ある最適多重層化二段抽出法について

——昭和45年国富調査のための法人および個人企業資産調査の標本設計の理論と実際(1)——

田　口　時　夫

(1970年9月受付)

On Some Optimum Multiple Stratified Two-Stage Sampling Design with Simple Cost Function

Tokio Taguchi

We report in this paper the sampling design applied to enterprise and uncorporated enterprise survey for national wealth survey in Japan (to be practised in 1970.)

In this design the anthor tried

(i) to apply the optimum allocation
(ii) to shorten the sampling work
(iii) to simplify the estimation and
(iv) to keep constant the sample size of secondary sanpling unit from each primary sampling unit.

The Institute of Statistical Mathematics

は し が き

昭和45年国富調査の実施にさいして，特に法人および個人企業部門には多くの問題があるので，その標本設計においても，かなりの基礎的，理論的再検討がせまられている。本稿はその結論であると共に設計の提案を示すものである。それは最終的には昭和30年国富調査の標本設計を一部踏しゅうしながら，なお，最適抽出比を適用するなど多くの改善を加えることとなった．

この小篇(I)においては調査主体側の条件下での標本企画の具体的提案までを扱い，次回に予定する(II)では更に調査目的に則して調査客体の条件を資料的に分析し乍らいわば内面的にこの企画案を検討補正する予定である．

1　標本設計の理論的基礎

以下に取り上げる問題は，現在調査員システムによる経済統計調査等で主として用いられる地域別，階級別の層化無作為二段抽出法である。

その際(1)において，まず，その一般論を示し，(2)において，実務上の諸制約をできるだけ加味することにする．

(1)　一般的，理論的観点

a)　母集団の構造

母集団は S 箇の地域階層と t 箇の階級により第1図のように二重に層化されているものとする。その各層 A_{hk} につき h はこの層の属する地域特性コードを，k は同じく階級特性コードを示すものとする。そのさい A_{hk} 層に属する第二次抽出単位（個体）の総数を N_{hk}，第一次抽出単位（地点）の総数を M_h とする。また，各層のたて横の累計はまわりの欄にあたえたように記号化する．

A_{11} (N_{11})	A_{12} (N_{12})	---------	---------	A_{1t} (N_{1t})	$A_{1}.$ ($N_{1};M_{1}$)
A_{21} (N_{21})	A_{22} (N_{22})	---------	---------	A_{2t} (N_{2t})	$A_{2}.$ ($N_{2}.;M_{2}$)
⋮	⋮				
⋮	⋮				
A_{s1} (N_{s1})	A_{s2} (N_{s2})	---------	---------	A_{st} (N_{st})	$A_{s}.$ ($N_{s}.;M_{s}$)
$A_{.1}$ (N_{1})	$A_{.2}$ (N_{2})	---------	---------	$A_{.t}$ (N_{s})	$N;M$

第1図 母集団の構造

注) ( ) 内は第一次および第二次抽出単位の数を示す。さらに $A_{h}.$ は $\frac{1}{p_h}$ なる地点抽出(第一次抽出)確率が，また A_{hk} は $\frac{1}{q_{hk}}$ なる第二次抽出確率があたえられているものとする.

b) 標本の構成

(a の母集団構造により各 $A_{h}.$ 層より m_h 箇の第一次標本(地点標本)が，また A_{hk} 層より n_{hk} 個の第二次標本(客体標本)がえられるものとする。この過程は第2図のように図解できよう.

第二次抽出比 $q_{hk}=\frac{N_{hk}}{n_{hk}}$ ／ 第一次抽出比 $P_h=\frac{M_h}{m_h}$	q_{h1}	q_{h2}	---------	---------	q_{ht}	第一次標本数
p_1	n_{11}	n_{12}	---------	---------	n_{1t}	m_1
p_2	n_{11}	n_{12}	---------	---------	n_{2t}	m_2
⋮	⋮	⋮			⋮	⋮
⋮	⋮	⋮			⋮	⋮
p_s	n_{s1}	n_{s2}	---------	---------	n_{st}	m_s
第一次標本数	n_1	n_2	---------	---------	n_t	$n;m$

第2図 標 本 の 構 成

注) まわりの欄は各層標本の累計を示すものである.

c) 推 計 方 式

b) の標本による a) の母集団の階級別合計 x_k および総合計 x の推定量を X_k および X と

し，A_{hk} に属する標本合計を X_{hk} で表わすと，単純推計方式により

(1) $$X_k=\sum_{h=1}^{s}p_h q_{hk} X_{hk}\ ;\quad k=1,2\cdots,t$$

(2) $$X=\sum_{h=1}^{s}\sum_{k=i}^{t}p_h q_{hk} X_{hk}$$

である.

