九州大学学術情報リポジトリ
Kyushu University Institutional Repository

よみ書き調査に於ける標本調査計畫の概要

林, 知己夫
統計數理研究所

https://hdl.handle.net/2324/12933

出版情報：統計数理研究. 3 (1/2), pp.1-10, 1949-12-20. Research Association of Statistical Sciences
バージョン：
権利関係：

論　　述

よみ書き調査に於ける標本調査計畫の概要

林　知　己　夫

統計數理研究所

(昭和24年4月30日受理)

此の研究は統計數理研究所の林知己夫，丸山文行，石田正次，堤正臣，高倉節子，田熊雅子，鈴木三千代及び白石一誠の一身同體の様な有機的協同研究になるものである。なほこの仕事は日本の國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會」の中でなされたものである。この委員會の運營と研究とは言語，社會，心理，教育，統計の各委員のティーム・ワークによつて行はれている。以下のべる概要の内容は統計方面の委員の企畫した部分の一部であることを諒承せられたい。

我々はこの「よみ書き能力調査」を計量社會學の建立のための一飛石と考え，この意圖の下に標本調査計畫をなしたのである。

(註)　計量社會學とは何であるか。

計量社會學は社會，人間現象に測度(メトリック)を入れ，そこにあらはれる諸要素ならびにその行動，態度，意欲心情を數量的に表現し，この現象のダイナミックな構造を實存的に思考し確率論的立場に立つて法則化しようとするものであると思惟する。

さて「よみ書き能力」とは何であるか．これは種々問題のあるところであるが委員會は「社會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」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) 大衆的な雜誌，(3) 政府と國民とに直接關係する文書，(4) 一般國民の職業及び經濟行爲に必要な基本的の文書，(5) 私信等であると考えられる。したがつてこの委員會の主として言語，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はれる調査事項を作りあげ多くの調査の結果適當と思は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，しからざれば0をあたえ能力を數量的に表現する事を企圖したのみならず，各問題固有の出來，不出來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對し同一條件の下に解答せしめる——知能檢査のテスト方法と同様に教示をあたえてそれに從つて解答させる——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。したがつてこれを實行するにあたつては被調査者を一個所に集合させそこで一定の教示にしたがつて一齊に調査を行はねばならない。言ひおくれたがこの調査の對象は通常の社會生活を營める數え年15歳から64歳迄の日本人である。この様な對象は全國で約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うすい事であると考えられる。

又簡單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經なければならずその數は自ら制限せられるであらう。(我々はこの様な調査員として調査地點のよく訓練された小學校の教員を極めて適當なものと考えた。)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。此等の事を考える時我々は調査の對象の一部を抽出調査を行はなければならない。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學的に信頼度の保證し得るところの 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。かくして調査實施にうつり，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點を制限しなければならず當然 Sub-Sampling 方法を用いなければならない），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）を考え合せ實際の被調査者の數を21008，調査地點を270と決定した。層化から如上の被調査者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當つては，如何なる方法でなすべきか，又なし得られるであろうか，又實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爲には，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つて諸種の困難を發見し，それに對する處置を考え全國調査の爲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本村，千葉縣君津郡長浦村を選定して調査を行つた。そしてその結果を分析し，全國調査を行う爲の基礎的經驗を得た。その一部を次にのべてみよう。

1. 調査を行う時にあたつて

被調査者に對して，當地へ移住した年，年齢，性格，產業，職業，その職についた年，その前の產業職業，家の產業，職業，學歷，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豐富にする様にした。

2. 被調査の抽出に關して

中央の計畫に於て調査地點を抽出し被調査者數を決定し各地點で調査員が被調査者たる個人を抽出することにした。この被調査者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臺帳がよいことがたしかめられた。小田原市に於ては臺帳の不備（轉出，住居不明等の比率）は2.4％程度，村に於ては0％であつた。又この臺帳から所定の被調査者をうるには，ランダム・スタートのンスティマティック・サムプリングが實際上好適な結果をあたえることが確かめられた。なほ中央では各地點のスタート番號，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ンステイマティックな偏倚を避けるのみならず各地點の調査員が誤りなく早く嚴密に抽出出來る様にする爲である。物資配給臺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險がある。準備調査の實狀，理論的考察から抽出間隔を相當の大きさ（50以上）にすること丶素數にすることが安全でありのぞましいと考えられた。

