改訂 経済統計学概論

瀧　　好英
里田武臣　著

八千代出版

改訂 経済統計学概論　　定価 2,800 円

昭和57年 5 月10日　初版印刷
昭和57年 5 月15日　初版発行

著者承認
検印省略

著　者　瀧　好英
　　　　里田武臣
発行者　茅沼　紘
印刷所　神谷印刷株式会社
製本所　三浦製本

発行所　八千代出版株式会社
東京都千代田区三崎町2—2—13
TEL 東京03（262）0420番

はしがき

本書の母体は，言うまでもなく『経済統計学概論』である。同書の初版は昭和46年4月に出版されたが，当時の「はしがき」には次のように記されている。

わが国に推測統計学が導入されてすでに久しい。にも拘わらず，いまなお，統計学は「記述統計学」と「推測統計学」の部門に厳然と二分されている。大げさな表現をするならば，両者は水と油のごとく交わらず，一冊の著書の中で論じられながら，互いにあくまでも一線を画している。

最近，各大学の経済学部では，「経済統計学」なる講座を設ける傾向がみられる。しかし，「経済統計学」の概念は必ずしも定かではない。思うに，「経済現象に関する統計学」および「経済統計についての方法論」といった二様の意味を有するのではあるまいか。前者は経済現象の測定ならびに分析，後者は経済指数・景気指標など二次加工的統計作成の実証的方法論，といった意味においてである。

経済統計学をこのように識認するとき，記述統計学・推測統計学というような区分を固執するのは，もはや適当ではない。両分野の理論体系は，経済統計学なる一連の体系の中に吸収し組み立てられる必要がある。こうした意図から，本書では，推測統計学を経済統計作成までの手段としてとり入れ，作成された経済統計の解析手法には記述統計学の分析理論をとり入れる，さらに，二次加工的経済統計の実際を別途論述する，という考え方で経済統計学の基本構造を組み立ててみた。したがって，たとえば推測統計学については，経済統計作成（データ収集）の段階では事実上ほとんど使用されていない「小標本の理論」や「検定の理論」は，思いきって削

除している。本来の推測統計学について一層の履修・研さんを志す向きは，本書の域に止まることなくさらに専門の著書に進むのが望ましい。

以上の「はしがき」が記されてすでに十余年の星霜が過ぎて行った。経済統計学に対する考え方には少しも変わりはないが，叙述の仕方や構成内容については，加除修正の筆を振いたいと思うところも少なからず生じている。とは言え，日頃の多忙にかまけて，思いきった措置もとれず今日に至ってしまった。今なお，全面的に手を加えるいとまのないまま，とりあえず第5章「統計的関係の分析」の書き直しを行なって『改訂 経済統計学概論』と題することとした。

本書の執筆は，第6章時系列分析 および 第8章景気観測 を里田武臣が，その他の章を瀧好英がそれぞれ担当した。

なお，本書の改訂発行にあたっては，八千代出版社社長茅沼紘氏の多大なお力添えをいただいた。厚くお礼申し上げたい。

昭和57年4月1日

著　者

目　　次

第１章　計算的基礎理論

11　平　　　均

一定の集団について経済量を観察するには，その集団を構成する単位の個々についての情報をまず入手しなければならない。この，個々の情報を入手するために観察することを統計調査という。統計調査によって得られた個々の情報は，これを統計データと呼ぶことにする。

観察した特定の集団について，これを定量的に端的に表現するのに平均値を用いることがある。平均値は，その集団を構成する各単位がどのように分布しているかを説明する手段の一つとして，その分布の中心の値をあらわしている。この意味で，平均値は特定の集団の代表値の一つということができる。しかし，これは，平均値があらゆる意味でその集団を完全に代表しうるということにはならない。平均値は，あくまでもその集団の分布の中心にすぎないからである。

