ある最適多重層化二段抽出法について（続）

——法人資産調査のための将来計画と昭和 45 年国富調査国有財産調査における標本設計——

田　口　時　夫

(1973 年 4 月受付)

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

Tokio Taguchi

In this paper, the author will mainly try to verify that the application of subsampling method with varying probabilities of selection at each stage to enterprise survey (subsampling system, in which units at each stage sampling are selected with replacement and with probabilities proportional to their capital sizes, and its improved system) brings more efficient and useful results than that of subsampling method with equal probabilities.

The Institute of Statistical Mathematics

は　し　が　き

我々は既に同一表題の下で昭和 45 年国富調査法人，個人企業資産調査における標本設計の理論と実際について二回に亙り報告した（筆者 [1] 及び [2] 参照）．

その形態は昭和30年調査に対し，第１次抽出単位の設定，層化方法及び抽出率の決定等について改善を加えたものであるが，此等の結果は猶満足な精度を与えるものではなく又此種調査に伴う諸問題について充分な解答を与えるものでもなかった．特に最適抽出率の算定結果は極めて利用性の乏しいものに終ったことを指摘せねばならない．

従って我々はこゝで改めて将来計画として基本的な解決の方法を示唆することにしたい．それは結論として従来の等確率抽出をもとにした体系に代り不等確率抽出を基本とすることであり，既にその具体的な適用を国有財産行政資産調査に対して行ったので §3 に於てその実施の概要の報告を兼ねることにした．

§1　各段階で不等確率抽出法を用いる復元副次抽出法について (Sukhatme [3] 参照)

我々が後掲 [2] の §3 以下で扱った標本設計上の諸問題は層化法及び抽出法の不適合性によるというよりは寧ろ経済量を特徴づけるＪ字型分布をもつ母集団に対して等確率抽出法を適用することの不合理性に帰因するものと考えられる．因みに不偏比推定方式による推定及びその精度は式が複雑で見通しが悪く不等確率抽出法による推定方式に優るものとは考えられない．

然し我々はこゝで不等確率抽出法の抽象的一般論（例えば Hanurav [4]）を直接展開するつもりはない．

我々は当面第１次及び第２次の各抽出段階に於てそれぞれ資本金額総計額及び資本金額に比例した撰択確率 (probability of selection) による法人抽出とその周辺の不等確率抽出法を考察すれば充分であると考える．

扨て当初は層化を特に考慮しないことにして二段階規模比例抽出法の為に次のように記号上の準備を行う．

第1表 記 号 一 覧

項目 ＼ 区分	母集団全体	第 i 番目の第1次抽出単位（地域）	第 i 単位中の第 j 番目の第2次抽出単位（企業）
size 第1次抽出単位数	M	—	—
size 第2次抽出単位数	$N=\sum_{i=1}^{M} N_i$	N_i	—
資産額総計又は資産額	x	x_i	x_{ij}
資本金総計又は資本金	y	y_i	y_{ij}
上記の比率 $=\dfrac{\text{資産額総計}}{\text{資本金総計}}$ 又は $\dfrac{\text{資産額}}{\text{資本金}}$	r	r_i	r_{ij}

抽出に於ては第2次標本数を各第1次抽出単位につき同数にすることは，実査上の大きな要請であった（筆者 [1] p. 85 及び [2] p. 10)．且この条件は単純集計を可能にするものである．

従って今第1次標本数を m' 又各第1次標本から抽出する第2次標本数を n' とする．この時 x, r に対する不偏推定量をそれぞれ X', R' とすれば

(1) $$X'=yR'=\frac{m'\,n'}{y}\sum_{m'}^{i=1}\sum_{i=1}^{n'} r_{ij}$$

であり又分散 $\sigma_{X'}^{'2}$ は容易に

(2) $$\sigma_{X'}^{'2}=y^2\left(\frac{\sigma_b^{'2}}{m'}+\frac{\bar{\sigma}_w^{'2}}{n'\,m'}\right)$$

となる。(Sukhatme [3], 8. 13, Subsampling with Varying Probabilities of Selection at each Stage pp. 404〜410)

こゝで

(3) $$\sigma_b^{'2}=\sum_{i=1}^{M}\frac{y_i}{y}(r_i-r)^2$$

又

(4) $$\bar{\sigma}_w^{'2}=\sum_{i=1}^{M}\frac{y_i}{y}\sigma_i^{'2}$$

及び

(5) $$\sigma_i^{'2}=\sum_{j=1}^{Ni}\frac{y_{ij}}{y_i}(r_{ij}-r_i)^2$$

である．特に $'$ を附したのは等確率抽出に基ずく Subsanpling 法による不偏推定量 X 及びそれに関係した諸分散

(6) $$X=N\bar{X}=\frac{1}{n\,m}\sum_{i=1}^{m}\sum_{j=1}^{n}x_{ij}$$

(7) $$\sigma_X^2=N^2\left(\frac{\sigma_b^2}{m}+\frac{\bar{\sigma}_w^2}{n\,m}\right)$$

(8) $$\sigma_b^2=\frac{1}{M}\sum_{i=1}^{M}(\bar{x}_i-\bar{x})^2$$

(9) $$\bar{\sigma}_w^2=\frac{1}{M}\sum_{i=1}\sigma_i^2$$

(10)
$$\sigma_i^2 = \frac{1}{N_i}\sum_{j=1}^{N_i}(x_{ij}-\bar{x}_i)^2$$

等と区別する為である．但し $\bar{x}=\frac{x}{N}$，$\bar{x}_i=\frac{x_i}{N_i}$ とする．

更に $\sigma'^2_{X'}$ は比率 r の標本分散

(11)
$$S'^2_b = \frac{1}{m'-1}\sum_{i=1}^{m'}(\bar{r}_{in'}-\bar{r}_{in'})^2$$

を用いると

(12)
$$\sigma'^2_{X'} = \frac{y^2E\left(S'^2_b\right)}{m'}$$

が算出される（前掲 [3]）．

我々は不等確率及び等確率の二つの抽出体系に基づくそれぞれの推計方式 (1) 及び (7) の精度を比較する為に其等の変動係数を考慮する必要がある．つまり

(13)
$$\frac{\sigma'_{X'}}{X'} = \frac{1}{\sqrt{m'}}\frac{\sqrt{\sigma'^2_b+\frac{\bar{\sigma}'^2_w}{n'}}}{r}$$

及び

(14)
$$\frac{\sigma_X}{X} = \frac{1}{\sqrt{m}}\frac{\sqrt{\sigma^2_b+\frac{\bar{\sigma}^2_w}{n}}}{\bar{x}}$$

が比較の基準を与える．此の比較は，例えば事業所センサスの調査区特性カードの活用により，最終的には数値によって行うべきものと思われるが，過去における国富調査及び法人統計の資料による次のような諸経験事実によって $m=m'$ とした場合 $\frac{\sigma'_{X'}}{X'}<\frac{\sigma_X}{X}$ を予想出来るように思われる．即ち

㈠ 調査区の資本金総計額による分布は，企業の資本金規模による分布と同様に J 字型である．

㈡ 調査区を単位として，資本金総計額は企業総数と可成りの相関をもつ（昭和 30 年国富調査設計資料）且又資本金 1 億円以上の企業数とも相関をもつ（昭和 45 年国富調査設計資料 前掲 [2] p. 4 第 2 表，法人企業数による調査区の分布の例）

㈢ 企業を単位として産業別に資本金と資産額は可成りの相関をもつ（昭和 30 年国富調査報告）又資本金と財務諸表における諸項目の金額との相関も，他標識と其等諸項目の金額との間の相関に比較して高度である（法人統計分析）．

㈣ 財務諸表における各項目の金額の変動係数（概ね >1）は，一般に財務諸比率の変動係数（概ね<1）より可成り大きい．且前者は企業規模の増大と共に増大するが，後者は逆に減少する（法人統計分析）

