偏差値，折れ線換算法および算術平均

永　谷　　彬

T-score, method of broken line conversion and arithmetic mean

Hitoshi NAGAYA

Some methods of conversion are regarded as monotone increasing real valued functions of a real variable. "T-score" is an example of such methods of conversion, and it is an application of the standard Gaussian distribution. We comment on them, and introduce the method of broken line conversion. Next, we introduce two methods of conversion; middle point conversion and initial point conversion, and give some applications of them.

序　文

度量衡や金額について，それらの単位を換える時の換算はよく知られている。また，現在流行の各種試験における得点を偏差値で表すのも換算の一つである。これらは，１実変数実数値函数（real valued function of a real variable）のうちの狭義単調増加函数（strictly monotone increasing function）といわれているものである。

すなわち，実数全体の集合を R とするとき，次の如き f を，S における狭義単調増加函数という。

$S\subset R$，$f: S\to R$，$x\in S$ に対して，$f(x)\in R$ かつ，$x_1, x_2\in S$ のとき，$x_1<x_2$ ならば，$f(x_1)<f(x_2)$ である。

例　$S\subset R$，$I_S: S\to R$，$x\in S$ に対して，$I_S(x)=x$ とする。この I_S は S における狭義単調増加函数であり，これを S における恒等函数ということにする。

上の第２番目の不等号＜を≦で書き直したときの f のことを単調増加函数（monotone increasing function）という。すなわち次のような f のことである。

$S\subset R$，$f: S\to R$，$x\in S$ に対して $f(x)\in R$，かつ x_1，$x_2\in S$ のとき，$x_1<x_2$ ならば $f(x_1)\leqq f(x_2)$ である。

函数が狭義単調増加ならば，単調増加である。

単調増加であるが，しかし狭義単調増加ではない函数による換算の例として，四捨五入法と端数切り捨て法がある。

この小論で扱う換算とは，単調増加函数のことをいう。ゆえに上記の恒等函数は一つの換算である。これを恒等換算ということにする。

２つの換算 $f_j: I_j\to R$，$j=1, 2$，に対して，$f_1(I_1)\subset I_2$，$f_1(I_1)=\{f(x)\mid x\in I_1\}$，なるとき，$f_1$ と f_2 の合成函数 $f_2\circ f_1: I_1\to R$，$(f_2\circ f_1)(x)=f_2(f_1(x))$，$x\in I_1$，もやはり換算である。

§１　ではガウス分布と偏差値の解説を行う。

§２　では折れ線換算法について論ずる。

１クラスを２つ以上に分割して，同一科目の授業を複数の教員が分担して行い，その試験における得点と他の科目の得点を加算する時，または，学校例えば大学における試験や入学試験，あるいは就職試験や資格認定試験等において，選択科目についての得点を，他の科目の得点と加算する時に役立てば幸いである。

§１や§２の換算法による換算値（理論値）を，通常，整数または小数点以下の部分が有限個であるような数値として算出しなければならない。その際の換算法として，四捨五入法と切り捨て法がよく用いられる。

§３　では，中点平均評価とその応用について記す。その中には四捨五入法がある。

§４　では，始点平均評価とその応用の１つとして切り捨て法について述べる。

この小論は，対象の一部分を調査した情報に依り，対象全体の情報を推測する方法に関するいわゆる推測統計学すなわち推計学（stochastics）に属するものではない。

小論作製の初期段階において，著者に対して援護的に意見を交換して下さった，同僚の樋口功氏，橋本有司氏，隅山孝夫氏に感謝の意を表する。

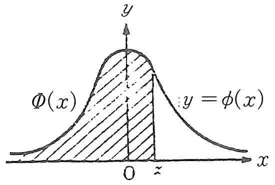
§１　標準ガウス分布と偏差値

（1. A）　標準ガウス分布[1)]

標準ガウス分布（standard Gaussian distribution）またの名を標準正規分布（standard normal distribution）というのは，密度函数（density function）$\phi(x)$，分布

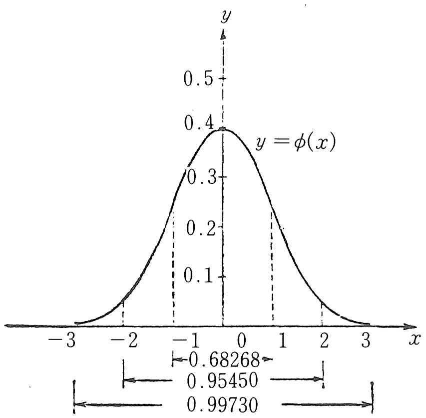
函数（distribution function）$\Phi(x)$ が次の様なものである。

$$\phi(x)=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-x^2/2},\quad \Phi(x)=\int_{-\infty}^{x}\phi(t)\,dt,\quad \Phi(\infty)=1,$$



x	$\phi(x)$	$\Phi(x)$	x	$\phi(x)$	$\Phi(x)$
0.0	0.39894	0.50000	2.0	0.05399	0.97725
0.1	0.39695	0.53983	2.1	0.04398	0.98214
0.2	0.39104	0.57926	2.2	0.03547	0.98610
0.3	0.38139	0.61791	2.3	0.02833	0.98928
0.4	0.36827	0.65542	2.4	0.02239	0.99180
0.5	0.35207	0.69146	2.5	0.01753	0.99379
0.6	0.33322	0.72575	2.6	0.01358	0.99534
0.7	0.31225	0.75804	2.7	0.01042	0.99653
0.8	0.28969	0.78814	2.8	0.00792	0.99744
0.9	0.26609	0.81594	2.9	0.00595	0.99813
1.0	0.24197	0.84134	3.0	0.00443	0.99865
1.1	0.21785	0.86433	3.1	0.00327	0.99903
1.2	0.19419	0.88493	3.2	0.00238	0.99931
1.3	0.17137	0.90320	3.3	0.00172	0.99952
1.4	0.14973	0.91924	3.4	0.00123	0.99966
1.5	0.12952	0.93319	3.5	0.00087	0.99977
1.6	0.11092	0.94520	3.6	0.00061	0.99984
1.7	0.09405	0.95543	3.7	0.00042	0.99989
1.8	0.07895	0.96407	3.8	0.00029	0.99993
1.9	0.06562	0.97128	3.9	0.00020	0.99995
2.0	0.05399	0.97725	4.0	0.00013	0.99997

$$\frac{1}{\sqrt{2\pi}}\int_{-k}^{k}e^{-x^2/2}\,dx=\Phi(k)-\Phi(-k)\fallingdotseq\begin{cases}0.68268, & k=1\\ 0.95450, & k=2\\ 0.99730, & k=3\end{cases}$$



