

# ある最適多重層化二段抽出法について

—昭和45年国富調査法人，個人企業資産調査に  
おける標本設計の理論と実際(Ⅱ)—

田 口 時 夫

(1972年4月 受付)

On Some Optimum Stratified  
Two Stage Sampling Design with Simple  
Cost Function (II)\*

Tokio Taguchi

At the first section in this paper we report the sampling design applied to enterprise and unincorporated enterprise survey for national wealth survey in Japan practised in 1970. The second section is concerned with some calculations on the allocation of sample and the predicted sampling errors.

Furthermore we treat the structure of population and examine effects of stratification at § 3.

The Institute of Statistical Mathematics

## は し が き

同上の標題をもつ前篇(1)(統計数理研究所彙報第18巻第2号 No. 34, p. 77~90)に於て筆者は法人及び個人企業抽出のための理論的検討と設計の構想を示した。

然し乍ら云うまでもなく構想を実現する場合には種々の現実的条件によって制約される。本篇では先ず § 1 に於て抽出の実行形態を示し，ついで § 2 に於て抽出に際して必要とした割当率の計算と予測精度を示した。

更に § 3 に於て設計の前提とした母集団の構造を資料によって確認すると共に今後の問題点，及び地域特性化についての要請を加えた。

§ 3 の分析は本来ならば設計前に行われるべきものであるが，これが事後的に抽出した標本をもとにして行われざるを得ない処にも又一つの問題があるが，こゝに聯が検討を加え今後の指針に供したい。

むすびに於ては，将来実施されうる抽出形態の予想を加えたが，その充分な展開は次回に行う予定である。

## § 1 昭和45年国富調査法人企業資産調査における法人及び個人企業の標本抽出法の具体的形態

### (一) 営利法人の標本抽出法

さきに発表した営利法人企業標本抽出に関する提案は(後掲[1]参照)，さらに実情に照して分析検討した結果，次のように具体化された。

#### 1. 抽出の方針

抽出の方法は，地域及び営利法人企業をそれぞれ第一次抽出単位及び第二次抽出単位とする層化二段抽出法に，単純費用関数を前提とする最適抽出法を適用した。その形態は原則的に第

\* see [1]

一次案 ([1] 参照) によるものである。

## 2. 調査対象

調査対象は、昭和44年度事業所センサスにおいて用いた事業所調査区及び同センサスにより把握された法人企業である。その所在は第1表に示される。

## 3. 抽出単位

上記調査対象における事業所調査区及び法人企業を、それぞれ第一次及び第二次抽出単位とする。

## 4. 標本数

第一次標本数は約2,000調査区とする。

第二次標本数は約20,000客体とする。

その細目は別表3のごとくである。

## 5. 層化基準

法人の層化は、先ず

(一) 地域ブロックによる層化

即法人の所在する地域により次の9ブロックに層化した。

(i) 北海道 (ii) 東北 (iii) 関東 (iv) 中部 (v) 北陸 (vi) 近畿 (vii) 中国 (viii) 四国 (ix) 九州を行い次に

(二) 資本金額による法人の層化

(i) 資本金額1億円以上の企業層

(ii) 資本金額1億円未満の企業層

とし、更に(二)(ii)層につき次のように地域層化を行った。

(三) 資本金額に基づく地域特性による層化

(i) 資本金額1千万円以上の企業の所在する調査区層 (A<sub>1</sub>)

(ii) 資本金額1千万円未満の企業のみ所在する調査区層 (A<sub>2</sub>)

とし、更に次の層化を加えた。

四 産業特性に基づく地域特性による層化

(三)の各層は更に製造業を中心とする産業特性により、次の三階級に分類する。

コード	特 性
05	製造業が全事業所の3割未満を占める調査区
06	製造業が3割~6割を占める調査区
07	製造業が7割以上を占める調査区

又(三)のA<sub>1</sub>に属する各層は、第二次抽出の際に更に、

第1表 昭和44年事業所統計調査特性別調査区数

(i) ブロック別構成

	05	06	07	合 計
1 北海道	3,751	86	6	3,663
2 東 北	8,059	438	132	8,629
3 関 東	16,655	3,144	535	20,334
4 北 陸	2,177	396	132	2,705
5 東 海	8,949	2,113	608	11,670
6 近 畿	9,162	2,117	644	11,923
7 中 国	4,885	387	131	5,403
8 四 国	3,024	173	19	3,216
9 九 州	8,221	308	105	8,634
全 国	64,703	9,162	2,312	76,177

## (ii) 府県別構成 (欄内の数字は調査区数)

区分 都道府県名	0 (資本金1千万円以上の法人が0である調査区)				1以上 (資本金が1千万円以上の法人がある調査区)				合計
	05 ( $\frac{1}{25}$ 抽出)	06 ( $\frac{1}{5}$ 抽出)	07 ( $\frac{1}{1}$ 抽出)	計	05 ( $\frac{1}{50}$ 抽出)	06 ( $\frac{1}{5}$ 抽出)	07 ( $\frac{1}{1}$ 抽出)	計	
北海道	89	8	2	99	27	9	4	40	139
青森	28	2	2	32	4	1	—	5	37
岩手	32	1	—	33	5	1	1	7	40
宮城	33	3	1	37	7	4	—	11	48
秋田	30	2	6	38	4	1	—	5	43
山形	27	5	—	32	6	7	1	14	46
福島	39	11	3	53	7	4	—	11	64
茨城	43	10	3	56	6	3	1	10	66
栃木	28	25	33	86	5	13	11	29	115
群馬	27	23	42	92	7	20	13	40	132
埼玉	40	39	16	95	11	44	20	75	170
千葉	48	4	—	52	9	7	5	21	73
東京都	118	128	80	326	81	261	188	530	856
神奈川県	59	7	5	71	24	27	20	71	142
新潟	46	26	87	159	10	19	31	60	219
富山	24	3	—	27	6	5	—	11	38
石川	16	24	47	87	6	20	21	47	134
福井	16	17	52	85	4	11	12	27	112
山梨	14	13	87	114	3	4	11	18	132
長野	37	12	5	54	10	15	4	27	83
岐阜	26	52	117	195	7	28	46	81	276
静岡	45	48	94	187	14	41	23	78	265
愛知	61	75	148	284	29	111	130	270	554
三重	30	15	8	53	7	10	2	19	72
滋賀	17	8	22	47	4	7	9	20	67
京都	29	34	92	255	12	35	60	107	362
大阪	79	76	68	223	40	166	193	399	622
兵庫県	62	26	56	144	21	37	34	92	236
奈良	13	16	19	48	3	12	8	23	71
和歌山	22	13	11	46	5	8	4	17	63
鳥取	13	1	—	14	3	2	—	5	19
島根	24	2	—	26	2	1	—	3	29
岡山	28	17	52	97	7	10	31	48	145
広島	38	21	30	89	12	19	18	49	138
山口	30	2	—	32	7	2	—	9	41
徳島	21	3	2	26	3	3	3	9	35
香川	15	5	10	30	5	7	2	14	44
愛媛	28	6	—	34	7	5	1	13	47
高知	22	4	1	27	2	2	—	4	31
福岡	68	16	14	98	17	11	6	34	132
佐賀	18	2	2	22	3	1	—	4	26
長崎	33	6	16	55	5	1	—	6	61
熊本	37	3	—	40	4	1	—	5	45
大分	27	1	—	28	4	1	—	5	33
宮崎	26	1	—	27	3	1	1	5	32
鹿児島	39	15	62	116	4	2	4	10	126
計	1,645	831	1,395	3,871	472	1,000	918	2,390	6,261

第2表 法人企業数による調査区の分布の例 (数値は企業数)