此の㈣は特に基本的と思われるので，此に関係する諸資料を第 2 表，第 3 表及び第 1 図から第 4 図に至る諸図表によって表現した．

扨て不等確率抽出を有利とする観点は，第 1 次標本数と第 2 次標本数の最適決定の面にも生ずるので次にこの問題を取上げよう．

今単純費用関数を総費用 C，第 1 次抽出単位の 1 箇当りの費用 c_1，第 2 次抽出単位 1 箇当りの費用 c_2 を用いて

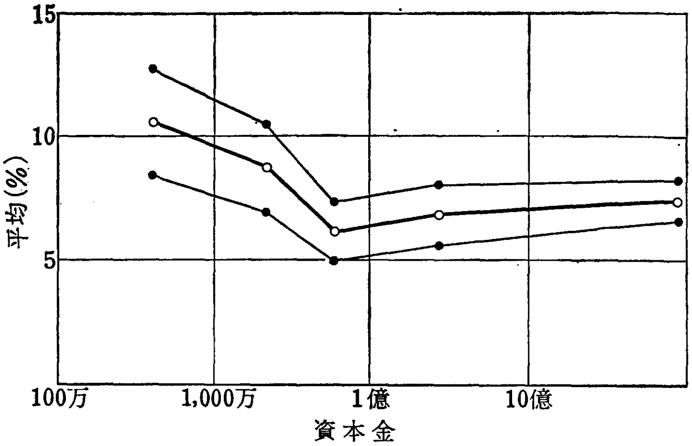
第2表 年次別及び四半期別調査による主要項目の変動係数

	法人数	資本金			売上高			資産総額			固定資産		
		平均	標準偏差	変動係数	平均	標準偏差	変動係数	平均	標準偏差	変動係数	平均	標準偏差	変動係数
年次調査													
1 (2百万未満)	520,058	0.4	0.6	1.50	38.8	73.4	1.89	17.5	36.5	2.08	4.8	20.6	4.30
2+3 (2～9)	207,522	4.6	2.0	0.44	151.4	262.2	1.73	86.0	142.8	1.66	27.2	56.8	2.09
4 (10～49)	41,951	19.0	9.9	0.52	496.4	724.8	1.46	297.0	382.4	1.29	95.4	114.6	1.20
5 (50～99)	5,370	61.7	14.0	0.23	1,195.2	1,829.3	1.53	836.9	1,037.8	1.24	293.1	341.4	1.16
6 (100～999)	4,735	250.3	173.0	0.69	3,882.1	7,127.7	1.84	2,837.6	4,148.4	1.46	1,137.7	1,525.8	1.34
7 (10億以上)	988	5054.6	10,960.0	2.17	53,770.5	159,374.7	2.96	47,107.2	103,693.2	2.20	18,908.6	41,125.4	2.17
計	780,797												
四半期調査													
2+3	226,535	5.0	2.1	0.43	43.7	67.4	1.54	85.1	120.5	1.42	28.5	73.4	2.57
4	47,352	19.3	9.9	0.51	159.5	256.9	1.61	368.7	621.6	1.69	117.3	217.0	1.85
5	5,942	62.6	14.4	0.23	385.9	571.5	1.48	1,001.6	1,150.9	1.15	355.6	604.6	1.70
6	5,438	251.3	177.8	0.71	1,082.9	2,621.4	2.42	3,037.8	5,293.2	1.74	1,112.7	1,307.1	1.17
7	1,096	4,803.5	9,269.0	1.93	15,122.3	51,344.3	3.40	46,896.7	101,059.2	2.15	18,363.7	35,587.7	1.94
計	286,363												

単位 百万円

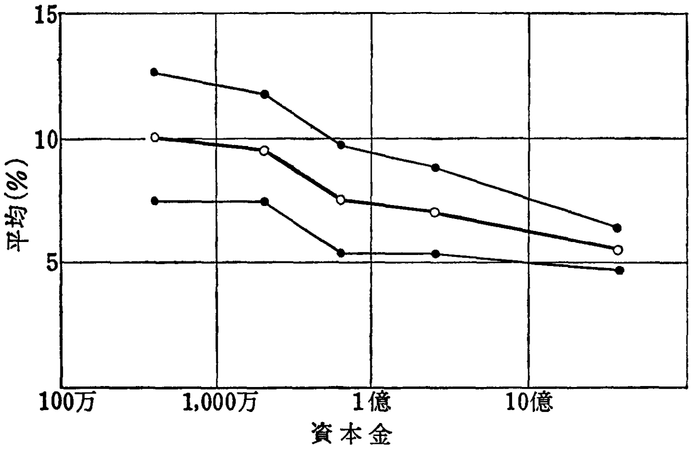
第3表　財務比率の産業別資本金階層別変動係数比較表

	売上高営業利益率$\left(\frac{営業利益}{売上高}\right)$					固定資産構成$\left(\frac{固定資産}{総資産}\right)$					人件費比率$\left(\frac{人件費}{売上高}\right)$					総資本回転率$\left(\frac{売上高}{総資産額}\right)$				
	2+3 (2百万円以上 1千万円未満)	4 (1千万円以上 5千万円未満)	5 (5千万円以上 1億円未満)	6 (1億円以上 10億円未満)	7 (10億円以上)	2+3	4	5	6	7	2+3	4	5	6	7	2+3	4	5	6	7
建設業	1.13	0.95	0.84	1.03	0.82	0.61	0.55	0.53	0.68	0.68	0.87	0.83	0.66	0.84	0.79	0.58	0.57	0.47	0.67	0.45
食料品	1.04	0.90	1.15	0.98	0.61	0.54	0.43	0.44	0.38	0.29	0.80	0.58	0.74	0.81	0.60	0.74	0.63	0.70	0.55	0.28
繊維	0.89	0.93	0.82	0.79	0.62	0.60	0.54	0.52	0.42	0.30	0.75	0.85	0.70	0.57	0.62	0.51	0.51	0.45	0.39	0.26
化学	0.56	0.65	0.77	0.80	0.73	0.55	0.49	0.47	0.42	0.34	0.87	0.58	0.62	0.50	0.49	0.75	0.48	0.54	0.44	0.35
鉄鋼	0.80	0.54	0.77	0.70	0.43	0.43	0.32	0.37	0.36	0.26	0.53	0.54	0.61	0.69	0.49	0.56	0.39	0.31	0.33	0.33
非鉄	0.84	0.86	0.56	1.04	0.69	0.72	0.58	0.35	0.42	0.38	0.77	0.65	0.59	0.59	0.38	0.26	0.56	0.41	0.36	0.31
金属	0.71	0.99	0.70	0.67	0.49	0.47	0.46	0.50	0.30	0.18	0.55	0.60	0.50	0.51	0.41	0.33	0.41	0.57	0.25	0.18
機械	0.78	0.69	0.65	0.62	0.55	0.51	0.47	0.45	0.45	0.35	0.54	0.62	0.45	0.41	0.38	0.42	0.33	0.60	0.33	0.29
電気機器	0.72	0.65	0.78	0.63	0.53	0.50	0.44	0.38	0.42	0.24	0.59	0.48	0.52	0.47	0.38	0.54	0.45	0.38	0.34	0.33
輸送用機器	0.82	0.80	0.76	0.68	0.41	0.48	0.30	0.46	0.39	0.36	0.76	0.55	0.37	0.61	0.45	0.56	0.38	0.49	0.49	0.53
卸売	1.03	0.87	0.96	1.37	1.53	0.82	0.74	0.75	0.80	0.85	1.06	0.88	1.06	1.17	1.44	0.88	0.90	0.17	0.20	0.30
小売	0.93	0.86	0.98	0.84	0.76	0.70	0.58	0.57	0.36	0.30	0.71	0.80	0.70	0.66	0.34	0.65	0.65	0.57	0.47	0.36
運輸通信	0.82	0.77	0.84	0.79	0.76	0.33	0.33	0.32	0.27	0.20	0.45	0.47	0.58	0.57	0.60	0.65	0.67	0.68	0.85	0.66
サービス	0.76	0.71	0.82	0.66	0.51	0.46	0.47	0.45	0.32	0.32	0.59	0.60	0.56	0.52	0.53	1.18	0.81	0.93	0.86	0.93

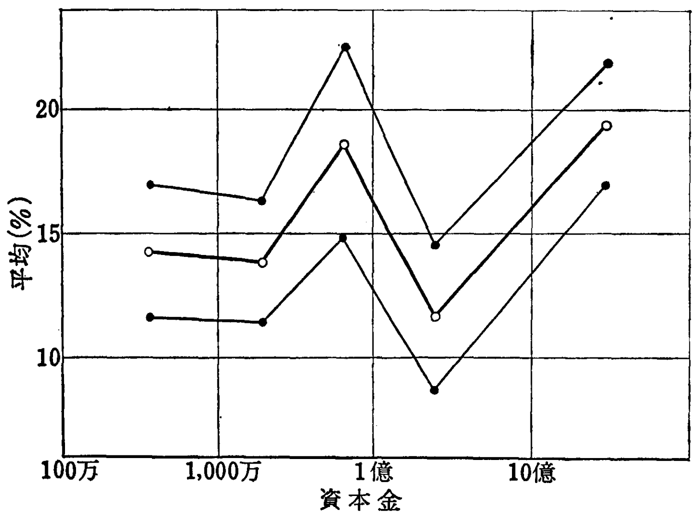


第1図 売 上 高 営 業 利 益 率

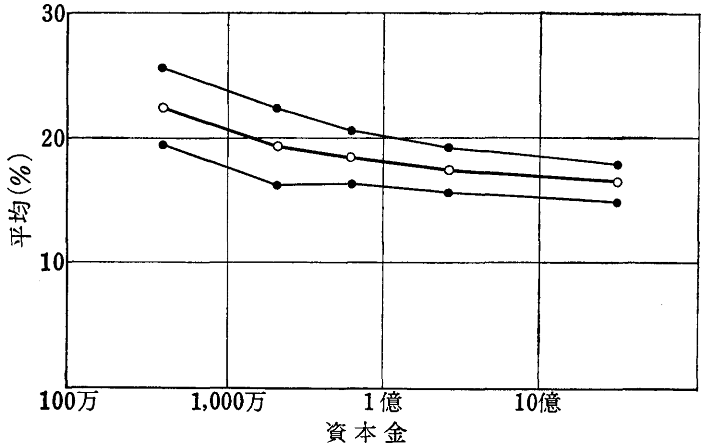
(i) 輸送用機器製造業 (中央の折線は階層別平均値を又上下の折線は平均値 ± $\frac{\sigma}{4}$ についての折線である)



