341	0.493	0.512
40–69 yrs. Old	504	0.468		805	0.543	
<High school	570	0.443	0.469	860	0.520	0.534
>High school	332	0.477		280	0.537	
Professionals+Managerials	68	0.473	0.470	131	0.533	0.532
Other	440	0.447		703	0.518	

If each subgroup agrees upon a different view, within-group correlations would be greater than a between-group correlation. None of the subgroups examined here follows that pattern. According to Table 5, there seems to be little difference between males and females for Japan, while Korean females show a higher level of agreement among themselves than Korean males. However, Korean males agree with females agree more than they do among themselves. Older respondents in Japan seem to agree more among themselves than younger individuals do, however, younger Japanese agree more with older Japanese than among themselves. We confirm in both countries that individuals with higher social status (higher educational level and occupational status) tend to agree more with each other than those with lower status, however, all within-group correlations are not higher than the between-group correlation. These results suggest that individual attributes and social status variables, to a certain degree, do account for the pattern of individual agreement and variation for both countries, however, distinct views do not seem to exist for various subgroups of individuals.

7. Korea-Japan comparison

Analyses up to this point suggested for both Korea and Japan that (1) there is a consensus on occupational prestige hierarchy, (2) such dominant view is shared by individuals with higher social status, compared to those with lower status, and (3) individuals with higher status show a stronger agreement with each other than lower status respondents do among themselves. Finally, in this section, I will focus on the cross-cultural differences in occupational evaluation based on the model of cultural consensus. I first compare Korea and Japan by assessing to what extent Koreans and Japanese share in their view of the occupational hierarchy. I aggregated the data from Korea SSM and Japanese SSM on the respondents' evaluations of 55 occupational titles that were common to both surveys. There are 2049 respondents; 903 Koreans and 1146 Japanese respondents. Table 6 shows the results from the singular value decomposition.

Table 6: Singular value decomposition (Korea and Japan combined)

Korea SSM + Japan SSM			
		Singular value	%
	1	1023.4	49.9
	2	71.0	3.5
	3	69.9	3.4
	4	46.6	2.3
	5	41.7	2.0
	n	2050	
	# of titles	55	

The overwhelmingly large first singular value in Table 6 indicates that there is a single prestige order extracted from all respondents' agreement, Koreans and Japanese combined. Scores from this first factor correlates with Korean prestige scores at .977 and with Japanese prestige scores at .989, validating that this factor corresponds to the conventional prestige scores. This first factor explains 49.9% of variance, leaving the other half for variation attributed to the difference between two countries, individuals and error.

In order to quantify the difference between two countries, I examined the 2049 by 2049 subject-by-subject correlation matrix. The analytical methods I apply here assume the following. First, the mean of subject-by-subject correlations within a group of individuals indicates the extent to which a common shared pattern exists. Second, the

correlation between two individuals is the product of the correlation of each individual with the shared societal pattern (Romney, Moore, Batchelder, and Hsia, 1999). These assumptions imply that the correlation between two individuals, i and j , r_{ij} , can be decomposed as a product of each individual's correlation with the collective pattern, $r_{ij} = r_{it} r_{jt}$, where t denotes the relevant shared pattern. An additional assumption is that there should be no negative correlation among subjects.

Based on the assumptions stated above, I divided the subject-by-subject correlation matrix into three parts; Korean-Korean, Japanese-Japanese, and Korean-Japanese raters, and Table 7 displays the mean inter-subject correlation for each part.

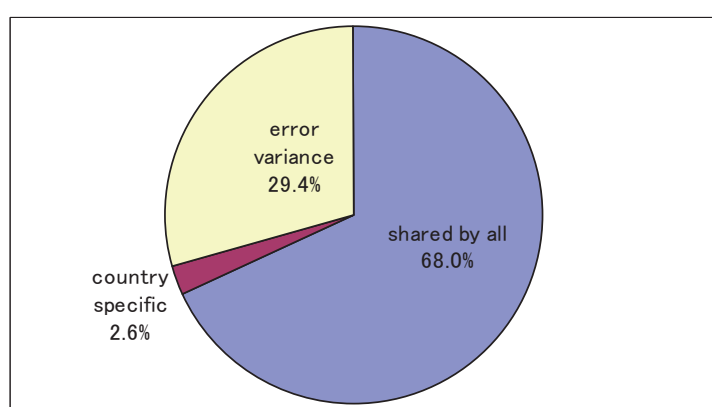
Table 7: Mean inter-subject correlation

Raters	Korean	Japanese	n
Korean	0.464	0.462	903
Japanese		0.526	1146

The mean correlation among the Korean subjects is 0.464, while Japanese show a mean of 0.526, slightly higher than that of the Koreans. It seems that Japanese agree among themselves more than the Koreans do. Note that the correlation between Korean respondents and Japanese respondents is smaller than either one of the mean within-group correlation, implying the existence of some between-group (Korea-Japan) differences. However small, this is an indication that Koreans and Japanese do hold a different view of the occupational structure. According to the process models of cultural consensus, the square root of the mean correlation within a group approximates the average conformity to the collective view (Romney, Weller, Batchelder, 1986; Batchelder and Romney, 1988). Thus, 0.681 for Koreans and 0.725 for Japanese would be the approximate estimate of the amount of conformity to an average Korean or Japanese would share with others of same nationality. The mean of the between-group correlations, i.e., Korean-Japanese correlation is 0.462. Following the decomposition of the correlation coefficient, $r_{ij} = r_{it} r_{jt}$, we can interpret the square root of 0.462, i.e., 0.680, as the approximate amount of social perception shared by Koreans and Japanese respondents. Comparing this value to what we obtained above would give us the relative portion that is specific to the groups. From the above analysis, we obtained 0.464 and 0.526 as the means of subject-by-subject correlations within groups, Koreans and Japanese, respectively. Since we have different group sizes for both samples, we

aggregate the two averages weighted by their group sizes, which equals to 0.499. The difference between the square root of this weighted mean, 0.706, and 0.680 obtained above would indicate the incremental portion that is specific to the groups. The difference is 0.026 ($= 0.706 - 0.680$), meaning that the group specific portion is merely 2.6%. The rest, $1 - 0.680 - 0.026 = 0.294$, then make up for sampling and error variance. Figure 1 graphically displays how the view of occupational prestige hierarchy is shared (1) by all, (2) country specific, and (3) sampling and error variance.

Figure 1: Partitioning of the occupational prestige evaluation shared by Koreans and Japanese



The above partitioning helps us to evaluate the relative size of the inter-group variation compared to the amount that is commonly shared. While 68.0% is commonly shared by both Koreans and Japanese, 2.6% is attributed to the difference in two countries.

Using the same logic, the inter-group differences were examined in terms of groups based on the raters' sex and age. For simplicity, individuals are grouped into two categories for age ("20-39 years old" and "40 years old or older"), education ("High school or less" and "More than high school"), and occupation ("Professionals + Managerials" and "Other"). Based on these mean inter-subject correlations, the relative amount of shared perception was computed for each variable and they are summarized in Table 8.

Table 8: Partitioning of the Shared Perception about Occupational Prestige

	Country	Sex	Age	Education	Occupation
Shared by all	0.680	0.693	0.693	0.699	0.704
Group specific	0.026	0.002	0.002	-0.003	-0.011
Error variance	0.294	0.305	0.305	0.304	0.307

Table 8 confirms that the inter-group variation is quite small for all subgroups, compared to what is commonly shared by all. The portion commonly shared across groups is consistently high, all at the level around 68% to 70%. Note, however, the country specific portion (2.6%) is greater than the differences between groups on other variables, where sex and age each accounts for only 0.2%. The difference between Korea and Japan is much greater than the variation among subgroups of respondents' with various attributes.

8. Summary and Discussion

Prior studies comparing occupational prestige scores among various societies consistently reported highly strong correlations, leaving little room to pursue any structural differences in occupational hierarchy. Instead of utilizing prestige scores that had already been aggregated as in those previous studies, this paper attempted to re-examine the variation among individuals' evaluation before making cross-cultural comparisons in occupational structure. Here I applied a theory of cultural consensus in conceptualizing the collective conscience of occupational prestige. Applying the statistical methods developed for the cultural consensus model and its related methods enabled us to elucidate the degree of consensus and individual variation in occupational prestige evaluations by focusing on the patterns of agreement among individuals' ratings.

The analyses in this paper found the following. First, it showed that there is a single occupational status ranking underlying people's perception about occupational prestige. This was shown both in Korea as well as in Japan. In both countries, the single underlying factor was so overwhelming in terms of the amount of explained variation found among individuals that we could call it a collective conscience or societal consensus. This elicited factor corresponds to the widely used conventional prestige scores, confirming its validity.

Second, the individual's level of conformity is related to the level of his/her

socioeconomic status, where conformity is defined as the degree to which an individual's responses correspond to the collective view. This is not because individuals with lower status hold a different view from those with higher status. Instead, it is perhaps due to the fact that the level of agreement differs according to the social status of the respondents. Higher status individuals agree more to each other than those of lower status do. Previous empirical studies also found such phenomena, and one of the explanations for this difference was offered by Wegener (1992). He suggested that higher status individuals tend to polarize the difference in occupational status, while those with lower status do not differentiate statuses of various occupations. This polarization tendency would possibly be responsible for higher correlation observed among high status individuals. Our data indicated, however, that was not the case. Standard deviation across ratings of all occupations for each individual is not necessarily negatively correlated with his/her socioeconomic index, such as occupational prestige. Based on the analyses in this paper, the theory of cultural consensus offers an alternative explanation. It suggests that the difference in the level of agreement between groups is related to the difference in the level of social knowledge of the respective group members about the relevant issue in question. Those occupying a higher location on a status hierarchy tend to have a better knowledge about the social reality. Yet, another interpretation would be possible from a conflict theoretical perspective -- individuals with high social status are responsible for creating the social reality since the aggregated view of occupational hierarchy reflects the dominant view of the respondents with higher socioeconomic status. If we conceptualize social reality to be something that is created in the minds of the members of the society collectively, then occupational ranking extracted may be called a social reality that tends to reflect the view of higher status individuals. Pursuing this issue inevitably would involve further research and discussion.

Finally, the difference between Korean and Japanese respondents was examined. Although the amount of explained variation is small (2.6%) compared to the shared portion (68%), the pattern of inter-subject correlation suggests that Koreans and Japanese do hold a different view and the difference is much greater than the negligible amount explained by the raters' characteristics. This leads us to believe that there are some structural differences between the two countries that cannot be explained by the respondents' attributes or social status. For example, a preliminary analysis of the criteria used by the respondents when evaluating the occupational status indicates that

there is a difference in the way Koreans and Japanese weigh various criteria. This also implies that differences in the structure of status hierarchy or in the factors determining one's status attainment, etc. must be accountable for making people in each country to evaluate the social standing of occupations differently. Although the similarity between the two countries is not deniable, such difference is certainly worthy of further investigation.

References

- Arita, Shin. 2005. "Kankoku ni okeru shokugyou hyoutei no bunseki: Nihon tonon hikaku wo chuushin ni (Analysis of Occupational Evaluation in Korea: Comparison with Japan)" in Sonoda eds. *Higashi Asia no Kaiso Hikaku* (Comparison of Stratification among East Asia) pp. 25-56. Chuo University Press.
- Balkwell, J.W., Bates, F.L., Garbick A.P. 1980. "On the intersubjectivity of occupational status evaluations: A test of a key assumption underlying the "Wisconsin Model" of status attainment." *Social Forces* 58:865-81.
- Balkwell, J.W., Bates, F.L., Garbick A.P. 1982. "Does the degree of consensus on occupational status evaluations differ by socioeconomic stratum? Response to Guppy." *Social Forces* 60:1183-9.
- Batchelder, W.H. and A.K. Romney. 1988. "Test Theory without an Answer Key." *Psychometrika* 53:71-92.
- Goldthorpe, J.H. and K. Hope. 1974. *The Social Grading of Occupations. A New Scale and Approach*. Oxford: Clarendon.
- Guppy, L.N. 1982. "On intersubjectivity and collective conscience in occupational prestige research: A comment on Balkwell-Bates-Garbin and Kraus-Schild-Hodge." *Social Forces* 60:1178-82.
- Guppy, L.N. and J.C. Goyder. 1984. "Consensus on Occupational Prestige: A Reassessment of the Evidence." *Social Forces* 62:709-25.
- Hodge, R.W., P.M. Siegel, and P.H. Rossi. 1964. "Occupational Prestige in the United States: 1925-1963."
- Hodge, R.W., Kraus, V., Schild, E.O. 1982. "Consensus in occupational prestige ratings: Response to Guppy." *Social Forces* 60:1190-96.
- Jencks, C.M., H. Acland, M.J. Bane, D. Cohen, H. Gintis, B. Heyns, and S. Michelson. 1972. *Inequality: A Reassessment of the Effect of Family and Schooling in America*.

New York: Basic Books.

- Moore, C.C., A.K. Romney, and T. Hsia. 2000. "Shared Cognitive Representations of Perceptual and Semantic Structures of Basic Colors in Chinese and English." *PNAS* 97:5007-10.
- Nakao, K. and J.K. Treas. 1994. "Updating Occupational Prestige and Socioeconomic Scores: How the New Measures Measure Up." *Sociological Methodology* 24:1-72.
- Romney, A.K., S.C. Weller, and W.H. Batchelder. 1986. "Culture as Consensus: A Theory of Culture and Informant Accuracy." *American Anthropologist* 88:313-38.
- Romney, A. K., C.C. Moore, W.H. Batchelder, and T. Hsia. 2000. "Statistical Methods for Characterizing similarities and Differences between Semantic Structures." *PNAS* 97:518-23.
- Sonoda, Shigeto. 2005. "Shokugyou hyouka no shakai rikigaku (Social dynamics of occupational evaluation: From China-Japan comparisons)" in Sonoda eds. *Higashi Asia no Kaiso Hikaku* (Comparison of Stratification among East Asia) pp. 3-23. Chuo University Press.
- Spearman, C. 1904. "'General Intelligence,' Objectively Determined and Measured." *American Journal of Psychology* 15:201-93.
- Treiman, D.J. 1977. *Occupational Prestige in Comparative Perspective*. New York: Academic Press.
- Wegener, B. 1992. "Concepts and Measurement of Prestige." *Annual Review of Sociology* 18:253-80.

Appendix: Prestige Scores

	Japan SSM 1995		Korea SSM 2004		KGSS 2005	
	mean	sd	mean	sd	mean	sd
Head of a large company	87.32	16.74	83.35	22.91		
Electrician	50.45	13.09	46.01	19.70		
Restaurant Cook	51.62	13.25	49.92	18.62		
Carpenter	53.13	15.39	37.19	23.09	33.50	21.33
Waitress	38.07	15.49	25.72	20.91		
Sales person in a large company	57.39	14.21	59.62	16.70		
Physician	90.12	14.85	83.91	19.22		
Owner of a small business	68.90	15.82	73.49	19.11		
Stewardess	70.00	17.24	64.36	19.60		
Architect	71.95	16.22	67.81	19.63		
Nurse	59.73	17.04	53.99	18.11	53.43	16.54
Machine assembler for a larger company	51.11	14.33	48.77	20.13		
Automobile sales person	47.25	11.91	42.69	17.57		
Miner	36.68	18.99	21.84	21.85		
College Professor	84.32	17.11	80.39	19.02	84.91	15.65
Fisherman	46.53	16.16	26.20	21.67		
Train station attendant	47.81	11.00	46.24	18.53		
Mail carrier	46.25	11.98	41.13	19.54		
Kindergarten teacher	52.94	12.95	50.83	17.15		
Construction foreman	56.70	13.70	54.58	17.13		
Fashion designer	64.55	15.29	62.02	17.33		
Police officer	57.92	15.29	55.52	18.88	51.02	16.09
Section chief in a small company	56.10	12.95	57.04	15.47	63.03	15.45
High-level government official	77.45	21.16	74.22	18.06	76.42	15.92
Grade school teacher	63.58	15.62	62.67	17.26		
Barber	49.68	11.46	45.87	19.19	34.26	18.75
Professional athlete	68.98	19.75	67.39	19.92		
Machine assembler for a small company	46.72	14.49	45.19	19.44		
Farmer	45.62	17.57	29.72	24.09	42.84	19.94
Printing factory worker	44.03	13.73	28.78	20.75		
Spinning factory worker	41.95	15.07	28.05	20.27		
Judge	86.91	16.94	86.06	19.54	90.71	14.30
Bus driver	48.92	11.54	35.63	20.48		
Guardsmen	39.90	16.12	23.80	21.17	22.38	20.00
Airplane pilot	82.46	17.87	75.72	19.37		
Sales clerk at a store	42.44	14.02	30.79	19.79		
Baker	44.62	13.22				

Appendix (continued)

	Japan SSM 1995		Korea SSM 2004		KGSS 2005	
	mean	sd	mean	sd	mean	sd
Priest	60.32	18.21	51.43	22.97		
Bank employee (except teller)	56.39	12.87	55.57	15.59		
Retail store owner	51.28	11.23	42.98	17.83		
Legislator	74.91	22.73	67.43	30.98		
Roadwork laborer	38.99	17.60	28.97	21.91		
Section chief in a municipal office	56.89	13.34	62.60	17.59		
Furniture maker	52.09	13.51	60.58	22.88		
Insurance sales person	44.27	13.83	38.39	19.69		
Automobile mechanic	46.81	11.86	40.70	20.42	41.08	18.66
Section chief in a large company	63.18	14.61	62.87	15.66		
Train conductor	51.32	10.70	51.39	17.56		
Office clerk in a small company	46.96	10.56	50.45	13.76		
Accountant	70.82	16.15	65.84	20.56		
Wholesale store owner	52.85	11.14	49.75	17.07		
Pharmacist	65.72	15.13	69.68	17.60	69.30	16.46
Automobile engineer	66.33	15.65	66.33	19.38		
Bank teller	49.38	10.00	49.25	15.58		
Food cannery worker	42.22	13.75	36.69	18.87		
Musician	66.61	16.01	64.13	19.15		
Dental Technician			58.14	17.95		
High school teacher			65.87	16.26		
Real estate agent			47.94	17.41		
Newspaper reporter			66.08	19.53	66.93	16.57
Street vendor			23.83	20.91	22.51	20.10
Physical therapist			49.39	17.38		
Instructor in a gym			46.00	15.96		
Restaurant owners			51.17	17.27	51.71	16.77
Genetic engineer			76.39	19.19		
Prep (cram) school teacher			56.19	17.27		
Receptionist for telephone shopping			37.03	19.35		
Maids and housecleaners					19.45	19.50
Junior high school teachers					67.08	15.81
Dry cleaners					40.59	18.40
Factory foremen					44.52	16.97
Factory workers					28.32	19.37
Assistant directors of bank					58.92	15.56
Entertainers					67.14	21.37

Appendix (continued)

	Japan SSM 1995		Korea SSM 2004		KGSS 2005	
	mean	sd	mean	sd	mean	sd
Local level civil servants					55.23	15.37
Section chiefs of a department in a small and medium company					63.03	15.45
Army Generals					79.17	19.40
Sewers					27.20	20.51
Sales persons in department store					36.21	17.85
Manual workers					15.04	19.09
Owners of electronic goods sales agents					59.61	16.89
Heads of departments in large companies					72.78	15.13

	n	56	56	66	66	30	30
mean		56.92	14.79	52.42	19.16	51.28	17.64
SD		13.57	2.67	16.23	2.54	20.61	2.05
min		36.68	10.00	21.84	13.76	15.04	14.30
max		90.12	22.73	86.06	30.98	90.71	21.37
range		53.44	12.73	64.22	17.22	75.67	7.07

社会関係に関連する指標の日本・台湾・韓国の 国際比較について

菅野 剛
(日本大学)

【要旨】

本研究では、社会関係や社会的ネットワークに関連する項目について、国際比較の基礎的な分析を試みる。その際に、(1)質問文と回答選択肢が若干異なる場合、(2)同一の測定がなされている場合、の双方について国際比較の試みを行った。まず、社会関係の客観的側面に関する指標として、頼れる相手の職業カテゴリーについての比較を行った(2005 年 SSM 日本調査、2005 年 SSM 台湾調査、2005 年 SSM 韓国調査)。地位ジェネレータによる測定という方法では一致しているが、質問文と選択肢には若干の違いがあり、全く同じ意味の測定指標ではないことに注意が必要である。これらの二値変数は概して割合が低かった。年齢、学歴、個人年収、職業ごとの違いを確認し、社会構造と社会的ネットワークの間にある程度の類似性を見出した。また、主観的側面に関するものとして、ほぼ同一の測定が行われている社会関係と社会階層に関連する指向(1995 年 SSM 調査、2004 年韓国威信調査)について比較分析を行った。これらにおいては、社会的属性との関連は弱かった。平均は韓国の方がやや高く、価値指向の項目同士の内部相関構造は類似しているが、若干異なる側面も見られた。その上で検証的因子分析を行い、また、多母集団解析で配置不変モデル、測定不変モデルについてよい適合度を確認した。

キーワード：社会的ネットワーク、地位ジェネレータ、関係指向、多母集団解析

1. 国際比較調査と比較項目について

社会関係や社会的ネットワークを測定する指標には様々なものがある。近年、ソーシャル・キャピタルや社会関係資源として、この分野は注目されている。しかし、社会的ネットワークの測定は、学歴、職業等の客観的属性や、階層帰属意識、権威主義等の主観的意識といった、測定が繰り返し実施され、研究が蓄積して安定した指標とは異なる。そこで、本研究では、社会関係に関連する主観的・客観的な側面について、国ごとの違いについて分析を試みることにする。その際に、国際比較分析においては、測定が等価であるように作成された質問項目群を用いることがある。これらにおいては、尺度が安定している場合、より正確な国際比較を行うことが可能となる。たとえば世界価値観調査、ISSP、アジア・パロメータなどは国際比較を念頭において調査が実施されている。本研究では、同じワーディングである質問項目として、1995 年社会階層と社会移動調査と 2004 年韓国職業威信調査に含まれ

ている、社会関係に関連する指向と社会階層に関連する指向を用いて分析を行う。他方で、質問や回答選択肢が異なっているデータを用いて、事後的に国際比較を行うこともある。例えば、Katz-Gerro(2002)は、それぞれの国で異なった質問と回答選択肢で測定された調査データを用いて、文化活動について比較研究を行っている。国や文化によっては、統一的な質問項目が意味を持たない場合もありえるので、国によって若干異なる質問項目を用いながらも、相対的な分布の違いを考慮しながら、分析を行うことが出来る。本研究では、求職や転職の際(日本・台湾)や、あるいは頼み事をする際(韓国)に頼りにする相手の職業について得られた、それぞれ異なる回答選択肢について分析を試みる。

