第１章　標本設計の概説

第１節　標本抽出の基本的な考え方

労働力調査が明らかにしようとするのは，雇用者数や完全失業者数など，ある属性を持つ15 歳以上の総数である。我が国の 15 歳以上人口は約１億１千万人（平成 29 年推計）であるが，労働力調査ではその約 1/1,100 の 10 万人を調査することにより全体を推定している。このように抽出した一部を調べて全体を推定する調査を標本調査という。また，属性を明らかにしようとしている対象全体を母集団（この場合は 15 歳以上人口），抽出されたものを標本と呼ぶ。

このような標本調査の結果から精度の高い推定をするためには，標本が母集団の良い縮図になっている必要がある。労働力調査の場合でも，大都市に住む者しか含まれていない標本や，収入の多い者しか含まれていない標本など，標本に偏りがあっては全体を正しく推定することはできない。

母集団の良い縮図を得る方法として，標本を無作為に抜き出す方法がある。いわゆるくじ引きの原理によって一人ずつ選んでいく方法で，この方法を採用した場合，自然に様々な属性の者が含まれるようになる。しかし，１億１千万人から直接，無作為に 10 万人選び出すというのは簡単ではない。くじ引きのためにはくじを作らなければならないのと同様，まず，１億１千万人のリストを作成する必要があり，そのためには国勢調査と同等の規模の調査が必要となる。しかも，そのリストが調査時点における母集団の姿を反映したものであるためには，毎月のメンテナンスが必要となる。これは労働力調査そのものより大変な作業となる。さらに，仮にリストが完成し，標本を抽出できたとしても調査の実施は大変なものとなる。全国に散在する 10 万人を調査するということは，調査員一人一人が広い地域に住んでいる数人を巡回して調査するということであり，非常に多くの調査員が必要になるからである。

このようなことから，労働力調査では，まず全国で約 100 万ある国勢調査の調査区から約2,900 調査区を抽出し，次に抽出した各調査区について調査対象の住戸を約 15 戸ずつ抽出するという，２段抽出法を用いて抽出を行っている。この方法は幾つかの利点を持つ。まず，第１段目の抽出である調査区の抽出においてリスト作りの困難がない。５年に１度行われる国勢調査において調査区の設定が行われており，このリストから必要な数の調査区を抽出すればよいからである。また，調査区は地面の区画であって，そこに住む人間がどう動こうとも変わることはないため，原則としてリストのメンテナンスが必要ない。次に，第２段目の抽出で用いられる住戸リストについても，調査区内の住戸のみをリストアップすればよいから，全国の住戸をリストアップすることに比べかなり容易である。さらに，実地調査

の面からみると，国勢調査の調査区はおおむね 50 世帯となるように設定されており，一人の調査員が担当するのに適当な世帯数になっている。

このように，国勢調査の調査区を用いた２段抽出法は抽出作業や実地調査上の利点が多い。一方で，標本に様々な属性の者が入るようにして良い縮図を得るという観点からみると，調査区という「かたまり」を抜き出しているため，例えば社会福祉施設だけからなる調査区のような同じ属性の者の集まりが偏って抽出されてしまうおそれがあるなど，全国から直接 10 万人抽出する場合より推定の精度は劣ると考えられる。

そこで，労働力調査では精度を上げるため，様々な手法を用いている。以下，そのような手法の解説も交えつつ，労働力調査における標本の抽出方法を説明する。

第２節　標本調査区の抽出

１　調査区の層化とその目的

調査区には，会社の独身寮があるもの，農家世帯の割合が高いもの，サラリーマン世帯の割合が高いものなど，様々なタイプがある。このことは，例えば産業別の就業者数を高い精度で推定しようとする場合，農家世帯の割合が高い調査区がたまたま多く抽出されるなどということが起こらないような工夫が必要であることを示唆する。そこで労働力調査では，調査区の抽出に層化抽出法を用いている。これは，調査区の持つ特性によって調査区を幾つかのグループに分けておき，各グループで独立に抽出するという抽出方法である。このグループを層といい，グループに分けることを層化という。層化抽出は，良い縮図を得るために非常に有効である。

