民國文獻資料叢編

《社會科學雜誌》全編

陶孟和　曾炳鈞
巫寶三　等　編

國家圖書館出版社

3

《社會科學雜誌》全編

陶孟和　曾炳鈞　主撰

第三冊

國家圖書館出版社

第三冊目錄

書籍評論

《社會科學雜誌》全編

第三冊

第三卷　第一期(一九三二年三月)
第三卷　第二期(一九三二年六月)
第三卷　第三期(一九三二年九月)
第三卷　第四期(一九三二年十二月)

社會科學雜誌

第三卷　民國二十一年三月　第一期

主撰者：　陶孟和　曾炳鈞

社會調查所

北平西安門內文津街

中國郵政總局特准掛號認為新聞紙類

社會調查所

社會調查所原名社會調查部,成立於中華民國十五年。至十八年六月,始改今名。乃一社會研究之學術機關。其事業計劃,約言之,爲:(一)關於社會問題行使學術上之研究與調查;(二)介紹國外調查社會問題及研究社會問題之新技術於中國;(三)將調查研究所得結果報告社會,以備解決國內社會問題之參考;(四)蒐集關於社會問題之圖書及資料以圖研究與閱覽之便利;(五)提倡社會研究之興趣,使專攻社會科學之人士致力於專門實際的研究;(六)與從事社會調查之機關謀合作,協力調查社會問題;(七)指導其他機關之社會調查事業。

目　　錄

第三卷　　第一期

I 論述　　頁數

II 書籍評論

統計學中分割數的問題

楊西孟

1. 中數的原定義

在這裏"分割數"一個名詞是用來包括中數(median)四分分割數(quartiles),十分分割數(deciles)………一類的數量。在一般統計學教本裏,中數歸入平均數一章,而其他的分割數往往附入離中趨勢(dispersion)一章中討論。實在說來,他們的性質相同,合併起來討論較為便利。

分割數中最簡單的一種自然是中數。它的意義最容易懂得。比如有十一個人依高低的次序站成一列,中間的一個便是第六那位,因為他的高度可作全隊高度的一種代表,所以用為平均數之一種,稱之為中數。但如人數不是單數而是偶數,如十二,就有兩個居中的人,第六和第七,問題便較複雜一點,於是統計學者說:假設在第六與第七之間添插一人,他的高度算為那二人的折中,這便用來作中數。

不錯,這種辦法是合理的。統計學者便依據這個思想的[illegible]下了一個中數的定義:把一些項目依小大的次序排成一[illegible]劃全列為二等分,則分割點所在的一項(實有的或假設的)[illegible]數項,這項的數量(如上例中之高度)為中數。

這個定義也算不錯。按照它的意義去求中數,總求得着,而且在一個題目之下,只有一個答案,至少在實用上是行了。

2. 其他分割數的原定義

依大小排成的行列既可分爲二等分,那末,分爲四等分,十等分,百等分,…………又何嘗不可? 又因爲分爲四等分,十等分……在統計學上也很有用處,於是就稱:—

(a)四等分的分割數爲"四分分割數"(quartiles),

(b)十等分的分割數爲"十分分割數"(deciles),

(c)百等分的分割數爲"百分分割數"(percentiles),

(d)三等分的分割數爲"三分分割數"(tertiles),

(e)五等分的分割數爲"五分分割數"(quintiles)。

統計學者並不多加思攷,於是輕輕便便地把中數的定義推廣一下,作爲其他"分割數"的定義。推廣的邏輯是這樣說的:—

中數既然指全列的正中一項。同樣的,四分分割數便是指分割全列數目爲四等分的各項,十分分割數便是指分割全列數目爲十等分的各項,百分分割數便是指分割全列數目爲一百等分的各項…………。[1]

3. 第一類錯誤的公式

這個推廣的定義,初看來似乎很對似的。但是我們且看它能否"兌現,"換句話說,看它在使用的時候是否行得通。

1 這段是仿金氏(King)的說法,其他各統計學者的解釋都與此相似。金氏的原文爲:——

"The median was defined as the middle item of the array. Similarly the quartiles are those items that divide the number of items in an array into fourths, the deciles those that divide it into tenths, the percentiles into hundredths, etc."

見 King: Elements of Statistical Method, p. 153.

更切實點說,假如一個行列中有 n 項,按照這個定義,試問第幾項是Q_1,第幾項是Q_3? 統計學者又輕輕便便地把應付中數的公式推廣一下。因爲中數是第$\frac{n+1}{2}$項,於是就認爲

