應用較差法求平均值

S. B. Solot, E. M. darting Jr. 原著
殷　來　朝　譯

一、緒　言

在氣象資料的統計分析工作中，我們往往會遇到那些不完整的資料，就發生計算的問題，例如，我們要想求出某地區上空 50mb 層的平均風，就可看到在 50mb 層的探空紀錄次數，不如那些較低層的紀錄來得多，很顯然的，如果仍使用簡單的平均法去計算這些不完整的紀錄，勢必產生差誤的結果，要想完全克服這種資料不全的困難，當然是不可能的，但可應用較差法來改進這種簡單平均法的結果，這方法就是取相連續二層（或一般相連接的二類）間實際紀錄數值的較差，予以平均即得，其結果比較那些只用不完整資料本身作簡單的平均所得的結果，要好許多。

這種方法並不是作者所創始的，本文的述說，只是爲了有鑒於一般氣象學者都似乎還不明瞭這種較差平均法，故決予以介紹，此方法在應用上近來曾經若干的研究，對於其結果的本質，已顯有改進。

二、較差法的解析

讓我們來研討幾組同時（或近乎同時）觀測的數值 y，牠是隨着某種自變數 x 之變化而變化的，在每組中的 y 諸值，是按照預定相同的各 x 值處測得，這些 x 值稱爲類別（Class），最好的例子，就是無線電探空儀測得隨高度而變的溫度值，此處所謂類別，就是各標準層次，如 1000mb, 850mb, 700mb, 等，在各組資料中的 y值（如溫度），在某類（如高度或標準層）以下是完整的，而在其餘各類則不完全，因爲探空紀錄向上逐漸減少，以致後續各類的資料，均係不完全整齊的，其排列有如下列第一表所示。

第一表　觀測紀錄的排列

令 (i) 爲, ＼ 類別(j)	x_1	x_2	x_3	……	x_a	x_{a+1}	…… $x_{j=L}$
1	$y_{1,1}$	$y_{1,2}$	$y_{1,3}$	……	$y_{1,a}$	$y_{1,a+1}$	…… $y_{1,L}$
2	$y_{2,1}$	$y_{2,2}$	$y_{2,3}$	……	$y_{2,a}$	$y_{2,a+1}$	…… $y_{2,L}$
3	$y_{3,1}$	$y_{3,2}$	$y_{3,3}$	……	$y_{3,a}$	$y_{5,a+1}$	…… $y_{3,L}$
⋮	⋮	⋮	⋮		⋮	⋮	⋮
i=n	$y_{n,1}$	$y_{n,2}$	$y_{n,3}$	……	$y_{n,a}$	$y_{n,a+1}$	…… $y_{n,L}$

表內：i 表示同時觀測所得的一組 y 數值。

j 指明每組 y 數值在某類下所測得的數值。

$y_{i,j}$ 表示 y 在第 i 組第 j 類的觀測數。

j=a 表示 y 紀錄屬於完整的最後一類，在本類中及其以前各類（$j \leqslant a$）有 n 個 y 數值。

j=a+1 表示 y 觀測紀錄不完整部份的第一類，其中有 m 個 y 值，而 $m<n$,（即其紀錄次數少於觀測次數）。

設其平均值之計算方程式如下：

$$_n\overline{y}_j = \frac{\sum\limits_{i=1}^{n} y_{i,j}}{n}$$

又

$$_m\overline{y}_j = \frac{\sum\limits_{i=1}^{m} y_{i,j}}{m}$$

在第 (a+1) 類中全部 y 的平均值爲

$$_n\overline{y}_{a+1} = {}_n\overline{y}_a + ({}_n\overline{y}_{a+1} - {}_n\overline{y}_a \quad \cdots\cdots\cdots\cdots\cdots\cdots(1)$$

我們可以用兩種方法來判斷 ${}_n y_{a+1}$ 的數值，

1, 簡單平均法 (Simple averaging)

$$({}_n\overline{y}_{a+1})_s = {}_m\overline{y}_{a+1} = {}_m\overline{y}_a + ({}_m\overline{y}_{a+1} - {}_m\overline{y}_a) \cdots\cdots\cdots(2)$$

2, 較差平均法 (The difference method)

$$({}_n\overline{y}_{a+1})_D = {}_n\overline{y}_a + ({}_m\overline{y}_{a+1} - {}_m\overline{y}_a) \cdots\cdots\cdots(3)$$

將上列 (2) 式與 (3) 式之右方加以比較，我們可以看出，第二項是相同的，所以二方程式的比較準確率，要從牠們的第一項來決定，因爲 ${}_n\overline{y}_a$ 乃是第 a 類 y 的確實平均值，而 ${}_m\overline{y}_a$ 只是一個估量的平均值，所以從第 (3) 式求得的估量 ${}_n\overline{y}_{a+1}$ 值，恒較第 (2) 式所求得者爲佳。

