

よみ書き調査に於ける標本調査計画の概要

林 知己夫

此の研究は統計数理研究所の林知己夫、丸山文行、石田正次、堀正臣、高倉節子、田熊雅子、鈴木三千代及び白石一誠の一身同体の様な有機的協同研究になるものである。

なおこの仕事は日本の国語をどうあらためたらよいかと言うことを考究する爲の一つの基本資料をうるために設けられた「よみ書き能力調査委員会」の中でなされたものである。

この委員会の運営と研究とは言語、社会、心理、教育、統計の各委員のチーム・ワークによつて行われている。

以下のべる概要の内容は統計方面の委員の企画した部分の一部であることを諒承せられたい。

さて、「よみ書き能力」とは何であるか。これは種々の問題のあるところであるが、委員会は「社会の一員の正常な生活に最少限度に必要な型及び度合の文字言語を使用する能力」とであると定義した。

そして其等の文字言語の媒介物は(1)新聞；(2)大衆的な雑誌、(3)政府と国民とは直接関係する文書、(4)一般国民の職業及び経済行爲に必要な基本的文書、(5)私信等であると考えられる。

したがつてこの委員会の主として言語、心理、教育の委員は上記の媒介物の主要なものから「よみ書き能力」を調査するに必要と思われる調査事を作りあげ多くの調査の結果適當と思われる問題を作成した。

この問題は75題ある。各問に正解ならば1, しかからざれば0をあてえ能力を数量的に表現する事を企図したのみならず、各問題固有の出来、不出来をみることによつて能力をダイナミックに把握する様につとめた。さてこの様な問題を被調査者に対し同一条件の下に解答せしめる、——知能検査のテスト方法と同様に教示をあてえてそれに従つて解答させる——のでなければ全般的に調査の結果をまとめることはできない。したがつてこれを実行するにあつては被調査者を一箇所に集合させそこで一定の教示にしたがつて、一斉に調査を行わねばならない。

喜いおくれたが、この調査の対象は通常の社会生活を営める数之年15才から64才迄の日本人である。

この様な対象は全国で約45,000,000人いるのであるから此等の人々を全部上の様な形式で調査することは不可能であり、且問題の性質上意味もろい事であると考えられる。

又簡単な調査でないため調査員は一定の素養をもち且一定の訓練を経なければならず、その数は自ら制限せられるであらう。(我々にこの様な調査員として調査地点のよく訓練された小学校の教員を極めて適当なものと考えた。) 加うるに、調査のための日時は少くなければならず費用も制限せられている。

此等の事を考へる時我々は調査の対象の一部を抽出調査を行わねばならない。

それならば、如何なる標本調査を行ふべきであるかと言うと科学的に信頼度の保証し得るところの Random Sampling が最もよいと考えられる。

我々の場合層化任意標本抽出法を採用することにした。かくして調査実施にうつり、標本理論上の精度、費用、日時、調査方法(被調査者を一箇所に集めて調査する以上調査地点を制限しなければならず当然、Sub—Sampling 方法を用いなければならぬ)、その他の諸種の事情(層化の事情ならびに細い分析結果を得たいと言ふ事情等)を考

之合せ実際の被調査の数を21008、調査地数を270と決定した。

層化から如上の被調査の決定迄以下順を追うてのべてみよう。

まずこの調査の様な全く新しい大規模の調査をなすに当つては、如何なる方法でなすべきか、又なし得られるであろうか。

又実際の制限の中で最適な方法は如何であろうかを調査する為には、或る規格を以て準備調査を行はねばならない。

この様な準備調査の結果の分析によって諸種の困難を発見し、それに対する処置を考へ全国調査の為の基礎的資料を得る必要がある。

我々の場合都市としては小田原市、郡部としては埼玉縣比企郡野木村、千葉縣君津郡長浦村を選定して調査を行った。そしてその結果を分析し、全国調査を行う為の基礎的体験を得た。その一部を次にのべてみよう。

