教育統計の諸問題 II

青山博次郎

(1956年11月受付)

On Some Statistical Problems in Education II

Hirojiro AOYAMA

In this papar we treat the problems of pre-screening in an achievement test, estimation of parameters of the distribution with regard to heterogeneous groups, and determination of the number limit of teachers in whole country. The former two of these problems were studied in cooperation with the Second Examination Section, Bureau of Recruitment, National Personnel Authority.

Institute of Statistical Mathematics.

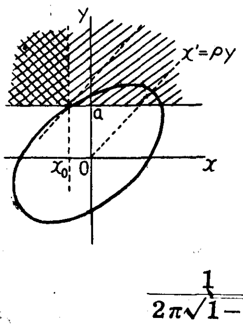
§1. 緒　言

さきに発表した教育調査における諸問題 I（彙報第1巻第1号）につづくものであるが，調査のみならず一般に教育統計に関する種々の問題をとり上げてみることにした．本稿では特に予備選択の問題と，集団の不均質性の影響，教員の定員決定の問題である．前二者は人事院任用局試験第二課においてとり上げられた問題であり，最後のものは文部省財務課においてとり上げられた問題である．

§2. 予備選択の方法 I

幾つかのテストがあるとき，最終テストの上位のものだけを残すように最初のテストで受験者を予めふるい落す方法について考えよう．これを予備選択といい，最終テストに多くの経費がかかるときに用いられている方法である．

先づテストが x,y の2つで何れも z スコア（平均 0, 分散 1）で表現されているものとする．x, y の同時分布はガウス分布とし，相関係数を r とする．$y \geqq a$ なるものを一定の損失（例えば $y \geqq a$ なるもののうち 5% は判定を誤る）の下に，テスト x によつて $x \geqq x_0$ となる者だけを選択するものとしよう．損失を $100\alpha\%$ 以下にするために回帰直線 $x' = \rho y$ を用い



$$x_0 = a\rho - \lambda_\alpha \sqrt{1-\rho^2} \qquad (1)$$

をとる．ここで λ_α はガウス分布の片側 $100\alpha\%$ 点である．これは

$$P_r(x < x_0 | y = a) = \alpha \qquad (2)$$

を満足する．

このとき実際 $x < x_0$ であるが，$y \geqq a$ となるものの割合は

$$\frac{1}{2\pi\sqrt{1-\rho^2}} \int_a^\infty \int_{-\infty}^{x_0} \exp\left[-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)\right] dxdy$$

$$= \int_a^\infty \varphi(y) \Phi\left(\frac{x_0 - \rho y}{\sqrt{1-r^2}}\right) dy \equiv P_1 \qquad (3)$$

但し
$$\varphi(y)=\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}},\quad \Phi(z)=\int_{-\infty}^{z}\varphi(y)dy$$

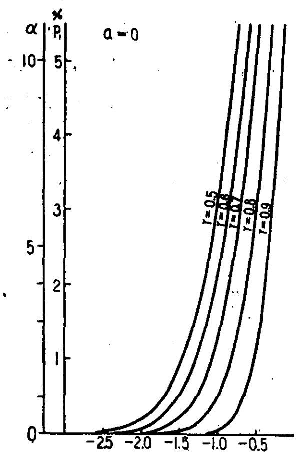
従つて $y\geqq a$ となる割合 $P=1-\Phi(a)$ が与えられると a が定まり，（1）より x_0 が求められる．このときの実際の損失は P_1 であつて，一般には αP より小である．

因みに P_1 を求めると

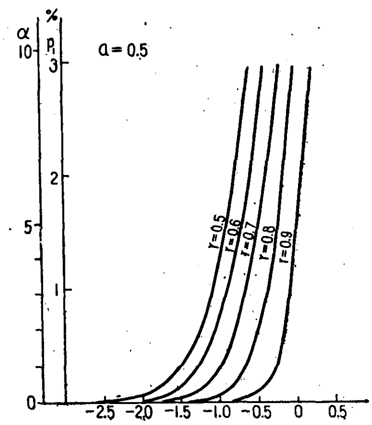
$$P_1<\alpha P-\rho\sqrt{1-\rho^2}\,\varphi(\lambda_\alpha\sqrt{1-\rho^2}-a\rho)[\varphi(\rho\lambda_\alpha+a\sqrt{1-\rho^2})$$
$$-(\rho\lambda_\alpha+a\sqrt{1-\rho^2})\{1-\Phi(\rho\lambda_\alpha+a\sqrt{1-\rho^2})\}]\qquad(4)$$

例えば $P=0.063$, $\alpha=0.05$, $\rho=0.98$ とすると $\lambda_\alpha=1.65$, $a=1.53$, $x_0=1.171$ となり，$x\geqq x_0$ の割合は 12.1％となる．

