1520

密勒氏統計方法論

上　册

徐堅譯

商務印書館發行

自　　　序

小學教師所要用的統計，並不十分多，也不必高深。但是師範學校裏學的統計，却很完備。學時因爲公式多，所以不容易記憶。學過後不能把公式一一實地應用，所以難熟習。某一部份常要用到的，或者反而嫌得簡略。沒有在師範學校裏學過的，往往想找些統計的書自修。自修統計學，比別的科目較難。前編測驗統計術（中華書局印行），在初學者，自修者，以及大多數只須某一部份的實用者，或者還嫌不適用。因爲那書裏仍舊包含統計中計算方法的各方面。方面多，所以敍述或者欠詳盡。現在這一本，不求完備，揀最重要，最常用的幾種，用更多的實例，作更詳細的說明。沒有學過的人可以拿來作自修的入門書；已經忘記了的或對於統計覺得繁瑣討厭的人，可以拿來作復習用。所須的算術，極淺近；不會得開平方的人，也沒有什麼困難。先看了本書，再讀測驗統計術，或者可以容易些。

二十五年八月　　俞子夷在浙江大學

密勒氏統計方法論

上册

F. C. Mills 著
徐 堅 譯

商務印書館發行

譯者序

一切精密之科學究研無不須以數量之材料爲根據,而處理社會現象者尤然蓋在社會現象之中,因子之錯綜複雜,淆人耳目,實較自然現象爲甚.且研究者又不能置身於利害關係之外,個人之情感與成見,每每蒙蔽其智慧,致摭拾片段事實,得似是而非之結論.苟不從事於客觀之數量材料之研究,而欲免除主觀判斷之錯誤,抑亦難矣.

然客觀之數字,充其極不過一現象之眞實紀載,並非卽爲眞理.未經適宜之選擇與組織,未經審愼之分析與解釋,吾人若以初次之表面之印象爲定論,或憑之以作臆測,則其中所包含之規律與法則非但不能表現,反將因此片面之印象而更爲模糊.故對於數量之材料,不能不有一種科學方法以嚴密處理之也.

密勒氏此書,詳明切實,固爲其特色;然最足稱道者猶不在此.任何科學方法,莫不有其條件與限制,其應用越出一定範圍,必生錯誤.故對於一種方法之研究,必須深切考慮其本質,不宜僅以熟習機械之運用爲滿足也.然脫離方法之實際運用,而談論原理原則,必使學者難於了解.卽使爲理至顯而

易明,而學者所得之印象,必不能深刻.密勒氏以說明方法爲主幹,復將每種方法在應用時須要注意之點,反復再四,時時提出,不厭其詳,使讀者得一愼密應用之科學精神,而使方法不致流爲若干死板之公式,其用意蓋至深也.

國內研究統計方法者,大抵以培養研究經濟現象之工具爲目的.密勒氏此書,在在引用經濟及實業方面之材料爲說明,讀者因此可得許多榜樣,作處理實際材料之參考.日來注意於經濟材料之蒐集與編製者日多,而關於此種材料之嚴密的科學分析,則尙爲罕見,此書之迻譯或非多事.譯者才學粗淺,掛一漏萬之處,在所不免.幸海內先達起而正之.

著者原序

近十年來,實業與社會科學方面對於數量方法之興趣,異常濃厚.在以前,實業上之判斷則以直觀,爲其主要之基礎,社會現象之研究則以無根據之假設爲其立論之方式;此種時代,似已過去.自然科學之歷史較久,其研究者所用方法亦素來較爲精確.社會科學家以及實業問題之嚴正的研究者,追隨自然科學,家之後,以事實之觀察及分析作研究之根據,現今此種研究方法之應用,其範圍之廣大實爲往昔所未見.當此等觀察爲數量的性質時.欲與以組織及解釋,必須有適當之方法.本書所述,卽綜合及分析此等觀察之法也.其中對於材料之屬於經濟及實業方面者,最爲着重.

所以限於此專門範圍之理由有二:其一,一般之統計方法實際上雖可以普遍應用,然每一研究部門莫不有其專門之問題,經濟部門中尤其如是,因其中特異之困難與問題爲數甚多也.現今已發明若干方法,在某程度上專爲應付此等特殊需要而設者,故須有當門範圍之書以論述之.其二,說明方法時,能取一特殊題材,則效力最大.蓋抽象之方法論對於平常讀者不免枯燥無味也.因此本書之作,對於從事經濟實

業問題之數量的研究者之特殊須要,曾特加注意焉.

嚴格的數學著作或須說理簡括,惟本書說明方法時未有此種企圖;蓋此書為學者而作,非為已有高深造詣之人;故作說明時,亦以前者為念.此書本屬導言性質,非欲將問題盡述無遺,故將若干定理之詳細證明省去.

現今普遍應用之數量分析方法,為長期發展之產物,積累許多部門之研究家之供獻而得者.欲列舉有功於統計學之發展之人名,實不可能之事.正文內間有個別指示;但現代統計家所受惠於其先輩者實多,決無一表可以為其詳盡之紀錄也.

本書材料之準備,有賴於友人之助者甚多. H. E. Anderson 氏及 H. B. Killough 氏為蒐集第十一、十三、十四各章中若干材料;並蒙哈佛經濟研究委員會 Warren M. Persons 教授允許應用其商品價格指數研究工作之若干結果;第七章之建築費用指數及第九章之商情指數乃美國電話電報公司統計部所編製,蒙該部統計主任 Seymour L. Andrew 氏慨允應用,一併誌謝.加利福尼亞大學教授 A. H. mowbray 氏之建議,曾使余改正最初複寫版中若干模糊不清之點.哥倫比亞大學教授 Henry L. Moore 氏及教授 Theodore H. Brown 氏及經濟研究院 Henry Schultz 先生助余仔細審讀原稿之若干部分, Herbert F. Tutt 氏助余製圖,至可寶貴.而 Donald H. Davenport 氏在本書著作之每一階段皆有幫助,自收集材料,

編製圖表,至獨力照料此書之印刷完成,有賴其力實甚鉅也.吾妻熱忱襄助,始終不懈,尤為余所感激不盡者也.

一九二四年十一月

Frederick Cecil Mills

目　次

統計圖索引

密勒氏統計方法論

第一章

統計方法與經濟實業問題

實業一語,(註)吾人今日常用以包括人類之多種活動,玆既以論述處理經濟材料以及解決實業問題之方法爲事,對於各種實業活動之適用統計方法者,自宜略加敍述也.

〔實業活動之種類〕

實業家所負擔之工作,若不過事苛細,可大致別爲三類.實業家中固甚多專門從事於一種活動者,但整個實業界則不得不對於下述三種場合中所發生之種種問題,盡予處理.按論理之程序列之,第一爲生產過程中之技術工作,包括化學,物理學,工程學,畜牧學及航海上諸問題.解決此等問題時所需之基本技術知識,爲吾人經濟生活之基礎.積苦心孤詣而求得之技術,所以處理原料與支配自然力者,卽屬於此領域者也.

(註) 原文以 Business 統括實際經濟活動之各方面,與日語"企業"二字之意義相似,故譯作"實業".

第二類之工作,為與各個實業單位之內部組織及管理上有關之一切活動.經管理機關,農場,礦山,工廠,鐵路,百貨商店等等,改裂有機無機物質以滿足人類慾望之技術作用得以遂行.當實業家組織此等單位,配合各部之工作,監督每一組織中各人之日常活動時,其所接觸之問題,自成一新部門.此等問題雖其重要性或次於生產技術問題,然於一般實業家則更為迫切而困難.蓋於此等問題之解決上,科學方法之發展較為落後.既不若前一場合之有系統之知識存在,亦且無同樣熟練之專家可委之以此等責任也.

上述之兩種經濟活動,在某一意義上為以自身為中心而可受支配者.鍊鋼家有其鎔化與精鍊之技術問題,並有其特殊之管理責任.農夫與礦主亦有此同類之問題,僅由於各人之情形不同而其形式有異耳.當進行此等場合之工作時每人所處理之種種問題,其一切因素多少可予以完全之支配.固亦有困難發生,然其困難常為此等工作所固有者,非由於問題中之構成因素實起變化,或由於有新因子突然加入也.在此點上,實業家所須完成之第三類工作,與前二者根本不同.蓋第三類工作所包含之問題,其最顯著之特徵,即其因素為不受直接有關之個人所支配者也.

此第三類之工作,包括購買,銷售,與一切附屬之活動其進行時以價格表示者.就現今經濟生活之組織情形言,此等工作對於實業家為其所作之一切工作中最重要者.生產上

以及內部組織及管理上之各種技術工作,僅爲達到目的之一種手段.由實業家觀之,經濟活動之目標爲以善價出售其產物.在此最後銷售以前之一切工作,必須居於其次,並以如何達此最後目標爲進行之準則.此地之要點爲實業家在買賣過程中所遇諸問題,含有爲彼所不能支配之許多因素.當收集原料,配合其他生產要件,以及最後之出售生產物時,彼皆周旋於市場之中——商品市場、勞動市場及金融市場——應付一主要變動全不受其支配之價格系統.實業家於其他較不重要之活動場合,固能發揮其高度支配力量,惟一至最後而最主要之一步,卽將其產物善價出售時,其支配力量卽形消滅.實業活動之動機在於牟利,欲牟利須買賣之成功,欲買賣成功須在不受支配之價格世界中遇一良好時機——實業之主要問題卽在於此.此亦價格系統所以在現代經濟生活中成爲左右一切之最重要因子者也.

現代企業家生存於價格環境之中.環境二字非不適當之比喻;此價格世界乃一互相倚賴諸部分相聯相應巧妙結合之系統,將企業家之一切實業活動全部包圍其中.此系統既超出個人支配能力之外,個人唯有使自己與之適應,力求對此系統之充分了解以爲活動之指針.苟無此了解,實業上之主要問題卽無從解決.

[經濟學及實業方面諸問題之數量的性質]

上述第一類諸問題,久已被認爲根本屬於數量的性質.

欲解決此等問題,須用自然科學所發展之精確方法.然屬於第二第三類之問題,成爲嚴格意義上之經濟實業問題者,其解決時須運用數量方法,實與前者相同.屬於質而不可以量計之考慮,固亦可參加此等問題之解決,惟此種考慮平常皆以數量之材料爲根據.事實經計量、權衡並與其他事實比較之後,即成爲實際之基礎,可根據之以判斷實業問題、與作經濟學上之推理.統計方法者,僅計量、權衡以及比較事實之方法而已.

前節所分別之三類問題,有兩類入本書討論範圍以內.對於生產技術問題之解決上,統計方法固亦部分適用,然本書之目的並不在討論此問題.統計分析方法.對於其他兩場合中諸問題之解決,(即問題之與實業單位內部組織及管理上有關者,以及問題之與使實業家接觸價格制度之買賣過程有關者)乃特別適用者也.

[統計方法與內部組織諸問題]

模範實業家於管理其組織時,須處理大量之數量材料,材料之所以稱爲"數量的"材料者,謂其爲用專門之單位計量表示者也.如煤若干噸,瓦斯若干立方呎,電力之消耗每小時若干瓩,生產銑鐵若干噸或鞋若干雙,機器與人工工作若干時,以及工資、成本、或售價若干圓角分,此即彼所處理者也.管理企業者所須處理之材料,隨實業單位之日益增大,亦愈趨繁多複雜,其眞正之意義愈難確定.單憑直覺或實際經驗

上得來之管理方法,而欲對於大量材料作一有效之分析,以及管理一實業單位之大於平均規模以上者,實爲不可能之事.實業中報酬遞減律 (law of decreasing returns) 之發生作用大半由於管理上之困難,乃過去之經驗所充分證明者也.

當吾人處理大量之材料時,莫不遇綜合與分析之問題——綜合材料而使之變爲簡單,使人類有限之才能足以予以處理;分析材料 (並比較之), 以辨別問題中所含之要素而估計其重要性.所以發展統計方法者,卽以便利大批數量材料之綜合與分析爲目的者也.

此種問題可舉成本費之分派 (allocation of cost) 爲一模範之例.此種工作亦稱爲成本會計.此問題中所包含之一切因子欲盡加分析,非用統計方法不可.因會計方法僅限於處理金錢單位,不適於作一各項費用之完全的分析也.除由成本費之分派所求得之知識以外,各項費用亦須應用統計方法加以分析,示其變動,予以比較,察浪費之所在.再如銷售記錄之分析,亦須綜合大量材料,以簡單易解之形式表示之,並決定此材料之意義.市場之分析,購買記錄與商品之研究,均須用一種數量方法,其應用不限於任何一種計量者.於內部管理之任何方面,統計方法皆可以補會計方法之不足,增拓執行業務者之知識,而對於業務之進行作更有效之統制.

[統計方法與對外諸問題]

當實業家進至市場而行買賣時,則遇許多新問題.價格、

系統,其變動爲實業家所如此深切關懷者,須要運用數量方法而作之分析.此地所處理之材料,其繁多複雜,使簡單化之事成爲一迫切之工作.且也,此地有商情循環之現象臨於實業家之面前,彼若欲使其營業方針能與市況之盛衰相適應,必須運用適當之工具以從事於分析此等現象.至於與經濟上之分配過程有關之一切問題,卽關於收入與財富在參加生產之各種人間之分配情形者,與實業家之直接利害固稍疏遠,然對於經濟學者則具有極大興味.此種問題之爲數量的性質.其解決之必須用數量的研究方法,正與其他重要經濟問題之與價值或價格之決定有關者相同者也.

此種數量之方法究爲何物?而應用此種方法之研究所以異於其他形式之研究者果何在乎?一切之科學研究,不論其特殊之方法爲何,皆經歷細密之觀察,合理之推論與確切之實證三步驟.數量方法與其他方法所不同之點,僅在其觀察,推論與實證皆以"計量"(measurement)爲根據;故較之非數量的分析方法(non-quantitative methods of analysis)更易精密確切.一種科學在未能應用計量以前,不論其研究者之查覺如何敏銳,工作如何辛勤,其觀察與發見之不能精確乃爲不可避免之事.應用計量方法以後,在分析問題中所包含之諸因子時,卽可用精確之單位表示;其對於一種科學所貢獻之種種便利,正如用一銳利之工具以代替一鈍缺不可靠之工具.數學與統計學及會計學,爲現代經濟學者手中之有力工

具,在實業方面,亦由於研究機關與方法之發達,其應用範圍亦日趨廣大.

統計家所用之工具,僅係爲特種研究而發展之若干數學方法.考統計方法之最初發展,並非用於經濟研究,而爲用於研究社會現象、政治現象以及人體測定 (anthropometric) 者;其一枝之發展(關於機率理論 theory of probabilities 者)可由論理之域追溯至賭徒之場.此種工具最初應用之範圍原極狹窄,但以後發見其大可應用於較廣之領域;經濟學即其新領域之一,將此等方法加以適當之改變與補充,而已獲得豐富之成果者也.因獲得此新工具之故,經濟學者之能力增強,其工作之精確性亦大爲增加.而在實業方面,亦與較深奧之經濟科學同受其惠.

關於統計方法所受之限制,此地或不必特別說明.統計方法固爲有限制者,且其限制並不能常爲應用此法者所注意.關於此事當於後文討論之.惟有一點在此可請特別注意者,統計方法有如一工具,須用之得法;由統計分析得出之結果,須予以審慎之解釋.人類猶未能創一完善之自動機,能一端接受雜亂之事實,另一端傾吐種種問題之答案也.審愼推理與精確判斷之才能,並不因統計研究方法之發達,而變爲陳舊之附屬品也.

第二章

圖示法

統計方法須對於實業與經濟學方面之種種事實加以綜合、分析與解釋.此等方法之說明可由若干基本問題之討論出發,此等問題之性質與其謂屬於統計學之範圍不如謂爲屬於數學之範圍.所以由此出發者,因吾人以爲對於後文各章中常常提及之數學觀念應首先加以說明;此等觀念即使已爲學者所熟知,此舉亦不能謂爲多事也.

統計分析原以處理根據計量而得之材料爲事.此種材料或以貨幣單位 (pecuniary unit) 表示,或以物理單位 (physical unit) 表示.哲學家笛卡兒 (Descartes) 氏所首創之解析幾何,使吾人整理與解釋此等材料時大爲便利.於此將解析幾何之基本原理摘要論述,諒非不適當者.

[垂直坐標] (rectangular coördinates).

設於平面上作垂直相交之兩直線,則該平面上之任何一點,其位置可以其與兩線交點之關係說明之.吾人稱垂直之一線(橫線)爲 $Y'Y$, 稱水平之一線(橫線)爲 $X'X$, 稱交點爲

O (原點 origin). (觀圖 1). 若 P 為該平面內一任意點, 吾人可引一線 PM 與 $Y'Y$ 相平行而與 $X'X$ 相交於 M, 更可引一線 PN 與 $X'X$ 平行而與 $Y'Y$ 相交於 N, 設以一單位量之, OM 之長等於 a, ON 之長等於 b, a 與 b 即成為 P 點之坐標 (coördinates). 此坐標指出該點對於原點 O 之關係以說明其位置. 如圖中 a 等於 6 而 b 等於 5. a 在 x 軸 (x-axis) 上之距離稱為 P 點之橫坐標 (abscissa), b 在 y 軸 (y-axis) 上之距離稱為 P 點之縱坐標 (ordi-

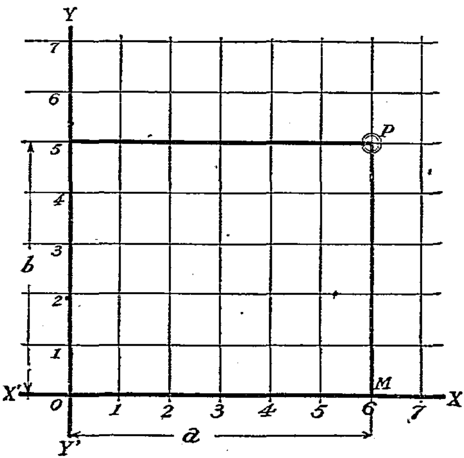


圖 1.—以垂直坐標定一點之位置

nate).（記數之規則,經常將橫坐標記於縱坐標之前).在同一平面上其他任何點之位置皆可用此法定之.反言之,若有任何兩實數 (real number), 取其一爲橫坐標,其另一個爲縱坐標,即可於平面上決定一點.

一點可在原點之左或在其右,可在原點之上或在其下.慣例以在原點右者爲正橫坐標,在原點左者爲負橫坐標;在原點以上者爲正縱坐標,在原點以下者爲負縱坐標.經濟統計所處理之數值,一般皆在右上方之一象限中.其中縱坐標與橫坐標皆爲正.

坐標爲數學上之基本觀念,對於統計工作有根本之重要性.用此法表現實業方面之材料,其效用可取一簡單之例以表明之.玆可以用下表之數字爲材料.

表　一

1923年各月棉花價格表

美國植棉者所得平均價格

月　份	每磅價格
正　月	24.5 ¢
二　月	25.9
三　月	27.7
四　月	28.4
五　月	26.9
六　月	25.6
七　月	26.2
八　月	23.5
九　月	24.1
十　月	27.2
十一月	28.8
十二月	31.0

(錄自 Weather, Crops and markets, Dec. 29,1923)

表中材料可用坐標系圖示之,將月份排列在x軸,價格則依y軸描之,如圖2之情形.給橫坐標時,以原點代表1923年正月,故1923年正月一數,在x軸上之值爲0,二月則爲1,其餘各月依次類推.代表1923年正月棉價之一點,其坐標爲0,24.5;代表二月者爲1,25.9;代表十二月者爲11,31.0.若各點之間用許多直線依次聯結之,(如圖),則全年棉花價格之趨勢更易明瞭.

[自變數與應變數] (independent and dependent variables)

用坐標定任何一點之位置時,須有兩個數值,既如前述,每點結合兩個因子,並說明其彼此間之關係.如前舉之一例,其中兩個因子,卽月份與棉花價格是也.棉花價格與時俱變,圖中之折線卽表示變動之方向與大小者也.時間與價格二

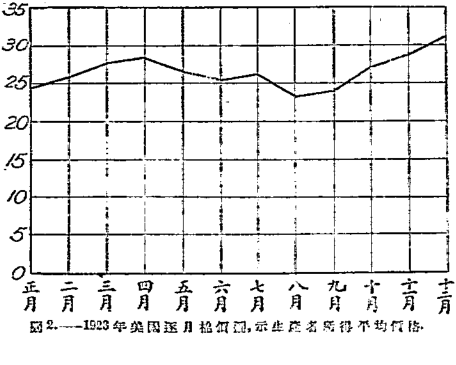


圖2.——1923年美國逐月棉價圖,示生產者所得平均價格.

者俱係變數 (variables),卽兩者在所討論之問題中,其數值不固定而以變動爲其特性.如在圖1中,橫坐標有固定之數值6,縱坐標有固定之數值5,惟圖2中之橫坐標與縱坐標兩者,其數值不一,其一小至0大至11,其一小至23.5大至31.0.習慣上以x與y爲記號以代表此種變數,前者常指變數之依橫軸描記者,後者常指變數之依縱軸描記者.(註)

圖2示棉花價格之隨時漲落.注意此地代表時間因子之變數,每次變化皆有一定單位,卽一月.吾人旣以時間因子作獨立之變化,於是可決定在如此任意劃分的一時期中價格因子所起之變動.變數之變動依吾人任意決定之一定數量者謂之自變數,常繪於x軸;其他之另一變數謂之應變數,常繪於y軸.應變數對於自變數之依從 (dependence) 或爲實在者,卽前者絕對爲後者之數值所決定;或僅係慣例,有如前舉之例.當時間因子成爲變數之一時,經常作爲自變數描繪.

[函數關係] (functional relationship)

若兩變數間有一完全依從之關係存在,卽y之值絕對爲x之值所決定時,吾人稱y爲x之函數.此種關係平常用$y=f(x)$之式表示之.如一落體在某一刹那間之速度,爲開始降落以後所經過之時間之函數;一定量氣體之壓力爲其溫度之函數;一定量之本錢按照固定利率之增益爲時間之函

(註) 最後之幾個字母常用以代表變數,最前之幾個常用以代表常數 (Constants)——卽在所討論之問題中數值不變之量.

數.若將自變數依x軸描繪,將與之相當之函數數值(卽應變數)依y軸描繪,則可得一函數圖示,其形式為一曲線(curve).(註)此函數關係之一觀念,在統計工作上異常重要.玆就若干簡單之函數略加討論如下.

[直線關係](linear relationship)

若兩變數之數值永遠兩兩相等,則其關係顯然可以$y=x$一式代表之.舉一極簡之例,如樹木之年齡與其年輪之關係,二十年之樹木有年輪二十個,……年齡與年輪之數永

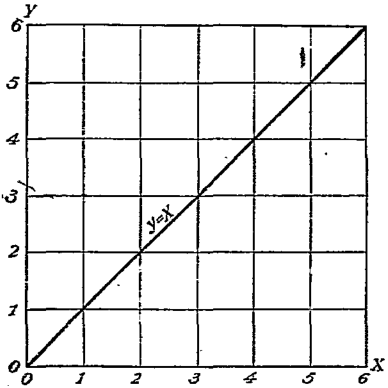


圖3.——方程式$y=x$之格爾圖線(graph).

(註) 坐標系中所繪之線,不問曲直,皆以"曲線"稱之.

適相等,吾人可抽取若干對之x與y之數值作樣,將此種關係繪於坐標圖上.如作一線聯結代表各對數值之各點,可得一直線.此直線通過原點.假定縱軸橫軸之尺度(scale)相同(卽劃分距離之單位相同)此直線平分XOY之一直角.(參考圖3)

同樣,任何之一次方程式(卽不包含xy或x與y之一次以上之方者),皆可以一直線代表之.其普遍式爲$y=a+bx$, a爲一常數,代表由原點至該線與y軸相交之交點間之距離, b爲一常數,代表該線之斜度(slope,卽該線與橫線所成之角之正切).常數a稱爲y截數(y-intercept).由直線方程式之普遍式,知x之數值爲0時,y應與此常數a相等.前舉之一例(圖3)中,a等於0,b等於1.一線之位置,視a與b之大小與其符號(正或負)而定.決定任何之直線時,所須解決之實際問題,爲從已有之材料中求出a與b之數值.在討論統計方法時,此問題將以種種形式出現.

茲可繪一簡單之一次方程式,以說明上述諸點.今有一函數$y=2+3x$.欲作一條摺幅線以表示之.先假定x爲某某若干數值.並決定與之相當之y數值.結果列如下表:

x	y $(2+3x)$
-4	-10
-2	-4
0	2
2	8
4	14

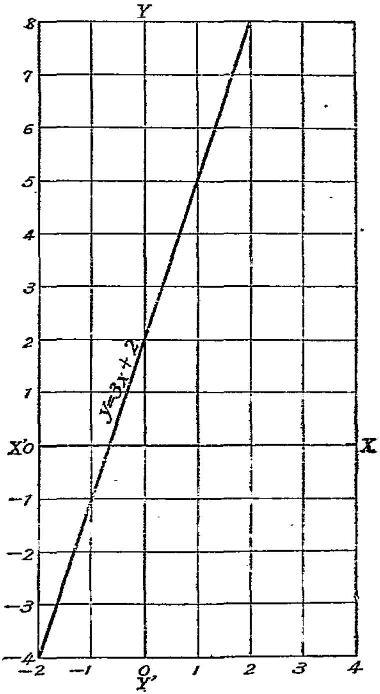


圖4.—方程式$y=3x+2$之格網曲綫

將表內各數值繪爲各點,以一線聯結之,得圖4中之格㭫幅線.因其爲直線式之函數,(卽其格㭫幅線之形式爲一直線),故只須有任何兩點,卽足以定全線之位置.y截數等於常數項2,該線與橫線所作之角之正切(卽該線之斜度)等於3(x之係數).因曲線上每一點之坐標皆合於此方程式,而每對數值之合於此方程式者皆爲曲線上之一點所代表,故知該線實足以代表此方程式.直線關係有一特點,卽當其中一個變數之增加按一固定之增量(constant increment)時,其他一個變數之增加亦按一固定增量.如上例中x之固定增量爲2,y之固定增量爲6.數列之數值按固定增量增大者,稱爲算術級數.(arithmetic series)

在自然科學中,變數間直線關係之例甚多.經濟現象方面可舉資金按單利率增殖爲例.吾人若以r代表單利率,x代表年數,y代表一圓之資金於x年底時所積之本利和.此關係可以下列方程式表示之:

$$y=1+rx$$

因此中r爲常數,故此關係爲一簡單之直線形式.在統計工作中,與此絕對相同之關係雖非絕無,然極少見;然近似於直線關係者,則常可遇見者也.

〔曲線關係〕(non-linear relationship)

曲線函數種類甚多,此地僅能就若干較爲普通者討論之.主要之非週期性曲線(non-periodic curves)中,一方以拋物

線(parabola)與雙曲線(hyperbola),另一方以指數曲線(exponential curve),爲最重要;學者須熟知其一般之特點.定羃級數(potential series)爲比較普遍之形式,應用甚廣,故於此提及.週期性函數(periodic functions)中,取正弦曲線(sine curve)作簡略之敍述,以其爲一基本形式也.

拋物線式及雙曲線式之函數關係,在自然科學中甚爲普通.此種曲線亦適合於某種經濟材料.其普遍式無常數項時爲$y=ax^b$.指數b爲正數時,此曲線爲拋物線,爲負數時,則曲線爲雙曲線.茲舉二例以說明此兩種形式.

題:作函數$y=x^2$之格欄幅線.

x	y (x^2)
-5	25
-4	16
-3	9
-2	4
-1	1
0	0
1	1
2	4
3	9
4	16
5	25

格欄幅線見圖5.

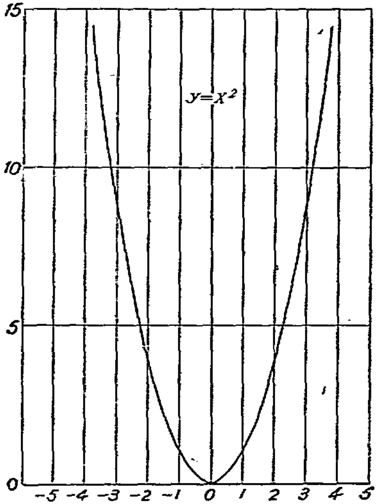


圖5. 拋物線：方程式 $y=x^2$ 之格柵幅線

題：作函數 $y=x^{-1}$ (x 為正數時)之格柵幅線.

x	y (x^{-1})
$\frac{1}{3}$	3
$\frac{1}{2}$	2
1	1
2	$\frac{1}{2}$
3	$\frac{1}{3}$
4	$\frac{1}{4}$
5	$\frac{1}{5}$

此函數之格柵幅線,為一等邊雙曲線,如圖6所示.此方程式亦可寫作$y=\frac{1}{x}$或$xy=1$.

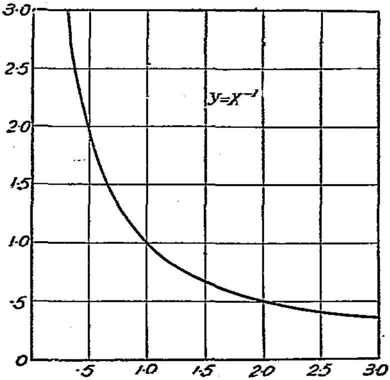


圖6.—等邊雙曲線:方程式$y=x^{-1}$(x為正數時)之曲線

此種形式之關係,其特點為x依照幾何級數(geometric progression)(註)增加時,y亦依照幾何級數增加.如前舉之拋物線之一例($y=x^2$),若以形成一幾何級數之數列為x之各值,則與之相當之y各值亦形成一幾何級數:

x	1	2	4	8	16	32
y	1	4	16	64	256	1024

(註)幾何級數者,為如此之一數列,其中每一項之值,皆為用一固定之乘數乘其前一項之值所求得者.

另一種函數其方程式爲$y=ab^x$.此種函數之方程式中,變量之一成爲指數(exponent),代表此種方程式之格槲幅線稱爲指數曲線.可舉下例爲說明:

題:作函數$y=2^x$(x爲正數時)之格槲幅線.

x	y (2^x)
0	1
1	2
2	4
3	8
4	16
5	32
6	64

格槲幅線見圖7.

兩變數增大時各按固定增量者,形成算術級數,其關係可以直線表示之;兩變數增大時各成幾何級數者,其關係或以拋物線或以雙曲線表示之;此前所已述者也.至於指數曲線則成一雜合之形式,其所表示之關係,一端爲按算術級數增大之變數,一端爲按幾何級數增大之變數.此種關係,觀上表之數字卽可明瞭.

根據下式而作之曲線,在統計研究中應用甚多;式爲

$$y=a+bx+cx^2+dx^3+\cdots\cdots$$

此種形式之方程式,稱爲定冪級數.此種曲線雖嚴格言之,非爲圓錐剖面式之拋物線,然根據此種方程式而作之曲線,常

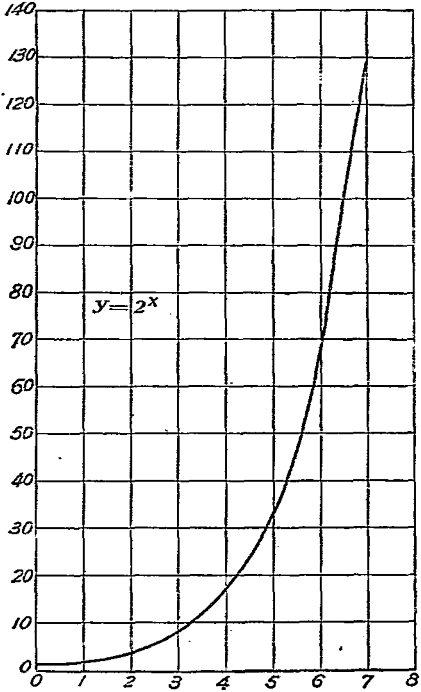


圖7——指數曲線：方程式$y=2^x$之曲線(x爲正數時之情形)

稱爲二次拋物線,三次拋物線等等,名稱各依方程式中x之最高次數而定.此種級數無一致而簡單之形式.後文將詳細討論之.

周期函數另成一類,電學與氣象學中之關係爲其顯著之例,但非僅僅限於此等領域.此種關係之特點,即應變數之各值,依自變數值中一固定之間距(Inte-val)而循環反復也.正弦曲線爲此種形式之曲線中之基本形式,茲舉下例說明之.

題:作函數$y=\sin x$之格欄幅線.

x (角之度數)	y ($\sin x$)
0°	.000
30°	.500
60°	.866
90°	1.000
120°	.866
150°	.500
180°	.000
210°	−.500
240°	−.866
270°	−1.000
300°	−.866
330°	−.500
360°	.000
390°	.500
......	......

格欄幅線見圖8.

求一代表兩變數間關係之數學式之事,其在統計工作

中之重要性,在未更進一步研究以前,難於完全說明.其基本

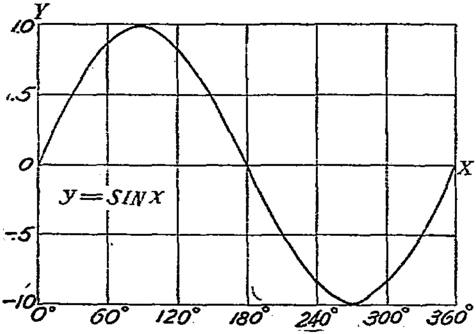


圖 8.——正弦曲線:方程式 $y=\sin x$ 之曲線

之目的,在決定潛伏於所觀察的現象之後之自然法則或經濟法則 (physical laws or economic laws); 另一更實際之目的,爲求一從一個變數之已知值可推算另一變數之近似值之公式.全書各例,皆用以說明如何能達到此種目的者也.(註)

〔對數〕

對數 (logarithms) 對於一般計算上所盡之作用極爲重要,而其對於整理統計原料上亦同樣重要.此地可略述對數之性質,與用之以便利計算之進行之方法,作爲溫習.討論之範圍限於以10爲底 (base) 之常用對數.

(註) 關於各種曲線形式之詳細討論見後文處理時間數列之分析處.

任何正數皆可以10之乘方(power)表示之.如

$$1{,}000=10\times10\times10=10^3,$$

$$10{,}000=10\times10\times10\times10=10^4.$$

兩式中10之指數(10之右上方所記之小數字),表示在連乘中10作爲因子出現之次數.代表10之整數乘方之指數爲一整數,惟代表其他數值之指數卽帶有分數.如100爲10之二次方或10^2;而110則爲10之2.04139次方或$10^{2.04139}$.

10之指數,或10求與某數相等時必須參加自乘之次數,稱爲該數之對數.100之對數爲2,110之對數爲2.04139.998之對數爲2.99913.此三對數皆以10爲底.然以任何一數爲底,皆可作成一對數系統;故普遍式爲

若 $$a=b^c$$

$$\log_b a=c.$$

卽"c爲以b爲底時之a之對數".應用之對數爲常用對數時,如記住下列之關係,則在一自然數,與底與對數間之關係自易回憶.

$$100=10^2$$

$$\log_{10}100=2.$$

任何數值之對數,皆有兩個部分,卽整數部分與小數部分是也.整數部分稱爲定位數(characteristic),小數部分稱爲定値數(mantissa).前者由觀察而得,後者可由對數表中檢出.定位數隨自然數中小數點之位置而變,定値數對於任何一

數字之結合體 (Combination of numbers), 不問其小數位置之前後,皆一成不變.此點可以下例說明之:

$$\log 8450 = 3.92686$$
$$\log 845 = 2.92686$$
$$\log 84.5 = 1.92686$$
$$\log 8.45 = .92686$$
$$\log .845 = 9.92686-10$$
$$\log .0845 = 8.92686-10$$

求一與某一對數相當之自然數時(此自然數名為逆對數 Antilogarithms),定值數定數字之順序,而整數或定位數決定小數點所在之位置.例如求2.17609之逆對數,檢對數表知與.17609相當之自然數為1500.定位數既為2,則所求之自然數必在100與1000之間,故必為150.

下列數字為與10之各乘方相當之數值,對之少加研究,可助吾人記憶10之倍數與其對數間之關係,使吾人易於決定所求之對數之定位數.

.0001	.001	.01	.1	1	10	100	1,000	10,000
10^{-4}	10^{-3}	10^{-2}	10^{-1}	10^{0}	10^{1}	10^{2}	10^{3}	10^{4}

下面一列之10之各指數,為上面一列之各數之對數.

注意由0至1間各數,其對數皆為負.如.845之對數為$-1+.92686$;書作$9.92686-10$.故以對數表示由0至無限大之間全部之正自然數時,對數已用盡所有一切之正負數值.故

一負之自然數既不能有正對數亦不能有負對數.

將各數如是表示爲10之乘方時,其便利之所在,即因此可省去平常乘法、除法、乘方、開方中許多麻煩手續也.

兩數之積的對數,等於兩數的對數之和,故欲使兩數相乘,可以兩數之對數相加.普遍式爲

$$a^b \times a^c = a^{(b+c)}$$

常用對數之特殊式爲

$$10^2 \times 1000^3 = (10\times10)\times(10\times10\times10) = 10^5 = 100{,}000$$

$$100\times1000 = 100{,}000$$

商的對數,等於被除數的對數減去除數的對數.故欲以一數除另一數,可由後者的對數減去前者的對數.普遍式爲

$$a^b \div a^c = a^{(b-c)}$$

特殊式爲

$$10^5 \div 10^2 = \frac{10\times10\times10\times10\times10}{10\times10} = 10^3 = 1000$$

$$100{,}000 \div 100 = 1000$$

某數幾次方的對數等於某數的對數乘其冪指數.普遍式爲

$$(a^b)^c = a^{bc}$$

特殊式爲

$$(10^3)^2 = (10\times10\times10)\times(10\times10\times10) = 10^6 = 1{,}000{,}000$$

$$1{,}000^2 = 1{,}000{,}000$$

某數之幾次方根的對數,等於方根指數除該數的對數所得之商.普遍式爲:

$$\sqrt[b]{a^c}=a^{\left(\frac{c}{b}\right)}$$

特殊式爲

$$\sqrt[3]{10^6}=10^{\frac{6}{3}}=10^2=100$$

$$\sqrt[3]{1{,}000{,}000}\qquad=100$$

總結:

$$\log(a\times b)=\log a+\log b$$

$$\log(a\div b)=\log a-\log b$$

$$\log a^b=b\times\log a$$

$$\log\sqrt[b]{a}=\log a\div b$$

對數之此等特殊便利之點,製造滑尺(sliding rule)時曾利用之.此減省日常計算之勞之工具,凡研究統計者皆應熟識之也.

[對數方程式] (logarithmic equations)

前面曾敍述應用垂直坐標系作材料之圖示之方法,並述此法之若干利益,有時與其給自然數不如給其對數;其原因,或爲使某種重要之關係更得明顯之表示,或爲使整理材料之手續大形簡便.特別當用對數後能將一複雜之曲線化爲一直線時,在處理材料與解釋結果上顯然可以簡便不少也.

前述直線方程式之普遍式爲 $y=a+bx$, a 與 b 爲常數, a

示該線之y截數,b示其斜度.用對數以使一方程式簡單化時,不外以$\log x$代x變數,或$\log y$代y變數,或兩變數俱以對數代替,因此使高次之方程式化爲較簡單之形式也.

此法可舉$y=x^2$一方程式說明之.此式繪於垂直坐標系時,成一如圖5之抛物線式曲線.改爲對數式時,成$\log y=2\log x$.

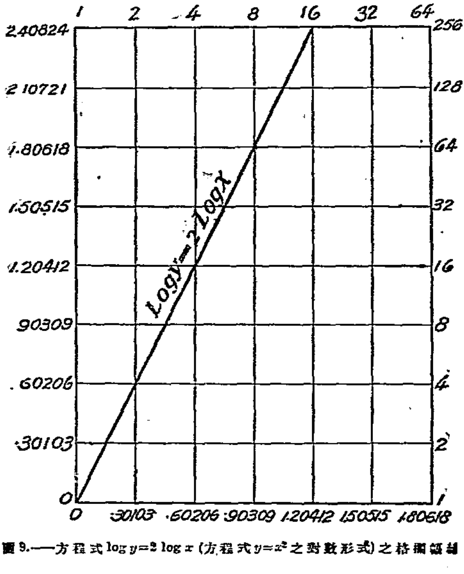


圖9.—方程式$\log y=2\log x$(方程式$y=x^2$之對數形式)之格欄繪製

此式爲一直線方程式,變數爲 $\log x$ 與 $\log y$. $\log x$ 爲正數時,此式之格欄幅線有如圖 9. 圖之上邊與右邊所註之數字,爲與下邊及左邊所列各對數相當之自然數,所以表示其中之關係也. 自然數成幾何級數,對數則成算術級數,圖上相等之一垂直距離,在對數尺度上爲相等之絕對增量 (absolute increments) 在自然數尺度上爲相等之百分比增量 (percentage increments).

同樣,方程式 $y=5x^3$ 可化爲 $\log y=\log 5+3\log x$, 而成一直線方程式; 一切方程式之如 $y=ax^b$ 式者,即繪圖時成一簡單之拋物線與雙曲線者,皆可化爲直線方程式 $\log y=\log a+b\log x$. 在圖上,即繪 y 諸值之對數與 x 諸值之對數也.

然指數曲線之方程式 $y=ab^x$, 其問題即不同. 化成對數式時,成 $\log y=\log a+x\log b$, 亦爲一直線方程式,式中兩常數爲 $\log a$ 與 $\log b$, 兩變數則爲 x 與 $\log y$. 此種方程式若繪圖時用 x 之自然數與 y 之對數,可得一直線. 此種形式之曲線將於後文討論並舉例說明之.

[雙對數圖與單對數圖]

繪對數亦有若干不利之處,若須繪之點數甚多,則檢表之工作或令人生厭. 且圖上不見爲主要興趣所在之原來之數值. 此種困難可用下法解決之:其法即依照對數尺度劃格,而記自然數於圖上. 如是格線上所見者雖爲自然數,然其距離則按各自然數之對數之大小爲比例者也. 圖10 即爲一此

種之圖.圖中之格欄幅線代表方程式$y=x^2$.

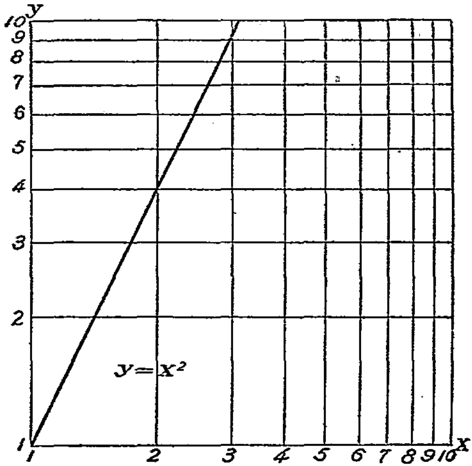


圖 10.——方程式$y=x^2$之格欄幅線(橫於對數格紙上)

(兩軸皆用對數尺度分格之圖謂之雙對數圖 double logarithmic chart).

尚有一種變相之圖,橫軸用算術尺度分格,縱軸用對數尺度分格,稱爲單對數圖 (semi-logarithmic chart),在統計工作上甚爲重要.在此種圖紙上作圖,卽等於在完全用算術尺度分格之紙上繪x之自然數與y之對數也.兼用此兩種尺度而

作圖時,可使指數式曲線變成一直線,此前所已述者也.此種單對數式或比例式格紙可用滑尺或對數自作之,或購買現成者.此種格紙於作經濟統計圖時有特殊之價值,因經濟方面之材料,時間因子常為其變數之一,而此變數最好用自然數尺度給之也.

此種形式之曲線可以複利公式為例.若以 r 代表利率,

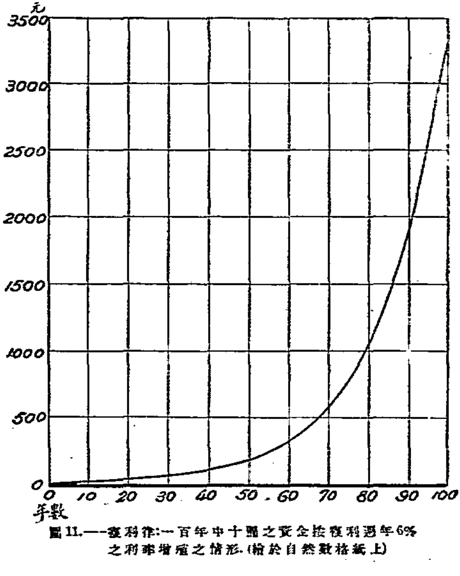


圖11.——複利律:一百年中十圓之資金按複利週年6%之利率增殖之情形.(繪於自然數格紙上)

x代表年數，p代表本金，y代表x年後之本利和，而利息則每年按複利計算(即每年之利息轉入下年之本金一同起息者)則得下列方程式：

$$y=p(1+r)^x.$$

化爲對數式，則成

$$\log y=\log p+x\log(1+r).$$

此爲一直線方程式.

將本金10元依複利六釐增殖之情形，繪於自然數格紙上，成圖11之曲線.此曲線爲指數方程式

$$y=10(1+.06)^x$$

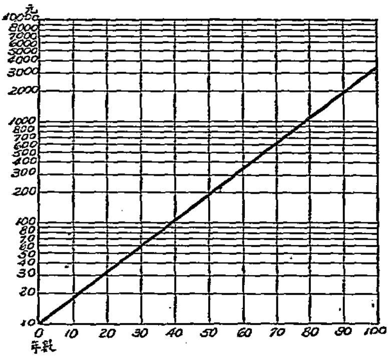


圖 12.——複利律：一百年中十圓之資金按複利週年6%之利率增殖之情形.(繪於對數格紙上)

之格欄輻線．y代表x年後之本利和．圖12所繪之材料與前圖相同，惟係繪於單對數格紙上，故指數曲線化爲一直線．

單對數格紙之應用，不限於化指數曲線爲一直線．有許多材料之意義，必須用此種格紙始能得最明白之表現．此種利益後文尙須詳論．

[作圖之意義]

由觀察或由統計調查而得之結果，若爲數量之形式，當吾人分析與解釋此等材料時，其第一步工作卽爲根據此等結果作圖．作圖不僅因開闢進一步硏究"關係"之途徑，而有其科學之價値，且與吾人一直接之實際目的相合，此目的卽獲得一材料之眞印象是也．蓋視覺刺激對於吾人之理解及想像上，其助力較之最新式之推理方法猶大過之．解釋一行數字之材料，可爲一困難工作，然將此等材料作成一圖時，可至簡而易明，因此在今日實業界之日常活動中，圖示方法所盡之作用，正如試驗室與打樣間然，乃極爲重要者．

本書限於範圍，不能對於今日工程師與統計員所用之種種圖示盡加詳述．僅能對於圖示原理之較重要者及圖之幾種主要形式略加說明耳，其他諸例散見後文各章．

[選擇圖式之標準]

每次作圖時應用何種圖式，其選擇時應考慮之總問題有二：第一爲其材料之性質若何．若某一問題之材料常可用幾種形式圖示者，其中常有一種最能適合於材料．當然亦有

若干圖式全不合於此種材料之表示者,故選定圖式時不可不牢記材料之特質為何也.

作圖之目的或為更重要之問題.常用之圖式中,每種各特別適合於某種目的,或顯示材料之某些特質,或著重其某些關係,然無一圖而可為一切目的之萬應靈丹者也.目的若未能明白確定,即無法選擇一最合宜之圖式.今敍述若干標準圖式以便作此種之選擇.

[適宜於繪時間數列之圖式]

時間數列(Time series)之圖示,其主要目的,在於表示材料中數值隨時間之過程而生之變動、此變動之總趨勢、以及在總趨勢上之波動.若目的在特別表示出絕對之變動,即數列中在不同時間之各值所表示之絕對單位之差異,則如圖13之簡單圖式即可合用.該圖所繪者為1901—1923年間美國

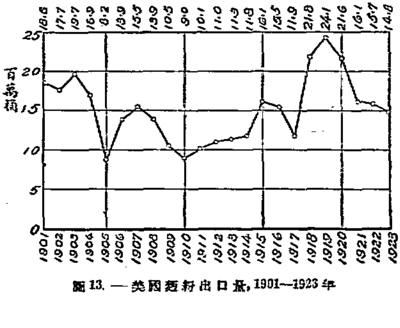


圖13.—美國麵粉出口量,1901—1923年

之每年麪粉出口量.縱軸橫軸皆以算術尺度分格,以各點示各年之數值,並爲便於理解起見,以一列直線連結之.此圖表示逐年之變動.1910年以前爲向下跌落之趨勢,自1911年至1919年向上增漲,1919年後又復下落.關於此圖之大致之構造,須注意下列幾點:

1. 標題明記所給材料之性質,及其所經歷之時期.

2. 縱線上之尺度自0起始,使人對於變動之大小能得一眞確印象.

3. 代表0之線與連結所給各點之線,較坐標圖上格線爲粗.

4. 表示尺度之數字記於圖之左方及下方.縱線上之尺度可重新記於右方以便閱讀.數字排列由小及大,縱線上由底邊起始,橫線上由右邊起始.

5. 所給各點之y值記於圖之上方,此法固佳,然非必要,因各值可另列一表也.

[比例圖之優點]

若吾人所研究之主要問題非絕對變動而爲相對變動,則宜用單對數式格紙:y軸用對數尺度分格,而x軸用算術尺度分格.此種格紙上,縱線上同樣長之距離,代表同樣大之百分比變動;而普通算術式格紙(即以自然數尺度分格者)則以同樣長之距離代表同樣大之絕對變動,此其相異之點也.應用單對數式格紙或比例圖以繪時間數列之理由,即一個

變動之意義,須視計量此變動之底數 (base) 之大小為斷.如基數為 100 時增加 100 與基數為 10,000 時增加 10,000,其意義相等;蓋兩者之增加皆為 100% 也.就絕對變動言,增加 10,000 為增加 100 之 100 倍,故在算術格紙上此兩變動將表現為一與一百之比例,但在單對數格紙上,吾人見此兩變動之意義恰為相等.

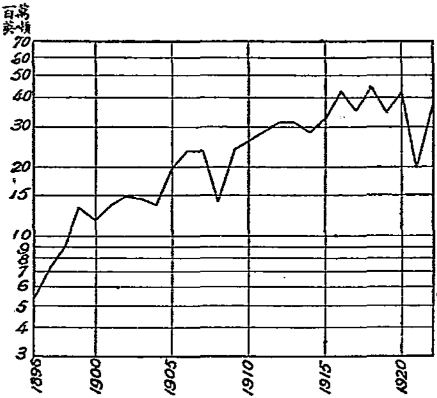


圖 14 美國鋼生產量, 1896—1922. (繪於單對數格紙上)

上圖即為單對數圖.表示 1896—1922 年間美國鋼產量.所繪者固為絕對數,但其縱線按對數尺度分格,使逐年變動之大小,與其相對量 (relative magnitude) 成正比而表現.

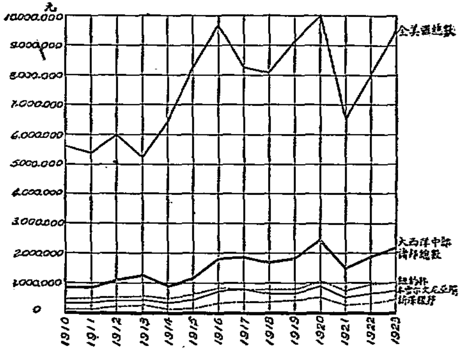


圖 15. —Acme 公司銷售量, 1910—1923 表示全美國總銷售量以及若干區域之銷售量. (繪於自然數格紙上)

將圖 15 與圖 16 兩者相比, 可見對數式或比例式分格法之若干顯著之優點.