いま $A_{h\cdot}$ うちの第 j 地点における k 階級に属する客体の分散を $\sigma^2{}_{hkj}$，また A_{hk} 層の地点間分散を σ_{bhk} とすれば，X の分散は

(3) $$\sigma_X{}^2 \fallingdotseq \sum_{h=1}^{s}\sum_{k=1}^{t}p_h(q_{hk}-1)\sum_{j=1}^{M_h}N_{hkj}\sigma^2{}_{hkj}+\sum_{h=1}^{s}(p_h-1)M_h\sigma^2{}_{bh}.$$

で示すことができる.

ここで N_{hkj} は A_{hk} 内の第 j 地点に属する客体数を表わすものである.

さらに計算の便宜上

(4) $$\sum_{j=1}^{M_h}N_{hkj}\sigma^2{}_{hkj}=\omega_{hk}$$

$$M_h\sigma^2{}_{bh\cdot}=\beta_{h\cdot}$$

とすれば，もちろん

(3′) $$\sigma_X{}^2 \fallingdotseq \sum_{h=1}^{s}\sum_{k=1}^{t}p_h(q_{hk}-1)\omega_{hk}+\sum_{h=1}^{s}(p_h-1)\beta_{h\cdot}$$

が成立する.

なお，(1)，(2) に代る比推方式，回帰推定方式については十分考慮の余地があるが，一般的にいって此等が不偏推定量あるいは分散計算を可能にする条件* はこの種の統計資料に対してはきびしすぎるように思われる.

エ　標本設計上の一般的制約条件

標本設計上の制約条件は通常標本地点の制約

(5) $$\sum_{h=1}^{s}m_h=\sum_{h=1}^{s}\frac{M_h}{p_h}=m$$

および客体標本数の制約

(5′) $$E\left(\sum_{h=1}^{s}\sum_{k=1}^{t}n_{hk}\right)=\sum_{h=1}^{s}\frac{N_{hk}}{p_h q_{hk}}=n$$

が考えられるが，これらの一方のみを採用すれば単純な層化抽出の場合に還元され，また，別々に独立な制約として加えると e) に述べる最適抽出比の決定方程式の解がえがたいことが容易に理解できる.

したがって (5) (5′) を関連づける制約条件の中で最も一般性を持つと思われる simple cost function

(6) $$\sum_{h=1}^{s}C_{h_1}\frac{M_h}{p_h}+\sum_{h=1}^{s}\sum_{k=1}^{t}C_{h_2}\frac{N_{hk}}{p_h q_{hk}}=C$$

但し　C_{h_1} は $A_{h\cdot}$ 層の一標本地点当りの費用

C_{h_2} は $A_{h\cdot}$ 層の単位 subsample 当りの費用

C は総予算額を示す。

を採用することにする.

* 文献［2］参照

e) 最適抽出比とその決定方程式*

(6) の制約下で (3′) を最小にするいわゆる最適抽出比

$$\left(p_h^{(0)},\quad q_{hk}^{(0)};\quad \begin{array}{l} h=1,2\cdots\cdot s \\ k=1,2\cdots\cdot t \end{array}\right)$$

の決定方程式は Lagrangian L:

$$L=\sigma_x^2+\lambda\left(\sum_{h=1}^{s} C_{h_1}\frac{M_h}{p_h}+\sum_{h=1}^{s}\sum_{k=1}^{t} C_{h_2}\frac{N_{hk}}{p_h q_{hk}}-C\right) \tag{7}$$

を用いて容易に表現できる. すなわち,

$$\frac{\partial L}{\partial p_h}=\sum_{k=1}^{t}(q_{hk}-1)\omega_{hk}+\beta_{h\cdot}-\frac{\lambda}{p_h^2}\left\{C_{h_1}M_h+\sum_{k=1}^{t} C_{h_2}\frac{N_{hk}}{q_{hk}}\right\}=0 \tag{8}$$

$$h=1,2\cdots\cdot,s$$

$$\frac{\partial L}{\partial q_{hk}}=p_h\omega_{hk}-\lambda\frac{C_{h_2}N_{hk}}{p_h q_{hk}^2}=0 \tag{8′}$$

$$h=1,2\cdots\cdot,s$$
$$k=1,2\cdots\cdot,t$$

したがって (8′) から直接

$$p_h^2=\frac{\lambda C_{h_2}N_{hk}}{q_{hk}^2\omega_{hk}};\quad h=1,2\cdots\cdot,s \tag{9}$$

がえられる. 一方 (8) の両辺に p_h を乗じたものから (8′) の両辺に q_{hk} を乗じ, さらに k について sum up したものを差引くと

$$\left(\beta_{h\cdot}-\sum_{k=1}^{t}\omega_{hk}\right)-\frac{\lambda}{p_h^2}C_{h_1}M_h=0$$

すなわち

$$p_h^2=\frac{\lambda C_{h_1}M_h}{\beta_{h\cdot}-\sum_{k=1}^{t}\omega_{hk}};\quad h=1,2\cdots\cdot,s \tag{9′}$$

がえられる.