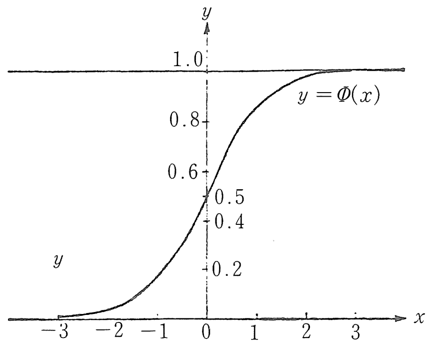


表1

Xが密度函数 $\phi(x)$ をもつ確率変数（random variable）であるとき，Xの1次変換による確率変数 $Y=\sigma X+\mu$ は平均 μ，分散 σ^2（すなわち標準偏差 σ）をもち，その密度函数 $g(x;\mu,\sigma^2)$ は次で与えられる：

$$g(x;\mu,\sigma^2)=\frac{1}{\sigma\sqrt{2\pi}}\mathrm{e}^{-(x-\mu)^2/2\sigma^2},$$

$$-\infty<x<\infty,\quad \sigma>0,\quad -\infty<\mu<\infty,$$

すなわち，Y の分布函数 $G(a;\mu,\sigma^2)$ は次で与えられる：

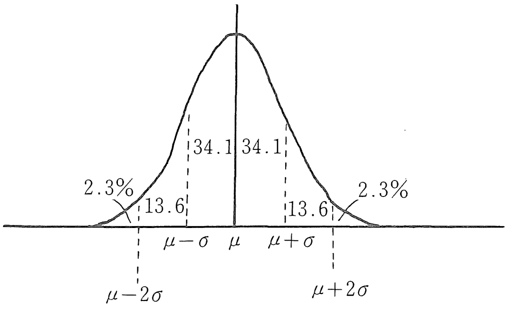
$$G(a;\mu,\sigma^2)=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{a}\mathrm{e}^{-(x-\mu)^2/2\sigma^2}\,dx$$

この積分において変数変換 $y=(x-\mu)/\sigma$ を行うことにより，G と Φ は次の関係で結ばれる：

$$G(a;\mu,\sigma^2)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{(a-\mu)/\sigma}\mathrm{e}^{-y^2/2}\,dy$$

$$=\Phi((a-\mu)/\sigma)\quad\cdots\cdots\cdots\cdots\cdots\cdots(*)$$

このYの分布を平均 μ，分散 σ^2（すなわち標準偏差 σ，$\sigma>0$）のガウス分布または正規分布といって，$N(\mu,\sigma^2)$ で表す。したがって標準ガウス分布とは，$N(0,1)$ のことである。$N(\mu,\sigma^2)$ の密度函数 $g(x;\mu,\sigma^2)$ のグラフは，$N(0,1)$ の密度函数 $\phi(x)$ のグラフを，横軸方向に μ だけ平行移動した上で，さらに横軸方向に σ 倍し，かつ縦軸方向を $1/\sigma$ 倍して得られる（図甲)。また $N(\mu,\sigma^2)$ の分布函数 $G(a;\mu,\sigma^2)$ のグラフは，$N(0,1)$ の分布函数 $\Phi(x)$ のグラフを横軸方向に μ だけ平行移動した上で，さらに横軸方向に σ 倍して得られる。



図甲

ガウス分布が統計学（statistics）において重要視される理由はいろいろあるが，その例を列挙する。

(1)　推計学で用いられる統計量（statistic）の分布で，標本（sample）の大きさがある程度大きくなると，ガウス分布によって近似されるものが多くある。しかし，このことは，ガウス分布あるいはガウス分布によって近似されるもの以外の分布が少ないというわけではない。[2)]

(2)　上記(1)と関連して，理論的分布の極限としてガウス分布が出て来ることが多い。

(3)　測定誤差の分布に関しては，誤差論（theory of errors）におけるガウスの誤差函数（error function）すなわち次式で与えられる函数 erf または erfc が，最少2乗法との関連において主役を演ずる。

$$\mathrm{erf}\ x = \mathrm{erfc}\ x = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}\,dt$$

ゆえに，

$$\mathrm{erf}\ x = \sqrt{2}\,(\Phi(x) - \Phi(-x))$$

歴史的には，ここで述べた通りガウスによって見出された N(0，1) を後続の人達が標準ガウス分布と名付け，それを一般化した $N(\mu,\sigma^2)$ を正規分布と名付けたものと思われる。[3)]

(4)　観測値（observed value）に対する母集団（population），すなわち着目している対象の全体，の分布としてガウス分布が想定できる場合が多くある。

(5)　観測値に簡単な変換を行なったものにガウス分布が想定出来る場合がある。

(6)　各種の試験の得点の分布は，ガウス分布に従うものと，ほとんど断定した状態で使われているものとして，いわゆる五段階評価がある。そしてこの場合，残念ながら往々にして誤用，悪用を見る。

推計学の観点より，ガウス分布を利用するに当たって，確率論（theory of probability）で得られた最有力な結果の一つは，いわゆる中心極限定理（central limit theorem）であろう：

確率変数 X_j，$j \geqq 1$，は独立で，平均 μ，分散 σ^2，$\sigma > 0$，をもつ同一分布に従うものとする。 $S_n = X_1 + X_2 + \cdots\cdots + X_n$ とするとき，$(S_n - n\mu)/(\sigma\sqrt{n})$ の分布函数は，$n \to \infty$ のとき，N(0，1) の分布函数に収束する。すなわち，N(0，1) の分布函数 $\Phi(x)$ に対して，次が成り立つ：

$$\lim_{n\to\infty} P_r\left\{\frac{S_n - n\mu}{\sigma\sqrt{n}} \leqq x\right\} = \Phi(x),$$

ここで，$P_r\{\cdot\}$ は“・”が成り立つ確率を表す。

推計学で利用するには，次の様にいい直す。

（CLT）平均 μ，分散 σ^2（すなわち標準偏差 σ，$\sigma > 0$）をもつ分布からの大きさ n の標本の標本平均を $\overline{X}$ とするとき，$(\overline{X} - \mu)/(\sigma/\sqrt{n})$ の分布函数は，$n \to \infty$ のとき，N(0，1) の分布函数に収束する。

この（CLT）における平均 μ，分散 σ^2 をもつ分布は，ガウス分布でなくてもよく，分布の型が同じであれば何でもよい。

（注意）　往々にして見られる（CLT）の誤用，悪用の原因は，“分布の型が異なるにもかかわらず，それらの分布の各々から一つずつ選んだ標本の大きさを大きくしさえすれば，その標本平均を（CTL）の $\overline{X}$ としてよい”と思い込むことにある。

（1.B）　偏　差　値

確率変数 x がガウス分布 $N(\mu,\sigma^2)$ に従うとき，x の偏差値（T－score）とは次の式で定義した $T(x) = T(x;\mu,\sigma)$ のことである[4)]：

$$T(x;\mu,\sigma) = 10 \times \frac{x-\mu}{\sigma} + 50$$

この x に対応する確率変数 t が標準ガウス分布 N(0，1) に従うならば，$x = \sigma t + \mu$ である。ゆえに $T(x) = 10t + 50$ ……………………………………（＊＊）

