統計平均數在經濟分析中的应用問題

·統計論文集之三·

統 計 出 版 社

統計平均数在經济分析中的应用問題

·統計論文集之三·

統計出版社

1958年·北京

統計平均数在經济分析中的应用問題

·統計論文集之三·

*

統 計 出 版 社 出 版

（北京复兴門外三里河）

北京市書刊出版業营業許可証出字第 075 号

国家統計局印刷厂印刷

新 华 書 店 發 行

*

書号：3006.95·787×1092耗 $^{1}/_{32}$·4 $\frac{5}{16}$印張·90,000字

1958年1月第1版

1958年1月第1次印刷

印数：1～2,060　　定价：（19）0.46元

目　录

分組平均数及差量对于制定先进平均定額的意义及应用*

（對計算先進平均數和平均差的几點改進意見）

徐 鍾 济

一、引 言

本文的目的在于說明分組平均数（先进平均数或落后平均数）及差量（平均差）等对于制定技术定額的意义及应用，并按統計工作多快好省的原則，提出一些建議，也是学習苏联的一点体会。

过去統計学者受了唯心主义观点的影响，常常孤立地来看平均数，而忽略了同时来看差量的重要意义。这样就使平均数的应用大大地受到限制了。事实上，統計学中的平均数和差量标志着总体的兩种不同的性質，而且平均数和差量二者之間是有密切联系的，二者联合使用，才能对总体提供全面的說明。先进平均数是国民經济各部門制定計划定額的基础。苏联部長会議当批准1947年度計划时，曾通令各部門規定計划数字不应以算术平均定額为基础，而应以先进平均定額为基础①。苏联統計学者們根据这一

* 1956年3月31日在浙江师范学院第一次科学討論会上宣讀的論文。

① 参看"苏联部長会議关于1947年国家計划的决議"。

指示，新穎地創造出先进平均数的理論。苏联奥斯特魯莫夫教授所著“新統計学概論”上册，关于先进平均数論述甚詳①。必須注意的是，先进平均数是一种平均数，同时在实質上也是一种差量。从后面正文將看出这种差量相当于先进平均数与一般平均数之差（也相当于先进平均数与落后平均数之差的一半），也將看出先进平均数相当于算术平均数加上（或减去）一倍这种差量的水平。所以先进平均数是把平均数与差量融合一起，这可算是苏联学者們对平均数的一种改革与利用，这样就使平均数能更好地为国民經济計划服务。

还是在五年以前，华伯泉同志因为苏联奥氏的“新統計学概論”一書对于分組資料的先进平均数的計算，有过高或过低的缺点，曾提出对分組的先进平均数及平均差的改正的計算方法②。作者因学習苏联并結合个人的体会，曾逐步补充，說明了分組平均数与平均差之間的密切联系，編入当时浙江大学高級統計講义內③，供教学参考并答复华君。后来虽和华君的論文合出“論先进平均数”一書（由立信会計圖書用品社出版），但沒有加以系統整理。今年(1956年)国家統計局召开第五屆全国統計工作会議，特别提出把統計工作做得又多又快又好又省的方針，使作者

