§1 調 査 の 計 画

§10 問題の設定

この調査は特につぎの2種類の問題を明らかにする．ひとつは読み書き能力に影響を及ぼす要因，もうひとつは読み書き能力を調べる手がかりである．

§100 読み書き能力に影響を及ぼす要因

読み書き能力を身につけ，失わないためにはつぎの三つの要因が大きな影響をあたえていると仮定される．

1　言語的要因

2　教育的要因

3　文化的要因

この調査の計画は，これらの要因の影響を測定できるように立てなければならない．そうでなければ，その結果は，勝手に解釈されないともかぎらないし，あとにつづくどういう具体的な計画に対してもあまり役立つものとなるとは思われないからである．

1　言語的要因．総合的な読み書き能力[1]はつぎの四つの言語的要素を使うことのできる能力に関係していると見ることができる．

i　数字（アラビア数字，漢数字）

ii　文字（ひらがな，カタカナ，漢字）

iii　語

iv　センテンス・パラグラフ（文・節）[2]

1）「総合的な読み書き能力」はこの本でいう「読み書き能力」とおなじものである．ただ，数字の読み書き能力などというような，ひとつひとつにわけられた能力とまぎれやすいときにだけ使われる．

2）センテンス・パラグラフについては §311.7 を見よ．

2　教育的要因．総合的な読み書き能力および1にあげた，それぞれの言語能力は，うけた教育レベルという要因に関連する．この関係がどういうものであるかを調べなけ

ればならない.

3　文化的要因. 総合的な読み書き能力およびいくつかの言語能力は，教育的要因以外の社会的経験にも関係する. これらの要因のうちおもなものは，

i　居住地

ii　性

iii　年齢

iv　産業・職業

v　新聞を読むかどうか，ラジオを聞くかどうか

などである. これらの関係を調べなければならない.

§101　読み書き能力を調べる手がかり

総合的な読み書き能力は，多くの別別の読み書きの能力を参考にしてきめられる. 総合的な読み書き能力は，こんどの場合，つぎのことがらがどの程度できるかによってきめるのが適当である.

1　数字（アラビヤ数字と漢数字）の読み[1]および書取り[1].

2　かな（ひらがなとカタカナ）と漢字との読みおよび書取り（変体がなを入れない).

3　語と漢字との意味の読みおよび書取り.

4　センテンス・パラグラフの理解（読みだけ).

1)「読み」とは，字の形を見て音(おん)を思い出す，いわゆる“読み”に，意味の理解がふくまれた，認知ともいうべきもの.「書取り」とは，音(おん)を聞いて形を再現する，いわゆる“書取り”に，意味の理解がふくまれた，想起ともいうべきもの. 以下，この本で「読み」,「書取り」というときはこの意味である. したがって,「読み書き能力」というときの「読み」,「書き」もこの意味で使われる. この本で“読み”,“書取り”と書くときは，意味の理解をふくまない，今までの意味である.

数字の読みおよび書取りの調査は，最初考えられていなかったが，日常の読み書きでもっとも必要とされるもののひとつであり，能力としても最低のものとして，どうしても調べなければならないと考え，計画のうちにつけたしたのである. アラビア数字と漢数字との両方について調べる必要がある.

かなを使う能力も特別に調べなければならない. この基本的な文字を使う能力があるかどうかは，いうまでもなく，調査ぜんぱんの結果に反映する. この能力を測定しなければ，結果ぜんたいで，あやまりのもとがどこにあるかを見ることのできない場合がある. ひらがなとカタカナとの両方について調べるべきである.