3. 各地點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はれる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。又調査曜日は原則として都市に於ては出勤通學の爲日曜がよいと思はれた。又調査時刻に對する集合狀況分析の結果定刻より30分すぎには殆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をそれに應ず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あつ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。被調査者の出席者と缺席者とが讀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つてしまう。唯徒に數を揃へたからよいのでなく調査すべき項目に關して調査しようとする對象と調査したものとが質のこと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，のこりの缺席者 18 % は概ね止むを得ない缺席理由であつた。この二つの集團が男女，年齢別構造上有意な差があるか否かをみるために χ^2 檢定をしたが有意な差はみられなかつた。しかしこれ丈では不十分であり，「よみ書き能力」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實施し第一日目に止むを得ず缺席したものは二日目に出席する樣にした。したがつて二日目に出席したものは準缺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間で「よみ書き能力」を示す正解答數の分布構造，平均解答數を比較してみ，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つた。以上の事を考え合せ誠に止むを得ず缺席するもの以外出席せしめ，可能なる限り出席率を增せばほゞ滿足すべき結果を得る（缺席者と出席者とが略々等質と考えられ或意味で調査結果に bias をあたえることが少い）ものであると言う豫想を得た。ために出席をよくする事を强調し，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樣努めることにした。

5. 標本調査計畫（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） 各產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畫をたてることである。

その前にまず問題になるのは調査對象を如何に明確に決定すべきかの問題である。大雜把に言えば全日本領土に居住する 15 歲から 64 歲までの日本人男女である。然し外海に點在する島嶼は如何にすべきであろうか。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，隱岐，壹岐，對馬，大島，能毛島である。この樣な箇所に迄亘つて調査することは本來必要なのであるが現在としては費用その他の點に於て極めて困難であり且又島嶼を調査する費用と勞力とがあるならばこれを內陸の調査にふりむける時，其の勞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ろうと考えられる。この島の人口を考える前に通常發表せられている人口數から第三國人の人口を除かねばならない。かくして島の人口と全體の人口との比を求めるとき 0.673 % なる數字を得た。したがつて島嶼の全體にあたえる影響はまず少いものと考えられる。如上の點を考え合せ此等島嶼を一應對象外として除外した。これが先ず我々の對象の母體をなすものである。

以上によつて定められた全對象から抽出によつてある人數丈をランダムに抽出したいのであるがこれを實行するのは不可能である。したがつて我々の場合所謂 Sub-Sampling の方法を用いざるを得ない。この爲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區をランダムに抽出するかと言うことが考えられるが市町村乃至は調査區に含まれる人口數が樣々であるから市區町村，調査單位で一樣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ことである。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團を層化し小さな集團とな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ない。この際に各市町村を層化して小集團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對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い精度をあたえるであろう。かつ又現在の調査の際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知識が少いこと及び標本數決定の立場を考慮し層化を嚴密に細密に行うことが得策であろうと考えられた。

今迄は市町村を單位にして話を進めて來たのであるが現在利用出來る基本的資料乃至は調査資料は區市郡についてのものが多い爲，實際我々としては層化の第一次單位として區市郡を用いざるを得なかつた。

(註)　層化の單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來ないが次の様な利點はある。

(i)　町村の數は12000に及び其等を統計資料（數字）について層化することは多くの勞力を要するのみならず，數多きため數字の質的分析を十分行い難く，ためにかえつて層化は形式的になかれ不測の誤りを冒す可能性があるが，郡の場合は數少きため地圖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かある。

(ii)　郡中に於ては相當性格のことなる集團か存するが同一層に屬する郡の間では垂直な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。

(a)　多くの複雜微妙な條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なつた性質をもつと考えられるから地域が廣大にわたる時其等の間に種々の事に關して地域的な差のあることは當然考えられる所であり地理的氣候的又文化發展段階の點乃至は歷史上から言つても寒冷の北海道から溫暖の九州にわたる間全體として極めて異質的なものがあると考えられる。此等の點から言つてまず地域區分を層化の第一としてとることは妥當であると思う。特に言語の問題である以上この層化は有利であろう。その區分を如何にして決定するかは難しい問題であり一應連續的に變化するものをある點で切ることは多少無理を伴うのは當然である。我々としては先人の多くの研究努力の結果一應決定せられている區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

