

ある“地区ぬき“抽出法による偏りに就て

林 知 己 夫

(1958 年 7 月 受付)

On the Bias by Area Sampling

Chikio HAYASHI

The survey on living patterns by area sampling method has been taken at the City of Iga-Ueno. These examples of the bias by area sampling are shown.

Institute of Statistical Mathematics

ここではある地区ぬき抽出法 (area sampling) によつて生ずる偏りの例について述べる。社会現象特に人を対象にして行う調査においては通常名簿 (台帳) によつて対象となる人々を抽出し調査を行うか、又はある地区を抽出しその中に常住している人々を対象に行うかの方法がとられている。然し此等の調査に当つては対象たるべき人の所謂追及調査 (call back) を必要とする。調査期日が至つて短いような場合、追及の時がないために偏りのある結果を出してしまうこともある。

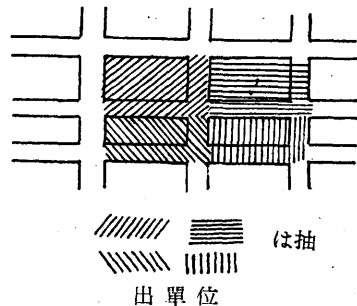
このような場合には、ある狭い地区を指定し、ある時点にそこに存在する人を対象にして調査を行うこともある。粗く言えば、多くの場合、前者は夜間人口を対象としての調査であり、後者は晝間人口を対象としての調査ともなる (この意味からは、特定の地点を限つて調査地点とした時は始めにのべたものと、今ここにのべたものとは調査対象が異つているのである)。こうして調査を行う場合には一応 call back の必要もなく、望ましいものとなるのであるが、調査時点の問題、調査対象の実際の把握等の問題等に関する困難が生じてくる。これが机上の通りに行われぬ所に調査の偏りが入りこんでくるのである。

さきに示した場合 (調査時間の至短の場合) の他に、如何にしても、現在する人々をもとにして調査を行わねばならぬこともある。例えばある時期、ある地区で行われている人間の行動模様の調査等はこれである。

ここでは伊賀上野市 (三重縣) で行つた生活場面に関する調査*結果について上の点を検討して、この種調査において生ずる偏りの一例について考えてみることにしよう。これはある一日、市民が如何なる生活模様を描いているかをしらべるものであり、言語調査のための基礎調査となるものである。

調査地区は次の様にして定められた。

まづ市の中を中心部 (市街区) と周辺部 (農業区) に層別した。両層とも道路によつて区切られた地区を抽出単位とした。この際道路は片側は地区に含ましめることにした。こうして層内はこの抽出単位地区に覆われるようにした。この結果、市街区は 70 単位、農業区は 97 単位となつた。調査員、調査時間の点からサンプルたるべき調査地区はかぎられたので、結局各層から夫々等しい確率を以て、市街区から 21 地区、農業層から 5 地区を抽出、これをサンプルとすること



* 国立国語研究所、統計数理研究所の文部省科学研究費による研究によつて行はれた「敬語現象に関する研究」の一部としてなされたものである。

になつた。一地区の広さは市街層では約(70 m×230 m) 農村層では約(300 m×300 m) 程度である。さて調査はこの地区について、生活模様を調べるのであるが、その内容から言つて調査時間の問題が入るので、ここでは「一日の時間平均をみる」と言う建前からサンプル地区を調査する時刻を等しい確率を以て按配することにした。

このような為、サンプルから我々の得べきものは、上野市民のa時よりb時に至る間の平均的生活場面模様の推定と言うことになるのである。

なおここで生活場面の模様と言つたが、これはいろいろな立場から分析されるのであるが、多くはある標識をもつものの比率と言つて表現されてくるものである。以上の意味から、我々の母集団は調査単位と共に時間的、空間的に明確にされてくるわけである。

