多段抽出における標本平均の分散について

大　石　　潔
多　賀　保　志

(1959 年 9 月 受 付)

On Variances of Sample Mean with Multi-Stage Sampling

Kiyoshi ÔISHI and Yasushi TAGA

In planning the survey, it is necessary to grasp the variances between and within the units in the population.

Here, the estimated variances of sample mean with three-stage sampling are calculated by using the results of the survey on the national character of the Japanese which the Institute of Statistical Mathematics carried out in 1953.

Institute of Statistical Mathematics

§ 1. 前　書　き

統計数理研究所においては，昭和 28 年度に国民性に関する全国調査を実施し，その結果は彙報その他によって発表された．それより丁度5年を経過した今秋（昭和 33 年度）国民性第二次調査を実施するため計画を立てている．

かような広範囲にわたる調査においては，全数調査はいうに及ばず，単純無作為抽出によるサンプル調査も不可能であろう．従って，当然多段抽出によるサンプルが選ばれることになる．この際，合せて層別が考えられる．さて，多段抽出であるが，第一次抽出単位としては，資料及びサンプリングの作業の関係上，行政上の市町村が採用されるのが常であろう．

所が，近年，町村合併の進行が著しく，一つの都市に近郊農村が併合されて膨張した都市や，多少都市的性格をもつ一つの町にその周辺の幾つかの村が合併して市制を施行した市も多く，あるいは，片田舎の数町村が合併して一つの新しい町を形成したところも多い*.

今，これらをサンプリング調査の立場から考えてみると，第一次抽出単位が膨張し，これに伴って，第一次抽出単位内の heterogeneity が増加することを意味する．従って単位内の分散は大きくなり，またこれらの単位間の分散も変化するであろう．ゆえに，新しい行政上の自治体を抽出単位として用いるためには，これらの分散の値がどの程度の大きさであるかを知る必要がある．

この報告の直接の目的は，前回（昭和 28 年度）の国民性の調査の結果を利用して，多段抽出による比率推定を行う場合の分散を計算し，今回（昭和 33 年度）の国民性の調査の層別・サンプリングの方針を決定するための資料を提供しようとするものであり，間接には，一般に同様の調査を実施する場合の層別・サンプリングの方針決定の手がかりを提供しようとするものである．

---

* §7. 補遺を参照せよ．

§2. 取り上げた項目

比率推定の分散を計算するために，前回の調査結果から取り上げた項目は次の通りである.

i) Q 8-3*； 職業の価値………職業に上下なしと答えた人の比率
ii) Q 25-1； 日本人と西洋人………日本人は西洋人とくらべて優れていると答えた人の比率
iii) Q 28-1； 相続の問題………養子に継せた方がよいと答えた人の比率
iv) Q 52-1； 新 聞………よく読むと答えた人の比率
v) Q 58-1； 支持政党………自由党または改進党（保守政党）と答えた人の比率

上に挙げた五つの項目のほかに

vi) 工業率…全人口に対する製造業従事者の比率（昭和 25 年度の国勢調査による）
vii) 商業率…全人口に対する卸売業及び小売業従事者の比率（同上）

について分散を計算した.

§3. 計算のための準備

1) 母集団の設定

前にも述べた通り，地方自治体の廃置分合―町村合併―にはいくつかのタイプがあるが，これらのタイプについて heterogeneity の変化を考えてみると，第一のタイプ即ち一つの都市に数個の近郊村が併合して膨張した場合が，最も多く増加しているであろう．従って，この場合が内分散の変化が最も著しく，また外分散の変化については，いづれのタイプでも略同程度の大きさであるように思われる．ゆえにこの場合の分散値を計算しておけば，他の場合の分散値はその値を越えないであろう．かかる理由と計算の手数を省くために，母集団としては市部を選んだ．即ち，昭和 30 年国勢調査時における市全体（但し六大都市を除く）―市部―を母集団として設定した.

2) 層 別

1) で設定した母集団から若干の市を抽出する前に，母集団を次に述べる三つの層に別けた．先ず，昭和 25 年国勢調査時に既に市であった自治体―旧市部―と，昭和 25 年国勢調査時より昭和 30 年国勢調査時までの間に新しく市となった自治体―新市部―とに別け，更に，旧市部については，昭和 25 年国勢調査時以後において，廃置分合により地域的，人口的に膨張した程度により，増加率の大なる自治体と増加率の小なる自治体の二層に別けた．これらの層及び新市部の層をそれぞれ層 P_{I}，層 P_{II}，層 N と呼ぶことにする.