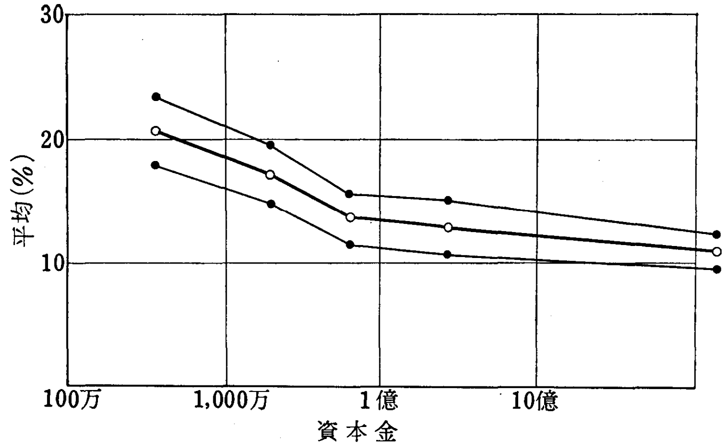
第1図 (ii) 食糧品製造業



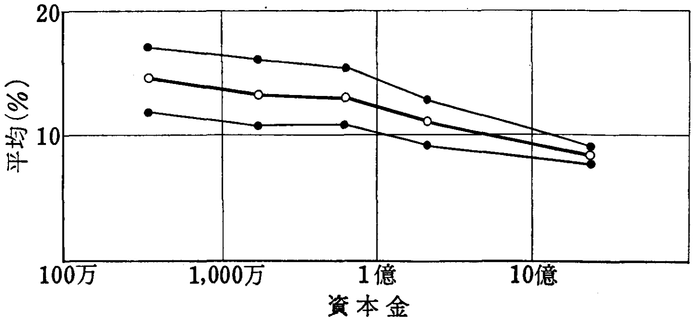
第1図 (iii) サ ー ビ ス 業



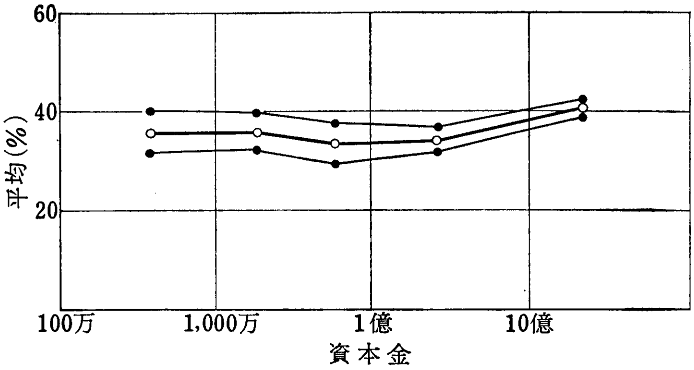
第2図　人　件　費　比　率
(i)　機 械 製 造 業



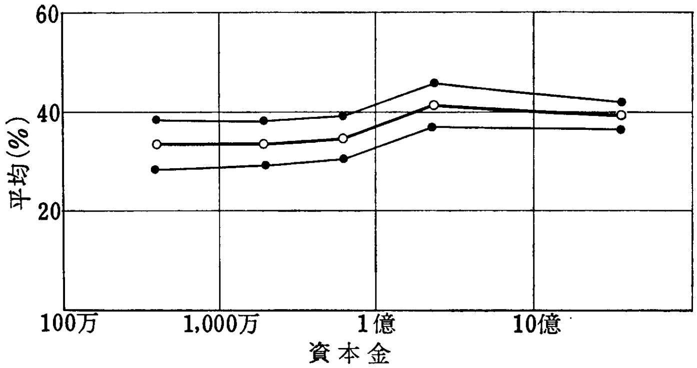
第2図　(ii) 鉄 鋼 製 造 業



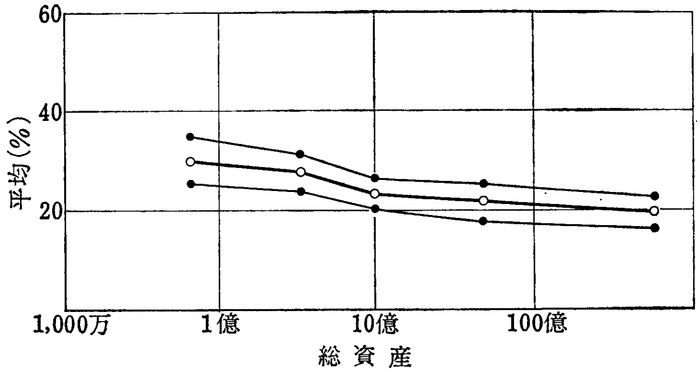
第2図　(iii) 小　売　業



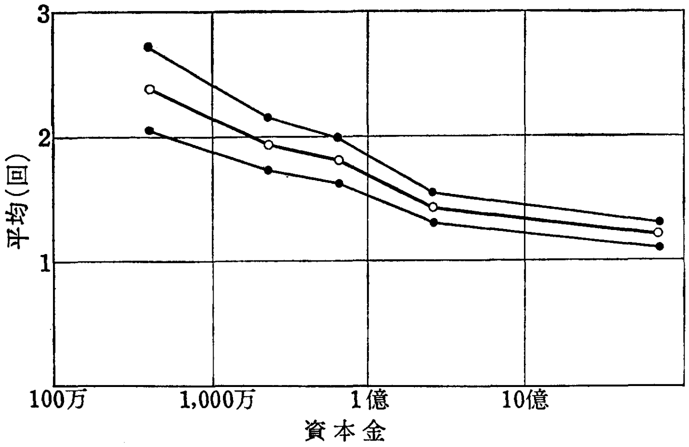
第3図 固定資産構成
(i) 金属製造業



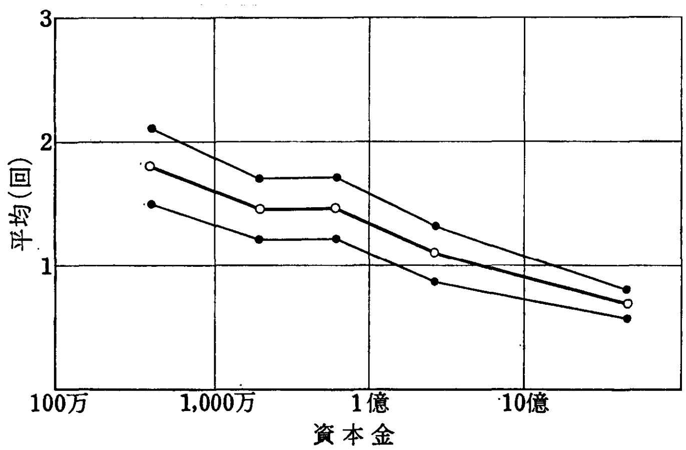
第3図 (ii) 繊維製造業



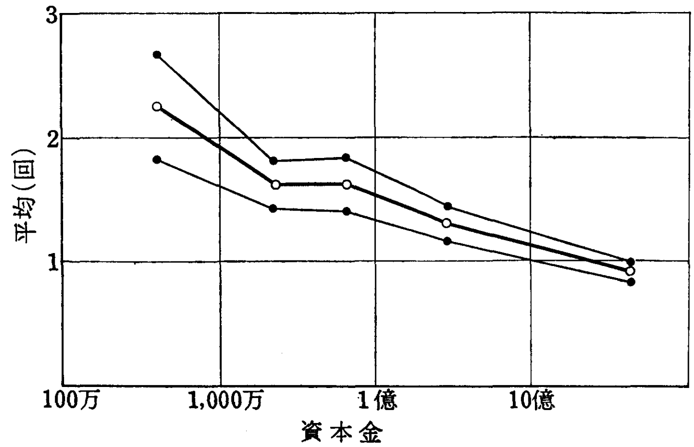
第3図 (iii) 建設業



第 4 図 総 資 本 回 転 率
(i) 電 気 機 器 製 造 業



第4図 (ii) 運 輸 通 信 業



第4図 (iii) 化 学 製 品 製 造 業

$$C = c_1 m' + c_2 m' n' \tag{15}$$

とする．この場合 C を一定として (2) の分散を最小にするような m_0', n_0' を算出すると，容易に

$$n_0' = \frac{\bar{\sigma}_\omega'}{\sigma_b'}\sqrt{\frac{c_1}{c_2}}, \quad m_0' = \frac{C}{c_1 + c_2 n_0'} \tag{16}$$

が得られる．

ここで上掲の諸資料から推して x の対資本金比率の規模による変動は観測値 x そのものの規模による変動よりも遙に緩慢であると予想されるから，一般に

$$\frac{\bar{\sigma}_\omega'}{\sigma_b'} > \frac{\bar{\sigma}_\omega}{\sigma_b} \tag{17}$$