地 域	調 査 区 数			1千万円以上の企業を有する調査区				1千万円以上の企業を有しない調査区	
				全 企 業		資本金額1千万円以上企業のみ			
	総 数	1千万円以上の企業を有する調査区	有しない調査区	平均企人数	標準偏差	平均企人数	標準偏差	平均企人数	標準偏差
宮 城	48	11	37	16.72	8.18	2.09	0.89	9.24	6.57
東 京 (区別)	12	11	1	50.73	46.52	5.91	4.48	57.00	0.00
	15	13	2	47.31	43.67	9.15	9.81	22.00	10.00
	12	11	1	38.82	32.37	4.73	4.39	40.00	0.00
	18	13	5	34.38	30.37	3.62	2.84	16.60	6.07
	26	19	7	32.95	33.63	1.84	1.38	24.86	9.42
	53	32	21	30.03	26.42	2.54	1.48	21.71	9.73
56	39	17	20.82	17.90	1.77	1.23	15.12	6.66	
岡 山	145	49	96	14.46	17.57	1.96	1.47	8.72	6.13

第3表 営利法人の標本調査区の構成

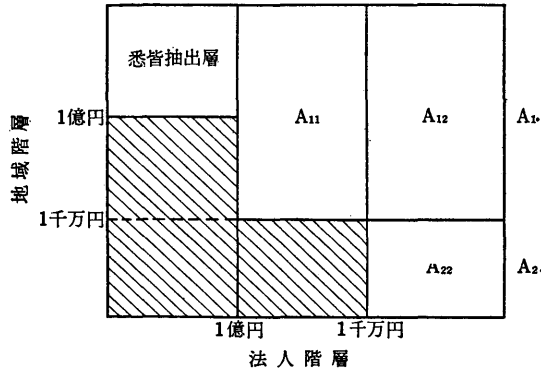
- (i) 資本金額1億円以上の企業の所在する標本調査区の抽出率 1/1  
約 6000 客体
- (ii) 資本金額1千万円以上の企業を有する調査区層からの第一次標本の抽出率

産業 code	標本調査区数	抽 出 率
05 (製造業が3割未満を占める調査区層)	196	1/100
06 (製造業が3割以上6割未満を認める調査区層)	370	1/10
07 (製造業が7割以上を占める調査区層)	317	1/2
計	883	

- (iii) 資本金額1千万円未満の企業のみを有する調査区層からの第一次標本の抽出率

産業 code	標本調査区数	抽 出 率
05	125	1/300
06	112	1/30
07	192	1/6
計	429	
総 計	1,912	

但し、東京、大阪、名古屋、京都は上記の抽出率×1/2とした結果である。



第1図 法人の層化

第4表 市区町村を抽出単位とした場合の営利法人の母集団規模

	$M_1$	$M_2$	$N_{11}$	$N_{12}$	$N_{22}$
北海道	147	72	2,112	20,640	977
東北	318	212	4,308	37,554	1,708
関東	441	138	17,199	201,114	2,655
中部	343	120	7,749	76,881	1,113
北陸	96	19	2,572	14,526	111
近畿	300	75	13,968	99,936	648
中国	212	144	3,120	30,192	1,371
四国	132	88	2,127	21,339	876
九州	275	260	4,365	49,275	2,420

注  $M_1$  は資本金額1千万円以上の企業を有する市町村数  
 $M_2$  は資本金額1千万円未満の企業をのみ有する市町村数  
 $N_{11}$  は  $M_1$  に属し、且つ資本金額1千万円以上の企業数  
 $N_{12}$  は  $M_1$  に属し、且つ資本金額1千万円未満の企業数  
 $N_{22}$  は  $M_2$  に属する企業数

第5表 各層の抽出比及び計算精度

	$p_1$	$p_2$	$q_{11}$	$q_{12}$	$q_{22}$	$r_1$	$r_2$	$e$ (計算誤差)
北海道	0.11728	8.68367	135.00	1017.60	13.74	15.8328	119.3289	0.00025
東北	0.11687	6.70836	132.08	1003.68	17.49	15.4361	117.3147	0.00025
関東	0.02526	4.67297	568.70	4921.98	26.60	14.3654	124.3151	0.00023
中部	0.06034	8.76634	229.99	1923.35	13.24	13.8775	116.0606	0.00010
北陸	0.05099	8.23944	277.44	2068.63	12.80	14.1466	105.472	0.00014
近畿	0.04432	8.05614	325.93	2483.26	13.66	14.4452	110.0525	0.00010
中国	0.10101	9.47800	150.24	1175.51	12.53	15.1757	118.7488	0.00018
四国	0.09727	10.09344	158.29	1240.99	11.96	15.3969	120.7143	0.00005
九州	0.08575	4.91494	172.58	1462.81	25.52	14.7987	125.4327	0.00005

註1  $p_1 p_2$  はそれぞれ  $M_1, M_2$  に対する第一抽出比

$q_{11}$  は  $N_{11}$  に対する第二抽出比

$q_{12}$  は  $N_{12}$  に対する第二抽出比

$q_{22}$  は  $N_{22}$  に対する第二抽出比

註2  $r_1 = p_1 q_1$

$r_2 = \frac{p_1 q_{12} + p_2 q_{22}}{2}$  とすれば理論的仮定は  $r_2 = p_1 q_{12} = p_2 q_{22}$  であるから

$e = \frac{|p_1 q_{12} - p_2 q_{22}|}{r_2}$  は計算精度を示す

第6表 調査旅費の変動による抽出比の変化

	$c_1=1000$ 円にした場合					$c_1=2000$ 円にした場合				
	$p_1$	$p_2$	$q_{11}$	$q_{12}$	$q_{22}$	$p_1$	$p_2$	$q_{11}$	$q_{12}$	$q_{22}$
北海道	0.21464	15.893	95.460	719.553	9.718	0.40111	29.700	67.501	508.801	6.872
東北	0.21389	12.277	93.392	709.715	12.364	0.39972	22.944	66.038	501.844	8.743
関東	0.04622	8.552	402.131	3480.363	18.810	0.08638	15.982	284.350	246.988	13.301
中部	0.11044	16.044	162.625	1360.015	9.362	0.20638	29.982	114.993	961.676	6.620
北陸	0.09332	15.080	196.179	1462.742	9.052	0.17439	28.180	138.720	1034.315	6.401
近畿	0.08112	14.744	230.467	1755.931	9.661	0.15159	27.553	162.965	1241.631	6.831
中国	0.18487	17.346	106.234	831.214	8.859	0.34548	32.416	75.119	587.757	6.264
四国	0.17803	18.473	111.929	877.512	8.457	0.33270	34.521	79.146	620.495	5.980
九州	0.15694	8.995	122.035	1034.364	18.047	0.29328	16.810	86.292	731.406	12.761

註  $c_1$  は一調査区当りの旅費を示す

(四) 第二次抽出の為の資本金額による層化

(i) 資本金額1千万円以上の企業層  $A_{11}$

(ii) 資本金額1千万円未満の企業層  $A_{12}$

とした。

以上の多重層化による各層のサイズは第1表のとうりである。又その図解は、第1図に示される。

## 6. 抽出率

抽出率の決定は、「1.抽出の方針」に従って算出した結果(第4,5,6表参照)に実施上の条件(1調査区当り10企業程度)を加味して第3表の如く第一次標本抽出率を決定した。

第2次標本の抽出は次のとうりである。

(i) 資本金額1千万円以上の企業層 1/1

(ii) 資本金額1千万円未満の企業層

(i)  $A_1$  層に属するもの 1/3

(ii)  $A_2$  層に属するもの 1/1

猶この抽出率の算出結果については §2 に於て詳論する。

## 7. その他

特に東京都、大阪、名古屋、京都については客体が集中するので調査能力を考慮して、第1次抽出率は第3表の抽出率  $\times 1/2$  とする。

又、第2次抽出率は同じく前項6(i)(ii)の抽出率  $\times 1/2$  とする。

以上の通りであるが、第5表の抽出比にみられるように各欄に算出された結果は、一市区町村当りの階級別法人数からみればその多くは実行不可能な数値である。

これは経済構造上階層間分散及び各階層内分散の開きが、極めて大きいことに由来するのである。従ってこうした場合には最適条件といった問題は極めて決定力に乏しいものとなり主たる決定要素は、調査員の単位調査区内での調査可能限界となる。