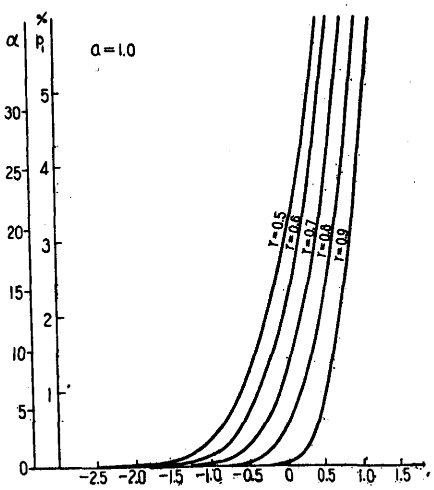
もし $P_1=\alpha P$ となる如く x_0 を求める場合は $P_1=\alpha P$ より x_0 を求めればよい．次の第1～4図は種々の ρ に対して，$a=0$, 0.5, 1.0, 1.5 の場合につき x_0 を求める図表である．



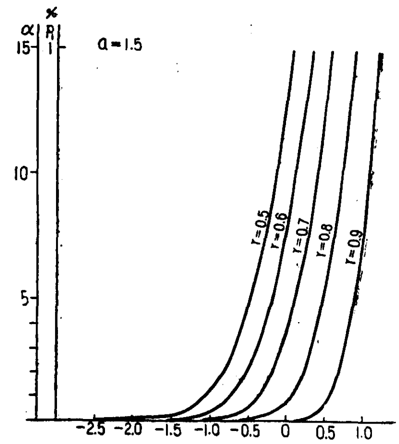
第　1　図



第　2　図



第　3　図



第　4　図

もしも **L. Chesire, M. Saffir and L. L. Thurstone** の **Computing Diagrams for the Tetrachoric Correlation Coefficient, 1933** があれば，四分相関係数を利用してもよい．この図表は $(b+d)/N$, $(c+d)/N$, d/N より ρ を求

b	a
d	c

められるようになつたものである．

（1）式利用の場合と，前の図表またはこの図表を用いて計算した結果は前例の $\rho = 0.98$ の場合に次のようになる．

	P	$x \geqq x_0$ の割合	実際の P_1 の人数	理論上の P_1 の人数
A 法	6.3％（13人）	9％（16人）	1	0.6人
B 法	〃	12.1（24人）	0	<0.6
A 法	16.0％（32人）	19（38人）	2	1.9
B 法	〃	29.7（59人）	0	<1.6
A 法	24.2％（48人）	27（54人）	2	2.4
B 法	〃	36.1（71人）	0	<2.4

（注）
A 法：四分相関
B 法：（1）式利用

次にテストが3つのときを考える．このときテストの得点を $x,\ y,\ z$ とするとき

$$U = w_1x + w_2y + w_3z = X + Y + Z, \qquad V = w_1x + w_2y = X + Y$$

とおく．w_i はテストの重みである．このとき問題となるのは U を最終的のテストの得点と考えるとき X だけ，Y だけ，或いは V の何れを用いて予備選択するのがよいかということである．この時も変数が2つの場合と同様に考えればよいが*，$Y \geqq Y_0$ の部分で完全に含まれる部分ができる限り大きくしたい要求がある（Z の平均，分散を推定するため，§4 参照）．

具体的な例によつて示そう．$\rho_{XY},\ \rho_{XZ},\ \rho_{YZ}$ をテスト間の相関係数，$\sigma_X^2,\ \sigma_Y^2,\ \sigma_Z^2$ をテストの分散とする．このとき

$$\rho_{UV} = \frac{\sigma_X^2+\sigma_Y^2+2\rho_{XY}\sigma_X\sigma_Y+\rho_{XZ}\sigma_X\sigma_Z+\rho_{YZ}\sigma_Y\sigma_Z}{\sqrt{\sigma_X^2+\sigma_Y^2+2\rho_{XY}\sigma_X\sigma_Y}\sqrt{\sigma_X^2+\sigma_Y^2+\sigma_Z^2+2\rho_{XY}\sigma_X\sigma_Y+2\rho_{XZ}\sigma_X\sigma_Z+2\rho_{YZ}\sigma_Y\sigma_Z}}.$$

$$\rho_{UY} = \frac{\rho_{XY}\sigma_X+\rho_{YZ}\sigma_Z+\sigma_Y}{\sqrt{\sigma_X^2+\sigma_Y^2+\sigma_Z^2+2\rho_{XY}\sigma_X\sigma_Y+2\rho_{XZ}\sigma_X\sigma_Z+2\rho_{YZ}\sigma_Y\sigma_Z}}$$

$$\rho_{VY} = \frac{\sigma_Y+\rho_{XY}\sigma_X}{\sqrt{\sigma_X^2+\sigma_Y^2+2\rho_{XY}\sigma_X\sigma_Y}}$$

