

《研究ノート》

『日本帝国統計年鑑』を電子化しての雑感

五十嵐 彰

【要旨】 本稿では、1882年（明治15年）から大戦によって中断される1940年（昭和15年）まで毎年継続的に収集された『日本帝国統計年鑑』について、データとしての特性と重要性、筆者が実際に電子化したデータの紹介、データを実際に電子化しての問題点、最後にその他の有用な戦前データと今後の展望について紹介する。今回は都道府県別外国人数をはじめとする14変数を電子化し、戦前の外国人数の推移について知見を得た。しかしながら、同時に電子化には多くの困難があり、それを1) 形式、2) 欠損、3) 一貫性の3つの観点から整理する。

キーワード：日本帝国統計年鑑、戦前、外国人数

I 本稿及び『日本帝国統計年鑑』概要

本稿では、『日本帝国統計年鑑』のデータとしての特性と重要性、筆者が実際に電子化したデータの紹介、データを実際に電子化しての問題点、最後にその他の有用な戦前データと今後の展望について記述する。

『日本帝国統計年鑑』とは、1882年（明治15年）から大戦によって中断される1940年（昭和15年）まで毎年継続して集められた統計情報をまとめた資料である。全59巻が刊行されており、図表1に示したような多岐にわたる統計情報を含んでいる。図表1では比較のために項目の順序を変えたが、多くの項目を第1巻から第58巻まで通して収集しているのがわかる。細目は省いたもののこれらの小項目も充実しており、例えば第1回の小項目を足し合わせると279項目、第29回では604項目、第58回では416項目となっている。小項目の中には例えば県別の牛車数（牛一頭引きか二頭引きかの区別あり）や難破船の数、病気ごとの患者数といったものまである。

戦前の日本に関する最も重要なデータといえるが、紙のデータとして保管されており、管見の限り電子化はなされていない。国会図書館デジタルアーカイブには画像情報として保存されたものがあるが、あくまで画像情報であり、かつ30巻以前のもの確認できない。電子化する必要のないデータであれば、現状でも特に問題ないといえる。しかしながら殊『日本帝国統計年鑑』に限っては、重要性が認識されているにもかかわらず電子化が進んでいないといえる。例えば『日本帝国統計年鑑』は海外の経済史研究に活用されており、こうした研究のいくつかはトップジャーナルに掲載されている（e.g., Grossman & Imai, 2008; Tang, 2014）。重要な知見を引き出すことのできる可能性を秘めているにもかかわらず、十分に活用できるよう未だ整えられていないといえるだろう。

II 実際のデータ

筆者が実際に入力・分析したデータを実例として記す。入力の際には社会学研究科大学院

生の小松恵さんと宮澤篤史さんに大いに助けてもらった。この場を借りて改めて感謝申し上げます。

項目内容	第1回	第29回	第58回
土地	土地	土地	土地
天候		気象	
人口	人口	人口	人口
第一次産業	農業 山林 漁業及び製塩	農業 山林及び狩猟 漁業及び製塩	農林及び水産
第二次産業	鉱山 工業	鉱業 工業及び賃金 築造	鉱業及び工業 労働及び職業紹介
第三次産業	通運 銀行及び金融 外国貿易	陸運 水運 郵便及び電信電話 内国商業及び会社 銀行及び金融 貯金及び保険 外国貿易	交通 商業及び金融 貿易
病院・衛生	衛生	衛生	
宗教	社寺	社寺及び教会	教育，神社及び宗教
教育	教育	教育 教育および慈恵	社会事業
治安	警察 監獄 司法	警察 監獄 裁判及び登記	警察，衛生及び災害 司法
軍	陸軍 海軍	陸軍 海軍	
政治	政事	議員選挙 官吏公吏及び恩給 財政	選挙，官吏公吏及び恩賞 財政 産業国体
財閥等	財閥	爵位勲章及び褒章	
開拓地	北海道	北海道	
植民地・領土		朝鮮 台湾 樺太	