1 調査を行う時にあつて

被調査者に対して、当地へ移住した年、年齢、性格、産業、職業、その職についた年、その前の産業職業、家の産業、職業、学歴、新聞のよむよまぬ、ラヂオのきく、きかぬ、調査場到着時刻を併せ調査し調査結果を豊富にする様にした。

2 被調査の抽出に関して

中央の計画に於て調査地点を抽出し被調査者数を決定し各地点で調査員が被調査たる個人を抽出することにした。この被調査抽出に用うる名簿としてはよく訂正されている物資配給台帳がよいことが反しかめられた。

小田原市に於ては台帳の不備(転出、住居不明等の比率)は2.4%程度、村に於ては0%であつた。又この台帳から所定の被調査を得るにはランカム・スタートのミスティマティック・サムフオリングが実際上好適な結果をあたえることが確かめられた。

なお中央では各地点のスタート番号、抽出間隔までも決定することに

した。これは抽出間隔の不用意の決定によつて生ずる抽出標本のシステマティックな偏倚を避けるのみならず、各地兵の調査員が誤りなく寧ろ厳密に抽出出来る様にする為である。

物資配給台帳は、各人の特定の配列を示しているのであるから抽出間隔を不用意に定めるのは危険がある。準備調査の実情、理論的考察から抽出間隔を相当の大きさ（50以上）にすることゝ兼收にすることが安全でありのやましいと考えられた。

3 各地点に於ける調査時刻の決定

小田原市の結果からみると女子は午後の一時が出席がよいと思われたのでこの様な時間は必ず調査時として入れねばならない。又調査曜日も原則として都市に於ては出勤通学の為日曜がよいと思われた。

又調査時刻に対する集合状況分析の結果定刻より30分前にはほとんど集合すべきものが集ることが判明したので調査開始時刻のそれに合わせる様心得ねばならない。

4. 被調査者の呼出し

被調査者を一箇所に集めて調査する以上被調査者の呼出しは最も重要な項目であり、これが大切なのは全く明らかなことである。

被調査の出席者と欠席者とが読み書き能力上差があれば調査は bias な結果をあたえ意味がうすくなってしまふ。

唯徒に数を揃えたからよいのでなく調査すべき項目に関して調査しようとする対象と調査したものとの質の異なるものがない様にすることが最も大切なことである。

小田原市の場合出席者は82%であり、のこりの欠席者18%は概ね止むを得ない欠席理由であつた。

この二つの集団が男女、年齢別構造上有意の差があるか否かをみるために χ^2 検定を行つたが有意な差はみられなかつた。しかしこれだけでは不十分であり、「読み書き能力」上の差は如何であるかをみなければな

らない。

小田原市では二日調査を実施し第一日目に止むを得ず欠席したものは二日目に出席する様にした。したがって二日目に出席したものは準欠席者とも言うべきものである。

この一日目、二日目出席のものの中で「読み書き能力」を示す正解答数の分布構造、平均解答数を比較してみ、その間に有意な差があるか否かをしらべた。その結果ともに有意な差はみられなかった。

以上の事を考へ合せ誠に止むを得ず欠席するもの以外出席せしめ、可能な限り出席率を増せばよい満足すべき結果を得る（欠席者と出席者とが略々等質と考へられ或る意味で調査結果に bias をあたえることが少い）ものであると言う予想を得た。ために出席をよくする事を強調し、役所との連絡、調査の傳達方法、其の他の努力によつて呼出しに成功する様努めることにした。

5. 標本調査計画（層化）に必要な資料

- (a) 市部、郡部でよみ書き能力上差のあることが了解せられた。
 - (b) 各産業別にみて差のあることがみられた。
 - (c) 新聞のよむ、よまぬによつて大いに差のあることがみられた。
- 次に此等の結果を利用して標本計画を立てることである。

その前にまづ問題になるのは調査対象を如何に明確に決定すべきかの問題である。大雑把に言えば全日本領土に居住する15才から64才までの日本人男女である。然し外海に点在する島嶼は如何にすべきであらうか。此等の島嶼は東松浦郡（五島列島等）、大島、八丈島、三宅島、佐渡島、隠岐、壹岐、対馬、大島、能毛島、である。

