全数調査と抽出調査を併用する場合のサムプリングについて

林　知己夫
多賀　保志
高倉　節子

(1954 年 10 月受付)

Some Methods of Stratification and Sampling.

Chikio HAYASHI
Yasushi TAGA
Setsuko TAKAKURA

In the present paper, we shall treat some problems in the stratified sampling survey: sample survey in some strata and complete survey in some strata. Here the method of stratification and sampling will be devised to diminish the size of sample as possible under the required precision.

When the labels are real number, for example, total income of an enterprise, total output of a maker, total amount of deposit in a bank etc. and we will estimate the total values in our universe, generally we have to decide the point x_0, over which complete survey is performed, under which sample survey (stratified sampling or sampling of regression type) is performed, in order to robtain the optimum result. This point has not been treated theoretically. As below, the methods are described with some examples.

Institute of Statistical Mathematics

サムプリング調査において，ある層は全数調査を行い，ある層は抽出調査を行うと言うような場合が多い．これは分析の特別の目的のために行うこともあるし，又全体への推定の精度を高めるために行うこともある．後者の場合については，まずいかなる層別を行うのがよいか，次にいかなる層で全数調査を行い，いかなる層で抽出調査を行うかを合理的な根拠を以てきめることが重要な問題となる（勿論この二つのこと，即ち層別の方法と抽出の方法とは現実的には相関聯させて考えなければならないのである）．

ここでは以上のことについての一つの考え方と方法を実例について示し，層別抽出法の気持をあらわしてみたいと思う．

特にここでは通常官庁統計等においてよく出会う問題，調査しようとする標識が実数で与えられている場合について考えを進めてみよう．

§1 序　論

過去のデータが存在し，この数値が今調査しようとするものと極めて高い関係があると予想せらる場合，過去のデータを用い，如何に層別を行えばよいかを考えてみることにする．

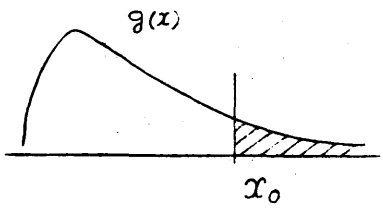
例えば金融調査における期末預金高の調査を行うものとして考えを進めよう．調査の単位は勿論銀行である．これが調査対象であり，これに等しい抽出確率を与えて第一段の母集団を構成するものとする．さて，ある過去の時期の預金高が次に調査しようとする時期の預金高と極めて関係が深いことは当然考えられるところである．なお他の要因のためにこの関係が乱されることもあるが今

はこの事は考えに入れずによく．乱されるべき要因が考えられるときには，別個に層別の要因としてとり入れ，以下にのべる考え方と併用して（或は採り入れて）企画をたてればよいのである．

§2 全数調査と層別調査を併用する場合

過去の資料としてある時期の期末の預金高が与えられ，我々は次に「次の時期の預金高の総額」を推定したいがためにサムプリングを行うとしよう．以下企画のためには，ある過去の時期の期末預金高をあらわす x の量を用いていろいろ精度を論ずることにする．これはデザインのため現在の所行いうる一つの妥当な考えと言うことができる．なお調査しようとするもの y が $y=ax+b$ (a, b は常数 $b \ll y$) である場合はここで得られた精度が全くあてはまることになる．($y=ax+b$, b が相当大のときは推定分散のみはこの方法で検討できるが変異係数については別に考えねばならない．つまり最後の精度の検討の時注意を要する)

以上のようなことは第一次近似として認められる所であろう (design の立場としてはこう考えてゆくのも妥当と思われる)．調査対象の総数は N であるとし，各調査対象に等しい抽出確率を与えて母集団を構成するものとする．こうして各銀行の預金高の分布が与えられているとする．これを分布函数の形で $F(x)$ で与えられているとする．これは近似的に密度函数の形 $g(x)$ で curve fit されるものとする．このような調査の時，ある預金高 x_0 以のものを全数調査し，それ以下のものについてはサムプリング上グを行うのが常道である．これは x_0 以上のものの数が少く且つ分散が大である場合が多いので当然考えられるところであるが，この x_0 をどこにきめるかは従来勘に委ねられている場合が多いので，ここでは x_0 をある立場から合理的にきめる一方法をのべてみることにする*．まず我々は標本の大きさ n を一定精度の下で最少にする立場で考えを進める．x_0 以上を全数調査とすると x_0 以上のサムプルの大きさは



第 1 図

$$N\int_{x_0}^{\infty} dF(x)$$

となる．x_0 以下のものを層別して標本を割当てるのであるが，割当ては分析の便宜から考えて比例割当てを用いることにする．ウェイトを用いる不便をさけるためである．R 個に層別したとし各層の大きさを N_i' とする．サムプル総数を n' とすれば各層に $n'\dfrac{N_i'}{\sum N_i'}=n_i'$ だけのサムプルが割あてられることになる．各層の分散は σ_i^2 とする．さてこれで調査を行い預金の総量 X を x で推定しようとする．

$$x=\sum N_i'\bar{x}_i+x'=\frac{\sum N_i'}{n}\sum x_i'+x'$$

$\begin{cases}\bar{x}_i \text{ は } i \text{ 層のサムプル平均，} x_i' \text{ は } i \text{ 層のサムプル総額}\\ x' \text{ は全数調査したものの総額}\end{cases}$

この分散を求めてみると $\sigma_x^2=\sum N_i'^2\sigma^2_{\bar{x}_i}$ となる．

$$\sigma^2_{\bar{x}_i}=\frac{N_i'-n_i'}{N_i'-1}\frac{\sigma_i^2}{n_i'}$$

したがつて総量推定の相対的なちらばりは

$$\frac{\sigma_x^2}{X^2}=\frac{\sigma_x^2}{N^2\bar{X}^2}=\frac{1}{\bar{X}^2}\sum_i\left(\frac{N_i'}{N}\right)^2\frac{N_i'-n_i'}{N_i'-1}\frac{\sigma_i^2}{n_i'}$$

* C. Hayashi, F. Maruyama, M. D. Ishida: On some Criteria for shatification, Annals of Statistical Mathematics, 1951, Vol. II, No. 2.

