中央統計聯合會

統計演講集

中央統計聯合會編

中華書局印行

民國二十六年十月印刷
民國二十六年十月發行

大學用書 中央統計聯合會統計演講集（全一冊）

◎實價國幣二元
（郵運匯費另加）

編者 中央統計聯合會
發行者 中華書局有限公司
代表人 路錫三
印刷者 上海澳門路 中華書局印刷所
總發行處 上海福州路 中華書局發行所
分發行處 各埠 中華書局

（本書校對者嚴敦傑）
（一一九〇一）

序

中央各院部會，前以鑒於國內統計事業，亟應通力合作，妥籌改進，爰於民國十九年組設中央統計聯合會．成立以來，既經隨時集會，致力於意見之交換，工作之溝通；而感於專門事業之處理，尤需學識之灌輸，復於二十二年七月起至翌年六月止，按月一次，分別邀約國內統計專家，假地中央大學及計政學院，舉行統計學術演講，繼以討論，藉資切磋．講題分為原理及實驗兩部份．計會期歷十二次，每次參加聽講者，除各會員機關全體統計人員外，尚有其他各機關之職員與學校之學生，凡三百餘人，誠為一時之盛舉．茲以演講竣事，前者臨時印發之演講大綱，業已陸續送罄，而各方需要參考，投函索取者，仍紛至沓來，爰請原來講師，各就初稿，詳予增訂，彙成專集，公諸同好．值茲付梓徵序，用述梗概焉．

中華民國二十五年二月吳大鈞序於國府主計處

中央統計聯合會統計演講集

目　次

甲　原理部份

序

1. 平均數　朱君毅……1

一、緒論　二、算術平均數　三、中數　四、衆數　五、幾何平均數　六、倒數平均數

2. 相差度　褚一飛……22

一、導言　二、相差度問題　三、全距離差距和四分位差　四、平均差與標準差　五、均互差　附錄——關於平均差計算公式的討論

3. 常態曲線與機誤　鄭堯柈……61

一、本講之目的　二、誤差法則之意義　三、常態曲線之面積　四、常態曲線之標準誤差　五、機誤　六、各種統計值之可靠性　七、計算可靠性時所使用各數值表

4. 相關　王書林……90

一、相關之意義　二、直線相關　三、非直線相關　四、等級相關

5. 時間數列　李成謨……113

一、時間數列之組合性及其分解　二、長期趨勢之決定　三、最小二乘方法　四、最小二乘方法之批評　五、其他決定長期趨勢之法　六、長期趨勢之去除　七、季節變動之性質及其重要　八、同月平均法　九、同月平均法校正法　十、環比中位數法　十一、移動

平均比率法　十二、長期趨勢比率法　十三、循環變動之重要　十四、循環變動之求法　十五、循環變動與不規則變動　十六、時間數列之相關

6. 表列與圖示法　鄭堯柈……163

一、本講之目的　二、製表　三、整理　四、統計圖

乙　實驗部份

7. 戶口普查　吳大鈞……212

一、世界人口實數之缺乏與我國之責任　二、戶口普查之意義　三、戶口普查之功用　四、外國戶口普查之沿革　五、我國戶口普查之沿革　六、我國正確戶口普查之迫切需要　七、各國戶口普查方法概述　八、普查之基本原則　九、我國過去各地戶口普查之方法概述

8. 農業淸查　唐啓宇……253

一、農業淸查之意義與目的　二、農業淸查之範圍　三、農業淸查之組織及人員　四、農業淸查之經費　五、農業淸查之步驟　六、農業淸查舉行時之懲罰　七、農業淸查方法之特點及其效果　附錄一——一九三〇年美國第十五屆普通農場淸查表　附錄二——江蘇句容縣農場淸查表