2. 日本・韓国・台湾における、頼りにする相手の職業(地位ジェネレータ)

本研究では、異なる質問項目・ワーディングであるため、解釈に制約が伴うものの、日本、台湾、韓国での比較を試みる。2005 年 SSM 日本調査の、就職や転職で力になってくれそうな相手についての質問項目では以下のように尋ねている。「かりに、あなたが、これから就職や転職をするとしたら、仕事の紹介や仲介で力になってくれそうな人はいますか。実際にあなたの力になってくれそうな人を思い浮かべ、その人の仕事をお答えください。つぎのなかから、あてはまる人すべてに○をつけてください。」 2005 年 SSM 台湾調査での、求職や転職をする際の紹介を頼む相手についての質問項目では、「もしあなたがいま、求職や転職する必要があったら、以下のどのような人に紹介を頼むと思いますか。(複数選択可)」と尋ねている。2005 年 SSM 韓国調査での、頼み事をできる相手についての質問項目では、「あなたが何か問題にぶつかった時、頼み事をできる次のような職業の知り合いはいますか。あてはまるものすべてに○をつけてください。」と尋ねている。それぞれにおける回答選択肢は表 1 となっている。

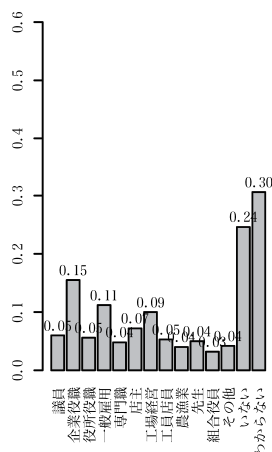
これらを用いて、記述的な比較分析を試みることにする。以上は、実際に助けてもらったり、サポートを得たことを測定しているわけではなく、期待・想定する社会関係ではあるが、個人が実際にアクセス出来る現実的な社会関係の広さを、間接的に示唆していると言える。これらいずれの質問項目も、回答選択肢が相手の職業であり、地位ジェネレータ形式となっている。1975 年 SSM 調査や 1995 年 SSM 調査において、つきあいのある相手の職業の有無を尋ねており、形式としてはこれに近い。地位ジェネレータを用いる場合は、たとえ回答者に有力な知り合いがいるとしても、選択肢にそのカテゴリーがなければ回答出来ないという問題がある。また、退職者、専業主婦、その他の社会的カテゴリーに属し、職業を有しない他者とのつながりについても測定出来ない。このため、男性の職業的ネットワークの測定には効率が良いが、専業主婦などの女性のネットワークについては測定ができていない側面がある。このような限界はあるが、通常大変コストがかかる社会的ネットワークについての測

定を大変効率よく実施出来るため、頻繁に使われる。日本と台湾においては就職・求職・転職などについての、職業に関連する質問であるが、韓国では、何か問題にぶつかった時に頼み事ができる職業の知り合いを尋ねている点が異なることに注意がいる。また、台湾では、回答選択肢に「考えたことがない」、「転職はありえない」が含まれている点も異なる。日本・台湾では求職・転職という場面が想定されていること、また、回答選択肢が職業名となっている地位ジェネレータであることから、ここでの分析においては、データを男性に限定して行うこととする。

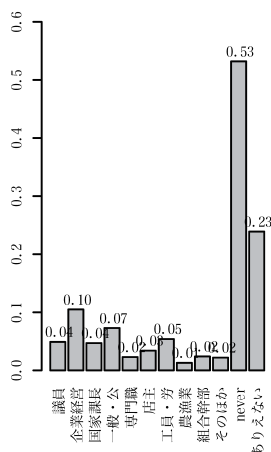
表 1: 頼りにする相手の職業の地位ジェネレータ

2005 年 SSM 調査	2005 年 SSM 台湾調査	2005 年 SSM 韓国調査
就職や転職で仕事の紹介や仲介	求職や転職で紹介	頼み事をできる職業の知り合い
1 国や地方の議員をしている人	議員	医師
2 民間企業で役職についている人	私営企業経営者・高級主幹	税理士・会計士
3 国や自治体の役所で役職についている人	国家機関課長級以上の人	弁護士
4 普通のサラリーマンや OL	一般職員・公務員	銀行員（課長以上）
5 医師・弁護士などの専門職の人	医者・弁護士などの専門職	公務員（課長以上）
6 商店や飲食店を営んでいる人	小商店の自営業者	民間企業の社長・役員
7 工場や工務店を営んでいる人	工業作業員・運転手・建築労働者など	政治家・政治家秘書
8 工場や商店で働いている人	農業・漁業従事者	聖職者
9 農業や漁業をしている人	同業組合幹部	労働組合や同業組合の役員
10 学校の先生	考えたことがない	警察官
11 同業組合や労働組合の役員	転職はありえない	誰もいない
12 その他（具体的に）	そのほか	
13 力になってくれそうな人はいない		
14 わからない		

日本男性：就職や転職で力になってくれそうな人



台湾男性：求職や転職で紹介を頼む人



韓国男性：頼み事をできる職業の知り合い

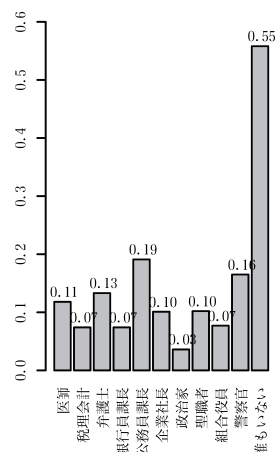


図 1: 頼りにすると回答された相手の職業の割合(日本、台湾、韓国)

2.1 頼りにすると回答された職業の割合(平均値)

まず、それぞれの国ごとに回答項目の平均を確認する(図 1)。なお、以降の探索的な分析については、主に R 2.6.1 for MacOS X を用いて行った(R Development Core Team, 2005)。

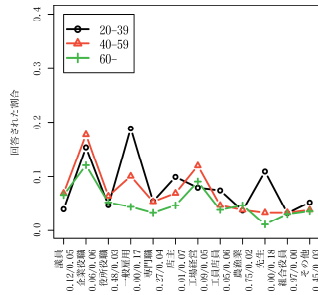
まず、日本における「いない」(0.24)、「わからない」(0.30)、台湾における「考えたことがない」(0.53)、「ありえない」(0.23)、韓国における「誰もいない」(0.55)など、いずれの国においても、職業名以外の項目の選択率が突出して高い。そして、頼りにすると回答された相手の職業については、概して平均値(割合)が低く、日本では 0.03-0.15、台湾では 0.01-0.10、韓国では 0.03-0.19 となっている。

まず、平均が高い、つまり選択率が高く、頼りにすると回答された職業について見ていく。日本の中では、企業役職(0.15)、一般雇用(0.11)が比較的に高めとなっている。台湾の中では、企業経営(0.10)、一般・公務員(0.07)が比較的高めとなっている。韓国の中では、公務員課長以上(0.19)が比較的高めであるが、銀行員課長(0.07)は相対的に高いというほどでもない。また、日本では工場経営者(0.09)も相対的に高めとなっている。韓国では、意味が近い項目として工員・労働者(0.05)があるが、回答には、経営者か労働者かによる違いも反映されていると考えられる。韓国では、警察官(0.16)が相対的に高い値を示している。

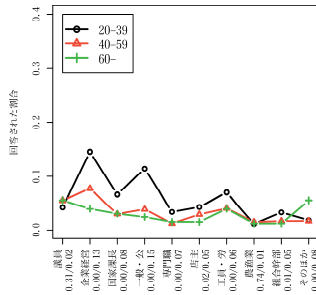
次に、平均が低い、つまりあまり選択がされず、頼りにするという回答が少なかった職業について見ていく。日本の中では、農漁業(0.04)、組合役員(0.03)が相対的に低くなっている。台湾の中でも、農漁業(0.01)、組合幹部(0.02)が相対的に低い。韓国の中では、農漁業に該当する項目はないが、組合役員(0.07)が相対的に低いと言える。また、日本では、議員(0.05)、専門職(0.04)も相対的に低い。台湾でも、議員(0.04)、専門職(0.02)が相対的に低いと言える。韓国では、政治家(0.03)が相対的に低く、日本・韓国と同様である。ただし、韓国では、専門職に該当するものとして、医師(0.11)、税理士・会計士(0.07)、弁護士(0.13)の 3 項目があるが、いずれも相対的に高めになっており、日本・韓国と異なる。韓国では求職・転職に限定せず、頼み事をできる相手を選択してもらっているためにこのような違いが生じた可能性が考えられる。平均が低い場合は、何らかの社会的障壁で社会関係が結びづらいということや、職業分布が原因で従事者が少なくアクセスしづらいということなど、いろいろな原因が考えられる。

また、日本と台湾に比べると、韓国における平均値が全体的に高めであることが図から分かる。これは、質問項目で問うている内容が若干異なることが一因であると考えられる。何らかの頼み事という幅広い内容でアクセス出来る社会的資源と、求職・転職に限定した場合に利用可能な社会的資源とでは、回答者から得られる回答の範囲が当然変わってくると言える。

日本男性：求職や転職で頼りにする人の仕事(本人年齢別)



台湾男性：求職や転職で頼りにする人の仕事(本人年齢別)



韓国男性：頼み事をできる知り合い(本人年齢別)

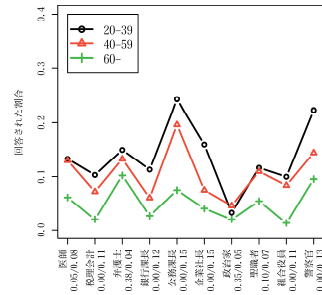
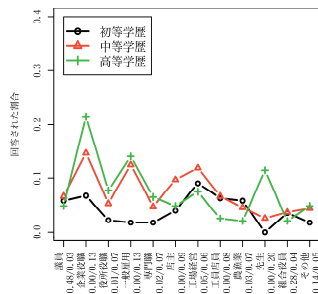
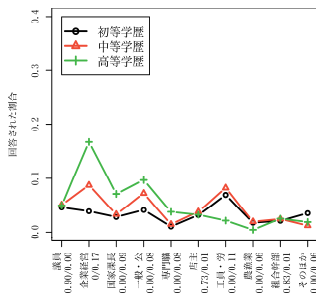


図 2: 頼りにすると回答された相手の職業の割合、年齢別(日本、台湾、韓国)

日本男性：求職や転職で頼りにする人の仕事(本人学歴別)



台湾男性：求職や転職で頼りにする人の仕事(本人学歴別)



韓国男性：頼み事をできる知り合い(本人学歴別)

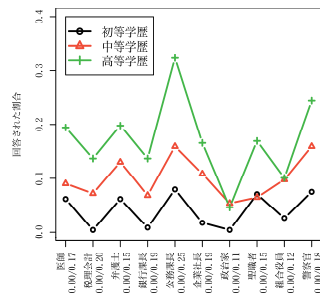
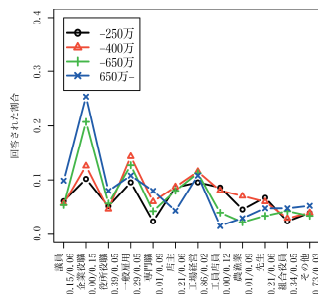
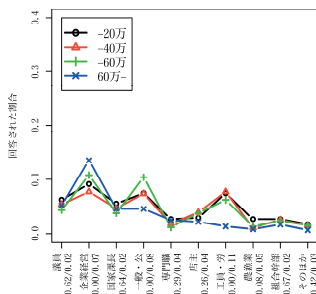


図 3: 頼りにすると回答された相手の職業の割合、学歴別(日本、台湾、韓国)

日本男性：求職や転職で頼りにする人の仕事(本人個人年収別)



台湾男性：求職や転職で頼りにする人の仕事(本人個人年収別)



韓国男性：頼み事をできる知り合い(本人個人年収別)

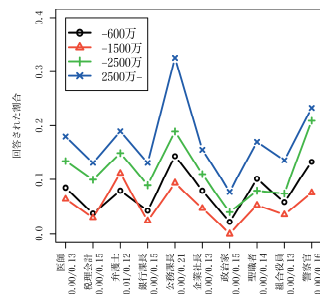


図 4: 頼りにすると回答された相手の職業の割合、個人年収別(日本、台湾、韓国)

日本男性：求職や転職で頼りにする人の仕事(本人職業別 (ISCO)) 台湾男性：求職や転職で頼りにする人の仕事(本人職業別 (ISCO))

韓国男性：頼み事をできる知り合い(本人職業別 (ISCO))

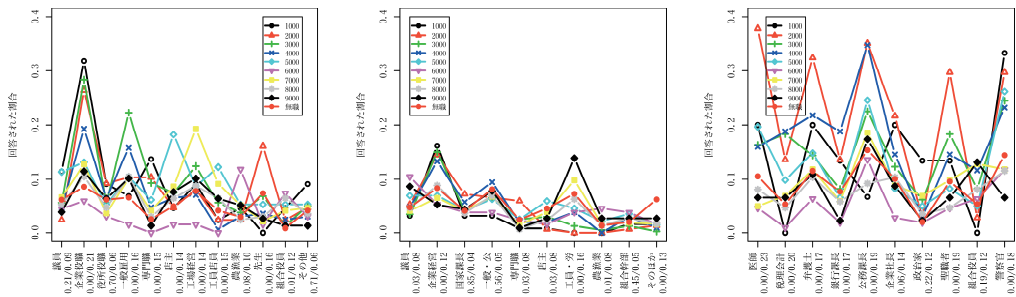


図 5: 頼りにすると回答された相手の職業の割合、職業別(日本、台湾、韓国)

2.2 属性ごとに見る頼りにする相手の職業の平均値

次に、年齢、学歴、個人年収、世帯年収、職業ごとの、回答の平均を見る。国際比較における学歴、職業、収入等の正確な分析は難しいが、それぞれの国の中における社会的資源の分布での相対的位置という視点から、社会階層とネットワークの関連の分析と比較ができる。まず、年齢ごとに回答項目の平均を確認する(図 2)。年齢は、20-39 歳、40-59 歳、60 歳以上の 3 カテゴリーとしている。図においては先ほどのグラフと同様に平均値を示している。ただし、情報が多く、プロットする点が多いため、同じ年齢グループごとに平均値を線で結んであり、結果的に折線グラフとなっている。また、X 軸のラベルには、いくつかの情報を示してある。それらは、頼りにすると回答された相手の職業名、年齢 3 分類と当該の二値変数との χ^2 二乗検定における有意確率、その関連の強さであり、職業名、p 値 / Cramer の V 係数の順に示してある。例えば、日本男性においては、年齢 3 分類と、頼りにする相手として議員をあげるかどうかの二値変数とのクロス集計において、 χ^2 二乗検定は有意でなく ($p=0.12$)、Cramer の V 係数(以下、V と表記)は 0.05 と関連も見られない。日本男性では、年齢が若いほど、一般雇用($V=0.17$)、先生($V=0.18$)を選択し、頼りにしている傾向がある。台湾男性においては、年齢が若いほど企業経営($V=0.13$)、一般・公務員($V=0.15$)を回答する傾向がある。また、関連は強くないが、全体的に若いほど回答が多い傾向が有意に見られる。韓国男性においては、年齢が若いほど、銀行課長($V=0.12$)、公務員課長($V=0.15$)、企業社長($V=0.15$)、警察官($V=0.13$)を回答する傾向がある。日本や台湾に比べ、韓国では有意な項目も多く見られるが、求職に限定されないサポートを問うていることが一因である可能性がある。

次に、学歴ごとに回答項目の平均値を確認する(図 3)。学歴は、日本、台湾、韓国のいずれにおいても、「初等学歴」、「中等学歴」、「高等学歴」の 3 カテゴリーとした。日本男性では、学歴が高いほど、企業役職($V=0.13$)、一般雇用($V=0.13$)、先生($V=0.20$)を回答している。台湾男性では、学歴が高いほど、企業経営($V=0.17$)を回答している。韓国男性では、

学歴が高いほど、医師(V=0.17)、税理士・会計士(V=0.20)、弁護士(V=0.15)、銀行課長(V=0.19)、公務員課長以上(V=0.25)、企業社長(V=0.19)、聖職者(V=0.15)、組合役員(V=0.12)、警察官(V=0.18)を回答する傾向がある。韓国では、頼み事をできるかどうかという質問について、学歴との関連が強い。いずれの国でも、学歴が高いほど、企業経営者や企業で役職に就いている相手へアクセスが出来る様子が伺える。

次に、個人年収ごとに回答項目の平均値を確認する(図 4)。それぞれ貨幣単位が異なるが、それぞれの国の男性において、標本数がおおよそ四等分になるようにカテゴリーを分割し、相対的な移り変わりを見ることにする。日本男性では、個人年収が高いほど、企業役職(V=0.15)を回答し、個人年収が低いほど、工員・店員(V=0.12)を回答する傾向がある。台湾男性では、年収との χ^2 乗検定で有意な項目は存在するが、関係の強さから見ると簡単には解釈が難しい結果となっている。ただ、個人年収が高い層(60 万-)は、企業経営を回答する一方で、一般職員・公務員工員・労働者を回答しない傾向は読み取れる。韓国男性では、個人年収が高いほど、医師(V=0.13)、税理士・会計士(V=0.15)、弁護士(V=0.12)、銀行課長(V=0.21)、公務員課長以上(V=0.21)、企業社長(V=0.13)、政治家(V=0.15)、聖職者(V=0.14)、組合役員(V=0.13)、警察官(V=0.16)を回答する傾向がある。韓国では、頼み事をできるかどうかという質問について、個人年収との関連が強い。なお、分析は省略するが、以上について同様の分析を世帯年収の観点からも行くと、個人年収の方が、関連が強い傾向がある。

次に、職業ごとに回答項目の平均値を確認する(図 5)。職業は、ISCO 職業分類の大分類を用いている。1000 番台 Legislators, senior officials and managers(立法議員、上級行政官、管理的職業従事者)、2000 番台 Professionals(専門的職業従事者)、3000 番台 Technicians and associate professionals(技術者及び準専門的職業従事者)、4000 番台 Clerks(事務的職業従事者)、5000 番台 Service workers and shop and market sales workers(サービス職業従事者、店舗及び市場での販売従事者)、6000 番台 Skilled agricultural and fishery workers(熟練の農林漁業従事者)、7000 番台 Craft and related trade workers(熟練職業及び関連職業従事者)、8000 番台 Plant and machine operators and assemblers(装置・機械操作員及び組立工)、9000 番台 Elementary occupations(初級の職業)となっている。カテゴリーが多いため、グラフが見にくいですが、本人の職業によって頼りにする相手の職業が異なる傾向が読み取れる。日本男性では、1000 番台、2000 番台、3000 番台において企業役職(V=0.21)、専門職(V=0.15)を回答する傾向がある。また、3000 番台、4000 番台では、一般雇用(V=0.13)を回答する傾向がある。5000 番台は店主(V=0.14)を回答する傾向がある。7000 番台は、工場経営(V=0.14)を回答する傾向がある。2000 番台は、先生(V=0.16)を回答する傾向がある。台湾男性では、1000 番台、2000 番台、3000 番台、4000 番台において企業経営(V=0.12)を回答する傾向がある。9000 番台、7000 番台では、工員・労働者(V=0.16)を回答する傾向がある。韓国男性では、2000 番台で医師(V=0.23)を回答する傾向がある。4000 番台、3000 番台、続いて 2000

番台で、税理士・会計士($V=0.20$)を回答する傾向がある。2000 番台、続いて 4000 番台、2000 番台においても弁護士($V=0.17$)を回答する傾向がある。4000 番台、続いて 2000 番台、1000 番台において、銀行課長($V=0.17$)を回答する傾向がある。2000 番台、4000 番台において、公務員課長以上($V=0.19$)を回答し、1000 番台において回答しない傾向がある。2000 番台、1000 番台において、企業社長($V=0.14$)を回答し、6000 番台において回答しない傾向がある。1000 番台において、政治家($V=0.12$)を回答する傾向がある。2000 番台において、聖職者($V=0.19$)を回答する傾向がある。9000 番台、7000 番台、4000 番台において組合役員($V=0.12$)を回答し、1000 番台で回答しない傾向がある。また、1000 番台、2000 番台などで警察官($V=0.18$)を回答し、8000 番台、6000 番台、7000 番台、9000 番台で回答しない傾向がある。必ずしも全ての結果において明瞭ではないものの、日本、台湾、韓国いずれにおいても、本人の職業階層によって生活圏が異なり、アクセス出来る職業に一定の傾向が存在することが分かる。

3. 日本・韓国における階層・関係指向の比較

次に、社会関係と社会階層に関連する価値指向について、比較分析を試みることにする。社会階層と価値についてはこれまで様々な分析が行われている(今田 1998, 2000; 井上 1998, 2000; 片瀬 1988; 片瀬・友枝 1990; 川端 1998; 木村 1998; 小堀・菅野 2007)。達成的地位や社会階層に関連する指向と、関係的地位や社会関係に関連する指向という枠組みが用いられることも多い。1995 年 SSM 調査と 2004 年韓国威信調査では、社会階層と社会関係に関連する価値指向について尋ねている(表 2)。以下では、「社会的評価の高い職業につくこと」=高い職業、「高い収入を得ること」=高い収入、「高い学歴を得ること」=高い学歴、「多くの財産を所有すること」=多くの財産、「高い地位につくこと」=高い地位、で表される価値志向を「階層指向」としてとりあげる。また、「家族から信頼と尊敬を得ること」=家族信頼、「ボランティア活動・町内会活動など社会活動で力を発揮すること」=ボランティア、「趣味やレジャーなどのサークルで中心的役割を担うこと」=趣味、で表される価値志向を「関係指向」としてとりあげる。また、回答選択肢については、重要でない=1 として、重要である=4 になるように、重要であるほど数値が高くなるように揃え、分析を行う。

3.1 日本と韓国における階層指向・関係指向の平均値

まず、階層指向と関係指向を表す項目について、分布を確認する(図 6)。図には、ボックスプロットが示しており、平均値も折線でプロットしている。また、X 軸のラベルには、変数項目名に合わせて平均値を表示し、続く括弧内に不偏標準偏差を表示している。日本では、

表 2: 階層指向と関係指向の項目

	1995 年 SSM 調査(日本)	2004 年韓国威信調査
階層指向	高い職業 高い収入 高い学歴 多くの財産 高い地位	社会的に評価が高い職業に従事すること 高い収入を得ること 高い学歴をもつこと 多くの財産を所有すること 高い地位につくこと
関係指向	家族信頼 ボランティア 趣味	家族から信頼と尊敬を受けること ボランティア活動、地域活動など社会活動に参加すること 趣味や余暇活動などの集まりで中心的な役割を担うこと