この様な箇所迄迄互つて調査することは本来必要なのであるが現在としては費用その他の点に於て極めて困難であり且つ又島嶼を調査する費用と労力とがあるならばこれを内地の調査にふりかける時、費用労力一定の下ではよりよい精度を示す結果が得られるであらうと考へられる。

この島の人口を考へる前に通常発表せられている人口数から第三国人

の人口を除かねばならない。

かくして、島の人口と全体の人口との比を求めるとき 0.673 % なる数字を得た。したがって島嶼の全体にあたる影響はまず小さいものと考えられる。如上の点を考へ合せ此等島嶼を一応対象外として除外した。これが先づ我々の対象の母体をなすものである。

以上によつて定められた全対象から抽出によつてある人数をランダムに抽出したいのであるがこれを実行するのは不可能である。

したがって我々の~~抽出~~^{所選} Sub-Sampling の方法を用いざるを得ない。この為個人を抽出する前に市町村をランダムに抽出するか又は統計局使用の調査区をランダムに抽出するかと云うことが考えられるが市町村乃至は調査区に含まれる人口数が様々であるから市区町村、調査単位で一様な確率を以て抽出したのでは各個人の抽出される確率が同一にならなくなる。この事は大規模の標本調査に於て困難な事である。

各人の抽出される確率を同一にするためには市町村の集団を層化し小さな集団としこれから人口に比例した確率を以て各市町村を抽出しなければならぬ。

この様に各市町村を層化して小集団を形成するに當つて我々の調査しようとするよみ書き能力に関してなるべく等質的なものを集めて層とすることが結果により精度をあたえるであろう。

かつ又、現在の調査の様に調査しようとする項目（但し同じ傾向をもつと考えられる）が多くさらにそれについての明確な予備知識の少いこと及び標本数決定の立場を考慮し層化を厳密に細密に行うことが得策であろうと考へられた。

今迄は市町村を単位にして話を進めて來たのであるが現在利用出来る基本的資料乃至は調査資料は区市郡についてのものが多い為、實際我々としては層化の第一次単位として区市郡を用いざるを得なかつた。

（註）層化の単位として町村を用いず郡を用いた結果どの位の^得失を得たかについて判定することは出来ないが次の様な利便はある。

(i) 町村の数は 12000 に及び其等を統計資料（数字）について

層化することは多くの労力を要するのみならず、数多きため数字の眞的分析を十分行い難く、ためにかえつて層化は形式的に於て不測の誤りを冒す可能性があるが、郡の場合は数少きため地図其他の調査資料と比較でき、よい層化がなされうる可能性がある。

(行) 郡中には於ては相当性格のことなる集団が存するが同一層に属する群の間では垂直なる構造をもちうる可能性がある。

層化に用いた資料は次の様なものであつた。

(Q) 多くの複雑微妙な条件から地域的に近い所はそれなりに近い性質をもつが遠い所は異つた性質をもつと考えられるから地域が広大にわたる時其等の間に種々の事に関して地域的な差のあることは当然考えられる所であり地域的氣候的又文化發展段階の点乃至は歴史上から言つても寒冷の北海道から温暖の九州にわたる間全体として極めて異質的なものがあると考えられる。

此等の点から言つてまず地域区分を層化の第一としてとることは妥当であると思う。特に言語の問題である以上この層化は有利であらう。その区分を如何にして決定するかは難かしい問題であり一応連続的に変化するものがある處で切ることは多少無理を伴うのは当然である。我々としては先人の多くの研究努力の結果一応決定せられている区分を主として尊重することにした。