層 P_{I}，P_{II} の区別は

$$\frac{\text{昭和30年国勢調査時における地域の人口}}{\text{昭和25年国勢調査時における地域の人口}} \gtreqless 1.3$$

に従った.

旧市 242 市の中，昭和 25 年国勢調査時より昭和 30 年国勢調査時までの 5 年間に地域に変更のあった市，及び地域に変更のなかった市の数はそれぞれ 181, 61 であるが，地域に変更のなかった旧市について上の商の値を計算してみると，その値が 1.3 未満であるのは，61 市中 59 市であった．この事実から，5 年間における人口の社会増・自然増の限度は略 0.3 であると見做し，上述の商をもって，増加の大小の基準とした.

* Q 8 は前回の国民性の調査に於ける質問 8 に対応し，3 は採用した比率に対応する選択肢の番号である．尚各項目の questionnaire 及び結果については §7. 補遺を参照せよ.

3) 抽出単位の構成

多段抽出における各段階の抽出単位はそれぞれ次の通りである.

第一次抽出単位……昭和 30 年国勢調査時における市.

第二次抽出単位……昭和 25 年勢調査時国における市，町または村.

第三次抽出単位……個人.

4) 地点（第一次抽出単位）の抽出

2) において述べたようにして得られた各層から，systematic random—等しい確率—に六分の一の市を抽出した．今，その内訳を示すと，次の通りである.

層	旧市部		新市部	計
	P_{I}	P_{II}	N	
層の市数	112	130	242	484
抽出数	18	22	40	80

5) 比率 p_i, p_{ij} の想定

前回の国民性の調査は，勿論サンプリング調査であるから，4) において抽出された地点に関するデーターが必ずしも存在する訳ではない．従って，これらの地点に関して前述の各項目の比率を，前回の調査結果を利用してどのように想定するかが問題になる.

さて，4) で抽出された第一次抽出単位の中には，一般に昭和 25 年国勢調査時の市・町・村—第二次抽出単位—がいくつかずつ含まれている．今昭和 30 年国勢調査時の i 市—第一次抽出単位—の中に，M_i 個の第二次抽出単位が含まれているとする．その第 j 番目の単位—昭和 25 年国勢調査時における市・町・村—が昭和 25 年国勢調査時における郡部に属する場合，その単位は，前回の国民性の調査における層別基準に従って，いづれかの層に含まれていた訳であるが，その層に対応する比率*を，第 j 番目の単位に与え，これを p_{ij} とする．また，第 j 番目の単位が昭和 25 年国勢調査時における市部に属する場合は，前回におけるデーターの得られている市を，東日本と西日本に大別し，更にこれらの各市の特性—昭和 25 年国勢調査による産業構成—を考慮して六つの層に別け，各層をそれぞれ一括して § 2 に挙げた各項目の比率を算定し，一方第 j 番目の市の特性を考慮して，これらのいづれかの層に対応せしめ，その層の算定比率を対応する比率として与えた.

かくして，M_i 個の p_{ij} が決定すれば i 市に対応する比率として，

$$p_i=\sum_{j=1}^{M_i}\frac{N_{ij}}{N_i}p_{ij}$$

ただし　N_i : i 市の人口

N_{ij} : 第 j 番目の市（町，村）の人口

を与えることができる.

ただし工業率及び商業率に関しては，p_{ij} は最初から exact な値が与えられている.

§ 4. 比率推定における分散の式

母集団は L 個の第一次抽出単位より構成され，それらの単位の大きさをそれぞれ N_1, $N_2, \cdots, N_L$ $(N=\sum_{i=1}^{L}N_i)$ とし，大きさ N_i の第一次抽出単位には M_i 個の第二次抽出単位が含まれ，その第 j

* 前回の調査に於いて，この層から必ず一つの町又は村が抽出され，その町（村）に関するデーターは得られている．この結果をもって，層に対応する比率とする.

