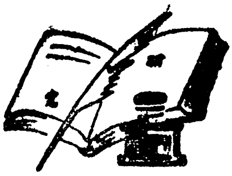

標本調査への手引 (8)

総理府統計局　高橋史朗

第2部　標本選定の技巧（つづき）

5　推定式と選びだす地域の数

問題(1)地域の選定の方法は，なお，解決しなければならない問題点を残しておりますが，ここでは，ひとまず，問題(2)選びだす地域の数へ，説明を進めてゆきたいとおもいます。

具体例による方が説明しやすいので，たとえば，この3月の従業員総数および1従業員当たりの平均売上高をもとめる場合をかんがえてみましよう，ここで，従業員総数は**算術平均**，また，1従業員当たりの平均売上高は**2個の算術平均の比率**の，それぞれ，一例になつております。

ところで，地域および飲食店の選定の方法ですが，それには，それぞれ，比例確率重複抽出および等確率非重複抽出が採られていると仮定します。実用からみると，比例確率系統抽出および等確率系統抽出の方が取り扱いやすいのですが，しかし，これらの方法には，理論からみて，非常に取り扱いにくい点があります。そこで，実際には，比例確率系統抽出および等確率系統抽出を採るが，理論では，比例確率重複抽出および等確率非重複抽出の結果を借りるというのが，一般のゆき方です。したがつて，この前者を採る場合には，それが，後者と同程度以上に，よく，その役割りを果たせるということを，まず，確かめておかなければなりません。

(1)　従業員総数をもとめる場合

従業員総数は，1次抽出単位である地域にたいして，算術平均とみなされますが，それは，従業員総数を，次のように

従業員総数＝1地域当たりの平均従業員数×地域総数

と表わしてみると，地域総数は既知であり，また，未知の1地域当たりの平均従業員数は算術平均であるからです。

これで明らかなように，従業員総数といつても，内容からいえば，そのうちの1地域当たりの平均従業員数だけを説明すれば十分なわけです。したがつて，ここでも，従業員総数ではなく，**1地域当たりの平均従業員数**をもとめる場合を説明してゆきたいとおもいます。

さて，まず，1地域当たりの平均従業員数の推定値ですが，それは，選びだされた飲食店の，3月の従業員数から，次の推定式によつて，もとめられます。

$$\text{従業員数の算術平均} \times \frac{\text{飲食店総数}}{\text{地域総数}}$$

この推定式が，$\frac{\text{飲食店総数}}{\text{地域総数}}$を除いて，第1部で説明した平均営業利益（算術平均の一例）の推定式と，まつたへ同じ形であることに注意してください。これは偶然ではなく，実は，推定式をこのような簡明な形にするために，逆に，地域や飲食店の選定の方法を，さきのように，決むたわけです。

次に，この推定値を実現値とする推定子ですが，いま，選びだす地域の数を ℓ として，それを $\overline{U}_\ell$ と表わしてみると，推定子 $\overline{U}_\ell$ の期待値E（$\overline{U}_\ell$）は，1地域当たりの平均従業員数Muに等しいという関係があります。すなわち，第1部第10節（9月号）の条件1は満たされているわけです。

$$E(\overline{U}_\ell) = Mu$$

また，標準偏差 $\sigma(\overline{U}_\ell)$，変動係数 CV（$\overline{U}_\ell$）は，従業員の地域間標準偏差σdx，地域内標準偏差σwx，地域間変動係数CVbx，地域内変動係数CVwx，ならびに，飲食店総数H，地域総数L，それぞれの地域の飲食店数の最大，最小間のある数N，それぞれの調査地域の調査飲食店数mによつて，次のように，

$$\sigma(\overline{U}_\ell) = \frac{1}{\sqrt{\ell}}\sqrt{\sigma_{bx}^2 + \frac{N-m}{N-1}\frac{\sigma_{wx}^2}{m}}\ \frac{H}{L}$$

$$CV(\overline{U}_\ell) = \frac{1}{\sqrt{\ell}}\sqrt{CV_{bx}^2 + \frac{N-m}{N-1}\frac{CV_{wx}^2}{m}}$$