關東地方

茨木，栃木，郡馬，埼玉，千葉，東京，神奈川，新潟，靜岡，長野

關西地方

富山，石川，福井，山梨，岐阜，愛知，三重，滋賀，京都，大阪，兵庫，奈良

中國四國地方

九州地方

(b)　市部，郡部別を用うる。

都部は複雜な要素，例えば學歷，產業，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るところがあり，其等かよみ書き能力に差をあたえるであろうと言うことは豫測に難くはない。更に準備調査の結果都鄙別がよみ書き能力を表現すると考えられる數字に關して差のあることが實證せられている。

(c)　市部に於ては人口數によつて層化をなした。これは人口があるいみで文化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)　產業構成率を層化の基準とした。

產業構成率とはここで有業者人口に對する各產業人口の比を言うのである。一般的區分に用うる各區市郡又產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「經濟的進步の諸條件」に述べられてある產業分類を基準として算出した。この分類は第一次產業，第二次產業，第三次產業でありその內容は次の通りである。第一次產業は農業，林業，水產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。第二次產業は鑛業，建設工業，製造工業，ガス・電氣業からなるものであり，これを今後唯單に工業と總稱することにする。第三次產業は商業，金融業，運輸・通信業，サービス業，自由業，公務團體，その他の產業を含むものであり唯單に商業と假稱することにする。この三つの區分には經濟學的には問題かあると言はれるが各產業構成の比率が各地點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

各地點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國は所謂文化的に發展した國が多く（工業率はそれに追隨している）農業率の高い國は未開な國が多く，又各國は發展するに從つて農業率が減少し商業率が增大して來ているものの樣である。更に商業は農業よりも國民所得が高く，それにともなう現象がおこつているものと考えられる。このいみから言つて上述の分類を層化に用うる事はいみをもつ。又この分類區分がよみ書き能力の點に關して直接的に意味をもつ事は準備調査の結果示されてをり，客觀的妥當性を實證された。この區分により大きく層化した後，更に產業の質的構造を考え，水產業の率，鑛業の率，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した。

(d)　次の層化の資料として昭和 15 年度のラヂオ普及率によつて示されてゐる數字を基礎として用いた。この理由は次の通りである。

(i)　新聞の普及率がよみ書き能力の點に關して重大な影響をあたえている事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關が高くしかも市區町村に付して得られる樣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ヂオの普及率である。縣別に見てラヂオの普及率（昭和15年度調査）と新聞の普及率との相關をとつてみるとこれは 0.753 であつた。(標本數 46) 因みに昭和21年度もこの相關をとつてみると 0.456 となつた。この樣な關係が縣別の時と郡市別の時と相似てあると考えるならばラヂオの普及率（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 年度のラヂオ普及率を用いない理由は上述の理由の外ラヂオの普及は戰時中に已むを得ぬ理由で膨脹せしめられてあるのは史的な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應あらはしているものと言えよう。よみ書き能力は現在の狀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ヂオ普及率を層化に用うることは更に理論的根據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圖書の販賣高との相關を考えてみる。販賣高としては昭和 22 年間の綜合的數字を用い一人あたり販賣高をもつて相關をとる對象とした。この結果相關係數は 0.639 となつた。