図表 1：3 時点における『日本帝国統計年鑑』掲載内容

今回は自分自身の専門である外国人研究のために『日本帝国統計年鑑』のデータを入力した。都道府県レベルの変数のみを用いている。研究目的としては、外国人の居住行動の規定要因を探ることであり、なぜある地域に外国人が多く居住するのかを検討することにある。外国人の数が増えることにより、経済や福祉など地域の状態が(良い方向にも悪い方向にも)変わると言われてきており、そのため外国人がどこに住むかを検討することは重要な課題であった。多くの先行研究では、事前の外国人住民の存在が新たな外国人移住者を生むことがわかっていた。しかしながらこれらの研究は、ではそれらの事前の外国人住民がどのように居住地選択をしたかという問いに答えることは出来ない。こうした問いに答えるには、外国人ネットワークがない(もしくはほぼない)状態にまで遡る必要がある。日本では鎖国解禁以後も外国人の居住地域を複数の地点に限定しており、1899年に至るまで外国人はこれらの地域以外に住むことは原則許されなかった。『日本帝国統計年鑑』には1899年およびそれ以降の県別外国人数が記載されており、この情報を使うことにより、外国人ネットワークが(ほぼ)存在しない状況において、なぜ外国人がある地域に移り住んだのかという問いに答えることができる。このようなモチベーションのもと、小松さんと宮澤さんにデータを入力してもらった。実際に集めたデータは図表2の通りとなる。

	年	欠損年
外国人数	1899-1938	1912, 1914, 1925
日本人人口	1898-1938	-
男子不就业率	1898-1937	-
女子不就业率	1898-1937	-
犯罪者検挙数	1898-1938	1924
米生産高	1898-1938	-
甘藷・馬鈴薯生産高	1898-1938	-
製糸工場数	1898-1937	1906
反物・帯値段	1898-1937	-
農業企業数	1898-1938	-
商業企業数	1898-1938	-
工業企業数	1898-1938	-
水陸運輸業企業数	1898-1938	-
所得税額	1902-1938	-

図表2: 『日本帝国統計年鑑』収集データ一覧

実際に収集できた年と、データが完全に欠損となっている年を示した。この他にもいくつかの県が欠損となっている場合がある。外国人数については、制度として外国人が開港地・開市地以外に居住できるようになったのが1899年であり、その年からの収集となっている。それ以前のデータは開港地・開市地のみのものであり、今回は収集しなかった。

所得税額については、制度として整備されたのが1902年であり、それ以降のデータとな

っている。所得税額は県ごとの生産高を分析するために収集している。本来であれば県民所得が望ましいが、県民所得の年別データは存在しないようであり、そのため代替変数として所得税を収集した。今回収集したのは所得税の課税対象となる県別所得総額に対し、税率をかけたもの、つまり実際に徴収された金額となる。しかしながらこれには二つの問題があり、一つは各年で税率が異なっており、経年比較に適さないことである（詳しくは大蔵省主税局、1988 参照）。二つ目の問題として、当時の所得税は現在でいう法人税、公債・社債利子、そして所得税の三つを混合したものとなっている（それぞれ第一種、第二種、第三種所得税と呼ばれていた）。そのため総額では所得を十分に反映したものにはならない。第三種所得税の対象、すなわち課税対象となる個人所得の県別総額を収集するほうがより適していると言えるが、こちらにも問題がある。所得税の課税対象は一定所得以上を稼いでいるものに限定され（300 円以上など）、その基準となる所得も何度か変動している（詳しくは大蔵省主税局、1988 参照）。加えて、課税対象となる所得のデータが収集され始めたのは 1903 年からであり、こちらのデータを用いると 1 年分のデータを捨てることとなる。ただ課税対象となる納税者人数もわかっており、一人あたり課税対象所得を計算できるという意味では、こちらの変数のほうが望ましいかもしれない。

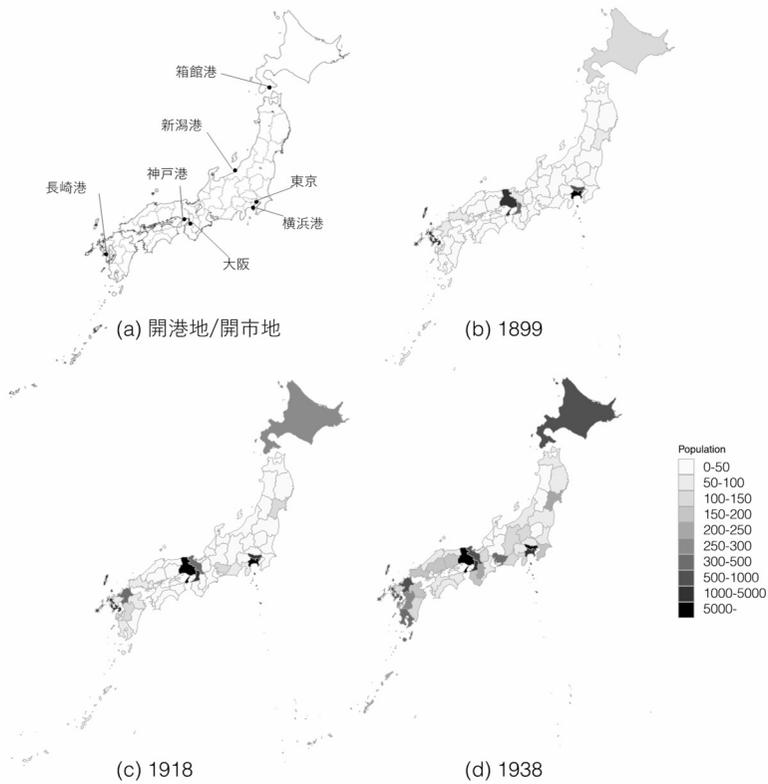
企業数についてだが、これ以上細かい分類はない。また本来であれば企業の生産高などを用いたいところだが、あいにく企業数もしくは資本金しかデータがなかった。後述するが、米は石高、甘藷・馬鈴薯（じゃがいもとさつまいも）は昭和 3 年までは貫で記載されており、それ以降はキロリットル（ヘクトリットル）で記載されていた。