と表わされますから，これを条件２に入れてℓについて解いてみると

$$\ell \geqq \left(\frac{\lambda}{\varepsilon}\right)^2\left(\sigma_{bx}^2+\frac{N-m}{N-1}\frac{\sigma_{wx}^2}{m}\right)\left(\frac{H}{L}\right)^2$$

$$\ell \geqq \left(\frac{\lambda}{\eta}\right)^2\left(CV_{bx}^2+\frac{N-m}{N-1}\frac{CV_{wx}^2}{m}\right)$$

となります。

そこで，いま，１地域当たりの平均従業員数を，たとえば，許容区間の相対誤差$\eta=5\%$，危険の確率$\lambda=2$でもとめることにしてみましよう，

まず，過去の資料から

$CVbx \fallingdotseq 0.27$

$CVwx \fallingdotseq 1.09$

ぐらいであり，また，それぞれの地域の飲食店数も，過去の資料から，最大で23，最小で11ぐらいであるので，もつとも安全をとつて，

$N = 23$

とし，次に調査員の活動能力をかんがえて，

$m = 8$

と決めると，選びだす地域の数ℓは

$$\ell \geqq \left(\frac{\lambda}{\eta}\right)^2\left(CV_{bx}^2+\frac{N-m}{N-1}\frac{CV_{wx}^2}{m}\right) = 278.\cdots\cdots$$

となります。したがつて，有効桁数をかんがえて，２桁までとると，選びだす地域の数は280になります。

(2) １従業員当たりの平均売上高をもとめる場合

まず，１従業員当たりの平均売上高の推定値ですが，それは，選びだされた飲食店の，３月の従業員数と売上高から次の推定式によつて，もとめられます。

$$\frac{\text{売上高の算術平均}}{\text{従業員数の算術平均}}$$

この推定式が，第１部で説明した１従業員当たりの平均売上高（２個の算術平均の比率の一例）の推定式と，まつたく同じ形であることに注意してください。これも偶然ではありません。

次に，この推定値を実現値とする推定子ですが，いま，選びだす地域の数をℓとして，それを$\dfrac{\overline{V}e}{\overline{U}e}$と表わしてみると，推定子$\dfrac{\overline{V}e}{\overline{U}e}$の期待値　$E\left(\dfrac{\overline{V}e}{\overline{U}e}\right)$は，１従業員当たりの平均売上高$\dfrac{Mv}{Mu}$に近似的に等しいという関係があります。すなわち，第１部第10節（９月号）の条件１は近似的に満たされているわけです。

$$E\left(\frac{\overline{V}e}{\overline{U}e}\right) \fallingdotseq \frac{Mv}{Mu}$$

また，標準偏差　$\sigma\left(\dfrac{\overline{V}e}{\overline{U}e}\right)$，変動係数 $CV\left(\dfrac{\overline{V}e}{\overline{U}e}\right)$は，従業員数の地域間変動係数CVbx，地域内変動係数CVwx，売上高の地域間変動係数CVby，地域内変動係数CVwy，従業員数と売上高との地域間相関係数ρbxy，地域内相関係数ρwxy，ならびに，飲食店総数H，地域総数L，それぞれの地域の飲食店数の最大，最小間のある数Nx，Ny，Nxy(注)それぞれの調査地域の調査飲食店数mによつて，近似的に，次のように

$$\sigma\left(\frac{\overline{V}_\ell}{\overline{U}_\ell}\right) \fallingdotseq \frac{1}{\sqrt{\ell}}\frac{Mv}{Mu}\sqrt{\left(CV_{bx}^2+\frac{N_x-m}{N_x-1}\frac{CV_{wx}^2}{m}\right)+\left(CV_{by}^2+\frac{N_y-m}{N_y-1}\frac{CV_{wy}^2}{m}\right)-2\left(\rho_{bxy}CV_{bx}CV_{by}+\frac{N_{xy}-m}{N_{xy}-1}\frac{\rho_{wxy}CV_{wx}CV_{wy}}{m}\right)}$$