例えば，抽出率 1/100（100 個に１個の割合で母集団から標本を選ぶこと。）で調査区を抽出する場合，地域ごとに層化してから，それぞれの地域において 1/100 の抽出率で抽出することで，たまたまある地域の調査区が多い標本となるおそれがなくなり，地域間のバランスのよい標本となる。同様に，第１次産業の割合が高い調査区，第２次産業の割合が高い調査区，第３次産業の割合が高い調査区というように，産業の特性により層化して抽出することで，たまたま農業人口の割合が高い調査区が多く選ばれてしまうというおそれがなくなる。

労働力調査では，地域別及び産業別表章において一定の精度を確保するため，国勢調査の結果から得られる調査区に関する詳細な情報を利用して，地域区分（11 地域）に加え，産業，従業上の地位により各調査区を分類した層化基準（第２章第３節）を作成し，利用している。

なお，この分類は国勢調査の調査時点の情報であるため，調査から月日が経つと次第に実態から離れてしまう。このため，５年に１度，国勢調査の調査区関連資料がそろった段

階で，新しいものに切り替えている。つまり，直近の国勢調査で設定された調査区を新たに層化し，これを抽出のためのリストとして使用している。

２　標本調査区の確率比例抽出

労働力調査では，リストから一つずつ無作為に調査区を抽出するのではなく，系統抽出法を用いて調査区を抽出している。系統抽出法とは，母集団に一連番号を付け，標本とする番号を等間隔に選ぶ方法である。例えば，200 調査区から 10 調査区を抽出する場合（抽出率 1/20）は，まず１から 200 までの番号を調査区に付ける。次に，１から 20 までの数字から無作為に一つの数字を選び抽出起番号とし，これに抽出間隔（抽出率の逆数）を次々に足して，得られる数の一連番号を持つ調査区を抽出する。仮に抽出起番号を７とすると，７に抽出間隔 20 を次々に足した 7，27，47，‥‥‥，187 を一連番号とする調査区が抽出される。

系統抽出法は抽出作業が容易であるという利点に加え，例えば調査区を市町村ごとに並べた上で一連番号を付与すれば，標本が一部の市町村に偏ることがなくなり，層化に似た効果も期待できる。

調査区の配列の効果

一般に調査区の特性が配列の順に従って単調な変化をするとき（例えば調査区の配列が都市的な地域から農村的地域の順になっているような場合）には層化に似た効果が表れ，推定の精度向上が期待できる。これに対し調査区の特性が配列の順に従って周期的な変化をするときには，調査区が特性の周期の長さに等しいか，あるいは近い場合，ある特性に偏った標本になるおそれがある。調査区の特性が配列順と無関係のときには，単純無作為抽出法の場合とほぼ同じと考えられる。

また，国勢調査の調査区は，１調査区当たりの世帯数がおおむね 50 世帯となるように設定されているが，実際には世帯数はかなりばらついている。これは推定の精度を低下させる原因となるが，抽出時に調査区の規模に関する情報を利用して確率比例抽出を行うことにより，精度を向上させることができる。

例えば，３調査区から一つ抽出し，その調査区に居住する者を全て調べて３調査区全体の就業者数を推定する場合を考えてみる。仮に，調査区（Ａ，Ｂ，Ｃ）の規模が次のようになっていたとして，Ａ，Ｂ，Ｃのいずれかを調べて 120 人という就業者数を推定してみる。

	A	B	C	計
総人口	100 人	60 人	40 人	200 人
就業者	60 人	40 人	20 人	120 人

無作為抽出や，先に述べた系統抽出の場合，A，B，Cそれぞれが選ばれる確率（抽出確率）は 1/3 である。仮にAが選ばれたとすると，全体の就業者数の推定値は，

$$60\text{ 人}\times\frac{3}{1}=180\text{ 人}$$

ということになる。ここで 3/1 倍したというのは，抽出率 1/3 の逆数を乗じたのであるが，抽出確率の逆数を乗じたともいうことができる。同様にB及びCが抽出された場合の推定値は，それぞれ 120 人，60 人となる。

一方，各調査区の総人口が事前に分かっていた場合，つまり，A，B，Cの人口規模の比が５：３：２であることが事前に分かっていた場合は，A，B，Cから一つ選ぶ場合の確率をそれぞれ 0.5，0.3，0.2 とすることができる。このような抽出法を確率比例抽出という。

