

⑦ シヤウブ勸告案についての世 論調査のためのサムプリング

(1949年10月)

西 平 重 喜

シヤウブ勸告案に対して、国民がどんな意見を持つたかをしらべるため、世論調査が行われた。

この世論調査は、一般国民を対象とするものと、ほかにいくつかの特別な対象をしらべるものとを、同時に行った。

ここではそのうち、一般国民を対象とする世論調査のサムプリングについて書き、付録では農民を対象としたものについて書く。

われわれがサムプリングを引き受けるとき、すでに予算と実施面から、調査は関東地方の1都6縣にかざられて、また調査員の数などもきまつていた。

なお、この方法は統計数理研究輯報第1号に報告したものと平行的な考え方もあり、重なりあいもあるが、この第2号の目的から、これをいともせず書くことにした。

さて、関東地方の住民といつても、たとえば職業の違いにより、また同じ職業でも場所によつて、それらの人々の意見には、大分くいちがひが出てくるだろう。そこで同じような意見をもつと思われる人々毎に、いくつかのグループを作り、そのグループからサンプルを選んだ方が、

よい調査ができる。

すなわち、ここでもまた層別をして調査をしたい。層別をした方がよいことは、これだけでない。それは先きに進むに従つて述べよう。

層別をするために、学歴構造とか、生活程度、年令分布、今までのこの種の調査結果などについての——いけば直接的な資料があればつづろがよい。このような資料は都とか縣毎には出ているものもあるが、そんな広い地域では、その中でもまたづい分いろいろな住民がいるので、都や縣を層別の単位にしたのでは、層別の効果はうすいだろう。

層別はその単位が小さい程、その効果は大きくなるが、個人についてはもちろん、字(アザ)についての資料はまずないといつてもよい。

たとえあつたとしても、ほんざつすぎ、かまつて層別の途中で誤りを起すだろう。ちよつと扱いやすい単位としては、市、區、町、村、がよい。

市、區、町、村——以下、市区町村をたゞ地吳と呼ぶ——をとるとすれば、これを層別するために、昭和22年臨時国勢調査の市区町村別の産業人口をもちうることをできる。これは地吳の住民を性格化するためには、間接的なものであるが、充分客観的な意味をもつであろう。

すなわち、われわれはこの資料を使つて、住民をその環境——地吳の性格——によつて、層別をすることにした。

たとえば、工業の發達した地吳では、住民全体も工員の空気に支配されようし；漁村では、漁夫の意見に影響されるだろう。というのである。

層別をするときは、もう一つ考へなければならぬことは、調査員にかかる経費のことである。

附録でのべるような、農民を対象とする調査のこともあつて、調査員ひとりて12人のサムアルをしらべるのが適當であろうと考へられた。

また精度と経済的條件から、調査員は65人位、したがつてサムアルは全部で780人ときめた。ところが、関東地方の住民は16,573,800だから*)

*) 脚註次頁

これを65人の調査員に振り分けると、1人の調査員あたり250,000人位となる。すなわち各層の人口が250,000人(又はその整数倍)となるように、層別をしておけば、どの層の調査地員——サンプルをとる市区町村——でも、12人(又はその整数倍)のサンプルがわりあてられる。うまくゆけば、各調査地点に調査員を1人(又はその整数倍)おくと、各調査員は調査能力としてちようどよい12人づつを調べることになる。もしこれを考えておかないと、ある地員では1人で18人とか20人もサンプルをしらべなければならなくなった。また他の地員では、わずか数人のサンプルのためにも、1人の調査員をおかすければならないようなことになる。

そうすると調査員に無理がおこつたり、時には調査員をましてやらなければならぬようなことが起る。しかしこれは第二義的なことであつて、サンプルの数をそろえるためには、層内の同質化がやぶれては、層別の効果がへつてしまう。

上のようなことを考えて、いよいよ層別をはじめる。まず地員を區、市、町村の三つに分け、それぞれ區部、市部、郡部とよぶことにする。

この三つの部分については、当然住民の意見がちがひ、大きな層別としてみとめてよいだろう。

この三つの部分の中を産業型というものを考へ、これによつて層別をする。その結果は第1表のとおりである。

前頁の*) この小さな0は計算結果の0でなく、桁数をあらわすにすぎない。すなわちここでは百人単位で計算したことを示す。

第 1 表

| 層の番号 | 地点数 | 総人口 | 農業人口 |
|------|-----|---------|---------|
| 11 | | 213,800 | 106,000 |
| 12 | 40 | 370,700 | 34,300 |
| 13 | 84 | 493,900 | 132,400 |
| 14 | 37 | 228,400 | 47,500 |
| 15 | 31 | 242,300 | 49,800 |
| 16 | 69 | 426,200 | 102,700 |
| 17 | 184 | 793,400 | 27,900 |
| 18 | 30 | 490,100 | 34,100 |
| 19 | 18 | 237,700 | 20,000 |
| 20 | 26 | 235,000 | 38,100 |
| 21 | 82 | 346,400 | 148,100 |
| 22 | 56 | 232,100 | 96,900 |
| 23 | 47 | 252,900 | 101,900 |
| 24 | 46 | 204,200 | 84,100 |
| 25 | 65 | 287,400 | 115,300 |
| 26 | 88 | 465,900 | 149,900 |
| 27 | 43 | 266,300 | 68,200 |
| 28 | 81 | 515,000 | 67,600 |
| 29 | 26 | 296,700 | 62,700 |
| 30 | 49 | 336,900 | 28,900 |
| 31 | 28 | 422,000 | 44,600 |
| 32 | 66 | 320,200 | 124,400 |
| 33 | 56 | 256,900 | 110,600 |
| 34 | 176 | 757,300 | 365,400 |
| 35 | 67 | 263,800 | 119,000 |