但し Nは総銀行数

$$N = N' + N\int_{x_0}^{\infty} dF(x),\quad N' = \sum N_i'$$

今 n_i は比例割当てによるものとすれば

$$\frac{\sigma_x^2}{\bar{X}^2} \fallingdotseq \frac{N'-n'}{N'-1}\frac{1}{n'}\left(\frac{N'}{N}\right)^2\frac{1}{\bar{X}^2}\sum_i \frac{N_i'}{N'}\sigma_i^2$$

$$= \frac{N'-n'}{N'-1}\frac{1}{n'}\left(\frac{N'}{N}\right)^2\frac{1}{\bar{X}^2}\sigma_w^2$$

σ_w^2 は $\sigma_w^2 = \sum_i \frac{N_i'}{N'}\sigma_i^2$ で内分散である.

一般的に考えて $(N'-n')/(N'-1) \fallingdotseq 1$ とすることはこの種のサムプリングの建前から言つて妥当であり，且つ $N'/N \fallingdotseq 1$ とすることも妥当である．後者の場合，全数調査すべきものが全体に比して十分小であることと考えるのは明らかな所であるからである.

そうすると

$$\frac{\sigma_x^2}{\bar{X}^2} \fallingdotseq \frac{\sigma_w^2}{n'\bar{X}^2}$$

となる.

我々としては

$$k\frac{\sigma_x}{\bar{X}} = \alpha,\qquad \frac{\sigma_x}{\bar{X}} = \frac{\alpha}{k} = \beta$$

として k, α, β をさだめ精度を決定してゆくのである.

さて，こうして β をさだめると精度がきまる.

我々の総サムプル数は

$$n = n' + N\int_{x_0}^{\infty} dF(x)$$

となる．或はこれを書きかえて

$$n = \frac{\sigma_w^2}{\beta^2\bar{X}^2} + N\int_{x_0}^{\infty} dF(x)$$

となる．これが最小になるように x_0 を定めればよいのである．ここで問題であるのは σ_w^2 である．これは勿論 x_0 の函数である．しかし同時に n' の函数であることも忘れてはならない．分点を求めても比例割当てをしようとする以上，$n'\frac{N_i'}{N'}$ が整数であり，且つ分散が小になるように層別するのでなければ意味がない．整数でない限り，当然必ず偏りが生ずるのであり——このように n_i' が1とか2となる場合（x_0 の近くの層における様に）が多いと予想せられる場合に於て——又ウェイトをかけるのであつてはよい所を失つて了い意味がない．このように考えを進める以上，偏りをなくすることを必須条件として考えを進める以上層別の方法（分散を小さくするように心掛けるのは言う迄もない）が調査すべきサムプルの大きさn'に依存するところが困難なところであり又興味のある所である.

σ_w^2 は詳しくは $\sigma_w(x_0, n')^2$ とかくべきである.

この $\sigma_w(x_0, n')^2$ は如何なる形をしているであろうか．これを実際の例を以てあてはめてみるに一応

$$\sigma_w(x_0, n)^2 = cx_0^a n'^b$$

の形をしているとみてよい事が判明した．こうすると

$$\frac{cx_0{}^a n'^b}{n'\bar{X}^2}=\beta^2=\frac{cx_0{}^a n'^{b-1}}{\bar{X}^2}$$

したがつて

$$n'=\left(\frac{\beta^2\bar{X}^2}{cx_0{}^a}\right)^{\frac{1}{b-1}}$$

となる．したがつて

$$n=\left(\frac{\beta^2\bar{X}^2}{c}\right)^{\frac{1}{b-1}}x_0{}^{-\frac{a}{b-1}}+N\int_{x_0}^{\infty}dF(x)$$

となる．ここで $\dfrac{\partial n}{\partial x_0}$ を求めると

$$Ng(x_0)=\left(-\frac{a}{b-1}\right)\left(\frac{\beta^2\bar{X}^2}{c}\right)^{\frac{1}{b-1}}x_0{}^{-\left(\frac{a}{b-1}+1\right)}$$

となる．

ここで実際のデータから $Ng(x)$ の曲線をかき，右辺の曲線をかき交点を求めることによつて実際的に最小の n を与えるべき x_0 を求め且つ n を知ればよいのである．

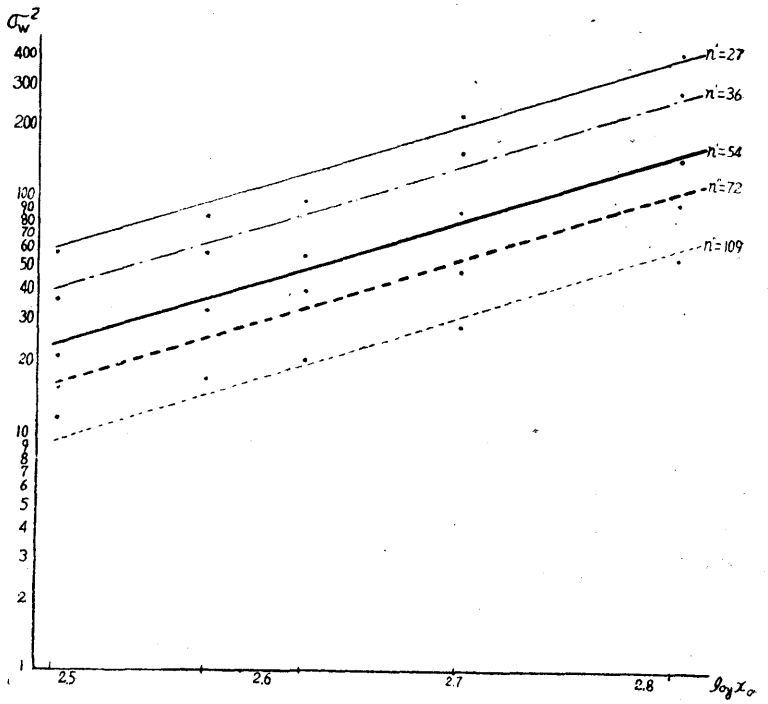
［実例］ 商業地域に在る銀行群についての例で述べてみよう．

これは昭和 28 年9月末日本銀行による全国銀行の期末預金高をもとにした数字である（このデータは日本銀行統計局の御好意により用いさせて頂いた．深く感謝の意を表するものである）．

