

吉田洋一監修
新数学シリーズ

8

統計調査法

統計数理研究所員

西平重喜著

補訂版

培風館

吉田洋一監修
新数学シリーズ
8

統計調査法

統計数理研究所員
西平重喜著

补丁版

培風館

西 平 重 喜 略 歴

1924 年 東京都に生まれる
1947 年 北海道大学理学部数学科卒業
現 在 統計数理研究所付属統計技術
員養成所長
この間東京大学文学部, 早稲田
大学政経学部非常勤講師 およ
び 第 5 大学 (人文科学) 客
員教授 (1975 年)

主 要 著 書

世論調査 (岩波書店, 吉田洋一共著)
日本人の国民性 (至誠堂, 共編)
日本人の意見 (誠信書房)
選挙の国際比較 (日本評論社)
日本の選挙 (至誠堂)
世論反映の方法 (誠信書房)

© 西 平 重 喜 1979

昭和 32 年 10 月 25 日 初 版 発 行
昭和 54 年 1 月 30 日 補 訂 版 発 行
昭和 57 年 4 月 10 日 補訂第 5 刷発行

新数学シリーズ 8

統 計 調 査 法

著 者 西平重喜

発行者 山本健二

発 行 所 株 式 公 司 培 風 館

東京都千代田区九段南 4-3-12・郵便番号 102

電話 (03) 262-5256(代表)・振替 東京 4-44725

定 価 ￥ 980.

日東紙工組版・中央印刷・三水舎製本

著者の承認をえて検印を省略しました

3341-0327-6955

新数学シリーズ 8・統計調査法

吉田洋一監修

新数学シリーズ

このシリーズは数学の各部門を手っとり早く知りたいと思う人たちのために企画された。したがって、数学専攻の学生ばかりでなく、自然科学、社会科学、技術方面にたずさわる人たちのためのよき入門書たることを志すものである。

上記の目的を達するために、各分冊とも準備的段階に多くの紙数を費すことをつとめて避け、単刀直入に本論に入る方針をとった。しかも、各分冊を読切とし、高校程度の数学以外に予備知識をもたない人にもたやすく理解しうよう、周到な工夫をこらすことを怠らなかった。といっても、また、読みやすいがゆえに初等的事柄ばかりを扱っているとはかぎらない。“高度の数学をわかりやすく応用しやすい形で”というのがシリーズ編集のモットーにほかならないのである。

なお、監修者は各分冊の題目と執筆者を選定するのみでその任務をおわったものとは考えなかった。すべての原稿に目を通したうえ、上記の方針をつらぬくために執筆者と十分に連絡をとり、もって監修の実をあげようとしてきたこともここに付言しておきたい。

1957年1月

監修者

培風館

補訂にあたって

アルバイトにきた学生が手回し計算機を見て、「ウワー、カッコイイ！」と叫びました。この本は 20 年間に 27 刷をかさねましたが、実は、その手回し計算機の使い方をのせたままでした。質問文の意味のとりちがえに注意すべき例として、第 2 次大戦の抑留者が出てきます。これでは読者が著者の意図をとりちがえてしまいます。すっかり書き直そうか、大修正、小修正と考えてみたのですが、同じ本を 2 度書く気にはなれませんし、大修正では木に竹をついだ感じになります。逆に、「これでいいじゃないか」と思います。いまでも年に 2 回は、この本に沿った講義をしますが、別にまちがいもありません。最近では統計学の本が書店の棚の数段を占めるほどですが、この種の、社会調査、世論調査を中心にした本はあまりないようで、やはり続刊の意義はあると思います。

そんなわけで、最小限の修正にとどめました。例題の中に SL（蒸気機関車）時代の停車時間や、東海道線の普通急行が出てきますが、緊張をほぐす御愛嬌と考えてください。ひとくちでいうと、数式の部分はほとんど変更がありません。調査、集計などに関連した部分を書き改めました。

初版当時、そのきざしが見えていた、統計データの新しい解析方法は、電子計算機の普及とともに、盛んに使われるようになりました。そしてまた、世論調査は経済成長が急速なときも、穏やかなときも、ますます重要視されています。著者はいま、世論調査や社会調査がきちんとした手続きでおこなわれ、データの精度に応じた解析がおこなわれるようにと、祈りたい気持です。