と考えられる．(因みに規模として資本金の代りに売上高を用いても，財務諸比率の規模効果は，上掲の諸図表と極めて類似した結果を与える)．従って又等確率抽出による最適標本数 m_0, n_0 に対して

$$m_0' < m_0, \quad n_0' > n_0 \tag{18}$$

が成立つであろう．このことは，不等確率抽出が調査実施を有利にすることを示すものであり，且 m_0, n_0 の非実際的な結果（筆者 [2] p. 5 第5表各層の抽出比及び精度計算参照）を補正して最適法の適用を有効にする方向を与えるものといえよう．

以上によって我々は此の種の統計調査に対しては，不等確率抽出を基礎にする副次抽出法が適切であることを原則的に確認出来る．

然しこの方法の実施に於ては猶多くの問題を生ずる．

以下其等を逐次具体的に考察することにしよう．

§2 各抽出段階に於て層化を伴う復元不等確率副次抽出法について

前節に述べた不等確率抽出法を適用する際にも，その精度をより向上させる為に産業及び資本金額による層化が必要である．但しこの抽出法による場合は資本金額による層化の効果が一部抽出法自体に吸収されるので，等確率抽出法に比べると産業による層化がより大きな比重を占めることになる．場合により財務諸比率は規模よりも寧ろ産業による変動が大きいものと考えられるから（上掲の諸図表参照），対資本金比率の標本平均をもとにして，総量を推計する不等確率抽出法に於ても，産業による層化の影響が大きいものと予想される．

今以上の観点の下で後掲の Cochran [5] とは異なる形で，各種の層化法を具体的に考察しよう．

(一) 第1次抽出単位に s 箇の層化を行った場合

層符号を h とし，第1表の諸記号の附数にこれを加えることによって層の所在を示すと (1), (2), (3), (4), (5) の各式に対応して不偏推定量

$$X' = \sum_{h=1}^{s} y_h R_h' = \sum_{h=1}^{s} \frac{y_h}{m_h' n'} \sum_{i=1}^{m_h'} \sum_{j=1}^{n'} r_{hij} \tag{19}$$

及びその分散

$$\sigma_{X'}'^2 = \sum_{h=1}^{s} y_h^2 \left(\frac{\sigma_{bh}'^2}{m_h'} + \frac{\bar{\sigma}_{wh}'^2}{h' m_h'} \right) \tag{20}$$

が得られる．こゝで

$$\sigma_{bh}'^2 = \sum_{i=1}^{M_h} \frac{y_{hi}}{y_h} (r_{hi} - r_h)^2 \tag{21}$$

$$\bar{\sigma}_{wh}'^2 = \sum_{i=1}^{M_h} \frac{y_{hi}}{y_h} \sigma_{hi}'^2 \tag{22}$$

$$\sigma'^{2}_{hi} = \sum_{j=1}^{N_{hi}} \frac{y_{hij}}{y_{hi}} (r_{hij} - r_{hi})^2 \tag{23}$$

である．

今これに費用関数

$$C = c_1 \sum_{h=1}^{s} m_h' + c_2 n' \sum_{h=1}^{s} m_h' \tag{24}$$

を加味して，最適標本数 m_{h0}', n_0' を求めると，それは Lagrangian

$$\begin{aligned} L(m_1' m_2' \cdots m_s', n') &= \sum_{h=1}^{s} y_h^2 \left(\frac{\sigma'^{2}_{bh}}{m_h'} + \frac{\bar{\sigma}'^{2}_{wh}}{n' m_h'} \right) \\ &\quad + \lambda \left(c_1 \sum_{h=1}^{s} m_h' + c_2 n' \sum_{h=1}^{s} m_h' - C \right) \end{aligned} \tag{25}$$

に対して

$$\frac{\partial L}{\partial m_h'} = 0; \ h = 1, 2, \cdots, s, \ \frac{\partial L}{\partial n'} = 0 \tag{26}$$

の解として与えられる．従って容易に

$$-\frac{y_h^2}{m_{h0}'^2} \left(\sigma'^{2}_{bh} + \frac{\bar{\sigma}'^{2}_{wh}}{n_0'} \right) + \lambda (c_1 + c_2 n_0') = 0; \ h = 1, 2, \cdots, s \tag{27}$$

及び

$$-\frac{1}{n_0'^2} \sum_{h=1}^{s} \frac{y_h^2}{m_{h0}'} \bar{\sigma}'^{2}_{wh} + \lambda c_2 \sum_{h=1}^{s} m_{h0}' = 0 \tag{28}$$

が成立する．こゝで

$$\sum_{h=1}^{s} ((27)\text{式} \times m_{h0}') - (28)\text{式} \times n_0'$$

を両辺について算定すると

$$-\sum_{h=1}^{s} \frac{y_h^2}{m_{h0}'} \sigma'^{2}_{bh} + \lambda c_1 \sum_{h=1}^{s} m_{h0}' = 0 \tag{29}$$

が得られる．此等の一般解を与えることは難しいが，今特に

$$m_h' = \kappa_{yh} \ ; \ h=1, \ 2\cdots, \ s \tag{30}$$

と制限しよう*．つまり各層からの第1次標本数を y の総額に比例して決定するのである．(24) に於て $m_h'=\kappa y_h$; $h=1, 2, \cdots s$ とおいた条件の下で (26) をとくと，(28) (29) に於て $m_{h0}'=\kappa y_h$, $h=1, 2, \cdots, s$ とおいた場合と同様な関係が成立するので容易に

$$n_0' = \sqrt{\frac{c_1}{c_2}} \sqrt{\frac{\sum_{h=1}^{s} y_h \bar{\sigma}'^{2}_{wh}}{\sum_{h=1}^{s} y_h \sigma'^{2}_{bh}}}, \quad m_{h0}' = \frac{C y_h}{(c_1 + c_2 n_0') \sum_{h=1}^{s} y_h} \tag{31}$$

が得られる．

以上の層化をより強化する為に次に層化の更に一般的な場合を考察しよう．

(二) 各抽出単位についてそれぞれ s 及び t 箇の層化を加えた場合，この場合は，層符号を第1次及び第2次各抽出単位に対しそれぞれ h 及び k とし，第1表の諸記号の付数にこれを加えることによって層の所在を示すと，(1), (2) の各式に対応して不偏推定量

$$X' = \sum_{h=1}^{s} y_h R_h' = \sum_{h=1}^{s} \frac{m_h'}{y_h} \sum_{i=1}^{m_h'} \sum_{k=1}^{t} \frac{1}{y_{hi}} \frac{y_{hki}}{n_{hki}'} \sum_{j=1}^{n_{hki}'} r_{hkij} \tag{32}$$

* 此の条件は $m_h'=\kappa' M_h$ 或は $m_h'=\kappa'' y_h \sigma_{bh}'$ 等とすることが出来る．

及びその分散

$$(33)\qquad \sigma_{X'}^{2}=\sum_{h=1}^{s} y_h^2\left\{\frac{\sigma_{bh}'^{2}}{m_h'}+\frac{1}{m_h'}\sum_{i=1}^{M_h}\frac{y_{hi}}{y_h}\sum_{k=1}^{t}\left(\frac{y_{hki}}{y_{hi}}\right)^2\frac{1}{n_{hki}'}\sum_{j=1}^{N_{hki}}\frac{y_{hkij}}{y_{hki}}(r_{hkij}-r_{hki})^2\right\}$$

が得られる．ここで

$$(34)\qquad \sigma_{bh}'^{2}=\sum_{i=1}^{M_h}\frac{y_{hi}}{y_h}(r_{hi}-r_h)^2$$

である．

然し此等は煩雑にすぎて実用的であるとは思われない．

従って今，第2次抽出単位の標本数を第1次抽出1標本当り一定数 n' に止め且その各層 k えの割当を，k 層の資本金総額 y_{hki} に比例して行えば

$$(35)\qquad \frac{n_{hki}'}{n'}=\frac{y_{hki}}{y_{hi}},\qquad \sum_{k=1}^{t} n_{hki}'=n'$$

が成立する．又此の時 (32), (33) 両式はそれぞれ

$$(36)\qquad X'=\sum_{h=1}^{s}\frac{y_h}{m_h' n'}\sum_{i=1}^{m_h'}\sum_{k=1}^{t}\sum_{j=1}^{n_{hki}'} r_{hkij}$$

及び

$$(37)\qquad \sigma_{X'}'^{2}=\sum_{h=1}^{s} y_h^2\left\{\frac{\sigma_{bh}'^{2}}{m_h'}+\frac{1}{m_h' n'}\sum_{i=1}^{M_h}\sum_{k=1}^{t}\sum_{j=1}^{N_{hki}}\frac{y_{hkij}}{y_h}(r_{hkij}-r_{hki})^2\right\}$$

となるから今便宜上

$$(38)\qquad \bar{\sigma}_{whk}'^{2}=\sum_{i=1}^{M_h}\sum_{k=1}^{t}\sum_{j=1}^{N_{hki}}\frac{y_{hkij}}{y_h}(r_{hkij}-r_{hki})^2$$