番目の単位の大きさを N_{ij} $(j=1, 2\cdots M_i)$ とする．また第三次抽出単位のもつ標識を X_{ijk} であらわすこととする．

1)　方式 A

母集団に含まれる L 個の第一次抽出単位の中から，各単位の大きさに比例した確率で，1個の単位を抽出し（比例確率抽出），その抽出された第一次抽出単位の中から，1個の第二次抽出単位を比例確率抽出により抽出する．さらにその中から n 個の第三次抽出単位を等確率で抽出すると仮定する．このとき，母集団平均の不偏推定式は

$$\bar{x}=\frac{1}{n}\sum_{n=1}^{k}x_{ijk}$$

となり，$\bar{x}$ の分散は次のようになる．

$$\sigma_{\bar{x}}^2=\sigma_{b1}^2+\sigma_{b2}^2+\frac{1}{n}\sigma_w^2$$

ただし　$$\sigma_{b1}^2=\sum_{i=1}^{L}\frac{N_i}{N}(\bar{X}_i-\bar{X})^2 \tag{4.1}$$

$$\sigma_{b2}^2=\sum_{i=1}^{L}\sum_{j=1}^{M_i}\frac{N_{ij}}{N}(\bar{X}_{ij}-\bar{X}_i)^2 \tag{4.2}$$

$$\sigma_w^2=\sum_{i=1}^{L}\sum_{j=1}^{M_i}\frac{N_{ij}}{N}\sigma_{ij}^2 \tag{4.3}$$

$$\sigma_{ij}=\frac{1}{N_{ij}}\sum_{k=1}^{N_{ij}}(X_{ijk}-\bar{X}_{ij})^2$$

$$\bar{X}=\sum_{i=1}^{L}\frac{N_i}{N}\bar{X}_i \qquad \bar{X}_i=\sum_{j=1}^{M_i}\frac{N_{ij}}{N_i}\bar{X}_{ij}$$

$$\bar{X}_{ij}=\frac{1}{N_{ij}}\sum_{k=1}^{N_{ij}}X_{ijk}$$

2)　方式 B

方式Aでは，第一段階においてただ1個の単位を抽出する場合を考えたが，ここでは，L 個の第一次抽出単位の中から，等確率で l 個を抽出し，これら l 個の第一次抽出単位の中から，それぞれ1個の第二次抽出単位を抽出，さらにその中から n_i 個 $(i=1, 2, \cdots, l)$ の第三次抽出単位を等確率で抽出する場合を考える．このとき母集団平均の不偏推定式は次のようになる．

$$\bar{x}=\frac{1}{N}\frac{L}{l}\sum_{i=1}^{l}\frac{N_i}{n_i}\sum_{k=1}^{n_i}x_{ijk}$$

今第三段階において抽出する単位の数 n_i を，それらが属する第一次抽出単位の大きさ N_i に比例するようにとる．すなわち，$n_i=\gamma N_i$ とすれば

$$\bar{x}=\frac{L}{Nl\gamma}\sum_{i=1}^{l}\sum_{k=1}^{n_i}x_{ijk}$$

となり，その分散は

$$\sigma_{\bar{x}}^2=\frac{L-l}{L-1}\frac{\sigma'^2_{b1}}{l}+\frac{\sigma'^2_{b2}}{l}+\frac{\sigma_w^2}{\bar{n}}$$

ただし　$$\sigma'^2_{b1}=\left(\frac{L}{N}\right)^2\cdot\frac{1}{L}\sum_{i=1}^{L}(X_i-\tilde{X})^2$$

$$\sigma'^2_{b2}=L\sum_{i=1}^{L}\left(\frac{N_i}{N}\right)^2\sum_{j=1}^{M_i}\frac{N_{ij}}{N_i}(\bar{X}_{ij}-\bar{X}_i)^2$$

$$\sigma_w^2=\sum_{i=1}^{L}\sum_{j=1}^{M_i}\frac{N_{ij}}{N}\sigma_{ij}^2$$

$$\sigma_{ij}{}^2=\frac{1}{N_{ij}}\sum_{k=1}^{N_{ij}}(X_{ijk}-\bar{X}_{ij})^2$$

$$\tilde{X}=\frac{1}{L}\sum_{i=1}^{L}X_i,\qquad X_i=\sum_{j=1}^{M_i}\sum_{k=1}^{N_{ij}}X_{ijk}$$

$$\bar{X}_i=\sum_{j=1}^{M_i}\frac{N_{ij}}{N_i}\bar{X}_{ij},\qquad \bar{X}_{ij}=\frac{1}{N_{ij}}\sum_{k=1}^{N_{ij}}X_{ijk}$$