① C.奥斯特魯莫夫：“新統計学概論”上册，1950年东北統計局譯印。

② 华伯泉：“先进平均数的正确計算方法及其应用”，“平均差的正确計算方法及其应用”。

③ 徐鍾济：“論中位数及平均差”，“論先进平均数之簡法及其在統計学上的应用”，“再論先进平均数及平均差”。

（注②、③所列五文曾合併編成“論先进平均数”一書，于1951年由上海立信会計圖書用品社出版。）

获得很多的啓發，因把前文重新整理，并提出更多的方法与建議。惟个人認識有限，謬誤必多，还請学者們多予批評并提出宝貴的意見为幸。

二、对先进平均数的計算及分組問題的商榷

先进平均数就是一个变数的一群变量中，先进于总体算术平均数的那部分的算术平均数。所謂"先进"，必須視所測度的变数而定：如欲測定單位时間內产品定額时，我們当然希望單位时間內的产品愈多愈好，那末"先进于总体的算术平均数"便应体会为"大于总体的算术平均数"；如欲測定單位产品所消耗的时間、材料或电力等有关成本的定額，我們当然希望單位产品的成本愈低愈好，那末"先进于总体的算术平均数"便应体会为"小于总体的算术平均数"。下面为了說明便利起見，我們不妨先以有关生产的定額为例。用公式表式，設$\bar{X}$为一群变量的算术平均数，X_a和X_b依次为大于和小于算术平均数的各个变量，N_a和N_b依次为大于和小于算术平均数各变量的个数。那么先进平均数的公式为：

$$\bar{X}_a = \frac{\Sigma X_a}{N_a} \qquad (1.1)$$

而落后平均数为：

$$\bar{X}_b = \frac{\Sigma X_b}{N_b} \qquad (1.2)$$

很明显，对于有关成本的定額，这两个公式的意义恰好相反。也就是以$\bar{X}_b$为先进平均数。

例1、某車間制造某种机器零件，10个工人每日的平均生产量（个数）如下：

15，16，17，17，18，19，19，20，22，25

兹求其先进平均数。

〔解〕（1）$\bar{X}=\frac{\Sigma X}{N}=\frac{188}{10}=18.8$个

（2）$\Sigma X_a=19+19+20+22+25=105$

（3）$\bar{X}_a=\frac{\Sigma X_a}{N_a}=\frac{105}{5}=21$个（先进平均数）

例2、如果上例只有9个工人，他們的每日平均生产量（个数）如下：

16，17，17，18，19，19，20，22，25

兹求其先进平均数。

〔解〕（1）$\bar{X}=\frac{\Sigma X}{N}=\frac{173}{9}=19.2$个

（2）$\Sigma X_a=20+22+25=67$

（3）$\bar{X}_a=\frac{\Sigma X_a}{N_a}=\frac{67}{3}=22$个（先进平均数）

例3、設有某厂制造每台車床所消耗的鋼材公斤数如下①：

325，350，352，355，358，360，363，370，387，420

这是材料消耗定額，現在求先进平均定額。

〔解〕（1）$\bar{X}=\frac{\Sigma X}{N}=\frac{3,640}{10}=364$公斤（每台車床）

（2）$\Sigma X_b=325+350+352+355+358+360+363=2,463$

（3）故先进平均定額为$\bar{X}_b=\frac{\Sigma X_b}{N_b}=\frac{2,463}{7}$

① 重工業部：“工業企業計划工作”，第40頁，1951年。

$=352$公斤（每台車床）

例4、假如上例只有9台車床，每台可消耗的鋼材公斤数如下：

325，350，352，355，358，360，363，370，387

現在求先进平均数

〔解〕（1）$\overline{X}=\frac{\Sigma X}{N}=\frac{3,220}{9}=357.8$公斤

（2）$\Sigma X_b=325+350+352+355=1,382$

（3）先进平均定額为$\overline{X}_b=\frac{\Sigma X_b}{N_b}=\frac{1,382}{4}$

$=345.5$公斤

比較例3和例4，可以看出如果少去一个突出的落后者，那么先进平均定額（材料消耗）就节約了6.5公斤（每台車床）。

例5、某工厂有工人200名，各工人之每日平均生产量之次数分配如下，試計算每人每日之平均生产量，假定所計算出的定額，在其他一切条件不变的情形下，我們將以它作为計划的基础。

表1　某厂工人每日平均生產量之次数分配

每日平均生产量个数	工人数
20—30	20
30—40	40
40—50	56
50—60	44
60—70	24
70—80	6
总計	200

这是一种分組資料，奥氏曾介紹一种方法，华氏則就理論上对于先进平均数的計算方法，提出了一些补充意見。我們首先介紹奥氏及华氏計算的根据。資料經分組后，計算先进平均数之大小，因每一个体之数值无从确知，故通常次数分配应用兩种假定之一：（1）每組距中全部次数之数值常集中于組距之中点，（2）各个次数之数值均匀分配于組距之間。奥氏系根据第一种假定，华氏系根据第二种假定。