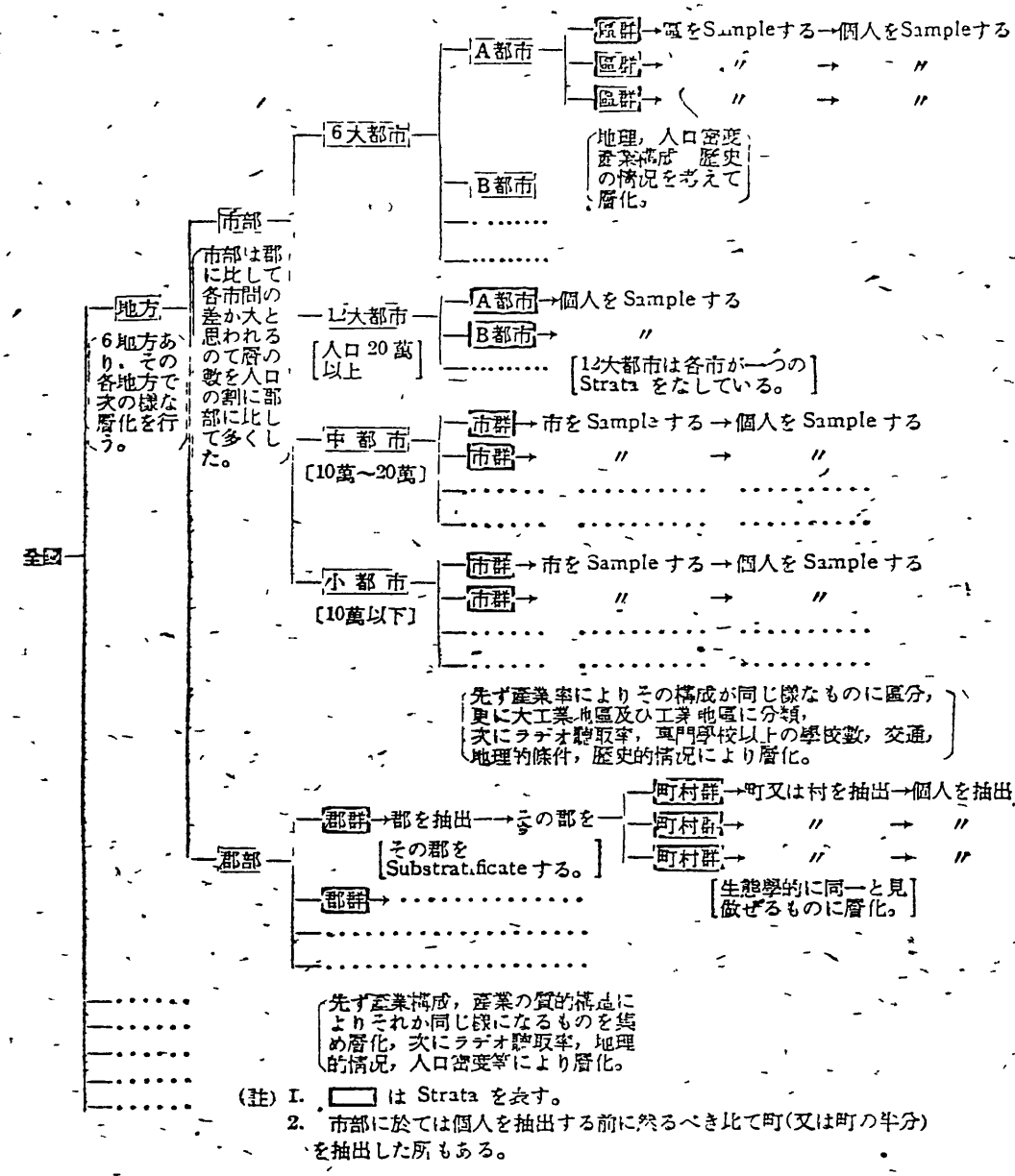
(iii)　次に人口密度との關係を考えてみる。人口密度は一般的に言つて都會的度合を反映しているものと考えられる。これと普及率との相關をもとめてみると相關係數 0.673 を得た。

(iv)　次に郡部に及ぼす都市の影響との相關を考えてみる。この影響度は都市の郡部に對する近接度，連繫度によつてあらわされこれは一應 $\frac{\text{近接する市の人口}}{\text{郡の人口}+\text{近接する市の人口}}$ で數字的に表現せられるものと考えてよかろう。郡の標識をこれによつてあたえ普及率との相關をとつてみた所，相關係數として 0.681 を得た。(標本數 63)

(v)　次に交通との關係を考えてみる。5 萬分の 1 の地圖とにらみあわせ交通の良好さと普及率との關係の深いことをたしかめ得た。

(vi)　經濟狀態即ち所得との關係を考えてみた。このため昭和 13 年度の一人當り所得と普及率との相關をとつてみた。この結果，相關關係として 0.834 を得た。

以上考え合せてみるとき昭和 15 年度のラヂオ普及率によつて表現せられているものは複雜なものであり，これはよみ書き能力と相當關係の



(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。

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。

(e)　人口密度を用いた所がある。

これは(c)項の補足をするいみてある。

(f)　學校數

都市の層化に專門學校以上の學校數を考慮に入れたところがある。

(g)　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かある。又Sub-Stratification にこれを用いたところは多い。