$$\bar{n}=\frac{Nl\gamma}{L}.$$

方式 A, B を通して，われわれの考えているのは比率推定の場合であるから

$$\sigma_{ij}{}^2=P_{ij}(1-P_{ij})=P_{ij}Q_{ij}\quad(P_{ij}+Q_{ij}=1)$$

$$\bar{X}_{ij}=P_{ij}$$

と考えればよい．ただし，P_{ij} は第 i 番目の第一次抽出単位に含まれる第 j 番目の第二次抽出単位のもつ比率をあらわす．

§5. 計算結果

§4 の方式 A 及び B に従って，サンプリングの各段階の分散の値を計算するのであるが，サンプリング調査の結果を利用するのであるから，ここで計算される値は勿論推定値である．$\sigma_{b1}{}^2$,, $\sigma_{b2}{}^2$, $\sigma_w{}^2$ の推定値をそれぞれ $s_{b1}'^2$, $s_{b2}'^2$, $s_w'^2$ とすれば，これらの推定式は次のようにあたえられる．方式 B の場合も全く類似の式となるから，方式 A についての推定式を掲げる．

$$s_{b1}'^2=\frac{L}{l}\sum_{i=1}^{l}\frac{N_i}{N}(\bar{x}_i-\bar{x})^2$$

$$s_{b2}'^2=\frac{L}{l}\sum_{i=1}^{l}\sum_{j=1}^{Mi}\frac{N_{ij}}{N}(\bar{x}_{ij}-\bar{x}_i)^2$$

$$s_w'^2=\frac{L}{l}\sum_{i=1}^{l}\sum_{j=1}^{Mi}\frac{N_{ij}}{N}s_{ij}'^2$$

$$s_{ij}'^2=\frac{1}{n_{ij}}\sum_{k=1}^{n_{ij}}(x_{ijk}-\bar{x}_{ij})^2=p_{ij}(1-p_{ij})$$

$$\bar{x}=\frac{L}{l}\sum_{i=1}^{l}\frac{N_i}{N}\bar{x}_i,\quad \bar{x}_i=\sum_{j=1}^{M_i}\frac{N_{ij}}{N_i}\bar{x}_{ij}$$

$$\bar{x}_{ij}=\frac{1}{n_{ij}}\sum_{k=1}^{n_{ij}}x_{ijk}=p_{ij}$$

ただし，工業率及び商業率については，サンプルによる値 p_{ij} ではなく，（国勢調査による）母集団の値 P_{ij} を使用した．

［方式 A］

項　目	層	$\sigma_{b1}{}^2$	$\sigma_{b2}{}^2$	$\sigma_w{}^2$	σ^2
	P_{I}	0.00284	0.00469	0.199	0.207
	P_{II}	0.00437	0.00046	0.211	0.216
Q 8-3	N	0.00175	0.00329	0.174	0.179
	P	0.00377	0.00246	0.205	0.212
	U	0.00420	0.00269	0.196	0.203

Q 25–1	P_{I}	0.00327	0.00357	0.195	0.201
	P_{II}	0.00107	0.00067	0.195	0.197
	N	0.00540	0.00576	0.197	0.209
	P	0.00214	0.00204	0.195	0.199
	U	0.00321	0.00313	0.195	00202
Q 28–1	P_{I}	0.00458	0.00889	0.226	0.239
	P_{II}	0.00651	0.00114	0.228	0.236
	N	0.00661	0.01164	0.229	0.247
	P	0.00575	0.00481	0.227	0.238
	U	0.00663	0.00681	0.228	0.241
Q 52–1	P_{I}	0.00368	0.00886	0.230	0.213
	P_{II}	0.0082	0.00138	0.209	0.219
	N	0.00546	0.01131	0.232	0.249
	P	0.00821	0.00492	0.219	0.232
	U	0.00929	0.00679	0.223	0.239
Q 58–1	P_{I}	0.00149	0.00268	0.238	0.243
	P_{II}	0.00202	0.00005	0.222	0.224
	N	0.00471	0.00437	0.224	0.233
	P	0.00322	0.00152	0.229	0.234
	U	0.00367	0.00235	0.228	0.234
工 業 率	P_{I}	0.00159	0.00101	0.0746	0.0772
	P_{II}	0.00269	0.00005	0.0862	0.0889
	N	0.00153	0.00091	0.0593	0.0618
	P	0.00219	0.00050	0.0807	0.0834
	U	0.00213	0.00062	0.0744	0.0772
商 業 率	P_{I}	0.00034	0.00060	0.0441	0.0450
	P_{II}	0.00015	0.00005	0.0580	0.0582
	N	0.00010	0.00052	0.0383	0.0391
	P	0.00029	0.00031	0.0514	0.0520
	U	0.00028	0.00037	0.0476	0.0482