9. 工業淸查　蔡正雅……293

一、工業之定義及其經營制度　二、工業淸查與工業分類法　三、我國工業調查之舉例及進行方法之商榷

10. 生命統計　陳長蘅……308

一、生命統計之意義及範圍　二、生命統計與人口統計之關係　三、

估計人口之方法　四、人口之性分配與年齡分配　五、生命登記之種類　六、生育率、死亡率、結婚率及疾病率等之計算方法　七、生命表之意義及製表方法之簡略說明　八、結論

11. 物價指數　盛　俊……………………………………………329

一、指數之意義　二、物價指數之功用　三、物價指數之沿革　四、物價指數之編製方法　附表一——我國之物價指數　附表二——各國物價指數所採用之計算公式

12. 測驗　沈有乾……………………………………………356

一、測驗之意義　二、測驗結果中之誤差　三、測驗分數之相對性　四、測驗之信度　五、測驗之效度　六、測驗分數之估計　七、統計方法在測驗之應用

中央統計聯合會

統計演講集

平均數

朱君毅

一. 緒論

平均數之功用,在用簡單數字表示繁複之事實. 萬千觀察,零亂無章,一經分類,則條理分明.然分類之後,雖綱領已列, 閱者猶難得其要點,故再進一步之手續,在將次數分配中之特徵,作簡單之量的界說.此種手續,卽平均方法;由此方法所得到之簡單數字,卽平均數.

平均數爲變量中之一値,其性質與變量相同. 故變量爲長度,平均數亦爲長度,變量爲重量,平均數亦爲重量.

1.平均數之性質 一妥善之平均數,應含有以下之性質:

(1)界說應嚴格 由估計而得之平均數,受觀察者之主觀影響,不能完全以客觀之事實爲根據.

(2)應根據全部觀察 若不能符合此點, 則平均數不能代表全體分配之特徵.

(3)不宜多含過於抽象之數學性質 若平均數性質簡明, 則讀者易於明瞭.

(4)易於計算 若有兩種平均數,其他各種條件相等,但甲種易於計算,乙種則非,則甲種自勝於乙種.但專重計算之簡易,而忽略他種條件,亦非所宜.

(5)宜固定而不多受"抽樣之變動"的影響　從同一組事實，而得各種抽樣，則各抽樣之平均數，鮮有彼此相同者。但用某種平均數時之彼此相差，必較用另一種平均數時之彼此相差爲大．此二種平均數，以相差較小者爲較妥善．

(6)可用代數方法計算　此爲妥善平均數之最要條件；意卽合併數列之平均數，可用各組合數列之平均數表出之：若一變量爲二個或三個以上之數列組合而成，則全部數量之平均數，應能以其組合部分之各平均數表出之．

2.平均數之種類　平均數之最通用者，計有三種：

(1)算術平均數(arithmetic mean；簡稱 M)，

(2)中數或中位數(median；簡稱 Mdn)，

(3)衆數或範數(mode；簡稱 Mo)．

此三種中尤以第(1)種爲最重要．此外尙有

(4)幾何平均數(geometric mean；簡稱 G)，

(5)倒數平均數或調和平均數(harmonic mean；簡稱 H)．

玆分別討論之：

二．算術平均數

若某變量之各值，X_1、X_2、X_3、……X_n，其次數爲N，則以N除各值之和，所得商數，卽算術平均數．

1.算術平均數之計算法　算術平均數之計算法，可用以下公式表出之：

$$M=\frac{1}{N}(X_1+X_2+X_3+\cdots\cdots+X_n)\cdots\cdots\cdots\cdots(1)$$

或用較簡明之公式，

$$M=\frac{1}{N}\Sigma(X)\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots(2)$$

若變量中之各值，有次數時，可用下式表出之，

$$M=\frac{1}{N}\Sigma(fX) \cdots\cdots\cdots\cdots (3)$$

在上式內：　N＝各值之總次數，

Σ＝各值之總和，

X＝變量中之各值．

倘各值爲之歸入次數分配表時，則各組距之中點值，代表各該組距內之各值．計算算術平均數之公式，可書之如下：

$$M=\frac{1}{N}\,(f\cdot\text{Mid-pt}) \cdots\cdots\cdots\cdots (4)$$

在上式內 Mid-pt 代表各組距之中點值．

關於公式(4)，吾人有兩種假定：(a)任何組距內之各值，均密集於各該組之中點，故可用各該組中點值以代表之；(b)任何組距內之各值，均均勻分配於各該組距之上，故組中點值爲代表數．