こうした場合，全体の推定値は等確率で抽出した場合と同様，抽出確率の逆数を乗じることによって得られる。例えばAが選ばれた場合は，

$$60\times\frac{1}{0.5}=120$$

となり，B，Cが選ばれた場合はそれぞれ

$$40\times\frac{1}{0.3}=133.3\cdots \qquad 20\times\frac{1}{0.2}=100$$

となる。各推定値の分布は抽出確率に従うので，例えば 10 回に５回は 120 という推定値に，10 回に３回は 133.3…という推定値になる。等確率（1/3）で選んだ場合は 180，120，60 という推定値がいずれも３回に１回の割合で得られるが，これと比べて確率比例抽出を行った場合はばらつきが小さくなり，しかも母集団の値に近い推定値が得られることが多くなる。

このように確率比例抽出を行うことにより一般に精度が向上する。この例では調査区の規模が完全に分かっていると仮定したが，ある程度近似的な状況でも同様の効果が期待できる。このような考え方により，労働力調査では，国勢調査時に得られる情報から換算世帯数（第２章第３節参照）を求めて抽出に利用している。

$$\text{換算世帯数}=\begin{pmatrix}\text{２人以上の}\\\text{一般世帯数}\end{pmatrix}+\frac{(\text{１人の一般世帯数})+(\text{施設等の世帯人員})}{3}$$

実際の抽出においては，換算世帯数は次のようにウエイト（第２章第４節参照）という形に集約して使用している。例えば，ウエイト２の調査区は同じ調査区を２個並べ，ウエイト４の場合は４個並べるというようにして調査区を１列に並べた上で，等間隔に抽出している。これは確率比例系統抽出とも呼ばれている。

換算世帯数	ウエイト
１～15	1
16～30	2
31～45	3
46～60	4
⋮	⋮

第３節　標本調査区内における住戸の抽出

１　住戸を抽出単位とする理由

第２段目の抽出，すなわち抽出された調査区における調査対象の抽出は，調査区のようにコンピュータのプログラムにより抽出するのではなく，調査員が実地に調査区を巡回してリストを作成し，指導員が抽出している。

この場合，何のリストを作成して抽出を行うかが問題となる。労働力調査は個人の属性を調べる調査であるから，①調査区内に居住する者のリストを作成し，個人を直接抽出する方法がまず考えられる。また，②世帯のリストを作成し，抽出した世帯の世帯員について調査する方法，③建物やアパートの部屋といった「入れもの」のリストを作り，抽出した「入れもの」に居住する世帯を調査する方法も考えられる。

労働力調査では，このうち③の方法，すなわち「入れもの」のリストを作成して抽出する方法を採っている。この「入れもの」を「住戸」と呼んでおり，抽出の際の単位となるものとして「抽出単位」（調査区を第１次抽出単位とみた場合は，住戸は第２次抽出単位）とも呼んでいる。抽出単位（住戸）は「調査区内にある住宅やその他の建物の各戸で，一つの世帯が居住できるようになっている建物又は建物の一区画」と定義され，例えば，一戸建住宅の場合はその建物全体が抽出単位（住戸）となり，アパート，マンションなどの場合は建物内の各区画それぞれが抽出単位（住戸）となる。

住戸を抽出単位とする大きな理由は，リストが劣化しにくい点にある。労働力調査では，同一の調査区を４か月継続して調査し，リストは開始月の前月に作成する。仮に世帯や個人のリストを用いた場合，転出及び転入などによりリストの内容が調査時点と合わなくなりやすく，また精度の高い推定を行うために必要なリストのメンテナンスも困難であ

る。住戸を抽出単位とした場合は，世帯や個人の移動にかかわらず抽出された住戸に調査時点で住む者を調査すればよく，またリストのメンテナンスも急に家が建ったり，取り壊されたりした場合などに限られるため，比較的容易である。

２　住戸の把握

まず，調査員が実地に調査区を巡回して調査区地図（付録６参照）を作成することにより，調査区内の全ての住戸を把握する。調査区地図には，抽出単位である住戸のほか、調査区の境界及び道路，河川，鉄道や建造物など目印となるものを記入する。また，調査員は把握した各住戸の名称や住所，居住者の有無を抽出単位名簿（付録５参照）に記入する。調査区地図は，抽出後，調査対象となった住戸に居住する世帯を訪問するときに必要となる。また，翌年の調査で，前年と同じ住戸に居住する世帯を確実に調査するためにも必要である。