| サンプル数 | 調査地点数 | 層をつくる組 |
|-------|-------|------------------------------|
| 10 | 1 | 1 |
| 17 | 1 | 2, 3, 5, 6, 7, 8 |
| 23 | 2 | 11, 19, 20, 21, 23, 28 |
| 11 | 1 | 16, 17, 18, 27, 32, 33, 34 |
| 11 | 1 | 36, 37, 38, 39, 40 |
| 20 | 2 | 65~73, 57, 58 |
| 37 | 3 | 59, 60, 41~48, 54, 50, 61~64 |
| 23 | 2 | 75, 74, 77 |
| 11 | 1 | 81 |
| 11 | 1 | 80, 82, 83 |
| 16 | 1 | 85 |
| 11 | 1 | 86 |
| 12 | 1 | 87 |
| 9 | 1 | 88 |
| 13 | 1 | 97, 98 |
| 22 | 2 | 105, 107, 103, 104, 106 |
| 12 | 1 | 4, 9, 100, 101, 102 |
| 24 | 2 | 108, 109, 114, 117, 119 |
| 14 | 1 | 115, 116, 118 |
| 16 | 1 | 120, 121, 123 |
| 20 | 2 | 124 ~ 128 |
| 15 | 1 | 138, 139, 147, 129, 130 |
| 12 | 1 | 134, 137 |
| 35 | 3 | 136, 143, 144, 150 |
| 12 | 2 | 133, 146, 148, 149 |

| 層の番号 | 地 区 数 | 総 人 口 | 農 業 人 口 |
|------|-------------|-------------|------------------|
| J26 | 3 | 198,200 | 12,400 |
| J27 | 4 | 236,600 | 26,700 |
| J28 | 3 | 256,600 | 13,400 |
| J29 | 4 | 245,400 | 8,300 |
| J30 | 3 | 235,400 | 9,800 |
| J31 | 3 | 253,500 | 15,200 |
| J32 | 3 | 205,900 | 5,700 |
| J33 | 3 | 244,100 | 19,200 |
| J34 | 5 | 244,500 | 14,800 |
| J35 | 1 | 252,900 | 13,200 |
| J36 | 1 | 261,800 | 6,000 |
| J37 | 4 | 329,900 | 35,200 |
| J38 | 2 | 262,800 | 5,600 |
| J39 | 3 | 581,200 | 7,900 |
| J40 | 2 | 239,300 | 1,900 |
| J41 | 3 | 537,900 | 500 |
| J42 | 2 | 269,000 | 1,000 |
| J43 | 4 | 501,800 | 1,100 |
| J44 | 1 | 313,700 | 2,200 |
| J45 | 2 | 335,000 | 1,600 |
| J46 | 2 | 294,500 | 600 |
| J47 | 2 | 285,600 | 800 |
| J48 | 2 | 452,700 | 4,100 |
| J49 | 3 | 588,600 | 25,300 |
| | | | |

| サンプル数 | 調査地点数 | 層をつくる組 |
|-------|-------|--------|
| 9 | 1 | 市 部 |
| 11 | 1 | " |
| 12 | 1 | " |
| 11 | 1 | " |
| 11 | 1 | " |
| 12 | 1 | " |
| 10 | 1 | " |
| 11 | 1 | " |
| 11 | 1 | " |
| 12 | 1 | " |
| 12 | 1 | " |
| 15 | 1 | 區 部 |
| 12 | 1 | " |
| 27 | 2 | " |
| 11 | 1 | " |
| 25 | 2 | " |
| 12 | 1 | " |
| 23 | 2 | " |
| 15 | 1 | " |
| 16 | 1 | " |
| 14 | 1 | " |
| 13 | 1 | " |
| 21 | 2 | " |
| 27 | 2 | " |
| | | |
| | | |