Q_1是第$\frac{n+1}{4}$的一項,

Q_3是第$\frac{3(n+1)}{4}$的一項,跟着便是:—

D_1是第$\frac{n+1}{10}$的一項,

D_2是第$\frac{2(n+1)}{10}$的一項,

D_3是第$\frac{3(n+1)}{10}$的一項,

⋮

D_9是第$\frac{9(n+1)}{10}$的一項。

除了好些統計學者在所著統計教本內不曾談到這裏以外,鮮明作此種主張的有 Secrist, King, Burgess 諸家。

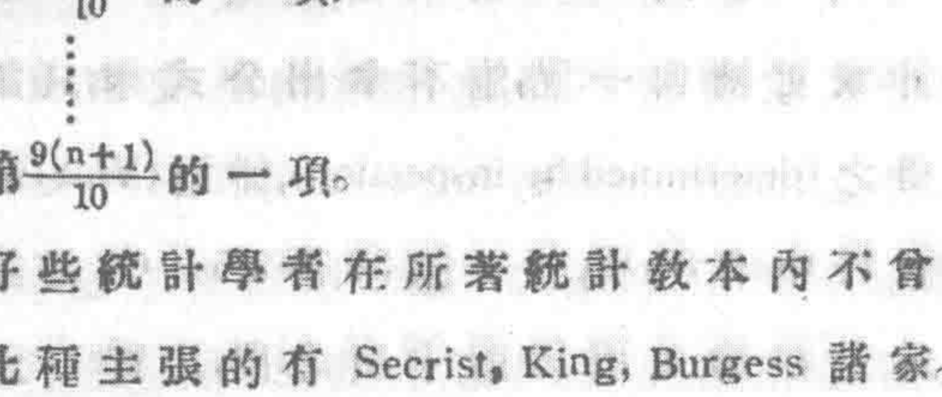
對於他們這套公式,暫且很簡單的試驗一下。比如有四個人排成一列[1],按照這套公式:

Q_1 是第$\frac{4+1}{4}$項,即第 1.25 項;

中數是第$\frac{4+1}{2}$項,即第 2.50 項;

Q_3 是第$\frac{3(4+1)}{4}$項,即第 3.75 項。

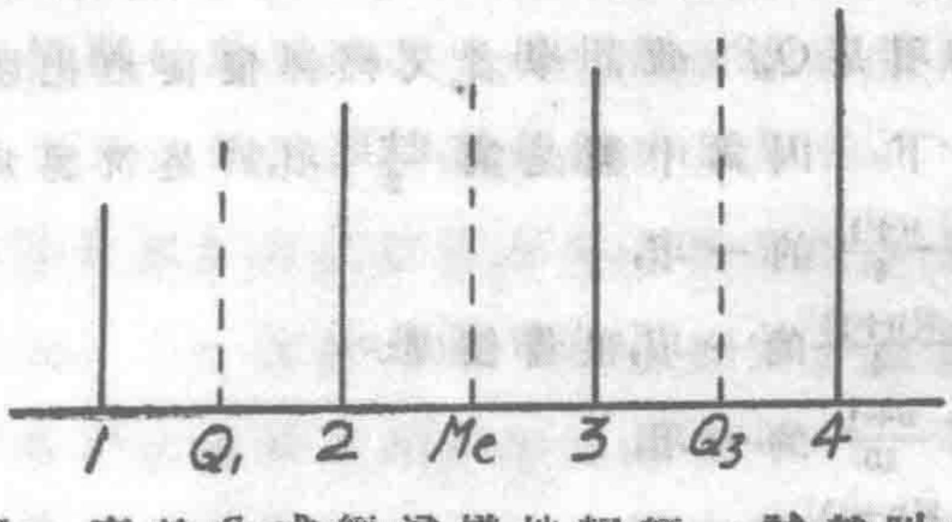
畫一簡圖,就是這樣:—



在這圖一看,我們便覺得Q_1和Q_3太靠兩邊了,依照上面求

1 所謂"列"是指按數量大小的次序排成的直行。下同。

中數的法則,Q_1應在1與2之正中,Q_3應在3與4的正中,如下圖:—



所以這一套的公式經這樣地輕輕一試,就叫佔不住。

4. 郁氏(Yule)的所謂觀察法

有的統計家更精密一點,他不拿出公式來,只說是Q_1和Q_3是可以觀察得之(determined by inspection),郁爾(Yule)就是這樣的。他說,"例如資料中有38項,以4除38得9.5。Q_1是第幾項呢?初學的人以為是在第九項與第十項之間,其實不對。因為那樣一來在Q_1之下的只有九項,而不是九項半。Q_1應該正在第十項。"[1]

郁氏這種觀察法,至多祇是他自己能夠使用,別人是不易懂得的。在他舉的這個例子,共38項,看出Q_1正在第十項,倒還不難。但如是39項,Q_1又往右邊(向第十一項的方向)移動幾分之幾呢?這豈是容易觀察的?看出四分點也就不易,看出十分點,百分點當更加困難。所以郁氏的方法並未給這個問題一個普遍的解決。

