㉗ 実験的研究に於ける有効な計画

序

本稿は，A. Wald 教授授が講義したものを学生が編輯した"The efficient design of experimental investigation"についての Note である。

原文はタイプ刷りで，トリビアルな誤植が割合にあるため，それを補正しつゝ Note をとりました。

内容は，標本論と実験計画法が主で，特に乱化法についての詳細な解説は斯学の入門者のみならず一般数理統計家にとつても御参考になる所が多いと信じます。

最後に，昨年11月，不慮の飛行機事故で亡くなられた，Wald 先生の生前数理統計学の分野に跡された大きな功績に敬意を表し，謹んで哀悼の意を表します。

1951.2.8

A. Hashizume

㉗実驗的研究に於ける有効な計画

目　次

㉗実験的研究に於ける有効な計画

1. はしがき

統計的研究の有用性は，適当な実験計画をすることによつて増大する。最近15年間に，この様な理論は二つの方向に発展して来た。その一つは農業実験に関聯して起つたものに又他は，社会科学的，経済的実験に関聯して起つたものにである。

そしてその二つの線も，理論的根底は共通しており，その方法的技術も夫々他へ応用できるものである。

2. 簡単な例

$x=$ 高さ，$y=$ 重さとし，x と y は二変数正規分布をしていると仮定したとき N けの標本から，x についての y の母集団回帰係数を推定する問題を考へる。

x_α，y_α を夫々 x, y からとられた α 番目の観測値とす。($\alpha=1, \cdots\cdots, N$)　こゝで標本をとる手続きは，特性 y のみについて無作為で，x についてはしからずとする。

即ち，調査者は，被調査の高さにのみ着目して標本をとる。$a_0, a_1, \cdots\cdots, a_k$ ($a_0 < a_1 < \cdots\cdots < a_k$) をきめて a_{i-1}，a_i の間にある標本の数 N_i（高さについて）を任意の大きさ（但し $\sum N_i = N$）に与る様にきめる事が出来るとする。

x についての y の回帰係数の最小自乗推定値の分散は $\sum(x_\alpha-\overline{x})^2$ に逆比例する事は良く知られてゐる。

こゝで $\overline{x}=(x_1+\cdots\cdots+x_N)/N$ である。よつて統計的研究の有効性は $\sum(x_\alpha-\overline{x})^2$ が最大になる様にきめれば，最良であることは明らかである。

他の簡単な説明は次の例で示される。 一杯の茶わんの中に最初にお茶かミルクかどちらが先に加えられたかを風味することによつて見分ける事が出來ると主張する婦人がいるとする。 彼女の鑑別力を調べるため，或るものはお茶を或るものはミルクが夫々最初に加えられてある8つの茶わんを用意してこの婦人に鑑別させる。 この際実験者は何個の茶わんが，ミルクを最初に，そして残りの何個がお茶を最初に加えてあるかを，あらかじめ婦人にしらせ，彼女に正しい分類をさせるのである。

こゝでその個数をいくつに決めたら有効な実験であるかと云ふ問題が起る。それを論じてみよう。

先づ，虚無假設（検定されるべき假説）は鑑定能力がないと云ふ事である。 即ち各茶わんに対して婦人がそれを正確に分類できる確率は½であると云ふ假説を立てる。

対立假説は，彼女が鑑別能力を持つ事であり，8つの茶わんを正確に分類できる確率は殆んど1である。

明らかに実験の有効性は虚無假説が正しいと云ふ前提のもとに8つの茶わんの正確に分類する確率が小さければ小さい程よい。

即ち，この場合では4つの茶わんがお茶，4つの茶わんはミルクが最初に加えられてあるならば一番有効な実験が得られる。

即ち，（虚無假説が正しいとゆう前提のもとに）正確な分類をする確率は

$$1/\binom{8}{4} = 4!\,4!/8! = 1/70$$

である。

もし二つの茶わんに最初にお茶又はミルクが加えられ，残りの6つの茶わんにミルク又はお茶を加えた様な場合には，実験の有効性の目印となる所の確率は $1/C^8_2 = \frac{1}{28}$ で，この場合，実験によつて得られた結果が，あまり有効でなくなる。

更に8ヶの茶わんから1ヶだけをお茶又はミルクを最初に注ぎ残りをミルク又はお茶を最初に注いだなら，例の確率は

$$1/C^8_1 = \frac{1}{8}$$

となり，実験の結果は，殆んど無価値の様に思はれる。

3　一般的問題　　πを母集団とし，Qを標本平均によって推定しようとする。　この母集団に於ける未知の母数とする。

問題は二つの観点から述べられる。　1) 標本はすでに抜かれてあり，或る意味で母数Qの最良の推定値である様な観測値の，函数 $Q^x(x_1, \cdots\cdots, x_N)$ を求める事

2). 標本はまだ抜き出されてなく，母集団からとられるべき観測値の数を決め，またいかなる方法で標本をとるかを決める。

我々はまとして問題2を取扱ひ，問題1は問題2を理解するに必要な範囲で議論することにする。

§2 で與へられた，最初の例に於て $\Sigma(x_\alpha - \bar{X})^2$ を最大ならしめる有意的選擇は無作爲的選擇より良い事は良く知られてある。

4. 標　本　論　　1). 母集団πに於ける各元素が標本に含まれるべき機会が皆等しいときの標本抽出法を單純無作爲標本抽出と云ふ。　2) もしも標本の方法が，母集団πが有限個の層に分割され，各層からあらかじめ定められた数の無作爲な標本を抽出するとき，層別標本抽出法と云ふ。

3). 標本に含まれるべき元素が規範と稱する，或る方法に従って，選ばれるとき，有意選擇法と云ふ。

（この事に関しては Ezekiel. Method of Correlation annalysis p.p. 252 ~ 255. 1930 を参照せよ）

高さに応じて選んだ §2 に於ける例は有意選擇法の一例である。

我々は此等の三つの方法についての損得を論じて見よう。
この為に最良線型推定値についてのマルコフの定理を掲げる。

5. 最良線型推定値の定義

統計量 $t(x_1, \cdots\cdots, x_N)$ は，もしもその期望値が未知の母集団の母数 Q に等しいならば，不偏な推定値と云ふ。

観測値 $x_1, \cdots\cdots, x_n$ の一次函数 $t = \sum \lambda_i x_i$ が Q の不偏な推定値で，また Q の他 いかなる不偏一次推定値よりも分散が小さいとき，t を Q の最良不偏推定値であると云ふ。

例へば，もしも $x_1, \cdots\cdots, x_n$ を母集団 Π からの N けの独立な観測値とすれば，$\bar{x} = (x_1 + \cdots\cdots + x_N)/N$ は母集団平均 Q の最良線型推定値である。明らかに，$\bar{x}$ は Q の不偏推定値であり $\bar{x}$ の分散は他のいかなる Q の不偏一次推定値より小さい。σ^2 を母集団分散とすれば，$\bar{x}$ の分散は σ^2/N に等しい。一次結合 $t = \sum_{i=1}^{N} \lambda_i x_i$ の分散は明らかに $\sigma^2 (\sum \lambda_i^2)$ に等しい。我々は t が Q の不偏推定値と云う條件，即ち $\sum \lambda_i = 1$ とゆう前提のもとに $\sigma^2 \cdot (\sum \lambda_i^2)$ を最小ならしめる樣に λ_i をきめよう。ラグランジュの方法を用ひるならば，容易に，その最小値は，$\lambda_1 = \lambda_2 = \cdots\cdots = \lambda_N = \frac{1}{N}$ でなければならない事が分るであらう。かくて $\bar{x}$ は最小の分散を持ち，最良線型推定値である事が分る。

6. 一般的問題に於ける最良線型推定値

$x_1, \cdots\cdots, x_N$ を N けの独立な確率変数とする。
x_α の平均値，分散を夫々 μ_α, σ_α^2 とする。そして

$$(1) \quad \mu_\alpha = g_{1\alpha}\beta_1 + \cdots\cdots + g_{p\alpha}\beta_p \quad (\alpha = 1, \cdots\cdots, N)$$

と假定する。こゝで $\beta_1, \cdots\cdots, \beta_p$ は未知で $\| g_{i\alpha} \|$ は，あらかじめ與へられた行列であるとする。

$b_i = \lambda_{i1} x_1 + \cdots\cdots + \lambda_{iN} x_N$ を β_i の推定値とする。
しからば，

(2)　$E(b_i) = (\sum_{\alpha} \lambda_{i\alpha} g_{1\alpha})\beta_1 + \cdots\cdots + (\sum_{\alpha} \lambda_{i\alpha} g_{p\alpha})\beta_p$

(3)　$\sigma^2(b_i) = \sum_{\alpha} \lambda_{i\alpha}^2 \sigma_\alpha^2$

推定値 b_i が不偏であるためには，次の條件が満足されなければならない。

(4)　$\sum_{\alpha} \lambda_{i\alpha} g_{j\alpha} = \delta_{ij} \equiv$ クロネッカーの δ 記号

