⑥

比率の一次形式パラメーターの推定について

遠藤　健児

§1. 標本調査による時間研究法

労働生理の研究において，作業の強度やエネルギーの代謝率などを算出するのに，時間研究法と呼ばれる調査法が知られている.

或る職種に従事する労働者は，定められた労働時間をいく種類かの作業を行うのに費すが，その各々の作業に従事する時間を測定するためには種々の観測方法が考えられる。

そのうちで Snap-reading Method と謂われるサムプリング調査の方式を応用する方法がいろいろな点で都合のよいものとして採り上げ上げられるようになっている（愛媛県立産業能率研究所資料 52-01，昭和27年1月）．この方法では，適当な個数の観測時点を予め作為なく選んでおいて，その時点に被観測者（労働者）がどの作業を行っているかを観測して，その頻度によって全作業時間を推定するのである。

この方法に依ると，被観測者に観察されているという意識を持たせずにすむ点，或は調査員に課せられる手間や労力が少くてすむといった点で好都合である.

勿論，サムプリング調査法の持つその他の長所を利用出来ることはいうまでもない。

§2 単純任意観測法

予め定められたT時間のうち，k種類の作業

$$W_1,\ W_2,\ \cdots\cdots,\ W_i,\ \cdots\cdots,\ W_k$$

の各々に従事すべき時間が夫々

$$T_1,\ T_2,\ \cdots\cdots,\ T_i,\ \cdots\cdots,\ T_k$$

$$(\Sigma T_i = T)$$

であるとしよう。 更に，各作業に従事する時間の割合を

$$P_i = T_i / T \quad (i = 1, 2, \cdots\cdots, k)$$

とおく。

ここで各作業について，エネルギーの代謝率といった或る種の係数

$$\lambda_1,\ \lambda_2,\ \cdots\cdots,\ \lambda_i,\ \cdots\cdots,\ \lambda_k$$

が対応して定められているものとする。 このとき知りたいのは

$$(1) \qquad R = \lambda_1 P_1 + \lambda_2 P_2 + \cdots\cdots + \lambda_k P_k$$

の形の式で与えられるRの値である。

このRの値を推定するには，時間比率 P_i の推定値が得られればよいが，このためには予め選ばれたn個の観測時点について，作業 W_i に従事していた時点の個数を x_i とするならば，その相対頻度

$$(2) \qquad p_i = x_i / n$$

をとればよい。 この統計量 p_i が，時間比率 P_i に対する不偏推定値であつて，観測時点の数がかなり大きいとき* にはその分布は正規型に近く，その標準誤差は

註 *) $nP_i > 5$ 且つ $n(1 - P_i) > 5$ となる程度にnが大きければよい。

(3) $$D(p_i) = \sqrt{P_i(1-P_i)/n}$$

で與えられることはよく知られている。 従つて

(4) $$r = \lambda_1 p_1 + \lambda_2 p_2 + \cdots\cdots + \lambda_k p_k$$

の形の一次形式統計量をとれば，これは R に対する不偏推定値となる。

一時点を観測すると，その時点がどの作業で占められているかということがわかるのであるが，この観測結果を W_i によつて占められているときには 1，占められていないときには 0 という値をとる変数 X_i $(i=1,2,\cdots\cdots,k)$ の組

(5) $$X = (X_1, X_2, \cdots X_k)$$

によつて示すことが出来る。 例えば W_1 によつて占められているときには，その観測結果を

$$X = (1, 0, 0, \cdots\cdots, 0)$$

と示すのである。 ここで X_i が 1 となる確率は P_i であるから，X は多項分布

$$(P_1 + P_2 + \cdots\cdots + P_k)'$$

に従うことになる。 従つて

(6) $$\begin{cases} \text{平　均：} \quad E(X_i) = P_i \\ \text{分散，共分散：} \quad E\{(X_i - P_i)(X_j - P_j)\} \\ \qquad\qquad\qquad\qquad = P_i(\delta_{ij} - P_j) \end{cases}$$

