# 第III部

# 調査対象の選び方

# 第6章　調査対象を選ぶ前に

## 6.1　標本調査用語の基礎知識

標本調査は、母集団から選んだ一部の調査対象のみを標本とし、標本における調査結果から母集団に関する情報を推定しようとする方法である (1.4 節参照)。

母集団から標本を選び出すときの単位を **抽出単位** と呼ぶ。抽出単位の例としては、個人、世帯、学校、施設、事業などが挙げられる。なお、抽出単位が常に調査対象に一致するとは限らない。例えば、公民館に勤務する職員の意識調査を行うため公民館を選び出す場合には、調査対象は公民館職員であるが、抽出単位は公民館である。

母集団における全ての抽出単位を並べたリストを **抽出台帳** あるいは **枠**（**フレーム**）と呼ぶ。個人を抽出単位とする場合には住民基本台帳や施設利用者名簿など、公民館を抽出単位とする場合には公民館のリストが抽出台帳となる。

標本として選び出す抽出単位の数の、母集団における抽出単位の総数に対する比率を **抽出率** と呼ぶ。

$$\text{抽出率 (\%)} = 100 \times \frac{\text{抽出単位の数 (標本)}}{\text{抽出単位の総数 (母集団)}}$$

全数調査 (1.4 節参照) とは、調査対象の抽出率が 100%である調査のことと言える。

標本調査では、標本に対する調査結果を基に母集団の情報を推定する必要がある (1.4 節参照)。標本における平均値が、そのまま母集団平均値の推定量 (8.1 節参照) となるような標本を **自動加重標本** と呼ぶ。

**ワンポイント アドバイス**

実際の調査では、台帳の不備や整備時期のズレなどによって、目的とする母集団と抽出台帳とが必ずしも完全に一致するとは限らない。そのため、抽出台帳に合わせて母集団を定義し直すこともある。定義し直した母集団を特に **枠母集団** と呼ぶ。

## 6.2　どういう標本がよい標本か

社会教育調査の目的は、母集団に関する情報を得ることである。標本調査では、母集団から選んだ一部の標本を調査し、その結果を基に母集団の様子を知ろうとする (1.4 節参照)。したがって、よい標本とは、その調査結果を基に母集団の情報を推定できるような標本、のことである。

標本となった調査対象については、その様子を調査結果によって直ちに知ることができる。しかし、そこから母集団の情報を推定することができなければ、標本を知ったというに過ぎない。

標本から母集団を推定するためには、標本が母集団の縮図・ミニチュアとなっていることが原則である。縮図となっていれば、標本について調べた結果は、母集団について調べた結果と似ている、と言えるからである。

**非確率抽出法** は、標本の特徴が母集団の特徴に一致するよう標本を選び出すことで、標本が母集団の縮図となることを狙う方法である (6.3 節参照)。例えば、母集団の男女比が 1 : 1 であれば、男女比が 1 : 1 となるように標本を選び出すのである。

ただし、基にした特徴については母集団の縮図となっていても、それ以外の特徴については母集団の縮図となっていることは全く保証されない。

これに対し、**確率抽出法** は、母集団の各抽出単位 (6.1 節参照) に、それが標本として選ばれる確率（これを **包含確率** と呼ぶ）を与え、その確率に従って標本を選び出す方法である。

確率抽出法では、調査対象の包含確率が全て等しい場合には、例えば、母集団の男女比が 1 : 1 であれば、標本の男女比も平均的には 1 : 1 となることが期待できる。このようなことは、男女比以外のあらゆる特徴について言えることであり、標本が母集団の縮図となっていると言える。

このハンドブックでは、主として確率抽出法を扱うこととする。

**ワンポイント アドバイス**

非確率抽出法のことを **有意抽出法** と呼ぶこともある。また、確率抽出法のことを **無作為抽出法** とも呼ぶ。ただし、単純無作為抽出法 (7.1 節参照) と混同しないよう注意する。

## 6.3　非確率抽出法のいろいろ

**割当法** は、標本として選ぶべき調査対象の数を、属性ごとに指定しておき、その条件が満たされるまで任意の調査対象を選び出す方法である。

表 6.1 は、200 人の標本を選び出すときの割当表の一例である。標本の属性の構成が表 6.1 の通りでありさえすれば、その選び出し方はどのようなものであってもよい。各セル（表の一マス）の人数は、表 6.1 ではどれも 20 人と等しいが、母集団での分布に合わせて変えることもある。