ゆえに，（1.A）で述べた通り確率変数 $T(x)$ はガウス分布 $N(50,\ 10^2)$ に従うことになる。

確率変数 x がガウス分布 $N(\mu,\sigma^2)$ に従うとき，$X(x;M,\ S) = X(x;M,\ S;\mu,\sigma)$ を次の式で定義し，x の一般偏差値（general T―score）ということにする。

$$X(x;M,S;\mu,\sigma) = S \times \frac{x-\mu}{\sigma} + M,\quad S > 0$$

標準ガウス分布 N(0，1) の確率変数 t により，$X(x;\ M,S) = St + M$ となるから，（1.A）で述べた通り一般偏差値 $X(x;M,S) = X(M,S)$ は $N(M,S^2)$ に従う。図甲において，$\mu = M$，$\sigma = S$ として，一般偏差値 $X(M,S)$ の分布表を得る。かくしていわゆる偏差値とは，X(50，10)のことである。ゆえに，偏差値 $T(x) = X(50,\ 10) = X(x;50,\ 10)$ の分布表は図甲において，$\mu = 50$，$\sigma = 10$ としたものである。

各偏差値 $T(x)$ が上位 $P(T(x))$％ の位置に在ることは，次の表2から分かる。この表は表1より次の様にして作製したものである。

$$P(T)=\begin{cases}(1-\Phi((T-50)/10))\times 100, & T\geqq 50,\\ 100-P(100-T), & T<50.\end{cases}$$

理由は，式(＊)，(＊＊)と図甲の対称性に依る。この式による理論値を，小数点以下第2位を四捨五入し，端数を切り捨てたものである。

表2　偏差値と序列との関係

偏差値	上位何%に位置するか	偏差値	上位何%に位置するか	偏差値	上位何%に位置するか	偏差値	上位何%に位置するか
75	0.6％	62	11.5％	49	54.0％	36	91.9％
74	0.8	61	13.6	48	57.9	35	93.3
73	1.1	60	15.9	47	61.8	34	94.5
72	1.4	59	18.4	46	65.5	33	95.5
71	1.8	58	21.2	45	69.2	32	96.4
70	2.3	57	24.2	44	72.6	31	97.1
69	2.9	56	27.4	43	75.8	30	97.7
68	3.6	55	30.8	42	78.8	29	98.2
67	4.5	54	34.5	41	81.6	28	98.6
66	5.5	53	38.2	40	84.1	27	98.9
65	6.7	52	42.1	39	86.4	26	99.2
64	8.1	51	46.0	38	88.5	25	99.4
63	9.7	50	50.0	37	90.3		

(1.C)　偏差値の性質

(I)　$P(T(x))$ は，x がガウス分布に従うとき，x の偏差値 $T(x)$ により，表2から求められた数である。このとき，x は上位から $P(T(x))$％ の位置にある。

(II)　x，y がそれぞれ $N(\mu_1, \sigma_1^2)$，$N(\mu_2, \sigma_2^2)$ に従うとき，それぞれの一般偏差値 $X(M, S)$，$Y(M, S)$ は共に $N(M, S^2)$ に従う。ゆえに x，y がそれぞれ上位何パーセントの処にあるかが分かり，この意味で x と y の双方を相対的に比較することが出来る。特に x と y の偏差値 $T(x; 50, 10)$，$T(y; 50, 10)$ は共にガウス分布 $N(50, 10^2)$ に従う。従って表2により，x と y がそれぞれ上位何パーセントのところに位置するかが分かり，双方の位置付けを相対的に比較することが出来る。例えば，入学試験において，満点が同一の値である選択科目の得点，平均，標準偏差が違っていても，これらの選択科目のそれぞれの得点がすべてガウス分布，あるいはそれに近い分布に従うならば，それぞれの得点がそれぞれの選択科目において上位何パーセントの位置に在るかを，相対的に比較するのに適している。

(III)　偏差値 $T(x)$ に他の数を加算することに意味があるか否か？　の問に対しては，次のような場合を除いて“無し”と答える。

試験を例にとって説明する。

同じメンバーによる同一科目の試験 t_j，$1\leqq j\leqq n$，の結果，いずれの場合においても得点の分布がそれぞれガウス分布 $N(\mu_j, \sigma_j^2)$ あるいはこれに近い分布に従うものとする。A君の各試験 t_j における偏差値を T_j とするとき，その算術平均を $T=(T_1+T_2+\cdots\cdots+T_n)/n$ とする。このとき，“A君のその科目の成績は，このメンバー全体の上位 $P(T)$％の位置に在る。”と消極的に判断する。

この様な場合以外において，上位からの位置を示す数としての偏差値に，他の数を加算することは，何の意味もない。

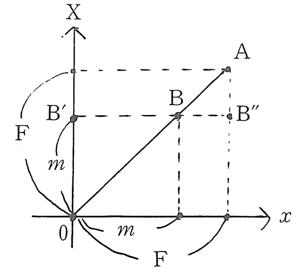
従って，複数科目の試験の得点の合計点による順位決定に際し，その為の加算に使用する数として，偏差値は何の役にも立たない。

例えば，入学試験における得点の合計を算出する必要があるとき，満点が同一である選択科目の間に著しい差があったとする。このときは，選択科目の素点を適当に換算した上で，他の科目の素点と加算することは，残念ながら止むを得ないことであろう。しかし，上記の如く，偏差値が使えないことにより，偏差値以外の換算法が必要となる。その為の一つの換算法として，折れ線換算法

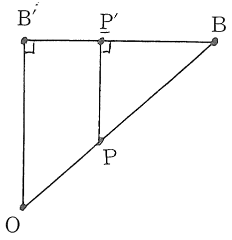
（仮称）を次の節で論ずる。

§ 2　折れ線換算法

閉区間〔0，F〕，F＞0，の恒等変換 $\mathrm{I}_{[0,\mathrm{F}]}$：〔0，F〕→〔0，F〕，をIと略記する。X＝I(x)＝ x， x∈〔0，F〕。ゆえにIのグラフは図乙1の線分OA，A＝(F，F)，で表される。OA 上の任意の点 B＝(m， m) を固定する。

