

統計分析シリーズ (II)

茨城大学教授 所 一 夫

II 標本調査

1. 標本の抽出

前回で、推計学が実際の統計調査に重要な意味を持つ1つは、一部調査によって全体を推知することができる事であると述べた。しかしその判断は確率判断であるから、その推定ができるためにはそこに集められた資料に対して確率が用いられるようになっていなければいけない。そのためには統計調査に先だって「どのように標本を抽出するか」という事を考えなければならない。

(i) フレームの設定

元来標本によって統計調査をする場合には、まず調査しようとする母集団が明確にはあくされていなければならない。たとえば市町村の世帯名簿とか選挙有権者名簿などのように台帳によってその母集団を明示しているものでなければ標本調査はできないのである。このように母集団を明確に示したものをフレームといっている。現在のように標本調査万能と見られている中で多額の費用と時間を費して国勢調査其の他の全数調査を行わなければならない1つの理由は、逆説らしく聞えるが、標本調査を可能ならしめるためであるといえることができる。

(ii) 任意抽出法 (ランダムサンプリング)

母集団から、どの単位も等確率で抽出されるような抽出法をランダムサンプリングといっている。これは公平なくじ引きによって標本を抽出する方法であるが、これを実際に行なうには乱数表とか乱数サイが用いられる。

乱数表

これは0から9までの数字の中からくじ引きの原理によって1つの数字を抽出してそれを記録し、次にまた0から9までの数字の中から同様にして1つの数字を抽出して記録し、以下同様にして次々と得られた数字を並べたものであり、次にその一部分の例を示す。

72 84 71 14 35 19 11 58 49 26 50 11 17 17 76
86 31 57 20 18 95 60 78 46 75 88 78 28 16 84
13 52 53 94 53 75 45 69 30 96 73 89 65 70 31

乱数表は統計数値表には必ず付いており、また多くの統計学の本には付表として付けてある。

次にこの表の用い方を例によって示そう。

例 大きさ70の母集団から大きさ10のランダムサンプリングによる標本を乱数表によって求めて見よう。

- (a) 母集団の各単位に1から70までの番号を付ける。
(フレームの確定)
 - (b) 母集団の大きさ70が100以下であることより、乱数表の数字を2つずつに区切っておく。(もし母集団の大きさが100より大で1000以下なら3つずつに区切る。)
 - (c) 表の任意の区切り(たとえば目を閉じて鉛筆の先を勝手に当てた場所)から2けたの数を読んで(02のように10の位が0の場合も2けたと見る)その数を番号とした母集団の単位を抽出する。もしその数が70より大であればその数ば捨てて乱数表のその次の区切りの数を番号とする母集団の単位を標本として抽出する。その数も70より大であれば次の区切りの数をとる。
 - (d) いま抽出した2けたの数の次の2けたの数に対応する番号の単位を抽出する。(もしこの番号が既出のものまたは70以上であればその数は捨てて次の区切りに対応する単位を抽出する。)
- このようにして次々と10個の単位を抽出すればこの標本はランダムサンプリングによる標本となる。

乱数サイ

これは正20面体の各面にそれぞれ0から9までの数字を2回ずつ書き入れたもので(日本規格協会から3個を1組として売り出されている)これを投げて上面に表われる数字を順次に記したものが前述の乱数表と同じものと考えられるように作ったものである。

したがってこれによって前例の標本を抽出するたは、70が2けたの数であることからこの乱数サイを2個用いる。その中の一方(色分けになっている)を10位の数字、他方を1位の数字として毎回この2個のサイを投げて2けたの乱数列を作り、これを乱数表の数列と見て前の場合と同様に2けたの数を番号とした母集団の単位を標本として抽出して行けばよい。

(iii) その他のサンプリング

母集団の大きさが非常に大である場合にはランダムサンプリングによる標本抽出が困難となるので、母集団を適当に分類しておき、全体の中から各類をランダムサンプリングによって抽出するとか、またその類の中から調査単位をランダムサンプリングによって抽出するとかなどが用いられる。これらの場合にもその基本原理となるものはそのランダム性によって出て来た結果に確率判断を用いようとするものであるが、ここではその細部に入る事はやめて、ランダムサンプリングにより抽出された標本からどのようにして母集団の性質が推定されるのか

