

## 標本調査を使用する際、注意すべき点について

行政管理庁 統計審査官 田 島 正

### 綿密な設計でランダム抽出を

現在、統計調査は標本調査と悉皆調査の方法で実施され、前者は経費面、時間面での効率性から多くの調査で使用されていることは周知の通りです。標本調査の場合、標本をランダム（無作為）に抽出することが前提になっているのですが、ランダムに抽出すればそれでよいと云うのではなく綿密な設計の下になされなければなりません。

例えば或る市で、市民の市政に対する意見を調査しようとして、市役所の市民相談室に来た市民の中から無作為に対象者を選び出して、世論調査を実施したとします。その結果から市政に対する市民の声を適格にキャッチできたとしたら、それは判断を誤まることとなります。

この場合、キャッチできたのは、何にか相談事があって、自発的に市役所にきた人達の意見であって、市民一般の意見と云う訳けには行かないのです。

市民一般の意見をキャッチするならば、別の方法を取らねばなりません。通常取られている方法は市役所や市の出張所で保管している管轄地域の住民票などから対象者を抽出する方法です。

### 人口比例による標本配分のメリット

例について説明しましょう。A市の人口は10

万人で、市役所本庁で管轄している住民票は5万人で、他は第1出張所が3万人、第2出張所が2万人管轄しているとします。

この市で2,000人を対象者に世論調査を実施するとした場合、2,000人の対象者を夫々本庁、第1出張所、第2出張所から何人ずつ抽出すればよいかを考えますと、普通、各所で管轄している人数に比例して抽出すれば良さそうだと思うでしょう。事実、それはそれで良いのですが、それだけが唯一の、抽出すべき対象者の決め方ではありません。ただそれは後で述べる様に優れたメリットがあるのです。

それはそれとして、今述べた方法で対象者数を決定しますと本庁では $2000人 \times \frac{5万人}{10万人} = 1,000人$ 、第1出張所では $2000人 \times \frac{3万人}{10万人} = 600人$ 、第2出張所では $2000人 \times \frac{2万人}{10万人} = 400人$ となります。

抽出すべき対象者数が決まったなら、各所で保管している住民票などから、所定数の対象者を無作為に抽出することになります。

さて、対象者が決定して、世論調査が実施され、或る質問に対して肯定的に答えたものが、本庁管轄分では600人、第1出張所管轄分では300人、第2出張所管轄分では300人であったとします。

この場合、対象者にはすべて会え、回答拒否もなかったとしますと、この質問に肯定的に答えた

人は、全部で、 $600人+300人+300人=1,200$ 人で、対象者全体は、 $2,000$ 人ですが、この場合の回答の "yes" の率は、 $\frac{1,200}{2,000}=0.6=60\%$ となります。

この様に調査結果が計算されるのは、対象者数が各所の管轄人員数に比例して決定されているからであって、もし管轄人員数に比例して決定されていない場合は、この様な計算はできません。

例えば、本庁、第1出張所、第2出張所に、 $800$ 人、 $600$ 人、 $600$ 人計、 $2,000$ の対象者数を割り当て、各所に於ける肯定的な答をした人の数が、前の例と同じく、夫々 $600$ 人、 $300$ 人、 $300$ 人であったとしても、この場合は、調査結果から "yes" の回答率が、 $\frac{600+300+300}{2,000}=\frac{1,200}{2,000}=60\%$ とはなりません。

	本庁管内での調査結果	第1出張所内での調査結果	第2出張所内での調査結果	全市での調査結果
第1回の調査 (比例配分した場合)	$\frac{600}{1,000}=60\%$	$\frac{300}{600}=50\%$	$\frac{300}{400}=75\%$	$\frac{60\% \times 5万人 + 50\% \times 3万人 + 75\% \times 2万人}{10万人} = 60\%$
第2回の調査 (比例配分によらない場合)	$\frac{600}{800}=75\%$	$\frac{300}{600}=50\%$	$\frac{300}{600}=50\%$	$\frac{75\% \times 5万人 + 50\% \times 3万人 + 50\% \times 2万人}{10万人} = 62.5\%$
各所で管轄している人員	5万人	3万人	2万人	