したがって (9) および (9′) から

$$q_{hk}^{(0)}=\sqrt{\frac{C_{h_2}N_{hk}}{C_{h_1}M_h}}\cdot\sqrt{\frac{\beta_{h\cdot}-\sum_{k=1}^{t}\omega_{hk}}{\omega_{hk}}}; \tag{10}$$

$$h=1,2\cdots\cdot,s$$
$$k=1,2\cdots\cdot,t$$

が成立する.

また, (9) (9′) を (b) に代入して

$$\sqrt{\lambda}=\frac{1}{C}\left\{\sum_{h=1}^{s}\sqrt{C_{h_1}M_h\left\{\beta_{h\cdot}-\sum_{k=1}^{t}\omega_{hk}\right\}}+\sum_{h=1}^{s}\sum_{k=1}^{t}C_{h_2}N_{hk}\omega_{hk}\right\} \tag{11}$$

したがって

* 文献 [1] 参照

$$(12)\quad p_h^{(0)}=\frac{1}{C}\sqrt{\frac{C_{h_1}M_h}{\beta_{h\cdot}-\sum_{k=1}^{t}\omega_{hk}}}\left\{\sum_{h=1}^{s}\sqrt{C_{h_1}M_h\left[\beta_{h\cdot}-\sum_{k=1}^{t}\omega_{hk}\right]}+\sum_{h=1}^{s}\sum_{k=1}^{t}\sqrt{C_{h_2}N_{hk}\omega_{hk}}\right\}$$

$$h=1,2\cdots,s$$

をうることになる．

以上が層化二段抽出法に関する要約であり，従来の抽出企画の一応の前提と云えよう．

(2)　標本企画に対する実践的諸問題

前節 d) で標本企画における予算上の一般的制約条件をあつかったが，こゝではさらに企画を実施に移した場合，その各段階で生ずるより具体的な諸制約を逐一検討することにする．そのためにあらかじめいくつかの記号上の定義を与えるのが説明上便利である．

a)　記号上の諸定義

(i)　母集団の各層の size によって構成される行列

$$(13)\qquad N=(N_{hk})=\begin{pmatrix}N_{11}\cdots\cdots\cdots N_{1t}\\ \vdots \\ N_{s1}\cdots\cdots\cdots N_{st}\end{pmatrix}$$

を母集団 size matrix と呼ぶことによる．

(ii)　同様に標本 size matrix は

$$(14)\qquad n=(n_{hk})$$

で表わすことができる．

(iii)　また，within variance

$$\sigma^2{}_{\omega hk}=\sum_{j=1}^{M_h}\frac{N_{hkj}}{N_{hk}}\sigma^2{}_{hkj}$$

および

between variance

$$\sigma^2{}_{bh\cdot}=\frac{1}{M_h}\sum_{j=1}^{M_h}(x_{h\cdot j}-\bar{x}_{h\cdot})^2$$

を要素とする行列及び vector

$$(15)\qquad V_\omega=(\sigma^2{}_{\omega hk})$$

$$(16)\qquad V_b=\begin{pmatrix}\sigma^2{}_{b1\cdot}\\ \sigma^2{}_{b2\cdot}\\ \vdots\\ \sigma^2{}_{bs\cdot}\end{pmatrix}$$

をそれぞれ within variance matrix および between variance vector と呼ぶ．

(iv)　以下の記述においては (15) あるいは (16) より ω_{hk}, $\beta_{h\cdot}$ を要素とする行列 W および vector $\boldsymbol{B}$ を用いるほうが便利である。したがって

$$(17)\qquad W=(\omega_{hk})$$

$$(18)\qquad \boldsymbol{B}=\begin{pmatrix}\beta_{1\cdot}\\ \beta_{2\cdot}\\ \vdots\\ \beta_{s\cdot}\end{pmatrix}$$

を特に W 行列および $\boldsymbol{B}$ vector と呼ぶことにする．

(v)　母集団の第一次抽出単位の size および第一次標本数の構成する vector をそれぞれ

(19) $$M=\begin{pmatrix} M_1 \\ \vdots \\ M_s \end{pmatrix}$$

(20) $$m=\begin{pmatrix} m_1 \\ \vdots \\ m_s \end{pmatrix}$$

で表わすことにする.

(vi) 最後に抽出比はそれぞれ

(21) $$q=(p_1\cdots\cdots, p_s)$$
$$q=(q_{hk})$$

によって表現する.

b) 抽出作業上の制約

抽出作業上において各層 A_{hk} ごとに異なる抽出比 p_h, q_{hk} を用いることは容易なようであるが, リストの質と量によって案外混乱をまねきやすい. したがって今 p_h と q_{hk} とが量として独立, すなわち,

(22) $$q_{1k}=q_{2k}=\cdots\cdots=q_{sk}=q_k \qquad k=1,2\cdots\cdots,t$$

とすれば, いくぶん負担を緩和することになるであろう.