又第3表(ii)(iii)にみられる産業特性に基づき各層の抽出率は、これ又精度向上の問題よりは寧ろ産業構造上の特徴から標本が勢い商業部門に従事する企業に偏ることを抑制する意味で決定された。具体的に製造業及び非製造業が略同数標本に加わる事を予想して05,06,07の各産業特性層は、略同数抽出された。この抽出率による最終結果については §3 に於て更に詳論する。

(二) 非営利法人の標本抽出法

非営利法人はその規模と所在の形態からみて、営利法人のための標本調査区内にある対象を悉皆抽出するのが有利と考えられた。

### ㊦ 個人企業の標本抽出法

後掲 [1] に述べた個人企業標本抽出に関する第一次案は、さらに分析検討した結果、次のように具体化された。

#### 1. 抽出の方針

抽出の方法は、地域および個人企業をそれぞれ第一次抽出単位および第二次抽出単位とする層化二段抽出法による。

#### 2. 調査対象

調査対象は、昭和 44 年度事業所センサスにおいて用いた事業所調査区および調査によりは握された個人企業である。その規模は第 7 表に示される。

第 7 表 個人企系の母集団の構成 (数値は調査区数)

ブロック名	コ ー ド			計
	05	06	07	
北海道	3,571	86	6	3,663
東北	8,059	438	132	8,629
関東	16,655	3,144	535	20,334
北陸	2,177	396	132	2,705
東海	8,320	2,039	577	10,936
近畿	9,791	2,191	675	12,657
中国	4,885	387	131	5,403
四国	3,024	173	19	3,216
九州	8,221	308	105	8,634
計	64,703	9,162	2,312	76,177

#### 3. 抽出単位

上記調査対象における事業所調査区および個人企業を、それぞれ第一次および第二次抽出単位とする。

#### 4. 標本数

第一次標本数は、約 2,000 調査区とする。

第二次標本数は、約 20,000 客体とする。その細目は第 8 表のごとくである。

第 8 表 個人企業の標本の構成 (数値は調査区数)

ブロック名	コ ー ド			計
	05	06	07	
北海道	36	6	6	48
東北	81	29	43	153
関東	167	207	178	552
北陸	22	27	44	93
東海	83	137	192	412
近畿	98	146	225	469
中国	49	26	44	119
四国	31	11	6	48
九州	82	22	35	139
計	649	611	773	2,033

#### 5. 層化基準

調査区の層化基準は、製造業を中心とする産業特性により、次の三階級に分類する。

コード	特 性
05	製造業が全事業所の3割未満を占める調査区
06	製造業が3割～6割を占める調査区
07	製造業が7割以上を占める調査区

この産業特性による調査区のブロック別分布は、第7表のとおりである。

猶小数標本によりこの産業特性による調査区の市町村別分布を調べた結果、市部において07が郡部において05が相対的に多いことが示された。

したがって、以上の層化は、地域特性は握において有効であると認められる。

## 6. 抽出比

抽出率は、資産総額、資産額平均および分散のおのおのにつき、特に製造業が全産業に対し占める地位が大である点から、5の各層ごとに次のように決定した。

コード	抽出率
05	$\frac{1}{100}$
06	$\frac{1}{15}$
07	$\frac{1}{3}$

さらに、各事業所調査区は概数60～70の事業所を含み、かつ、その構成は平均して約9割が個人企業に属すると考えられるので、第二次客体抽出においては、各調査区ごとに、建設、製造、卸売、小売、サービスの各産業につき $\frac{1}{5}$ 抽出その他産業は悉皆抽出とし、各調査区標本数がほぼ10企業程度になるように決定した。

## 7. その他

特に東京都においては必然的に標本数が集中するので、第一次標本をさらに $\frac{1}{2}$ 抽出し、それを標本調査区とした。

また、北海道については、地域表章のため07の調査区の抽出率を $\frac{1}{1}$ とした。

## §2 法人企業の標本設計における標本の割当と精度の予測

§1で決定した層化方法は必要最少限の要請によるものと云えるが、この極度に簡略化した方式に於ても特に調査区に関する資料が事前に得難い為多くの制約を伴うのである。

従って抽出比の決定及び予測精度の計算に於ては実際の標本構造に対し次のような若干の変更を加えざるを得なかった。

1. 算出の基礎に用いたのは資本金額である。
2. 第一次抽出単位は事業所調査区の代りに市区町村を用いて算出した。
3. その際標本市区町村に含まれる調査区数及び法人数が§1の設計通りそれぞれ約2,000調査区及び20,000客体を占めるように市町村標本数を決定した。

4. 以上の計算に於ては地域産業特性を考慮していない。

又東京、大阪、名古屋、京都について標本抽出比を高めることを考慮していない。

この第一次抽出単位の変更がどの程度結果に影響するかについては§3に述べることにする。更に計算方法の細部については以下の方式に従った。

5. 各計算に於ける各資本金額階級の代表値は昭和40年度事業所センサスによる階級別資本金額の全国平均を用いた。

従って例えば第10表における北陸地方の推定精度が他地方に比較して若干低下しているが、それはこの地方の法人数が少ないことと共に以上の点を考慮する必要がある。

6. 第10表における推定精度は資本金額1億円未満の企業層のみについてであり、従って



第9表 ブロック別地域階層別市町村間分散  $V_{bih}$  及び資本金階級別内分散  $V_{wihk}$  (単位1万円)

	$V_{w11}$	$V_{w12}$ $V_{w22}$	$V_{b1}$ $V_{b2}$
北海道	1,632,353	28,987 22,364	$0.1488105 \times 10^{11}$ $0.3012752 \times 10^7$
東北	1,714,869	30,110 20,566	$0.1498395 \times 10^{11}$ $0.4705412 \times 10^7$
関東	1,980,622	26,526 20,025	$0.32037483 \times 10^{12}$ $0.97409377 \times 10^7$
中部	2,121,515	30,472 20,805	$0.5616224 \times 10^{11}$ $0.2851393 \times 10^7$
北陸	2,041,838	36,740 35,165	$0.7864317 \times 10^{11}$ $0.3214739 \times 10^7$
近畿	1,957,812	33,775 26,258	$0.10409183 \times 10^8$ $0.3374673 \times 10^7$
中国	1,774,092	29,278 22,385	$0.2005195 \times 10^{11}$ $0.2487317 \times 10^7$
四国	1,723,366	28,476 17,374	$0.2160668 \times 10^{11}$ $0.2178262 \times 10^7$
九州	1,865,591	26,209 21,060	$0.2781754 \times 10^{11}$ $0.8653180 \times 10^7$

第10表 地域別推定標本精度 (単位1万円)

	( $p_1=2, p_2=6, q_{11}=10, q_{12}=30, q_{22}=10$ とした場合 )			
	$X$	$\sigma_x^2$	$\sigma_x$	$\sigma_{x/z}$
北海道	8,877,676	$0.228645 \times 10^{18}$	1,512,100	0.17033
東北	17,620,000	$0.497035 \times 10^{18}$	2,229,427	0.12653
関東	74,906,721	$0.142217 \times 10^{15}$	11,925,497	0.15920
中部	33,896,131	$0.196984 \times 10^{14}$	4,438,288	0.13094
北陸	9,490,741	$0.767574 \times 10^{18}$	2,770,513	0.29192
近畿	54,139,009	$0.319177 \times 10^{14}$	5,649,579	0.10435
中国	13,189,192	$0.440536 \times 10^{18}$	2,098,896	0.15914
四国	8,854,778	$0.295721 \times 10^{18}$	1,719,655	0.19421
九州	19,013,175	$0.788531 \times 10^{18}$	2,808,079	0.14769
全体	239,987,423	$0.224014 \times 10^{15}$	14,967,098	0.06237

1億円以上の悉皆抽出による企業層を考慮すれば全精度は大巾に向上する。

又事業所単位で地域を集計する場合は、地方支社は1億円以上の本社(関東, 近畿に集中)に附随して悉皆調査される率が高いからこの集計結果についても精度は第10表よりも高いものと思われる。

7. 因みに抽出比, 及び精度の計算に用いた各式は次の通りである。

$$(1) \quad p_h^{(0)*} = \sqrt{\frac{\lambda C_{h1} \sum_{i=1}^9 M_{ih}}{\sum_{i=1}^9 \left[ \beta_{ih} \cdot - \sum_{k=1}^2 w_{ihk} \right]}}$$