表の如く、第1回の調査では、市全体の "yes" の比率は、 $60\%$ 、第2回のそれは $62.5\%$ と計算されます。第1回の $60\%$ と云う結果は先に、 $\frac{1,200}{2,000}=60\%$ と出した結果と一致していることに気づかれるでしょう。これは偶然一致した訳けではなく、第1回では各所で管轄している人員に比例して対象者数を割り当てた結果そうなったのです。

例えば、比例配分して対象者数を決めた時は、

これは例えば、ある学校の1学年にはA、B、Cの3組あり、それぞれの生徒数が $80$ 人、 $60$ 人、 $40$ 人であったとし、ある試験の結果、各組の平均点が夫々 $70$ 点、 $75$ 点、 $60$ 点であったとすると、学年の平均点は、各組の人数が異なるので、各組の平均点に人数を重味とした加重平均として計算されることを思い出せば判り良いでしょう。

この学年の平均点は

$$\frac{70 \times 80 + 75 \times 60 + 60 \times 40}{80 + 60 + 40} = \frac{12,500}{180} = 69.4$$

と計算されます。

この考え方から、世論調査の結果(肯定的に回答した者の比率)について類推しますと、

各所の肯定的に回答した人数を合計して、全対象者数で割れば、全市での肯定的に回答した者の比率が得られると云うことです。これを「単純集計が可能である」と云います。これが前にのべたメリットなのです。世論調査では、通常1回の調査で相当数の質問をしますが、対象者数を比例配分法で決めておけば、各質問についての回答の比率を単純集計で出せると云う訳けです。

これは集計作業の上から云えば大きな省力化と

云えましょう。

### 特定地域の

#### 標本精度をよくするには

さて、話を少し変えまして、標本調査の結果には、必然的に所謂標本誤差があると云うことは周知の通りです。これは今の例で云えば、第1回の調査で "yes" の回答率が $60\%$ と出たとしてもこれは $10$ 万人の市民の中から $2$ 千人を標本として取った結果であって、 $10$ 万人を全部調査した場合の結果とは当然食い違う筈で、この違いの程度を表わすものが標本誤差です。容易に考えられる様に、対象者数が増えれば標本誤差は少なくなります。

そこで今の例に戻れば、全市での "yes" の率は $2,000$ 人の対象者数で $60\%$ と出た訳けですが、 $2,000$ 人と云う数は対象者数として結構妥当な数です。と云うことは、 $2,000$ 人を対象者数とした時の標本誤差はそれ程大きくはないと考えられるからです。勿論この標本誤差の大きさを計算する方法はありますが、今はそれに深入りしないで話を進めます。さて、市全体の "yes" の比率に対する誤差はそれ程大きくないのは判かったとして、本庁、出張所別の "yes" の比率を、標本誤差を相当小さい所で知りたいと云う必要が起こった時、今の例について見るとどうでしょう。

本庁管内では $1,000$ 人の対象者で "yes" の比率が $60\%$ 、第1出張所管内では $600$ 人で $50\%$ 、第2出張所管内では $400$ 人で $75\%$ となっていますが、 $500$ 人未満の対象者数では、標本誤差が相対的に大きく第2出張所管内での肯定的な答の比率をもつと正確に知りたい時は、第2出張所に割り当てられる対象者数を増やさな