と表わすことにしよう．

(37) 式を (20) 式と比較すると $\bar{\sigma}_{wh}'^{2}$ と $\bar{\sigma}_{whk}'^{2}$ の差に過ぎないことが容易に理解されるから

$$(39)\qquad e_{hk}=\frac{\bar{\sigma}_{wh}'^{2}}{\bar{\sigma}_{whk}'^{2}}$$

が第2次層化の効果を表わすことになる．一方最適標本数の決定は(一)の場合と全く同様な条件 (30) と計算過程によって

$$(40)\qquad n_0'=\sqrt{\frac{c_1}{c_2}}\sqrt{\frac{\sum_{h=1}^{s} y_h\bar{\sigma}_{whk}'^{2}}{\sum_{h=1}^{s} y_h\sigma_{bh}'^{2}}},\qquad m_{h0}'=\frac{Cy_h}{(c_1+c_2 n_0')\sum_{h=1}^{s} y_h}$$

が得られる．従って e_{hk} が各 h につき大である程 n_0' は (一) の場合より小となる．

実際問題として n_0' は逆に調査員の能力によって先決されるべきであって，それに応じて，第2次抽出単位の大きさ及びその層化の基準が(40)及びそれに関連する諸式によって決定されるのが順序であろう．例えば前篇 II の諸資料によれば h, k 共に産業特性 (2～3 層)×資本金規模特性 (2～3 層) とした場合，n=10 程度に止めると，第2次抽出単位は現在の事業所調査区の 4～5 箇の大きさを必要とするのではあるまいか．

猶，(35)式について第1次層の性格に応じて

$$\frac{n_{hki}'}{n_h'} = \frac{y_{hki}}{y_{hi}}, \quad \sum_{k=1}^{t} n_{hki}' = n_h', \quad n' = \frac{1}{\sum_{h=1}^{s} y_h} \sum_{h=1}^{s} n_h' y_h \tag{41}$$

とする時，(30) の仮定の下で $C = c_1 \sum_{h=1}^{s} m_h' + C_2 \sum_{h=1}^{s} m_h' n_h'$ として容易に

$$n_{h0}' = \sqrt{\frac{c_1}{c_2}} \frac{\bar{\sigma}_{whk}'}{\sigma_{bh}'}, \quad m_h{}'_0 = \frac{C y_h}{(c_1 + c_2 n_0') \sum_{h=1}^{s} y_h} \tag{42}$$

が成立つ，等条件によって異なる解が得られる．

以上§1，§2を通じて特に企業抽出に於ては単位の規模（資本金等）に比例した確率抽出法が精度の向上はもとより第1次，第2次の各層化の特性と基準，又調査区の単位の大きさを決定するに当って有効に作用する事を示した．この抽出法はそれのみではなく或種の推計に極めて適したものと思われる．次節に於て，これを実際に施行した具体例について解説し，更にこの方式の一般化を試みよう．

§3　復元不等確率抽出法による集計効果

(一)　昭和45年国富調査行政財産調査えの適用例について最初に国有財産調査に当っての当初案を略原文のまま示すことにしよう．

昭和45年国富調査のための国有財産調査標本企画案　要綱

1.　抽出の対象と推計の方針

国有財産調査のための標本設計については各種の方法が考えられるが，こゝでは精度及び実査，集計の各面からみて最も妥当すると思われる二，三の企画について概説する．

一般的に抽出は単位箇所の抽出と項目抽出に大別されるが，前者の抽出は此の際所轄機関の状況からみて好ましくなく，従って，抽出企画は後者の抽出に限定すべきものと思われる．

又推定方式は標本平均による推定よりは，不偏比推定によるのが此の際効率的であり且精度が高いものと思われる．

2.　資産項目の抽出方法と推計方式

(1)　層化1段抽出法に基ずく標本平均による推定方式

この抽出法を具体的に資産項目に適用すると，まず行政，普通財産別及び資産大分類（建物，工作物，機械，器具，備品，船舶，車輛運搬具）に層化し，更に場合により取得時期（40年再評価時以前及び以後）の層化を加える．更に取得価格或は40年評価額に応じて悉皆抽出部分と，無作為抽出部分に層化する．この数量層化に際しては，例えば取得価格或は評価額につき

1億円以上の資産	悉皆
1千万円～1億円	$\frac{1}{5}$
百万円～1千万円	$\frac{1}{50}$
百万円未満	$\frac{1}{500}$

等，資産額のオーダーによる層化が妥当と思われる．

此の場合，数量階級による層化を簡易化する，或は推定精度をたかめるのが次に示す不偏比推定方式である．

(2)　層化1段抽出法による不偏比推定方式

此の場合資産の層化及び抽出法は (1) に準ずるものとする．但し，数量・層化を例えば

1億円以上	悉皆
1億円未満	$\frac{1}{100}$

等幾分簡易化することが可能である．

推計方式を考える場合は，問題を簡単にする為に母集団 (size N) を単一の層と看做しこれから標本 (size n) を無作為に抽出するものとする．更に口座 i の取得価格或は評価額を母集団及び標本につきそれぞれ x_i, X_i とし，又残価率及び物価倍率を x_i, X_i に対してそれぞれ r_i, R_i; p_i, P_i とすれば，国富評価額 y_i, Y_i は各次式で示される．

$$y_i = P_i r_i x_i \quad \cdots\cdots (1)$$

$$Y_i = P_i R_i X_i \quad \cdots\cdots (2)$$

今総資産評価額

$$y = \sum_{i=1}^{N} y_i \quad \cdots\cdots (3)$$

の推定量 $\tilde{y}$ に次式の不偏比推定形式を与えることにする．即ち

$$\tilde{y} = \sum_{i=1}^{N} x_i \frac{1}{n} \sum_{i=1}^{n} \frac{Y_i}{X_i} + \frac{N-1}{n-1}\left(\sum_{i=1}^{n} Y_i - \frac{1}{n}\sum_{i=1}^{n}\frac{Y_i}{X_i}\cdot\sum_{i=1}^{n} X_i\right) \quad \cdots\cdots (4)$$

これは所謂 Hartley-Ross 推定方式であり (4) 式の第1項で示される通常の比推定方式の bias を是正した方式である．

この (4) 式に (1) 式を代入すると

$$\tilde{y} = \sum_{i=1}^{N} x_i \frac{\sum_{i=1}^{n} P_i R_i}{n} + \frac{N-1}{h-1}\left(\sum_{i=1}^{n} P_i R_i X_i - \frac{1}{n}\sum_{i=1}^{n} P_i R_i \cdot \sum_{i=1}^{n} X_i\right) \quad \cdots\cdots (5)$$

が得られる．

この推計に必要な $\sum_{i=1}^{N} x_i$ を抽出時に算出し同時に精度の向上と層化及び集計の簡易化を求めるものが次に示す不等確率抽出法である．

(3) 項目の金額規模に比例した確率抽出法とそれによる不偏推定方式

この抽出法に於ては (1), (2) の諸抽出法における数量階級による層化を省略することが出来る．抽出は例えば1億円以上資産を悉階とし，未満の資産につき帳簿の記載順に，取得価格又は 40 年評価額を積算し，ランダムに選んだ起点金額に達した項目と以後積算額が1億円，2億円，等に達する毎にその項目を標本として採択することである．

今標本，母集団に関する記号は (2) を踏襲することにすると，この場合の1億円未満の資産総評価額 y の不偏推定量は復元標本と考えた場合単に

$$\tilde{y} = \frac{1}{n}\sum_{i=1}^{N} x_i \sum_{i=1}^{n} \frac{Y_i}{X_i} = \frac{1}{n}\sum_{i=1}^{n} P_i Q_i \cdot \sum_{i=1}^{N} x_i \quad \cdots\cdots (6)$$

で与えられる．(5) 式と比較するとこの方式は遙に簡便であることが分る．

因みに (6) 式の分散は

$$\sigma \tilde{y}^2 = \frac{1}{n}\left(\sum_{i=1}^{N} x_i\right)^2 \sigma_r'^2; \quad \sigma_r'^2 = \frac{1}{N}\sum_{i=1}^{N} \frac{x_i}{\bar{x}} r^2{}_i - r^2, \quad r = \frac{\sum_{i=1}^{N} y_i}{\sum_{i=1}^{N} x_i} \quad \cdots\cdots (7)$$

で与えられる．

以上の推計は，net の推計であったが gross の推計に関しても取得価格が既知のものについては，同様の手順（単に $R_i=1$ とする）によって推計することが出来る．

(6) 式における $\frac{1}{n}\sum_{i=1}^{n} P_i Q_i$ は，資産の平均修正率であり，又 gross の場合，これは平均物価倍率を示すことは明かであるから式の意味が明瞭である．

因みに昭和 30 年度国富調査に於ては再評価限度額と国富評価額との間の相関は全体及び各

項目を通して 0.8 を超える結果が得られた.