\* 後掲 [1] p.85 (45) (47) (48) 式参照

$$(2) \quad q_{hk}^{(0)*} = \sqrt{\frac{\sum_{i=1}^9 \sum_{h=1}^2 C_{h2} N_{ihk}}{C_{h1} \sum_{i=1}^9 M_{ih}}} \sqrt{\frac{\sum_{i=1}^9 \left\{ \beta_{ih\cdot} - \sum_{k=1}^2 w_{ihk} \right\}}{\sum_{i=1}^9 \sum_{h=1}^2 w_{ihk}}}$$

$$(3) \quad \sqrt{\lambda}^* = \frac{1}{C} \left[ \sum_{h=1}^2 \sqrt{\sum_{i=1}^9 \left( \beta_{ih\cdot} - \sum_{k=1}^2 w_{ihk} \right) C_{h1} \sum_{i=1}^9 M_{ih}} \right. \\ \left. + \sum_{k=1}^2 \sqrt{\left( \sum_{i=1}^9 \sum_{h=1}^2 w_{ihk} \right) \left( \sum_{i=1}^9 \sum_{h=1}^2 C_{h2} N_{ihk} \right)} \right]$$

$$(4) \quad **X = \sum_{i=1}^9 \sum_{h=1}^2 \hat{p}_h \sum_{k=1}^2 q_{hk} X_{ihk} + X_0$$

$$(5) \quad **\sigma_x = \sum_{i=1}^9 \sum_{h=1}^2 \sum_{k=1}^2 \hat{p}_h (q_{hk} - 1) \sum_{j=1}^{Mih} N_{ihkj} \sigma_{ihkj}^2 \\ + \sum_{i=1}^9 \sum_{h=1}^2 (\hat{p}_h - 1) M_{ih} \sigma_{bih}^2$$

である。

こゝで  $l$  はブロック,  $h$  は資本金額による地域階層,  $k$  は資本金額による法人階層を示す。

8. 第10表の計算は第9表の結果に基づいて

$$\hat{p}_1 = 2 \quad \hat{p}_2 = 6 \\ q_{11} = 10 \quad q_{12} = 30 \quad q_{21} = 10$$

を採用した結果である。

つまり市町村を単位としたこの結果は調査区を単位とした第3表の実行抽出率と当然異なるが次の比率

$$(6) \quad \frac{\hat{p}_2}{\hat{p}_1} = 3, \quad \frac{q_{12}}{q_{11}} = 3$$

は同一であるとして

産業特性及び調査限界を考慮した上で事業所を抽出単位とする実行標本抽出率の決定に採用したことによる。

### §3 法人母集団の性質と層別の効果

本篇及び前篇に於て我々は母集団の性格に対して二、三の仮定を置くことにより標本の設計を考察した。

此等の仮定は事前の実証することが困難であったが、抽出された標本についてその二、三を吟味することが可能となった。

以下はその結果の概説である。

#### (一) 地域抽出単位の大きさについて

今回の法人調査に用いた地域抽出単位は事業所調査区である。

一般に第一次抽出単位については、その地域的拡がりがある一定の範囲に止まること及び各単位内の第二次抽出単位の数が基準数量に整一されていることが利用上有用であると認められている。

従って事業所調査区に於ても60乃至70の事業所を含むという設定基準をもつものではあるが、これが直ちに法人数の整一性を表現しているわけではない。事実第2図及び第3図はそれ

