実体分布と度数分布

関　弥三郎

はしがき

統計学では統計集団の量的構造—量的標識について明らかにされた統計集団の構造—を解明する場合、個々の集団構成単位のもつ量的標識を配列した統計系列を直接取り扱うのでなく、個々の単位を量的標識によってグループ分けし、量的標識の値（変量値）の変化に伴う単位数の変化すなわち度数分布としてとらえ、度数分布の特徴を分析するのが普通のやり方である。そして度数分布の特徴の分析は次の手続きによって行われる。

(1) 度数分布は対称分布か正または負の非対称分布、単峰分布か複峰分布等の数種の分布形態に分類することができるから、いずれのタイプに属するかを明らかにする。そして

(2) 度数分布の中数値（Mittelwerte）を規定することにより統計系列の中央の位置を求め、あるいは補償値を得る。次に

(3) 度数分布の撒布の程度を表す撒布度（Streuungsmaß）を計算する。更に

(4) 度数分布が非対称の時はその歪みの方向と程度を歪度（Schiefe）によって測定し、また時には度数分布の峰の尖りの程度を尖峰度（kurtosis）によって測定することもある。

しかしこれで統計集団の量的構造の解明という問題が完全に処理されているか、というとそうではないのであって、まだ分析されないで残されている問題がある。今量的標識を所得、単位数を人数とすると、何万円台の所得者が何人であって全体の何％を占めている、というだけでは所得分布の問題の解明には不十分であって、何万円台の所得者の所得金額はいくらであって所得総額の何％を得ているか、といった所得そのものの分配状態の分析にまで立ち入らなければならないことは明らかであろう。一般的に言えば、各量的グループごとにそれに属する単位がもつ変量値の合計に着目して、変量値の変化に伴うこの合計の変化の特徴を分析することである。

この量的グループに属する単位がもつ値の合計の分布は、単位数の分布を度数分布と呼ぶのに対して、一般に慣用の名称は与えられていない。このような合計は変量xとそれをもつ単位数fとの積xfにほかならないから、xをウエイトとした加重度数とみることができ、故にそれは一種の加重度数分布であると言えるであろう。しかし、この合計は先の所得分布の例からもわかるように、所得金額のような実質的な経済量・社会量であって、加

重度数という表現は適切でないと言わねばならぬ。この分布について特別の考察を加えている米沢治文氏はこれを実体分布と呼んでいるのであるが[1]、それの方が実情により即していると思われるので、やや熟さないきらいはあるがここでも実体分布と言うことにする。

しかし、ここで注意すべきことは実体分布の分析はどのような量的標識の場合でも必要になるのではないのであって、量的グループに属する単位がもつ変量値の合計が社会経済的（事物的）に有意義である場合でなければならず、変量値の合計が事物的に何等の意味も持ち得ない時は実体分布の分析は問題にならないことを忘れてはならぬ[2]。人口や所得は実体分布の分析が必要な典型的な場合であるが、身長や年令はそれが無意味な代表的な例である。人口規模別市町村分布において人口五万未満の市町村の人口は何千万人であって全人口の何%を占めているとか、年収一〇〇万円未満の給与所得者の所得金額は何億円であってそれは総給与所得の何%に当る、等のことは社会的経済的に有意義な知識であり、それ故に人口総数を二等分する位置にある市町村の人口規模や総給与所得を折半する位置の所得額（境界値）、あるいはまた人口や所得金額が最大であるクラスの変量値（最重値）といったものが必要になるのである。ところが、二〇―二九才人口の年令合計や身長一六〇―一六九cmの学生の身長合計は何等の事物的意義をも持ち得ないから、年令の総合計、身長の総計を二等分する位置にある年令、身長あるいはまた年令合計、身長合計が最大になるクラスの変量値は問題にならないのである。

実体分布の分析は特に経済統計、人口統計の分野で必要になる場合が多い。所得分布の場合には早くから所得人員分布（度数分布）の分析だけではなく所得金額分布（実体分布）の分析も併せ行われてきたのであって、実体分布の中数値の一つである境界値は既に一九世紀中頃には用いられていたのである[3]。そして、実体分布の中数値で

ある境界値と最重値が統計方法論において始めて取上げられたのは、心理学者、哲学者であるフェヒナー（Gustav Theodor Fechner, 1801—1887）の著書「集合論」（Kollektivmaßlehre, 1897. 彼の死後出版された）においてであるが、フェヒナー自身はそれらに余り重要性を認めなかったし、その後の統計学者によって不当に軽視されてきたのである。(4) ドイツ社会統計学の碩学チチェク（Franz Zizek, 1876—1938）の名著「統計的中数値論」（Die statistischen Mittelwerte, 1908）においても、ただ一言フェヒナーが境界値と最重値を中数値として挙げていることを述べているにすぎず、何等の理論的考察は加えられていないのであって、(5) 統計的中数値としてのこれらの重要性を理論的に述べたのは、フラスケンパー（Paul Flaskämper）の論文「統計的中数値の論理への寄与」（Beitrag zur Logik der statistischen Mittelwerte, Allg. Stat. Arch., 21 Bd., 1931）が最初である。(6) しかし、フラスケンパーにおいても度数分布に対する実体分布としては取り上げられておらず、ただ統計的中数値の一形式としてのみ考察されているのである。その後はメンゲス（Günter Menges）の最重値に関する研究を除いては、(7) 実体分布の研究には見るべきものがなかったのであるが、最近米沢氏がその著「経済統計計量分析」（一九七二年）において、度数分布に対する実体分布を明確に意識して、実体分布の分析方法及び実体分布と度数分布の関連について検討を加えているのであり、(8) その理論統計学における意義は高く評価さるべきであると考える。しかし、米沢氏の実体分布の研究はもう一つ組織的でなく、特にフラスケンパーが加えた分析の視角—事論理と数論理の併行論—を欠いており、更に実体分布の分布形態についての説明も不十分であると思われる。本稿はフラスケンパーやメンゲス、米沢氏によって拓かれた実体分布の研究をできるだけ簡単な形に整理し、更に必要な補足、拡張を付け加えて統計分析の実際における利用の便宜に供すると共に、実体分布の分析の今後の発展に役立てんとするものである。

（1）米沢治文「経済統計計量分析」一九七二年、六〇ページ。
（2）P. Flaskämper, Beitrag zur Logik der statistischen Mittelwerte, Allg. Stat. Arch., 21 Bd., 1931, SS. 392—393.
（3）日本統計学会編「国民所得とその分布」昭和十九年、二一五ページ。但し、そこでは境界値を二等分値（equator）と言っている。
（4）フラスケンパー、大橋、足利訳「一般統計学」昭和二八年、一二四ページ。
（5）ジィジェーク、岡崎文規訳「統計的中数値論」大正十五年、二〇〇ページ。
（6）P. Flaskämper, Beitrag, SS. 392—397.
（7）G. Menges, Über den Schwersten Wert T, Allg. Stat. Arch., 37 Bd., 1953.
（8）米沢治文、前掲書、六〇—七一ページ。