註 1) U は市部全体を表し，P は旧市部を表す．方式 B においても同様．

註 2) σ^2 は母集団から単純無作為抽出により，サンプルを抽出したときの分散に相当する．即ち母集団における比率を P とすれば

$$\sigma^2=P(1-P)$$

$$\text{但し}\quad P=\frac{1}{N}\sum_{i=1}^{L}N_iP_i$$

又簡単な計算により σ^2 は $\sigma_{b1}{}^2$, $\sigma_{b2}{}^2$ 及び $\sigma_w{}^2$ の和に等しいことが分る．

$$\sigma^2=\sigma_{b1}{}^2+\sigma_{b2}{}^2+\sigma_w{}^2$$

然し，吾々の推定値の場合には，必ずしも同様な関係が成立するとは限らない．

［方式 B］

項　目	層	σ'^2_{b1}	σ'^2_b	σ^2_w
Q 8-3	P_{I}	0.0369	0.00729	0.199
	P_{II}	0.0775	0.00004	0.211
	N	0.0509	0.00334	0.174
	P	0.0614	0.00356	0.205
	U	0.0838	0.00416	0.196
Q 25-1	P_{I}	0.0302	0.00410	0.195
	P_{II}	0.0470	0.00065	0.195
	N	0.0805	0.00589	0.197
	P	0.0338	0.00236	0.195
	U	0.0464	0.00337	0.195
Q 28-1	P_{I}	0.281	0.00987	0.226
	P_{II}	0.0610	0.00089	0.228
	N	0.1954	0.01173	0.229
	P	0.0486	0.00530	0.227
	U	0.0734	0.00732	0.228
Q 52-1	P_{I}	0.112	0.00132	0.230
	P_{II}	0.329	0.00115	0.209
	N	0.273	0.01113	0.232
	P	0.240	0.00128	0.219
	U	0.328	0.00317	0.223
Q 58-1	P_1	0.0421	0.00385	0.238
	P_2	0.0733	0.00047	0.222
	N	0.1217	0.00425	0.224
	P	0.0646	0.00214	0.229
	U	0.0909	0.00287	0.228
工業率	P_{I}	0.00222	0.00107	0.0746
	P_{II}	0.0129	0.00004	0.0862
	N	0.00198	0.00094	0.0593
	P	0.00829	0.00055	0.0807
	U	0.0107	0.00071	0.0744
商業率	P_{I}	0.00113	0.00079	0.0441
	P_{II}	0.00248	0.00005	0.0580
	N	0.00017	0.00052	0.0383
	P	0.00196	0.00041	0.0514
	U	0.00272	0.00050	0.0476

§6 あとがき

前§に掲げた分散の計算結果は，順を追って述べてきたような方法で求めたものである．すなわち，各段階の分散は母集団のもつ標識から直接計算したものではなく，一部分を抽出してそれより得られた推定値である．更にその推定値を計算するために用いた第二次抽出単位—昭和 25 年国勢調査時における市・町・村—のもつ比率 p_{ij} は，（工業率・商業率を除いて）実際にそれらの市町村で調査を実施して得られた結果ではなく，市町村を市部郡部に大別し，おのおのを地理的地方・人口・産業的特性等を考慮して層別をなし，同一の層に属する市乃至町村は，すべて同一の比率をもつと仮定して，対応する比率をもって代用したものが大半である．従って前回の調査における層別の効果を相当に認めない限り，分散の計算結果の信頼性は薄くなるように思われる．しかし，たとえ層別の効果が余りないとしても，市の膨張及び対応する比率の代用を人為的な操作と考え，計算結果を構成されたモデルにおけるものとすることにより，利用上十分意味をもつであろう．ここでは，方式Aに限定して論を進める．