しかし, 反面においてこの場合 (6), (8), (8′) はそれぞれ

(23) $$\sum_{h=1}^{s} C_{h_1}\frac{M_h}{p_h}+\sum_{h=1}^{s}\sum_{k=1}^{t} C_{h_2}\frac{N_{hk}}{p_h q_k}=C$$

(24) $$\sum(q_k-1)\omega_{hk}+\beta_{h\cdot}-\frac{\lambda}{p_h{}^2}\left\{C_{h_1}M_h+\sum_{k=1}^{t} C_{h_2}\frac{N_{hk}}{q_k}\right\}=0$$
$$h=1,2\cdots\cdots,s$$

(24′) $$\sum_{h=1}^{s}\omega_{hk}p_h-\frac{\lambda}{q_k{}^2}\sum_{h=1}^{s} C_{h_2}\frac{N_{hk}}{p_h}=0 \qquad k=1,2\cdots\cdots,t$$

となり, $p_h{}^{(0)}$, $q_k{}^{(0)}$; $h=1,2\cdots\cdots, s$, $k=1,2\cdots\cdots,t$ を求めることは一般に困難となる.

念のため, 以下 $p_h{}^{(0)}$, $q_k{}^{(0)}$ の算出を容易にする諸条件を列挙してみる.

(i) size matrix N の行数あるいは列数を制限すること, 例えば $t=1$ とすれば特殊な層化二段抽出法になるが, この場合も $q_{hk}=q_k$ としたためにその解法はむずかしい.

今逆に $s=1$ とすればこれは第二次抽出単位に層化を加えた double sampling design となり, M が大なる場合には, かなり有力な方法として用いられている. この場合地域符号 h を省略すると (23), (24), (24′) の解はそれぞれ

(25) $$p^{(0)}=\frac{C_1 M}{C}+\frac{\sqrt{C_1 M}}{C}\ \frac{\sum_{k=1}^{t}\sqrt{C_2 N_k \omega_k}}{\beta_{\cdot}-\sum_{k=1}^{t}\omega_k}$$

(26) $$q_k{}^{(0)}=\sqrt{\frac{C_2 N_k\left\{\beta_{\cdot}-\sum_{k=1}^{t}\omega_k\right\}}{C_1 M \omega_k}}$$

(27) $$\lambda=\frac{\left\{\sqrt{C_1 M\left\{\beta_{\cdot}-\sum_{k=1}^{t}\omega_k\right\}}+\sum_{k=1}^{t}\sqrt{C_2 N_k \omega_k}\right\}^2}{C^2}$$

によって定まる,

(ii) p および q を制限*すること

今

(28) $$p_1=p_2=\cdots\cdots=p_s=p$$

とする。このとき (24′) の両辺を k について sum up すると

(29) $$p^2=\lambda\frac{\sum_{h=1}^{s}\sum_{k=1}^{t}C_{h_2}N_{hk}}{q_k^2\sum_{h=1}^{s}\sum_{k=1}^{t}\omega_{hk}}$$

をうる．一方

$$\sum_{h=1}^{s}(24)\times p-\sum_{k=1}^{t}(24)\times q_k$$

を考えると，

$$\sum_{h=1}^{s}\left\{\beta_{h\cdot}-\sum_{k=1}^{t}\omega_{hk}\right\}p-\frac{\lambda}{p}\sum_{h=1}^{s}C_{h_1}M_h=0$$

すなわち，

(30) $$p=\sqrt{\frac{\lambda\sum_{h=1}^{s}C_{h_1}M_h}{\sum_{h=1}^{s}\left(\beta_{h\cdot}-\sum_{k=1}^{t}\omega_{hk}\right)}}$$

となり (29) と比較して

(31) $$q_k=\sqrt{\frac{\sum_{h=1}^{s}\sum_{k=1}^{t}C_{h_2}N_{hk}}{\sum_{h=1}^{s}C_1M_h}}\sqrt{\frac{\sum_{h=1}^{s}\left(\beta_{h\cdot}-\sum_{k=1}^{t}\omega_{hk}\right)}{\sum_{h=1}^{s}\sum_{k=1}^{t}\omega_{hk}}}$$

さらに，(23) に (29)，(30) を代入して

(32) $$\lambda=\frac{1}{C^2}\left[\sqrt{\left\{\sum_{h=1}^{s}C_hM_h\right\}\left\{\sum_{h=1}^{s}\left(\beta_{h\cdot}-\sum_{k=1}^{t}\omega_{hk}\right)\right\}}+\sqrt{\left\{\sum_{h=1}^{s}\sum_{k=1}^{t}C_{h_2}N_{hk}\right\}\left\{\sum_{h=1}^{s}\sum_{k=1}^{t}\omega_{hk}\right\}}\right]^2$$

をうる。

(30)，(31)，(32) は (25)，(26)，(27) の拡張になっている．

(iii) cost function の改良

(23) より複雑な cost function，例えば

(33) $$\sum_{h=1}^{s}\left(C_{h_0}\sqrt{\frac{M_h}{p_h}}+C_{h_1}\frac{M_h}{p_h}\right)+\sum_{h=1}^{s}\sum_{k=1}^{t}C_{h_2}\frac{N_{hk}}{p_hq_k}=C^*$$