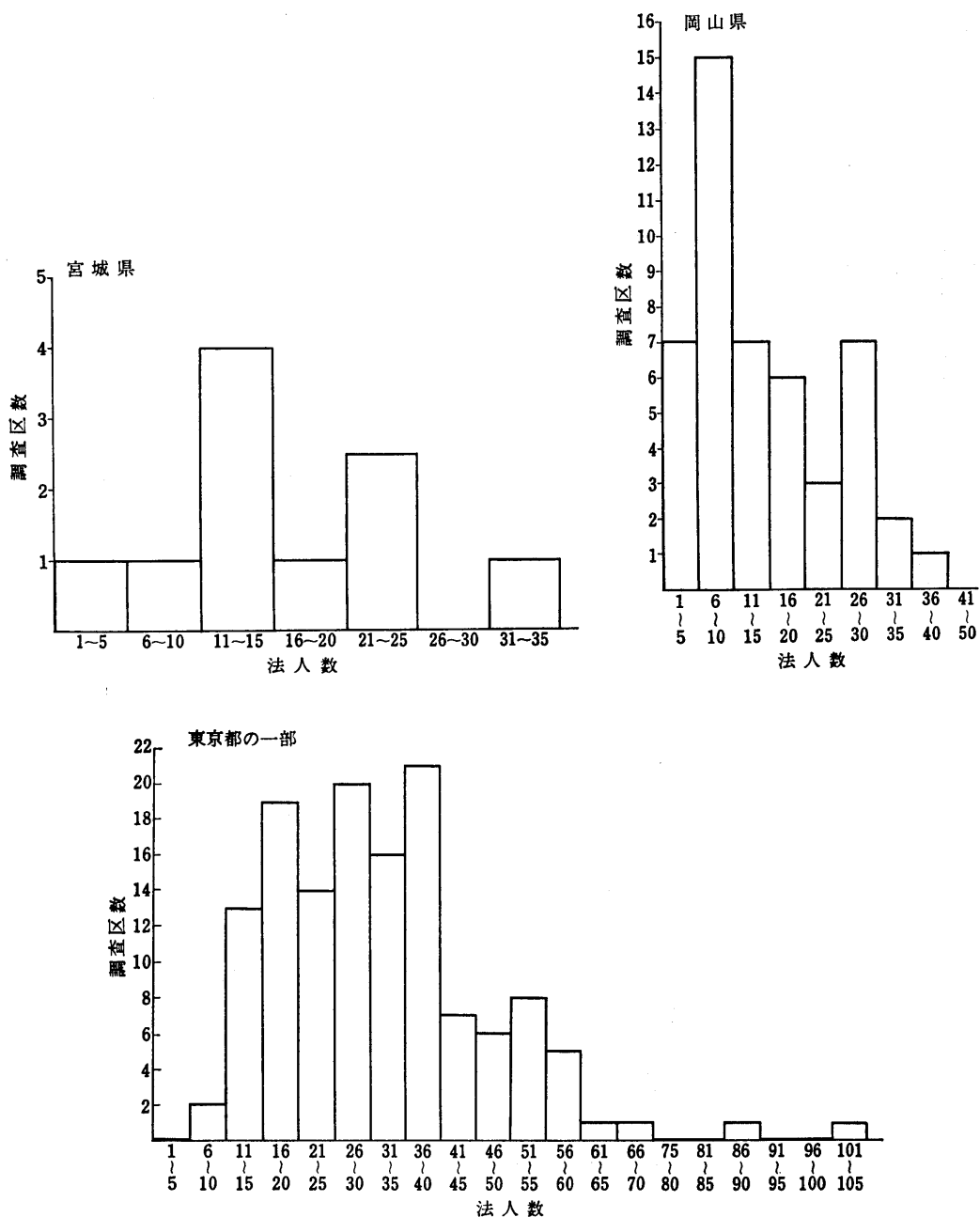
\*\* 後掲 [1] p. 89 (54) (55) 式参照

第11表 所在法人数による調査区の分布表

1) 資本金額1千万円の法人を以上を含む調査区

(a) 法人総数による調査区の分布

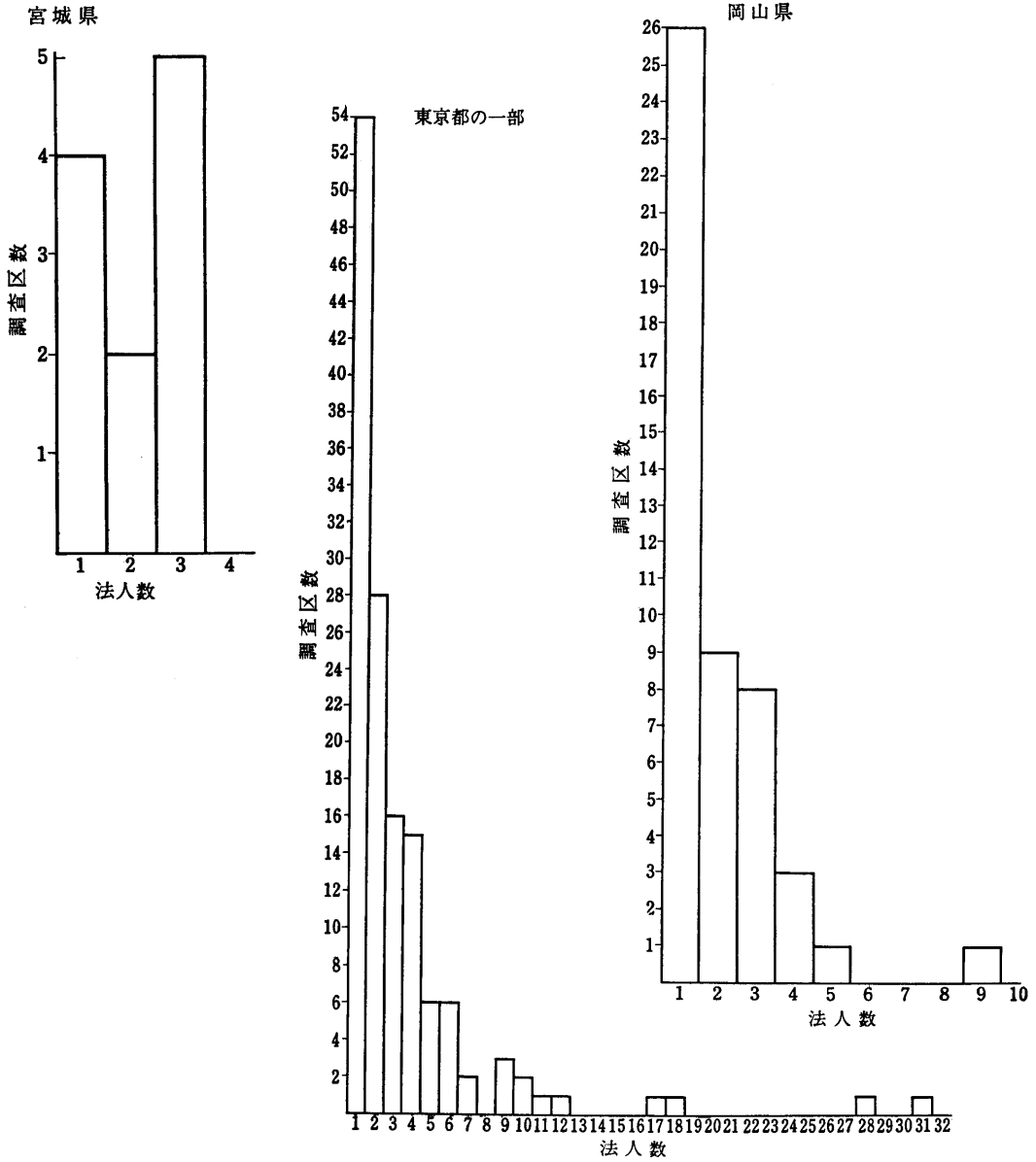
地 域	法人数																
	1-5	6-10	11-15	16-20	21-25	26-30	31-35	36-40	41-45	46-50	51-55	56-60	61-65	66-70	86-90	101-105	
宮 城 県	1	1	4	1	3		1										
東 京 都		2	13	19	14	2	16	21	7	6	8	5	1	1	1		
岡 山 県	7	15	7	6	3	7	2	1								1	



第1図 (i) 第11表の1) a)のグラフ

(b) 資本金額1千万以上の法人数による調査区の分布

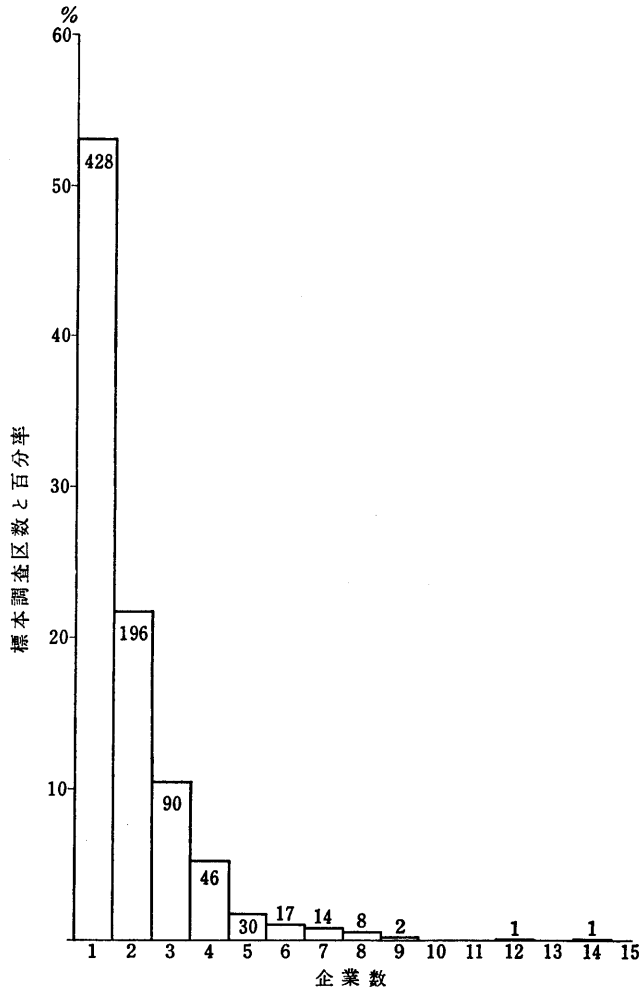
地域		法人数															
		1	2	3	4	5	6	7	8	9	10	11	12	17	18	28	31
宮	城	4	2	5													
東	京	54	28	16	15	6	6	2		3	2	1	1	1	1		
岡	山	26	9	8	3	1				1							



第3図 (ii) 第13表の1) bのグラフ

2) 資本金額1千万円以上の法人を含め調査区  
法人総数による調査区の分布

地域	法人数									
	0	1~5	6~10	11~15	16~20	21~25	26~30	31~35	36~40	
宮城	2	11	11	7	1	4	1			
東京	1	2	7	10	9	9	7	6	1	
岡山	4	31	31	17	8	3	2	1		



第2図 (iii) 資本金額1千万円以上の企業数による調査区の分布 (全国の資本金額1千万以上の法人を有する調査区を対象とするもの)

を裏書きするもので特に大都市に於て格差が著しい。従って直接法人数を基準内に揃えた法人調査区の設定は逆に調査区の地理的広狭の格差を著しく増大させることになる。又、資本金による内分数の都市部と郡部での格差はこれによって消滅する訳ではない。然しそれ等の事情にも拘らず2~4程度の事業所調査区を合併した法人調査区の設定は、標本抽出比  $p_1, p_2, q_{11}, q_{12}, q_{22}$  の決定を理論的算定値に幾らかでも近づける可能性を与える点で精度上からは望ましいことといえる。一方調査区の拡大は調査員の負担を増大させるものであるから、その点を考慮した現実的規模決定がなされねばならない。

□ 地域抽出単位の資本金額特性と問題点

前篇 [1] p.86 の (51) (52) (53) 式及び (52) (53) 式の補正 (第 19 巻第 1 号 No.35 p.43) に於て筆者は法人調査に用いた資本金額による調査区の層化について次の予想を行った。