但用公式(4)計算算術平均數時，手續頗繁．欲使之簡便，可採用以下兩種方法：(a)將組距作爲測量之單位計算；(b)求眞正平均數與假定平均數之差，而不直接求眞正平均數之值．

設 A 爲假定算術平均數(arbitrary mean)，X 爲各變量，ξ(讀若Xi)爲假定算術平均數與變量中各值之差，則

$$X-A=\xi,$$

$$X=A+\xi,$$

$$\Sigma(fX)=\Sigma(fA)+\Sigma(f\xi),$$

$$\frac{1}{N}\Sigma(fX)=\frac{1}{N}\Sigma(fA)+\frac{1}{N}\Sigma(f\xi);$$

但 A 爲常數，故

$$\frac{1}{N}\Sigma(fX)=A+\frac{1}{N}\Sigma(f\xi),$$

因

$$M=\frac{1}{N}\Sigma(fX),$$

故 $$M = A + \frac{1}{N}\Sigma(f\xi) \cdots\cdots(5)$$

若以 i 爲組距內之單位數，則

$$M = A + \frac{1}{N}\Sigma(f\xi)i \cdots\cdots(6)$$

故用公式(5)(6)時，所計算者，非爲 $\Sigma(f\cdot\text{Mid-pt})$，而爲 $\Sigma(f\xi)$. ξ 爲差數，因其數目較小，故計算便利.若各值歸入組距，則每組距可視爲單位，其手續尤爲簡捷.欲使 ξ 之差數盡量變小，則 A 之值，應從全距之中點處假定之.

茲用公式(6)顯示算術平均數之計算法如下：

第一表　算術平均數之簡捷算法

X	f	ξ	$f\xi$
0 — 5	2	−4	−8
5 — 10	4	−3	−12
10 — 15	5	−2	−10
15 — 20	6	−1	−6
20 — 25	8	0	−36
25 — 30	7	1	7
30 — 35	6	2	12
35 — 40	5	3	15
40 — 45	3	4	12
45 — 50	1	5	5
	47		51

$$M = A + \frac{1}{N}\Sigma(f\xi)i = 22.5 + \frac{1}{47}(51-36)\times 5$$
$$= 22.5 + \frac{75}{47} = 22.5 + 1.6 = 24.1$$
$$\therefore M = 24.1$$

2.算術平均數之性質　算術平均數爲各平均數中之最妥善者，茲

將其性質略述如下：

(1)界說嚴格.

(2)根據全部觀察.

(3)不多含過於抽象之數學性質，故易於明瞭.

(4)易於計算.

(5)予極端量數以相當之權.

(6)若知總和與次數，即可求得之.

(7)變量中各值與算術平均數之差之代數和等於零.

證明：

$$M=\frac{X_1+X_2+X_3+\cdots\cdots+X_n}{N},$$

$$\therefore\quad NM=X_1+X_2+X_3+\cdots\cdots+X_n,$$

$$\therefore\quad (X_1-M)+(X_2-M)+(X_3-M)+\cdots\cdots+(X_n-M)=0.$$

(8)可用代數方法計算：合併數列之算術平均數，可用各組合數列之平均數表出之.

$$M=\frac{N_1M_1+N_2M_2+\cdots\cdots+N_nM_n}{N_1+N_2+\cdots\cdots+N_n}.$$

在上式內： M_1 = 第一數列之算術平均數，

M_2 = 第二數列之算術平均數，

....，

M_n = 第N數列之算術平均數，

M = 合併數列之算術平均數，

N_1 = 第一數列之次數之和，

N_2 = 第二數列之次數之和，

....，

N_n = 第N數列之次數之和，

$N=N_1+N_2+\cdots\cdots+N_n$ = 合併各數列次數之和.