その様子を明らかにしてゆこうと思う。

2. 平均と標準偏差

(i) 平均の意味

母集団の持つ特性は適当に数量化することにより、(a)その総和Tを求めること、(b)その平均Mを求めること、(c)その相対比率Pを求めることにまとめられる。そしてこの(a)(b)(c)については母集団のフレームが確定して居ればその大きさNがわかっているから、平均がわかれば総和は $T=MN$ として直ちに算出できる。

また比率Pは、その属性を持った単位には1の数値を持たせ、その属性を持たない単位には0の数値を持たせると、それらの数値の平均Mが比率Pを示すことになる。

たとえば7人のうち男子3人の場合には男子を1、女子を0としてこれらの(1,1,1,0,0,0,0)の平均を求めると

$(1+1+1+0+0+0+0) \div 7 = 3/7$ となり、これは全体に対する男子の比率を示している。

以上より私達は母集団の平均(これを母平均という)で母集団の性質が表わされるものと考えてゆく。

(ii) 標準偏差

母平均によって母集団の特性がつかめるとしても、それは1つの数字であるから、次に問題になるのは母集団で各単位の持つ数値 (x_1, x_2, \dots, x_n) がどのように散布されているかという事である。この程度を示す1つの方法が標準偏差(これをSで表わす)である。これは平均をMすなわち $M = (x_1 + x_2 + \dots + x_n) \div N$ とするとき

$$S = \sqrt{\frac{(x_1 - M)^2 + (x_2 - M)^2 + \dots + (x_n - M)^2}{N}}$$

によって示される。実際計算では上式を少し変形して

$$S = \sqrt{\frac{x_1^2 + x_2^2 + \dots + x_n^2}{N} - M^2}$$

とした方がよい。

例 ある町でコシヒカリで炊いた米飯の味について各地区から24人を選んで試食をしてもらい、各人に-3, -2, -1, 0, 1, 2, 3の評点をつけてもらった。A地区24人の評点は次のとおりであった。この場合の評点の平均と標準偏差を求めて見よう。

評点	-1	0	1	2	3	計
人数	4	1	6	11	2	24

$$M = (-1 \times 4 + 0 \times 1 + 1 \times 6 + 2 \times 11 + 3 \times 2) \div 24 = 1.25$$

$$S^2 = \frac{(-1)^2 \times 4 + 0^2 \times 1 + 1^2 \times 6 + 2^2 \times 11 + 3^2 \times 2}{24}$$

$$= 1.25^2 = (72/24) - 1.5655 = 1.4375,$$

$$\text{故に } S = \sqrt{1.4375} = 1.2$$

3. 標本平均の分布

大きさNの母集団から大きさnの標本を抽出した場合に、その標本の各単位の持つ数値 (x_1, x_2, \dots, x_n) の平均を標本平均と言い次式のように \bar{x} でこれを表わす。

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

この \bar{x} は同じ母集団から抽出された標本の平均であつても、一般には各標本ごとにその値は異なるものであろう。しかしこれらに関して数理統計学はこれらの標本平均の分布が次の法則に従っていることを教えている。

(i) 標本平均は標本のとり方によって種々異なった値をとるがそれらの値の平均は母平均Mに等しい。

(ii) またこれらの種々の標本平均の標準偏差(これを標本標準偏差と言い $s(\bar{x})$ で表わす)は母集団の標準偏差をSとすると次のようになっている。

$$s(\bar{x}) = \sqrt{\frac{N-n}{N-1}} \frac{S}{\sqrt{n}}$$

実際問題ではnに比べてNが十分大となっているので $\sqrt{(N-n)/(N-1)}$ はほとんど1に近く

$$s(\bar{x}) = \frac{S}{\sqrt{n}}$$

と見積る場合が多い。

(iii) \bar{x} の分布はnが大きいか場合には(勿論Nはそれよりもはるかに大きい)正規分布と言われる分布と見てよい。この(iii)の性質を中心極限定理と言っている。

正規分布については次号で述べるつもりであるが、これらの法則は結局「ランダムサンプリングによって母集団の性質はどのように標本に反映されるか」を述べたものであり、「身体の実態はX線でフィルムにどのように写されるか」がわがればX線のフィルムを見て身体の実態が推知できる事と同じものである。

そうすると次の問題はこの3つの法則があればどうして標本調査の結果から母平均が推定できるのかをしらべる事であり、またその実際の解析の様子を示すことである。

統計ニュース

○2日～3日 県民所得推計業務研究会

○6日～7日 統計職員基礎研修

○9日～10日 統計実務講習会

○20日～21日 工作機械設備等統計調査ブロック会議

○23日～24日 漁業センサス本調査全国会議