北海道地方

東北地方

青森、岩手、宮城、秋田、山形、福島

関東地方

茨木、栃木、群馬、埼玉、千葉、東京、神奈川、新潟、静岡、長野

関西地方

富山、石川、福井、山梨、岐阜、愛知、三重、滋賀、京都、大阪、兵庫、奈良

中国四国地方

九州地方

(b) 市部・郡部別を用うる。

都部は複雑な要素、例えば学歴、産業、人口密度、交通等に於て差があり、その生活形態及び生活模様は大いに異なるところがあり、其等がよみ書き能力に差をあたえるであろうということは予期に難くはない。

更に準備調査の結果都部別がよみ書き能力を表現すると考えられる数字に関して差のあることが実証せられている。

(c) 市部に於ては人口数によつて層化をなした。これは人口がある意味で文化程度を表現しているものを考えたからである。原則として6大都市、20萬以上、10萬以上、10萬以下に分類した。

(d) 産業構成率を層化の基準とした。

産業構成率とはここで有業者人口に対する各産業人口の比を言うのである。一般的区分に用うる各区市郡又産業構成率としては見透しをよくする爲、コーリン・クラークの「経済的進歩の諸條件」に述べられている産業分類を基準として算出した。この分類は第一次産業、第二次産業、第三次産業でありその内容は次の通りである。

第一次産業は農業、林業、水産業を含むものであり我々はこれを今後唯に農業とよぶことにする。

第二次産業は鉱業、建設工業、製造工業、ガス、電気業からなるものであり、これを今後唯に工業と総称することにする。

第三次産業は商業、金融業、運輸・通信業、サービス業、自由業、公務団体、その他の産業を含むものであり唯に商業と仮称することにする。この三つの区分には経済学的には問題があると言はれるが各産業構成の比率が各地異について二次元的に表現せられるので見透しが極めてよく各地点の特色を第一次的に把握するのに好適なものである。

しかも、コーリン・クラークの説く所によつてみれば商業率の高い国は所謂文化的に発展した国が多く（工業率はそれに追隨している）農業率の高い国は未開拓国が多く、又各国は発展するに従つて農業率が減少し商業率が増大して來ているものの様である

更に商業は農業よりも国民所得が高く、それに伴う現象がおこっているものと考えられる。この意味から言つて上述の分類を層化に用いることはいみをもつ。又この分類区分がよみ書き能力の点に關して直接的に意味をもつ事は準備調査の結果示めれてをり、客観的妥当性を実證された。この区分により大きく層化した後、さらに産業の構造を考へ、水産業の率、鉱業の率、製造工業の率、商業の率の高い特異なものについて他と分離して別層となすことには。

(e) 次の層化の資料として昭和15年度のラゲオ普及率によつて示されている数字を基礎として用いた。この理由は次の通りである。

(1) 新聞の普及率がよみ書き能力の点に關して重大な影響をあたえていることが了解せられこれを層化に用うるのが極めて有利なのであるが新聞普及率に關する資料が市區郡について得られなかつた。新聞普及率によつて得られている資料は昭和22年度縣別のものである。したがつて新聞普及率と相関が高くしかも市區町村に付して得られる様な資料を探さねばならない。この資料として考えられるのはラゲオの普及率である。

縣別に見てラゲオの普及率(昭和15年度調査)と新聞の普及率との相関をとつてみるとこれは0.753であつた。(標本数46) 因みに昭和21年度もこの相関をとつてみると0.456となつた。この様な關係が縣別の時と郡市別の時と相似であると考えるならばラゲオの普及率(昭和15年)を用うるのが妥當であることが了解せられよう。昭和21年度のラゲオ普及率を用いない理由は上述の理由の外ラゲオの普及は戦時中に已むを得ぬ理由で膨脹せしめられてあるので歴史的不いみを含めての文化指示度を示していないと考えたからである。たしかに昭和15年度の普及率は自然的に發展した過去の文化の段階を一応あらわしているものと言へよう。よみ書き能力は現在の状況のみによつて決定せられるものでなく過去の諸種の要因に左右せられているものであると考える時このラゲオ普及率を層化に用うることは更に