ここで、とくに郡部について出来上った層は25はすぎないが、われわれはすでに全国の町村を29郡に分け、さらに150組に分けてあるので、層別作業にはこれを利用した。

ここでも少しくわしく産業型とか、上の29郡、150組について説明しよう。各地員毎に、第2表のようなカードを作った。⁰これは昭和

第2表 産業別人口カード

| | 就業人口 | 産業率 |
|--------------------------------|------|------|
| 1. 農 業 | 1381 | 17.3 |
| 2. 林 業 | 1 | 0.0 |
| 3. 水産業 | 1 | 0.0 |
| 4. 鉱 業 | 0 | 0.0 |
| 5. 建設工業 | 32 | 0.4 |
| 6. 製造業 | 211 | 2.6 |
| 7. <small>ガス、電気</small> 水道業 | 1 | 0.0 |
| 8. 商 業 | 45 | 0.6 |
| 9. 金融業 | 2 | 0.0 |
| 10. 運輸通信業 | 32 | 0.4 |
| 11. サービス業 | 67 | 0.8 |
| 12. 自由業 | 3.3 | 0.4 |
| 13. 公務団体 | 40 | 0.5 |
| 14. 其 他 | 2 | 0.0 |
| 合 計 | 1848 | 23.0 |
| 完全就業計 | 1641 | |
| 部分就業計 | 207 | |
| 人口数 | 4011 | |

22年臨時国勢調査の産業(大分類)別人口と、それらの産業別人口が、その地員人口に対する%を出したものである。この%を産業率とよぶことにしておく。

結局、この%のもようが似ている地点を一つの層にするのだが、全国で10,500,関東地方でも2,000近くある町村を、1枚1枚くらべることは、できるものでない。それに反とえば第2表にあげた三箇村をみれば、農業率が17%で、たんぜん他の産業率をおさえている。

そこで三箇村は内部構造から見れば農村だが、全国の町村の中での相対的な地位はこれだけからは分らない。事実、全国の半分以上の町村が、36%以上の

農業率を示している。

つぎに高い製造工業率は2.6%である。しかし全国で1/4以上の町村が、6%より高い製造工業率をもっている。

すなわち、地元の産業率の大小だけから、その地元は農村だとか、工業的町といい切ることばできない。

そのためには、地元の産業形態を、全国の町村と相対的にきめなければならぬ。このため、われわれは全国の地元の各産業率について、度数分布をしらべることにした。

それも全地元では大変だから、全国の地元を一列にならべ、出発点をランダムにきめて、20番目毎の等間隔に527地元をえらび出した。

このようにして作ったのが第3表(次頁)である。

この表のAは農業、Bはガス水道電気業、Cは製造工業ほか金融、運輸通信、サービス、自由、公務団体業の分布はB型で；林業、鉱業水産業、建設工業、商業はC型である。

さて、A型では、分布が連続的に広い範囲にわたっているから、ある%で農業率を組分することは不自然である。

それに農業率は、他の産業と逆比例的な傾向が予想されるから、始めから農業率で層別するより、各段階で適当に考えた方がよい。

つぎに、B型の分布では、逆に範囲がせまく、どの地元でもほとんど一定の率(B表では0%)を示している。

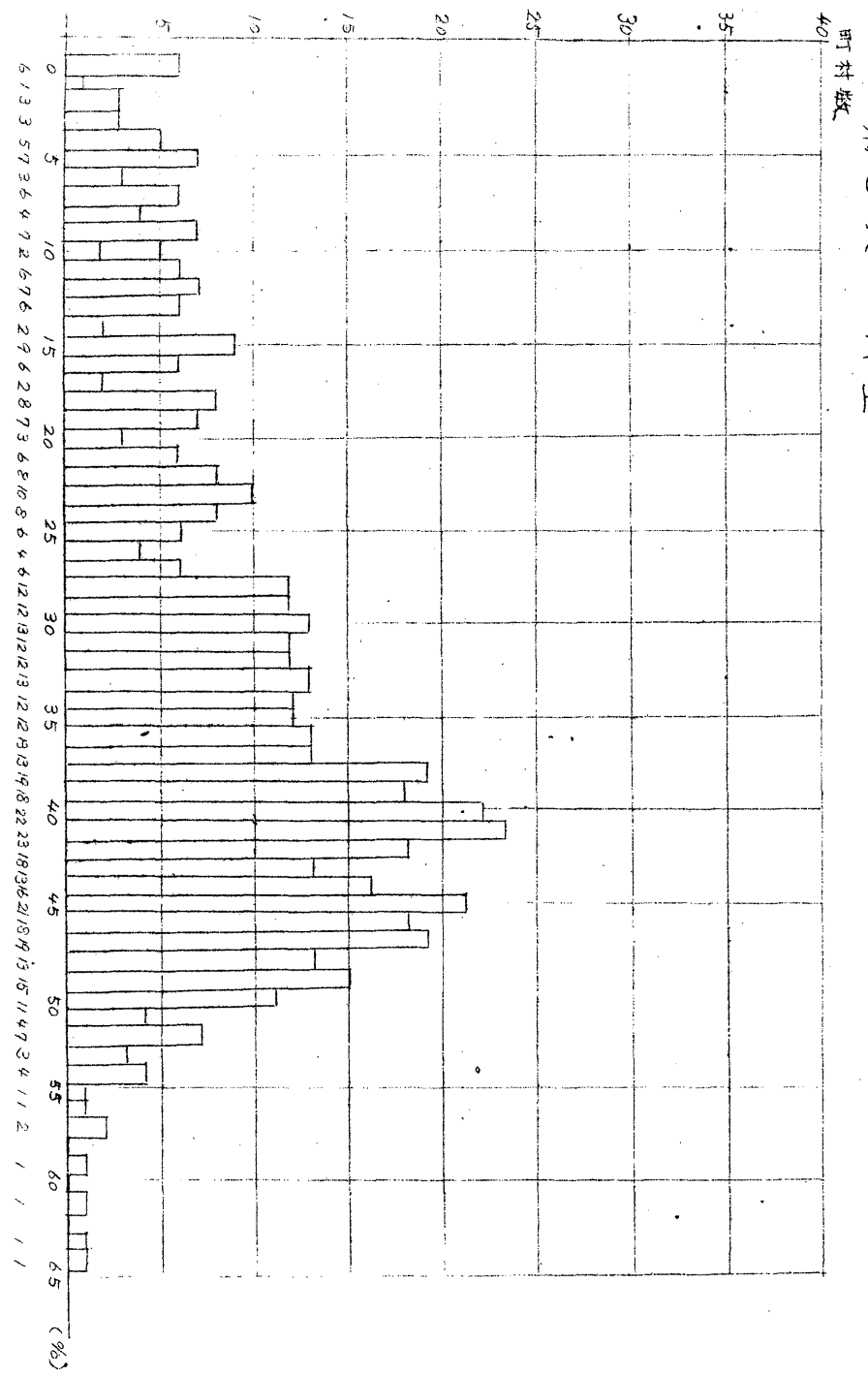
それに、これらの産業は他の産業に従属して存在するので、これらだけをとり出して、層別の基準とするより、他の産業に含めて考えた方がよい。