$$CV\left(\frac{\overline{V}_\ell}{\overline{U}_\ell}\right) \fallingdotseq \frac{1}{\sqrt{\ell}}\sqrt{\left(CV_{bx}^2+\frac{N_x-m}{N_x-1}\frac{CV_{wx}^2}{m}\right)+\left(CV_{by}^2+\frac{N_y-m}{N_y-1}\frac{CV_{wy}^2}{m}\right)-\left(\rho_{bxy}CV_{bx}CV_{by}+\frac{N_{xy}-m}{N_{xy}-1}\frac{\rho_{wxy}CV_{wx}CV_{wy}}{m}\right)}$$

と表わされますから，これを条件２に入れてℓについて解いてみると

$$\ell \geqq \left(\frac{\lambda}{\varepsilon}\right)^2\left(\frac{M_v}{M_u}\right)^2\left[\left(CV_{\theta x}^2+\frac{N_x-m}{N_x-1}\frac{CV_{wx}^2}{m}\right)+\left(CV_{\theta y}^2+\frac{N_y-m}{N_y-1}\frac{CV_{wy}^2}{m}\right)-2\left(\rho_{\theta xy}CV_{\theta x}CV_{\theta y}+\frac{N_{xy}-m}{N_{xy}-1}\frac{\rho_{wxy}CV_{wx}CV_{wy}}{m}\right)\right]$$

$$\ell \geqq \left(\frac{\lambda}{\eta}\right)^2\left[\left(CV_{\theta x}^2+\frac{N_x-m}{N_x-1}\frac{CV_{wx}^2}{m}\right)+\left(CV_{\theta y}^2+\frac{N_y-m}{N_y-1}\frac{CV_{wy}^2}{m}\right)-2\left(\rho_{\theta xy}CV_{\theta x}CV_{\theta y}+\frac{N_{xy}-m}{N_{xy}-1}\frac{\rho_{wxy}CV_{wx}CV_{wy}}{m}\right)\right]$$

となります。

（注）それぞれの地域の，従業員数と売上高との相関係数は，すべて，同符号と仮定します。

そこで，いま，１従業員当たりの平均売上高を，たとえば，許容区間の相対誤差η＝５％，危険の確率λ＝２でもとめることにしてみましょう。

まず，過去の資料から

$$CV_{\theta x} \fallingdotseq 0.27 \qquad CV_{\theta y} \fallingdotseq 0.61 \qquad \rho_{\theta xy} \fallingdotseq 0.87$$
$$CV_{wx} \fallingdotseq 1.09 \qquad CV_{wy} \fallingdotseq 1.65 \qquad \rho_{wxy} \fallingdotseq 0.83$$

ぐらいであり，また，それぞれの地域の飲食店数も，過去の資料から，最大で23，最小で11ぐらいであるので，もつとも安全をとつて

$N_x = 23 \qquad N_y = 23 \qquad N_{xy} = 11$

とし，次に，調査員の活動能力をかんがえて

$m = 8$

と決めると，選びだす地域の数ℓは

$$\ell \geqq \left(\frac{\lambda}{\eta}\right)^2\left[\left(CV_{\theta x}^2+\frac{N_x-m}{N_x-1}\frac{CV_{wx}^2}{m}\right)+\left(CV_{\theta y}^2+\frac{N_y-m}{N_y-1}\frac{CV_{wy}^2}{m}\right)-2\left(\rho_{\theta xy}CV_{\theta x}CV_{\theta y}+\frac{N_{xy}-m}{N_{xy}-1}\frac{\rho_{wxy}CV_{wx}CV_{wy}}{m}\right)\right]=607.\cdots\cdots$$

となります。したがつて，有効桁数をかんがえて，２桁までとると，選びだす地域の数は610になります。

6　選びだす飲食店の数と推定式の修正

問題(1)地域の選定の方法は，飲食店についての比例確率抽出になつておりますが，これを「完全」におこなうためには，調査する３月の，それぞれの地域の飲食店の数があらかじめ分かつていなければなりません。