さて，條件（4）のもとに らぐらんじゆの方法で（3）を最小にならしめるならば

(5)　$\lambda_{i\alpha} \sigma_\alpha^2 = \sum_{j} k_{ij} g_{j\alpha}$　（k_{ij}：らぐらんじゆの乗数）

かくて，

$$\lambda_{i\alpha} = \frac{1}{\sigma_\alpha^2} \sum_{j} k_{ij} g_{j\alpha}$$

之を（4）に代入すると

(6)　$$\sum_{\alpha} \frac{(\sum_{r} k_{ir} g_{r\alpha}) g_{j\alpha}}{\sigma_\alpha^2} = \delta_{ij}$$

$\sum_{\alpha} g_{r\alpha} g_{j\alpha} / \sigma_\alpha^2 = a_{rj}$　とおけば，（6）から

(7)　$\sum_{r} k_{ir} a_{rj} = \delta_{ij}$

かくて

(8)　$K_{ir} = C_{ir}$　但し　$\| C_{ir} \| = \| a_{ir} \|^{-1}$

よつて（5）と（8）から

(9)　$$\lambda_{i\alpha} = \sum_{r} \frac{C_{ir} g_{r\alpha}}{\sigma_\alpha^2}$$

$$(10)\quad b_i = \sum_\alpha \sum_r c_{ir} \frac{g_{r\alpha} x_\alpha}{\sigma_\alpha^2}$$

が得られる。

$$I_r = \sum_\alpha \frac{g_{r\alpha} x_\alpha}{\sigma_\alpha^2}$$ とおけば

$$(11)\quad b_i = \sum_r c_{ir} I_r$$

さて（11）が最小自乗推定値と同じである事を示さう。

$\beta_1, \beta_2, \cdots\cdots, \beta_p$ の最小自乗推定値 $b_1^*, \cdots\cdots, b_p^*$ は

$$\sum_\alpha \frac{(x_\alpha - \beta_1 g_{1\alpha} - \cdots\cdots - \beta_p g_{p\alpha})^2}{\sigma_\alpha^2}$$

を最小ならしめる値である。 かくて最小自乗推定値 $b_1^*, \cdots\cdots, b_p^*$ は次の方程式を満足しなればならない。

$$(12)\quad a_{i1} b_1^* + \cdots\cdots + a_{ip} b_p^* = I_i \qquad (i = 1, \cdots\cdots, p)$$

こゝで、 $a_{ij} = \sum_\alpha \frac{g_{i\alpha} g_{j\alpha}}{\sigma_\alpha^2}$ 　 $I_i = \sum_\alpha \frac{X_\alpha g_{i\alpha}}{\sigma_\alpha^2}$

（12）の解は（11）式に於ける右辺と一致する。 a_{ij} と I は次数 -1 の $\sigma_1^2 \cdots\cdots \sigma_N^2$ について斉次函数で同様に c_{ij} は次数 1 の斉次函数である。 従って， b_i は次数 0 の斉次函数となる。

かくて b_i の値は，もしも，$\sigma_1^2, \cdots\cdots \sigma_N^2$ に同じ常数 λ^2 を掛けても不変である。

7. マルコフの定理

$k_1, \cdots\cdots, k_p$ を與へられた常数とし，$z = k_1\beta_1 + \cdots\cdots + k_p\beta_p$ とする。

もしも b_i が β_i の最良線型推定値とすれば，

$t = k_1 b_1 + \cdots\cdots + k_p b_p$ は τ の最良線型推定値である。

證　明： $\| k_{ij} \|$ $(i, j = 1, \cdots\cdots, p)$ を前章に於て $g_{1\alpha}, \cdots\cdots, g_{p\alpha}$ $(\alpha = 1, \cdots\cdots, n)$ を與へたとき、それから導かれる。或る定まった，特異でない対稱な行列とする。今

$$\tau_i = \sum_j k_{ij} \beta_j \qquad (i = 1, \cdots\cdots, p)$$

とおく。しからば

$$E x_\alpha = \mu_\alpha = g'_{1\alpha} \tau_1 + \cdots\cdots + g'_{p\alpha} \tau_p \quad (\alpha = 1, \cdots, n)$$

と成る。但し $g'_{i\alpha}$ は $g_{i\alpha}$ の k_{ij} を係数とする一次線合式である。前章の結果から，$\tau_1, \cdots\cdots, \tau_p$ の最良線型推定値は，最小自乗推定値 $t_1 \cdots\cdots t_p$ に一致する。

このとき $t_i = \sum_j k_{ij} b_j$ である事が分る。

こゝで勿論 $b_1, \cdots\cdots, b_p$ は夫々 $\beta_1, \cdots\cdots, \beta_p$ の最良線型推定値である。今例へば $k_1, \cdots\cdots, k_p$ を夫々 $k_1, \cdots\cdots, k_p$ に等しくする様に $(g_{1\alpha}, \cdots\cdots, g_{p\alpha})$ を適当にきめれば，我々の定理は證明された事になる。

8.　有限母集団からとられた二つの共分散

M を母集団に於ける元素の總数とし，それらの元素の或る標識値を $\alpha_1, \cdots\cdots, \alpha_M$ とする。

しからば母集団平均は $\mu = (\alpha_1 + \cdots\cdots + \alpha_M)/M$ で，又母集団分散は $\sigma^2 = (\alpha_1^2 + \cdots\cdots + \alpha_M^2 - M\mu^2)/M$ である。

X_1 と X_2 をこの母集団から　れた二つの確率変数とする。しからば $E(x_i) = \mu$, $\sigma_{x_i}^2 = \sigma^2$ $(i = 1, 2)$ である。最初に引出されたものが i 元素で二番目のものが j 元素 $(i \neq j)$ である確率は $1/M(M-1)$ である。よって

$$(13) \qquad E X_1 X_2 = \frac{1}{M(M-1)} \sum_{i \neq j} \sum \alpha_i \alpha_j$$

かくて

$$(13') \quad \sigma_{x_1x_2} = \frac{1}{M(M-1)} \sum_{i \neq j}\sum \alpha_i \alpha_j - EX_1 EX_2$$

$$= \frac{(\alpha_1 + \cdots\cdots + \alpha_M)^2 - (\alpha_1^2 + \cdots\cdots + \alpha_M^2)}{M(M-1)} - \mu^2$$

$$= \frac{M^2\mu^2 - M\sigma^2 - M\mu^2}{M(M-1)} - \mu^2$$

$$= \frac{M\mu^2 - \sigma^2 - \mu^2}{M-1} - \mu^2 = -\frac{\sigma^2}{M-1}$$

9. マルコフの定理の層別標本論への応用

母集団Πは層$\Pi_1, \cdots\cdots \Pi_k$に分割されてあるとする。$\mu_i, \sigma_i^2$を夫々$\Pi_i$の母平均及び母分散とする。そして$\Pi_i$から大さ$m_i$の無作為標本を抽出する，$x_{ij}$を$i$一層からとられた$j$一番目の観測値とする。$M_i$を$\Pi_i$の元素の数とするとき$\beta = \sum_{i=1}^{k} M_i \mu_i$の最良線型推定値を求める問題を取扱はう。

パラメーター$\mu_1, \cdots\cdots, \mu_k, \sigma_1, \cdots\cdots, \sigma_k$は未知で$M_1, \cdots M_k$は既知であるとする。線型推定値

$$b = \sum_{i=1}^{k} \sum_{j=1}^{m_i} \lambda_{ij} x_{ij}$$

を考へる。推定値bはもしも

$$E(b) = \sum_{i=1}^{k} \sum_{j=1}^{m_i} \lambda_{ij} \mu_i = M_1\mu_1 + \cdots\cdots + M_k \mu_k$$

が，$\mu_1, \mu_2, \cdots\cdots \mu_n$に関して恒等的に成り立つならば不偏である。よつて

$$(16) \quad \sum_{j=1}^{m_i} \lambda_{ij} = M_i \qquad (i = 1, \cdots\cdots, k)$$

故に我々は

$$(17)\quad \sigma_b^2=\sum_{i=1}^{k}\sigma^2\Big(\sum_{j=1}^{m_i}\lambda_{ij}x_{ij}\Big)=\sum_{i=1}^{k}\Big\{\sigma_i^2\Big[\sum_{j=1}^{m_i}\lambda_{ij}^2-2\frac{1}{M_i-1}\sum_{j<j'}\lambda_{ij}\lambda_{ij'}\Big]\Big\}$$

$$=\sum_{i=1}^{k}\left\{\frac{\sigma_i^2\Big[M_i\sum_{j=1}^{m_i}\lambda_{ij}^2-\Big(\sum_{j=1}^{m_i}\lambda_{ij}\Big)^2\Big]}{M_i-1}\right\}$$

之に (16) を使えば

$$(17')\quad \sigma_b^2=\sum_i\sigma_i^2\,\frac{M_i\sum_{j=1}^{m_i}\lambda_{ij}^2-M_i^2}{M_i-1}$$