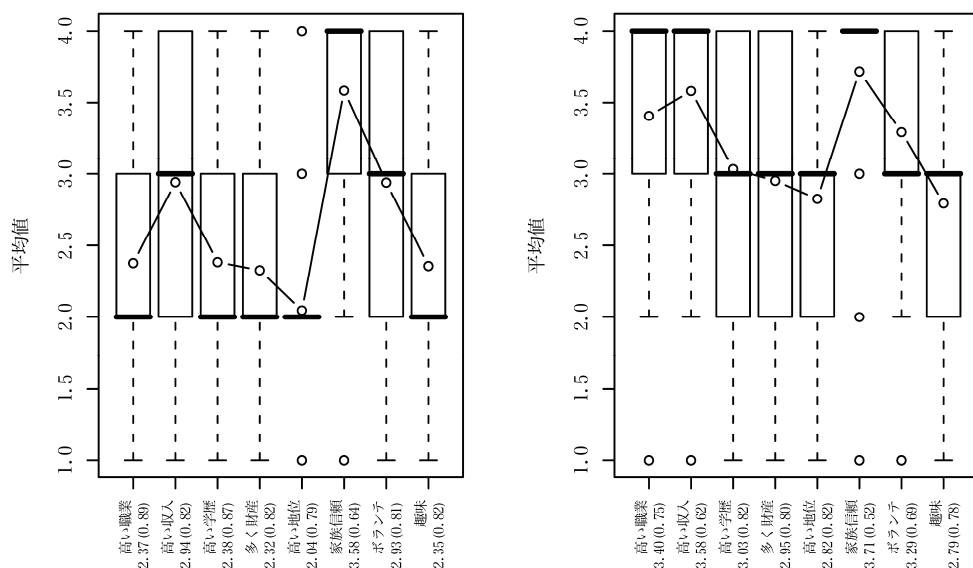


図 6: 階層指向・関係指向の項目のボックスプロットと記述統計（日本、韓国）

平均を見ると、家族との信頼尊敬が突出して高い(3.58)。続いて、高い収入(2.94)、ボランティア(2.93)が高めの平均値を示している。そして、高い職業(2.37)、高い学歴(2.38)、趣味・レジャー(2.35)、多く財産(2.32)が低めとなり、高い地位(2.04)が最も低い項目となっている。また、家族との信頼尊敬の不偏標準偏差が最も小さく(0.64)、その他の項目の標準偏差は 0.79-0.89 程度である。

韓国では、家族との信頼尊敬の平均値が日本と同様に最も高い(3.71)。高い収入(3.58)、評価の高い職業(3.40)、ボランティア(3.29)などの項目も平均値が高く、重視されている。そして、高い学歴(3.03)、多く財産(2.95)、高い地位(2.82)の平均値が低めとなっており、趣味・レジャー(2.79)が最も低い平均値を示している。日本と同様に、家族信頼の標準偏差が最も小さい(0.52)。その他の項目の標準偏差は 0.62-0.82 程度となっている。

全般的に、日本に比べて、韓国における方が平均値が高い。分散については、標準偏差の

値は韓国の方がやや小さめとなっており、平均が高めであるために、ばらつきが押さえられている可能性がある。また、平均値の相対的な高低については、どちらの国においても類似していることが分かる。例えば、高い学歴、高い財産、高い地位、趣味レジャーサークルは、相対的に見て、他の変数に比べて平均が低めである。また、家族との信頼と尊敬は、平均が際立って高い。高い収入も、他の項目に比較して、やや高めとなっている。評価の高い職業については、他の項目との相対的な比較では、日本におけるよりも韓国において重視されている傾向が分かる。

3.2 日本と韓国における社会的属性と階層指向・関係指向の平均値

次に、先ほどと同様に、これらの分布と、社会的属性ごとの平均値の違いを確認する。先ほどと同じように、グラフには X 軸にそれぞれの価値項目の名称と、クロス集計を行った場合の χ^2 乗検定の p 値、参考として Cramer の V 係数を示してある。前節で分析を行った、頼りになる相手の職業についての回答は、回答が少なく、平均（割合）が全体的に低めであった。しかし、年齢、学歴、収入、職業などごとに、回答には有意な差が見出され、かつ、中には関連が強いものも見出された。これに対して、ここで分析を行っている階層指向と関係指向については、標本数が多いために属性ごとに有意な違いが出てくるが、関連自体はあまり強くないものが多い。

日本では、多くの財産を除く全ての価値指向において、性別ごとに有意な差はあるが、大きな違いがあるとは言えない(図 7)。年齢別に見ると、価値指向全ての項目で有意差がある。ただし、大きな違いは見られない。学歴別に見ると、家族信頼尊敬を除く全ての項目で有意差が見られる。しかし、学歴によって、価値観が大きく異なるというほどの強い関連ではない。個人年収ごとに見ると、半数の価値項目において有意差はなく、有意であったとしても、大きな違いは確認出来ない。韓国では、性別ごとに見ると、有意でない項目が多く、大きな違いはないが、女性において、高い学歴($V=0.12$)を重視していることが分かる(図 8)。年齢ごとに見ると、年齢が高いほど、高い職業($V=0.18$)、高い学歴($V=0.12$)、家族信頼尊敬($V=0.12$)、ボランティア($V=0.11$)を重視する傾向がある。学歴ごとに見ると、初等学歴($V=0.12$)で高い職業を重視する傾向がある。個人年収ごとに見ると、高い学歴のみにおいて有意差があるが、大きな関連は見られない。

以上より、特に日本では、性別、年齢、学歴、個人年収による価値指向の違いはほとんど見られない。社会的属性との関連が弱いことは今田(1998)によって既に分析されている。韓国では、日本と異なり、年齢が若干効果を持っている点が興味深いが、性別、学歴、個人年収による価値指向の違いはあまり見られない。クロス集計を行うための、カテゴリーの

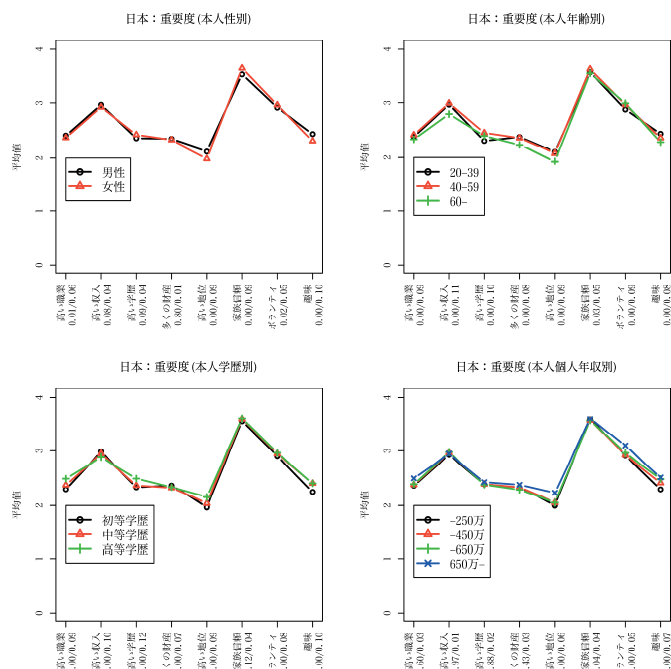


図 7: 日本：属性ごとの価値指向の平均値、 χ^2 乗検定 p 値、Cramer の V 係数

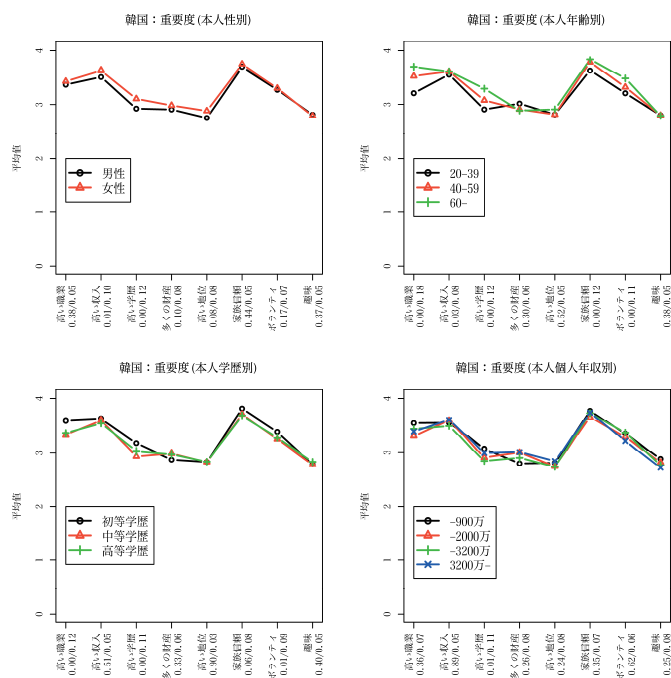


図 8: 韓国：属性ごとの価値指向の平均値、 χ^2 乗検定 p 値、Cramer の V 係数

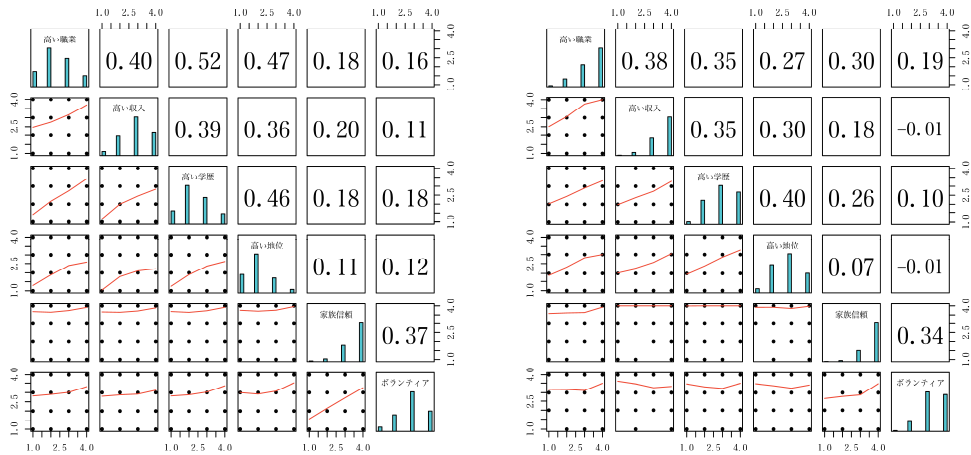


図 9: 階層指向／関係指向の項目の散布図と相関係数（日本、韓国）

区分けの仕方によって結果が変わる可能性はあるが、少なくともこれらの大まかな分析を行った範囲では、階層指向・関係指向の価値指向は、社会的属性によってあまり異なっていないと言える。

3.3 日本と韓国における階層指向・関係指向の相関構造

次に、これらの相関構造を探索(図 9)。図には、階層指向・関係指向の項目の散布図行列と相関係数行列、対角にそれぞれの度数分布が示してある。日本では、高い職業、高い収入、高い学歴、高い地位の間での相関は 0.36-0.52 と高めであり、家族信頼とボランティアの相関は 0.37、それ以外の相関は 0.2 以下と小さめになっている。韓国では、高い職業、高い収入、高い学歴、高い地位の間での相関は 0.27-0.40 であり、家族信頼とボランティアの相関は 0.34 となっている。それ以外の相関では、日本と同じく、多くが -0.01-0.19 以下で小さめである。これらをもとに、探索的因子分析と検証的因子分析を行った。どちらの国においても、階層指向と関係指向の二つの因子について分析を行うことが出来る。個別に探索的因子分析を行うと異なる結果の可能性もあるが、比較可能性をなるべく確保する方針で分析を行った。韓国においては、日本に比べて、階層指向と関係指向の関連が強い点が異なっている。

日本データでのモデル適合度は、 $\chi^2 = 17.62$ 、 $d.f. = 6$ 、 $p = 0.007$ 、 $GFI = 0.998$ 、 $AGFI = 0.992$ 、 $RMSEA = 0.028$ 、 $CFI = 0.996$ 、 $SRMR = 0.012$ である。韓国データでのモデル適合度は、 $\chi^2 = 36.43$ 、 $d.f. = 6$ 、 $p = 0.000$ 、 $GFI = 0.987$ 、 $AGFI = 0.954$ 、 $RMSEA = 0.076$ 、

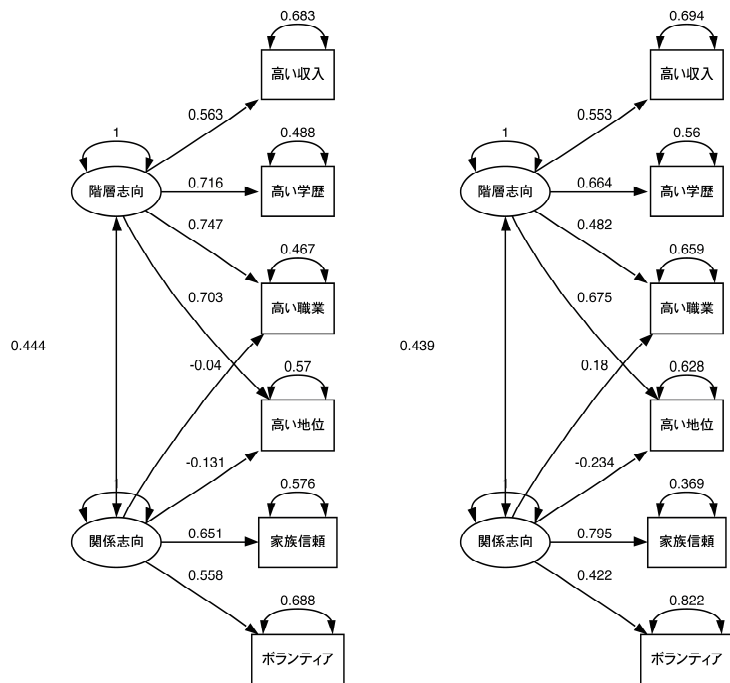


図 10: 階層指向と関係指向の個別 CFA、標準化解、日本(左)と韓国(右)

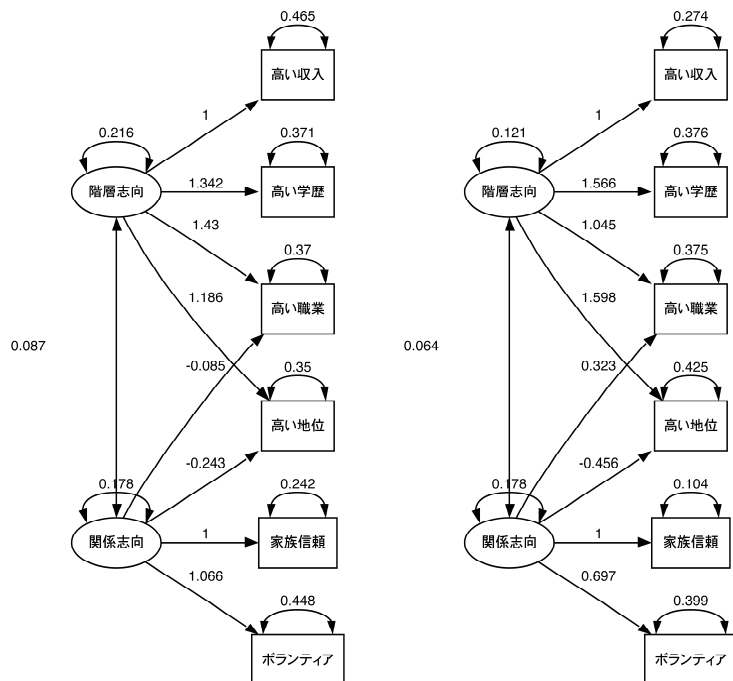


図 11: 階層指向と関係指向の個別 CFA、非標準化解、日本(左)と韓国(右)

CFI = 0.959, SRMR = 0.034 である。分析は R と AMOS で行い、同等の結果を得た。ここ

では、R library(sem) (Fox, 2006)で行った結果を図示している。適合度は χ^2 乗検定によって棄却されるが、サンプル数の多さを考えると特に問題はない。また、GFI、AGFI、RMSEA、CFI、SRMR それぞれの指標を見ても適合度は悪くないと言える。

日本の分析結果を見ると、図に示してある標準化解より、高い収入、高い学歴、高い職業、高い地位に階層指向因子が影響を与えている。関係指向は、家族信頼、ボランティアに影響を与えており、高い地位に対しては弱い負の効果を有している。図 9 でも確認が出来るが、高い職業、高い収入、高い学歴、高い地位の間の相関は 0.4 程度以上であり、階層指向因子としてまとまる。また、これらと家族信頼、ボランティアとの相関は 0.2 以下程度と相対的に低くなっており、関係指向因子として分かれる。しかし、高い地位と、家族信頼、ボランティアとの相関は 0.1 程度と、さらに一段と低くなっており、さらに、階層指向因子と関係指向因子に 0.4 程度の正の相関があるために、関係指向因子から高い地位へ負のパスが通っている。この傾向は、韓国においてさらに顕著となっている。

日本では、関係指向因子から、評価が高い職業へのパスの値が標準化解で -0.04 と小さく、有意でない。パスを外した場合の適合度は、 $\text{chisq} = 19.52$ 、 $\text{d.f.} = 7$ 、 $p = 0.007$ 、 $\text{GFI} = 0.997$ 、 $\text{AGFI} = 0.992$ 、 $\text{RMSEA} = 0.027$ 、 $\text{CFI} = 0.996$ 、 $\text{SRMR} = 0.013$ であり、その他のパスにほとんど変化はない。階層指向因子と関係指向因子の間の相関は、日本で 0.444、韓国で 0.439 であり、階層指向と関係指向の間にそれなりの関連が確認出来る。韓国の分析結果も日本と類似の傾向を示しているが、パスの値の大小はやや異なっている。特に、韓国では関係指向因子から、高い地位と高い職業へのパスの値の絶対値が大きめであり、日本の方が単純構造に近い。

日本と韓国の比較のために、非標準化解を確認すると、関係指向の分散は、日本でも韓国でも 0.178 とほぼ同じであるが、階層指向の分散は、韓国の 0.121 に比べ、日本は 0.216 と大きい。また、両国で非標準化解での因子負荷量は異なるものの、個別の検証的因子分析を行うことに加え、日本と韓国の二つの母集団で、母集団間に母数制約をつけずに同時に解析する配置不変モデル、母集団間でそれぞれ対応する観測変数に対する因子パターンに等価制約をつける測定不変モデルについても AMOS で分析を行った。多母集団の同時解析の結果、配置不変モデルでは、 $\text{chisq} = 54.548$ 、 $\text{d.f.} = 12$ 、 $p = 0.000$ 、 $\text{GFI} = 0.995$ 、 $\text{AGFI} = 0.982$ 、 $\text{RMSEA} = 0.032$ 、 $\text{CFI} = 0.988$ 、 $\text{SRMR} = 0.012$ 、 $\text{AIC} = 114.548$ であり、適合はよいと言える。また、測定不変モデルでは、 $\text{chisq} = 98.977$ 、 $\text{d.f.} = 18$ 、 $p = 0.000$ 、 $\text{GFI} = 0.990$ 、 $\text{AGFI} = 0.977$ 、 $\text{RMSEA} = 0.037$ 、 $\text{CFI} = 0.977$ 、 $\text{SRMR} = 0.018$ 、 $\text{AIC} = 146.977$ である。モデル間の検定を行うと、 $\text{chisq} = 44.430$ 、 $\text{d.f.} = 6$ 、 $p = .000$ であり、適合がやや落ちるが、測定不変モデルにおいても、全体的な適合度はよいと言える。日本と韓国とで、階層指向と関係指向における類似の価値指向の構造を見出し、共通の尺度を用いて比較分析を進めていくことは可能であると思われる。

4. 議論

本研究では、求職・転職の際に頼る相手・あるいは頼み事をする相手についての地位ジェネレータによる社会的ネットワークの測定と、関係指向・階層指向という、価値指向における主観的な社会的ネットワークのあり方について、国際比較分析を試みた。社会的ネットワークを測定する際に用いられる地位ジェネレータという枠組みについては、日本・台湾・韓国で共通だが、それぞれ問いかけ内容、回答選択肢が異なっている。このように、単に国際比較を適用することには注意が必要である場合をとりあげ、それぞれの指標について、記述的に社会的属性との基礎的な関連について確認した。また、日本・韓国で、ほぼ同じ問いかけ内容・ほぼ同じ回答選択肢であり、もともと国際比較が念頭に置かれている場合として、階層指向と関係指向について比較分析を行った。

これらの分析結果より、日本と台湾での、求職・転職の際に頼りにする相手の職業を尋ねた地位ジェネレータによる測定方法では、回答選択肢が異なるとはいえ、ある程度、社会階層や社会構造、また、実際の職業分布やアクセスのしやすさを反映していることが示唆される結果が確認されたと言える。実際の就職や転職の場面において、これらの回答がどのような効果を有しているのか、また、日本と台湾で効果に違いがあるのか、等の分析には他の変数との関連をさらに探る必要がある。韓国においては、頼み事をする際の相手の職業を尋ねており、回答選択肢も日本、台湾と若干異なるため、さらに判断は難しいが、地位ジェネレータによる測定の方法自体は、社会的属性による違いを浮かび上がらせることが有効であると確認出来たと言える。国際比較調査において、質問内容や回答選択肢を機械的にまったく同一にすることが必ずしも最適な測定法であるとは限らない。得られた回答をそれぞれの国・文化の中での相対的位置付けに置き直すことによって、より理解が深まる可能性があるといえる。また、職業威信スコアは国際比較研究において安定していると言われているが、ネットワークの測定において、地位ジェネレータにどのような職業を含めればよいのかという点について国際比較研究はあまりなされていないため、これらの分析結果は多少なりとも今後の研究につながると言える。

また、価値指向について、日本と韓国で分析を行った結果として、社会階層に関連する価値指向と、社会関係に関連する価値指向に大きく分かれる点で類似点が見出された。多母集団解析により、配置不変と測定不変をある程度確認出来たことは、今後の分析での一つのよりどころとなる。その際に、因子分析の過程で、なるべく国際比較が可能になるように共通の結果が表れるように分析を進めているため、「多くの財産」、「趣味・レジャーで中心的役割」については、分析で必ずしも奇麗にまとまらず、最終的な分析では外した結果を今回は報告している。国際比較分析に限らないが、ある程度共通の物差しを構築して複数の母集

団を比較する際には、調査票に含まれる全ての項目を分析に含めることはなかなか困難な作業であるといえる(Sugano 2005)。これら二つの変数を入れることで、共通の相関構造に奇麗にまとまらないというのは、ある意味では、これらの変数が、両国におけるそれぞれの独自の要因や関連構造の違いを強く表しているからという可能性がある。また、韓国においては若干年齢が効果を持っていたものの、両国において関係指向と階層指向へ効果を持つ有効な変数の探索も必要であり、今後さらに検討をする必要がある。価値観については、様々な国際比較研究蓄積してきているが、1995 年 SSM 調査で採用され、今田(1998)、井上(1998)等での分析枠組みの国際比較研究についてはまだあまり研究が進んでいないと言える。特に、社会関係に関連する項目の比較は文化的な違いから困難な側面もあるが、今後さらに詳細な分析が望まれる。

文献

- Fox, J. 2006. "Structural Equation Modeling With the sem Package in R." *Structural Equation Modeling*, 13:465-486.
- 今田高俊. 1998. 「社会階層の新次元: ポスト物質社会における地位変数」今田高俊編『社会階層の新次元を求めて』1995 年 SSM 調査シリーズ 20. 1995 年 SSM 調査研究会: 25-43.
- 今田高俊. 2000. 「ポストモダン時代の社会階層」今田高俊編『社会階層のポストモダン』東京大学出版会: 3-53.
- 井上寛. 1998. 「社会階層指向かそれとも脱階層指向か」今田高俊編『社会階層の新次元を求めて』1995 年 SSM 調査シリーズ 20. 1995 年 SSM 調査研究会: 45-56.
- 井上寛. 2000. 「脱-階層指向の状況と構造」今田高俊編『社会階層のポストモダン』東京大学出版会: 57-81.
- 片瀬一男. 1988. 「社会階層と価値志向: ライフスタイルにおける階層志向と私生活志向」『1985 年社会階層と社会移動全国調査報告書』第 2 巻. 社会階層と社会移動全国調査委員会: 201-223.
- 片瀬一男・友枝敏雄. 1990. 「価値意識: 社会階層をめぐる価値志向の現在」原純輔編『現代日本の階層構造』第 2 巻. 東京大学出版会: 125-147.
- Katz-Gerro, T., 2002. "Highbrow Cultural Consumption and Class Distinction in Italy, Israel, West Germany, Sweden, and the United States." *Social Forces* 81(1): 207-229.
- 川端亮. 1998. 「階層意識に関わる変数の探索: 地位的変数と関係的変数を中心に」尾嶋史章編『ジェンダーと階層意識』1995 年 SSM 調査シリーズ 14. 1995 年 SSM 調査研究会: 81-92.
- 木村好美. 1998. 「高齢者の社会的地位とライフスタイル: 役割喪失と社会階層が価値意識・自己評価に及ぼす影響」白倉幸男編『社会階層とライフスタイル』1995 年 SSM 調査シリーズ 17. 1995 年 SSM 調査研究会: 125-134.
- 小堀真・菅野剛. 2007. 「達成的地位志向・関係的地位志向の社会的ネットワークへの影響」『研究紀要』73. 日本大学文理学部人文科学研究所: 47-61.
- R Development Core Team. 2005. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Revelle, W. 2007. "User manual for the psych package Version 1.0.27." <http://www.personality-project.org/>.