理論的根拠をあたえるものであろう。

- (ii) 次に一種の文化的指標と考えられる図書の販売高との相関を
考えてみる。販売高としては昭和22年間の総合的数字を用
い、一人あたり販売高をもつて相関をとる対象とした。この
結果相関係数は0.639となつた。
- (iii) 次に人口密度との関係を考えてみる。人口密度は一般的に
言つて都会的度合を反映しているものと考えられる。これと
普及率との相関を求めてみると相関係数0.673を得た。
- (iv) 次に郡部に及ぼす都市の影響との相関を考えてみる。この影
響は都市の郡部に対する近接度、連繋度によつてあらわされ
これは一応

近接する市の人口

郡の人口 + 近接する市の人口

で数的に表現せられるものと考えよからう。郡の概識を
これによつてあたへ普及率との相関をとつてみた所、相関係
数として0.681を得た。(標本数68)

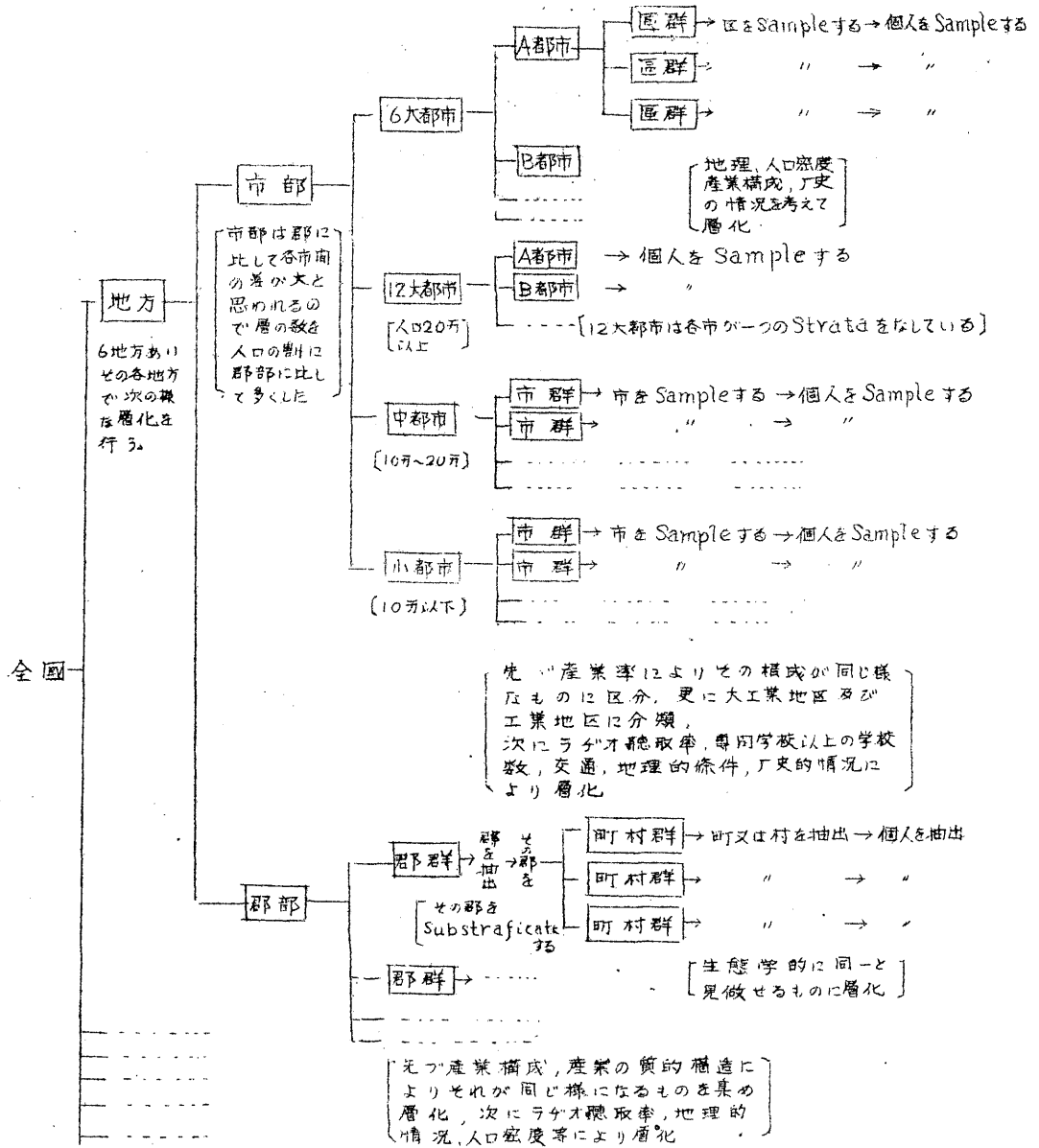
- (v) 次に交通との関係を考えてみる。5万分の1の地図とにらみ
合せ交通の良好さと普及率との関係の深いことをたしかめ得
た。
- (vi) 経済状態即ち所得との関係を考えた。このため昭和
13年度の一人当り所得と普及率との相関をとつてみた。
この結果、相関係数として0.834を得た。