根据第一种假定求先进平均数，則算术平均数$\bar{X}$所在組之N_m个数值均集中于該組之中值X_0，可設Xa为算术平均数所在組以上（大）各組之中值，f_a为以上各組之次数，N_a为以上各組次数之和。若組中值X_0大于算术平均数$\bar{X}$，則先进平均数为

$$\bar{X}_a=\frac{\Sigma Xafa+X_0N_m}{Na+N_m} \qquad (2)$$

若組中值X_0小于算术平均数，則先进平均数为

$$\bar{X}_a=\frac{\Sigma Xafa}{Na} \qquad (3)$$

本例（計算見下頁表2）$\bar{X}=\frac{\Sigma Xf}{N}=\frac{9,500}{200}=47.5$个

算术平均数所在組之中点为$X_0=45$，适小于算术平均数，故先进平均数应为：

$$\bar{X}_a=\frac{\Sigma Xafa}{Na}=\frac{5,080}{84}=60.5\text{个}$$

此法虽符合先进平均数之条件，然每因次数表中所选擇之組距及組限不同，而算术平均数所在組之中值亦随之变

动，有时大于算术平均数，有时小于算术平均数，求得之先进平均数因而迴異，有过高或过低之弊，故結果粗略，不甚可靠。

表2　某厂工人每日平均生產量先进平均数之計算

組距（个）	f	X	Xf
20—30	20	25	500
30—40	40	35	1,400
40—50	56	45	2,520
50—60	44 ⎫	55	2,420 ⎫
60—70	34 ⎬ 84	65	2,210 ⎬ 5,080
70—80	6 ⎭	75	450 ⎭
总　計	200		9,500

根据第二种假定，設 u 为算术平均数所在組之上限，c为組距，f_u为算术平均数文至上限u間所含之次数，即

$$f_u=\frac{(u-\overline{X})}{c}N_m \qquad (4)$$

X_u为算术平均数至上限u之中点，即

$$X_u=\frac{\overline{X}+u}{2} \qquad (5)$$

则先进平均数为

$$\overline{X}_a=\frac{\Sigma X_a f_a+X_u f_u}{N_a+f_u} \qquad (6)$$

由此法求得之先进平均数所受次数表中組距及組限变动之影响甚微，故較为精密。

上例（例5）$\overline{X}=47.5$，$u=50$，$N_a=84$，$c=10$，

$N_m=56$

$$f_u=\frac{(50-47.5)}{10}\times 56=14,\quad X_u=\frac{47.5+50}{2}$$

$$=48.75$$

先进平均数可由表3直接計算，即得

表3　某厂工人每日平均生産量之先进平均数之計算

組　　距	f*	X*	X*f*
47.5—50	14	48.75	682.5
50—60	44	55	2,420
60—70	34	65	2,210
70—80	6	75	450
总　　計	98		4,762.5

* 上表之 f 为各f_a及f_u，X为各X_a及X_u

$$\bar{X}_a=\frac{\Sigma X^*f^*}{\Sigma f^*}=\frac{\Sigma X_af_a+X_uf_u}{N_a+f_u}=\frac{4,762.5}{98}=58.8\text{个}$$

实則公式(6)尚可更为化簡，設$d=\dfrac{\bar{X}-X_0}{c}$，則

$$\frac{X_u-X_0}{c}=\frac{1}{c}\left(\frac{\bar{X}+u}{2}-X_0\right)=\frac{u-X_0}{2c}$$

$$+\frac{\bar{X}-X_0}{2c}=\frac{1}{2}\left(\frac{1}{2}+d\right),$$

$$X_u=X_0+\frac{c}{2}\left(\frac{1}{2}+d\right)$$

$$f_u=\frac{u-\bar{X}}{c}N_m=\frac{X_0+\frac{1}{2}c-\bar{X}}{c}N_m=\left(\frac{1}{2}-d\right)N_m$$