これらの変数の中で、外国人数について少し掘り下げて論じるために、図を示す。図表 3 では年別総外国人数の推移を、図表 4 では日本地図を用いて外国人数の広まりを示している。1912 年、1914 年、1925 年はデータが記載されていないので空欄となっている。この図は大きく分けて 2 点の特徴があるといえる。1 点目は日本に居住する外国人数が右肩上がりの傾向にあることである。確かに年に応じて増減はしているものの、基本的には減少後すぐに回復し増加するというパターンをたどっているといえる。2 点目は大きな災害や恐慌の際には外国人数が減少しているということである。例えば図表 3 からは、1923 年の関東大震災や、1931 年以降の昭和恐慌、1937 年以降の日中戦争時に外国人数が大幅に減少している。こうした傾向は現在にも通じるものがあり、例えば 2009 年の金融危機や 2013 年の東日本大震災では外国人数は大幅に減少している。関東大震災後は復興特需ともいえるべきものが起きているのだろう、外国人数は V 字カーブを描いて増加している。恐慌後は時間を置いて回復しているが、これも現代に通じる傾向といえる。

次に、図表 4 に外国人数をプロットしたものを示す。図表 4a は開港地、開市地の場所を示したものである。開港地は箱館、新潟、横浜、神戸、長崎であり、開市は東京と大阪である。図表 4b から 4d は 1899 年、1918 年、1938 年それぞれの時点における外国人数を日本地図にプロットしたものである。開港地・開市地ではやはり外国人数が多いが、その中でもバリエーションがある。横浜や神戸は多いものの、北海道や新潟になるとその他の地域とそこまで変わらない。1899 年の時点では当然ながら開港地・開市地に外国人が集中しているが、年を追うごとに、特に 1938 年になると全国に外国人が散在している様子がわかる。では、こうした移住過程はどのように説明できるのだろうか。現在論文としてまとめているところであり、今後はそちらの論文にて結果を報告する予定である。



図表 3 : 1899 年から 1938 年にかけての総外国人人数



図表 4 : 1899 年から 1938 年の 3 時点における外国人人数プロットと開港・開市場の位置

III 実際に電子化しての雑感

本節では上述したデータを電子化した際の雑感について簡潔に述べる。まずもっとも関心の高いと思われる、どの程度時間を要するかという点である。小松・宮澤両氏の入力作業時間に関する情報も掲載することを考えたが、お二人に入力してもらったのはあくまで必要な変数の抜き書きであるため、1ページ当たりにかかる時間の計算が少々ややこしいことになる。そこで改めて自分で1ページ分(15行×8列プラス変数名)を入力してみたところ、9分半で入力できた。巻ごとに揺れはあるものの、入力の際に使った年鑑は約1200ページであるため、1冊を入力し終えるのに190時間かかる計算となる。また今回入力した情報は行数も列数も比較的少ないものであるため、これ以上時間を要すると見てよいだろう。スキャンデータとOCRを活用して取り込むという可能性もある。しかし紙自体が非常に薄く、裏写りが激しいため、十分に質の高い機材が必要となるだろう。