割当法は、マーケティングの分野などにおいて、よく使われる方法である。その理由としては、母集団情報の推定よりも商品に関心がある人の考えを知ることが目的であること、属性間の比較のためにどの属性についても十分な標本サイズを確保したいこと、が挙げられる。

**典型法** は、母集団の中で平均的、典型的と思われる調査対象を標本とする方法である。

**募集法** は、例えば広報誌などに調査協力者募集を公告し、応募してきた人を標本とする方法である。

**縁故法** は、友人・知人などの縁者を標本とする方法である。つてを頼ってさらに多数の人に調査依頼をつないでいく方法を **雪だるま法**（**スノーボール法**）と呼ぶ。

**ワンポイント アドバイス**

歴史的には、確率抽出法よりも割当法の方が古くから用いられていた。1948 年のアメリカ大統領選挙において、割当法が選挙結果の予測に失敗し、その経験から様々な確率抽出法が工夫されるようになったことは、よく知られている。

表 6.1: 割当表

| | 20 歳代 | 30 歳代 | 40 歳代 | 50 歳代 | 60 歳以上 |
|---|---|---|---|---|---|
| 男性 | 20 人 | 20 人 | 20 人 | 20 人 | 20 人 |
| 女性 | 20 人 | 20 人 | 20 人 | 20 人 | 20 人 |

# 第7章　確率抽出法のいろいろ

## 7.1　抽出方法の基本（その1）──単純無作為抽出法──

**単純無作為抽出法** は、最も基本的な抽出方法である。

**単純無作為抽出の方法**

1. 大きさ $N$ の母集団の各抽出単位に、$1, \ldots, N$ の番号をつける。
2. 1 から $N$ の間の乱数を一つ作る。
3. 「その乱数」番目の抽出単位が、まだ選ばれていなければ標本とする。
4. ステップ 2. とステップ 3. を、あらかじめ定めておいた標本サイズ $n$ の抽出単位が得られるまで繰り返す。

単純無作為抽出法は、確率抽出法の基本である。そのため、本ハンドブックでは、推定量 (8.1 節参照) やその標準誤差の算出式 (8.2 節参照)、さらに標本サイズを決めるための早見表 (8.3 節参照) では、単純無作為抽出法を前提としている。

単純無作為抽出法では、少なくとも $n$ 回は乱数に従って順番を数えなければならない。そのため、母集団が小さかったり、抽出台帳がコンピュータ上で整理されている場合に主として用いられる。

実際には、系統抽出法 (7.2 節参照) が用いられることが多い。

**ワンポイント アドバイス**

ステップ 3. で既に選ばれたかどうかを気にせず、ステップ 2. とステップ 3. を $n$ 回繰り返す方法を **復元抽出法** と呼ぶ。これに対し、本文中の方法を **非復元抽出法** と呼ぶ。復元抽出法では、重複のため、最終的に標本サイズが $n$ になるとは限らない。そこで、普通は非復元抽出法が用いられる。

## 7.2　抽出方法の基本（その２）—系統抽出法—

**系統抽出法**（あるいは**等間隔抽出法**）は基本的な抽出方法の一つで、調査で実際によく用いられる。

**系統抽出の方法**

1. 母集団の各抽出単位に、$1, \ldots, N$ の番号をつける。
2. 1 から $N$ の間の乱数（**スタート番号** と呼ぶ）を一つ作り、スタート番号の抽出単位を標本とする。
3. 標本として選ばれた抽出単位から数えて $D$（これを **抽出間隔** と呼ぶ）番目の抽出単位を標本として選ぶ。もし最後の抽出単位まできたら、抽出台帳の最初に戻って数え続ければよい。
4. 3. を、$n-1$ 回繰り返し、最終的に大きさ $n$ の標本を得る。

系統抽出法の利点は、抽出台帳 (6.1 節参照) の上での並び順を工夫することで、単純無作為抽出法 (7.1 節参照) に比べ、標本誤差 (1.5 節参照) を小さくできることである。