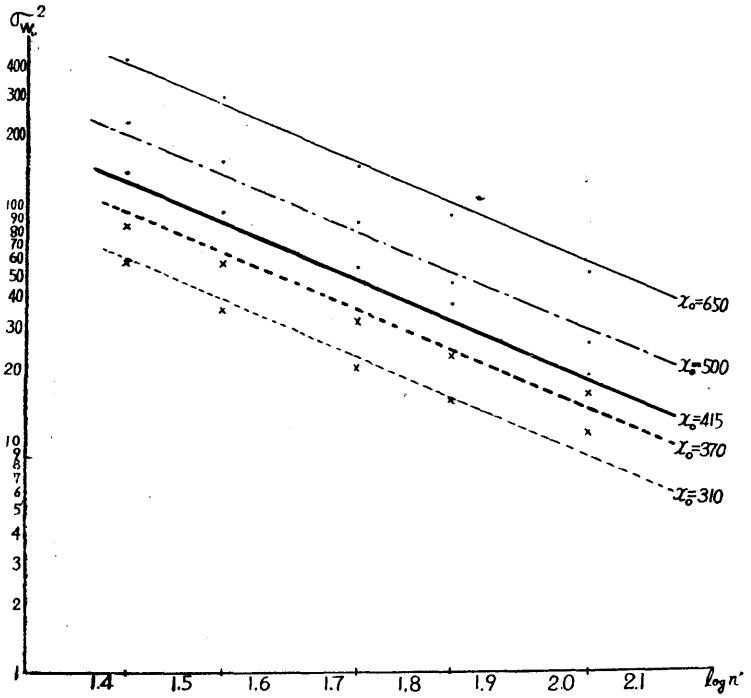
$$\alpha=0.051,\quad k=3,\quad \beta=0.017,$$

とする．一応 n' が相当大きくなり標本平均の分布がガウス分布に近いと考えられるから 99.7 % の信頼度で相対数差5％と言うねらいになる．

さて $\sigma_w(x_0,n')^2$ のグラフを実際のデータでかいてみると



第 2 図



第　3　図

となり，$cx_0^a n'^b$ で当て嵌りの良さそうなのを見てとり最小自乗法により係数をきめると

$$\begin{cases} \log c = -2.7287 \\ a = 2.5704 \\ b = -1.3265 \end{cases}$$

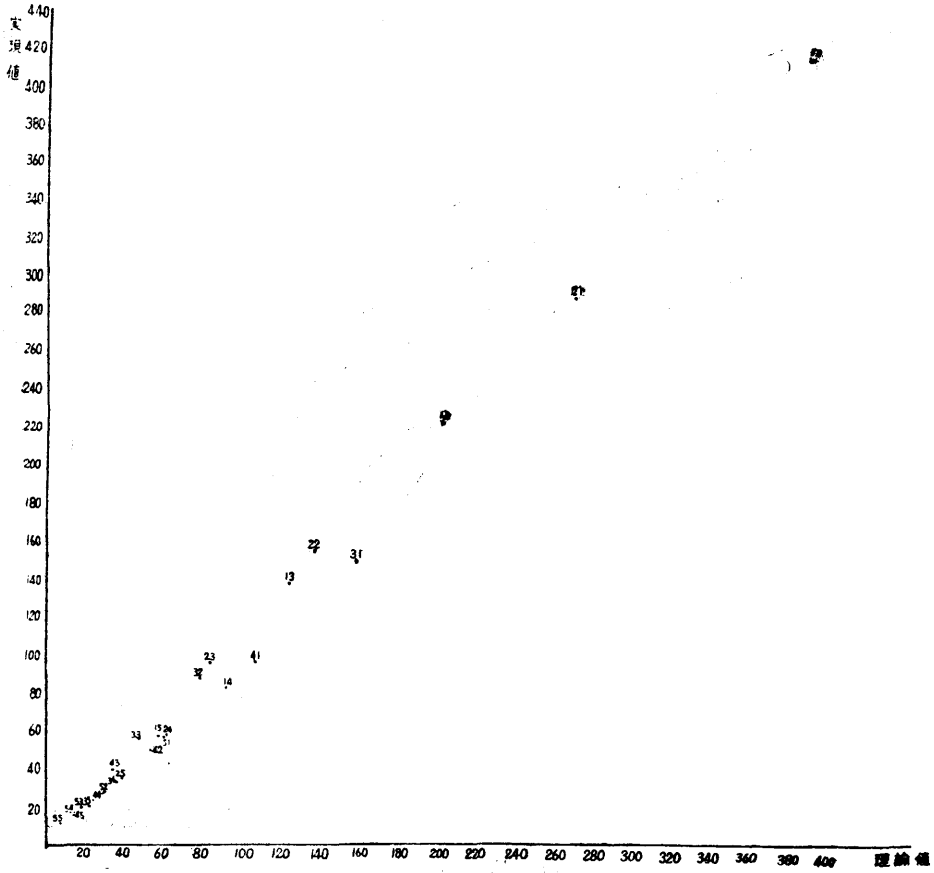
となり，かなりよく当て嵌つていることがわかる．因に理論値と実際値との関係は第4図の様になる．

なお $Ng(x_0)$ を求めるのに累積の曲線 $NF(x_0)$ から微係数をよみとることにした．このような方法はこのような場合に望ましいと考えられる．因みにもとのデータからヒストグラムをかいてみると第5図の様になつていた．