Sugano, T. 2005. "Leisure Activities in Germany and Japan: Comparative Analysis with ALLBUS 1998 and JIS 2001." 社会学論叢 153: 1-24.

Cross Cultural Comparative Analysis of Measurements on Social Network between Japan, Taiwan and Korea

Tsuyoshi SUGANO

Nihon University

The aim of this paper is to conduct a cross-cultural comparative analysis on both of subjective and objective measurements on social network. Position generator measurements on social network are analyzed among 2005 SSM Survey in Japan, 2005 SSM Survey in Korea and 2005 SSM Survey in Taiwan. Although there are several differences in questionnaires and response items, similar relationships between social structure and social networks are suggested. Social orientation on status achievements and social relationships are also compared between 1995 SSM Survey in Japan and 2004 Prestige Survey in Korea. Means of social orientations are higher in Korean than in Japanese, and the correlation structure is similar in both countries. Structural equation modeling with multi-population analysis clarified that goodness-of-fit in metric invariant model is acceptable.

Keywords and phrases: Social Network, Position Generator, Value Orientation on Social Relationship, Metric Invariant

サンプル選択の影響分析

－欠損値・打ち切りデータ－

高田 洋
(札幌学院大学)

【要旨】

社会調査においては、どんなデータであっても得られたものは偏りを有しているかもしれない。特に、社会科学のデータにおいては、標準的なサンプリングの手法に基づいていても、すべてのサンプルについて、すべての変数の情報を得られることはむしろ少なく、通常は、欠損値やデータの利用の困難さのために、欠けているのが通例である。本論文では、Tobit モデルの Heckman による 2 段階推定法を利用することにより、サンプルが選択される危険性を示すハザード比を推定し、どのような属性を持ったサンプルが脱落する可能性が高いかを、SSM データで得られた変数間を比較することによって、間接的に検討する。その結果、脱落する可能性の高いサンプルは、(1) 不安定雇用であること、(2) 学歴が高いこと、(3) 大企業ホワイトカラー、(4) 住居が一戸建てでないこと、(5) 年齢が若いことであった。このようなサンプルは、欠損する可能性が高い。ただし、欠損する可能性が高いということは、回収されなかったということではない。

キーワード： サンプル選択、欠損値、Tobit モデル

1. サンプル選択の影響評価

社会調査は、データの一般性を得るために、母集団の縮図であるようにサンプリングされている。しかし、実際に得られるデータは、どんなに注意深く調査を設計し実査を行ったとしても、サンプリングされたすべての対象者について回答を得ることは不可能である。また、近年では、調査そのものに対する信頼の低下や、プライバシー意識の高まりによって、以前より協力を得られにくくなっている。回収率の問題は、以前とは比べものにならないくらい大きな問題になっている。

ただし、調査に協力するのをためらったり、批判的になったりすることは、国民全体の権利意識の高まりや、学術研究そのものに対する健全な批判的精神であり、むしろ、研究者にとって歓迎すべきことでもありうる。回収率の低下は社会調査にとって重要な問題ではあるが、学術研究そのものについては研究の意義や方法が、改めて責任を持つものとして問われることになり、長期的に見れば、学問的な発展にむしろ寄与することもある。以前の工夫をしなくても協力を得られたような時代にはあり得なかった調査研究の意義や結果の説明を問われることになり、それはそれでよい面もある。

しかしながら、標本調査設計どおりではない欠損したデータを分析することは、一般的には注意を有する。通常の統計的分析では、母集団から無作為に抽出されたサンプルであることが仮定されているので、回収率が低いデータをそのまま分析することは誤った結果になる場合がある。それでもデータの欠け方がランダムである場合には問題にならない。問題は、欠け方がランダムでない場合である。その場合は、どのような偏りがあるのかを知らなくてはならない。分析がうまくいかなかった場合、その原因を回収率の低さに求めてしまいがちであるが、その前にバイアスがどのような影響を与えているのかを評価する必要がある。

さらに、社会調査においては、どんなデータであっても得られたものは偏りを有しているかもしれないと考えるべきである。特に、社会科学のデータにおいては、標準的なサンプリングの手法に基づいていても、すべてのサンプルについて、すべての変数の情報を得られることはむしろ少なく、通常は、欠損値やデータの利用困難さのために、欠けているのが通例である。

得られたサンプルは、一種の選択されたデータである。その選択がランダムであったり、ランダムに近い場合は問題がない。しかし、何らかの選択の偏りがある場合、分析によって予測された推定値はその偏りに影響を受けているかもしれない。従って、その影響の度合いがどれくらいかを明らかにすることは重要である。推定された値がサンプルの偏りに強く影響されているならば、推定値は一般性を持たない。本研究では、欠損の傾向を評価し、それがどのように分析結果に影響を与えるのかを分析する。

2. 打ち切りのあるデータ

本論では、特に線型モデルにおけるバイアスを評価する。回帰分析は、一つの従属変数について複数の独立変数があることが通常であるが、独立変数と関連しないで生ずる欠損や打ち切りは、分析結果に影響を与えない。影響を与えないということは、欠損がランダムなものとして処理されるということである。この場合、誤差の相関が 0 であるという回帰式の仮定は保たれたままである。問題は、独立変数と関連して生ずる欠損や打ち切りであり、それは分析結果に影響を与える。そのようなデータを通常の最小自乗法を基礎とした線型モデルで分析すると誤った結論を導くことがある。典型的な例として、打ち切りのあるデータ(truncated data)と抑制されたデータ(censored data)の例を述べる。

回帰分析において従属変数のある値についての打ち切りがあるサンプルは、回帰係数に影響を与える。その例として、Richard A Berk (1983)は、貧困層の教育と収入の関係を挙げている。貧困層の研究において、貧困線以下の層にかぎられたサンプルを対象とした場合、収入を従属変数、教育年数を独立変数とした回帰分析では、貧困線以上のサンプルがないために、推定された回帰係数は小さくなり、教育と収入は見かけ上弱い関係となる。これは、独立変数の収入の低さに関連して、従属変数の打ち切りが起こるために生ずる偏りである。このように独立変数と関連して従属変数につ

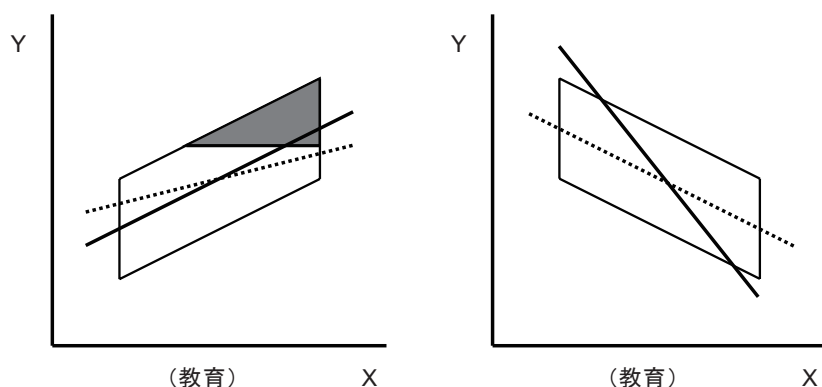


図 1 打ち切りのあるデータ

いて打ち切られたデータを有するサンプルの分析は、そのサンプルの偏りのために誤った結果を導く。

また、被説明変数の選択肢が抑制されているために、回帰式の推定を誤る場合がある。たとえば、教育を独立変数、市民運動の参加の程度を従属変数とする回帰分析を考えてみよう。SSM では、市民運動の参加は、「いつもしている、よくしている、ときどきしている、めったにしない、したことがない」の選択肢が設けられている。この選択肢の中で、たとえば、「したことがない」は抑制された選択肢の一例である。この中には、「積極的にしない」から、「参加はしたいがやむを得ない理由でできない」までが含まれるが、それが一つの選択肢に圧縮されるために、回帰分析が過少評価されているかもしれない。圧縮されることを抑制された (censored) と言うが、もしも、その抑制が独立変数である教育と関連して起きるならば、通常の回帰分析の結果は誤った結果となる。

前者の例は図 1 の左の図、後者は右の図である。打ち切りデータがないときの推定された回帰直線は実線で、打ち切りデータがあるときの回帰直線は破線で示してある。灰色の部分とは、打ち切りデータの部分である。このようにサンプルに選択の偏りがあるデータに基づいた分析は、間違った推定結果となる可能性がある。

3. Tobit モデルと Heckman 推定

このようなサンプルの選択の偏りがある場合、通常の回帰分析によって推定をおこなうことはできない。バイアスを考慮した分析が必要となる。ここでは、Tobit モデルの Heckman 推定を利用する。Heckman 推定は、2 段階推定と呼ばれ、第 1 段階で選択モデルを第 2 段階で目的モデルの推定をおこなう。はじめに、1 段階目でサンプル選択・非選択の 2 値変数を従属変数とした推定をおこない、その推定値を基にハザード比を計算する。次に、2 段階目で、そのハザード比を目的とする回帰分析の一つの独立変数とすることにより、回帰分析の偏りをコントロールする。

通常の回帰式、

$$y_i = \mathbf{x}_i' \boldsymbol{\beta} + u_i \quad (1)$$

において、 $\mathbf{x}$ と $\boldsymbol{\beta}$ は独立変数と回帰係数のベクトル、 u は分散 σ^2 、平均 0 の誤差である。通常の最少自乗法では、誤差の期待値は 0 であるので、 y の期待値は次になる。

$$E(y_i) = \mathbf{x}_i' \boldsymbol{\beta} \quad (2)$$

サンプル選択や打ち切りが y のある値 c で起きた場合、 y が c を越えたときデータは観測され、 y が c を越えないときデータは観測されない。このとき、 $c=0$ とおけば、 $y>0$ のときの期待値は次のようになる。

$$E(y_i) = \mathbf{x}_i' \boldsymbol{\beta} + E(u_i | y_i > 0) \quad (3)$$

つまり、誤差の期待値は 0 にならない。(3) に (1) を代入して変形すると、

$$E(y_i) = \mathbf{x}_i' \boldsymbol{\beta} + E(u_i | u_i > -\mathbf{x}_i' \boldsymbol{\beta}) \quad (4)$$

となる。 u は、分散 σ^2 、平均 0 の正規分布に元々従っているので、右辺第 2 項の期待値は、 $-\infty$ から $\mathbf{x}_i' \boldsymbol{\beta}$ までの切断された正規分布の期待値 ($pr(u_i > -\mathbf{x}_i' \boldsymbol{\beta}) = pr(u_i < \mathbf{x}_i' \boldsymbol{\beta})$) ということになる。標準正規分布関数を Φ 、密度関数を ϕ とすると

$$E(u_i | u_i > -\mathbf{x}_i' \boldsymbol{\beta}) = \sigma \frac{\phi_i}{\Phi_i} \quad (5)$$

となり、 ϕ/Φ はハザード比であるから、それを λ とすると、

$$E(y_i) = \mathbf{x}_i' \boldsymbol{\beta} + \sigma \lambda_i \quad (6)$$

となる。

λ は未知であるので、それを推定するために、サンプルに選択されているか選択されていないのかの 2 値変数に対してプロビットモデルで推定をおこないその推定をおこなう。各サンプルに対するプロビットモデルの予測値が Φ にあたるので、そのときの ϕ を計算し、比を取る。その予測されたハザード比を $\hat{\lambda}$ とすれば、次のようになる。

$$E(y_i) = \mathbf{x}_i' \boldsymbol{\beta} + \sigma \hat{\lambda}_i \quad (7)$$

Heckman 推定の詳細については、Richard Breen (1996) や G. S. Maddala (1992) にまとめられている。

本論文の分析では、1 段階目のサンプル選択の推定値を基に、2005 年 SSM 調査における各変数との関係を分析し、そのバイアスの評価を試みる。

4. サンプル選択ハザードの推定

4.1 サンプル選択ハザードの推定

2005 年 SSM データの欠票データを用い、回収された／されないを従属変数とする。その従属変数に対し、性別、住居形態、居住都道府県、居住地域の都市規模、年齢階層を独立変数としたプロビット分析をおこなった。その結果、次のような予測式が推定された（高田、

2008)。

$$\begin{aligned}\hat{\Phi}_i = & .547^* - .168^* s - .268^* h \\ & - .005r_1 - .095^* r_2 - .374^* r_3 - .400^* r_4 - .030r_5 + .113r_6 \\ & - .077^* r_7 - .360^* r_8 - .322^* r_9 - .128^* r_{10} - .106r_{11} \\ & - .336^* u_1 - .273^* u_2 - .227^* u_3 - .087^* u_4 \\ & - .498^* a_1 - .303^* a_2 - .235^* a_3 - .226^* a_4\end{aligned}\quad (8)$$

(* $p < .10$)

s は性別（男性=1、女性=0）、 h は住居形態（一戸建て=0、それ以外=1）、 r は地域（順に、北海道、東北、関東、京浜、甲信越、北陸、東海、近畿、阪神、中国、四国、参照カテゴリーは九州）、 u は都市規模（順に、政令都市、25 万人以上の市、10 万人以上の市、10 万人未満の都市、参照カテゴリーは郡部）、 a は年齢層（順に、20 代、30 代、40 代、50 代、参照カテゴリーは 60 代以上）である。

予測値から λ を計算する。この値は、サンプルに脱落しやすい危険性を示す変数である。このハザード比を様々な変数において比較することにより、間接的ではあるが、どのような人が脱落しやすく、どのような人を選択しやすいのかを明らかにすることができる。ただし、注意してもらいたいのは、これはバイアスそのものの評価ではないということである。示されるのは、各変数の値ごとの脱落の可能性であり、実際に脱落しているわけではない。ハザード比が高いのにも関わらず、回収できたサンプルは、脱落する可能性は高いのに、脱落しなかったことを示している。ここでは、各変数の値ごとに危険可能性を比較することにより、間接的にどのような属性をもつサンプルが選択されない可能性が高いのかを明らかにする。

4.2 2005 年 SSM データにおけるサンプル選択の評価

(1) 基本属性

まず、基本的な属性ごとのサンプル選択ハザードを見ていく。性別ごとのサンプル選択ハザードの平均は、危険率 1% 以下での有意な差があった。男性の方が女性よりも、サンプル脱落の危険性が高い。就業の有無については、同じく 1% 以下有意で、無職者の方がサンプル脱落の可能性が高い。留め置き票の種別については、有意な差はない。

年齢が若いほど脱落する可能性が高い。統計的にも有意である。年齢との相関係数は、-.508 であり、これも 1% 以下有意である。ただし、年齢は、ハザード比を計算するプロビットモデルにおいて有意な要因であったので当然の結果ではある。兄弟姉妹数は、9 人以上になるとサンプルが各 50 以下になり小さいが、それでもきょうだい数が多いほど、ハザード比が低くなるという傾向が見られる。3 人までと 4 人以上の差が大きい。これは部分的には時代効果によるものといえる。3 人まではそれほどの差がない。住居形態においては、一戸建てと

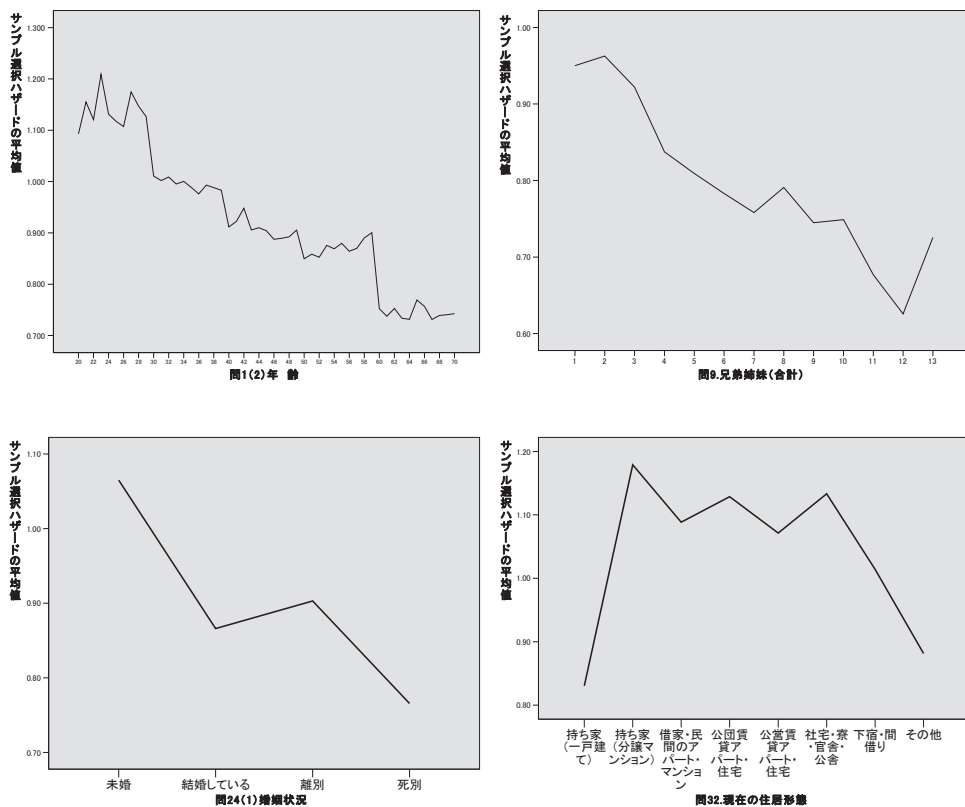


図2 年齢・きょうだい数・婚姻状況・住居形態
によるサンプル選択ハザードの平均値

左上：年齢、右上：きょうだい数、左下：結婚状況、右下：住居形態

すべて分散分析の結果 $P < .001$ で有意

他のカテゴリーとの差が顕著である。一戸建ては捕まえやすく、サンプル脱落の可能性が低い。

(2) 職業

次に職業の項目について、男女別にハザード比を比較してみた。まず、従業上の地位であるが、男女とも、ハザード比が低く脱落しにくいのは、経営者、臨時雇用、自営業、脱落しやすいのは、派遣社員、仕事を探している無職、学生である。男女ともほぼ同じような傾向であるが、女性において派遣社員が脱落しやすいのが特徴的である。配偶者の従業上の地位についても同じ傾向が見て取れる。派遣社員と休職中の無職者が、サンプルに脱落する可能性が高い。不安定雇用者が捕まえにくいという傾向が出ている。