C型では、適当な広がりとし、ちょうどよい山をもった分布をしている。

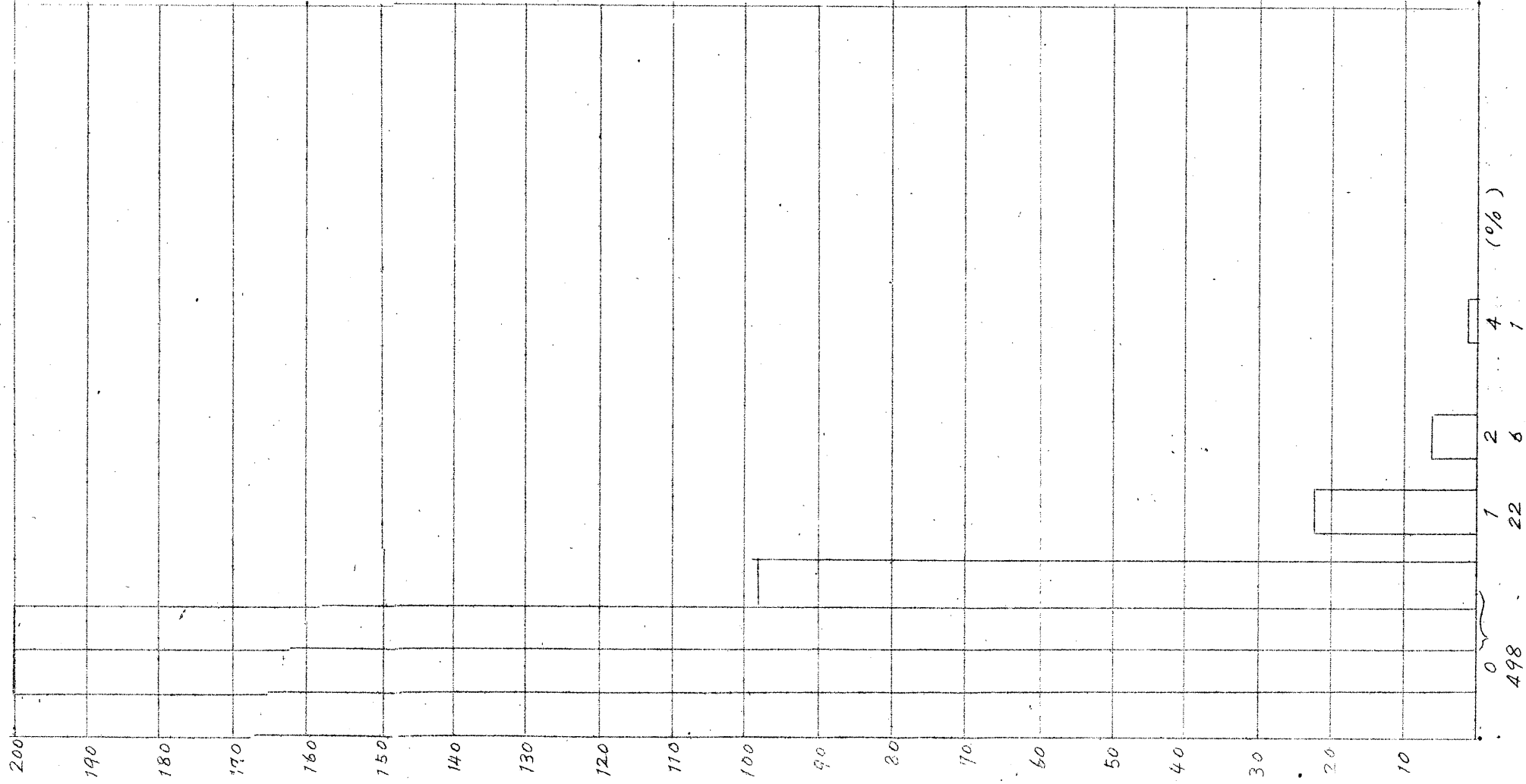
この山———度数が最大となる(%の)値をモードという———を目安にする。このモードを超えた%を示す地元は、全国でその産業が普通以上；これに達しないときは、普通以下と考えられる。