住戸の把握に当たっては，調査時に人が居住している可能性のあるものは全て把握して調査区地図及び抽出単位名簿に記入しなければならない。居住部分のない事務所や工場は，人が住む可能性がないのでその必要はないが，空き家は人が住む可能性があるので記入する。また，建築中の家についても，調査時に完成している可能性があれば記載する。調査区地図及び抽出単位名簿の作成は，正確な調査を行うために極めて重要な作業である。

なお，病院，高齢者介護施設のような社会福祉施設，建設従業者宿舎などでは，部屋ごとに抽出単位（住戸）としているが，１室が 10 人以上収容できるようになっている場合，柱や通路などの目印によって更に小さく分割することとしている（第２章 表２－３参照）。これは，後述する住戸の抽出において，住戸内の人口の大きさを均等とみなして等確率としており，精度を考えた場合，各抽出単位内に居住する者の数が均等に近い方が好ましいためである。

３　標本とする住戸の抽出

調査対象となる住戸は，調査員が作成した抽出単位名簿から指導員が抽出する。この抽出は，住戸に一連番号を付して等確率で系統抽出を行っている。一連番号は，把握時に居住者のなかったものから番号を付け，次に居住者があったものに番号を付ける。これは層化と同じ効果を狙ったもので，この方法により調査区内における居住者がいない住戸，居住者がいる住戸の比に応じて，住戸が抽出されるようになる。

抽出率は，調査区のウエイトの逆数を用いている。ウエイトは，換算世帯数 15 を単位に定めるため，例えば国勢調査時に換算世帯数が 50 であった調査区は，ウエイトは４で，その調査区における住戸の抽出率は 1/4 となる。このとき，抽出単位名簿で抽出の起点

（抽出起番号）を２とすると，２・６・10…46・50 の計 13 世帯が抽出される。この方法を採ると，調査区における抽出単位数が多くなるに従い抽出率は小さく（抽出率の分母は大きく）なり，どの調査区も 15 世帯程度調査されるようになる。これは，調査員の事務量が平均化するという利点と，本章第５節で述べるように推定式が簡単になるという利点を持っている。

第４節　標本の交代

１　標本交代の方法

継続して標本調査を行う場合，各回の標本（標本調査区，調査標本）の決め方は，

① 最初に代表性の高い標本を選定して，それを長期間固定する方法

② 毎回全面的に標本を交代する方法

③ 毎回部分的に標本を交代していく方法

が考えられ，これらの一般的な特徴は次のとおりと考えられる。

①　最初に代表性の高い標本を選定して，それを長期間固定して調査する方法

これは，標本調査区のような調査標本の外的条件を固定する場合と，人や世帯のような調査標本自身を固定する場合で特徴が異なる。

ア　標本調査区を固定する場合

〔長　所〕

・ 標本調査区を毎月交代する場合より，調査標本の均質性が保たれるので月々の時系列の精度が高い。

〔短　所〕

・ 毎月の標本を累積して求める年平均結果などについては，標本調査区を毎月交代する場合より精度が低い。

・ 抽出単位名簿は初めに作り，その後は抽出単位の異動に応じて名簿上での追加，削除を行うことになると，既に作成した名簿に依存しがちになり，特に転入者など新たに調査対象とするべき標本を漏らすおそれがある。

イ　調査標本を固定する場合

〔長　所〕

・ 調査標本の均質性が保たれるので月々の時系列の精度が高い。

・ 調査員と調査標本となった報告者は大体顔見知りになり，調査に対する報告者の抵抗が比較的少ない。

・ 調査員の交代が少ないので新任調査員の訓練費用が少なくて済む。

〔短　所〕

・ 調査標本となった報告者の居住地が移動した場合に追跡が困難になり，調査から漏れるおそれがある。
・ 時間の経過とともに生じる標本の質的変化（高齢化など）及び量的変化（死亡による減少など）により，母集団の代表性の低下を招きやすい。
・ 同一の報告者に毎月同じ調査を繰り返し行うため，特定の報告者に負担を負わせることになる。また，一般的に，調査員が同じ調査標本を継続的に調査することにより事務の簡略化が生じ，調査の正確性が損なわれるおそれがある。

② 毎回全面的に標本を交代する方法

〔長　所〕

・ 毎月の標本を累積して求める年平均結果などについては，毎月標本を継続する場合より精度が高い。

〔短　所〕

・ 月々の時系列については，毎月標本を継続する場合より結果の安定性が乏しくなる。
・ 調査のたびに抽出単位名簿を作る必要がある。
・ 毎月報告者が交代することとなるため，調査員の報告依頼の労が多くなるほか，標本の交代に伴う調査員の交代が多くなり，新任調査員の訓練費用もかかる。