1978 年 7 月

西 平 重 喜

は じ め に

この本の構成は、‘もくじ’からもわかるように、実際に役立つようにという立場から、まず調査の企画や方法(第Ⅰ章)から始め、サンプリングの理論は第Ⅳ章にまわして、サンプリングの実際のやり方を第Ⅱ章で述べる。つぎに第Ⅲ章では集計や計算について、これも計算を実際にする方法を中心に説明をした。第Ⅳ章のサンプリングの理論もかんたんな例をあげて説明することにつとめたが、ここでは数式を出さないわけにはいかなかった。しかし、この章をとばして読んでも、第Ⅴ章のサンプリングの結果から全体についての推定や検定をする方法についての説明を理解するのに、さまたげとならないであろう。この第Ⅴ章も実例を中心に説明を進めた。

つぎに、もうすこしこの本のねらいをはっきりさせるために、この本がどんな読者を対象として書かれたかを述べよう。第Ⅰにはある大会社の人事課の O 君である。著者はかれのように統計調査の知識が実際に必要な人に役立つような本にしたいと考えた。そうして、O 君に最新の技術を気にするまえに、これだけのことはどうしても知っておいてもらいたいと思うのである。第Ⅱには社会学や心理学あるいは政治、経済、法律などの文科系の学生諸君の教科書として使える本にしたいと考えた。事実、この本は著者がそのような人たちに対しておこなった講義をもとにして書かれているし、こんごはこれを教科書に使うつもりである。第Ⅲには数学科の学生諸君である。ふつう、数学科では統計が実際に役立つところまで講義がおこなわれていない。一方、世間では数学的な考え方と、実際に数学を使える人を望む声がおきているから、数学科の学生にとっては、副読本のような意味をもつ本にしたいとも考えた。

以上のような読者を対象としたので、著者自身の経験をもとにして、実例ないしは実際におこりうる例によって、書くことにつとめた。したがって、この本は統計調査法という題ではあるが、その例や解説は著者が関心をもつ社会心理学的な諸問題によっている。しかし、社会心理学的といってもべつに

むずかしい概念は持ち出してはいない。いわば新聞の社会面の記事を例にしたようなものであるから、他の領域の読者の理解も深めることになると思うのである。

さきに、実際に役立つようにと念願して、この本を書いたとはいったが、この本はけっして統計調査法の公式集でも、ハンド・ブックそのものでもない。著者が読者に知ってもらいたいと思うことは、統計調査の技術的な結果そのものよりは、その根本にある考え方である。であるから、統計調査の考え方の骨組がすいて見えるような、例をもち込むことにつとめたのである。もし、読者が統計調査の考え方に慣れることができるならば、この本では述べなかった最新の統計理論や、調査技術に出会っても、おおよその見通しがつくであろう。なお、それらの説明を与えるためには、この本でいどのことは十分にマスターしておいてもらわなければならないのである。それだけではなくて、この程度のことをマスターすれば、現実の統計調査に関する諸問題の多くを解明することができるであろう。

なお、この本を書くにあたっては、多くのかたがたのおかげをうけている。ここで、そのお名前をあげつくせないことを残念に思う。

一般的、科学的あるいは数学的な考え方については、吉田洋一先生のお教をうけている。それだけではなく、先生は古屋茂氏、赤撰也氏とともに原稿を何度も読んで、かずかずの注意を与えてくださった。さらに統計学、技術的諸問題やその処理方法については林知己夫氏をはじめ統計数理研究所の諸氏に負っている。釣谷純子さん、石川温子さんはこの本で扱った事例のほとんどすべての集計をしてくださった。また美恵子は相談相手となり、計算をしてくれた。みなさんに深く感謝の気持を表わすしだいである。

1957年 春

著 者

も く じ

補訂にあたって	i
は じ め に	iii

I. 調査の企画と方法

§ 1. 実験・観測・調査	1
§ 2. 調査の構成	3
§ 2.1. 調査の進め方	3
§ 2.2. パネル調査	7
§ 3. 調査実施の方法	10
§ 3.1. 面接調査法	10
§ 3.2. 留め置き調査法	13
§ 3.3. 郵便調査法	14
§ 3.4. 集合調査法	16
§ 4. 質問のつくり方	17
§ 4.1. 質問および調査票をつくるときの一般原則	17
§ 4.2. 質問の形式	20
§ 4.21. 自由回答法	20
§ 4.22. 二項選択法	22
§ 4.23. 多項選択法	22
§ 4.24. 順位を決定するための方法	25
§ 5. 調査票のていさいと調査に必要な書類	27
§ 6. 調査の誤差	28
§ 6.1. 調査の誤差	28
§ 6.2. 調査の信頼性と妥当性	30
§ 6.3. サンプル調査	31