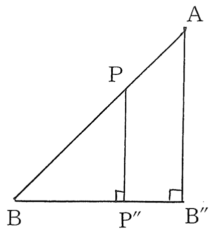


図乙.1



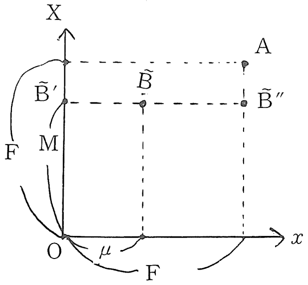
$0 < x < m$

図乙.1.1

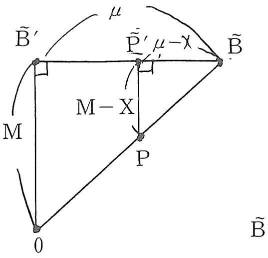


$m < x < \mathrm{F}$

図乙1.2

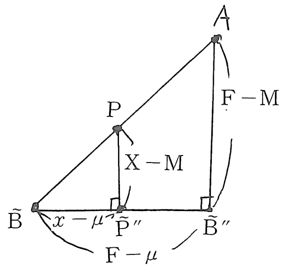


図乙.2



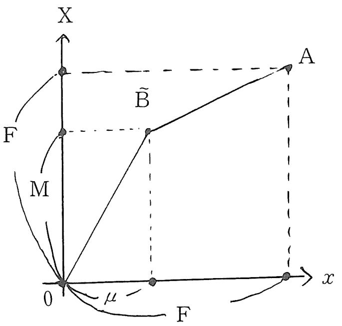
$0 < x < \mu$

図乙2.1



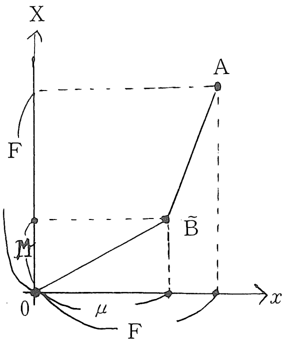
$\mu < x < \mathrm{F}$

図乙2.2



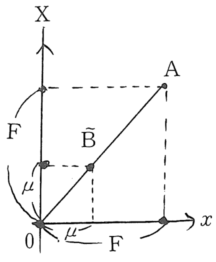
μ<M

図乙3.1



μ>M

図乙3.2



μ=M

図乙3.3

図乙

図乙1.のように2点 B′，B″ を定める。

$0 < x < m$ のときは，図乙.1.1 のような2点 P，P′ により，

$$\overline{\mathrm{PP'}}\,/\,\overline{\mathrm{OB'}} = \overline{\mathrm{BP'}}\,/\,\overline{\mathrm{BB'}} \quad \cdots\cdots(1)$$

$m < x < \mathrm{F}$ のときは，図乙1.2のような2点 P，P″ により，

$$\overline{\mathrm{PP''}}\,/\,\overline{\mathrm{AB''}} = \overline{\mathrm{BP''}}\,/\,\overline{\mathrm{BB''}} \quad \cdots\cdots(2)$$

つぎに，函数 f :〔0，F〕→〔0，F〕を次のように定義する。 μ，M∈〔0，F〕 のとき， f :(0, μ, F) $\mapsto$ (0, M, F)， 点(μ, M) を $\tilde{\mathrm{B}}$ と書く。

図乙2.のように2点 $\tilde{\mathrm{B}}'$， $\tilde{\mathrm{B}}''$ を定める。X＝ f(x) とし，点 (x，X)＝(x， f(x)) を P と書く。

$0 < x < \mu$ のときは，P を線分 O $\tilde{\mathrm{B}}$ 上にとる。図乙2.1のような2点 P， $\tilde{\mathrm{P}}'$ に対して，式(1)と同様に

$$\overline{\mathrm{P}\,\tilde{\mathrm{P}}'}\,/\,\overline{\mathrm{O}\,\tilde{\mathrm{B}}'} = \overline{\tilde{\mathrm{B}}\,\tilde{\mathrm{P}}'}\,/\,\overline{\tilde{\mathrm{B}}\,\tilde{\mathrm{B}}'} \quad \cdots\cdots(\bar{1})$$

が成り立つ。

$\mu < x < \mathrm{F}$ のときは，P を線分 $\tilde{\mathrm{B}}$A 上にとる。図乙2.2のような2点 P， $\tilde{\mathrm{P}}''$ に対して，式(2)と同様に，

$$\overline{P\bar{P}''}/\overline{A\bar{B}''}=\overline{\bar{B}\bar{P}''}/\overline{\bar{B}B''} \quad \cdots\cdots\cdots\cdots(\tilde{2})$$

が成り立つ。

$(\bar{1})$より $(M-X)/M=(\mu-x)/\mu$

$$\therefore \quad X=\frac{M}{\mu}\times x, \qquad 0<x<\mu.$$

$(\tilde{2})$より $(X-M)/(F-M)=(x-\mu)/(F-\mu)$

$$\therefore \quad X=\frac{F-M}{F-\mu}\times(x-\mu)+M, \qquad \mu<x<F.$$

このグラフは図乙3.1と図乙3.2および図乙3.3のようになる。この関数は狭義単調増加である。ゆえに単調増加である。ゆえにこの関数は換算である。この換算のグラフは，$\mu\neq M$のときは折れ線 $O\bar{B}A$ である。このことから，この換算法を折れ線換算法（method of broken line conversion）ということにする。すなわち，

$$X=\begin{cases}\dfrac{M}{\mu}\times x=\dfrac{M}{\mu}\times(x-\mu)+M, & x\leqq\mu, \\[2ex] \dfrac{F-M}{F-\mu}\times(x-\mu)+M, & \mu<x\leqq F.\end{cases} \quad \cdots\cdots(A)$$

折れ線換算法による $X=f(x)$ を $X=X(x)$ と書くことにする。

定　　理

閉区間〔0，F〕の有限個の元に着目し，その任意の一つを x と表し，x の個数を，統計学に倣って度数ということにする。それを $g(x)$ で表す。折れ線換算法 $X:[0,F]\to[0,F]$；$(0,\mu,F)\mapsto(0,M,F)$ において，μを x の算術平均とするとき，$X(x)$ の算術平均 $E(X)$ は次で与えられる。

$$E(X)=(\sum_X X\,G(X))/N$$
$$\underline{=M+(\mu-M)FR_m/(\mu(F-\mu)N)} \cdots\cdots\cdots(B)$$

ただし，

$$G(X)=G(X(x))=g(x),$$

$$N=\sum_x g(x)=\sum_X G(X),$$

$R_m=\sum_{x>\mu}(x-\mu)g(x)$，（平均$\mu$の右側1次モーメント），

$$R_m\geqq 0,$$

である。

（証　　明）

$$X=\begin{cases}X_L=M(x-\mu)/\mu+M, & x\leqq\mu, \\ X_R=(F-M)(x-\mu)/(F-\mu)+M, & \\ & \mu<x\leqq F,\end{cases}$$

とし，$\tilde{X}_L=$(X_Lの右辺の第1項)，$\tilde{X}_R=$(X_Rの右辺の第1項)，と書くことにする。

$$L_m=\sum_{x\leqq\mu}(x-\mu)g(x)\leqq 0, \qquad R_m=\sum_{x>\mu}(x-\mu)g(x)\geqq 0,$$