語と漢字との意味はふつうに使われている狀況で調べなければならない. このためには，いくつかの語あるいは漢字のうちから刺激語に対して，シノニム(同意語)，アン

トニム(反意語)，あるいはおなじカテゴリイ(部類)の語あるいは漢字を選ばせる方法がある．

語はスタイル[1]という面からつぎのように分類される．

i 専門語および特殊な語

ii ふつうの文章語

iii 口頭語および俗語

しかし，はっきりした分類は，その語の使われている資料の綿密な研究によらなければならない．

1) この本でいう「スタイル」(style) はある作家のスタイルなどというときの，いわゆる“文体”と違い，語あるいはセンテンス・パラグラフについて，そのあらわれる場面および位置からくる性質を意味する．

言語テストでいう，意味と理解とは，言語能力の働きであると同時に，一般知能の働きでもある．じっさい，テストには，言語を理解する能力を調べて，その点数で一般知能を測定するものが多い．これらふたつの要因のあいだに直接関係があるかぎりは，テストは不当に一方へかたよることなく，一般知能によって言語を理解する能力を測定することが確かにできる．

調査者の経験にもとづいて，テスト問題に使われる概念の型およびその複雑さのレベルを，被調査者にたやすく理解できるようなものにするということが，誤差を防ぐ方法として，こういうテストにふつうとり入れられている．したがって，被調査者の年齢，教育および社会的カテゴリイで，その人にふつう予期されているよりは，テスト問題のレベルを少しひくめるように心がけることが必要である．しかし，特殊な型の文字言語を読む能力に関係しては，literacy の定義を考えれば，言語テストのこの一般的要件はある程度制限される．したがって，テストの問題は，使われている資料のなかの概念の型とレベルとを処理する能力を調べることになる．しかし，literacy をどう定義したところで，それには言語上およびテストをうけるうえのある程度の知識が予想されているのである．そうして，ここでとられた定義は，literacy のいろいろな働きという点でなっとくの行くものであると考えられる．そのうえ，使われる資料の概念の型およびその複雑さの範囲は比較的広いのである．したがって，テスト問題には，すべてをふくませることはできないから，これらの概念の型およびその複雑さのおもなものを反映させなければならない．どんな場合でも，テスト問題に使われる言語とそのテスト問題があらわす内容とのレベルは近いものでなければならない，という一般方針が採用されるべきである．この方針が採用されてはじめて，簡単な観念が複雑な言語であらわれていたり，あるいはその逆の関係にあるものが，テスト問題資料のなかで，上に述べたものから区

別され，テストの問題とすることが不適当とされるのである．しかし，もし，資料となる語およびセンテンス・パラグラフを分析してみて，簡単な観念が複雑な言語であらわされることがふつうであるということがわかれば，こうした場合を反映するいくつかの別別の問題を出して，言語があまりにも複雑であるために，その資料の内容を伝えることができなかった度合をはっきり示さなければならない．

意味と理解とは，具体的な方面でも，抽象的な方面でも測定することができる．どちらの方面でも，語とセンテンスとの意味，ならびに理解を調べるには，つぎのような型の読みの反応を必要とする．すなわち，刺激語に対して，シノニム，アントニム，およびおなじカテゴリイの語を一群の語のなかから選ばせ，または問答形式で答えさせて測定する方法がある．パラグラフの意味の理解は，どちらの方面でも，1 語で答えさせるか，または問答形式で答えさせるか，あるいはフレーズまたはセンテンスで文章を構成させるかによって測定される．語および漢字の意味の理解を調べるには，問題の穴(ブランク）をうめるという方法を使うといい．

上に述べたどの場合であっても，すぐ答が点数になるように，被調査者に，いくつかの答のうちからそのひとつを選ばせるような問題を出さなければならない．自由に答を書かせる問題はつぎのような場合にだけ許される．すなわち，

i　文字あるいは語の書取りを調べる“穴うめ式テスト”で，入れるべき語が，ほかにそれと並べて選ぶべき相手がない場合．

ii　数学に関する問題の場合．

ごく広い範囲にわたって書取りを必要とするような問題，たとえば被調査者じしんの言語で述べたり総合したりするようなものはさけなければならない．そのような問題に対する答は，被調査者の過去の興味と経験とを反映するものであり，今のところ，その答を点数にしてくらべることのできるような便宜な方法はない．

§11　調査の設計

§110　Mass Communication Media の準備

テスト資料は，§003.1 にあげた 4 種類の mass communication media から，つぎのことを基礎にして選ばなければならない．

1　それを受取る者およびそれを読む者

2　その内容のスタイルおよび事項

新聞については，つぎのことをよく代表するように選ばなければならない．i できるだけ広い地域，ii 男女両性，iii すべての年齢，iv すべてのおもな職業．

届・通達については，政府の職員および代行者と市民（civilian）ぜんたいとのあいだの情報および行動によって，ふだん必要とされるものを選ばなければならない．市民