II. サンプルングのやり方

§ 1. サンプルング	34
-------------	----

§ 2. 乱 数 表	37
§ 3. 乱数表の使い方, 単純ランダム・サンプリング	38
§ 4. 等間隔サンプリング	40
§ 5. 2 段サンプリング (サブ・サンプリング)	43
§ 6. 層別サンプリングのかんたんな場合	46
§ 7. 多段サンプリングの場合の層別	52

III. 集 計, 計 算

§ 1. 度 数 分 布 表	57
§ 2. グ ラ フ	60
§ 3. 代 表 値	62
§ 4. ちらばりの度合	65
§ 5. モーメント, 歪度, 尖度	69
§ 6. 平均値, 分散, 標準偏差の計算方法	72
§ 7. 相 関 係 数	82
§ 7.1. 問題のおこり	83
§ 7.2. 回 帰 直 線	85
§ 7.3. 相関係数の意義	88
§ 7.4. 相関係数の計算方法	90
§ 8. 相 関 比	98

IV. サンプル理論

§ 1. 例によるサンプリング理論のあらまし	103
§ 2. サンプルの平均値	108
§ 2.1. サンプルの平均値の理論分布	108
§ 2.2. 分 布	112
§ 2.3. ガウス曲線の性質	114
§ 2.4. 平均値のサンプリング誤差	117
§ 3. 比率(パーセント)のサンプリング誤差	119
§ 4. サンプルの分散などについて	128

§ 5. 層別について	130
§ 5.1. 層別の場合のサンプリング誤差	130
§ 5.2. 各層へのサンプルの割当法	135
§ 5.3. 層別の効果	139

V. 推定と検定

§ 1. サンプルでの平均値から, 母集団での平均値を推定すること	145
§ 2. サンプルでの比率(パーセント)から 母集団での比率を推定すること	147
§ 3. サンプル数の決定方法	150
§ 4. 平均値の差の検定	151
§ 4.1. 平均値の差の分布	152
§ 4.2. 有意差	156
§ 4.3. 平均値の差の検定の手順	158
§ 5. 2組のサンプルの間の比率(パーセント) の差の検定	159
§ 5.1. その考え方	159
§ 5.2. その手順	160
§ 6. 1組のサンプルにおける比率(パーセント) の差の検定	162
§ 6.1. その考え方	162
§ 6.2. その手順	164
§ 7. 分布の間の差の検定— χ^2 検定	165
§ 7.1. どんな場合に必要か	165
§ 7.2. 適合度の検定	166
§ 7.3. 分布の一致性の検定	171
§ 7.4. 項目の独立性の検定	175
§ 8. χ^2 分布, t 分布	180
§ 9. ノン・パラメトリックな検定	181
§ 10. 順位相関係数, 一致係数	183

追記	188
あとがき	191~192
参考書	193~196
問題の答	197~199
付表	
付録第1表 乱数表の一部	201
付録第2表 パーセントのサンプリング誤差の早見表	202
付録第3表 $1/n = 1/n_1 + 1/n_2$ の n の表	203
付録第4表 χ^2 の表	204
索引	205

I. 調査の企画と方法

この本では、調査とは統計調査、しかも社会現象の諸問題についての調査をさすことにする。統計調査というのは集団についての情報を得ることを目的としている。たとえば税務署、警察あるいは福祉関係の事務所の調査、知能検査などによって個人についての情報を得て、その個人に対して対策をとるための調査は、統計調査ではない。また、統計調査ではあまりにも細かいことは問題にすべきではない。

まず第 I 章では調査全体の計画や運営方法について述べる。ただ、被調査者のきめ方については、すこしくわしく話をするために、第 II 章にまわすことにするが、第 I 章でも出てくることばの意味だけはここで述べておこう。結果を出したい対象者全体を実際に調査できないとき、一部分をえらび出して被調査者をきめることがしばしばある。そのえらび出すことを **サンプリング** といい、えらび出された被調査者を **サンプル** という。とくにサンプリングのうち、最もすぐれている方法である **ランダム・サンプリング** というのは、クジ引きと同じ原理で、被調査者をきめることである。この本でとりあつかうのは主としてランダム・サンプリングによる調査法なのである。

§1. 実験・観測・調査 調査の意義とか価値ということを考えたすと、大きな問題になりそうである。ここでは実験について、ポアンカレ (H. Poincaré, 1854-1912) がいっていることを引用してみよう*。