平均値には，その使用目的に応じていろいろな値を考えることができるが，大別すれば

$$\begin{cases}\text{計算的に誘導された値}\\ \text{中心的単位の位置を示す値}\end{cases}$$

に分けられる。さらに，これらは次のように細分することができる。

計算的誘導値

$$\begin{cases}\text{算術平均}\\ \text{幾何平均}\\ \text{調和平均}\end{cases}$$

位置による平均値

$\begin{cases}\text{中位数}\\ \text{4分位数}\\ \text{モード}\end{cases}$

11.1　計算的に誘導する平均法

(1)　算術平均

一つの集団についての統計データが与えられたとき，それらデータの総和をデータの個数で割って得られる平均値を算術平均（arithmetic mean）という。

いま，それらのデータを x_1，x_2，……，x_n とし，その個数を n 個とすれば，算術平均 $\bar{x}$ は次式で定義される。

$$\bar{x}=\frac{x_1+x_2+\cdots\cdots+x_n}{n}=\frac{1}{n}\sum_{i=1}^{n}x_i \qquad 11\text{—}1$$

11.1表　世帯の人員別分布表

世帯人員 (x)	世帯数 (f)	[総人員 (fx)]
1人	1,761	1,761人
2	2,977	5,954
3	3,959	11,877
4	4,866	19,464
5	3,637	18,185
6	2,303	13,818
7	1,295	9,065
8	539	4,312
9	215	1,935
10	88	880
計	21,640	87,251

資料：厚生省「昭和40年国民生活実態調査報告」
（注）原資料では最高の階層を「世帯人員10人以上」としているが，ここでは便宜上「10人」とした。

ここで，x_i は各統計データを表わすものでそれぞれの大きさをもっている。x_i の内容は金額であることもあり数量の場合もあろうが，いずれにせよ統計量を意味している。そこで，統計データを表わす x_1，x_2，……，x_n（略して x_i）を「変量 x」と呼ぶことにする。

たとえば，ここに次表のような統計データが与えられたとする。同表は，家族人員別の世帯数をあらわしているもので，各人員に対応する世帯数（これを度数という）を頻度と

して表示している。このような統計表を度数分布表という。問題は，この統計集団における平均世帯人員を求めるものとする。その場合，変量xを単純に平均しては意味がない。各x_iに，対応する世帯数（f_i）を乗じたものの合計値 $\sum fx$ を求めればそれはその集団に属する総人員であるから，これを世帯総数$\sum f$ で除せば1世帯平均人員が得られる。すなわち

$$\bar{x}=\frac{f_1x_1+f_2x_2+\cdots\cdots+f_nx_n}{f_1+f_2+\cdots\cdots+f_n}=\frac{\sum fx}{\sum f} \qquad 11\text{—}2$$

上表の例について計算するならば，表右の別欄のようにfx，$\sum fx$を計算し，それを〔11—2〕式に代入して

$$\bar{x}=\frac{\sum fx}{\sum f}=\frac{87,251}{21,640}=4.03\text{人}$$

となる。したがって，この集団における1世帯平均人員は4.03人ということになる。4.03人という世帯人員は実際には存在するわけがないが，計算的に誘導したこの集団の代表値と考えればいろいろの点で利便が多い。

算術平均は，理論的にきわめてすぐれた性質を有しているが，実際上も，理解しやすく便利な代表値とされている。そのため，算術平均は「平均中の平均」と言われるほど代表的な平均値で，単に平均というときは算術平均を指すことが多い。本書でも，とくに明記しない限り「平均」は「算術平均」を意味するものとする。

次に，特定の集団の部分集団に関する平均値 $\bar{x}_j$ ($j=1, 2, \cdots\cdots, m$) が与えられたとき，その部分集団の平均値から全集団の平均値を求める方法を考える。

各部分集団の度数がわかっているときは，$f_j\bar{x}_j$ はj集団の総数に等しいから，その合計値 $\sum_j f_j\bar{x}_j$を度数の合計 $\sum_j f_j$ で除せばよい。すなわち，〔11—2〕式のxを$\bar{x}$と置き換えた形になる。もし，各部分集団の度数が相対数（割合，比率）で与えられているならば，その比率を w_j と置けば全集団の平均値$\bar{x}$は次式で求められる。

$$\bar{x}=\frac{\sum \bar{x}_j w_j}{\sum w_j} \qquad 11\text{—}3$$

w_jは第j集団（部分集団）の重み（weight）＝重要度を意味する。重要度を加味して平均する〔11—3〕式の平均方法を加重算術平均法という。〔11—1〕式による平均をとくに区別する必要があるときは，単純算術平均法と呼んで加重方式と区別する。加重算術平均法は，$\bar{x}_j$をx_i，w_jをf_iと置き換えて考えれば，度数分布における平均法（度数平均法）と同じ原理であることが知れよう。