例えば、図 7.1 で調査対象が地域ごとにまとまって並んでいるとする。単純無作為抽出法では、乱数によっては左の方の（地域の）調査対象ばかり、あるいは右の方の（地域の）調査対象ばかりが標本として選ばれる可能性がある。これに対し、系統抽出法では、全ての地域からまんべんなく標本を選ぶことができる。

**ワンポイント アドバイス**

住民を調査対象とした調査では、選挙人名簿や住民基本台帳を抽出台帳とすることが多い。このとき、抽出間隔を偶数とすると、得られた標本が男性あるいは女性に偏ってしまうことがある。これは、台帳での並びが、男・女・男・女・男・・・となっていることがあるためである。このため、抽出間隔は奇数とすることが多い。

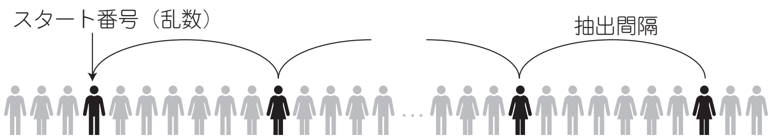



図 7.1: 系統抽出法

## 7.3　グループを選び出す—集落抽出法—

母集団がいくつかのグループ（**集落** と呼ぶ）に分かれている場合には、集落を抽出し、選ばれた集落に含まれる全ての調査対象を標本とする **集落抽出法** がある。

例えば、ある地域の児童・生徒を調査対象とする調査を行いたいものとする。子どもを直接、単純無作為抽出する代わりに、その地域の学校をいくつか抽出し、選ばれた学校に通う児童・生徒全員を標本とするのが集落抽出法である。この場合、抽出単位 (6.1 節参照) は児童・生徒ではなく、学校となる。

集落の抽出は、単純無作為抽出法や系統抽出法などによる。

集落抽出法は、調査対象の台帳が集落ごとに整理されている場合によく用いられる。集落抽出法では、全ての調査対象の台帳は不要であり、標本として選び出された集落内の台帳さえあればよいからである。

例えばある地域の児童・生徒を調査対象とするときに、全ての児童・生徒が並んだ名簿は用意できなくとも、各学校には在籍児童・生徒の名簿が整理されている。また、ある地域の公民館利用者を調査対象とする場合、その地域の利用者全体の名簿はなくとも、各公民館には利用者の名簿があるであろう。

集落抽出法では、一般に、同じ標本サイズの単純無作為抽出法に比べ、標本誤差 (1.5 節参照) は大きくなる。単純無作為抽出法の場合と同じ標本誤差を得るには、標本サイズをより大きくする必要がある。つまり、標本サイズを決めるための早見表 (8.3 節参照) にある人数よりも標本サイズを大きくする必要がある。

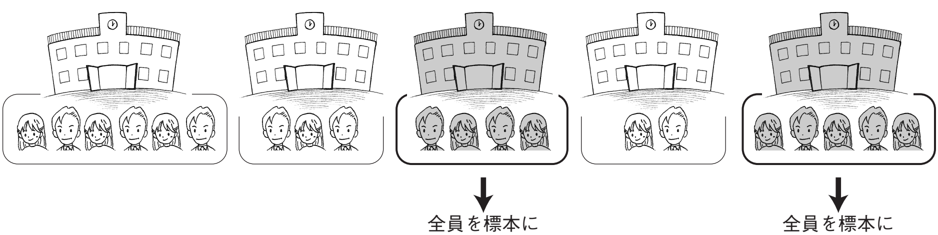



図 7.2: 集落抽出法

## 7.4　選んだグループごとにさらに選び出す (その 1 )—多段抽出法—

**多段抽出法** は、選び出された集落 (7.3 節参照) ごとに、その集落の中からさらに標本を選び出していく方法である。

集落抽出法 (7.3 節参照) では、例えば集落として学校を選び出し、その学校に在籍する全ての児童・生徒を標本とする。一校あたりの在籍数が非常に多く、全員を調査するのは難しい場合、選ばれた学校の中で、標本とする児童・生徒をさらに選び出すことがある。このような抽出方法は、**二段抽出法** と呼ばれる。