いま　$\rho_{XY} = 0.721,\ \rho_{XZ} = 0.522,\ \rho_{YZ} = 0.750,\ \sigma_X = 60,\ \sigma_Y = 90,\ \sigma_Z = 50$　のときについて考えると

$$\rho_{UV} = 0.980, \qquad \rho_{UY} = 0.956, \qquad \rho_{UX} = 0.846$$

前と同様に $U \geqq U_0$ の割合が 0.06 になるようにし，損失を 0.15％（$\lambda_\alpha = 3$）とすると，$V \geqq V_0$ となる割合は 0.176,

$$V_0 = \bar{V} + (1.56\rho_{UV} - 3\sqrt{1-\rho^2_{UV}})\sigma_V = \bar{V} + 0.9318\sigma_V$$

このとき $V \geqq V_0$ なる部分に殆んど完全に含まれる $Y \geqq Y_0$ の割合は

$$Y_0 = \bar{Y} + \sigma_Y(0.9318\rho_{VY} + 3\sqrt{1-\rho^2_{VY}})$$
$$= \bar{Y} + 1.8054\sigma_Y$$

であるから 0.0355 となる．

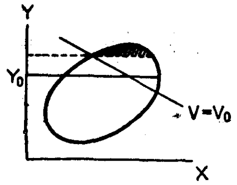
もし直接に U と Y の関係より $Y \geqq Y_0$ なる選択を行えば

$$Y_0 = \bar{Y} + (1.56\rho_{UY} - 3\sqrt{1-\rho^2_{UY}})\sigma_Y = \bar{Y} + 0.6113\sigma_Y$$

となり，$Y \geqq Y_0$ なる割合は 0.271 となる．

これらの関係を表示すると次頁の表のようになる．これによつてみると，V_0 による選択の方が効率はよい（0.176）が，$Y \geqq Y_0$ を殆んど完全に含む割合が小さい．従つて $Y \geqq Y_0$ による選択

* §3の表ができると四分相関係数を利用したと同様の方法で正確な分割点を求めることができる．



$P_r(U \geqq U_0)=0.06$	$P_r(V \geqq V_0)$	$Y \geqq Y_0$ が殆んど完全に含まれる割合
V_0 で分割	0.176	0.0355
〃	0.271	0.0668
Y_0 で分割	—	0.271

は効率（0.271）は悪いが最初の要求には合している．このことから実際には採点能力を考え，$Y \geqq Y_0$ による選択をどの位の割合で行えばよいかを決定しなければならない．

§3. 予備選択の方法 II

前の方法は最終テストによる合格者を目標として予備選択を如何にすればよいかということであつた．ここでは幾つかのテストの各々について夫々独立に予備選択を行うときは，これらの選択を合成すると，結局どれだけのものが選択されたことになるかを求める問題を考えよう．

先づテストはすべて z スコアで表わされており，同時分布はガウス分布としておく．

もしテストが x_1, x_2 の2つであれば §2 の四分相関係数の図表を用いれば直ちに結果が判る．即ち $\dfrac{a+c}{N}\left(\text{または } \dfrac{b+d}{N}\right)$, $\dfrac{a+b}{N}\left(\text{または } \dfrac{c+d}{N}\right)$ を与えれば表より直ちに $\dfrac{a}{N}$ が求められるからである．

そこで図表を利用してテストが x_1, x_2, x_3 の3つのとき選択について考えよう．仮定により x_1, x_2, x_3 の同時分布は

$$\varphi(x_1,x_2,x_3)=\frac{1}{(2\pi)^{\frac{3}{2}}\sqrt{\varDelta}}\exp\left\{-\frac{1}{2\varDelta}(\varDelta_{11}x_1^2+\varDelta_{22}x_2^2+\varDelta_{33}x_3^2+2\varDelta_{12}x_1x_2+2\varDelta_{13}x_1x_3+2\varDelta_{23}x_2x_3)\right\} \tag{1}$$

但し

$$\varDelta=\begin{vmatrix} 1 & r_{12} & r_{13} \\ r_{12} & 1 & r_{23} \\ r_{13} & r_{23} & 1 \end{vmatrix}$$