しかるに，

$L_m+R_m=\sum_x(x-\mu)g(x)=0$，（$\mu$のまわりの1次モーメント）．

ゆえに，　　$L_m=-R_m.$

かくして，

$$\sum_X X\,G(X)=\sum_{X_L}X_L\,G(X_L)+\sum_{X_R}X_R\,G(X_R)$$
$$=\sum_{X_L}(\tilde{X}_L+M)G(X_L)+\sum_{X_R}(\tilde{X}_R+M)\,G(X_R)$$
$$=\sum_{X_L}\tilde{X}_L\,G(X_L)+\sum_{X_R}\tilde{X}_R\,G(X_R)+M(\sum_{X_L}G(X_L)$$
$$+\sum_{X_R}G(X_R))$$
$$=\frac{M}{\mu}\sum_{x\leqq\mu}(x-\mu)\,g(x)+\frac{F-M}{F-\mu}\sum_{x>\mu}(x-\mu)\,g(x)$$
$$+M\sum_X G(X)$$
$$=\frac{M}{\mu}\times L_m+\frac{F-M}{F-\mu}\times R_m+MN$$
$$=\left(\frac{-M}{\mu}+\frac{F-M}{F-\mu}\right)R_m+MN$$
$$=\frac{(\mu-M)FR_m}{\mu(F-\mu)}+MN$$

$$\therefore\ (\sum_X X\,G(X))/N=M+(\mu-M)FR_m/(\mu(F-\mu)N)$$

（了）

系1.

等式(B)は M の1次関数であるから，M は一意的に次で表される　：

$$M=\frac{\mu((F-\mu)N\,E(X)-FR_m)}{\mu(F-\mu)N-FR_m} \quad \cdots\cdots\cdots\cdots$$

ただし，$R_m=\sum_{x>\mu}(x-\mu)\,g(x).$ $\cdots\cdots\cdots\cdots$ (C)

系2.

$\max_{0\leqq x\leqq F}|X(x)-x|=|M-\mu|$，かつ

$$|M-\mu|=\frac{\mu(F-\mu)N|E(X)-\mu|}{\mu(F-\mu)N-FR_m}\geqq|E(X)-\mu|.$$

換算手順（procedure of conversion）

(1)　$E(X)$ を指定して系1の式(C)によりMを求める。

(2)　このMを用いて式(A)により，x の換算値 $X(x)$ を計算する。

(3)　$X(x)$ の算術平均は，定理の式(B)により $E(X)$ となるはずであるが，これは理論値の範囲内のことである。

ここの(1)〜(3)の各段階において，実用的には，適当な換算法，例えば四捨五入法あるいは切り捨て法を続いて行うことになる。

$\mu=\mathrm{M}=\mathrm{E(X)}$ のとき，$\mathrm{X}(x)$ のグラフは図乙3.3において線分OAである。この換算は，序文で述べた恒等換算である。

$\mathrm{M}>\mu$，$\mathrm{M}<\mu$ のときの $\mathrm{X}(x)$ のグラフは図乙3.1と3.2において，折れ線 $\mathrm{O\tilde{B}A}$ である。

§§E(X)の指定方法の例

必須科目がm個あり，その外に選択科目のグループがk個あり，各グループ G_j の満点が F_j 点，$1\leqq j\leqq k$，であるような試験に着目する。

受験者は各 G_j において1科目ずつ選択するものとする。

必須科目のm個の試験の得点 t_i，$1\leqq i\leqq m$，と G_j の得点 S_j，$1\leqq j\leqq k$，の合計：

$$\sum_{i=1}^{m} t_i+\sum_{j=1}^{k}\mathrm{S}_j$$

の大小により，順位を決定するものとする。

一般性を失わないので，$k=1$として述べる。

選択科目同志の得点に著しい違いがあったとき，選択科目の素点（得点）を適当に換算し，その換算値をm個の必須科目の素点と合計するとき，選択科目の素点の換算に上記の折れ線換算法を利用するときの案を，以下に列挙する。

一般性を失わないから，$m=2$とする。

同じく一般性を失わないから，1グループ内の選択科目数を2とする。

必須科目を A_1，A_2 で表し，選択科目を B_1，B_2 で表す。

以下いずれの場合分け case N も，E(X) の指定の方法で行うことにする。

a_i ：A_iの素点の算術平均，$1\leqq i\leqq 2$，

b_j ：B_jの素点の算術平均，$1\leqq j\leqq 2$，

F_i ：A_iの満点，$1\leqq i\leqq 2$，

$\mathrm{F_B}$ ：B_jの共通の満点，$1\leqq j\leqq 2$，

r_1，r_2，r_3，r_4：順に a_1，a_2，b_1，b_2 の対満点比：
$\mathrm{r}_i=a_i/\mathrm{F}_i$，$1\leqq i\leqq 2$，$\mathrm{r}_{j+2}=b_j/\mathrm{F_B}$，$1\leqq j\leqq 2$.

Case I　A_i，$1\leqq i\leqq 2$ と関連づける。

（I.1）　A_i，$1\leqq i\leqq 2$，B_j，$1\leqq j\leqq 2$ のうちの1つを基準として，他の3つを換算する案：

$$\mathrm{r}=\max_{1\leqq i\leqq 4}\{\mathrm{r}_i\},\qquad \mathrm{r}=\mathrm{r}_{i_0},$$

$$\mathrm{E(X}_l)=\mathrm{crF}_l,\ 0<\mathrm{c}\leqq 1,\ 1\leqq l\leqq 4,\ l\neq i_0.$$

$$\mathrm{F}_l=\mathrm{F_B},\ l=3,\ 4.$$

（付記）　X_l，$1\leqq l\leqq 2$，はA_i，$1\leqq i\leqq 2$，の素点の換算点を表し，X_l，$3\leqq l\leqq 4$，はB_j，$1\leqq j\leqq 2$，の素点の換算点を表す。

（I.2）　A_i，$1\leqq i\leqq 2$ のうちの1つを基準として，残りの3つを換算する案：

$$\mathrm{r}=\max_{1\leqq i\leqq 2}\{\mathrm{r}_i\},\qquad \mathrm{r}=\mathrm{r}_{i_0}\ とする。$$

$$\mathrm{E(X}_l)=\mathrm{crF}_l,\ 0<\mathrm{c}\leqq 1,\ 1\leqq l\leqq 4,\ l\neq i_0.$$

$$\mathrm{F}_l=\mathrm{F_B},\ l=3,\ 4.$$

Case II　A_i，$1\leqq i\leqq 2$，と無関係。

（II.1）　B_j，$1\leqq j\leqq 2$，のうちの1つを基準として，他の1つを換算する案：

$$m=\max_{1\leqq j\leqq 2} b_j,\ m=b_{j_0}とする；$$

$$\mathrm{E(X}_l)=\mathrm{c}m,\ 0<\mathrm{c}\leqq 1,\ l\neq j_0+2,\ 3\leqq l\leqq 4.$$