一　実体分布の特性値

実体分布の特徴の分析は度数分布の場合と同様に

(1) 実体分布の分布形態を明らかにし、
(2) 実体分布の中数値を規定して統計系列の中央の位置を求め、次に
(3) 実体分布の撒布の程度を計算し、更に
(4) 実体分布の歪みの方向と程度を測定する、

等によって行うことが考えられる。説明の便宜上実体分布の中数値の規定からみて行くことにする。

度数分布の場合は算術平均、中央値（中位数）、最頻値（並み数）等の中数値によって統計系列の中央の位置を表

すのであるが、実体分布の場合も同様に中数値を求めて統計系列の中央の位置を決めることが必要である。(1) フラスケンパーによれば実体分布の中数値には境界値と最重値があり、前者は中央値に類似し後者は最頻値と同じ性質のものである。

a　境界値

まず境界値（Scheidewert）であるが、これはそれよりも大きい変量値の和がそれよりも小さい変量値の和と丁度相等しくなる位置にある変量値であって、Rで表すのが普通である。これを Scheidewert というのはフェヒナーの命名によるのであって、これはまた médiale, equator とも呼ばれ、わが国では加重中位数と言われたことがある。

中央値Zが単位総数を折半する位置にある変量値であるのに対して、境界値Rは単位数ではなしに変量値の総計を二等分する位置にある変量値であり、いずれも統計系列を二つの等しい部分―但し違った観点のもとで―に分割する位置の値であって統計系列の中点を表すのである。このようなRはZと並んで事物的意義を持ち得ることは十分首肯し得るところである。けだし、Zは単位の個数に関して中点であるが、変量値に差がある以上変量値についてみる時はそれはもはや中点ではなく、従って事態の正確な認識のためには更に変量値に関する中点であるRが必要になるからである。そしてRもZも共に統計系列の中点を表しそれによって統計系列全体の水準を代表的に示すのであるが、その場合ZよりもRの方がより良く統計系列全体の水準を反映するのである。なぜならば、所得分布の場合で言うと、低位及び中位の所得者の所得額が不変でただ高位の所得者の所得額のみが上昇

する、というやり方で社会全体の所得水準が変動したと仮定すると、Zとしては不変であり社会全体の所得水準の上昇を表さないが、Rは直ちに上昇し且つ高位の所得者の所得上昇の程度に応じて変動するであろうからである。このようにZは単位の個数しか問題にしないために変量値の変動に比較的鈍感であるのに対して、Rは変量値を問題にするためにそれの変動に敏感に反応するのであり、この点で算術平均に類似している。

RはZよりも常に値が大きい。すなわち

$$R > Z \tag{1}$$

なぜならば、Z以上の変量値の個数とZ以下の変量値の個数は同じであるが、それぞれの変量値の合計は当然Z以上の方が大きく、従って変量値の和が上下相等しくなるためにはその限界値はZよりも大きくなければならないからである。しかし、Rと算術平均mとの関係はそのように簡単なものではないのであって、度数分布の形態のいかんによって変化する。すなわち、度数分布が

$$\left.\begin{array}{ll} \text{正の非対称分布の場合は} & R \gtreqless m \\ \text{対称分布及び負の非対称分布の場合は} & R > m \end{array}\right\} \tag{2}$$

であって、一番多く現れる正の非対称分布の場合はRとmの関係は一義的には決らないのである。これは次のようにして証明することができる。(2) 今mより大きい方の変量をx_0、小さい方の変量をx_uで表しそれぞれの個数をN_0 N_u $(N=N_u+N_0)$ とすると、Rに対応する $\frac{1}{2}\sum^{N} x$ は次のように書き換えることができる。

$$\frac{\sum^{N} x}{2} = \frac{\sum^{N_u} x_u + \sum^{N_0} x_0}{2} = \sum^{N_u} x_u + \frac{\sum^{N_0} x_0 - \sum^{N_u} x_u}{2} \tag{i}$$

(i)の右辺第1項はmに対応する値であるから、第2項の正負のいかんによりRとmの大小関係が決まることがわかる。すなわち、第2項が正の時は $\frac{1}{2}\sum^{N} x$ が $\sum^{Nu} x_u$ よりも大きいのであるからRはmよりも値が大きいことになり、第2項が0の時は $\frac{1}{2}\sum x$ と $\sum x_u$ が等しくなるからRとmは一致し、第2項が負の時は $\frac{1}{2}\sum x$ が $\sum x_u$ よりも小さくなるからRはmよりも値が小さいことになる。第2項をtと置くとその符号は次のようにして決めることができる。mよりの偏差を $d=x-m$ で表すと

$$\sum^{N_0} x_0=\sum^{N_0}(x_0-m+m)=\sum^{N_0} d_0+N_0m \qquad \text{(ii)}$$

$$\sum^{Nu} x_u=\sum^{Nu}(x_u-m+m)=-\sum^{Nu}|d_u|+N_um \qquad \text{(iii)}$$

(ii)−(iii)

$$\sum^{N_0} x_0-\sum^{Nu} x_u=\sum^{N_0} d_0+\sum^{Nu}|d_u|+(N_0-N_u)m$$

$$\therefore \quad t=\frac{\sum^{N_0} x_0-\sum^{Nu} x_u}{2}=\frac{\sum^{N}|d|+(N_0-N_u)m}{2} \qquad \text{(iv)}$$

対称分布の時は $Z=m$ であるから $N_0=N_u$ 故に $t=\frac{\sum|d|}{2}>0$

負の非対称分布の時は $Z>m$ であるから $N_0>N_u$ 故に $t=\frac{\sum|d|+|N_0-N_u|m}{2}>0$

正の非対称分布の時は $Z<m$ であるから $N_0<N_u$ 故に $t=\frac{\sum|d|-|N_0-N_u|m}{2}\gtreqless 0$

かくて(2)が得られる。（証明終り）

表1　世帯人員別普通世帯数と人口

（昭和40年）

世帯人員	世帯数	人口	累積世帯数	累積人口
人	千世帯	千人	千世帯	千人
1	1,863	1,863	1,863	1,863
2	3,292	6,585	5,155	8,448
3	4,207	12,621	9,362	21,069
4	5,148	20,590	14,510	41,659
5	3,733	18,665	18,243	60,324
6	2,456	14,737	20,699	75,061
7	1,397	9,779	22,096	84,840
8	569	4,550	22,665	89,390
9	246	2,219	22,911	91,609
10〜	174	1,873	23,085	93,482
合計	23,085	93,482		

資料：昭和40年国勢調査，20%抽出集計

$m=4.05$人

$D=4$　　$T=4$

$Z=4$　　$R=5$

表2　人口階級別市町村数と人口

（昭和40年）

人口階級	市町村数	人口	累積市町村数	累積人口
千人		千人		千人
〜 2	52	70	52	70
2〜 5	320	1,181	372	1,251
5〜 10	1,144	8,663	1,516	9,914
10〜 20	1,006	13,957	2,522	23,871
20〜 30	276	6,607	2,798	30,478
30〜 40	167	5,781	2,965	36,259
40〜 50	110	4,862	3,075	41,121
50〜100	170	11,431	3,245	52,552
100〜500	119	22,921	3,364	75,473
500〜	12	22,802	3,376	98,275
合　計	3,376	98,275		