$\varDelta_{ij}$ は $\varDelta$ の i, j 要素の余因数

である．

この相関曲面を $x_3=\text{const.}$ で截るとき，x_3, x_3+dx_3 の間に入る確率要素は

$$\varphi(x_3)dx_3=\frac{1}{\sqrt{2\pi}}e^{-\frac{x_3^2}{2}}dx_3 \tag{2}$$

截口の密度函数は

$$\frac{1}{(2\pi)^{\frac{3}{2}}\sqrt{\varDelta}}\exp\left\{-\frac{1}{2\varDelta}(\varDelta_{11}\xi_1^2+\varDelta_{22}\xi_2^2+2\varDelta_{12}\xi_1\xi_2+C)\right\}\Big/\frac{1}{\sqrt{2\pi}}e^{-\frac{x_3^2}{2}} \tag{3}$$

ここで ξ_1, ξ_2 は

$$\left.\begin{aligned} \varDelta_{11}x_1+\varDelta_{12}x_2+\varDelta_{13}x_3=0 \\ \varDelta_{12}x_1+\varDelta_{22}x_2+\varDelta_{23}x_3=0 \end{aligned}\right\} \tag{4}$$

を満足する (x_1^0, x_2^0) を原点とする新座標系 $O'\text{-}\xi_1\xi_2$ による座標を表わし，

$$C=\varDelta_{11}(x_1^0)^2+\varDelta_{22}(x_2^0)^2+2\varDelta_{12}x_1^0x_2^0+2(\varDelta_{13}x_1^0+\varDelta_{23}x_2^0)x_3+\varDelta_{33}x_3^2=\varDelta x_3^2 \tag{5}$$

従つて（3）式は

$$\frac{1}{\sqrt{\varDelta_{22}}}\xi_1=\eta_1, \qquad \frac{1}{\sqrt{\varDelta_{11}}}\xi_2=\eta_2$$

とおくとき

$$\frac{1}{2\pi\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}(\eta_1{}^2+\eta_2{}^2-2\rho\eta_1\eta_2)\right\} \tag{6}$$

と変形できる．

ここで ρ は偏相関係数 $r_{12\cdot3}$ で

$$\rho = r_{12\cdot3} = -\frac{\Delta_{12}}{\sqrt{\Delta_{11}\Delta_{22}}} \tag{7}$$

である．

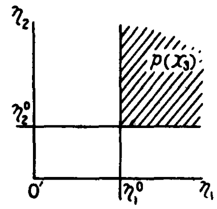
ところで　$x_1 = h$　を $O'\text{-}\xi_1\xi_2$ 系で表わすと

$$\eta_1{}^0 = \frac{1}{\sqrt{\Delta_{22}}}(h - r_{13}x_3) \tag{8}$$

同様に　$x_2 = k$　は

$$\eta_2{}^0 = \frac{1}{\sqrt{\Delta_{11}}}(k - r_{23}x_3) \tag{9}$$

となるから，相関係数が ρ であるとき $\eta_1 = \eta_1{}^0$，$\eta_2 = \eta_2{}^0$ によつて分割した四分相関係数の表を利用して，$\eta_1 \geqq \eta_1{}^0$，$\eta_2 \geqq \eta_2{}^0$ の部分の比率 $p(x_3)$ を読みとる．



そうすれば

$$x_1 \geqq h,\ x_2 \geqq k,\ x_3 \geqq l \tag{10}$$

の部分の比率は

$$\int_l^\infty p(x_3)\varphi(x_3)dx_3 \tag{11}$$

を計算して求めることができる．

実際の計算には，(11) を Simpson 法則によつて近似計算してやればよい．

例．$r_{12} = 0.4$，$r_{13} = 0.5$，$r_{23} = 0.6$，$P_r(x_1 \geqq h) = A$，$P_r(x_2 \geqq k) = B$，$P_r(x_3 \geqq l) = C$ とおくとき次のような結果が得られる．

$C=0.7$ のとき

A＼B	0.3	0.5	0.7
0.3	0.139	0.196	0.237
0.5	0.198	0.293	0.364
0.7	0.242	0.369	0.469

$C=0.3$ のとき

A＼B	0.3	0.5	0.7
0.3	0.102	0.133	0.149
0.5	0.137	0.182	0.209
0.7	0.158	0.215	0.258

これらの数値はもとの四分相関係数表のグラフから読みとつたので，ごく概略値（また $x_i \geqq 3$ 以上は近似的に $\varphi = 0$ と考えて計算しなかつたので実際より小さい値となつている．このため最終選択比率を少なめに即ち安全に見積ることができる）にすぎない．しかし 3 変数を利用して予備選択をしたときの最後に残される人数を推定するには十分役立つものである*．

以上を 4 変数に拡張することは容易である．ただし 3 変数についての数表が完成しているものとしての話である．これは r の組合せと，A，B，C の組合せとで尨大なものとなろう．