(h)　其他生態學的條件

以上の他に數學的に表現するのにむつかしい質的差異を以つて區分した所もある。例えば産業の質的構造の差異卽ち小工業地帶，大工業地帶の別等。

以上のべた資料を基にして層化を實行することになる。層化するにあたつて先ずその個數を問題にすべきである。このため層化を純粋に生態學的に言つて滿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 との釣合がとれる樣にとも考えて層化を綿密に行う必要がある）と共に抽出すべき標本數を各調査地點に配當する時に調査地點に標本數があまりにも小にならぬ樣にすること——あまりにも小になると一地點毎の推定が不確かになりすぎ調査結果が貧弱となる。相當數標本がゆきわたればその地點毎によみ書き能力を調査出來全國の推定と共に興味ある結果が得られるであろう——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。この結果層の數は市部で87，郡部で76計163となつた。層化のための手段の一覧圖式を書いてみると前頁の表の樣になる。

within

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。抽出された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樣に標本をとることにした。層化の基準には5萬分の1地圖，土地利用圖，産業分布圖，ラヂオ普及率，人口表を用いた。こうして一郡中に2乃至は3個の第二次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。此等抽出された全國の市町村は次に示す通りである。

層化ならびに抽出地點表の一例（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 內人口及び%	割當サムプル數
20萬≦				高專以上の學校數 14	◉仙臺	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	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	青	0.772 132777	39
	40%＞		40%≦		●米澤 郡山 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

都部

商業率	特殊産業率	ラヂオ普及率	工業率			Strata 内人口及び%	抽出町村	割當サムプル数
19%≦		20%≦		●栃	宮	0.406	船岡町	54
				名	〃	312393	金ケ瀬村	28
				宮	〃			
				志	〃			
				玉	〃			
19%＞≧13%	水産業率 10%≦			東	青	0.702		
				津	〃			
				軽	岩	540607	岩泉村	71
				北	〃		船越村	71
				下	宮			
				氣	〃			
				●下				
				閉				
				伊				
				牡				
				鹿				
				吉				
				本				
	鑛業率 10%≦			鹿	秋	0.445		
				角		342420	川部村	43
				●石	福		泉村	46
				城				
	その他	20%≦	10%≦	稗	岩	1.228		
				貫	宮	946062		
				刈	〃			
				栗	秋		上山町	51
				田	〃		中川村	196
				南	秋			
				原	〃			
				●南	〃			
				秋	山			
				東	〃			
				村	〃			
				西	〃			
				村	〃			
				信	福			
				夫	〃			
				相	〃			
				馬	〃			
			10%＞	中	青	0.818	黒石町	39
				津	〃	630409	石川村	64
				軽	〃		尾崎村	62
				●南	〃			
				津	宮			
				北	〃			
				津	〃			
				遠	〃			
				軽	〃			
				登	〃			
				田	〃			
				豆	〃			
				米	〃			
				加	〃			
				理	〃			
				美	〃			
		20%＞	14%≦	上	岩	1.298	大館町	83
				閉	秋	999474	早口町	59
				●北	山		澤口村	120
				秋	〃			
				最	〃			
				西	福			
				西	〃			
				伊	〃			
				田	〃			
				岩	〃			
				上	〃			
				耶	〃			
				賀	岩	1.435		
				澤	〃	1105197		
				西	〃			
				白	〃			
				川	〃			
			14%＞	和	岩		本荘町	
				賀	〃		鮎川村	102
				澤	〃		西滝澤村	90
				戸	秋			96
				利	〃			
				●由	〃			
				仙	〃			
				北	山			
				鹿	〃			
				山	福			
				賜	〃			
				津	〃			
				沼	〃			
				平	〃			
				村	〃			
				置	〃			
				賀	〃			
				北	〃			
				河	〃			
13%＞		20%≦		柴	岩	0.504	一關市	45
				海	〃	388023	日形村	57
				●西	宮			
				磐	〃			
				井	秋			
				伊	宮			
				具				
				川				
				奥				
				河				
				邊				
				桃				
				生				

13%>		20%>	10%≦	上岩雄飽安安南大鰺	北手勝海蓬積津沼葉會	青岩秋山福〃〃〃〃	1 140 877800	三本松町 上川崎村 石井村	35 95 96
			10%>	西三江西山南東東石田丸	津磐盛田白輕戸刺井本賜川川川村戸	青〃岩〃秋山〃福〃〃岩	1.281 986666	田子町 名久井村 川内村	108 75 75
				合計			9.25 7129656		1865
				地域合計			11.138 3518801		2295