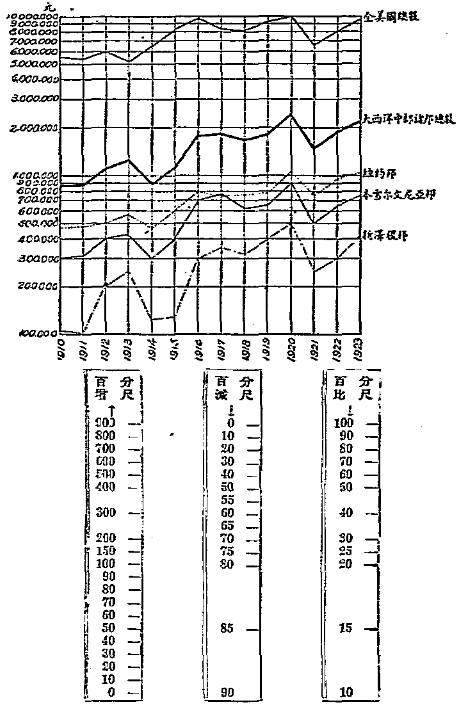


圖16.——Acme公司銷售量, 1910—1923. 表示全美國總銷售量以及若干區域之銷售量.(繪於單對數格紙上)附列增尺減尺與比尺

此兩圖所繪之材料如下:

表　二

Acme公司銷售量, 1910—1923

	本雪爾文尼亞	新澤西	紐　約	大西洋中部諸邦總數	全美國總數
1910	$ 305,000	$ 105,000	$ 465,000	$ 875,000	$ 5,600,000
1911	310,000	100,000	480,000	890,000	5,400,000
1912	400,000	200,000	500,000	1,100,000	6,000,000
1913	425,800	250,000	575,000	1,250,000	5,200,000
1914	300,000	125,000	465,000	890,000	6,400,000
1915	400,000	130,000	600,000	1,130,000	8,200,000
1916	700,000	300,000	800,000	1,800,000	9,700,000
1917	760,000	350,000	740,000	1,850,000	8,300,000
1918	630,000	320,000	750,000	1,700,000	8,100,000
1919	650,000	400,000	775,000	1,825,000	9,200,000
1920	900,000	500,000	1,050,000	2,450,000	10,000,000
1921	500,000	250,000	750,000	1,500,000	6,500,000
1922	650,000	300,000	950,000	1,900,000	8,000,000
1923	750,000	425,000	1,025,000	2,200,000	9,500,000

若將表中五個數列皆繪於一圖,而此圖爲用算數尺度分格者,則必須擇一尺度之能包含其中最高數値 10,000,000 元者.如此之一尺度將使最小諸數之相對的重要性(relative importance)減輕.由圖 15 上觀之,似乎在此圖所包括之時期內,全國銷售總數之變動甚大,中部大西洋諸邦之總數變動較小,而其他單獨表示之三邦之銷售量實際上幾無變動.此圖實與人一錯誤之印象.實際之眞相應當如圖 16 單對數尺度之圖上所表示者,在此圖上,單獨表示之各邦,其銷售量之變動,反相對的較全國總量之變動爲大.故諸數列其各項之

大小相差甚遠者,欲彼此比較時,算術尺度之圖毫無用處,反足以掩蓋眞相,表現一歪誤之景象.

圖16下方所印之尺度,特別表示對數尺度之若干最有用之特點.用增尺 (scale of increase) 計量某數列之數值在時間上兩點之間增加之量,結果可相當正確.前曾指出單對數圖上一定之垂直距離,在圖上無論何處皆代表一固定之百分數增量.如 100,000 元至 200,000 元間之垂直距離,與 200,000 元至 400,000 元之間者相同.增尺之讀法永遠由底線 (bottom) 向上.任何垂直距離皆可以用之計量,將其築某一距離而讀之,即可決定該距離所代表之百分數增量.例如欲決定新澤稷邦在 1912—1913 年間其銷售量增加幾何,以增尺計量在代表此兩年數值之兩點間之垂直距離,尺上表示增加 25%.

減尺 (scale of decrease) 之用法相同,用以計量在對數尺度上兩點間之垂直距離,而決定其所代表之百分數減量 (percentage decrease).減尺之讀法爲由頂向下.圖上之箭頭表示各種尺度讀數之方向.

比尺 (scale of comparison) 用於決定在任何時間一數列對於另一數列之百分關係 (percentage relation).例如吾人欲求 1912 年中部大西洋諸邦之總銷量與美國全國之總銷量間之百分關係,以比尺計量代表總銷量之兩點間之垂直距離,由頂向下讀之,得數約爲 18%.

上述各種尺度,欲作於一圖上,不難用按比分段之法作

之.當將許多之圖作於單對數格紙時,最好將此種尺度作成較爲固定之形式,成爲專門尺.

用單對數尺度作圖之主要利益可綜述之如下:

1. 用單對數尺度時,指數曲線化爲一直線.例如代表依複利增殖之任何數額之資金之曲線,在此種圖上成一直線.

2. 在此種圖式上,任何數列其增減之率若固定不變則成一直線.

3. 同樣大小之相對變動爲斜度相同之線所代表.故兩數列其增減之率 (rate) 相同者表現爲平行線.

4. 欲比較兩個或兩個以上之數列之變動之率,可將代表各數列之諸線之斜度互相比較.

5. 用單對數尺度分格之圖,可在同一時間旣繪絕對量而又比較其相對之變動.

6. 在單對數圖上,諸數列其各項之大小相差甚遠者亦可互相比較.

7. 吾人可直接由圖上讀出變動之百分比,並決定量與量間之百分關係.

〔頻數比較圖〕

頻數 (frequency) 者,不同各組之事或物之數也.若吾人目的爲將諸頻數相比較,則須用另一形式之圖.玆舉下列之人口統計之數字作爲本問題之說明.

表　三

美國新英格蘭諸邦農戶數1920

邦名	農戶數
梅因	48,227
新罕布什爾	20,523
威爾蒙	29,075
馬塞諸塞	32,001
路得島	4,083
康涅狄格	22,655

圖17爲一條形圖 (bar diagram), 爲1920年六邦農戶數之比較.此圖形式雖簡單,但甚合於頻數比較之目的.

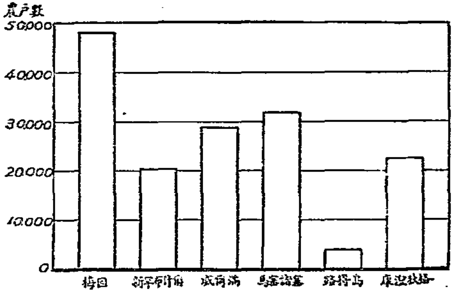


圖17. 新英格蘭諸邦農戶數, 1920年

下章中表示頻數分配時,尚有幾個條形圖之例.彼處吾人將論及在處理某種材料時,此簡單之條形圖如何變爲頻數多邊圖 (frequency polygon) 或頻數曲線 (frequency curve). 此種頻數曲線爲一非常重要之圖式,宜留待後文詳細討論之.

〔成分圖〕

吾人作圖表示材料時,有時常欲將一個總數分爲若干成分(component parts),使總數之變動與其各成分之變動同時得一表現.下表爲此種材料之一簡單之例.

表　四

*XYZ*公司生產成本費,1923

(示生產物每一單位中所含成本)

月　份	原料費	勞動費	普通費	總成本費
正　月	$ 32.00	$ 12.00	$ 6.00	$ 50.00
二　月	31.00	11.00	6.00	48.00
三　月	38.00	12.00	6.50	51.50
四　月	31.00	13.00	6.00	50.00
五　月	27.00	13.00	7.00	47.00
六　月	24.00	13.00	7.00	45.00
七　月	23 00	16.00	7.00	46.00
八　月	22.00	18.50	8.00	48.50
九　月	23.50	18.00	7.50	49.00
十　月	24.00	16.00	7.50	47.50
十一月	24.00	17.00	8.00	49.00
十二月	24.00	17.00	8.50	50.50

圖18爲此表數字之圖.由之可見當總成本費相當穩定時其中有若干之成分變動甚大.故僅知總數尙嫌不足,非知各項成分之情形不可.此種圖式非但將總數之變動明白表示,且將每一成分之變動亦明白表示.

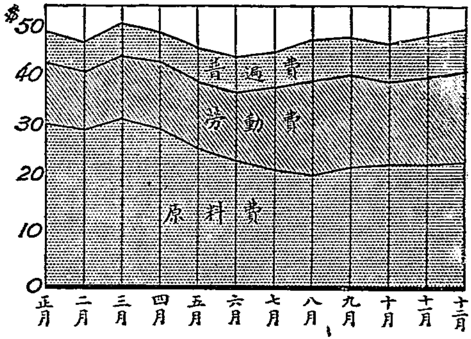


圖18.——XYZ公司1923年每月生產之本費之成分

[累積圖] (cumulative charts)

在研究一數列之發展時,在許多情形中,吾人之興趣不在前後相接之各項之個別數值,而在若干如此之項累積而得之總值.作全年之生產計畫時即可爲如是之情形.此處之問題在截至某時爲止之累積生產量與計畫中之生產量之關係.故宜有一種圖式以比較之.下列材料乃適於用累積圖表示之一例:

表　五

累積計劃生產量與累積實際生產量

Speedwell 汽車公司, 1924

月份	計劃中生產車輛數	累積數	實際生產車輛數	累積數
正月	2,000	2,000	750	750
二月	2,000	4,000	1,250	2,000
三月	2,000	6,000	1,250	3,250
四月	3,000	9,000	2,000	5,250
五月	3,000	12,000	1,500	6,750
六月	3,750	15,750	3,750	10,500
七月	3,750	19,500	4,250	14,750
八月	5,000	24,500	6,000	20,750
九月	4,000	28,500	5,750	26,500
十月	4,000	32,500		
十一月	4,000	38,500		
十二月	4,000	40,500		

表中數字表示截至九月底之情形:

圖 19 上有兩條累積曲線 (cumulative curve), 表示每屆月底時實際生產量與計劃生產量之間之關係, 且由格線可知實際生產量落後於計劃生產量之數值大致幾何. 參照表內數字 (數字之表與所給之圖常應放在一處), 即可決定其確切之關係, 此種累積圖用途甚多, 下章將述其一二,

[甘脫氏進行圖] (Gantt progress chart)

若用甘脫氏(H.L. Gantt)所發明之圖式表示上例之材料, 可得最有效之結果. 此種圖式之性質與其種種用途, 爲篇幅所限, 不能於此與以適宜之敘述, 惟可就其若干特點說明之.

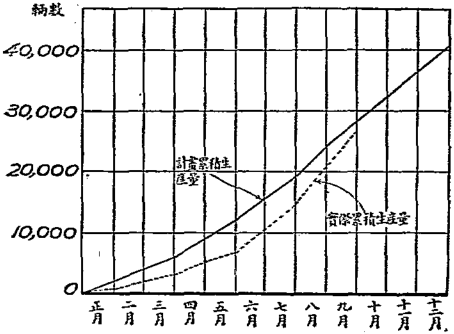


圖19.——計畫生產量與實際生產量之比較 (累積圖)
Speedwell 公司, 1924

計劃一經作成,即可用甘脫氏進行圖以校核實際之成就與計劃之關係.若已有一計劃如表五所示者,吾人可將各月與各年之數值記入一種表格,其式如圖20在每月之空格內,左方記該月之計劃生產量,右方記入截至該月底爲止之累積計劃生產量.圖20表示最初兩個月之記錄,粗黑線表示累積實際生產量,計車2000輛.上面之細黑線在正月二月之格內者,爲各該月之實際生產量之表示.若任何月之實際生產量與計劃生產量相等,狹線將延長至充滿該月之一格.若超過計劃生產量時,則表現爲兩條細黑線.

此地應注意者,每月所占之格,其代表之時間固相等,但

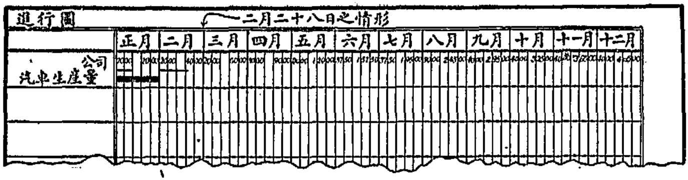


圖20.—計劃生產量與實際生產量之比較：甘脫氏進行圖(表示截至二月二十八日之情形)

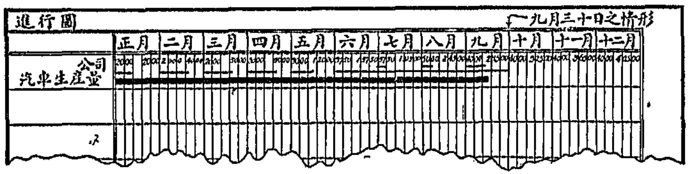


圖21.—計劃生產量與實際生產量之比較：甘脫氏進行圖(表示截至九月三十日之情形)

代表之實際生產量則不等.如正月一格之五分之一,代表實際生產量車400輛(正月全格代表2000輛),但八月之一格,其五分之一代表1000輛(八月全格代表5000輛).故欲知實際生產量之絕對數值,須參照該月計劃中原定之生產量始能決定.

圖21表示九月底之情況,圖頂之箭頭表示實際上達到之時間.由箭頭與粗線末端間之關係,可知實際生產量落後於計劃之規定凡半月.最後三月之狹線表示實際生產量每月皆超過計畫中所規定之量.

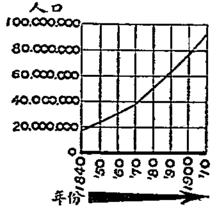
甘脫氏圖在政府機關與實業組織中有種種不同之應用.此種圖式將若干部門或若干區域之情況之發展綜合於一圖,故可經濟無數地位.用於表示目下工作進行之狀況,作實際成績與預定計畫之比較時,此種圖式爲現今所知之一切圖式中之最簡單而有效者.對於增進管理與統制之效率上,俾益實多也.

[圖示之標準規則]

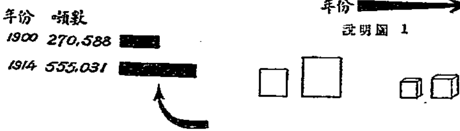
圖示之方法,旣廣泛應用於自然科學、社會科學以及實業各方面,故其法千差萬殊,難立一實用之標準.欲矯此弊,美國爰有一聯合委員會之組織,網羅關心於此問題之各團體代表,以研究所得作成報告,提供若干標準之方法以供圖示工作之用.此報告乃一簡潔明白之圖示之模範,雖已散見於許多刊物,惟轉載於此必不爲多事,蓋其傳播惟恐不遠也.玆

引委員會之建議於下:(註)

1. 圖之一般排列,應自左而右.



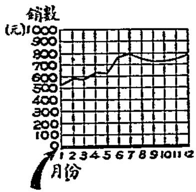
說明圖 1



說明圖 2

2. 量之能以直線之長短表示者,用面積或體積表示時易致誤解.

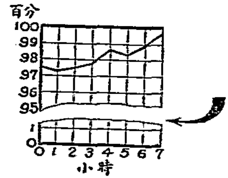
3. 繪曲線時,選垂直尺度宜儘量求0線(zelo line)現於圖上.



說明圖 3

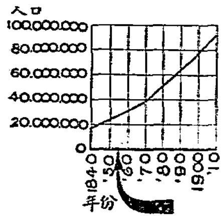
(註) 此報告爲圖示標準委員會所發表,主席爲伯林托氏(Willard C. Brinto) 載 Quarterly Publications of the American Statistical Association, Vol, 14, 790—795, 1915.

4. 若垂直尺度上之零線不能正常表現於圖上,零線仍應表現,用橫裂之圖示之.

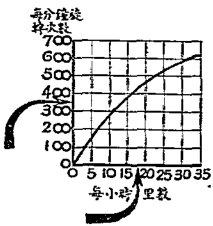


說明圖 4

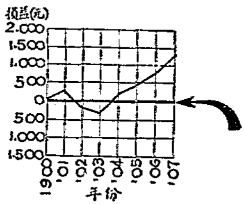
5. 零線務須與其他坐標線顯然有別.



說明圖 5 (a)

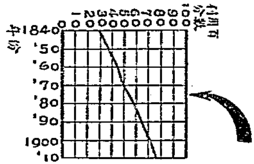


說明圖 5 (b)

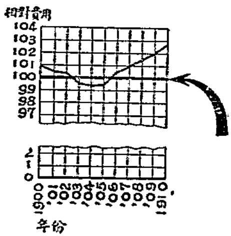


說明圖 5 (c)

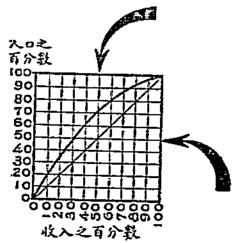
6. 代表百分數之尺度上,100%之一線或其他線之代表比較之底數(base)者,宜特別表示.



說明圖　6 (a)

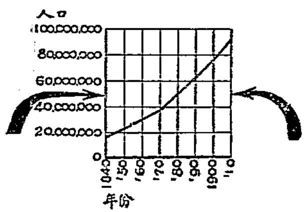


說明圖　6 (b)



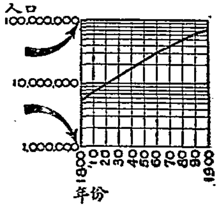
說明圖　6 (c)

7. 當圖上之尺度指示日期,而所包括之時期不成一完全之單位時,最好不將最初與最後兩縱線特別加粗;因此圖不能代表時間之終始也.

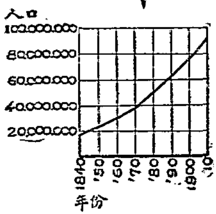


說明圖　7

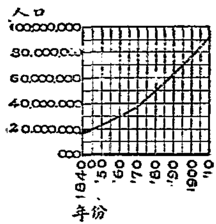
8. 將曲線繪於對數式坐標紙上時,在圖之邊限之線,應代表在對數尺度上之10之某次方.



說明圖 8



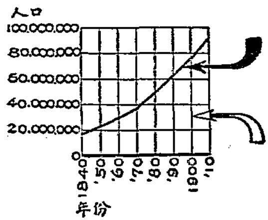
說明圖 9 (a)



說明圖 9 (b)

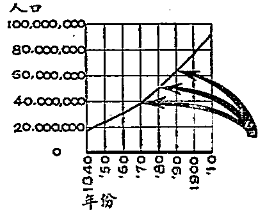
9. 除讀圖時所必須者外,勿宜多作坐標線.

10. 圖中曲線應與格線顯然有別.

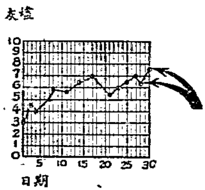


說明圖 10

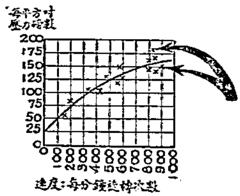
11. 給一代表許多觀察值之曲線時,須儘可能使代表個別觀察之所有各點全部明白表現於圖上.



說明圖 11 (a)

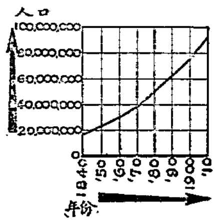


說明圖 11 (b)



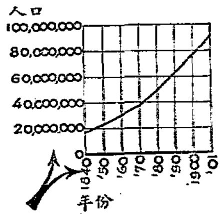
說明圖 11 (c)

12. 曲線圖之水平尺度宜常由左向右讀之,垂直尺度宜常由下向上讀之.

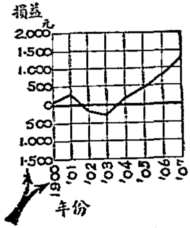


說明圖 12

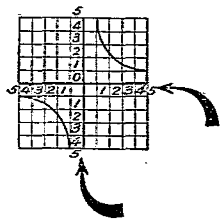
13. 標明圖上尺度之數字應排列於圖之左方及下方，或各沿縱橫兩軸.



說明圖 13 (a)

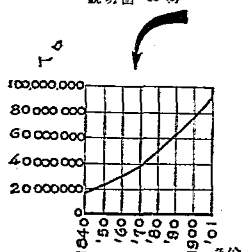


說明圖 13 (b)

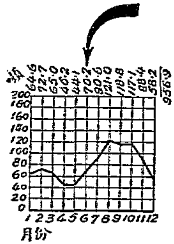


說明圖 13 (c)

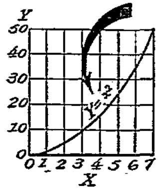
14. 數字之材料及公式，常宜包括於圖形之中.



說明圖 14 (a)

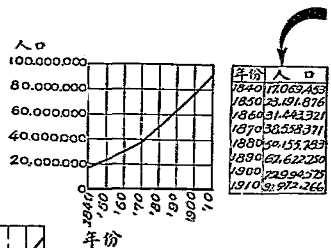


說明圖　14 (b)



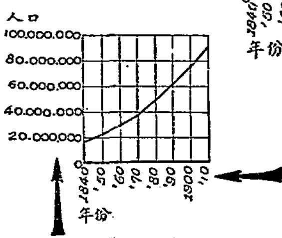
說明圖　14 (c)

15. 數字之材料如不能包括於圖內,則宜將其作成表之形式隨列圖後.



年份	人口
1840	17.069.453
1850	23.191.876
1860	31.443.321
1870	38.558.371
1880	50.155.783
1890	62.622.250
1900	75.994.575
1910	91.972.266

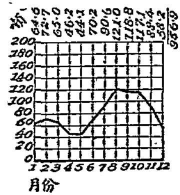
說明圖　15



說明圖　16

16. 圖上一切字與一切數之位置,宜使閱者易於由圖之底邊讀起,或由圖之右邊讀起.

17. 一圖之標題須力求明白與完全,爲求明白起見,必要時可另加副標題或註解.



一九一四年逐日第二機生產之鋁鑄物之量生產量以美噸計鋁屑錠數不在內

說明圖 17

參考書

(一) 關於數學觀念者.

關於圖示以及運用對數時所涉及之若干基本數學觀念,任何標準的代數教本與解析幾何教本皆有論述.下列各書中關於此等問題之研究,當爲有助於研究統計之學生者.

Griffin, F. L. Introduction to Mathematical Analysis.

Karsten, Karl G. Charts and Graphs.

Lipka, Joseph. Graphical and Mechanical Computation.

Mellor, J. W. Higher Mathematics for Students of Chemistry and Physics.

Schultze, Arthur. Graphic Algebra.

Steinmetz, C. P. Engineering Mathematics.

Whitehead, A. N. An Introduction to Mathematics.

(二)關於作圖法者.

Brinton, W. C. Graphic Methods fer Presenting Facts.

Clark. Wallace. The gantt Chart.

Field, J. H. Some Advantages of the Logarithmic Scale in Statistical Diagrams. Journal of Political Economy, October, 1917.

Fisher, Irving. The "Ratio" Chart. Quarterly Publications of the American Statistical Association, June, 1917.

Haskell, A. C. Graphic Charts in Business.

Haskell, A. C. How to Make and Use Graphic Charts.

Karsten, Karl. Charts and Graphs.

(出版所在及日期載本書卷末書目中)

第 三 章

統計材料之組織：頻數分配

從事實業或研究經濟之統計家,其任務包括組織、分析、及解釋有關於業務與經濟情況之數量的材料.在此基本工作以外,原始材料之蒐集亦可在其任務之內;然材料之直接由初級來源(primary source)或次級來源(secondary source)編集而得者較為多見.

首先吾人須分別兩種問題,一為分析時間數列時所遇者,一為組織及分析不包含時間因子之材料時所遇者,研究時間數列時,其主要目的為計量及分析變數數值之依時變動(chronological variation),如數年間銷數之變動,煙煤產量之增減,一般物價水平之漲落等等.然研究某時之所得分配等問題時,所用之方法卽大不相同.此時吾人將按所得之多少分為若干組,以研究每組所包含之人民各有若干.組織此後一種材料時,其普遍之問題為決定變數各值重複出現之次

數,與各値分配之情形.由此種材料組織而成之數列,名爲頻數數列(frequency series),以與時間數列或歷史數列相區別.適用於分析此兩類數列之方法大異,故將分別敍述.本節所研究之材料,其中卽或有時間要素存在,亦不成爲因子.茲述予此種材料以組織及初步分析之方法.

[未組織材料]

統計家所處理之數量材料,在其原始狀態時,爲一團無組織之原料,無形式或結構.此等原料,或取自一個實業機關之生產紀錄或銷售紀錄,或爲各種價目之混雜集合.材料若爲他人所蒐集,或已整理爲一總表之形式,然此形式或全不合於研究者之特殊目標.統計家之第一任務,卽編列數字而成一種形式,俾其意義能適合於當前目標而得表現,俾便於以之與無數相類之材料作比較,俾更進一步的分析成爲可能.前言科學方法包括觀察,推理,及實證三過程.材料爲觀察之結果,必待其有一確定之形式與條理結構,始能入推理之過程也.

下表數字示某一週內某工廠210工人所得計件工資,茲用以作此種材料之原始狀態之例:

210個工人各人在一週中所得之工資

$ 26.25	$ 28.70	$ 24.15	$ 29.75	$ 29.20	$ 30.60	$ 23.40	$ 24.75
26.70	24.35	25.75	27.20	28.30	25.25	27.75	27.60
28.20	27.30	27.80	26.35	27.40	28.30	26.60	25.75
27.70	28.60	25.30	27.80	26.40	27.30	28.35	27.00
24.30	27.80	27.60	26.30	27.40	23.50	29.60	27.80
27.60	25.35	27.55	29.00	24.10	27.00	24.50	27.25
26.15	29.30	23.10	27.10	28.50	27.45	26.15	28.35
29.95	25.55	27.55	26.60	24.25	30.00	28.55	28.00
27.30	27.90	25.25	24.10	27.45	24.55	26.55	27.55
26.75	31.00	24.00	25.35	26.50	28.30	27.95	25.55
30.25	38.55	26.75	24.60	25.75	26.55	27.80	28.90
29.55	30.00	24.60	25.75	26.30	27.00	28.25	25.25
25.75	26.25	26.30	26.75	27.90	28.30	25.70	26.30
26.60	27.00	30.75	28.60	28.10	23.50	24.75	25.15
26.30	27.25	28.15	29.10	30.10	29.90	28.55	27.30
26.55	27.55	23.00	24.50	22.55	26.55	27.55	28.10
30.70	28.60	27.90	26.80	24.10	25.25	26.30	27.90
26.90	25.30	25.80	28.85	27.55	27.30	25.00	26.00
26.55	27.80	28.60	30.55	29.50	24.10	25.15	27.15
28.10	26.30	27.10	24.60	27.80	26.30	27.90	29.80
24.10	25.15	27.50	24.25	25.70	26.80	30.15	29.30
28.15	28.65	24.55	25.85	26.10	27.00	26.80	27.55
29.00	23.00	28.60	29.30	28.55	28.80	27.55	23.60
26.10	27.15	25.75	26.80	27.15	26.30	28.55	25.80
24.55	25.80	26.75	27.30	27.55	28.25	25.60	26.30
26.85	27.30	28.10	32.00	28.15	26.30	27.75	26.25
28.60	26.00						

[序列]

若將此等數字按大小爲序而排列之,可得一較有條理之結構.數量變動之範圍與在此範圍內分配之一般狀態,可因之而明顯,使吾人能進一步加以組織.如此排列時,得下列之序列(Array):

序列: 210 個工人各人在一週中所得之工資

$ 22.55	$ 25.15	$ 26.15	$ 26.75	$ 27.45	$ 27.95	$ 28.60
23.00	25.15	26.15	26.75	27.45	27.95	28.65
23.00	25.15	26.25	26.80	27.50	28.00	28.70
23.10	25.25	26.25	26.80	27.55	28.10	28.80
23.40	25.25	26.25	26.80	27.55	28.10	28.85
23.50	25.25	26.30	26.80	27.55	28.10	28.90
23.50	25.25	26.30	26.85	27.55	28.10	29.00
23.60	25.30	26.30	26.90	27.55	28.15	29.00
24.00	25.30	26.30	27.00	27.55	28.15	29.10
24.10	25.35	26.30	27.00	27.55	28.15	29.20
24.10	25.35	26.30	27.00	27.55	28.20	29.30
24.10	25.55	26.30	27.00	27.55	28.25	29.30
24.10	25.55	26.30	27.00	27.60	28.25	29.30
24.10	25.60	26.30	27.10	27.60	28.30	29.50
24.15	25.70	26.30	27.10	27.60	28.30	29.55
24.25	25.70	26.30	27.15	27.70	28.30	29.60
24.25	25.75	26.35	27.15	27.75	28.30	29.75
24.30	25.75	26.40	27.15	27.75	28.35	29.80
24.35	25.75	26.50	27.20	27.80	28.35	29.90
24.50	25.75	26.55	27.25	27.80	28.50	30.00
24.50	25.75	26.55	27.25	27.80	28.55	30.00
24.55	25.75	26.55	27.30	27.80	28.55	30.10
24.55	25.80	26.55	27.30	27.80	28.55	30.15
24.55	25.80	26.55	27.30	27.80	28.55	30.25
24.60	25.80	26.60	27.30	27.80	28.55	30.55
24.60	25.85	26.60	27.30	27.90	28.60	30.60
24.60	26.00	26.60	27.30	27.90	28.60	30.70
24.75	26.00	26.70	27.30	27.90	28.60	30.75
24.75	27.10	26.75	27.40	27.90	28.60	31.00
25.00	26.10	26.75	27.40	27.90	28.60	32.00

[頻數表]

數字而成此序列,較之前舉任意排置者便利於研究固已多矣;然猶未能一望而知材料之全部意義,工廠經理可知

一週內最微之工資為二十二元五角半,最高為三十二元,大部分工人所得者,介於二十五元與二十九元之間;然僅為此材料之模糊印象耳.用分組法,將獲得工資之額數在某限以內者共入一類,則可得一簡潔之形式以表示工資分配.下表為分組之結果,每組之範圍(即組距 class-interval)為二元.

表 六

工人之額數分配

依一週工資分組(組距為二元)

一週之工資	得左列工資之人數(額數)
\$ 22.00 至 \$ 23.99	8
24.00 ,, 25.99	48
26.00 ,, 27.99	96
28.00 ,, 29.99	47
20.00 ,, 31.99	10
32.00 ,, 33.99	1

此表為原來數字之簡約的總結,不僅使吾人得知工資之大致範圍,並表示 210 工人之工資在此範圍內如何分配.然此中損失細節極多.如觀此表僅知該週得工資自二十四元至二十五元九角九分之間者,有四十八人,惟不能知此四十八人如何分配於此兩元之範圍內,僅就表上觀之,四十八人或可人人恰得二十四元也.此種細節之損失,乃用分組法以綜合與簡化材料時所不可避免者.

若組距縮小,損失之細節亦隨之減少,然組數增加足以使表格冗長,致複雜不易領解.下列各表所表示之材料與前表同,用以分組之組距為一元,五角及二角五分:

工人之頻數分配

(按一週所得之工資分組)

表七　(組距=$1)

一週所得之工資	頻數
$22.00—$22.99	1
23.00— 23.99	7
24.00— 24.99	21
25.00— 25.99	27
26.00— 26.99	42
27.00— 27.99	54
28.00— 28.99	34
29.00— 29.99	13
30.00— 30.99	9
31.00— 31.99	1
32.00— 32.99	1
	210

表八　(組距=$.50)

一週所得之工資	頻數
$22.50—$22.99	1
23.00— 23.49	4
23.50— 23.99	3
24.00— 24.49	11
24.50— 24.99	10
25.00— 25.49	12
25.50— 25.99	15
26.00— 26.49	22
26.50— 26.99	20
27.00— 27.49	24
27.50— 27.99	30
28.00— 28.49	17
28.50— 28.99	17
29.00— 29.49	7
29.50— 29.99	6
30.00— 30.49	5
30.50— 30.99	4
31.00— 31.49	1
31.50— 31.99	0
32.00— 32.49	1
	210

表九　(組距=$.25)

一週所得之工資	頻數
$22.75—$22.99	1
23.00— 23.24	3
23.25— 23.49	1
23.50— 23.74	3
23.75— 23.99	0
24.00— 24.24	7
24.25— 24.49	4
24.50— 24.74	8
24.75— 24.99	2
25.00— 25.24	4
25.25— 25.49	8
25.50— 25.74	5
25.75— 25.99	10
26.00— 26.24	6
26.25— 26.49	16
26.50— 26.74	10
26.75— 26.99	10
27.00— 27.24	11
27.25— 27.49	13
27.50— 27.74	14
27.75— 27.99	16
28.00— 28.24	9
28.25— 28.49	8
28.50— 28.74	14
28.75— 28.99	3
29.00— 29.24	4
29.25— 29.49	3
29.50— 29.74	3
29.75— 29.99	3
30.00— 30.24	4
30.25— 30.49	1
30.50— 30.74	3
30.75— 30.99	1
31.00— 31.24	1
31.25— 31.49	0
31.50— 31.74	0
31.75— 31.99	0
32.00— 32.24	1
	210

前列之四表(連表六在內)表示同一材料之四個不同程度之綜合.表六表七與表八表示同一普遍的特色:即極端各組頻數甚小,漸近分配之中心時,頻數逐漸有規則地增大.所分之組數越多,增減越無規則.表九以二角五分爲組距;共有三十八組.此表中分配狀態極無規則,大不對稱.而其他各表之結構則有秩序而多少近於對稱狀態.每表皆以綜合簡潔之形式表示關於工資之材料,使觀此等表格者,較之參考最初所示之雜亂數字,更易明瞭該工廠一週工資之多少及其分配狀態.此種業經整理之材料稱頻數分配 (frequency distributions), 其目的乃如此名所示,以簡約的形式表示在一變數諸值所及之範圍內該變量之分配之性質也.頻數分配表之製作,爲組織及分析可用前例代表之一類數量材料之第一步.

[頻數表製作步驟]

前文僅爲研究頻數表 (frequency tables) 之一般導言,猶未論及與製表有關之諸重要問題也.其細節猶當綜括敍述之.吾人宜知此地所取之第一步,即按數字之大小予以排序之工作,在實際作表時並非必要.若已檢視材料而決定全表之上限 (upper limit) 與下限 (lower limit), 吾人只須決定組數幾何,將組距書於適宜之空白紙上,而劃記應入各組之次數.此事旣畢,即可該計頻數,作成如前例所示之表式.惟雖在此等簡單之手續中,亦有若干點須考慮者.

[組距之大小]

決定組距(與決定組數之作用相同)之大小時,吾人須牢記一基本條件:即組之排列不可使各例於組內之分配與平均分配 (even distribution) 之狀態大相違背也.如此之排列甚爲必要.蓋於解釋頻數表時,以及將來用頻數表作根據而進行各種計算時,皆假定每組中點之值足以代表該組各例之值者也.如根據表八而作計算時,即假定在26.00元與26.50元之間之22例,皆可以該組中點之值26.25元爲其值之代表.此一假定難遇絕對眞確之時.試一翻原表,即可知其不確.若求絕對之確切,除非使表上每一數值各成一組.但簡約既屬必要,組之排列,僅能求其在不與其他要件相違背之情形下,減少其錯誤至於最小限度耳.表六爲組距對於所處理之材料過大之一例.

前節所述之要件,明示組數宜多.但第二個要件常與之矛盾,即謂組數之決定,宜使結果所得頻數成一有次序有規則之數列.若對於所處理之材料言爲分組過細,即不能得此規則性,頻數表將無結構與秩序.表九之分組不合於此要件,吾人前已指出矣.此外,爲使材料易於整理及其意義易爲了解起見,組數宜有限制.通常組距之選定,宜使組數不大於25,不小於10,究用何數,宜視材料之性質而定.上例之表八以.50元爲組距,似爲最能合於各要件者.

[組限之位置]

組距之位置非常重要,若加注意,可使製表簡捷,並利於將來之計算.組距與組限若皆為整數,製表時最為便利.各組中點值若為整數,則利於平均數及其他統計計量 (statistical measure)之計算.材料若允許用5或5之倍數作組距時,常可得適當之組限與中點,然如此之排列並非必要者.

若干材料,在作成分配時,表示有集中於某某若干數值之傾向.玆將1921年美國聯邦準備銀行貼現之商業票據數照會員銀行所負貼現率或利率分組時所得之分配,取其一部分列如下表.

貼現率(%)	票據數
6	18970
$6\frac{1}{4}$	697
$6\frac{1}{2}$	4616
$6\frac{3}{4}$	135
7	17362
$7\frac{1}{4}$	10

此處顯見整數為頻數集中之點,而$\frac{1}{2}$%處為頻數之次級集中點,而在$\frac{1}{4}$之諸值間(如$6-6\frac{1}{4}$, $6\frac{1}{4}-6\frac{1}{2}$之間)竟無一例.可見將此種材料分組時,應使組中點合於諸例集中之值,而組限亦須依此目的而決定.因根據頻數表而作之一切計算,如前所已述,乃假定每組中各項皆集中於組中點者也.故若以$\frac{1}{2}$%為前例之組距,各組應為$5\frac{3}{4}$至$6\frac{1}{4}$(但不包含$6\frac{1}{4}$), $6\frac{1}{4}$至$6\frac{3}{4}$等等,不宜為6至$6\frac{1}{2}$, $6\frac{1}{2}$至7等等也.

[觀察之精度與組之定義]

作頻數表時必須明定各組之意義,使每組各有確定之範圍,某一事例之應入某組,毫無疑問.有時可見有排列如下式之表:

組距	頻數
0至10	3
10至20	8
20至30	15
30至40	6
40至50	2

若無說明,立卽遇一問題,卽一例之具有10之數値者應入第一組乎?抑第二組乎?爲避免混淆起見,各組之範圍宜指示如下表之式.

組距	頻數
0至 9.9	3
10至19.9	8
20至29.9	15
30至39.9	6
40至49.9	2

但此法所能解決之困難,僅限於處理觀察精確至小數點下一位爲止之數値時所遇者.若精確之程度僅至個位爲止,(卽諸例之位於9.5與10.5之間者,其値皆爲10),則僅改變組限之記法並不能解決如何處置恰在組限上之事例之問題.在此種情況之下,位於組限上之一例可分割爲二,半屬前組,半屬後組.

游爾(Yule)氏在"統計理論入門"一書中,曾定一有用之

原則,即決定組限時用於記明組限之數字,宜較原料中各例項之數值多一位小數或含有更小之分數.就前例言,若觀察所得數值精確至小數點下一位,即表示9.9之一數值實際為位於9.85與9.95間之各值,則確切記明組別時,組距應以0至9.95, 9.95至19.95等表示之.用此等組限時,組中點之值將為4.95, 14.95等等.吾人在製作及應用如前舉數表之表時,宜時時記憶其組限之眞正意義.無論何處組限之決定,皆必須視觀察之精度而定.

指定各組之範圍時,若能將各組之上限與下限完全記明,如前舉各表,則製表之工作可因而簡單化.單獨記明各組之下限或其組中點之值,皆易生錯誤.惟各組中點之值,特別於以表為根據而作計算時,宜專設一行以記明之.

[其他要件]

組距之大小須全表一律,以便比較.間有取不同之組距製表者,於是在表之一部,以相差在5以內之數值為一組,而另一部之各組則以10為數值之範圍.此兩部顯然不能相比.吾人如對於某一部分欲有一特別細密之記錄時,最宜專門為此特別之一部分另作一附表,如此可無損主要表之效用.

另一要件之性質相同,即不應有無定限之組 (indeterminate class) 存在也.無定限之組,亦稱"空端"之組 (open-end class),即組之範圍未與規定者也.若將取得計件工資30元及30元以上之人悉歸入一組,而名之曰"30元及30元以上",則

此組之上限全不確定.此表作爲計算之根據時,此缺點卽成爲致命之傷.吾人若遇十分特殊之情況,不得不將一如此之無定限之組列入表中時,應將實際之數值附記於表後.

前兩節中所述之錯誤,可由下例見之.

表 十

ABC 公司雇工之頻數分配

根據每日工資分組

工資之組別(元)	每組人數(頻數)
3.00 以下	15
3.00 至 3.99	30
4.00 至 4.99	40
5.00 至 6.49	30
6.50 至 8.99	10
9.00 以上	5
	130

例中空端之組,其範圍不能確定;中間各組,其範圍有兩者爲1元,有一者爲1.50元,有一者爲2.50元,此種之表實無何等價值.

[統計表之結構]

前所述者,限於製作頻數表時所遇之多少爲技術性之問題.而對於全表之形式,橫列與縱行之排列,標題以及記數等則未着一字.製表總原則尙未規定.此等原則雖不能於此詳述,但與統計表之結構有關之普通條件則可提示如下.

統計表爲以摘要之形式,表示大批數量材料之方法.此

摘要如不清楚,重要,簡括,而易解,則吾人殊未因製表分組之工作而實有所獲.蓋冗雜無組織之表猶如散漫難解之語言也.每表應含有一定之目的,其編製須使此目的明白實現.達此目的之方法須各視其有關之特殊條件而定.然在可能範圍內應合於標準的慣例.下列爲有助於決定統計表形式與排列之普通原則.

(1) 總標題宜清楚簡括,爲表中材料之完全的描寫.

(2) 各行各列之標題宜簡潔,其意義須確定無疑.

(3) 變量須由左而右由上而下遞增(若可能時).

(4) 直行與橫列可分別註以數字以便參照.

(5) 所用之計量單位應明白記出.

(6) 務須註明材料之出處.

(7) 每表應自成一單位,自足自明.解釋此表時所必須之一切說明,必須成包括於表內爲本體之一部,或作爲附註而附於表後.

[頻數分配之圖示]

頻數分配,如上述者,摘材料之要,使合於進一步之整理,其所完成之統計作用,至重且鉅.此種之頻數分配不僅可以用表表示,且可用前面說明之坐標系一般原則,作圖以表示之.頻數分配之多種特色,常因圖示而得最明顯之表現.

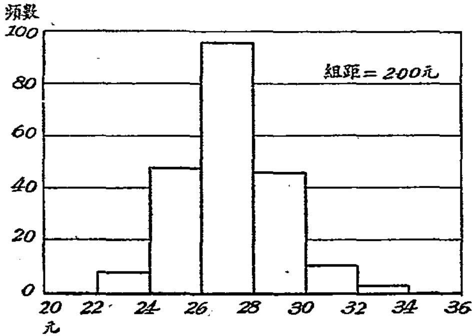


圖22—直方圖:210個工人按一週所得工資分組時之分配(組距=$2.00)

表六表示210人一週工資分配,以2.00元爲組距.圖22爲其圖示.組距沿x軸描記,相當之分組頻數(class-frequencies)沿y軸描記,兩軸自選適宜之尺度.應注意者,橫坐標尺度不起於0而起於20元,由0至20之一段業已省去,爲求作圖之便利也.然學者欲得兩變數間關係之真確印象,必須牢記此點.此圖名爲直方圖(column diagram 或 histogram).作圖時須將代表各組距之上限與下限之兩點,用許多橫線連結之.圖上各長方形之面積,與其所代表之例數成比例,而總面積則代表210例.故此圖足以明白表示分配之真相,確切顯出每一工資組中所包括之工人之相對數.

但圖22組距過大,稍違事實,删抹細節過多,觀之不能得一各項配列之眞實概念.圖23之直方圖,乃描繪以1元爲組距時之分配情形者,組距略小,圖形較有規則而接近於對稱之狀態.圖24以.50元爲組距,情形相同.至於圖25則以.25元爲組距,前已指出,此種分組對於所處理之材料失之過細,結果得一甚不規則之結構.(注意,四個圖之垂直尺度其大小不等,故須參考尺度上數碼,始能將彼此之分組頻數互相比較).

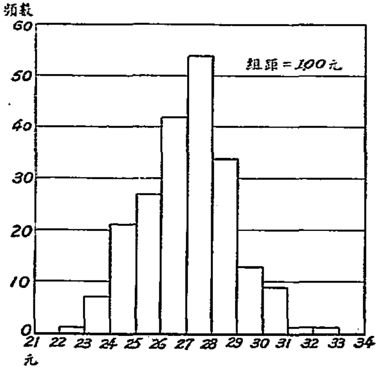


圖23——直方圖:210個工人按一週所得工資分組時之分配(組距=$1.00)

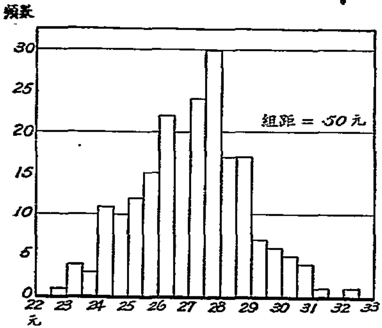


圖24——直方圖:210個工人按一週所得工資分組時之分配(組距=$.50)

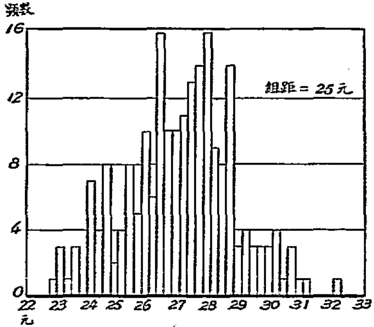


圖25——直方圖:210個工人按一週所得工資分組時之分配(組距=$.25)

下列圖26,27及28乃相當於圖22,23及24各直方圖之

頻數多邊形 (frequency polygons)

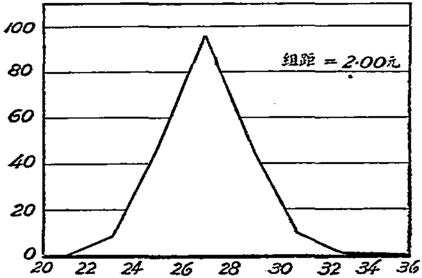


圖 26——頻數多邊形：210 個工人按一週所得工資分組時之分配（組距=$ 2.00）

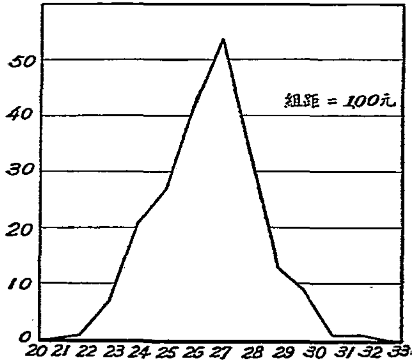


圖 27——頻數多邊形：210 個工人按一週所得工資分組時之分配（組距=$ 1.00）

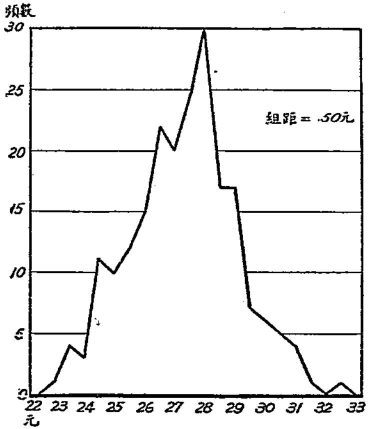


圖28——頻數多邊形:210個工人按一週所得之工資分組之分配(組距=$.50)

上列各多邊形之作法,係將各組中點爲橫坐標,諸分組頻數爲縱坐標,更以一折線連結所得各點.爲完成此形,故將表中最低組以下之一組及最高組以上之一組包括在內,此兩組之分組頻數爲0.於是多邊形之兩端接於底線,與外加二組之中點相連.頻數多邊形曲線下之全面積,代表全體例數.但某一組距所得之面積,因該組內兩端分配不平均之故,

不能謂與該組距所包例數成比例.惟各組中點縱坐標之高,當然代表該組之頻數.

[曲線之修勻]

試再觀因組距不同而產生之結果.將組距漸次縮小,直至某點爲止,直方圖與多邊圖漸次愈加光滑而規則,但過此點後,卽有缺口出現,組距較大時分組頻數所表現之有規則的變化,遂因不規則而大異於常態之組之出現而破壞,圖25中此種不規則之組異常明顯.就常態言,依工資多少而分之不同各組,其工人數目,由工資之最低點起遞增,至27.50元附近達最多數,過此又復遞減,至工資之最高點32元時只有工人一人;不規則之組顯然爲此常態之例外.蓋210人旣全體從事於同一工作,而其工資之多少僅視其熟練迅速之程度而定,則各組人數之分配應有規則,若吾人之材料不僅爲一週之紀錄,而爲五十二週之紀錄;卽若取210人每人在一年中每週平均所得額爲材料,則用小組距時所得分組頻數,當較現今實際所得者較有規則;蓋意外變動之爲一週所獨有者將因而除去也.或者,吾人若取10920人(210人之52倍)各人在一週間所得工資之紀錄爲材料,結果亦將相同.故吾人欲得光滑而有規則之曲線,不僅須縮短組距,且須增加所包括

之例數,以除去影響於少數觀察之偶然不規則性.若用精密之分組處理少數事例,結果可以圖25代表之.然增加例數,常爲實際上所不可能之事.如能得一實際可行之法,能用以求一近似之分配,近似於用極小組距處理極多事例時所可得者,則吾人所最希望者也.曲線修勻法 (curve smoothing) 卽求此近似分配之工具.吾人使用此法,能得一光滑之頻數曲線,無因小變化而起之不規則性.

此光滑之頻數曲線足以代表材料之眞正的潛在分配狀態.前曾指出頻數多邊形之面積並不與所包含之例數成比例,其原因卽由於材料之不規則性,經修勻以後之頻數曲線,此種不規則性業已除去,故立於橫軸尺度上某兩點之縱線間之面積,可假定爲與該兩值間之理論頻數 (theoretical frequency) 成比例.並且,吾人已得一代表趨勢之光滑曲線後,可以補插法 (interpolation, 或譯內推法) 求原表未載之中間數值之頻數(註 1)

下列材料(註 2)代表1918年美國個人所得額 (income) 在4,000元以下者之分配,以作修勻法之例.

(註 1) 因實際上之限制,統計工作之材料必然具有許多缺漏,變數之已知諸值不相連續,補插法乃估量在變數之兩已知值間之數值或於曲線上兩已知點間作第三點之方法.補插之數值苟能與已知數值相符合,則此法之運用可稱爲精確.

(註 2) 見 Vol. I, Income in the United States, National Bureau of Economic Research. New York, Harcourt, Brace & Co., 1921, 132—33.

表十一

美國個人所得額分配狀態,1918年

(包括所得額在4000元以下者)

所得額分組(註1)	人數(註2)
0至 100元	62,809
100 ,, 200	103,704
200 ,, 300	209,037
300 ,, 400	489,963
400 ,, 500	961,991
500 ,, 600	1,549,974
600 ,, 700	2,154,474
700 ,, 800	2,668,466
800 ,, 900	3,013,034
900 ,, 1,000	3,144,722
1,000 ,, 1,100	3,074,351
1,100 ,, 1,200	2,850,526
1,200 ,, 1,300	2,535,285
1,300 ,, 1,400	2,205,728
1,400 ,, 1,500	1,832,230
1,500 ,, 1,600	1,512,649
1,600 ,, 1,700	1,234,397
1,700 ,, 1,800	999 996
1,800 ,, 1,900	811 236
1,900 ,, 2,000	663,789
2,000 ,, 2,100	549,787
2,100 ,, 2,200	463 222
2,200 ,, 2,300	395,115
2,300 ,, 2,400	340,141
2,400 ,, 2,500	295,490
2,500 ,, 2,600	258,650
2,600 ,, 2,700	227,731
2,700 ,, 2,800	201,488
2,800 ,, 2,900	178,901
2,900 ,, 3,000	154,499
3,000 ,, 3,100	142,802
3,100 ,, 3,200	128,217
3,200 ,, 3,300	115,583
3,300 ,, 3,400	104,504
3,400 ,, 3,500	94,803
3,500 ,, 3,600	86,405
3,600 ,, 3,700	79,023
3,700 ,, 3,800	72,562
3,800 ,, 3,900	66,900
3,900 ,, 4,000	61,894

(註1) 各組之上限屬於其上之一組,如個人之所得額為100元者歸入第二組是也.