③ 毎回部分的に標本を交代していく方法

この場合の長所及び短所は，前二者のほぼ中間程度になる。

２　標本交代の工夫

雇用・失業動向などをみるために労働力調査の結果を利用する場合，前月差（比）や前年同月差（比）によって動向を把握することが多い。このため，各月の推定値の精度向上だけではなく，前月，前年同月との比較上の安定性のための工夫も必要となる。

そこで，労働力調査の標本の交代においては，前月，前年同月との比較の安定性の向上を図るため，標本調査区は４か月継続して調査し，毎月 1/4 ずつ新しい調査区に交代している。また，標本調査区は，１年後の同じ時期にも調査を行う。

同様に，比較の安定性の向上を図るため，調査区を継続して調査する４か月について２か月ずつ前期と後期に分け，前期と後期との入れ替わりにおいて住戸（調査標本）を交代し，前期と後期はそれぞれ２か月同じ住戸（調査標本）を継続して調査している。

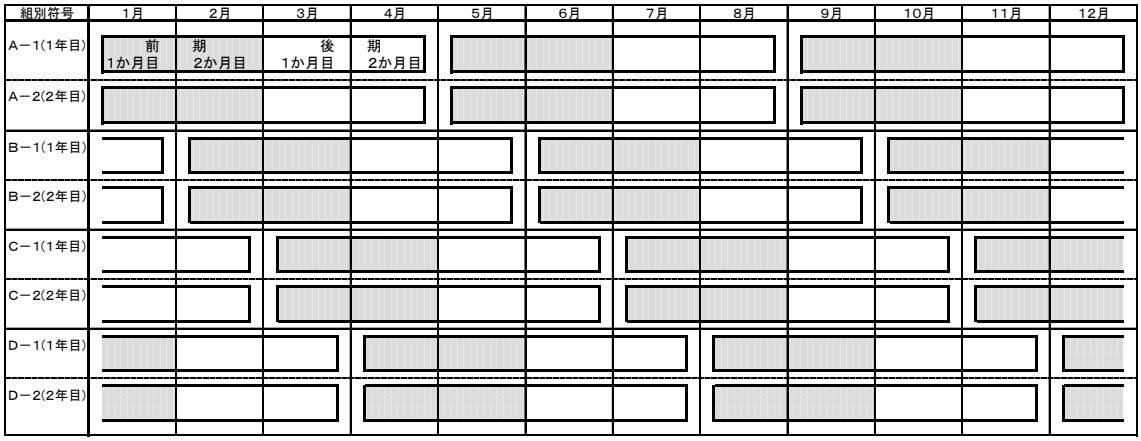
３　標本交代の効果

この標本の交代方法を図示すると図１－１－ａのようになる。同図に示すように，ある月をみた場合，調査を行う調査区は８組に分けられる。まず，調査区が今年初めて標本調査区となった調査区（これを「１年目調査区」という。）か，前年同月に調査し，再び調査を行う調査区（これを「２年目調査区」という。）かにより二つに分かれる。次に，調査区を４か月継続して調査するうちの前期か後期か，住戸を２か月継続するうちの１か月目か２か月目かにより四つに分かれ，計８組となる。

図１－１－ｂに示すように，前年との継続を見ると，２年目の調査区（A-2，B-2，C-2，D-2）が常に半分含まれており，調査を行う住戸のうち半分は前年の同じ月に調査を行っている。また，前月との継続をみると３／４が同じ調査区となっており，調査を行う住戸のうち半分は前月にも調査を行った調査区（B-1，B-2，D-1，D-2）である。これらは，それぞれの前年同月との比較，前月との比較の安定性を向上させる効果を持っている。