われわれの計算では，市部のみをとりあげ，これを三つの層 P_{I}，P_{II}，N に別けただけであるから，各層の大きさは非常に大きいが，これらの層を更に小さい層に別けて，各層に方式Aのサンプリングを適用したとき，新しい層に関する各段階の分散値と §5の計算結果との関係について考えてみる．たとえば層 P_{I} を a ケの小さい層に別け，第 j 層 $(j=1, 2, \cdots, a)$ の大きさを $N^{(j)}$，その比率を $\bar{X}^{(j)}$ とし，その層に含まれる抽出単位の数を $L^{(j)}$ (但し $\sum_{j=1}^{a} L^{(j)}=L$)，おのおのの単位の大きさ及び比率をあらためて $N_i^{(j)}$，$\bar{X}_i^{(j)}$ $(i=1, 2, \cdots, L^{(j)})$ とすれば，(4.1) 式は次のように書き換えられる．

$$\sum_{i=1}^{L}\frac{N_i}{N}(\bar{X}_i-\bar{X})^2=\sum_{j=1}^{a}\frac{N^{(j)}}{N}\sum_{i=1}^{L}\frac{N_i^{(j)}}{N^{(j)}}(\bar{X}_i^{(j)}-\bar{X}^{(j)})^2+\frac{1}{2}\sum_{j\neq k}^{a}\frac{N^{(j)}}{N}\frac{N^{(k)}}{N}(\bar{X}^{(j)}-\bar{X}^{(k)})^2 \tag{6.1}$$

上の層別において，a ケの層の大きさが等しくなるように別けられたとすれば，上式の右辺は

$$\frac{1}{a}\{\sigma_{(1)}{}^2+\cdots+\sigma_{(a)}{}^2\}+\frac{1}{2a^2}\sum_{i\neq j}^{a}(\bar{X}^{(j)}-\bar{X}^{(k)})^2 \tag{6.2}$$

となる．ここで $\sigma_{(j)}{}^2=\sum_{i=1}^{L(j)}\frac{N_i^{(j)}}{N^{(j)}}(\bar{X}_i^{(j)}-\bar{X}^{(j)})^2$ は第 j 番目の層内の第一次抽出単位に関する分散を表わす．(6.2) 式によって新しい層の分散と計算結果との関係は明らかにされるが，もし層別の効果があまり期待できないならば，(6.2) の第二項の order は小さくなり，層 P_{I} の第一段階の分散は新しい各層の分散の算術平均になる．この場合，各層の分散の値が同程度の大きさであるならば，新しい層に関する第一段階の分散は，おおよそ $\sigma_{b1}{}^2$ に等しいと見做すことができる．また (4.1) 式，(4.2) 式についても同様な変形ができ，$\sigma_{b2}{}^2$，$\sigma_w{}^2$ はそれぞれ第二段階及び第三段階の分散に関して，各層の分散の算術平均になっていることが容易に分る．

今までわれわれの取扱ってきた推定式は偏りのあるものであるが，偏りのない推定式について考えてみよう．各段階の分散 $\sigma_{b1}{}^2$，$\sigma_{b2}{}^2$，$\sigma_w{}^2$ の偏りのない推定値をそれぞれ $s_{b1}{}^2$，$s_{b2}{}^2$，$s_w{}^2$ とすれば，分散の計算に実際使用した推定値 $s_{b1}'^2$，$s_{b2}'^2$，$s_w'^2$ との間には次の関係が成立する．

$$s_{b1}{}^2=s_{b1}'^2-\frac{L-l}{L(l-1)}\bar{x}^2+\frac{L}{l}\frac{L-l}{l-1}\sum_{i=1}^{l}\left(\frac{N_i}{N}\right)^2\bar{x}_i{}^2-\frac{L}{l}\sum_{i=1}^{l}\frac{N-N_i}{N}\sum_{j=1}^{M_i}\frac{N_{ij}}{N}\frac{N_{ij}-n_{ij}}{N_i}\frac{p_{ij}q_{ij}}{n_{ij}-1}$$

$$s_{b2}{}^2=s_{b2}'^2-\frac{L}{l}\sum_{i=1}^{l}\sum_{j=1}^{M_i}\frac{N_i-N_{ij}}{N}\frac{N_{ij}-n_{ij}}{N_i}\frac{p_{ij}q_{ij}}{n_{ij}-1}$$

$$s_w{}^2=s_w'^2+\frac{L}{l}\sum_{i=1}^{l}\sum_{j=1}^{M_i}\frac{N_{ij}-n_{ij}}{N}\frac{p_{ij}q_{ij}}{n_{ij}-1}$$