(註2) 國立經濟研究局報告上謂:"以下之數直計至個位為止,此種計算上之精確,除技術的意義外,未敢謂有其他意義也".

下列圖29,30及31爲此材料之直方圖,其組距各爲500元,200元及100元.組距縮小時,直方圖之形式趨規則與一致,原來材料只許將組距縮小至100元爲止.然吾人之問題,爲求一將組距不斷縮小時所漸漸接近之潛在的分配狀態,

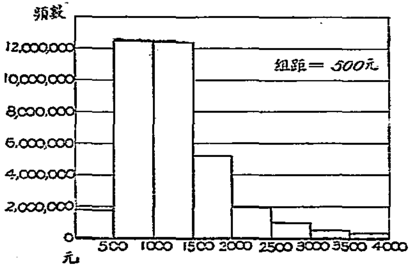


圖29.—直方圖:1918年美國個人所得額分配包括所得額在4000元以下者.(組距=500元)

若吾人以一光滑之曲線,包合之面積與直方圖相等,而其由直方圖中每一長方形割去之部分能與加於該長方形之部分約略相等者,如此之曲線即代表吾人所求之分配之頻數曲線.此曲線所包含之面積必須與直方圖相等之一點,爲基本要件;而不改變各長方形面積一點,常因大異尋常之組之存在而有例外,然仍爲有助於實際工作之原則也.(配合光滑曲線於材料之比較精細之方法,須待後面再與討論.然單以觀察修勻之法如前面所述者,所得曲線,與吾人所求者已可十分近似).

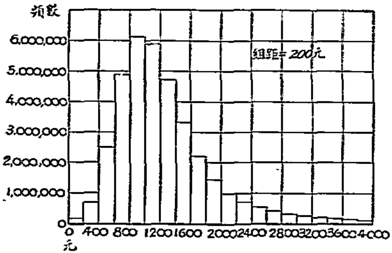


圖30——直方圖：美國1918年個人所得額分配.包括所得額在4000元以下之一切個人.(組距=$ 200)

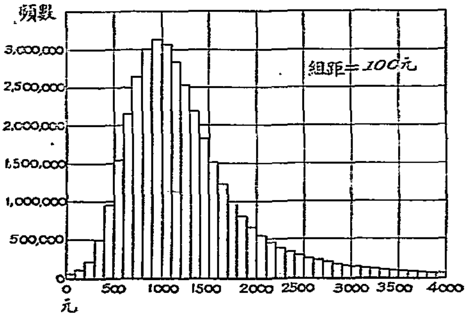


圖31——直方圖：美國1918年個人所得額分配.包括所得額在4000元以下之一切個人.(組距=$ 100)

圖32爲將圖31之所得額分配之直方圖修匀之結果,所得額各組間陡漲陡落頗不自然之處業已修平,而得一以無

限小之增量漸變之曲線,即吾人若取數百萬人之所得額而作一圖時所可預期之代表眞實的分配狀態之曲線也.吾人所求者卽此,——爲一眞實的潛在分配狀態之近似,而消去因分組而生之尖銳的屈折者也.

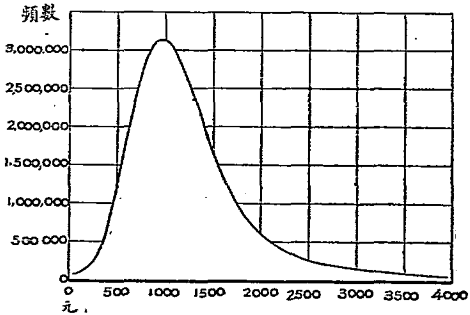


圖32 —頻數曲線:1918年美國個人所得額分配,包括所得額在4000元以下之一切個人.(根據圖31化出)

[連續數列與非連續數列]

修勻法是否合於論理,視所整理之材料之性質爲斷.由此觀點論之,頻數數列之如上述者,可分爲兩類:卽連續數列(continuous series)與非連續數列(non-continuous series或discrete series)是也.連續數列者,自變數之值增減時,其增量爲無限小者也.非連續數列者,自變數所代表之現象當其值變化

時永有一定之數量者也.代表其潛在數値 (underlying value) 之曲線,非如連續數列之曲線之爲光滑的形式,而爲跳躍的形式.

應請特別注意者,作此連續與非連續之分時,吾人所指之數値,乃就實際材料所由取汲之潛在的現象界 (underlying universe of phenomena) 而言.任何一例,不論其代表連續數列或非連續數列,其自變數之諸値間常有間隙 (breaks) 存在.連續數列之所以亦有間隙者,由於吾人在計量時所用之工具以及感官皆爲受限制者.如量百人之身長,頻數數列之自變數(身長)將依一定數量而增加;吾人或量至寸爲止,或量至分爲止,但若集萬人或千萬人依長短而排列之,其相鄰兩人間身長之差,將極微細,長度,一連續變數也.雖所舉之例其値不相連續,亦無損其爲連續性之本質.

變數之如利率或貼現率者,其分配情形乃完全不同.若取100個行情順大小爲序排列之,其變動固與前述計量人體長短之例相同,將爲不連續的形式;然在身長之例,其潛在的數値,若取千百萬人而定之,結果將爲一連續之變動;而利率及貼現率之行情,縱集例多至於無限,其潛在數値間終有間隙.蓋貼現率之漲落,或以$\frac{1}{4}$%爲定量,或以$\frac{1}{2}$%爲定量,而不以無限小之數爲量者也.故此種數列名爲非連續數列.

修勻法使吾人可就有限之例而求得一近似於由無限大之例所能求出之分配狀態.此法乃假定吾人實際研究之

例之不規則性純起於偶然,而其潛在數值應表現爲一連續不斷之變動者.因此,修勻法顯然僅在應用於連續數列時始爲合理.吾人可修勻身長之直方圖以表示全體人民之眞正的潛在的分配狀態.而根據此修勻法而作之補插亦爲有效.但修勻法應用於顯著之非連續數列則甚不合理.如作一光滑之曲線表示貼現率之分配以求貼現率4.3675%之理論頻數乃毫無意義之事.雖然,在實際統計工作中,固常有宜於將非連續數列作爲連續數列同樣處理者,遇此種情形,修勻法亦可應用.惟當解釋及應用修勻曲線(smoothed curves)時,連續變動與非連續變動之重要的理論上之差異,應時時明記於心中也.(註)

[統計材料之累積式的排列]

爲某種目的,材料之取累積式排列者,較之上述頻數表之各組獨立者更合於用,下列材料可以說明此種累積式之排列之若干利益.

調查電話桿之經用程度得結果如下.

(註) 欲知修勻法之詳細請參考 G. H. Knibbs, "The Theory and Justification of Curve Smoothing," in H. Secrist, Readings and Problems in Statistical Methods. N. Y., Macmillan, 1920, 278—82.

表十二

248707個電話桿頻數分配表

(按經用年數分組)

經用年數	桿數(頻數)
0— 0.9	1,150
1— 1.9	4,221
2— 2.9	10,692
3— 3.9	13,966
4— 4.9	16,633
5— 5.9	18,211
6— 6.9	19,011
7— 7.9	19 260
8— 8.9	20,909
9— 9.9	19,879
10—10.9	20,764
11—11.9	15,454
12—12.9	14,237
13—13.9	13,779
14—14.9	9,764
15—15.9	8,534
16—16.9	7,654
17—17.9	6,918
18—18.9	4,591
19—19.9	1,798
20—20.9	815
21—21.9	313
22—22.9	102
23—23.9	47

上表示第一年中電話桿損壞者1,150根,滿一年以上而未及二年即損壞者4,221根,等等.此乃平常之頻數表.若將其中數字取累積式排列之如下表,其意義與作用將大勝此表.

表十三

248,707電桿累積類數分配表

(按經用年數分組)

經用年數	桿數(類數)
少於 1年	1,150
,, 2	5,371
,, 3	16,063
,, 4	30,029
,, 5	46 662
,, 6	64,873
,, 7	83,884
,, 8	103,144
,, 9	124,053
,, 10	143,932
,, 11	164 696
,, 12	180,150
,, 13	194,387
,, 14	208,166
,, 15	217,930
,, 16	226,464
,, 17	234,123
,, 18	241,041
,, 19	245 632
,, 20	247,430
,, 21	248,245
,, 22	248,558
,, 23	248,660
,, 24	248,707

吾人必須注意類數數列可用兩種方式累積.由上表求經用不及若干年數之桿數甚易,但更可作一相反之表,表示超過若干年尚存在之電桿總數,此表常較前者更爲便利.電桿總數作反累積時得表如下:

表十四

248,707電話桿累積頻數表

(按經年數分組,反累積)

(1) 經用年數	(2) 桿數(頻數)	(3) 百分比
0年以上	248,707	100.0
1	247,557	99.5
2	243,336	97.8
3	232,644	93.6
4	218,678	88.0
5	202,045	81.2
6	183,834	73 8
7	164,823	66.3
8	145,563	58.3
9	124,654	50.1
10	104,775	42.1
11	84 011	33.8
12	68,557	27.6
13	54,320	21.8
14	40,541	16.3
15	30,777	12.4
16	22,243	8.9
17	14,584	5.9
18	7,666	3.1
19	3,075	1.2
20	1,277	0.5
21	462	0.2
22	149	0.06
23	47	0.02
24	0	0.00

上列之兩種累積頻數表,對於多種材料之處理上常極有利益.統計壽命之表常取此式排列.用科學方法研究折售(depreciation)時,對於各種設備皆須製詳盡之"死亡率表"(mortality table),而此種表亦以作成累積式者最為有用.有時

宜將頻數化爲百分數,如列於表十四之第(3)項者.但吾人宜常記百分比之意義須視其所根據之絕對數而定.

[累積頻數曲線]

此種累積形式之材料,其普遍適用之性,爲綜合材料時不得不用之分組方法所限制.吾人除非用數學方法實行補插,否則僅能限於兩表上所實際指出之各點.因此宜有一普遍化的累積頻數曲線,與前節所述之修勻頻數曲線相類似者.若將表十三各值繪於坐標紙上(以經用年數爲橫坐標,以與之相當之桿數爲縱坐標),通過所得各點而作一光滑之曲線,即得圖 33 所示之累積頻數曲線 (ogive, or cumulative frequency curve).

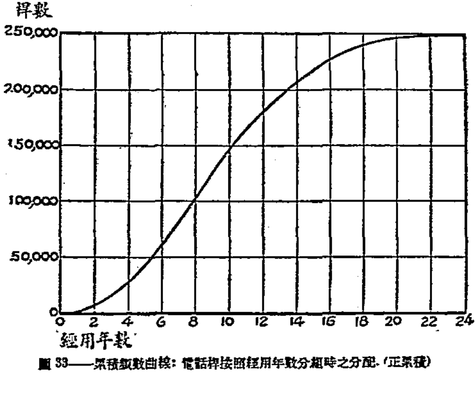


圖 33——累積頻數曲線：電話桿按照經用年數分組時之分配.（正累積）

圖34所給者爲表十四之材料.

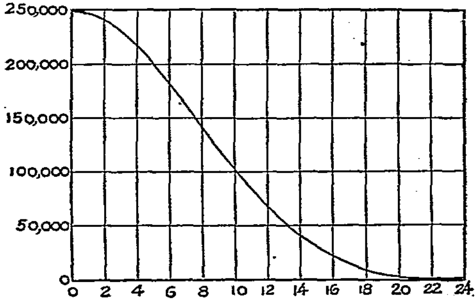


圖34——累積頻數曲線：電話桿按照經用年數分組時之分配(反累積)

此種曲線乃頻數數列圖示中最有效最有用者之一.因特殊之組距而受之種種限制,頗已大部除去;組距組數可變,曲線之形式根本如一.平常形式之頻數曲線,若非分組之法相同,則不能相比;累積頻數曲線則不受此種限制.不特此也,全圖組距不一致時,足使平常頻數曲線改其正當形式,惟於累積頻數曲線則不生影響也.

累積曲線特宜於補插.若欲知經用未滿$15\frac{1}{2}$年之桿數,由圖33上橫坐標$15\frac{1}{2}$處求曲線之縱坐標之值,得補插之數222,000.欲知經用至$8\frac{1}{2}$年以上之桿數,可就圖34以同法求之,得補插之數135,000.

對於此種曲線,更有一種補插法,即決定某一間段(interval)中之例數是也.此法可使吾人不受原表組距之限制.例如,欲知經用$10\frac{1}{2}$年以上15年以下之桿數,讀表或觀圖而知15年以下之桿數爲217,930.用前述方法補插,得經用不足$10\frac{1}{2}$年者154,000.從前數減去後數得63,930,此即$10\frac{1}{2}$年至15年之一間段中損壞之桿數也.此數當然僅爲實際數值之一近似值,(用修勻法與補插法求得之一切數值,莫不爲近似值).

吾人須知累積頻數曲線可直接由序列作出,不須先作頻數表以爲過渡.此種曲線,實際上可視作僅爲序列之圖示,代表統計上組織材料時最簡單方式之一,且爲整理數量材料之一最有效方法.

[累積頻數曲線與頻數曲線之關係]

累積頻數曲線與頻數曲線,僅同一材料之兩種不同排列,各有其特殊優點,若知其間結構關係,兩者之特點更可明瞭.圖35示此結構關係:

此圖根據下列頻數表而作,爲美國木廠之分配,以每生產一千尺木板所需之勞動費用爲分組之標準.(根據Monthly Labor Review, 1923,1,14.所載Ethelbert Stewart氏"木廠之勞動效率與生產力"一文,9元以上分散之七例删去).

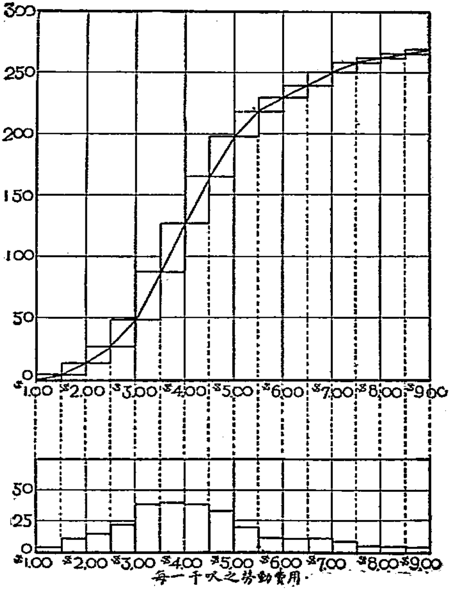


圖35——美國木廠之分配(依照1920年勞動費用分組)

表示累積頻數曲線與頻數曲線間結構關係

表十五

美國269所木廠頻數分配

(依1921年勞動費用分組)

木材千尺之勞動費用(包括全體雇員在內)	廠數(頻數)
1.00—1.49元	3
1.50—1.99	10
2.00—2.49	14
2.50—2.99	22
3.00—3.49	38
3.50—3.99	40
4.00—4.49	38
4.50—4.99	33
5.00—5.49	20
5.50—5.99	11
6.00—6.49	10
6.50—6.99	11
7.00—7.49	8
7.50—7.99	4
8.00—8.49	4
8.50—8.99	3
	269

圖35上段表示作累積頻數曲線之方法.各長方形之面積與該組所括例數成比例,若直方圖然.既用累積法,故各長方形之底各爲其前一組之累積頻數.故第一長方形之y值(頻數)爲3,以0爲底;第二組之y值爲10,以3爲底,餘可類推.連結諸長方形之曲線,當頻數低時其斜度甚和緩,頻數漸大而斜度亦陡;近分配之上限時頻數復小,曲線亦趨平易.此乃累積頻數曲線也.

當代表分組頻數諸長方形皆以0線爲底,而x諸值與前

相同時,卽得以前所敍述之直方圖.由此直方圖可化出頻數多邊形與修勻之頻數曲線.

參考書

Bowley, A. L. Elements of Statistics (52—81)

Chaddock, R. E. Principles and Methods of Statistics. (Chaps IV. V.)

Day, Edmund E. Standardization of Construction of Statistical Tables. Quarterly, Publications, American Statistical Association, March, 1920.

Jones, D. C. A. First Course in Statistics. (5—21)

Kelley, Truman L. Statistical Method. (1—37)

King, W. I. Elements of Statistical Method. (83—107)

Pearl Raymond, Medical Biometry and Statistical Method (116—157)

Rugg, H. O. Statistical Methods Applied to Education (57—94)

Secrist, Horace. Introduction to Statistical Methods (116—157)

Secrist, Horace. Readings and Problems in Statistical Methods (242—271)

Yule, G. Udney. An Introduction to the Theory of Statistics (75—105)

第四章

頻數分配之敍述：平均數

數量材料之分類以及頻數分配表之編製,在組織與分析工作上爲一重要階段.藉分類之助,材料之內在結構得以顯示,無數材料本質上的統一亦可暴露.惟此僅統計分析工作之第一步耳.欲與一批材料之重要特質以更確切之計量與表述,其方法猶待發展.爲某些目的,頻數分配之本身更須綜合與約縮,擷取精華,鑄成三四個重要之數字.

每一頻數分配若自成一新奇而獨特之問題,自依其特殊之規律,則此種分配之研究與敍述將甚困難.幸情形並非如此.數量材料,縱其代表之領域大相懸絕,當集爲頻數分配之時,仍表示若干共同之特性,依據若干普遍之規律.於是在一領域中獲得之經驗,可爲其他領域工作之嚮導.無數材料之形態旣有此一致之性,吾人乃得發展一普遍之方法,以組織分析及比較來自科學研究之無數領域之種種計量.

[頻數分配之比較]

數量事實之全領域中,有一共同之排列法則存在.欲明此事,莫如取代表各種不同材料之頻數分配而作比較.茲請

注意下列頻數分配與頻數曲線之特性,並比較之.

圖36之曲線根據表十六之材料而作,示某軍隊中18,780人身高之分配.

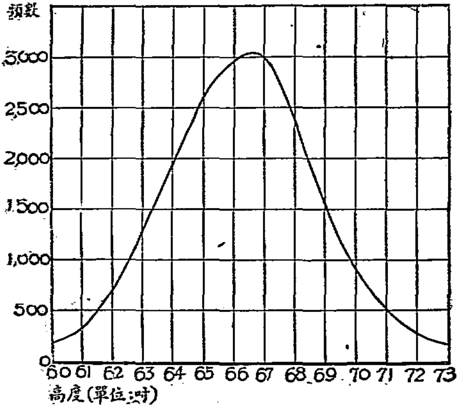


圖36——頻數曲線: 18,780個兵士依照身高之分配.

表十六

兵士身高分配

身高(吋)	人數	身高(吋)	人數
60+	197	67+	3,071
61+	317	68+	2,287
62+	692	69+	1,599
63+	1,289	70+	878
64+	1,961	71+	520
65+	2,613	72+	262
66+	2,974	73+	174
		總數	18,780

(根據G. C. Whipple. Vital Statistics, 377)

圖37之頻數曲線根據勃拉特萊(Bradley)氏470次天文計量中觀測之差誤而作,材料列於圖下表中.

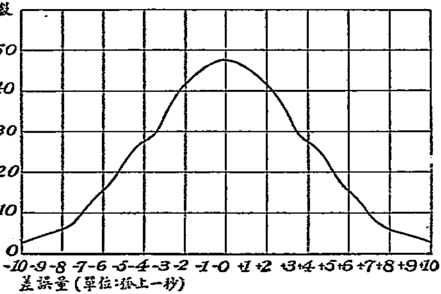


圖37——頻數曲線:某種天文觀測上之差誤之分配.

表十七(註)

某種天文計量上觀測值差誤之分配

差誤量 (單位:弧上一秒)	差誤次數
0.0與0.1間	94
0.1 ,, 0.2 ,,	88
0.2 ,, 0.3 ,,	78
0.3 ,, 0.4 ,,	58
0.4 ,, 0.5 ,,	51
0.5 ,, 0.6 ,,	36
0.6 ,, 0.7 ,,	26
0.7 ,, 0.8 ,,	14
0.8 ,, 0.9 ,,	10
0.9 ,, 1.0 ,,	7
1.0以上	8
總數	470

(註) 錄自 Mellor, Higher Mathematics for Students of Chemistry and Physics, 514.

材料中未將差誤之高於眞值者與低於眞值者分開.故根據之以作圖時,僅能得圖上曲線之半.其他一半之曲線乃根據一假定而得,此假定即低於眞値之差誤,若其量與高於眞值之差誤同大,則兩者之次數亦相等也.此種假定用於此地甚爲合宜.

將一砲正對一靶(一點)精確校準後發射百次,百彈所着之點將環繞分佈於靶之周圍.不論礮及其校準爲如何精確,恰能中的者僅屬少數.惟着彈之點,其分佈將甚有規則.若繪一長方形以表示全部中彈之點,並劃分此長方形(或散佈帶 zone of dispersion)爲八個相等部分,各部分中彈之點將如下圖.(若任舉一實例,情形或與此稍有出入,惟積長期之經驗終可得如此之分配).

2	7	16	25	25	16	7	2

圖 38—砲彈散佈帶:中彈次數之理論百分分配.

一切鎗砲莫不與此普遍之原則相合.鎗砲越精,則散佈帶越小,惟帶內之理論的頻數分配莫不相同.用於校準砲位之發射規則乃根據此事實而定者也.

吾人可將發射之實際紀錄與此理論分配對照比較.下表乃一砲正對二百碼外一個固定之靶之中點發射一千次之紀錄(註)靶上橫分十一等分.

(註) 此紀錄載 Report of Ihi Chief Ordnance, 1878 附錄 S. 結果載 Mauofield M--riman 氏 The Methio of Least Squares, 1897.

表十八

從一個砲發出一千彈之分配

靶上第?格	中彈次數
1(上端)	1
2	4
3	10
4	89
5	190
6	212
7	204
8	193
9	79
10	16
11(下端)	2
總數	1000

圖39爲此結果之圖示.

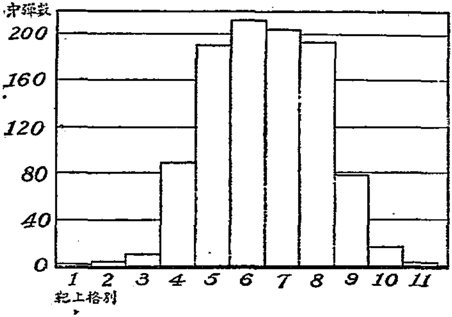


圖39——直方圖：從一個砲發出一千彈之分配.

此處散佈帶分爲十一等分,而前述理論分配時則分爲八等分,彼此不同,故不能從直接相比.然此處之分配,其一般形式與前述諸例相同.其中稍微偏集於靶之下半之傾向,無疑因架砲時未能完全矯正重心之故.

向上拋擲錢幣,其下落時所得正反面之分配,吾人認爲純由偶然所決定.在一次實驗中,將10錢幣上擲100次.正面同時出現之幣數及其頻數,(在此種情形下,在每一次拋擲中,正面最多能有10個,亦可全無正面).列於下表:

表十九

拋擲錢幣的結果之分配

(表示10個錢幣之100次試驗)

正面向上之幣數	出現次數
10	0
9	1
8	4
7	7
6	23
5	30
4	20
3	9
2	5
1	1
0	0
	100

圖40為上列之頻數分配之圖.

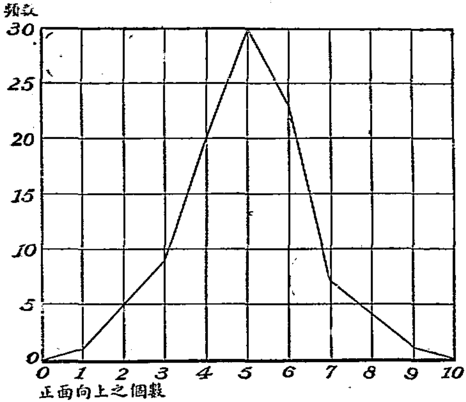


圖40—頻數多邊形:拋擲錢幣試驗中正面向上之幣數分配.

[經濟材料之分配]

在前述之四個互相懸殊的領域中,吾人見數量材料之排列,似有近於一致之規律存在.然前數例所代表者,非經濟界之事實.經濟材料果表示有同樣之普遍之性質乎?請參照第三章諸例,即關於工人每週所得工資,電話桿經用年數,木廠之勞動費用,以及美國個人所得額在4000元以下者之頻數分配,與前舉之四例比較之.(注意所得額一例若將4000元以上者包括在內,表示頻數分配之曲線將有一長尾

伸於右端).今更舉若干經濟材料之例於下:

圖41表示物價漲落之分配情形,根據米契爾(W.C.Mitchell)氏之研究,米氏集前後兩年間商品批發物價變動之例凡5578個.如1912年紐約中等內地棉每磅之平均價格為$0.115,1913年平均價格為$0.128;即漲價11.3%.此數即成為漲價表中之一項,歸入"10—119%"一組.全表包含與此同樣之項凡5578個.此材料繪於圖41,成多邊形,曲線未經修勻.

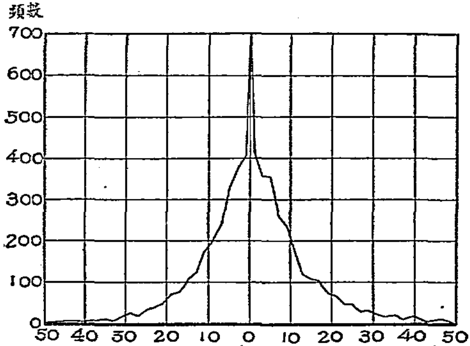


圖41——頻數多邊形:5,540個商品批發價格變動(前後兩年間之變動)之分配.

(註)錄自美國勞工統計局公報284號"The Making and Using Index Numbers".圖上表示之物價變動僅限於漲價51%與跌價51%之範圍以內者.尚有跌價55%之一例與漲價自52%—104%之間之三十七例未包括在內.

表二十所表示者,爲倫敦紐約間1882—1913年金鎊匯兌行情.此中不計時間因子,僅以各月行情分組,表示三十二年中各種行情出現之月數.圖42爲其圖示.

表二十

1882—1913年倫敦紐約間各月匯兌行情之分配

組距	月數
$4.8275—$4.8324	1
4.8325— 4.8374	6
4.8375— 4.8424	11
4.8425— 4.8474	21
4.8475— 4.8524	23
4.8525— 4.8574	24
4.8575— 4.8624	25
4.8625— 4.8674	40
4.8675— 4.8724	45
4.8725— 4.8774	49
4.8775— 4.8824	35
4.8325— 4.8874	45
4.8875— 4.8924	33
4.8925— 4.8974	16
4.8975— 4.9024	8
4.9025— 4.9074	1
4.9075— 4.9124	1
	384

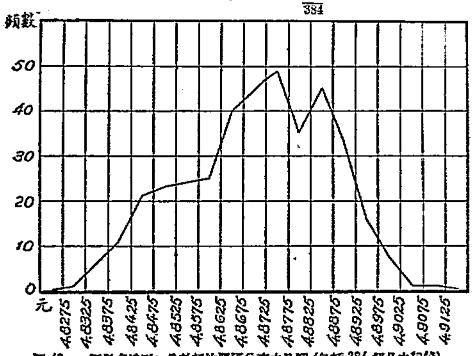


圖42——頻數多邊形：倫敦紐約間匯兌率之分配(包括384個月之紀錄)

可見頻數曲線與直方圖之根據經濟材料而作者,不必皆如根據自然現象之材料而作者之具有對稱而有規則之形式也.其中若干有偏左偏右之傾向;在某些例中,頻數增減之時有折裂存在,破壞其有規則之發展.惟雖有此種種差異,計量之來自經濟學、天文學、人體測量學、彈道學、及純偶然等種種不同領域者,彼此間仍有同族相似之點在.茲述其若干共同之特性如下:

[頻數分配之一般特性]

第一特性為計量所得之各值有變動 (variation): 如人體之高度各不同;天文上測量同一之量而得之結果有差異;發彈時儘量使條件不變,然子彈仍不能集於一點;各人之收入有多少;匯兌行情每星期每月皆有漲落.在某一情形中所得之各種不同之觀察結果或數值,沿尺度而分配於兩個極端數值之間.

考此等數值沿尺度(x軸)而分配之情形,設吾人由一個極端移向另一極端逐點考察之,則可見尺度上依次排列各點所得之例數(即依次排列各組之頻數),先逐漸增大至一最高點,然後減小,增減之情形皆多少帶有規則性.故數值雖有變動,吾人仍可得一集中趨勢 (central tendency), 即例數集中於某某數點也.此為可注意之第二個特性,為一切頻數分配所共有者.

若計量尺度上依次排列之各組與最集中點間互相差

離 (deviation) 之量，可見小差離出現之次數較之大差離遙爲衆多；而極端之差離出現之次數則極少；集中點左右兩方之差離，在自然科學與純偶然性之領域中所得之各例，達於完全（或幾乎完全）相等之狀態；在經濟材料之分配中亦近於相等.（最集中點兩端差離不近於相等之例外亦非少見，如所得額分配之例乃頗爲顯著者也）.

圖 43 之曲線名爲"機率曲線 (probability curve)" 或"差誤正態曲線 (normal curve of error)，"其特性將於後文詳細討論之.此處僅舉之爲一基本形式.前述諸例中有若干與之甚爲接近，而其他則對之多少有顯著之差離.應請特別注意者，與此種形式相異者亦多而重要，惟此差誤正態曲線因其爲基

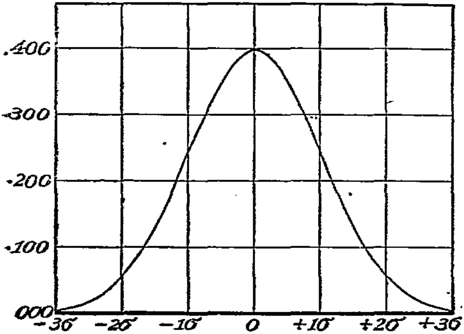


圖 43——差誤正態曲線.

本形式之故,在統計工作上有極端之重要性耳.蓋分配形式中即與之有極大差離者,仍與之有充分相似之點,足以證明用一普遍之方法以敍述頻數分配爲正當.數量材料之分配各各不同,而彼此間之差異與對於某種標準形式之差異,皆有極大之意義,惟雖有差異,仍有同族相似之性貫通所有各例.每一新的頻數分配並非一孤立之現象,乃大家族之一員,欲與以敍述及分析,凡方法之適用於其他各例者,未始不可爲解決此新問題之助也.

已知有此大致共同之形式,則吾人將如何敍述一頻數分配,並使與其他頻數分配相別乎?由前面之討論中,已可見若干方法.

[敍述頻數分配之方法]

前曾指出,觀察結果所得一切數值,乃沿尺度分佈者.欲敍述其頻數分配,吾人可於尺度上選擇一能徹底代表整個分配之數值.諸數值之頻數既各各不同,最中選者自爲出現次數最多之某數值,換言之,即尺度上集中程度最高之一點.此數值即該分配之集中趨勢量 (measure of central tendency).如所得額分配即可取其中包含人數最多之一組之中點(在表十一上爲950圓)作爲全分配之代表.須注意者,一個頻數分配可有幾種集中趨勢量,皆名爲平均數(averages),此最普通之數值僅爲其中之一耳.

如此之一個單獨的代表數值,因有甚多用途,惟獨立應

用時,顯然尙有關於分配之多種事實未能表現.平均數左右之分配之性質,乃有重大意義者.列於表內之一切事例,其數值密集一處乎?抑其散佈之範圍極廣乎?任何平均數?其代表之資格,須視其他諸數值對之集中之程度而定,卽須視集中趨勢點周圍之集中程度而定.故敍述一分配時,必須以離中趨勢量(measure of variation)補充平均數之不足.此計量乃表示諸數值在集中值周圍"散佈"之程度者.

精密之敍述並當述及分配之對稱程度.在最大集中點兩方例數之分配是否相等?頻數曲線是否如前舉之所得額分配之例而偏於一方?此乃非常重要之問題.若曲線不爲對稱之形式,則宜決定其不對稱之程度,計量此種程度者爲偏態量(measure of skewness).

最後,更可以差誤正態曲線爲標準,與頻數曲線相比較而定其尖峻之程度.若將圖41上代表物價漲落之頻數多邊形修勻,所得曲線顯然將比較正態曲線尖峻甚多.此種集中值處諸例密集之事實非常重要.頻數曲線之此種特性名爲峯態(kurtosis),而峯態之計量爲敍述頻數分配之最後之一步.

已得上述之幾個計量後,統計分析之工作乃可順利進行.吾人由一堆混亂之材料出發,作成頻數表而使材料變爲可以加工之形式,更將表中所顯示之基本事實進而鑄成三四個意義重要之計量,此種過程,不僅顯露該分配之特性且

使其便於與其他類似之分配相比較.例如有美國數千萬人之個人所得額數字,未經組織,欲與英國之同樣之材料相比較,此乃不可能之事.惟吾人若能就每國求一平均數或最足以代表之所得額,並描寫個人所得額在此集中值周圍之分配,吾人卽有一合理之根據以作比較研究.當整理與分析大量材料時,不問研究之目的爲何,吾人必須儘量利用敍述類數分配之各種計量,以盡綜合,化簡與比較之能事.

下節討論此敍述方法之一,卽關於集中趨勢之計量者.在討論過平均數問題以後,再處理與離中趨勢計量與編態計量有關之問題.

[平均數]

吾人知所以一平均數——單獨之標本數字——可代表一類數分配者,因多數之數字有集中於一集中值周圍之傾向,而一切觀察事例遠離此集中值之狀態,多少具有光滑與規則之性也.全因事例之集中於尺度上集中點周圍,此種代表性數字始有意義.故平均數之所以代表整個分配,僅因其爲一標本的數值也.若分配中各項之數值大相懸殊而無集中趨勢,卽無從求一單獨之數值以代表之.如下列三個數目:3, 125; 1000, 其算術中數 (arithmetic mean) 爲 376, 但 376 之一數並不能爲其所根據之三數值之代表.平均數欲眞能有代表之意義,其不得不滿足之基本要件,卽在集中值之周圍必須有集中趨勢存在是也.

若回思頻數分配之一般的性質,則對於下述之一種平均數不難瞭解其理.前已指出,在x尺度上集中程度最高之一點,卽出現之次數最多之一個數值,吾人可取以作爲整個分配之標本.此數值名爲衆數(mode),其所在之一組名爲衆數組(modal group).若將一分配繪成頻數曲線,其衆數卽x值之與最大縱坐標相當者.(嚴格言之,一分配之衆數,爲x值之與配合於該分配之理想的頻數曲線上之最大縱坐標相當者).至於最大縱坐標自身,乃計量衆數組之頻數者也.學者於決定衆數時每將此兩數相混.故吾人須牢記代表衆數值之距離,非在y軸,而在x軸.而縱坐標所代表者,爲屬於各組之例數;而非屬於各組之例之數值也.

吾人亦可選x尺度上平分總例數之一點,作爲一分配之標本.此點之數值多爲中位數(median),此值較一半事例之數值爲大,較另一半事例之數值爲小.如1918年美國個人所得額之中位數,爲1140元.蓋三千七百萬人中,有半數之每年收入不足此額,另有半數則超過此額也.若以頻數曲線表示一分配,由x軸上與中位數之值相當之一點作縱坐標線,則此線必將曲線下所包括之面積平分爲二.何也,因中位數爲平分總例數各半之值,而頻數曲線下之全面積乃代表分配中所包括之總例數者也.

算術中數爲第三種平均數,亦所以代表分配者.此種平均數由計算之法求得,故受分配中之每一項數值之影響,此

點顯然與衆數及中位數不同,蓋後兩者完全決定於各項之例在頻數表中之相對的位置,而不受個別各項之數值所影響者也.算術中數爲一分配之重心,若將頻數曲線杜塞而作成立體形式,算術中數將爲此曲線之平衡點之x值.

幾何中數 (geometric mean) 與倒數中數 (harmonic mean) 另爲兩種平均數,至其特質待後文討論之.

計算此種種平均數或確定其位置時,若事例極多,手續可極爲麻煩.惟此種計算之類,可用適當之方法減輕之.爲便於說明此種方法起見,可用下列之符號.

M :算術中數.

Mo:衆數.

Md:中位數.

m:代表一個觀察之數值;在頻數分配中爲組中點之值.

f:頻數分配之一組中所含例數.

N:某一數列或頻數分配所含之總例數.

Σ(sigma):總和符號.

[算術中數之計算]

用上述符號,可列算術中數之公式如下:

$$M=\frac{\Sigma m}{N}.$$

如2,5,6,7諸數之中數,等於諸數之和被除於4所得之商,卽$\frac{20}{4}$

或5是也.故算術中數之計算,當各數代表實在值時,僅爲一簡單之加法與除法.前曾有一表記述210個工人一週所得之工資.若得該表諸數相加而除以210,即得一週工資之中數爲26.920元.此地將210項相加之事已覺厭倦,若欲求三千七百萬個個人所得額之中數,應用加法幾不可能.因此,由於實際上的原因,計算平均數時常不得不根據頻數分配而不根據未分組之材料.茲取美國北達各塔邦(North Dakota)小麥產量之材料說明此法.美國農業經濟局記載1911—1921年北達各塔邦53縣中每縣每畝小麥平均產量,綜括其數字成下表:

表二十一

1911—1921年北達各塔邦諸縣小麥產量算術中數之計算(註)

組距（每畝斛數）	中點 m	頻數 f	fm
0—1.9	1	3	3
2—3.9	3	26	78
4—5.9	5	78	390
6—7.9	7	107	749
8—9.9	9	113	1,017
10—11.9	11	65	715
12—13.9	13	40	520
14—15.9	15	22	330
16—17.9	17	45	765
18—19.9	19	41	779
20—21.9	21	21	441
22—23.9	23	8	184
		569	5,971

$$M=\frac{\Sigma(fm)}{N}=\frac{5971}{569}=10.49\text{斛}.$$

(註) 此時期由各縣平均產量之記載見北達各塔邦農業大學試驗站公報165期(1922)上"Cost of Production and Farm Organization on 126 Farms in North Dakota, 1921", 120—21.

前在討論分組方法之一節中,曾舉出若干注意之點,其重要性可觀此例而益明.當時曾指出當選定組距時,應使各例在組內平均分配之假定,不與事實大有出入.蓋每組內之各例成一平均分配時,組中點可為各例之代表;然不成平均分配之形式時,組中點即不能為其代表.由頻數分配表計算算術中數之法,乃假定各例在組內成一平均分配,假定組中點之值與組內例數相乘時其乘積大致等於組內個別各項相加之總和者,故求算術中數之公式為$M=\frac{\Sigma(fm)}{N}$.表二十一說明詳細之算法.

用此法求得之數值有時稱為加權算術中數 (weighted arithmetic mean),實際上吾人為求m行十二個數字之算術中數.但非直接由此十二個數目求一簡單之平均數,乃先按十二個組內例數之多寡為比例,以加權於十二個組中點值,然後再予平均者.正如求五人之所得額之中數,其中兩人之所得額為2000元,三人之所得額為3000元,求中數時當然非以2000元與3000元相加而以2除之者;2000元之數須以2加權,3000元之數須以3為權數加權,得總和為13000元而以5除之,其商始為所求之算術中數也.雖然,由頻數分配計算算術中數之法固取加權形式,然"加權平均數" (weighted averages)一詞尚有更嚴格的意義,後文將論及之;故此名詞不應一般地應用於由頻數分配計算而得之平均數也.

[計算算術中數之簡法]

由頻數分配表計算算術中數,比較直接由未分組材料計算時一般上已大爲便利.惟當包括之例數極大時,雖用上述之法尤覺煩勞.計算之手續固尙可大行減省也.

由算術中數之計算方法,可知一數列之各項與其中數間之差離 (deviation) 之代數和爲零.如以 $m_1, m_2, m_3 \cdots\cdots m_n$ 代表數列之各項,以 M 代表其算術中數,以 $d_1, d_2, d_3 \cdots\cdots d_n$ 代表個別各項與中數間之差離,於是

$$M=\frac{m_1+m_2+m_3+\cdots\cdots+m_n}{N} \tag{1}$$

$$NM=m_1+m_2+m_3+\cdots\cdots+m_n \tag{2}$$

當然 N 等於項數,故

$$(m_1-M)+(m_2-M)+(m_3-M)+\cdots\cdots+(m_n-M)=0 \tag{3}$$

但

$m_1-M=d_1,\ m_2-M=d_2, \cdots\cdots$ 等,故方程式 (3) 可寫作

$$\Sigma d=0.$$

旣然如此,吾人可任意取一數値作假定中數,計算數列中各項對於此假定中數之差離,而求得差離之代數和;由此和卽可確定假定中數與實在中數間之差數.蓋此差數卽各項對於假定中數之差離之中數也.若以 M' 代表假定中數,而 $c=M-M'$, $d'_1, d'_2, d'_3, \cdots\cdots d'_n$ 代表各項對於 M' 之差離(卽 $d'_1=m_1-M', d'_2=m_2-M', \cdots\cdots$) 則

$$d'_1=d_1+c,\ d'_2=d_2+c,\ d'_3=d_3+c, \cdots\cdots d'_n=d_n+c.$$

$$\Sigma d'=\Sigma d+Nc.$$

但

$$\Sigma d=0.$$

$$\therefore\quad \Sigma d'=Nc.$$

故

$$c=\frac{\Sigma d'}{N}.$$

因$M=M'+c$,故可由已知值M'與c求得實在算術中數之值.其法可以下列簡單之例說明之.

表二十二

算術中數之計算

(簡法,未分組材料)

m	f	d'
5	1	-15
15	1	-5
25	1	$+5$
35	1	$+15$
45	1	$+25$
	5	$+25$

$$M'=20$$

$$c=\frac{\Sigma d'}{N}=\frac{+25}{5}=5$$

$$M=M'+c=20+5=25$$

當以20為假定中數而計算差離時,若以根據實在中數而計算之差離為標準,則每一例皆有一常數之差誤.此差誤等於實在中數與假定中數之差.依假定中數而計算之差離,其代數和即為此常數差誤之N倍,蓋包括之事例每多一項,此差誤亦多重複一次也.以N除代數和即可定差誤之數量,而實在中數之值於是乎求得.

數量之取頻數分配形式者,計算中數之工作尚可更進一步減省,其法即以組距為單位而計量差離也.在最後與以

必要的校正時,實在中數與假定中數間之差可仍以原來單位表示.前曾舉以計算中數之小麥產量材料,復可利用以詳細說明此法.

表二十三

1911—1921年北達各塔州各縣小麥產量算術中數之計算

(簡法)

組距	中點	頻數	差離 (以組爲單位)	fd'	
(斛/畝)	m	f	d'	−	+
0— 1.9	1	3	−4	12	……
2— 3.9	3	26	−3	78	……
4— 5.9	5	78	−2	156	……
6— 7.9	7	107	−1	107	……
8— 9.9	9	113	0	……	……
10—11.9	11	65	1	……	65
12—13.9	13	40	2	……	80
14—15.9	15	22	3	……	66
16—17.9	17	45	4	……	180
18—19.9	19	41	5	……	205
20—21.9	21	21	6	……	126
22—23.9	23	8	7	……	56
		569		−353	+778

算　法

$M'=9$

依 M' 算得差離之代數和：

$$\begin{array}{r} +778 \\ -353 \\ \hline -425 \end{array}$$

求 c,(以組距爲單位)：

$$c=\frac{425}{569}=.7469$$

代 c 爲原來單位：

組距=2

c(以原來單位)$=.7469\times 2$
$=1.4938$

決定 M 之值：

$M=M'+c$

$M=9+1.4938=10.4938$ 斛.

計算之術中數之簡法其步驟可以綜括之如下：

1. 組織材料成頻數分配形式.

2. 取附近分配中心之一組之中點,爲假定中數.

3. 以組距爲單位,將每組各行對於假定中數之差離,列爲一行.假定中數所在之一組,其中各項之差離爲 0,較低之一組,其各項之差離爲 −1,較高之一組,其各項之差離爲 +1,餘按此類推.

4. 將每組之頻數與該組之差離相乘,並記其符號,將乘得之積列爲 fd' 一行.

5. 求 fd' 行中諸項之代數和.

6. 將頻數總和 (N) 除代數和,得商爲校正數 (c),此 (c) 乃以組距爲單位者.

7. 以組距之值乘此校正數 (c),得積爲校正數之以原來單位表示者.

8. 將此校正數加於假定中數 (M') 之上,所得之和即實在中數 (M).

[中位數位置之決定:未分組材料]

中位數爲 x 變數之如此之一數值,即當事例按值之大小排列時,全部中有50% 在其下另有50% 在其上者也.對於許多頻數分配,此數爲有用而重要之數值.

當處理之材料未經分組,未成頻數分配之形式時,中位數位置之決定爲簡易之事.只須將材料按大小爲序排列,然後從數列之一端數起,數至平分總例數爲兩個相等部分之一點,此即中位數所在也.茲作一簡單之例以說明之:假定[750 元, 975 元, 1,128 元, 1,450 元, 1,475 元, 1,825 元, 1,950 元] 七數爲七人在一年中各得收入,數值之尺度自 750 元至 1950 元,其上列有七數.取 1000 元之一數而觀之,其一端有二項,另一端有五項,顯然不合中位數之定義.惟 1450 元一值,兩端各有三項,與七人中一人之收入額相符,乃此地之中位數也.若吾

人更假定將中心一項分割爲二,則此點之兩端各得$3\frac{1}{2}$項

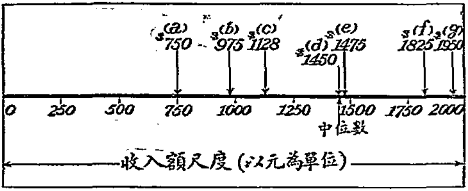


圖44——中位數位置之決定:未分組材料之處理(七個個人收入額).

情形如圖44所示.此圖明白表示中位數爲尺度上如此之一點,即其位置將頻數平分爲二者也.

若材料所包含之例數爲偶數,則問題稍有不同.下表之材料爲1921年美國22邦中每邦木廠鋸匠之平均工資,茲取以爲例.

表二十四

1921年各邦鋸匠平均工資

邦名	每人一小時平均工資
維基尼阿	$ 0.586
本雪爾文尼亞	.621
吞奈西	.627
北加羅林那	.647
西維基尼阿	.658
梅因	686
南加羅林那	.721
威士康新	.729
密支干	.730
阿拉巴馬	.749

(續前)

明尼索達	.755
塔克薩斯	.761
喬治亞	.779
亞堪薩斯	.785
密西西比	.798
路易西阿那	.824
南卡羅利達	.825
伊達阿	.833
加利福尼亞	.864
蒙塔那	.904
華盛頓	1.045
俄勒岡	1.106

(錄自 Momthly Labour Review, Tan. 1923, 9.)

此處之中位數,必須兩端各有11邦之一數值,故任何數值,凡介於$.755(明尼索達邦平均工資)與$.761(塔克薩斯邦平均工資)之間者,皆能合於中位數之定義.在此種情形下,實在之中位數爲一不定值.習慣上常取位於上下兩限之中點之數值爲中位數,在本例即爲$.758.

此中位數不與任何一邦之工資額相當.例數爲偶數時之情形常爲如是.若處理之材料爲非連續數列,中位數將爲不合於材料性質之一個虛數值(unreal value).如1900—1922年紐約市上60—90日期票貼現率之中位數,其值爲4.865%,此率乃從不出現於市上者,純爲虛擬的性質也.

[中位數位置之決定:已分類材料]

當材料已成頻數分配形式時,決定中位數位置之工作,實質上未有變更.惟由於分組之結果,個別各項之眞正數值

已不可復識,致問題稍形複雜耳.下列材料代表連續的276個月間紐約市上60—90日期票平均貼現率,可用以說明此法.

表二十五

中位數位置之決定,貼現率,紐約

組距 (貼現率,%)	中點 m	頻數 f
2.75—3.24	3.00	8
3.25—3.74	3.50	21
3.75—4.24	4.00	52
4.25—4.74	4.50	38
4.75—5.24	5.00	39
5.25—5.74	5.50	42
5.75—6.24	6.00	35
6.25—6.74	6.50	13
6.75—7.24	7.00	5
7.25—7.74	7.50	4
7.75—8.24	8.00	9
		276

$$\frac{N}{2}=\frac{276}{2}=138$$

$$Md=4.75+(9/39\times.50)$$
$$=4.75+.115$$
$$=4.865$$

此處中位數位置之決定,即在乎決定一數值其兩端各能得138項者.假定由尺度之下端算起,逐組向上,算至第一組(包含2.75—3.25之各項)之上限時,已越過8例,上端尙餘268例;算至第二組上限時,越過39例;至第四組上限時,越過129例;至第五組上限時,越過168例;故知兩端各得136項之數,必在第五組上下限之間某處.第五組爲4.75—5.25,吾人所求之點應在此距離中何處乎?

前曾指出在計算上有一假定,即任何一組所包括之各項事例,在該組組距內作平均之分配是也.未算至第五組前,

在第四組上限處,吾人已超過129例,故只須從第五組中再取9例作爲補充,即可成所求之138例之數.根據平均分配之假定,此9例當在尺度上相當於組距之9/39之距離以內.組距爲.50;.50之9/39等於.115.當自尺度下端向上計算時,已至4.75之後,更須向上移動相當於.115之距離.在尺度上具有4.865之數值之一點,即兩端各得138例項之分界線.此數值即中位數.

頻數表右方之算式,即表示計算之方法.茲將決定中位數位置之步驟摘述如下:

1. 將材料作成頻數分配.

2. 以2除材料所包含之總例數;得數爲中位數所在點兩端各項具有之例數.

3. 由尺度之下端起始,按次將各組頻數加併,直至求得包含中位數值之一組之下限.

4. 決定在中位數所在組中尙須取幾例加於業已加併之分組頻數上始能得等於$N/2$之結果.

5. 將中位數所在組之例數除尙須追加之例數,分數爲追加各例在組距內應處之地點.

6. 以此分數乘組距.

7. 將(6)之乘積加於中位數所在組之下限,其和爲中位數之值.

最後之三個步驟,僅爲補插法之簡單形式.

此法亦可倒轉,由尺度之上端起始而向下計算,如是則最後之一步,爲由中位數所在組之上限減去(6)之乘積.

N若爲奇數,$N/2$將包含一分數.計算之法與上述者完全相同.