が出る。各々定められた i について，$\sum_{j=1}^{m_i}\lambda_{ij}^2$ を條件(16)のもとに最小ならしめれば，分散 σ_b^2 は最小となる。ラグランジュの方法を用いれば，我々は容易に最小値は

$$(18)\quad \lambda_{ij}=\frac{M_i}{m_i}$$

とおく事によって得られる。よって最良線型推定値は

$$(19)\quad b_i=\sum_{i=1}^{k}\sum_{j=1}^{m_i}\frac{M_i}{m_i}X_{ij}=\sum_{i=1}^{k}M_i\bar{x}_i$$

$$但し\quad \bar{x}_i=\Big(\sum_{j=1}^{m_i}X_{ij}\Big)\Big/m_i$$

となる事が分る。(17')と(18)から

$$(20)\quad \sigma_b^2=\sum_i\sigma_i^2\,\frac{\frac{M_i^2}{m_i}-M_i^2}{M_i-1}=\sum_i\sigma_i^2M_i\,\frac{M_i-m_i}{m_i}\,\frac{M_i}{M_i-1}$$

が得られる。

10. m_iの最良の選び方

$M=\sum M_i$, $m_i=\sum m_i$

$s_i^2=M_i\sigma_i^2/(M_i-1)$ とおく，しからば

$$(21)\quad \frac{M-m}{m}\sum_i M_i S_i^2+\sum_i m_i\left[\frac{M_i S_i}{m_i}-\frac{\sum_i M_i S_i}{m}\right]^2$$

$$-\frac{M}{m}\sum_i\left\{M_i\left[s_i-\frac{\sum_i M_i S_i}{M}\right]^2\right\}=\frac{M-m}{m}\sum_i(M_i s_i^2)$$

$$+\left\{\left(\sum_i\frac{M_i^2 S_i^2}{m_i}\right)-\frac{(\sum_i M_i S_i)^2}{m}\right\}-\frac{M}{m}\left[\left(\sum_i M_i S_i^2\right)-\frac{(\sum_i M_i S_i)^2}{M}\right]$$

$$=\sum_i s_i^2\left[M_i\frac{M-m}{m}+\frac{M_i^2}{m_i}-\frac{M_i M}{m}\right]=\sum s_i^2\left[-M_i+\frac{M_i^2}{m_i}\right]$$

$$=\sum_i \sigma_i^2\frac{M_i}{M_i-1}M_i\frac{M_i-m_i}{m_i}=\sum_i\sigma_i^2 M_i\frac{M_i-m_i}{m_i}\frac{M_i}{M_i-1}=\sigma_i^2$$

が成立つ。　(21) の左辺の項のみは m_i に関係するから，σ_i^2 を最小にするには，この中の項のみを $\sum m_i=m$ なる條件のもとに最小にすればよい。　またこの項は

$$\sum_i\left(\frac{M_i^2 S_i^2}{m_i}\right)-\frac{(\sum M_i S_i)^2}{m}$$

に等しいから　$\sum\frac{M_i^2 S_i^2}{m_i}$　を最小ならしめればよい。
しからばその最小値は，

$$(22)\quad m_i=\frac{m M_i S_i}{\sum_i M_i S_i}\sim\frac{m M_i S_i}{\sum m\sigma_i}$$

とおく事によつて得られる事がすぐ分る。
さて次に (21) の二番目の式の第二項と第三項に着目すれば

$$(23)\quad \sigma_b^2=\frac{M-m}{m}\sum M_i S_i^2+\sum\left(\frac{M_i}{m_i}-\frac{M}{m}\right)M_i S_i^2$$

なる事がたゞちに分る。　もしも $\sigma_1, \cdots\cdots, \sigma_k$ が未知のとき恐らく m_i を M_i に比例させたなら最良であらう。

もし m_i を M_i に比例させるなら（23）から勿論

$$\sigma_l^2 = \frac{M-m}{m} \sum M_i S_i^2$$

が得られる。

11.　単純無作爲標本論との比較

M けの元素からなる母集団 Π を考へる。

j 層に於ける i- 元素の標識値を α_{ij} とする。 M_i を i 層の元素の数とすれば勿論 $\sum_{i=1}^{k} M_i = M$ である。　単純無作爲標本論の場合，即ち，各元素が同じ確率で抜きられる場合 Π の母集団分散は

$$(24) \quad \sigma^2 = \left(\sum_{j=1}^{M_i} \sum_{i=1}^{k} \alpha_{ij}^2 - M\alpha^2 \right) / M$$

但し

$$\alpha = \left(\sum_i \sum_j \alpha_{ij} \right) / M$$

で與へられ i 層の分散は

$$(25) \quad \sigma_i^2 = \left(\sum_{j=1}^{M_i} \alpha_{ij}^2 - M_i \alpha_i^2 \right) / M_i$$

で與へられる。

（24）と（25）から

$$(26) \quad \sigma^2 = \frac{\sum_i M_i \sigma_i^2 + \sum_i M_i \alpha_i^2 - M\alpha^2}{M} = \frac{\sum_i M_i \sigma_i^2 + \sum_i M_i (\alpha_i - \alpha)^2}{M}$$

なる関係が得られる。

もしも我々は層別確率比例抽出法を採用するならば，

$\beta = M_1\alpha_1 + \cdots\cdots + M_k\alpha_k$ の最良線型推定値は（23'）によつて

與へられる分散を持つ，即ち

$$(27)\quad \sigma_b^2=\frac{M-m}{m}\sum M_i S_i^2 \sim \frac{M-m}{m}\sum_{i=1}^{k} M_i\sigma_i^2 \quad (\sum m_i = m)$$

もしも我々が單純無作爲比例抽出法によつて m けの標本を抽きとるならば，即ち $k=1$ のとき $\beta=M_1\alpha_1+\cdots\cdots+M_k\alpha_k=M\alpha$ の最良線型推定値 b^* は $b^*=M\bar{x}$ によつて與へられる。
但し $\bar{x}$ はすべての m けの觀測値の算術平均である。

b^* の分散は(27)の右辺に於て $k=1$, $M_1=M$, $S_1^2\frac{M}{M-1}\sigma^2$ とおく事によつて得られる。　かくて

(26)によつて $M\sigma^2=\sum M_i\sigma_i^2+\sum M_i(\alpha_i-\alpha)^2$ であるから

$$(28)\quad \sigma_{b^*}^2=\frac{M-m}{m}\frac{M^2}{M-1}\sigma^2\sim\frac{M-m}{m}M\sigma^2$$

$$=\frac{M-m}{m}\left[\sum_i M_i\sigma_i^2+\sum_i M_i(\alpha_i-\alpha)^2\right]$$

(27)と(28)から我々は $\sigma_{b^*}^2-\sigma_b^2>0$ で，その差は $\sum_i M_i(\alpha_i-\alpha)^2$ である事を知る。

12.　層ごとに調査費用が異なる場合の m_i の最良の選び方

(20)によつて我々は

$$(29)\quad \sigma_b^2=\sum_i \sigma_i^2 M_i\frac{M_i-m_i}{m_i}\frac{M_i}{M_i-1}$$

である事を知つた。　今調査に要する全費用があらかじめ與へられてあり，例へばそれを A とする。　i 層で一標本あたり調査に要する費用を g_i とし，又そこからとられる標本の大さを m_i とすれば，次の條件

$$(30)\quad \sum m_i g_i = A$$

が成立つ。　かくてこの條件のもとに(29)の右辺を最小ならし

める様に，ラグランヂュの方法で m_i を決定すれば

$$\frac{\partial \sigma_b^2}{\partial m_i}+\lambda g_i=-\sum_i \sigma_i^2\frac{M_i^2}{m_i^2}\frac{M_i}{M_i-1}+\lambda g_i=0$$

かくて

$$m_i^2=\frac{\sigma_i^2 M_i^3}{\lambda g_i(M_i-1)}$$ 即ち m_i は

$$\frac{\sigma_i M^{3/2}}{\sqrt{g_i(M_i-1)}}\sim\frac{\sigma_i M_i}{\sqrt{g_i}}$$

に比例する事が分る。

13. 母集団 Π の元素が単一な元素の代りに元素の集りである場合の標本論

人口母集団は通常群をなしてゐる。 我々が標本調査をしようとする母集団は，通常各人が離れ離れに住んでいる母集団ではない事が多い。

まとまつた各人についてある資料が記せられてある，カード又は帳面のあつまりである事が多い。

それらの資料は通常或る方法で一まとめにしてある。 例へば一般せんさすの資料は，各人の住所に従つてまとめてある。
即ち同じアパートに住むすべての人は一枚の用紙に書きならべてある。 かゝる情況のもとでは，実際問題として各個人個人を無作為に引き抜く事は困難である。

かくて我々は母集団が単一な元素の集りの代りに部分集団の集りとして取扱ほふ方が便利である。 この場合の標本論は今迄述べて来た標本論と全く同じである。

母集団 Π が M けの群（元素）から成つてゐるとしそしてそれが k 個の層に別れてゐるとする。

M_iを i 層に於ける群の数とし，α_{ij} ($i=1,\cdots,k,\ j=1,\cdots,M_i$) を i 層 j 群に属する各人の標識 x の和であるとする。