となる。

また

$$\lambda X = \lambda_1 X_1 + \lambda_2 X_2 + \cdots\cdots + \lambda_k X_k$$

の平均は

$$\lambda_1 P_1 + \lambda_2 P_2 + \cdots\cdots + \lambda_k P_k = R \ ,$$

その分散は

$$(7) \quad D^2(\lambda X) = E\left\{\sum \lambda_i (X_i - P_i)\right\}^2 = \sum_{i,j} \lambda_i \lambda_j P_i (\delta_{ij} - P_j)$$

$$= \sum P_i (\lambda_i - R)^2$$

$$= \frac{1}{2} \sum_{i,j} P_i P_j (\lambda_i - \lambda_j)^2$$

となることがわかる。　この式はまた次のように書きかえて評価することも出來る。

$$E\left\{\sum \lambda_i (X_i - P_i)\right\}^2$$

$$= \sum \lambda_i^2 P_i (1 - P_i) - \sum_{i \neq j} \lambda_i \lambda_j P_i P_j$$

n 個の時点について観測を行つたとき，例えば W_i に関する相対頻度は，各観測結果を表わす X の成分 X_i の和（ $S(X_i) = x_i$ とかく）を n で割つたものである。　即ち

$$p_i = x_i / n = S(X_i) / n$$

このことから，p_i の平均，標準誤差が夫々 (2) (3) のようになることは明らかである。

同様に考えれば

$$r = \sum \lambda_i p_i = \frac{1}{n} S(\lambda_1 X_1 + \cdots\cdots + \lambda_k X_k)$$

であるから，r の平均は R，その分散は

$$(8) \quad D^2(r) = \frac{1}{n^2} \cdot n D^2(\lambda X)$$

$$= \frac{1}{n} \sum_i P_i (\lambda_i - R)^2$$

$$= \frac{1}{n} \left\{ \sum_i \lambda_i^2 P_i (1 - P_i) - \sum_{i \neq j} \lambda_i \lambda_j P_i P_j \right\}$$

となる。

このような推定を行う際に，予め要求される程度の精度が得られるようにするには，(3) 或は (5) にもとずいて n を適当に定めればよい。

例えば，時間比率を % で表わすとき，特定の時間比率 $100P$ % に関して，β % の確率で期待される誤差の幅を α % 以内に抑えたいならば，正規分布の ($100-\beta$) % 点を $t(\beta)$ とすれば

$$t(\beta) \cdot \sqrt{\frac{P(1-P)}{n}} \leqq \frac{\alpha}{100}$$

となるような n を以て観測時点の数とすればよい。

こゝで P は未知であるが，その大体の値を以て計算を行えばよい。（第一図参照）

一般に $P(1-P)$ は，$P=0.5$ のとき最大 ($=1/4$) で，P が 0.5 から隔るに従って単調に減小するから，$P=0.5$ とおいたときの n (或は $P_1, P_2, \cdots\cdots, P_k$ のうちで最大と考えられる値に対する n) をとれば，どの時間比率の推定においても同程度，或はそれ以上の精度が保たれることになる。

また (1) の如き R の値を (4) の r で推定する際には，個々の時間比率 P_i に対する誤差をその絶対的な数値によって抑えるだけでは，R に関して r が持つ誤差の，相対的な大きさを抑えることが出来ない。 例えば r の持つ誤差の大さの大きさを，真の価 R の 10 分の 1 程度に抑えたいような場合には，p_i の絶対的な誤差を考えるだけでは n が定まらない。