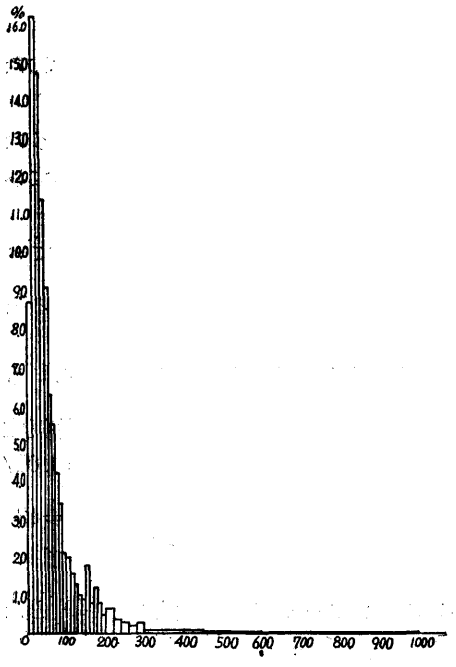
こうして両者の曲線を描き交点を求めると（第6図）x_0 として41億5千万円を得た．こうして n' を求めると　$n'=43$,　$N\int_{x_0}^{\infty} dF(x)=42$　となる．n' としてこの程度であり，且つ分布の状態をみるとき*，サンプル平均の分布はガウー分布をなすとみることは妥当である．又　$\frac{N'-n'}{N'-1}\fallingdotseq 1$, $\frac{N'}{N}\fallingdotseq 1$ も当然みとめられる．したがつてこれで所期の精度を期待できるが，いろいろ曲線のあてはめ等仮定があるので念のため検算してみると β として 0.01735 を得て，少し n' が不足するので $\beta=0.017$ とするために $n'=45$ とすることが望ましいことがわかつた．これで　$n'=45$,　$N\int_{x_0}^{\infty} dF(x)=42$,　$n=87$ としてサムプルの大きさが決定されることになる．

さて $n'=45$ であり，且つ総数 $N'=2181$ である（$N=2223$）から一つの層の大きさが 49 とな

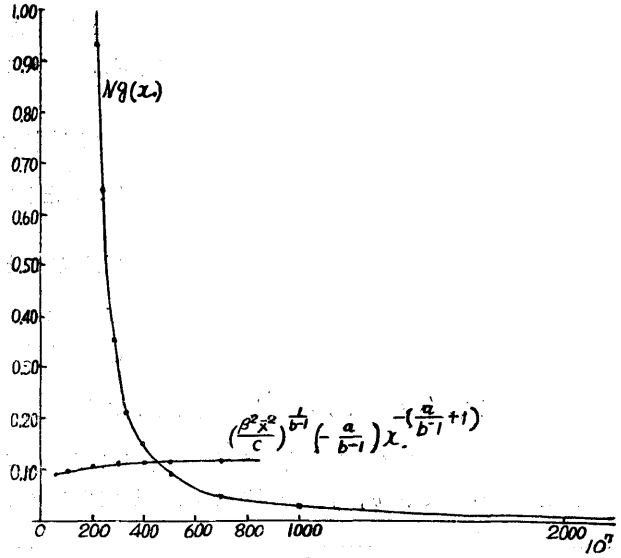
* 林知己夫：サムプリング調査はどう行うか，東大出版，p.48


実現値
理論値

第 4 図


%

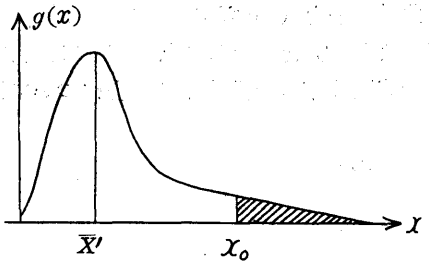
第 5 図


Nq(x)
$(\frac{\beta^2\bar{x}^2}{c})^{\frac{1}{b-1}}(-\frac{a}{b-1})x^{-(\frac{a}{b-1}+1)}$
10^7

第 6 図

るようにしなければならない．このために x の大きい順にならべ 49 となるところで分点をさだめ，層をつくり，等確率で抽出を行えばよいことになる．なお大さ 49 の層を固執することは x の大きい所（x_0 に近い所）では重大な意味をもつので，絶対必要であるが，x の小さいところでは全体に与える偏りの影響を考えて少しの狂いは問題とならない．N/n' が整数でない限り必ず最後に問題がおこるが x の小さい所では必ず層の大きさを 49 で切らずにある程度づつ加減してゆくか或は層の大きさを大きくして 49 の倍数になるようにして適宜補正して大きい層をつくり，割当サムプルの大きさをそれに比例させて偏りの量が小さくなるように心がけてゆけばよい．

さて実際に x_0 や n' の解を上に述べた様にして求めるにはかなりの手間を必要とする．それは主として実際のデータからの計算が大変な為であるから，この点を簡略化する便法も考えられるのでこの考え方を以下に説明してみよう．先ず問題となつている確率変数 X の密度函数を $g(x)$（分布函数は $F(x)$）とし，母集団を $X\leqq x_0$ なる層と $X>x_0$ なる層とに分ける（この x_0 を分割点と呼ぶことにする）．そこで第1の層（$X\leqq x_0$）から n' 個のサムプルを層別することなく抽出し，第2の層（$X>x_0$）は全数調査を行つて，母集団平均 $\bar{X}$ を推定するものとし，それに応じて合理的な分割点 x_0 を次の様に決めることにする．



第 7 図

サムプル平均 $\bar{x}$ の分散 $\sigma_x^2=c$（一定）という条件の下に，総サムプル数，$n=n'+N\int_{x_0}^{\infty}dF(x)$（$N$は母集団の大きさ）を最小ならしめるように分割点 x_0 をきめる．

x_0 の近似解は次のようにして求められる：　サムプル平均 $\bar{x}=p'\bar{x}'+(1-p')\bar{x}''$ を母集団平均 $\bar{X}$ の推定値とすれば，その分散 $\sigma_{\bar{x}}^2$ は

$$\sigma_{\bar{x}}^2=p'^2\frac{Np'-n'}{Np'-1}\frac{\sigma'^2}{n'}\fallingdotseq\left(\frac{p'}{n'}-\frac{1}{N}\right)p'\sigma'^2$$

となる．但し $p'=\int_{\infty}^{x_0}dF(x)$ は第1層のウェイト，σ'^2 は第1層内の分散で，$Np'\gg 1$ と仮定する（この仮定は無理なものではない）．そこで

$$S=n'+N(1-p')+\lambda\left\{\left(\frac{p'}{n'}-\frac{1}{N}\right)p'\sigma'^2-c\right\}$$

とおき

$$\frac{\partial S}{\partial n'}=0,\quad \frac{\partial S}{\partial x_0}=0,\quad \left(\frac{p'}{n'}-\frac{1}{N}\right)p'\sigma'^2=c \tag{*}$$