證明：　$M=\frac{\Sigma X}{N}$, $M_1=\frac{\Sigma X_1}{N_1}$, $M_2=\frac{\Sigma X_2}{N_2}$, ……, $M_n=\frac{\Sigma X_n}{N_n}$;

但　$\Sigma X=\Sigma X_1+\Sigma X_2+\cdots\cdots\cdots+\Sigma X_n$,

$\therefore \quad NM=N_1M_1+N_2M_2+\cdots\cdots+N_nM_n$,

$$\therefore \quad M=\frac{N_1M_1+N_2M_2+\cdots\cdots+N_nM_n}{N_1+N_2+\cdots\cdots+N_n}.$$

以上所述算術平均數之性質，可舉一實例說明如下．設中國歷代帝王之人數，及其平均年壽如下表所列，試問各代帝王平均年壽多少？

朝代	帝王人數	平均年壽(以年爲單位)
漢	25	33.70
魏晉	26	40.75
宋至隋	52	32.79
唐五代	33	45.65
宋	42	48.90
元	9	40.80
明	14	45.00
清	9	49.95
	210	

$$M=\left(\frac{1}{25+26+52+33+42+9+14+9}\right)(25\times 33.70+26\times 40.75$$
$$+52\times 32.79+33\times 45.65+42\times 48.90+9\times 40.80+14\times 45.00$$
$$+9\times 49.95)$$
$$=\frac{8614.08}{210}$$
$$=41.0194.$$

(9)算術平均數，或不發現於實際資料之中．

(10)倘極端量數缺乏，則算術平均數不能精確決定．

(11)算術平均數,不能在圖示上指出.

(12)在品質測量上,不能求得算術平均數.

3.何時應用算術平均數

(1)倘數列之各值,應各得其相當之權.

(2)倘吾人欲得到最高之可靠性.

(3)倘以後須計算標準差及 Pearson 相關係數等.

三. 中　　數

倘數列中之各值,由小而大,順序排列,中有一點,其上下之量數均爲百分之五十,此點卽爲中數.

1.中數之計算法

(1)倘數列爲間斷者,則中數爲順序排列之第$\frac{N+1}{2}$個量數之值.

$$Mdn=\text{第}\frac{N+1}{2}\text{個量數之值}\cdots\cdots\cdots\cdots(7)$$

(a)倘N爲奇數時,則計算之示例如下:

25,26,27,28,29,30,31,32,33,34,35=11個量數,

$$Mdn=\text{第}\frac{11+1}{2}\text{個量數之值}=\text{第}6\text{個},$$

$\therefore\ Mdn=30.$

(b)倘N爲偶數時,則計算之示例如下:

25,26,27,28,29,30,31,32,33,34,35,36=12個量數,

$$Mdn=\text{第}\frac{12+1}{2}\text{個量數之值}=\text{第}6.5\text{個},$$

$\therefore\ Mdn=30.5.$

(2)倘數列中各量數已歸入次數分配,則中數爲第$\frac{N}{2}$個量數之值.茲舉示例,說明算法如下:

第二表 中數之計算法

X	f	計算
0 — 5	2	$\frac{N}{2}=\frac{47}{2}=23.5$
5 — 10	4	向下遞加：$2+4+5+6=17$
10 — 15	5	$23.5-17=6.5$
15 — 20	6	$\therefore\ Mdn=20+\frac{6.5}{8}\times 5$
20 — 25	8	$=20+4.06=\underline{24.06}$
25 — 30	7	覆驗：向上遞加
30 — 35	6	$1+3+5+6+7=22$
35 — 40	5	$23.5-22=1.5$
40 — 45	3	$\therefore\ Mdn=25-\frac{1.5}{8}\times 5$
45 — 50	1	$=25-0.9375=\underline{24.06}$
	47	

(3)倘應用簡寫符號，第二表之算法，可以公式表出之如下：

$$Mdn=l+\frac{\frac{N}{2}-F}{f}i \quad\cdots\cdots(8)$$

或

$$Mdn=L-\frac{\frac{N}{2}-F'}{f}i \quad\cdots\cdots(9)$$

在上式內：　l＝含有中數組距之低限度，

L＝含有中數組距之高限度，

f＝含有中數組距之次數，

i＝組距之單位數，

N = 總次數,

$F = l$ 以下之次數之和,

$F' = L$ 以上之次數之和.