C型の六つの産業とそのモードは、(表 58頁)

第3表 A型



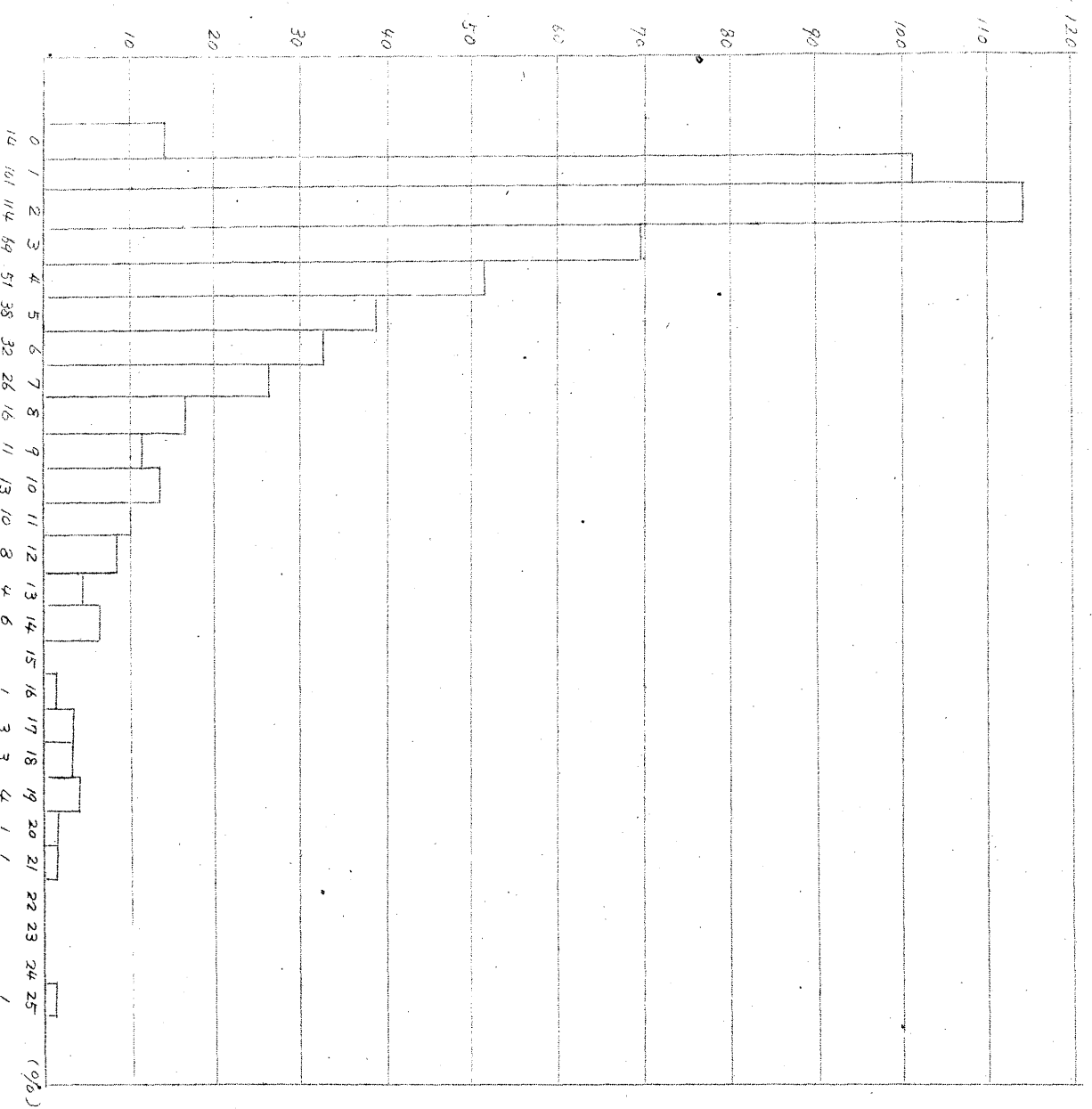
第三表 B 型



第三表 B 型、外
五七頁，次折込

第三表 C型

(町村数)



第三表C型、内七員、次打込

| 産業 | 林業 | 水産業 | 鉱業 | 建設工業 | 製造工業 | 商業 |
|--------|----|-----|----|------|------|----|
| モード(%) | 1 | 0 | 0 | 1 | 2 | 1 |

実は、林業のモードは0%だが、わが国の農業と林業はきりはなせないから、もう少し高く分類基準をもうけた方がよいと考え1%にした。
*)