特に時間比率が異常に小さな作業 W_i に対して，係数 λ_i がか

なり大きいような場合には，への誤差としてその作業 W_i がかなりの寄與をすることになる。

この様な場合には，時間比率 P_i に対する推定値 p_i の相対的な誤差を考慮する必要がある。

ところで，100 P % という時間百分率の推定に際して，（相対的）誤差の大きさが真の価の ε % 以下となることが β % の確率で期待出來るために必要な観測時点の数 n は，正規分布における 100−β % 点を $t(\beta)$ とするならば

$$(9)\qquad t(\beta)\sqrt{\frac{1-P}{nP}} \leqq \frac{\varepsilon}{100}$$

によって求めることが出來る。（第二図参照）

一般に $(1-P)/P$ は P について單調減少の函数だから，全程度の精度を保つためには，P が小さいほど n は大きくとらねばならない。

いまどの作業の時間比率の推定においても，上と同程度或はそれ以上の精度が保たれているように n を選んだものとすると，(9) をかきかえた。

$$\frac{1}{n}P_i(1-P_i) \leqq \left(\frac{\varepsilon P_i}{100\,t(\beta)}\right)^2$$

$$(i=1, 2, \cdots\cdots, k)$$

という関係が成立つわけだから，(8) によれば

$$D^2(r) = \frac{1}{n}\left\{\sum \lambda_i^2 P_i(1-P_i) - \sum_{i\neq j}\lambda_i\lambda_j P_i P_j\right\}$$

$$\leqq \left(\frac{\varepsilon}{100\,t(\beta)}\right)^2\left\{\sum_i \lambda_i^2 P_i^2 + \sum_{i\neq j}\lambda_i\lambda_j P_i P_j\right\}$$

$$-\left\{\frac{1}{n}+\left(\frac{\varepsilon}{100\,t(\beta)}\right)^2\right\}\sum_{i\neq j}\lambda_i\lambda_j P_i P_j$$

$$< \left(\frac{\varepsilon}{100\,t(\beta)} \right)^2 \left(\Sigma \lambda_i P_i \right)^2$$

従って

$$t(\beta)\,\frac{D(r)}{R} < \frac{\varepsilon}{100}$$

という関係が成立つことになる。　rの分布はやはり正規型に近いから，このことはrの相対的誤差の大きさが，真の価のε %以下であることがβ %の確率で期待出來ることを示している。

P も λ も共に小さい作業については，rの誤差を評価する(8)の右辺で，この項による寄與は無視出來るので，これを除いた他の作業（のうちで時間比率が最小と考えられるもの）について精度を考慮してれを定めれば十分である。

観測時点の個数は，必ずしも上述の如く予め計画された精度が得られるように定めるものとは限らない。　調査員の人数や，調査の手間や費用に関する考慮から定められる場合も少くない。

このようなとき，実際に到達し得る精度は，(3)或は(8)の式から計算される。

仝一職種に従うものが幾人かあるときには，定められた個数の観測時点をその幾人かに割当てることにすれば，被観測者一人当りの観測時点の個数をそれだけ少くてすませることが出來る。

§3. 層別観測法

全時間を予めいくつかの部分に分割した上で，その各々に適当な個数の観測時点を按配する方法が考えられる。

いま全時間Tをk個の部分

$$S_1,\ S_2,\ \cdots\cdots,\ S_\nu,\ \cdots\cdots,\ S_k$$

に分割するものとして，各S_νの長さ（時間）をN_ν，S_νにおける作業W_iの時間比率を$P_{\nu i}$とすれば

$$N_1 + N_2 + \cdots\cdots + N_k = T$$

$$P_{\nu 1} + P_{\nu 2} + \cdots\cdots + P_{\nu k} = 1$$

$$(\nu = 1, 2, \cdots\cdots, k)$$

であって，全体におけるW_iの時間比率P_iは

$$P_i = (N_1 P_{1i} + N_2 P_{2i} + \cdots\cdots + N_k P_{ki}) / T$$

となっている。

更にS_νで考えられる(1)のRに対応するパラメーターを

$$(10) \qquad R_\nu = \lambda_1 P_{\nu 1} + \lambda_2 P_{\nu 2} + \cdots\cdots + \lambda_k P_{\nu k}$$

とおくと

$$(11) \qquad R = \sum_i \lambda_i P_i = \frac{1}{T} \sum_\nu N_\nu R_\nu$$

となる。

こゝでn個の総観測時点をk個の部分に夫々

$$n_1,\ n_2,\ \cdots\cdots,\ n_k$$

個宛按分するものとして，S_νにおいてn_ν個の観測結果から得

られる W_i に関する相対頻度を $p_{\nu i}$ とすれば，

$$r_\nu = \lambda_1 p_{\nu 1} + \lambda_2 p_{\nu 2} + \cdots\cdots + \lambda_k p_{\nu k}$$