[四分位數與十分位數]

爲許多目的,吾人常欲在形成頻數分配之各項所排列之數值尺度上,決定數點之位置,此數點乃以其他形式平分總例數爲幾個相等之部分者.中位數平分總數爲二,四分位數 (quartiles), 十分位數 (deciles) 及百分位數 (percentiles) 與之相似;在尺度上平分事例總數爲四等分之各點,其數值爲四分位數;平分事例總數爲十等分之各點,其數值爲十分位數;平分事例總數爲一百等分之各點,其數值爲百分位數如第一四分位數爲尺度上之一點,有四分之一之事例位於其下,其餘四分之三之事例則位於其下.第二四分位數,其值與中位數相等.第三十分位數爲尺度上一點,有十分之三之事例位於其下,另外十分之七之事例則位於其上.計算一律由尺度之下端起始.

例:　決定貼現率之第一四分位數之位置.

(參觀表二十五)

$$N/4=69$$

$$Q=3.75+(30/52\times .50)$$

$$=4.038$$

例：決定貼現率之第四十分位數之位置.

(參觀表二十五)

$N/10=27.6 \qquad D_4=4.25+19.4/38\times.50$

$4N/10=110.4 \qquad =4.505$

用圖決定中位數,四分位數,十分位數及百分位數之法,待後文再加以說明.

[衆數位置之決定]

衆數爲與頻數曲線上最大縱坐標相當之x變數值.此觀念極易了解,如最普通之工資,最普通之收入,最普通之高度等等皆是也.衆數所在之點,集中趨勢最強.費希奈爾(Fechner)氏稱之曰"Dichtester Wert"即最密集值,誠予此平均數之特質以最佳之表現.然欲確定一個材料之實在衆數値之位置,良非易事.一般統計工作中僅得其近似之值,然對於大部分之用途,此近似已相當精確,可以應付矣.(較精確之法詳見第十五章附註).

茲可取下列之分配,以說明求近似衆數值之法.

表二十六

5%債券之頻數分配

(根據紐約股票交易所1923年10月31日鐵路及工業5%債券之行情)

行情組距	中點 m	頻數 f
不足71.5		30
71.5—73.49	72.5	2
73.5—75.49	74.5	5
75.5—77.49	76.5	5

(續前)

77.5—79.49	78.5	5
79.5—81.49	80.5	12
81.5—83.49	82.5	13
83.5—85.49	84.5	15
85.5—87.49	86.5	11
87.5—89.49	88.5	17
89.5—91.49	90.5	17
91.5—93.49	92.5	33
93.5—95.49	94.5	45
95.5—97.49	96.5	51
97.5—99.49	98.5	59
99.5—101.49	100.5	23
101.5—103.49	102.5	9
103.5—105.49	104.5	1
		353

不足71.5之30例分佈甚廣,此表有此"空端"之組存在,卽無從計算其中數之數值.故此處以應用衆數爲平均數較爲適當.

組限爲97.5—99.49之一組包含例數最多.此組似爲衆數組,其組中點98.5可權且作爲近似衆數之數值.惟分組之法若不同,則所求得之衆數值亦各異,如將原來之債券行情依不同組距列表,可得下列之結果.(此地不必刊列全表,故僅列中心數組之頻數).

(a) 組距 (單位=1)	頻數 f	(b) 組距 (單位=2)	頻數 f	(c) 組距 (單位=3	頻數
93.5—94.49	17	94.5—96.49	54	91.5—94.49	50
94.5—95.49	28	96.5—98.49	50	94.5—97.49	79
95.5—96.49	26	98.5—100.49	52	97.5—100.49	77
96.5—97.49	25				
97.5—98.49	25				
98.5—99.49	34				
99.5—100.49	18				
100.5—101.49	5				

以1爲組距,所得之衆數爲99;以2爲組距,但組限與表二十六中所用者不同,所得之衆數爲95.5;以3爲組距,所得之衆數爲96.若再改變其分組方法,更可得其他諸值.故衆數似爲一奇怪而不可捉摸之平均數,對於同一材料,其數值似隨組距之大小及組限之位置而俱變.

困難之根本原因,在吾人所研究之材料爲有限.實在的衆數,爲材料無限大時出現次數最多之數值;若吾人能無限增加事例之總數,自能得其確切之位置.蓋既有充分之事例,縮小組距即可使衆數之近似值接近其實在值.組之範圍大,所被抹殺之細點愈多;若組距縮小,則能保留之細點越多,而實際的分配狀態之反映亦越眞.但大部分實際工作既必須根據相當小量之材料,縮小組距增加組數之結果,將使分配上呈現缺裂及不規則狀態,失去對稱與秩序,以致無法確定集中趨勢最大之點究在何處.表示債券價格之不同諸表爲說明此點之佳例.

吾人可用數學方法求得一代表實在衆數之數值,而不須收集無數事例.從前已將修勻法簡略說明,修勻法之一種,爲根據某一頻數分配材料配給一適當形式之理想的頻數曲線.在理論上,如此修勻之結果,足以代表當無限的縮小組距兼增加例數時所可得之分配狀態.相當於此理想曲線上最大縱坐標之x變數數值,即爲實在的衆數.

爲應大部分實際之須要,有衆數之近似值已可適用,而此種近似值大可用簡單之方法求得.第一種粗略之近似值可取頻數最大之一組之中點代表之,此法已經前述.吾人若能遵守以前列舉之分組之總原則,此法就一般言之可不致發生重大錯誤.

若分配之狀態尙爲規則,則在衆數所在組用插補法,可得一近似值較之僅以此組中點爲代表者更爲眞切.參觀表二十六中債劵價格表之編製,可見衆數所在組兩端之分配並非對稱.衆數所在組中點之值爲98.5.其較低之一組中點爲96.5者,包含51例,其較高之一組中點爲100.5者,包含23例.而此種不平均狀態在相連之較高較低諸組中繼續存在,較低之諸組所含例數多於較高諸組.吾人在他處皆假定在每組之上下限間例數成爲平均的分配,惟此處之衆數所在組內,其實情或非如此.由組外之分配情形觀之,集中趨勢最大之點似應在組距之下半,即介於97.5與98.5之間,因此衆數所在之一點,與其謂恰合於組中點,毋寧謂其落於中點98.5

之下.吾人可用加權法決定其在組內之位置,假定引向尺度下端之力為51(其較低一組之例數),引向尺度上端之力為23(其較高一組之例數),則可應用下列符號,作公式表其情形:

l =衆數所在組之下限.

f_1=在數值上較低一組之頻數.

f_2=在數值上較高一組之頻數.

i =組距.

補插之公式為

$$Mo=L+\frac{f_2}{f_2+f_1}\times i.$$

將此公式用於表二十六之債券價格材料,得

$$Mo=97.5+\frac{23}{74}\times 2=97.5+.62=98.12.$$

若權數(f_1及f_2)為較低之兩組或三組之頻數總和與較高之兩組或三組之頻數總和,用此公式求得之近似衆數值,有時可更為精確.在本例,若在衆數所在組高低兩端各取三組之頻數總和為權數,則代表債券價格之衆數值為97.91.

在上面說明中,吾人假定一材料唯有一集中點,然有時幾個集中點同時存在,決定衆數位置之問題於是變為複雜.如以.25元為組距表示工資分配之表九,有兩個確定之衆數點.此種形式之分配名為雙峯(bi-modal);因其繪成曲線時有兩峯也.若材料為同質(homogereous),雙峯為材料短少及分組不適當之結果.或由於組距對於材料所包含之例數為過

小.遇此種情形可將組限移動並增大組距,直至確立一衆數組後,衆數之近似值即可決定.此法與例數極多時決定實在衆數之位置之方法恰恰相反,蓋後者之法,爲使組距縮至極小也.惟若例數有限而欲求集中趨勢最高之點,則不得不增大組距,以消除因材料短少而生之不規則性.

若組距與組限不論如何改變,而分配仍爲雙峯,則此材料或非同質,恐有兩個不同之分配誤合爲一.此種情形在生物統計中非爲少見.從此可知原認爲同種之動物中實有兩個不同之種存在.材料若非同質,頻數分配即失去其全部之意義.此理不論於經濟統計之場合,或其他任何場合,皆爲適用者.

[根據算術中數及中位數決定衆數之法]

在某種情形之下,可以應用另一種方法求近似衆數之值,此法係根據算術中數,中位數及衆數三者之關係而來者.在完全對稱之分配中,中數,中位數與衆數三者相同.分配不成對稱,則三者分立.偏斜程度不大時三者間有一頗爲固定之關係,衆數與中數相去最遠,中位數位於從中數向衆數之距離上三分之一之處.但偏斜程度甚高時,此種關係即不復存在.故在偏斜程度尚微之分配上,若已知此三平均數中之兩種,即可求得第三種之近似值.然實際上此法僅用於衆數值之決定.因其餘兩種平均數皆有其他之方法可更精確的計算也.即在求衆數值時,亦僅在其他較精確之方法不適用

或不必用時,始應用此法.

根據上述之關係,可得公式如下:

$$Mo=M-3(M-Md).$$

將此公式應用於表十二之電桿材料,結果為

$$Mo=9.33-3(9.33-9.015)=8.385.$$

此數較衆數所在組之中點值8.5略小,較之用加權法(每端取四組之頻數總和加權)從衆數值求得之值8.49亦小.

此地應注意者,所有此等衆數之數值,其正確性皆屬虛擬的性質.以上討論之一切方法,關於衆數位置之決定者,所得結果不過為近似值;此點在解釋與應用其結果時宜切記勿忘.

[用圖推算衆數中位數四分位數及十分位數之法]

以上所述之幾種統計計量可用作圖法推算,對於此法略加討論,可使吾人對於頻數曲線及累積頻數曲線得一更深刻之理解.

平常之頻數曲線上,衆數值極易決定.蓋由定義觀之,衆數即在水平尺度上與此種曲線上最大縱坐標相當之一點.若就頻數多邊形求衆數值,則可得一大致之數值,即具有最大頻數之一組之中點數值也.用觀察或數學方法將多邊形修勻為一光滑之曲線後,求得之衆數值更與實在衆數相接近.圖45為根據觀察修勻之曲線,為表八工資材料之圖示,所以說明用作圖決定衆數之法.水平尺度上與此曲線之最大

縱坐標相當之一點,其值爲27.50元,此爲衆數之近似值.同一材料用加權法求得之衆數值爲27.69元,由中數及中位數求得者爲27.347元,可以相比.

曲線上並表示中數及中位數之位置,前曾指出當分配之不對稱程度尙微時,三種平均數間有一固定關係,中位數居其他兩者之中,約當由中數向衆數之距離上三分之一之處.此地頗可用此種關係代表,故可由修勻之曲線求衆數之近似值.惟原來材料之不規則性,使根據觀察以修勻曲線之法未免勉強耳.

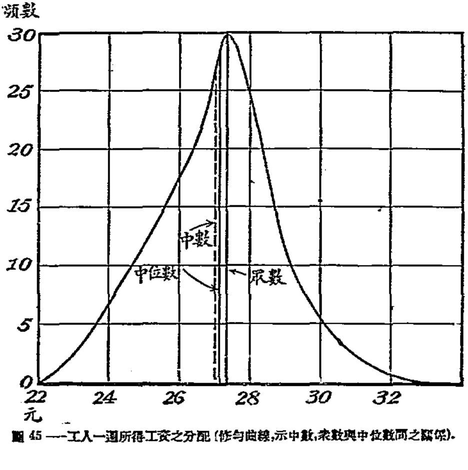


圖 45——工人一週所得工資之分配(修勻曲線,示中數,衆數與中位數間之關係).

玆用上述之工資材料排列成累積頻數表,如表二十七。更給圖46以表示之:

表二十七

某廠傭工之累積分配表

(根據每週工資分組)

每週工資	得此工資之傭工數(頻數)
不足$22.50	0
,, ,, 23.00	1
,, ,, 23.50	5
,, ,, 24.00	8
,, ,, 24.50	10
,, ,, 25.00	29
,, ,, 25.50	41
,, ,, 26.00	56
,, ,, 26.50	78
,, ,, 27.00	98
,, ,, 27.50	122
,, ,, 28.00	152
,, ,, 28.50	169
,, ,, 29.00	186
,, ,, 29.50	193
,, ,, 30.00	199
,, ,, 30.50	204
,, ,, 31.00	208
,, ,, 31.50	209
,, ,, 32.00	209
,, ,, 32.50	210

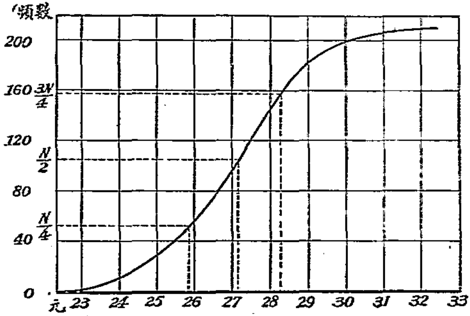


圖 46——工人一週所得工資之累積分配表（示作圖以決定中位數及四分位數之法）.

一段累積頻數曲線,其傾斜度之大小,視水平尺度上與之相當之一段距離內所增加之例數之多少而定.因此曲線初甚和緩,旣而峻急,終復平坦以達於最高之一端.故衆數之值,爲水平尺度上與傾斜度最大點(卽頻數之增加最速之點,亦卽頻數分配中集中程度最高之點)相當之一點.欲由一修勻累積頻數曲線而決定衆數近似值,可先決定傾斜度最大之點(轉向點),而讀 x 尺度上相當點之值.就現今之例言,用此法可求得衆值近似值爲27.50元左右之一數.

中位數,四分位數及十分位數亦可由累積頻數曲線用作圖法求其數值.此種曲線經修勻後,可使補插法頗爲圓滿.圖之規模若相當的大,此法可得出精確之數值.由垂直尺度

(累積頻數之尺度)上距底邊$\frac{N}{2}$之處之一點,作一水平線通至累積曲線,則其交點之橫坐標卽爲中位數之值.此值僅須由交點作一垂直線至x軸卽可決定.圖46卽表示此法之應用者.本例用此法求得之中位數,其值爲27.125元.而直接用補插法求得者爲27.1458元.用完全同樣之方法,亦可決定四分位數.卽將垂直尺度分爲四等分,更由各點作水平線以通至累積曲線,求與交點相當之x值是已.

在某些用途上,尤其在吾人所平均者非絕對數而寧爲率或比率時,前述之平均數無一種爲適當者.在此一方面,幾何中數與倒數中數爲特別合宜之平均數形式,故吾人宜熟識之.

[幾何中數]

幾何中數(geomtric mean 爲n個數目之連乘積之n次方根;故可以下式表示之:

$$Mg=\sqrt[n]{a_1\cdot a_2\cdot a_3\cdots\cdots a_n}$$

如2,4,8三數,其幾何中數爲

$$\begin{aligned}Mg&=\sqrt[3]{2\cdot4\cdot8}\\&=\sqrt[3]{64}\\&=4.\end{aligned}$$

根據計算之法,可顯見數列中有一數爲0,則該數列之幾何中數亦爲0.

幾何中數之實際計算,可因應用對數而大爲便捷式爲:

$$\log Mg=\frac{\log a_1+\log a_2+\log a_3+\cdots\cdots+\log a_n}{N}.$$

此式表示幾何中數之對數等於個別各數之對數之算術中數.

求若干數值之加權幾何中數時,權數各成爲其所加權之一項數值之指數.如權數有n個,分別爲$W_1, W_2, W_3\cdots\cdots W_n$,加權於$a_1, a_2, a_3\cdots\cdots a_n$諸項,則求幾何中數之公式爲:

$$Mg=\sqrt[N]{a_1^{w_1}\cdot a_2^{w_2}\cdot a_3^{w_3}\cdots\cdots a_n^{w_n}}.$$

此卽等於每項重複若干次,次數與權數之值相當.恰如求加權算術中數時也.用對數表示之,則加權幾何中數之公式爲

$$\log Mg=\frac{w_1\log a_1+w_2\log a_2+w_3\log a_3+\cdots\cdots+w_n\log a_n}{N}.$$

玆取下表之材料以說明計算幾何中數之法.表中所列者爲紐約股票交易所115種紅利優先股票之價格,爲1922年8月26日收盤時情形.

表二十八

幾何中數之計算

優先股票價格

組距	m	f	$\log m$	$f\log m$
\$35—\$44.9	\$40	1	1.602060	1.602060
45— 54.9	50	6	1.698970	10.193820
55— 64.9	60	8	1.778151	14.225208
65— 74.9	70	5	1.845098	9.225490
75— 84.9	80	14	1.903090	26.643260
85— 94.9	90	22	1.954243	42.993346
95—104.9	100	27	2.000000	54.000000
105—114.9	110	18	2.041393	36.745074
115—124.9	120	14	2.079181	29.108134
		115		224.736792

$$\log Mg=\frac{224.736792}{115},\quad \log Mg=1.954233,\quad Mg=\$90.00.$$

[幾何中數之特性]

幾何中數之本質,可就其以平均數的資格對於其所代表之各項數值之關係而研究之.

若將算術中數代入其所代表之數列之各項,則所得之總和不變.如2,4,8之和爲14,其算術中數爲$4\frac{2}{3}$;將中數代入以前之三項,總和依然爲14.若將幾何中數代入其所代表之數列之各項,則各項之連乘積不變,如2,4,8連乘其積爲64,三項之幾何中數爲4,以4代入各項,連乘積依舊爲64,此爲幾何中數之特性.

此外,就算術中數言,大於該數之各項之差離,其總和若不計正負號,與小於該數之各項之差離總和相等,而就幾何中數言,則其與各項之比之乘積兩端相等.即幾何中數對較小諸項之比的連乘積,等於較大諸項對幾何中數之比的連乘積.例如3,6,8,9之幾何中數爲6,方程式爲:

$$\frac{6}{3}\times\frac{6}{6}=\frac{8}{6}\times\frac{9}{6}.$$

此例所表示者,爲幾何中數之最重要特性,幾何中數實比率之平均數,其主要之應用,在經濟統計之範圍內,爲物價指數之一方面.蓋此中最重要之點爲物價變動之比率也,物價從50而變爲100之變動,其重要程度,實與物價從100而變爲200之變動相等.然此點在算術中數即無從表示.蓋後一變動所表示之絕對值之差爲100,爲前一變動所表現者50之兩

倍.故由算術中數表現之,後者之重要性將兩倍於前者.茲更引一常見之例,如有兩個物價變動,其一爲由100增至1000,增加10倍;其一爲由100減至10,減至原價$\frac{1}{10}$.1000與10之算術中數爲505,其幾何中數則爲$\sqrt{1000\times10}$,即100.兩變動之比率相同,在幾何中數則表現彼此平衡;而算術中數505乃完全不適於表示此物價變動之平均比率者也.此點在論指數之一章中尙須詳加討論.

前曾指出,在某種用途上,對數圖之應用可得若干利益;此等利益於應用幾何中數時亦可得之.幾何中數有時稱爲對數中數,因其對數僅爲其所代表之各項數值之算術中數也.凡遇所欲平均者爲百分變動,或要點不在絕對數之差而在比率之差時,皆宜應用幾何中數.

按複利增殖之資金,欲求增殖之平均率,即一應用幾何中數之問題.若p_0代表時期開始時之原本,p_n爲時期末積得之本利和,r爲利率,n爲時期之年數,每年將利息轉入本金一次,則p_0在n年後增殖得之本利和可以下列方程式代表之.

$$p_n=p_0(1+r)^n.$$

由此

$$r=\sqrt[n]{\frac{p_n}{p_0}}-1.$$

如以1000元依複利增殖十二年後之結果爲1600元,即增加60%.其算術中數爲5%,但5%並非實際之利率,實際之利率應爲:

$$r=\sqrt[12]{\frac{1600}{1000}}-1,$$

$$=\sqrt[12]{1.60}-1,$$

$$=1.04-1,$$

$$=.04\text{或}4\%.$$

凡欲求增加之平均比率或減少之平均率時,問題卽與此例完全相同,若用算術中數必致錯誤.

[作爲集中趨勢量之幾何中數]

何種形式之頻數分配,其集中趨勢最宜以幾何中數代表之乎?此問題曾經引起饒有興味之討論.當絕對值依算術尺度描繪時,若能得一相當對稱之分配,則用算術中數代表此種材料之集中趨勢,自勝於應用幾何中數.但絕對值依算術尺度描繪之結果,若不能得稱之頻數曲線,而描繪絕對數之對數時此種不對稱之狀態卽可除去,而得一對稱之曲線時,幾何中數似更爲適當.蓋在此種分配中,成爲對稱者非各值對集中值之絕對差離 (absolute deviation), 而爲其相對差離 (relative deviation), 或以比率表示之差離.而最能代表集中趨勢者,爲各值之對數的算術中數(前已指出,此值卽原來各值之幾何中數).如是描繪之曲線,其對稱中心將爲幾何中數之對數值.一頻數曲線之代表物價變動之百分比者,將以此等變動之幾何中數的對數爲中心,表示一對稱之狀態.此種百分變動,當其爲自然數時,呈現不對稱形式;其對於算術

中數之差離之範圍,在數值較高之一端者,大過於在數值較低之一端者(參考圖 49 與 50). 蓋商品之價格,可由一基數增加至 1,000% 或 1,000% 以上,然無有能跌落至 100% 以上者也. 下文討論指數之一節中,將有關於此專門問題之詳細討論.(註)

此種不根據各項之自然數而根據其對數所作之頻數曲線,曾用於表示所得額分配之材料,得良好之結果.所得額分配範圍極大,欲繪其自然數,必須以一圖而能表示曲線各部分之特性,此乃事實所不許.用雙對數紙作圖(即等於繪 x 值及 y 值之對數)可解決此困難,而予人以一全分配之真實印象,顯示其各部分間相互之關係,同時暴露若干爲自然尺圖所掩蔽之重要特相.特別因此法可使乘數以上之一段曲線修勻爲一直線,使首先應用此法於所得額材料之第一人 Vilfredo Pareto 氏提出其關於所得額分配之理論,即所謂巴

(註) 華爾施氏在其"估量問題"(C. M. Walsh, Problem ol Estimation, 35)一書中,曾舉應用平均數之準則如下:

(一) 數列中各項之值,吾人無從測度或指定其上限或下限者,宜用算術平均數.

(二) 若下限有一固定之值,或爲 0,或大於 0,而上限無從測度或指定者,宜用幾何平均數.物價變動之情形即爲如是,故華爾施氏認爲編製物價指數時應用幾何平均數爲正確.

(三) 當在實際經驗上或在事物之本質上,具有某種上限或下限之存在,而上列準則不能應用時,必須研究材料之實際分配狀態.此時衆數若與算術平均數較爲接近,宜用算術平均數;若與幾何平均數較爲接近,宜用幾何平均數.

爾多氏律(Pareto's law)者.在美國因因對於所得額分配之深刻研究,使國立經濟研究局之主事者對巴氏定律引申而得之若干結論提出異議.然用雙對數圖以表現所得額材料之價值,仍爲世所公認.國立經濟研究局出版之報告"美國之所得"(Income in united states)一書中載有關於此種形式之曲線之討論,甚有興味.

[倒數中數]

倒數中數爲平均數之一種,其應用之範圍甚有限.惟對於某種材料,吾人須用之以避錯誤.求時間比率(time rate)之平均數時,吾人必須用之;用以整理若干其他材料,亦有特別有利之點.下例說明應用此平均數之方法.

一汽車以每小時20哩之速率行4哩,復以每小時30哩之速率行4哩;試求其平均速度.兩種速率之算術中數爲每小時25哩,惟此爲錯誤之解答.事實上機器以每小時20哩之速率行12分鐘,以每小時30哩之速率行8分鐘,於是以20分鐘行8哩,平均速率應爲每小時24哩.此法等於用加權法求20與30之平均數,前者權數爲12,後者爲8.直接由兩比率求倒數中數可得同樣結果.一數列諸值之倒數中數,爲個別各值之倒數之算術中數之倒數.若吾人所欲平均之一列比率以$r_1, r_2, r_3 \cdots\cdots r_n$,代表,倒數中數($H$)之公式爲

$$\frac{1}{H}=\frac{\frac{1}{r_1}+\frac{1}{r_2}+\frac{1}{r_3}+\cdots\cdots+\frac{1}{r_n}}{N}$$

用適纔引過之數字,則

$$\frac{1}{H}=\frac{\frac{1}{20}+\frac{1}{30}}{2}$$

$$=\frac{5}{120}$$

$$H=24$$

如求 a, b 二數之倒數中數,可用較簡之式 $H=\frac{2\,ab}{a+b}$; 求 a, b, c 三數之倒數中數,可用公式 $H=\frac{3\,abc}{ab+bc+ac}$. 此兩公式皆與第一公式相等.應用倒數表可使計算一列數值之倒數中數之工作大爲便捷.

玆從價格方面舉一例以說明倒數中數用於處理經濟材料之情形.若某種商品其開價爲"每元若干",則就幾個如此之價目而求其平均數時,若用簡單之算術中數,其結果可致錯誤.如已知之價目爲"每元 4 個","每元 5 個"與"每元 20 個",求每元平均可購個數(即平均價格).根據已知之數目(4, 5 與 20)求得之算術中數爲 $9\frac{2}{3}$,表面上似爲每元平均可購個數,平均價格爲每個 10.34 分.但原來價目實等於每個 25 分, 20 分,及 5 分.此幾個價格之算術中數爲 $16\frac{2}{3}$ 分,以原來之形式表示之,即平均每元可購 6 個.假定吾人欲平均加權於每一價目,此平均價格即爲正確之數.當價目之形式爲"每元若干"時,其算術中數爲一加權平均數,某價目所包括

之商品單位越多,其權數亦因之而越大.惟用倒數中數時,正確之結果可直接由原來價目求得. 4, 5,與20三數之倒數中數爲 6.故 6 爲每元平均可得之個數.

[各種平均數之關係]

一列數值之各種平均數計算決定之後,可見其彼此之間常有某種關係存在.

1. 算術中數,中位數與衆數三者,在對稱之分配中相合爲一.

2. 稍微不對稱之分配,其中位數位於中數與衆數之間,約在自前者至後者之距離上三分之一之處.因此在此種形式之分配中,其關係略如下列公式.

$$Mo=M-3(M-Md)$$

3. 任何一列數值之算術中數,大於幾何中數.

4. 任何一列數值之幾何中數,大於倒數中數.此 3, 4 兩點唯一之例外,爲各項數值各各相等之時.此時算術中數,幾何中數與倒數中數皆相等.

5. 任何二項之幾何中數,等於此二項之倒數中數與算術中數之幾何中數.如二項數值爲 2 與 8, 其倒數中數爲 $3\frac{1}{5}$, 幾何中數爲 4, 算術中數爲 5. 但 4 亦爲 $3\frac{1}{5}$ 與 5 之幾何中數.若項數超過二項,此關係卽不復存在.

6. 若材料之離中趨勢合於算術定律,衆數與中位數常較接近於算術平均數而較遠於幾何平均數.若材料之

離中趨勢合於幾何定律，衆數與中位數常較接近於幾何平均數而較遠於算術平均數.

[主要平均數之特色]

算術中數

1. 算術中數之值，數列中各項之值皆影響之.在某種用途上，此平均數受差離極大諸項之影響過甚.

2. 算術中數計算甚易，無論何處皆有一定之值.

3. 算術中數爲計算而得之平均數，故可用代數方法處理.

中位數

1. 中位數之值，不受對此平均數差離極大諸項之值所影響.

2. 當數列中各項不便以量計時，中位數之位置仍可決定.

3. 中位數之位置，當材料不完全時亦可決定；吾人僅須知全分配共有若干例，其一般的位置如何，並對於附近分配中心之各項有確切之知識.

4. 中位數對於用代數方法處理上，不及算術中數，幾何中數及倒數中數之爲適宜.

衆數

1. 衆數之值，不受對於此平均數差離極大各項之值所影響.

2. 衆數之近似值易於確定,惟求其實在值須經繁複計算.

3. 分配所包括之例數若非甚多且有顯著之集中趨勢,則衆數毫無意義.

4. 衆數位於集中趨勢最大之點,故此平均數最足爲分配之標本.

5. 衆數不能用代數方法處理.

幾何中數

1. 幾何中數所受差離極大各項之影響,不及算術中數之大.

2. 無論何處,幾何中數皆絕對確定,

3. 求變動之率,或量與量間之比之平均數時,變動之比率相同則重要性相等,故宜用此幾何式平均數.此種平均數特別適宜於表示價格變動之平均比率.

4. 幾何中數可用代數方法處理.

倒數中數

1. 倒數中數適於平均時間比率與某種類似之項.在經濟統計之領域內用於價格材料之處理.

2. 倒數中數計算時費力,且不爲人所熟知,致其在日常統計分析中,其效用未免減色.

3. 倒數中數可用代數方法處理.

由前列摘要,可見每種形式之平均數各有其特別適用

之領域;對於某種用途與條件,各爲一切平均數中之最佳者.吾人須了解每一種形式之特性與其所受之限制,始能應用適當.欲完全的敍述一頻數分配,除其他統計計量外,常須有兩種或三種主要平均數.單獨用一種平均數時,最有用者或爲算術中數,因其計算簡單,且可用代數方法處理,而其意義又絕對確定,爲人人所共知,故在統計工作中效用甚著,但其可以應用之範圍,只能限於一定條件之下,現今對於幾何中數之特殊優點明瞭愈多,故在各種統計工作上應用之範圍亦擴大.平均數之審慎的應用,乃正確之統計分析之基本要件也.

參考書

Bowley. A. L. Elements of Statistics (82—109, 138—139)

Chaddock, R. E. Principles and Methods of Statistics (Chaps. VI, VII, VIII.)

Jones, D. C. A first Course in Statistics (22—41)

Kelley, Truman L. Statisticae Method (121—140)

Pearl, Raymond. Medical Biometry and Statistics (264—272)

Rugg, H. O. Statisticae Methods Applied to Education (97—148)

Secrist, Horace. Introduction to Statisticae Methods (234—292)

Walsh, C. M. The Problem of Estimation (1—67)

Yule G. U. An Introduction to the Theory of Statistics (106—132)

Zizek, Franz. Statisticae Averages (82—109)

第五章

頻數分配之敍述：離中趨勢量與偏態量

在前數章中,吾人首論整理材料之方法,研究如何將大批數量材料彙爲一種形式,使全部材料之特性易於決定;其次論及已經彙集之材料應如何予以敍述.頻數分配之作成,第一目的已經達到;當求得一平均數,作爲單獨之標本數值,以代表頻數分配之集中趨勢時,第二目的亦已部分告成.惟任何平均數單獨應用時,並不能予頻數分配以一完全之敍述.如欲將某一分配之主要特質盡述無遺,欲使其與其他分配能相比較,首先須有另外三種數值,第一離中趨勢量(measure of variation or dispersion),所計量者爲分配中各項離開集中值之程度,或可謂"分散"之程度.第二爲分配之對稱量或偏態量 (measure of symmetry or skewness),所計量者爲分配在集中值兩端之平衡與否.第三爲峯態量 (measure of kurtosis),所計量者爲衆數所在之處事例菌集之程度.茲僅就計量離中趨勢與偏態之法討論之.計量峯態之法待稍後再談.

[離中趨勢之本質與重要性]

數量材料之集合中有離中趨勢存在;此事吾人前已提及,並曾指出其與統計家之工作有何關係.就實際言之,任何數量材料之集合,不同其來自社會的,生物學的,或經濟的領域,莫不有離中趨勢,莫不以個別單位間量之差異爲其特徵.此離中之事實與同族相似之事實同樣重要.生物學上之"離中"(變異),爲進化過程中之基本因素.計量一種屬之生理之特性(如身長)時,若不計及該種屬中在此方面之平均的離中趨勢,此計量即非完全.一國之平均所得額,其意義或反不若社會各階級間所得額之差異(所得額之離中趨勢)之爲重要.價格變動擾亂經濟組織之正常活動,使一部分人受損失,另一部分人得意外之財,此乃由於價格制度之種種成分所受影響各有不同之故;引起擾亂者,非價格水平之一般的變動,乃各種價格變動間之差異也.

若不知某一頻數分配離中趨勢之程度若何,單獨一平均數意義甚小.若離中趨勢之大,竟使分配中無明顯之集中趨勢存在,則平均數將毫無意義.離中趨勢越小,平均數越有意義,因此,不論單獨敍述一頻數分配,或以之與其他頻數分配相比較,集中趨勢之計量必須以離中趨勢之計量爲補充.

[計量絕對離中趨勢之各種方法]

離中趨勢可用原來材料之計量單位表示,亦可用百分數等與原來單位無關之抽象數字表示之.當應用原來單位

時,所計量者爲絕對的離中趨勢(absolute variability);當用抽象數字時,所得者爲相對的離中趨勢(Relative vareability),後者較前種形式更適於比較.茲請先研究絕對的離中趨勢之計量.

[全距]

全距(Range)爲離中趨勢之一種粗率的計量,指分配中所包括之最小項與最大項之兩數值間之差.表三十爲1882—1913年倫敦紐約間各月匯兌率之分配,原表數字中最小項爲4.83元;最大項爲4.908元.故全距爲4.908元至4.83元,或.078元.尺度上相當於.078元之距離將全體各項統包在內.若不能得原始之材料,則可由頻數表求其近似值.此時全距爲最低組下限與最高組上限間之差,在現今所舉之例中應爲.085元.

全距之值,顯然祇繫於兩極端之事例,只須有一變態之項存在,其值即可大變,此游移不定之性質,與不足代表各項分配之實際狀態之危險,使其少見用統計工作之中.全距常用於計量股票市場行情之漲跌,然其是否適合甚可疑也.

[平均差]

計算各項對於集中值之離中趨勢時,一較精確之方法,爲計算各項對此集中值之差離之數而平均之.下列一簡單之例以明此計算之方法:($M. D.$ 爲平均差 mean deviation 之記號).

表二十九

平均差之計算

m	f	d
3	1	6
6	1	3
9	1	0
12	1	3
15	1	6
		18

$M=9$

$M.D.=\frac{18}{5}=3.6$

平均數爲 9.（此地算術中數與中位數相合爲一）將各項差離之數,不分正負而相加,所得之和復以項數除之.此法可記爲下式:

$$M.D.=\frac{\Sigma d}{N}$$

以平常之言語表之,所謂一列數值之平均差者,各數值對於一平均數值（中數或中位數）之差離之算術中數也.當相加與平均時,差離之正負號置而不計在實際計算上,差離之根據中數計算或中位數計算無甚分別.在理論上,應以中位數爲計算之標準,因以中位數爲基數時,所得平均差爲數最小也.

表三十說明從已經作成頻數分配之材料計算平均差之方法.在計算一頻數分配之各項之算術中數時,曾假定每一組內所包括之各項皆集中於該組之中點;換言之,該組中點之值,可視爲所包括各項每項之值.計算平均差時,吾人仍作此同一之假定.

表三十

平均差之計算

1882—1913年倫敦紐約間匯兌率

(此表根據每月初之行情而作)

組距	中點	類數	根據假定原點所得差離	
	m	f	d'	fd'
$4.8275—$4.8324	$4.830	1	$.040	$.040
4.8325—4.8374	4.835	6	.035	.210
4.8375—4.8424	4.840	11	.030	.330
4.8425—4.8474	4.845	21	.025	.525
4.8475—4.8524	4.850	23	.020	.460
4.8525—4.8574	4.855	24	.015	.360
4.8575—4.8624	4.860	25	.010	.250
4.8625—4.8674	4.865	40	.005	.200
4.8675—4.8724	4.870	45	......	......
4.8725—4.8774	4.875	49	.005	.245
4.8775—4.8824	4.880	35	.010	.350
4.8825—4.8874	4.885	45	.015	.675
4.8875—4.8924	4.890	33	.020	.660
4.8925—4.8974	4.895	16	.025	.400
4.8975—4.9024	4.900	8	.030	.240
4.9025—4.9074	4.905	1	.035	.035
4.9075—4.9124	4.910	1	.040	.040
		384		$5.020

假定原點=4.8700元(美金)
中位數=4.8721
相差= .0021元
196例每例差離過小.0021元
188例每例差離過大.0021元
上下對消剩8例過小.0021元
根據假定原點所求之差離之總和=5.020元
校正數(.0021元之8倍)=.0168元
根據中位數所得之差離之總和=5.020+.0168元
=5.0368元

$$M.D.=\frac{5.0368\text{元}}{384}=.01312\text{元}$$

用表二十九所示之方法,亦可由此表求得平均差.即先求每組中點對於中位數之差離,(此地中位數為4.8721元.第一組之一項,其差離為.0421元,第二組之六項,每項差離為.0371元)以之與該組項數相乘,然後以全分配之例項總數除

所得乘積之總和是也.惟實在的差離有小數存在,處理費力.若差離由一假定原點計算,求各例項對此假定中位數(或中數)之差離總和,更以一校正數校正其因根據假定原點不根據實在平均數而引起之差誤;則計算上可較爲便利.表三十所應用之法,即此簡便之法也.

由算術中數計算平均差時,亦可應用此法.而此兩種計算又皆可有更簡捷之法,即如前章某些計算之例,用組距爲單位以計量差離也.

表三十中,差離非由中位數之值 4.8721 元計算,而爲由 4.870 元計算,後者乃任意選定作爲原點者.對於此假定原點之差離總和,不計正負記號,爲 5.020 元.此和與對於中位數之差離總和相差幾何?以假定原點 4.870 元爲中點之組所含 45 例,每例之 d' 值(對於假定原點之差離)各爲 0, 但每例對於中位數之差離,爲 4.870 元與 4.8721 元兩數之差,即 .0021 元.故此地計算得之差離(0)過小 .0021 元.計算數值較低各組每項之差離時亦同犯此錯誤.總計對於假定原點之差離較對於中位數之差離小 .0021 元者共 196 例.但數值高於假定原點之各組中,各例所犯錯誤方向相反.如假定原點所在組之上一組,其中之 49 項,對於假定原點之差離各爲 .005 元,而對於中位數之差離則爲 .0029 元;故計算得之差離各爲過大 .0021 元.總計犯此方之錯誤者共 188 例.抵消結果,剩有 8 例過小 .0021 元.因此全體差離之總和 5.020 過小 8×.0021 元即

.0168 元.於是得對於中位數之差離之總和爲 5.0368 元,平均差爲.01312 元.

此法可以下列之公式總括之:

$$M.D.=\frac{\Sigma(fd')+(N_s-N_l)c}{N}$$

其中

N_s= 由假定原點計量而得之差離小於由中位數(或中數)計量而得之差離之例數.

N_l= 由假定原點計量而得之差離大於由中位數(或中數)計量而得之差離之例數.

c = 假定原點與中位數(或中數)間之差.

欲應用此公式,假定原點與中位數(或中數)必須在同一組距內.(註)

計算平均差之法,可分下列各步驟:

1. 決定中位數之值.

2. 以中位數所在組之中點作爲假定原點,求各組中每項對此原點之差離,與各組之頻數相乘,於是置乘積之

(註) 上述之計算方法,假定在中位數所在組中之若干例項爲集中於該組之中點者.若假定此等例項在該組距內成一平均之分配,因而於計算方法上改以修改以適合此新假定,則所得結果之精確性可略有所增益.(參考H. L. Riety 氏 Hand book of mathematical Statistics, 29—31.關於此法之說明)

正負號不計,相加而求其總和.

3. 決定對於假定原點之差離比較對於中位數之差離爲大之例數,並決定對於假定原點之差離比較對於中位數之差離爲小之例數,然後求此兩數之差,更將此差乘中位數與假定原點間之差,其乘積爲對於假定原點諸差離之總和與對於中位數諸差離之總和間之差,於是以此差校正對於假定原點諸差離之總和.

4. 將例數除校正總和,其商卽對於中位數之平均差.(求對於中數之平均差時其法相同.)

[標準差]

計算平均差時,曾置正負號而不計,故在代數意義上言,此法爲背理者.計算標準差 (Standard deviation) 時,此種錯誤可以避免,而獲得一更有確切數學意義之計量.標準差平常用小寫之希臘字母 σ (Sigma) 代表之.

計算標準差之方法,首先將各項對於算術中數之差離作成平方;求平方差離之總和而以項數除之,得許多平方差離之算術中數;更將此中數開方求方根,得數卽爲標準差.故標準差爲平方差離之算術中數之方根,英文中另有一名爲"Root-mean-square deviation",實將計算之法詳述無遺,計算差離時,永遠以算術中數爲原點,因如此計算時,所得之値永遠爲一最小數也.茲舉一簡單之例於下,以說明此法之運用.

表三十一

標準差之計算法

m	f	d	d^2
3	1	−6	36
6	1	−3	9
9	1	0	0
12	1	+3	9
15	1	+6	36
	5		90

$M=9$

$\sigma=\sqrt{\frac{90}{5}}$

$=\sqrt{18}$

$\sigma=4.24$

由未經分組之材料計算標準差時,(如此地之例),其公式爲:

$$\sigma=\sqrt{\frac{\Sigma d^2}{N}}.$$

若材料已經分組,計算標準差之方法較爲複雜.此時不得不用假定原點以計算差離,因如此可使計算大爲簡捷也.求標準差之普遍式爲:

$$\sigma=\sqrt{\frac{\Sigma fd^2}{N}},$$

此中 f 代表分組頻數,d 爲對於算術中數之差離,N 爲包含之總例數.由前式可得

$$\sigma^2=\frac{\Sigma fd^2}{N},$$

若以 d' 代表對於假定原點之差離,由此而得之標準差以 s 爲代表,則得

$$s^2=\frac{\Sigma f(d')^2}{N},$$

由算術中數求得之標準差(σ),較之由尺度上任何點所可求得之標準差爲小.故 s^2 大於 σ^2.吾人可以 c 代表實在之中數與假定原點間之差,吾人極易證明下列之關係:

$$\sigma^2=s^2-c^2$$

—— —— —— ——

因 $\sigma^2=\frac{\Sigma d^2}{N}$　　　但 $\Sigma d=0$

$s^2=\frac{\Sigma(d')^2}{\mathrm{N}}$　　　$\therefore \Sigma(d')^2=\Sigma d^2+Nc^2$

$d'=d+c$　　　$\frac{\Sigma(d')^2}{N}=\frac{\Sigma d^2}{N}+c^2$

$(d')^2=d^2+2cd+c^2$　　　$s^2=\sigma^2+c^2$

$\Sigma(d')^2=\Sigma d^2+2c\Sigma d+Nc^2$　　　$\sigma^2=s^2-c^2$

(參考 Yule, Introduction to the theory of Statistics)

—— —— —— ——

故標準差用計算 s^2 與 c^2 之方法計算最易確定.其手續詳細說明於表三十二,此表爲384個月倫敦巴黎匯兌率之分配.

表三十二

標準差之計算法

1882—1913年倫敦巴黎匯兌率

(此表根據每月月初之匯兌率，原文載經濟學案)

組距	中點	組數	由假定原點計算之離差				
	m	f	d′	fd′	$f(d')^2$	$(d'+1)^2$	$f(d'+1)^2$
25.07—25.089 法郎	25.08 法郎	1	−8	−8	64	49	49
25.09—25.109	25.10	4	−7	−28	196	36	144
25.11—25.129	25.12	14	−6	−84	504	25	350
25.13—25.149	25.14	20	−5	−100	500	16	320
25.15—25.169	25.16	45	−4	−180	720	9	405
25.17—25.189	25.18	60	−3	−180	540	4	240
25.19—25.209	25.20	40	−2	−80	160	1	40
25.21—25.229	25.22	43	−1	−43	43	0	
25.23—25.249	25.24	42	0			1	42
25.25—25.269	25.26	32	1	32	34	4	128
25.27—25.289	25.28	26	2	52	104	9	234
25.29—25.309	25.30	21	3	63	189	16	336
25.31—25.329	25.32	20	4	80	320	25	500
25.33—25.349	25.34	4	5	20	100	36	144
25.35—25.369	25.36	6	6	36	216	49	294
25.37—25.389	25.38	2	7	14	98	64	128
25.39—25.409	25.40	2	8	16	128	81	162
25.41—25.429	25.42	2	9	18	162	100	200
		384		−372	4,076		3,716

$N=384$

組距 $=.02$ 法郎

c (以組距為單位) $=\frac{-372}{384}=-.969$

c^2 (以組距為單位) $=.9390$

s^2 (以組距為單位) $=\frac{\Sigma f(d')^2}{N}=\frac{4,076}{384}=10.6146$

σ^2 (以組距為單位) $=s^2-c^2=10.6143-.9390=9.6756$

σ (以組距為單位) $=3.11$

σ (用原來單位) $=3.11\times.02=.0622$

請注意,此地全部之計算皆用組距爲單位,直至最後始化爲原來之單位. c 爲眞正之算術中數與假定原點間之差,計算時先求差離之代數和,然後以例數除之,將 c 化爲原來單位,加在選爲假定原點之值上,卽可決定算術中數,但此舉並非必要.蓋在計算標準差時,算術中數之眞正數值可以不必知道者也.

表三十二第 (7) (8) 兩行,乃所以供吾人校核計算之正確與否者(薛立愛氏校核法 Charlier check).(註)若計算差離時,所用之原點較之現今所用者低一個組距,則得等於 $d'+1$ 之一列數值,第 (7) 行爲其平方;乘以相當之頻數則得第 (8) 行各值.其總和爲 3716, 此總和與以前諸值有固定關係如下:

$$\Sigma f(d'+1)^2=\Sigma f[(d')^2+2d'+1]$$
$$=\Sigma f(d')^2+2\Sigma fd'+\Sigma f$$

或 $$\Sigma f(d'+1)^2=\Sigma f(d')^2+2\Sigma fd'+N$$

將表三十二中計算得之各值代入最後一式,可得校核式如下:

$$3{,}716=4{,}076+2(-372)+384$$
$$=3{,}716$$

各項數值業已作成頻數分配時計算其標準差之方法,

(註) 參考 Vorlesungung über die Grundzüge der Mathematischen Statistik, C. V. L. Charlier, 19.

其步驟可綜括之如下:

1. 在分配中心附近選定一組以其中點作為假定原點.

2. 計算各組中每項對此原點之差離,並以各該組之頻數乘之.

3. 求乘積之代數和,以 N 除之,得 c (以組距為單位).再計算 c^2 之值.

4. 求各組中每項對於假定原點之差離的平方,並以各該組之頻數乘之.

5. 求平方差離之總和,以 N 除之,得 s^2 (以組距為單位).

6. 根據公式 $\sigma^2=s^2-c^2$, 計算 σ^2, 開方得 σ (以組距為單位).

7. 如此得來之 σ,復以組距乘之,結果為以原來單位表示之 σ.

標準差之若干特質,及其對於其他離中趨勢量之關係,將於後文討論之.

〔四分位差〕

前章討論如何決定平均數之位置時,關於四分位數與十分位數,曾加以敍述.在分配中各項所分佈之價值尺度上,前者為平分項數為四等分之諸點,後者為平分項數為十等分之諸點.若知四分位數與十分位數各點之位置,則吾人對

於一個頻數分配的離中趨勢之程度與性質，自可作一確切之敍述。雖然，此種知識對於頻數分配之敍述固爲有用，然爲簡明之敍述與比較計，反不及平均差與標準差合宜，蓋幾個相關聯之數字，其意義反不及一單獨之數字顯豁易解也。惟吾人可根據四分位數計算一種數值，即四分位差（quartile deviation），此種離中趨勢量，以計算簡便及意義明顯爲其優點。

兩個四分位數間之距離，當然包括全體例數之半，集中之程度越高，此距離越小。因此吾人可以兩個四分位數間之關係計量離中趨勢，相當正確。四分位差爲第一四分位數與第三四分位數間距離之半。故英文又名 Semi-inter quartile. range. 如 $Q.\ D.$ 代表四分位差，Q_1 代表第一四分位數，Q_3 代表第三四分位數，則

$$Q.D.=\frac{Q_3-Q_1}{2}$$

若以 K 代表第一四位位數至第三四分位數間之中點，則頻數分配之全部例項中應該有半數在 $K\pm Q.D.$ 之範圍以內引用前舉之倫敦巴黎匯兌率之材料，得

$$Q_3=25.262$$

$$Q_1=25.174$$

$$Q.D.=\frac{25.262-25.174}{2}$$

$$=.044$$

$$K=25.174+.044$$
$$=25.218$$

故此地有全體例項之半數位於 25.218±.044 之範圍內.此種記述與匯兌率之平均數(中數中位數或衆數)並用,成為分配之有效的敍述.一完全對稱之分配,其 K 值將與中位數之值相合(即中位數在 Q_1 與 Q_3 之距離之中點也).匯兌率之分配為不對稱,中位數之值為 25.214,而 K 之值為 25.218.

〔機差〕

計量天文現象或物理現象時,觀察家對同一固定量計量所得之結果,各各不同.但此種不同之結果,其分配有一定規則,繪成曲線時與差誤正態曲線相似.遇此種情形,顯然迫切需要一種離中趨勢之計量,以指示結果之可靠與否.若許多觀察家所得之結果或一個觀察家許多次觀察之結果相差甚遠,此結果自不能認為可靠;相反,若相差甚微,則此結果可以認為可靠,在此種情形下平常應用之離中趨勢之計量,名為機差 (probable error). 機差者,在某一事件中,恰為半數觀察之差誤所超過之某一數值也.由若干次觀察所得之一列數值,其中正確之機會最大者,為其算術中數;故機差總依算術中數計算,此數值之所以名為機差,蓋由於某一觀察與全部各次觀察之算術中數間,其相差之值大於機差之機會恰為 $\frac{1}{2}$ 也.因此各次觀察之結果,若列為頻數分配形式,由算術中數起上下兩端各取相當於機差之數值,則事例總數之

半將盡包其內.

此種離中趨勢量除原來所應用之領域以外,亦應用於其他某種領域,其中若仍用原來之名詞 probable error (嚴格譯之應爲機誤)似易滋誤會;不若視爲 probable deviation (機差)也.此卽對於算術中數之一距離,爲全部差離中之半數所超過者.

機差此離中趨勢計量,惟有應用於依照差誤正態曲線之分配時始有完全之意義.在此種情形下,彼固具有確定而嚴格的意義;但應用於編態分配則不然,故遇編態分配不宜應用.四分位差於正態分配中其値與機差相等,應用於編態分配之敍述,較之機差其意義更爲直接,故宜用於此種編態之分配.應用機差爲統計結果之可靠量(measure of reliability)之事,將於後文詳述之.

某一事件之機差數値,若遇正態分配,可由標準差之數値決定之,因兩者間常有一固定之關係存在也.此關係可以下列公式示之:　($P.E.=$ 機差)

$$P.E.=0.6745\,\sigma.$$

[各種離中趨勢量間相互之關係]

玆將各種離中趨勢量作一總比較,並略述其相互之關係,俾其意義易於了解.

1. 全距爲尺度上包括全體觀察値之距離.

2. 四分位差爲尺度上如此之距離,卽將第一四分位

數與第三四分位數間之中點爲起點，將此距離兩方展開時，可包括全體觀察値之半數者.

3. 平均差之計算，以算術中數爲原點者，在對稱或微偏之分配中，約爲標準差之$\frac{4}{5}$. $7\frac{1}{2}$倍於平均差之距離，以算術中數爲中點時，約可包括全體例項之99%.

4. 若以中數爲起點兩方展開等於標準差之距離，在正態或微偏之分態中，約可包括全體例項之$\frac{2}{3}$.（在正態分配中恰可包括全體例項之68.26%）.若以中數爲起點兩方展開兩倍於標準差之距離，約可包括全體例項之95%.（在正態分配中恰可包括95.46%）. 若以中數爲起點兩方展開三倍於標準差之距離，約可包括全體例數99%.（在正態分配中恰可包括99.73%）.六倍於標準差之距離而以中數爲中點，約包括全體例項之99%，此爲一通則，對於校核計算之正誤上，甚有用處.