多段抽出法では、一段目の抽出単位 (6.1 節参照) のことを **第一次抽出単位**（**PSU**；Primary Sampling Unit）と呼ぶ。また、二段目の抽出単位を **第二次抽出単位**（**SSU**；Secondary Sampling Unit）と呼ぶ。さらに、最終段の抽出単位を **最終抽出単位**（**USU**；Ultimate Sampling Unit）と呼ぶ。先の例では、学校が PSU であり、児童・生徒が SSU かつ USU である。

各段の抽出単位の抽出は、単純無作為抽出法や系統抽出法などによる。

多段抽出法は、非常に大きな母集団に対して用いられることが多い。例えば、日本全国の個人を母集団とする場合、一段目では市区町村を選び出し、二段目では選ばれた各市区町村ごとに個人を選び出す、などといった二段抽出法がよく用いられる。

多段抽出法では、一般に、段の数が多くなるほど標本誤差 (1.5 節参照) は大きくなる。例えば、単純無作為抽出法（つまり一段抽出法）よりも、二段抽出法の方が標本誤差は大きい。日本全国から標本を選び出すよりも、選ばれた一部の市区町村からのみ標本を選び出す方が、標本誤差が大きくなることは、容易に想像できよう。

そのため、多段抽出法では、不必要に段の数を増やさないようにするのがよい。

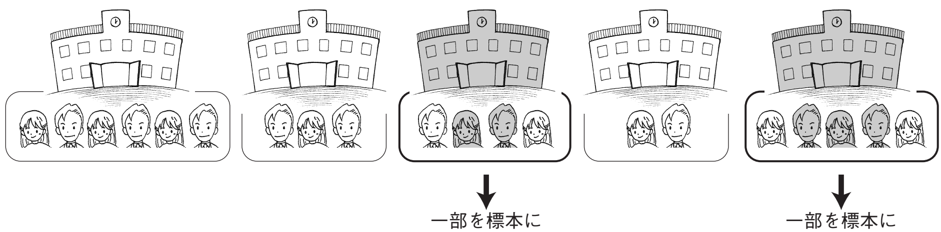



図 7.3: 多段抽出法

## 7.5　選んだグループごとにさらに選び出す（その２）──確率比例抽出法──

多段抽出法 (7.4 節参照) では、普通、自動加重標本 (6.1 節参照) が得られず、標本の平均値がそのまま母集団平均の推定値 (8.1 節参照) とはならない。そこで、各抽出単位 (6.1 節参照) の包含確率 (6.2 節参照) を調整することで、自動加重標本を得る方法が、 **確率比例抽出法** である。

つまり、確率比例抽出法を用いると、多段抽出法の場合にも、自動加重標本が得られ、標本の単純集計値がそのまま母集団平均の推定値となる。

例えば、三つの学校があり、各学校の在籍数は、学校Ａが二人、学校Ｂが三人、学校Ｃが五人であるとする。また、一段目として一つの学校を選び、二段目としてその学校から二人の生徒を選ぶことにする。学校Ａの生徒は、学校Ａが選ばれれば必ず標本として選ばれるのに対し、学校Ｃの生徒は、学校Ｃが選ばれても標本として選ばれるとは限らない。

仮に、一段目の学校を等確率（1/3 の確率）で選ぶとすると、各学校の生徒が最終的に標本となる確率（包含確率）は、

学校Ａ　：　$1/3 \times 2/2 = 1/3$

学校Ｂ　：　$1/3 \times 2/3 = 2/9$

学校Ｃ　：　$1/3 \times 2/5 = 2/15$

となり、学校Ａの生徒ほど標本として選ばれやすいことになってしまう。そのため、母集団の推定値を得るためには、標本の結果を重みづける必要がある (16.2 節参照)。

そこで、学校を等確率で選ぶのではなく、その大きさ（在籍生徒数）に応じて、大きな学校ほど選ばれやすく、小さな学校ほど選ばれにくくする。具体的には、学校Ａの包含確率を 2/10、学校Ｂの包含確率を 3/10、学校Ｃの包含確率を 5/10 とすると、各学校の生徒が標本として選ばれる確率は、