によって (ii) をさらに一般化することはできない．

逆に (23) をより単純化して

(34) $$C_{11}=C_{21}=\cdots\cdots=C_{s1}=C_1$$

あるいは

$$C_{12}=C_{22}=\cdots\cdots=C_{s2}=C_2$$

* 文献［1］参照

等としても結果は同じである。

(iv) size matrix N の形を特異化すること

N を対角行列とする．この場合われわれは (1) であつかった一般の層化二段抽出法に準じて optimum values を算出することができる．

この時 (23), (24), (24′) はさらに

$$\sum_{h=1}^{s} C_{h_1}\frac{M_h}{p_h}+\sum_{h=1}^{s}\frac{C_{h_2}}{p_h}\frac{N_{hh}}{q_h}=C \tag{35}$$

$$(q_h-1)\,\omega_{hh}+\beta_{hh}-\frac{\lambda}{p_h^2}\left\{C_1 M_h+C_2\frac{N_{hh}}{q_h}\right\}=0 \tag{36}$$

$$h=1,2\cdots\cdots,s$$

$$\omega_{hh}\,p_h-\lambda\frac{C_{h_2}}{q_h^2}\frac{N_{hh}}{p_h}=0 \qquad h=1,2\cdots\cdots,s \tag{36′}$$

と変形され，解は

$$p_h^{(0)}=\sqrt{\frac{\lambda C_{h_1}M_h}{\beta_{hh}-\omega_{hh}}} \tag{37}$$

$$q_h^{(0)}=\sqrt{\frac{C_{h_1}N_{hh}(\beta_{hh}-\omega_{hh})}{C_{h_1}M_h\,\omega_{hh}}} \tag{38}$$

$$\sqrt{\lambda}=\frac{1}{0}\sum_{h=1}^{s}\left\{\sqrt{C_1M_h(\beta_h-\omega_{hh})}+\sqrt{C_2M_{hh}\omega_{hh}}\right\} \tag{39}$$

となる．

c) 集計面における制約条件

いま (28) に代り

$$p_1q_{1k}=p_2q_{2k}=\cdots\cdots=p_sq_{sk}=r_k \qquad k=1,2\cdots\cdots,t \tag{40}$$

とするならば，階級別に単純集計が行なえることになり，集計面でいくぶん簡約化することができる．

この場合 (3′) および (6) はそれぞれ

$$\sigma_x^2\fallingdotseq\sum_{k=1}^{t}\left(\sum_{h=1}^{s}\omega_{hk}\right)r_k+\sum_{h=1}^{s}\left\{\left(\beta_{h\cdot}-\sum_{k=1}^{t}\omega_{hk}\right)\right\}p_h-\sum_{h=1}^{s}\beta_{h\cdot} \tag{41}$$

$$\sum_{h=1}^{s}\frac{C_{h_1}M_h}{P_h}+\sum_{k=1}^{t}\frac{\sum_{h=1}^{\beta}C_{h_2}N_{hk}}{r_k}=C \tag{42}$$

となりこの面でもいくらか能率の向上が期待できる．

また，最適抽出比の条件は (8), (8′) に代り

$$\frac{\partial L}{\partial p_h}=\left(\beta_{h\cdot}-\sum_{k=1}^{t}\omega_{hk}\right)-\lambda\frac{C_{h_1}M_h}{p_h^2}=0 \tag{43}$$

$$h=1,2\cdots\cdots,s$$

$$\frac{\partial L}{\partial r_k}=\sum_{h=1}^{s}\omega_{hk}-\lambda\frac{\sum_{h=1}^{s}C_{h_2}N_{hk}}{r_k^2}=0 \tag{44}$$

となるからよういに

$$(45)\qquad p_h^{(0)}=\sqrt{\frac{\lambda C_{h_1} M_h}{\beta_{h\cdot}-\sum_{k=1}^{t}\omega_{hk}}}$$

$$(46)\qquad r_k^{(0)}=\sqrt{\frac{\lambda\sum_{h=1}^{s}C_{h_2}N_{hk}}{\sum_{h=1}^{s}\omega_{hk}}}$$

をえ，したがって (40) により

$$(47)\qquad q_{hk}^{(0)}=\sqrt{\frac{\sum_{h=1}^{s}C_{h_2}N_{hk}}{C_{h_1}M_h}}\sqrt{\frac{\beta_{h\cdot}-\sum_{k=1}^{t}\omega_{hk}}{\sum_{h=1}^{s}\omega_{hk}}}$$

が成立する．

また (45), (46) を (42) に代入して

$$(48)\qquad \sqrt{\lambda}=\frac{1}{C}\left[\sum_{h=1}^{s}\sqrt{\left(\beta_{h\cdot}-\sum_{k=1}^{t}\omega_{hk}\right)C_{h_1}M_h}+\sum_{k=1}^{t}\sqrt{\left(\sum_{h=1}^{s}\omega_{hk}\right)\left(\sum_{h=1}^{s}C_{h_2}N_{hk}\right)}\right]$$