算術平均が理論的にきわめてすぐれたいくつかの特性を有することは前述したが，統計的方法においてはこの特性を利用して理論の展開を行なうことが多い。ここでは，そのうちの主要な特性について述べておく。

特性Ⅰ：$\sum(x-\bar{x})=0$ 11—4

算術平均は，すべての変量を一様にならした場合の値を意味するから，個々の変量と平均値との差の総和は零となる。すなわち，

$$\sum(x_i-\bar{x})=\sum x_i-n\bar{x}=n\bar{x}-n\bar{x}=0$$

$$\because \bar{x}=\frac{1}{n}\sum x_i \qquad n\bar{x}=\sum x_i$$

特性Ⅱ：$\sum(x_i-\bar{x})^2<\sum(x_i-x_a)^2$ 11—5

任意の値x_a（$\neq\bar{x}$）からの偏差の自乗和は，つねに，平均値（$\bar{x}$）からの偏差の自乗和より大である。すなわち

$$\begin{aligned}\sum(x_i-x_a)^2&=\sum\{(x_i-\bar{x})+(\bar{x}-x_a)\}^2\\&=\sum(x_i-\bar{x})^2+2(\bar{x}-x_a)\sum(x_i-\bar{x})+n(\bar{x}-x_a)^2\\&=\sum(x_i-\bar{x})^2+n(\bar{x}-x_a)^2\\&\because\sum(x_i-\bar{x})=0\end{aligned}$$

$$\therefore\sum(x_i-\bar{x})^2<\sum(x_i-x_a)^2$$

$$\because n(\bar{x}-x_a)^2>0$$

(2) 幾何平均

変量 x_i（i=1, 2, ……, n）の相乗積の n 乗根を幾何平均（Geometric mean）という。すなわち，幾何平均は次式で定義される。

$$G=\sqrt[n]{x_1\cdot x_2\cdots\cdots x_n}=\sqrt[n]{\prod_{i=1}^{n}x_i} \qquad 11\text{—}6$$

この式は，手計算はかなり厄介であろうが，対数を利用すれば容易になる。それには，上式両辺の対数を求めれば

$$\log G=\frac{1}{n}(\log x_1+\log x_2+\cdots\cdots+\log x_n)$$

$$=\frac{1}{n}\sum_i(\log x_i) \qquad 11\text{—}7$$

となるから，これを真数に直せば回答Gが得られる。

幾何平均は，変量x_iが比率で与えられているときの平均法としてすぐれた性質をもっている。たとえば２個の商品A，Bの価格が下のように与えられたとしよう。

	5年前の価格	現在価格	倍　　率
商　品　A	100円	200円	2倍
商　品　B	300円	150円	0.5倍

５年前に比べた倍率はそれぞれ２倍，0.5倍で，その平均倍率を求めるものとする。算術平均ならば（2+0.5）÷2=1.25倍となる。しかしこの平均値には不合理な面がある。逆にして現在価格に対する５年前の倍率を平均してみれば容易に納得できよう。５年前の倍率はAが100/200=0.5倍，Bが300/150=２倍であるから，平均倍率は（0.5+2）÷2=1.25倍。どちらの倍率をとっても1.25倍に騰貴しているというのは完全に不合理である。このように，変量が比率であるときは幾何平均を用いるのがよい。幾何平均ならば，５年前に対する現在の倍率の平均は$\sqrt{2\times0.5}$=１倍つまり不変となる。現在価格に対する５年前の倍率を平均しても同じく１倍となる。つまり，商品Aが２倍に騰貴した一方で商品Bは半値に下落したため，平均すれば変わらずと出るのが道理にかなっている。

個々の変量にウエイトを加味して幾何平均を求めるときは

$$G_w = \Sigma w \sqrt{\prod_{i=1}^{n} x_i^{w}} \qquad 11\text{—}8$$

の方法による。これを加重幾何平均という。加重幾何平均の計算にも対数を利用するのが便利である。対数計算は次式によればよい。

$$\log G_w = \frac{1}{\sum_i w_i} \sum_i (w_i \cdot \log x_i) \qquad 11\text{—}9$$