を聯立させて，n' 及び x_0 の解を求めるとよい．初めの2つの式より容易に

$$n'=\frac{Np'\sigma'^2}{(x_0-\bar{X}')^2} \tag{**}$$

を得る*（但し $\bar{X}$ は第1層の母集団平均）．

---

* 増山元三郎氏がこれと同様な考えの下に全数調査の分割点を求めている（「連続分布で近似できる有限母集団で一半は全部，他半は一部調査する境目の推定法」，講究録，第5巻，第2号）．

それによると総サムプル数 n を一定として，第1層についてのサムプル平均 $\bar{x}'$ の分散 $\sigma_{\bar{x}'}^2$ を最小ならしめる分割点 x_0 を求めると $x_0=\bar{X}'+\sigma'\sqrt{\frac{N'}{n'}+\frac{1}{N'-1}+2}$ を得る．然しながら，$\bar{X}'$，σ' は x_0 についての函数であり，n' は x_0 とは独立な変数であるから，上の x_0 は explicit な解ではない．且つこの解は，n を一定にしたという条件下に $\sigma_{\bar{x}}^2$ を最小ならしめるようにして求められているが，本来はこの稿に述べた如く総平均 $\bar{X}=p'\bar{X}'+(1-p')\bar{x}''$ の分散 $\sigma_{\bar{x}}^2$ を最小にするように x_0 をきめなければ，推定の方法としての意味がなくなつてしまう．更に第1の層を更に細かく層別する段階迄進まなければ，実際的なものとはいえない．

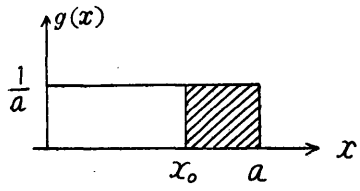
$g(x)$ の函数形が与えられなければ，これ以上 explicit にはとけないから，一般には（*）と(**) を聯立させて x_0 を求める．その一方法としては，(**) を（*）に代入すると

$$p'(x_0 - \bar{X}')^2 = cN + p'\sigma'^2$$

を得るから，この左辺及び右辺を x_0 の函数と考えて，両者の graph を描き，その交点を求めればよい．この式からみると，X の分布が著しく左に偏つていて，且つ右の方に長い尾をひく場合に x_0 の増加に伴う左辺の増加は，右辺のそれよりも急激であるから，その様な場合に限つて特に有効であることがわかる．

又右辺の cN は x_0 に無関係な定数であるから，実際には $y=p'(x_0-\bar{X}')^2$ と $y=p'\sigma'^2$ の2本の曲線を描き（いづれも原点を通る），cN の値に応じて $y=p'\sigma'^2$ の曲線をスライドさせて両者の交点より x_0 を求めることができて便利である．

母集団分布が一様分布の場合については explicit にとけるから先ずそれについては述べる．



第 8 図

$$g(x) = \frac{1}{a} \quad (0 \leqq x \leqq a)$$
$$= 0 \quad (x<0 \text{ 或は } x>a)$$

とすれば，

$$\bar{X}' = \frac{x_0}{2}, \quad p' = \frac{x_0}{a}, \quad \sigma'^2 = \frac{x_0^2}{12}$$

だから，(**) に代入して

$$n' = \frac{Nx_0}{3a}$$

を得る．これを更に（*）に代入して

$$x_0^3 = 6acN \text{ 或は } x_0 = \sqrt[3]{6acN}$$

を得る．従つて

$$n' = \frac{N}{3a}\sqrt[3]{6acN}$$

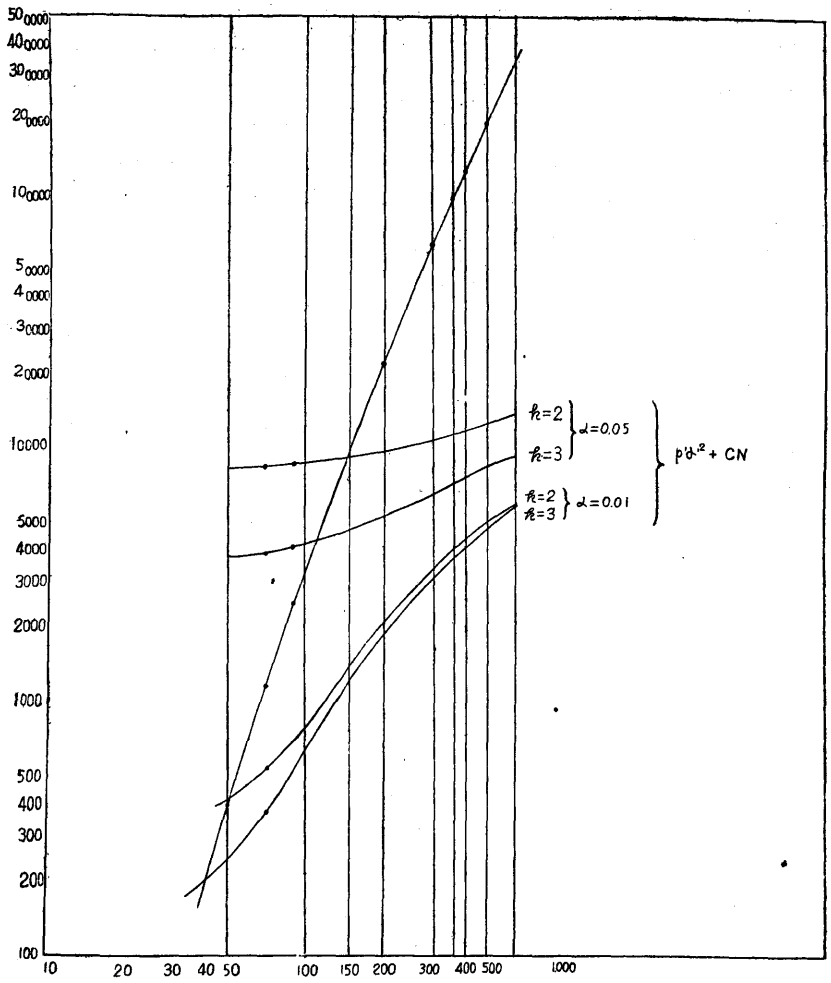
となる．

次に前の金融調査(商業)の例にあてはめてみると，第9図のようになり，分点x_0及び n', $(1-p')N$の値は第1表のようになる．これを前の結果と比べると，($x_0=41.5$ 億，$n'=45$，$(1-p')N=42$，$n=87$ であるから)，x_0 はずつと小さくなり，サムプル数は大巾に増している．これは第1の層を更に小さい層に細分しない為である．従つて先ず上のように分点 x_0 を定め，次に x_0 より小さい所を R 個の層に分けたとしよう．その時の内分数を $\sigma_w^2 = \sum_{n=1}^{R} p_i\sigma'^2_i$ とし，$\frac{\sigma_w^2}{\sigma^2}=L$ とおく．サムプル数 n' を各層の大きさに比例して割当てるものとすれば，精度を元の場合と同じにする為には，サムプル数が元の L 倍となればよい．即ちサムプル数は $(1-L)n'$ だけ少くてすむ．