は (10) の R_ν に対する不偏推定値であって，従ってまた

$$r = \frac{1}{T} \Sigma N_\nu r_\nu$$

は (11) の R に対する不偏推定値となる。　各部分での観測結果は独立であるから，この r の分散は

$$\begin{cases} D^2(r) = \dfrac{1}{T^2} \Sigma N_\nu^2 D^2(r_\nu) \\ D^2(r_\nu) = \dfrac{1}{n_\nu} \sum_i P_{\nu i} (\lambda_i - R_\nu)^2 \end{cases}$$

となる。

これに対して，部分に分けない場合とは，予め $n_1, n_2, \cdots, n_k$ を指定しないで，これを作為なくとられた n 個の時点のうち各部分におちたものの頻度を考えて，R の推定値として

$$r = \frac{1}{n} \sum_\nu n_\nu r_\nu$$

をとることに他ならない。

ところで，特定の（$n_1, n_2, \cdots\cdots, n_k$）なる頻度が得られた上で考えると，上述の r の平均と分散は夫々

$$\frac{1}{n} \sum_\nu n_\nu R_\nu$$

$$\frac{1}{n^2} \sum_\nu n_\nu^2 D^2(r_\nu) = \frac{1}{n^2} \sum_\nu n_\nu \Sigma P_{\nu i} (\lambda_i - R_\nu)^2$$

となる。　次に頻度が種々に変る場合を考慮すると，（$n_1, n_2, \cdots, n_k$）は多項分布

$$\left(\frac{N_1}{T} + \frac{N_2}{T} + \cdots\cdots + \frac{N_k}{T} \right)^n$$

に従う変量だから

$$E(n_\nu) = n \cdot N_\nu / T$$

$$E\{n_\nu - E(n_\nu)\}\{n_\mu - E(n_\mu)\} = n(N_\nu/T)\{\delta_{\mu\nu} - N_\mu/T\}$$

となることから

$$\begin{cases} D^2\left(\frac{1}{n}\sum_\nu n_\nu R_\nu\right) = \frac{1}{n}\sum_{\mu,\nu} R_\mu R_\nu \frac{N_\mu}{T}\left(\delta_{\mu\nu} - \frac{N_\nu}{T}\right) \\ \qquad\qquad = \frac{1}{n}\Sigma N_\nu (R_\nu - R)^2 / T \\ E\left(\frac{1}{n^2}\sum_\nu n_\nu^2 D^2(r_\nu)\right) = \frac{1}{n}\sum_\nu \frac{N_\nu}{T}\Sigma P_{\nu i}(\lambda_i - R_\nu)^2 \end{cases}$$

となる。 従って一般の場合の r の分散は

$$D^2(r) = \frac{1}{n}\left[\frac{1}{T}\sum_\nu N_\nu (R_\nu - R)^2\right] + \frac{1}{n}\sum_\nu \frac{N_\nu}{T}\sum_i P_{\nu i}(\lambda_i - R_\nu)^2$$

となる。

扨て部分に分けたときの推定値 r を求める際に，各部分への割当て時点の数 n_ν を，S_ν の長さ N_ν に比例するように定めてあるものとすれば *

$$\frac{n_1}{N_1} = \frac{n_2}{N_2} = \cdots\cdots = \frac{n_k}{N_k} = \frac{n}{T}$$

即ち

$$\frac{N_\nu}{n_\nu} = \frac{T}{n}$$

$$(\nu = 1, 2, \cdots\cdots, k)$$

註 * この場合には，推定値 r は次のように表わされる。

$$r = \sigma_1 p_1 + \lambda_2 p_2 + \cdots\cdots + \lambda_k p_k$$

となっているから

$$D^2(r) = \sum_\nu \frac{N_\nu^2}{T^2} D^2(r_\nu)$$

(13)

$$= \sum_\nu \frac{N_\nu^2}{T^2} \frac{1}{n_\nu} \sum_i P_{\nu i}(\lambda_i - R_\nu)^2$$

$$= \frac{1}{n} \cdot \sum_\nu \frac{N_\nu}{T} \sum_i P_{\nu i}(\lambda_i - R_\nu)^2$$