（II.2）　B_j，$1\leqq j\leqq 2$，を無条件に双方とも換算する案：

$$\mathrm{E(X}_l)=a,\ 3\leqq l\leqq 4,$$

a：B_j，$1\leqq j\leqq 2$ に無関係に，全く独立に指定した数。

注意1　（I.1)，（I.2）および（II.1）における c は適当に決める。例えば，$\mathrm{c}=1$，0.95，0.9，0.8，……，etc.

c が $0<\mathrm{c}<1$ のとき，この様に選んだ理由，すなわち $\mathrm{c}\neq 1$ とした理由には，いろいろある。

一般性を失わないから，$\mathrm{c}=0.9$ のときについて述べる：

(イ)　基準とした科目の素点を既得権として尊重する。

(ロ)　受験者が選択した以上，有利か不利かの判断力ないし着眼力を無視出来ない。

(ハ)　問題作製側に対する反省を促す。

この外にも，その意味付けがあるものと思われる。

注意2　（II.1）のときの欠陥：

$\min_{1\leqq i\leqq 2}\{a_i\}>m$ のとき，その差が著しいときには，A_i，$1\leqq i\leqq 2$，の素点により順位が決められかねない。

反対に $\max_{1\leqq i\leqq 2}\{a_i\}<\mathrm{c}m$ のとき，その差が著しいときには，B_j，$1\leqq j\leqq 2$，の素点により順位が決められかねない。

（II.2）のときについても同様の考察が出来る。

注意3　$\mathrm{r}=\max_{1\leqq i\leqq 4}\mathrm{r}_i$，$\mathrm{r}=\mathrm{r}_{i_0}$，とする。$\mathrm{r}_{i_0}$と$\mathrm{r}_i$，$i\neq i_0$，の差が著しいとき，$\mathrm{r}_{i_0}$を基準にして，他の科目の素点の換算点を引き上げることが適切になることもあろう。その為には，（I.1）または（I.2）が適している。

例1.

$$\max_{1\leqq i\leqq 2}\mathrm{F}_i=300,\qquad \mathrm{F_B}=200\ とする。$$

$$\mathrm{F}_{i_0}=300,\qquad i_0=1\ または\ 2，とする。$$

$a_{i_0} \fallingdotseq 0.4\times \mathrm{F}_{i_0}=120$, $b_\mathrm{B}=\min\limits_{1\leq j\leq 2} b_j \fallingdotseq 0.6\times \mathrm{F}_\mathrm{B}=120$,

となるようなことも時には起こるであろう。このとき，$a_{i_0}/b_\mathrm{B}\fallingdotseq 1$，したがって，$\mathrm{F}_{i_0}=300$，$\mathrm{F}_\mathrm{B}=200$ としたことが無意味となる。このようなときには，（Ⅰ.1)が適当と思われる。

例2.

$\min\limits_{1\leq i\leq 2} \mathrm{F}_i=200$,　　$\mathrm{F}_\mathrm{B}=200$ とする。

$\mathrm{F}_{i_0}=200$,　　$i_0=1$ または 2，とする。

$a_{i_0}/b_j\fallingdotseq 1$ が理想的であるが，$a_{i_0}\fallingdotseq 0.6\mathrm{F}_{i_0}=120$,

$\min\limits_{3\leq j\leq 4} \mathrm{r}_j\fallingdotseq 0.4$，$\max\limits_{3\leq j\leq 4} \mathrm{r}_j\fallingdotseq 0.5$，とすると，

$\min\limits_{1\leq j\leq 2} b_j\fallingdotseq 0.4\times \mathrm{F}_\mathrm{B}=80$

$\max\limits_{1\leq j\leq 2} b_j\fallingdotseq 0.5\times \mathrm{F}_\mathrm{B}=100$

となるようなことも時には起こるであろう。このような時には（Ⅰ.2）が適当と思われる。

例3.

F＝200 なる選択科目A，Bについて，（Ⅱ.2)を利用した例を挙げる。

μ_A，μ_B　：各々A，Bの素点の平均，

$\mathrm{N_A}$，$\mathrm{N_B}$　：各々A，Bの受験者数，

$\mathrm{M_A}$，$\mathrm{M_B}$　：折れ線換算法により，系1の式(C)により算出した各々A，Bに対するMの値，

$\mathrm{R_{m,A}}$, $\mathrm{R_{m,B}}$　：各々A，Bの平均 μ_A，μ_B の右側1次モーメント，

$\mathrm{X_A}$，$\mathrm{X_B}$　：各々A，Bの素点の換算点，

$\tilde{\mathrm{X}}_\mathrm{A}$，$\tilde{\mathrm{X}}_\mathrm{B}$　：各々 $\mathrm{X_A}$，$\mathrm{X_B}$ の小数点以下第1位を四捨五入し，端数を切り捨てた整数値，

$\mathrm{E(X_A)=E(X_B)}=a=80$,　　$\mu_\mathrm{A}=50$,　　$\mu_\mathrm{B}=90$,

$\mathrm{N_A=N_B}=10$,

表3

素点 x	10	20	30	40	50	60	70	80	
換算点$\tilde{\mathrm{X}}_\mathrm{A}$	18	36	55	73	91	98	105	113	小数第1位四捨五入
人数$\mathrm{g_A}$	1	1	1	1	1	2	1	2	

$$\mathrm{R_{m,A}}=\sum_{x>50}(x-\mu_\mathrm{A})\,\mathrm{g_A}(x)$$

$$=(60-50)\times 2+(70-50)\times 1+(80-50)\times 2$$

$$=100,$$

$$\therefore\quad \mathrm{M_A}=\frac{50\times(80\times(200-50)\times 10-200\times 100)}{50\times(200-50)\times 10-200\times 100}$$

$$=\frac{5\times 10^6}{55\times 10^3}=\frac{10^3}{11}=90.\dot{9}\dot{0},$$

∴　換算式：

$$\mathrm{X_A}=\begin{cases}\dfrac{90.9}{50}\times x, & x\leq 50,\\[2ex] \dfrac{200-90.9}{200-50}\times(x-50)+90.9 & \\[2ex] =\dfrac{109.1}{150}\times(x-50)+90.9, & x>50.\end{cases}$$

この式で算出して，上の表3を得る。この表から，実際には，$\mathrm{E}(\tilde{\mathrm{X}}_\mathrm{A})=80.0$ となる。