幾何平均は，変量が比率で与えられたときの平均法としてすぐれていることは特色の一つをなしているが，もう一つの特色として，一般に算術平均より小さい値をとる点をあげるべきであろう。つまり，算術平均をM，幾何平均をGで表わせば，一般に

$$M > G$$

の関係を有する。この関係は次の方法で証明することができる（平均はxとyについて計算するものと仮定する）。

$$\frac{M}{G} = \frac{\frac{1}{2}\left(x+y\right)}{\sqrt{x \cdot y}} = \sqrt{\frac{x^2+2xy+y^2}{4xy}}$$

$$= \sqrt{1 + \frac{(x-y)^2}{4xy}} > 1$$

つまり$\frac{M}{G}>1$から$M>G$が証明される。ただし，$x=y$のときは$\frac{M}{G}=1$で$M=G$となる。要は，$x=y$のときのように，同じ大きさの変量について平均値を求めるときは，算術平均の方法でも幾何平均の方法でも結果には変わりがないということである。しかしそのようなケースは実際にはほとんど起こり得ないし，かりに起こったとしても変量がすべて同値ならばわざわざ平均値を計算する必要はないことになる。したがって，一般的に$M>G$の関係を認めてよい。

なお，上の証明過程でも知れるとおり，幾何平均法を用いるときは変量の中に負の値が混っていてはならない点を付言しておく必要があろう。

(3) 調和平均

変量が逆数の形で与えられたときの平均方法は，まず，各変量 x_i の逆数 $1/x_i$ を求め，これを算術平均したのち再び逆数に変換する必要がある。たとえば，2種類の砂糖についての貨幣の購買力（第7章参照）を，「貨幣1円で購入しうる砂糖の量」6グラムおよび7グラムのように与えられたとする。この貨幣の購買力は，それぞれの砂糖の価格の逆数であるから，平均購買力を求めるには，まず平均価格を計算しなければならない。すなわち砂糖Aのグラム当り価格$\frac{1}{6}$円と砂糖Bのグラム当り価格$\frac{1}{7}$円から，

$$\left(\frac{1}{6}+\frac{1}{7}\right)\div 2=0.155\ (円)$$

をうる。この逆数1/0.155＝6.45(g)を求めればA，B2種類の砂糖に関する1円の平均購買力となる。

これら一連の計算過程を一般式で書けば

$$H=\frac{1}{\frac{1}{n}\left(\frac{1}{x_1}+\frac{1}{x_2}+\cdots\cdots+\frac{1}{x_n}\right)}=\frac{n}{\sum_{i=1}^{n}\left(1/x_i\right)} \qquad 11\text{—}10$$

となる。このような平均値を調和平均 (harmonic mean) という。

各変量に度数があるときは，

$$H_f=\frac{\Sigma f}{\Sigma f\frac{1}{x}} \qquad 11\text{—}11$$

度数が相対数（割合い）で与えられたときの加重調和平均は〔11—11〕式の f を w に置き換えた形でよい。

調和平均の利用範囲は広くない。物価指数や生産指数などの総合算式（第7章）としてまれに利用される程度であろう。しかし，経済統計以外では利用度はそれほど低くはない。たとえば平均時速の算出などはその好例である。いま，「1台の自動車が最初の5キロメートルを時速60キロで走り，次の10キロメートルを時速40キロで走ったときの平均時速」を求める問題を考

えてみる。平均時速とは，「走行距離の合計」を「総所要時間数」で割った値であるから，上の例だと「走行距離の合計」は5＋10＝15（km）を使えばよい。所要時間は

$$\begin{cases} \text{最初の区間} & \frac{1}{60}\text{時}\times 5=\frac{5}{60}\text{時} \\ \text{次の区間} & \frac{1}{40}\text{時}\times 10=\frac{10}{40}\text{時} \end{cases}$$