前と同様に　$x_4 = \text{const.}$　で相関曲面を截り，

$$\Delta_{1i}x_1 + \Delta_{2i}x_2 + \Delta_{3i}x_3 + \Delta_{4i}x_4 = 0 \qquad (i = 1,\ 2,\ 3) \tag{12}$$

を満足する点 $(x_1{}^0, x_2{}^0, x_3{}^0)$ に原点を移動し，新座標系を $O'\text{-}\xi_1\xi_2\xi_3$ とする．ここで勿論 Δ_{ij} は

* 詳しい数表は人事院任用局試験第二課の高橋達郎氏によって計算され，試験研究第 17 号に掲載されている．

$$\Delta = \begin{vmatrix} 1 & r_{12} & r_{13} & r_{14} \\ r_{12} & 1 & r_{23} & r_{24} \\ r_{13} & r_{23} & 1 & r_{34} \\ r_{14} & r_{24} & r_{34} & 1 \end{vmatrix} \tag{13}$$

の i, j 要素の余因数である．

また

$$\Delta^* = \begin{vmatrix} 1 & r_{12.4} & r_{13.4} \\ r_{12.4} & 1 & r_{23.4} \\ r_{13.4} & r_{23.4} & 1 \end{vmatrix} \tag{14}$$

とおき，この i, j 要素の余因数を Δ^*_{ij} とおく．ここで

$$\sqrt{\frac{\Delta_{11}}{\Delta}}\xi_1 = \sqrt{\frac{\Delta_{11}^*}{\Delta^*}}\eta_1 \tag{15}$$

などの変換式によつて $O'-\eta_1\eta_2\eta_3$ 系に変換するとき，$x_4 = \mathrm{const.}$ による截口の密度函数は

$$\frac{1}{(2\pi)^{\frac{3}{2}}\sqrt{\Delta^*}}\exp\left(-\frac{1}{2\Delta^*}\sum_i\sum_j \Delta^*_{ij}\eta_i\eta_j\right) \tag{16}$$

となり，$x_1 = h$, $x_2 = k$, $x_3 = l$ は夫々

$$\left.\begin{aligned} \eta_1^0 &= \frac{1}{\sqrt{1-r_{14}^2}}(h-r_1\,x_4) \\ \eta_2^0 &= \frac{1}{\sqrt{1-r_{14}^2}}(k-r_{24}x_.) \\ \eta_3^0 &= \frac{1}{\sqrt{1-r_{14}^2}}(l-r_{34}x_4) \end{aligned}\right\} \tag{17}$$

に変換される．

従つて $r_{12.4}$, $r_{13.4}$, $r_{23.4}$ を3変数の場合の相関係数に相当するものと考え，$\eta_1 \geqq \eta_1^0$, $\eta_2 \geqq \eta_2^0$, $\eta_3 \geqq \eta_3^0$ に含まれる比率 $p(x_4)$ を前に作成した数表から読取ることができる．これを用いて

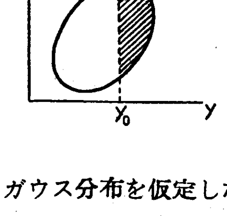
$$\int_m^\infty p(x_4)\varphi(x_4)dx_4 \tag{18}$$

を計算すれば，求める　$x_1 \geqq h$, $x_2 \geqq k$, $x_3 \geqq l$, $x_4 \geqq m$　に属する部分の比率を得る．

§4.　集団の不均質性の影響

予備選択によつて一部分の受験者がふるい落されるとき，残りのものについてのテストの得点の平均，分散，相関係数などへの影響を考えるのが目的である．ここで取上げるのは特に2変数 y, z の場合である．例えば y は択一式のテストの得点，z は記述式のテストの得点という場合に当る．

もし y, z の同時分布が2次元のガウス分布であると仮定できれば，$y \geqq y_0$ 内における相関係数 r, 標準偏差 s_y, s_z, 全集団における標準偏差 σ_y が判つているとき，全集団における相関係数 ρ, 標準偏差 σ_z は



$$\rho = \frac{r\sigma_y/s_y}{\sqrt{1-r^2+r^2(\sigma_y^2/s_y^2)}} \tag{1}$$

$$\sigma_z = s_z\sqrt{1-r^2+r^2(\sigma_y^2/s_y^2)} \tag{2}$$

によつて与えられる．(H. Gulliksen: Theory of Mental Tests, 1950, New York, John Wiley & Sons Inc.)