把 X_u，f_u代入(6)式，即得

$$\bar{X}_a=\frac{\Sigma X_a f_a+X_0 f_u+\frac{c}{2}\left(\frac{1}{4}-d^2\right)N_m}{N_a+f_u} \quad (7.1)$$

如再設 $x'=\frac{X-X_0}{c}$ 或 $X=X_0+cx'$,則

$$\bar{X}_a=X_0+\frac{c\left[\Sigma x'_a\ fa+\frac{1}{2}\left(\frac{1}{4}-d^2\right)N_m\right]}{Na+fu} \quad (8.1)$$

同样，設 f_l 为下限 l 至算术平均数的次数，即

$$f_l=\frac{\bar{X}-l}{c}N_m=\frac{X-X_0+\frac{1}{2}c}{c}N_m=\left(\frac{1}{2}+d\right)N_m,$$

即得落后平均数

$$\bar{X}_b=\frac{\Sigma X_b f_b+X_0 f_l-\frac{c}{2}\left(\frac{1}{4}-d^2\right)N_m}{N_b+f'_l} \quad (7.2)$$

和

$$\bar{X}_b=\frac{c\left[\Sigma X'_b\ f_b-\frac{1}{2}\left(\frac{1}{4}-d^2\right)N_m\right]}{N_b+f'_l} \quad (8.2)$$

上例(例5)由表2，已計算得$\bar{X}=47.5$, $X_0=45$, $c=10$

$$d=\frac{47.5-45}{10}=0.25,\quad Na=84,\quad N_m=56,$$

$$N_b=20+40=60$$

$$f_u=\frac{50-47.5}{10}\times 56=14,\quad f_l=\frac{47.5-40}{10}\times 56=42,$$

故由公式(7.1)

$$\overline{X}=\frac{5,080+45\times14+\frac{10}{2}\left(\frac{1}{4}-0.0625\right)\cdot56}{84+14}$$

$$=\frac{5,080+630+\frac{1}{2}(140-35)}{98}=\frac{5,762.5}{98}$$

$=58.80$个（先进平均数）

由（7.2），

$$\overline{X}_b=\frac{1,900+45\times42-\frac{10}{2}\left(\frac{1}{4}-0.0625\right)\cdot56}{60+42}$$

$$=\frac{1,900+1,890-52.5}{102}=\frac{3,737.5}{102}$$

$=36.64$个（落后平均数）

学者如用公式（8.1）和（8.2）計算，亦可得同样結果，但計算则更为簡單（計算見表4）。

表4　某厂工人每日平均生產量之分組平均数（簡法）

組距（个）	中值	f	x′	x′f
20－30	25	20	－2	－40 } －80
30－40	35	40	－1	－40
40－50	45	56	0	
50－60	55	44 } 84	1	44 } 130
60－70	65	34	2	68
70－80	75	6	3	18
总計	—	200	—	$\Sigma\|x'\|f=210$

由上表，$\overline{X}=45+\frac{130-80}{200}\times10=47.5$

$$X_0=45,\quad c=10,\quad d=\frac{47.5-45}{10}=0.25$$

$$N_a=84,\quad N_m=56,\quad N_b=20+40=60,$$

$$f_u=\frac{50-47.5}{10}\times 56=14,\ f_l=\frac{47.5-40}{10}\times 56=42$$

$$\Sigma x_a' f_a=130,\ \Sigma x_b' f_b=-80$$

故由公式（8·1）

$$\overline{X}_a=45+\frac{10\left[130+\frac{1}{2}\left(\frac{1}{4}-0.0625\right)\cdot 56\right]}{84+14}$$