以上が当初案の内容であるが，実際に (3) に当る抽出集計法を適用したのは行政財産の内工作物等一部に止めた．然もその場合の主な推計は net から gross を推計することに主眼がおかれたので場合によっては，残価率の逆数の標本平均が適用された.

扨て企業資産の評価は (一) の (1) 及び (3) 式と同様である．従ってこの具体例が将来企業部門に対しても適用さるべき理由は充分存在する．尨大な資産の集積の結果 45 年調査に於ては既に 30 年調査時のような調査員による転写は不可能となり，更に資料の蒐集，及び集計に大きな遅滞を示している現状からすれば，もし将来此の種の調査を実施するとすれば，企業抽出に加えて企業単位に資産項目分類別に項目抽出を行うことは殆ど不可避といえるであろう．次にこの場合についての標本形態を示すことにする.

(二)　企業抽出と項目抽出との結合形態

前節 1 及び 2 の企業抽出に対して更に取得価格又は評価額に比例した項目抽出を加えるならば標本論上では標題の範疇を超えた三段抽出法として扱うことが出来る (Sukhatme 前掲書 p. 409).

先前節 1 で扱ったように層化を加えない副次抽出標本から更に層化を加えない復元項目抽出を行うものとしよう .

記号は第 1 表をもとに 1 標本企業の資産額 x_{ij} を推定する為にその企業の資産項目総数 L_{ij} から l_{ij} 箇の標本項目を項目規模（帳簿価格）z_{ijp} に比例して抽出するものとしよう．その場合の不偏推定量は

$$X' = y\,R' = \frac{y}{m'\,n'} \sum_{i=1}^{m'} \sum_{j=1}^{n'} \frac{z_{ij}}{l_{ij}'\,y_i} \sum_{p=1}^{l_{ij}'} r_{ijp} \tag{43}$$

である．但し $r_{ijp} = \dfrac{x_{ijp}}{z_{ijp}}$ とする

又その分散は

$$\sigma_{X'}^{2} = \frac{\sigma_b'^2}{n'} + \frac{\bar{\sigma}_w'^2}{n'm'} + \frac{\sum_{i=1}^{M} \frac{y_i}{y} \sum_{j=1}^{N_i} \frac{y_{ij}}{y_i} \cdot \frac{1}{l_{ij}'} \frac{z_{ij}^2}{y_{ij}^2}\,\sigma_{ij}'^2}{n'\,m'} \tag{44}$$

である．但し $\sigma_{ij}'^2 = \sum_{p=1}^{L_{ij}} \dfrac{z_{ijp}}{z_{ij}} (r_{ijp} - r_{ij})^2;\ r_{ij} = \dfrac{x_{ij}}{z_{ij}}$　とする.

単純集計を可能にする為には明かに企業当りの平均項目数 l' をきめて

$$\alpha \frac{l_{ij}'}{l'} = \frac{z_{ij}}{y_{ij}};\quad \alpha = \frac{1}{\sum_{i=1}^{M} N_i} \sum_{i=1}^{M} \sum_{j=1}^{N_i} \frac{z_{ij}}{y_{ij}} \tag{45}$$

（簿価合計額対資本金比率の企業平均）

によって調査項目を割当てるのがよい．但しこの α は既存資料によって推定するのがよい．この場合 l_{ij}' には業種や企業施設の新旧等が大きく影響するであろう.

(45) を (43), (44) の各式に代入するとそれぞれ

$$X' = \frac{\alpha\,y}{m'\,n'\,l'} \sum_{i=1}^{m'} \sum_{j=1}^{n'} \sum_{p=1}^{l'} r_{ijp} \tag{46}$$

及び

$$\sigma_X'^2 = y^2 \left(\frac{\sigma_b'^2}{n'} + \frac{\bar{\sigma}_w'^2}{n'\,m'} + \frac{\bar{\bar{\sigma}}_{wz}'^2}{n'\,m'\,l'} \right) \tag{47}$$

が得られる．こゝで

$$(48)\qquad \bar{\bar{\sigma}}'^2_w=\sum_{i=1}^{n}\frac{\alpha}{y}\sum_{j=1}^{N_i}\sum_{p=1}^{L_{ij}}z_{ijp}(r_{ijp}-r_{ij})^2=\sum_{i=1}^{n}\frac{\alpha}{y}\sum_{j=1}^{N_i}z_{ij}\sigma_{ij}^2$$

となる．

以上をより一般化して前節2で扱った二重層化副次抽出標本から u 箇の大項目別に $l'_{hkijv};v=1,2,\cdots,u$ 箇の標本項目を規模比例で抽出するものとしよう．

但し単純集計を可能ならしめる為

$$(49)\qquad \frac{n'_{hki}}{n'}=\frac{y_{hki}}{y_{hi}}$$

又

$$(50)\qquad \alpha\frac{l'_{hkijv}}{l'}=\frac{z_{hkijv}}{y_{hkij}};\quad \alpha=\frac{1}{N}\sum_{h=1}^{s}\sum_{i=1}^{M_h}\sum_{k=1}^{t}\sum_{j=1}^{N_{hki}}\sum_{v=1}^{u}\frac{z_{hkijv}}{y_{hkij}}$$

と仮定する．

この時不偏推定量は

$$(51)\qquad X'=\frac{\alpha}{l'n'}\sum_{h=1}^{s}\frac{y_h}{m_h'}\sum_{i=1}^{m_h'}\sum_{k=1}^{t}\sum_{j=1}^{n_{hki}'}\sum_{v=1}^{u}\sum_{p=1}^{l'_{hkijv}}r_{hkijvp}$$

である．又その分散は

$$(52)\qquad \sigma'^2_{X'}=\sum_{h=1}^{s}y_h^2\left\{\frac{\sigma'^2_{bh}}{m_h'}+\frac{\bar{\sigma}'^2_{whk}}{m_h'n'}+\frac{\bar{\bar{\sigma}}'^2_{whkvz}}{m'_hn'l'}\right\}$$

となる．

但し

$$(53)\qquad \bar{\bar{\sigma}}'^2_{whkvz}=\sum_{i=1}^{M_h}\sum_{k=1}^{t}\sum_{j=1}^{N_{hki}}\sum_{v=1}^{u}\sum_{p=1}^{L_{hkijv}}\alpha\frac{z_{hkijvp}}{y_h}(r_{hkijvp}-r_{hkijv})^2$$

である．

こゝで改めて最適標本数を考えてみよう．

今費用関数として(24)式の費用に加えて調査員が転写作業を直接行い，或は指導するものとして一項目当りの単価 c_3 で企業平均項目数 l' についてその費用を計上し

$$(54)\qquad C=c_1\sum_{h=1}^{s}m_h'+c_2\sum_{h=1}^{s}m_h'n'+c_3\sum_{h=1}^{s}m_h'n'l'$$

を用いることにする．

この場合に於ても m_h' についての一般解を得ることは困難であるから

$$(55)\qquad m_h'=\kappa y_h;\ h=1,\ 2\cdots,=s.$$

として Lagrangian

$$(56)\qquad L(\kappa,n',l')=\frac{1}{\kappa}\sum_{h=1}^{s}y_h\left\{\sigma'^2_{bh}+\frac{\bar{\sigma}'^2_{whk}}{n'}+\frac{\bar{\bar{\sigma}}'^2_{whkvz}}{n'l'}\right\}$$
$$+\lambda\kappa\sum_{h=1}^{s}y_h(c_1+c_2n'+c_3n'l'-C)$$

について $\dfrac{\partial L}{\partial\kappa}=0,\ \dfrac{\partial L}{\partial n'}=0,\ \dfrac{\partial L}{\partial l'}=0$ の解を求める．

つまり

$$(57)\qquad -\frac{1}{\kappa_0^2}\sum_{h=1}^{s}y_h\left\{\sigma'^2_{bh}+\frac{\bar{\sigma}'^2_{whk}}{n_0'}+\frac{\bar{\bar{\sigma}}'^2_{whkvz}}{n_0'l_0'}\right\}+\lambda\sum_{h=1}^{s}y_h(c_1+c_2n_0'+c_3n_0'l_0')=0$$

$$(58)\qquad -\frac{1}{n_0'^2\kappa_0}\sum_{h=1}^{s} y_h\left\{\bar{\sigma}'^2_{whk}+\frac{\overline{\overline{\sigma}}'^2_{whkvx}}{l_0'}\right\}+\lambda\kappa_0\sum_{h=1}^{s} y_h(c_2+c_3 l_0')=0$$

$$(59)\qquad -\frac{1}{l_0'^2\kappa_0}\sum_{h=1}^{s} y_h\frac{\overline{\overline{\sigma}}'^2_{whkvx}}{n_0'}+\lambda\kappa_0\sum_{h=1}^{s} y_h c_3 n_0'=0$$

によって最適標本数 m_0', n_0', l_0' を決定しよう．

先 (57) $\times\kappa-$(58) $\times n'$ 及び (58) $\times n'-$(59) $\times l'$ によって我々は

$$(60)\qquad -\frac{1}{\kappa}\sum_{h=1}^{s} y_h\sigma'^2_{bh}+\lambda\kappa_0 c_1\sum_{h=1}^{s} y_h=0$$