資料：昭和40年国勢調査

$m=29.1$千人

$D=7.7$　　$T=9.1$

$Z=11.7$　　$R=85.1$

最後にRの決め方について説明しておこう。度数分布の量的階級に級間隔がない場合は、単位のもつ変量値を小さい方から累積して行き $\frac{1}{2}\sum^{N} x$ を含むクラスの変量値をもってRとすればよい。表1の世帯人員別普通世帯分布の場合は $\frac{1}{2}\sum x=46,741$ 千人であり、それは五人世帯までの累積和の中に含まれているからRは五人である。（なお $\frac{N}{2}=11,542.5$ 千世帯であるからZは4人である。）しかし、度数分布の量的階級に級間隔がある時は、それによってはRが存在する区間を知り得るだけであってRの値そのものは直ちには得られない。表2の人口階級別市町村分布の場合 $\frac{1}{2}\sum x=49,137.5$ 千人であり、それは人口五―一〇万階級までの累積和の中に含まれるために、Rは五―一〇万人の間にあることがわかるのみである。そこで、このような場合にZを決めるのと同様のやり方で計算的にRの値を求めることが必要であり、その算式は次のようである。[3]

$$R=\underline{x}+\frac{c}{s}\left(\frac{S}{2}-S_R\right) \qquad (3)$$

但し，$\underline{x}$：R所在クラスの下限値
　　c：　〃　　級間隔
　　s：　〃　　変量値の合計
　S：変量値の総計（$\sum^{N} x$）
S_R：$\underline{x}$以下の変量値の合計

これによって表2のRを求めると八・五万人となる。（Zは同様の計算により一・二万人となる。）なお、実体分布の累積図表が与えられている時は $S/2$ の点を通る水平線を引き、それが累積多角形と交る点から下した垂線の足の横座標の値を読み取ることによって、Rの値をグラフから求めることができる。

なお、Zと並んで単位総数を四等分する二個の四分位数Q_1 Q_3を規定することができるのと同様に、Rと並んで変量値の総計を四等分する値を求めることも有意義であろう。すなわち、それ以下の変量値の合計が変量値の総計の $1/4$ 及び $3/4$ であるような位置にある変量値であって、今それを仮りにR_1 R_3で表わすことにしよう。しかし、R_1 R_3は中数値ではなく後に述べる実体分布の撒布度として役立つのである。これらの四分位数には特別の名称は付けられていないのであるが、Rを邦語で加重中位数とも言うのに準じて加重四分位数と呼ぶことが考えられる。はしがきで述べたように、実体分布を加重度数分布と言うのが不適当であるのと同様の理由から、この名称は実情によくマッチしないきらいはあるが、他に適当な名称がないのでここではこれを用いることにする。(4)

（1）度数分布の中数値は統計系列の中央の位置を与えると同時に補償機能をも果す（但し算術平均の場合のみ）のであるが、実体分布の中数値は（算術平均に類似の中数値がないために）前者のみであって後者は問題にならない。

（Flaskämper, Beitrag, SS. 385 ff.）

（2）Flaskämper, Beitrag, S. 396.

（3）水谷一雄「統計学」昭和二五年、一〇七ページ。

（4）R_1R_3を第1及び第3四等分値と言う場合がある（日本統計学会編、前掲書、一七八—九ページ）が、Q_1Q_3も度数の四等分値であるから、実体量の場合のみを特に四等分値と名付けるのは適当ではなく、むしろ加重四分位数の方が良いと思う。

b　最重値

次に最重値（schwerster Wert）であるが、これは変量値とそれの個数の積が最大になる位置にある変量値であって、Tで表される。これを schwerster Wert と言うのはフェヒナーの命名によるのである。

最頻値Dが最大単位数に対応する変量値であるのに対して、最重値Tは単位数ではなしにクラスの変量値の和が最大であるところの変量値である。ZやRと違ってDもTも統計系列を二つの等しい部分に分割するものではないから厳密な意味における統計系列の中点ではないが、Dはその個数が一番多い変量値従って最も普通の変量値であり、またTはその値の和が一番大きいという意味において最もウエイトの大きい変量値であることから、いずれも統計系列の特徴的な中央の位置を示すのである。このようなTはDと並んで事物的に有意義であることは明らかであろう。例えば、消費支出階級別世帯分布において一番多くの世帯が支出する金額（D）を知ると共に、消費支出の合計が一番大きい従って消費購買力が最大のクラスの支出金額（T）を知ることは経済的に有意義だからである。なお、度数分布において複数個の山が現れDを一義的に決めることができない場合があるのと

同様に、実体分布においても数個の山が生じてTを一義的に規定し得ない場合があることは断るまでもないであろう。

次にTとDの関係を調べると、一般に

$$D \leqq T \tag{4}$$

である。[1] なぜならば、D以下では変量xも度数$f(x)$もDのクラスより小さくなり、従って積$xf(x)$も減少するからTはDよりも小さくなることはあり得ない。そしてD以上ではxは大きくなり$f(x)$のみが小さくなるから、積$xf(x)$が直ちに減少する場合（$D=T$）と、しばらくは増加を続けやがて減少に向う場合（$D<T$）とが考えられるからである。なおこの関係を図によって説明すると（図1参照）、Dの右側で度数曲線$y=f(x)$がDに対応する極大点Pを通る直角双曲線$xy=Df(D)$よりも小さい時は、積$xf(x)$が$Df(D)$よりも小さくなるから$D=T$であり、$y=f(x)$が$xy=Df(D)$よりも大きい時は、積$xf(x)$が$Df(D)$よりも大きくなるから$D<T$である。そして、後者の場合$xy=Df(D)$よりも大きい$f(x)$上の点を通る双曲線群のうち一番外側にある双曲線を$xy=x_k f(x_k)$とすると、$T=x_k$である。

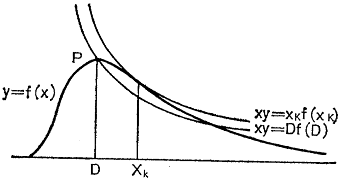

図1

しかし、度数曲線$y=f(x)$がすべてのxに対して連続で微分可能の時は

$$D < T \tag{5}$$

であって$D=T$はあり得ない。[2] それは次のようにして証明することができる。まず度数曲線$y=f(x)$を$x=D$

において微分すると

$$f'(x)=0 \qquad \text{(i)}$$

であり、次に積 $xf(x)$ を $x=T$ において微分すると

$$\frac{d}{dx}\{xf(x)\}=f(x)+xf'(x)=0 \qquad \text{(ii)}$$

である。今 $x=D=T$ と仮定すると(i)を(ii)に代入することにより

$$f(x)=0 \qquad \text{(iii)}$$

が得られる。しかし度数曲線が D において0になることはないから $D=T$ はあり得ない。従って $D<T$ となる。

（証明終り）

以上の T と D の関係は度数分布が対称分布であっても正または負の非対称分布であっても同様に成立する。しかし、T と m の関係は度数分布の形態のいかんによって異なるのである。すなわち、度数分布が

$$\left.\begin{array}{ll} \text{正の非対称分布の場合は} & T \gtreqless m \\ \text{対称分布及び負の非対称分布の場合は} & T > m \end{array}\right\} \qquad (6)$$