第 1 表

	$k=2$	$k=3$
x_0	15.0億	11.0億
n'	210	259
$(1-p')N$	256	363
n	466	622
n_0 (単純抽出の場合)	1693	1951

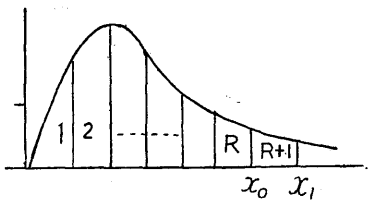
この減少した分がかなり大きいと，x_0 以上の全数調査をする層の大きさ $N(1-p')$ は，Ln' に比べてずつと大きくなるから，この場合は更に分点 x_0 を右側にずらす必要があると考えられる．何となれば，この場合には更に全数調査すべき層を細分すると，全数ではなくサムプリングをした方が有利な層が出てくるであろうからである．そしてこの際全数調査を行うべき層の大きさと，サムプ



第　9　図

リングを行うべき層からのサムプル数の和が，ほぼ釣合う様に分点をとつた時分析の上からも能率的であると考えられる．

さてその一方法としては，新しい分点 x_1 を，$(1-L)n'=\int_{x_0}^{x_1} dF(x)$ となるように定める．そして区間 $x<_0, x_1>$ に相当する $(R+1)$ 番目の層を設ける．そして元の R 個の層と併せて新たに内分散を計算し，精度を同じとして第 $(R+1)$ 層からとるべきサムプル数 n'' を計算する．その値は大体非常に小さいと考えられ，x_1 より小さい層よりとるべきサムプル数 $Ln'+n''$ と，x_1 より大きい層よりとるべきサムプル数 $(1-p')N-(1-L)n'$ とがほぼ釣合うならば，x_1 が定数の求める分点となる．この操作を繰返すことによりはやく分点をきめることも可能となる．



第 10 図

さて以上に述べたことは，片側に長く尾を引く分布の場合に特に有効であるから，応用範囲はかなり広いものと考えられる．例えば昭和 28 年度の福岡県における災害報告によると第2表のようになつている．この分布の型はほぼ指数型に近いから(第 13 図参照)，丁度右側に

第　2　表

工事種別	県 工 事	市町村工事
$\bar{N}$	3843	7448
$\bar{X}$	148.83(万円)	112.57(万円)

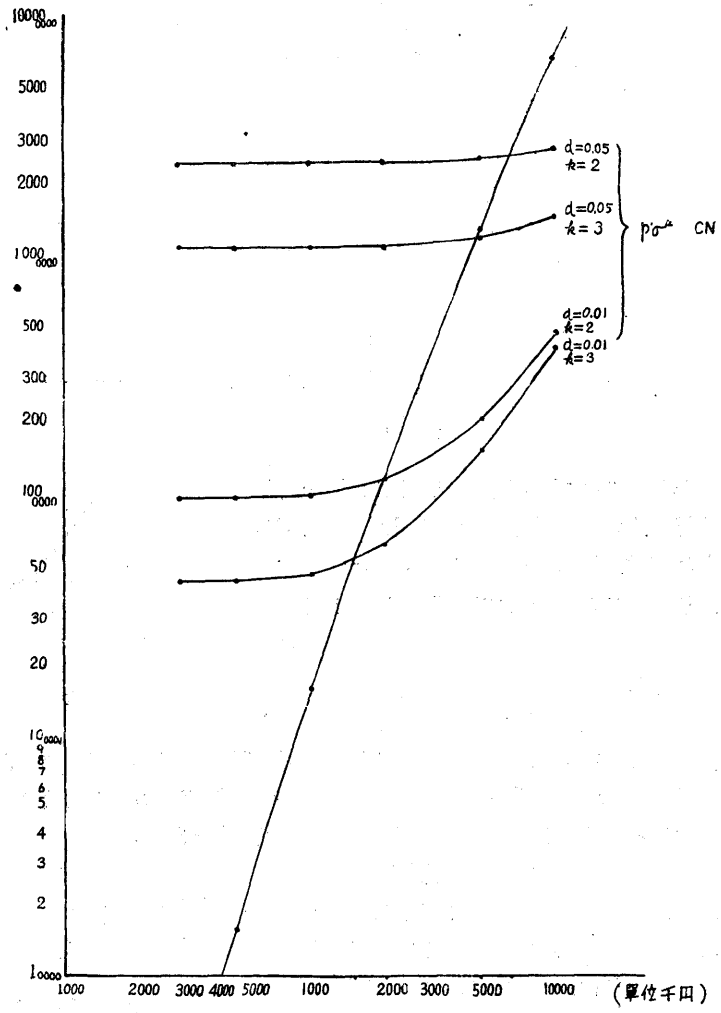
長く尾をひいている分布の好例である．そこで $y=p'(x_0-\bar{X}')^2$ と $y=cN+p'\sigma'^2$ とのグラフを描いて，その交点を求めると，第11～12図のようになる．これからみると，相対精度 α を 0.05 とすれば，分点 x_0 は第3表となる．実際に 30 万円，50 万円，100 万円，200万円，500 万円，1000 万円に分点を設け，相対精度を 0.05 とする為に必要なサムプル数を計算してみると第4表となり，上の結果とよく一致する．($k=2$ の場合)