従業員数別のハザードは、全体の傾向が男女別にしても同じであった。つまり、官公庁を

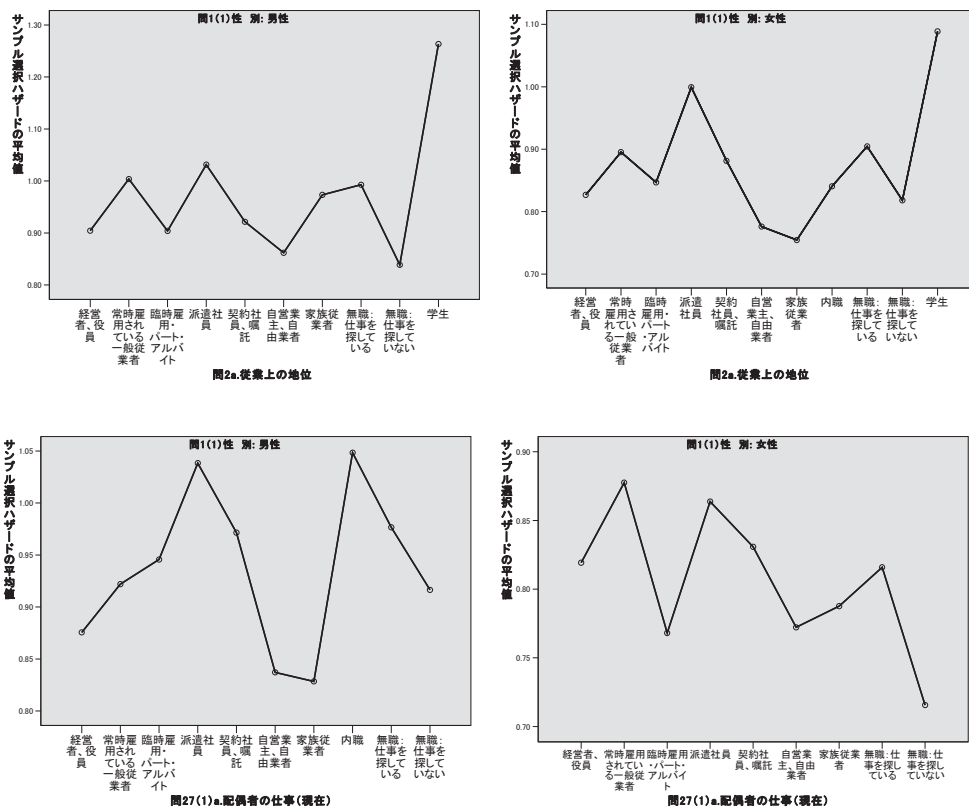


図 3 従業上の地位によるサンプル選択ハザードの平均値

左：男性、右：女性、上：本人、下：配偶者

すべて分散分析の結果 $P < .001$ で有意

除けば、従業員数が多いほど、サンプルが脱落する可能性が高い。役職別には、M 字型に近い台形方であり、役職なしと経営者が脱落する可能性が低く、中間の役職が脱落する可能性が高い。労働時間は、外れ値を除けば労働時間が長いほど、脱落しやすいという傾向がある。忙しい人は、調査に答えてもらいにくい。ただし、男女別に見ると、これは男性だけの傾向で、女性は労働時間にかかわらず一定のレベルを示している。女性の場合、外れ値を除けば労働時間による差は有意でなくなる。

職業分類は、大企業、ホワイトカラーが脱落しやすく、小規模、ブルーカラーが脱落しにくい。また、農業が脱落しにくいことが顕著である。職業威信との相関は、すべて正で有意である。職業威信が高いほど、脱落しやすい傾向がある。

(3) 教育・文化資本、出身階層

本の冊数が多いほど、ハザード比が高いというほぼ線型の関係が見られる。学歴が高いと、

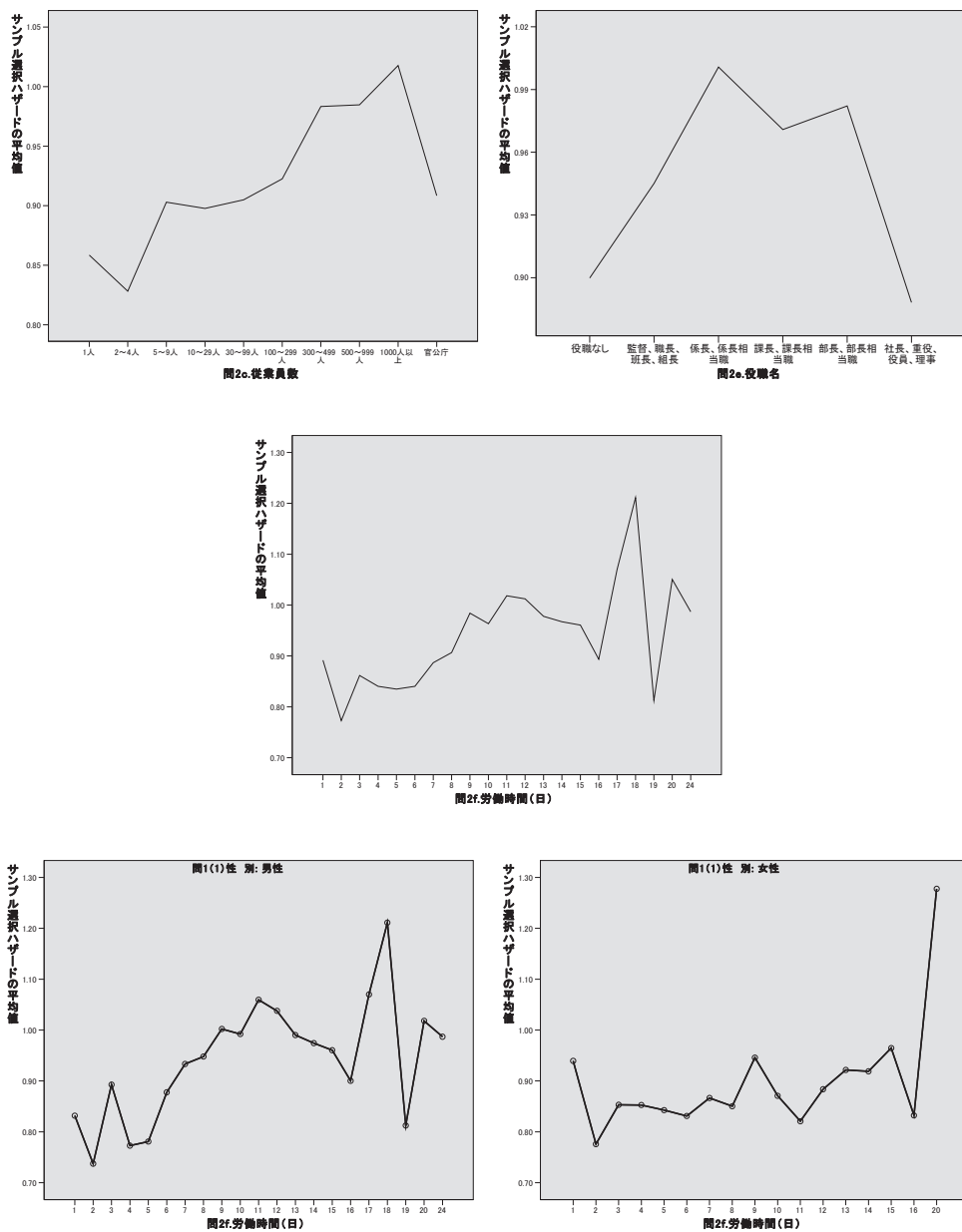


図4 従業員数、役職、労働時間によるサンプル選択ハザードの平均値

上左：従業員数、上右：役職、中：労働時間、下左：男性、下右：女性、

右下を除いてすべて分散分析の結果 $P < .001$ で有意、右下のみ $P < .01$ で有意

ハザード比が高くなる。教育を受けた人ほど、調査に協力しにくい傾向がある。また、父親の学歴と母親の学歴においても同様な傾向が見られる。

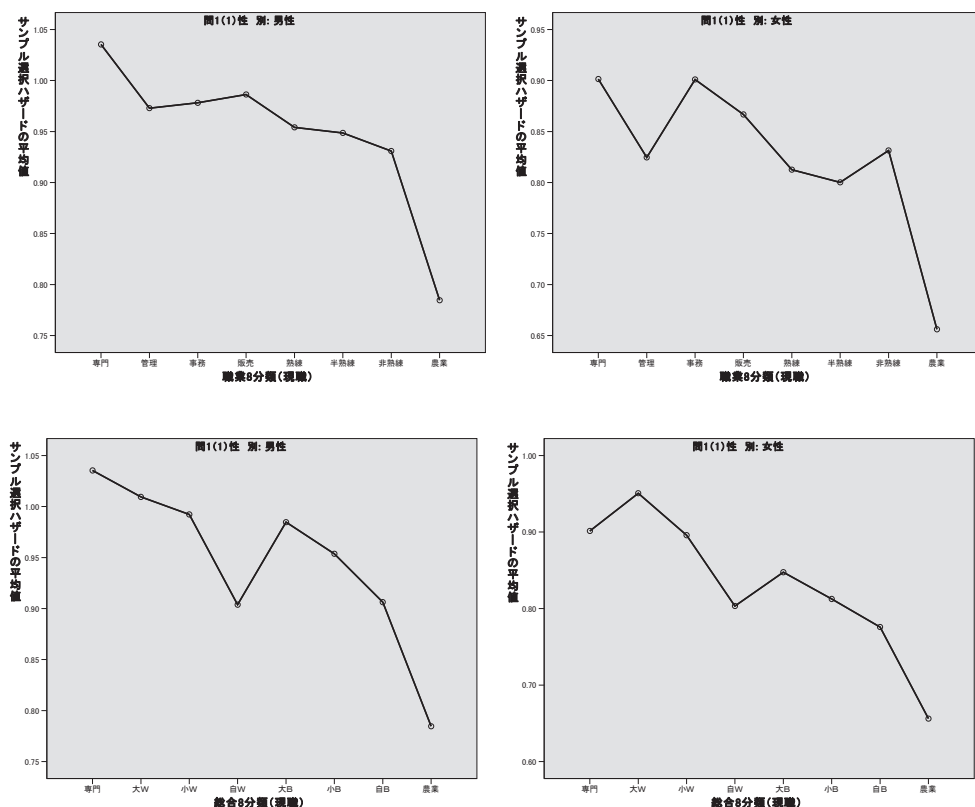


図 5 職業分類によるサンプル選択ハザードの平均値

上：SSM 職業 8 分類、下：SSM 総合職業 8 分類、左：男性、右：女性

すべて分散分析の結果 $P < .001$ で有意

表 1 ハザード比と職業威信スコアの相関係数

		現職	初職	15歳時父職	父主職	母職	配偶者現職	配偶者結婚時職
男性	<i>r</i>	.101	.080	.182	.190	.090	.109	.096
	<i>p</i>	.000	.000	.000	.000	.000	.000	.001
	<i>N</i>	2196	2575	2225	2324	1672	1018	1297
女性	<i>r</i>	.117	.110	.168	.160	.151	.097	.093
	<i>p</i>	.000	.000	.000	.000	.000	.000	.000
	<i>N</i>	1932	2961	2574	2678	2090	1799	2094

収入との関係では、男女別の相関係数を見た場合、男性の場合有意ではない。女性の場合は、有意であり、収入が高いほど脱落しやすいという傾向がみられる。それぞれのカテゴリーごとのプロットを見ると、実は単純な線型関係ではなく、収入が高い人と低い人が脱落しにくく、中間が脱落しやすいという傾向が見て取れる。男性は、本人の収入が U 型にハザードとの関係を持つのに対し、女性は配偶者の収入がほぼ正の線型関係でハザードの関係を持っている。

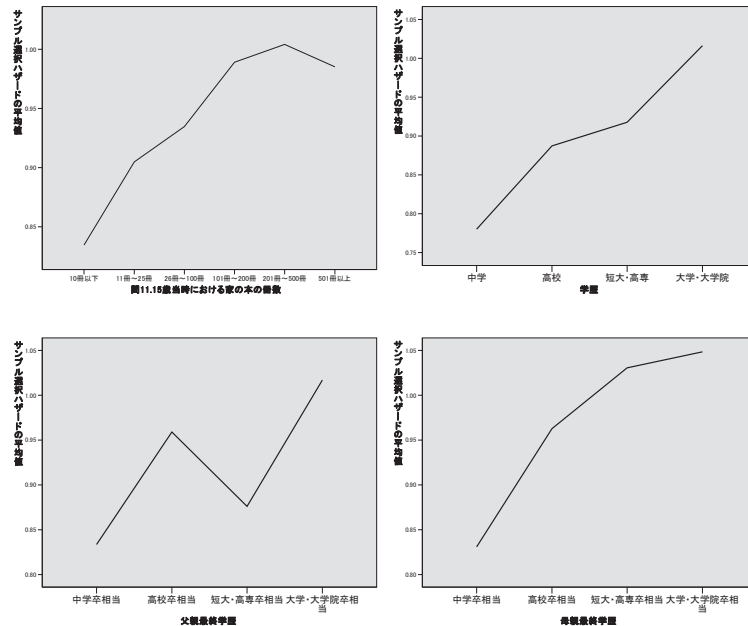


図 6 文化資本と教育によるサンプル選択ハザードの平均値

左上：文化資本、右上：本人学歴、左下：父最終学歴、右下：母最終学歴

すべて分散分析の結果 $P < .001$ で有意

表 2 ハザード比と収入の相関係数

		対象者の年収	配偶者の年収	世帯の年収
男性	r	.001	-.018	.036
	p	.965	.486	.129
	N	2278	1556	1797
女性	r	.040	.221	.069
	p	.039	.000	.003
	N	2694	1649	1857

5. 結論

ここまでの分析によって、間接的ながらも、サンプル選択されにくいケースを明らかにすることができた。脱落する可能性の高いサンプルは、(1) 不安定雇用であること、(2) 学歴が高いこと、(3) 大企業ホワイトカラー、(4) 住居が一戸建てでないこと、(5) 年齢が若いことであった。このようなサンプルは、欠損する可能性が高い。ただし、欠損する可能性が高いということは、回収されなかったということではない。

欠損の可能性がどのように分析結果に影響を与えているのかの 1 つの例として、簡単な地位達成モデルを用いて、1 段階目で推定されたハザードを、目的となる回帰モデルの説明変数として加えたモデルを評価してみよう。まず、本人の職業威信を従属変数 y とし、年齢、性別ダミー、父職業威信、父親学歴、母親学歴、本人学歴を独立変数とした重回帰分析を評

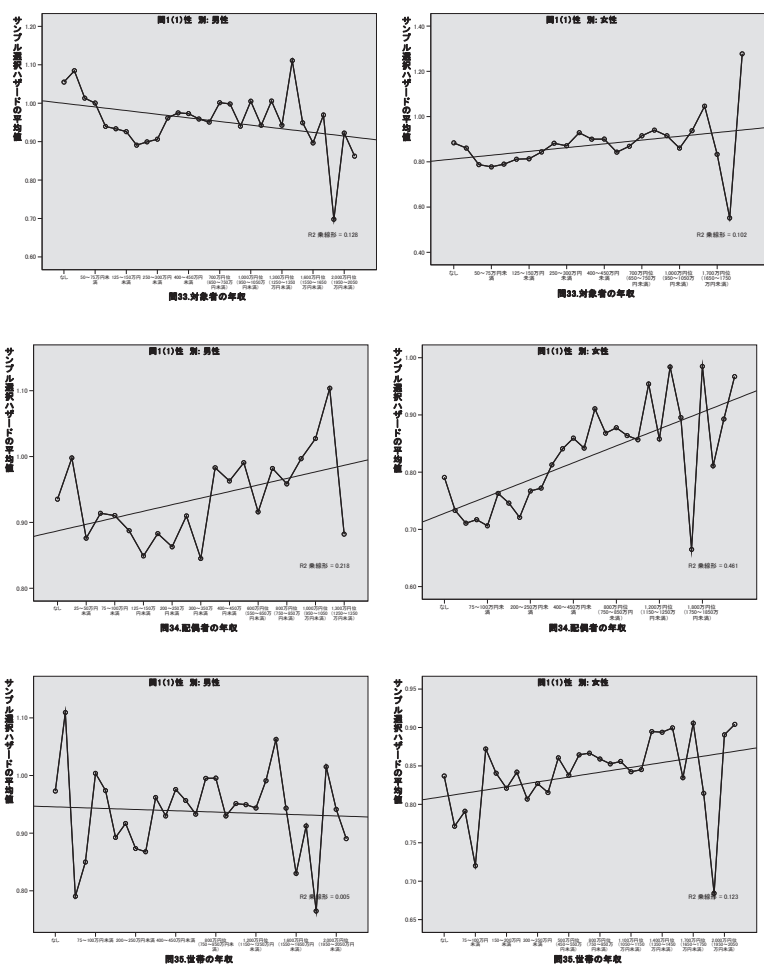


図 7 収入によるサンプル選択ハザードの平均値

左：男性、右：女性、上：本人収入、中：配偶者収入、下：家族収入

右下は非有意、その他は、分散分析の結果 $P < .01$ で有意

価してみた。次の結果を得られた（標準化回帰係数）。

$$\hat{y} = .071^* a - .121^* s + .105^* fp + .067^* fe + .017me + .372^* e \quad (9)$$

$$(R^2 = .231, N = 2876, ^* p < .01)$$

ここで、 a は年齢、 s は性別（男性=0、女性=1）、 fp は父親の職業威信、 fe は父親学歴、 me は母親学歴、 e は本人学歴である。

ここで、ハザード比を独立変数に加えることにより、バイアスをコントロールしてみよう。次の結果になった（標準化回帰係数）。

$$\hat{y} = .079^* a - .116^* s + .104^* fp + .065^* fe + .017me + .371^* e + .019h \quad (10)$$

$$(R^2 = .231, N = 2876, ^* p < .01)$$

ここで、 h は付け加えられたハザード比である。結果を見ると大きな違いはない。変数の組み合わせによるが、このようにバイアスの影響がほとんど見られない場合もある。

分析結果に、サンプル選択がどのように影響を与えるのかは、分析する変数の組み合わせと、バイアスがどのように起こっているかによる。回収率が低くても、ランダムな欠損の場合は、元々問題にならないし、偏ったデータであっても分析する変数の組み合わせによっては、分析結果に影響を与えない場合もある。社会調査においては、回収率を上げるような努力は最大限するべきではあるが、欠損値があることが直ちに分析結果に影響を与えることではないということも理解するべきである。

【文献】

- Berk, Richard A. 1983. "An Introduction to Sample Selection Bias in Sociological Data." *American Sociological Review* 48: 386-398.
- Breen, Richard. 1996. *Regression Models: Censored, Sample Selected or Truncated Data*. SAGE Publications.
- Maddala, G. S. 1992. *Introduction to Econometrics, Second Edition*. Englewood Cliffs: Prentice-Hall. 和合肇訳. 1996. 『計量経済分析の方法（第2版）』. 東京: シーエーピー出版.
- 高田洋. 2008. 「投票行動と社会意識－発展した民主主義社会における投票態度－」. 土場学（編）『公共性と格差』（2005年SSM調査シリーズ7）. 2005年社会階層と社会移動調査研究会 所収.

Influential Analysis of Sample Selection: Missing data and Truncated or Censored Data

Hiroshi TAKADA

Sapporo Gakuin University

Social survey data generally has several types of bias or incompleteness in sample. In this article, sample selection bias in the 2005 SSM Japan survey is estimated by Heckman two stage method of Tobit model. Using the estimated hazard, what attribute of the sample contributes to high selection hazard is indirectly verified. We found that the samples with high possibility of dropping out are (1) instable employment, (2) lower educational status, (3) large enterprise white-collar worker (4) person who does not live in a detached house, and (5) younger person. The missing possibility of such sample is high. However, one could also argue that the higher missing probability does not always leads to unit non-response and that whether or not sample selection affects the result of the analysis depends on the combination of variables involved.

Keywords and phrases: Sample Selection, Missing Value, Tobit model

多重対応分析を用いた社会空間の構築

ー階層研究の新たな展開を求めてー

近藤博之

(大阪大学)

【要約】

本稿は、『ディスタンクシオン』におけるブルデューの分析枠組みを階層研究に応用し、1965-2005年の5時点のSSM調査データに多重対応分析（MCA）を適用することから日本の階層構造を吟味している。欧州の分析例と比較可能なように、諸々の社会経済的変数から客観的な社会空間を構築する形で分析を進めたが、その結果、日本では第1-2軸の分散が欧州の分析例よりも大きくなること、社会空間における各職業の布置はブルデューが描く関係と似ているものの、経済資本と文化資本が同じ方向に展開し主軸の意味が異なること、そうした社会空間の基本構造が時間的に安定であること、さらにそれが階層帰属意識や政党支持などの主観的変数の分布に有意な予測をもたらすこと、等が確認された。

キーワード： 社会空間，ブルデュー，多重対応分析

1. 階層研究における社会空間のアイデア

社会学における階層研究の基礎をつくったソローキンは、20世紀の早い時期に社会空間（social space）についてつぎのような記述を行っている。1）社会空間は人間社会の全体である。2）人の社会的位置は人間社会のすべての集団及び個別集団のなかの成員に対する関係の総体である。3）それらの関係を特定することで全体における人の位置づけが与えられる。4）そうした集団全体及び特定集団における位置の総体が社会的座標系を構成し、すべての人を社会的に位置づけることを可能にする（Sorokin, 1927, P. 6）。

現実の多様な関係から人の社会的位置を特定するという発想は階層研究の核心をなすものと言えるが、ソローキン自身において社会空間のこれ以上の展開はみられず、すべての関係は単純に垂直と水平の2次元に分類できるとして、もっぱら世間の基準に依拠した階層と移動の分析に向かっている。日本の階層研究でも社会空間そのものが考察の対象となることはほとんどなかったように思われる。たとえば安田はソローキンの社会空間のアイデアを肯定的に評価し、それを垂直次元に偏らせないために特性空間（property space）と呼ぶのが相応しいとしたものの、〈空間〉の用語は多くの人にとって馴染みにくいものであるから、〈社会空間における個人の位置〉というかわりに〈個人の社会的地位〉という語法を用いたいとして空

間の用語を放棄している（安田 1971, 42 頁）。

その後の階層研究は、世代間の職業移動や世代内の地位達成過程を因果的観点で吟味する膨大な数の実証分析を生み出したが、ソローキンが示唆したような幾何学的な視点で社会空間を描こうとする関心はほとんど表に現れることがなかった。それに近いものとしては、世代間の職業移動表や婚姻による職業結合表に正準相関分析を適用して職業間の距離を測定する試みがあるが¹、分析の主な目的は職業カテゴリーから量的成分を取り出すことに向けられており、必ずしも従来の階層研究の枠組みに対抗するものではなかった。クラスター分析による地位の一貫性・非一貫性に関する分析（今田・原 1979）も、空間的な発想を含んでいるものの目的は階層化の断面を明らかにすることに限られている。そうしたなかこれらの潮流とは別のところで社会空間のアイデアを前面に出して階級分析を行ったのがブルデューの『ディスタンクシオン』（1979=1989）であった。その試みはマルクスやウェーバーの古典的な階級理論に対するアンチテーゼを意味していたが（Weininger 2005）、既成の職業分類や職業的地位尺度を用いて因果分析一辺倒の分析を繰り返している英米の階層研究に対しても痛烈な批判を含んでいた。

ブルデューが行なったのは、まずさまざまな社会経済的指標を用いて客観的な社会空間を描くことであった。その社会空間と慣習行動からつくられた生活様式空間とが対応することをハビトゥスの概念を用いて説得的に論証したのが『ディスタンクシオン』の画期的な点であったが、ここでは地位達成研究（たとえばパス解析による分析）と同様の変数構成をとりながら従来の階層研究とは異なるアプローチが採用されている点に注目しよう。とくに注意したいのは、職業の並びや序列が所与ではなく、あくまでもデータ内在的に導かれている点である。既存の職業分類を用いて各々の機会状況を検討する通常の研究法では、分析結果によって階層体系上の職業の位置づけが変わることはないが（高等教育進学率が 30% から 60% に上昇しても出身階層としての職業 A は職業 A としての意味を担わされる）、ブルデューの採用したアプローチではそうではない。職業どうしの関係が似ていれば近くに、似ていなければ遠く離れた位置に、データに依存しつつ配置されるのである。伝統的な階級論の立場からはこの点がなかなか理解されないが²、ブルデューと同じ観点に立つなら、階級は経験世界のデータから便宜的に切り取られるものに過ぎず、データに先行して理論的に与えられるものでも、データを越えて実体化されるものでもないのである。