である。なぜならば、対称分布の場合は $m=D$ であるから(5)より $m=D<T$ また負の非対称分布の場合は $m<D$ であるから $m<D<T$ であり、いずれも T は m よりも大きくなるが、正の非対称分布の場合は $D<m$ であるから T と m の関係は一義的に決らないからである。しかし、余り対称分布に近くない限り正の非対称分布の場合は $T<m$ である。そして、正負いずれの場合でも非対称の程度が著しい程 T と m との差は拡大するのである。

度数曲線が正規の場合Tの値は次の式によって求めることができる。(3)

$$T=\frac{m+\sqrt{m^2+4\sigma^2}}{2} \tag{7}$$

これの証明は次のようである。すなわち、度数曲線は

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-m)^2}{2\sigma^2}}$$

であり、これを微分すると

$$f'(x)=-\left(\frac{x-m}{\sigma^2}\right)f(x)$$

となる。これを(ii)に代入すると

$$\frac{d}{dx}\{xf(x)\}=\left\{1-\frac{x(x-m)}{\sigma^2}\right\}f(x)=0$$

ところが正規曲線の性質から $f(x)\neq 0$ であるから

$$1-\frac{x(x-m)}{\sigma^2}=0 \quad \text{すなわち} \quad x^2-mx-\sigma^2=0$$

$$\therefore \quad x=\frac{m\pm\sqrt{m^2+4\sigma^2}}{2} \tag{iv}$$

しかし、$\sigma\neq 0$である限り $m<\sqrt{m^2+4\sigma^2}$ であるから、平方根がマイナスの符号をもつ根の場合はTがmよりも小さくなり、(6)の条件に反することになる。故に平方根がプラスの符号をもつ根のみが可能であり、かくて(7)

が得られる。(証明終り)

なお(7)より、度数曲線が正規の場合mとTの差は標準偏差σの大小に依存するのであって、xの撒布が大きい程両者の開きは著しくなることがわかる。また標準正規分布の時は$m=0$, $\sigma=1$であるから$T=1$となる。

最後にTの値の決め方について説明しよう。度数分布の量的階級に級間隔がない時は、(4)より明らかなようにTがDより小さくなることはあり得ないのであるから、最大度数のクラスから始めてそれ以後のクラスごとに変量値と単位数の積を比較して、それが一番大きくなるクラスの変量値をもってTとすればよい。しかし、度数分布の量的階級に級間隔がある場合は、それではTが存在する区間を知り得るのみであって、Tの値そのものは直ちには分からないのである。そこでDの値を決めるのと同様のやり方で、計算によってTの値を求めることが必要であり、その算式は次のようである。

$$T_1=\underline{x}+c\frac{(S_0-S_{-1})}{(S_0-S_{-1})+(S_0-S_{+1})} \tag{8}$$

または

$$T_2=\underline{x}+c\frac{S_{+1}}{S_{-1}+S_{+1}} \tag{9}$$

但し, $\underline{x}$: T所在クラスの下限値

c : 〃 級間隔

S_0 : 〃 変量値の合計

S_{-1} : その下のクラスの変量値の合計

これらの式は、隣接するクラスはその変量和に比例してTの位置に影響を及ぼすのであって、$S_{+1}=S_{-1}$ならばTは級間隔の中央にあり、$S_{+1}>S_{-1}$ならば上限値に接近し、$S_{+1}<S_{-1}$ならば下限値に近づく、との仮定から出発して導かれたものである。(4)

S_{+1}：その上のクラスの変量値の合計

これによってTの値を計算する場合、クラスによって級間隔が異なる時は隣接クラスの変量和の取扱いに注意すべきである。これを表3a）の例で説明すると、その場合Tは三〇—四〇万円クラスに所在することがわかる。そこで(8)または(9)によってTの値を決めるのであるが、ここで問題になるのは、T所在クラスの下のクラスにおいて級間隔が半分になっていることである。先に述べたように、Tの計算式は隣接クラスがその変量和に比例してTの位置に影響を及ぼす、との仮定から導かれているのであるが、このように級間隔が$\frac{1}{2}$であれば当然そのクラスの変量和は少くなり、従ってその変量和を使って計算すると、Tの値は隣接クラスの変量和の大きさだけではなく級間隔の違いの影響をも受けて決まることになり、正確な値が得られないのである。そこで、級間隔の違いをなくした場合の変量和を用いて計算しなければならないのであるが、そのためには標準間隔当り変量和を用いる方法と、クラスの合併によって級間隔の違いを是正する方法とが考えられる。標準間隔を一〇万円とすると二五—三〇万円クラスのみを修正すればよいから、その所得金額割合六・〇%に級間隔の比率$10/5=2$を乗じて一〇万円間隔当り割合一二・〇%に修正し、これをS_{-1}としてTを計算するのである。次は二〇—二五万円クラスと二五—三〇万円クラスを合併すると一〇万円間隔のクラスになり級間隔の差がなくなるので、それぞれの所得金額割合の合計$5.2+6.0=11.2\%$をS_{-1}としてTを計算するのである。しかし、標準間隔当りに修正する時は

人工的な変量和となるから、与えられたデータがクラスの合併による級間隔の差の是正を許すならばまずその方法によるべきであり、それが不可能な場合に始めて標準間隔当り変量和の方法を用いるべきであろう。

(1)(2)(3) Menges, a. a. O., SS. 36—37.

(4) この式のDの場合についての証明は、フラスケンパー「一般統計学」一二二ページ、または F. E. Croxton and D. J. Cowden, Applied General Statistics. 2nd ed., 1955, pp. 190—191. 参照。

c　非調和平均

以上で度数分布の中数値である中央値、最頻値に準ずる実体分布の中数値の境界値と最重値を説明したのであるが、今度は度数分布で一番多く使われる算術平均に相当する実体分布の中数値を考えてみよう。度数分布の算術平均が $m=\frac{\sum xf}{\sum f}$ すなわち変量xを度数fで加重して平均したものであるから、それと同じ計算操作を実体分布の場合に適用すると、変量xを変量和xfで加重して平均すればよいことになる。今これをA_hで表すと

$$A_h=\frac{\sum x\cdot xf}{\sum xf}=\frac{\sum x^2f}{\sum xf} \qquad (10)$$

である。これはフェヒナーによって非調和平均 (antiharmonisches Mittel) と呼ばれた算式に外ならない。(10)は次のように変形することができる。