しかし，そのような資料は，一般に，手に入りませんから，したがつて，勢い，地域の選定は，過去の資料から，やむをえず，それぞれの地域の飲食店の数に見当をつけて，「不完全」におこなうしかないわけです。

ところで，この「不完全」が，「完全」とどれだけ食い違つているか，実際に分かるのは，地域の選定が終わり，それぞれの調査地域に調査員が配置されて，飲食店の名簿が作成された時です。したがつて，「不完全」を「完全」に切り替えるチヤンスは，問題(4)選びだす飲食店の数にしかありません。

では，それぞれの調査地域から，どれだけの数の飲食を選びだせば，「不完全」は「完全」に変わるかといいますとその算式は次のようになります。ここで，〔設計〕は，設計のさいにもちいた近似値をしめします。

$$\text{ある調査地域の調査飲食店数}=\text{その調査地域の〔設計〕調査飲食店数}\times\frac{\text{その調査地域の飲食店数}}{\text{その調査地域の〔設計〕飲食店数}}$$

たとえば，地域の総数を12，選びだす地域の数を３とし，過去の資料から，それぞれの地域の飲食店の数に見当をつけて，第12表のように，地域４，地域６，地域10を選びだし，それぞれの調査地域に調査員を配置して，飲食店の名簿を作成したとしてみましょう。（次頁表参照）

まず，地域４は，設計では，過去の資料から，21軒と見当をつけましたが，実際もそのとおりなので，設計どおり，そのうち，８軒を調査することになります。次に，地域６は，設計では20軒でしたが，実際は17軒なので，算式にしたがつて，

$$\text{調査飲食店数}=\text{〔設計〕調査飲食店数}\times\frac{\text{飲食店数}}{\text{〔設計〕飲食店数}}=6.8$$

地　域 通し番号	〔設計〕 飲食店数	調査 地域	飲食 店数	調査飲 食店数
1	23			
2	15			
3	16			
4	21	○	21	8
5	23			
6	20	○	17	7
7	19			
8	17			
9	20			
10	13	○	24	15
11	22			
12	11			

注　○は調査地域をしめしています。

となり，７軒を調査することになります。また，地域10は，設計では13軒でしたが，実際は24軒なので，算式にしたがつて，15軒を調査することになります。

ところで，このように修正すると，地域４は問題ありませんが，地域６は調査飲食店数が予定した８軒から７軒に減るので，もしも，調査員に設計どおりの手当を支給すれば，手当が割高になります。しかし，そうかといつて調査飲食店数に合わせて，手当を減らせば，調査員の不満が問題として残りましよう。また，地域10は，逆に，調査飲食店数が８軒から15軒に増しますが，もしも，調査員に設計どおりの手当を支給すれば，手当が割安になりますし，また，たとえ，調査飲食店数に合わせて，手当を増しても，仕事の量が問題として残ります。なお，地域６では，調査飲食店数が８軒から７軒に減りましたが，これをもつと極端にかんがえると，実際には飲食店がないために，調査飲食店数が０になる場合もありうるわけです。

こうして見てくると，この修正も，また，なかなか多くの問題点をふくんでいることに気付かれるでしよう。そこで，これをどのように取り扱つたらよいかですが，それは調査，調査によつて非常に異なるので，ここに述べる余裕がありません。なお，実際におこなわれている標本調査で，これをどのように取り扱つているかを，できるだけ広く見ることは，非常に参考になるとおもいます。

この選びだす飲食店の数の修正は，推定式にも修正をもたらします。いまそれを第５節の具体例でみると，次のようになります。

(1)　1　地域当たりの平均従業員数

$$\text{従業員数の算術平均} \times \frac{\text{調査飲食店数}}{\text{〔設計〕調査飲食店数}} \times \frac{\text{〔設計〕飲食店総数}}{\text{地域総数}}$$

(2)　1　従業員当たりの平均売上高

$$\frac{\text{売上高の算術平均}}{\text{従業員数の算術平均}}$$

これでお分かりのように，１地域当たりの平均従業員数（算術平均の一例）の推定式には，新たに $\frac{\text{調査飲食店数}}{\text{〔設計〕調査飲食店数}}$ が加わりますが，１従業員当たりの平均売上高（２個の算術平均の比率）の推定式には，変化がありません。