我々は k 個の層から抽き出した無作為標本をもとゝして $\beta=\sum_{i=1}^{k}\sum_{j=1}^{M_i}\alpha_{ij}$ を推定しようとする事を考へる。

i 層から無作為にとった群の数を m_i とし，i 層 j 群に属する人の標識の和を x_{ij} ($i=1,\cdots,k,\ j=1,\cdots,m_i$) とする。しからば β の最良線型推定値は

$$b=\sum_{i=1}^{k}\sum_{j=1}^{m_i}\frac{M_i}{m_i}x_{ij}=\sum_{i=1}^{k}M_i\bar{x}_i$$

$$\text{但し}\quad \bar{x}_i=\left(\sum_{j=1}^{m_i}x_{ij}\right)\Big/m_i$$

σ_b^2 に対する公式も又，m_i の最良の選び方に対するものも母集団の元素が単一元素である場合と同じである。

さて我々は群の数 m_i の大きさが b の分散にどんな影響を及ぼすかについて一寸述べて見よう。

もし m_iを M_iに比例してとるならば，分散 σ_b^2 は (23′) によつて與へられる。　即ち

$$\sigma_b^2=\frac{M-m}{m}\sum M_iS_i^2\sim\frac{M-m}{m}\sum_i M_i\sigma_i^2$$

である。

14. yについての x の回帰が直線である様な管理標識 y が存在する場合

xを標識値とし，その平均値は母集団 Π から推定し得るとする。次に y を管理標識と呼ばれるもう一つの標識値とし，それについては我々は y についての x の回帰線が線型であると知ってゐるものとする。

今 y が k この値 $y_1, \cdots, y_k$ をとり得るとする。
Π を k この層 $\Pi_1, \cdots, \Pi_k$ に分割し，且つ Π_i は y_i を含むようにする。　次に α_{ij} を i 層に於ける j-元素とし，

$\alpha_i = \left(\sum_{j=1}^{M_i} \alpha_{ij}\right)/M_i$ とする。　こゝで M_i は i 層に於ける元素の総数である。　y についての x の回帰線が線型であると云ふ事より

$$(31) \qquad \alpha_i = A + B y_i \qquad\qquad (i = 1, \cdots, k)$$

が成り立つ。　但し A と B は或る常数である。
M_i と y_i $(i = 1, \cdots, k)$ は既知であると仮定し，我々は

$$\alpha = \frac{1}{M} \sum_{i=1}^{k} \sum_{j=1}^{M_i} \alpha_{ij} = \frac{1}{M} \sum_{i=1}^{k} M_i \alpha_i$$

$$\text{但し} \quad M = \sum_{i=1}^{k} M_i$$

を推定する事を問題にする。
　さて一般性を失ふ事無しに $\sum M_i y_i = 0$ と仮定して差支ない。　実際，C を任意の常数とするとき，新しい変数 $y^* = y + C$ を導入すれば，y^* についての x の回帰線を線型であるから C を適当に選んで $\sum M_i y_i^* = 0$ とすることが出来るからである。　(31) から

$$\alpha = \frac{1}{M} \sum_i \sum_j \alpha_{ij} = \frac{1}{M} \sum M_i \alpha_i = \frac{1}{M} \sum M_i (A + B y_i)$$
$$= A + B \frac{\sum M_i y_i}{M}$$

$\sum M_i y_i = 0$ であるから

$$\alpha = A$$

となる。　かくて我々はパラメーター A を推定しなければならない。

15. A の最良線型推定値

我々は i 層 ($i=1,2,\cdots\cdots,k$) から大さ m_i の無作為標本を抽き出し， x_{ij} をこの標本の j 番目の値とする。
しからば A の線型推定値は

$$(32)\quad t = \sum_{i=1}^{k}\sum_{j=1}^{m_i}\lambda_{ij}x_{ij}$$

によつて與へられる。

$$(33)\quad Et = \sum_i\sum_j\lambda_{ij}\alpha_i = \Sigma\Sigma\lambda_{ij}(A+By_i)$$

$$= A\sum_i\sum_j\lambda_{ij} + B\Sigma\Sigma\lambda_{ij}y_i$$

となり之が不偏推定値であるためには A に一致しなければならない。　よつて

$$(34)\quad \sum_i\sum_j\lambda_{ij} = 1$$

$$(35)\quad \sum_i\sum_j\lambda_{ij}y_i = 0$$

が成り立つ。　i 層に於ける x の母集団分散を σ_i^2 とすれば
(13′) 式を用い， t の分散は

$$(36)\quad \sigma_t^2 = \sum_i\sigma_i^2\left[\sum_{i=1}^{m_i}\lambda_{ij}^2 - \frac{2\sum_{j<k}\lambda_{ij}\lambda_{ik}}{M_i-1}\right]$$

で與へられる。　條件 (34), (35) のもとに (36) を最小ならしめる様にきめよう。　例によつてラグランジュの方法を用ひれば

$$(37)\quad 2\lambda_{ij} - (2\sum_{k\neq j}\lambda_{ik})/M_i-1 - \alpha - \beta y_i = 0$$

(α と β はラグランジュの乗数)

が得られこれは次の如く書き直せる。

$$(38)\quad \left(2+\frac{2}{M_i-1}\right)\lambda_{ij}-\left(2\sum_{k=1}^{m_i}\lambda_{ik}\right)\Big/(M_i-1)-\alpha-\beta y_i=0$$

よつて

$$(39)\quad \lambda_{i1}=\lambda_{i2}=----=\lambda_{im_i}\equiv\lambda_i \qquad (i=1,----,k)$$

となる。　よつて

$$t=\Sigma\Sigma\lambda_{ij}x_{ij}=\Sigma\Sigma\lambda_i x_{ij}=\sum_i\lambda_i^*\bar{x}_i$$

ただし　$\bar{x}_i=(\sum_j x_{ij})/m_i \qquad \lambda_i^*=m_i\lambda_i$

又（13′）から容易に，

$$\sigma_{\bar{x}_i}^2=\sigma_i^2\left\{\frac{1}{m_i}-\frac{m_i(m_i-1)}{(M_i-1)m_i}\right\}=\sigma_i^2\left[\frac{M_i-m_i}{m_i(M_i-1)}\right]$$

が得られる。

$$(41)\qquad W_i=\frac{m_i(M_i-1)}{\sigma_i^2(M_i-m_i)}$$

とおく，　しからば　$\bar{x}_1$ ----，$\bar{x}_k$　は独立で $\bar{x}_i$ の分散は $\frac{1}{W_i}$ であるからマルコフの定理により $\bar{x}_1$,---- $\bar{x}_k$ の一次式の推定値の内 A と B の最良推定値は

$$(42)\qquad \Sigma(\bar{x}_i-A-By_i)^2W_i$$

が最小ならしめる如くきめればよい。　もし t と v が夫々 AとBの最良線型推定値とするなら，即ち t と v が（42）を最小ならしめる A と B の値なら，次の二つの正規方程式を満足しなければならない。

$$(43)\quad \begin{cases} t\Sigma z_i^{*2}+v\Sigma y_i^*z_i^*=\Sigma x_i^*z_i^* \\ t\Sigma z_i^*y_i^*+v\Sigma_i y_i^{*2}=\Sigma x_i^*y_i^* \end{cases}$$

但し　$x_i^* = \bar{x}_i\sqrt{w_i}$　　$y_i^* = y_i\sqrt{w_i}$　　$z_i^* = \sqrt{w_i}$

かくて x_i^* の分散は 1 に等しいから，最小自乗法から良く知られる如く，t の分散は C_{11} に等しい。但し

$$\begin{Vmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{Vmatrix} = \begin{Vmatrix} \Sigma z_i^{*2} & \Sigma y_i^* z_i^* \\ \Sigma y_i^* z_i^* & \Sigma y_i^{*2} \end{Vmatrix}^{-1}$$

よつて

$$(44)\quad \sigma_t^2 = C_{11} = \frac{\Sigma y_i^{*2}}{(\sum_i z_i^{*2})(\sum_i y_i^{*2}) - (\Sigma y_i^* z_i^*)^2}$$

$$= \frac{\Sigma W_i y_i^2}{(\Sigma W_i)(\Sigma W_i y_i^2) - (\Sigma W_i y_i)^2}$$

16. m_i の最良の選び方

我々は $\sigma_1^2 = \sigma_2^2 = \cdots\cdots = \sigma_k^2 = \sigma^2$ なる場合を考へる。m_i は M_i に較べて非常に小さいと考へられるから，$\frac{M_i - 1}{(M_i - m_i)} \sim 1$ として差支へ無い。

しからば（41）から

$$W_i \sim \frac{m_i}{\sigma}, \quad \sigma_t^2 \sim \sigma^2 \frac{\Sigma m_i y_i^2}{m\Sigma m_i y_i^2 - (\Sigma m_i y_i)^2}$$

$$= \sigma^2 \frac{1}{m - \dfrac{(\Sigma m_i y_i)^2}{\Sigma m_i y_i^2}}$$

左辺は明らかに $\Sigma m_i y_i = 0$ のとき最小となる。

かくて，標本の大さ，$m_1, \cdots\cdots, m_k$ を $\Sigma m_i y_i = 0$ なる

如く定めれば t の最小の分散が得られる。

さて，これらを導くに際して我々は $\bar{y}=(\sum M_i y_i)/M=0$ である事を假定してゐた。もし $\bar{y}\neq 0$ なら t の分散は容易に分る様に $\sum m_i(y_i-\bar{y})=0$ なるとき最小となる。