の特殊な集団にだけ関する文書や，政府内部にだけとりかわされる文書のようなものはのぞくことが必要である．

ビラの類は，それぞれ特別の集団に関係するものではあるが，全体として，おもな職業のすべて，および経済的企業の全領域にわたるように選ばなければならない．たとえば，商業通信に関する文書および広告，または労働組合の基礎的な書式や文献．

個人的な手紙は，専門的あるいは文学的なものよりは，むしろ概して消息を主とするもので，男女を問わず日本のいくつかのおもな地域の成年者によって書かれたものを選ぶべきである．

mass communication media は，いったん選ばれたら，こんどは内容によって分類される．その第一歩は言語学者およびこのような文書を読むことにかけて深い経験を持っている専門家の意見にもとづいて，内容をいっぱんにスタイルによって分類することである．ここで使うべきスタイルのおもな基準はつぎのように分類されるであろう．

1　専門語および特殊な語

2　ふつうの文章語

3　口頭語および俗語

いろいろなスタイルの要素の割合を特別に，かつ注意深く分析するよりも，むしろ記事でのスタイルぜんたいの調子をはかることに努力を払わなければならない．この方法は印象的であるが，もし最後の決定が，いくにんかの専門家の協力によって得られるならば，主観的とはいっても，ただ最小限度に主観的であるにとどまるであろう．

mass communication media をスタイルによって分類したが，そのおもな部類のなかで，記事および文書を事項のいろいろなカテゴリイに分析しなければならない．語集[1]および，もっと小さい範囲でではあるが，スタイルもまた，事項によって違うから，最後に整理したテスト資料のなかに，これらの違いをよく代表することをあやまらないようにするために，この分類をするのである．事項による分類は，すべての記事にわたって語集の多くのものがおなじであり，また，しばしばいくつかの違うことがらがおなじ文書（たとえば新聞記事でのように）でとり扱われているので，解釈がどうであるかによることが多い．しかし，ここでもまた，その記事あるいは文書の全体の調子をまっさきに考えるべきで，どちらともつかない場合は合議できめなければならない．どんな場合でも，カテゴリイは media に押しつけるよりは，むしろ media から出てくるものである．いいかえれば，カテゴリイは media を検討してから立てるべきものである．

1）「語集」はいわゆる“語彙”のことである．われわれはこの場合語彙のかわりに，「語集」を使うことにする．

§111 テスト資料の構成

選び出された mass communication media は，テスト資料になるように分析されることが必要である．

まず，つぎの三つの基準によって，おおざっぱに分類されたいくつかの文書を選ばなければならない．

1 mass communication media の型（新聞など）

2 スタイルの全体の調子

3 事項

これらの資料から，それぞれの相違をよく代表するような sample（サンプル）[1] をひきぬかなければならない．最後の sample では，mass communication media のあらゆる型，スタイルのおもなレベル，およびこの調査が機能的な能力をはかるという方向をとるものである，という見地から，それぞれの事項にもっとも本質的であると考えられる語集（key-word）がひきぬかれていなければならない．

1) §113.0 を見よ．

この点から，テストの問題を作るための資料をあつめるのに，ふたつの方法が可能である．すなわち，

1 テストの問題を sample から直接，センテンスおよびパラグラフの形で選ぶか，あるいは，テストの問題を sample の一般的な印象を基礎にして作る．以上のどちらを選ぶかは，テストを準備する人たちぜんたいの討議と意見の一致とにまかされる．

2 語を，ある sampling 比率によってかぞえ，スタイルおよび事項の完全に客観的な切断面を示すような語集表を作り，この語からテストの問題を作る．

まえの方法は時間もあまりかからないし，もし数人の専門家の意見にまかされれば，ほぼ正確であろう．あとの方法は，はるかに多くの時間と人とを必要とするであろうが，客観的に見て，いっそう信用のおけるものになるであろう．

ついで，§101 で述べた原理にしたがってテスト資料を構成すべきである．

§112 調査の方法

ここで，調査の方法について考えておこう．

調査には，1 回の被調査者の群のわけかたによって，個人調査法と団体調査法とがある．

個人調査法には，

1 解答のしかたや調査をうけているときの態度までもこまかく観察し，記録できること．

2 文字を十分に書いたり読んだりできない者に対しても，それがどの程度にでき

ないのかを知ることができること.