先ず地域抽出比について

$$(7) \quad p_1 < p_2 < \dots < p_i \text{ ([1] における (51) 式)}$$

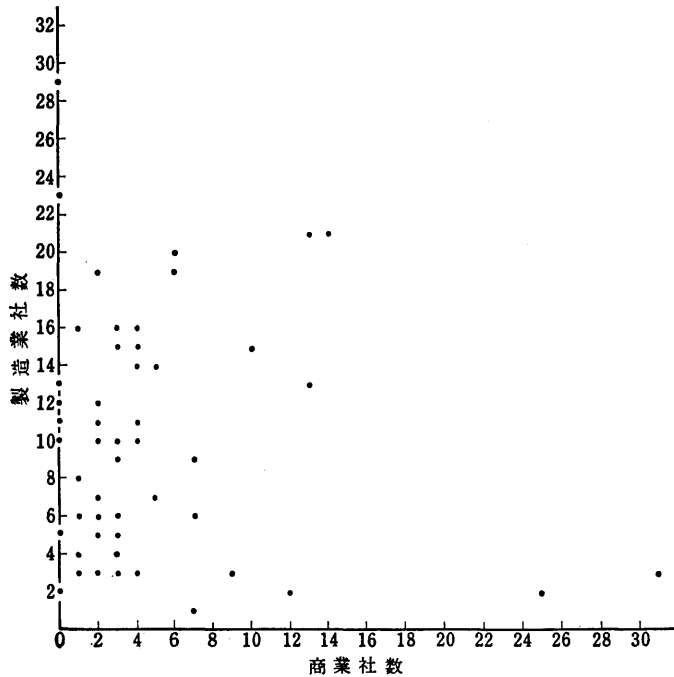
であり、第二次抽出比について

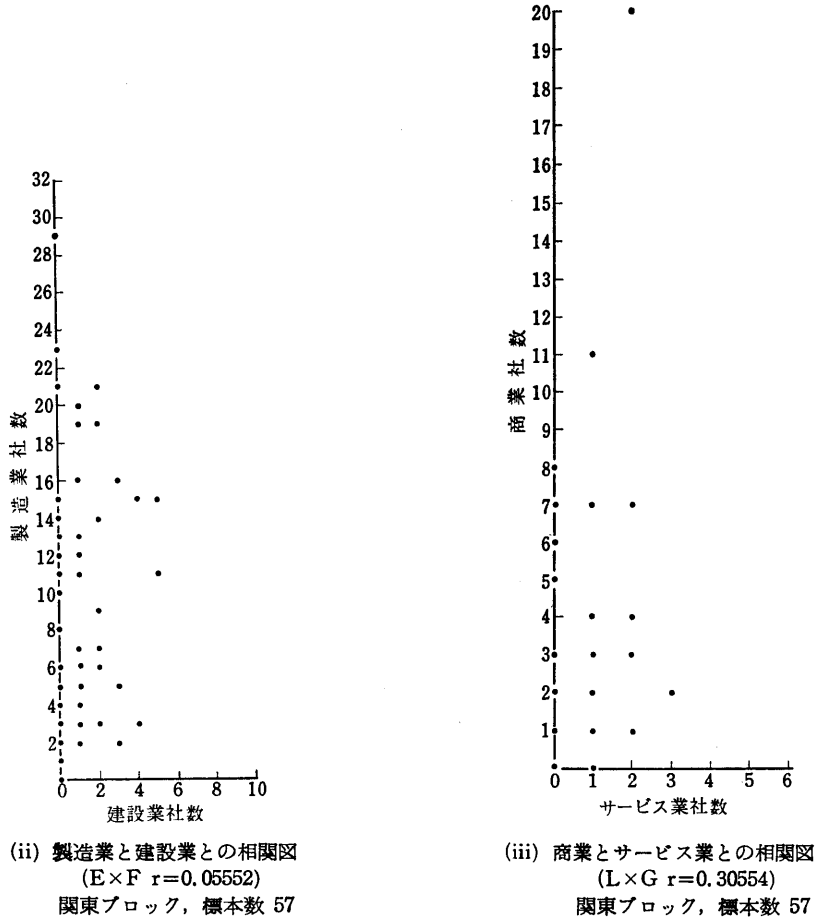
第 12 表 ブロック別調査区内の各産業法人数間の相関

	製造業×商業	製造業×建設業	商業×サービス業	製造業×その他の業種	サービス業×その他の業種
1 北海道 (27)	-0.34186	-0.05997	0.45484	-0.21550	-0.13836
2 東北 (115)	-0.03299	0.17478	0.35077	-0.03794	-0.03520
* 関東 (326)	-0.06030	0.03560	0.29887	-0.00978	0.18248
4 北陸 (37)	-0.05677	-0.10558	0.48513	-0.02233	-0.07332
5 中部 (256)	-0.04161	0.12827	0.32740	-0.03377	-0.46658
6 近畿 (297)	0.03110	0.18788	0.07182	0.09143	-0.10621
7 中国 (83)	-0.12336	0.16356	0.35545	-0.02815	0.12553
8 四国 (34)	-0.05643	0.11499	0.23508	-0.116015	-0.08497
9 九州 (65)	0.11410	0.32461	0.26195	0.13421	0.02637

注 ( ) 内は標本調査区数

(i) 製造業と商業との相関 (G×F  $r = -0.02637$ )  
 関東ブロック, 標本数 57





第3図 第12表の相関図

第13表 ブロック別産業別標本構成 (抽出法人数とその構成比率)

ブロック名 (括弧内は標本調査区数)	製造業	商業	建設業	サービス業	その他の業種	合計
1 北海道 (27)	87. (28.618%)	130. (42.763%)	23. (7.566%)	24. (7.895%)	40. (13.158%)	304. (100.000%)
2 東北 (115)	415. (45.109%)	334. (36.304%)	58. (6.304%)	62. (6.739%)	51. (5.543%)	920. (100.000%)
* 関東 (326)	3145. (59.139%)	1349. (25.367%)	332. (6.243%)	257. (4.833%)	235. (4.419%)	5318. (100.000%)
4 北陸 (37)	198. (63.462%)	65. (20.833%)	14. (4.487%)	18. (5.769%)	17. (5.449%)	312. (100.000%)
5 中部 (256)	1786. (71.241%)	436. (17.391%)	142. (5.664%)	103. (4.108%)	40. (1.596%)	2507. (100.000%)
6 近畿 (297)	2021. (70.443%)	505. (17.602%)	130. (4.531%)	105. (3.834%)	103. (3.590%)	2869. (100.000%)
7 中国 (83)	399. (59.910%)	147. (22.072%)	53. (7.958%)	36. (5.405%)	31. (4.655%)	666. (100.000%)
8 四国 (34)	175. (52.239%)	68. (20.298%)	20. (5.970%)	12. (3.582%)	60. (17.910%)	335. (100.000%)
9 九州 (65)	199. (34.489%)	233. (40.381%)	38. (6.586%)	37. (6.412%)	70. (12.132%)	577. (100.000%)

(8)  $q_{1k} > q_{2k} > \dots > q_{kk}$  ([1]における(52)式の補正)である。

市区町村を第一次抽出単位として算出した前掲第5表及び第6表は  $t=s=2$  の特殊な場合であるが、cost の如何に関係なくこの関係が成立することを明瞭に読みとることが出来る。

処で以上 (7) (8) の予想は更に遡ると母集団に関する次の予想が根底になるのである。

つまりまづ各層の客体数につき

$$(9) \quad N_{hh} < N_{hh+1} < \dots < N_{ht}; h = 1, 2, \dots, t$$

が成立し、又調査区層のサイズにつき

$$(10) \quad M_1 < M_2 < \dots < M_t \text{ ([1] (53) 式の補正)}$$

が成立する。

一方各層内の分散につき

$$(11) \quad \bar{\sigma}_{whh} > \bar{\sigma}_{whh+1} > \dots > \bar{\sigma}_{whi}; h = 1, 2, \dots, t$$