図１－１ 標本の交代方法

ａ ８組の標本調査区における標本交代の例



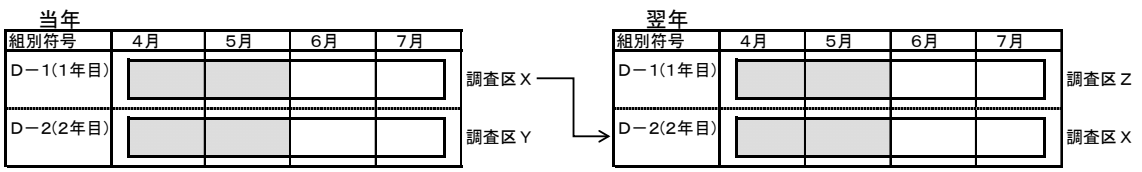
ｂ １・２年目，前期・後期及び１か月目・２か月目の別 ８組の標本調査区（５月の例）

	組別符号			
1年目	A-1	B-1	C-1	D-1
2年目	A-2	B-2	C-2	D-2

	組別符号			
前期	A-1	A-2	D-1	D-2
後期	B-1	B-2	C-1	C-2

	組別符号			
1か月目	A-1	A-2	C-1	C-2
2か月目	B-1	B-2	D-1	D-2

ｃ １・２年目標本調査区の交代の例



第５節　結果の推定方法

１　線型推定の考え方

標本調査は，一部を調査して全体を推定しようとするものであるが，全体の推定値は，標本から得られた値に，抽出率（抽出確率）の逆数を乗じることで得られる。このような推定を線型推定という。

労働力調査は調査区を層別に抽出しているため，各層で独立に推定値を算出し，これを足し合わせて全体の推定値を得ることになる。以下，第l層の就業者数X_lを推定する場合を例に説明する。

第l層で抽出された調査区がm_l個，各調査区のウエイトが$w_{li}(i = 1,\ \ 2,\ \ \cdots,\ \ m_l)$であるとし，第$i$調査区において，抽出された住戸全体で就業者が$X_{li}(i = 1,\ \ 2,\ \ \cdots, m_l)$人居住していたとすると，第$l$層の就業者数の線型推定値$X_l$は次のようにして求めることができる。

① まず，抽出された調査区内の就業者数の合計を推定する。抽出率はウエイトの逆数としていたから，抽出率の逆数はウエイトそのものになる。したがって，$X_{li}w_{li}$が第i調査区の就業者数になる。これは，ウエイト３の調査区の場合，三つに一つの割合で住戸を調査するから，調査した住戸に居住する就業者の合計が50人であったならば，50×３＝150人をその調査区の就業者数と推定するというものである。別の見方をすれば，抽出された各住戸は，３戸の住戸を「代表」しているのであるから，各住戸の就業者数を３倍し，それを足し合わせれば推定値が得られるともいうことができる。この「代表」の度合いがつまり抽出率の逆数なのである。

② 次に，各調査区の就業者数の推定値から層全体の就業者数X_lの推定値$\hat{X}_l$を求める。調査区の抽出は確率比例抽出であるため，各調査区の就業者数の推定値に調査区の抽出確率の逆数を乗じて層全体の就業者数の推定値を求める。

第i調査区の就業者数は$X_{li}w_{li}$と推定されていることから，第i調査区一つから層内全体の就業者数$\hat{X}_l$を推定しようとするとき，層内の全調査区のウエイトの合計をw_lとすれば，第i調査区からの推定値$\hat{X}_{li}$は，

$$\hat{X}_{li} = (X_{li}w_{li}) \times \frac{w_l}{w_{li}} = X_{li}w_l$$

となり，第i調査区のウエイトw_{li}によらない値となる。さらに，第l層の就業者数$\hat{X}_l$は，$\hat{X}_{li}(i = 1,\ \ 2,\ \ \cdots,\ \ m_l)$の平均値と考えることができるから，

$$\hat{X}_l = \frac{1}{m_l}\sum_{i=1}^{m_l}\hat{X}_{li} = \frac{1}{m_l}\sum_{i=1}^{m_l}X_{li}\,w_l = \frac{w_l}{m_l}\sum_{i=1}^{m_l}X_{li}$$

となり，第l層の各調査区の抽出単位に居住する就業者数を合算してw_l/m_l倍することで推定できる。抽出率をウエイトの逆数としたことにより，このw_l/m_lは調査区によらない定数になっている。このw_l/m_lを線型推定用乗率という（第３章第４節参照）。