3　筆答のほかに，必要な場合には口頭，作業などによって解答させることができるので，よりこまかい分析もできること.

4　時間的な制限を加える必要がないので，被調査者にとって時間のムダが起こらないこと.

5　調べられるべき者が，もっともつごうのいい場所で（たとえば家庭人はめいめいの家庭で），調査をうけることができるので，調べようとする対象をきめられた場所にあつめるための処置がはぶかれ，また，予定している対象を確実に追求して調べることができること.

などの利点がある

団体調査法には，

1　多数の者を短時日に検査できること.

2　施行条件を一定にすることがたやすく，したがって，その面からくる影響による誤差を少なくできること.

などの利点がある.

しかし，われわれの場合のように，多くの被調査者が全国的にひろがっているときには，個人調査法によると，

1　調査をみじかい時日に終えようとすれば，非常に多くの調査者を必要とすることになり，また，

2　調査者を少数ですますつもりならば，調査の時日を極めて長く必要とする.

3　したがって，どちらの場合も，たくさんの費用を必要とする.

4　予定しているそれぞれの対象を戸別訪問するとしても，各家庭でほかの家族その他から分離して調べることは，日本の現在の家庭状況ではむずかしく，個人調査法の利点を十分に活用できない.

したがって，この読み書き能力調査では，団体調査法によらなければならない，という結論に達した.

つぎに，団体調査法をとる場合には，時間をどの程度にしたらいいか，という問題がある．学生生徒などのように，試験になれた人たちに対しては，相当の長時間を予定してもいいが，対象が一般の人である場合には，これまでされた知能検査などの例から考えても，1 時間以上にわたることはむずかしいと思われる．また，小田原市での準備調査のときに，テストが終ったのち，面接（interview）によって聞きとった結果によれば，50 分程度が適当のようであったので，テストのための実働（正味）時間が 50 分となるような問題を作ることにした.

§113 Sampling

§113.0 あらまし

この調査が，知能検査のように，おなじ問題をおなじ条件のもとに，しかも一定の教示 (instruction) をあたえてされる以上，調べられる者は，ひとところにあつめられなければならない．この調査は，日本国民を対象とするのであるから，あとから述べるように，ふつうの社会生活を営んでいる，と理論的に考えられる年齢層だけをとっても，調べなければならない人の数は約四千万人であって，これらの人を全部調べることは時間，費用，労力の点から不可能に近く，かつ，この調査の目的からいっても無意味に近い．なお，やや複雑な調査であるため，実際に調査者はかなりの素養があり，しかも一定の訓練をうけなければならない，と考えられる．したがって，そのような人を極めて多く得ることは期待できないから，被調査者の数もおのずから制限されるであろう．さらに，調査のための日時は，費用および調査の目的からいっても，なるべくみじかいことが望ましい．そのうえ，調査に使える費用もかぎられている．

以上のいくつかの事情から，調査は調査対象の全部ではなく，その一部についてするよりほかはない．つまり，いわゆる，sampling[1] の方法をとるのである．

1) いくつかの対象からあるきまった比率で組織的に対象をひきぬくことを「sample する」という．sample することの意味で使うときは sampling となる．

sampling は，いっぱんにつぎのような場合にとられる方法である．

1　日時，費用および調査者の事情などのために，全数調査ができない場合．

2　調査上のいろいろな誤差を少なくしたい場合．微妙な調査の場合には，調査者および被調査者の数がふえればふえるだけ調査の実施にともなう誤差が大きくなるから，むしろ sampling による調査のほうが正確な結果の得られることがある．また，この場合のほうが，集計・整理にともなう誤差が少なくなることが多い．