$$A_h=\frac{\frac{1}{N}\sum x^2f}{m}=\frac{m^2+\sigma^2}{m} \quad \left(\because\ \sigma^2=\frac{1}{N}\sum x^2f-m^2\right) \qquad (11)$$

$$=m(1+c^2),\quad c=\frac{\sigma}{m} \qquad (12)$$

(12)から$\sigma \neq 0$である限り

$$A_h > m \qquad (13)$$

であって、xの撒布が大きい程A_hとmの差はより拡大することがわかる。

米沢氏はこれを境界値、最重値と共に実体分布の特性値としている。[1] しかし、算術平均は単位がもつ変量値の総計を単位総数で除して一単位当りに均らした値を与える、という事物的意味をもつ（算術平均の補償機能）[2]のに対して、非調和平均の場合は単位のもつ変量値の自乗の総計を変量値の総計で割るという計算操作の事物的意味が不明であり、また、算術平均はそれより小さい変量x_uについてとった偏差$m-x_u$の合計と、それより大きい変量x_0についてとった偏差x_0-mの合計とが相等しくなる位置にある値であり、この意味において統計系列を二つの等しい部分に分割し、統計系列の中央の位置を示すのに対して、[3] 非調和平均にはそのような数学的性質は認められないために、統計系列の中央の位置を表わす特性値とは言えないのである。従って、フラスケンパーは非調和平均を「その数学的性質はいかに興味があろうとも……統計学にとってはさらに重要でない。[4]」として、実体分布の中数値とはしないのであるが、われわれもフラスケンパーのこの見解に従うものである。

それでは非調和平均は全く無意味なものであって何等の役割も果し得ず単なる数学的興味のみから存在するにすぎない、と言い切れるかと言うと必ずしもそうではないのであって、間接的ながら度数分布の撒布の程度を測定し得るのである。けだし、分散$\sigma^2=\frac{1}{N}\sum(x-m)^2$が算術平均のまわりの二次の積率であって、算術平均を基準として変量の撒布を測定するのに対して、(11)の分子は原点のまわりの二次の積率であって、原点を基準とし

て変量の撒布を測定しているからである。そして、それを算術平均で除したものが非調和平均であるが、分子が変量値の自乗であるからそれを相対化するためには分母は算術平均の自乗でなければならず、この点で非調和平均は撒布度としては不完全なものであると言わねばならぬ。そこで非調和平均を算術平均で割って相対化すると、(12)より明らかなように $\frac{A_h}{m}=1+c^2$ となり変動係数 c によって規定され、より完全な撒布度となるのである。

（1） 米沢治文、前掲書、六三—六四ページ。
（2）（3） Flaskämper, Beitrag, SS. 387—388, 390—391. フラスケンパー「一般統計学」一〇〇—一〇五ページ。
（4） フラスケパー「一般統計学」一二五ページ。

d 撒布度

今度は実体分布の撒布の程度の測度を考えてみよう。撒布というのは中数値のまわりにおける個々の値の分布であって、撒布の程度を測定する撒布度は個々の値の中数値よりの偏差 d の中数値として得られる。すなわち、偏差 d の中数値として中央値を用いた場合が四分位偏差であり、算術平均によった場合が平均偏差、平方平均をとった場合が標準偏差である。[1] 四分位偏差は、より詳しく説明すると、偏差をとる中数値に中央値 Z を用い、そして偏差 $d=x-Z$ の個数に着目して、Z より大きい変量 x_0 についてとった偏差 $d_0=x_0-Z$ の中央値—それは明らかに第3四分位数 Q_3 の Z よりの偶差 Q_3-Z に外ならない—を求め、次に Z より小さい変量 x_u についてとった偏差 $d_u=Z-x_u$ の中央値—それは第1四分位数 Q_1 の Z よりの偏差 $Z-Q_1$ である—を求め、両者を平均したものである。[2] 故に四分位偏差は

$$\frac{(Q_3-Z)+(Z-Q_1)}{2}=\frac{Q_3-Q_1}{2}$$

であり、これをQ_1とQ_3の平均値で割って相対化するのである。すなわち

$$\frac{\frac{1}{2}(Q_3-Q_1)}{\frac{1}{2}(Q_3+Q_1)}=\frac{Q_3-Q_1}{Q_3+Q_1}$$

今のは度数分布の場合であるが、実体分布の撒布の程度を測定するには偏差をとる中数値に境界値Rを用い、そして偏差 $d=x-R$ の個数ではなしにそれに対応する変量xに注目して、Rより大きい変量x_0についてとった偏差 $d_0=x_0-R$ の境界値―それは第3加重四分位数R_3のRよりの偏差 R_3-R に外ならない―を求め、次にRより小さい変量 x_u についてとった偏差 $d_u=R-x_u$ の境界値―それは第1加重四分位数R_1 の R よりの偏差 $R-R_1$ である―を求め、両者を平均すればよい。従って実体分布の撒布度は

$$Q=\frac{(R_3-R)+(R-R_1)}{2}=\frac{R_3-R_1}{2} \tag{14}$$

であり、これをR_1とR_3の平均値で割って相対化すればよいであろう。すなわち

$$\frac{\frac{1}{2}(R_3-R_1)}{\frac{1}{2}(R_3+R_1)}=\frac{R_3-R_1}{R_3+R_1} \tag{15}$$

この撒布度はフラスケンパーによっても考察されておらずただ暗示されているのみであって、(3) 特別の名称は与え

られていないのであるが、先にR_1R_3を加重四分位数と言ったことから、ここでは仮りに⒁を加重四分位偏差、⒂を加重四分位偏差係数と呼ぶことにする。

四分位偏差は偏差の個数について全体の $\frac{1}{2}$ を含む中央部分の範囲をもって撒布の程度を測るのであるから、変量値の変動に極めて鈍感にしか反応しないのに対して、この加重四分位偏差は偏差に対応する変量値に関してそれの総和の $\frac{1}{2}$ を含む中央部分の範囲をもって撒布の程度を測るのであるから、変量値の変動には比較的敏感に反応するのであって、この点で平均偏差、標準偏差に類似している。

以上は偏差dの中数値として境界値を用いた場合であるが、この外にdの中数値として算術平均、平方平均を用いることが考えられるであろう。しかし実体分布に関しては算術平均の計算操作が事物的意味を持ち得ないのであるから、当然偏差dに対しても変量xを考慮して算術平均及び（それの拡張である）平方平均を適用することは無意味であり、故に度数分布の場合の平均偏差、標準偏差に相当する実体分布の撒布度は、計算的には可能であるにしても実際上は存在しないと言ってよいであろう。

（1）（2） フラスケパー「一般統計学」一三〇—一三八ページ。

（3） Flaskämper, Beitrag, S. 395.

e 歪　度

最後に実体分布の歪みの方向と程度の測定の問題があるが、当然この場合も度数分布の歪度の算式に準じた方法が考えられる。度数分布の歪度には ⑴算術平均と最頻値の比較による方法 ⑵中央値と四分位数の比較によ