２　比推定の考え方

先に算出した線形推定値について，補助的な情報を利用することで，より推定の精度を高めることができる。

先の例で考えてみると，第l層の就業者数X_lの線型推定値$\hat{X}_l$を求める方法と同様の方法により，第l層の総人口P_lの線型推定値$\hat{P}_l$を求めることができる。すなわちi番目の標本調査区において抽出された住戸にP_{li}人が居住していたとすると，層全体の総人口$\hat{P}_l$は，

$$\hat{P}_l = \frac{w_l}{m_l}\sum_{i=1}^{m_l} P_{li}$$

となる。$\hat{P}_l$も$\hat{X}_l$も標本からの推定値であるから，標本の選ばれ方によって，実際の値であるP_lやX_lより大きくなったり小さくなったりする。しかし，この二つの推定値の実際の値からのずれ方は，同じ方向であることが多いと考えられる。例えば世帯規模の大きい世帯からなる調査区がたまたま数多く抽出された場合，$\hat{P}_l$も大きい値となるが，同時に$\hat{X}_l$も大きい値となる可能性が高いからである。つまり，$\hat{P}_l$と$\hat{X}_l$の比は，$\hat{P}_l$や$\hat{X}_l$そのものに比べ，より安定することが予想される。

そこで，仮に[注1)]人口の大きさP_lが，別の資料により正確に知ることができたとすると，$\hat{X}_l$そのものを推定値とするより，

$$\tilde{X}_l = P_l \times \frac{\hat{X}_l}{\hat{P}_l} = \hat{X}_l \times \frac{P_l}{\hat{P}_l}$$

を推定値とした方が安定した値を得ることができる。この方法は比推定と呼ばれ，P_lをベンチマーク人口という。この方法を用いると，$\tilde{X}_l$の誤差は，$\hat{X}_l$と$\hat{P}_l$の誤差が同じ方向に向かう場合，$\hat{X}_l$の誤差と比べてかなり縮小する。このように，別途正確な数値が得られるものと高い正の相関を持つものを推定する場合，比推定は非常に有効である。

注 1）ベンチマーク人口は必ずしも層別とする必要はない。労働力調査においては，ベンチマーク人口は男女，年齢５歳階級（15 区分）及び地域（11 区分）別としている（第３章第３節参照）

第６節　推定値の誤差

１　誤差とは

ある時点における就業者数を標本調査から推定する場合，結果数値が必ずしも真の値に一致するわけではない。この差を誤差といい，一般的に標本調査であることに起因する「標本誤差」と，それ以外の実地調査における調査票の誤記入などに起因する「非標本誤差」とに分けて考えられている。結果数値をみる場合，誤差の存在を常に認識しておく必要がある。

２　標本誤差

労働力調査では標本を調べて全体を推定しており，単純化して考えると，抽出率が1/1,000で10万人調査した場合に，そのうち５万人が就業者であったとすると，

$$5\text{万人} \times \frac{1000}{1} = 5000\text{万人}$$

という算式によって就業者数は5000万人と推定しているといえる。

このとき，この5000万人という結果数値は真の値に等しいとは限らない。なぜなら，調査された10万人という集団は全国民の完全な縮図とは限らないからである。同じ時点において標本の抽出をやり直して10万人調査できたとしても，再び10万人中５万人が就業者数となるとは限らず，５万1000人かもしれないし，４万9000人かもしれない。つまり，真の値は一つであっても，推定値はそれより大きくなるかもしれないし，小さくなるかもしれないのである。このように，標本から推定することによって生じる誤差を標本誤差という。

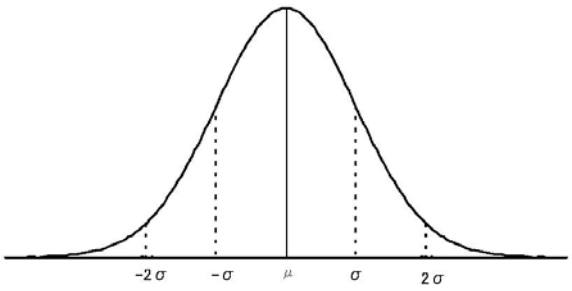
労働力調査は，このような標本誤差が避け得ないものであるから，結果数値をみる場合には注意が必要である。例えば，先月5000万人であった就業者数が今月5001万人になったとした場合，季節変動がないものとしても，これをもって直ちに就業者数が増加したとは判断できない。先月の数値も今月の数値も共に標本誤差を含んでいるため，真の値は逆に減少しているかもしれない。