就圖43而研究之，可助吾人理解標準差在正態分配中之意義.

5. 在正態分配中，機差等於0.6745σ，兩倍於機差之距離而以中數爲中點，將包含全例數50%，八倍於機差之距離而以中數爲中點，約包括全例數之99%.

［各種主要離中趨勢量之特色］

全距

1. 全距易於計算，其意義亦易了解，作爲一粗糙之離

中程度之計量,全距乃有用者也.

2. 全距之值決定於兩極端項,因此頗不確定.一數之增減即可使其值大變.

3. 介於兩極端項之間之各項之分配性質,全距不能表現.

四分位差

1. 四分位差爲一易求而易解之離中趨勢量;作爲一粗糙之離中趨勢量,較全距爲佳.

2. 四分位差非對於任何一特殊之平均數之離中趨勢量.

3. 在第一四分位數與第三四分位數間各項之分配.與在兩個四分位數外之各項之分配,皆不影響四分位差.兩分配之情形儘可大相懸殊,只須其四分位數相同,四分位差亦相同.四分位差因不爲個別各項之差離所影響,故不能視爲一確切之離中趨勢量.

4. 四分位差不適用代數計算方法.

平均差

1. 平均差受每項之數值所影響.此數爲一個頻數分配之中位數(或中數)及個別各項間之差之平均數,有確切的意義.

2. 平均差所受極端差離之影響,較標準差爲小.

3. 就數學之理論言之,平均差作離中趨勢之計量,不

若標準差之爲合理與方便.

標準差

1. 標準差受每項之數值所影響.

2. 先求差離之平方然後相加,故免去不顧正負號之代數意義上之錯誤.

3. 標準差有確定的數學意義,且十分適宜於用代數方法處理.

4. 標準差就一般言之,較之其他離中趨勢之計量少受抽樣變動(fluctuations of sampling)之影響.

5. 差誤正態曲線係皆以標準差單位分析,如此得來之知識大大增加標準差之效用.

機差

1. 若分配情形按照差誤正態定律,機差有確定之意義.惟對於其他情形之分配卽無此確切之意義,故不宜用於後者之敍述.

2. 機差對於其適用之分配爲極有效之計量.其最重要之用處爲指示量的計量之可靠性.

3. 正態分配之機差與標準差間有固定關係存在,使機差易於求得.

所有上述各種離中趨勢之計量,皆可用於某種特殊目的.惟以標準差爲最普遍適用,無論何處須要高度之精確時宜應用之.機差實際上僅爲標準差之一部分($P.E.=0.6745\sigma$),

其適用之範圍確定而有限止.

〔相對離中趨勢之計量〕

前節所討論者爲絕對離中趨勢.所求得之離中趨勢量,皆以絕對單位敍述材料之離中趨勢.如以法郎計倫敦巴黎間匯兌率之標準差,以噸數計生鐵生產之標準差,諸如此類.若吾人之目的在敍述一單獨之頻數分配,自宜一貫應用原來單位;但若有兩個不同的頻數分配而欲比較其離中趨勢量時,乃有困難發生.單位若原非一致,困難固爲必然之事,然即使單位相同,仍有同樣之困難存在.如狗之體重與馬之體重,可同以磅計也,然不可因馬之體重之標準差,較大於狗之體重之標準差,即謂前者離中之程度較大.絕對離中趨勢量須與計算差離時所根據之平均數相聯繫始有意義;離此平均數而單獨應用則其意義全失.因此,欲彼此比較時,必須先化爲相對的形式,其法顯然爲將離中趨勢量化爲計算差離時所根據之平均數之百分數.於是求得之數量爲一抽象數,爲材料之相對離中趨勢量,可與其他分配中所得之相似項彼此比較.

〔離中係數〕

相對離中趨勢量中應用最廣者,爲皮爾生氏 (Pearson) 所發明之離中係數 (*Coefficient of variation*),以字母 V 代表之.欲求此計量,只須將標準差化爲算術中數之百分數.故

$$V=\frac{\sigma}{M}\times 100$$

前面分析倫敦巴黎匯兌率材料所得之結果,用此公式求其離中係數得

$$V=\frac{.0622}{25.2206}\times 100$$

$$=.25\%$$

根據表三十之材料,在此同時期內倫敦紐約匯兌率之離中趨勢可以求得,更求其離中係數得.33%.可見在此時期內,倫敦紐約間匯兌率之變動,較之倫敦巴黎間爲大.

若將其他任何種離中趨勢量,化爲計算差離時所根據之平均數之百分數,尙可求得與此種離中係數相似之其他離中性之指數(Index of variability),但平常所採用者爲皮爾生氏之係數,亦唯有此一種係數受廣汎之應用也.

[偏態量]

前文說明如何敍述一個頻數分配之集中趨勢,以及如何計量集中趨勢周圍之集中與分散之程度.吾人尙須一種計量,即指示該頻數分配之偏態或不對稱之程度者也.分配中之各項在集中值周圍分佈排列之狀態爲對稱的形式與否,此事甚關重要.如吾人更有一此種之數字,則一個頻數分配之特質可盡納於三個簡單之數字中.此三個數字即平均數,離中趨勢量及偏態量是也.計量偏態之法有幾種:

若頻數曲線之形式爲完全對稱者,則中數,中位數及衆數相合爲一.分配之情形愈不對稱,三者相去越遠.中數與衆數間相差最大,故可用此差數爲偏態之計量.爲便於與其他頻數分配之相似之偏態量相比較起見,亦如相對離中趨勢之計算然,宜於求一抽象數.皮爾生氏主張以標準差除中數與衆數間之絕對值之差,以其商爲偏態量;其公式爲:(偏態量以 sk 代表之.)

$$sk=\frac{M-Mo}{\sigma}$$

在正態(對稱)之分配中,中數與衆數相合爲一,此數之値當然爲 0. 其他情形中,數値可正可負,視兩平均數在尺度上相互之位置如何而定.

偏斜不甚之分配,用下列公式計算偏斜程度較爲便捷:

$$sk=\frac{3(M-Md)}{\sigma}$$

此公式幾與前式相符合,因在偏斜不甚之分配中,中位數位於中數及衆數之間,在前者向後者之距離上約當三分之一之處.

衆數之位置既難用簡單方法決定,有時甚願有計算較易之偏態量代替皮爾生氏之方法.蒲蘭(Bowley)氏曾提出一種方法,根據第一及第三四分位數與中位數之關係:若分配爲對稱形式,則兩個四分位數與中位數間之距離相等;分

配若不對稱,情形卽非如是.故若以 q_2 代表較高之四分位數與中位數間之差,q_1 代表中位數與較低之四分位數之差,則可用下列公式

$$sk=\frac{q_2-q_1}{q_2+q_1}$$

作爲計量偏態之辦法.此數將在 0 與 ±1 之間變化.因爲在完全對稱之情形中,$q_2=q_1$,故此數爲 0;若不對稱之程度大至中位數與兩四分位數之一相合,則 q_2 或 q_1 必有一者爲 0,由公式求得之答數爲 +1 或 −1,蒲蘭以爲數值爲 .1 者爲輕微的偏態,數值爲 .3 者爲顯著的偏態.

用此法求得之數值當然與用皮爾生公式求得者不能相比.

〔峯態〕

前曾提及頻數曲線之第四特質可以計量者,卽以之與差誤正態曲線相比較時,其頂端銳鈍之程度也.專門名詞名之曰峯態 (kurtosis).計量峯態之方法當於第十五章中討論之.

參考書

Bowley, A. L. Elements of Statistics (110—117)

Chaddock, R. E. Principles and Methods of Statistics (Chap. IX)

Davies, G. R. Introduction to Economic Statistics (29—46)

Jones, D. C. A First course in Statistics (42—51)

Kelley, Truman L. Statistical Method (70—82)

King, W. I. Elements of Statistical Method (141—158)

Pehrl, Raymond. Medical Biometry and Statistics (272—278)

Rietz, H. L. (Editor) Handbook of Mathematical Statistics (27—33)

Rugs, H. O. Statistical Methods Applied to Education (149—179)

Secrist, Horace. An Introduction to Statistical Methods (377—424)

West, Carl J. Introduction to mathematical Statistics (45—58)

Yule, G. H. Introduction to the Theory of Statistics (133—156)

第六章

物價指數

[指數之性質]

"指數"(Index number)之一名詞,吾人用以代表分析統計數列時所應用之多少相似之若干方法.研究物價變動時,應用最廣.惟對於其別種用途若略加考慮,則此計量之基本特質自可明瞭.指數之最簡單之形式,爲時間數列中以比值(Relative)表示之各項.如下表爲美國棉花消費量之指數:

表三十三

1913—1923美國國內棉花消費量

(1913年=100)

年份	棉花消費量(單位包)	棉花消費量比值
1913	5,583,463	100.0
1914	5,448,760	97.6
1915	6,008,984	107.5
1916	6,620,415	118.6
1917	6,815,811	122.0
1918	6,176,547	110.5
1919	5,919,520	106.0
1920	5,843,200	104.5
1921	5,405,721	96.8
1922	6,087,520	109.0
1923	6,513,698	116.7

一種商品之價格同樣可以比值表示,以某一日期或某一時期之價格作爲基數.

表三十四

敏尼阿波利埠第一號北方春小麥
之平均價格,1913,1919--1923
(1913年=100)

年份	每蒲平均價格	價比(價格之比值)
1913	$0.8735	100
1919	2.5600	294
1920	2.5581	293
1921	1.4660	168
1922	1.3450	154
1923	1.1810	135

時間數列之各項,依照一固定之基數化成比值後,日期不同之諸數值易於比較;觀察整個數列之趨勢,亦比較材料表現爲原來之形式時便利甚多,且使不同諸數列之趨勢之互相比較輕而易舉.

指數之一名詞,固亦用於指此種之比值,然最好留以稱呼代表若干數列的結合之數字.所結合之數列或爲關於物價者,或爲關於生產者,或爲關於消費者,或爲關於工資者,或爲關於貿易量者,或爲關於其他任何種依時變動之因子者.編製此種種特殊形式之指數時,莫不遇有十分複雜之問題.然主要之目標處處相同:即獲得一單獨之簡單數字,以指示許多力量之淨餘合力;此許多力量者即影響其所結合之若干數列者也.

美國煤及煤油生產之變動,可作一簡單之指數以表示之.欲作一如此之指數,必須設法將煙煤.白煤及煤油之生產數字結合爲一.下表爲1910—1923年三數列之生產數字及比值:

表三十五

1910—1923年煙煤、白煤及煤油之生產(美國)

(1910年=100)

年份	煙煤產量(單位:百萬美噸)	比值	白煤產量(單位:百萬美噸)	比值	煤油產量(單位:百萬桶)	比值
1910	417.1	100	84.49	100	209.6	100
1911	405.9	97	90.46	107	220.4	105
1912	450.1	108	84.36	100	222.9	103
1913	478.4	115	91.52	108	248.4	118
1914	422.7	101	90.82	107	265.8	127
1915	442.6	106	89.00	105	281.1	134
1916	502.5	120	87.58	103	300.8	143
1917	551.8	132	99.61	118	335.3	160
1918	579.4	139	98.83	117	355.9	170
1919	465.9	112	88.09	104	378.4	180
1920	568.7	136	89.60	106	442.9	211
1921	415.9	100	90.44	107	472.2	225
1922	404.5	97	52.90	63	557.5	266
1923	545.3	131	95.2	113	725.7	346

欲根據此三個數列求一粗略之指數以代表燃料生產量,顯然不能將原來數字相加因,所用之單位不同也.此困難可用比值以避免之,可將三個比值之簡單平均數作爲各年之指數.茲列如此求得之指數於下:

表三十六

1910—1923年美國煤及煤油之生產指數

(1910年之生產100)

年份	指數	年份	指數
1910	100	1917	137
1911	103	1918	142
1912	105	1919	132
1913	114	1920	151
1914	112	1921	144
1915	115	1922	142
1916	122	1923	196

在求此指數時,將每年之三個比值相加而除以三,即視三個數列之重要性皆相等.此種與各比值以同權求得之指數,有人名之曰未權指數(unweighted index),但此名易滋誤會,蓋此中未嘗不用權數,特權數各各相等耳.此根據同權而求得之指數,顯然不能忠實反映此處結合之三數列.因三數列之重要不同,不宜以同權處之也.下列數字示1921年交易所中煙煤,白煤及煤油之批發價格,可指示三個數列之相對的重要性.(根據美國勞工統計局所編之數字)

礦產	1921年交易所批發價格
烟煤	$1,948,000,000
白煤	731,000,000
煤油	712,000,000

粗算之,三者彼此之關係爲3,1,1之比,吾人可以此爲三數列之權數,用以計算每年之指數,下例舉1910及1923年指數之求法以說明之:

表三十七

煤及煤油生產之加權指數計算法

礦物	生產量之比值 1910	權數	權數×比值	生產量之比值 1923	權數	權數×比值
烟煤	100	3	300	131	3	393
白煤	100	1	100	113	1	113
煤油	100	1	100	146	1	346
		5	500		5	852

1910年燃料生產指數 =500÷5=100

1923年燃料生產指數 =852÷5=170

如此求得之1910—1923年十四年中每年之指數數值如下:

表三十八

1910—1923年美國煤及煤油生產之加權指數

年份	指數	年份	指數
1910	100	1917	135
1911	101	1918	141
1912	106	1919	124
1913	114	1920	145
1914	107	1921	126
1915	111	1922	124
1916	121	1923	170

前後兩指數數列(表三十六與表三十八)顯有重大差

別後者先經合理的加權,自較前者爲精確,足以較爲忠實地表現影響三數列之各種力量之共同的結果.

指數之另一種形式,乃將諸數列之各項相加,以所得總值爲全體之代表,而不以平均數爲代表.此種形式之指數,唯有各數列所用之單位相同時始可求得.常用於指示物價水平之變動,以及比較同樣之一批貨物在不同日期之成本.下列數字爲此種形式之指數之一例:

表三十九

Bradstreet 氏 1913—1923 年美國之

批發物價指數

年份	指數	年份	指數
1913	$ 9.21	1919	18.66
1914	8.90	1920	18.81
1915	9.85	1921	11.37
1916	11.82	1922	12.12
1917	15.64	1923	13.40
1918	18.71		

上列各年總值爲該年96種商品平均批發價格之和.未加以前各價格一律化爲每磅之值,使在某種程度上可以互相比較.此種指數極易改爲比值形式,只須將任何一年總值作基數而以其他各年總值表示爲基年總值之百分數也.

指數之形式極多,上述諸例僅指示其中之數種.此名詞可指簡單之比值言,可指若干比值之平均數言,亦可指若干比值或絕對數之總值言.在上述諸例中,指數之目的無非計量一時期之變動,而爲指示時間數列中各值的變動之工具.然指數之意義不以此爲限,其應用之範圍甚廣.欲計量推銷員之能力,則可用數值表示各種決定其效率之因子,然後求各因子數值之平均數而作指數.企業中各部門之效率,亦可同樣以指數示之.無論在何種情形之下,指數之編製,皆先將若干不同之因子化爲可以比較之項,復代以一單獨之數字作其代表.不論以此時與彼時比較,或將從單位相同各項平均而得之幾種指數相比較,皆因有指數之故,其工作大爲便利.故無論何種形式之指數(除簡單之比值外),皆爲統計上之平均數之一種.而以前所述關於平均數之一切規律及限制,莫不可以適用於指數者也.

本書僅欲討論指數方法對於時間數列之應用.但應用於不同形式之時間數列時,其規律與辦法甚不一致,故其中有若干形式必須單獨論述.茲先就批發物價指數討論之.

[物價之變動]

當仔細觀察物價變動時,難見其有秩序或固定趨勢,吾人僅見其錯綜複雜互相矛盾之一團耳.下列物價表乃隨便集成者,當吾人爲比較逐月物價變動而檢視一切物價時,所見之情形大致可以此表爲代表.

表四十

商品批發價格(註)

商品	單位	批發價格 1923年十月	批發價格 1923年十一月
磚(平常建築用,平均在廠價格)	一千塊	$14.752	$14.746
基性銑鐵(谷密)	英噸	23.500	20.875
水泥(波特蘭,在廠價格)	桶	1.893	1.842
亞麻仁原油(紐約)	加侖	.943	.910
鋼條(貝西末爾,匹茲堡)	英噸	40.000	40.000
馬口鐵(匹茲堡)	一百磅	5.500	5.500
電氣用銅(在廠價格)	磅	.126	.123
鉛(紐約)	磅	.069	.069
鋅(紐約)	磅	.067	.067
白煤(紐約)	英噸	11.471	11.478
煙煤(支加哥)	美噸	4.600	4.525
粗製煤油(賓西爾法尼亞,在廠價格)	桶	2.500	2.388
馬達用汽油(紐約)	加侖	.185	.170
中等棉(紐奧爾良)	磅	.292	.339
二號紅冬牌小麥(支加哥)	蒲	1.097	1.081
糖(紐約)	磅	.090	.087

表內十六種商品,在1923年十月與十一月間,有四種之價格未起變化,三種漲價,九種跌價.漲跌之程度,或為甚微,或甚為顯著.此處所見之縮影,可代表整個物價系統之實際現象.一切物價之上漲下跌或固定不變,決無絕對一致之理.在任何國家之市場上或全世界市場上交易之數千種商品,各因其所受影響之不同而自為變動.然各種商品價格之變動亦非孤立.此物之價變則影響他物,而他物之價變亦影響此

(註) 美國勞工統計局編製.

物,除專門影響某種商品之力量外,又有較普遍之力量其作用及於整個價格系統而影響一切商品者.經濟統計家之任務,即爲由發生於某一時間之混亂複雜之物價變動中求出秩序,由無數細小之變動中選取影響整個經濟系統之普遍趨勢也.

引起物價變動之力量繁多而複雜,惟可略述其大概.第一爲生產與消費情形中專與個別商品有關之一切變動.直接影響個別商品之價格,新銷場之開拓,個別之生產技術之改良,時尙之趨捨,需要之由一商品而轉移於他商品,需要與供給之季節變動,——一切皆引起物價之不斷變動不斷適應.此種變動在平時最爲明顯,直接爲商人與消費者所覺察.此種變動之影響於整個物價系統,已如前述,惟一般言之,不能使整個系統起向上向下之總變動也.

引起此種總變動之力量,其範圍較廣.生產技術之一般的進步,及由此而來之人類勞動生產力之增強,使消費品之供給大增,而影響於物價;貨幣制度之變化,特別爲金銀供給量之變化,影響流通貨幣之供給,故對於物價起直接之作用.銀行及信用制度之變化,及商業慣例之改變,影響信用票據之運用,及貨幣與信用之流轉速度,故對於物價所起之作用,性質亦同,所有此種種力量皆作用於物價,然其影響之範圍,不若直接影響個別商品之諸因子之爲專門也.

[一般批發物價指數之目的]

上述種種力量不能使其互相隔離而分別與以估量.其聯合之作用使價格變動錯綜複雜.欲研究此種變動可由若干不同之觀點下手.研究之目的或爲觀測物價系統內部之改組,確定爲環境改變引起的變動之性質與程度.此種研究使吾人對於物價之動態及其相互間的關係之性質獲得寶貴之知識.惟目前所欲研究之問題,爲如何決定種種力量之純粹合力.物價之一切變動能否彼此抵消?一部上漲一部下跌時純結果能否無變?在某一時間,其變動能否偏重一方,召致一般物價水平之抬高或降低?若有此種趨勢存在,其內容如何?應如何計量之?以前各節所討論之統計方法可否應用於此問題之解決?

在此研究中,請先答覆最後提出之一問題.吾人已有綜合統計材料之種種方法,惟此種種方法祇能應用於一定條件之下,前文曾指出一平均數所代表者,必須爲無數同質材料之顯著的集中趨勢,否則此數將毫無意義.平均數之各種型式,應以何者最爲適用,亦應視其所代表之分配之性質如何而定.對於原來材料之分配狀態未經研究,而欲適當運用平均數或其他統計量,殆爲不可能之事.故吾人首須決定問題之原料爲何,並研究從此原料整理而得之頻數分配.

現今假定一極普通之目的,即決定在兩個指定日期間一般批發物價水平之變動是也.當然,此即等於計量在彼兩

日期間貨幣價值之變動.此問題之原料,爲在此互相比較之兩日期之若干商品價目.每一對價目指示一種商品之價格變動.此變動爲多種力量相互作用之結果.若集合無數如此之價目,則所集成之材料,可以代表無數力量之相互作用;其中若干力量之作用限於個別而專門之範圍,另有若干力量之作用則普遍影響於大批之商品或全體之商品.吾人所欲決定者,爲所有影響物價之種種因子之純粹合力.吾人對於引起個別的物價變動之無數力量,思設一數以計量其共同合作之結果.此數卽成爲批發物價之指數.

吾人所須處理之每一單位,代表物價之一個變動.以前討論之各種統計方法,是否適宜於組織及分析許多個如此之單位乎?此點須視此種單位在成集團時其動態如何而定.下列諸例說明由此種材料分組而得之頻數分配.

[價格比率之頻數分配]

每一個價格變動,當然成一比率(ratio),卽一種商品在此一時之價格對於彼一時之價格之比率也.若用以前曾述及之法,將若干之比率一律化爲比值(relative)之形式,(如取表四十中之一對價格爲例,1923年十一月生鐵價格與同年十月生鐵價格之比率爲$20.875:$23.500,化爲比值形式得88.8:100.)則彼此可以比較.下列頻數分配表,表示346種商品在1914年之批發價格;皆以1913年之價格爲基數(base),化成比值.此346個比值之分配如下:

表四十一

346種商品之價格比值(價比)之類數分配，1914年

(1913年平均價格=100)

價比	中點 m	例數 f	占總例數 %
62.5— 67.4	65	1	.3
67.5— 72.4	70	1	.3
72.5— 77.4	75	5	1.5
77.5— 82.4	80	7	2.0
82.5— 87.4	85	20	5.6
87.5— 92.4	90	35	10.0
92.5— 97.4	95	51	14.5
97.5—102.4	100	134	39.0
102.5—107.4	105	50	14.5
107.5—112.4	110	21	6.0
112.5—117.4	115	12	3.5
117.5—122.4	120	3	1.0
122.5—127.4	125	2	.6
127.5—132.4	130	2	.6
132.5—137.4	135		
137.5—142.4	140	1	.3
142.5—147.1	145		
147.5—152.4	150	1	.3
		346	100.0

(註) 表中所包括之346種商品係美國勞工統計局用於作批發物價指數者.原來數目及其比值載該局公報第269號: "1890—1919年之批發物價."

圖47爲表示上述分配之頻數多邊形,此圖根據百分數之分配而作,以便與相似之分配互相比較.此處顯然可見此頻數分配與前面所描寫之標準形式相符合.在集中趨勢周

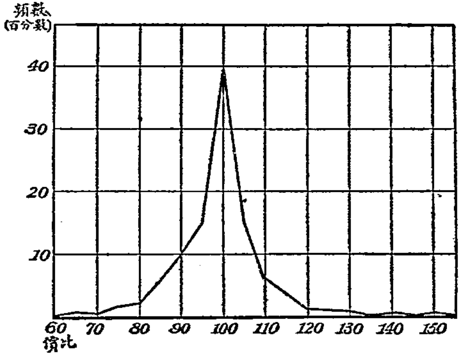


圖47.——頻數多邊形:1914年346種商品之價比之分配(1913平均價格=100)

圍同有明顯之集中狀態.本例全部各項中有39%其變動不超過基年價格之25%者;故價格有穩定之勢.集中趨勢兩端之分配,衆數以上之距離大於其以下之距離此點雖甚重要,然兩端尚相當對稱.此分配之集中趨勢究以何種平均數代

表之爲最適當,玆姑不置論;然必有某種平均數可十分合理的應用於此,乃顯而易見者也.

方纔所擧之例,其根據之材料爲由一年至次年之物價變動;在其所處之時期中,一般之物價水平未有變動.米契爾氏曾擧一例,較前者大爲廣博,彼在1890—1913年之一時期內,抽取由一年至次年之價格變動凡5,578個,而觀其分配.結果情形與前例相同.衆數以上之距離更大於衆數以下之距離.吾人宜注意此23年中大部分均爲物價高漲之時期也.米契爾氏求得之分配見圖41.

研究逐年之價格變動時,價格之惰性(inertia of prices)最爲明顯.故最好再研究長期之價格變動之性質,觀其是否有同樣型式之分配.玆擧價格變動之二例,其一以十年爲期,選擇此時期時,曾注意時期之末之價格水平大致與時期之初相同;其一以五年爲期,此五年間以物價飛漲爲特色.下列之一表記載222個價格變動之分配,1900年之價格乃以1890年之價格爲基數而化爲比值者.(注意,一般物價水平,由1890年至1896年向下跌落,由1896年至1900年向上回漲,1900年之物價約低於1890年者1%.)

表四十二

1900年222種商品價比之分配

(1890年平均價格=100)

價　　比	中點 m	個數 f	占總個數 %
42.5— 47.4	45	2	1.0
47.5— 52.4	50	1	.5
52.5— 57.4	55	3	1.5
57.5— 62.4	60	3	1.5
62.5— 67.4	65	3	1.5
67.5— 72.4	70	3	1.5
72.5— 77.4	75	8	3.5
77.5— 82.4	80	12	5.0
82.5— 87.4	85	19	8.5
87.5— 92.4	90	20	9.0
92.5— 97.4	95	30	13.5
97.5—102.4	100	27	12.0
102.5—107.4	105	24	11.0
107.5—112.4	110	14	6.0
112.5—117.4	115	18	8.0
117.5—122.4	120	10	4.5
122.5—127.4	125	7	3.0
127.5—132.4	130	9	4.0
132.5—137.4	135	2	1.0
137.5—142.4	140	4	2.0
142.5—147.4	145	1	.5
147.5—152.4	150	1	.5
152.5—157.4	155	1	.5
		222	100.0

圖48之頻數多邊形表示此材料之分配之百分數.(圖47與圖48之垂直尺度不同,比較時請注意.)

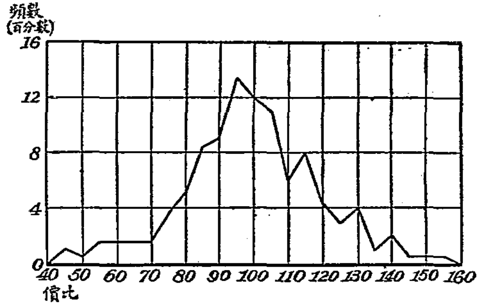


圖48——頻數多邊形:1900年222種商品價比之分配
(1890年平均價格=100)

圖47與圖48所繪之分配相差甚鉅.第一,後者之變動範圍較大,包括之時期既長,此種情形乃意料之事也.第二,本例雖仍有一顯著之集中趨勢,然例數之集中於衆數組者,其所占總例數之百分數,與前例相比甚爲微小.兩分配繪於算術尺度格紙上,皆表現相當對稱之狀態.(關於此點,請注意此地互相比較之兩日期之物價水平,用算術中數表示時實際相等;前後兩例皆如是.)在前一例中,集中趨勢周圍集中之程度較之後例高過甚多,而各價比對於算術中數之差離則較小.此分配與十分精確之物理測量或十分精確之鎗礮發射所得之結果相似.後一例之頻數分配曲線相當於較不精確之物理測量或陳舊不精確之鎗礮發射所得之觀察值或子彈之分配狀態;其中衆數之頻數較小,各項對於算術中數之

差離較大.所比較之價格相隔之時間越長,則近於此地後一例之曲線之傾向愈爲明顯;最大縱坐標之值減小,分配之全距增大;曲線越形平坦而左右擴展;此爲業經確證之事.隨時期之增長,任何一種平均數之代表資格顯然減少.然集中趨勢周圍若無顯著集中之情形,平均數僅一抽象無具體之意義.

於此可作一嘗試性之結論曰:價格變動爲適用統計計量方法者.時期若非過長,一平均數值足以適當代表之.計量價格變動時所可採取之最長時期爲何,固無固定之標準.然指數之有確切重要之意義者,其所根據者必爲相當短期之價格比較,而最精確者必爲逐年價格變動之比較.指數之僅以指示價格變動之一般趨勢爲目的者,可以採用較長之時期,雖然,此種指數之作者與其應用者,對於其限制所在,應明白認識也.(註1)

上舉兩例中所比較之價格,在其相比之兩日期,用算術中數代表之,其水平彷彿相同.此種當然爲特殊之例,茲請進而就一般價格水平大變之時期,研究其個別價格變動之分配狀況.下表所列乃1918年1437種商品價比之分配,以1913年七月至1914年六月間平均價格爲基數.(註2)

(註1) 本段結論大致根據米契爾氏之研究.請參考"指數作法及用法"(美國勞工統計局公報第284期批發物價叢書.)

(註2) 戰時工業局物價部計算;轉載於公報第284期第一部70頁.

表四十三

1918年1437種商品價比之分配

(以1913年七月至1914年六月平均價格為100)

價比	中點 m	例數 f	占總例數 %
36	36	1	*
49	49	1	*
50— 69	60	4	.3
70— 89	80	17	1.2
90—109	100	61	4.3
110—129	120	64	4.5
130—149	140	130	9.0
150—169	160	212	14.7
170—189	180	219	15.2
190—209	200	164	11.4
210—229	220	135	9.4
230—249	240	104	7.2
250—269	260	76	5.3
270—289	280	54	3.8
290—309	300	42	3.0
310—329	320	30	2.1
330—349	340	31	2.1
350—369	360	16	1.1
370—389	380	13	.9
390—409	400	7	.5
410—429	420	7	.5
430—449	440	8	.6
450—469	460	4	.3
470—489	480	4	.3
490—509	500	4	.3
510—529	520	5	.4
530—549	540	3	.3
550—569	560	4	.3
587	587	1	*(不足.1%者以*示之)
627	627	1	*
727	727	1	*
730	730	1	*
743	743	1	*
761	761	1	*
784	784	1	*
826	826	1	*
848	848	1	*
900	900	1	*
1165	1165	1	*
1358	1358	1	*
1585	1585	1	*
1764	1764	1	*
2049	2049	1	*
2863	2863	1	*
3009	3009	1	*
		1437	

圖49爲此分配之圖示.(此圖所用尺度不與前二圖同.)

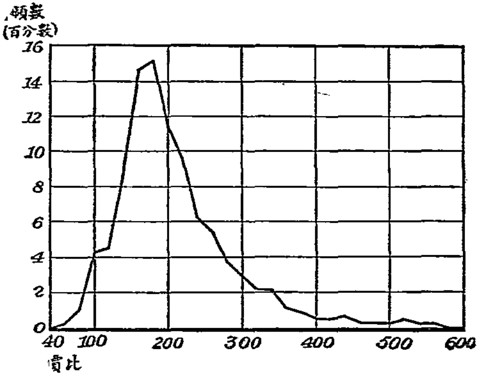


圖49——頻數多邊形:1918年1437種商品價比之分配

(1913年七月至1914年六月平均價格=100)

就此分配而研究之,可以證實由以前二例所得之結論.此分配中有一充分顯著之集中趨勢足爲一平均數所代表.且此地条數組爲以180爲中點之一組,故此集中趨勢不能歸源於惰性,而應歸源於若干影響全部物價系統之外力之存在.惟此一分配與以前兩分配之間,亦有一顯著差異之點,即偏態之傾向曾見於第一例者,此處更爲明顯也.依算術尺度而作出之曲線,顯然偏於一方.集中最強之點近尺度之下

限;而一長尾向右延伸,實際上超過圖邊甚遠.此種情形,參觀原表更可知其詳細.試由100之基數起,最小之價比爲36,在尺度上絕對低落64點;最高之價比爲3009,絕對增加2909點.誠然,此種情形之發生,正值一般物價水平飛漲之時;(戰時工業局之指數表示1918年超過基期94%),此點固足以說明衆數組之位置,惟不能說明分配之性質.說明頻數曲線之形式時,吾人須準備答覆一編製指數上之基本問題.此問題者何?即一個物價變動分配之集中趨勢,須以何種形式之平均數爲其代表數字是也.

[求物價變動之平均數之問題]

物價之增漲,以比值表示之,並無上限.100%,500%甚至1000%之增漲,非難於想像而不可能者,戰時工業局對於大戰時期價格增漲之研究中,指出最高數爲Acetiphenetidin(阿西非丁)之價格,其百分數爲4981.惟物價之減落,以100%爲最大限度,蓋此點表示商品之價格已跌至零也.此事足以說明前數曲線之偏態.吾人若集合大量之價格比率而合爲一表,依算術尺度繪其頻數曲線,此曲線必表示此種特色,而材料之來自物價飛漲時期者,此種特色最爲顯明.

此事對於選擇適當平均數之關係,在前文討論平均數時業已道及.平均若干比率時以應用幾何中數最爲適當.吾人取上列分配之各種平均數而比較之,即可明瞭.1437個價格比率,其算術中數爲217,而衆數爲180,—中數而離衆數甚

遠,自不免減損其代表之資格.此處幾何中數爲194,此値與衆數及中位數(191)大爲接近.其他理由一切不計,即據此各種平均數間之關係言,此地幾何中數實勝於算術中數.

前在討論平均數一章中,曾指出一頻數分配依對數尺度描繪時,若較之依算術尺度描繪時更近似於正態曲線,即表示其材料之集中趨勢合於幾何定律而不合於算術定律.圖50表示以1913—14年爲基年時1918年1437個價比之分配依對數尺度描繪之結果.依算術尺度描繪時所得顯著之偏

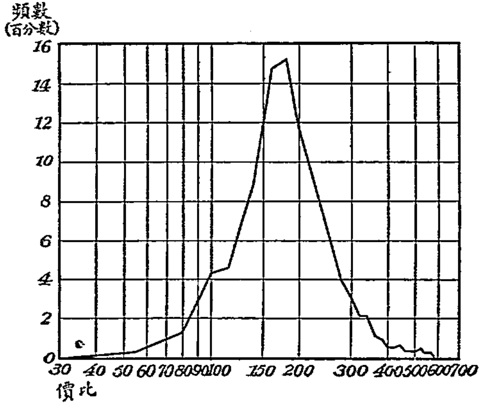


圖50.——頻數多邊形:1918年1437種商品價比之分配,價比依對數尺描繪(1913年七月至1914年六月平均價格=100)

態此處毫不存在,惟得一非常近似於正態曲線之曲線.

物價比率趨於按幾何定律而分配之傾向,固爲應用幾何中數以平均此種比率之一強固理由,然並非謂物價指數絕不能以他法求之.就一般言之,幾何中數爲應用於此之最合理之平均數,惟正確之指數固尙可有他法求之.此點將詳之於後文.

請將前數頁所討論者摘要綜述之:在開始硏究編製指數之實際方法以前,余以爲對於原料之性質,以及此原料經組合後所得之分配之性質,應先加硏究,以決定平常之統計方法是否合用.吾人已知原料之內容爲個別的物價變動,表現爲比率.集合此無數之比率乃得一頻數分配.此頻數分配與差誤正態定律多少相似.物價變動之分配中,顯然有一集中趨勢存在,可以平均數爲其正當代表.但互相比較之價格若其相隔之時期較長,集中趨勢卽較不顯著而離中趨勢則反見顯著;致平均數之代表性隨時期之增長而日益減低.此外,尙有偏態傾向之存在,且可見在物價增高之時期此種傾向最爲明顯.其原因由於吾人所處理之比率,其下限有定而上限則爲無定者.參照以前關於平均數之討論,吾人知物價變動之材料,爲屬於依照幾何定律而分佈之一種.若欲求此種變動之平均數,宜用幾何中數.此點就一般言之固爲正確,然就若干目的言,吾人知並非無其他方法適於指數之編製也.

在編製指數之理論上,關於平均數種類之選擇,以及與此相關之問題——孰爲最佳之加權方法兩者,自來爲討論之中心問題.假定批發物價指數之目的,在計量一般物價變動之平均比率,則關於第一問題在理論上最佳之解決,前已言之;惟加權問題則爲一新因素.

編製指數時所以須用加權法者,其理由前已指出:指數所計量者,不論其爲生產量.貿易量或物價之變動,其所根據之各種價目與比率,重要性並非相等者.編製生產指數時,煤之重要勝於鉛,編製批發物價指數時,小麥價格之重要勝於亞麻仁.在一個指數中,包含有各種不同之商品,計量此種種商品之相對重要性之問題,其解決之法有多種.大致權數之選擇,一方面決定於編製該指數之特殊目的,一方面決定於獲得決定權數時所須之材料時所有之困難.玆舉若干例以表示各種權數,然後繼以較深刻之討論.

[用於編製指數之各種方法]

過去與現在編製指數之法有多種.其所以不同原因甚多.蓋諸法中在理論上以何者最佳,迄無一致之意見.且從事於此種工作者,才力不齊,而此中必須克服之若干實際困難,使應付上必然發生差別;最後,編製指數之目的不同,亦召致若干差別之原因也.

編製指數時各種實際方法之差別,及所得結果之差別,欲加說明,最有效之法,或爲取同一之材料,而應用此種種方

法求其指數下表爲1910—1923年十二種主要農產物在十二月一日之平均農場價格,(註)茲取以爲原料.

表四十四

美國1910—1923年十二種主要農產物在十二月一日之平均農場價格

農產品名	單位	1910	1911	1912	1913	1914	1915	1916
玉蜀黍	嘝	$ .480	$ .618	$ .457	$ .691	$ .644	$ .575	$ .889
棉花	磅	.141	.088	.119	.122	.068	.113	.196
乾草	短噸	12.14	14.29	11.79	12.43	11.12	10.63	11.22
小麥	嘝	.883	.874	.760	.799	.986	.919	1.603
雀麥	嘝	.344	.450	.319	.392	.438	.361	.524
白番薯	嘝	.557	.799	.505	.687	.487	.017	1.461
棓	磅	.0393	.0494	.0405	.0354	.0392	.0520	.0569
大麥	嘝	.578	.889	.505	.537	.543	.516	.881
菸草	磅	.093	.094	.108	.128	.098	.091	.147
亞麻仁	嘝	2.317	1.821	1.147	1.199	1.260	1.740	2.486
稞麥	嘝	.715	.832	.663	.634	.865	.834	1.221
米	嘝	.678	.797	.935	.858	.924	.906	.889

農產品名	單位	1917	1918	1919	1920	1921	1922	1923
玉蜀黍	嘝	$ 1.279	$ 1.365	$ 1.347	$ .677	$ .423	$ .657	$ .727
棉花	磅	.277	.276	.356	.140	.162	.238	.310
乾草	短噸	17.09	20.13	19.55	16.72	12.11	12.11	14.07
小麥	嘝	2.008	2.042	2.151	1.443	.926	.926	.923
雀麥	嘝	.666	.709	.715	.472	.302	.302	.415
白番薯	嘝	1.228	1.193	1.606	1.164	1.101	1.101	.823
棓	磅	.0672	.0728	.0728	.0577	.0370	.0370	.0747
大麥	嘝	1.137	.917	1.210	.707	.419	.419	.540
菸草	磅	.240	.280	.390	.211	.199	.199	.203
亞麻仁	嘝	2.966	3.401	4.383	1.766	1.451	1.451	2.108
稞麥	嘝	1.660	1.516	1.345	1.278	697	.692	.647
米	嘝	1.896	1.918	2.668	1.189	.952	.934	1.103

(註) 1910—20年之數字係潘桑(W. M. Person)氏所編;參觀"費暄氏之指數公式"一文,載Review of Economic Statistics Prel. vol III, 105. 1921—23年之數字根據惠萊氏"農産與市場"一書,各年棓價乃十二月租棓批發價格,因無相當之農場價格也.

[符號說明]

計算各種不同之指數時所用符號,其意義列後:

p_0': "0" 時(基期)某一商品之價格.

q_0': "0" 時同一商品之量.

p_1': 計算期"1"時同一商品之價格.

q_1': 計算期"1"時同一商品之量.

p_0'': "0" 時第二種商品之價格.

q_0'': "0" 時第二種商品之量.

p_1'': 計算期"1"時第二種商品之價格.

q_1'': 計算期"1"時第二種商品之量.

$\frac{p_1'}{p_0'}$: 價比(某一商品在計算期"1"時之價格對於同一商品在"0"時之價格之關係)

$\frac{q_1'}{q_0'}$: 量比.

P_0: "0" 時之物價水平.

P_1: 計算期"1"時之物價水平.

[簡單物價指數]

費暄氏在其對於物價指數各種編製法之詳盡的分析中(書名"物價指數作法"(The Making of Index Numbers),1922年 Houghton Mifflin Co. 出版)分別爲六個基本形式:卽總值式(aggregative,或 price aggregate),算術式(arithmetic),倒數式(harmonic),幾何式(geometric),中位數式(median),與衆數式(mode)是也.最後一種從未實際應用,可置不論.關於其餘五種形式之特質,可先論其最簡單之形式;次及較複雜之結合.

[實際物價之總值]

編製簡單總值式指數時,將屬於某一日期之物價相加;一般的物價變動,乃以如此算出之各不同日期之結果相比較而計量之.用上列之符號可得公式如下:

$$\frac{P_1}{P_0}=\frac{\Sigma p_1}{\Sigma p_0}$$

根據此法計算表四十四之指數,得結果如下,第二行爲實際總值;第三行乃根據1910年之總值爲基數將各年實際總值化成之比值,以便於比較.

表四十五

農產農場價格之指數

(實際價格之總值)

(1) 年份	(3) 指數 (實際物價總值)	(3) 指數,價比 (1910=100)
1910	$16.9653	100
1911	21.5814	114
1912	17.3785	92
1913	18.5124	98
1914	17.4722	92
1915	17.3540	91
1916	21.5739	113
1917	30.5142	161
1918	33.8198	178
1919	35.7938	189
1920	25.8247	133
1921	18.7790	99
1922	20.0230	106
1923	21.9437	116

吾人不久將以應用此法編製物價指數之結果,與應用其他方法由此同一材料求得之結果相互比較.此種形式之指數其主要缺點顯而易見.蓋彼旣非未權指數,亦非平權指數(Equally weighted index).每種商品對於指數之影響,爲其每單位價格所決定;然此貿易中此各商品究以何者爲單位,全係偶然之事.在本例之指數中,乾草因以噸爲單位之故,致所有其餘之十一種商品全體所得之權,猶不及其一項所得之權之爲大.次於乾草者爲亞麻仁,其所得之權亦較其他各項特大.故以如此之各項價目相加而得之指數,加權全不合理,吾人不能視之爲農產物農場價格之眞實的反映也.

有人曾用一種方法以避免因各種商品在貿易中所用單位不同而引起之不平均的加權,此法卽將一切價目先化爲同一之單位.如乾草,米,玉蜀黍,棉花及其他一切商品,先以磅爲單位開列價目,然後將此種價目相加而求指數.編製勃拉特斯忒利氏指數時曾用之.但此種加權法之不合理實與前法相同.吾人果欲作平均之加權,則此法並未能予一切商品以平均之權.如1910年乾草之價爲每磅$.00607,棉花爲每磅$.141,米爲每磅$.015,則在每磅價格之總值中,棉花所得之權,實9倍於米,23倍於乾草.

[價比之算術平均數]

編製指數之另一種方法,爲先將一商品在某一日期之價格作基數,將該商品之每一價目一律化爲比值,然後用任

何一種慣用之方法求諸比值之平均數.下例說明此種方法之第一種形態,用兩年之材料,以1910年爲基年.

表四十六

計算價比編製指數之法

(1) 商品	(2) 單位	(3) 物價 (1910)	(4) 價比	(5) 物價 (1911)	(6) 價比
玉蜀黍	斛	$ .480	100	.618	128.8
棉花	磅	.141	100	.088	62.4
乾草	短噸	12.14	100	14.29	117.7
小麥	斛	.883	100	.974	99.0
雀麥	斛	.344	100	.450	130.9
馬鈴薯	斛	.557	100	.799	143.5
糖	磅	.0393	100	.0494	125.6
大麥	斛	.578	100	.869	150.2
煙草	磅	.093	100	.094	101.1
亞麻仁	斛	2.317	100	1.821	78.7
裸麥	斛	.715	100	.832	116.2
米	斛	.678	100	.797	117.5
			1200		1371.6

由此諸數極易求出兩年價比之算術平均數.每一價比之公式爲$\frac{p_1'}{p_0'}$若有N個價比,在計算期"1"時之指數公式爲

$$\frac{\Sigma\left(\frac{p_1'}{p_0'}\right)}{N}$$

在此地卽爲

$$指數(1910年)=\frac{1200}{12}=100$$

$$指數(1911年)=\frac{1371.6}{12}=114.3$$

由1910年至1923年各年之指數用此法計算而得者列於表四十九第(3)行中.

[價比之簡單平均數中暗中包含權數]

此種形式之指數常稱爲價比之未權指數.然正與前兩例所說明之形式相同,此中亦無嘗不加權.每一項用爲權數之數量,實爲每一商品在基年售價$100之量,如上例之權數爲下列諸數量:

玉蜀黍	208.3	嘝
棉花	710.0	磅
乾草	8.24	噸
小麥	113.3	嘝
雀麥	291.0	嘝
馬鈴薯	180.0	嘝
糖	2650.0	磅
大麥	173.2	嘝
煙草	1076.0	磅
麻仁	43.2	嘝
稞麥	140.0	嘝
米	147.7	嘝

計算價比之簡單平均數時,實際上先決定上列諸量在十一年中各年可售之總值.在1910年,上列諸量各售$100,故總值爲$1,200;在1911年則上列諸值可售之總值爲$1371.60.

此種總值爲12所除,得表四十九第(3)行諸指數:即1910年爲100;1911年爲114(114.3),等等.於是所謂"價比之未權平均數"者,實爲實際價格之加權平均數.用爲權數之每一商品之量,其值在基年1910年時皆爲$100;在此意義上爲平均的加權.(註)

[價比之中位數]

求每年價比之平均數時,可不用算術中數而用中位數代之.將表四十六第(6)行中諸價比按大小爲序而排列之,可得下列之分配:

62.4	117.7
78.7	125.6
99.0	128.8
101.1	130.9
116.2	143.5
117.5	150.2

價比之最小者爲62.4,最大者爲150.2;中位數之值爲117.6.吾人即以此中位數爲1911年之指數.如此以價比之中位數代表之各年指數,列於表四十九之第(4)行.

[價比之幾何平均數]

茲計算每年諸價比之幾何平均數,並將所得結果與以前諸法所得之結果比較之.單獨之一個價比爲$\frac{p_1'}{p_0'}$,故N個

(註)價比之簡單平均數之此種特質,麥柯萊(F. R. Macaulay)氏曾特別指出之.參考American Economic Review.1925年十二月號,P.928.

價比之幾何中數其公式如下：

$$mg=\sqrt[N]{\frac{p_1'}{p_0'}\times\frac{p_1''}{p_0''}\times\frac{p_1'''}{p_0'''}\times\cdots\cdots}$$

計算幾何中數時平常皆用對數以利其計算；在此地則

$$\log mg=\frac{\log\left(\frac{p_1'}{p_0'}\right)+\log\left(\frac{p_1''}{p_0''}\right)+\log\left(\frac{p_1'''}{p_0'''}\right)+\cdots\cdots}{N}$$

茲取1910年及1911年之材料計算其幾何中數以說明其方法.下表中各商品之價比乃由表四十六轉載而來者.

表四十七

價比之幾何平均數之計算法

(1) 商品	(2) 價比 1910	(3) (2)行諸數之對數	(4) 價比 1911	(5) (4)行諸數之對數
玉蜀黍	100	2.0	128.8	2.10992
棉花	100	2.0	62.4	1.79518
乾草	100	2.0	117.7	2.07078
小麥	100	2.0	99.0	1.99564
雀麥	100	2.0	130.9	2.11694
白番薯	100	2.0	143.5	2.15685
糖	100	2.0	125.6	2.09899
大麥	100	2.0	150.2	2.17667
烟草	100	2.0	101.1	2.00475
亞麻仁	100	2.0	78.7	1.89597
裸麥	100	2.0	116.2	2.06521
米	100	2.0	117.5	2.07004
		24.0		24.55694

$$\text{Log } Mg\ (1910)=\frac{24}{12}=2$$

$Mg=2$ 之逆對數 $=100$

$$\text{Log } Mg\ (1911)=\frac{24.55694}{12}=2.04641$$

$Mg=2.04641$ 之逆對數 $=111.3.$

此111.3即1911年之指數.由此法求得各年之結果列於表四十九第(5)行中.

[價比之倒數平均數]

倒數平均數之特質前文曾加討論,當猶記倒數中數之倒數,即其所平均各值之倒數之算術中數.在此處,其包含之各項爲$\frac{p_1'}{p_0'}$式之價比.此種價比之倒數爲$\frac{p_0'}{p_1'}$.求N個價比之倒數中數,其公式爲:

$$\frac{1}{H}=\frac{\frac{p_0'}{p_1'}+\frac{p_0''}{p_1''}+\frac{p_0'''}{p_1'''}+\cdots\cdots}{N}$$

或

$$H=\frac{N}{\Sigma\left(\frac{p_0}{p_1}\right)}$$

計算之法以下表說明之:

表四十八

價比之倒數平均數之計算法

(1) 商品	(2) 價比 1910	(3) (2)行諸數之倒數	(4) 價比 1911	(5) (4)行諸數之倒數
玉蜀黍	100	.01	128.8	.007763975
棉花	100	.01	62.4	.01602564
乾草	100	.01	117.7	.008496177
小麥	100	.01	99.0	.010101010
雀麥	100	.01	130.9	.007639419
白番薯	100	.01	143.5	.006968641
豬	100	.01	125.6	.007961783
大麥	100	.01	150.2	.006657790
煙草	100	.01	101.1	.009891197
亞麻仁	100	.01	78.7	.01270648
稞麥	100	.01	116.2	.008605852
米	100	.01	117.5	.008510638
		.12		.111328602

$$H(1910)=\frac{12}{.12}=100$$

$$H(1911)=\frac{12}{.111328602}=107.8$$

用此法求得之各年指數列於表四十九第(6)項內.

編製上述之五種形式之指數,未曾用一合理的加權方法.每個皆稱爲"未權"平均數,此名頗易引起誤會.第一種根據實際物價總值而編製之指數,實爲曾加重權之指數,然其權數爲不合理者.至於其他四種,其權數爲1910年可用$100購買之量.今可集五種結果而比較之指數各計至整數個位爲止.

表四十九

1910—1923年農產物農場價格之指數

(1910=100)

(1) 年份	(2) 實際價格之總值(化為比值)	(3) 價比之算術平均數	(4) 價比之中位數	(5) 價比之幾何中數	(6) 價比之倒數平均數
1910	100	100	100	100	100
1911	114	114	118	111	108
1912	92	95	93	92	90
1913	98	104	98	100	97
1914	92	101	102	97	91
1915	91	104	104	102	101
1916	113	156	152	151	147
1917	161	208	206	204	198
1918	178	215	209	210	205
1919	189	252	236	241	231
1920	136	151	143	145	139
1921	99	115	99	107	101
1922	103	129	114	124	119
1923	116	142	134	135	129

此表中各種指數繪圖如下

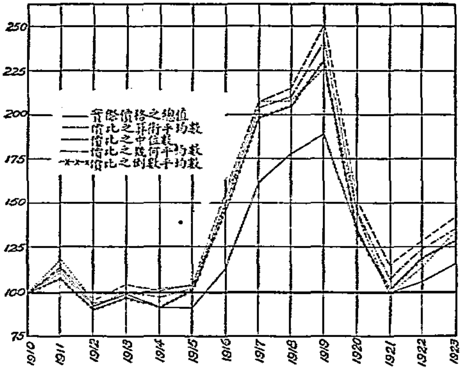


圖51.——1910—1923年農產物價之五種簡單指數之比較(1910=100).