をうる．

つまりこの方法においては常に最適抽出比の算出が可能であり，かつ (10), (11), (12) と比較するならば，$\sqrt{\lambda}$ の算出したがって $p_h^{(0)}$ の算出において，はるかに有利であることがわかる．つまり b) の面，特に，企画面に対しても有利に作用している．

d) 実査に関する制約条件

この面での制約は予算条件 (6) とならび重要と思われるが，特に単位第一次（地点）標本内の第二次（客体）標本数がだいたい斉一に保たれることが必要であろう．

この条件は (28) あるいは (40) のごとく単純に formulate するには複雑に過ぎるので，次のように使用上さしつかえない程度に各層の基準を具体化することにする．

つまりまず条件として

(i) 階級層化は，x と相関度の高い数量 y の大小による数量階級であるとする。具体的にその階級分点を $y_1>y_2>\cdots>y_{t-1}$ とすれば

$A_{\cdot 1}, A_{\cdot 2}, \cdots, A_{\cdot t}$ はそれぞれ

$$(49)\qquad y\geqq y_1,\ y_2\leqq y<y_1,\ \cdots\cdots,\ y_{t-1}>y$$

によって決定される．

(ii) 各単位地点に属する客体数は，だいたい一定であるとする．

これに対する標本設計の方針は

(iii) 地域階層 $A_{i\cdot}$ は $A_{\cdot j}$; $1\leqq j\leqq i-1$ に属する客体を含まず $A_{\cdot i}$ に層する客体を少くとも一つ含むように構成するのがよい．したがってこの時 $s=t$ が成立する．また，この層化基準によると母集団 size matrix N は

$$(50)\qquad N=\begin{pmatrix} N_{11} & N_{12} & \cdots & N_{1t}\\ & N_{22} & & N_{2t}\\ & & \ddots & \vdots\\ 0 & & & \vdots\\ & & & N_{tt}\end{pmatrix}$$

で示される上三角行列となる．

(iv) 以上の各層に対する Sample の割当は c) で示した，つまり (40) を満足する条件下の最適割当法を適用する．以上によるとこの層化が b), c) の条件に有効であることがただちにわかる． さらに，A_{hk} の内に特に異常な分布形態を示すものがないかぎり，(iii) の層化に (iv) の割当を与えると一般に

$$(51) \qquad p_1 < p_2 < \cdots\cdots < p_t$$

が予想されるのが (ii) の条件下で (iv) の方式 $p_h q_{hk} = r_k$ を適用すれば，さらに

$$(52) \qquad q_{1h} < q_{2h} < \cdots\cdots < q_{th}$$

が成立し，結局上位地域層 (h 小) に属する上位階級 (k 小) 標本の増大は，同地域階層の下位階級 (k 大) 標本の減小により相殺され，各標本地点内の標本客体数が均一化するような作用が認められる．

以上はなお具体的な調査区の data の解析によって，十分検討する必要があるが，一般的に集団が経済量分布にみられる歪み型分布を示し，かつ，階級分点が後述2の設計にみるように対数 scale の下でだいたい等間隔に近い場合には，(iii) の層化により

$$(53) \qquad \begin{array}{l} N_{h_1} < N_{h_2} < \cdots\cdots < N_{ht} \qquad h = 1, 2 \cdots\cdots, s \\ M_1 < M_2 < \cdots\cdots < M_s \\ \bar{\sigma}_{wh1} > \bar{\sigma}_{wh2} > \cdots\cdots > \bar{\sigma}_{wht} \qquad h = 1, 2 \cdots\cdots, s \\ \sigma_{bh1} > \sigma_{bh2} > \cdots\cdots > \sigma_{bht}, \ \sigma_{b1.} > \sigma_{b2.} > \cdots\cdots > \sigma_{bs.} \ \text{等} \end{array}$$

が期待できよう．もちろんこの結果についても data による判定が必要である．

また，(ii) の条件は一般に保障されず，したがって地域を客体数によりあらかじめ層化する必要が予想される．

しかしそれにもかかわらず，われわれは以上 (i), (ii), (iii), (iv) に掲げた層化ならびに抽出の方針が b), c), d) の条件に適合し，かつ2に述べる具体的な設計問題に対しこれまでの諸方法の中では最も情報に富み且適切であると結論できるように思われる．

2 標本設計の実際的形態*

(1) 昭和 30 年国富調査における法人企業標本設計の概説*

昭和 30 年調査においては，だいたい1の結論を予想して法人を資本金額により，(i) 1億円以上，(ii) 5千万円以上1億円未満，(iii) 1千万円以上5千万円未満，(iv) 500 万円以上1千万円未満，(v) 500 万円未満の5階級に分類し，さらに地域階層は，市町村（大都市にあっては区）を単位に (イ) (i) 階級の法人が所在する．(ロ) (i) 階級が所在せず，(ii) 階級が所在する．(ハ) (i), (ii) 階級が所在せず．(iii) 階級が所在する．(ニ) (i), (ii), (iii) 階級が所在せず．(iv) 階級が所在する．(ホ) (i), (ii), (iii), (iv) 階級が所在せず．(v) 階級が所在する．の5層に設計時において (ヘ) 法人の所在しない層を加えて層化した．