以上考へ合せてみると昭和15年度のラゲオ普及率によつて表現
せられているものは複雑なものであり、これはよみ書き能力と相当関
係の深い指標であることが考えられ、これを層化に使うことは意味が
あると思う。

(f) 人口密度を用いた所がある。

これは(g)項の補足をする意味である。

(g) 学 校 数



(註) 1. □ は Strata を表す

2. 市部に於ては個人を抽出する前に然るべき比で町(又は町の半分)を抽出した所もある

都市の層化に専門学校以上の学校数を考慮に入れたところがある。

(Ⅳ) 地域的條件

地形的、氣候的、地域的に同一と思われるものはこれを層化に用いたところがある。又 Sub-Stratification にこれを用いたところは多い。

(Ⅴ) 其他生態学的條件

以上の他に数学的に表現するのはむづかしい質的差異を以つて区分したところもある。例えば産業の質的構造の差異即ち小工業地帯、大工業地帯の別等。

以上のべた資料を基にして層化を実行することになる。
層化するにあつて先づその個数を生態学的に^{向應すべきである。このため層化を正確に}満足すべき程度に行おうとする。(Sub-Sampling System)によつて標本調査を行うので between の variance, with の variance との釣合がとれるようにとも考へて層化を綿密に行ふ必要がある)と共に調査地点に標本数があまりにも小にならぬ様にする——あまりにも小になると一地点毎の推定が不確かになりすぎ調査結果が貧弱となる。相当数標本がゆきわたればその地点毎によみ書き能力を調査出来全国の推定と共に興味ある結果が得られるであらう。——及び調査の可能性等を考察して層化法を決定したつもりである。この結果層の数は市部で87, 郡部で76計163となつた。層化のための手段の一般図式を書いてみると前頁の表の様になる。

此等の立場によつて層化して後各層から一つの市區郡を人口に比例した確率を以つて抽出した。市、區の場合はこれから直接個人を抽出するのであるが郡の場合はさらに第二次的に層化して後町村を抽出することにした。抽出される郡についてこれを層化しその郡そのものをよく表現する様に標本をとることにした。層化の基準には5萬分の1地図、土地利用圖、産

業分布図，ラヂオ普及率，人口表を用いた。

こうして一郡中に2乃至3個の第二層がつくられこれから夫々人口に比例した確率を以つて町村を抽出した。

此等抽出された全国の市町村は次に示す通りである。

層化ならびに抽出地点表の一例（東北地方）

市 部

人口	商業率	ラヂオ普及率	工業率	その他	市	縣	Strata内人口及び%	割当セル数
20萬 ≤				高等以上の学校数 14	⊕ 仙台	宮	0.380 292784	87
20萬 > ≥10萬				同上 7 4	盛岡 ⊕ 秋田	岩 秋	0.289 222859	66
10萬 >	40% ≤	45% ≤		同上 4 3 6	弘前 山形 ⊕ 福島	青 山 福	0.323 248688	74
		15% >	40% ≤		鶴岡 ⊕ 酒田 若松 平	山 福 	0.237 182402	54
			40% >		青森 塩釜	青 宮	0.1772 132777	39
	40% >		40% ≤		⊕ 米沢 郡 釜石 能代	山 福 岩 秋	0.253 195162	58
			40% >		⊕ 石巻 八戸 宮古	宮 青 岩	0.227 174473	52
合 計							1.881 1449145	430