これらのモードをたくみに組合せて、層別の基準を作った。この表は、輯報1号の第4表である。

この表をつかつて、29群に分けることが出来る。これでは充分でない点があるので、群を適当に組合せさらに150組に分けた。

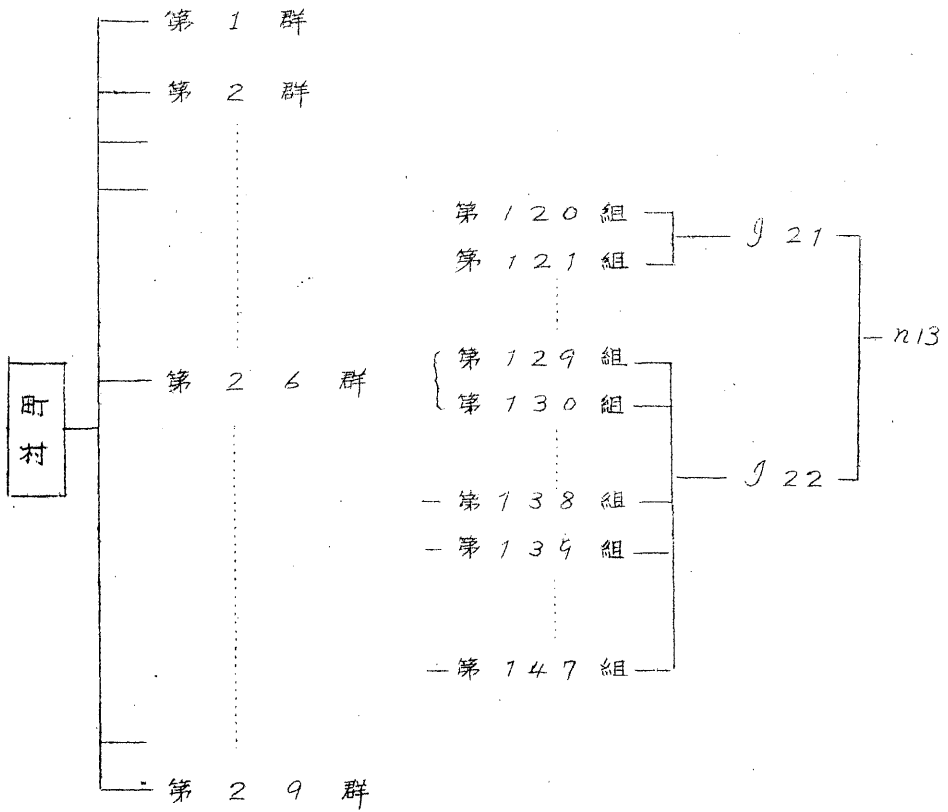
たとえば、第26群を二つに分け、なにか一つでも産業率がモードを越えていれば、その地点は第130組に入れ；そうでないものは第129組とした。そこで第2表の三箇村は第130組に入った。

こうして出来上った組を切りはりして、今度の調査のための層をつつた。三箇村は第130組だから、他の五つの組とあわせて出来ている922層に入っている。

この関係を画にかくと、第4表のようになる。(次頁)

註. *) このモードや分布は、まことに使ったことのある全国の地点によつたが、この調査は関東地方であるから、関東地方のモードをしらべた方がよかつた。しかし全国を渡つても意味はある。

第 4 表



(n 13 については付録を見よ)

この層番号のまゝの J は付録でのべる特別の層と區別するため、
一般 (J_{ppan}) をいみする。なお農業人口も付録で必要となる。

これで層別は完成した。そこで次に調査地点を各層からえらぶ。

これには地点の人口に比例してえらばれるようにした。このことについては、“パンに関する世論調査”にくわしく書いてある。

少しちがう点は、調査をする市區町村がきまつてから、その市區町村全体にわたって、サンプルをとるのでは、調査員の労力が大変であり、また、あまりなれていない調査員に、その地点の1,000から30,000もある名簿(このときは配給台帳から12人前後を決めることは、むずかしいと思われる。さらに、いいかげんに、調査しやすい処だけやってくる危険がある。そこで字(アザ)まで指定することにした。

このときも、もし字別人口が分つていて、それに比例するような確率で、調査する字が決められればよいが、それまで分つていない。

しかし、上のような危険にくらべれば、同じ割合で字をきめても、誤差はそれ程大きくなさそうである。

とくに、ここでは役場のある字にはウェイト2をつけてみた。すなわち役場のある字の人口は多いと思われるから、その字だけ2回数えることにした。

最後にサンプル600人を各調査地点に割り当てる。これにはいわゆる比例割当法によつた。

すなわち

$$\frac{\text{層の人口}(N_i)}{\text{関東地方の人口}(N)} \times \text{サンプル総数}(S) \\ = \text{調査地点のサンプル数}(S_i)$$

たとえば、第1層の調査地点のサンプル数は、

$$N = 16,573,800, \quad N_1 = 213,800, \quad S = 768 \text{ 名から}$$

$$S_1 = \frac{213,800}{16,573,800} \times 768 = 10$$

となる。

これで中央におけるサムプリングは終わったのであるが、調査員はそれ
をきめられた市區町村の役場などに行き、サンプル(個人)をきめら
れた数だけとらなければならない。