表4

素点 x	50	60	70	80	90	100	110	120	
換算点$\tilde{\mathrm{X}}_\mathrm{B}$	43	52	60	69	77	89	100	111	小数第1位四捨五入
人数$\mathrm{g_B}$	1	1	1	1	1	2	1	2	

$$\mathrm{R_{m,B}}=\sum_{x>90}(x-\mu_\mathrm{B})\,\mathrm{g_B}(x)$$

$$=(100-90)\times 2+(110-90)\times 1+(120-90)\times 2$$

$$=100,$$

$$\mathrm{M_B}=\frac{90\times(80\times(200-90)\times 10-200\times 100)}{90\times(200-90)\times 10-200\times 100}$$

$$=\frac{612\times 10^4}{79\times 10^3}=\frac{6120}{79}=77.468\cdots\cdots$$

$$\fallingdotseq 77.47,$$

∴　換算式：

$$\mathrm{X_B}=\begin{cases}\dfrac{77.47}{90}\times x, & x\leq 90,\\[2ex] \dfrac{200-77.47}{200-90}\times(x-90)+77.47 & \\[2ex] =\dfrac{122.53}{110}\times(x-90)+77.47, & x>90.\end{cases}$$

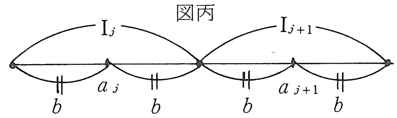
この式で算出して，上の表4を得る。この表から，実際には，$\mathrm{E}(\tilde{\mathrm{X}}_\mathrm{B})=80.1$ となる。

§3　中点平均評価

x：小数部分が有限の数で，符号の正負を問わない。このような x を有限個扱う。

I_j：x の属する区間で，$\mathrm{I}_j=[a_j-b,\ a_j+b)$。すなわち，中点を a_j，巾を $2b$，$b>0$，とし，図丙のように，$a_{j+1}=a_j+2b$，$1\leq j\leq r-1$，とする。r はこの様な区間の個数とする。

図丙



$f(x)$：xの度数，$\sum_x f(x)=n$ とする。

x^*：$x\in \mathrm{I}_j$ のとき，I_j の中点 a_j を表す。すなわち，$x^*=a_j$，$a_j-b\leqq x<a_j+b$，

E：xの算術平均，E^*：x^*の算術平均。

定　　理

上記の記号の下で，次が成り立つ：

$|\mathrm{E}^*-\mathrm{E}|\leqq b$ ……………………………（E^*.1）

正確にいえば，次の通りである：

$\mathrm{E}-b<\mathrm{E}^*\leqq \mathrm{E}+b$，　…………………（$\mathrm{E}^*$.2）

i.e.　$\mathrm{E}^*-b\leqq \mathrm{E}<\mathrm{E}^*+b$.　…………………（$\mathrm{E}^*$.3）

（証明）　x^*の定義より，$-b<x^*-x\leqq b$.

$\therefore\quad -b\sum_x f(x)<\sum_x(x^*-x)f(x)\leqq b\sum_x f(x)$，

$\therefore\quad -b<(\sum_x x^*f(x))/n-(\sum_x xf(x))/n\leqq b$，

i.e.　$-b<\mathrm{E}^*-\mathrm{E}\leqq b$.　（了）

付記1.　等号の成り立つのは，各 $x\in \mathrm{I}_j$ が I_j の左端にあるとき，かつそのときに限る。

付記2.　φ：$x\mapsto x^*$ は単調増加函数であるから，φは1つの換算である。

応用1.　四捨五入法

（1.a）　xの小数点以下第k位，$k\geqq 1$，を四捨五入し，第（$k+1$）位以下の端数を切り捨てた数値 x^{**} の算術平均 $\mathrm{E}(x^{**})$ と，xの算術平均$\mathrm{E}(x)$ に対して，$|\mathrm{E}(x^{**})-\mathrm{E}(x)|\leqq 5\times 10^{-k}$ なることは，よく知られている。上の定理の応用として確認してみる：$2b=10^{-(k-1)}$ であるから，$b=5\times 10^{-k}$。ゆえに式（E^*.1）に依る。特に，$k=1$ のときは $|\mathrm{E}(x^{**})-\mathrm{E}(x)|\leqq 0.5$ となる。例えば，§2の表3，表4にある $\tilde{\mathrm{X}}_{\mathrm{A}}$，$\tilde{\mathrm{X}}_{\mathrm{B}}$ に対して，$\mathrm{E}(\tilde{\mathrm{X}}_{\mathrm{A}})$，$\mathrm{E}(\tilde{\mathrm{X}}_{\mathrm{B}})$ はそれぞれ $\mathrm{E}(\mathrm{X}_{\mathrm{A}})$，$\mathrm{E}(\mathrm{X}_{\mathrm{B}})$ と $|\mathrm{E}(\tilde{\mathrm{X}}_{\mathrm{A}})-\mathrm{E}(\mathrm{X}_{\mathrm{A}})|\leqq 0.5$，$|\mathrm{E}(\tilde{\mathrm{X}}_{\mathrm{B}})-\mathrm{E}(\mathrm{X}_{\mathrm{B}})|\leqq 0.5$ なる関係にあるはずである。実際§2で見た様に，$\mathrm{E}(\tilde{\mathrm{X}}_{\mathrm{A}})=\mathrm{E}(\mathrm{X}_{\mathrm{A}})=80$，および$\mathrm{E}(\tilde{\mathrm{X}}_{\mathrm{B}})=80.1$，$\mathrm{E}(\mathrm{X}_{\mathrm{B}})=80$，であった。このことは偶然なことではなかったことが理解されよう。

（1.b）　xの整数部分の10^lの位，$l\geqq 0$，を四捨五入し，かつその右側の整数部分をすべて0とし，さらに小数点以下の数字をすべて切り捨てて出来る整数値をx^{***}とするとき，$|\mathrm{E}(x^{***})-\mathrm{E}(x)|\leqq 5\times 10^l$ なることも見易いことであろう。

応用2　ある企業が，従業員に対して，某月の家計に関するアンケートを行ったとき，その結果の一部を表5のようにまとめた。

ただし，黒字，赤字ともに0の人達の総数は黒字1万円未満の区間の人数に繰込んでおいた。実際の平均をE万円とする。すなわち，E＝2.3 ならば，1人平均

	金額の区間	人数	区間の中点の金額（万）
黒字	⋮	⋮	⋮
	2万円から3万円未満	*	2.5
	1万円から2万円未満	**	1.5
	1万円未満	***	0.5
赤字	1万円未満	****	−0.5
	1万円から2万円未満	*****	−1.5
	2万円から3万円未満	******	−2.5
	⋮	⋮	⋮

表5　　表6

23,000円の黒字であることを意味し，E＝−0.7ならば，1人平均7,000円の赤字であることを意味する。しかし，表5からはEの値を計算出来ないから，各金額の区間における中点の金額の表6を使って，Eの概算を行うことにした。その結果が*E 万円であった。このとき，*EによりEのことがどの程度分かるか？

x：単位は万円で，黒字または赤字の金額の数値として，黒字の場合は正の値，赤字の場合は負の値で表す。

$f(x)$：x に対応する度数（人数），$x=0$ の場合は既に述べた通りである。

$\sum_x f(x)=n$，E：xの算術平均。

黒字の場合の区間の条件は“中点平均評価”における区間の条件を満たしているが，赤字の場合はそうではない。

例えば，赤字が1万円から2万円未満の区間は，中点は -1.5 であるが，この区間のx は，$-2<x\leqq -1$ となるからである。この場合は，x^* の代用として，次の様に$^*x$を定義して用いる：