“実験は真理の唯一の根源である。実験のみが我々に何か新しいことを教える、実験のみが我々に確実性を与える。

“観測するだけでは充分ではない。これらの観測を利用しなければならないし、それには一般化を行わなければならない。

“よい実験とは何か。それは一つの孤立した事実とは別のことを我々に知らしめるものであり、我々に予見することを得しめる、換言すれば我々に一般化することを得しめ

* 科学と仮説 (河野伊三郎訳) 第 9 章物理学に於ける仮説、実験及び一般化の役割。

るものである。なるほど、一般化がなければ予見は不可能である。人が仕事をした環境はいつになっても、一度に全部がそのまま再現することはない。だから一度観測した事実はいつになっても再び起こることはあるまい。人の断定しうることは、類似の環境においては類似した事実が生ずるだろう、ということだけである。だから予見するためにすくなくとも類似に助けを藉りなければならない。換言すればこれはもう一般化することである。”

これらの引用はどこまでつづけてもきりがない。実験とか観測ということばを調査ということばに読みかえれば、調査の意義や役割がはっきりとすることであろう。ただ実験と調査とは多少ちがう点がある。実験とは十分に——完全に——条件を整えておいてこれに人為的な操作をくわえ、その前後の変化と操作との関係をみようとするものである。自然科学は実験をくりかえすことによって進歩してきたといえるであろう。しかし自然科学のなかでも、とくに天然現象ととりくんでいる天文学、気象学、地震学あるいは海洋学という部門では、実験はごく限られた場合にしかおこなえない。これらの科学は観測によって進歩してきたといつてよいであろう。この観測というのは各種の現象を計量したり、現象があらかじめ考えた理論のとおりに起こるか否かを、実際にしらべることである。たとえば、南極における観測という場合、南極で漠然と自然現象を眺めていて、珍しいことが起ったら、それをしらべようというのではない。あらかじめ起るべきことを予想しておいて、それに関する観測をつづけているわけである。

こう考えてみると、実験も観測もあらかじめ仮説をたてて、その仮説が成り立つか、成り立たないかを検証する点では、同じである。ただその検証を人為的な変化によってためすか、自然におこる変化を待ちうけて、ためすかのちがいがあるようである。

社会現象の研究や、社会の実態を知るためにも、実験や観測の態度が必要である。すなわち、漠然と事件の起ることを待っていても、新聞記事のようなものしか書けないし、急にある場所に出かけて行っても、ルポルタージュ文学にしかならないであろう。やはり仮説をたてて、その真偽をたしかめるという態度でのぞむべきである*。自然科学のように、問題を細分化してひとつひとつ

* 仮説については、ボアソナレの前出書、第9章のうち、仮説の役割、をみよ。

検証して積み上げてゆくというわけにはゆかず、総合的に考えてゆかなければならないので、こんなんではある。しかし他によい方法はない。

§ 2. 調査の構成 まず、一般的な社会調査の場合の調査の進め方を、順を追って説明し、それから、時間的変化を追う調査の例を述べよう。しかし、被調査者のきめ方（サンプリング）や調査の方法については、あとからくわしく述べるから、それらについては順序としてあけておくだけで、説明ははぶいてある。

§ 2・1. 調査の進め方

1° 調査のねらいをはっきりさせる。いままでの文献や資料、または知識を活用して、問題点をうきばりにしてゆき、質問の形に持ち込むのが第1段階の仕事である。調査の範囲や対象、サンプリングの台帳（被調査者をきめるための名簿）、時期、その他の準備をととのえることも、この段階できめてしまうことになる。

2° サンプリング（被調査者の選定）。広い地域——たとえば、東京都23区とか日本全国——のような場合には、サンプル（被調査者）個人を中央でえらぶことはできない。したがってサンプリングは2段階（以上）に分かれ、中央では第1段のサンプリング——たとえば、投票区とか市区町村のサンプリング——をおこない、調査地点を決定する。このとき、必要があれば、サンプリングの精度をあげるために、等質なものからなるグループをつくること——これを層別するという——もある（層別については第II章をみよ）。そうして調査地点で第2段のサンプリング——個人のサンプリング——をおこなう。第2段のサンプリングは、東京での調査のような場合は、企画者があらかじめ区役所などをまわってやってしまうが、全国調査のような場合は、4°の現地作業にふくまれる。