3　おなじ目的の調査で，いろいろなことを調べたい場合．全数調査に要するとおなじ費用と日時とを sample 調査にかけるならば，全数調査で計画されるよりも多くのことがらについて調査でき，調査そのものを効果的にすることができる．

4　時とともに移り変わるような対象を調べる場合．全数調査のために集計・整理に多くの時日を費していると，その結果が無意味になるようなことがある．このような場合には，簡単に sampling によって調べるほうがいい．4年ごとにする国勢調査（全数調査）の詳しい項目の集計・整理には，4年以上かかることが少なくない．

5　あまり高い精度を必要としない調査の場合．

さて，sampling はどのような方法をとるべきであろうか．ここでは，科学的に，つまり確率論的に信頼度を保証できる random sampling（ランダム・サンプリング）[1]

がもっともいいと考えられる．random sampling といっても，それは，読み書き能力に影響を及ぼすと考えられる，いろいろな要因を基礎にして，対象をなるべく同質な集団にわけ，そこから random に sample するという，stratified random sampling method（層別してから random sample をひきぬく方法）をとるべきである．ただし，このような sample の方法をとる以上，実際に調査する人に，対象をひきぬく sampling の技術を訓練しなければならない．

1） これについては §301.0 で述べる．

調査の対象である sample[1] を作るには，つぎのことを考えなければならない．

1　読み書き能力に影響を及ぼす要因の決定

2　層別[2] の準備

3　sample の決定とその割当方法

以下，それぞれについて，そのあらましを述べよう．

1） ここでこの本で使われる“調べられる人”の呼びかたについて定義しておこう．「sample」とは，sampling によって母集団（§301.1 を見よ）からひきぬかれた，調べられるべき対象をいう．実際に調査をうけるかどうかはまだ問題ではない．この章のはじめに述べた「sample する」にはいつも「する」がついているから，この「sample」とまぎれるおそれはないと思われる．つぎに述べる「sampleさん」と区別した理由は，たとえば §451「欠席者の分析」で明らかであろう．

「sample さん」とは，調査をうけた sample をいう．‘被調査者，被験者’，‘ひきぬく，抽出する’ などの語を使わなかったのは，これらの語が，「統計学的に厳密にひきぬく」という意味をふくんでいないと考えたからである．ただし，「sample を sample する」とでもいうべきときだけは「sample をひきぬく」とした．

この本で使った「被調査者」は，調査の一般論として調査される者を論ずる場合，または，われわれの調査であっても，調査される者が sampling によらないで得られた場合，たとえば，予備テストについてだけ使われる．

「調査される（べき）者」というときには，まだ「被調査者」であるか，「sample（さん）」であるかがわからないことを示す．

2） §113.2 を見よ．

§113.1　読み書き能力に影響を及ぼす要因の決定

読み書き能力を身につけ，また失わないことは，つぎのような言語以外の要因にも影響されるものと考えられる．これについては §100 で簡単に述べたが，もう一度整理しておこう．すなわち，1 居住地，2 性，3 年齢，4 個人の経歴および生活状態，である．これらを概括的に考察するとつぎのようである．

1　居住地．このカテゴリイには四つの要因が組み合わさっている．すなわち，i 市部と郡部，ii 地理的・文化的地域，iii 経済的・生態学的地域，iv 交通の便．以上のうち，iv はもともとほかの三つのものといっしょにまざり合っていると考えられるから，

特別に調べる必要はない．

i　市部と郡部．市部と郡部とはひとつのカテゴリイとして，都会化の程度，人口密度および西洋化の程度といったような特長を反映する．いっぱんに，このカテゴリイに選ばれる下位の等級は，まず，大都市，中都市，地方都市，町，村という5重の分類を反映しなければならない．つぎに反映すべきことは，基準になると思われる，これら各地域内の産業構成，文化普及度あるいは人口密度のなんらかの違いである．なお，大都市的地域は生態学的観点からさらにこまかくわけることが必要であろう．外海に点在している島島は，調査を実施する場合の能率という点から，のぞくことが必要である．