$^*x$：区間 $-2<x\leqq -1$ における x に対しては，$^*x=-1.5$ すなわちこの区間の中点を表す。すなわち，

$-0.5\leqq {}^*x-x<0.5$ ……………………………(1)

他の赤字の各区間においても全く同様である。

一方黒字の場合は，xの属する区間の中点を x^* として，

$-0.5 < x^* - x \leqq 0.5$ ……………………………(2)

黒字（≠0）の人，赤字（≠0）の人が，それぞれ少なくとも1人は居る。

このアンダーラインの部分を満たすならば，式(1)と式(2)より次のようになる：

$$-0.5\times\sum_x f(x) < \sum_{x<0}({}^*x - x)f(x) + \sum_{x\geqq 0}(x^* - x)f(x) < 0.5\times\sum_x f(x) \cdots\cdots\cdots(***)$$

一方，$^*\mathrm{E}=(\sum_{x<0}{}^*xf(x)+\sum_{x\geqq 0}x^*f(x))/n$，$\mathrm{E}=(\sum_x xf(x))/n$。 ゆえに（＊＊＊）より，

$-0.5 < {}^*\mathrm{E} - \mathrm{E} < 0.5$,

すなわち，$|{}^*\mathrm{E}-\mathrm{E}| < 0.5$.

かくして，上記アンダーラインの部分が満たされているならば，*EとEの違いは，0.5未満であることになる。すなわち，

{ *E×10,000円とE×10,000円の違いは，5,000円未満である。

上記アンダーラインの部分が満たされていることを仮定して，2つの例を挙げる。

例1. *E＝3.2

このとき，“実際の平均は黒字であり，その額は，27,000円より多いが，37,000円より少ない”と理解出来る。しかし，“実際の平均が2万円台であるのか，あるいは3万円台であるのか”ということについては不明である。

例2. *E＝0.3

このとき，“平均は黒字でその額は3,000円であり，赤字ではない”というのは正確ではない。“実際の平均は，黒字ならば，その額は8,000円未満であり，一方赤字ならば，その額は2,000円未満である”と理解すべきである。この場合，“実際の平均が黒字であるのか，あるいは赤字であるのか”ということについては不明である。

注意1.

この2つの例から分かるように，統計の名の下で，“ウソ”をつく可能性のあることを理解されよう。[5]

注意2.

中点平均評価と同様に，応用2のような問題の解答は，区間の巾$2b$，$b>0$，で表されることになるから，bの選び方すなわちデーターの“まとめ方”により，概算の結果が変わって来る。この応用2におけるアンケートのまとめ方も，5,000円巾あるいは，1,000円巾とすることにより，より精密な数値が出て来る。しかし，その上でも尚，注意1で述べたような，算出結果の誤用，悪用の危険性は残る。

§4 始点平均評価

x：小数部分が有限の数で，符号の正負を問わない。このようなxを有限個扱う。

I_j：xの属する区間で，$\mathrm{I}_j=[a_{j-1}, a_j)$，$a_j=a_{j-1}+\mathrm{c}$，$\mathrm{c}>0$，$1\leqq j\leqq \mathrm{r}$。すなわち，等間隔cの区間r個が続いて並んでいる。$f(x)$：xの度数，$\sum_x f(x)=n$。

$\tilde{x}$：$x\in \mathrm{I}_j$のとき，$\tilde{x}=a_{j-1}$。すなわち区間の始点を表す。

$\tilde{\mathrm{E}}$：$\tilde{x}$の算術平均，　E：xの算術平均。

定　理

上の記号の下で，次が成り立つ：

$0\leqq \mathrm{E}-\tilde{\mathrm{E}} < \mathrm{c}$, ……………………（$\tilde{\mathrm{E}}$.1）

すなわち　$\mathrm{E}-\mathrm{c} < \tilde{\mathrm{E}} \leqq \mathrm{E}$, ………………（$\tilde{\mathrm{E}}$.2）

$\tilde{\mathrm{E}}\leqq \mathrm{E} < \tilde{\mathrm{E}}+\mathrm{c}$. ……………………（$\tilde{\mathrm{E}}$.3）

すなわち，$\tilde{\mathrm{E}}$はEを越えることはなく，かつ$\tilde{\mathrm{E}}$とEの違いはcを越えない。

（証明）　$\tilde{x}$の定義により，$0\leqq x-\tilde{x} < \mathrm{c}$

ゆえに，$0\leqq \sum_x f(x)(x-\tilde{x}) < \mathrm{c}\sum_x f(x)=\mathrm{c}n$

ゆえに，$0\leqq(\sum_x f(x)(x-\tilde{x}))/n < \mathrm{c}$

すなわち（$\tilde{\mathrm{E}}$.1）～（$\tilde{\mathrm{E}}$.3）を得る。　（了）

注意　$\varphi: x\mapsto\tilde{x}$ は単調増加函数であるから，換算の1つである。

応用　小数部分が有限の数を有限個扱うものとする。xの小数点以下第k位，$k\geqq 1$，以下の端数を切り捨てて得られる数 $\tilde{x}$ の平均を$\tilde{\mathrm{E}}_{-k}$，$k\geqq 1$，とすれば，

$0\leqq \mathrm{E}-\tilde{\mathrm{E}}_{-k} < 10^{-(k-1)}$.

一方，整数部分の10^kの位以下，$k\geqq 0$，のすべての整数部分の数字を0で表し，かつ小数部分を切り捨てて得られる整数の平均を$\tilde{\mathrm{E}}_k$，$k\geqq 0$，とするとき，

$0 < \mathrm{E}-\tilde{\mathrm{E}}_k < 10^{k+1}$，$k\geqq 0$.

ともに，式（$\tilde{\mathrm{E}}$.1）より得られることは見易い。

§3の応用2で挙げたアンケートに，そのまとめ方を上記始点平均評価の応用として使用する方法も見易いと思われる。

参考文献

(1) 野本久夫：確率・統計，朝倉書店，昭和48年，第2版P.P.24～29。

(2) 竹内啓・藤野和建：2項分布とポアソン分布，東大出版会。

(3) C.F.ガウス：誤差論（飛田武幸，石川耕春訳），紀伊国屋書店。

(4) 杉山高一：偏差値と合計点をめぐって（青山博次郎編：暮しのなかの統計学，第8章），東洋経済新報社。

(5) ダレル・ハフ：統計でウソをつく法（高木秀玄訳）ブルーバックス，B120，講談社。

（受理　昭和59年1月17日）