この結果は，第3図に概括されるが，この各部分 $\pi_1, \pi_2, \pi_3, \pi_4$ は調査形態を変えざるをえなかった．つまり

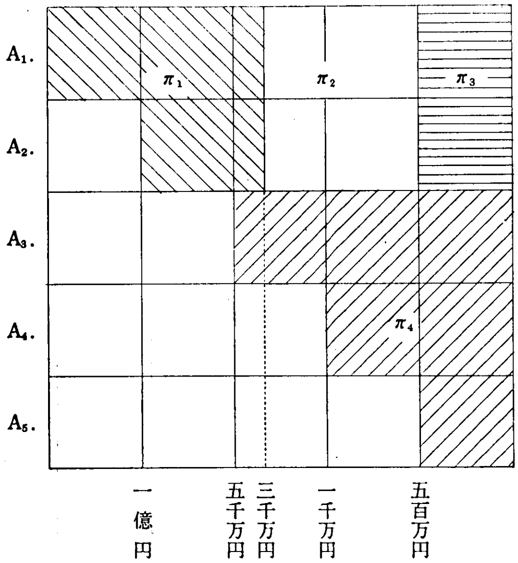
(a) π_1 部分　全数調査
(b) π_2 部分　産業別法人層化比例抽出法
(c) π_3 部分　労働力調査区単位層化抽出法
(d) π_4 部分　市町村単位層化抽出法

のごとくである．

これは当時まだ事業所設査区の設定がなく市区町村および労働力調査区を適宜併用せざるをえなかった事情によるが，その条件下では比較的地域の産業構造に対応し，調査実施を可能に

---

* 文献 [3] 参照



第3図 昭和30年調査における設計法人階層（資本金額階級）

したものである．ただし以上の性質からわかるように，この層化法は不完全な上三角形の多重層化で最適抽出法を部分的に適用するにとどまった．

(2) 昭和 45 年度法人企業標本設計方針案

(A) 昭和 45 年調査の標本抽出は原則として多重層化二段抽出法による．

(B) そのさい．単位調査区は原則として事業所調査区による．

(C) 各種の層化基準は結果表章および 1 における理論的検討に基づいて次のごとくする．

イ 法人に関する層化

(a) 第一次層化は産業大分類（ただし，製造業に関しては中分類）による．

(b) 第二次層化は資本金額により

(i) 1 億円以上

(ii) 3 千万円以上 1 億円未満

(iii) 1 千万円以上 3 千万円未満

(iv) 3 百万円以上 1 千万円未満

(v) 3 百万円未満

の五階級による．

ロ 地域に関する層化

(a) 第一次層化は欠の 7 地域ブロックによる．

(i) 北海道 (ii) 東北 (iii) 関東 (iv) 中部・北陸 (v) 近畿 (vi) 中国・四国 (vii) 九州

(b) 第二次層化は製造業を中心にした調査区の産業特性により次の三階級に分類する．

(i) 製造業が 6 割以上の調査区

(ii) 製造業が 3 割〜6 割の調査区

(iii) 製造業が 3 割未満の調査区

(c) 第三次層化はその単位調査区に所在する法人階級により次の 6 層に分割する．

(i) 資本金額 1 億円以上の法人が所在する調査区

(ii) (i) 階層に属さず，且資本金額 3 千万円以上 1 億円未満の法人が所在する調査区

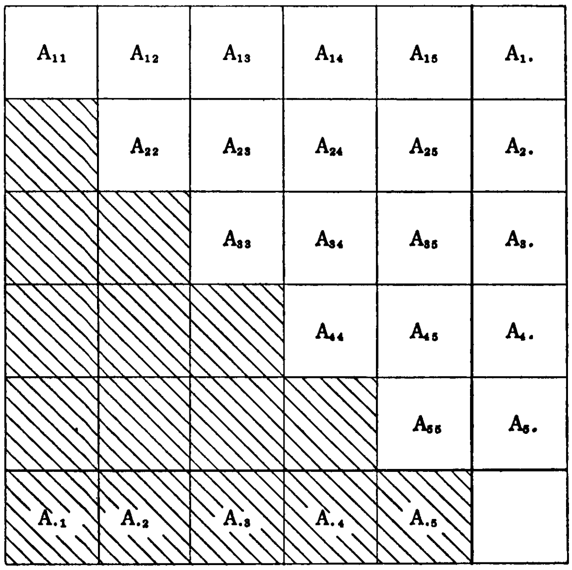
(iii) (i), (ii) 階層に属さず資本金額 1 千万円以上 3 千万円未満の法人が所在する調査区