第　2　表

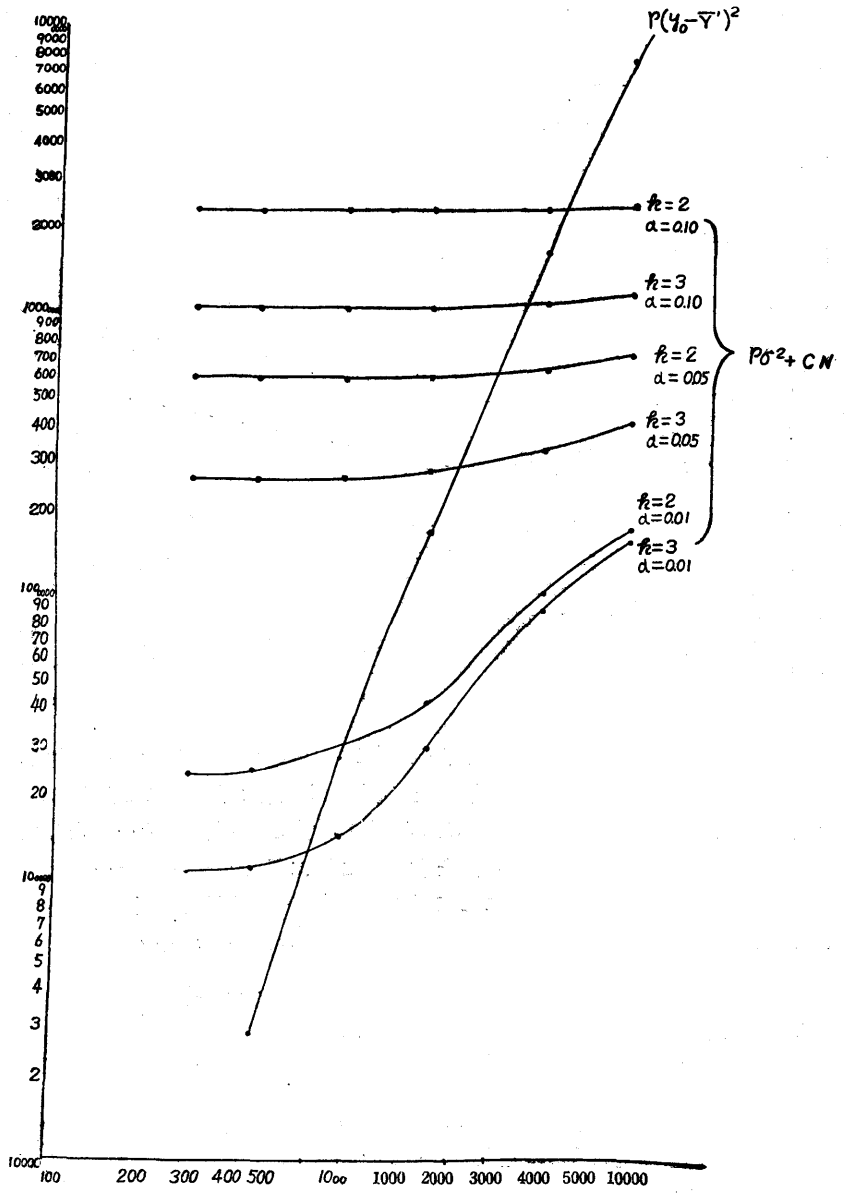
	県工事	市町村工事
$k=2$	670(万円)	580(万円)
$k=3$	490(万円)	430(万円)

第　4　表

	県　工　事			市町村工事		
分点(万円)	n'	$N(1-p')$	n	n'	$N(1-p')$	n
30	1	3100	3101	1	3106	3101
50	1	2562	2563	1	2562	2563
100	1	1845	1846	3	1845	1848
200	6	1132	1138	26	1127	1153
500	38	512	550	146	512	658
1000	133	256	389	477	256	733
—	1298	0	1298	2578	0	2578



第　11　図



第　12　図　　(単位千円)

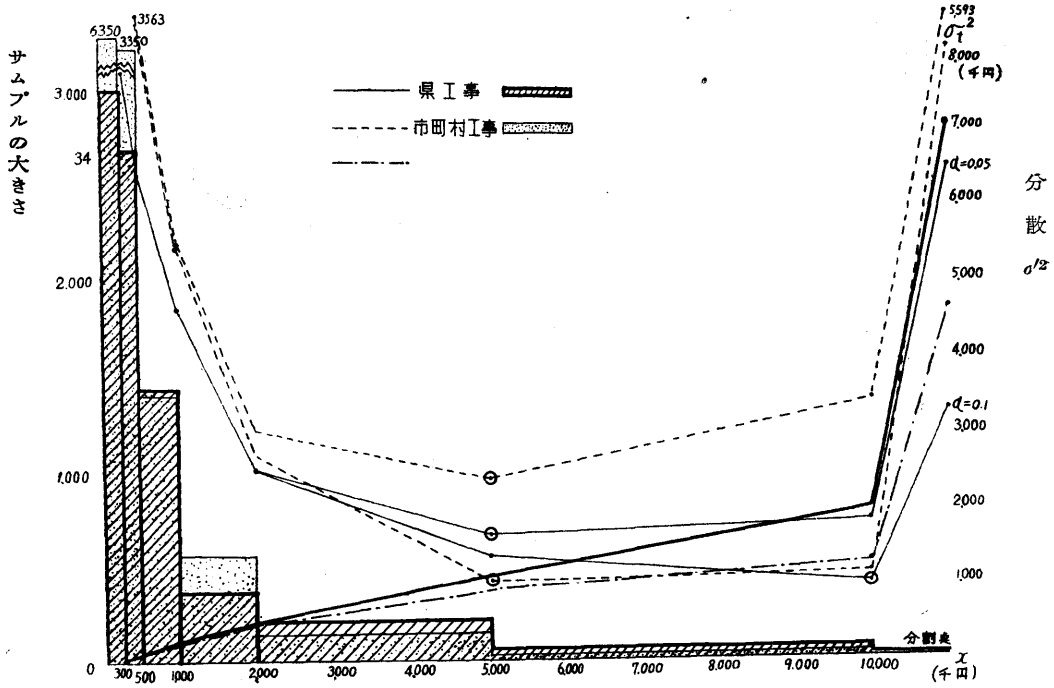
§3　全数調査と regression estimate を併用する場合

前に示した考え方に沿うものであるが，サムプリングを行うにあたつて，regression estimate を用い得る場合もあると考えられるのでそれについてのべてみよう．