即ち，

$$\sum m_i y_i = m\bar{y} = m\ \frac{\sum M_i\ y_i}{M}$$

が満足されるならば上の等式を m で割れば，

條件

$$(46)\qquad \frac{\sum m_i y_i}{m} = \frac{\sum M_i\ y_i}{M}$$

即ち，管理標識の標本平均についての條件が得られる。

この方法は伊太利の統計学者 C. Gini と L. Galvani によつて與へられたものである。

この方法を proposive selection 法と呼んでゐる。

條件（46）は $\sigma_1=\text{-------}=\sigma_k$ であるときにのみ σ_t^2 を最小ならしめる事に注意せよ。

この理論は又母集団の元素が單一元素の代りに群である場合にも拡張出来る。

17. 二重標本抽出法（The method of double sampling）

問題の構成

例へば家族が食料に費す金の総高 M を推定するとゆうやうに，母集団の各元素の標識 x の平均値を調べる問題を考える。

このやうな場合，それを調べるための資料を集める事は良く訓練された調査を長期に必要とし，相当金がかゝるであらう。調査に要する費用があらかじめ與へられた予算で抑えられその費用の

もとで調査を実施するならば，非常に小さな標本数しかとれず，望ましい精度の推定が得られない。

もし標識 x が他の標識 y （例えば収入）に高い相関をもっているとし，y は又 x に較べて少い費用で調査が出来るとする。このやうな場合，次の手続きで調査を行ふならば最初の標識のもっと正確な推定が得られるであらう。

先づ第二の標識 y を母集団から，十分良い精度の推定を得られる程度の大きさの標本を抜き出す．

かくて，第二の標識値に従つて，いくつかの層を作り各層に対して第一の特性に関係する資料を得られる様抜き出す。

この方法で我々は第二の標識を使はないで行ふ調査よりも，同じ調査費用でよりよい精度の推定が得られるであらう。

問題は与へられた費出のもとで最初の標本数と最初の標識の最も正確な推定を生ぜしめる部分標本をいかに決定するかである．

y の変動の範囲は既知であるとし，それを s けの区間に分割する。例えばそれを $(y_0\ y_1)(y_1\ y_2)$ ------ $(y_{s-1}\ y_s)$ とする。標識 y について大さ N の標本 S_1 を抜き出す。

この標本は s けの層に分け，その i 層は (y_{i-1}, y_i) にあるすべての観測値からなる様にする。標本 S_1 の i 層 $(i=1$ ------ $,s)$ から，我々は標識 x について，大さ m_i の部分標本 S_{2i} をとり出す。さて Π_i を $y_{i-1}<y<y_i$ なる母集団のすべての元素の集合．Π を全集合とし，Π_i を Π の i 層と名付ける事にする．p_i を Π_i に属する Π の元素の数との比とし，更に $\bar{x}$, x_i を夫々全母集団 Π, Π_i に於ける X についての平均値とする。

我々は標本 S_1，S_{2i} をもとにして $\bar{x}=\sum_{i=1}^{s} p_i x_i$ を推定しようとする事を試みる。標本 S_1 から比 p_1 ---- p_s を推定する事が出来，又標本 S_{2i} から x_i を推定する事が出来る．

$m_0=\sum_{i=1}^{s} m_i$ とし，A，Bを夫々標識 x, y についての

一元素についての調査費用とすれば全費用は

$$(47)\qquad Am_0 + BN = C$$

で與へられる。 今 C は固定されてゐるとし條件 (47) のもとに, m_1 ……, m_s ; N のもつともよい値を探す問題を取扱ふのがこの目的である.

18. $\bar{x} = \sum_{i=1}^{s} p_i x_i$ の最適推定値

n_i を i 層に属する S_1 の元素の数とすれば p_i の最良推定値は勿論 $r_i = n_i/N$ である。

x_{ij} を標本 S_{2i} に於ける j 元素の x の値とする。

我々は x の推定値が

$$(48)\qquad F_1 = \sum_{i=1}^{s}\sum_{j=1}^{s}\sum_{k=1}^{m_i} \lambda_{ijk}\, r_i\, x_{jk}$$

なる形をしていると予め制限しておく。 こゝで λ_{ijk} は常数である。 推定値 (48) が次の二つの条件を滿足するとき最良であると云ふ。

(i) $E(F_1) = \bar{x}$

(ii) F_1 の分散が性質 (i) をもつ他のいかなる (48) の形の推定値のそれり小さい.

かくて我々は條件 (i) (ii) が滿される様に, 係数 λ_{ijk} を決めよう。 $E(F_1)$ を計算するために母集団に於ける各層 Π_1 --- Π_s の元素の総数は標本に於ける各層のそれに較べて非常に大きいと假定する。 従つて, 標本の一こ一このとり方は互ひに独立であると假定してしまう.

S_{2i} は標本 S_1 の i 層からとられ Π_i から直接ではないけれども, 變数 r_i と x_{jk} は独立である.

これは次の様にして知る事が出來る。

全母集団 Π から無作為標本 S_1 を抜く代りに各層 Π_i から大さ n_i の無作為標本を抜く。 たゞ $n_i, \cdots\cdots, n_s$ は與へられた数で $\sum n_i = N$ であると考へられる。*

この標本は x の値に関係なく抜かれたものであるから x_{jk} の分布は，数 $n_1, \cdots\cdots, n_s$ に無関係である.

かくて

> * 註. 之は誤りである。n_i は確率変数であるから $n_i = 0$ なる値をとり得る。 よつて，後で m_i を決める時 $n_i \geq m_i$ でなければならないから，m_i を決定出來なくなる場合がある。
>
> もし，この様な事が起る場合，標本抜取りをやりなほせば偏倚な推定値となる。
>
> 統計数理研究所輯報 No. 2 林知己夫，石田正次，白河市言語調査に於けるサムプリング調査計画を参照されたい。

固定された $r_1, \cdots\cdots, r_s$ についての x_{jk} の條件附確率は $r_1, \cdots\cdots, r_s$ に関係しない。 その結果，r_i と x_{jk} は独立に分布する。 從つて

$$E(F_1) = \sum_{i=1}^{s} \sum_{j=1}^{s} \sum_{k=1}^{m_i} \lambda_{ijk} E(r_i) E(x_{jk})$$

$$= \sum\sum\sum \lambda_{ijk} p_i x_j$$

故に條件 (i) によつて

$$\sum_{i=1}^{s} \sum_{j=1}^{s} \sum_{k=1}^{m_i} \lambda_{ijk} p_i x_j = \bar{x} = p_1 x_1 + \cdots\cdots + p_s x_s$$

かくて

$$(49)\quad \begin{cases} \sum_{k=1}^{m_i} \lambda_{ijk} = 0 & i \neq j \\ \sum_{k=1}^{m_i} \lambda_{iik} = \dfrac{1}{m_i} \end{cases}$$

が得られる。

さて次に F_1 の分散 V_1 を求めよう。先づ r_1, ----- , r_s が固定された常数であると云ふ假定のもとの F_1 の分散 V_1^* を求めてみる。σ_i^2 を π_i に於ける x の分散とすると

$$(50)\qquad V_1^* = \sum\sum\sum (\lambda_{ijk})^2 r_i^2 \sigma_j^2$$

が得られる。任意の確率変数 z に対し r_1 ----- , r_s が固定された時の z の條件附期望値を $E(z / r_1 \text{----} r_s)$ とすると明らかに次式が成立つ

$$(51)\qquad E(F_1^2 / r_1 \cdots\cdots r_s) = V_1^* + \{E(F_1 / r_1 \cdots r_s)\}^2$$

$$= \sum\sum\sum \lambda_{ijk}^2 r_i^2 \sigma_j^2 + \{\sum\sum\sum \lambda_{ijk} r_i x_j\}^2$$

(49) から

$$\sum\sum\sum \lambda_{ijk} r_i x_j = \sum_{i=1}^{s} r_i x_i$$

である事が分るから

$$(52)\qquad E(F_1^2 / r_1, \cdots\cdots r_s) = \sum\sum\sum \lambda_{ijk}^2 r_i^2 \sigma_j^2 + \left(\sum_{i=1}^{s} r_i x_i\right)^2$$

F_1^2 の條件の附かない期望値 $E(F_1^2)$ は r_1 ----, r_s を確率変数であるとしたときの (52) の右辺の期望値に等しい。

$$E(r_i r_h) = p_i p_h \left(1 - \frac{1}{N}\right) \qquad (i \neq h)$$

$$E(r_i^2) = p_i^2 + p_i(1-p_i)/N$$

であるから

$$(53)\quad E(F_1^2) = \Sigma\Sigma\Sigma\, \lambda_{ijk}^2 \left[p_i^2 + p_i(1-p_i)/N \right] \sigma_j^2$$

$$+ \sum_i \left[p_i^2 + p_i(1-p_i)/N \right] x_i^2 + \sum_{i \neq k}\sum p_i p_k x_i x_k \left(1 - \frac{1}{N}\right)$$