[幾種簡單指數之比較]

價比之四種平均數,彼此之間關係較爲密切,而與總值式指數間則甚疏遠.總值式指數之不適於作爲物價變動之計量,其理由前已言之矣.至於其他四種指數中,算術式指數,幾何式指數,與倒數式指數之間,有一固定之關係存在,此乃由於三種平均數之性質而來者.除基年外,幾何式指數常大於算術式指數,而倒數式指數常小於幾何式指數;而彼此相差之數,則隨價格之離中趨勢之增大而俱大.中位數式指數

因此地所平均者僅十二項,故帶有若干不穩定之性質,其對於其他平均數之關係,亦不能始終如一.

在此不同之種種結果中應選取何種乎?此數種"未權指數",無一能稱爲完全者.因其暗中所用之權數,實未能計量其所包括之各種商品之相對的重要性也.現今且置加權問題不論.吾人能否測驗各種指數方法對於計量上列物價變動之適當性乎?

[時間互換測驗法]

爲達到此種目的,費暄氏所應用者即所謂"時間互換測驗法"(time reversal test).此法不過測驗一種指數方法在時間上能否向前向後雙方應用.如自1910年至1911年,糖之價格由每磅四分漲至每磅八分,則1911年之價格將爲1910年之200%.而1910年之價格將爲1911年價格之50%.一數各爲他數之倒數;兩數相乘之積(2.00×.50)等於一.同樣,若用某一種編製指數之方法求得某年之物價總水平爲其前一年物價總水平之200%,則用此法在時間上倒算之亦應得正確之結果;此結果即第一年之物價水平應爲第二年物價水平之50%.當用同一方法處理兩年之材料,以兩年交替爲基年時,則所得兩指數應互爲倒數.兩指數之乘積應常等於一.若結果不然,則此方法中必有固有之偏誤在.

試將此測驗應用於上述諸方法,取1910與1911兩年之物價,以1910年爲基年,則結果如下:

年份	實際價格之總值(化爲比值)	價比之算術平均數	價比之中位數	價比之幾何平均數	價比之倒數平均數
1910	100	100	100	100	100
1911	113.79414	114.3	117.6308	111.3	107.8

以1911年爲基年,則結果如下:

年份	實際價格之總值(化爲比值)	價比之算術平均數	價比之中位數	價比之幾何平均數	價比之倒數平均數
1910	87.87799	92.76	85.0177	89.85	87.47
1911	100	100	100	100	100

將1911年之各指數與相當之1910年之各指數兩兩相乘,可得下表諸值.(相乘時指數所取之形式爲比之形式,非百分比之形式.)

實際價格之總值	價比之算術平均數	價比之中位數	價比之幾何平均數	價比之倒數平均數
1.00	1.0602	1.00	1.00	.9429

合於此時間互換測驗法者有三種.算術平均數與倒數平均數皆不合.前者有顯著之向上偏誤(upward bias),因其

1910年與1911年之錯誤兩相重疊,向上偏誤之量達百分之六以上,而倒數平均數則向下偏誤(downward bias),其量幾與前者相等.若無法矯正此固有之偏誤,則此種平均數不應用之於指數之編製.

〔指數之加權〕

上節已敍述五種簡單之指數.引用加權法後,可能之結合其數大增,惟此地與吾人有關者僅有數種.

欲編製物價變動之精確計量,必須用合理的權數,此權數必須正確反映所包括各商品之相對的重要性.若忽略加權之問題,則不論吾人之覺察與否,必然有胡亂而不合理之權數暗存乎其中.

前舉諸例所用之材料,亦可用以說明加權方法,表示各種權數對於指數數值之影響.用於編製農產品農場價格指數時所用之權數,或為農產物生產之量,或為其值,各依其所選之指數之種類而定.1910—1923年各年之產量列於下表:(註)

(註) 1910—1919年之材料取自旦氏(E. E. Day)"An Index of the Physical Volume of Production"(轉載自Review of Economic Statistics, 1920年九月號); 1920—1923年取自 Weather's Crops and Markets.

表五十

1910—1923年十二種農產品每年產量

年份	玉蜀黍(百萬蒲)	棉花(百萬包)	乾草(百萬短噸)	小麥(百萬蒲)	燕麥(百萬蒲)	白番薯(百萬蒲)	糖(百萬磅)	大麥(百萬蒲)	菸草(百萬磅)	亞麻仁(百萬蒲)	裸麥(百萬蒲)	米(百萬蒲)
1910	2886	11.61	69.38	635.1	1186	.49.0	3318	173.8	1103	12.72	34.90	21.51
1911	2531	15.69	54.92	621.3	922	.92.7	4296	160.2	905	19.37	33.12	22.93
1912	3125	13.70	72.69	730.3	1418	420.6	3764	223.8	963	28.07	35.66	25.05
1913	2447	14.16	6 .12	763.4	1122	331.5	3941	178.2	854	17.85	41.38	25.74
1914	2673	16.14	70.07	891.0	1141	409.9	4134	195.0	1035	13.75	42.78	23.65
1915	2 95	11.19	85.92	1025.8	1549	359.7	4230	228.8	1062	14.03	54.05	28.95
1916	2567	11.45	91.19	636.3	1252	287.0	4671	182.3	1153	14.30	48.86	40.86
1917	3065	11.30	88.31	6 6.7	1593	442.1	3950	211.8	1249	9.16	62.93	37.74
1918	2503	12.04	76.66	921.4	1538	411.9	4220	256.2	1439	13.37	91.04	38.61
1919	2917	11.42	91.33	941.0	1248	357.9	3634	165.7	1390	8.92	88.46	41.06
1920	3209	13.44	87.85	833.0	1497	403.3	4065	189.3	1580	10.77	60.5	52.07
1921	306	7.95	82.38	841.9	1078	361.7	……	154.9	1070	8.03	61.7	37.61
1922	2906	9.76	95.88	857.6	1216	453.4	…	182.1	1247	10.37	103.4	41.40
1923	3054	10.08	84.90	785.7	1230	412.4	……	198.2	1475	17.43	63.0	33.26

(註₁) 每包毛重五百磅.

(註₂) 糖產量代表美國隣接諸邦各收穫年甜菜株與蔗糖總產量加上由不隣接諸邦該收穫年七月一日起始之合計年度內輸入之量. 1921—23年數字統計算時以此數列之估量值爲根據.

[實際價格之加權總值]

將實際價格相加而以總值爲指數,所得結果全不合理,此點已經前述.然在相加之前各價格若已經適當之加權,吾人卽不能同樣予以反對.吾人若以基年(在"0"時)生產之物量(q_0)作權數,則加權總值之公式爲:

$$\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0}$$

美國勞工統計局所用之方法實即此法,惟其權數之量非取自基年而為基年以外之一年耳.加權總值式指數之計算法可以下表說明之:

表五十一

實際價格加權總值之計算法

(1) 商品	(2) 單位	(3) 價格 1910	(4) 權數(1910年生產量,單位百萬)	(5) 價格×權數 (p_0q_0)	(6) 價格 1911	(7) 權數(1910年生產量,單位百萬)	(8) 價格×權數 (p_1q_0)
玉蜀黍	嘝	$ .480	2,886	$1,385,280,000	.618	2,886	$1,783,548,000
棉花	磅	.141	5,805	818,505,000	.088	5,805	510,840,000
乾草	噸	12.14	69.38	842,273,200	14.29	69.38	991,440,200
小麥	嘝	.883	635.1	560,793,300	.874	635.1	555,077,400
雀麥	嘝	.334	1,186	407,934,000	.450	1,186	533,700,000
白番薯	嘝	.557	349.0	194,393,000	.799	349.0	278,851,000
糖	磅	.0393	3,618	142,187,400	.0494	3,618	178,729,200
大麥	嘝	.578	173.8	100,456,400	.869	173.8	151,032,200
菸草	磅	.093	1,103	102,579,000	.094	1,103	103,682,000
亞麻仁	嘝	2.317	12.72	29,742,240	1.821	12.72	23,163,120
裸麥	嘝	.715	34.90	24,953,500	.832	34.90	29,036,800
米	嘝	.678	24.51	16,617,780	.797	24.51	19,534,470
				$1,625,494,820			$5,158,634,380

求得第(5),(8)兩行之總值,即可算出吾人所須之比值形式之指數.兩年中無論何年皆可作為基年,而以他年之價

格總值化爲其比值.如以1910年之總値爲基數,則1911年之指數爲111.5其他各年之指數用同樣方法求得者,皆列於表五十四第(2)行中.

更有另一種加權總值,其權數非取自基期而取自互相比較之兩時期中較後之一時期.即以計算期"1"時之價格與"0"時之價格互相比較時,取q_1("1"時之物量)爲權數;以計算期"2"時之價格與"0"時之價格相比較時,則以q_2("2"時之物量)爲權數是也.以代數式表示之計算期"1"時之指數爲:

$$\frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}.$$

計算之法,除權數逐年改變外,其餘一切皆如前例.用此法求得各年之指數列於表五十四第(3)行內.

此二法所用權數皆爲物量(Quantities),因以物量乘物價,則所得爲物價之總值也.惟加權於個別之價比時,不復能用物量.此抽象之比值必須以價值(values)加權,所得之乘積始可相比.蓋價值共以圓爲單位,而物量之單位則千差萬殊也,用爲權數之價值可以數法求得之.

費暄氏("指數作法"第54頁)列舉下列四法,其中第二第三兩種形式爲混合形式.

I.　每一權數＝基年物價×基年物量(p_0q_0)

II.　每一權數＝基年物價×計算年物量(p_0q_1)

III.　每一權數＝計算年物價×基年物量(p_1q_0)

IV. 每一權數＝計算年物價×計算年物量(p_1q_1)

正如某些平均數式指數之有固有的偏誤(即型偏誤 Type bias),價值加權法之每一種無不有其顯著之權偏誤(Weight bias).(此種固有的偏誤於物量加權法中不存在.)I,II兩種加權法中,所用者爲基年物價,結果生向下偏誤.III,IV兩種加權方法中,所用者爲計算年物價,結果生向上偏誤.此種情形一部分可由數學方法證明,一部分則曾爲無數的試驗所確證(註)

(註) 以第III種方法加權之指數必然大於第I種方法加權而得之指數.用第III式對某一商品之價比加權時,其式爲

$$\frac{p_1}{p_0}\times p_1q_0$$

而用第I式時,其式爲

$$\frac{p_1}{p_0}\times p_0q_0$$

若p_1大於p_0(若價比大於100),則第III式之權數(p_1q_0)大於第I式之權數(p_0q_0).即所有各價比之高於100者,其由第III式所得之權較重於由第I式所得者.惟p_1若小於p_0,則第III式之權數(p_1q_0)小於第I式之權數(p_0q_0).即所有各價比之低於100者,其由第III式所得之權輕於由第I式所得者.於是第III式對於一切漲價之影響過於加重,而對於跌價之影響則表現不足,結果經常大於第I式.以第IV式與第II式相比較情形亦同.至於第I第IV兩式之間,無必然關係,惟一般言之,以第IV式加權之指數,常大於以第I式加權之指數,基年加權法生向上偏誤,而計算年加權法則生向下偏誤.(欲知權偏誤之詳細討論,請讀費暄氏"指數作法"第五章第384—387頁.)

以下之若干例中,吾人所處理者僅限於基年(1910)所生產之物量之價値,此種價値悉列於表五十二第(3)行中,計算直至百萬爲止以應加權之須要.

〔價比之加權算術平均數〕

計算此種形式之指數時,每一價比各以一適當之權數乘之.乘積之總和復以權數之總和除之.此法以下表說明之:

表五十二

價比之加權算術平均數之計算法

(1) 商品	(2) 價比 1910	(3) 權數	(4) 價比×權數	(5) 價比 1911	(6) 權數	(7) 價比×權數
玉蜀黍	100	$1,385	$138,500	128.8	$1,385	$178,388.0
棉花	100	819	81,900	62.4	819	51,105.6
乾草	100	842	84,200	117.7	842	99,103.4
小麥	100	561	56,100	99.0	561	55,539.0
燕麥	100	408	40,800	130.0	408	53,407.2
白番薯	100	194	19,400	143.5	194	27,839.0
糖	100	142	14,200	125.6	142	17,835.2
大麥	100	100	10,000	150.2	100	15,020.0
菸草	100	103	10,300	101.1	103	10,413.3
亞麻仁	100	29	2,900	78.7	29	2,282.3
裸麥	100	25	2,500	116.2	25	2,905.0
米	100	17	1,700	117.5	17	1,997.5
合		$4,625	$462,500		$4,625	$515,835.5

(權數爲1910年生產之物量之價値,以百萬爲單位)

$$加權算術中數(1910)=\frac{\$462,500}{\$4,625}=100$$

$$加權算術中數(1911)=\frac{\$515,835.5}{\$4,625}=111.5.$$

此處求得之1911年之指數,其值與從表五十一計算得者相同.彼處之指數爲實際價格之加權總值,所用之權數爲基年所生產之物量.凡價比之以基年價值加權者,其算術中數常與此種總值化成之比值相等.(註)

〔價比之加權幾何平均數〕

計算加權幾何中數之法,與計算未權幾何中數之法同,所異者,其每一價比之對數須被乘於一定之權數,而此加權對數之總和爲權數之總和所除,如此求得之結果爲所求之指數之對數.此法以下表示之:

(註) 此點可用代數方法立即證明之.任何商品之基年價值,爲 p_0q_0,而次年之價比爲 $\frac{p_1}{p_0}$.此種價比之加權算術中數等於

$$\frac{\frac{p_1'}{p_0'}\times q_0'p_0'+\frac{p_1''}{p_0''}\times p_0''q_0''+\frac{p_1'''}{p_0'''}\times p_0'''q_0'''+\cdots\cdots}{p_0'q_0'+p_0''q_0''+p_0'''q_0'''+\cdots\cdots}$$

化成

$$\frac{\Sigma p_1q_0}{\Sigma p_0q_0}$$

此即加權總值式指數之公式.

倒數式指數以次年之全值作權數者可以同法化爲

$$\frac{\Sigma p_1q_1}{\Sigma p_0q_1}$$

此式即實際價格加權總值之以二年物量爲權數者,此指數式吾人前已見之矣.(第207頁)

表五十三

1911年價比之加權幾何平均數之計算法

(1910=100)

商品	價比 1911	價比之對數	權數 (1910生產價值)	價比之對數×權數
玉蜀黍	128.8	2.10992	1,385	2922.23920
棉花	62.4	1.79518	819	1470.25242
乾草	117.7	2.07078	842	1743.59676
小麥	99.0	1.99564	561	1119.55404
雀麥	130.9	2.11694	408	863.71152
白番薯	143.5	2.15685	194	418.42890
棉	125.6	2.09899	142	298.05658
大麥	150.2	2.17667	100	217.66700
菸草	101.1	2.00475	103	206.48925
亞麻仁	78.7	1.89597	29	54.98313
裸麥	116.2	2.06521	25	51.63025
米	117.5	2.07004	17	35.19068
			4,625	9401.79973

$$\log Mg=\frac{9401.79973}{4,625}=2.032822$$

$$Mg=107.9$$

以1910為基年,則1911年之指數為107.9.用此法求得各年之指數數值列於表五十四第(5)行中,與其他方法求得之結果並列比較.

此三種指數吾人將如何判別其優劣乎?吾人首先可用

時間互換測驗法,卽前次應用於五種單簡指數之測驗者.測驗之結果,此三種指數無一能合格者.簡單之幾何式指數,前次能合於測驗之條件,然此處加權以後,竟與其他方法同樣不免於錯誤;加權法之應用使結果多得一種偏誤.若單以時間互換測驗爲標準,三者無一能使人滿意.於此吾人可試用費暄氏所發明之第二種基本測驗方法,卽所謂"因子互換測驗法"("Factor reversal test")者是也.

[因子互換測驗法]

任何商品在任何一年之總價值,當然爲其生產量與每單位價格之乘積,以代數表示之,等於$p_0'q_0'$.某年之總價值與其前一年總價值之比率爲$\frac{p_1'q_1'}{p_0'q_0'}$.若由第一年至第二年,價格與產量皆加倍,則價比爲200,量比(物量之比值)爲200價值之比爲400.卽第二年之總價值將爲第一年總價值之四倍.價值之比等於價比與量比之乘積.就一種商品言,此種關係極爲明顯.

吾人對於若干商品,若編一代表其由一年至次年之價格變動之指數,又編一代表由一年至次年產量變動之指數,則兩個指數之乘積依理應等於第二年總價值對於第一年總價值之比率.若乘積不等於此價值比率,則兩指數中至少有一個有錯誤.

請應用此測驗方法於第一種加權總值式指數$\left(\frac{\Sigma p_1q_0}{\Sigma p_0q_0}\right)$,

求物量指數時,只須交換公式中q與p之位置;其公式為

$$\frac{\Sigma q_1 p_0}{\Sigma q_0 p_0}$$

吾人所欲計量者旣僅為物量變動之結果,故同樣之價格因子(p_0)出現於分母與分子之中.代入已知之十二種農產物價值,乃得

$$\text{物量指數}(1911)(1910=100)=\frac{\$4,446,264,630}{\$4,625,494,820}.=.96125.$$

用百分數之形式,則生產量指數(1911)為96.125,(以1910年為基年).更用同樣公式$\left(\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0}\right)$求與之相當之物價指數,得111.5.兩指數之乘積為

$$.96125\times1.115=1.0718$$

卽若價格增高11.5%,同時物量減少3.875%.總價值將增加7.18%.

然價值比率為

$$\frac{\Sigma p_1 q_1}{\Sigma p_0 q_0}=\frac{\$4,748,718,320}{\$4,625,494,820}=1.02664,$$

彼此之間相差$4\frac{1}{2}$%,故此公式不合於因子互換測驗,未為美滿.

就第二種加權總值式指數而測驗之,用1910年為基年,則得1911年之數值如下:

$$物價指數=\frac{\Sigma p_1q_1}{\Sigma p_0q_1}=106.8.$$

$$物量指數=\frac{\Sigma q_1p_1}{\Sigma q_0p_1}=92.05.$$

$$乘積=.9205\times 1.068=.9831.$$

(求乘積時指數爲比率之形式,非百分數)

此地偏誤之方向相反,偏誤之量爲4%以上(1.02664−.9831=.04354).

加權幾何平均數同樣不能合於此基本的因子互換測驗.觀此總值式與幾何式兩者,可知原來在簡單形式時未有偏誤之指數,皆因加權而生偏誤.然欲能正確表現事實,吾人又不得不用權數.一切簡單指數與簡單指數之加權形式,又無一能同時合於此兩種基本測驗者.費暄教授曾試驗四十六個如此之公式,結果能適合於時間互換測驗者僅有四個(即簡單幾何式、中位數式、衆數式與總值式四種),其中能適合於因子互換測驗者竟無之.

["標準"指數]

欲解脫此種困難,可用交叉法(crossing process)將公式修正;即取偏誤方向相反之公式求其幾何平均數是也.費暄教授對於一切可能的公式全用此法試驗,結果得十三個公式能同時合符於兩種測驗.更以精確與計算便捷爲原則選取其一爲標準,此即"標準指數"(ideal index)之由來也.此標準指數爲上述之兩種總值式指數之幾何中數.公式爲

$$\sqrt{\frac{\Sigma p_1q_0}{\Sigma p_0q_0}\times\frac{\Sigma p_1q_1}{\Sigma p_0q_1}}$$

本例根據已得之結果可立卽求標準指數(1911)如下

$$標準指數=\sqrt{111.5\times106.8}$$
$$=109.14$$

此指數對於時間互換測驗法以及因子互換測驗法皆能適合.請先試前者.

$$物價指數,1911年(1910=100)=109.14$$
$$物價指數,1910年(1911=100)=\ 91.63$$
$$1.0914\times.9163=1.00$$

更以因子互換測驗法試1911年之材料(以1910爲基年),則得

$$物價指數=\sqrt{\frac{\Sigma p_1q_0}{\Sigma p_0q_0}\times\frac{\Sigma p_1q_1}{\Sigma p_0q_1}}=109.14.$$

$$物量指數=\sqrt{\frac{\Sigma q_1p_0}{\Sigma q_0p_0}\times\frac{\Sigma q_1p_1}{\Sigma q_0p_1}}=94.07.$$

$$價值指數=\frac{\Sigma p_1q_1}{\Sigma p_0q_0}=102.664.$$

物價指數與物量指數之乘積$=1.0914\times.9407=1.02668.$

自1910年至1923年每年之標準指數,與構成標準指數之兩種指數,以及以基年之價值爲權數而得之加權幾何式指數,皆列入下表並繪圖52以表示之:

表五十四

1910—1923年農產物農場價格諸加權指數之比較

(1) 年份	(2) 總值式 (以基年物量為權數)	(3) 總值式 (以計算年物量為權數)	(4) 標準指數 (2),(3)兩列指數之幾何中數	(5) 加權幾何平均數 (以基年價值為權數)
	$\frac{\Sigma p_1q_0}{\Sigma p_0q_0}$	$\frac{\Sigma p_1q_1}{\Sigma p_0q_1}$		
1910	100.0	100.0	100.0	100.0
1911	111.5	106.8	109.1	107.9
1912	94.4	93.7	94.0	94.0
1913	112.4	109.4	110.9	109.9
1914	102.8	98.9	100.8	97.0
1915	102.3	102.3	102.3	101.0
1916	156.8	151.3	154.0	151.2
1917	212.5	210.9	211.7	207.2
1918	224.2	220.3	222.2	219.3
1919	241.3	238.2	239.7	235.1
1920	139.9	137.9	138.9	137.2
1921	104.3	103.6	104.0	101.6
1922	131.6	127.9	129.7	128.9
1923	149.3	145.2	147.2	144.1

(註) 表中1910—1919兩列總值式指數之值為潘森氏所計算(載"費雪氏之指數公式"一文 Review of Economic Statistics, Prel, Vol. III 107).

各種簡單指數間,其值相差甚大,將各種加權指數相比較時卽不見有此種情形.重要之差別固仍存在,然無若干簡單指數之有越出常軌的動態也.

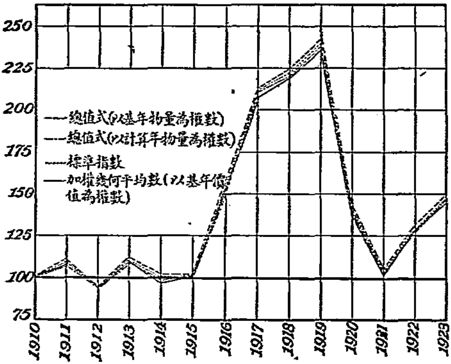


圖 52.——1910—1923 年農產物價格之四種加權指數之比較(1910=100)

在表中所列之四種指數中,最能正確計量 1910 年與每一計算年間之平均價格變動者,自爲"標準"指數.惟須注意者.此指數專爲計量兩個指定日期間之變動而設,非用以比較中間時期之變動者.如 1923 年之指數,其值爲 1910 與 1923 年之物價與物量間之關係所決定.其中包含兩種加權,而權數每年各各不同.若欲以 1923 年與 1922 年相比較,吾人必須

重新作一指數,而其中僅僅包含1923年與1922年之物價與物量.若直接以上表中"標準"指數各值互相比較,則因與編製此指數時所用之權數之性質不合,必致錯誤.

幾何式指數因其權數有一定,故每年之指數不僅可與基年指數直接相比,且可與其他各年之指數直接相比,此為其優點之一.其基期之位置,可直接根據比值而移動.如此求得之結果,與從原來材料直接求得者相同.若就同一材料求其"標準"指數以與之相比,不致有大出入,惟不能極端與確耳.

有兩種主要之障礙,使標準指數不被普遍採用.其一,求權數時須用每年或每月之產量,然此等材料常不易蒐集.其二,計算頗費時間;雖然,在須要十分精確之處,此點因不能視為一嚴重之困難也.費暄氏曾提出一公式以代替之,較標準指數公式簡便甚多.式為:

$$\frac{\Sigma(q_0+q_1)p_1}{\Sigma(q_0+q_1)p_1}$$

此式亦為愛奇渥斯(Edgeworth)氏與馬夏爾(Marshall)氏所推重,費暄氏則認為"蓋舉其結果之精確,計算之迅速,差誤之最小,形式之簡單四點而觀之,此為最佳最實際而面面俱到之公式."用此公式求得之結果,平常較"標準"公式求得者小$\frac{1}{4}$%.下表用1910年與1911年之材料,以說明其計算之

方法.

表五十五

以兩期物量總和作權數之加權總值式指數計算法

(1) 商品	(2) 單位	(3) 物價 1910	(4) 1910年物量＋1911年物量(以百萬為單位)	(5) 1910年 物價×物量之和 (3)行×(4)和	(6) 物價 1911	(7) 1911年 物價×物量之和 (6)行×(4)行
玉蜀黍	蒲	$ .480	5,417	$2,600,160,000	$ .618	$3,347,706,100
棉花	磅	.141	13,650	1,924,650,000	.088	1,201,200,000
乾草	噸	12.14	124.30	1,509,002,000	14.29	1,776,247,000
小麥	蒲	.883	1,265.4	1,109,401,200	.874	1,098,093,600
燕麥	蒲	.344	2,108	725,152,000	.450	948,600,000
白番薯	蒲	.557	641.7	357,426,900	.799	512,718,300
糖	磅	.0393	7,914	311,020,200	.0494	390,951,600
大麥	蒲	.578	334	193,052,000	.869	290,246,000
菸草	磅	.093	2,008	186,744,000	.094	188,752,000
亞麻仁	蒲	2.317	32.09	74,352,530	1.821	58,435,890
裸麥	蒲	.715	68.02	48,634,300	.832	56,592,640
米	蒲	.678	47.44	32,164,320	.797	37,809,680
				$9,071,759,450		$9,907,352,710

$$\frac{\Sigma(q_0+q_1)p_1}{\Sigma(q_0+q_1)p_0}=\frac{\$9,907,352,710}{\$9,071,759,450}$$

$=109.2$(以1910為基期之1911年指數)

(註) 若吾人之主要目的在作逐年之比較,則"標準"指數可用連鎖式編製之.先用每一年作其下一年之基期而求各年之指數,此種指數名為環比指數(link index numbers),諸環可就一固定之基期而"連鎖"之.潘森氏指出用此法所生之錯誤為累積的性質,許多年環比指數連鎖之結果,此錯誤可甚為嚴重.

(此地指數取百分數形式)

此公式所需材料,與"標準"公式所需者同,此種材料平常無從蒐集,廣博豐富之物量材料惟有逢全國總調查時期始可得到,其他諸年不得不應用固定之權數.在此種情形之下,加權總值式指數

$$\frac{\Sigma p_1q_0}{\Sigma p_0q_0}$$

或可謂為最普遍有用之形式.加權幾何式指數之優點甚多,惟犯有一種固定之權偏誤.若遇權數甚至權數之近似值皆無從決定時,簡單幾何式及簡單中位數式兩種指數,較之指數之其他簡單形式遙為優良.而幾何式較之中位數式尤為普遍適用.

[各種形式之指數之可靠性]

吾人由整個價格領域抽取一範樣(sample),根據對此範樣之研究而編製指數以為全領域之代表.若由同一領域繼續抽取其他範樣,編製而得之指數不免有所差別,是為抽樣變動(fluctuation of sampling).因此由各種形式求得之指數並非絕對穩定絕對可靠.其可靠之程度吾人思設法以計量之,各種指數中孰者因抽樣變動而起之差為最小乎(註1)?

凱萊(Iruman L. Kelley)氏(註2)曾計量各種主要之指數

(註1) 抽樣問題與統計計量之可靠性之關係詳後.

(註2) Statistical Method, 334—346 頁.

形式之機差,根據其在此方面成績之高下而按序排列之.加權幾何式與加權中位數式之兩種指數成績最佳,最爲可靠,最少受抽樣變動之影響.費暄氏之"標準"指數反見落後,然仍在價比之加權算術式與倒數式平均數之前.價比之簡單未權算術平均數成績最差.

凱萊氏有鑒於加權幾何中數之可靠性.伸縮性與其一般優點,選之爲物價指數之最佳形式.加權總值式

$$\frac{\Sigma(p_1w)}{\Sigma(p_0w)}$$

之應用精選之權數(不必恰爲市場上之物量或消費量)者,根據優良指數之基本要件觀之,與加權幾何中數可有同等之記錄,其成績且超"標準"指數而過之.權數所以不用實際物量者,求加權之能有伸縮性也.

[編製物價指數之其他問題]

前所述者,爲平均一批材料以求物價變動指數之技術問題.若干方法已證明爲錯誤,若干方法則各合於一定之目的.欲精明地應用指數,應清楚明瞭其編製時所用方法,俾可確知該數字所欲計量者爲何事,其可靠之程度爲如何也.

編製指數者所遇之問題,反應用指數者所應考慮之點,猶未盡於此也.與求平均數之問題及加權之問題同樣重要者,爲與選擇範樣有關之問題.一般物價水平之絕對確切之計量,惟有取在某一時期中一切流通貨幣單位(包含信用在

內)與同一時期中一切與貨幣交換之物量單位而定其比率,始能求得.故計量在兩個時期間之一般物價變動時,必須對於此兩因子在每個時期中之一切情形洞悉無遺.此當然爲不可能之事.故不得不用抽樣方法.而抽樣時主要之問題,爲決定範樣中應包括之商品之數目與種類,此等商品之價格乃物價指數之根據也.

[範樣中應包括之商品數目]

前曾提及方法與用途間之關係,在此地吾人又遇一同樣之問題.欲決定某一範樣中應包括之商品之數量及種類,必須問編製指數之目的爲何.假定指數之目的在計量物價水平之一般變動,則應包括商品多少之問題甚易答覆——曰範樣之規模越大,則結果越能代表眞相.根據大規模範樣而作之頻數多邊形,較之根據小規模範樣而作者,更接近於代表一切物價之理想曲線.美國勞工統計局之指數根據404個價目,勃拉特斯忒利氏之指數,根據96個價目,雖後一種指數自有其優長之點,然作爲一般物價變動之計量,寧視前一種指數爲較可靠者,雖然,指數之根據少數價目而作者,亦未必毫無價值.在物價系統與指數特性之研究上有極多貢獻之米契爾氏,曾取若干指數而仔細比較之,此等指數所根據之價目項數爲各各不同者;結果出人意外,蓋諸指數間彼此甚爲相似也.根據有限之價目求得之指數,在反映大量之物價變動上,與指數之根據數百種商品之價格而編製者,情形

極爲相似.惟在重要之細節上有差異,此種差異足使人關於某一年之物價變動發生懷疑.凡遇如此之情形,指數之根據多數價目而作者,設其所包括之商品能公平代表物價系統中之各種因素,則吾人必須認之爲較能忠實代表一般物價變動之數.

然爲他種目的,根據少數價目而作之指數或反受歡迎,"敏感"(sensitive)指數尤然,蓋此種指數乃用以預測一般的物價變動,非作爲一般物價水平之變動之確切計量也.紐約聯邦準備銀行所編之指數即屬此類,過去根據十二種基本商品(原料),之價目,今則根據二十種商品之價目.潘蓀氏所編批發物價指數之性質亦少許相似,其根據爲十個價格數列.此種情形中,目的爲求一表示商情循環(business cycle)之物價指數,計量產業興衰之趨向.欲求一敏感指數,與其廣包無數商品,反不如精選少數價格易於激變之商品.惟此種形式之指數用途有限,廣包大量商品之指數所表現之"惰性"(sluggishness),爲物價系統所固有,而爲代表一般物價之忠實指數所必須反映者也.

應包括之商品應有幾種?商品其價目足爲指數之根據者,其性質爲如何?此前一問題不能離後一問題而單獨討論.一指數之代表的性質,一部固依據其所包括之物價數列之多少而定,然此等數列之性質或更爲重要.當選擇商品價目時,須知各種不同之商品集團,其價格之動態彼此之間顯

有特殊之差異.此等價格集團與其彼此間相互之關係,其動態,以及其對於經濟制度之活動上對於繁榮蕭條之趨向上之關係,在經濟學者及實業家觀之,皆事之有迫切而實際之重要性者也.

〔物價系統之因素〕

物價系統內部之關係,錯綜複雜,欲以圖示之法表示其與實之概念,殆爲不可能之事.圖53僅就範圍較廣之關係與意義較重要之因素中取其若干而表示之耳,大別爲生產要素之價格.批發物價.另售物價及服務價格各門.最後一門包括以營利爲目的之企業與政府機關兩者之服務.私人服務

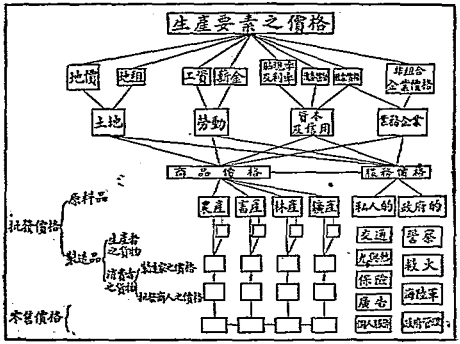


圖53.——價格系統中若干因素間之關係的圖示

之價格一部入批發物價;另一部分,當其在代表向最後消費者完盡之最後的服務之範圍內,入零售物價.每一門中,相關連諸價格.又各成重要集團.在第一部門(生產要素價格)中,地租與地價薪金與工資.貼現率與利率.債劵價格與股票價格.當漲跌之時,一切皆具特殊之動態.在商品價格部門,此種差別同樣存在.至於最後一門,則公私團體所盡各種之服務亦各有其特殊之價格變動者也.

物價系統內各單位各集團間之關係線,無從用圖示法全部表示.若此爲可能之事,吾人將見每一單位或集團各發出無數關係線連結於任何其他單位或集團.此種關係中有若干爲直接與重要者,若干則疏遠而爲實際上所難於捉摸者.

用一平面以表示價格系統亦非正當辦法.平面之圖,全部表現於一刹那間.而實際之情形,則與過去將來皆有無數之關係.向後視之,乃連結每一價格於過去價格之無數連繫,向前視之,連繫更多,蓋將來之價格與價格關係爲直接由現在之價格轉化而來者.例如製造品今日之價格,其受過去原料,勞動力等等之價格之影響,甚明顯也.

精確之貨幣購買力指數,或精確之一般物價水平變動指數,必須包括價格系統中所有如此一切門類之代表.惟現今發表之指數無一能如是者.若干批發物價數自名爲一般價格變動之計量,然吾人不能以此視之.

目前吾人所研究者爲編製批發物價指數之法,上文所述,所以明示此種指數眞正之範圍及意義,指示其所計量者,雖或爲價格系統中最重要之一因素,然亦僅止於一因素也.下文將特別就編製指數之問題,研究在價格系統之此一部門內若干互相關連的價格之不同集團.

[批發物價部門內之各種價格集團]

批發物價指數旣必須依據價目之範樣,此範樣必須足以代表全體,必須其所包括之商品之價格,足爲價格系統內種種因素之標本.爲此目的,因素之分類必須依據各不同集團價格變動之特殊性質.如此分別而得之集團中,最顯著者莫如代表各種不同產業之集團.紡織品之價格與鋼鐵之價格,皮革之價格與化學藥品之價格,其所受之影響各不相同,市況之衰落與好轉,並不同時影響於一切工業,影響之方式亦不同;故批發物價指數必須包括一切重要工業集團之價目.若某數種商品之價格對於指數之影響特別重大,此指數卽在此程度上失去其代表之資格.勃拉特斯忒利氏之指數,現已證明其過於重視棉織物,皮革與乾肉以致超過此等商品在貿易中實際上之重要性,此事雖不能損害其在若干方面之效用,然作爲一代表批發物價之指數,則不免減色也.

取1913年以來農產與家具之批發價格數而比較之,可知不同產業集團之商品,其價格變動間之差異程度爲何如

也.

欲求一指數能有眞正之代表資格,僅就一切工業取適當數量之代表納於範樣之中猶爲不足.原料與製造品之價格變動亦有顯明的差異,對於每一個此種集團,必須予以合宜之代表,一般言之,前者之價格反應市況較爲敏銳,其變動起於製造品之先,變動之程度較爲激烈,此事可以兩個相關之理由說明之.原料之買賣所以供給製造與銷售,市況經一度衰落而好轉時,消費者對於製造品須要之增加(或意料中之增加)使互相競爭之製造家競購原料,各出高價.須要之增加,其最先之壓力中心,爲原料市場,競買之結果,使此種市場之價格,在其他貨物之價格尙未受影響之前首先高漲(然此非一定不易之論,1919年之好轉卽爲例外)同樣,恐慌與破產時期之到來,先爲大量製造品之堆積,一見市況衰疲之最初徵象,對於原料之須要立卽減少.在原料市場上,較之製造品市場,單純之貿易力量較能自由活動,故此種市場價格之變動亦早於其他商品市場之價格.

製造品價格之所以較爲穩定,其第二理由爲此種價格中包含較多之固定的費用因子,卽普通費用(overhead charges)及工資是也.工資、利息、地租之變動,較商品價格遲緩,而範圍亦小.此種因子加入商品價格之中,足使價格趨於穩定.當商品由原料之階段走近完全製品之階段時,其價格中所包含之穩定因素日見增多,故其價格之變動亦較爲微弱.

(註1)製造品之不同各類間所以有重要差異者,其源卽在於此.

原料集團與製造品集團之內,又各可分爲若干小集團,其商品價值各具顯著之特性.原料集團之內,農產、畜產、林產、礦產四者,其價格變動之情形,相差甚大.農產受市面之影響,然亦受氣候及收成之影響;雖受某程度之價格變動,然其反映市面情況,不如礦產價格之正確.(註2)畜產與林產,就其價格變動反映市面狀况之眞實與否而言,似居於前二者之間.故在選擇原料納入價格之範樣以便根據而編製代表性質之指數時,對於此種種小集團必須與以一公平之權數.(註3)

再者,製造品在其價格變動上亦不成一致之集團.當其供給更進一步之加工,或用於其他物品之生產時,對於製造家競買之關係上,其地位與原料品相似,故其價格之變動比較爲大.若其需要爲供給最後之消費,則純粹市面之影響較小,而其價較格爲穩定,前面所述當商品費用中包含之工資,普通費用等穩定因素增大時價格亦日趨穩定之理,於此亦

(註1) 例見米契爾氏"指數之作法及其用法",(美國勞工統計局公報284期44—45頁).

(註2) 吾人不可因此推論謂農業生產、農產品價格、與一般之市面狀况間毫無關係.吾人有充分根據相信在農業狀況與市面狀況之間有一直接之因果關係存在.惟直接之價格關係時或取矛盾之形式耳.

(註3) 參觀"指數之作法以及用法"第47頁.

有關係.故爲編製批發物價指數的根據之價格範機,必須包括生產者之貨物消費者之貨物之代表,必須同樣包括在製造之中間階段與最後階段之貨物之代表.(註1)

價格系統尙有其他分類方法,惟就吾人目前之觀點言之,前舉者乃最爲重要.批發物價之代表指數,須以由前舉之各商品集團所取得之價目爲根據,視每一集團內各種商品在貿易中之相對重要性,決定各該集團之權數.

美國之批發物價指數(註2)

[美國勞工統計局之指數]

美國有權威之批發物價指數爲勞工統計局所編製者.開始於1902年,直至1913年皆取價比之未加權平均數形式,每一價比皆以1890—1899年十年間各該商品之平均價格爲基數.1914年之報告在方法上劃一新時代,編製指數發表時所用基期亦與前不同.就目前之情形言,任何一時期(年或月)之指數取實際價格之加權總值形式,總值各以1913年爲基期化爲比值以便比較.

該局發表之指數,現今爲根據404價格數列而編製.(同

(註1)　同上書45,49兩頁.

(註2)　關於此種指數之敍述與比較,詳載美國勞工統計局公報第284期"美國與外國之批發物價指數",該局經常出版之公報;"批發物價"上,對於該局批發物價指數之價格有詳細記載.

一種商品可用數個價目代表,表現不同之品質與不同之市場,如原棉之價目有二,一爲紐奧爾蘭市場上中等原棉之價目,一爲紐約市場上中等內地原棉之價目).每一商品價格用一固定之權數乘之,權數爲1919年市上各該商品之數量.

下表表示此法應用於棉花之情形:

商品	平均價格 1919 (每磅)	1919年上市數量	1919年平均價格×1919年上市數量
棉花,中等,紐奧爾蘭	$.319	3,806,921,000磅	$1,214,407,799
棉花中等,內地,紐約	.325	1,903,461,000磅	618,624,825

將404個價格數列全用此法算過後,求最後一行之總和,則得該時期之指數,如此地即1919年之指數.發表時此總和以1913年之總和爲100求出比值而表示之.此種編製之法,使基期可移動至任何一年或任何一月,而根據此新基期改編各比值爲其百分數.

故此指數係根據一單貨物之批發原價而編成.貨單不變,原價總值則因各種商品之價格變動而生變動,指數則計量此種個別商品變動對於原價總值之影響.全部商品更分爲九個集團,分別求其總值,除總表之指數外,更公佈各集團個別之指數.所分之商品集團如下:

農產物

食品

衣服與衣料

燃料

金屬與金屬製品

建築材料

化學藥品

家具

雜項

[聯邦準備局指數]

聯邦準備局所編美國批發物價指數有不同之兩組.第一組根據勞工統計局所編之材料而重加分類,使物價系統中若干因子更易確切研究,所用之法及權數與勞工統計局所用者同.下列為重加分類後之商品集團:

原料

　農產

　畜產

　林產

　礦產

生產者之貨物

消費者之貨物

此種分類之效用首為米契爾氏所證明;其意義已於前數頁中說明.分類指數最初發表於1918年十月之聯邦準備公報(Federal Reserve Bulletin),並有一表開列每一商品集團中之各件名稱.

聯邦準備局已開始編製主要商業國之批發物價指數以便比較國際之價格變動,因各國編製指數之方法與所包括商品之性質大有不同,難於比較也.現今此種國際性物價指數通常發表者有美國、加拿大、英國、法國及日本等國.

編製此種指數之方法,一般上與上述勞工統計局編製指數時所用方法相同.以適當權數乘每一商品之價格,求諸乘積之總和而得貨幣總值,以1913年之總值為100而化各總值為比值.此種指數與勞工統計局之指數差異之點,在其所包括之商品數目,其所用之編類方法,以及其所加之權數,兩相不同.

每一國際價格指數,根據90個至100個價目而作,代表70種不同商品.分類之方法有兩種.其一即前面已經說明者,分全部商品為原料、生產者貨物、與消費者貨物之三類,同樣之商品復分為"生產貨物"與"進口貨物"之兩類.最後,又編製一種"出口貨物"之指數,選取上述諸商品集團中在出口貿易上佔重要位置之商品,以其價格作為此指數之根據.此六種分類指數包括一切商品在內之總指數各分國別公佈.而對於美國以外各國之得編此指數者,更公佈一種指數以表示將一切價格改為金本位後之情形.

編製此種分類指數時所應用之權數,為1913年之生產量、進口量或出口量.編製代表"一切商品"之總指數時,"生產貨物"之貨幣總值與"進口貨物"之貨幣總值相加為一.

此種指數已編製者有1913年通年之指數,及1919年正月起直至現在之逐月之指數.(關於法國者起自1920年之正月.)(註)

[戰時工業局之指數]

美國現有之一切批發物價指數中蒐羅材料最廣者,爲戰時工業局物價處爲研究大戰時期物價而作之指數.此研究廣被1913—1918年六年長之時期.研究之結果詳細記載於戰時物價史(History of Prices During the War)一書,爲戰時工業局所出版,摘要載物價公報(Price Bulletin)第一期.

該局局員蒐集該時期中1474種商品之價格數列,其中應用於編製指數者有1366個.分爲五十個商品組及七個產業集團.五十個商品組,七個主要集團,"全體商品,"及組內更加細別之諸小組,各有其單獨之指數.

此種指數皆實際價格之加權總值以比值表示者,形式與勞工統計局所編者相同.所用權數爲1917年(歷年)美國國內之生產量及該年之進口量.某一商品組內各商品在某一月之實際價格乘以適當權數而總加之,所得之和即爲該組之貨幣總值.在綜合各組之總值而求七個主要集團或全體商品之指數時,每組各依其相對重要性而加權,其法與加權於個別商品時所用者相同.

(註) 關於此種指數之詳細記述,請讀聯邦準備局1924年出版之小冊子"Price in the United States and Abroad".

將總值化爲比值時所用之基期,選定爲大戰爆發前之十二個月,即自1913年七月至1914年六月之一時期.但基期可任意移動而無損於其正確性.

除此種政府機關所發表之指數外,復有數種爲私人機關所編製者.其中以勃拉特斯武利氏與鄧(Dun)氏之指數爲最有名.

[勃拉特斯武利氏之批發物價指數]

勃拉特斯武利氏之指數爲實際價格之總值.其編製之法爲取96種大宗商品之價格一律化爲每磅之價格,然後將此"每磅"價格相加而求其總數.不用加權方法.其編製每月指數時所根據者,爲每月初一日之價格,非如勞工統計局指數之取每月平均價格也.價格既不以比值表示,故無基期.此發表之價格總值可任取一年或一月爲基期而化成比值表示之.此種指數起自1892年以迄於今.

前文中曾指出一切指數無不爲加權者,不論此加權之爲自覺者與否也.勃拉特斯武利氏之指數對於某些商品組加以重權,1897年七月十日發表之一表中,焦炭之每磅價格爲$0.0007,酒精之每磅價格爲$0.34,羊毛爲$.50.由此可見將"每磅"價格相加時,後兩者所得之權較重於前者甚遠,潘蓀氏曾作下表以示勃拉特斯武利指數(1921年九月一日之指數)之結構.(註)

(註) 載 Review of Economic Statistics, Prel. Vol. III. 365—366.

表五十六

1921年九月一日之勃拉特斯次利指數之

權數之分配

商品組	權數(百分數分配)
織品(原料與製品)	
棉織物	16.7
羊毛(原料)	5.6
其他	3.3
食品	
牛奶,蛋,牛油,牛酪	8.6
乾製豬肉,牛肉,豬油	6.9
糧食	12.4
生熟皮貨	13.3
建築用料	1.5
煤及焦炭:	
煤	*
焦炭	*
金屬:	
鐵與鋼	.7
其他金屬	4.3
化學品及藥品	9.6
牲口	3.6
油類	4.0
菜類	3.2
製麪包原料	1.0
海軍儲備	.9
雜貨	4.2

(*不足.1%)

此權數分配中,棉織物、生熟皮貨,化學藥品及乾肉之地位特別重要,此指數之特質大部分爲此種加權方法所決定,

且其權數不固定,因價格之變動影響於權數之分配而使之亦起變動.如1920年二月一日織品之權數爲全體權數之34.46%,在其後之十八月間織品價格之跌落較諸一般物價大爲迅速.故1921年九月一日織品之權數,只合全體權數之25.58%.

勃拉特斯忒利氏指數之加權方法雖不合理,然自有其價值在,特別在用爲商情之氣壓表時,彼爲最有用之批發物價指數之一.

[鄧氏批發物價指數]

R. G.Dun公司之商業通信社逐月發表於Dun's Review上之批發物價指數,與勃拉特斯忒利氏之指數相似之點,在其皆以實際價格之總值表示而無基期.惟就編製之方法,權數之分配,以及所被之時期言之,兩指數固大有不同也.

鄧氏指數,爲以元角分表明若干主要商品對於每人一年之供給量之價值之報告.其每月之指數,爲以所包括之每一主要商品之價格,與該商品每人每年之平均消費量相乘而得.如1921年正月一日之指數爲$198.600,此數爲根據1921年正月一日之普通價格計算所包括之一切商品對於每人一年供給量的價值之結果.鄧氏指數以每月第一日之普通價格爲根據.編製時約用三百種商品之價格,然其確切數目與商品之名稱從未發表.應用之價目僅以生活必需品爲限,奢侈品除外.雖其權數根據每人之消費量而定,此指數並非

以計量生活費用爲目的.大宗商品之如銑鐵,煤,建築材料等等非爲個人所直接消費者亦包括在內.故此指數乃用以計量批發物價水平之一般變動者.最初發表於1901年,惟其計算直回溯至1860年.

在邏輯意義上言之,鄧氏用商品之每人平均消費量爲權數以加權於每一價格,誠爲妙法,以惜疑者此法在實際上應用之效力爲如何耳.在美國,除關於比較少數之商品外,關於商品消費之適當材料尚付缺如.且就發表之數字觀之,其加於食物之權數幾佔全體權數50%,吾人根據事實可知其不免太過.此對於食物之過度加權一點,足以說明鄧氏指數之若干特殊動態,說明其所以與其他一般性指數永不能密切相合之故.

[費暄氏批發物價星期指數]

費暄氏於1923年正月開創一種新的批發物價指數,每星期計算而發表.綿密之星期物價指數以此爲始.

此種指數發表爲價比形式,以1913年爲基期,惟此價比乃由實際物價之加權總值而化成.所用權數爲人口局所記載之1919年市上之商品量.故所用公式爲上述第一種加權總值式指數公式:

$$\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0}$$

僅其權數非取自基期而取自1919年耳.此法與勞工統計局

所用方法實際相同.費暄氏認爲此法乃材料所允許之範圍內最接近於"標準"指數者.

此指數於加權方法上顯示一有趣味之革新.此指數所根據者爲取自鄧氏評論之205種商品價格.對於最主要之各種商品,用1919年之實際數量爲權數,爲使各商品組之權數適當,亦略有修改,惟對於其他之每一種商品,用1000,100,10,1,.1等整位數爲權數以便於計算,各處皆取整位數之最接近(在幾何級數上最接近)於實際物量者.此法於未採用前,先經證明對於較不重要之商品若用成數爲權數時,結果之錯誤不出1%.

1923年之費暄氏指數,以1922年十一月之總值爲基,惟星期指數接合於勞工統計局之指數上.勞工統計局所編1922年十一月之指數爲156,計算1923年各指數時,某一星期之總值對於基期總值(1922年十一月)之比率被乘於156.目前(1924年)之基期總值爲1923年十一月之總值,但用同樣之接合法,可化指數爲以1913年爲基期之比值.

每星期之指數發表於下星期一之日報上,發表之迅速使其多一種價值,此指數乃目前編製之一般性的物價指數中最有用者之一也.

[潘菘氏商情循環物價指數]

上述之各種指數,雖其實在意義限於批發物價之範圍,然皆以計量一般物價水平之變動爲目的.尚有若干物價指

數爲專門目的而編製者,此種限於專門目的而作之指數,在將來或將大增,其中最有興味者爲潘蓀氏及柯爾(Eunice S. Coyle)氏所編製之"商情循環物價指數"(Commodity price index of business cycles)(註)

編製此指數之目的爲"計量相互交射之繁榮時期與蕭條時期中一般商情之變動",換言之,問題爲"選擇及結合批發物價之數列以求一商情循環之指數".商情循環者,首先爲一價格現象.其他因子固亦甚爲重要,惟一切因子皆藉其作用於價格系統之影響而表現;藉價格之變動,此循環始最直接的爲商人,雇員及消費者所洞悉.價格系統之各因素雖全部受到影響,然證券價格與批發物價循環變動(cyclical change)所受影響最爲直接與猛烈.潘蓀教授有志於編製一足以反映及計量直接有關於商情循環之物價變動之批發物價指數.故其問題與編製"用於一般目的"之指數者之問題根本不同,其所用之方法自亦相異.