る方法及び　(3)算術平均よりの偏差の3乗平均による方法があるが、実体分布の場合は算術平均の計算操作が有意義に適用し得ないことから、それらのうち算術平均を利用する方法は使えない。従って中央値と四分位数による算式のみが準用可能となるのである。この算式は、Q_1 Q_3 の Z よりの偏差は度数分布が対称分布の時は相等しく $(Z-Q_1)=(Q_3-Z)$ であるが、非対称分布の場合は撒布範囲の広い方の値が大きくなり、正の非対称の時は $(Z-Q_1)<(Q_3-Z)$ 負の非対称の場合は $(Z-Q_1)>(Q_3-Z)$ であることから、$(Q_3-Z)-(Z-Q_1)=Q_1+Q_3-2Z$ は非対称の方向と程度を測定することができ、これを四分位偏差の二倍 Q_3-Q_1 で割って相対化したものである。(1)すなわち

$$S_k=\frac{Q_1+Q_3-2Z}{Q_3-Q_1}$$

実体分布の場合も R_1 R_3 の R よりの偏差は実体分布が対称分布の時は $(R-R_1)=(R_3-R)$ 正の非対称の場合は $(R-R_1)<(R_3-R)$ 負の非対称の時は $(R-R_1)>(R_3-R)$ であるから、同様の次の算式を歪度とすることが考えられる。すなわち

$$\frac{(R_3-R)-(R-R_1)}{R_3-R_1}=\frac{R_1+R_3-2R}{R_3-R_1} \qquad (16)$$

(16)の値は-1と$+1$の間で変動し、実体分布が対称分布の場合は0になり、正の非対称分布の時はプラス、負の非対称分布の時はマイナスの値をとる。

（1） Arthur Lyon Bowley, Elements of Statistics, 6th ed., 1937, p. 116.

二　分布形態

以上実体分布の特性値について考察してきたのであるが、次に実体分布の分布形態と度数分布の分布形態を比較してみよう。実体分布と度数分布を比べる場合変量和と度数とでは測定単位が異なるために、それぞれの構成比率を計算してそれによって比較しなければならない。今変量をx、度数をfとすると変量和はxfである。そして、相対度数を$f(x)=\dfrac{f}{\sum f}$相対変量和を$g(x)=\dfrac{xf}{\sum xf}$で表し両者の差をとると、$g(x)=\dfrac{xf(x)}{m}$と書けることから

$$g(x)-f(x)=\left(\frac{x-m}{m}\right)f(x) \qquad (17)$$

となり

$x \lesseqgtr m$ の時は　$g(x) \lesseqgtr f(x)$

であることがわかる。すなわち、xが算術平均mよりも小さい範囲では相対変量和は相対度数よりも小さく、xがmと等しいところでは両者は一致し、xがmよりも大きくなると逆に相対変量和の方が相対度数よりも大きくなるのである。そして(4)より$D \leqq T$であるから、実体分布の山は度数分布の山と一致するかまたは多くの場合それの右側に現れるのである。

以上のことから実体分布と度数分布の分布形態そのものはほぼ同じであるがただ山が多少ずれており、mよりも小さいxの変域では実体分布の方が度数分布よりも低く、mよりも大きいxの変域ではその逆であることがわかった。なお、度数分布が正規分布の場合でも実体分布は対称分布にはならないで極く軽度の正の非対称分布に

なるのである。それは次のことから分かる。今 m よりも小さい変量を x_u、大きい変量を x_0 で表わすと、度数分布が正規であるから m から等距離にある x_0 と x_u に対応する度数は相等しく $f(x_0)=f(x_u)$ である。ところが $x_0>x_u$ であるから変量和は等しくならず $x_0f(x_0)>x_uf(x_u)$ となる。従って、m より上の変量和と m より下の変量和が等しくなるのは x_0 が x_u 以上に m より離れている場合であって、そのことは m より大きい変量の方が m より小さい変量よりも撒布の範囲が広い、すなわち正の非対称分布であることを意味する。なおこのことから、度数分布が軽度の負の非対称の時に始めて実体分布は対称分布になるのであり、また度数分布が負の非対称であっても実体分布は極く軽度の正の非対称になる場合がある、と考えられる。

次に仮想例によって実体分布と度数分布を比較してみよう。度数分布のタイプごとに実体分布の分布形態及び度数分布と実体分布の特性値の関係を知るために、正の非対称分布、正規分布及び負の非対称分布の場合について見ることが必要である。それぞれを仮想例について計算した結果は表3のとおりであって、それを図示したものが図2である。図2の吟味から今までに述べたことが確認される。なお、事物的には無意味であるが参考までに非調和平均の値を計算して示しておいたのであるが、量的標識が連続変量の場合は実体分布が

正の非対称分布の時は　$T<R<A_h$

対称分布に近似する時は　$T \fallingdotseq R \fallingdotseq A_h$

負の非対称分布の時は　$T>R>A_h$

であり、度数分布の場合の DZm の関係と同様の関係が成立することがわかる。

また、図2を見るとa）正の非対称分布の場合が度数分布と実体分布の差の合計（両曲線に挟まれた部分の面積）は

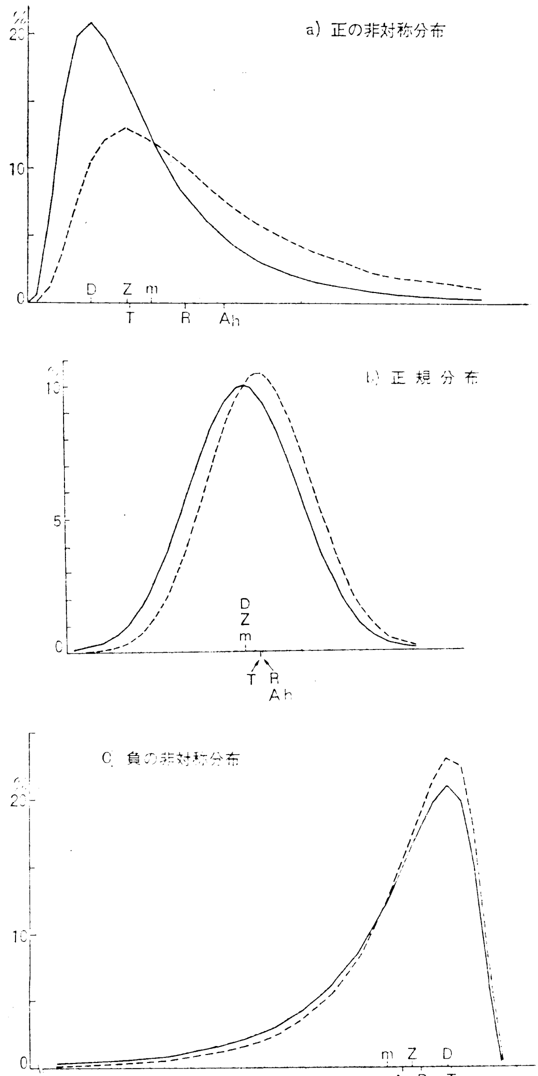

図２　所得階級別人数（実線）と所得金額（破線）

一番大きく、b）正規分布次いでc）負の非対称分布に移るに従ってそれは減少することがわかる。今これらの分布の変動係数 $c=\frac{\sigma}{m}$ を求めると a）77.0%　b）30.8%　c）16.4% であって、変量の撒布はa）が一番大きくc）が一番小さいのである。従って、度数分布と実体分布の差の合計が小さい程変量の撒布はより小さく、中数値のまわりにより密集していることがわかる。そして、その場合は当然実体分布の撒布もより小さくなると考えられる。今加重四分位偏差係数 $\frac{R_3-R_1}{R_3+R_1}$ を計算すると a）0.435　b）0.183　c）0.068 であって、a）が一