5. 包氏(Bowley)的繁公式

於是包氏(Bowley)有精密的方法出來。他給我們下面一套的公式[2]:—

1 Yule: Introduction to Theory of Statistics, p. 147.

2 Bowley: Elements of Statistics, p. 107.

項數	中數	Q_1	Q_3
4n	$\frac{1}{2}\left(2n^{th}+\overline{2n+1}^{th}\right)$	$\frac{1}{2}\left(n^{th}+\overline{n+1}^{th}\right)$	$\frac{1}{2}\left(3n^{th}+\overline{3n+1}^{th}\right)$
4n+1	$\overline{2n+1}^{th}$	$\frac{1}{4}n^{th}+\frac{3}{4}\overline{n+1}^{th}$	$\frac{3}{4}\overline{3n+1}^{th}+\frac{1}{4}\overline{3n+2}^{nd}$
4n+2	$\frac{1}{2}\left(\overline{2n+1}^{th}+\overline{2n+2}^{nd}\right)$	$\overline{n+1}^{th}$	$\overline{3n+2}^{nd}$
4n+3	$\overline{2n+2}^{nd}$	$\frac{3}{4}\overline{n+1}^{th}+\frac{1}{4}\overline{n+2}^{nd}$	$\frac{1}{4}\overline{3n+2}^{nd}+\frac{3}{4}\overline{3n+3}^{rd}$

我把這套公式試驗一番,覺所得的答數沒有毛病。比如仍用上面試驗 $\frac{n+1}{4}$ 一類的公式所用的例子,人數爲4,故4n=4,即n=1。以此代入中數的公式得$\frac{1}{2}(2+2+1)=2\frac{1}{2}$,代入$Q_1$的公式得$\frac{1}{2}(1+1+1)=1\frac{1}{2}$,代入$Q_3$的公式得$\frac{1}{2}(3+3+1)=3\frac{1}{2}$:三個答案都對的。

但是,我在這裏發生了兩個問題:第一,隨便設幾個數來試驗公式,不是澈底的辦法,最緊要的還是從理論上去求證明。第二,就假定在理論上包氏的公式是對的,還要看它在實用上是否便利。

現在我們先討論第二個問題。我們看看包氏的公式共有12個,即中數四個,Q_1四個,Q_3四個。每個公式的形式並不簡單,全體也顯不出什麼便於記憶的規律。假如按照包氏的方法來求十分點,我可以仿照他推出下面一套的公式:—

項數	中數	D_1	......	D_8	D_9
10n	$\frac{1}{2}(5n+\overline{5n+1})$	$\frac{1}{2}(n+\overline{n+1})$		$\frac{1}{2}(8n+\overline{8n+1})$	$\frac{1}{2}(9n+\overline{9n+1})$
10n+1	$5n+1$	$\frac{4}{10}n+\frac{6}{10}\overline{n+1}$		$\frac{7}{10}\overline{8n+1}+\frac{3}{10}\overline{8n+2}$	$\frac{6}{10}\overline{9n+1}+\frac{4}{10}\overline{9n+2}$
⋮	⋮	⋮		⋮	⋮
10n+8	$\frac{1}{2}(\overline{5n+4}+\overline{5n+5})$	$\frac{7}{10}\overline{n+1}+\frac{3}{10}\overline{n+2}$	......	$\frac{1}{10}\overline{8n+6}+\frac{9}{10}\overline{8n+7}$	$\frac{3}{10}\overline{9n+7}+\frac{7}{10}\overline{9n+8}$
10n+9	$5n+5$	$\frac{6}{10}\overline{n+1}+\frac{4}{10}\overline{n+2}$	......	$\frac{3}{10}\overline{8n+7}+\frac{7}{10}\overline{8n+8}$	$\frac{4}{10}\overline{9n+8}+\frac{6}{10}\overline{9n+9}$

我們數一下,求十分分割數的公式有 9×10=90。假如要求百分分割數,就應該有如下的公式:—

項數	中數	…………	P_{99}
100n	$\frac{1}{2}(\overline{50n}+\overline{50n+1})$	…………	$\frac{1}{2}(\overline{99n}+\overline{99n+1})$
100n+1	50n+1	…………	$\frac{51}{100}\overline{99n+1}+\frac{49}{100}\overline{99n+2}$
⋮	⋮		⋮
100n+98	$\frac{1}{2}(\overline{50n+49}+\overline{50n+50})$	…………	$\frac{48}{100}\overline{99n+97}+\frac{52}{100}\overline{99n+98}$
100n+99	50n+50	…………	$\frac{99}{100}\overline{99n+98}+\frac{1}{100}\overline{99n+99}$

照此辦法,求百分分割數公式就該有 99×100=9900。這麼簡單一件事——求百分分割數——要用上約近一萬個的公式!不用說,這種公式是太不合實用了。所以上面提出的第二個問題,即是,包氏的公式是否合於實用,可說是完全得一個反面的答案。

至於上面提出的第一問題,包氏的公式是否在理論上能成立,我們暫且不忙去證明它;因爲包氏並未把他推演公式的程序給我們,我們難以着手。我們還是自己去探討,待有正確解答之後再以我們的解答與包氏的公式相較。

6. 原定義中含有之兩疑點

現在起始我們自己探討的工作。首先要批評的是各種分割數的定義。原前的定義大意是這樣:—

如果把觀察所得的各項依小大的次序排成一列,那嗎,把項數平分的一項爲中數項,把項數四等分的三項爲四分項,把項數十等分的 9 項爲十分項,…………。

我要質疑的是兩點:(一)項數,(二)觀察所得的項數。

(一)求分割數的時候,我們所分割爲幾等分的東西眞是項