学校Ａ　：　$2/10 \times 2/2 = 2/10$

学校Ｂ　：　$3/10 \times 2/3 = 2/10$

学校Ｃ　：　$5/10 \times 2/5 = 2/10$

となり、どの学校の生徒も等しくなる。そのため、標本の平均値がそのまま母集団平均の推定値となる。

包含確率を、抽出単位の大きさに比例させることから、確率比例抽出法という名称がある。

実際に、確率比例抽出を行う際には、抽出単位の大きさを累積し、そこから系統抽出法 (7.2 節参照) によって標本を選び出すことが多い。

## 7.6　どのグループからも選び出し、誤差を減らす——層化抽出法——

**層化抽出法**（**層別抽出法**）は、母集団をいくつかのグループ（**層**と呼ぶ）に分け、それぞれの層において標本を選び出す方法である。層は、互いに重なり合うことがなく、全てを合わせると母集団に一致しなければならない。

例えば、個人を調査対象とする場合、性別を層として母集団全体を男性と女性とに分け、男性からも女性からも標本を選び出す方法や、公民館を選び出すために、設置年度が古い公民館と新しい公民館とに分け、それぞれから標本を選び出す方法が層化抽出法である。

さらに、日本全国の個人を調査対象とする場合、まず全国の市区町村をその規模によって層化し、それぞれの規模の中で二段抽出 (7.4 節参照) を行う方法は、層化二段抽出法と呼ばれる。

各層に割り当てる標本サイズを決める方法としては、どの層も等しくする方法（同数割当）や、層の大きさに比例させる方法（**比例割当**）などがある。比例割当とすると、自動加重標本 (6.1 節参照) が得られる。さらに、標本誤差を最小とするための**ネイマン割当**（**最適割当**）もある。

層化抽出を行うと、層化を行わないときよりも、標本誤差 (1.5 節参照) が小さくなることが多い。そのため標本抽出では、できる限り層化を行うのがよい。特に、層内では調査対象の回答が類似し、異なる層の間では調査対象の回答も大きく異なる、というような層化を行うと、標本誤差を小さく抑えることができる。

また、例えば男女別の推定を行いたいときなど、あらかじめ男女で層化し、各々必要な標本サイズを確保しておくとよい。

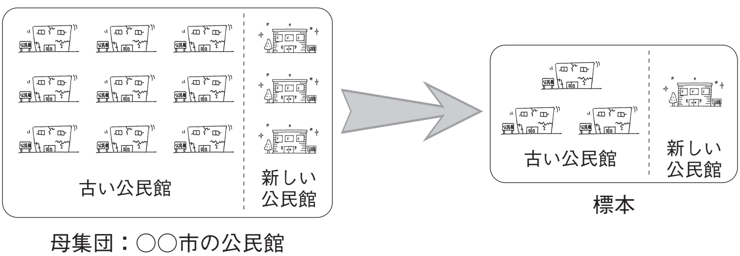



図 7.4: 層化抽出法

## 7.7　前回調査の標本から選び出す—二相抽出法—

**二相抽出法** は、他の調査の標本の中から、さらに少数の標本を選び出す方法である。一般に、標本の中からさらに標本を選び出すことを繰り返す抽出方法を **多相抽出法** と呼ぶ。

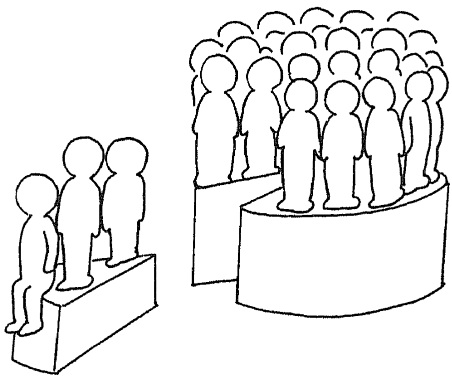
二相抽出法は、標本抽出のための情報があらかじめ十分に得られていない場合によく用いられる。つまり、一相目の情報を利用することで、二相目において、標本サイズは小さくとも標本誤差の小さな標本を得るのである。

例えば、最初に大きな標本サイズで簡便な調査を実施する。その結果を基に標本を層化 (7.6 節参照) して二相抽出を行い、より小さなサイズの標本に対して、詳細な調査を実施する。

また、時系列的な変化を調べるため、調査を毎月実施することにしたとする。毎回、大規模な調査を実施することはコストがかさむので、まず一年に一回大規模な調査を実施する。その結果を基に標本を層化した上で、毎月、標本を二相抽出する。これにより、各回の標本サイズを小さくし調査実施コストを抑えられると同時に、標本誤差の増大も抑えることができる。