となる。　(12)と(13)を比較すれば，部分に分けて推定する方が r の分散に関しては，

$$\frac{1}{n} \sum_\nu \frac{N_\nu}{T} (R_\nu - R)^2$$

だけ小さくなることがわかる。　従つて，各部分相互間の R_ν の値が出来るだけ異るように，換言するならば各部分内部では時間比率を重みにして考えたとき，λの変動が小さくなるように分割すれば，それだけ精度が向上することになる。

このように分けた各部分を層と呼ぶ。

個々の時間比率，例えば P_1 を推定することが問題となつている場合には，その精度を評価するのに必要な，推定値 p_1 の分散を与える式は，上述の結果の式で

$$\lambda_1 = 1, \quad \lambda_2 = \lambda_3 = \text{-----} = \lambda_k = 0$$

従つて

$$r_\nu = p_{\nu 1}, \quad r = \frac{1}{T} \sum_\nu N_\nu p_{\nu 1}, \quad R_\nu = P_{\nu 1}, \quad R = P_1$$

とおけば求められる。

観測時点を比例的に割当てたときには，R の推定値 r は，標本全体からの相対頻度によつて(4)と全じ形にかける。

また n 個の観測時点を何人かの同職種のものに分割して観測を行うのは，層別法の特殊な場合と考えることができる。

§4. 系統的観測

調査計画に際して，種々の観点に基ずく考慮を行つて適当な観測時点の個数が定められたとき，それだけの観測時点は作為なく抽出するのであるから，全時間の極く限られた一部分にしか撒布されないこともあり得るし，また前後の順によつて時点を排列したとき，隣り合つた二時点間の隔りは區々まちまちとなるのが一般である。　この点は調査を能率的に行う上には聊か具合が悪いと思われる。

このために，系統的観測法を代用することが考えられる．

この方法は，定められた観測時点の個数nによつてまづ抽出間隔

$$d = T/n$$

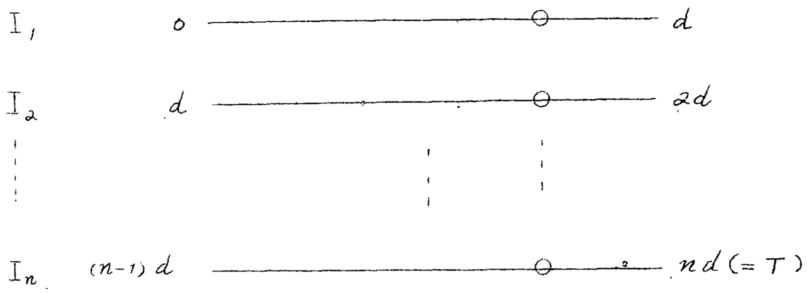
を定めて，最初からd時間までの時刻のうちから作為なく一時点（時には数個の時点，例えばm個の時点）を選ぶ。　そしてこの時点を出発点として以後d時間毎に観測時点をとつて観測を行うのである。　この方法によると，抽出される観測時点の数は勿論n（m個の時点から出発したときにはmn）である。

時間比率に対する推定の方法は今までと同様に観測された相対頻度によればよいのであるが，系統的に観測を行うためにその精度は異つてくることなる。

いまT時間に亘る全区間をd（$=T/n$）時間毎に区切つてできるn個の部分区間

$$I_1,\ I_2,\ \cdots\cdots,\ I_n$$

をとつて，これを次のように重ね合わせた図式で示そう．



各区間の起点から全一の経過時間後に起るn個の時点を一纏めにして考えると，選ばれるべきn個の観測時点は，このような組の全体のうちの一組ということになる。

いま観測されるべき時点の各々に(5)の如き変量を対応させて考えるならば，実際に観測されたn個の時点の組から得られた観測結果から

$$r_s = S\left(\lambda_1 X_1 + \lambda_2 X_2 + \cdots\cdots + \lambda_k X_k\right)/n$$