これらの関係から，$\sigma_{b1}{}^2$ に関しては直接には明らかでないが，二三の項目について偏りを近似計算することにより，$s_{b1}'^2$ は $\sigma_{b1}{}^2$ を過大に推定していることが推測される． また $s_{b2}'^2$ は明らかに

σ_{b2}^2 を過大に推定したものであり，$s_w'^2$ は σ_w^2 を過小に推定したものであって，かつその偏りは約 $\dfrac{1}{n_{ij}-1}\fallingdotseq\dfrac{1}{20}$ であることが分る．しかし工業率及び商業率に関しては，第二次抽出単位のもつ比率 p_{ij} に母集団の値 P_{ij} を用いているから，各段階の分散 $s_{b1}'^2$，$s_{b2}'^2$，$s_w'^2$ はそれぞれ σ_{b1}^2，σ_{b2}^2 及び σ_w^2 の偏りのない推定値になっている．

昭和 25 年及び昭和 30 年の国勢調査時における市町村の大きさは，廃置分合により著しく変化していることは前にも述べたが，それぞれの年時における市町村を第一次抽出単位とし，個人を第二次抽出単位とする二段抽出を考えると，昭和 25 年時の市町村を第一次抽出単位としたときの単位間の分散は $\sigma_{b1}^2+\sigma_{b2}^2$ に等しく，単位内の分散は σ_w^2 である．また昭和 30 年時の市町村を第一次抽出単位としたときの単位間の分散は σ_{b1}^2 であり，単位内の分散は $\sigma_{b2}^2+\sigma_w^2$ に等しい．計算結果をみると，σ_{b1}^2 と σ_{b2}^2 の大きさは略等しく，σ_w^2 の大きさは σ_{b1}^2, σ_{b2}^2 の大きさと比べて 1～2 桁高いから，両年時における内分散はほとんど同じであるが，外分散では昭和 30 年時の方が昭和 25 年時よりも小さく，ほぼ 1/2～1/3 になる．このことは，昭和 25 年時の場合には，昭和 30 年時には市部に含まれているが，昭和 25 年時には郡部に属していた町村がそのまま第一次抽出単位に含まれているから，単位間の heterogeneity が高く，寧ろ当然の結果であろう．もしそれらの町村を除外して，昭和 25 年時の市部のみを考えるならば，外分散，内分散は共に同じ位の値になることが予想される．

実際の調査に当ってその精度を高めるためには，調査費用を考慮して分散が小さくなるように各段階の抽出単位数を決定しなければならないが，精密な考察は今後に譲るとして，一応各段階の分散の大きさをそろえることを考える．層 P_{II} においては，直接分散の計算に関係した 22 市の中，昭和 25 年国勢調査時以後 5 年の間に廃置分合による影響がないものが過半数（14 市）を占めており，これらの市は第二段階の分散への貢献が全くない．従って σ_{b2}^2 の値は層 P_{I}，層 N のそれに比べて一般に小さくあらわれているから，各段階の分散値の大きさは層 P_{I} 及び層 N の二層について考察すればよいであろう．さて σ_{b1}^2 と σ_{b2}^2 とは同じ大きさであって，後者は前者のおおよそ 1～2 倍の値を示し，σ_w^2 は σ_b^2 のほぼ 20～60 倍の値であるから，各段階の分散の大きさをそろえるようにするには，各段階において抽出すべき単位数の比を

$$1:(1\sim2):(20\sim60)$$

とすればよいことが分る．今第二次抽出単位を投票区におきかえることを考えると，一般に市町村よりも投票区の方が局所性が強いと考えられるから，第二段階の分散は前の場合より大きくなり，第三段階の分散は小さくなるであろう．このために，第二段階の分散が k 倍 $(k\sigma_{b2}^2)$ になったとすれば，各段階の分散の和は一定 $(=\sigma^2)$ であるから

$$\sigma^2=\sigma_{b1}^2+\{\sigma_{b2}^2+(k-1)\sigma_{b2}^2\}+\{\sigma_w^2-(k-1)\sigma_{b2}^2\}$$

が成立して，第三段階の分散は $(k-1)\sigma_{b2}^2$ だけ減小する．よって

$$\sigma_{\bar{x}}^2=\sigma_{b1}^2+k\sigma_{b2}^2+\frac{1}{n}\{\sigma_w-(k-1)\sigma_{b2}\}$$