表3　所得階級別人数と所得金額（仮想例）

a）正の非対称分布

所得階級	人　数	金　額
万円	%	%
〜 5	0.2	0.0
5〜 10	3.3	0.6
10〜 15	7.5	2.1
15〜 20	9.9	3.9
20〜 25	10.4	5.2
25〜 30	9.8	6.0
30〜 40	16.5	12.9
40〜 50	12.0	12.1
50〜 60	8.5	10.5
60〜 70	6.0	8.7
70〜 80	4.2	7.1
80〜 90	3.0	5.7
90〜100	2.2	4.7
100〜110	1.6	3.8
110〜120	1.2	3.1
120〜130	0.8	2.2
130〜140	0.6	1.8
140〜150	0.5	1.6
150〜160	0.4	1.4
160〜170	0.3	1.1
170〜180	0.2	0.8
180〜190	0.2	0.8
190〜200	0.1	0.4
200〜	0.6	3.4
合　計	100.0	100.0

注1） 人数は　$\log x \in N(1.5441, 0.3010^2)$　として求めた。

2） 所得金額＝中央の値×人数により求めた。

m=44.6万円　Q_3=56.3
D=22.3　σ=34.4
Z=35.4　c=77.0%
Q_1=21.9

$$\frac{Q_3-Q_1}{Q_3+Q_1}=0.439$$
$$\frac{Q_1+Q_3-2Z}{Q_3-Q_1}=+0.219$$

A_h=71.1万円　R_1=35.5
T=36.6　R_3=90.3
R=56.8

$$\frac{R_3-R_1}{R_3+R_1}=0.435$$
$$\frac{R_1+R_3-2R}{R_3-R_1}=+0.223$$
$$\theta=2\{F(m)-G(m)\}=0.537$$

b）正規分布

所得階級	人　数	金　額
万円	%	%
〜 10	0.1	0.0
10〜 20	0.2	0.0
20〜 30	0.3	0.1
30〜 40	0.6	0.2
40〜 50	1.0	0.4
50〜 60	1.7	0.7
60〜 70	2.7	1.4
70〜 80	3.9	2.2
80〜 90	5.3	3.5
90〜100	6.8	5.0
100〜110	8.2	6.6
110〜120	9.3	8.2
120〜130	9.9	9.5
130〜140	9.9	10.3
140〜150	9.3	10.4
150〜160	8.2	9.8
160〜170	6.8	8.6
170〜180	5.3	7.1
180〜190	3.9	5.6
190〜200	2.7	4.0
200〜210	1.7	2.7
210〜220	1.0	1.6
220〜230	0.6	1.0
230〜240	0.3	0.5
240〜250	0.2	0.4
250〜	0.1	0.2
合　計	100.0	100.0

注1） 人数は　$x \in N(130, 40^2)$　として求めた。

2） a）の注 2） 参照。

m=130万円　Q_3=157.1
D=130　σ=40
Z=130　c=30.8%
Q_1=102.9

$$\frac{Q_3-Q_1}{Q_3+Q_1}=0.209$$
$$\frac{Q_1+Q_3-2Z}{Q_3-Q_1}=0$$

A_h=142.2万円　R_1=116.1
T=141.3　R_3=167.9
R=141.9

$$\frac{R_3-R_1}{R_3+R_1}=0.183$$
$$\frac{R_1+R_3-2R}{R_3-R_1}=+0.003$$
$$\theta=2\{F(m)-G(m)\}=0.245$$

c）負の非対称分布

所得階級	人　数	金　額
万円	%	%
〜 50	0.6	0.1
50〜 60	0.1	0.0
60〜 70	0.2	0.0
70〜 80	0.2	0.1
80〜 90	0.3	0.1
90〜100	0.4	0.2
100〜110	0.5	0.2
110〜120	0.6	0.3
120〜130	0.8	0.5
130〜140	1.2	0.8
140〜150	1.6	1.1
150〜160	2.2	1.7
160〜170	3.0	2.4
170〜180	4.2	3.6
180〜190	6.0	5.4
190〜200	8.5	8.1
200〜210	12.0	12.0
210〜220	16.5	17.3
220〜225	9.8	10.6
225〜230	10.4	11.5
230〜235	9.9	11.2
235〜240	7.5	8.7
240〜245	3.3	3.9
245〜	0.2	0.2
合　計	100.0	100.0

注1） 人数はa）の人数を上下入れ替えて求めた。

2） a）の注 2） 参照。

m=205.5万円　Q_3=228.1
D=227.7　σ=33.6
Z=214.6　c=16.4%
Q_1=193.7

$$\frac{Q_3-Q_1}{Q_3+Q_1}=0.081$$
$$\frac{Q_1+Q_3-2Z}{Q_3-Q_1}=-0.219$$

A_h=211.0万円　R_1=200.3
T=228.7　R_3=229.6
R=217.8

$$\frac{R_3-R_1}{R_3+R_1}=0.068$$
$$\frac{R_1+R_3-2R}{R_3-R_1}=-0.193$$
$$\theta=2\{F(m)-G(m)\}=0.114$$

表４

	a）	b）	c）
$\frac{T-D}{T+D}$	0.244	0.042	0.002
$\frac{R-Z}{R+Z}$	0.233	0.044	0.007
$\frac{A_h-m}{A_h+m}$	0.229	0.045	0.013

番大きくｂ）ｃ）と順次小さくなっておりこのことが証明される。なお、度数分布と実体分布の差の合計が小さいことは両分布はより良く近似していることを意味し、従って、同じ性質の分布特性値であるDとT、ZとR及びmとA_hの差は相対的により小さくなるのである。図２の場合でこれを比較すると表４のとおりであって、今述べたことが確認される。

今見たように度数分布と実体分布の差の合計は変量の撒布の程度を反映するのであるが、それでは両者の差の合計はどのような値になるのであろうか、次にそれを求めてみよう。[1] 今変量xを連続変量とし、相対度数（正確には度数密度）を$f(x)$、相対変量和（変量和密度）を$g(x)$とすると

$$\int_{-\infty}^{\infty} f(x)dx=1,\quad \int_{-\infty}^{\infty} g(x)dx=1$$

であるから、度数分布と実体分布の差（但し正負の符号は保存しておく）の合計は

$$\int_{-\infty}^{\infty}\{f(x)-g(x)\}dx=\int_{-\infty}^{\infty}f(x)dx-\int_{-\infty}^{\infty}g(x)dx=0$$