若將公式(8)、(9)應用於第二表之材料時,則

$$Mdn = l + \frac{\frac{N}{2} - F}{f} i = 20 + \frac{\frac{47}{2} - 17}{8} \times 5 = 24.06,$$

或

$$Mdn = L - \frac{\frac{N}{2} - F'}{f} i = 25 - \frac{\frac{47}{2} - 22}{8} \times 5 = 24.06.$$

(4)倘次數分配,變爲累積圖,則中數可用圖算求得之.其法卽在縱坐標上之中點作一橫線,與橫坐標平行,直至與累積曲線相交切.由交切點作一垂直線,與橫坐標相交切,此交切點,卽中數之值.

2.中數之性質

(1)容易計算,且較算術平均數爲易.

(2)倘所量之物,能依次排列,且最中之數量得知,則各物雖不能盡數測量,或兩端之量數未明,亦可求得中數.

(3)有時品質之估計,完全不能應用數量表示;此時祇有應用中數,而算術平均數完全不能應用.例如英國 Galton 之測量智力,卽用中數表示平均智力者.

(4)中數因不受兩端最大或最小量數之影響,故有時勝於算術平均數.

(5)有時數列之中段量數,雖有密集之象,而仍乏明晰之衆數,此時中數實有勝於衆數之處.舉例如下:

X	f
0 — 5	2
5 — 10	3

10 — 15	6
15 — 20	6
20 — 25	6
25 — 30	4
30 — 35	1

(6)中數之值不因數列中多加幾個數量而大受變更，關於此點中數亦勝於衆數.

〔例〕

1,2,3,7,7,12,20: 中數＝7; 衆數＝7.

1,2,3,7,7,12,20,20,20: 中數＝7; 衆數＝20.

(7)不能用代數方法計算:合併數列之中數,不能用各組合數列之中數列表出之.

〔例〕

1,2,3,4,5之中數＝3,

11,18,19,20,41之中數＝19,

但 3與19之中數＝11,

而 1,2,3,4,5,11,18,19,20,41之中數＝8.

(8)中數之決定,固亦根據全部觀察;但所謂全部觀察者，祇就各量數之比較等第或大小而言,而非若算術平均數之根據精確數量者.

(9)中數有時亦不能代表事實上之實際數值.

〔例〕

15,16,17,18,21,23,25,27: 中數＝19.5.

(10)中數之值,有時毫無代表性質.

〔例〕

15,16,16,16,17,19,19,19,21: 中數＝17,有代表性質者,爲16

與19二值.

(11)以中數乘總次數,並不能得到總和.

3.何時應用中數

(1)倘極端量數對於平均數所發生之影響過大,而吾人欲免除之.例如計算工廠工人之平均工資,或計算某地方人民之平均資產,則欲免除經理或巨富者之影響,可用中數(或衆數)以代表平均數.

(2)倘事實之一部或全部缺乏精確數量.

(3)倘以後有計算四分位點及各百分位點之必要者.

4.四分位點及百分位點　四分位點及百分位點之性質,與中數不同,更與平均數無關,似不宜在此處討論.但其求法,與中數之求法,原理相同,故附述如下:

(1)第一四分位點(first quartile point, 簡稱 Q_1) 倘數列中之各值,由小而大,順序排列,內有一點,其下爲百分之二十五量數,其上爲百分之七十五量數,此點卽爲第一四分位點.其求法可用以下公式表出之:

$$Q_1 = l + \frac{\frac{N}{4} - F}{f} i \qquad (10)$$

或

$$Q_1 = L - \frac{\frac{3N}{4} - F'}{f} i \qquad (11)$$

在上式內之各符號已見公式(8)(9).

(2)第三四分位點(third quartile point, 簡稱 Q_3) 倘數列中之各值,由小而大,順序排列,內有一點,其下爲百分之七十五量數,其上爲百分之二十五量數,此點卽爲第三四分位點.其求法可用以下公式表出之:

$$Q_3 = l + \frac{\frac{3N}{4} - F}{f} i \qquad (12)$$