データとして入力する際にも、他にも問題点がある。1) 形式、2) 欠損値、3) 一貫性、の観点からまとめる。

1. 形式

『日本帝国統計年鑑』の形式は現代の統計表の形式と大幅に異なっているため、入力(手入力、機械入力にかかわらず)大きな困難が予想される。まず第1巻(1882年、明治15年)から第23巻(1904年、明治37年)までは漢数字でデータが表記されている。これらの巻はすべて縦書きであるため、右開きである。アラビア数字に慣れている分、入力作業においては多少困難となるかもしれない。地の文も漢字カタカナ混じりである。また都道府県の順番も異なっている。官庁データを利用したことがある人ならわかると思うが、都道府県の並びは北海道を1番、沖縄を47番とした並びで統一されている。しかしながら『日本帝国統計年鑑』では、第1巻から第9巻までは東京からはじまり京都、大阪、神奈川、兵庫、長崎といった都市圏の順になっており、第1巻では都道府県の数も40である。第10巻から第32巻までは東京、神奈川、埼玉、千葉、といった地域ごとの順になっている。第33巻から第42巻までは北海道からはじまる現在の都道府県の並びに近いものとなるが、宮城県などの位置が微妙に異なっている。第43巻からは現在の並びとなっている。あくまでこれらは慣れの問題であるものの、順応するまで多少時間を必要とするだろう。

またその他の形式の問題点として、非常に複雑な表の作り方をしている。図表5に明治37年の獣医の数を示したデータを例示したが、変数が何重にも入れ子になっていることがわかる。例えば「獣医」の下に「免状下付人員」と「仮免状下付人員」があり、さらに「免状下付人員」は3つのカテゴリに分か

部中州本		地方		
神奈川	東京			
一	四	及試	免状 下付 人員	獣 医
一	一	第 験		
	五	校 官		
一	八	卒 立 業 学		
一	二	校 府	免 員 状 下	
三一	四六	卒 県 業 立 学		
一	一	計		
一	二			
六	九			
四	七			
二	一	合		
〇	三	計		
	六			

図表5：第23回日本帝国統計年鑑の獣医数に関する表

れている。加えて、こういった表では一セル一情報が原則であるが、なぜか「府県立学校」の項には二つの情報が入っている。この情報については特に説明がない。これ以外にも「0」は四捨五入の結果1未満になったものを示し、「|」は該当数なし（実質的に0）を示し、「…」は不詳、などといった独特の記法を用いている。これに加えて、ある年では不詳の項目それ自体を書かないことで示す場合もある（例えば外国人のいない都道府県は、表に記載されない）。電子化の際にはこうした困難を乗り越える必要があるだろう。

最後に、各年鑑は「明治〇〇年刊行」などと表紙に記載されているが、この年が必ずしも各種統計値が収集された年と同一であるとはいえない。例えば第29回日本帝国統計年鑑は明治43年に刊行されているが、「現在本籍人口」は明治41年の値、「本邦在留外国人」は明治42年の値、「国有鉄道職員数」は明治43年の値、といった具合である。これは一機関が統一的にデータを収集するのではなく、各省庁・機関が収集したデータを一冊にまとめるという『日本帝国統計年鑑』の性格からくるものと思われるが、それでもなおこの点に関しては自覚的に利用すべきだろう。

2. 欠損値

欠損値自体は避けられない問題だが、公的統計として欠損値が存在するのは比較的珍しいのではないと思われる。例えば1899年の外国人に関する表では、「データが期日までに届かなかった」という理由で京都の値が欠損となっている。理由は様々だが、ある年の一地域のデータが欠損となっているケースは『日本帝国統計年鑑』において比較的多く見られる。

また、ある変数の値が年によってまるまる収集・収録されていない場合もある。これは制度的な理由と、非制度的な理由の二つから生じている。制度的な理由として、例えば所得税という制度が導入されていない1902年以前には、当然所得税という項目は存在しない。また外国人が決められた地点以外に居住することが許されなかった時代（1899年以前）には、決められた地点以外における外国人数というデータは存在しない。他方、こうした制度的な理由とは関係なく、ある年の変数が収集されていない場合がある。例えば図表2で示したような3つの変数（外国人数、犯罪者検挙数、製糸工場数）はそれぞれある年度のデータがそっくり欠損になっている。

このような単年の欠損であればよいが、十数年にわたって収集していたにもかかわらず、それ以降は収集しなくなる変数もある。例えば県別の石炭消費量は、都道府県別の経済とエネルギー消費に関して重要な知見を与えてくれる変数と言える。しかしながらこの項目は第9巻から第25巻まで継続的に収集されていたにもかかわらず、第26巻以降は日本全体の値を収集するにとどまっており、その後第33巻以降は年次データすら収集されなくなっている。こうした十年単位の欠損は分析の幅を狭めるため、自分の関心のある変数がどの程度の期間にわたって収集されているかを事前に確認する必要があるだろう。