を求めれば，これがパラメーターRの推定値となる。何故ならば，例えばν番目の区間I_νの作業構成を，前に考えた層S_νと同様の記号で表わすならば（この場合各部分の長さN_νはいずれもT/nで，$h=n$），

$$E\left\{S(\lambda_i X_i)/n\right\} = \frac{1}{n} \underset{\nu}{S} \lambda_i P_{\nu i} = \lambda_i P_i$$

$$\left(\because\ P_i = \sum_\nu \frac{T}{n} P_{\nu i} \div T = \frac{1}{n} \sum_\nu P_{\nu i}\right)$$

となるからである。

いま観測されたn個の点の組のうちで作業W_iによって占められるものの相対頻度を

$$p_i = S(X_i)/n$$

とおくと，上述のr_sは，

$$r_s = \lambda_1 p_1 + \lambda_2 p_2 + \cdots\cdots + \lambda_k p_k$$

となる。 形の上では，この推定値は（4）と全く同様である．

上に述べたように r_s は R の不偏推定値であるが，こゝでその分散を

$$D^2(r_s) = E(r_s - R)^2$$

とおくことにする。

いま一時点の観測結果から得られる統計量

$$\lambda X = \lambda_1 X_1 + \lambda_2 X_2 + \cdots\cdots\cdots + \lambda_k X_k$$

の分散は，（7）で与えられるが，前述の n 個の点の一組が得られた上で考えると

$$E(\lambda X) = \lambda_1 p_1 + \lambda_2 p_2 + \cdots\cdots + \lambda_k p_k = r_s$$

$$D^2(\lambda X) = \sum_i p_i (\lambda_i - r_s)^2$$

となっている。 従って（7）で示される λX の全分散は次のように二つの成分から成る。

$$\sum P_i (\lambda_i - R)^2 = E(r_s - R)^2 + \sum_i E\left[\, p_i (\lambda_i - r_s)^2 \,\right]$$

この右辺の第一項は，系統的観測で得られる推定値 r_s の分散に他ならない。

従って系統的観測を行わない通常の場合の r（(4)）の分散は，（8）を参照して

$$D^2(r) = \frac{1}{n} E(r_s - R)^2 + \frac{1}{n} E\left\{ \sum_i p_i (\lambda_i - r_s)^2 \right\}$$