及び

$$(61)\qquad -\frac{1}{n_0'\kappa_0}\sum_{h=1}^{s} y_h\bar{\sigma}'^2_{whk}+\lambda\kappa_0 c_2 n_0'\sum_{h=s}^{s} y_h=0$$

を得るから，(60), (61) 及び (59) 式によって

$$(62)\qquad \kappa_0=\frac{\sqrt{\sum_{h=1}^{s} y_h\sigma'^2_{bh}}}{\sqrt{\lambda c_1\sum_{h=1}^{s} y_h}},\quad \kappa_0 n'_0=\frac{\sqrt{\sum_{h=1}^{s} y_h\bar{\sigma}'^2_{whk}}}{\sqrt{\lambda c_2\sum_{h=1}^{s} y_h}},$$

$$\kappa_0 n_0' l_0'=\frac{\sqrt{\sum_{h=1}^{s} y_h\overline{\overline{\sigma}}'^2_{whkvx}}}{\sqrt{\lambda c_3\sum_{h=1}^{s} y_h}}$$

が成立する．

又 (54) 式から

$$(63)\qquad \sqrt{\lambda\sum_{h=1}^{s} y_h}=\frac{1}{C}\left(\sqrt{c_1\sum_{h=1}^{s} y_h\sigma'^2_{bh}}+\sqrt{c_2\sum_{h=1}^{s} y_h\bar{\sigma}'^2_{whk}}+\sqrt{c_3\sum_{h=1}^{s} y_h\overline{\overline{\sigma}}'^2_{whkvo}}\right)$$

が得られるから，α を既存資料から推定すれば (62), (63) によって逐次 m_0', n_0', l_0' を決定することが出来る．

n_0', l_0' は予め調査能力によって定められるから，以上の 結果によって逆に 単位調査区の設定，産業及び項目分類の基準と示すことが出来よう．

以上によって我々は法人資産調査標本設計の可成り具体的な将来計画を与えることが出来た．現実には猶多くの問題が伏在するに違いないが，法人及び資産の規模比例抽出によって，一つの可能性が与えられるように思う．

§4　その他の復元不等確率抽出法と最適撰択確率について

§1 から §3 に至る以上の諸節に於て我々は専ら抽出単位の 規模に 比例する 撰択率のみを扱ったが，猶その他の撰択率や其等の併用等について考慮する必要がある．

(一)　単位の規模 y についての相対度数 $f(y)$ の逆数に比例した撰択率について

一般に size N の母集団から撰択率 p_i をもって復元的に n sample を抽出した時 x の総額の不偏推定量 X' は

$$(64)\qquad X'=\frac{1}{n}\sum_{i=1}^{n}\frac{x_i}{p_i}$$

で与えられる．又この時 X' の分散は

$$(65)\qquad \sigma^2_{X'}=N^2\frac{\sigma'^2}{n}$$

第4表　等確率抽出法と不等確率抽出法との分散による比較

産業符号	資本金階級 / 比較分数 / 産業	2百万未満		2百円以上1万未満		1千万以上1億未満		1億以上10億未満		10億以上		計		註*			
		$\sigma_x^2<\sigma_x'^2$	$\sigma_x>\sigma^2{}_x'^2$	$\sigma_x^2<\sigma_x'^2$	$\sigma_x^2>\sigma_x'^2$	$\sigma_x^2<\sigma_x'^2$	$\sigma_x^2>\sigma_x'^2$	$\sigma_x^2<\sigma_x'^2$	$\sigma_x^2>\sigma_x'^2$	$\sigma_x^2<\sigma_x'^2$	$\sigma_x^2>\sigma_x'^2$			2倍	3倍〜9倍	10倍〜99倍	100倍〜
01	農業	0	10 (1)	17	3	11	9 (4)	2	8	—		30	30(5)				
06	林業	1	9 (7)	8	12	19	1	8	2	—		36	24(7)				
08	漁業	6	4	9	11	10	10	2	8 (3) 内3倍1	0	10 (10*)	27	43(13)	2	7	1	1
10	鉱業	10 (5) 内3倍1	0	15	5	9	11	8	2	0	10 (9*)	42(5)	28(9)	2	6		
11	石炭鉱業	?		?		10	0	10 (6) 内3倍1	0	10	0	30(6)	0				
15	建設業	4	6	7	13	0	20	0	10 (8) 内3倍1	1	9 (9*)	12	58(17)	1	6	2	
18	食品製造業	6	4	2	18	5	15	0	10	1	9 (6*)	14	56(6)	3	3		
20	繊維工業	4	6	8	12	6	14 (4)	10	0	0	10 (10*)	28	42(14)		7	3	
24	紙パ製造	4	6	5	15	16	4	2	8	0	10 (10*)	27	43(10)	1	9		
26	化学工業	10 (2)		4	16	3	17 (1)	10	0	0	10 (10*)	27(2)	43(11)		7	3	
30	窯・土石	8 (1)	2	13	7	5	15	3	7	1	9 (8*)	30(1)	40(8)		6	2	
31	鉄鋼業	5 (2)	5	17	3	1	19	3	7	0	10 (10*)	26(2)	44(10)		1	8	
32	非鉄製造	9 (7) 内3倍以上5	1	16	4	10	10	9	1	0	10 (10*)	44(7)	26(10)	2	6	2	
33	金属製品	9 (1)	1	9	11	10	10	1	9	0	10 (9*)	29(1)	41(9)		9		
34	一般機械	9	1	9	11	6	14	6	4	0	10 (10*)	30	40(10)	1	9		
35	電気機械	0	10 (3)	3	17	1	19	3	7	0	10 (10*)	7	63(13)			10	
36	輸送用機械	7 (1)	3	10	10	7	13	2	8 (1)	0	10 (10*)	26(1)	44(11)		1	9	
37	精密機械	1	9	2	18	8	12	7	3	5(1)	5 (1)	23(1)	47(1)				
38	武器製造	9 (3)	1	7	13	1	19 (5)	8	2	0	10 (10*)	25(3)	45(15)	1	3	5	1
39	その他製造	6	4	9	11	3	17	5	5	8(6)	2 (1)	31(6)	39(1)				
40	卸売業	8 (4)	2	3	17	1	19	0	10	0	10 (9*)	12(4)	58(9)		8	1	
49	その他の小売	4	6	4	16	13	7	2	8 (3)	6	4 (1*)	29	41(4)		1		
58	不動産業	1	9	4	16	14	6	7	3	0	10 (6*)	26	44(6)	3	3		
59	その他の不動産	6 (1)	4	18	2	13	7 (1)	9	1 (1)	1	9 (9*)	47(1)	23(11)	1	6	2	
61	道路運送	7	3	10	10	2	18 (3)	2	8	0	10 (10*)	21	49(13)	1	9		
64	航空運輸	0	10	10	10	10	10	6	4	0	10 (10*)	26	44(10)	2	4	4	
69	その他の運輸	10	0	9	11	8	12	0	10 (6) 内3倍2	0	10 (9*)	27	33(15)		5	4	
70	電気業	—		—		—		9 (2)	1 (1)	0	10 (10*)	9(2)	11(11)		1	5	4
71	ガス業	—		—		16	4	0	10 (7) 内3倍以上2	0	10 (8*)	16	24(15)		1	5	4
80	娯楽業	3	7	10	10	9	11	0	10	4	6 (5*)	26	44(5)		1	5	2
81	放送業	—		—		—		2	8	4	6 (2)	6	14(2)		4	1	
	計	147(27)	123(11)	238	302	227	343(18)	136 (8)	164(30)	41 (7)	249(222)	789(42)	1,181(281)	20	123	72	12

欄内の数字は，$\sigma_x^2<\sigma_x'^2$ 或は $\sigma_x^2>\sigma_x'^2$ をとる項目数を表わす

更に（　）内の数字はその比が2以上の項目数を表わす

但し

$$\sigma'^2 = \frac{1}{N}\sum_{i=1}^{N}\frac{x_i^2}{p_i} - \bar{x}^2 \tag{66}$$

である.

従って今 x が J 字分布に従い且 x, y の相関が高いとき，もし

$$\frac{1}{H_{f(y)}} = \sum_{i=1}^{\alpha}\frac{1}{f(y_i)} \tag{67}$$

が過大でなければ $p_i = \dfrac{H_{f(y)}}{f(y_i)}$ として σ'^2 が σ^2 より小となる場合が生じよう．この場合明かにこの撰択率が有効であるが，この抽出は y の規模を等確率に抽出することを意味し，従って乱数によって規模を指定した後に該当する対象を見出すことになり技術的に検討を要する.

㈡　複数箇の規模を用いて撰択率を決定する場合

資本金 y によって比例抽出を行う場合猶低額階層について不充分であった．従って，この場合更に従業員数 z を用いて，yz 或は $\sqrt{yz}$ 等に比例した抽出を考えることが出来よう.

抽出方法は稍複雑化するきらいがあるが，それに見合う有効な結果を期待することも出来よう.