よつて

$$(54)\quad V_1 = E(F_1^2) - (p_1 x_1 + \cdots + p_s x_s)^2$$

$$= \Sigma\Sigma\Sigma\, \lambda_{ijk}^2 \left[p_i^2 + p_i(1-p_i)/N \right] \sigma_j^2$$

$$+ \frac{1}{N} \sum_i p_i(1-p_i) x_i^2 - \frac{1}{N} \sum_{i \neq k}\sum p_i p_k x_i x_k$$

(54)から条件(49)のもとに λ_{ijk} についての V の最小値は，もしも

$$(55)\quad \begin{cases} \lambda_{ijk} = 0 & i \neq j \\ \lambda_{jjk} = \dfrac{1}{m_j} \end{cases}$$

なるとき成立つ事が分る。これらの値を(54)に代入すれば

$$(56)\quad V_1 = \sum_{i=1}^{s} \left\{ \left[p_i^2 + \frac{p_i(1-p_i)}{N} \right] \frac{\sigma_i^2}{m_i} + \frac{p_i(1-p_i)}{N} x_i \right\}$$

$$- \overline{N}^{2} \sum_{i=1}^{s-1} \sum_{j=i+1}^{s} p_i p_j x_i x_j$$

が得られる。(48)と(55)から $\bar{x}$ の最良推定値は

$$(57)\quad F_1 = \sum_{i=1}^{s} r_i x_i$$

但し $\quad x_i = \sum_{j=1}^{m_i} x_{ij} / m_i$

によつて與へられる。

19. m_i と N に対する最適な與へ方

任意の変数の組 $(a_1, \cdots\cdots, a_s)$ に対して次の恒等式が成立つ。

$$(58) \quad \sum_{i=1}^{s} m_i \left(\frac{a_i}{m_i} - \frac{\sum_i a_i}{m_i} \right)^2$$

$$\equiv \sum \frac{a_i^2}{m_i} - \left(\sum_i 2a_i \right) \frac{\sum a_i}{\sum m_i} + \frac{(\sum a_i)^2}{\sum m_i}$$

$$\equiv \sum \frac{a_i^2}{m_i} - \frac{(\sum a_i)^2}{\sum m_i} = \sum \frac{a_i^2}{m_i} - \frac{(\sum a_i)^2}{m_0}$$

$a_i = \sigma_i \sqrt{p_i^2 + p_i(1-p_i)N^{-1}}$ とおけば (58) から

$$(59) \quad \sum_i \left[p_i^2 + p_i(1-p_i)N^{-1} \right] \frac{\sigma_i^2}{m_i}$$

$$= \frac{\left[\sum_i \sigma_i \sqrt{p_i^2 + p_i(1-p_i)N^{-1}} \right]^2}{m_0}$$

$$+ \sum m_i \left(\frac{\sigma_i \sqrt{p_i^2 + p_i(1-p_i)N^{-1}}}{m_i} - \frac{\sum_i \sigma_i \sqrt{p_i^2 + p_i(1-p_i)N^{-1}}}{m_0} \right)^2$$

更に

$$(60) \quad \sum p_i(1-p_i)x_i^2 - 2\sum_{i=1}^{s-1} \sum_{j=i+1}^{s} p_i p_j x_i x_j = \sum_i p_i x_i^2 - (p_1 x_1 + \cdots + p_s x_s)^2$$

$$= \sum p_i x_i^2 - \bar{x}^2 = \sum p_i (x_i - \bar{x})^2$$

(56), (59), (60) から

$$(61) \quad V_1 = \frac{\left(\sum \sigma_i \sqrt{p_i^2 + p_i(1-p_i)N^{-1}}\right)^2}{m_0} + \sum_i m_i \left(\frac{\sigma_i \sqrt{p_i^2 + p_i(1-p_i)N^{-1}}}{m_i} \right.$$

$$\left. - \frac{\sum_i \sigma_i \sqrt{p_i^2 + p_i(1-p_i)N^{-1}}}{m_0} \right)^2 + \frac{1}{N} \sum_i p_i (x_i - \bar{x})^2$$

この式の右辺の第一項と第三項は m_i に無関係で m_o と N に関係する。よつて m_i $(i=1,\cdots\cdots,s)$ は中央項が 0 になる様に選ぶと，即ち

$$(62)\qquad m_i=\frac{m_o\sigma_i\sqrt{p_i^2+p_i(1-p_i)N^{-1}}}{\Sigma\sigma_i\sqrt{p_i^2+p_i(1-p_i)N^{-1}}}$$

とおく。しからば (61) は

$$(63)\qquad V_1'=\frac{\left(\Sigma\sigma_i\sqrt{p_i^2+p_i(1-p_i)N^{-1}}\right)^2}{m_o}+\frac{1}{N}\Sigma p_i(x_i-\bar{x})^2$$

に簡約される。更に $p_i(1-p_i)N^{-1}$ を無視すれば

$$(64)\quad V_1'\sim V_1''=\frac{(\Sigma\sigma_i p_i)^2}{m_o}+\frac{1}{N}\Sigma p_i(x_i-\bar{x})^2$$

我々は條件

$$Am_o+BN=C$$

のもとに (64) の右辺を最小なる如くきめれば最適な選び方となる。

$$(\Sigma p_i\sigma_i)^2=a^2,\ \Sigma p_i(x_i-\bar{x})^2=b^2$$

とおけば

$$(65)\qquad \frac{a^2}{m_o}+\frac{b^2}{N}$$

を條件 $Am_o+BN=C$ のもとに最小ならしめればよい。ラグランジユの方法を用いて

$$-\frac{a^2}{m_o^2}=-\lambda A,\quad -\frac{b^2}{N^2}=\lambda B$$

（ λ＝ ラグランヂュの乗数）

である事を知る。

よつて m_0, N は夫々 $\dfrac{a}{\sqrt{A}}$, $\dfrac{b}{\sqrt{B}}$ に比例させればよい事が分る。 即ち

$$m_0 = K\frac{a}{\sqrt{A}}, \quad N = K\frac{b}{\sqrt{B}}$$

比例項 K の値は

$$(68) \qquad K(a\sqrt{A} + b\sqrt{B}) = C$$

から求まる。（67）（68）から

$$(69) \quad m_0 = (a/aA + b\sqrt{AB}) : N = (b/(a\sqrt{AB} + bB))$$

が出る。

上の結果から知る通り，我々は m_0, N, 又 m_i の最適な値を求めるために，母数，p_i, σ_i, x_i が既知であるとした。しかしそれらの値は，初等的不等式から推定しうるものである。

p_i, σ_i, x_i の第一次近似値は得られる。 それらを（69）（62）式に代入すれば m_0, N, m_i の近似値が得られる。

二重標本調査法を用ふべきか否かを決める前に，我々は，それが直接的な単純無作為標本調査法よりよりよい結果を與へるであらうか。どうかについて研究すべきである。

同じ費用 C のもとに，大きさ $M = C/A$ の単純無作為標本 S_0 を採用し，母集団平均 $\bar{X}$ の線型推定値を求めるならば，其の最良推定値は標本平均 $\bar{x}$ で，分散 V_0 は

$$(70) \qquad V_0 = \frac{1}{M}\left[\Sigma p_i \sigma_i^2 + \Sigma p_i (x_i - \bar{x})^2\right] \quad 但し\ M = \frac{C}{A}$$

で與へられる。 式（70）は単一観測の分散が式（26）で與へら

れると云ふ事から従ふ。　かくて二重標本抽出法を用ふるかどうかを決定するためには，我々は V_i ，σ_i ，x_i に対して，既知の知識又は基礎的不等式から得られる基礎的値を代入して（70），(69)(64)式で夫々與へられる値 V_0 ，　m_0 ，N ，V'' を計算してみる。　そして V_1'' が V_0 より小さいか大きいかによつて二重標本抽出法が単一単純無作為標本抽出法を用ひるより有利であるか，ないかを決めるべきである。

20.　或る例

$s=3$ ，$p_1=\frac{1}{4}$ ，$p_2=\frac{2}{4}$ ，$x_1=1$，$x_2=3$，$X=6$

$\sigma_1=1$，$\sigma_2=2$，$\sigma_3=4$，$A=4$，$B=1$，$C=500$

とすれば

$$a=\Sigma p_i\sigma_i=2.25$$

$$\bar{x}=3.25$$

$$\sigma^2=\Sigma p_i(x_i-\bar{x})^2=3.1875\doteqdot(1.7854)^2$$

を得る。（61）から最適値

$$m_0=59\quad N=142\qquad（最も近い整数値）$$

更に（62）式を用ひて　$m_1=10$　$m_2=39$　$m_3=40$

が得られる。　分散 V_1 は（63）から

$$(71)\qquad V_1=.079555$$

この値が、m_i を（62）に従つてきめない値から得られる分散と比較してみるのは興味がある。　例へば $m_1=29$　$m_2=m_3=30$

とおけば

$$V_1=.0919$$

を得る。 かくて(71)で與へられた値よりも，15％も多くなり精度に於て，考へなければならない、又不必要な損失を招く(70)から又

$$(72) \quad V_0 = .0755$$

を得る。 かくてこの場合 $V_1 > V_0$ で単純無作爲標本抽出法を用ひた方がよりよい精度を得る事が分る。

この結果は変数 x についての層間の違いが層内に於ける x の変異にしらべて小さいとゆう事と，x, y について資料を得るための費用の差があまり大きくないという事実によるものである。