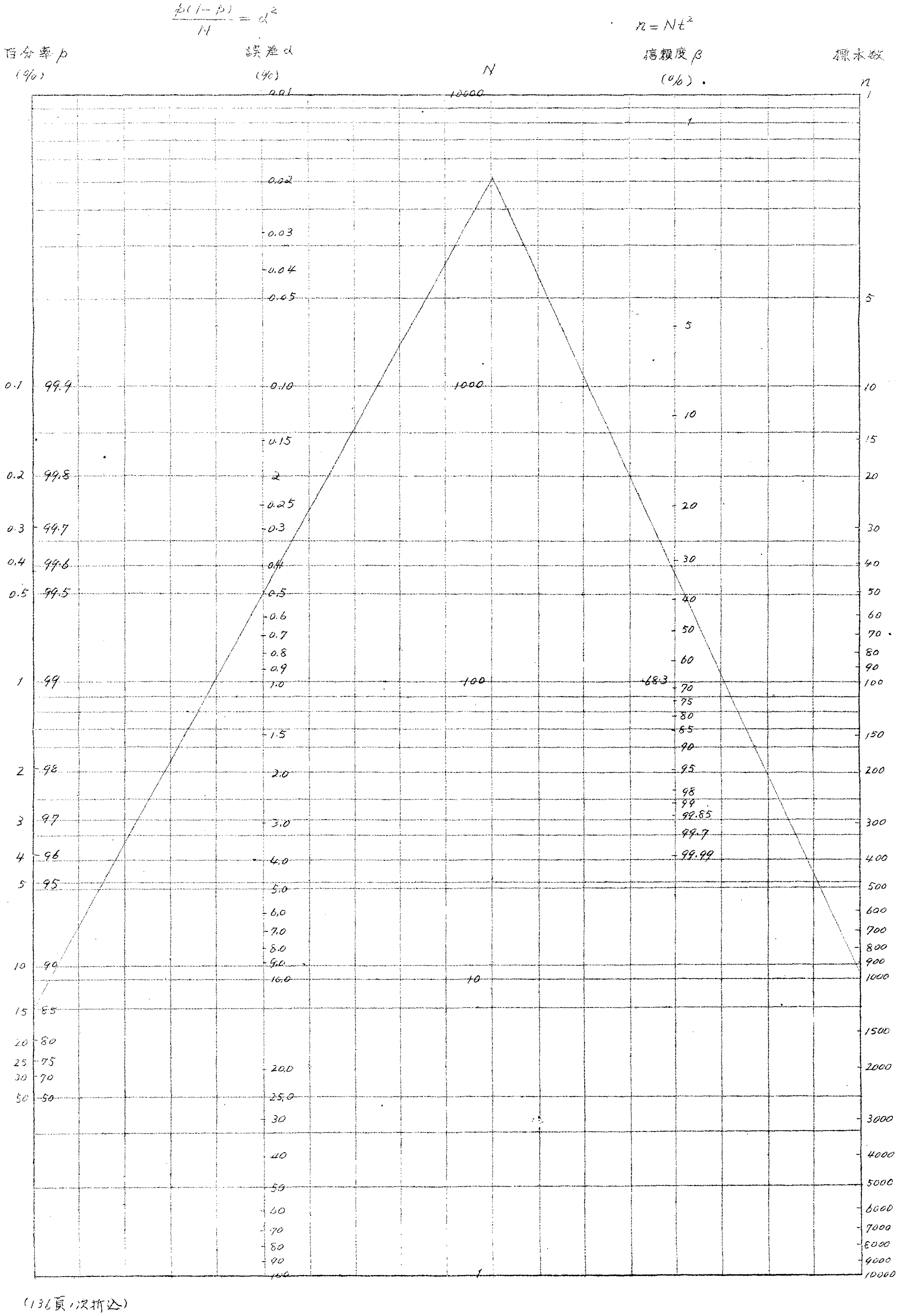
となることがわかる。 この結果を，系統的観測結果 r_s の分散

第 一 図

推定しようとする百分率 p (%) に対して，β % の確率で期待出来る誤差の幅が α (%) であるために必要な n を求めるノモグラム．

用法：(i) 左の三本によつて，p と α の値から N を求める．

(ii) 右の三本によつて，(i) で求めた N と，β の値から n を求める．



(136頁ノ次折込)