しかし，標本誤差の存在は，結果数値が信頼できないということにはつながらない。推定値が真の値に「近い」ということ，つまり大きく真の値から離れることは少ないということがいえるからである。上の例でいえば，もし5000万人から5100万人に増加していれば，真の値も増加しているであろうということはかなりの確率でいえるのである。このような判断に根拠を与えるものとして標本理論がある。標本理論は，標本調査から得られた推定値が真の値からどの程度離れる可能性があるかを理論的に示してくれる。

標本理論において，真の値からの距離を測定する際の物差しとなるのが「標準誤差」

である。抽出を何度も繰り返した場合，それら抽出結果による推定値は真の値の周りに，ある分布を示すであろう。この分布の広がりが小さければ精度の良い推定ということができる。分布の広がり具合は，一般的に分布の標準偏差σで示され，この標準偏差を標本理論では標準誤差と呼んでいる。標準誤差は精度を示す指標であると同時に，誤差を測る尺度となる。また，標準誤差を真の値に対する比率で示したものを標準誤差率という。標本理論によれば，推定値$\bar{x}$は真の値μの周りにほぼ正規分布をしていると考えられることから，$\bar{x}$とμの差がσ未満となる確率すなわち$|\bar{x}-\mu|<\sigma$となる確率は約68%，$|\bar{x}-\mu|<2\sigma$となる確率は約95%となる。つまり，推定値の誤差は３回中２回はσの範囲に収まっており，2σの範囲を超えることは20回に１回程度しかないと考えられる（図１－２）。

図１－２ 推定値の分布



労働力調査の場合，ある月の推定値が5000万人のとき，標準誤差は約25万人であると推定されているから，調査結果を正確に記述しようとするなら，単に推定値が5000万人であるというのではなく，例えば5000万人±25万人の間に真の値が2/3の確率で存在するというように記述する必要があろう。

標準誤差は，標本の抽出方法あるいは結果の推定方法が複雑な場合，簡単には求められないが，10万人を無作為に抽出したと考えると，労働力調査のような人数の推定の場合，15歳以上人口をN，ある属性を持つ人口をX，標本数nの標本によるXの推定値を$\bar{x}$としたとき，$\bar{x}$の標準誤差$\sigma(\bar{x})$は，

$$\sigma(\bar{x}) \fallingdotseq N \times \sqrt{\frac{p(1-p)}{n}} \qquad p = \frac{X}{N}\ (X\text{の15歳以上人口に占める割合})$$

となり，標準誤差$\sigma(\bar{x})$を真の値Xで割った標準誤差率は，

$$\frac{\sigma(\bar{x})}{X} \fallingdotseq \frac{N\sqrt{\frac{p(1-p)}{n}}}{X} = \frac{N\sqrt{\frac{p(1-p)}{n}}}{Np} = \sqrt{\frac{1-p}{pn}}$$

となる。この式を見てわかるように，標準誤差は標本数の平方根に反比例して小さくな

る。また，全体に占める割合pが小さい場合，標準誤差は小さくなるが，標準誤差率は逆に大きくなる。

なお，労働力調査は実務上の制約などから２段抽出法を採っているため，実際の標準誤差はこの式で示される値よりやや大きくなる。

３　非標本誤差

非標本誤差とは，誤差の要因のうち標本抽出（偶然性）に起因するものを除いた全ての要因により生じる誤差をいう。それは更に，その要因により幾つかに分けることができる。例えば，報告者が質問を誤解したり懸念したりして，事実と異なる回答をする場合の誤りや，無回答，調査員の面接の拙さによる誤り，不慣れによる標本の脱落・把握誤り，連絡・指導の不徹底による誤り，調査票の処理及び集計上の誤りなどである。また，このように，非標本誤差は調査のあらゆる段階で発生する可能性がある。

非標本誤差の特徴は，標本誤差のそれとは対照的である。すなわち，標本誤差が標本の大きさと密接な関係があり，その制御が標本の大きさを通じて可能であること，避けられないものであること，量的な測定ができることに対して，非標本誤差は標本の大きさと直接関係がなく，標本の大きさを通じた制御ができないこと，原因を究明すれば避けられるものもあること，量的な測定が難しいことなどである。

調査が大規模になり関係者の数が増えると，非標本誤差の発生源も増えるものである。労働力調査に限らず，調査の各段階で誤りをできるだけ少なくし，非標本誤差を小さく抑えるためには，調査関係者の努力と回答者の統計に対する協力・理解が最も重要である。