多段抽出法 (7.4 節参照) と多相抽出法との形式的な違いは、多段抽出法では、抽出単位がそれぞれの段によって異なる（一段目は学校、二段目は個人など）のに対し、多相抽出法では、それぞれの相によって必ずしも異なるとは限らない（一相目は個人、二相目も個人など）という点である。

より本質的な違いは、多相抽出法では、ある相の調査結果を利用して、次の相の抽出を行う、という点である。したがって、例えば単純無作為抽出 (7.1 節参照) した標本の中から、さらに単純無作為抽出することは、形式的には二相抽出と言える。しかし一相目で選び出した標本の情報を二相目の抽出で用いないので、結果的には単純無作為抽出を一度行うことと同等である。



## 7.8　どの抽出方法がよいのか

抽出方法を決めるにあたっては、利用できる抽出台帳 (6.1 節参照) とコストを考慮し、その中で最も標本誤差が小さくなる抽出方法とするのがよい。

層化抽出法は、ほとんどの場合、標本誤差を小さくするので、利用するのがよい。層の数はなるべく多い方がよいが、標本サイズに対してあまり多くなりすぎると、比例割当 (7.6 節参照) ができず自動加重標本 (6.1 節参照) ではなくなるため、注意が必要である。

母集団が大きく、抽出台帳が集落ごとに整理されている場合には、多段抽出法を用いることになる。各 PSU (7.4 節参照) の大きさが分かっていれば、自動加重標本を得るために、確率比例抽出法を用いるとよい。多段抽出法では、段の数が多くなると標本誤差が大きくなるため、なるべく段の数は減らすのがよい。また、最終的な標本サイズが同じであれば、抽出する PSU の数を増やし、各 PSU 内で抽出する SSU の数を減らす方が、標本誤差は小さくなる。

自然な集落があり、集落内の全てを調査できるのであれば、単純無作為抽出を行うよりも集落抽出を行う方がよいことが多い。標本サイズが同じであれば、集落抽出法よりも単純無作為抽出法の方が標本誤差は小さくなる。しかし、集落を無視して単純無作為抽出を行うよりも、集落を利用した方が、少ないコストで標本サイズを大きくできることが多いからである。ただし、標本誤差を小さくするためには、なるべく一つの集落の大きさは小さくし、抽出する集落の数を多くするのがよい。

抽出台帳が整理されている場合には、単純無作為抽出よりも系統抽出を用いる方がよいことが多い。ゆるやかな層化の効果が得られ、標本誤差が小さくなるからである。その際、抽出間隔 (7.2 節参照) はなるべく大きくとるのがよい。

以上の抽出方法は、組み合わせて用いることが多い。例えば、日本全国の個人を調査対象とする場合、まず全国の市区町村を人口規模等により層化する。次に、各層内で、国勢調査区を確率比例抽出する。最後に、選ばれた調査区内で個人を系統抽出する。この抽出方法は、層化二段系統抽出法と呼ばれる。

# 第8章　標本サイズを決めるには

## 8.1　推定量とは

社会教育調査の目的は、母集団における平均値や総計値（これらを **母集団特性値** と呼ぶ）を知ることである。全数調査では、調査によって母集団における全ての調査対象の値が分かるから、ただちに母集団平均値や母集団総計値を求めることができる。

一方、標本調査では、母集団から選ばれた標本の値しか分からないため、母集団特性値を推定する必要がある (1.4 節参照)。標本の値を基に母集団特性値を推定する方法を **推定量** と呼び、その方法に従って推定し、得られた具体的な値を **推定値** と呼ぶ。

推定量は、標本抽出方法により異なる。ここでは、標本が単純無作為抽出法により得られた自動加重標本 (6.1 節参照) である場合の推定量を紹介する。

- **母集団比率の場合**：標本における比率が、そのまま母集団における比率の推定量となる。

$$\text{母集団比率の推定量} = \text{標本比率}$$

例えば、標本における比率が 25%であれば、母集団推定値も 25%である。

- **母集団平均値の場合**：標本における平均値が、そのまま母集団における平均値の推定量となる。

$$\text{母集団平均値の推定量} = \text{標本平均値}$$

例えば、標本における 1 日の学習時間の平均が 1.2 時間であれば、母集団推定値も 1.2 時間である。

- **母集団総計値の場合**：標本平均値に母集団サイズ（人数）を乗じた値が、母集団総計値の推定量となる。

$$\begin{aligned}&\text{母集団総計値の推定量}\\ &\quad = \text{母集団サイズ} \times \text{標本平均値}\end{aligned}$$