ガウス分布を仮定しない代りに，線形回帰を仮定してもよいし，部分と全体に於ける回帰係数が同一で，かつ一方の変数から他方を推定したときの誤差分散が同一ということを仮定しても同じで

ある．

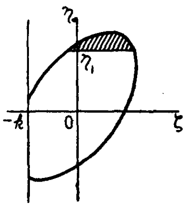
ところが z の分布がガウス分布でない場合にはどのようにしたらよいかというのが我々の問題である．そこで z の分布を調べてみると，trunkated normal に似ている．そうすると，K. Pearson, A. Lee, A.C, Cohen 等の論文* によつて，truncation の行われない場合の $\bar{z}$, σ_z を推定することができる．従つて前述の方法と，この方法を結合すればよいことになる．

y, z は T スコアで表わされているとき，η, ζ を夫々対応する z スコア（平均 0, 分散 1）とする．このとき η, ζ の同時分布の密度函数は

$$f(\eta,\zeta)=\frac{1}{A}\frac{1}{2\pi\sqrt{1-\rho^2}}\exp\left[-\frac{1}{2(1-\rho^2)}(\eta^2-2\rho\eta\zeta+\zeta^2)\right] \qquad (3)$$

但し

$$A=\int_{-\infty}^{\infty}\int_{-k}^{\infty}\frac{1}{2\pi\sqrt{1-\rho^2}}\exp\left[-\frac{1}{2(1-\rho^2)}(\eta^2-2\rho\eta\zeta+\zeta^2)\right]d\eta d\zeta \qquad (4)$$



$-k$ は truncated point の座標，ρ は truncation がないとしたときの相関係数である．

いま ζ の η に対する回帰直線を

$$\zeta=\alpha\eta+\beta \qquad (5)$$

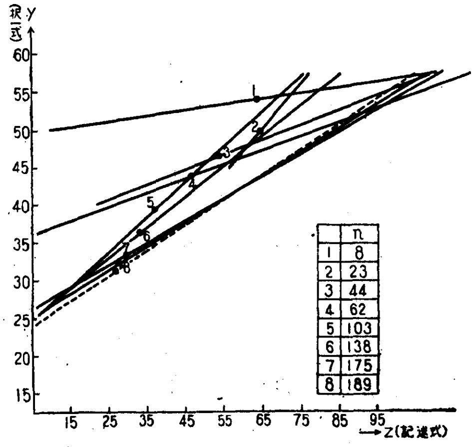
とおくと，η と ζ の全域（truncated）における相関係数 R を用いて

$$\alpha=\frac{R\sigma_\zeta}{\sigma_\eta},\qquad \beta=\bar{\zeta}-\alpha\bar{\eta} \qquad (6)$$

但し

$$\left.\begin{aligned}\bar{\eta}&=\int_{-\infty}^{\infty}\int_{-k}^{\infty}\eta f(\eta,\zeta)d\eta d\zeta,\quad \bar{\zeta}=\int_{-\infty}^{\infty}\int_{-k}^{\infty}\zeta f(\eta,\zeta)d\eta d\zeta\\ \sigma_\eta{}^2&=\int_{-\infty}^{\infty}\int_{-k}^{\infty}\eta^2 f(\eta,\zeta)d\eta d\zeta-\bar{\eta}^2\\ \sigma_\zeta{}^2&=\int_{-\infty}^{\infty}\int_{-k}^{\infty}\zeta^2 f(\eta,\zeta)d\eta d\zeta-\bar{\zeta}^2\\ R\sigma_\eta\sigma_\zeta&=\int_{-\infty}^{\infty}\int_{-k}^{\infty}\eta\zeta f(\eta,\zeta)d\eta d\zeta-\bar{\eta}\bar{\zeta}\end{aligned}\right\} \qquad (7)$$

従つて図の如く $\eta\geqq\eta_1$ に於いて推定した回帰直線の勾配 ρ は真の勾配 α と異るのである．実例について種々の η_1 に対して回帰直線の勾配の変化を図示すると右図のようになる．



そこで実際に z, σ_z を推定するには次のような手続をふめばよい．先づ

$$\eta=\frac{y-y_0}{c},\qquad \zeta=\frac{z-z_0}{d} \qquad (8)$$

とおく．ここで y_0, z_0, c, d は truncation のないときの平均，標準偏差である．

（i）k の推定が行えたとする．(これは $y<y_1$ なるサンプルを要するので，結局は，$\bar{z}$ σ_z が判つていることと同じである．しかしこ

* K. Pearson and A. Lee, "On the generalized probable error in multiple nolmal correlation", Biometrika, vol. 6, 1908. A. Lee, "Table of Gaussian 'tail' functions when the 'tail' is larger than the body", Biometrika, vol. 10, 1915. A.C. Cohen, J.R. "On estimating the mean and standard deviation of truncated normal distributions", J.A.S.A. vol. 44, 1949.