(iv)　(i), (ii), (iii) 階層に属さず且資本金額3百万円以上1千万円未満の法人が所在する調査区

(v)　(i)～(iv) 階層に属さず資本金額3百万円未満の法人が所在する調査区

ただし，調査実施上必要ならば設計時において法人なき層を (v) より分離して (vi) 層を構成する．

(D)　以上により法人および地域に関する最終層化はつまり
イ(b) および ロ(c) については 1, (2) d) の理由により上三角型に層化されている．
従ってこれを図解すると第4図のごとくである．



階級分点 y_i 1億円　3千万円　1千万円　3百万円

第4図　昭和45年調査における母集団構造
法人階級（資本金額(y)階級）

(E)　抽出比は，法人に関する層化 イ(a) についてつまり産業大分類別にはすべて1とする．
また，地域に関する層化 ロ(a) つまりブロック別にはすべて1とする．
また，ロ(b) つまり産業特性に関しても原則として1とする．
ただし，第4図に示した最終層化 イ(b) および ロ(c) に関しては 1 (2) d) (iv) に基づいて同じく c) に示した最適抽出比を適用する．

(F)　以上の原則は，事業所調査区カードの分析の結果によっては次のような変更が生ずる可能性がある．すなわち，

○ 事業所調査区を適宜併合し，新調査区を設定する．

○ イ(a) および ロ(b) の層を外しまたは簡略化する．

○ 各調査区における法人数がきわめて不均質の場合，この数によりさらに層化する．

(G)　(A)～(E) で原則どおり行われた場合，例えば total の推計は

$$(54)\qquad X=\underbrace{\sum_{l=1}^{u}}_{\text{(地域ブロックの総和)}}\underbrace{\sum_{m=1}^{w}}_{\text{(調査区産業特性層の総和)}}\underbrace{\sum_{h=1}^{s}}_{\text{(調査区資本金特性層の総和)}}p_h^{(0)}\underbrace{\sum_{k=h}^{t}}_{\text{(法人資本金階級の総和)}}q_{hk}^{(0)}\underbrace{\sum_{i=1}^{v}}_{\text{(法人産業階級の総和)}}X_{lmhki}$$

$$=\sum_{l=1}^{u}\sum_{m=1}^{w}\sum_{k=1}^{t}r_k^{(0)}\sum_{h=1}^{k}\sum_{i=1}^{v}X_{lmhki}$$

であり標本誤差は

$$(55)\qquad \sigma^{(0)}{}_x{}^2=\sum_{l=1}^{u}\sum_{m=1}^{w}\sum_{h=1}^{s}p_h^{(0)}\sum_{k=h}^{t}(q_{hk}^{(0)}-1)\sum_{i=1}^{v}\sum_{j=1}^{Mlmhi}N_{lmhkij}\sigma^2{}_{lmhkij}$$

$$+\sum_{l=1}^{u}\sum_{m=1}^{w}\sum_{k=1}^{t}(p_h^{(0)}-1)\sum_{i=1}^{v}M_{lmhi}\sigma^2{}_{blmh\cdot i}$$

によって評価する.

なお，(54) に代り資本金-資産額比率を用いた比推定方式あるいは資本金による資産の回帰推定方式についても考慮しうる.

(3) 昭和 45 年個人企業標本設計方針案

個人企業に関する標本設計においては多重層化二段抽出法による.

そのさい，個人企業階級は従業員数階級のみである，

また，地域階層は，事業所調査調査区を単位地点とし，ブロック別および調査区産業特性のみによる.

その抽出率の決定は 1(2)b により特に (28) の条件つまり地域階層別には同一の抽出率をあたえることにより (30), (31), (32) により p, q_k を決定する.

したがって total の推計は (54) の記号にしたがい

$$(56)\qquad X=p^{(0)}\sum_{l=1}^{u}\sum_{m=1}^{w}\sum_{k=1}^{t}q_k^{(0)}X_{lmk}$$

(ただし k は従業員階級を示す)

また標本誤差は

$$(57)\qquad \sigma^{(0)}{}_x{}^2=p^{(0)}\sum_{l=1}^{u}\sum_{m=1}^{w}\sum_{k=1}^{t}(q_k^{(0)}-1)\sum_{j=1}^{Mlm}N_{lmkj}\sigma^2{}_{lmkj}$$

$$+(p^{(0)}-1)\sum_{l=1}^{u}\sum_{m=1}^{w}M_{lm}\sigma^2{}_{blm}.$$

で評価する.

なお，従業員数-資産額の相関により比推定，回帰推定を考慮することも可能である.

文　献

[1] Hansen, Hurwitz and Madow; Sample Survey Methods and Theory, Vol I, II, Wiley & Sons, Inc, New York, 1953.

[2] 林 知己夫，サンプリング調査はどう行うか，東京大学出版部，1951 年

[3] 中山伊知郎監修，日本の国富構造，東洋経済新報社，昭和 34 年

統計数理研究所