が成立し、又各層間の分散につき

第14表 ブロック別標本調査区当り産業別平均標本法人数及び標準偏差

	製 造 業 (F)		商 業 (G)		建 設 業 (E)		サ ー ビ ス 業 (L)	
	M(平均)	S(偏差)	M(平均)	S(偏差)	M(平均)	S(偏差)	M(平均)	S(偏差)
1 北海道 (27)	3.89048	3.22222	5.90064	4.81481	0.97006	0.85185	1.16534	0.88889
2 東北 (115)	4.28688	3.60870	3.83116	2.90435	0.87848	0.50435	1.07369	0.53913
* 関東 (326)	7.79618	9.64724	5.00515	4.13804	1.34972	1.01840	1.31345	0.78834
4 北陸 (37)	4.75461	5.35135	2.38696	1.75676	0.91094	0.37838	1.10580	0.48649
5 中部 (256)	5.11426	6.97656	2.56588	1.70313	1.47271	0.55469	1.50463	0.40234
6 近畿 (297)	6.15913	6.80471	2.47161	1.70034	0.77220	0.43771	0.81497	0.37037
7 中国 (83)	4.98784	4.80723	2.42142	1.77108	1.12559	0.63855	0.73128	0.43373
8 四国 (34)	4.44663	5.14706	2.46147	2.00000	0.80869	0.58824	0.76244	0.35294
9 九州 (65)	5.00731	3.06154	3.81452	3.58461	0.99061	0.58462	0.96028	0.56923

第15表 ブロック別産業別田集団構成 (法人数とその構成比率)

	全産業	製 造 業	商 業	建 設 業	サービス業	その他の産業
北海道	30,835	4,956 (16.073)	14,034 (45.513)	3,830 (12.421)	523 (1.696)	7,492 (24.297)
東北	31,506	7,286 (23.005)	13,806 (43.992)	3,529 (11.201)	482 (1.530)	6,382 (20.272)
関東	257,586	78,408 (30.439)	105,768 (41.061)	21,418 (8.315)	5,970 (2.318)	46,025 (17.868)
中部	86,353	32,750 (37.926)	32,614 (37.768)	6,620 (7.666)	1,271 (1.472)	13,098 (15.168)
北陸	8,860	3,001 (33.871)	3,292 (37.156)	900 (10.158)	121 (1.366)	1,546 (17.449)
関西	110,146	41,574 (37.744)	40,890 (37.123)	8,139 (7.389)	2,695 (2.447)	16,848 (15.296)
中国	36,421	10,499 (28.827)	15,122 (41.520)	3,628 (9.961)	551 (1.513)	6,621 (18.179)
四国	21,538	5,617 (26.079)	9,250 (42.947)	1,840 (8.543)	208 (0.966)	4,623 (21.464)
九州	44,901	9,733 (21.677)	19,952 (44.436)	5,015 (11.169)	801 (1.784)	9,400 (20.935)
全国合計	628,148	193,786 (30.850)	254,782 (40.561)	54,919 (8.743)	12,622 (2.009)	112,040 (17.837)

下段は百分率



第 16 表 産業別平均抽出率 =  $\frac{\text{標本法人数 (資本金額 1 億円未満のみ)}}{\text{母集団法人数}}$

	全 産 業		製 造 業		商 業		建 設 業		サ ー ビ ス 業		そ の 他 の 産 業	
	抽出率	標本サイズ 母集団サイズ	抽出率	標本サイズ 母集団サイズ	抽出率	標本サイズ 母集団サイズ	抽出率	標本サイズ 母集団サイズ	抽出率	標本サイズ 母集団サイズ	抽出率	標本サイズ 母集団サイズ
北 海 道	0.0099	304 30,835	0.0176	87 4,956	0.0092	130 14,034	0.0060	23 3,830	0.0459	24 523	0.0053	40 7,492
東 北	0.0292	920 31,506	0.0570	415 7,286	0.0241	334 13,860	0.0164	58 3,529	0.1286	62 482	0.0080	51 6,382
関 東	0.0206	5,318 257,586	0.0401	3,145 78,408	0.0128	1,349 105,768	0.0155	332 21,418	0.0430	257 5,970	0.0051	235 46,025
中 部	0.0290	2,507 86,353	0.0545	1,786 32,750	0.0134	436 32,614	0.0215	142 6,620	0.0810	103 1,271	0.0031	40 13,098
北 陸	0.0352	312 8,860	0.0660	198 3,001	0.0197	65 3,292	0.0156	14 900	0.1488	18 121	0.0110	17 1,546
近 畿	0.0261	2,874 110,146	0.0486	2,021 41,574	0.0124	505 40,890	0.0160	130 8,139	0.0400	105 2,695	0.0067	113 16,848
中 国	0.0183	666 36,421	0.0380	399 10,499	0.0097	147 15,122	0.0146	53 3,628	0.0653	36 551	0.0047	31 6,621
四 国	0.0156	335 21,538	0.0312	175 5,617	0.0074	68 9,250	0.0109	20 1,840	0.0577	12 208	0.0130	60 4,623
九 州	0.0124	577 44,901	0.0204	199 9,733	0.0117	233 19,952	0.0076	38 5,015	0.0462	37 801	0.0074	70 9,400
全 国	0.0220	13,813 628,148	0.0280	5,425 193,786	0.0128	3,267 254,782	0.0147	810 54,919	0.0518	654 12,622	0.0059	657 112,400

について、抽出率の算出に当たっては、母集団の抽出率を

(12)  $\sigma_{b1} > \sigma_{b2} > \dots > \sigma_{bt}$ . ([1] (53) 式の補正)

が成立する。

実際に市区町村を第一次単位として算出した第4表は(9), (10)の成立が各ブロック別にも明白である。又第9表の結果をみれば(11) (12)の各式も完全に明白である。

然し乍ら、市区町村を単位とした場合成立する関係が調査区を単位とした場合にも成立するという主張は可成り曖昧である。

実際に第1表によると、資本金額1千万円以上の企業の所在する調査区層のsizeと1千万円未満の企業のみを含む調査区層のsizeは略同数で約3万5千調査区である。然しこの場合に於ても第1図のように資本金額1億円以上を含む調査区層をA<sub>1</sub>層から独立させるならば(9)式の関係が成立する事は全体としては明かである一方、第18表の宮城、岡山両県についてM<sub>h</sub>欄をみれば、この関係の成立は殆ど認められないのである。従って、母集団概念は元来技術的相対的たりうるが、特に以上の(9) (10) (11) (12)の関係は第一に地域単位に依存し、第二に階級分点の設定に依存することが判明する。

然し他方調査区の設定基準及び資本金額についての階級分点は、経験的に或程度規準化されており、その規準に基づく限り以上の各式の成立は保障されうるのである。

此の事実を追求するならば、(一)と同様に分布法則の成立の条件及び過程に遡るのでここでは詳論を避けるが、以上の結果からみて次の諸点を差当っての結論とすることが出来る。

即 事業所調査区特性カードにおいて

(i) 資本金額1千万円以上の本社を含む調査区層

(ii) 資本金額1千万円未満の本社のみを含む調査区層

を一つの分類基準としているがこの分類は極めて不十分であり将来第(ii)層から少くとも

(iii) 資本金額1億円以上の本社を含む調査区層

第17表 ブロック別調査区内の全企業数と各産業別企業数との相関

	製 造 業	商 業
1 北 海 道	0.16932	0.77072
2 東 北	0.67056	0.66613
3 関 東	0.75310	0.55471
4 北 陸	0.75895	0.56448
5 中 部	0.82244	0.48446
6 近 畿	0.90162	0.41532
7 中 国	0.80541	0.41848
8 四 国	0.44601	0.22535
9 九 洲	0.76137	0.65389

第18表 宮城・岡山両県に関する調査区を単位とする母集団の構造 (σの単位1万円)

調査区	N <sub>hk</sub>		$V_{whk} = \sum_{j=1}^{Mh} \frac{N_{hkj}}{N_{hk}} \sigma^2_{hkj}$		M <sub>h</sub>	$V_{bh} = \frac{1}{M_h} \sum_{j=1}^{Mh} (x_{hj} - \bar{x}_h)^2$
	N <sub>11</sub>	N <sub>12</sub> N <sub>22</sub>	$\hat{V}_{w11}$	$\hat{V}_{w12}$ $\hat{V}_{w22}$	M <sub>1</sub> M <sub>2</sub>	$\hat{V}_{b1}$ $\hat{V}_{b2}$
宮 城	189	887 315	0.10482400 × 10 <sup>7</sup>	0.22061050 × 10 <sup>8</sup> 0.12511923 × 10 <sup>8</sup>	93 (38) 51	0.43268403 × 10 <sup>8</sup> 0.49675041 ~ 10 <sup>8</sup>
岡 山	161	948 314	0.92742335 × 10 <sup>6</sup>	0.20399932 × 10 <sup>8</sup> 0.19343884 × 10 <sup>8</sup>	91 (40) 57	0.21524405 × 10 <sup>8</sup> 0.44488895 × 10 <sup>8</sup>