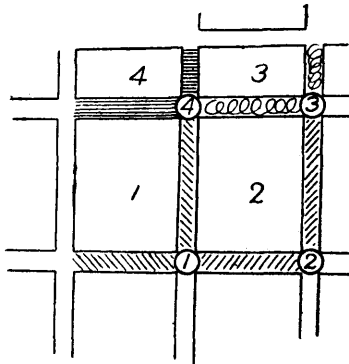
この調査のために次の様な教示を示した。

生活場面の調査の教示

a 目的

この調査の目的は人間の社会生活の場面にはどういふ種類のものが、どの位の頻度でおこつているのかと言ふことをしらべて、これからの調査の基礎とすることにある。社会生活の縮図をつくる時の資料とするものである。

このためにはもしも出来るならば、いかなる陰蔽物をも透して写る写真装置があり、これをつかつて、時間毎に、空中写真をとり、それを分析すればよいのである。しかしこのことは不可能であるから、この気持を汲んで調査を行わなければならない。実際に限られた調査員が出向いて調査を行うのであるから、当然上へのべたところをそつくりそのまま行うことは出きないが、そのくいちがいが偏りをあたえないようにすべきである。



b 調査方法

一つの調査地点は道路でくぎられた一区劃である。これが町の中のこともあるし、田舎の中のことでもあるのである。それにこだわらなければならない。

この区劃(道路を含む)の中におこりつつある場面を描き出せばよいのである。このために原則として、道路は観察で、家の中(わづかな外を含む)は面接調査によるものとする。同じく事場等においては観察がよいかもしれない。

こうしておこりつつある場面とそれを構成している人々をすべて記録するのである。

調査の手順をあげてみる。

(i) 道路はどこまでふくめるか。

四番の道路を含めるのはよくない。あらかじめ含める道路をきめておく。たとえば上図の様に

①は1へ ②は2へ ③は3へ ④は4へ

含ませればよい。あらかじめこれをきめておく必要がある。

(ii) 調査にゆくところ

道路は観察によるが家の中は観察しきれない。しかし観察できる場合もある。たとえば店で買物をしている場合などである。観察できるところは観察し、質問しなければならないときは質問して、「この私が訪問した時刻」になにをしていたかを記録しなければならない。工場、事務所等では中に案内してもらい、そこで観察によつてすべての人の行動場面を記録すればよいのである。この時の空地のあつかいも道路上の観察と同様である。田舎の中の行動も道路に準じ行動場面を記録すればよい。

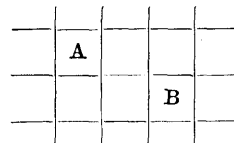
(iii) 何時の状況を記録すべきか

これは厳密に言えば調査場所に到達した瞬間である。しかし現実的にはわづかのゆとりをもたせてもよい。しかし観察は対象へ目をやつた瞬間をとるものとする。質問のとき「今」とこちらが言うだけでよい。

(c) 記録の仕方

記録は忠実に行情測や想像をまじえておならない。又意図的な構えを以て観察してはならない。「私は非情なもの」として観察しなければならない。又調査は一区劃内において各場所（一定の面積を有する）唯一回とする。この調査時刻はその場所（面積）へ目を向けた瞬間である。それが一度すぎればその場所はすでに再び調査してはならない。場所から場所への着目はなるべく一樣になる様に移動させるのが望ましい。

例えば、まづ道路をしておわつたならばもう道路は調査対象となるべき場所とわなつていない。目をつけた時に人がいなくて、家を調査して道路へ出たとき人がいても、それは調査にとつてはならない。事務所等では場所（面積）を一つづつみてゆくのであるが目をつけた瞬間丈をとるのであつて、調査をしてしまえばその場所はもう調査すべき対象ではない。目をつけたとき、Aの場所で甲という人が事務をとつていた。しかしづつと調査してゆきBへ目をつけたとき甲と言う人はBの場所で乙の人と話していた。この時には甲はAの所で調査されるし、Bの所でも調査されてよい。つまりこのようなとき甲は二回調査されるのである。対象は人ではない、家と言ひ、道路と言ひ場所の上におこつている事柄を記録すると言ひ立場でゆかねばならない。（早取り写真の如く）。



ただ家の中の見えない所は人に面接して質問してそれを書き入れると言ひ便宜的手段をとるのである。これとてもなるべく上述の気持で行うべきである。

この調査はあわてず、さわがずゆつくり行えばよいのであるから、焦る必要はない。その場面に目をやつた時の行動を記録すればよいのである。唯時刻をはつきり記入すればよいのである。