となり，このとき，第三段階の分散と第二段階の分散の比の値は

$$\frac{\sigma_w^2-(k-1)\sigma_{b2}^2}{k\sigma_{b2}^2}=\frac{\sigma_w^2}{k\sigma_{b2}^2}-\frac{(k-1)}{k}$$

となるから，各段階において抽出すべき単位数の比は

$$1:(k\sim2k):\left(\frac{21-k}{k}\sim\frac{61-k}{k}\right)$$

となる．もし第一次抽出単位を層別すれば，(6.1) 式から分るように，第一段階の分散を多少は減小させ得るかも知れないが，第二段階の分散を減少させる効果は期待できない．ゆえに第二段の抽出において十分考慮を払い，第二次抽出単位（投票区）を層別する方が効果的であろう．これらの

事柄は，分散の推定値の偏りが小さいときにいえることであって，もし偏りが大きい場合には，各段階の抽出単位数が異ってくる．工業率及び商業率の推定値は偏りのないものであるから，これらについて各段階の分散値の比をみると，第一段と第二段階との比は 1：(2/3～5) であり，第二段と第三段のそれはほぼ 1：70 である．Q の項目について偏りを補正したとき，各段階の分散の比が，この程度の大きさに近い値をもつならば，第二段の抽出については，前と同様なことがいえるであろうし，また第三段の抽出単位数は $\frac{70}{2k-1}$ 位にすればよいであろう．

ここでは市部について取扱ってきたが，郡部について考えてみると，第二段の抽出単位の間の heterogeneity は低くなるから，第二段の分散と第一段階の分散との比の値は小さくなるであろう．（実際郡部についての分散値を計算し，その比をみなければ，確かなことは分らないが）この比の値が1に近い値をもつことも考えられうるから，市部の場合ほど，第二次抽出単位の層別に考慮を払う必要がないかも知れない．

この報告には幾多の点に不足があるが，今後，資料の集積をまって，より信頼性の高い分散値が計算され，一層厳密な考察がなされることを期待したい．

統計数理研究所

§7. 補　遺

（1）　昭和25年及び昭和30年の国勢調査時における地方自治体数を表示すれば次の通りである．

	区部(区　数)	市部(市　数)	郡部(町村数)
昭和 25 年	84	242	10166
昭和 30 年	84	484	4323

（2）　§2 で取り上げた各項目について，その questionnaire 及び結果を参考までに列挙する．ただし，以下に表示する結果は全国平均であることを断っておく．

Q 8.　実際に必要な物を作ったり，売り買いする仕事をしている人と，学者や芸術家などのような人とは，どちらが社会的に見て価値が高いと思いますか？

1. 実際の仕事の方が高い	2. 学者や芸術家の方が高い	3. 職業に上下なし	4. 一がいに云えぬ	5. 無　答	6. その他	計
30.6	20.6	24.6	14.2	9.6	0.4	100.0%

Q25　日本人は西洋人とくらべて，一口でいえばすぐれていると思いますか，それとも劣っていると思いますか？

1. 優れている	2. 劣っている	3. 西洋人と同じ	4. 一口では云えぬ	5. 無　答	6. その他	7. 優劣両方ある	計
20.3	28.3	14.0	21.2	15.5	1.0	0.1	100.0%

Q28.　子供がないときは，たとえ血のつながりがない他人の子供でも，養子にもらって家をつがせた方がよいと思いますか，それともつがせる必要はないと思いますか？

1. つがせた方がよい	2. つがせる必要はない	4. 場合による	4. 無　答	5. その他	計
73.5	15.8	7.1	2.7	0.9	100.0%

Q52.　あなたは新聞をお読みになりますか？

1. よくよむ	2. 時々よむ	3, よくよまぬ	4. その他	計
55.9	24.8	18.9	0.4	100.0%

Q58. 最後にこの調査は選挙とは全く関係がありませんから，誰れに投票するかというようなことはうかがいませんが…，あなたは何党を支持していらっしゃいますか？

1. 自由，改進（保守党）	2. 右社，左社（革進党）	3. 支持政党なし	4. その他	計
40.7	23.1	20.3	15.9	100.0%

なお，昭和 25 年の国勢調査による工業率・商業率は次の通りである．

工業率	商業率
6.8 %	4.8 %

参考リポート

統計数理研究所；Stratify についての一資料

〃 ；Stratify についての資料（その 2）