ii　地理的・文化的地域．過去の社会調査の経験が教えることは，地理的・文化的特長の基準としては，都府県よりはむしろ，ある諸地域が考えられなければならない，ということである．これら諸地域の正確な区画とその数とは，多くの方面の専門家の判断によっていろいろであるかもしれない．しかし，この調査では，ほかの要因と相関させるために，地域の数は最小限まで減らさなければならないが，つぎの6地域以下に減らしては危険と思われる．すなわち，北海道，東北，関東，関西，中国・四国，九州．

iii　経済的・生態学的地域．この分類は産業および文化的カテゴリイの反映である．考えられるおもな下位の等級は，まず全国産業総計での産業構成比率，産業の質的構造の差異および文化普及率である．産業構成のうち，食糧生産業の型は，日本の社会学でふつう使われている分類からとられるであろう．すなわち，農業，水産業，鉱業と林業．

2　性．男女について，さまざまの違いが，理論的にも常識的にも予想される．

3　年　齢．年齢と読み書き能力との関係を，どういう限界内に求める必要があるかをきめなければならない．literacy の機能的定義をひろげれば，読み書き能力が，文化的活動に参加するうえに影響を及ぼす，と理論的に考えられる年齢だけを考えればいいと思われる．したがって，14 歳以下および 65 歳以上の男女ははぶかれるわけである．これらの年齢層は肉体的特長に左右されている．

すなわち，14 歳以下は，読み書き能力を現に得つつある期間であって，まだ世間から，完全にその能力を得たとは考えられていない期間である．一方，65 歳以上は，ふつう，その個人が社会的活動からのぞかれる期間である．これらの理由によって，調査は 15 歳以上 65 歳未満のあいだの年齢についてだけするのがいいと考えられる．相関関係を見る目的から，調べられる人人を，たとえば 15 歳から 19 歳，20 歳から 24 歳などというように，おなじ年齢間隔を持った下位期間にわける[1]．

1)　年齢のかぞえかたは，調査のおこなわれた 1948 年には，まだいわゆる“数え年”の方法によっていたので，この本で何歳というのは，すべて数え年によったもので

ある．なお，“満”によるかぞえかたは 1950 年から公式のものとなった．

4　個人の経歴および生活狀態．個人の過去の経歴および現在の生活狀態が読み書き能力に影響を及ぼすことは理論的に当然考えられるが，特にそのうちで注意しなければならないのは，つぎの三つのことであろう．

i　職業

ii　習得した教育レベル

iii　新聞を読む程度

つぎに，それぞれについて述べよう．

i　職　業．職業については，1947 年 10 月の国勢調査がとった職業分類を使う場合には，別別のカテゴリイを結び合わせなければならない．使われるべき正確な分類は，しかるべき社会学者，統計学者，経済学者および教育学者によってきめられるであろうが，いっぱんにつぎのような分類に近づけなければならない．

a　事務的のもの

b　技術的のもの

c　自由業的のもの

d　肉体労働的のもの

なお，これをさらに明確にとらえるためには，産業構成とこの職業とを合わせ考えることが必要である．また，文字言語の使用と経験とがおなじであると考えられるかぎり，このようなカテゴリイの数は，最小限に減らされなければならない．

ii　習得した教育レベル．教育は，読み書き能力を身につけ失わないことに影響する範囲をきめるものであるから，sample はいくつかの教育的経験のレベルを代表するものでなければならない．大切な問題は，literacy がどんな教育レベルで確実にあたえられるかということである．ほかの要因がおなじならば，読み書き能力は義務教育の 6 年間（旧制），または 9 年間（新制），あるいはほかの教育のレベルによってどんなに影響されるかということである．調査をする場合に考えなければならないカテゴリイはつぎのようである．

a　まったく教育をうけなかった者

b　小学校を中途退学した者

c　小学校を卒業した者

d　高等小学校以上の教育をうけた者

さらに d はつぎのようにわけられるであろう．

a　高等小学校を中途退学した者

b　高等小学校を卒業した者

c　中等学校，実業学校，青年学校を中途退学した者

d　中等学校，実業学校に在学中の者

e　中等学校，実業学校，青年学校を卒業した者

f　大学，高等専門学校を中途退学した者

g　大学，高等専門学校に在学中の者

h　大学，高等専門学校を卒業した者

おそらくf以下の教育レベルは無視していいであろう．いいかえれば，ここで定義されるような literacy は高等教育の期間中にいっそうみがきをかけられるであろうが，それはこの期間以前に最低限得られるであろうと考えられる[1]．