こではこのようなサンプリングを行わず前回の結果から推定し得られるものとする).

(ii) $y \geqq y_1$ なる領域において s_y, s_z, r_{yz} を求める.

$$\left.\begin{aligned} \rho &= \frac{r_{yz}c/s_y}{\sqrt{1-r^2_{yz}+r^2_{yz}(c^2/s_y{}^2)}} \\ \sigma_\eta{}^2 &= \frac{\sigma_y{}^2}{c^2} = 1-\frac{\rho^2}{A^2}\{\varphi(k)\}^2-\frac{\rho^2}{A}k\varphi(k) \end{aligned}\right\} \quad (9)$$

但し
$$\varphi(k)=\frac{1}{\sqrt{2\pi}}e^{-\frac{k^2}{2}}$$

より ρ, c を求める. 実際には

$$B=\frac{k}{A}\varphi(k)+\left\{\frac{\varphi(k)}{A}\right\}^2$$

とおいて

$$(B-1)r_{yz}{}^2c^4+c^2(\sigma_y{}^2r_{yz}{}^2+s_y{}^2r_{yz}{}^2-s_y{}^2)+\sigma_y{}^2s_y{}^2(1-r_{yz}{}^2)=0$$

を解けばよい.

(iii) $\bar{\eta}=\dfrac{\bar{y}-y_0}{c}=\dfrac{\rho}{A}\varphi(k)$ より y_0 を求める.

(iv) $\alpha=\dfrac{\rho(1-B)}{1-\rho^2B}$ より α が判る.

(v) $d=\dfrac{c}{\rho}\dfrac{r_{yz}s_z}{s_y}$ より d を求める.

(vi) $y \geqq y_1$ に於ける回帰直線と $y=y_0$ の交点より z_0 を求める.

(vii) $\bar{\zeta}=\dfrac{\bar{z}-z_0}{d}=\dfrac{1}{A}\varphi(k)$ より $\bar{z}$ を求める.

(viii) $\zeta=\alpha\eta+\beta$ は $z=\alpha\dfrac{d}{c}(y-\bar{y})+\bar{z}$

(ix) $\sigma_\zeta{}^2=\dfrac{\sigma_z{}^2}{d^2}=1-B$ より $\sigma_z{}^2=(1-B)d^2$

以上の方法は k についての推定に大きく左右される. 従つて実際問題としては余り適当な方法ではない. むしろ短刀直入に $\bar{z}$, $\sigma_z{}^2$ をサンプリングによつて求めた方が, 採点の費用さえ許せば, 精確であろう.

サンプリングでやる場合は, 上位群: $y \geqq y_1$ の人数を N_1 人, 下位群: $y<y_1$ の人数を N_2 人とするとき, 下位群より n_2 人のサンプルを抽出して

$$\hat{\bar{Z}}=\frac{1}{N}(N_1\bar{z}_1+N_2\bar{z}_2) \quad (10)$$

$$\hat{\sigma}_z{}^2=\frac{1}{N}\left(N_1\sigma_{1z}{}^2+(N_2-1)\frac{n_2}{n_2-1}s^2_{2z}\right)+\frac{N_1N_2}{N^2}\left\{(\bar{z}_1-\bar{z}_2)^2-\frac{N_2-n_2}{N_2-1}\frac{s_{2z}{}^2}{n_2-1}\right\} \quad (11)$$

によつて推定する.

このとき (10), (11) は共に $\bar{z}$, $\sigma_z{}^2$ の不偏推定値であつて, 有限母集団修正を無視すれば

$$D^2(\hat{\sigma}_z{}^2) \doteqdot \frac{1}{n_2}\left[\frac{N_2{}^2}{N^2}(\mu_4(2)-\sigma_{2z}{}^2)+\frac{4N_1{}^2N_2{}^2}{N^4}\sigma_{2z}{}^2(\bar{Z}_1-\bar{Z}_2)^2+\frac{4N_1N_2{}^2}{N^3}\mu_3(2)(\bar{Z}_2-\bar{Z}_1)\right] \quad (12)$$

但し $\mu_i(2)$ は下位集団における i 次の中心モーメント.

例. $N_1=44$, $N_2=154$, $n_2=20$ のあるテストの結果についてみると右表のようになり, サンプリングによる方法の方がずつとよいこと

	$\bar{Z}$	$\sigma_z{}^2$
実際の値	26.24	418.14
k を用いる推定法	25.08	540.72
サンプリングによる推定法	25.67	418.94

が看取される.

§5. 教員の定員の決定

現在教員定数の算定基準は政令106号によつて都道府県単位として次のように定められている.