ればなりません。その時、前に述べた比例配分法による単純集計も可能の様にするにはどうすれば良いでしょうか。

例えば第2出張所に割り当てられた $400$ 人を倍にして $800$ 人にしたとします。勿論、第2出張所管内についての "yes" の比率についての精度は上がりますが、全市についての比率を単純集計によって出すにはどうすれば良いでしょうか。先ず考えられるのは本庁、第1出張所に割り当てられた対象者数を倍にすれば良いのではないかと云うことです。しかし、本庁、第1出張所に割り当てられた対象者数は $500$ 人以上あり、増やす必要はないし、それに経費的にも制約があるとしたらどうしますか。考えられることは、全市の比率を求める時は、(A) 第2出張所管内で後で追加した $400$ 人の回答は除いて、単純集計するか、又は、(B) 本庁、第1出張所で得られた回答を2倍にし、第2出張所の $800$ 人の回答と合せて単純集計するかと云うことでしょう。当然のことながら、方法が違えば、全市での比率も違ってきますが、どちらの方法が望ましいのでしょうか。今、第2出張所管内で追加した $400$ 人の対象者の中では肯定的な答をした人が $280$ 人いたとしますと、(A) の方法では全市の "yes" の率は勿論 $60\%$ となり、(B) の方法では

$$\frac{600 \times 2 + 300 \times 2 + (300 + 280)}{2,000 \times 2} = \frac{2,380}{4,000} = 59.5\%$$

となります。どちらかが推定値として正しいのかと云えば、どちらも推定値として正しいと云えます。標本調査の結果は、前に述べた様に標本誤差がつきもので、調査毎にその結果が違うのが通常と云えます。標本調査の結果に要求されるのは、標本理論で云う「不偏推定値」であることと云うことが第一ですが、その点から云うと、今、述べた2つの結果はともに「不偏推定値」であるのでど

らも正しいと云う訳けです。

どちらもその意味で正しいとして、どちらの結果がより精度が高いのかと云うと、それは、優劣がつき、容易に推測される様に、実際に調査された対象者の回答をより多く含んでいる(B)の方が良いのです。ここで精度が高いと云うことは、標本誤差がより小さいと云うことです。

なお、ついでに云いますと、前に述べた第2回の調査で"yes"の率が62.5%と出ましたが、この結果も不偏推定値です。しかし単純集計はできないと云う所が、(A)、(B)の方法とは違う所です。又(A)即ち第1回の調査と比べてどちらが精度がよいかと云うことは、ケース・バイ・ケースで一概に云えませんが、今の例で計算してみますと第2回の調査の方が僅かながら精度が良いと云う結果が出ます。

さて、前に戻って付け加えて置きますと、(A)と(B)とでは、(B)の方が精度が良いと云いましたが、これは対象者数即ち標本数が増えれば、精度が良くなることからはきているのですが、この場合注意しなくてはならないことは(B)の場合の対象者数は $2,000人 \times 2 = 4,000人$ ではなく $1,000人 + 4,000人 = 1,400人$ と云うことです。

## 標本数をきめる

### 目安としての変異係数

前に標本数が増加すれば精度が高くなると思いましたが、どの程度標本数をとれば良いかと云うことは、調査を企画した側で、標本誤差をどの程度にしたいかと関連し、一概に決められません。

1つの目安としては、標本調査の結果得られた推定値、今迄の例で云えば第1回の調査の"yes"の率の60%、第2回の62.5%と云ったものですが、これでもって、その時の標本誤差の大きさを示す量として用いられる推定値の標準偏差を割ったものが5%以下になることがあげられます。今出した量のことを変異係数と云いますが、上のことは言葉で云えば変異係数が5%以下であることが一応の目安であるとうことになります。

$$\text{変異係数} = \frac{\text{推定値(平均値)の標準偏差}}{\text{推定値(平均値)}} \leq 0.05$$

この目安は測定したい推定値の夫々について適用されるもので、市全体についての推定値については勿論のこと、先にふれた様に本庁、第1、第2出張所管内の夫々の結果についても相当の信頼度をもって得たい時は、今の目安が適用されることになります。その場合の推定値は各所に於ける推定値であり、全市のそれではなく、標準偏差も各所に於けるそれであることは云うまでもありません。