

第1章 統計用語－2

センサスと標本調査

<いみ>

統計調査で、対象をすべて調査する全数調査（^{しっかい}悉皆調査）のことをセンサスという。一方、対象となる集団全体（母集団）から一部（標本、サンプル）を抽出して調査を行い、調査結果に基づいて母集団の特性を推定する調査方法を標本調査という。

<たとえば>

表 センサスと標本調査の例

区分	調査名	周期	調査対象	標本の選び方(例)
センサス	国勢調査	5年	10月1日午前0時現在で、日本に3か月以上住んでいる人及び3か月以上にわたって住むことになっている人。	
	経済センサス-活動調査	5年	以下に掲げる事業所を除く全国全ての事業所及び企業 <ul style="list-style-type: none"> ・国・地方公共団体の事業所 ・日本産業分類に基づく次の事業所 <ul style="list-style-type: none"> ・農業、林業に属する個人経営の事業所 ・漁業に属する個人経営の事業所 ・生活関連サービス業、娯楽業のうち、家事サービス業に属する事業所 ・サービス業のうち、外国公務に属する事業所 	
標本調査	労働力調査	毎月	日本に住む15歳以上の者及びその世帯	【層化2段抽出法】 ① 第1次抽出（調査区の抽出）：国勢調査の調査区を特性により層化し、各地域の各層毎に所定の抽出率により抽出 ② 第2次抽出（住戸の抽出）：1次で抽出された標本調査区の全ての住戸から、1調査区当たりほぼ15戸となるように所定の抽出率により住戸を抽出

<かんどころ>

1. センサス(全数調査)の特徴

センサスは、母集団全体を対象とするので、精度の高い集計結果が得られる。また、母集団の全体構造を明らかにするので、標本調査の標本抽出の際や、標本調査の結果から全体を推定する際の資料にもなる。反面、大きな費用と労力、時間を要する。

2. 標本調査の特徴

標本調査は、センサスに比べ、対象数が少なくて済み、調査のための労力、費用、時間が節約できる。反面、標本数が少ないと結果の信頼性に問題が残る。従って標本抽出の適否が調査に大きく影響する。国の基本的な統計を所管する総務省の基幹統計14統計のうち、全数調査により作成されているのは僅か3統計であり、主要な統計調査でも標本調査は大きなウエイトを占めている。

3. 標本抽出の方法

無作為に標本を抽出し、客観性があるといわれる無作為抽出法と、知識・経験を活かして抽出する有意抽出法がある。無作為抽出法には主に次のものがある。

(1) 層化多段抽出法

よく用いられている方法であり、層化抽出法と多段抽出法を組み合わせたもの。

層化抽出法：母集団をあらかじめいくつかのグループ(層)に分け、各層から必要な標本数を無作為に抽出する方法。各層内はできるだけ等質となる分け方が望ましい。層化の目的は少ない標本数で高精度の母集団推定を行うことである。

多段抽出法：複数段階に分けた抽出方法。例えば、1段階で市町村を、2段階で市町村の中の地区を、3段階で地区の中の世帯を抽出すれば、3段階抽出となる。抽出段数が増し推定精度の劣化を防ぐため、層化抽出法と併用し、層化多段(2~3段が多い)抽出法として用いられることが多い。<例：労働力調査(層化2段階抽出法)家計調査・個人企業経済調査(層化3段階抽出法)、毎月勤労統計調査(第二種事業所)(層化2段階抽出法)>

(2) 集落抽出法

調査対象の集まり(集落)を抽出単位とし、抽出された集落をすべて調査する方法。散在する対象を絞って調査するため、少ない費用で済むことが多い。

<例：毎月勤労統計調査特別調査(層化集落抽出法)>

(3) 系統抽出法

母集団から一定間隔で抽出する方法。抽出率50分の1の場合、1~50から数値(仮に23)を選び、以下逐次50を加えた番号(73, 123, ...)を抽出する。作業が簡単で、リスト配列の工夫により精度の向上を図ることができる。<例：学校保健統計調査>

4. 母数の推定と標本誤差

標本調査の目的は、そこから得られる情報から全体(母集団)についての何か(平均等)を推定することである。しかし、標本は無作為に抽出された母集団の一部でしかなく、誤差(標本誤差)を避けることはできない。母集団の平均(母平均) μ を推定したいとしよう。標本の平均(標本平均)と μ が等しいという保証はない。そのため、標本誤差がどの程度あるかに気を配る必要がある。標本誤差の指標としては、標準誤差がよく用いられており、標本平均に関しては、

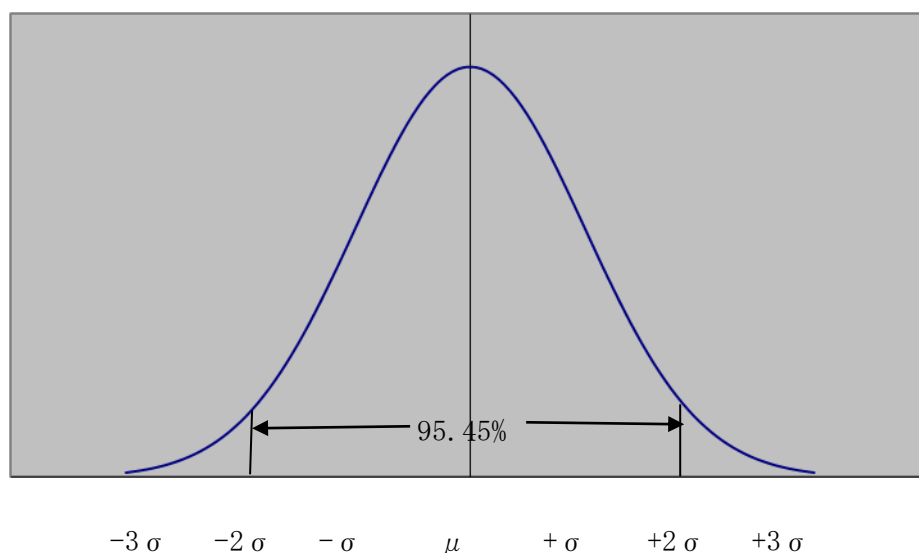
$$\text{標準誤差} = \text{母集団の標準偏差 } \sigma \div \sqrt{\text{標本の大きさ } n}$$

により求められる。また、標準誤差÷標本平均を誤差率という。

母集団の分布が正規分布でないことも多いが、標本が大きければ(n が大きければ)、標本平均の分布は正規分布と見なせる(中心極限定理。その平均は μ 、標準偏差は標準

誤差に等しい)。普通は計算される標本平均は1つだが、仮想的に何回も標本調査を行った場合に得られる多くの標本平均の分布が、ここでいう「標本平均の分布」である。標本平均の分布が正規分布であれば(上述のように実際そうみなせる)、図のように(母集団平均 μ) $\pm 2 \times$ (標準誤差)の範囲に、仮想的に得られる多くの標本平均のうちの95.45%が入る。言い換えれば、手元に1つだけ得られている標本平均は、この範囲に0.9545の確率で入る。逆に、推定したい母集団の平均 μ は0.9545の確率で、(標本平均) $\pm 2 \times$ (標準誤差) = (標本平均) \times { $1 \pm 2 \times$ (誤差率) } の範囲に入る。標本誤差のために誤りを犯すことは避けられないが、どの程度の確率で誤りを犯すかを把握しておくことが重要である。

図 正規分布表



「第1章 統計用語-3」平均と標準偏差、変動係数を参照