（以下略）

調査員によつてなされた調査時間は次に示す通りになつた。

調査時間は9時—12時、2時30分—5時30分の間であつた。なおこの間一人を調査するに要した時間（調査時間/その時間内に調査された人数）は平均的にみて

時 刻	0900— 1000	1000— 1100	1100— 1200	1430— 1530	1530— 1630	1630— 1930
一人当りに 要した平均 調査時間	0.38 ^分	0.57 ^分	0.43 ^分	0.52 ^分	0.57 ^分	0.71 ^分

であつた。

この結果の細目は別に発表することになつている*。

ここではその内容についてではなく、この調査によつて計上されてくる上野市における人口、家屋数**と実際の人口と家屋数との比較を行つてみることにする。

この調査によつて推定される人口は11236人となつた。推定すべきものは上野市の上述調査時間内における平均人口である。言わばうごく空間的人口の時間的平均である。これに関する信頼中を求めてみよう。

この分散をもとめてみると

$$\sigma_x^2 = \sigma_{x_1}^2 + \sigma_{x_2}^2$$

ここに $\sigma_{x_1}^2$ は地区間の分散に基因するサンプリング分散

$\sigma_{x_2}^2$ は調査時刻に関するもので、地区を単位とした場合、そこに存在する人数の時間の間の分散（人の集散状況）に基因するサンプリング分散

$$\sigma_{x_1}^2 = (167)^2 \sum p_i^2 \frac{N_i - n_i}{N_i - 1} \frac{\sigma_i^2}{n_i}$$

* この一部は、柴田武“言語生活の場面調査”・言語生活、第17号、昭和28年2月に述べられている。

** この調査では上野市と言つても、近接農村部（新しく編入されたもの）をのぞく地区を調査地点に限つてある。この中心的上野市の中にも前述の様に市街地区と農村地区とがあるのである。

$$p_i = \frac{N_i}{\sum N_i} \quad N_i \text{ は } i \text{ 層の抽出単位の数}$$

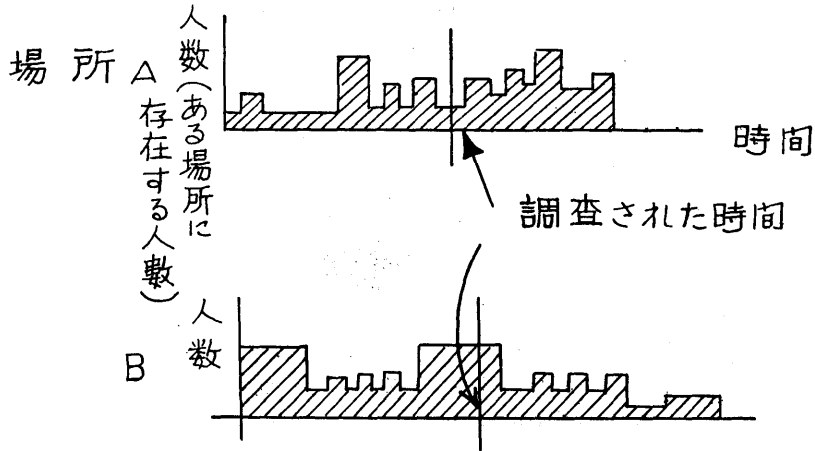
n_i は抽出サンプル数

σ_i^2 は i 層の地区間の分散

ここに $167 = \sum N_i$, $N_1 = 70$, $N_2 = 97$, $n_1 = 21$, $n_2 = 5$

これによると我々の調査の場合 $\sigma_{x_1} = 1004$ となる。

σ_{x_2} を説明してみよう。これはある着目した場所を一日中（当該調査時間中）すべてしらべ、その時間平均をとらないためにおこつてくるものである。



晝間人口を推定するのに、この晝間人口を一応平均的意味（時間についての）でのものと解しておこう。こうすると調査された時間は一つの瞬間点であるから、平均人口（ある場所の）の推定はこの調査瞬間の数字で推定しなければならなくなる。このために各場所毎に時間をサンプルするためにおこる分散を考えなくてはならない。なおこの時間のサンプリングの所は実際的にみると、厳密には疑問の余地はあるが一応こう考えるのも無理ではないと思われる。各場所毎に、この時間による分散は調査していないので、時間による分散を一応各場所間の分散で推定してその order を出してみることにしよう。この推定はきわめて大き目の推定であろうと思われる。なおこの様に場所間の分散でおきかえるとき、道路は道路、家のあるところは家のあるところで別々に分けて層別式による内分散の推定の形で行つた。ここに道路の場合は各地区において一つの単位としてとりあつた。