再將第 (1) 式與第 (3) 式的右方加以比較，我們可以看見其唯一不同之點，在於第二項，其不同之故，在於第 (a+1) 類中有若干次資料不全，當然我們無法克服這種資料不全的困難，不過，如果第 (1) 式及第 (2) 式中的右端第二項與共同的第一項相較，是比較的很小的話，我們也可從第 (3) 式求得到 ${}_n y_{a+1}$ 爲一近似值，其條件就是要連續的二個 y 值之間的變化數小於 y 值本身。

不完整資料的形式有二種：即偏向於一方 (Biased) 者和非偏向於一方 (Unbiased) 者，偏向於一方的資料中，y 值是有系統的偏離前面的數值，(就是最大值或最小值有遠離頭一類的數值之趨勢)，至於非偏向於一方者，則 y 值是不規則的偏離前一數值：

在偏向一方的資料中，第 (2) 式中的 ${}_m\overline{y}_a$ 和第 (3) 式的 ${}_n\overline{y}_a$ 數值的差異是有系統的，偏頗愈大，則以上二值間相差之值亦愈大，而較差平均法比簡單平均法的優越性也更爲增大，在非偏向一方的資料中，則上列二值間的差異是不規則的，因此較差平均法的優越性也不如在偏向一方資料那樣顯著。

至少在理論上，第 (3) 式右邊第二項的絕對值，能由於我們在選擇時，減少 x 的增加量，(例如在相連二類間的間隔) 而使減至最小，當然在實用上，我們很少能控制 x 之增加的，不過較差平均法可靠的適當條件，就是相連二類資料間的 y 值改變量，定要小於 y 值本身的一半才行。

三、舉　　例

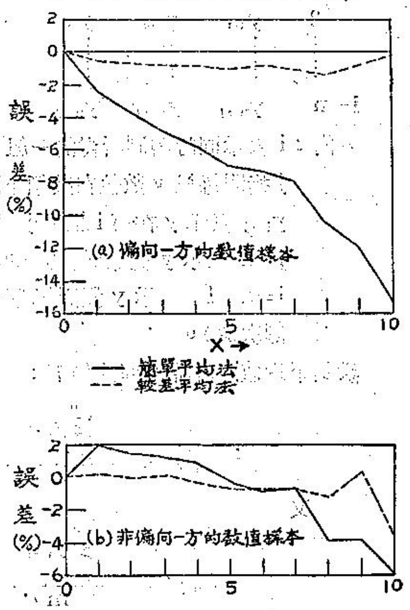
爲表明計算步驟和顯示較差平均法所得的結果，特在此處舉一個統計實例，在一張表內列有三十個 y 組數值，每組都按照 x 值分爲11類，其範圍如下：

$$y = (\overline{y}_0 + y') + (\overline{a} + a')x$$

$$\overline{y}_0 = 50 \qquad \sigma(y) = \pm 10$$

$$a = 10 \qquad \sigma(a) = \pm 3$$

第1圖：統計實驗之結果



上式中右上角帶有撇 (，) 號的字母係代表與各平均值的偏差數，假定 y_0 和 a 都是正常的分佈 (Normal distribution) 從平均值 $\overline{y}_0$ 和 $\overline{a}$ 的偏差數，乃是任意用一個數字表來指定的，在這張總表裏面，有二組不完整的 y 值，從 x=1 起，以後相繼的各 x 直行內 y 值，都予剔除，對于各組，偏向一方的資料，則先後將所餘之 y 最高值略去，在處理非偏向一方的資料時，則將各直行中觀則相同的數值任意予以捨棄，這樣繼續抽剔下去，直至在 x=10 的直行中，僅留下一次觀測數，爲舉例明顯起見，此處只舉出偏向一方的資料 (見第二表)。

在第二表內有兩種變換計算的詳細方法，任一種都可以使用，因爲牠們所得結果完全相同，在缺漏紀錄較少的地方使用第一法較爲方便，在缺漏大部份紀錄時則以使用第二法爲宜。

由上述實驗，我們可以將較差法與簡單平均法在技術上的誤差，加以比較，此種比較的誤差可以百分率表示，如第一圖所示，由圖中可見在偏向一方的資料的誤差，曲線對於所有各 x 值都是負值，這是由於偏向值增多的影響，二條曲線間顯著的不同，是在於簡單平均法中的，誤差是累增的，但在較差平均法則不然，在任意指定的資料中的誤差，是正負兩方都