参考・サイコロの目の出る経験的確率の実験

サイコロを振つた回数	1　の　目		2　の　目		3　の　目		4　の　目		5　の　目		6　の　目	
	回	%	回	%	回	%	回	%	回	%	回	%
1,000	185	18.50	173	17.30	146	14.60	182	18.20	167	16.70	147	14.70
5,000	918	18.26	826	16.52	841	16.82	803	16.06	808	16.16	809	16.18
8,000	1,457	18.21	1,332	16.65	1,331	16.64	1,247	15.59	1,319	16.49	1,314	16.43
9,000	1,631	18.12	1,419	16.57	1,475	16.39	1,420	15.78	1,484	16.49	1,499	16.66
10,000	1,808	18.08	1,636	16.36	1,639	16.39	1,588	15.88	1,660	16.60	1,669	16.69

ここに掲げた表は，本稿筆者である高橋先生が，かつて実験的にサイコロを振られた結果の１部を示したものである。この回数は，更に１万５千回とか２万回と累積されていくが，ここには１万回までの結果を表示した。回数が多くなるに従つて，サイコロが全く均質で理想的に作られているとすれば，１万回振つた場合の１の目，２の目……６の目のでる回数は先験的に$10,000 \times \frac{1}{6}$回でなければならないし，またその相対頻度は16.66%である筈である。然しながら，この表の１の目の相対頻度は，18%を上回つて多過ぎる結果を示し，４の目は振る回数が多くなるに従つて，16%を下回つた相対頻度を示している。すなわちこれは，実験に用いたサイコロが全く理想的均質の状態に作られてない証拠と見るべきである。

編　集　部

確　　率（数学・統計用語）

ある事象が生起するかしないか確実に知られていない場合に，その生起の確からしさを数量的にあらわしたものを確率という。確率は実験または経験によつて定義される経験的確率と，先験的に定義される先験的確率とがある。

経験的確率の定義は次のとおりである。Nが非常に大きい数であるとき，N回の実験または経験において事象Eの起つた回数がrであるとき$\frac{r}{N}$（または$\lim_{N\to\infty}\frac{r}{N}$）を一回の実験または経験においてEの起る確率という。

具体的にサイコロを例にすると，サイコロを振つて1の目が出たときを1，1の目以外を0としてn回振つたとすると，n回を分母とし1の目の出た回数を分子とすると，nが大きい程，その値は$1/6$に近づきその極限値は$1/6$ということになる。即ちサイコロの1の目の出る確率は経験的に$1/6$であることがわかる。

先験的確率は次のように定義される，同じ確からしさで起りうる場合の数が全部でnそのうち事象Eの起る場合の数がrであればEの起る確率は$\frac{r}{n}$である。

例えば，1個のサイコロを振つたときの1の目のでる確率はサイコロが正六面体で等質に作られたものとすれば，場合の数nが6で，1の目がでるに都合のよい場合rは1である。この場合経験しなくても1の目のでる確率r/nは$1/6$であることが先験的にわかる。

このように定義された確率は次の性質をもつ。

(1)　確率は0より小でなく，1より大でない。

(2)　すべての事象のうち，そのいずれかが起る確率は1である。

(3)　n個の排反事象のいずれかが起る確率は，おのおのの事象のおこる確率の和に等しい。

(4)　n個の独立事象が同時におこる確率はおのおのの事象の起る確率の積に等しい。

排反事象とは，いくつかの事象があつて，その中の一つが起れば，他のものが起り得ないとき，これらの事象は互に排反するという。互に排反する事象のうち，そのいずれか一つが起る確率はおのおのの事象の起る確率の和に等しい。

独立事象とは，いくつかの事象があつて，そのうちのあるものの生起が他のものの生起に何の影響も与えないとき，これらの事象は独立事象である。または確率的に互に独立であるという。n個の互に独立な事象が同時に起る確率は，各事象が起る確率の積に等しい。

――参考資料・統計小辞典――

編　集　部