註<sup>1</sup>) M<sub>1</sub>欄でカッコ内の数値は資本金額1億円以上の企業が所在する調査区数。

註<sup>2</sup>) wave は推定量であることを示す。

註<sup>3</sup>)  $\hat{V}_{bh}$  は資本金額1億円以上の企業を含んでいない。

第 19 表 宮城・岡山両県に関する調査区を単位とする標本抽出比の算定結果  
( $c_1=1000$  円とした場合)

	$\lambda$	$\hat{p}_1$	$\hat{p}_2$	$q_{11}$	$q_{12}$	$q_{22}$
宮 城	0.8492	0.0030	0.0293	8.8368	64.6927	6.5493
岡 山	0.6777	0.0034	0.0261	6.5131	44.2007	5.7968

第 20 表 標本精度と計算誤差

	宮	城	岡	山
$\sigma^2_T$	0.1339207253 $\times 10^{12}$		0.6713175968 $\times 10^{11}$	
$\left. \begin{matrix} T_1 \\ (1 \text{ 億円以上,} \\ \text{を含む}) \end{matrix} \right\} \left. \begin{matrix} T_2 \\ (1 \text{ 億円未満} \\ \text{のみ}) \end{matrix} \right\}$	4294667 3	662951. 5	3401484. 0	606064. 2
$\frac{\sigma_T}{T_1}, \frac{\sigma_T}{T_2}$ (変動係数)	0.085107449	0.5520038907	0.0761720384	0.4275091205
$e = \frac{ \hat{p}_1 \cdot q_{12} - \hat{p}_2 \cdot q_{22} }{r^2}$	0.0000		0.0000	

(σ の単位 1 万円)

註 第 20 表は  $p_1=100/3, p_2=100$  及び  $q_{11}=1, q_{12}=3, q_{22}=1$  として算出した結果である。然し実際の抽出に於ては産業特性を考慮した為宮城については  $p_1=74, p_2=260$ , つまり上記の抽出比の略 2 倍であり逆に岡山については  $p_1=18, p_2=56$  となり上記の抽出比の略 1/2 である。従って全国的にみた場合略表中の  $\sigma_x/T$  を期待することは不可能ではない。

を独立させるのが有利である。更に出来れば

(iv) 資本金額 10 億円以上の本社を含む調査区層

を分離させるのが望ましい。

この結論の根拠は、極めて限定された資料に関してではあるが、調査区を第一次抽出単位として (7)~(12) が成立することを第 18 表から読みとることが出来るからである。

## ㊦ 地域抽出単位の産業特性と問題点

§1 の標本設計に於ては資本金額による層化と共に産業特性による層化を行った。この種の層化は、産業別分布のアンバランスの是正と精度の向上 (例えば製造業の平均資産額はサービス・商業のそれ等より高額である) にあるが、その効果はどのように実現されているであろうか。

## 1. 調査区を単位とした産業の地理的聯関

調査区抽出の段階で各産業別の標本数を control する為には、各産業法人数の相関が一つの指標となる。第 12 表及び第 3 図は §1 ㊦の層化法についての一つの批判を与えるものである。

各ブロックの共通の傾向を辿るとまず

- (i) 製造業と商業は微弱な逆相関が認められる
- (ii) 製造業と建設業は逆に微弱な正相関が認められる
- (iii) 商業・サービス間には中程度の正相関が認められる
- (iv) 製造業とその産業 (製造, 商業, 建設, サービスを除く産業) 間には微弱な逆相関が存在する

(v) サービス業とその他産業間には略完全に独立な状態が認められる。

以上の結果から我々は製造業の比率のみによって設定した層化に対し次の層化法の妥当性を主張しうる。

つまり産業別にバランスのとれた標本を求めるには地域産業特性は製造業、商業、建設業、サービス業、その他の各産業につきそれ等の何れかが5割を越えるといふ条件により5階層に分割するのが望ましい。

或は各産業の分布からみれば製造業、商業、建設業を除くと他は少数であるから、5割を越えるという条件で4階層に分割する。又は製造、商業、その他の三階層に分割するのが抽出上有利であるといえる。

以上のように稍不十分な産業特性により層化を行ったが、その結果は第13表及び第14表のような標本構成をもつ結果が得られた。

これを母集団構成(昭和44年度事業所調査報告)と比較してみよう。(第15表参照)

これによって製造業・サービス業については母集団に対し標本の構成が高いことが分る。この産業特性による層別の効果は第16表によってより端的に示されることになった。即製造、サービスについての高い抽出率とその他の各産業の低い抽出率が対蹠的である。

問題は商業法人(略50%を占める)を抑制する目的で行った製造業を中心とした層化法が、製造・サービスを除く本来構成比率の低い非商業法人の抽出率をも低下させている点にある。

従って製造業標本を増加させる事を別として単に商業標本を抑制し、非商業標本一般を増大させ、標本の産業別バランスを得る為には、更に現行通りの産業特性カードを利用する場合には、製造ではなく商業を中心とした構成比によって層化の方が有効であるといえよう。

然し乍ら反面第17表によれば北海道を除き調査区内の企業総数は製造業企業数に対しより高い相関をもつものであり、従って調査区内及び調査区間分散を抑え全体としての精度を高め又単位調査区当りの標本数を把握して調査実施の条件を整える上では製造業を中心とした層別法が有効であるといえる。

## む す び

以上各節のデータを基にした解説を通して経済量を調査の対象とする時、そのサンプリング形態を決定するものは地域間では企業密度の分布であり、同一地域内では企業規模分布であることは明かであるが、同時に此等の分布が過大な格差を伴う時所謂 Optimality の観点は殆ど幻想に近くなる事も理解しうるのであり、更に遡るとランダム・サンプリング・システムを手放しで礼讃する訳にはいかなくなるのである。

つまり市町村を単位として設計時に算出した第9表における過大分散及びそれをもとに算出された第5表の最適抽出比の異常性は、調査区を単位として事後に検討した第18表、第19表にもそのまま反映されているからである。又上記の抽出比の決定による精度は差程満足すべきものではなく特に1億円未満層について不十分である。これは一重に標本調査区数の増量によってカバーする他はなく、例えば第一次標本数3000~4000、標本客体数30000~40000とするならば可成りの精度の向上が期待出来よう。更に、現在の事業所調査区を2~4程度合併した法人調査区を設定することにより、 $p_1$ 、 $p_2$ 及び $q_{ij}$ 間の比率を理論的計算値に接近させることが出来るから一層の効果が期待出来る。又この設定に伴う負担は産業特性による層化を廃止し代りに第二次抽出に於て商業、非商業によって異なる適当な抽出比を決定すれば、可成り軽減することが出来よう。

勿論こうした古典的サンプリング・システムと稍異質であるが幾分の改良案と思われるものも幾つか存在する。例えば比推定方式による場合、規模に比例した不等確率抽出法を採択する場合、或は補助変量の総額に比例した各層への割当法を適用する場合等である。

此等の方針は実際に国富調査、国有財産調査行政財産の抽出に適用したので次回は副題を改めて、上記の具体例に基づいて此等を検討することにしたい。この過程を経た上で改めてランダム・サンプリング・システムを検討するのが妥当であろう。

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 ——昭和45年国富調査のための法人及び個人企業資産調査のための標本設計の理論と実際 (I)—— 統計数理研究所彙報, 第18巻, 第2号, 1971
- [2] R. G. D. Allen : Sampling for current Economic Statistics J. R. S. S. series A. Vol. 127, 1964.  
T. V. Hanurav; Some aspects of unified sampling theory, Sankyā, Vol 28, 1966