この点を説明するために、次の例を考へよう。

例 II. p_i, σ_i, x_i は，例 I. に於けると同じとする。$A = 40$, $B = 1$, $C = 5000$ とすれば，C/A は例 I と同じ値をとるから V_0 も例 I と同じになる。 しかしこの場合

$$V_1 = .05147$$

となり，二重標本抽出法を用ひた方が大きな利益を得る。

例 III. p_i, σ_i, A, B, C は，例 I. と同じとする。$x_1 = 1$, $x_2 = 6$, $x_3 = 11$ とする。 層間の変差は例 I より大きくなつてゐる。 必要な計算を実行すると

$$V_0 = .15 \qquad V_1 = .129$$

となり、この場合も二重標本抽出法の方が直接的な単純無作爲標本抽出法より 15 ％ も精度が良くなつてゐる。

21. 同じ精度が要求される場合の各層の標本数の配分の決め方

k この層，Π_1, ----- Π_k に層別された母集団 Π を考へる。x を或る標識とし，α_{ij} を i 層に於ける j 元素の値とす

る。

M_i を Π_i に於ける元素の総数 $M=\sum_{i=1}^{k} M_i$ とし，

$$\alpha_i = \left(\sum_{j=1}^{M_i} \alpha_{ij}\right)\Big/ M_i$$

とおく前の講義に於て我々はパラメーター

$$B=\sum_{i=1}^{k}\sum_{j=1}^{M_i} \alpha_{ij}=\sum_{i=1}^{k} M_i\alpha_i$$

を推定する問題を考へた。　このために k この層の間の標本数の最良の配分は（22）で與へられた。

しかしながら，B を推定するための目的でなく層 $\Pi_1,\cdots\cdots,\Pi_k$ の平均 $\alpha_1,\cdots\cdots,\alpha_k$ そのものを調べ，それらを互ひに比較する事が目的である場合もたまたま起る。

このやうな場合に於て我々は $\alpha_1,\cdots\cdots,\alpha_k$ が皆同じ精度で推定される事が望ましい。　m_i を i 層からとられた標本数 $\bar{x}_i$ をその標本平均とする。

又 σ_i^2 を i 層の母集団分散とすれば $\bar{x}_i$ の分散は

$$\sigma_{\bar{x}_i}^2=\frac{\sigma_i^2(M_i-m_i)}{m_i(M_i-1)}$$

で與へられる。　もし標本の大さが決まってゐるならば，例へばそれを m とするとき $m_1,\cdots\cdots,m_k$ は次の方程式を満足しなければならない。

$$m_1+\cdots\cdots+m_k=m$$

(73)
$$\frac{\sigma_1^2(M_1-m_1)}{m_1(M_1-1)}=\cdots\cdots=\frac{\sigma_k^2(M_k-m_k)}{m_k(M_k-1)}$$

もし M_i が m_i に較べて非常に大きいならば $\dfrac{M_i-m_i}{M_i-1}$ は、

1に近似するから，この場合上のこつの方程式は

$$m_1 + \cdots\cdots + m_k = m, \quad \frac{m_1}{\sigma_1} = \cdots\cdots = \frac{m_k}{\sigma_k^2}$$

即ち，標本の大さ $m_1, \cdots\cdots, m_k$ は分散 $\sigma_1^2, \cdots\cdots, \sigma_k^2$ に比例してとるべきである。 (73)式の取扱ひに関しては Stock and Frankel Annals of Math. Statistic p. 238. Sep. を参照せよ。

22. 乱化法の原理

統計的研究計画に於ける重要な道具の一つとして，所謂乱化法と呼ばれる方法がある。それを簡単な例で説明してみよう。

ミルクが注がれてあるお茶を、ミルクかお茶かどちらが最初に加えられたかを風味する事によつて鑑別出来ると主張する一人の婦人がいるとする。 そこで，ミルクが最初に注がれた茶わん4けと，お茶が最初に加えられた茶わん4け，合計8けの茶わんを用意せよ。

虚無假説として，この婦人は鑑別能力がないと云ふ假説を立てる。 即ち，8けの茶わんを4けづつのこつの群に分けるいかなる分類の仕方も，同様な可能性をもつとする。 このやうな分類の仕方は $8!/4!4! = 70$ 通りであるから、虚無假説のもとで、正しい分類が出来る確率は $1/70$ に等しい。 虚無假説が真であるときすべての分類の仕方の確率は各々相等しい事を保証さすために，婦人の判断に影響を及ぼすやうなすべての他の事柄に関して、同じ様な條件をつけなければならない。

例へば，ミルクを最初に注いだ茶わんに砂糖を加え，他の茶わんにそれを加えないならば，それらを風味する事によつて婦人の判断力に影響を及ぼす明りような違いがある。

その他結果を損うような多くの起り得る誤差がある。

例へば，ミルクの量を茶わんごとにいろいろ変へる等。

我々が検定しようとする事柄——即ちミルクとお茶とどちらが先に加えられてゐるか——以外の事で８けの茶わんについて完全に同じやうな状態に，実験を計画する事は不可能である。

しかし，それらの有意な変異は乱化法で，かなり広範囲まで消去する事が出来る。　それらの有意因子をあらかじめ決めてしまつてから，８けの茶わんから，無作爲に４けを決めてしまう。

しからば，虚無假説が眞であるとゆう前提のもとに，有意因子の存在の如何に係はらず婦人が正しい分類をする確率はやはり，$1/70$ である。

これは次の様にして知る事が出来る。

８けの茶わんを４けづつの二つの群に分類する仕方の数は70通りある。　それを $C_1, C_2, \cdots\cdots, C_{70}$ と記してみる。

検定すべき事柄については婦人は鑑別力はないが，しかし８けの茶わんの分類についての婦人の判定に影響を及ぼす，例へば，茶わんごとに砂糖の分量を変へるとゆう様な或る有意因子がきまつているとする。

彼女が分類 C_i を決定する確率を p_i とすれば，p_i は有意因子が存在するから一般に $1/70$ と異なるであらう。

実験者は有意因子をどの茶わんに施すかを夫々きめて，それからその８けの茶わんから４けを無作爲に４け選び，それに最初にお茶を注ぎ，残りに最初にミルクを注ぐ。

実験者は70通りの分類の仕方 $C_1, \cdots\cdots, C_{70}$ から一つを無作爲に選擇する事が云へた。

虚無假設が正しいと云ふ前提のもとづけは，婦人によつてなされた分類は実験者によつて選ばれた分類と独立である。　よつて実験者が C_i を選び又婦人がこれと同じ分類 C_i を決める確率は積 $p_i \dfrac{1}{70}$ に等しい。　よつて婦人が正しい分類をなす確率は婦人の分類と実験者のそれが一致する場合でそれは，

$$\sum_{i=1}^{70} p_i \frac{1}{70} = \frac{1}{70}$$

に等しい。 かくて彼女が正しい分類をする確率は，もしもすべての有意因子がきめられてしまつた後，実験者による4けの茶わんの無作為選擇（無作意にどの茶わんにお茶又はミルクを最初につぐかをきめる事）をするならば有意因子が存在しない場合と同じになる。

乱化法の重要性は，乱化法が適用される事によつて，有意因子の存在によつて起される所の検定手続に於けるいかなる偏倚も消去されると云ふ事実による。

23. 他の例

我々は農作物のいろいろの変種の作高の平均値が取扱ひ，u と v によつて同じであるかどうかとゆう假説を検定したい。

このために我々は15対の実験用プロットを用意する。
そして各一対のプロットに取扱ひ，u と v を施す事にする。

u_α，v_α を α 番目のプロットに夫々取扱ひ u，v による作高とする。 $P_{\alpha 1}$，$P_{\alpha 2}$ を夫々 α 番目の左側，右側プロットとする。 $\gamma_{\alpha i}$ $(i=1,2)$ をプロット $P_{\alpha i}$ に於ける取扱ひが施されない場合の作高の平均値とする。

$\mu+\delta_{\alpha i}$；$\nu+\beta_{\alpha i}$ を夫々取扱ひ u，v が施されたときのプロット $P_{\alpha i}$ に於ける作物平均高とする。

さて30けの取扱ひ，$u_1, \cdots, u_{15}$；$v_1, \cdots v_{15}$ は独立に分布しているとする。 作高は一般に土壌肥沃度の差によつて異なるであらう。 もしもプロットの対を（例へば硬貨を投げる等によつて）無作為に決めるならば，次の式が成り立つと假定してみる。