注意

1. 人口は昭和 22 年 10 月 1 日現在の國勢調査による內地人人口。(三國人をのぞく)
2. イタリック數字は全人口に對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 年の國勢調査による。
4. 人口密度は昭和 22 年の國勢調査による。
5. ラジオ普及率は昭和 15 年放送協會業務要報による。
6. 高專以上の學校數は昭和 23 年 2 月文部省の資料による。
7. その他の資料については「334 層化法」參照。
8. * じるしは，層の條件に反するが，いろいろの理由によりその層に入れたことを示す。
9. ●じるしは抽出された區市郡支廳を示す。抽出町村欄の町村は●じるしの郡市より抽出したものである。

次に考えるべきことは調査すべき標本數の決定である。

我々はまず對象たる日本人の各人のよみ書き能力が調査の結果點數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點に關して對象たる日本人全體はある分布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うことを調査する立場に立つて標本數を決定することにしよう。今母集團の標識を $A_1 \quad A_l$ とし，比率(構造)が夫々 $p_1 \cdots p_l$ $(\sum p_i=1)$ であるとする。

今この母集團から標本調査を行い $(p_1 \cdots p_l)$ と言う構造をある精度を以つて推定する問題を考えることになる。今

$$\frac{N-n}{N-1} \cdot \frac{p(1-p)}{n} \leqq \varepsilon \qquad p=\min(p_1 \cdots p_l)$$

N 母集團の總數
n 標本數
ε ある常數〔推定の精度に依存する〕

によつて標本數 n を決定するとするとき，この標本數があれば $p_1 \cdots p_l$ の推定は一定の信頼精度以上（標本調査によつて算出される變異係數をある一定數より小にすること）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數を決定するために母集團で 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數を決定することにした。

この時 $n=17100$ を得た。なおこの標本數があれば大體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實施しているため精度は上述のものよりはるかによくなつている事は確實な事である。次に實際に標本抽出をなすときSub-Samplingを行つ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つて生ずる variance は $\sigma^2 \leqq \sum_{i=1}^{R} k_i^2 \sigma_{bi}^2 + \frac{\sigma^2}{n}$

σ_{bi} は i 層の berween の variance

k_i は i 層の總數の全體に對する比率

n は 標本數

R は 層の數

となるので實際の場合につき豫想的にあらい評價を市部，郡部別に行つてみると

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-Sampling による精度は信頼幅の意味で約 $\sqrt{2}$ 倍大になるものと考えられる。この様な評價は層化の效果も考えに入れぬのであり，相當控え目なものであると言えよう。(なお實際の場合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落下は一應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數及び精度が推定されたのであるが次にその標本數を如何に配當するかの問題である。我々としては複雜な分析を行うこと，構造決定のための計畫であること，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，層の總人口に比例して標本數を配當した。即ち所謂割當法を採用した。この方式は町村に至るまでもおしすすめることにした。

さいごに理論的に決定せられた標本數にある常數を乘じて實際の割當數となした。これは止むを得ず缺席する無害の缺席者のあることを見越し，準備調査の結果を參酌して都市では $\frac{1}{0.75}$ 郡部では $\frac{1}{0.65}$ となした。この結果算出された標本數は前に示した抽出市區町村名の標本數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點で個人を抽出するのに必要な數字Random Start の番號，抽出間隔を指定した。抽出間隔は $\frac{\text{地點人口}\times 0.585}{\text{標本數}}$ を以つてした。0.585 は對象內總人口の全人口に對する比である。

次にこの様な數字を用いて嚴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 と言う言葉を使用しなかつた。Random と言うと「勝手に，恣ふまゝ」と言う感じをいだき各地點の調査員が恣意的に行動するおそれがあり，かえつて全體に不測の系統的な偏倚をあたえる危惧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餘地なく嚴密に指示通りしたがう様にした。

〔補〕

以上が調査計畫の概要であつて調査はこれにそつて昨年8月一齊に施行された。この時調査が確實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，調査方法の不確實であつた箇所，被調査者の不協力であつた箇所に對しては再調査を實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對して 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學校5年から高等學校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。

此等の細部に關してはやがて發表の機會があるであろうと思う。既にその一部は雜誌「教育」1949 年 2 月號柴田武「讀み書き能力調査について」に發表せられてある。