例えば、1 公民館あたりの平均事業数が 2.3 であり、母集団における公民館の数が 50 であれば、母集団における総事業数の推定値は $50 \times 2.3 = 115$ となる。

## 8.2　標準誤差とは

標本調査では、選ばれた調査対象が標本に応じて推定値は異なってくる。この違い・バラツキのことを標本誤差と呼ぶ (1.5 節参照)。標本誤差の大きさは、一般に推定量の標準偏差（これを **標準誤差** と呼ぶ）で表すことが多い。

選ばれる標本によって推定値はバラつくものの、統計理論から、その 95% は、

$$母集団特性値 \pm (1.96 \times 標準誤差)$$

の範囲に収まることが分かっている。

標準誤差の算出方法は、標本抽出方法および推定量により異なる。ここでは、単純無作為抽出法の場合の標準誤差を紹介する。なお、

| | | |
|---|---|---|
| $N$ | ： | 母集団サイズ（人数） |
| $n$ | ： | 標本サイズ（人数） |
| $p$ | ： | 母集団における比率 |
| $\sigma^2$ | ： | 母集団における分散 |

とする。

- 母集団比率の場合：

$$標準誤差 = \sqrt{\frac{N-n}{N-1}\frac{p(1-p)}{n}}$$

- 母集団平均値の場合：

$$標準誤差 = \sqrt{\left(1-\frac{n}{N}\right)\frac{\sigma^2}{n}}$$

- 母集団総計値の場合：

$$標準誤差 = N\sqrt{\left(1-\frac{n}{N}\right)\frac{\sigma^2}{n}}$$

例えば、母集団サイズ $N = 50,000$、標本サイズ $n = 1,000$、母集団における比率 $p = 0.4$ (40%) の場合、

$$\begin{aligned} 標準誤差 &= \sqrt{\frac{49,000}{49,999}\frac{0.4 \times 0.6}{1,000}} \\ &= 0.0153 \end{aligned}$$

である。したがって、仮に標本抽出と母集団推定を繰り返すと、そのたびに推定値はバラつくものの、その 95%は、

$$0.4 \pm (1.96 \times 0.0153) = 0.4 \pm 0.03$$

の範囲、すなわち 0.37〜0.43 の範囲に収まる。

標準誤差は、

- 標本サイズが大きいほど小さい
- 標本サイズの平方根にほぼ反比例する
- 比率の場合には、母集団における比率が $p = 0.5$ のとき、標準誤差が最も大きい

といった性質を持つ。

## 8.3　標本サイズを決めるための早見表

8.2 節では、推定値の 95%は、

母集団特性値 ± (1.96 × 標準誤差)

の範囲に収まると説明した。

逆に、調査結果を基にある推定値が得られたとき、

推定値 ± (1.96 × 標準誤差)

という範囲を考えることができる（この範囲を 95% **信頼区間** と呼ぶ)。当然、母集団特性値がこの範囲に含まれる場合もあれば、含まれない場合もある。しかし 95%の場合は、母集団特性値がこの範囲に含まれているということである。

標本サイズ（調査対象の人数）$n$ は、結果に必要な信頼区間の幅から逆算する。信頼区間の幅が狭い（精度が高い）推定値が必要であれば、標準誤差の大きさを小さくするために、 $n$ は大きくしなければならない (8.2 節参照)。信頼区間の幅が広くても（精度が低くても）よければ、$n$ は小さくてもよい。

ここでは、母集団における比率の 95%信頼区間を $\pm e$ に抑えたいものとする。

$N$ ： 母集団サイズ（人数）

$n$ ： 標本サイズ（人数）

$p$ ： 母集団における比率

とすると、標準誤差の式から逆算し、

$$n = \frac{N}{\left(\frac{e}{1.96}\right)^2 \frac{N-1}{p(1-p)} + 1}$$

が得られる。

なお、多くの場合、母集団における比率 $p$ は分かっていないため、上式において $p = 0.5$ とする。母集団サイズ $N$ や標本サイズ $n$ が同じときには、$p = 0.5$ のときに標準誤差は最も大きくなる。そのような誤差が最大の場合を想定しておけば、実際には誤差はそれよりも小さく抑えられるだろう、ということである。