有發生，但此處仍能顯示出用較差法較爲符合於優良的結果。

由於極多數的氣象資料，都是有若干偏向於一方的差異的，能適用較差法去處理，使用此法之一般結果，都可獲得如第一圖中 a 二曲線本質所示之比較。

第二表　偏向一方資料 y 的排列

X　數：	0	1	2	3	4	5	6	7	8	△	9	△	10	△
觀 測 數														
1	40	51	56	67	79	93	106	117	—	—	—	—	—	—
2	60	(60	(60)	(60)	(60)	(60)	(60)	(60)	—	—	—	—	—	—
3	55	69	82	91	(91)	(91)	(91)	(91)	—	—	—	—	—	—
4	50	61	70	86	(86)	(86)	(86)	(86)	—	—	—	—	—	—
5	50	54	64	77	87	98	102	111	122	11	—	—	—	—
6	35	47	58	63	69	82	91	97	107	10	116	9	127	11
7	40	49	60	72	81	90	102	112	125	13	—	—	—	—
8	40	45	57	64	73	82	95	105	117	12	122	5	—	—
9	45	55	64	75	86	98	110	121	—	—	—	—	—	—
10	55	66	79	(79)	(79)	(79)	(79)	(79)	—	—	—	—	—	—
11	45	55	61	70	77	83	94	104	115	11	127	12	—	—
12	65	74	80	(80)	(80)	(80)	(80)	(80)	—	—	—	—	—	—
13	50	60	68	76	82	96	103	111	118	7	—	—	—	—
14	55	64	68	76	87	94	106	116	121	5	—	—	—	—
15	40	50	66	70	80	88	94	103	114	11	127	13	—	—
16	45	51	61	76	92	100	108	120	—	—	—	—	—	—
17	50	56	65	75	83	88	99	110	118	8	126	8	—	—
18	50	58	72	81	93	102	110	(110)	—	—	—	—	—	—
19	65	72	83	89	100	(100)	(100)	(100)	—	—	—	—	—	—
20	60	68	76	87	91	106	115	122	—	—	—	—	—	—
21	70	(70)	(70)	(70)	(70)	(70)	(70)	(70)	—	—	—	—	—	—
22	35	43	51	58	72	76	87	100	106	6	122	16	—	—
23	30	37	47	55	66	82	98	106	118	12	—	—	—	—
24	55	66	78	90	99	(99)	(99)	(99)	—	—	—	—	—	—
25	45	54	69	82	90	100	110	(110)	—	—	—	—	—	—
26	50	62	73	79	92	98	107	116	—	—	—	—	—	—
27	55	70	77	87	97	104	(104)	(104)	—	—	—	—	—	—
28	45	58	68	82	92	102	(102)	(102)	—	—	—	—	—	—
29	60	73	(73)	(73)	(73)	(73)	(73)	(73)	—	—	—	—	—	—
30	60	74	(74)	(74)	(74)	(74)	(74)	(74)	—	—	—	—	—	—
a	30	28	26	24	22	20	18	16	11	11	6	6	1	1
b	1500	1772	2030	2264	2481	2674	2855	3009	—	—	—	—	—	—
c	—	272	258	234	217	193	181	154	—	—	—	—	—	—
d	—	—	—	—	—	—	—	—	—	106	—	63	—	11
e	—	9.7	9.9	9.8	9.9	9.7	10.1	9.6	—	9.6	—	10.5	—	11.0
f	50	59.7	69.6	79.4	89.3	99.0	109.1	118.7	128.3	—	138.8	—	149.8	—

〔註〕本表計算步驟：

較差平均法係自資料不全之 x 值直行開始應用。(本列中 $x=1$)

第一法：

1. 將每行之空格內，以其最後一記錄數值填滿，並用括弧表示之。
2. 在表底 (d) 行中塡入各直行的確實測量次數，不算有括弧者。
3. 將各直行紀錄數相加成總數 (包含有括號各數)， 並塡入 (b) 行。
4. $c_1=b_1-b_0$,　$c_2=b_2-b_1$,　餘依此類推。
5. (d) 行留空白不塡。
6. $e_1=\dfrac{c_1}{a_1}$,　$e_2=\dfrac{c_2}{a_2}$,　餘依此類推。
7. $f_1=f_0+e_1$,　$f_2=f_1+c_2$;　餘依此類推。

f_n $(n=1, 2, \cdots\cdots, 10)$ 即爲第 n 直行內 y 的平均值，係由較差法估計而得。

第二法：

1. 將相連續各直行間紀錄之較差，塡寫於△直行內。
2. 將各△直行之總數塡寫於 (d) 行內。
3. $l_1=\dfrac{d_1}{a_1}$,　$l_2=\dfrac{d_2}{a_2}$,　餘依此類推。
4. $f_1=f_0+e_1$,　$f_2=f_1+e_2$,　餘依此類推。

(見第一法註7)