であるから，両区間を通じての延べ所要時間は，

$$\frac{5}{60}+\frac{10}{40}$$

したがって両区間を通しての平均時速は

$$\frac{5+10}{\frac{5}{60}+\frac{10}{40}}=45\ (\text{キロ})$$

となる。最後の計算式が調和平均式であることは言うまでもない。

一般に，調和平均は，算術平均，幾何平均のいずれよりも小さい値をとる。したがって，前述の$M>G$の関係とあわせれば，算術平均（M），幾何平均（G），調和平均（H）の間には，

$$M>G>H$$

の関係が成立することになる。$M>G$の関係はすでに証明されたから，ここでは$G>H$が証明されればよい。例によって，両者の比を展開すれば，

$$\frac{G}{H}=\frac{\sqrt{x\,y}}{\frac{2}{\frac{1}{x}+\frac{1}{y}}}=\frac{\sqrt{x\,y}\left(\frac{1}{x}+\frac{1}{y}\right)}{2}=\sqrt{\frac{(x+y)^2}{4xy}}$$

$$=\sqrt{1+\frac{(x-y)^2}{4xy}}>1$$

$$\therefore G>H$$

となる。ただし，$x=y$のときは $G=H$ である点は前述したところと同じである。

計算的に誘導する平均法としては，上述の平均のほかに次式で定義される

平方平均 (RMS――root mean square) がある。

$$RMS=\sqrt{\frac{\sum x^2}{n}} \qquad 11\text{—}12$$

しかし，この方法は特殊な場合に利用が限定されるので，必要に応じてそのつど述べることとする。

11.2 位置による平均

(1) メジアン（中位数）

変量x_i ($i=1, 2, \cdots\cdots, n$) を大きさの順に並べたときの中央に当る数値をメジアン (median) という。中位数あるいは中央値などの呼び名もある。

与えられた変量が奇数個のときは，メジアンは，当然 $(n+1)\frac{1}{2}$番目の値$x_{(n+1)/2}$であるが，もし変量が偶数個のときは，中央２項の平均をもってメジアンとする。そのときの中央２項とは$\frac{1}{2}n$ 番目と$\frac{1}{2}n+1$番目であるから，メジアンM_eは次式のように書くことができる。

$$M_e=\frac{1}{2}(x_{n/2}+x_{n/2+1}) \qquad 11\text{—}13$$

統計データが度数分布表で与えられたときは，各階級に属する変量は，その階級の度数という情報しか提供してくれず，同一階級内の変量の大小関係はわからない。したがって，それらのすべてを大きさの順に並べることもできない。そのようなときは，次の要領でメジアンを決定する。

たとえば，次表の度数分布表が与えられたとする。変量は全部で21,640個 ($=n$) であるが，それらすべてを所得の順に並べることはできない。なんとならば，各同一階級に属している世帯は所得順位が不明だからである。かりに，21,640世帯が所得の少ない順に並んでいると考えれば，nは偶数であるから，中央値は21,640/2=10,820番目と10,820+1=10,821番目の平均所得ということになる。表右別欄のように累積度数を計算すれば，それら中央２項に該当する世帯は480,000～599,999円の階級に含まれていることがわかる。

11.2表 所得階級別世帯数

所得			世帯数	累積度数
円 0	～	円 59,999	178	178
60,000	～	119,999	488	666
120,000	～	179,999	723	1,389
180,000	～	239,999	985	2,374
240,000	～	359,999	2,496	4,870
360,000	～	479,999	4,139	9,009
480,000	～	599,999	3,493	12,502
600,000	～	799,999	3,961	16,463
800,000	～	999,999	2,105	18,568
1,000,000	～	1,499,999	2,140	20,708
1,500,000	～	1,999,999	512	21,220
2,000,000	～		420	21,640
計			21,640	—

資料：厚生省「昭和40年国民生活実態調査報告」

そこで，それぞれの階級内では，そこに含まれる世帯はすべて等間隔（所得額が等差）に分布しているものと仮定して次のような直線補間法でメジアンの位置を推定する。上例の場合ならば，480,000～599,999円の級幅は120,000円で，その度数は3,493であるから，各世帯の所得差は一様に120,000円/3,493＝34.3544円であるとみなす。一方，中央2項は，集団全体での順位は10,820番および10,821番であるが，所属階級内だけで考えれば10,820－9,009＝1,811番目およびその次の1,812番目に当たる。したがって，それら世帯の推定所得額は

480,000円＋34.3544円×1,811＝542,216円
480,000円＋34.3544円×1,812＝542,250円

となる。そこで，平均値（542,216円＋542,250円)/2＝542,233円がメジアンとなる。

以上の計算過程を一般式で書けば次のようになる。

$$M_e = x_a + C\left(\frac{\frac{n}{2} - F}{f_m}\right) \qquad 11\text{—}14$$