① 港區区長選挙に關する世論調査のサムプリング計画

水　野　坦

選挙の世論調査というものは，選挙前に選挙人の中の，どの位がどの候補者に投票しようと思っているかを予想しようとするものであつて，多くの関心をもたれている。

われわれが行つた選挙の世論調査は，たゞ單に區長に誰が当選しそうかという予測をするだけではなくて，浮動票の追求，言廻し（wording），性別，年令別，職業別，学歴別などによる投票の相違，無回答者（Non-response）や到達不能者（unarraival）の分析，被調査者のウソの分析などによつて，この種の調査の精度をあげ，また一般の予測方式の研究にも役立たせようとするものであつた。

調査全体については，いづれこの彙報の続刊で報告されるであらうし，一部はすでに一・二の学会で報告ずみである。

ここではたゞ等間隔抽出（Systematic sampling）の例として，とりあげる。

上にのべたようなくわしい分析をしようとするのであるから，相当多くのサンプルをとらなければならない。

はじめには，１７００人位のサンプルをとりたいと思つたが（文末の註をみよ），予算の都合で，調査員の数が制限をされ，４７人しか得られない。

また調査のできる期間もわずか２，３日しかなかつたので，調査員に

徹底的な追求（call back）を要求することからも，24人以上は無理と考へ，1200人のサンプルをとることにした。

しかし調査員を増す——— 従つてサンプル数をふやす——— 可能性も予想されたので，1200人以上にふやす場合にも，かんたんにしかもひずみのないようなサンプルがとれるやうな方法をとつた。

この調査の対象は港区の選挙権をもつ者全体である。

ところが，有権者については，選挙人名簿があつて，われわれの調査対象全部の名前がのつている。

その総数は約120,000人であつた。この選挙人名簿を見ると，大体町毎に一冊にまとまつていて，20行の罫紙を使つていた。

すなわち，各頁に2　　　の選挙人の氏名，住所，生年月日等が書いてある。　われわれはかんたんにしかもひずみのないサンプルをとるために，このどのページも20行という点に注目した。

すなわち，抽出間隔を20 —— 20人目毎にサンプルをとる —— ことにすれば，いちいち間隔を数えなくても，各ページの同じ位置の人をサンプルとしてとればよい。———

すなわち，始めに20より小さな数を乱数表からきめる。この数がたとへば3だつたとすると，どのページでも，上から（始めから）3番目の人をとればよい。

もし，よごしてよい名簿だつたら，つみ重ねておいて，3番目のところを千枚通しで穴をあけることもできるだらう。

このように抽出間間を名簿の週期と一致させておけば，あやまりなく簡単に早くサンプルを抽出できる。

ところがこのまゝでは約120,000人の選挙人の$1/20$をとることになるので，約6000人ものサンプルを得てしまう。

そこで3ページおきにして，まず$1/60$の約2000人をとつた。——

すなわち，実際には抽出間隔を20でなくて60にした。

その方法は，まず60より小さな数を乱数表からきめる。

たとへばそれが２４だつたとすると，第１にえらばれるのは第２頁の上から４番目の人である。つぎには，　３頁目の第５頁の上から４番目の人をとる。といつた風に３頁目の４番目をとつてゆけばよい。

しかも，このようなとり方をすれば，人手があれば，何人でも一せいに別々の名簿を分擔することができる。

すなわち，始めの出発番号の頁を，間隔の頁数で割つたときの剰余と同じ剰余の頁の上から一定の位置をとればよい。

上の例では，第２頁で間隔は３頁で，上から４番目だから，各人は受取つた名簿の各頁を３で割つて２が剰る頁（２，５，７，----26,29----110，113，------- など）の第４番目を書きぬけばよい。

このようにして書き抜いてみると，１６８１人の選挙人をとることができた。

ところがわれわれはせいぜい１２００人しかサンプルをとることができないから

$$1200 \fallingdotseq 1681 \times 5/7$$

から，5/7 にへらさなければならない。　このへらし方もつぎのようにすれば，間違いなく簡単に実行出來る。

$$5/7 = 2/3 + 1/3 \times 1/7$$

そこでまず１６８１人の内 1/3 をすててしまい，そのうち 1/7 を生かせばよい。

まず３より小さな数を乱数表できめる。　たとえば２だつたとすると，１６８１人の名簿のうち２番目，５番目，８番目，------と青線で消してゆく。

つぎは７より小さな数を乱数表からきめる。　たとえばこれが４だつたとすると，今青線を引いて消したもののうち，４番目，１１番目，１８番目，--------に赤丸をつけて，これを生かし，これと青線で消されなかつたもの（すなわち青線だけを引かれたものをすてた残り全部）をサンプルとすれば，われわれの要求を充してくれる。

これを画いてみると，下図のようになる。

1 ~~2~~ 3 4 ~~5~~ 6 7 ~~8~~ 9 10 ~~11~~ 12 13 ~~14~~ 15

16 ~~17~~ 18 19 ~~20~~ 21 22 ~~23~~ 24 25 ~~26~~ 27 28 ~~29~~ 30

31 ~~32~~ 33 34 ~~35~~ 36 37 ~~38~~ 39 40 ~~41~~ 42 43 ~~44~~ 45

46 ~~47~~ 48 49 ~~50~~ 51 52 ~~53~~ 54 55 ~~56~~ 57 58 ~~59~~ 60

61 ~~62~~ 63 64 ~~65~~ 66 ………………………………… ~~74~~ 75

76 78 ……………………………………………………

……………………………………………………

……………………………………………………

……………………………………………………

…………………………………… 1645 ~~1646~~ 1647 1648 ~~1649~~ 1650

1651 ~~1652~~ 1653 1654 ~~1655~~ 1656 1657 ~~1658~~ 1659 1660 ~~1661~~ 1662 1663 ~~1664~~ 1665