小学校		$\left(\text{学級数} \times \dfrac{13}{12} + \text{学校数}\right) \times 1.03$
中学校		$\left(\text{学級数} \times \dfrac{13}{9} + \text{学校数}\right) \times 1.03$
高等学校	300人以下のとき	$\dfrac{\text{生徒数}}{50} \times \text{週当り授業時数} \times \dfrac{1}{12}$
	300~750人	$\dfrac{\text{生徒数}}{50} \times \text{週当り授業時数} \times \dfrac{1}{15}$
	750人以上	$\dfrac{\text{生徒数}}{50} \times \text{週当り授業時数} \times \dfrac{1}{18}$

一般に定数算定の基準として考えられるべき要素は, 学校数, 学級数, 児童生徒数の3つがあるが, 小, 中学校では児童生徒数の考慮がされていない. 従つて1学級当りの生徒数を減少させることによつて定数増加を図る懼れもあり, 各都道府県によつて学級編成もまちまちとなつている. また一方分校のような小規模の学校を各地に散在させざるを得ない地理的条件下にある府県では1学級当り生徒数の減少も已むを得ない. そこでこれを理論学級という立場から規定していくものとして取扱つてみよう. 例えば各都道府県毎に次表のような場合を考えると, (生徒数 z_i には学級数が

生徒数	中央値 (z)	学校数 $p(z)$	理論学級数 $Y=g(z)$
$z_1 - h \sim z_1 + h$	z_1	x_1	a_1
$x_1 + h \sim x_1 + 2h$	z_2	x_2	a_2
$z_1 + 2h \sim x_1 + 3h$	z_3	x_3	a_3
....	⋮	⋮	⋮
計		x	

$g(z_i) = a_i$ が適当であると定める) 平均1学校当り生徒数 $\bar{z}$ は

$$\bar{z} = \sum_i p_i z_i \qquad \text{但し} \qquad p_i = x_i/x \tag{1}$$

となる.

各都道府県毎に生徒数の分布 $p(z)$ を求めれば定数はきまるので問題は解決されて了うのであるが, 一々この分布を調査によつて求めることは面倒なことが多い. そのような場合においての考え方が本節の主眼とするところである.

さてこの場合 $\bar{z}$ のみを調査することは容易であることに注目して, これを用いて理論学級数を求める方法を考えればよい. 1理論学級当り生徒数は

$$\eta = \frac{\bar{z}}{a_1 p_1 + a_2 p_2 + \cdots\cdots}$$

と書けるから, これを求めておき, 生徒総数を η で除して学級総数を求めるようにすればよい. 所がこの式の形から判る如く, $p(z)$ の分布の形によつて, 同一の $\bar{z}$ であつても η が異つてくるから, その変動の限界を調べておかなければならない.

かくて (1) が一定という条件の下に

$$\sum_i a_i p_i \tag{2}$$

の max., min. を求めておけば η の max. min. が判る. これは linear programming の特殊の場合に当る.

（1）より
$$\sum_i p_i z_i/\bar{z} = 1 \tag{3}$$
いまベクトル $\mathfrak{a}_i$ を
$$\mathfrak{a}_i = \begin{pmatrix} \frac{z_i}{\bar{z}} \\ a_i \end{pmatrix} \tag{4}$$
とおくと，$\sum \mathfrak{a}_i p_i$ の表わす端点は $\mathfrak{a}_i$ の端点が作る convex polyhedron の内点となる．従つて（4）のベクトルで丁度 $\begin{pmatrix} 1 \\ c \end{pmatrix}$ となるものをとり，c の max., min. が求める（2）の max., を与えるのであるが，これは図の1を通る縦線と，多角形の辺との交点によつて与えられる．

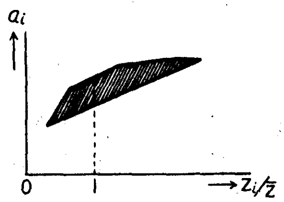


圖 1

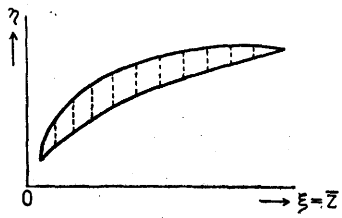


圖 2

このようにして各都道府県に対する $\xi = \bar{z}$ に対して1理論学級当り生徒数
$$\eta = \frac{\int z p(z) dz}{\int g(z) p(z) dz} = \frac{\text{生徒総数}}{\text{理論学級数}}$$
を巾をもたせて定めておくことができる（勿論法律的には η を一定値にしておく）．かくて各都道府県の教員は
$$\text{（その都道府県の生徒数）} \div \eta \tag{5}$$
によつて決定することができる．

統計数理研究所