郡部

商業率	特殊 産業率	ラ子才 普及率	工業率	郡	縣	Strata内 人口及び%	抽出町村	割当サ ポル数
19% ≤		20% ≤		◎ 栄田 名古屋 志田 五造	富山	<u>0.406</u> 312898	船岡町 金ヶ瀬村	54 28
19% > ≥ 13%	水産業率 10% ≤			◎ 東津輕 下北 下野 下本	青森	<u>0.702</u> 540607	岩泉村 船越村	77 71
	鉱業率 10% ≤			◎ 鹿石	秋田	<u>0.445</u> 342420	川部村 泉村	43 46
その他	20% ≤	10% ≤	10% ≤	◎ 稗栗 川市 南東 西信 相馬	岩手	<u>1.228</u> 946062	上山町 中川村	51 796
		10% >	10% >	◎ 中津輕 南津輕 北津輕 遠田 登水 加美	青森	<u>0.818</u> 6307	黒石町 石川村 尾崎村	39 64 62
	20% >	14% ≤	14% >	◎ 上伊豆 北秋田 最上 西置賜 西伊達 岩手 耶麻 西台	岩手	<u>1.298</u> 999474	大館町 早口町 沢口村	83 59 120
		14% >	◎ 和賀 二戸 由利 仙北	岩手	<u>1.435</u>	本荘町 船川村 西郷町	102 90 96	

				平鹿 北村山 東置賜 北会津 河沼	秋山 川碓 川	1105197		
13% >		20% ≤		④ 茨城 西伊 伊黒 河那	岩宮 川秋宮 川	0.504 368023	一関市 日形村	45 57
13% >		20% >	10% ≤	④ 上北 岩手 雄勝 飽海 安運 南大 大雙	青森 山 川 川	1.140 877800	二本松町 上川崎村 石千村	35 95 96
			10% >	④ 西三 江 山 市 東 石 川	青 川 川 川 川 川 川 川	1.281 986666	西子町 紅久井村 川内村	108 75 75
				合 計		9.257 7129656		1565
				地域合計		11.138 3515801		2295

注 意

1. 人口は昭和22年10月1日現在の国勢調査による内地人人口（
三国人をのぞく）
2. イタリック 数字は全人口に対する各層の人口比率
3. 産業率は有業者比率をあらわし、昭和22年の国勢調査による
4. 人口密度は昭和22年の国勢調査による。

5. ラジオ普及率は昭和15年放送協会業務要報による。
6. 高専以上の学校数は昭和23年2月文部省の資料による。
7. * じるしは、層の条件に反するが、いろいろの理由によりその層に入れられたことを示す。
8. ② じるしは抽出された区市郡支廳を示す。抽出町村欄の町村は② じるしの郡市より抽出したものである。

次に考へるべきことは調査すべき標本数の決定である。

我々はまず対象たる日本人の各人のよみ書き能力が調査の結果点数によつて表現せられるものとする。この様にするとよみ書き能力の得点に関して対象たる日本人全体はあつたて構造をもつものと考えられる。この様な分布構造がまず如何なるものであるかと言ふことを調極する立場に立つて標本数を決定することにしよう。

今母集団の標識を A_1, \dots, A_k とし、比率(構造)が夫々 p_1, \dots, p_k ($\sum p_i = 1$) であるとする。

今この母集団から標本調査を行い (p_1, \dots, p_k) とする構造をある精度を以つて推定する問題を考へることになる。今

$$\frac{N-n}{N-1} \cdot \frac{p(1-p)}{n} \leq \varepsilon \quad p = \min(p_1, \dots, p_k)$$