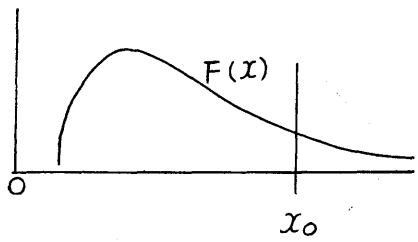
過去のデータと現在得らるべきデータとの間において母集団に直線的相関があると見做され得る場合（Nが十分である必要はある），（なおその直線のまわりの分散が一定である場合はさらに有利である*）にこの方法が用えられる．今これがみたされているとしよう．

x_0 以上を全数調査するとすればこのサムプルの大きさは $N\int_{x_0}^{\infty} dF(x)$ によつてあたえられる．

* 森林調査における統計数理的問題，統計数理研究所，第1巻，第2号．



第　13　図



第　14　図

x_0 以下の所で regression estimate を用いるのである（前の場は層別抽出を行つた）．ここから n' のサムプルを等しい確率で抽出し，現在のデータとして y_j を調査したとする．これに見合う過去のデータを x_j とする．ここで (x, y) 間の回帰直線と引く，これを

$$\tilde{y}=b(x-\bar{x}')+\bar{y}'$$

但し
$$\begin{cases}\bar{x}'=\dfrac{1}{n}\sum x_j\\ \bar{y}'=\dfrac{1}{n}\sum y_j\end{cases}$$

$$b=\frac{\sum(x_j-\bar{x}')(y_j-\bar{y}')}{\sum(x_j-\bar{x}')^2}$$

とする．ここで

$$\bar{y}=b(\bar{X}'-\bar{x}')+\bar{y}'$$

但し $\bar{X}'$ は x_0 以下の母集団での総平均とする．

として平均を推定し，しかる後

$$y=N'\bar{y}+y''$$

但し
$$\begin{cases}N'=N\int_0^{x_0}dF(x)\\ y''\text{ は全数調査したものについての総額}\end{cases}$$

によつて金額を推定する．

さてこの y の分散は近似的に（n' があまり小でない限り，この程度については前前の 21 頁脚註の論文参照）

$$\sigma_{\bar{y}}^2 = N'^2\sigma_{\bar{y}}'^2 \fallingdotseq N'^2\frac{\sigma'^2(1-\rho'^2)}{n'}$$

によつて与えられるとみてよかろう．ここに σ'^2 は分散で

$$\sigma'^2 = \frac{\int_0^{x_0}(y-\bar{Y}')^2 dG(y)}{\int_0^{x_0} dG(y)}$$

但し $G(y)$ は y の分布函数，$\bar{Y}'$ は母集団平均をあらわす．ρ' は y と x との母集団相関係数をあらわす．

ここにこれらは x_0 の函数となつていることに注意すべきである．

さてここで相対的にみると

$$\frac{N'^2\sigma'^2(1-\rho'^2)}{N^2\bar{Y}^2 n'} \fallingdotseq \frac{\sigma'^2(1-\rho'^2)}{\bar{Y}^2 n'}$$

但し $\bar{Y}$ は総平均，N は総数であるが一般に第一次近似として $N' \fallingdotseq N$ と考えてよい．これについては前の所参照．

となる．ここで $\frac{\sigma'^2}{\bar{Y}^2}$ は $\frac{\sigma_x'^2}{\bar{X}^2}$ と略同一の様子を示すと考えてみる．（これは X と Y との相関が略分点を通ることつまり比例関係にあることを意味する）これも第一次近似として妥当であろう．ここに $\sigma_{x'}^2$ は x_0 以下のものについての x の分散である．そうすると

$$\frac{\sigma_y^2}{N^2\bar{Y}^2} \fallingdotseq \frac{\sigma_{x'}^2(1-\rho'^2)}{\bar{X}^2 n'}$$

となる．これを前の場合と同様に

$$k^2\frac{\sigma_{x'}^2(1-\rho'^2)}{\bar{X}^2 n'} = \alpha^2 \quad , \qquad \beta = \frac{\alpha}{k}$$

とすると

$$n' = \frac{\sigma_{x'}^2(1-\rho'^2)}{\bar{X}^2\beta^2}$$

となる．

そこで

$$n = n' + N\int_{x_0}^{\infty} dF(x)$$

を考え n を最小にする x_0 を $\frac{\partial n}{\partial x_0} = 0$ から求めればよい．

さて第一項が問題となるが

$$\sigma_{x'}^2 = \frac{\int_0^{x_0}(x-\bar{X}')^2 dF(x)}{\int_0^{x_0} dF(x)}$$

で x_0 の函数であり ρ' も亦 x_0 の函数である．これは当然考えられる所である．$\sigma_{x'}^2$ については $F(x)$ を知れば x_0 の函数として容易に求められるが実際のデータにあたつて求め適宜曲線の当て嵌めを行い $\sigma_{x'}^2$ を x_0 の函数としてあらわすのがよい．又 ρ'^2 についても実際のデータから x_0 の

値の函数として ρ^2 をプロットしこれに曲線のあてはめを行うのがよい．或は又 $\sigma_{x'^2}(1-\rho^2)$ を一まとめにしてプロットし曲線のあてはめを行うことも実際的な考え方である．

註　曲線としては操作の容易なものたとえば $\log x$ とか e^{-bx} とか cx^a とかを含む簡単なものが望ましい．

かうして $\sigma_{x'^2}(1-\rho^2)=\varphi(x_0)$ と書けるならば

$$n=\frac{\varphi(x_0)}{X^2\beta^2}+N\int_{x_0}^{\infty}dF(x)$$

となり，前と同様の方法を用いて実際に最高の精度を得べき x_0 及び n を求めることが可能となる．

こうして x_0 及び n をさだめその精度を計算し，前にのべた場合と比較し，いづれが得策であるかを決定するのが企画の立場からは望ましい所である．

なお，さらに企画を進め double sampling を用いてより妥当性ある企画をつくり，妥当な意味に於てサムプルの大きさを減ずることも考えてよい．

（統計数理研究所）