このときは等間隔法によつた。

すなわち、配給台帳のうち、指定された字の総人口(A人)をしらべ
る。きめられたこの地点のサンプル数をa人とすれば

$$A \div a = p \text{ 余り } q$$

を出し、pより小さい数(pを含む)を乱数表できめる。

今それをαとする。このαを出発番号という。

台帳の始めから数えて、α番目の人をまずサンプルとしてとり、次に
p人目毎に(p-1人目おきに)サンプルをとつてゆけばよい。

但し、こうすると、αとqの関係から、ちょうどa人とれるときと、
a+1人とれてしまうことがある。

もし、a+1人とれてしまったら、このa+1より小さい数β(a+1
を含む)を乱数表できめ、その人をのこせばよい。

たとえば、画にかけるように、A, aを小さくして A = 28人、
a = 5人の地点を考えてみよう。

$$28 \div 5 = 5 \text{ 余り } 3 \quad (\therefore p=5, q=3)$$

今、乱数表をみて、出発数(α)をきめる(もちろん α ≤ 5)。

今、α = 4 だったとすれば、

○○○◎○○ ○○○◎○○ ○○○◎○○ ○○○◎○○ ○○○◎○○ ○○○

この◎の5人をサンプルにすればよい。

もし、出発数αが2だったら、

○◎○○○○ ○◎○○○○ ○◎○○○○ ○◎○○○○ ○◎○○○○ ○◎○○

この◎全部をサンプルにすれば、6人になってしまう。

6より小さい数 β を乱数表からとる、($\beta \leq 6$)
 $\beta = 3$ だったとする。

このとき、5人のサンプルとしては3番目の◎をのぞけばよい。

付 録

農民だけを対象とする世論調査では、地点の農民や農産物などが分つていれば、層別に利用したいが、それがない。

また、第1表で分るように、層の間では全人口と農民数との間には相関がないから、一般国民のための層をそのまま使うわけにはゆかない。

そこで、一般用の層で、その産業型の似ているものを新らしく一つの層としてまとめるという方法をとった。*)

こうしておけば、層の中では、農業率は大体一定の値を示すだろう。

この農民の調査のためには30地点以内、600人のサンプルというワケがある。

関東地方の農民は2,968,700人であるから、一つの層の農民人口は1,000,000人位になる。

すなわち、一般層の農民人口が1,000,000人位になるように、しかも産業型が似ているようにして、第5表(次頁)のような農民用の層を作った。

このようにして、2種類の層別をするとき、片方の層をそのまま——時には二つか三つに分けた部分層を——他の層を作るときの単位にすれば、つぎのように調査地点を両方の調査で共通にすることもできる。

註. *) 必要なときは、一つの層を細く分けてもよい。

第 5 表

| 層 番 号 | 特 性 | 性 | 農 業 人 口 | サ ム プ ル | 調 査 地 点 数 | 含 ま れ る 層 |
|-------|-------------------------------|---|-------------------|---------|-----------|-------------|
| 22 1 | 水 産 業 | | 1799 ^人 | 36 | 1 | 9 3, 4 |
| 22 2 | 鉱 業 | | 498 | 6 | 1 | 9 5 |
| 22 3 | 林 業 (近 代 化) | | 1027 | 21 | 1 | 9 6 |
| 22 4 | ” (原 始 的) | | 3131 | 63 | 3 | 9 7, 8 |
| 22 5 | 農 業 率 高 く, 高. 工 業 普 通 (茨 木) | | 2062 | 42 | 2 | 9 9, 10, 11 |
| 22 6 | ” (東 京, 千 葉, 神 奈 川) | | 469 | 20 | 1 | 9 12 |
| 22 7 | ” (群 馬, 栃 木) | | 1019 | 20 | 1 | 9 13 |
| 22 8 | ” (埼 玉) | | 841 | 17 | 1 | 9 14 |
| 22 9 | 農 業 率 高 く, 工 業 率 も 高 い | | 1153 | 24 | 1 | 9 15 |
| 22 10 | 農 業 率 は 普 通, 工 業 率 が や や 高 い. | | 2195 | 44 | 2 | 9 15, 13 |
| 22 11 | い ち ん な 農 業 率 が 平 均 し て い る | | 1025 | 21 | 1 | 9 2, 17 |
| 22 12 | 工 業 率 が 高 く, 農 業 率 低 い | | 1073 | 22 | 1 | 9 19, 21 |
| 22 13 | 平 凡 | | 2133 | 43 | 2 | 9 20, 22 |
| 22 14 | 農 業 率 高 く, 工 業 率 高 し | | 1106 | 23 | 1 | 9 23 |
| 22 15 | ” 商 業 率 高 し | | 3654 | 74 | 3 | 9 24 |
| 22 16 | ” 工, 商 業 率 と 同 じ | | 1190 | 24 | 1 | 9 25 |
| 22 17 | 水 産 地 (昭 和 22 年 10 月) | | 1060 [*] | 21 | 1 | 9 1 |
| 22 18 | 市 部 | | 3772 | 76 | 3 | 市 と 區 |
| | | | 29687 | 598 | 27 | |