表 8.1 は、母集団サイズ（$N$）と 95%信頼区間の幅（$e$）ごとに、必要な標本サイズ（$n$）を表にしたものである。

例えば、母集団が 10,000 人で、推定量の 95%信頼区間を $\pm 0.02 = \pm 2$ ポイント 以内に抑えたいときには、表の $N = 10,000$ と $e = 0.02$ の交わるところを見て、$n = 1,937$ となる。

同様に、$\pm 0.01 = \pm 1$ ポイント 以内に抑えたいときには、$n = 4,900$ となり、母集団の半分近くの人を標本としなければならないことが分かる。

表 8.1: 標本サイズを決めるための早見表

| | 95%信頼区間の幅（$e$） | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 母集団 ($N$) | 0.10 | 0.05 | 0.04 | 0.03 | 0.02 | 0.01 | 0.005 | 0.001 |
| 100 | 50 | 80 | 86 | 92 | 97 | 99 | 100 | 100 |
| 200 | 66 | 132 | 151 | 169 | 185 | 196 | 199 | 200 |
| 300 | 73 | 169 | 201 | 235 | 267 | 291 | 298 | 300 |
| 400 | 78 | 197 | 241 | 292 | 343 | 385 | 396 | 400 |
| 500 | 81 | 218 | 274 | 341 | 414 | 476 | 494 | 500 |
| 1,000 | 88 | 278 | 376 | 517 | 707 | 906 | 975 | 999 |
| 2,000 | 92 | 323 | 462 | 697 | 1,092 | 1,656 | 1,902 | 1,996 |
| 3,000 | 94 | 341 | 501 | 788 | 1,334 | 2,287 | 2,783 | 2,991 |
| 4,000 | 94 | 351 | 523 | 843 | 1,501 | 2,825 | 3,623 | 3,984 |
| 5,000 | 95 | 357 | 537 | 880 | 1,623 | 3,289 | 4,425 | 4,975 |
| 10,000 | 96 | 370 | 567 | 965 | 1,937 | 4,900 | 7,935 | 9,897 |
| 20,000 | 96 | 377 | 583 | 1,014 | 2,144 | 6,489 | 13,153 | 19,593 |
| 30,000 | 96 | 380 | 589 | 1,031 | 2,224 | 7,276 | 16,846 | 29,092 |
| 40,000 | 96 | 381 | 592 | 1,040 | 2,266 | 7,745 | 19,597 | 38,401 |
| 50,000 | 96 | 382 | 594 | 1,045 | 2,292 | 8,057 | 21,725 | 47,526 |
| 100,000 | 96 | 383 | 597 | 1,056 | 2,345 | 8,763 | 27,755 | 90,570 |
| 200,000 | 96 | 384 | 599 | 1,062 | 2,373 | 9,164 | 32,227 | 165,530 |
| 300,000 | 97 | 384 | 600 | 1,064 | 2,382 | 9,307 | 34,056 | 228,595 |
| 400,000 | 97 | 384 | 600 | 1,065 | 2,387 | 9,379 | 35,050 | 282,388 |
| 500,000 | 97 | 384 | 600 | 1,065 | 2,390 | 9,424 | 35,676 | 328,815 |
| 1,000,000 | 97 | 385 | 600 | 1,066 | 2,396 | 9,513 | 36,995 | 489,901 |
| 2,000,000 | 97 | 385 | 601 | 1,067 | 2,399 | 9,559 | 37,693 | 648,832 |
| 3,000,000 | 97 | 385 | 601 | 1,067 | 2,400 | 9,574 | 37,931 | 727,503 |
| 4,000,000 | 97 | 385 | 601 | 1,067 | 2,400 | 9,581 | 38,051 | 774,454 |
| 5,000,000 | 97 | 385 | 601 | 1,067 | 2,400 | 9,586 | 38,124 | 805,651 |
| 10,000,000 | 97 | 385 | 601 | 1,067 | 2,401 | 9,595 | 38,269 | 876,246 |