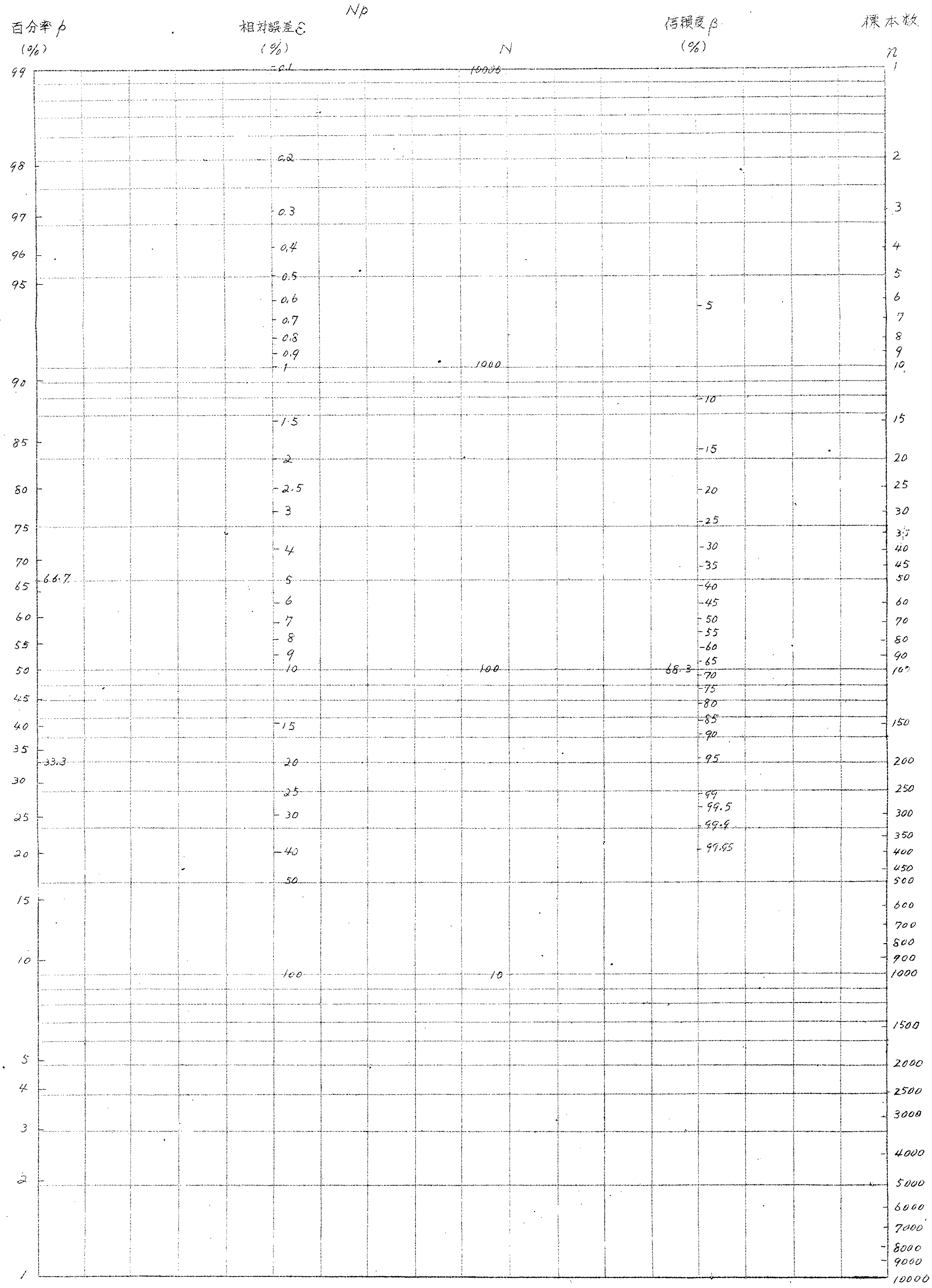
第　二　図

推定しようとする百分率 p(%)に対して，相対的誤差の大きさが真の価の ε % 以下であることが β %の確率で期待出来るために必要な n を求めるノモグラム．

用法：（i）左の三本を使って，p と ε の値から N を求める．

（ii）右の三本によって，（i）で求めた N と，β の値から n を求める．

$$\frac{1-p}{Np} = \varepsilon^2 \qquad n = Nt^2$$



136頁ノ次折込

と比較すると，

$$D^2(r_s) - D^2(r) = \frac{1}{n}\left[(n-1)E(r_s-R)^2 - E\left\{\sum_i p_i(\lambda_i - r_s)^2\right]\right.$$

となる。

従って，この結果から判断するまでもないけれども，系統的観測を行つてなお且つ精度を下げないようにするためには，各組毎の r_s が余り違わないようになつていなければならない。

これとは逆に，各組における $\sum p_i(\lambda_i - r_s)^2$ が大きいとき，換言すれば各組間の作業構成が非常に異るような場合には系統的観測を行うと可成り精度が悪くなる。

例えば，作業が週期的に循還していて，その週期が抽出間隔と一致するような場合には，系統的観測による誤差は重大なものとなる。

系統的観測によつて失われる精度は，適当な層別法と併せて調査するならば或る程度回復出来る筈である。