1) この調査をした 1948 年は，ちょうど教育制度が新しくなる入れかわりの時期にあたっていたので，整理上に多少の無理をしなければならなかった．たとえば，中等学校を旧制度では小学校につづく 5 年（または 4 年），新制度ではおなじく 3 年で卒業するのであるが，集計にあたっては，制度の新旧を問わず，卒業ならば一様に中等学校卒業としてまとめた．§410を見よ．

iii　新聞を読む程度．社会生活で新聞は大切な報道機関であって，これを読んでいるかいないか，または読めるか読めないか，によって読み書き能力に差異の考えられるのは当然であろう．

§113.2　層別の準備

層別（stratification）[1]をするには，まず，以上のいろいろの要因に関する分布を示す統計資料を準備しなければならない．このような統計資料（統計表）は，層別のためにはもちろん，strata（ストラータ）[1]から random に得られた sample が偶然的なものであるかどうかを明らかにするためにも，さらに，調査後，要因のあいだの相関関係を見るのにも必要である．

1) 「層別」とは，ある要因，要素（標識，しるし）について同質と考えられる，ひとつのもの，あるいはその集団をひとまとめにしてさらに大きい集団にまとめることである．こうしてできた新しい集団を「strata（層）」という．strata の単数は stratum であるが，この本では単数，複数を問わず，すべて strata とした．なお，「層」という語を使わなかったのは，これが漢字1字であらわされているために，つかみにくいと考えたからである．

これらの表の大部分は 1947 年（昭和 22 年）10 月の国勢調査およびその他の見積りを基礎にして作ることができる．

前項に述べたことを考え，また実際に起こる困難をも考え合わせて，つぎのように層別するべきであろう．まず，日本全国民を読み書き能力に関して同質と考えられる集団にわけなければならない．そのために，i 地理的・文化的地域，ii 市部・郡部，iii 経

済的・生態学的地域，iv 文化普及率，v 産業構成，vi 人口密度などによって，それらのほぼ等しい構成を持つような集団にわけるのである．

§113.3 Sample の決定とその割当方法

sampling によって sample をきめるには，sample をとるべき母体をはっきりさせなければならない．この母体は母集団といわれるものであるが，母集団はわれわれが調べようとする調査対象から，ある確率をあたえて作られるものである．それでは，調査対象とはどんなものであろうか．はじめにこれをきめてかからなければならない．

まず，14 歳以下および 65 歳以上の人は調査対象からのぞいた．これは，上にも述べたように，かれらはまだ，あるいはもう正常な社会生活を営んでいない，とだいたい認められるからである．おなじようなわけで，外国人，めくら，つんぼ，おし，および精神的あるいは肉体的に無能な人はのぞくべきであろうと考えた．

さらにまた，調査実施上の便宜を考えて，外海にある島島に住む人人をも対象外と考えることにした．

以上のように考えて調査対象をきめたならば，対象となる日本人の各人におなじ確率をあたえて sample をひきぬくということにして，母集団を作るのがいいであろう．

それから sample をきめるには，sampling の理論にしたがって，高い信頼度で，日本国民ぜんたいの読み書き能力の分布を推定できるように心がけることが必要である．

こうしてきめられる sample の数を全国に割当てる方法がこんどは問題となる．われわれは，ただ読み書き能力だけではなく，この機会に，まだできていないいろいろな統計（学歴，性，年齢，職業，ラジオ普及率，新聞を読む者の率，居住年数，集合状況）のまざり合った結果（統計表）をも得たいと思うから，ただひとつのしるし[1] にだけつごうのいい割当方法は適当ではなく，また集計のときに，ただ平均点とか，あるいはひとつのしるしだけにかぎらず，こまかい分析をもしなければならないので，そのときの便宜をも考えて，strata の人口に比例して sample をふりわける方法をとるべきであろう．

1) 「しるし」はふつう使われる“標識”のかわりに使った．