データに含まれる情報から社会空間を構築し、そこから社会の諸々の分化と対立の構図を柔軟に読み取るブルデューの方法は、初めに紹介したソローキンの社会空間のアイデアをもっとも押し進めたものといえるだろう。それは、階級の消滅や階層流動化が論点となる現代

1 代表的なものとして、Klatzky and Hodge (1971), Duncan=Jones (1972), Featherman, Jones and Hauser (1975), Domanski and Savinski (1986), Rytina (1992) などがある。詳しくは近藤 (2006) を参照のこと。

2 たとえば Oesch (2006) は、ブルデューの階級分析はマルクス主義やウェーバー派のものに比べて概念的に曖昧であり、階級地図の描き方も操作化が困難で、恣意的なものであるとして退けている。

の社会状況により相応しいアプローチを与えてくれるようにみえる。窮乏化理論や産業化理論のような大きな仮説が失われ、状況が見えにくくなればなるほど、焦点を絞った仮説＝検証型の分析よりも関連の全体的な把握を目指す探索型の分析手法が有用となる。実際、同じ社会空間を共有しながら系統的に分析を進めていく方が、分野ごとの断片的な検証結果をつなぎ合わせるよりも全体像はつかみやすくなる。

以下、社会空間のアイデアをもとに行なわれたブルデューの分析を社会研究のユニークな方法として特徴づけ、社会空間アプローチが階層研究において持ち得る意義と可能性について SSM 調査データの分析を通して示してみる。

2. 多重対応分析の利用

1) ブルデューの分析枠組み

『ディスタンクシオン』のなかでブルデューが用いたのは、対応分析の1つの発展型である多重対応分析（Multiple Correspondence Analysis, 以下 MCA）であった。まず、その背景について簡単に説明しておこう。

対応分析についての代表的な解説書のなかで、ルバールら（Lebart et al., 1977, 大隅・ルバール他 1994）は、特定の事象を諸々の人口学的変数と組み合わせて吟味する場合、それらの変数の相関を考慮した総合的な観点で取り組むのが効率的であるとして、人口学的変数による MCA の効用を説いている。実際、彼らは性別、学歴、持家、株式保有、不動産所有、年齢、都市規模の8変数に MCA を当てはめ、それより作成された2次元の平面に意見や行動の調査結果（各カテゴリーの重心座標）をプロットして、社会経済的地位との関連が一望できる図を提示している。

『ディスタンクシオン』もこのアイデアを採用し、そうした形の分析にさらに積極的な意味を与えている。ブルデューは、諸々の要因関連を社会空間のなかで同時に考慮し、共通の生活条件をもつ人々を客観的な階級として取り出すのは、個別の分析では必然的に解体されてしまう集合生活の体系性を再構成し、特定の慣習行動を集合体のメカニズムとして明らかにするためであると説明する。この目的のためにすでに確立されている有力な変数に頼るのは、統計的装いをもって現れる見かけ上の効果が真に重要な差異や変異を見え難くしてしまうので適当ではない。この観点から、ブルデューは、余暇（労働）時間、父親の学歴と職業、本人の学歴、所得総額、消費支出額、洗濯機と電話の所有、居住形態、居住地区の人口といった社会経済的諸指標を用いて、1960年代フランスの社会空間の模式図を描いたのであった。そして、職業を目印とした社会空間が「経済資本」と「文化資本」という2つの資本タイプの組合せで構成されること、慣習行動の特徴的な差異もそれらの資本機能の違いに結び付けて解釈可能なことを、対応分析のイメージを用いて示したのであった。さらに、趣味や文化

的行動についての MCA の分析から支配階級や中間階級の生活様式空間をつくり、そこに学歴や職業の社会経済的指標を投影して〈意識〉と〈存在〉の対応を確認するという逆方向の検討も行っている。

ブルデューのこうした分析は、階層研究のなかに文化をキーワードとする新しい研究関心を生み出したが、そこでの方法を社会研究の系統的な方法として継承・発展させようとする動きは、対応分析に対する理解が社会学において一般に低いこともあり、すぐには登場しなかった。ブルデューが用いた方法を社会の構造や変化を分析するための用具として一般化し、MCA による比較社会的分析を進めようとする動きが現れたのは、ようやく 1990 年代の後半になってからのことである。その方向の代表的な研究例として、ローゼンルンド (Rosenlund 1996, 2000, 2001) のノルウェー社会の分析を挙げることができる。彼は、『ディスタンクシオン』の方法を丁寧にとり、そこに自らのアイデアを付け加えながら 1970 年代以降のノルウェー社会の変化について興味深い分析を行っている。最近になり、そうした試みはブルデューの関心を共有するフランスの統計学者たちの後押し (Le Roux and Rouanet, 2004) を受けて、さらに洗練されたものになっている (Prieur and Rosenlund 2005, Hjellbrekke et al., 2007)。ブルデューのモデルを、ブルデュー自身が期待していたような形で比較社会的に探求していく道がようやく拓けてきたといえる (ブルデュー 1990, 81 頁)。

『ディスタンクシオン』の分析を範例とした階層研究の社会空間アプローチは、具体的につぎのような手順で進められている。1) 社会経済的変数を活性変数³とした社会空間の構築、2) MCAの基本情報を用いた社会空間の構造的特徴の把握、3) 意識及び行動変数の社会空間への補充的投影、1') 意識及び行動変数を活性変数とした生活様式空間の構築、2') MCAの基本情報から生活諸領域における類型や対立関係の抽出、3') 生活様式空間に対する客観的変数の補充的投影。1～3) の社会経済的変数を中心とする分析も、1'～3') の意識及び行動変数を中心とする分析も、それ自体がまとまった課題を構成するが、それぞれの所を換えて空間の構築と補充変数による解釈を相互的に行うことは〈存在〉と〈意識〉の対応を吟味するきわめて有効な方法となる。それはルバルらにおいて相互的アプローチとして推奨されたものであり、ブルデューやローゼンルンドにおいて実際に試みられたものでもある。

2) 交差配列の仮説

ブルデューは、社会空間を構成する基本的次元として、周知のように「資本総量」、「資本タイプの構成」、「軌道」の3つをあげている。教育や職業や所得の社会経済的変数が互いに関連することは言うまでもないが、それらの変数にMCAを適用すると、全体の平均的な相関

3 MCA では、空間構築に直接関わる変数を活性変数(active variables)といい、空間構築には直接関わらないものの空間の特徴を把握するために個人の主座標を利用して事後的に投影される変数を補充変数または例示変数(supplementary or illustrative variables)という。

をもっとも大きくするスコア体系が自動的に導かれることが知られている⁴。したがって、第1軸には各変数のカテゴリーが予想された通りの順序で並び、文字通り「資本総量」の大小を表わす軸となる。この明白な関連が除去された後に、第2軸以下でやはり最大の相関をもたらず変数間の関連が探索されるのだが、そのパターンに注目して軸の意味を特徴づけたのが「資本タイプの構成」である。ブルデューの描く社会空間では、たしかに経済資本（所得や資産で捉えられる）が増加すると文化資本（教育が代表的変数となる）が減少し、反対に文化資本が増大すると経済資本が減少するような並びで、すべての職業が空間内に配置されているのが分かる。この座標平面の特徴を簡潔に表現したのが、「交差配列構造」（石井1993）の仮説である。

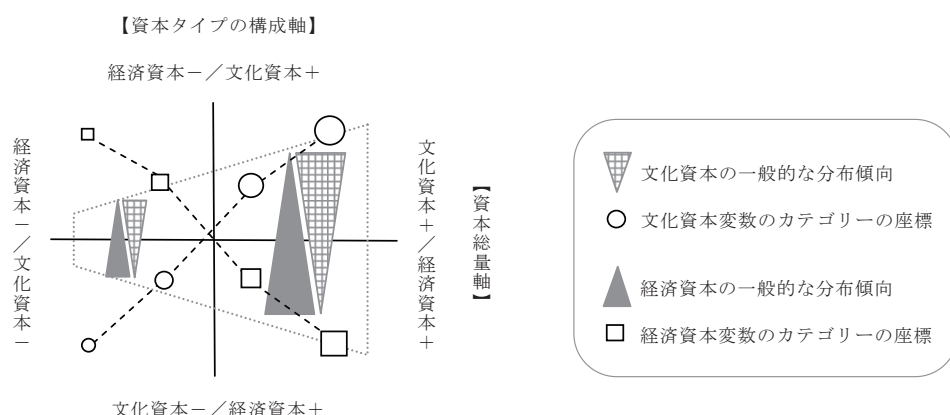


図1 社会空間の交差配列構造

上の図は、2つの資本タイプが交差する様子を、その背景を含めて図示したものである。この仮説のポイントは、社会空間が一次元のハイパーキカルな構造（横軸の資本総量）によってのみ規定されるわけではなく、依拠する資源タイプの特徴（縦軸の違い）によっても有意味な区別が成り立ち、それぞれの水準で生活条件や行動様式に系統的な差異がつくりだされるとするところにある。ここでの対比が示唆しているのは、言うまでもなく所有する経済資本を頼りに世代的な地位継承を図ろうとする集団と、学校教育や資格制度を次世代の地位形成に利用しようとする集団の対立である。

ノルウェーとデンマークのデータからブルデューと同様の社会空間を構築したローゼンランドも、これと同様の交差配列のパターンが抽出できることを経済資本（世帯所得や所有する車の価値）と文化資本（父親の学歴）のカテゴリーの布置から確認している。ただし、ノルウェー社会について1970年、1980年、1990年の3時点で検討してみると、1970年のデー

4 たとえば Greenacre (2006) が、そのことを MCA の正準相関による解釈から示している。

タでは経済資本と文化資本の間に明確な差異が認められず、1980年、1990年と時代が進むにつれて両者の距離が広がってくることから、交差配列のパターンは資本総量の実質的な増加から新しい配分原理が定着することによって出現した最近の社会構造の特徴であると、たいへん興味深い指摘を行っている（Rosenlund 2000）。

さらに、社会空間の3番目の次元としてブルデューがあげるのが「軌道」である。それは本人世代と親世代の社会経済的地位の変化に言及したもので、平面で捉えられた社会空間に奥行きの違いをもたらすものとなる。ブルデューによると、経済資本と文化資本の同系列内の移動は相対的に容易であるのに対し、異なる系列間の移動は通常は困難であるという。したがって、階級間の障壁を越えようとする意欲が強い場合や結果的に障壁を越えて移動を成し遂げた場合は、周囲とのハビトゥスの違いがいつそう際立ってくる。社会空間のなかにそうした傾向が集散的なものとして存在し、生活様式空間にも際立った陰翳を与えるというのがブルデューの観察である。

MCAによる社会空間分析の焦点は、こうしたブルデューの分析を踏まえて、とりあえず第1軸における全体の相関と、第2軸以下の変数間関連のパターンに置かれることになる。

3) 社会空間における個人の分布

ところで、『ディスタクシオン』のなかに社会科学におけるMCAの典型的な活用例を見出しているルアネラ（Rouanet, Ackermann and Le Roux 2000, Rouanet 2006）は、ブルデューが変数の関連ではなく、あくまでも社会空間のなかの個人の態様に興味を寄せていた点を強調している。「個人×変数」を基本データ・セットとするMCAでは、個人のつくりだす分布からそうした検討がなされ得る。それを可能にするのが、座標空間において個人と変数（の各カテゴリー）を相互に関連づけるつぎの推移方程式である。

$$y_{\alpha}^i = (1/\sqrt{\lambda_{\alpha}}) \left(\sum_{k \in K_i} y_{\alpha}^k / Q \right) \quad (1)$$

$$y_{\alpha}^k = (1/\sqrt{\lambda_{\alpha}}) \left(\sum_{i \in I_k} y_{\alpha}^i / n_k \right) \quad \cdots \quad \left\{ y_{\alpha}^k = (1/\sqrt{\lambda_{\alpha}}) \hat{y}_{\alpha}^k \right\} \quad (2)$$

λ_{α} は、「個人×変数」の基本データ・セットに特異値分解を当てはめたときの α 番目の固有値、 y_{α}^i は対応する固有ベクトルから定義される個人 i の主座標、 y_{α}^k は同じくカテゴリー k の主座標、 $\hat{y}_{\alpha}^k$ は個人分布の主座標から求めたカテゴリー k の平均点、 Q は変数の数、 n_k はカテゴリー k を選択した個人の数である。集合 $\{K_i\}$ は個人 i が選択したカテゴリーの全体、同じく $\{I_k\}$ はカテゴリー k を選択した個人の全体である。

式から読み取れるように、行（個人）空間における個人の座標 y_{α}^i は各人が選択したカテゴリーの主座標 y_{α}^k がもとになっており、反対に列（カテゴリー）空間における各項目の座標 y_{α}^k は k を選択した個人の主座標 y_{α}^i がもとになっている。よって、同じ変数におけるカテゴリー

の位置の相違には、他の全ての変数における個々人の選択結果の違いが集約されているとみなすことができる。つまり、座標空間における学歴変数のカテゴリーAとカテゴリーBの位置の相違は、学歴Aを有する人の職業や所得や財産の状況と学歴Bを有する人の職業や所得や財産の状況の違いを反映したものとなる。このように、各々の社会的特徴で区分された下位集団の平均的な配置は、他の諸々の特徴の総体として決まってくるのである。これが、階級や階層を何らかの固定的な構造としてではなく、現実の諸関係からそのつど実態的に把握していこうとする社会空間アプローチの最大の特徴となる。

実際に、MCAを用いた社会空間アプローチでは、座標 y_{α}^i による個人の分布が社会分析のために大いに活用される。ローゼンルンドも、座標平面への個人の投影図を、広場に集まった人々が自分に似た人の近くに座った状態にたとえて説明している。たしかに個人の位置は各人に該当するすべての変数の特徴（カテゴリー点 y_{α}^k ）の平均として与えられるので、任意の2人を比べたときにある1つの特徴が大きく異なると、地位変数どうしの関連により他の変数の特徴も違っている可能性が高くなり、両者の位置は結果的に遠く離れたものとなる。逆に、近くに位置する人たちは多くの点で似ている人たちであるとみなすことができる。それも、社会経済的変数どうしの相関が高いと判断できるからだが、その相関に重要な変化が生じれば、その動きは個人の位置の違いに新たな差異をもたらし、推移方程式に表された関係を通してカテゴリーの布置に影響していくことになる。

4) MCAによる趨勢の検討

本論では時点間の変化に焦点を当てたいので、それについても簡単に触れておく。

まず、変数の構成やカテゴリーの区分はすべての時点で可能な限り揃えておく必要がある。その上で、各時点のデータを個別に分析することが出発点となる。個別データのMCAによる分析結果を比較して、まずは全体の様子が似ているか似ていないかを見極める必要がある。とりあえず少数の主軸で全体の分散の何%が説明されているか、その分散に対する変数やカテゴリーの寄与はどうか、各カテゴリーの空間的な布置に目立った変化がないかなどが比較すべきポイントとなる。

しかし、カテゴリーどうしの位置関係を比較しても、カテゴリーに関する距離の定義（周辺度数が関係する）や主軸の抽出はデータによって異なるので、複数の結果を見比べても同じ基準で比較しているという「根拠」は弱い。ちなみに、MCAでは、個人 i によるカテゴリー k の選択状況を $x_{ik}=1$ または $x_{ik}=0$ として、任意の個人 i と i' の多次元空間における距離（カイ2乗距離）が次式で与えられる。 n はサンプル数、 Q は変数の数である。

$$d^2(i, i') = \frac{n}{Q} \sum_k \frac{1}{n_k} (x_{ik} - x_{i'k})^2 \quad (3)$$

i と i' でカテゴリーの選択内容が異なるたびに両者の間に距離が発生し、その距離は違いのあるカテゴリーの周辺度数の大きさ (n_k/n) に影響を受ける。同じくカテゴリー k と k' の距離はつぎの式(4)ように表わされる。2つのカテゴリーが同じ変数に属していて、個人 i による同時選択がない場合は第3項が0となる⁵。

$$d^2(k, k') = \frac{n}{n_k} + \frac{n}{n_{k'}} - \frac{2}{n_k n_{k'}} \sum_i x_{ik} x_{ik'} \quad (4)$$

このように、分析のもととなる距離の定義に周辺度数が直接関わってくるので、異なるデータを比較するときはそれをどう扱うかが問題となる。実際、座標平面上の各カテゴリーの位置関係が2時点で異なっても、その違いは単純に周辺度数の違いからきているのかもしれないし、あるいはまたカテゴリー点を投影する僅かな角度の違いが座標平面に大きな距離を作りだしているのかもしれない。つまり比較の「根拠」を担保するには、空間構築に関係する距離の定義と主軸の方向を何らかの仕方で揃えておく必要がある。同じ素材で分析の対象を作り、同じ角度を切り口にするというイメージである。そのときに浮かび上がるカテゴリーの布置に明確な違いがあれば、この間に社会空間の構造に本質的な変化があったとみなしてよいだろう。

このことを実際に行うには2つの方向がある。1つは、特定のデータで空間構築を行い、そこでの情報を利用して他のデータを同じ空間のなかに補充的に配置することである。その場合は、先行変数がつくる主軸と各カテゴリーの座標を共有することになるので、補充変数の追加によって空間構造そのものが影響を受けることはない。立場を換えて互いが作り出す空間に他方のデータを補充的に配置したときに同じ傾向が観察されるようであれば、たしかにそれがこの間に起こった変化だとみなすことができる。もう1つは、合併データで全体の傾向を押さえておき、その上で内部の違いを吟味するやり方である。この場合は、合併データという平均像が空間構築の材料となり、各時点のデータはすべて補充的に扱われることになる。ここでも、距離に関する定義と主軸が共有されており、比較の根拠が与えられる。個々の特徴が平均的に薄められる欠点はあるが、個別データの特徴が似ている場合は効率的な方法となる。以下では、この合併データからアプローチし、共通性を確保した上で内部的な差異に注目してみる⁶。

3. MCA 用のデータと変数

⁵ したがって、同一変数のカテゴリー間の距離は個人の選択状況に関係なく周辺度数の大きさだけで与えられることになるが、その点がMCAの難点として指摘されている。

⁶ この他に、ラ・ルーとルアネ (Le Roux and Rouanet 2004) は距離の定義を共有する空間で下位集団ごとに異なる主軸を導いて比較する方法を工夫している。

ここでの分析は、1965 年～2005 年SSM調査データを用いる⁷。それより、各々のデータに共通の変数として、本人の現職、従業先の規模（民間と官公庁の別を含む）、父親の主な職業、父親の最終学歴、本人の最終学歴、本人の年間所得、所有する財項目、年齢の 8 つを取りだし、社会空間を構築するための材料とした。居住地域の特徴を含んでいないなど若干の違いはあるが、学歴、職業、所得の基本変数を確保している点はブルデューやローゼンルンドの設定と同じである。

8 つの変数について具体的に説明しよう。まず、本人と父親の職業は国勢調査の社会経済分類を用いている。ただし、「教員・宗教家」のカテゴリーを教員のみとし、宗教家を「その他の専門職」に含めるなどの変更を加えている。参考までに国勢調査の分布と SSM 調査の分布を対比したものを図 2 に示しておく。

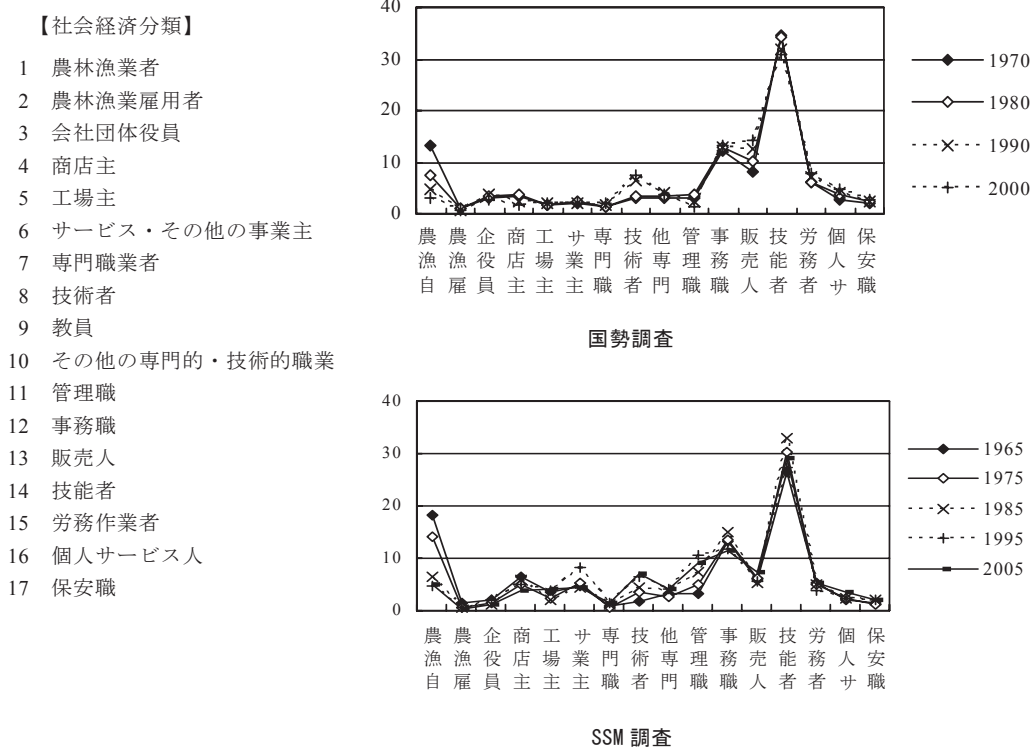


図 2 対象者の職業の比較

全体によく似た分布となっているが、SSM 調査データでは業主と管理職の比率がやや高く、販売人、技能者、労務作業者の比率がやや低いことが分かる。他方、父職については煩雑さを避けるために、専門（7, 9, 10）、管理（3, 11）、雇用ホワイト（8, 12, 13, 17）、自営業主（4,

⁷ 1955 年 SSM 調査を除外したのは、農業従事者に関する所得データが得られていないことによる。

5, 6), 雇用ブルー (14, 15, 16), 農業 (1, 2) の 6 分類とした。技術者を専門ではなく雇用ホワイトとしたのは、もとの 17 分類を用いた社会空間の分析からその位置づけがより適当であると判断したからである。従業先の従業員規模は、1-4 人、5-29 人、30-299 人、300-999 人、1000 人以上、官公庁の 6 分類である。