3° 調査員への指示(instruction)。調査票 その他必要書類を調査員に渡し、また、調査のしかたや、ときにはサンプリング（第2段）などについて、調査員に指示を与える。このとき、口頭で説明するほかに、簡単明瞭な「調査員のてびき」(interviewer guide) を渡したほうがよい。指示事項を

複雑にしても調査の現場ではその通りにはできないから、必要最小限にとどめなければならない。この指示は、調査の企画者が直接調査員に接しておこなうのがよいわけである。ある全国調査では各地(大学)に調査員を集めてもらって、中央から数名の指示者(instructor)を派遣し、指示者が調査員に直接指示を与えた。しかし別の全国調査では、各地の責任者(調査員を何人か監督する人、大学の先生)を東京と大阪に集め、調査員への指示は責任者を通じて与えた。前者のほうが指示が徹底するし、企画者と調査員の人間的接触は重要である。また費用もほとんどかわりない。

4° 現場での作業(field work). 指示を受けた調査員は、必要があればまずサンプリングをおこない、つぎにそれらのサンプルを調査する。

5° 回収. 調査がすんだら、指示に従って調査員に調査票や必要書類を返させる。指示者が調査員といっしょに泊り込んでいるような場合には、毎日、調査がすんだ分を回収し、不明をただしたり、必要があれば再調査を命じることもある。しかし、そうでないときは、たんに誤字や読めない字をなおすことにとどめるべきで、調査票を回収する人の主観や調査員の記憶にたよって、調査結果をなおすのはよくない。たとえ答の記録にむじゅんしていることがあるとしても、中央で統一ある処理をするようにしたほうがよい。

なお、この段階で調査不能となったサンプルに対しては、その理由を明瞭にしておかなければならない。

6° 集計への準備. 最近では、集計は電子計算機でされることが多い。その手順を述べよう。まずサンプルの答をコード(code, 符号)にする。このことをコーディング(coding)という。ふつうは調査票を印刷する前に、あらかじめ予想される答を印刷し、そのそばに数字のコードを印刷しておく。これをプリ・コーディング(pre-coding)といい、これに対して調査がすんでからコードを定めることを、アフター・コーディング(after-coding)という——答が全く予想できない場合は、本格的調査にとりかからず、準備調査(pre-test)をおこなう——。この答のコードは調査票の最下段にまとめたり、質問文の欄外などのコード・ボックスに記入し、これを見ながら電子計算機用カードやテープに

パンチをする場合と、調査票からパンチ用のデータ・シート(data-sheet)用紙に転記して、パンチする場合がある。とくに大量のパンチを要するときは後者をつかう。

このような方法だと、パンチ・カードやテープを目で見てもなんのことかわからず、あとから試行錯誤によって、新しい分類などを行うことがこんなんである。そこで研究調査などでは、調査票から‘手集計カード’にコードを転記し、ときにはコード化する前の文章のままの答を記入しておいて、研究者が目で見、このカードをさばきながら、集計方針のアイデアをねる必要がある。そうしてコードをなおしたり、追加したりして、この‘手集計カード’からパンチする。なお、データを保存する際には、調査票はかさばるし、データ・シート用紙はやぶれるし、電子計算機用カード、テープ、ディスクなども破損しやすいので、この‘手集計カード’と両方保存しておくといふ。手集計カードを使って集計するときは、カードの四辺の場所を定めて、性別、年令別、職業別、地域別などに、マジック・インキをぬっておくといふ。なおカードの中心部は見にくい。また目で追いやすいように、コード・ボックスの形を変えておくといふ。だいたい1,000枚を越えれば、手集計はこんなんである。

7° 集計。 まず各項目ごとに、どういう答(もの)がいくつずつあったかをしらべることを、単純集計という。そうして2つ(以上)の項目の関係をみるために、こちらでこう答えたものが、あちらでどうなっているか、という関係がわかるような集計をすることを、クロス集計(相関集計, cross-tabulation)という。クロス集計はどの項目間でおこなうかをあらかじめ計画しておいて、順序よく進めてゆかなければならない。

ふつうはフェース・シートといわれる(cf. p. 27)、性別、年令別、学歴別、職業別、階層別、地域別(大都市とか町村の別)、あるいは支持政党別などを基本項目とよび、すべての質問とクロス集計をする。電子計算機を使うとクロス集計はなんでもないが、あまり多すぎると、とても見られない。

以上が調査実施の各段階での、一般的な仕事および注意であるが、さら