1	u_1	v_1
2	u_2	v_2
3	u_3	v_3
4	u_4	v_4
5	u_5	v_5
6	u_6	v_6
7	u_7	v_7
8	u_8	v_8
9	u_9	v_9
10	u_{10}	v_{10}
11	u_{11}	v_{11}
12	u_{12}	v_{12}
13	u_{13}	v_{13}
14	u_{14}	v_{14}
15	u_{15}	v_{15}

(74)
$$\begin{cases} u_\alpha = \mu + \varepsilon_\alpha + \eta_\alpha \\ v_\alpha = \nu + \varepsilon'_\alpha + \eta'_\alpha \end{cases}$$

こゝで $\varepsilon_1, \cdots, \varepsilon_{15}$, $\varepsilon'_1, \cdots, \varepsilon'_{15}$ は平均 0 をもつて独立に分布する確率変数で，η_α は γ_{21}, γ_{22} のいづれかの値を確率　でとる。又

$$\eta'_\alpha = \gamma_{\alpha 1} + \gamma_{\alpha 2} - \eta_\alpha \quad \text{である。}$$

よつて

(75)

$$u_\alpha - v_\alpha = (\mu - \nu) + (\varepsilon_\alpha - \varepsilon'_\alpha) + (2\eta_\alpha - \gamma_{\alpha 1} - \gamma_{\alpha 2})$$

かくて

$$E(u_\alpha - v_\alpha)$$
$$= E(\mu - \nu) + E(\varepsilon_\alpha) - E(\alpha\varepsilon'_\alpha) + E(2\eta_\alpha - \gamma_{\alpha 1} - \gamma_{\alpha 2})$$
$$= \mu - \nu$$

よつて，虚無假説が正しいならば，即ち

$$\mu = \nu$$

ならば，$u_1 - v_1, \cdots\cdots, v_{15} - v_{15}$ は平均 0 をもつて独立に分布する。もしも $(u_\alpha - v_\alpha)$ $(\alpha = 1, \cdots\cdots 15)$ が，同じ分散をもつて正規分布に従ふならば，良く知られた t 分布

$$t = \frac{\sqrt{15}\ \bar{z}}{s}$$

但し　$\bar{z} = (z_1 + \cdots\cdots + z_{15})/15 \qquad z_\alpha = u_\alpha - v_\alpha$

$$s^2 = \frac{\sum_{\alpha=1}^{15} (z_\alpha - \bar{z})^2}{14}$$

をなす。　t 検定は，もとの分布が正規でないときも，その標本数があまり小さくなければ，近似的に之を用いても差支へない。

もしも t 検定の適用性がうたがはしい時には R.A.Fisher によつて導かれた所の次の正密検定を用ひる事が出来る。

確率変数，$\varepsilon_1, \cdots\cdots, \varepsilon_{15}$ ；$\varepsilon_1', \cdots\cdots \varepsilon_{15}'$ については，その平均が 0 で独立にしかも同じ分布法則に従つてゐる事以外に何も知られてゐないとする。　しからば $\varepsilon_\alpha - \varepsilon_\alpha'$ の分布は，縦軸に関して対称で，又 $(2\eta_\alpha - \gamma_{\alpha 1} - \gamma_{\alpha 2})$ の分布も縦軸に関して対称で且つ η_α は ε_α と ε_α' と独立であるから (75) から，$u_\alpha - v_\alpha$ の分布はもし虚無假説が眞である，即ち $\mu - \nu = 0$ ならば，縦軸に関して対称である。

我々は $|u_\alpha - v_\alpha|$ があるきまつた値，例えばそれを r_α を持つ所の部分母集団を考へよう。

$u_\alpha - v_\alpha$ の分布の対称性から與へられた部分母集団に於て，$u_\alpha - v_\alpha$ が r_α に等しい確率も又 $-r_\alpha$ に等しい確率も共に $1/2$ である。

次に $(k_1 r_1 \cdots\cdots k_{15} r_{15})$ （但し $k_1 \cdots\cdots k_{15} = \pm 1$）なる 2^{15} 組の数列を考えよう。　しからば與えられた部分母集団に於て 2^{15} の数列の各々は皆等しい確率を持つから，かくて又，それに対応する $\bar{z} = (\sum_1^{15} k_\alpha r_\alpha)/15$ の値は同じ確率を持つて起る。　$\bar{z}$ のそれらの 2^{15} の値を大きの順に並べ，この並べた両端に於ける棄却域を決める。　即ち標本から計算された

$$\sum_{\alpha=1}^{15} (u_\alpha - v_\alpha) \Big/ 15$$

が，この並べられた $\bar{z}$-数列の或る片側に落ちれば，$\mu - \nu = 0$ とゆう假説を棄却する事にする。

さて棄却域に含まれる Ξー数列の Ξ の数を a とすると，もし虚無仮説が眞であるとき，之を誤つて棄却する確率は $a/2^{15}$ に等しい。　t 検定と較べて，この検定方法の利点は，虚無仮説が眞であるとき，これを棄却する確率が，ε_α の未知の分布の型に関係しない事である。

24. 乱化法の利点と欠点

乱化法の利点は，それが有意因子の存在の有無に係はらず不偏な検定が出来るところにある．

例えば，§23 で挙げた例では，有意因子はプロットの相違による土壌沃肥度である。　即ち α と i の函数である所の $\gamma_{\alpha i}$ がそれである。

$$\gamma_{\alpha 1} - \gamma_{\alpha 2} = \delta \quad (\alpha = 1, \cdots\cdots, 15)$$

δ を或る正の常数とすれば，取扱ひ，u，v を夫々プロットの左側，右側に施したとき，虚無仮説，即ち $\mu = \nu$ が眞であつてさえ，差 $\sum u_\alpha/15 - \sum v_\alpha/15$ に有意を見出すであらう。

この場合，虚無仮説が正しいと云ふ前提のもとでも $u_\alpha - v_\alpha$ の期望値が 0 でなく，δ に等しいから検定は相当な偏倚をもたらす。　しかしながらもし実験を §23 で述べた方法で乱化するならば（$u_\alpha - v_\alpha$）の期望値は $\gamma_{\alpha i}$ のいかんに係はらず常に $\mu - \gamma$ に等しい。　（もしも虚無仮説が正しいならば，それは 0 である。）　　かくて，もし $\gamma_{\alpha i}$ が未知であるとき，乱化法は検定を不偏にする利点を持ち，他のいかなる系統的並べ方も偏倚検定をもたらす事になる。

しかしながら乱化法は次に述べる或る欠点をも持つてゐる。それは統計的検定のもとゝなる分散の増大によつて検定の感度を減小せしめる事である．

この事をもつと詳しく説明しよう。

$\gamma_{\alpha 1} - \gamma_{\alpha 2} = \delta > 0$ である場合を考える。

しからば $\gamma_{\alpha 1}+\gamma_{\alpha 2}-2\eta_\alpha$ は $+\delta$ か $-\delta$ かいづれかを夫々，確率 $1/2$ でとる。 ε_α, ε'_α の分散は皆 σ^2 に等しいとする。しからば（75）から（$u_\alpha - v_\alpha$）の分散は $2\sigma^2+\delta^2$ に等しい事が分る。よつて $\Sigma(u_\alpha - v_\alpha)/N$ の分散は $(2\sigma^2+\delta^2)/N$ に等しい。こゝで N は，我々が実験に対して用ひられるプロットの対の数とする。さて我々は次の無作為な並べ方と，系統的な並べ方とを比較してみよう。

取扱ひ，u はプロット $P_{1,1}\cdots\cdots P_{\frac{N}{2},1}$，$P_{\frac{N}{2}+1,2},\cdots P_{N,2}$ に施こし，v は残りのすべてのプロットに施すとする。（N は偶数であるとしておく）しからば $(\Sigma u_\alpha - \Sigma v_\alpha)/N$ の期望値は $\mu - \nu$ に等しく（もし虚無假説が正しければ 0 に等しい。）又，$(\sum_\alpha u_\alpha - \Sigma v_\alpha)/N$ の分散は $2\sigma^2/N$ である。
又 σ^2 は

$$s^2 = \frac{\sum_{\alpha=1}^{N/2}(z_\alpha - \bar{z}_1)^2 + \sum_{\alpha=\frac{N}{2}+1}^{N}(z - \bar{z}_\alpha)^2}{N-2}$$

$$\text{但し}\quad z_\alpha = u_\alpha - v_\alpha \qquad \bar{z}_1 = \frac{2}{N}\sum_1^{N/2} z_\alpha$$

$$\bar{z}_2 = \frac{2}{N}\left(\sum_{\frac{N}{2}+1}^{N} z_\alpha\right)$$

で推定される。かくて，この場合系統的並べ方は

無作為な並べ方よりもずつと有効な検定が得られる。

我々は $\gamma_{\alpha 1} - \gamma_{\alpha 2} = \delta$ $(\alpha = 1,\ \cdots\cdots, N)$ が既知であると假定してゐるからこそこの方法で有効な検定が得られたわけである。しかしながら，大低の実際的場合では，$\gamma_{\alpha i}$ について何も知る事が出來ない事が多い。それで，かゝる場合系統的並べ方は検定を偏倚にするから，無作爲な並べ方の方が最も安全である。

結論として，次の事が云へる（以下第二巻第1号へつづく）