1666 ~~1667~~ 1668 1669 ~~1670~~ 1671 1672 ~~1673~~ 1674 1675 ~~1676~~ 1677 1678 ~~1679~~ 1680

1681

このようにした結果1182人のサンプルを得た。

もし，調査員をますことができ，サンプル数をふやせるときは，青線で消したものから，てきとうな間隔に生かせばよいのである。

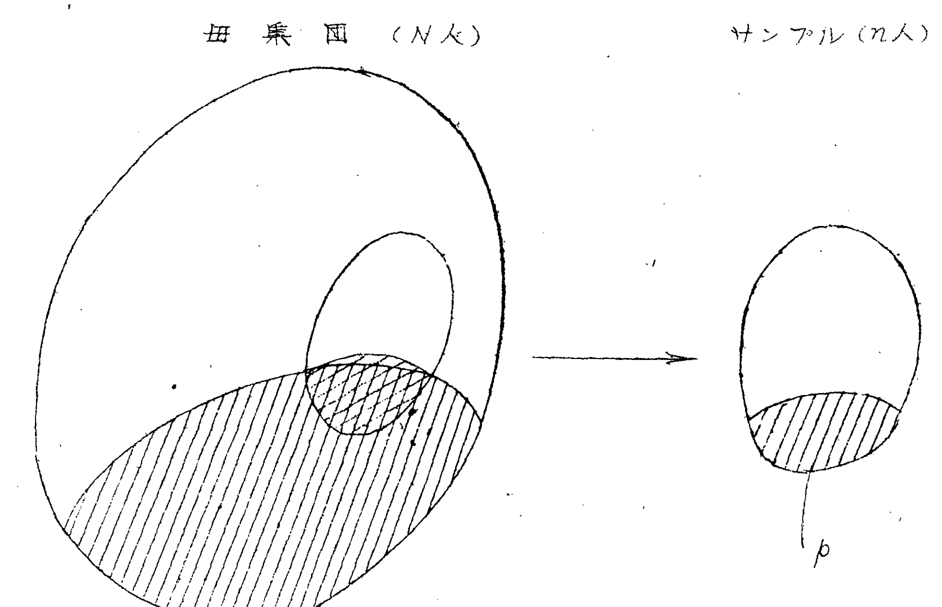
なおこの調査は，林知己夫所員，橋爪淺治，石田正次，木村等の諸君及び小島嘉江，村岡充子，高倉節子，田熊雅子，鈴木三千代の諸嬢におうものである。

またこれは，文部省科学試験研究費によるものである。

なお，この報告は，水野所員出張中のため西平重喜がすべての責任をもつ。

［注］　サンプルの数は，調査の精度と共に考えなければならない。その関係はつぎのような式であらわせる。

$$(1)\qquad a = k\,\frac{N-n}{N-1}\sqrt{\frac{p(1-p)}{n}}$$



全斜線の割合P

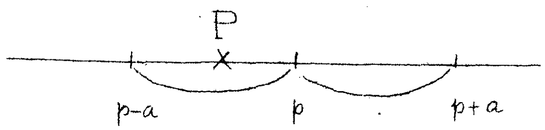
この式のいみは，N 人の母集団の中で，ある標識（例えば自由党支持率，自由業に従事している人の率，投票率などなんでもよい。）に属するものの割合を P とする。この母集団から n 人のサンプルをとつて来たとき —— このサンプル n 人の組合は $\binom{N}{n}$ 通りあつて，そのひと組をとつて来たとき —— 母集団で考えた標識に属するものの割合が p であつたとする。この p から P を推定するとき，

$$(2) \quad P_r\{p-a \leqq P \leqq p+a\} = \alpha$$

すなわち，サンプルの割合 p に a という幅をつけたとき，母集団の P は $p-a$ と $p+a$ の間に入っている確率が α である。

但し，この α によつて（1）式の k がきまる。

また（1）から分かるように，この方法に



よれば，相対精度 $\frac{a}{p}$ は p が増す程小となるから，一番少ない割合の標識について，n を決めておけば，それより多い割合の標識についての相対精度はよくなる。

このことをこの調査について書いてみよう。

区長当選者の支持や投票，棄権，各産業，学厂，年令別などの各カテゴリーが港区民の中で占める割合はどれも5%より大きいであろう。

一方，今迄の調査の経験から有権名簿のあやまりは5%ぐらいである。

また，5%位のところを問題にすれば，大体この調査の目的は達せられると思われる。

そこで $p = 0.05$ をとることにする。a をきめるためには，相対精度 $\frac{a}{p}$ を 0.1 で満足することにして，$a = 0.1 \times 0.05 = 0.005$

ときめる。

また（2）式の α すなわち信頼度として 95％をとることにして，$k = 3$ をとる。 Nは有権者名簿から約120,000である。

これらを（1）式に代入すれば

$$0.005 = 3 \times \frac{120{,}000 - n}{120{,}000 - 1}\sqrt{\frac{0.05 \times (1 - 0.05)}{n}}$$

これから n をきめればよいのであるが，

$$\frac{N - n}{N - 1} = \frac{120{,}000 - n}{120{,}000 - 1} \fallingdotseq 1$$

と考えれば，

$$n = 3^2 \times \frac{0.05 \times (1 - 0.05)}{0.005^2} = 1710$$

こうして，1700のサンプルがほしいこときめた。

終