本人と父親の最終学歴は、旧制と新制の学校段階を区別したSSM教育分類を用いている。旧制は、尋常小学校、高等小学校、中学校・実業学校・師範学校、高等学校・専門学校・高等師範学校、大学の 5 分類、新制は中学、高校、短大・高専、大学の 4 分類である。ただし、父親の学歴に新制短大・高専が現れることはほとんどないので、実際の分析では学歴不明とともにミッシングとして扱っている⁸。また、1965 年データで父親学歴に新制の学校はなく、2005 年データで本人学歴に旧制の学校がないなど、調査年によってもデータ構成に若干の違いがある。そうした違いはあるものの、客観的な社会空間を構築する上で教育カテゴリーがきわめて重要な役割を果たすので、あえてカテゴリーを合併することはせず、それぞれの時代性を反映させることにした。

他方、経済変数を代表する本人の所得と所有する財項目は、各調査年の分布がだいたい同じとなるように 5 区分したものを用意した。所得の場合は調査票にプリコードされた所得カテゴリーを用いて全体を 5 区分し、所有する財項目の場合は各調査年データに含まれている項目から総合的なスコアを導いてから全体を 5 区分している。具体的には、それぞれの財項目から耕運機など所有者の職業が限られるものと普及率の極端に高いもの (90%以上) を除き、残った項目に対応分析⁹を適用した結果から、第 1 軸への相対的寄与率が高いものを各々 10 項目選び、それらの項目の合計点により全体を 5 区分した。

最後に、年齢は 20~70 才を 10 才刻みとする 5 カテゴリーとした。社会空間を構築する変数に年齢を加えたのは、年齢自体も 1 つの地位変数であるということのほかに、2 番目の主成分にも実質的な意味をもたせて曖昧な要素を積極的に取り除いておくという理由からである。実際、年齢を加えた場合は、1-2 軸平面でのカテゴリーの布置に大きな違いは生じないものの、第 2 主成分の分散が大きくなることを各調査年度のデータから確認している。

上記 8 変数のカテゴリー数の総計は 66 となるが、一部の変数の無回答・不明を非活性カテゴリーとして距離の計算に含めない特殊な形式の MCA を用いたので、実際に社会空間の構築に関わったカテゴリー数は合併データで 61 個となった。個別年のデータでは 56~61 個の範囲である。

⁸ ここでの分析はラ・ルーとルアネの Specific MCA (Le Roux and Rouanet 2004, 203-213) を用いて行なっているが、そこではサンプルそのものをミッシングとするのではなく、非活性 (passive) カテゴリーとして距離の計算に含めない扱いがとられる。式 (3) で一方が活性 (active) で、他方が非活性であれば、距離計算には前者の要素のみが含まれる。

⁹ 職業を共通のグローバル変数とし、5 時点の財項目の所有状況をクロス表で積み上げた特殊な形式をもつ対応分析による。Pages and Becue-Bertaut (2006)を参照のこと。

4. 社会空間の基本構造

1) イナーシャの分解

表 1 が、個別データと合併データの各々に MCA を適用した結果をまとめたものである。全イナーシャは行と列が作り出すデータの全分散を意味し、行点または列点を一次元の直線に投影してできる分散を大きな順に取り出したのが固有値である¹⁰。

表に示されるように、MCA では全イナーシャに対する固有値の割合がきわめて小さなものになる。ただし、それは MCA の構造的な特徴からきており、各々の軸の寄与が絶対的に小さいということではない。その点の過小評価を是正し、寄与率についてより現実的な基準を与えるものとして、MCA ではしばしばベンゼクリの修正比率が計算される。下段に示した数値がそれであるが、いずれの年度も第 1 軸の寄与が全体の半分程度、1 軸と 2 軸の合計ではだいたい 3/4 程度の累積比率となることが分かる。どの調査年のデータも、2 次元平面の描写で社会空間の基本的な特徴が十分に捉えられているようにみえる。ちなみに、ここでの変数構成と比較的よく似た条件で社会空間分析を行っているノルウェーのローゼンルンドの例では、いずれの分析でもベンゼクリの修正比率がこれよりも低い値となっている。今後の検討課題に属するが、社会経済的変数の相関が日本では相対的に高いと言えるかもしれない¹¹。

表 1 主軸の分散(固有値)と Benzecri の修正及び累積比率

	1965	1975	1985	1995	2005	1965-05
固有値						
1 軸	0.3734	0.3798	0.3766	0.3609	0.3499	0.3635
2 軸	0.3087	0.3113	0.2998	0.3187	0.2928	0.2881
3 軸	0.2311	0.2086	0.2226	0.2450	0.2130	0.2082
全イナーシャ	6.2816	6.2787	6.6673	6.6674	6.0614	6.6659
カテゴリー数	58	58	61	61	56	61
Benzecri の修正比率						
1 軸	48.1	52.7	50.0	42.4	49.7	54.5
累積比率						
1-2 軸	74.3	80.8	74.1	71.1	77.4	79.9
1-3 軸	83.1	86.5	81.7	82.0	85.0	86.5
1-5 軸	92.3	92.6	91.2	91.4	92.0	94.5
サンプル数	1,952	2,521	2,214	2,180	2,196	11,063

つぎに、8 つの変数のうちどの変数が主軸形成により深く関わっているかを各軸の分散に対する寄与率から検討してみよう。MCA では、行点または列点の主座標 (y_{α}^k 及び y_{α}^i) の重みつき 2 乗和が対応する主軸の分散に等しくなり、つぎの関係が成り立つ。

$$\lambda_{\alpha} = \sum_k (n_k/nQ)(y_{\alpha}^k)^2 = \sum_i (1/n)(y_{\alpha}^i)^2 \quad (5)$$

¹⁰ 全イナーシャは「個人×変数」のデータからピアソンの χ^2 値を求めて総数 (Qn) で割ったものに等しく、MCA の場合、それは変数の数 Q と全体のカテゴリー数 K から一律に定まる。しかし、ここでは一部のカテゴリーを個人間の距離の計算に含めない特殊 MCA を採用しているのでデータごとに少しずつ異なった値となる。

¹¹ 世代間職業移動表の正準相関分析でも同じ傾向が確認されているので (近藤 2006)、筆者は日本の階層構造はかなり単純な構造をもつのではないかと推測している。

第1軸と第2軸について、それぞれの固有値に対する寄与を変数ごとにまとめたのが表2-1、表2-2である。ただし、MCAの場合、各変数がつくりだす分散はカテゴリー数によって自動的に決まり、変数 q のカテゴリー数を k_q として、 $(k_q-1)/Q$ で表される。したがって、もともとカテゴリー数が大きい現職はどの軸に対しても高い寄与をもたらすものとして立ち現れることに注意する必要がある。この変数の寄与が小さければ注目に値するが、ここでのように寄与が大きいときは当然と受け取っておくのがよい¹²。

表 2-1 第 1 軸に対する変数の寄与

	1965	1975	1985	1995	2005	1965-05
現職	0.2097	0.2022	0.1956	0.1943	0.2099	0.2113
学歴	0.2044	0.2044	0.2126	0.2281	0.1951	0.2127
父職	0.1418	0.1324	0.1404	0.1434	0.1420	0.1518
父学歴	0.1268	0.1342	0.1209	0.1279	0.1373	0.1381
規模	0.1300	0.1172	0.1015	0.1089	0.1275	0.1121
所得	0.0682	0.0804	0.0787	0.0702	0.0689	0.0800
財所有	0.0776	0.0928	0.0958	0.0364	0.0259	0.0685
年齢	0.0415	0.0365	0.0545	0.0909	0.0933	0.0254
全体	1.0000	1.0001	1.0000	1.0001	0.9999	0.9999

表 2-2 第 2 軸に対する変数の寄与

	1965	1975	1985	1995	2005	1965-05
学歴	0.2893	0.2546	0.2510	0.2667	0.2491	0.2460
（本人学歴）	(0.2593)	(0.2444)	(0.2021)	(0.0755)	(0.0401)	(0.1529)
（父学歴）	(0.0300)	(0.0102)	(0.0488)	(0.1913)	(0.2090)	(0.0932)
年齢	0.2376	0.2364	0.2176	0.214	0.2192	0.1818
現職	0.2077	0.2035	0.2053	0.1941	0.2060	0.2318
所得	0.1335	0.1245	0.1145	0.1331	0.1160	0.1176
財所有	0.0451	0.0573	0.0740	0.0841	0.0853	0.0660
父職	0.0407	0.0521	0.0709	0.0633	0.0682	0.0733
規模	0.0461	0.0715	0.0668	0.0448	0.0561	0.0834
全体	1.0000	0.9999	1.0000	1.0002	0.9999	1.0000

第1軸での寄与が大きいのは、明らかに本人及び父親の学歴と職業である。それらを合計すると、カテゴリーがつくる第1軸の分散の6割前後を占めることが分かる。したがって、第1軸では、とりわけ教育達成の違いとそれを通じた職業達成の違いが集約的に捉えられているとみなすことができる。他方、第2軸で寄与が大きいのは年齢であり、学歴の寄与も旧制と新制の対比が原因であると考えられる。ただし、古い世代の父学歴に新制の学校が含まれないこと、新しい世代の本人学歴に旧制の学校が含まれないことから、学歴の寄与率は調査年ごとに大きく変動するので、ここでは便宜的に両者の寄与率を合計して表示している。

¹² こうした分析でそれぞれの職業が空間の全域に位置づくことはブルデューの分析にも見られた通りであり、かえってその布置から空間の特徴を読むことが可能となる。

また、第2軸では所得と財所有の寄与も相対的に大きくなっている。こうしたことから、第2軸は加齢を中心に所得や資産の差異を反映していると解釈することができるだろう。

なお、ここに示した各変数の寄与率は、個人の分布においてそれぞれの変数カテゴリーから級間及び級内分散を求めて分散比 (η^2) を計算したときの値に等しいことを付け加えておく。たとえば、1965 年データの現職では第1軸の分散に対する相対的寄与が 0.2097 となっている。これは、表1に示した固有値に対する割合なので、寄与の絶対値は $0.3734 \times 0.2097 = 0.0783$ となる。MCA における各変数のカテゴリーのウェイトはもとの比率に $1/Q$ をかけてつくられているので、この値を8倍してもとに戻してやると、第1軸の分散に関する現職の絶対的寄与が $0.0783 \times 8 = 0.6264$ となる。この値は、個人空間の全体の布置から17個の現職カテゴリーを区別し、各々の平均値と分散から分散比 η^2 を計算したときの値に等しい。つまり、現職単独でこの空間の個人分布を説明した場合、全体の分散の63%程度がそれによって説明されるという意味になる。同じ観点で2005年データについて現職の η^2 を計算すると、その値は0.5877となる。社会空間における職業の説明力はこの間に6%ほど低下したということになる。

これらの変数の寄与をグラフで比較したのが図3である。寄与の平均は0.125 (=1/8) であり、この水準を上回る変数が主軸形成に相対的に大きな寄与をなしているとみなすことができる。調査年により若干の変動はあるが、40年の時間幅のなかでみれば社会空間の基本構造はきわめて安定していると言えるのではないだろうか。以下では、それを「根拠」として合併データを用いた分析を行うことにするが、確認のために個別データのMCAと合併データのMCAから個人の主成分得点を取り出し、各々の相関係数を求めてみると表3の結果が得られた。どの個別データをみても合併データとの相関は非常に高く、この間の階層構造がきわめて安定的であることを示している。

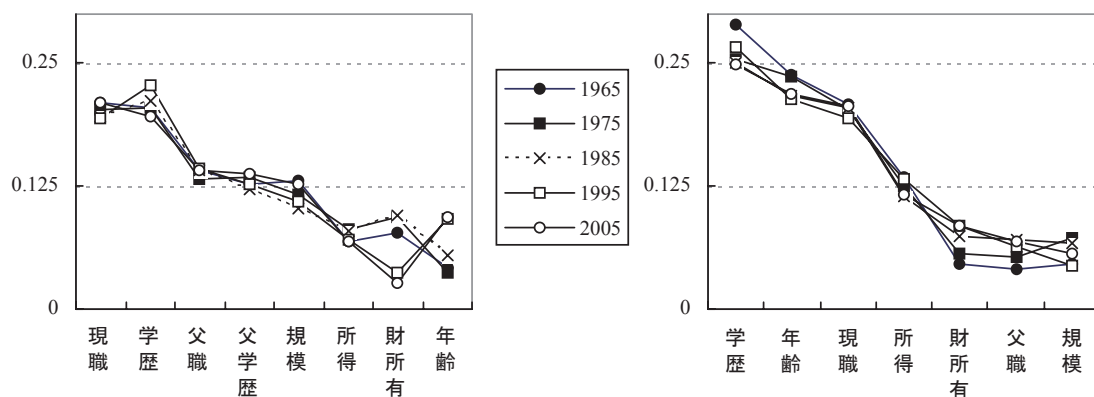


図3 各調査年データにおける変数の寄与の比較、第1軸(左)と第2軸(右)

表 3 合併データと個別データにおける
個人の主座標得点の相関係数

個別データ	合併データ	
	第 1 軸	第 2 軸
1965	0.989	0.962
1975	0.996	0.981
1985	0.994	0.981
1995	0.977	0.986
2005	0.978	0.982

2) カテゴリーの寄与からみた空間の特徴

今度は、各変数のどのカテゴリーが主軸形成に寄与しているかを検討してみよう。その観点から合併データにおける特殊MCAの分析結果をまとめたのが表 4-1、表 4-2 である¹³。分析に用いたカテゴリー数は 61 なので、ここでは原点を挟んだ対称の位置関係で、相対的寄与が 1/61 を上回るカテゴリーを拾いだしている。合計欄にあるように、ここに挙げてあるカテゴリーの寄与で第 1 軸の分散の 83%、第 2 軸の分散の 78%を占めている。第 1 軸における職業の対比は農林漁業と事務職及び管理職によるもので、教育資格に基づいた職業達成の違いに対応している。他方、第 2 軸のそれは自営業と技能者の対比が中心であり、生産手段の所有・非所有に関係しているのが明らかである。同様に、第 2 軸の学歴の寄与は旧制と新制の

表 4-1 第 1 軸に対する寄与の大きなカテゴリー（1965-2005 合併データによる）

第 1 軸	変数の寄与	寄与の大きいカテゴリー		相対的寄与	
		(-)	(+)	(-)	(+)
学歴	0.2127	旧尋小 旧高小 新中学	新大学	0.0196 0.0463 0.0275	0.1060
現職	0.2113	農漁自	管理職 事務職	0.0700	0.0291 0.0254
父職	0.1518	父農業	父専門 父管理 父雇W	0.0575	0.0372 0.0365 0.0174
父学歴	0.1381	旧尋常	旧中学 旧高専 旧大学 新大学	0.0405	0.0219 0.0214 0.0289 0.0184
規模	0.1121	1～4 人	1000 人以上 官公庁	0.0590	0.0171 0.0232
所得	0.0800	所得 1	所得 4 所得 5	0.0337	0.0222 0.0198
財所有	0.0685	財所有 1	財所有 5	0.0205	0.0261
年齢	0.0254				
合計	0.9999			0.3746	0.4506

¹³ MCA の結果のこうした取り扱い扱いは、Le Roux and Rouanet（1998, 2004）を参考にしている。

表 4-2 第 2 軸に対する寄与の大きなカテゴリー (1965-2005 合併データによる)

第 2 軸	変数の寄与	寄与の大きいカテゴリー		相対的寄与	
		(-)	(+)	(-)	(+)
現職	0.2318	農漁自 会社役員 商店主 他業主	技能者	0.0455 0.0224 0.0225 0.0220	0.0536
年齢	0.1818	50-59 才 60-69 才	20-29 才 30-39 才	0.0296 0.0397	0.0875 0.0184
学歴	0.1529	旧高小 旧中学 旧高専 旧大学	新高校	0.0267 0.0253 0.0175 0.0190	0.0430
所得	0.1176	所得 5	所得 2	0.0763	0.0230
規模	0.0834	1~4 人		0.0553	
財所有	0.0660	財所有 5		0.0452	
父職	0.0733		父雇 B		0.0423
父学歴	0.0932		中卒 高卒		0.0288 0.0377
合計	1.0000			0.4470	0.3343

対比によるもので、年齢要素が中心であることが確認される。なお、MCA の座標空間における左右、上下の方向は任意であり、それらの方向に絶対的な意味はない。あくまでもカテゴリーの相対的な位置関係から内容上の解釈が与えられることに注意して欲しい。

つづけて、ブルデューの交差配列仮説を検討するために、これらの寄与の大きなカテゴリーに目印をつけて 1-2 軸の座標平面に配置してみよう。図 4-1、図 4-2 がそれである。すでに示したように、第 1 軸に沿ったカテゴリーの対立は制度化された学歴資本が中心であること、また第 2 軸に沿ったカテゴリーの対立は年齢や世代の違いが中心であり、それに有資産の自営業と雇用者の対比、及び所得の高低が重なっていることなどが改めて確認される。

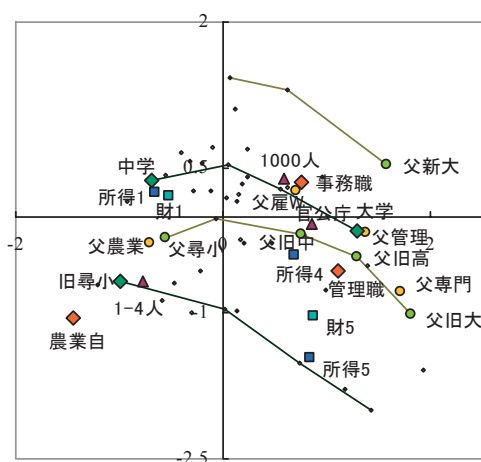


図 4-1 第 1 軸の中心的対立関係

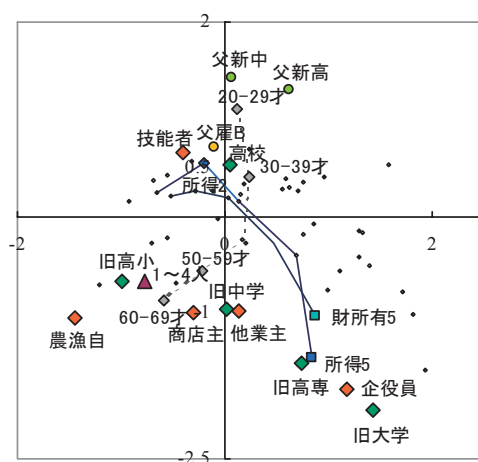


図 4-2 第 2 軸の中心的対立関係

ブルデューは、これと同様の社会空間図で第1軸を「資本総量」の軸と特徴づけたわけだが、ここでも横軸に沿った序列関係はすべての変数で一致している。しかし、第2軸が「資本タイプの構成」を反映した軸であるとはみなしがたい。なぜなら、経済資本が第2象限（左上）から第4象限（右下）に向けて直線的に増大して行くとするなら、文化資本は第3象限（左下）から第1象限（右上）の方向にやはり直線的に増大し、上下対称の関係でなければならないのに、そうはなっていないからである。実際、文化資本の代理指標である学歴資本の系列と経済資本の代理指標である所得や財の系列は2つの図に示すように軌跡がほとんど重なってしまうのである。

ローゼンルンドは、われわれの場合と同様に教育を文化資本の指標に用いて、ノルウェーとデンマークのデータからブルデューが示した交差配列のパターンが導けることを確認している。しかし、日本のデータでは必ずしも同じパターンが得られない。5時点のデータを個々に検討しても結果は同様なので、これが数千人規模の全国サンプル調査から社会空間を構築したときの日本社会の実態ということになる。違いの原因は、1) 伝統的な学歴社会論が示唆するように日本では文化資本と経済資本の重なりが大きいからなのか、2) 教育が文化資本の代理指標として不完全だからなのか、同じく 3) 経済資本が資産所有の面から十分に捉えられていないからなのか、それとも 4) ヨーロッパの先行研究に何らかの問題があるのか、のいずれかであろう。また、先に紹介したようにローゼンルンドは、かつてはここでのパターンと同じく高学歴と高所得が重なり文化資本と経済資本が対立軸を形成していなかったが、その後の資本総量の増加が資源配分原理の変化を伴い、2つの資本系列が対比的なパターンを示すようになってきたと分析している。われわれの場合は、今のところそのような変化も確認できない。ここでの変数構成にMCAを適用する限り、社会空間の基本構造は5時点ではほとんど同じなのである。社会空間と象徴空間の再生産を保証するメカニズムは社会によってかなり異なっているようである¹⁴。

5. 個人の分布による変化の検討

1) 社会空間のミクロな描写

「個人×変数」を基本データ・セットとするMCAの利点は、式(1)に示されたように変数カテゴリーがつくる同じ空間において個々のサンプルの位置を視覚的に表示できるところにある。今度は、個人の分布をもとに社会空間の内容的な検討を進めてみよう。まず、図5に分布の全体像を示しておく。

この図は、8変数61カテゴリーで作られた多次元空間における個人の分布を第1-2軸の座

¹⁴ 1995年SSM調査データから生活様式空間の分析を行った片岡（2003）によると、日本でも経済資本と文化資本の交差配列のパターンが観察できるという。こうした齟齬を詳しく検討することによって日本社会の特徴を見極めることができるだろう。

標平面に投影したものである。投影された個人点は全サンプル（11,063）で、まず 1975-1995 年データを小さな点として敷き詰め、その上に 1965 年データの黒点、さらにその上に 2005 年の白点を乗せている。図を概観して分かるのは、この 40 年間の変化がかなり小さなもので、全体の形状を概ね保持したまま重心が少しだけ右上に移動してきたということである。実際、下敷きにした中間年のデータも、1965 年と 2005 年の 2 つのデータの範囲内にほぼ収まっているのが分かる。表 1 の結果に戻るなら、この図的表示が多次元空間における個人分布の分散の 8 割程度をとらえているとみなすことができる。