決定此種指數之結構時,首先對於若干產業集團之分類指數及若干商品之個別價目作一測驗,以選擇價格數列中其變動與1903—1914年一時期內一般商情及價格之"標準循環變動"(typical cyclical fluctuations)相符合者.過去研究之結果,證明勞工統計局與勃拉特斯忒利氏之物價指數以及許多非價格數列(non-price series)在該時期中之循

(註) 參考 Review of Economic Statistics, Prel Vol. III. 353—369 頁.

環變動有大致共同之點,即1907,1910,1912三年皆爲波浪式變動之波頂所在.取此各種數列所共有之波浪式變動爲標準,測驗各種價格數列以摒除其中變動之不合於標準狀態者.

最初編製者爲各種產業集團之物價指數,此種指數以上法測驗之.結果證明礦業及鋼鐵、皮革、化學藥品、紡織品之製造業之物價指數表現標準循環變動;其變動與商情之循環趨勢相合符.農業、食品工業、以及石、泥、玻璃之製造集團,其變動不合標準;此種數列之價格變動與一般物價一般商情之趨勢不相合符.

研究且進一步而以同樣方法測驗個別商品之價格變動.一切價格數列其變動不合標準者以及價格數列之固定不變者,因其不合於現有之目的,一律摒除之.在變動不合標準之產業集團中發現若干數列之變動合於標準;且發現個別商品之價格較之整個商品集團之指數爲敏感易變,因此決定最後的商情循環物價指數之編製,不應以集團指數爲根據,而應以個別之價格數列爲根據.

最後選定參加於指數之編製之商品爲數僅十;包括棉子油、焦煤、鋅塊、銑鐵、棒鐵、上等豬肉、皮革、印花布、被單布及毛絲絨線.此種商品之價格變動甚大,反映商情變化極靈敏.且其所代表之產業在實業上亦屬重要.

蒐集此十種商品之價格直至1890年以前,然後編製所

求之"商情循環物價指數".每月每年之價格,以1890—99為基期而化為比值.(其後已將基期移至1919年).指數為此種價比之簡單幾何中數.

[其他批發物價指數]

紐約聯邦準備銀行編製之二十種主要商品之批發物價指數,雖不明白作為商情循環之指數,然其性質與前者大略相似.此指數為主要商品市場上價格變動之敏感指數.發表時取比值形式,以1913年為基期,載該銀行之月報(Monthly Review)上.業已編製者,僅限於自1913年至今日.

Annalist雜誌上每週發表一種食品批發價格指數,根據25種主要食品之價目而編製.將價目以1890—99年為基年化成比值,取簡單算術中數形式結合之,此指數之目的,僅限於計量一理論的家庭食品預算中所需要之食品在批發時之價格之變動.

另一種食品批發價格指數為紐約湯麥斯•吉伯生(Thomas Gibson)氏所發表於其每週商情通信中者.以22種主要食品之價格為根據.雖謂採用鄧氏之一般方法,然對於其所用之加權方法無詳細報告.發表之指數根據加權價比而作,各價比之基數為1890—99年一時期之平均價格.惟所得結果復以一固定因子乘之,此法原為使此指數與鄧氏指數相調和,而以前者為後者之續者也.

[指數與指數間之差異]

就上述諸批發物價指數取若干年而比較之,(註)可見在物價變動之一般方向上彼此符合之程度頗足驚人,惟在詳細之點亦有若干重要差異.如由一月至次月之價格水平,在一種指數表示上升,在另一種指數可表示不動甚或下降.此種差異尤其在物價變動劇烈之時期如1916—1921年者特別明顯.此種差異之原因,在討論編製指數之方法時已暗示及之,無論何時,此種矛盾之動態可歸源於下列原因之一種或數種:

甲、包括商品之數目不同.

乙、包括商品之種類不同.

丙、蒐集價目之方法不同.(如或爲契約價目,或爲公開市場之價目;或爲每月第一日之價目,或爲一月之平均價目).

丁、所用權數不同.

戊、求平均數之法不同.

[批發物價指數結論]

玆根據上述各種不同之點,對於各種批發物價指數之主要特質及用途綜述之如下:

1. 美國現今發表之各種指數中,用於一般目的之批發物價指數其最佳者當推勞工統計局所作者.惟其所計量者

(註) 米契爾氏在"指數之作法及其用法"一文中將美國之此三種主要指數作一比較.載勞工統計局公報284期第一部,第94—112頁.

非物價變動之平均比率,而爲一定數量貨物之貨幣價值之變動.

2. 對於1913—1918年之一時期,戰時工業局所作之指數,爲一般性之批發物價指數中最佳者.

3. 鄧氏指數固爲批發物價變動之良好計量,惟因其對於食品方面加權過重,且其作者不宣布所包括之商品爲何其所用權數爲何,故其價值爲之減色.

4. 勃拉特斯武利氏指數爲批發物價變動之良好計量,惟因其對於原料加權過重,有損於其作爲代表之價值.未用合理之權數亦爲相當重要之缺點,然因其過分加權於原料,作爲商情之氣壓表觀之,反勝於其他兩種主要指數.

5. 潘蓀氏指數非爲計量一般物價變動而作,用爲商情循環之指數甚爲適當,編製之目的本在於此.

6. 紐約聯邦準備銀行所作包括十二種商品之指數,爲計量基本原料價格變動之有用工具.

7. 聯邦準備局爲比較國際價格而作之指數,在便利各國價格變動之比較上爲一異常有用之作用.各國分別之指數爲各該國一般物價變動之良好計量.

8. 費暄氏星期批發物價指數,其根據之材料廣博而計算又精確,甚爲有用.其發表之迅速使其效用更爲增高.

9. Annalist 雜誌之食品批發價格指數價值有限.由其編製之法(求價比之簡單平均數)易得可疑之結果,基期(1890

—99)與計算期相隔過遠.

[其他物價指數]

用指數計量價格變動之法,不限於批發價格.此法尙有多種變形應用於計量其他領域之價格變動者.於此略就其中數種之性質稍加說明,當爲有益之事也.

[零售物價之指數]

美國勞工統計局經常出版一種零售食物價格指數.所用之方法與該局計算批發物價指數之方法大致相仿,僅因材料之性質不同不免有所差異耳.

根據之材料,爲全美國51個代表城市中在每月十五號時43種食品之實際零售價格,由商人處蒐集而來.雖各處之習慣有不同,儘可能使所蒐集之材料可以互相比較.將每一種食物之無數價格相加,其總和以報告此種食物之價格者之人數除之,即取得該種食物於全美國之平均價格.於是復根據該局所作各種預算研究中1918年之平均家庭消費量,視每一種食物在此消費量中之相對的重要性,決定該種食物之權數,將此權數乘平均價格,諸乘積相加即得實際價格之加權總值.此種求實際價格加權總值之方法與編製批發物價指數時所用之法相同.此零售價格之總值亦以1913年爲基期化爲比值,此種比值通常載於每月勞工評論(Monthly Labour Review)上.

對於51個城市,亦各各用此同樣手續求出每一城市之

另售食物價格指數而發表之.而全美國與每一城市之平均實際價格,更發表於不定期之零售價格公報上,此種辦法使勞工統計局之工作之價值大爲增加.

關於計算批發物價變動問題之內在的困難,前面已有較爲詳細之討論.欲編製零售物價指數之如上述者,所遇困難更多.舉凡編製批發物價指數時之一切理論問題旣須解決,加以選取適當權數,獲得眞確價格數字及可以互相比較之價目等實際困難,於此更爲嚴重.商品標準化(commodity standardization)之缺乏,商業慣例與地方習俗之差異,使編製零售物價指數之困難特爲重大.因此現今發表之零售物價指數,其可信之程度無一可比之最佳之批發物價指數者.

[生活費指數]

在編製零售物價指數時已爲嚴重問題者,在計量生活費時,其解決之難且加倍.欲就普通家庭預算所包括之各項如食物價格,房租,另售衣料價格,燃料費,另售家具價格,與其他一切雜項開支之價格而求其平均數,並作一指數以計量此各項費用之變動,首須克服無數統計上之困難.關於最適當之平均方法與加權方法之選擇之理論問題固屬其一部分,然最主要者乃如何蒐集正確而豐富之實際材料一問題.故生活費指數在其可靠性未經證明,比較現在已大有進步,應用時不得不特別謹愼也.

經常編製之生活費指數有二種,一爲勞工統計局所作,

一爲全國工業會議局(National Industrial Conference Board)(雇主聯合會之一機關)所作.前者在每月勞工評論上發表,後者則在全國工業會議局之刊物上發表.家庭開支之主要項目,皆依其在預算中之相對的重要性而加權,總結果用1913年之價値爲基數化爲比値而表示.(全國工業會議局之指數用1914年七月爲基期).

[農產物之價格及購買力之指數]

美國農業部編製之一組指數甚爲有用.其中有一種指數逐月計量主要農作物在農場上可售得之價格之變動.另一種指數計量家畜產物農場價格之變動.第三種指數合併前兩指數爲一,計量一切農產物之平均價格之變動.此三者皆用1913年爲基期作成比値形式而發表.

農產物價格指數之變動,其意義更因第四種農產物購買力指數(Index of purchasing power of farm products)之編製而愈爲明瞭.單獨一貨幣價格之指數意義甚少,因貨幣購買力經常在變動中也.故必須藉一指數,其目的爲計量農人所購之貨物之價格變動者,以說明貨幣價格指數之意義.此種指數由改組勞工統計局批發物價指數,排除其中包括之農產物與食物而得.農產物購買力指數之計算,可用1923年正月之數字爲說明.該月之農產物價指數爲116(1913=100).除去農產物與食物以外之商品批發價格指數爲170.此卽表示農人出售其生產物之所得較1913年增加16%,惟一般商品之

批發價格較1913年增加70%.故其購買力以1913年爲基期時等於$\frac{116}{170}\times 100$,或68.2%.

1923年各月之各種指數,其數值列於下表:

表五十七

商品價格指數

1923年農產物價格購買力指數(1913=100)

月份	在農場上之價格			*商品批發價格	†農產物購買力
	農作物(每月15號)	家畜(每月15號)	農作物與家畜		
正月	126	106	116	170	68
二月	130	107	118	172	69
三月	134	106	120	175	69
四月	139	107	123	176	70
五月	140	105	123	172	71
六月	139	100	120	168	71
七月	136	102	119	165	72
八月	136	102	119	163	73
九月	138	109	123	164	75
十月	139	103	121	161	75
十一月	137	97	117	160	73
十二月	137	94	116	155	73

*農產物及食物除外 † 藉其他生產物表示

應知上列之購買力指數,乃根據一重要假定而計算者,此假定卽農人所購之貨物,其價格變動,可以將勞工統計局批發物價指數改組而計算之也.然此點不無疑問,因農人平常購物非以批發價格,而所購之種類亦非恰爲勞工統計局指數所包含者,所以利用該指數者,當然因無更適當之指數可用之故,然對其可能發生之錯誤,吾人應加認識也.

此農業部之各種指數,已編製者有自1913年至現在之一時期內各月之指數.載該部出版之"收穫與市場"(Crops and Market)上.

[貨幣工資與實在工資之指數]

當工業中頻起波瀾之時,工資之變動特別爲人所注意.對於此問題之注意,雖因此種時期而加強,然此問題乃具有永久之重要性者.政府機關雇主聯合會工會及私人研究者曾作有許多之工資指數,對此種種指數此地不擬與以估價,僅欲特別說明此種指數之編製上與了解上之若干重要之點.

編製一貨幣工資指數時所遇者,本質上亦爲求諸價格的平均數之問題;惟此地所取之價格非商品之價格而爲勞動之價格.因勞動"價格"開價之方法不同,吾人乃立遇困難.編製工資變動指數之根據,應爲每小時工資(Hourly rate of pay)乎,應爲每週全時工資(Full time weekly earnings),卽以每週額定工作時數與每小時工資相乘之積)乎?抑爲每週實得工資(Actual weekly earnings)乎?指數之根據每小時

工資而作者,對於勞動時間之增減,失業,半失業(under-employment),額外工作及特殊津貼,皆不能表現.根據第二種材料而作者固能表現勞動時間之增減,然其他缺點依然存在.惟每週實得工資之指數之既包括工作鬆懈時期又包括工作緊張時期者,似爲最有用之形式.但編製此種形式之指數時,其材料特別難得,此點於實踐上異常重要.美國紐約邦之勞工部曾根據由全邦若干代表工廠蒐集而得之材料作成如此之一指數.其形式爲以圓角分計之實際平均數,吾人可任取一基期而化此諸數爲比值.

除選取何種形式之工資數字一問題外,尚有取何種方法加權與求平均數之問題,此地不必加以討論.然"代表性"之一事或爲最重要者,此與編製批發物價指數時應包括何種商品種類之問題相當.若僅就熟練工人之工資而平均之,其結果不能作爲一般工資率之指數.再者,若將工頭之工資及報酬甚高之技術工人之工資一併包括在一般工資指數之內,則此兩種工資所得之權數應與其實際上之重要性相合,並對於結果所得之指數其性質如何應有明白之說明.

實際上,處理工資材料與處理批發物價相同,最好能對於工人之不同各組分別各作一指數,然後再將諸指數結合,而爲一切工資經過適當加權之平均數.研究工資變動,亦與研究物價變動同,與其將無數異質之種類集合爲一,而使成爲一單獨之指數,倒不如將材料分割爲若干重要而同質之

集團,然後再求其平均數,蓋後法更能增長吾人對於變動之本質之認識也.

現今已發表之各種工資指數間,彼此之差異既多而且重要,其原因爲編製時所用方法在上述諸點上有不一致處.此種指數,若其編製時所用方法未經詳細說明,所用材料之性質未經完全記述,吾人對於其所表示之工資變動,不能無條件信任也.

單獨一貨幣工資指數,惟有在生活費不變動時始有意義.而當物價與生活費上昇之時,工資指數上升所表示之利益或純屬幻象;情形相反時所表示之損失亦然.欲比較各時期生活之苦樂,必須將貨幣工資指數化爲實在工資指數,此後者乃兼顧生活費之變動者也.

編製實在工資指數之法可舉下例說明之,工資指數所根據之材料,爲紐約邦475,000左右之工人每週平均所得之工資,以1914年十二月之工資爲基數.生活費指數乃美國勞工統計局所編,代表全國32個城市,以1914年十二月之生活費爲100(註)

(註)此地若以僅限於代表紐約邦諸城市之生活費指數校正紐約邦工資指數,則更爲合理,但結果與現今之數字不致相差過鉅.

表五十八

1914—1923年美國紐約邦實在工資指數

(1914十二月=100)

日期	紐約邦工人每週工資額之比值	生活費	實在工資(即貨幣工資之購買力)
1914年十二月	100	100	100
1915 ,,	107	102	105
1916 ,,	123	115	107
1917 ,,	140	138	101
1918 ,,	185	169	109
1919 ,,	209	193	108
1920 ,,	225	195	115
1921 ,,	198	169	117
1922 ,,	210	165	127
1923 ,,	222	168	132

以某一時期之生活費指數除該時期之貨幣工資指數,則得實在工資之指數.單單一工資顯然不能表示工人之境遇.經過校正以後,實在工資固亦表示增加,惟其增加之程度大不如貨幣工資所表現者之高.惟實在工資之指數,乃根據另兩組數而作,而此兩組指數在編製時,錯誤之餘地極大;故最後所得結果,吾人僅能以近似值視之.表五十八所求得者:

吾人僅能視為計量紐約邦工廠工人實際收入變動之一組粗糙的指數也.

參考書

Bowley, A. L. Elements of Statistics (196—213)

Chaddock, R. E. Principles and Methods of Statistics (Chap. X)

Fisher, Irving. Revision of the Weekly Index Number Journal of the American Statistical Association, Sept., 1924.

Fisher, Irving. the Making of Index Numbers.

Flux, A. W. The Measurement of Price Changes. Journal of the Royal Statistical Society, March, 1921 (167—215)

Kelley, Truman L. Statistical Method (331—347)

Mitchell, W .C. The Making and Using of Index Number. Part I, Bulletin 284, U. S. Bureau of Labor Statistics. Oct., 1921.

Persons, Warren M., and Coyle, Eunice. A Commodity Price Index of Business Cycles. Review of Economic Statistics, Prel Vol III. (353—369)

Persons, Warren M. Fisher's Formula for Index Numbers. Review of Economic Statistics. Prel. Vol. III (103—113).

Walsh, C.M. The Measurement of General Exchange Value.

Walsh, C. M. The Problem of Estimation.

Young, Allyn A. The Measurement of Changes in the General Price Level. Quarterly Journal of Economics Aug., 1921 (557

—573).

Young, Allyn A. Index Numbers (in Rietz, H. L., Handbook of Mathematical Statistics, (181—194)

第七章

時間數列之分析：長期趨勢之計量

前文各節主要在處理頻數數列以及組織與敍述此種數列時所遇之各種問題.吾人現今所處理之材料,乃以分析隨時間而起之變動 (chronological variations) 爲研究之中心問題者.此種數列在經濟統計之領域最爲重要,因經濟學及實業上之材料,其大部分皆爲時間變數 (variables in time) ——如銀行淸算額,鋼鐵生產額,銷售量等等.在統計研究之任何其他領域中,時間數列 (time series 或 series in time) 並不占如此第一等的重要地位,故適於分析時間數列之方法,其發展僅爲最近之事,隨統計方法在經濟領域內較廣之應用而起者.

無論對內管理之日常事務上或一般經濟狀況之分析上,皆有與時間數列相關之問題發生.一方面如銷售,購買,與利潤,他方面如股票價格,利率,與破產,皆爲隨時變動之變數.在分析此種數列時,吾人常欲知其發展之率與性質,並將週期變動(periodic fluctuation)與意外變動(accidental fluctuation)

隔離而單獨研究.推銷主任欲知銷售量之變動如何,察其變動之時間與原因,並以之與生產量相比較.經濟學家欲知價格之長期趨勢(trend)如何,仔細考察物價水平之升降.不論為規定一小規模之營業計劃,或為作一經濟預測之最精細之方略,莫不根據此種對於過去之趨勢與變動之研究,與根據相關諸時間數列之變動之比較.商情循環之科學的研究,唯有應用此種方法始為可能.現在吾人之任務為說明適於分析時間數列之各種方法.

[時間數列之初步組織]

時間數列之材料所須之初步組織工作,常較統計材料之須化為頻數分配之形式者為少.不論初級材料或次級材料,在其出處常已為可供分析之形式.但有數事宜予注意者.

對於數字所指之日期,應認識清楚並確切說明.每月之材料可為每月第一日者,(如勃拉特斯脫利氏批發物價指數),可為每月之平均數(如勞工統計局之物價指數),或為每月之總和(如棉花消費量之數字).又可為每月之累積數字,則每一項所代表者為至某一日期為止之該年之總數,如若干表示煤產量之材料即如此.若材料為一月或一年之平均數,必須問此平均數如何求得之法.

其次,根本重要之點,為任何時間數列中,其代表不同日期之材料間,必須有嚴格之比較性.時間數列若非同質者,則任何分析之嘗試,必歸錯誤而無效.但此種數列並非少見,如

商會或政府機關所發表之商品生產量消費量之數字,常根據或多或少之行家之報告.物價數列中不同日期之價目,其所指之商品單位或等次稍有變更,或收集材之市場不同,即失去比較性.調查時分類方法變化,則調查得之材料亦無相比之資格.據報告云,編製商品輸出入數字之方法,直至最近,皆足以使不同日期之數字互相比較時發生錯誤之結果.又如推銷員之區域一有變動,可使其報告根本不同.據謂美國鋼公司之"未交貨噸數"之數字,其表示之意義時時不同.如1922年七月之數字,所指之契約其裝貨日期乃隨顧客之便利而決定者;而其前一年之數字,則有許多買賣其裝貨日期乃確切規定者.此等爲時間數列之缺點之例;此等缺點足以使分析工作爲無效.故一個數列必須經過嚴格之測驗後,始能認爲正確與同質也.

[時間數列之圖示]

爲洞察一時間數列之形狀及準備進一步分析起見,普通第一步手續爲根據材料作圖.用圖示法,最能理解一數列之長期趨勢與一般特性.繪圖時平常之算術格紙以及單對數格紙雙對數格紙(較少)皆可應用.後兩種對於某些用途之優點前已言之.三者之選擇視材料之性質與研究之目的而定.注意點在銷路,物價,生鐵產量或其他變動之絕對量,或在比較諸數列之絕對差數,則宜用平常之直線圖.注意點在百分數變動(percentage variation)或作相對變動之比較時,

則宜應用單對數格紙.一般言之,吾人若有瞭解此後一種圖式之習慣者,總以用之為宜.蓋經濟材料之以時間因子為其變數之一者,當繪於單對數格紙時,其關係常得一較明白而正確之表現,而數列亦可作更有意義之比較.

對於若干目的,材料一經繪圖,則研究時間數列之過程即已告終.一般長期趨勢,可由圖上決定.季節變動及其他變動之存在亦可確斷.長期趨勢及變動亦可作大致之比較.惟吾人應知如此獲得之一切知識,其性質為"非數量的" (non-quantitative),所作之比較,亦為嘗試與近似的性質.雖然如此,吾人恭想長期趨勢與關係時,根據此種之圖較之根據原來數字大為清楚.如此獲得之知識雖非確切,然為某些目的已足敷用.至於對其他目的,則尚須更確切之計量與更精密之分析.茲可述若干適宜之方法於下.

[影響時間數列之各種力量]

分析一時間數列之一般目的,為就作用於該數列之各種力量中隔離其一種或幾種所發生之影響.其目的,或為瞭解此單獨一數列在過去之動態以便預言其將來,或為使兩個或兩個以上之數列可以互相比較.然欲絕對正確的隔離各種力量之影響,並非在任何情形中皆為可能之事,有時甚至連一近似之結果亦不可得.惟一般言之,若有充分長期之數字,則吾人可以相當正確的計量各種力量所加於該數列的動態之影響.此種計量方法應用於經濟分析之可能性,決

未爲人所完全認識也.

作用於時間數列之種種力量究爲何物乎?在此種力量中,其獨特而僅影響於一數列者,或有其例;然一般言之,吾人可將作用於此種數列之種種影響,納入如下之少數範疇.

[長期趨勢]

第一,經濟統計中大部分之數列,皆表示一確定之長期趨勢.此長期趨勢或爲有固定之方向者,或爲依固定之率而改變其方向者.其方向與率,甚或因新因素之參加,而發生突兀之變動者.例如一個商號之生產量與銷售量,就若干年之一時期而觀之,常表示一頗有規則之發展.又如人口數目.基本礦物生產量.汽車登記數等等亦莫不如此.在某些情形中,發展之率可爲一負數,如過去半世紀內美國之利率即其例也.故長期趨勢 (secular trend) 之觀念,同樣包含此正負兩方面之變動.

分析一時間數列時,某一日期之長期趨勢值 (trend value) 即爲該日期之正則值 (normal value). 此正則值者何,即一切意外及糾纏之力量若能除去而僅僅餘下正則的發展之結果時該數列所應有之數值也.任何數列在一定日期皆有其正則值,吾人欲判斷發展因子以外之一切力量的結果,可用之爲基數或參照點 (point of reference). 此正則值之觀念在經濟分析中爲根本重要者.斯奈德 (Carl Snyder) 氏曰:"其他任何方法,皆不能使吾人將經濟上之事件作一確

當之配列,有若是之迅速者."

於此應特別指出者,長期趨勢之意義,指統計數列之規則,光滑而長期之變動而言,故不論在絕對量上或在增減之率上,凡頻起而突兀之變動,皆與長期趨勢之觀念甚不相容.新因素之參加與舊因素之消滅,固可召致偶然之變動;然此時若分割一時間數列所代表之時期爲若干部分,個別決定其趨勢,此事實已破壞長期漸變之整個觀念.影響數列諸因子中若屢起劇烈之更迭者,固無長期趨勢之可言也.

吾人不能由此推論以爲一切時間數列中皆有確定之向下或向上之長期趨勢存在.許多數列僅繞一固定水平而變動,而此水平非與時俱變者,一定地點之氣壓紀錄即其例也.

[循環變動]

當研究一時間數列之圖示時,長期趨勢可於其總的向上或向下之傾向見之.惟因有無數其他變動外加於此趨勢之上,故吾人不能僅用觀察法決定其確切之位置.此無數外加於長期趨勢之變動,或規則或不規則,或劇烈或和緩,或簡單或複雜;牽引變數,使或高於或低於長期趨勢之值.任何日期之變數數值所代表者,即此種種力量與長期趨勢互相作用所淨得之合力也.此等對於漸次而正則之變動成爲擾亂因子之力量,可分爲幾種形式.

季節變動 (seasonal variations),於經濟統計數列之包含

每月之數值者,大部分皆可見之.商品之生產量及消費量、利率、銀行清算額、鐵路貨運量及其他多種之材料,皆有季節變動,每年僅少有不同耳.季節變動果若存在,則其性質必然爲週期性者,常以十二個月爲固定之週期.另有一種週期性較不顯著然而頗有規則之變動,即循環變動(cyclical fluctuation)是也.與經濟循環或商情循環有關之力量,影響於數列,卽生此種結果.物價、工資、工業生產量、交易所貿易額以及與個別實業單位之活動有關之數列之大部分,皆受市況之盛衰交替所影響.盛衰之時期雖長短不一,然此變動之一般結果,在過去尙爲規則,足以使此種變動成爲可研究之對象.

[意外變動]

與此種稍有規則之變動糾纏一起者,爲無規則之意外變動之結果,——如災變(如舊金山之地震)、戰爭、水災、火災以及無數細小事變皆是也.細小事變引起之擾亂雖較不猛烈,然其爲出於意外則與大事變同.此種無規則之變動,亦如季節變動與循環變動然,牽引某一日期之變數之實際數值,使脫離正則數值,——正則數值者,卽因長期趨勢而生之規則而一致之變動若單獨存在時所可得之數值也.由此等擾亂因子所純得之結果,決定某一日期之實際數值高低於正則數值之數.

分析時間數列時,須儘量使此各種力量之結果相互隔離.問題或須單獨研究一個因子,或須將已知之數值完全分

劃,所用之材料若爲按年而編者,當然無季節因素參加其內.玆先就包括此種材料之問題加以研究,以爲說明種種方法之開端.

[長期趨勢之計量]

此種問題之材料,可舉下列數字爲其一例.爲求計算簡便計,以十萬萬爲單位.

表五十九

1860—1923年紐約之銀行清算額

(單位:十萬萬)

1860	$ 7.2	1876	$21.6	1892	$36.7	1908	$ 79.3
1861	5.9	1877	23.3	1893	31.2	1909	103.6
1862	6.9	1878	19.9	1894	24.4	1910	97.3
1863	14.9	1879	29.2	1895	29.9	1911	92.4
1864	24.1	1880	38.6	1896	28.8	1912	100.7
1865	26.0	1881	49.4	1897	33.4	1913	94.6
1866	28.7	1882	46.9	1898	42.0	1914	83.0
1867	26.7	1883	37.4	1899	60.8	1915	110.6
1868	28.5	1884	31.0	1900	52.7	1916	159.6
1869	37.4	1885	28.2	1901	79.4	1917	177.4
1870	27.8	1886	33.7	1902	76.3	1918	178.6
1871	29.3	1887	33.4	1903	66.0	1919	235.8
1872	33.8	1888	31.1	1904	68.6	1920	243.2
1873	35.5	1889	35.9	1905	93.8	1921	194.4
1874	22.9	1890	37.4	1906	104.7	1922	217.9
1875	25.1	1891	33.7	1907	87.2	1923	214.0

前曾指出,任何一年之數字,(如1912年$100,700,000,000之數),爲可以分作長期趨勢,循環變動及無規則或意外變動之各種力量共同合作而得之純粹結果.吾人之第一問題爲決定此各種力量中第一種力量之結果.

圖 54 中所給者,爲自 1860 年起至 1923 年爲止紐約之銀行清算額.由此圖可見有一固定之長期趨勢存在,並有離趨此長期趨勢之差離,明顯而多少帶有規則性.欲求此長期趨勢之近似值,其法有數種.應用移動平均數 (moving average),可以除去一時的變動之結果,而求得代表經常活動的發展因子之結果之數值.假定在時間因子與另一變數之間有一固定之函數關係存在,至少在經驗上如此,則長期趨勢之近

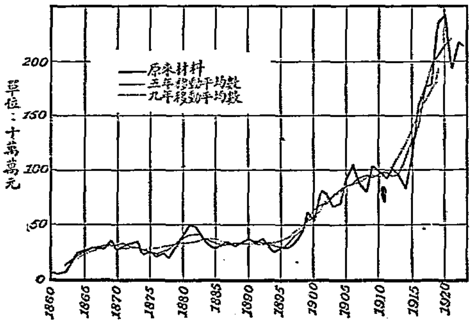


圖 54.——1860—1923 年紐約銀行清算額及其移動平均數

似值,可配給一適當之曲線於已給之材料以求得之.若用就此材料而修勻之,所得結果大致相同,惟如此求得曲線,自屬於近似的與經驗的性質.在某種研究中,可用一個統計數列作爲另一同性質的數列之基數或長期趨勢線(trend line).

[移動平均數]

用移動平均數以決定長期趨勢之法,先求若干年(或月,或星期)之平均數,以此平均數作為計算此平均數時所取之時期之中點之正則數值或長期趨勢值.就紐約銀行清算額(1900—1923)作如此之計算而求得之三年、五年、七年及九年移動平均數,其結果如下表:

表六十

1900—1923年紐約銀行清算額之
三年、五年、七年、九年移動平均數
(單位:十萬萬)

年份	原來材料	三年移動平均數	五年移動平均數	七年移動平均數	九年移動平均數
1900	$ 52.7				
1901	79.4	$ 69.5			
1902	76.3	73.9	$ 68.6		
1903	66.0	70.3	76.8	$ 77.4	
1904	68.6	76.1	81.9	82.3	$ 78.7
1905	93.8	89.0	84.1	82.3	84.3
1906	104.7	95.2	86.7	86.2	86.3
1907	87.2	90.4	93.7	90.6	88.1
1908	79.3	90.0	94.4	94.0	92.0
1909	103.6	93.4	92.0	95.0	94.8
1910	97.3	97.8	94.7	93.6	93.6
1911	92.4	98.8	97.7	93.0	94.3
1912	100.7	95.9	93.6	97.5	102.3
1913	94.6	92.8	96.3	105.5	113.2
1914	83.0	96.1	109.7	116.9	121.6
1915	110.6	117.7	125.0	129.2	137.0
1916	159.6	149.2	141.8	148.5	153.7
1917	177.4	171.9	172.4	169.7	164.1
1918	178.6	197.3	198.9	185.7	177.8
1919	235.8	219.2	205.9	201.0	192.4
1920	243.2	224.5	214.0	208.8	
1921	194.4	218.5	221.1		
1922	217.9	208.8			
1923	214.0				

1904年之三年移動平均數為1903—4—5三年之數字之

平均數,1904年之五年移動平均數爲 1902—3—4—5—6 五年之數字之平均數.其他移動平均數之算法亦相同.在每種情形中,移動平均數皆代表所包括之時期之中心;作爲該時期之中點之正則數値.時期之年數若爲奇數,則中點適爲一年,計算手續簡便;然此非必要之事.因時期之年數爲偶數時,雖中點跨於兩年之間,仍可就求得之諸移動平均數,更求此移動平均數之"兩年移動平均數",將中點移置於各年之中也.圖54上所給者,爲整個時期之五年及九年移動平均數以及原來之材料.

就圖上可見平均之結果,能得一較爲光滑之曲線,減弱牽引各年數値脫離長期總趨勢之諸變動之影響.求每一平均數時所取之時期愈長,則所得之曲線愈光滑.雖然,決定時期之長短時猶有其他因子須加考慮者,玆就若干此種因子而論之.

[移動平均數之特質]

設已知循環變動爲繞一前後一致之水平,或爲繞一前後以一致之斜度而上升之線,而循環時期之長短與變動程度之大小皆固定不變時,則一移動平均數,其包括之時期與循環時期相等(或爲循環時期之倍數)者,將成一直線,此直線即爲長期趨勢之完全之表現.在同樣條件之下,一移動平均數,其包括之時期爲大於或小於循環之時期者,結果不爲一直線,而爲一新循環,其時期與原來之循環相等,惟變動之

程度較小.且如此求得之新循環,其循環之最小數與最大數之位置不必與原來之循環之最小數與最大數之位置相合.一般言之,當求得如此之新循環時,移動平均數所根據之時期愈長,則變動之程度愈小.(註)

上述諸理可以下列任意選得之數字說明之.第一例中五個數字相繼重複出現,繞一公共之水平變動.

表六十一

移動平均數之應用之說明

(1) 循環材料	(2) 五項移動平均數	(3) 十項移動平均數(中點經移置)	(4) 三項移動平均數	(5) 八項移動平均數(中點經移置)
2				
6			$5\frac{1}{3}$	
8	$6\frac{1}{5}$		8	
10	$6\frac{1}{5}$		$7\frac{2}{3}$	
5	$6\frac{1}{5}$		$5\frac{2}{3}$	$6\frac{3}{8}$
2	$6\frac{1}{5}$	$6\frac{1}{5}$	$4\frac{1}{3}$	$6\frac{13}{16}$
6	$6\frac{1}{5}$	$6\frac{1}{5}$	$5\frac{1}{3}$	$6\frac{3}{8}$
8	$6\frac{1}{5}$	$6\frac{1}{5}$	8	$5\frac{6}{8}$
10	$6\frac{1}{5}$	$6\frac{1}{5}$	$7\frac{2}{3}$	$5\frac{11}{16}$
5	$6\frac{1}{5}$	$6\frac{1}{5}$	$5\frac{2}{3}$	$6\frac{3}{8}$
2	$6\frac{1}{5}$	$6\frac{1}{5}$	$4\frac{1}{3}$	$6\frac{13}{16}$
6	$6\frac{1}{5}$	$6\frac{1}{5}$	$5\frac{1}{3}$	$6\frac{3}{8}$
8	$6\frac{1}{5}$	$6\frac{1}{5}$	8	$5\frac{6}{8}$
10	$6\frac{1}{5}$	$6\frac{1}{5}$	$7\frac{2}{3}$	$5\frac{11}{16}$
5	$6\frac{1}{5}$	$6\frac{1}{5}$	$5\frac{2}{3}$	$6\frac{3}{8}$
2	$6\frac{1}{5}$		$4\frac{1}{3}$	$6\frac{13}{16}$
6	$6\frac{1}{5}$		$5\frac{1}{3}$	
8	$6\frac{1}{5}$		8	
10			$7\frac{2}{3}$	
5				

((3)行與(5)行中各項皆用兩項移動平均數移置)

(註)但變動程度之遞減非有規則的而為循環性的.

(2) 行與 (3) 行之移動平均數代表循環已被全部除去之材料.移動平均數之時期不與循環之時期相等,或不與該時期之倍數相等時,循環不能除去,觀表中 (4) (5) 兩行卽明.

上述之結論,不論循環變動所圍繞之直線爲如何,皆可適用.下面一例仍用前例之材料,惟加一固定不變之增量 3.此卽等於將同一循環加於斜度爲 +3 之一線上.

表六十二

移動平均數應用於直線長期趨勢之數列之説明

(1) 循環材料	(2) 五項移動平均數	(3) 十項移動平均數(中點經移置)	(4) 三項移動平均數	(5) 八項移動平均數(中點經移置)
2				
9			$8\frac{1}{3}$	
14	$12\frac{1}{5}$		14	
19	$15\frac{1}{5}$		$16\frac{2}{3}$	
17	$18\frac{1}{5}$		$17\frac{2}{3}$	$18\frac{3}{8}$
17	$21\frac{1}{5}$	$21\frac{1}{5}$	$19\frac{1}{3}$	$21\frac{13}{16}$
24	$24\frac{1}{5}$	$24\frac{1}{5}$	$23\frac{1}{3}$	$24\frac{3}{8}$
29	$27\frac{1}{5}$	$27\frac{1}{5}$	29	$26\frac{6}{8}$
34	$30\frac{1}{5}$	$30\frac{1}{5}$	$31\frac{2}{3}$	$29\frac{11}{16}$
32	$33\frac{1}{5}$	$33\frac{1}{5}$	$32\frac{2}{3}$	$33\frac{3}{8}$
32	$36\frac{1}{5}$	$36\frac{1}{5}$	$34\frac{1}{3}$	$36\frac{13}{16}$
39	$39\frac{1}{5}$	$39\frac{1}{5}$	$38\frac{1}{3}$	$39\frac{3}{8}$
44	$42\frac{1}{5}$	$42\frac{1}{5}$	44	$41\frac{3}{4}$
49	$45\frac{1}{5}$	$45\frac{1}{5}$	$46\frac{2}{3}$	$44\frac{11}{16}$
47	$48\frac{1}{5}$	$48\frac{1}{5}$	$47\frac{2}{3}$	$48\frac{3}{8}$
47	$51\frac{1}{5}$		$49\frac{1}{3}$	$51\frac{13}{16}$
54	$54\frac{1}{5}$		$53\frac{1}{3}$	
59	$57\frac{1}{5}$		59	
64			$61\frac{2}{3}$	
62				

((4)行與(5)行中各項皆用兩項移動平均數移置.)

所取之移動平均數其時期若與循環之時期相等或與其倍數相等,則能盡去循環變動之影響而得長期趨勢值.移動平均數所根據之時期不等於循環之時期時,循環依然存在,其時期照舊,變動之程度則變小,觀(4)(5)兩行卽明.

當此種理想的簡單情形不復存在,循環時期之長短與變動程度之大小不復爲固定不變時,則移動平均數之意義較爲模糊而其解釋亦較爲複雜.循環之時期長短不一,則選擇移動平均數之時期較難,一般言之,宜選等於或大於循環時期的平均長度之時期,根據較短之時期而作之移動平均數,結果所得之線將有顯著之循環;計算移動平均數時包括之時期愈長,則循環變動之範圍愈狹小.

當循環之時期不變而其範圍不一時,則移動平均數其時期與循環之時期相等者,結果得一具有較小的循環之長期趨勢線.此次級之循環,當移動平均數之時期與原來循環之時期相等(或與其倍數相等)時,其變動之範圍達最小限度.當此兩不規則性相合,卽材料中之循環,其時期之長短與範圍之廣狹皆不一致時,最足以代表長期趨勢之移動平均數,爲其時期等於循環時期之平均長度,或此長度之倍數者.

當長期趨勢非一直線時,則吾人又當考慮一新因子.若一數列中潛在之長期趨勢爲一凹面向上之曲線,則一移動平均數將永遠大於實際長期趨勢值;若相反而爲一凸面向

上之曲線,則移動平均數將永遠小於實際長期趨勢值.

此種情形於下面之例中描寫之.表六十三中之數字,表示一時期與範圍皆爲固定之循環加於一凹面向上(即數值增大時其率亦不斷遞增者)之長期趨勢線上所得之諸數值.若移動平均數能全部肅清循環之影響,則結果所得之諸數值,將等於五項之平均數值$\left(6\frac{1}{5}\right)$加上$y=x^2$一方程式之諸數值所得之和.

表六十三

移動平均數應用於曲線數列之說明

(遞增率)

(1) x	(2) x^2	(3) 循環材料	(4) (2)行+(3)行	(5) 五項移動平均數	(6) 實際長期趨勢值 $(x^2+6.2)$
0	0	2	2		
1	1	6	7		
2	4	8	12	12.2	10.2
3	9	10	19	17.2	15.2
4	16	5	21	24.2	22.2
5	25	2	27	33.2	31.2
6	36	6	42	44.2	42.2
7	49	8	57	57.2	55.2
8	64	10	74	72.2	70.2
9	81	5	86	89.2	87.2
10	100	2	102	108.2	106.2
11	121	6	127	129.2	127.2
12	144	8	152	152.2	150.2
13	169	10	179	177.2	175.2
14	196	5	201	204.2	202.2
15	225	2	227	233.2	231.2
16	256	6	262	264.2	262.2
17	289	8	297	297.2	295.2
18	324	10	334		
19	361	5	366		

此地移動平均數之值永遠大於實際長期趨勢值,在此種形式之數列中,此種偏誤之形式永遠存在.

表六十四表示以同樣之循環數值加於一凸面向上(卽增大時其率不斷遞減者)之長期趨勢線上所得之結果.此地若有方法完全取消循環之影響,所得結果將等於五項之平均數值 $\left(6\frac{1}{5}\right)$ 加上 $y=\sqrt{x}$ 一方程式之諸值所得之和.

此地移動平均數各值永嫌過小,其與長期趨勢值不相合之情形,在 x 爲較小之值時最爲明顯,因增率之遞減於此時最爲明顯也.

表六十四

移動平均數應用於曲線數列之說明

(遞減率)

(1) x	(2) $\sqrt{x}$	(3) 循環材料	(4) (2)行+(3)行	(5) 五項移動平均數	(6) 實際長期趨勢值 ($\sqrt{x}+6.2$)
0	0	2	2.00		
1	1.00	6	7.00		
2	1.41	8	9.41	7.428	7.61
3	1.73	10	11.73	7.876	7.93
4	2.00	5	7.00	8.166	8.20
5	2.24	2	4.24	8.414	8.44
6	2.45	6	8.45	8.634	8.65
7	2.65	8	10.65	8.834	8.85
8	2.83	10	12.83	9.018	9.03
9	3.00	5	8.00	9.192	9.20
10	3.16	2	5.16	9.354	9.36
11	3.32	6	9.32	9.510	9.52
12	3.46	8	11.46	9.658	9.66
13	3.61	10	13.61	9.800	9.81
14	3.74	5	8.74	9.936	9.94
15	3.87	2	5.87	10.068	10.07
16	4.00	6	10.00	10.194	10.20
17	4.12	8	12.12	10.318	10.32
18	4.24	10	14.24		
19	4.36	5	9.36		

前面研究直線長期趨勢之移動平均數時,曾指出移動平均數所根據之時期,一般言之,至少應等於循環之時期,苟材料有絲毫不規則時,最好取循環時期之較高的倍數.移動平均數所取之時期愈長,則其性質愈爲穩定.但潛在之長期趨勢若與直線形式相差過遠,而取向上或向下之曲線形式時,則應用任何移動平均數所發生之錯誤,隨其所取之時期之長而增加.在此種情形中,應用移動平均數以計量長期趨勢時,應在足以消除循環影響之範圍內選取最短之時期,此時期等於一個循環之平均長度.

但在實際上,此各種情形組成複雜之結合:循環時期之長短與變動之大小各異,則須根據一頗長之時期以求移動平均數.材料之長期趨勢常爲曲線形式,則須一根據短時期而求得之移動平均數以減少向上或向下之偏誤.尚有一考慮之點在實際工作上亦爲重要者,即吾人不能計算移動平均數直至最後之時期也.若移動平均數所取之時期短,則不能計算之時期亦短,每次選擇時期時,應根據上述諸考慮之點對於實際材料加以研究.

前面之討論中,吾人視移動平均數爲長期趨勢之表現.然移動平均數尚可用於去除無規則之變動.若以此爲目的,則移動平均數應根據較循環之平均時期爲短之時期.

玆可回至紐約銀行清算額之問題.觀圖54,可見不同之移動平均數所作諸線,彼此間有重要之差異存在.五年平均

數最接近於原來材料之曲線,正如吾人所預期.九年平均數成一最光滑之長期趨勢線,惟在另一方面,其離原來材料亦最遠.此種情形在 1893—1898、1911—1915 兩時期最爲明顯,因在此兩時期內數列之發展率具有重大變化也.除此種偏誤外,九年平均數似乎最能確切代表一般之長期趨勢.

吾人可參考關於該時期商情趨勢之知識,以決定此各種移動平均數之價值.紐約銀行清算額,本爲代表一般商情之做成指數,直接反映投機活動與產業活動之變動,其數列中,巨大與細小之商情循環咸有反映.吾人若已知最後半世紀內 (1870—1920) 商情循環之次數,即可決定此各種平均數中,究以何種最宜於作爲計算循環差離 (cyclical deviations) 之標準.此法實由已知之結果倒推,非常可應用者也.

若以每次蕭條以後恢復常態之前一年爲每一商情循環之起點,則可分成下列諸循環,代表一般實業活動之情形:(註)

1870—1878	1904—1908
1879—1885	1908—1911
1886—1897	1912—1914
1898—1904	1915—

用三年平均數時,所得之循環個數太多,難以列舉.其實

(註) 此等循環,以奧格本 (W. F. Ogburn) 及湯麥斯 (Dorothy Thomas) 氏所作之指數爲根據;請參考 Quarterly Publications of the American Statistical Association, Sept., 1922, 324—340.

材料對於此平均數之差離,主要爲細小而意外之變動,不能視爲眞循環者也.材料對於七年移動平均數以及對於九年移動平均數之差離,則表現有下列諸循環:

對於七年移動平均數之諸差離之循環	對於九年移動平均數之諸差離之循環
1871—1878	1871—1878
1879—1884	1879—1887
1885—1888	1888—1897
1888—1897	1898—1900
1898—1900	1900—1903
1900—1903	1904—1907
1904—1907	1908—1914
1908—1911	1915—
1911—1914	
1915—1918	

如此決定之循環與循環之間,其某些差異無疑的爲所分析之材料之差異所引起.惟其他差異中有若干應加注意者.對於七年移動平均數之差離,較其他兩數列各多表示三個循環.在其他兩數列中,1885—1888 及 1915—1918 兩循環皆作爲小變動而不另行單獨表出. 1898—1900 之循環,在對於九年移動平均數之差離中亦存在,惟在前面用爲比較之標準之指數中則無之. 1911—1914 之循環,在對於七年平均數之差離中存在,惟在對於九年平均數之差離中則無之.大戰年間清算額之急速增加,使九年平均數發生極大偏誤,竟致材料中之一眞正之循環爲其抹殺.

總言之,對於目前之材料,移動平均數其時期短於七年

者,不足以作爲長期趨勢之計量.九年平均數除在時期之最後一部分——增率急速變化之時期外,可爲其長期趨勢之有效之計量.七年之時期似嫌過短,因其將若干變動之不甚重要者亦一併表出也.

一般言之,移動平均數之主要優點,爲其伸縮自如之適應性.用數學曲線表現長期趨勢時,常須分一時期爲數段,逐段配合一單獨之曲線.當情況改變與增減之率變化劇烈時,此爲必然之結果.然移動平均數遇此種變動則有伸縮自如適應新情況之優點,故作爲長期趨勢之計量,常較之費盡大力配合而得之曲線,更爲有效.

復有一種加權移動平均數,在英文稱 progressive mean (進行中數) 者,計算時對於所包括之各項,各以不同之權數加權,近於中心之各項則權數特大.權數則以二項展開式中各項之係數爲之.如前例中有用五項移動平均數者,如加權時,其權數卽 1, 4, 6, 4, 1. 加權移動平均數自有其用處,惟對於消去時間數列之週期性或修削此種數列之不規則變動,似不甚適宜.

〔用數學曲線表現長期趨勢之法〕

對於許多種材料,其長期趨勢可不用根據移動平均數而作之線,而寧用一數學曲線表現之.如增加 (或減少) 時依一固定之絕對數增量 (或減量), 則一直線可爲長期趨勢之確切表現.增加或依一固定之百分數(如本金之依複利率,而

增殖者，則一固定之數學曲線爲此長期趨勢最好之表現.在許多經濟統計數列中，其材料之增加與減少似依照一定之法則.遇此種情形時,吾人若能得一數學式以代表此潛在之法則,則分析,說明及作圖之工作,皆可因之而大爲便利.當然,在無論何種事件中,皆有不合於法則之情形存在,對於長期趨勢線,皆有向上向下之差離.惟代表潛在的發展法則之方程式,其價值並不能因此而破壞也.

作爲長期趨勢之計量之移動平均數,與此種數學曲線間有一根本之差異在.前者並不暗示材料爲合於任何一定之法則,其所根據者僅爲此已知之材料;一般之長期趨勢若有改變,移動平均數卽追隨此新長期趨勢.故移動平均數爲伸縮自如之長期趨勢計量,適應於變更之情況;僅欲作一經驗性的近似值,表現一數列之趨勢.至於配合於經濟數列之數學曲線,實際上固亦不過爲一經驗性的近似值,惟其意義稍爲不同.因此種曲線乃根據一假定,卽材料在表面上之種種變動(意外變動或其他)之後,有一"法則"存在也.所假定者固仍不過爲一經驗性的法則,但其認定長期趨勢有前後一致而確定之性質並可以數學式表示,則其特點也.此種假定若欲有效,則在此法則持續之時期內,凡影響所研究之數列之條件,必須無重大變動.如金之生產量,其長期趨勢可用一方程式代表之.一旦開採之法大變,金生產量亦將大變.以前之方程式卽不能適用.在此變動以前之材料,與在此變動以

後之材料,其性質不同;故無從求得一方程式而能單獨代表整個時期之長期趨勢者.不同,代表整個時期之長期趨勢者將無從求得.

着手實際解決決定長期趨勢之問題時,第一步工作爲選擇適當之曲線形式.此工作或爲最困難之一部分;的確,個人判斷之成分在此一部分中最爲明顯.因選擇最適當之曲線一事,並無客觀之規律可以遵循,無固定不變之標準存在.關於此點,吾人於研究曲線之主要形式之特性與其配製之方法以後猶須談及,目前且假定已經選定一種曲線,與第二章中所描述之諸形式中之一種或其相關之形式相似,吾人乃研究如何將此曲線配合於材料之方法.

[配合一直線之法;最小平方法]

若材料經描繪後,表示之長期趨勢最適於以一直線代表者,則配合之工作僅爲決定方程式 $y=a+bx$ 中常數數值之問題,吾人所求之 a, b 之數值,須能使此方程式代表一最切合於材料之長期趨勢之直線.茲可舉一簡單之例以說明各種可用之方法.圖 55 上繪有九點 (1, 3; 2, 4; 3, 6; 4, 5; 5, 10; 6, 9; 7, 10; 8, 12; 9, 11).試配合一直線於此各點.

a, b 之近似值可根據觀察決定之.吾人可作一直線經過諸點之間而其所取方向似最能接近於長期趨勢者,於是計算此直線之斜度,決定此直線與 y 軸之交點,而得所求之方程式之近似值.此法顯然爲粗率而不確切,每個人所求得

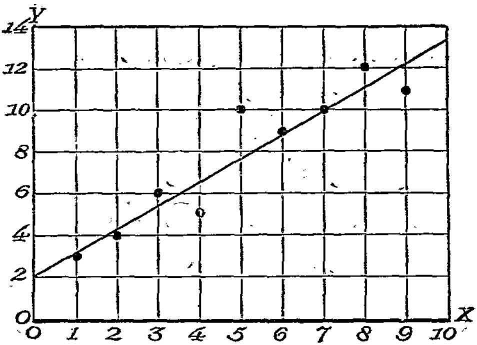


圖55.——表示對於九點配合一直線之情形.