3. 一貫性

年ごとに一貫性のない収集をしている場合がある。一貫性のなさは、同一の内容を異なる項目名で指している、同一の項目名だが年ごとに内容が異なる、という二点にまとめられるだろう。

まず、同一の内容を指していると思われるが、項目名が異なっている場合がある。例えば「検挙された犯罪件数(犯罪者数)」という項目は年によって「全国犯罪者捕拿及自主」「就捕犯罪者違警罪犯罪諸規則違反者」「道府県検挙犯罪人及警察犯慮罰令諸犯罪人員」「司法警察官ノ取扱ヒタル犯罪者及犯罪嫌疑者」「犯罪検挙件数」など、様々な呼称で呼ばれている。対象となる犯罪の種類はおおまかに重複しているので同一項目と捉えて差し支えないと思われるが、利用する際には注意が必要だろう。

次に、同一の項目名だが内容が異なる場合について記述する。先述のように、日本式の度量衡(石や貫など)から西洋式の度量衡(キログラムなど)に昭和4年頃から変換される。変換式自体は一貫しているものの、分析や記載の際には注意が必要だろう。第3章で収集した県別外国人数だが、これも内容が年によって異なっている。具体的には、1910年の韓国併合以降は韓国人が外国人数に入らなくなっている。これは制度としては一貫性をもっているが、経年比較の際には考慮すべき事項だろう。所得税も同様の問題を抱えている。各年で課税免除額の下限が異なっており(1887年以降300円以上、1913年400円以上、1918年以降500円以上、1920年以降800円以上、1926年以降1200円以上、1938年以降1300円以上)、比較には限界がある。

IV その他有用な戦前のデータ

本稿では『日本帝国統計年鑑』について紹介した。しかしながら戦前のデータはこれに限らない。他のデータと組み合わせて分析することで、よりロバストな分析となるだろう。例えば物価は年によって大きく異なっているといえる。戦前の物価水準に関しては(1901年以降ではあるものの)日本銀行がデータを提供しており(参考文献参照)、これを用いることで賃金や流通額の物価を補正することができる。県別ではないものの、戦前の経済に関する包括的な研究を一橋の研究グループが行っており、その成果は『長期経済統計』としてまとまっている(大川・篠原・梅村, 1965)。また当時の天気については年鑑降雨量の合計値や日照時間などのデータが気象庁により提供されている(なぜか奈良県だけ戦前のデータがない)。第2章で示した研究関心に照らし合わせると、天気のよい地域に外国人が移住するという仮説の検定を行える。また天気の情報を操作変数に用いることもできるだろう。

V まとめ

本稿では『日本帝国統計年鑑』の概要、使用例、データ収集の際の問題点について述べた。『日本帝国統計年鑑』は未だに電子化が進んでいないが、重要性は述べたとおりであり、戦前の日本社会に関する新たな知見を多く与えてくれる。特に経済学や社会学において計量手法を通じた歴史研究が盛んになっている昨今において、同データは大きな意義を持つだろう。本稿で述べた使用例では一部のデータしか収集できなかったが、今後は統計年鑑全体を電子化することで、より柔軟な統計分析をのぞむことができるだろう。

参考文献

- 大川一司・篠原三代平・梅村又次監修, 1965, 『長期経済統計』東洋経済新報社.
- Grossman, R. S., & Imai, M, 2008, The evolution of a national banking market in pre-war Japan. *Explorations in Economic History*, 45(1): 17-29.
- 日本銀行, 2019, 「戦前基準の物価指数」(2020年1月7日アクセス) <https://www.stat.go.jp/library/faq/faq17/faq17a07.html>.
- 大蔵省主税, 1988, 『所得税百年史』, 大蔵省主税局.
- Tang, J. P, 2014, Railroad expansion and industrialization: evidence from Meiji Japan. *The Journal of Economic History*, 74(3): 863-886.

Summary

Digitizing “Statistical Yearbook of the Japanese Empire”

Akira Igarashi

This study introduces the characteristics and importance of “Statistical Yearbook of the Japanese Empire”, a series of statistics collected annually from 1882 to 1940. In this study I describe the experience of digitizing a part of the book, difficulties faced during digitalization, future usage of the book, and useful datasets from the pre-war period. For the purpose of my research, I digitalized information regarding the number of foreigners in each prefecture as well as 13 other variables, obtaining insights for foreigner situations during pre-war Japan. However, I experienced three types of problems during the digitalization including the format of the book, missing variables, and consistency of variables across years.

Keyword: Statistical Yearbook of the Japanese Empire; pre-war; number of foreigners