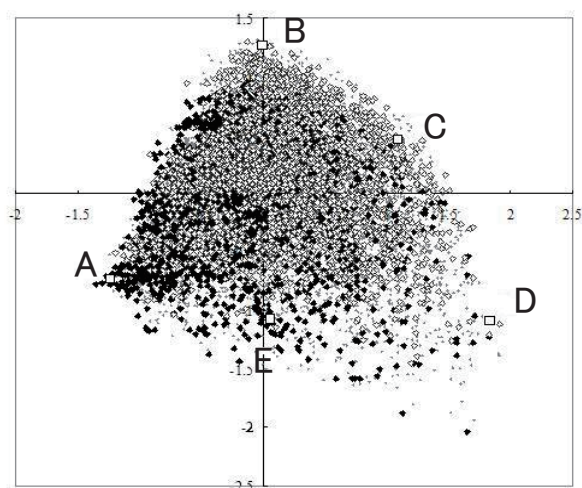


図 5 第 1-2 平面への個人の投影図

表 5 8 変数からなる 5 つの典型

	A	B	C	D	E
本人現職	農林漁業者	技能者	事務職	専門職業者	他の事業主
本人従業先	1-4 人	30-299 人	官公庁	1000 人以上	1-4 人
父職	父農業	父雇 B	父管理	父専門	父自営
父学歴	旧制高小	新制高校	新制高校	新制大学	旧制高小
本人学歴	中学卒	高校卒	大学卒	大学卒	中学卒
所得 5 区分	I	II	III	IV	V
財 5 区分	I	II	III	IV	V
年齢	60-69 才	20-29 才	30-39 才	40-49 才	50-59 才

なお、図の中に記した A～E のアルファベットは、表 5 に示したような典型的な特徴をもった人の位置を、座標平面の目印となるように配置したものである。8 つの特徴のどれかが異なるごとに目印の位置が上下左右に移動することになるが、そうして現実に出現したすべての位置をプロットしたのがこの図の意味するところである。

2) 職業の空間的配置

この個人の分布から比較したい特徴について下位集団の平均点 ($\hat{y}_a^k$) を求め、社会空間の「白地図」の上にプロットすることができる。それは空間構築に関わった変数でも、関わらなかった変数でも構わない。また、平均点とともに線や面で特徴を表示することも、数値に要約しただけでは分からない関係を浮かび上がらせる効果をもつだろう。

図 6 は、この手続きから 17 個の職業を社会空間のなかに配置したものである。図 5 と図 6 を見比べることから、左上のブルーカラー層の領域が密集していて差異が小さいこと、右下の専門職及び会社経営者の層が薄く差異が相対的に大きいこと、時間とともに分布の重心が左下の農林漁業から右上の事務・販売・技術といった組織のホワイトカラー層に移動してきたことなどの基本傾向を確認することができる。また、この図では目印 D が右下の位置にあるが、それを頂点として D-C-B と D-E-A の系列を辿れば、ブルデューが模式図として描いた職業の社会空間図（ブルデュー 1979=1989, 193-94 頁）とよく似た配置が得られていることが分かる。ただし、日本のデータでは教育変数と所得変数の並びが交差配列をなさないので、経済資本系列と文化資本系列の間の幅が狭く、重なりが大きなものになっている。

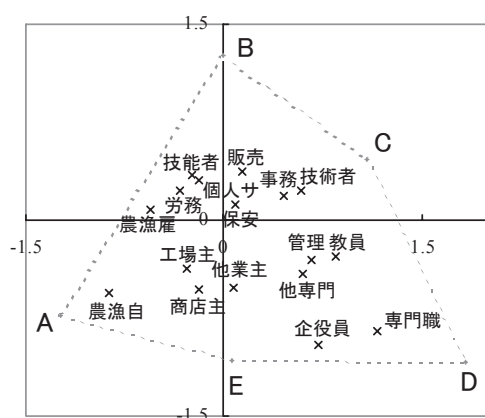


図 6 第 1-2 平面への現職の配置

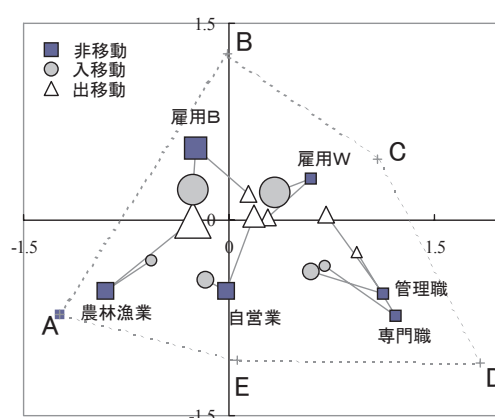


図 7 第 1-2 平面における移動履歴の投影

また、図 7 は本人の現職と父親の職業が同じ場合を「非移動」、現職が同じで父職が異なる場合を「入移動」、父職が同じで現職が異なる場合を「出移動」として区別し、6 つの職業分類でそれぞれに該当する個人の重心を表示したものである。マーカーの大小はサンプル数の違いを反映しており、農林漁業からの出移動と雇用 B 及び雇用 W の入移動でその規模が大きくなっている。つまり、この間の農業層からブルーカラー層及び下級ホワイトカラー層への移動が A-B ラインの方向で生じた様子が捉えられている。

さて、この図の最も重要な特徴は非移動層が（ある意味で当然のことながら）A~D で囲まれた領域の周縁に位置し、2 世代が同じ職業に就くことで社会空間上の特徴を強めている

ことである。逆に、出移動者の中心は他の職業群が重なり合う中央付近に集まり、その移動先は各々の出身の教育水準を反映して、第1軸の並びに沿ったものとなる。また、入移動者は同じく中央の交錯する領域から牽引されてくるが、自営業、雇用B、雇用Wの人材給源はそれぞれのコア（非移動者）に近いところにあり、管理職、専門職の給源はそれらのコアから離れたところにある。したがって、ブルデューが行ったように目印Dを頂点とした一部の領域を「支配階級」として取り出したなら、『ディスタンクシオン』の分析と同様に「古参層」と「新参層」の違いが浮かび上がってくることだろう。

このように所定の変数によって社会空間の輪郭が定まり、そこに空間構築に関わったすべての個人が位置づけられたなら、彼らの所持する情報を用いて社会空間の輪郭をさらに詳しく描いていくことができる。同様に、社会空間を「白地図」として下位集団の位置をプロットすることから各々の変数の意味を解釈していくことができる。そのようにして算出されたカテゴリー平均点を対応する固有値の平方根でスケールしなおしたものが、式(2)に示したようにカテゴリー空間におけるその変数の主座標となるのである。

3) 意識変数の補充的プロット

つぎに、この社会空間に意識変数を投影して別の角度から空間のリアリティを確認してみよう。まず、5時点の「階層帰属意識」を取り上げ、図8の左図に各々の平均点のプロットを示す。ただし、「上」の回答はどの時点のデータでも少ないので「中の上」と合併している。図にあるように、カテゴリーの並びが直線的であること、マニュアル職業が集中する左上の領域から専門職や会社経営者が散在する右下の領域に向かって伸びていることから、まずは予想通りの結果が得られているといえるだろう。各時点の並びはほぼ並行的で、この間に重心の移動はあったものの空間の基本構造に大きな変化がなかったことが示唆されている。

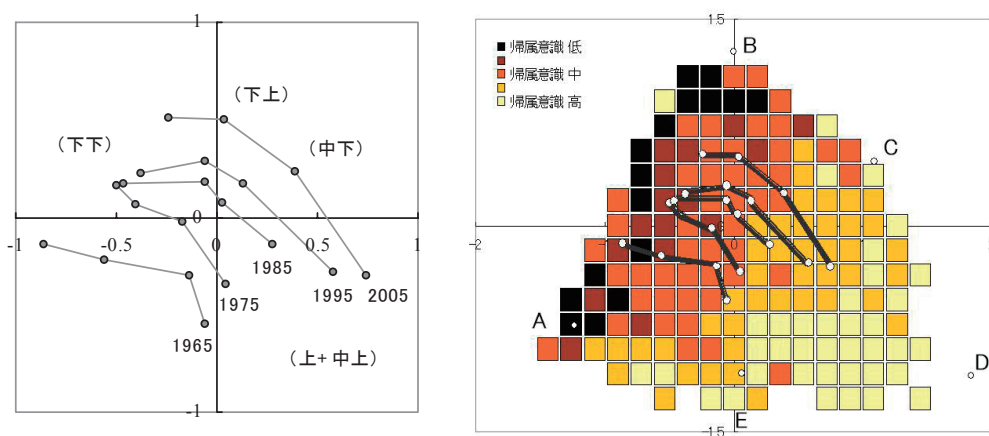


図8 社会空間における「階層帰属意識」の投影

さらに、右図は社会空間の全体を小単位に区分けし、階層帰属意識を数量化¹⁵したときの平均点で各々の枠目の色を塗り分けたものである。いずれかの枠目に位置を占めた個々人が「階層帰属意識」を尋ねられて、「上」や「中の下」などの旗を掲げた状態を集約的に表現しているとみなして欲しい。個人単位でみればどのカテゴリーも空間の全域に観察することができるが、こうして近隣の様子を集約していけば自ずと社会空間の局所の特徴が浮き上がってくる。こうした意識の連続した分布からも、社会空間のリアリティを確認することができるだろう。ここでは、とくに「下の下」や「下の上」がA-Bの並びに密集し、分布域の幅がかなり狭くなっていることが分かる。また、高度成長期以後に差が小さくなっていた階層帰属意識の分布が近年になって左端をA-Bラインに置きながら右端をD方向に伸ばしてきている様子も見て取れる。つまり、高学歴化の進展とホワイトカラー的職業の増大により、農林漁業及びブルーカラー的職業から図の右方向に大量に人が動き、それとともに社会空間上の位置が多様化してきたことが、階層帰属意識が再び広がってきていることの背景になっていると推察できる。

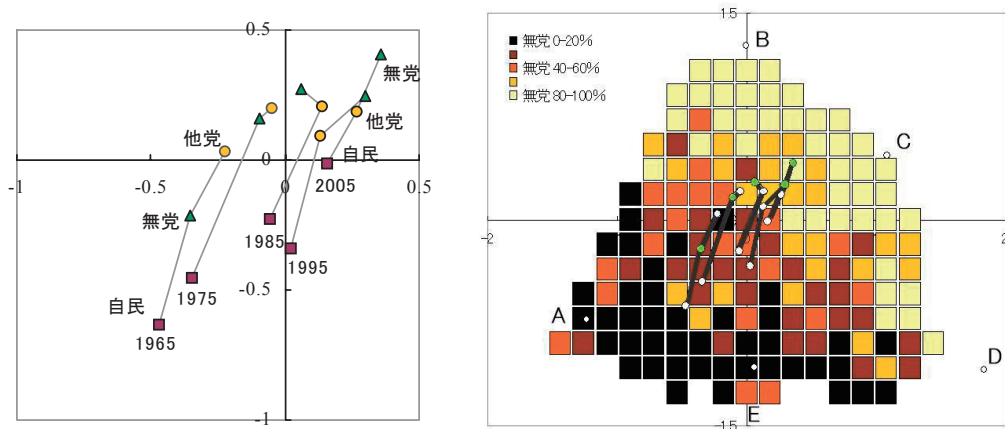


図9 社会空間における「支持政党」の投影

同様の投影図を、「支持政党」について作成したのが図9である。この間の政党の離合集散から政党名による比較が不可能なので、この期間の主たる政権政党であった「自民党」（自民）と「それ以外の政党」（他党）および「支持政党なし」（無党）の3つの下位集団を設け、それぞれについて各調査年の平均的な位置を示している。まず気付くのは、これらの分布が階層帰属意識に直交している点である。つまり、支持政党の基本対立軸は年齢と資産所有の第2軸にあり、階層帰属意識の対立には必ずしも重なってこなかったといえる。また、支持政

¹⁵ 連続したカテゴリーの累積比率をもとに標準正規分布の臨界点 z を与えて平均を求めている。なお、サンプル数が10人に満たない枠目は表示から除いている。（図9に表示した枠目も10人以上に限っている。）

党なしの無党派層が高度成長期以降は政党対立の枠外に逃れてしまった点も興味深い。さらに右の図は、先ほどと同じ樹目で「支持政党なし」の比率を 20% ずつ 5 段階に色分けして示したものである。ここでも、政治意識は社会空間上で連続した分布をなしている。無党派が D-C-B 系列の高学歴若年層を中心とする事務・技術・販売のホワイトカラー層の周縁により多く集まり、D-E-A の経営・管理、自営業主、農業の高年層でその比率がきわめて低いことは、予想通りといえるだろう。図 6 及び図 7 の結果と比べると興味深いのが、ここでは A~D 枠の周縁になるほど傾向がはっきりしており、政治意識が社会空間の軌道と多いに関係のありそうなのが示唆されている。

5 時点のデータすべてに共通する変数はこの 2 つに限られるので意識変数の投影はここまでするが、個別時点の部分的な情報でも社会空間の特徴をつかむ目的で例示的に利用する限りとくに不都合があるわけではない。初めにも述べたように、同じ社会空間を共有しながら多様な側面を同時に捉えることで、階層化の状況がより理解しやすくなるのである¹⁶。また、こうした形の意識変数の検討は、それらの意識が社会空間上に傾向的な広がりをもっていか否かを明らかにしてくれるので、階層意識の分析において広義の階層意識を弁別するのに有効だろう。従来の SSM 研究は、「客観的階層」を前提としない意識でも階層に関連すると予想される意識はすべて広義の階層意識に含まれるとの立場をとってきたが（たとえば原・盛山 2000）、研究対象の定義としてはいささか拡散的である。ここに描いたような社会空間を「客観的階層」とみなすことで、階層意識についてもう少し限定的な（あるいは系統的な）取組みが可能となるはずである¹⁷。

6. まとめと課題

ブルデューの分析を社会研究の方法論として継承・発展させようとしている欧州の社会学者及び統計学者たちの仕事を参考に、MCA を用いて SSM 調査データの分析を行った。その方法は、階級・階層を所与とする従来の構造的アプローチとは異なる発想を採用しており、社会空間アプローチとして階層研究に新しい分析視角を持ち込むものといえる。また、階級・階層の通俗的理解から自由であるということの他に、特定の社会空間が構築されたなら補充変数の技法を通してあらゆる調査項目の布置を共通基盤の上で検討することができるという体系性も、この方法がもつ優れた特徴といえる。

他方、本稿における SSM 調査データの分析結果は、この 40 年間の日本の階層構造の巨視的な安定性を示すものであり、そしてまた西欧の分析例に比べたときの日本の社会空間がもつ特徴的な差異を示唆するものでもあった。交差配列仮説のようにブルデューの見方が成立

¹⁶ ここでは個人分布の特徴づけを「線」と「面」の平均で行なったが、ラ・ルーとルアネは下位集団の中心とパラツキを同時に表示する楕円の利用を勧めている。透視図の効果を期待するなら、たしかに楕円の作図が有効だろう。

¹⁷ これについては、近藤（2008）も参照のこと。

しない面もあれば、職業の空間的布置や世代間移動がつくる軌道の特徴などブルデューの議論に符合する面もあった。ともかく、そうした検討を通して、限られた時空間で実践されたブルデューの分析が階層研究の一般的な文脈に十分に乗り、彼自身が期待していたように新しい知識を生成するモデルとして通用することが確認できた。

この先の分析としては、全体をいくつかのクラスター（階級）に分けて個別に階層化の動向を吟味することや、さまざまな行動・意識変数を用いて社会的な分化や対立の関係を捉えることなどがあげられる。趣味や文化的活動に焦点をあてた生活様式空間の分析も当然視野に入ってくる。さらに、この方法を東アジアや西欧諸国の国際比較に広げて、社会構造や文化の差異を捉えるための方法として鍛え上げていくことが今後の大きな課題となる。いずれにせよ、社会空間アプローチは階層研究の新しい可能性を示すものであり、内外の研究者による専門分野を超えたさまざまな連携が望まれる。

【文献】

- ブルデュー, P., 1989, 『ディスタンクシオン』, 石井洋二郎訳, 藤原書店. (原本の出版年は 1979)
- ブルデュー, P., 1990, 「差別化の構造 - 日本で『ディスタンクシオン』を読む」, 加藤晴久編集『ピエール・ブルデュー超領域の人間学』, 藤原書店.
- Domanski, H. and Z. Savinski, 1986, "Occupational mobility in Poland, 1972-82," *Social Science Information* 25-4.
- Duncan=Jones, P., 1972, "Social mobility, canonical scoring, and occupational classification," in *The Analysis of Social Mobility: Methods and Approaches*, edited by Keith Hope, New York: Oxford University Press.
- Featherman, D., F. L. Jones, and R. Hauser, 1975, "Assumptions of social mobility research in the U.S.: the case of occupational status," *Social Science Research* 4.
- Greenacre, M. J., 2006, "From simple to multiple correspondence analysis," in M. Greenacre and J. Blasius eds., *Multiple Correspondence Analysis and Related Methods*, Chapman & Hall/CRC.
- 原純輔・盛山和夫, 2000, 『社会階層－豊かさの中の不平等』, 東京大学出版会.
- Hjellbrekke, J., B. Le Roux, O.Korsnes, F. Lebaron, L. Rosenlund and H. Rouanet, 2007, "The Norwegian field of power Anno 2000," *European Societies*, 9(2).
- 石井洋二郎, 1993, 『差異と欲望－ブルデュー『ディスタンクシオン』を読む』, 藤原書店.
- 今田高俊・原純輔, 1979, 「社会的地位の一貫性と非一貫性」, 富永健一編『日本の階層構造』東京大学出版会.
- 片岡栄美, 2003, 「「大衆文化社会」の文化的再生産 - 階層再生産, 文化的再生産とジェンダー構造のリンケージ」, 宮島喬・石井洋二郎編『文化の権力－反射するブルデュー』, 藤原書店.
- Klatzky, S. R. and R. W. Hodge, 1971, "A canonical correlation analysis of occupational mobility," *Journal of the American Statistical Association*, 66-333.
- 近藤博之, 2006, 「移動表による職業的地位尺度の構成－オーディネーション技法の応用－」, 『理論と方法』, Vol. 21, No.2: 313-332.
- 近藤博之, 2008, 「社会空間アプローチによる階層と教育の分析」, 米澤彰純編『教育達成の構造』SSM 調査シリーズ 5, 2005 年社会階層と社会移動調査研究会 所収.
- Lebart, L. A. Morineau and K. M. Warwick, 1977, *Techniques de la Description Statistique: Methodes et Logiciels pour l'Analyse des Grands Tableaux*, Bordas, Paris. (邦訳は, 翻訳者の論文を加えた新版と

- して出版されている。大隅昇・L. ルバール・A.モリノウ・K. M. ワーウィック・馬場康維『記述的多変量解析法』, 日科技連, 1994)
- Le Roux, B. and H. Rouanet, 1998, “Interpreting axes in multiple correspondence analysis: Method of the contributions of points and deviations,” in Blasius, J. and M. Greenacre eds., *Visualization of Categorical Data*, San Diego: Academic Press.
- Le Roux, B. and H. Rouanet, 2004, *Geometric Data Analysis: From Correspondence Analysis to Structured Data Analysis*, Kluwer Academic Publishers.
- Oesch, D., 2006, *Redrawing the Class Map: Stratification and institutions in Britain, Germany, Sweden and Switzerland*, Palgrave, Macmillan.
- Pages and Becue-Bertaut, 2006, “Multiple factor analysis for contingency tables,” in M. Greenacre and J. Blasius eds., *Multiple Correspondence Analysis and Related Methods*, Chapman & Hall/CRC.
- Prieur A. and L. Rosenlund, 2005, “Social and cultural distinctions in Aalborg, Denmark,” paper to Rethinking Inequalities – 7th ESA conference, Torun, Poland.
- Rosenlund, L., 1996, “Cultural Changes in a Norwegian Urban Community: Applying Pierre Bourdieu's Approach and Analytical Framework,” *International Journal of Contemporary Sociology*, 33(2).
- Rosenlund, L., 2000, “Social Structures and Change in Norway: Economic and Cultural Dimensions,” *International Journal of Contemporary Sociology*, 37(2).
- Rosenlund, L., 2001, “Postmodern disorder or social regularities of the advanced and differentiated society,” a discussed paper at the workshop “Cultural Capital and Social Classes,” Uppsala university, 12-14, Oct, 2001. <<http://www.skeptron.uu.se/broadly/sec/p-rosenlund-postmod-01.doc>>
- Rouanet, H., 2006, “The Geometric Analysis of Structured Individuals×Variables Tables,” in M. Greenacre and J. Blasius eds., *Multiple Correspondence Analysis and Related Methods*, Chapman & Hall/CRC.
- Rouanet, H. W. Ackermann and B. Le Roux, 2000, “The geometric analysis of questionnaires: The lesson of Bourdieu's *La Distinction*,” *Bulletin de Méthodologie Sociologique*, 65.
- Rytina, S., 1992. “Scaling the intergenerational continuity of occupation: Is occupational inheritance ascriptive after all?” *American Journal of Sociology*, 97-6.
- Sorokin, P. A., 1959, *Social and Cultural Mobility*, The Free Press of Glencoe, Illinois.
- Weininger, E., 2005, “Foundations of Pierre Bourdieu's class analysis,” in E. O. Wright ed., *Approaches to Class Analysis*, Cambridge University Press.
- 安田三郎, 1971, 『社会移動の研究』, 東京大学出版会.

Constructing Social Space by Multiple Correspondence Analysis: A New Orientation in Social Stratification Research

Hiroyuki KONDO
Osaka University

This paper applied Pierre Bourdieu's approach and analytical framework in *Distinction* to examine the class and stratification system of Japan. Five dataset of the periodical SSM survey from 1965 to 2005 were used in the comparative design, and the relationship among socioeconomic variables was analyzed through constructing a social space by multiple correspondence analysis (applying Le Roux and Rouanet's Specific MCA). Major findings are the following. First, the ratio of total variance explained by the first two axes is larger than

European comparatives, which seems to be consistent with a disproportional emphasis in educational credentials in Japanese society. Second, although the relative positioning of occupations in the social space is found to be similar to Bourdieu's configuration, the axes have different meaning so that economic capital and cultural capital are distributed along the same direction, not the opposite direction as showed in *Distinction*. Third, we confirm that the basic structure of social space has been stable for at least an examined term and it can predict a distributional pattern of subjective items, for instance self-evaluation of strata attribution or political parties which one supports, as suggested by the homology hypothesis.

Keywords and phrases: social space, Bourdieu, multiple correspondence analysis

既発表成果一覧

1. 学会等報告

元治恵子. 2006. 「職業評定の日韓比較」 日本社会学会第 79 回大会発表要旨集録.

元治恵子・中山厚穂. 2006. 「職業評定の構造－日本と韓国の比較－」 日本分類学会研究報告会.

2005年SSM調査シリーズ 12
社会調査における測定と分析をめぐる諸問題
前田 忠彦 編

2008年3月10日発行

発行

2005年SSM調査研究会

〒980-8576 仙台市青葉区川内 27-1 東北大学大学院文学研究科行動科学研究室内
2005年社会階層と社会移動調査研究会事務局（発行担当）

〒980-8576 仙台市青葉区川内 27-1 東北大学大学院文学研究科行動科学研究室内
2005年社会階層と社会移動調査研究会事務局（事務局）