となる。算術平均をmとすると

$$\int_{-\infty}^{\infty}\{f(x)-g(x)\}dx=\int_{-\infty}^{m}\{f(x)-g(x)\}dx-\int_{m}^{\infty}\{g(x)-f(x)\}d(x)=0$$

であるから

$$\int_{-\infty}^{m}\{f(x)-g(x)\}dx=\int_{m}^{\infty}\{g(x)-f(x)\}dx \qquad \text{(i)}$$

すなわち，mより小さいxについての両分布の差の合計とmより大きいxについての両分布の差の合計は相等しいのである。(i)の右辺に(17)を代入すると

$$\int_{m}^{\infty}\{g(x)-f(x)\}dx=\frac{1}{m}\int_{m}^{\infty}(x-m)f(x)dx \qquad \text{(ii)}$$

(ii)の右辺の積分はmより大きいxについて，mよりの偏差$x-m$を相対度数をウエイトとして加重合計したものであるから，それは平均偏差の$\frac{1}{2}$に等しいことになる。なぜならば，平均偏差はmよりの偏差の絶対値を平均したものであるが，算術平均の性質からmより大きいxについての偏差の合計と，mより小さいxについての偏差の合計とが相等しいからである。今平均偏差を算術平均で除して相対化した平均偏差係数をθで表すと，(ii)は次のように書ける。

$$\int_{m}^{\infty}\{g(x)-f(x)\}dx=\frac{\theta}{2},\ \text{但し}\quad \theta=\frac{1}{m}\int_{-\infty}^{\infty}|x-m|f(x)dx \qquad \text{(iii)}$$

(iii)を(i)に代入することにより

$$\int_{-\infty}^{m}\{f(x)-g(x)\}dx=\frac{\theta}{2} \qquad \text{(iv)}$$

(iii)と(iv)の和をとると

$$\int_{-\infty}^{m}\{f(x)-g(x)\}dx+\int_{m}^{\infty}\{g(x)-f(x)\}dx$$
$$=\int_{-\infty}^{\infty}|f(x)-g(x)|dx=\theta\ . \qquad (18)$$

すなわち、度数分布と実体分布の差の合計（両曲線の挟む部分の面積）は平均偏差係数に等しいのであり、算術平均を中心に測定した変量の撒布の程度を表すものであることがわかった。なお(iv)はまた次のように書くことができる。

$$\int_{-\infty}^{m}\{f(x)-g(x)\}dx=\int_{-\infty}^{m}f(x)dx-\int_{-\infty}^{m}g(x)dx=F(m)-G(m)=\frac{\theta}{2} \tag{19}$$

ここに$F(x)$と$G(x)$は相対度数と相対変量和の累積分布を表わす。(19)から相対度数と相対変量和の累積分布において、$x=m$に対応する値の差によって$\theta/2$が得られることがわかる。従って、その差を二倍することにより平均偏差係数の値を容易に求めることができる。

（1）米沢治文、前掲書、六八—七〇ページ。

むすび

以上統計集団の量的構造の解明において従来未開拓のまま残されてきた問題である実体分布の分析の方法について、フラスケンパーによって示された理論的方向に従って整理を行い私見を述べてきたのであるが、考え方においても用語についても不完全な点が多くあり、また更に考察すべき問題も残されており、今後の研究によって補足しなければならないと考える。最後に、度数分布の分析方法と並んで実体分布の分析方法を確立することは、理論統計学のより以上の整備のために必要な事柄であるばかりでなく、従来から行われてきた所得分配の不平等

の測定の問題により完全な理論的基礎を与える点に、その実践上の重要性があることを強調しておきたい。

所得分配の不平等の測定の場合所得人員分布（度数分布）と所得金額分布（実体分布）の関係が考察されるのであるが、それぞれの累積分布を対数に変換して両者の関連を一次式でとらえ、そのパラメーターをもって不平等の測度とするのがジニの集中指数であり、またそれぞれの累積分布を相対化して百分率で表し両者の関連を示したものがローレンツ曲線であって、それと均等分布線（対角線）とが囲む部分の面積をもって不平等の測度とするのである。そして、度数分布と実体分布そのものを比較して両者の特性値の隔りをもって不平等の測度とすることも行われているのであるが、その場合には特に実体分布についての知識がその基礎として必要になるのである。はしがきで述べたように、境界値が所得金額分布の特性値として用いられたのは一九世紀中頃であるが、それが所得分配の不平等の測度に利用されたのは一九世紀末であった。すなわち、ホルムス（G. K. Holms）が一八九二年に境界値と中央値の差$R-Z$をもって所得分配の不平等の測度として提唱したのである。その後ボルトキウィッツ（Ladislaus von Bortkiewicz, 1868—1931）はこれを算術平均で除して相対的測度としたのであるが、田村市郎氏は歪度のいかんにより算術平均は境界値より大きい場合もあればまた中央値より小さいこともあるから、この場合の除数に用いるのは不適当であるとして $\eta=\frac{R-Z}{R+Z}$ をもって不平等の測度とすることを提唱している。[1] しかし、これらの式の不平等の測度としての意義は、度数分布と実態分布の形態を比較した場合に指摘したように、両分布の差の合計が小さい程RとZはより接近するのであり、そしてその時は所得（変量）の撒布はより密である―従って所得の分配はより平等である、という知識を基礎として始めて良く理解し得るのである。更にまた、所得分配の不平等の測度として度数分布の相対的撒布度を用いる方法があるが、その場合には当然同様の権

利をもって実体分布の相対的撒布度も利用さるべきであり、むしろ後者の方がすぐれていると考えられる場合すらあり、そのためにも実体分布の分析方法の確立の大切なことがわかる。

米沢氏は「経済統計計量分析」において実体分布を取り上げることの意義を次のように述べている。(2)「われわれが度数分布のほかに、いわばその背後に隠れていた実体分布を明るみに引き出したのは、この実体分布こそはたんなる度数分布だけの表面的な観察では把握されないような、経済量の分配を明らかにするについての有力な手がかりを与えるものとの問題意識をもって臨んだからであった。」ところが、米沢氏は度数分布と実体分布の隔りを測定する方法として、非調和平均と算術平均の差を算術平均で除して相対化する方法を検討した結果、それが変動係数の自乗に等しいことを知り—これは⑿を c^2 について解くことによって証明することができる—「これではわざわざ実体分布の領分まで立ち入って検討したことが無益であったことになり、やはり度数分布をとり扱うだけで事足りたことを物語っている。」と述べて失望の意を表明している。しかし、われわれは実体分布の領分に立ち入ることが無益であるとは思わないのであって、それの理論統計学における意義及び所得分配の不平等の測定における重要性は先に指摘したとおりである。

(1) 日本統計学会編、前掲書、一八一ページ。
(2) 米沢治文、前掲書、六六ページ。