* 推 定

こうすれば、経済的に得をするばかりか、すぐれた種々の調査員ですませることもできて、調査の精度を高めることもできる。

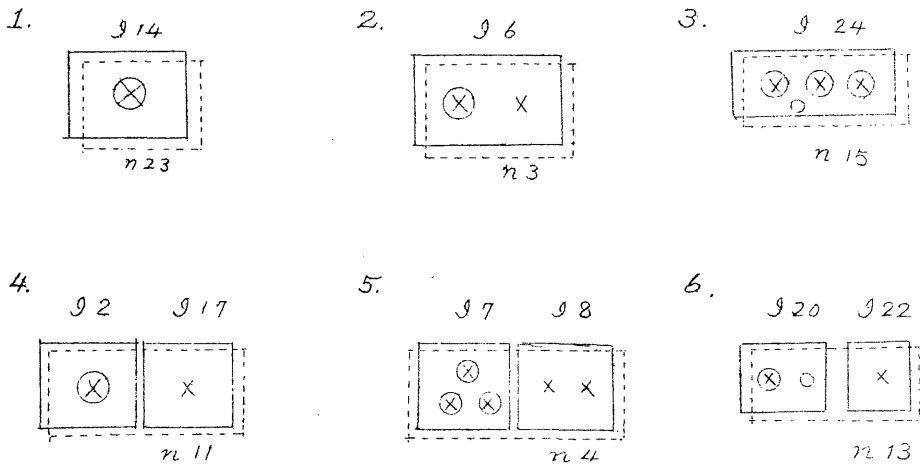
この地点を共通にする方法を、例をあげてのべよう。


1) $n 14 \equiv J 23$ 両方とも1地点。

このようなとき、全く同じ層と考へたゞ抽出の確率を総人口でなくて、農民数によらなければならない。

ところが、前にことわつたように、農業用の層の中では産業型が仮定しているはずであり、農業率は一定と考へているのだから、同じ地点をとることができる。

第 6 表



 は J の層 X は J の調査地点

 は n の層 O は n の調査地点

2) $n_3 = J_6$, 但し J では2地点, n では1地点。
 J で2地点をきめるときのことを考へれば, *) 始めにえらん
だ方の地点だけで農民の調査をすればよい。

3) $n_{15} = J_{24}$; J で3地点, n で4地点。
2) のときと同じ理由から, J の3地点をのこした残りから,
同じ確率 (*equally likely*) にきめればよい。

4) $n_{11} = J_2 + J_{17}$ J_2, J_{17} 各1地点, n で1
地点。

n_{11} が二つの部分層からなると考へ, J_2, J_{17} の農業人口
に比例する確率で, n_{11} の調査地点はどちらの部分層からと
るかをきめる。たとへば J_2 であつたとする。この J_2 の中
では, 既に確率付で抽出した地塊があるので, この地点を n_{11}
の調査地点にする。

5) $n_4 = J_7 + J_8$; J_7 は3地点, J_8 は2地点で,
 n_4 は3地点のとき, ここでも J_7, J_8 を n_4 の部分層と考
へ, まず3地点の組が J_7, J_8 からどのようにとられるか,
—— J_8 からばかりとられるか; それとも J_7 から一つ,
 J_8 から二つとられるか, など —— をしらべる。**))

このときは3地点とも J_7 からとられることがわかつた。

そこで n_4 の3地点として, J_7 の3地点をとる。

6) $n_{13} = J_{20} + J_{22}$; J_{20} は1地点, J_{22} も1地

**) 統計数理研究輯報第1号48頁をみよ。

註 *) そのまえに, 一つの層から二つ以上の地点の組をその大きさに
比例した確率できめるときは, まず一つの地塊を大きさに比例した
確率でとり, つぎに残りの中から, どの地点も同じ確率で抽いてゆ
けばよいことを注意しておく。こうして, 順序をつけてきめてお
けば, 調査地点の数をへらすときには, あとから抜いた方から捨て
ゆけばよい。(講究録 5巻 12号 (42))

点； n_{13} から 2 地点。

5) と同じようにして， n_{13} の 2 地点が J_{20}, J_{22} にどのように含まれるかをみる。 このときは J_{20} に 2 地点ともとることになった。ところが J_{20} には一つの調査地点がとれていない。そこで第二の地点は， J_{20} から調査地点をのぞいた地点の中から，同じ確率で一地点を改めてきめればよい。

5), 6) についてはもつといろいろな場合が起るだろうが，方法は全く同じである。

このようにすれば， 3), 6) のようにときにだけ新しい地点をとらなければならない。

この調査では新しくとつたのはたゞ 2 地点にすぎなかった。

なお，この計画は，水野 坦所員におうものであり，訂正をこまめに細いことは，小島嘉江，鈴木三千代さんにしてもらった。

(1950¹¹ 9²⁵ 20¹¹)