㈢　抽出法の併用

大蔵省法人統計調査の資料によって作製された第4表によると，財務諸表上の諸項目の平均金額の推定に於ては，一般的にみて，低額規模階層に於ては等確率抽出法が有利であり，逆に高額規模階層に於ては，規模比例抽出法が遙かに有利であると語められる．更に産業別に効果が一様でない事が認められる．このことから両抽出法の併用形態が望ましいものといえるであろう．そしてこのことは適宜 $\sigma'^2_b, \bar{\sigma}'^2_w$ 等を $\sigma^2_b, \bar{\sigma}^2_w$ 等と置きかえることによって，これまでの各節で扱った最適抽出率の決定に何等支障を来さないのである.

因みに第4表について更に詳しく解説すると，この表は二つの分散即ち

$$\sigma^2_X = \frac{1}{N}\sum_{i=1}^{N} x_i^2 - \bar{x}^2 \tag{68}$$

と

$$\sigma'^2_X = \frac{1}{N}\sum_{i=1}^{N}\frac{\bar{y}}{y_i}x_i^2 - \bar{x}^2 \tag{69}$$

を産業別資本金額階級別に財務諸表の各項目毎に算出した結果の一つの集括表である．つまりその欄内の各数値はそれぞれ $\sigma^2_X < \sigma'^2_X$ 及び $\sigma^2_x > \sigma'^2_x$ を示した項目の箇数を表わしている．処で同数箇の標本抽出を前提にすると，等確率による場合と規模比例確率による場合との結果の優劣は将に σ^2_x と σ'^2_x の大小によって決定され，それが小なる値を示す程有利な抽出法であることは既定の事実である（例えば[3]参照）．従って又表の欄内の数値が大である程一方の抽出法が有利であることを示すことになる．こうしてみると第4表の結果は規模別には低額階層，産業別には農林業，鉱業，石炭鉱業，非鉄製造等に関しては等確率抽出法が有利であり，逆に高額階層そして建設，製造等の主要産業に於て規模比例抽出法が有利であるといえる．特に将来規模が増大する傾向にあることや産業の近代化を考慮するならば，全体として後者の抽出法が優位を示すことが予想出来る.

㈣　最適撰択率の予想形態

不等確率抽出法はミクロ的観点え一歩移行することによりいわばより質的な抽出法に近づいたともいえる．それではその撰択確率は如何に決定さるべきであろうか．この決定にこそ寧ろより本質的な最適問題が潜むのであって，標本数の決定等に関する抽出過程の細部に関する最

適性に二義的な問題となるであろう．従って我々は次にこれを取上げるべきであるが，標本加重平均を用いた x の total の不偏推定を基礎にすれば，それは x に関する関数モデルの選択の如何に拘わるのである．

例えば生産関数形式

(70)
$$x = \alpha y^p z^q$$

を採択するならば

(71)
$$\sigma'^2_x = \frac{1}{N}\sum_{i=1}^{N}\frac{\overline{y^p z^q}}{y_i^p z_i^q}x_i^2 - \bar{x}^2$$

但し
$$\overline{y^p z^q} = \frac{1}{N}\sum_{i=1}^{N} y_i^p z_i^q$$

を最小ならしめる p_0, q_0 を決定し $y^{p_0} z^{q_0}$ に比例した選択確率をとることになる．

又もし線形モデル

(72)
$$x = \alpha y + \beta z + \cdots + \gamma w$$

がよく適合する場合

(73)
$$\sigma'^2_x = \sum_{i=1}^{N}\frac{\overline{\alpha y + \beta z + \cdots + \gamma w}}{\alpha y_i + \beta z_i + \cdots + \gamma w_i}x_i^2 - \bar{x}^2$$

但し
$$\overline{\alpha y + \beta z + \cdots + \gamma w} = \frac{1}{N}\sum_{i=1}^{N}(\alpha y_i + \beta z_i + \cdots + \gamma w_i)$$

を最小にするような $\alpha_0, \beta_0, \cdots, \gamma_0$ を決定し，$\alpha_0 y + \beta_0 z + \cdots + \gamma_0 w$ に比例した抽出確率をとることになる．

つまり規模比例確率抽出に於ては，σ'^2_w, σ'^2_b 等の式から明かなように，x と比例関係に近い量，従がって x との比が一点分布に近いものを抽出基準量に採択することが望ましいからである．

処で関数モデル自体の妥当的根拠は扨置き，この x との比が一点分布に近いという基準は重要であり，従来の最少二乗法的回帰推定方式によるモデルの決定が，必ずしもこの条件にあった結果を与えるとは限らないことは一考に値するであろう．

む　す　び

これまで三回に亙り検討を重ねた本題の取扱いは，主として標本精度を中心にした計量的側面に関するものであった．然し本稿のむすびに当って確率標本抽出を実質的に考察することは，問題を発展させる上で意義あることであろう．その為に既述の企業抽出に用いた規模比例抽出法を再び取上げることにしよう．そこでは我々は資本金に比例した抽出，即各資本金単位量についての等確率抽出を採択した．こゝで資本金を基準に用いたのは，それが input の代表量として output 諸量を或程度まで規定するものとの予想に基づくものである．他方消極的にはこれ以外の事前情報量は従業員数を除き一般に把握困難である点によるのである．更に各層内で資本金単位量について等確率抽出を行うことは，各企業の産出力の蓋然的同等性ではなく各単位資本量について，その所在の如何に拘りなく，産出力の蓋然的同等性を是認することである．今もしその所在を離れて抽象的に捉えるならばこのことは潜在的には妥当するであろう．然し現実に例えば対資本金比率で表わされた産出力が，例えば第 1, 2, 3, 4 の諸図に認められるような企業規模によって異なる傾向を示す事実や，企業間に平均化を阻む相互作用が認められる以上方法としての仮定的近似的性格を免かれないであろう．この点に又§4 の後半に述べたような他の方法を考慮する余地が生ずるのである．

更に今適用の場を移して例えば社会統計調査に於て社会状態を或種の属性をもつ人員数の比率 (%) によって把握しようとする場合に，各階層内の各人がその属性の所有に関して自由撰択性をもち且その人員比率が略均衡していると予想される限り，各抽出単位に属する人員数を

規模として比例抽出する所謂確率比例抽出法が極めて効果的であるとされているが，これは前記の法人抽出の例と直線的な関係にあるといえる．つまりこの場合は個人を第2次抽出単位として (19) 式に於て $y_{hij}=1$ とすることに他ならないからである．

但しこの場合に於ても各人がその属性に支配或は束縛される，或はその所有に関して各人が対立，競合又は協調等の関係によって影響され合っていることが具体的に明かな場合（現実は必ずこうした状態を伴っていると思われるが）この方法は必ずしも好ましいものとはいえないであろう．

更に人員数と異なる測定単位（年数，回，度数，点，件数，金額等）によって平均或は総額を求める場合は，各人間の格差の大小や全体的な移動の大きさによって必ずしもこの方法が好ましいものとはなり得ないであろう．第1次抽出についていえば，この場合地域の面積や人口密度，人口流出入率更に出来れば所得水準等の利用が規模標識として有効であるかもしれない．要は当然の事乍ら先づ対象のもつ構造とメカニズム，その運動の法則性と不確定性等についての認識の深浅の程度，又統計実務機関の調査能力や資料の蓄積と整備の状況，更にいうまでもなく調査目的によって抽出の system が具体的に決定さるべきであって，単に慣行や使用の多寡によって特定の方法を選定するには現象は複雑にすぎるようである．

その意味に於て本稿で扱った不等確率抽出法が単に経費の節減や，集計の簡約化，精度の向上といった技術的改善に止まらず，必然性と蓋然性の包括的客観的認識の上に立つ柔軟な抽出法を見出す為に必要な一階梯となる事が望ましい．そして筆者は現にこの方法を媒介として以上の展望を得た事に於て既にその要望を充したものと考えるのである．

補　　遺

不等確率抽出法に関しては，T. Mizuno 氏等の先駆的業績は欠かせないものであるが，下記の参考文献中に於て既に充分紹介されているので敢て言及しなかった．こゝで非礼をお詫びする次第である．

又参考資料の面で大蔵省証券局資本市場課の方々特に伊藤，花井の両氏に一方ならぬ御援助をうけた事を厚くお礼したい．

参　考　文　献

[1] 田口時夫： ある最適多重層化二段抽出法について (I) 1971，統計数理研究所彙報，第18巻，第2号．

[2] 田口時夫： ある最適多重層化二段抽出法について (II) 1972，統計数理研究所彙報，第20巻，第1号．

[3] P. V. Sukhatme : Sampling Theory of Survey with Applications, 1953, The Iowa State College Press.

[4] T. V. Hanurav : Some aspects of unified sampling theory, 1966, Sankhyä, Vol. 28.

[5] W. G. Cochran : Sampling Techniques, 1963, John Wiley & Sons, Inc.
鈴木，高橋，脇本訳，第2巻，1972，東京図書株式会社．

参　考　資　料

大蔵省法人統計季報及び年報，財政金融統計月報及びその基礎資料
昭和48年10月23日付蔵証第2894号
昭和48年10月13日付行管承統第339号