$$=45+\frac{10(130+5.25)}{98}=58.80\text{个(先进平均数)}$$

由（8·2）

$$\overline{X}_b=45+\frac{10\left[-80-\frac{1}{2}\left(\frac{1}{4}-0.0625\right)\cdot 56\right]}{60+42}$$

$$=45+\frac{10[-80-5.25]}{102}=36.64\text{个（落后平均数）}$$

上面已經介紹了先进平均数的計算，現在所欲提出討論的，則为計算先进平均数时之分組問題。

在方法上，最初系以算术平均数为中心，將总体划分成兩个組，第一个組是超过了总体的部分，卽先进部分，第二个組是落后部分。然后需要算出分組了的各組平均指标，先进平均数就是代表先进部分的指标，它不反映总体中落后部分的工作水平。奥氏特别指出，在这个方法中，存在着平均法与分組法之間的不可分的联系，和它們間辨

証的統一，这也就是先进平均数的科学方法論的基础。

在这里發生下面几个問題：

(1)需要先計算算术平均数，然后分組，最后再計算先进平均数，手續較繁。

(2)算术平均数易受兩極端的影响，如果总体中存在一二个突出的落后工作者，那末它的算术平均数就大大降低，先进平均数也随之降低。

(3)以算术平均数为中心来分組，随着地点不同和时期不同，先进部分与落后部分的比例也在变动，这样計算出的先进平均数，就不便于作不同地点和不同时期的比較。

由于以上的問題，使我們想起，計算平均数的方法有数种，我們是否亦可用其他平均数为中心，將总体划分成兩个組計算各組平均数。

奥氏的“新統計学概論”(上册72頁)提及列宁在其經典著作“俄国資本主义的發展”一書中，曾利用革命前俄国用極其复杂的标志編制的且不能直接相互比較的資料，將其重新分組，把俄国全部农民分成三个阶層：一为占全体农戶数百分之五十的貧农低級組，一为占全体农戶数百分之二十的富农高級組，又一則为介于二者間的百分之三十的中級組。这种特殊分組法把这些資料重新分組，一点也沒有改变农民的高級阶層和低級阶層之間的实际关系，但能使我們从極复杂标志中得出具有明确标志的三个大組，才能把不同地区不同农民的分化資料加以比較。显然这个划分貧农阶層之标准是以全部农戶役畜分配之中位数为中心。

得到以上的啓示，我們的問題就更为具体化，即計算

先进平均数时，是否亦可以中位数为中心，將总体划分为兩个組，一个是先进部分，另一个是落后于总体的部分，兩个組各占百分之五十。这种分組法不仅因中位数較算术平均数更自然更具体，把全国資料，或同一地方不同时期的資料，譬如生产定額資料，一致的划分为相等的兩个部分，各占百分之五十，使相互比較时更便利些；且以中位数为中心来分組，每組各占百分之五十，即使偶然有一二个突出的落后工作者，先进平均数也不会受其影响，也就是不会降低；同时，我們必要时还可計算兩种先进平均数，以便选擇其中較高的一个。至于計算方法，如以中位数为中心，將总体划分为兩个組，也有不少便利之处。

三、建議一个新先进平均数——百分之五十的先进平均数

基于上节討論的理由，我們建議一个以中位数为中心来分組的新的先进平均数。中位数乃平分一群变量为兩部分，使大于中位数之項数与小于中位数之項数相等（各占百分之五十）的一个数值。設以数序表示，則N个变量按大小次序排列时（設为 X_1，X_2，……，X_N）；中位数的数序为 $\frac{N+1}{2}$。若N为奇数，那末中位数即为最适中的变数（設 $N=2m+1$，那末中位数是 $M_d=X_m$），若 $N=2m$ 为偶数，那末按次序排列时，并沒有最适中的一个变量存在。这时候居中兩个变量（X_m 和 X_{m+1}）間的任何数值，均符合于中位数的条件，惟習慣上常以居中一对变量的中点的数值为中位数〔也就是 $M_d=\frac{1}{2}(X_m+X_{m+1})$〕，