N 母集団の総数

n 標本数

ε ある常数(推定の精度に依存する)

によつて標本数 n を決定するとき、この標本数があれば p_1, \dots, p_k の推定は一定の相対精度以上(標本調査によつて算出される変異係数がある一定数より小にすること)でなされることはただちに了解せられる。

我々としてはまず粗く標本数を決定するために母集団で5%を示す標識のものを信頼度99%、信頼幅0.5%で推定しうる様に標本数を決定す

ることとした。

この時 $n = 17100$ を得た。なおこの標本数があれば大体細かい分析も可能であり、他の量の推定にも十分であることが知られた。

しかし我々の場合は層化を実施しているため精度は上述のものよりはるかによくなっている事は確実な事である。

次に実際に標本抽出をなすとき Sub-Sampling を行っているの
でこのための影響を考えておこう。Sub-Sampling による精度即ち
標本調査によって生ずる variance は

$$\sigma^2 \leq \sum_{i=1}^R k_i^2 \sigma_{bi}^2 + \frac{\sigma^2}{n}$$

σ_{bi} は i 層の between の variance

k_i は i 層の総数の全体に対する比率

n は 標本数

R は 層の数

となるので実際の場合につき予想的にあらい評価を市部、郡部別に行っ
てみると

第一項は 0.0000030

第二項は 0.0000028

の Order をもつ事が言える。この様に考えるとき Sub-Samp-
ling による精度は信頼幅の意味で約 $\sqrt{2}$ 倍大に圧るものと考えられ
る。この様な評価は層化の効果も考へに入れぬのであり、相当控之目
なものであると言えよう。(なお実際の場合にはさらに郡を第二次層に分
ち町村を抽出しているのであるがこれにもとづく精度の誘下は一応等閑
に附しうるものと思われ、上述の $\sqrt{2}$ 倍の精度のおちると言う結論で
十分であろうと考えられる) 以上の様にして標本数及び精度が推定さ
れたのであるが次にその標本数を如何に配当するかの問題である。

我々としては複雑な分析を行うこと、構造決定のための計画であるこ

と、その他の多くの項目の調査であること、計算の簡便さから、層の総人口に比例して標本数を配当した。即ち所謂割当法を採用した。この方式は町村に至るまでおしよめることにした。

さいごに理論的に決定せられた標本数にある常数を乗じて実際の割当数とした。これは止むを得ず欠席する無害の缺席者のあることを見越し、準備調査の結果を参酌して都市では $\frac{1}{0.75}$ 郡部では $\frac{1}{0.85}$ となした。この結果算出された標本数は前に示した抽出市區町村名の標本数の欄に示されてある通りである。

以上の様にして標本調査の計算をたて次に各地点で個人を抽出するのに必要な数字 Random Start の番号、抽出間隔を指定した。

抽出間隔は $\frac{\text{地点人口} \times 0.585}{\text{標本数}}$ を以つてした。

0.585 は、対象内総人口の全人口に対する比である。

次にこの様な数字を用いて厳密に抽出を行い標本を決定するための指示書を作成した。これには Random という言葉を使用しなかつた。Random と云うと「勝手に、思うまゝ」という感じをいだき各地点の調査員が恣意的に行動する恐れがあり、かえつて全体に不測の系統的な偏倚をあたえる危険があつたからである。とにかく末端の調査員は全く個人採量の余地なく厳密に指示通りしたがる様にした。

〔補〕

以上が調査計画の概要であつて調査はこれにそつて昨年8月一斉に施行された。この時調査が確実に行われたか否かをみるために各地に視察員を派遣した。かくして調査の行われた後、調査方法の不確実であつた箇所、被調査者の不協力であつた箇所に対しては再調査を実施し調査のチェックを行つた。なおこの外本調査の問題がよみ書き能力調査に対して validity あるものであるか否かをみるために validity テストを行い、且全國民との比較のために小学校5年から高等学校3年に至るまで同様の規格で標本調査を行うことにした。