これによると

$$\sigma_{x_2} = 419$$

となつた。この小さくなるわけは、場所（観察単位）の数が多くなるからである。

こうすると一応

$$\sigma_x = 1088 \text{ 人}$$

推定される。ここで95(99)%の信頼度で巾をつけてみると ($k\sigma_x$ において $k=2.7, k=3.9$ を用いる)

$$\pm 2940$$

$$\pm 4240$$

となる。

即ち

$$(14170, 8300)$$

$$(15480, 6990)$$

となる。この上野市の夜間人口をみると25400なのである。この開きは何であろうか。ここで考えられるのは晝間上野市から外へ出てゆく人々である。これを切符、定期の売上枚数から推定してみ

ると

汽車電車の切符によるもの	1300
自動車 "	300
定期 "	1000
計	2600

となる。この外歩行により市外へ出るものも考えられるが近村に大工場等の機関のないところよりみて僅少のものと考えられる。

しかし外部から上野市へ来ているものもあることに注意しなければならぬ。さてこの 25400 に見合うものの推定としてこの後者を無視し、晝間の平均的意味の推定人口に上述の数を加えて、見積つてみるとそれは一応

(16800, 10900)

(18100, 9600)

となる。この見積りの方法はひとまづ妥当なものとしてよい。これによつても高々 18100 しかない。流入している人口のあること(市であるから相当あると思われる)、平均的意味がみる際に外出者の推定の多目なことを考えてみると一応この数字は最高の推定を示しているのではなからうか。

因に、調査されたものの性別比率をみると男 46.3%, 女 53.7% となり、上野市の人口調査によつて夫々 47.5%, 52.5% に近いのである。この性別による標識の比率の近さと推定人口のくいちがいは甚だ興味があるのである。

さて以上はうごく人を対象としたので一応心配がのこるので今度は戸数(家の)の推定を行つてみた。これによると、推定戸数は 2810 となる。この分散(この時は時間的変動が考えられないから地区間の分散に基因するものだけでよい)によつて 95(99)% で信頼巾をつけてみると ± 950 , ± 1370 となるから、推定の最高は一応 4180 と考えられる。しかし公式の統計の数字から求めてみると 5300 である。人口の場合と同様に約 80% 程度のもとなつている。

以上の事を考え合せてみるに、一地区内に存在するものを素手で調査する事は——あらかじめ良い航空写真か完全な地図をもたない限り——きわめて困難であることを物語つていてと考えてよいであろう。調査をよく行つているものの常識として「行きにくい、目につきにくい、きたない所」がぬけているのではないかと思われるのである。我々の一瞥による町の印象と、リストによる町の調査とのくいちがいの大きい事はよく経験される所なのである。

call back をせず又迅速な調査のための、又必然的にこれによらねばならない現在する人々をもとにする調査、名簿の次第に少なくなつてゆく将来、賞用してゆかなければならぬこの様な調査の難しさは以上の様な偏り(しかも極めて大きな偏り)を如何にして除くかにあるのである。

性別、年齢別等の比率が調査対象のそれに等しいことも保証にはならないのである。この様な調査法による調査を読むとき、この種の問題が如何に取扱われているかを見ることなしには、その調査結果は信じ難いものなのである(所謂地区の名簿作成—リストイング—における偏りの問題もこれと同様のものであろう。この偏りの大きいことは比較的地区ぬき法の容易なアメリカにおいても報告されている)。

ここでは地区ぬき法による偏りの一つの例をあげ、この種問題の解決の一つの手掛りとなれば幸である。

発表の際御注意をいただいた統計局の水野坦大兄に深く感謝するものである。