之結果彼此間可相差甚大.惟最能確切適合於圖上之材料者祗有一條直線,此最適合之一線,其常數可用最小平方法(method of least squares)求之.

最小平方法所根據之理論,此地不必細述,可略述其要點如下:設關於某一數量,已得若干觀察值(observation values),欲根據之求該數量之最可能之值,則用數學方法,可以證明該數量之最可能之值爲如此之一數值,卽將其與每一觀察值之差(此差名爲餘值"Residual")、作成平方而相加時,所得之和爲一最小數.諸觀察值之算術平均數卽如此.如某一距離,由若干人計量之而結果若各各不同,則最可能之值爲諸

不同結果之算術平均數.計算此平均數之手續有下列各步,為便於下文之說明起見,列舉如下.吾人所求者,為所計量之距離之最可能之値,其式為

M=(一個常數)

設此數値已有近似値三個:

$$\begin{array}{rr} & M = \ \ 5672\text{呎} \\ & M = \ \ 5671\text{呎} \\ & M = \ \ 5676\text{呎} \\ \hline (\text{相加}) & 3M = 17019\text{呎} \end{array}$$

因只有一未知數M,故由此方程式可直接求得

M=5673呎

此數値與各個近似値之差,其平方之和為一最小數.

若計算兩個變數間之關係,問題亦相同.此時則吾人以求一適當之方程式,以正確代表此關係為目標.吾人以觀察之結果作成關係方程式時,式中常數彼此不能合符.換言之,吾人以結果繪圖時,諸點不能在一直線上.則吾人所求之關係方程式中之常數,其最近似之値為何乎?此地之答復,與前面求單獨的一個數量之近似値時之答復全同.吾人所求之常數爲如此之數値,即將其代入方程式而繪為直線時,各已知點對此直線之差離,其平方之總和成為一最小數者是也.設每一對數目皆為兩變數間實在關係之近似値,吾人欲求最可能之關係,則代表此關係之線,為諸差離之平方總和為

一最小數者.（最小平方法及其計算時之覆驗詳見附錄一）

在此地有 x 與 y 之九對數目.將此各對數目代入直線方程式之普遍式 $y=a+bx$, 得下列諸觀察方程式:

$$
\begin{aligned}
3&=a+1\,b\\
4&=a+2\,b\\
6&=a+3\,b\\
5&=a+4\,b\\
10&=a+5\,b\\
9&=a+6\,b\\
10&=a+7\,b\\
12&=a+8\,b\\
11&=a+9\,b
\end{aligned}
$$

任取上列諸方程式之兩個作聯立方程式,可求得 a 與 b 之值.惟此值不能適合於其餘諸方程式.吾人須將此九個觀察方程式合併爲兩個正則方程式 (Normal equation), 聯立解之,可得 a, b 之最可能之值.求第一個正則方程式之法,爲以各觀察方程式中第一未知數(a)之係數乘該方程式,如此求得之諸方程式相加.在此處之例中 a 之係數全爲 1, 故九個觀察方程式如舊.求第二個正則方程式之法,爲以各觀察方程式中第二未知數(b)之係數乘該方程式,如此求得之諸方程式相加.故第一方程式全體皆爲 1 所乘,第二方程式則被乘於 2, …… 等等.茲將求兩個正則方程式之法說明於下:

表六十五

由觀察方程式求正則方程式之方法

$3=a+1b$	$3=1a+1b$
$4=a+2b$	$8=2a+4b$
$6=a+3b$	$18=3a+9b$
$5=a+4b$	$20=4a+16b$
$10=a+5b$	$50=5a+25b$
$9=a+6b$	$54=6a+36b$
$10=a+7b$	$70=7a+49b$
$12=a+8b$	$96=8a+64b$
$11=a+9b$	$99=9a+81b$
$70=9a+45b$	$418=45a+285b$

兩個正則方程式為

$$70=9a+45b$$

$$418=45a+285b$$

(作正則方程式之一般規則詳見附錄一.)

現今可解此方程式求 a, b 之值.以 5 乘第一方程式,所得結果由第二方程式中減去之,則 a 可消去;得 b 之值為 $\frac{68}{60}$, 或 1.133. 將此值代入兩個方程式之任何一個中,求得 a 之值為 2.111. 故代表最適合之直線之方程式為

$$y=2.111+1.113x$$

實際計算時不必如上面之例,將方程式寫出而求其總和,只須將適當之數值代入下列二方程式即可求得常數之值

$$\Sigma(y)=na+b\Sigma(x)$$

$$\Sigma(xy)=a\Sigma(x)+b\Sigma(x^2)$$

所用符號之意義如下:

$\Sigma(y)$: y諸值之總和.

$\Sigma(x)$: x諸值之總和.

$\Sigma(xy)$: 各對x值與y值之乘積之總和.

$\Sigma(x^2)$: x諸值之平方之總和.

n: 數值之"對"數;在圖上即爲點數.

列表計算較爲便利,其例如下:

表六十六

計算配合一直線所需數值之法

x	y	xy	x^2	
1	3	3	1	
2	4	8	4	
3	6	18	9	$n=9$
4	5	20	16	$\Sigma(x)=45$
5	10	50	25	$\Sigma(y)=70$
6	9	54	36	$\Sigma(x^2)=285$
7	10	70	49	$\Sigma(xy)=418$
8	12	96	64	
9	11	99	81	
45	70	418	285	

將此五數值代入前舉之兩正則方程式,則得所求之正則方程式,結果與自觀察方程式計算而得者相同.

既得一代表最適合的直線之方程式,則可計算相當於x之一定數值之y值,而以此值與觀察值相比較:

表六十七

一個變量之觀察值與計算值之比較

x	y (觀察值)	y (計算值)	d	d^2	xd
1	3	$3.2\frac{4}{9}$	$-.2\frac{4}{9}$	.0597	$-.2\frac{4}{9}$
2	4	$4.3\frac{7}{9}$	$-.3\frac{7}{9}$	.1427	$-.7\frac{5}{9}$
3	6	$5.5\frac{1}{9}$	$+.4\frac{8}{9}$	.2390	$+1.4\frac{6}{9}$
4	5	$6.6\frac{4}{9}$	$-1.6\frac{4}{9}$	2.7041	$-6.5\frac{7}{9}$
5	10	$7.7\frac{7}{9}$	$+2.2\frac{2}{9}$	4.9381	$+11.1\frac{1}{9}$
6	9	$8.9\frac{1}{9}$	$+.0\frac{8}{9}$	.0079	$+.5\frac{3}{9}$
7	10	$10.0\frac{4}{9}$	$-.0\frac{4}{9}$	.0020	$-.3\frac{1}{9}$
8	12	$11.1\frac{7}{9}$	$+.8\frac{2}{9}$	.6760	$+6.5\frac{7}{9}$
9	11	$12.3\frac{1}{9}$	$-1.3\frac{1}{9}$	1.7190	-11.8
總和			0.0	10.4885	0.0

各點對於所得直線之差離,其總和等於零.諸差離各以其相當之 x 值乘之,諸乘積之總和亦等於零.計算之正確與否可以此法覆驗.諸差離之平方之總和為 10.4885, 為一最小數.以任何其他數值作為 a, b 之值,所得直線,其差離之平方之總和必大於10.4885.

[配合一直線之特殊情形]

聯立兩個正則方程式而解之,在任何情形之下,皆可可得 a, b 之最可能之數值.在某種特殊情形下,計算之手續可以簡單化,在處理經濟材料時,此種特殊情形並非少見.若 x 之諸值為依次相繼之數,(如描繪一個中間無間斷之時間數列時,x 之諸值無有不依次相繼者.)則可取其中位數值為原點,觀察之數為奇數時,當然即為居中之一項.於是 $\Sigma(x)$ 之值

爲 0,而正則方程式爲

$$\Sigma(y)=na$$

$$\Sigma(xy)=b\Sigma(x^2)$$

如有一個時間數列.由1900—1920年逐年之數字排成,吾人可取1910年爲原點,相當於1909年之 x 值爲 -1, 1911年爲 $+1$, 諸如此類.當材料可以如此處理時,求 a, b 之值大爲便利.若年數爲偶數時,此法亦可應用,時間(x變數)以半年爲單爲而計算之.

再者, x 之諸值若爲以 0 爲起點之依次相繼諸正數, $\Sigma(x)$ 與 $\Sigma(x^2)$ 之值易於決定.最初之 n 個自然數之總和等於 $\frac{n(n+1)}{2}$.如由 1 至 10 各數之總和,等於 $\frac{10(10+1)}{2}$,或55.此項可以代替正則方程式中之 $\Sigma(x)$.同樣,最初之 n 個自然數之平方之總和,等於 $\frac{2n^3+3n^2+n}{6}$.例如由 1 至 5 各數之平方之總和爲 $\frac{250+75+5}{6}=55$.此式可代替正則方程式中 $\Sigma(x^2)$,而得

$$\Sigma(y)=na+b\left(\frac{n(n+1)}{2}\right).$$

$$\Sigma(xy)=a\left(\frac{n(n+1)}{2}\right)+b\left(\frac{2n^3+3n^2+n}{6}\right).$$

有時,用此種形式之方程式計算,較之最初之方程式爲方便,時間數列之材料可以如此處理之,各年以由 1 起而依次相繼之諸數代表之.

[配合定冪級數曲線之法]

上文所討論者限於直線長期趨勢.此種曲線形式時能與材料吻合,惟在許多情形中,亦有不能確切適合於材料者.在實際上克服此種困難之法,有時爲分裂一數列爲若干段,對於每段所代表之時期各單獨作一線以表示之.數列中有眞正之缺口存在時,則就整個之時期言,其性質並不一致,此分段表示之法固爲正當.然整個時期在本質上若果爲同一之性質,則長期趨勢之整個觀念,將因此分段配線之法所破壞.在許多不合以直線表示之情形中,一定冪級數(potential series)之曲線,可能確切的代表長期趨勢.配合此種曲線之法可略述之.

定冪級數曲線之方程式,其普遍式爲$y=a+bx+cx^2+dx^3$……. 此種形式之方程式當然不代表一抛物線式之曲線,然平常常用以稱呼定冪級數.若其中之x以二方爲止,則稱爲二次抛物線;以三方爲止,則稱三次抛物線;餘可類推.平常應用時,此種曲線不宜超過x三方以上.如以二方爲止,則有三個未知數;須聯立三個正則方程式而解之,始得所求之數值.

計算之手續與上述關於直線形式者同.每一觀察方程式以該式中第一未知數之係數乘之,結果所得之各方程式相加,得第一個正則方程式.用同法施之第二第三未知數,解如此求得之三個正則方程式而求a, b與c之值.結果爲此三常數之最可能之值.下列爲三個正則方程式之普遍式:

$$\Sigma(y)=na+b\Sigma(x)+c\Sigma(x^2).$$

$$\Sigma(xy)=a\Sigma(x)+b\Sigma(x^2)+c\Sigma(x^3).$$

$$\Sigma(x^2y)=a\Sigma(x^2)+b\Sigma(x^3)+c\Sigma(x^4).$$

茲舉例說明此計算之手續.已知 1, 2; 2, 6; 3, 7; 4, 8; 5, 10; 6, 11; 7, 11; 8, 10; 9, 9 諸點,欲配合一二次拋物線.配合曲線時與一切複雜之計算相同,最有實際之重要意義者,一切工作之進行必須有一確定而系統之形式,每一步皆須與前後有確定之關係,可用覆驗之處必須覆驗,因最小心之計算猶不免發生錯誤也.列表計算常為有益,因可使每一步手續及每一組結果皆得明白之表現也.

現今之材料可列為下表:

表六十八

計算配合一條二次拋物線所須諸數值之方法

x	y	xy	x^2	x^2y	
1	2	2	1	2	$n=9$
2	6	12	4	24	$\Sigma(x)=45$
3	7	21	9	63	$\Sigma(x^2)=285$
4	8	32	16	128	$\Sigma(x^3)=2,025$
5	10	50	25	250	$\Sigma(x^4)=15,333$
6	11	66	36	396	$\Sigma(y)=74$
7	11	77	49	539	$\Sigma(xy)=421$
8	10	80	64	640	$\Sigma(x^2y)=2,771$
9	9	81	81	729	
45	74	421	285	2,771	

當 x 之諸值為自 1 起而依次相繼之諸整數時(如目前之例), $\Sigma(x)$, $\Sigma(x^2)$, $\Sigma(x^3)$, 與 $\Sigma(x^4)$ 諸值可查現成之表而得之.(參考 Pearson, Tables for Statisticians and Biometricians, 第二

十八表, Cambridge University Press, 1914.)

將各値代入上面之方程式,得下列正則方程式:

$$74 = 9a + 45b + 285c$$

$$421 = 45a + 285b + 2{,}025c$$

$$2{,}771 = 285a + 2{,}025b + 15{,}333c$$

聯立此數方程式而解之得三個常數之下列數值:

$$a = -.929.$$

$$b = +3.523.$$

$$c = -.267.$$

所求之曲線,其方程式爲

$$y = -.929 + 3.523x - .267x^2$$

此曲線與九個已知點繪於下圖

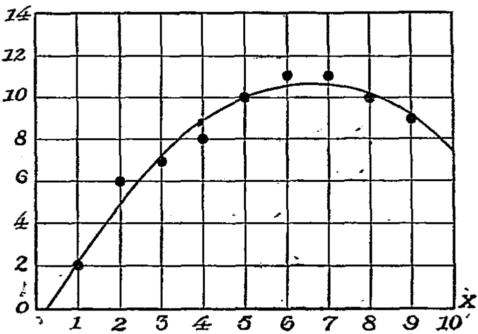


圖56.——表示對於九點配合一條二次拋物線之情形

若 x 諸值依次相繼有如此例者,取其中位數為原點,可減省計算之煩.此種情形中,$\Sigma(x)$ 與 $\Sigma(x^3)$ 皆為0,而正則方程式為

$$\Sigma(y)=na+c\Sigma(x^2).$$

$$\Sigma(xy)=b\Sigma(x^2).$$

$$\Sigma(x^2y)=a\Sigma(x^2)+c\Sigma(x^4).$$

當欲配合一 $y=a+bx+cx^2+dx^3$ 式之三次拋物線於某種材料時,須決定四個常數之值,須有四個正則方程式.此四式如下:

$$\Sigma(y)=na+b\Sigma(x)+c\Sigma(x^2)+d\Sigma(x^3).$$

$$\Sigma(xy)=a\Sigma(x)+b\Sigma(x^2)+c\Sigma(x^3)+d\Sigma(x^4).$$

$$\Sigma(x^2y)=a\Sigma(x^2)+b\Sigma(x^3)+c\Sigma(x^4)+d\Sigma(x^5).$$

$$\Sigma(x^3y)=a\Sigma(x^3)+b\Sigma(x^4)+c\Sigma(x^5)+d\Sigma(x^6).$$

解方程式而求四個或四個以上之常數時須用大量之算術計算,以此種形式之方程式表現長期趨勢是否合宜,頗有疑問.若有充分之常數,可作一曲線適應於材料中之每一變動.惟如此之曲線殆不能取為長期趨勢之代表.(註) 經濟

(註) 關於應用上述形式之定冪級數以代表經驗曲線一事,許登梅茲氏以為在下列各種情形中,其應用始為正當:(1)若 a, b, c……等相繼之係數,其值遞減甚速,以致在觀察之範圍內,高次各項之值急遽變小而居於不重要之地位者.(2)若相繼諸係數遵循一定法則,指示有一代表某種函數(如雙曲函數、三角函數等)之收斂級數(Convergent series)存在者.(3)若所有之係數其數值皆甚小,數值大者僅有少數,故僅此少數須加研究者.參考 Steinmetz, Engineering Mathematics, 214—215.

材料,其與一簡單而一貫之長期趨勢(不論其爲直線或其他形式)小有出入,原爲意中之事;惟若有一眞正之長期趨勢存在,而材料與一相當簡單之曲線形式大有出入者,實屬罕見.若此種出入爲情況激變所致,則無一單獨之長期趨勢線可以適合,最好分割一時期爲若干部分,每一部分單獨作一長期趨勢線,以表示之.許登梅玆(Steinmetz)氏曾言:"惟有在觀察範圍以內之物理情況不變時,經驗曲線始可爲一單獨之方程式所代表."此語雖爲配合曲線於自然科學之材料而發,然同様可適用於經濟材料也.

[一個標本的問題:破產之長期趨勢之決定]

前數節說明配合某種形式之曲線於簡單材料之法,在討論少許不同之各種形式以前,先在此插一具體之例,想必有益.吾人已知1897—1921年美國破產之數字,欲配合一直線,一二次拋物線,及一三次拋物線.同時配合此三種曲線者,爲使其結果可以比較而示其重要之差別也.

爲計算便利計,取時期之中心一年1909爲原點.正則方程式所須之若干數值按下表計算而得,x 諸值代表時間因子,而 y 諸值爲與之相當之破產數.

表六十九

1897—1921年英國破產數
計算配合長期趨勢線所需之數值

(1) 年份	(2) x	(3) y	(4) xy	(5) x^2y	(6) x^3y
1897	−12	13,083	−156,996	1,883,952	−22,607,424
1898	−11	11,615	−127,765	1,405,415	−15,459,565
1899	−10	9,642	− 96,420	964,200	− 9,642,000
1900	− 9	9,912	− 89,208	802,872	− 7,221,848
1901	− 8	10,648	− 85,184	681,472	− 5,451,776
1902	− 7	9,973	− 69,811	488,677	− 3,420,739
1903	− 6	9,775	− 58,650	351,900	− 2,151,400
1904	− 5	10,417	− 52,085	260,425	− 1,302,125
1905	− 4	9,967	− 39,868	159,472	− 637,888
1906	− 3	9,385	− 28,155	84,465	− 253,3[illegible]5
1907	− 2	10.274	− 20,548	41,096	− 82,192
1908	− 1	14,066	− 14,066	14,066	− 14,066
1909	0	11,872	....	....	....
1910	1	11,588	11,588	11,588	11,588
1911	2	12,679	25,358	50,716	101,432
1912	3	13,832	41,496	124,488	373,464
1913	4	14,553	58,212	232,848	931,392
1914	5	16,780	83,900	419,500	2,097,500
1915	6	19,035	114,210	685,260	4,111,560
1916	7	16,498	115,486	808,402	5,658,841
1917	8	13,073	104,584	836,672	6,693,376
1918	9	9,331	83,979	755,811	6,802,299
1919	10	5,515	55,150	551,500	5,515,000
1920	11	8,463	93,093	1,024,023	11,264,253
1921	12	19,982	239,784	2,877,408	34,528,896
總和	0	301,958	+188,084	15,516,228	+9,881,156

$n=25$ $\Sigma(x^3)=0$

$\Sigma(x)=0$ $\Sigma(x^3y)=9,881,156$

$\Sigma(y)=301,958$ $\Sigma(x^4)=121,420$

$\Sigma(xy)=188,084$ $\Sigma(x^5)=0$

$\Sigma(x^2)=1,300$ $\Sigma(x^6)=13,471,900$

$\Sigma(x^2y)=15,516,228$

〔直線之配合〕

原點旣爲 x 諸値之中點,則求常數之値時,需解之方程式爲

$$\Sigma(y)=na$$

$$\Sigma(xy)=b\Sigma(x^2)$$

將已知値代入公式,得

$$301,958=25\,a$$

$$188,084=1,300\,b$$

因此

$$a=12,078$$

$$b=114.7$$

故代表最適合之線之方程式爲

$$y=12,078+144.7\,x$$

〔二次拋物線之配合〕

定一二次拋物線時,須解之正則方程式爲

$$\Sigma(y)=na+c\Sigma(x^2)$$

$$\Sigma(xy)=b\Sigma(x^2)$$

$$\Sigma(x^2y)=a\Sigma(x^2)+c\Sigma(x^4)$$

代入適當數値,則得下列方程式:

$$301,958=25\,a+1,300\,c$$

$$188,084=1,300\,b$$

$$15,516,228=1,300\,a+121,420\,c$$

解上列方程式求其常數值

$$a=+12,258$$
$$b=+144.7$$
$$c=-3.45$$

故所求之方程式爲：

$$y=12,258+144.7\,x-3.45\,x^2$$

〔三次拋物線之配合〕

定一三次拋物線時若原點爲 x 諸值之中點則求常數值時須解之正則方程式爲

$$\Sigma(y)=na+c\Sigma(x^2)$$
$$\Sigma(xy)=b\Sigma(x^2)+d\Sigma(x^4)$$
$$\Sigma(x^2y)=a\Sigma(x^2)+c\Sigma(x^4)$$
$$\Sigma(x^3y)=b\Sigma(x^4)+d\Sigma(x^6)$$

代入已知數得下列方程式：

$$301,958=25\,a+1,300\,c$$
$$188,084=1,300\,b+121,420\,d$$
$$15,516,228=1,300\,a+121,420\,c$$
$$9,881,156=121,420\,b+13,471,900\,d$$

聯立解之得下列諸值：

$$a=+12,258$$
$$b=+482$$
$$c=-3.45$$
$$d=-3.61$$

故所求之方程式爲：

$$y=12,258+482x-3.45x^2-3.61x^3$$

原來之材料及此三方程式所代表之長期趨勢線，給於圖 57，y (破產數) 之實在值，及其在此三種情形中之計算值或正則值 (normal value)，以及每種情形中實在值對於正則值之百分數差離皆列於下表：

表七十

1897—1921 年美國之破產數

實在值，三長期趨勢線所代表之正則值，

實在值對於正則值之百分數差離.

年份	實在值	直線 正則值	二次拋物線 正則值	三次拋物線 正則值	差離 對於直線	差離 對於二次拋物線	差離 對於三次拋物線
1897	13,083	10,342	10,025	12,215	+26.5%	+30.5%	+ 7.0%
1898	11,615	10,486	10,294	11,344	+10.8	+13.2	+ 2.3
1899	9,642	10,631	10,466	10,703	− 9.3	− 8.0	−10.0
1900	9,912	10,776	10,676	10,272	− 8.0	− 7.0	− 3.8
1901	10,618	10,920	10,880	10,030	− 2.5	− 2.0	+ 6.5
1902	9,973	11,0 5	11,076	9,953	− 9.9	−10.0	+ 0.2
1903	9,775	11,210	11,266	10,022	−12.8	−13.2	− 2.5
1904	10,417	11,354	11,448	10,213	− 8.3	− 9.0	+ 2.1
1905	9,967	11,499	11,624	10,506	−13.3	−14.3	− 5.1
1906	9,385	11,644	11,793	10,878	−19.4	−20.4	−13.7
1907	10,274	11,789	11,955	11,313	−12.9	−14.1	− 9.2
1908	14,066	11,933	12,110	11,776	+17.8	+16.0	+19.4
1909	11,872	12,078	12,258	12,258	− 1.7	− 3.2	− 3.2
1910	11,588	12,223	12,400	12,733	− 5.2	− 6.6	− 9.1
1911	12,679	12,367	12,533	13,179	+ 2.5	+ 1.0	− 3.8
1912	13,832	12,512	12,660	13,575	+10.5	+ 9.2	+ 1.9
1913	14,553	12,657	12,781	13,900	+15.0	+13.9	+ 4.6
1914	16,780	12,801	12,895	14,131	+31.0	+30.0	+18.6
1915	19,035	12,943	13,002	14,246	+47.0	+46.2	+33.5
1916	16,493	13,091	13,102	14,225	+26.0	+25.9	+15.9
1917	13,073	13,236	13,194	14,045	− 1.2	− 0.9	− 7.0
1918	9,331	13,380	13,281	13,685	−30.2	−29.8	−31.8
1919	5,515	13,525	13,330	13,123	−59.1	−58.7	−57.8
1920	8,463	13,670	13,432	12,338	−38.0	−37.0	−31.4
1921	19,952	13,814	13,498	11,307	+44.6	+48.0	+76.8

〔幾條長期趨勢線之比較〕

由上面之討論與圖57,可知在某一事件中用不同形式曲線代表長期趨勢所得之結果可相差甚遠.在目前一例中,直線與二次拋物線所取之途徑甚爲接近,惟三次拋物線則表示一完全不同之長期趨勢.此例乃選以特別表示此種差

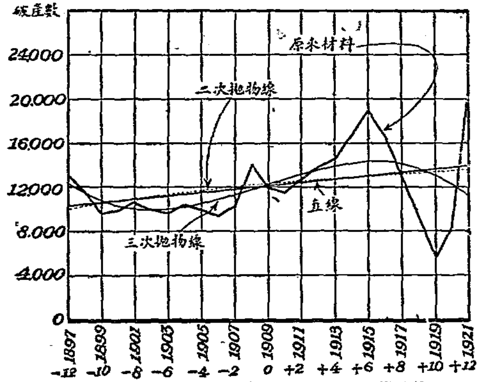


圖 57.——1897—1921 年美國破產數與三條長期趨勢線.

異者.每年之正則值,旣以長期趨勢線上該年之縱坐標作計量,則應用不同之長期趨勢線,結果將有十分不同之代表正則狀態之標準.並且,實際材料對於長期趨勢線之差離所表示之循環,亦因所取之曲線形式不同而不同.在目前之例中,此種情形可於表七十之差離百分數之數字見之.

實際材料對於直線之差離,與其對於二次拋物線之差離甚相切合,惟對於三次拋物線之差離則顯然不同.後者之符號與前二者之符號不同者有四年,其餘尚有許多年之差離,其量相互懸殊.若延長此三次拋物線於材料之範圍以外,所得結果將非常不合理.作外推時,此三種長期趨勢線固皆含有錯誤之危險,然在此種用途上直線與二次拋物線當較三次拋物線為佳.

決定採取何種曲線時所應考慮之點,後面尚須詳述.就本例言,選擇曲線時有若干點須注意者.若實際材料有長時期高於或低於長期趨勢線,則此線之配合或不適當.對於破產之實際材料,無論直線或二次拋物線皆不能與之密合,自1899--1907年間一連九年之實際材料,皆低於此兩長期趨勢線所代表之標準.三次拋物線與材料較為接近,僅就觀察而判斷之,可謂在此材料之限度以內,此線乃較為適合者.

[配合曲線時應用對數之法]

上述之一種曲線為簡單而極有用之形式,然單對數式之幾種曲線,在分析時間數列時,其效用或更為普遍.前文對於用單對數格紙(或稱比率格紙 ratio paper)描繪許多材料的數列之利益,曾加說明.此種作圖法之基本優點在其能真確表示量與量間之相對的變動 (relative variations) 或比率.在分析經濟材料時,此種形式之關係平常為注意中心所在,故長期趨勢之決定應在此同一之基礎上進行.

如此進行時,吾人可利用幾種曲線,其形式與上述之各種曲線大致相同,所不同者,僅爲 y 之地位盡爲 $\log y$ 所替代而已.即直線方程式變爲 $\log y=a+bx$, 而定冪級數之普遍式變爲 $\log y=a+bx+cx^2+dx^3+\cdots\cdots$ 求得之曲線可給於算術格紙上,記 x 之自然數及 y 之對數;或可用單對數格紙,沿 y 軸劃分對數尺度,而 x 與 y 之值皆以自然數給記.兩法中以後一種較爲簡單.

茲歷述配合一條 $\log y=a+bx+cx^2$ 形式之曲線之各種步驟,以說明其方法.問題爲求1908—1922年美國煤油生產額之長期趨勢,正則方程式所需之各數值由下表求之.

表七十一

1908—1922年美國之煤油生產額

計算配合長期趨勢線所需數值

年份	x	y 生產額 (單位：百萬桶)	$\log y$	$x, \log y$	$x^2, \log y$
1908	−7	178.5	2.25164	−15.76148	110.33036
1909	−6	183.2	2.26269	−13.57614	81.45684
1910	−5	209.6	2.32139	−11.60695	58.03475
1911	−4	220.4	2.34321	− 9.37284	37.49136
1912	−3	222.9	2.34811	− 7.04433	21.13299
1913	−2	248.4	2.39515	− 4.79030	9.58060
1914	−1	265.8	2.42456	− 2.42456	2.42456
1915	0	281.1	2.44886	…………	…………
1916	1	300.8	2.47828	2.47828	2.47828
1917	2	335.3	2.52543	5.05086	10.10172
1918	3	355.9	2.55133	7.65399	22.96197
1919	4	378.4	2.57795	10.31180	41.24720
1920	5	442.9	2.64631	13.23155	66.15775
1921	6	472.2	2.67413	16.04478	96.26868
1922	7	557.5	2.74624	19.22368	135.56576
			36.99528	9.41834	694.23282

$n=15$　　　　$\Sigma(x)=0$

$\Sigma(\log y)=36.99528$　　　　$\Sigma(x^2)=280$

$\Sigma(x,\log y)=9.41834$　　　　$\Sigma(x^3)=0$

$\Sigma(x^2,\log y)=694.23282$　　　　$\Sigma(x^4)=9352$

應解之三個正則方程式如下

$$\Sigma(\log y)=na+b\Sigma x+c\Sigma x^2$$

$$\Sigma(x,\log y)=a\Sigma x+b\Sigma x^2+c\Sigma x^3$$

$$\Sigma(x^2,\log y)=a\Sigma x^2+b\Sigma x^3+c\Sigma x^4$$

代入已知值,則得

$$36.99528=15\,a+280\,c$$

$$9.41834=280\,b$$

$$694.23282=280\,a+9352\,c$$

解而求常數,

$$a=+2.450508$$

$$b=+\ .033637$$

$$c=+\ .0008488.$$

故所求之曲線其方程式爲(以1915爲原點)

$$\log y=2.450508+.033637\,x+.0008488\,x^2$$

將代表任何一年之 x 值代入此方程式,即可求得該年長期趨勢值或正則值之對數.於是可得長期趨勢值之自然數.下表列每一年之正則值,以及實在值對於正則值之百分數關係.

表七十二

美國煤油生產額之長期趨勢
及實在值與長期趨勢值之比較
（長期趨勢線爲配合於生產額之對數者）

年份	x	y（實在值）生產額（單位：百萬桶）	長期趨勢之對數	y（計算值）長期趨勢值（單位：百萬桶）	實在值對於長期趨勢值之關係(%)
1908	−7	178.5	2.256639	180.6	98.8
1909	−6	183.2	2.279242	190.2	96.3
1910	−5	209.6	2.303543	210.2	104.2
1911	−4	220.4	2.329540	213.6	103.2
1912	−3	222.9	2.357236	227.6	97.9
1913	−2	248.4	2.386629	243.6	102.0
1914	−1	265.8	2.417720	261.6	101.6
1915	0	281.1	2.450308	282.2	99.6
1916	1	300.8	2.484994	305.5	98.5
1917	2	335.3	2.521177	332.0	101.0
1918	3	355.3	2.559058	362.3	98.2
1919	4	378.4	2.598636	396.9	95.3
1920	5	442.9	2.639913	436.4	101.5
1921	6	472.2	2.682886	481.8	98.0
1922	7	557.5	2.727557	534.0	104.4

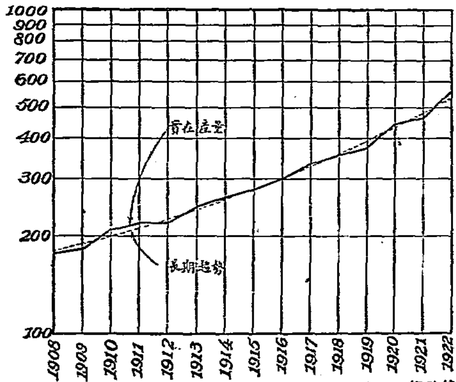


圖58.——1908—1921年美國煤油生產量,及配合於對數之長期趨勢線.

代表實在生產額之各點及長期趨勢線繪於圖58.在此例中,求得之方程式,其曲線爲長期趨勢之極好代表.

欲明此種曲線之優點,可配合一同樣之曲線(二次拋物線)於自然數,而將所得結果與上列之結果比較之.諸常數之值用最小平方法求之,得下列之方程式:

$$y=279.387+24.262\ x+1.650\ x^2\ (\text{以}1905\text{爲原點})$$

由此方程式算得之長期趨勢值,實在生產額,與兩者間之百分比關係列如下表.

表七十三

美國煤油生產額之長期趨勢

及實在值與長期趨勢值之比較

(長期趨勢線爲配合於自然數者)

年份	x	y (實在值) 生產額 (單位:百萬桶)	y (計算值) 長期趨勢值 (單位:百萬桶)	實在值對於長期趨勢值之百分比關係
1908	−7	178.5	190.403	93.7
1909	−6	183.2	193.215	94.8
1910	−5	209.6	199.327	105.2
1911	−4	220.4	208.739	105.6
1912	−3	222.9	221.451	100.7
1913	−2	248.4	237.463	104.6
1914	−1	265.8	256.775	103.5
1915	0	281.1	279.387	100.6
1916	1	300.8	305.299	98.5
1917	2	335.3	334.511	100.2
1918	3	355.9	367.023	97.0
1919	4	378.4	402.835	93.9
1920	5	442.9	441.947	100.2
1921	6	472.2	484.359	97.5
1922	7	557.5	530.071	105.2

此曲線與實在生產數目繪於圖59.

此例中用以代表煤油生產之長期趨勢之兩個方程式,皆含有三個常數,故其結果可作合法之比較.觀察此兩曲線,可知配合於材料之對數者較之配合於自然數者更能代表長期趨勢,惟此僅一般之印象耳,最好能有一確切之計量,此種計量即標準差.計算此標準差之法與處理頻數分配時所

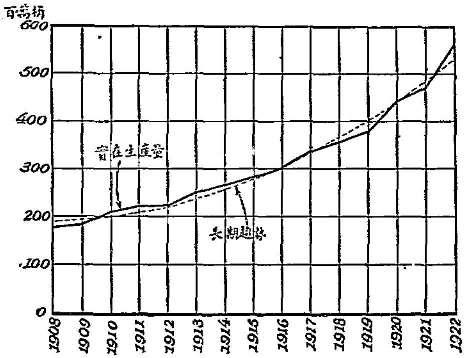


圖59.——1908—1922年美國煤油生產量,及配合於自然數之長期趨勢線

用之法完全相同,不過彼處差離之計算以算術中數爲標準而此處則以長期趨勢線爲標準耳.

當曲線配合於自然數時,標準差爲12.46,配合於對數時,

標準差為 9.44.(兩者皆以百萬桶為單位). 可見在後一條曲線週圍之離中趨勢,較之在配合於自然數之曲線之周圍者小;故在此例中,在計量長期趨勢上對數曲線遙為優勝.(註)

前數節中討論之兩族曲線,足以應經濟統計家大部分之需要.大部分時間數列,其長期趨勢可以定冪級數曲線描寫之,或配合於自然數,或配合於材料之對數(即 y 值之對數;至於為 x 變數之時間因子,在兩種曲線上皆以自然數表現.) 此兩族曲線為伸縮自如而應用甚廣之曲線形式.尚有數種曲線形式於時間數列上應用較狹而在特殊情形中可得完善之結果者,茲略述如下.

平常拋物線式之曲線 ($y=ax^b$), 一般上不適用於取時間數列形式之經濟材料,因若應用時,代表時間之一變數須照幾何級數處理也.前文曾說過,此種曲線作於雙對數格紙上成一直線之形式.惟若對於某一數列之長期趨勢,可用此種形式之曲線確切描寫時,則就經驗之意義上言,此種曲線之應用當為合理.

(註) 上述之標準差係根據差離之實在數計算而來,以煤油百萬桶為單位;若根據差離之百分數計算,對數曲線之優越性更為明顯,此乃意中之事也.如此求得之標準差,在配合於自然數之曲線為 4.01, 在對數曲線為 2.69.

應注意者,比較標準差之大小以測驗長期趨勢曲線之法,須參加比較之曲線,其方程式所包含之常數個數相等時,方為合法.若常數多至與所擬之點數相等,則所作之曲線可通過每一點,此時標準差等於零,惟此種曲線決然非長期趨勢之計量.

配合此種曲線之最簡便之方法爲應用對數而用一直線形式之方程式.方程式

$$y=ax^b$$

取對數形式時,變爲

$$\log y=\log a+b\log x.$$

配合此種曲線時所須之兩個正則方程式爲

$$\Sigma(\log y)=n\log a+b\Sigma(\log x)$$

$$\Sigma(\log x.\log y)=\log a\Sigma(\log x)+b\Sigma(\log x)^2$$

以根據材料計算而得之數值代入上列方程式,可解之而求得 $\log a$ 與 $\log b$,其法與配合一平常之直線之法相同.(註)

簡單之指數曲線($y=ab^x$)在分析時間數列時可應用甚廣.取對數之形式時變爲

$$\log y=\log a+(\log b)x.$$

配合此種曲線時所需之兩個正則方程式爲

$$\Sigma(\log y)=n\log a+\log b\Sigma(x)$$

$$\Sigma(x\cdot\log y)=\log a\Sigma(x)+\log b\Sigma(x^2)$$

此爲前文曾經敍述之關於單對數族的曲線之直線形式.解正則方程式而得之兩常數,再求其逆對數,則吾人可使所求之方程式取指數曲線形式.

(註) 萊蒙潘爾 (Raymond Pearl) 氏所作 Medical Biometry and Statistics, (Philadelphia, Saunders, 1923.)之附錄中,有一表包括自一至一百諸自然數之對數之和,甚爲有用.

貢波玆曲線(Gompertz curve)亦曾用於經濟統計之說明.其方程式原爲保險業上之計算而作.方程式爲

$$y=ab^{c^x}$$

應用此式於經濟統計之分析,其理由卽人口之增加依照一普遍之法則,而在有些工業,其生產量成爲人口增加之直接函數者,亦有此法則存在也.(註1)

有一多少與前者相似之增長曲線,其方程式爲

$$y=d+\frac{b}{e^{-ax}+c}$$

者,萊蒙潛爾氏與李德(Lowell J. Reed)氏曾用以預測人口之增加.(註2)此種曲線亦曾用以描述某種經濟材料之長期趨勢.

[配合曲線之幾種簡法:平均數法]

求高度精確之結果時,曲線之配合宜用最小平方法.若僅求一近似之值,則配合曲線尙有其他方法可用.

用平均數法(method of averages)配合一直線時,先分數列爲兩個等分或近於相等之兩部分.由此觀察值之兩集團,

(註1) 潑萊斯各脫(Raymond B. Prescott)氏在"Law of Growth in Forecasting Demand"一文(載 Journal of American Statistical Association, 1922年十二月號147—79)中對於此曲線之特質曾加敍述.

(註2) 見潛爾氏與李德氏合著之"Predicted Growth of Population of New York and its Environs", 1923年 Committee of Plan of New York and its Environs 出版.

得兩個方程式,其形式皆如下式:

$$\Sigma(y)=na+b\Sigma(x)$$

聯立解之,得 a 與 b 之值.此法等於就數列之每一半各求一點,其坐標等於 x 諸值與 y 諸值之算術平均數者,而以一直線連結之.平常如此求得之線不與用最小平方法求得者相合.

此法可利用表六十九之破產數材料以說明之.數列共包括25年,可將13個觀察值歸入第一集團,以12個歸入第二集團,求得之兩個方程式爲

$$140{,}629=13\,a-78\,b$$

$$161{,}329=12\,a+78\,b$$

解之得

$$a=12{,}078$$

$$b=210$$

長期趨勢線之方程式爲

$$y=12{,}078+210\,x$$

用最小平方法,其方程式爲

$$y=12{,}078+144.7\,x$$

此後一個方程式所代表之曲線自最爲適合.此例中用簡法求得之方程式與正確之結果相差甚大.當原來材料與長期趨勢線相離甚遠有如此例者,情形常爲如是.

[選點法]

配合於某種材料之曲線,其形式已經決定時,常數之近似值,可以選點法 (method of selected points) 求之.所選點數與所須之常數之個數相等.將此各點之坐標代入正則方程式,解而求常數.如欲配合一三次拋物線於破產材料,必須選定四點.求選擇適當,可先作照觀察作一試驗曲線.下列諸點可作計算之基礎.(參考表六十九及圖57等關於破產之材料).

x	y
−10	10,000
− 5	10,000
+ 5	14,000
+10	13,000

將上例 x, y 諸值代入所求曲線之正則方程式得

$$10{,}000=a-10\,b+100\,c-1{,}000\,d$$

$$10{,}000=a-5\,b+25\,c-125\,d$$

$$14{,}000=a+5\,b+25\,c+125\,d$$

$$13{,}000=a+10\,b+100\,c+1{,}000\,d$$

聯立解之得下列之常數值:

$$a=+12{,}166$$

$$b=+483$$

$$c=-6.7$$

$$d=-3.3$$

故所求之方程式為

$$y=12{,}166+483\,x-6.7\,x^2-3.3\,x^3$$

由此可見此方程式與用最小平方法求得之方程式之差異.

用此等簡法處理平常經濟材料,錯誤之範圍甚大;蓋經濟材料大部分不與長期趨勢線極爲接近也.經濟材料相當於觀察差誤甚大之物理上的計量.當物理材料依據一定法則,而觀察差誤爲微小時,平均數法與選點法之應用,尙不致引起重大錯誤.惟經濟材料常與任何之長期趨勢線相差甚巨,故一般上不宜應用簡法.

[代表長期趨勢之曲線之選擇]

上文敍述可以配合於一時期之經濟材料以代表其長期趨勢之各種曲線形式.但此許多形式中,在每一種情形之下應選擇何種?何種形式能使所包括之各年每年皆得一最佳之正則標準?前數節中對此問題亦曾提及,惟未曾規定其一般的原則.而實際上對此基本問題亦無一般的原則可以答復,在此種情形中適合與否並無絕對標準可供測驗.在每一例中究以何種曲線最能代表長期趨勢,大都決定於個人之判斷,在如此之判斷中,經驗盡一主要之作用.惟仍有若干平常應加考慮之點,足以助吾人挑選一適當之曲線形式,有時並能使吾人單獨挑選一種而確信其爲最佳者.玆述如下.

1. 將材料繪成圖式,爲選擇曲線形式之第一步.作圖既成,吾人常可由觀察而決定適當之形式.描繪材料之法可有四種,處理經濟材料時以前二者最爲重要.

a. x 自然數, y 自然數.(即將材料之數字繪於平常

算術格紙上.)

b. x 自然數,y 對數.(依自然數尺度給 x 諸値,依對數尺度給 y 諸値.即用單對數格紙.)

c. y 自然數,x 對數.(用單對數格紙,x 一邊用對數尺度.)

d. y 對數,x 對數.(用雙對數格紙.)

若在幾種情形中有一種得一直線長期趨勢,則可以選擇如此一種方程式,卽在作圖時在一定條件下(參觀第二章)成爲直線者.若不能得一直線方程式,則圖上之材料或可提示其他某種簡單形式.硏究原來材料之曲線以選擇一代表長期趨勢之曲線時,吾人須熟識代表一切簡單方程式之種種曲線.(註)

2. 適當之曲線又可硏究 x,y 兩變數間之關係而決定之.最簡單之情形中,有下列各種關係:(注意,算術級數爲按固定之絕對增量而變動者,幾何級數爲按固定之百分數而變動者.)

a. 若當 x 諸値形成算術級數時相當之 y 諸値形成一幾何級數,其關係爲指數式,此種形式之關係用下列方程式代表之:

(註) 下列各書中可見各種曲線之圖:
Empirical Formulas, T. R. Running, N. Y. Wiley, 1917.
Engineering Mathematics, C. P. Steinmetz, N. Y. MacGraw-Hill, 1917.
Graphical and Mechanical Computation, Joseph Lipka, N. Y. Wiley, 191[illegible]

$$y=ab^x.$$

b. 若當 x 諸值形成幾何級數時相當之 y 諸值形成一幾何級數,其關係爲簡單之拋物線式或雙曲線式,用方程式代表之,其式爲

$$y=ax^b.$$

c. 若當 x 諸值形成算術級數時相當之 y 諸值之"第一級差" (first differences or first order differences) 爲一常數,其關係爲一直線形式,用方程式表示之,其式爲

$$y=a+bx$$

當 x 諸值形成一算術級數時, y 之依次相繼之值間之差,名爲"第一級表", 用 Δy 之記號代表之.依次相繼之第一級差間之差,名爲"第二級差", 以 $\Delta^2 y$ 代表之.其他較高級之差亦同樣求得.下表示諸差之構成:

x	y	Δy	$\Delta^2 y$	$\Delta^3 y$
1	11	29		
2	40	61	32	12
3	101	105	44	12
4	206	161	56	12
5	367	229	68	12
6	596	309	80	12
7	905	401	92	12
8	1306	505	104	12
9	1811	621	116	
10	2432			

d. 若當 x 諸值形成算術級數時相當之 y 諸值之第

n 級差爲一常數,則變數間之關係可用包括 x^n 之定冪級數方程式代表之,其式爲

$$y=a+bx+cx^2+dx^3+\cdots\cdots qx^n$$

如前例之以第三級差爲常數者, x 與 y 之關係可用下列方程式代表之:

$$y=a+bx+cx^2+dx^3.$$

此等測驗與其他與此類似之測驗,可應用於無數其他形式之材料者,在倫寧氏及呂潑加氏書中(書名見前節註釋)曾有敘述.在此應特加聲明者,當選擇一曲線用以分析經濟材料時,此等測驗恰能完全適合於材料之情形,雖有之亦甚少.此種情形僅於所選之曲線通過一切已知點時始能存在.惟每一種材料平常總能近似於上述諸種情形之一,而使吾人知何者爲最適當之曲線形式.

3. 若研究原來材料時不能得一確斷,則可先配合幾種形式之曲線,然後比較其結果而決定之;有如前面所舉破產及煤油生產之例.若兩個曲線之方程式皆含有同個數之常數,則取兩者之標準差而比較之,其結果,在材料之限度以內,爲配合的密切程度之確實有效的測驗.

應用下式,標準差甚容易求得:

$$\Sigma(d^2)=\Sigma(y^2)-a\Sigma(y)-b\Sigma(xy)-c\Sigma(x^2y)-\cdots\cdots$$

式中 $\Sigma(d^2)$ 爲長期趨勢線周圍之差離之平方總和.(此方程式之由來見附錄一,該處可見其普遍式.)若方程式所包含之

常數其個數各不相等,此種測驗乃爲無效,而吾人僅可恃觀察比較之.遇此種情形,欲決定某一曲線最能確切代表長期趨勢,必須以個人之判斷爲依據.

此地應重新指出,在材料範圍以內之配合之密切程度,單就其本身言並非一最後之標準.取一方程式,其所包含之常數個數與代表材料之點數相等者,可作一曲線通過每一已知點,惟如此之一曲線不一定代表長期趨勢.長期趨勢之觀念爲一有規則而光滑之潛在運動,雖有與之不合之差離存在,但能顯示一數列之長時期的傾向.因此,一般言之,曲線應取簡單之形式,始能合於長期趨勢之觀念.然此非謂一複雜之長期趨勢,可以用一不合於已知材料之簡單曲線代表之也.

4. 作曲線之前,有一重要問題首須答復者,卽長期趨勢線用於何種限度以內是也.若僅用於代表已知材料之範圍以內(卽用於內推 Interpolation)之情形者,決定選擇何種曲線之條件爲一種.若用以推至材料之範圍以外,用爲決定下一時期的正則之根據者,則條件爲另一種.前者只須曲線能合理地適合於材料,後者不僅如此,且須求延長線之趨勢合理,與過去之紀錄相一致.此點在上文中討論破產之長期趨勢時曾經特別指出.

宜清楚認識者,延長或外推(Extrapolation)爲一種猜測,僅於假定已得一適當之長期曲線,且作用於數列之情況在

將來與過去者相同之時,始為合理.情況之變更,新因素之加入,足使此延長歸於無效.並且在處理經濟統計時,除回顧過去之事實外,常不能知何時曾起變動.延長長期趨勢線而求得之結論,因此常致錯誤.在實際統計工作中,應用此種延長方法,其理由為過去發展之途徑乃將來發展之最大可能之途徑.延長於較遠之將來者,較之延長於短期之將來者,其錯誤之域自為更大.故當取得新材料時,長期趨勢線宜時加修改.

欲延長一曲線時,包含常數之個數較少之簡單曲線比較複雜之曲線為佳.在某一情形中,一三次或四次拋物線可非常適合於材料,惟不宜與以延長.潘林(Perrin)氏曾言,一曲線之適於內推者,可全不適於外推,毋忘斯言.

一般言之,配合於 y 諸值之對數之簡單曲線,比較配合於 y 之自然數者,當延長時,所得結果較為可恃.加爾斯登(Karl G. Karsten)氏在一關於此點之有趣味之討論中,(註)主張按定率變動之現象,較之按定量變動之現象,較能維持其固有之長期趨勢,最能計量變動之率者,當然為單對數曲線.

5. 當配合一長期趨勢線之目的為研究與商情循環有關之差離時,另有一法可以測驗配合之確當與否,此法卽以差離之循環與許多其他數列所表示之實際商情相比較也.

(註) Charts and Graphs, 423-425.

一般上長期趨勢線上所表示之循環,能與一般之商情之循環相合符者,此線即爲最佳.今可取配合於破產材料之數曲線而以此法測驗之.在其所包括之時期內,商情之大循環反映於對於直線及二次拋物線之差離中,小循環則否.此兩條長期趨勢線表示在1899—1907年整個時期內破產之數皆低於正則値,當長期趨勢用三次拋物線表示時,1901與1904兩年破產之數略高於正則值.此表示1899—1907年整個時期雖一般上爲繁榮,然此兩年在小循環中爲蕭條之局面.於此可見高次曲線與材料更爲接近,使較小之變動表現爲循環.至於在某一情形中此較小之變動是否宜於作爲循環,則須視研究之性質而決定之.

6. 在吾人所欲研究之整個時期中,有時其已知之數列竟無一曲線可以與之配合者.此種情形亦非罕見.其原因或由於情況之變更,致長期趨勢改變.如美國批發物價長期趨勢,自南北戰爭結束以後直至1896年,爲向下的發展,可以一直線確切代表之.自1896年至世界大戰之初,長期趨勢爲向上的發展,可以一二次拋物線代表之.與此同樣之變動,存在於許多經濟數列之中.將整個時期分爲幾段,可就如此劃分之各段各作一長期趨勢線.惟此法應用過度可得完全不合理之結果.長期趨勢之觀念爲一漸次發生而長期之變動,分割一數列爲幾段而配合許多長期趨勢線,實違反此整個觀念.此法惟在情況中有眞正的變化時始爲應用適當,惟在無

論何種情形中,總宜力求以單獨一線代表整個時期之長期趨勢.

[用一相關之統計數列代表長期趨勢之法]

一數列之長期趨勢,有時不可能以一數學曲線代表.遇可得之材料僅限於短期,或因新因子之參加而舊時數列之長期趨勢今已顯然變更,則此種情形即可發生.大戰開始以後之商品價格,即其例也.戰前之長期趨勢業已大大改變,更無從確斷將來之長期趨勢.潘蓀氏遇此困難,乃創一新法以應付之,此法即用一數列爲基,而計算另一相關之數列之循環變動也.(註 1)

關於物價方面,所用之兩個數列爲勃拉特斯脫利氏批發物價指數與潘蓀氏商情循環物價指數.此兩種指數構成之法前文業已詳述.(註 2)勃拉特斯脫利氏指數依據96種商品而作,潘蓀氏指數依據10種商品而作,兩者之變動在時間上大致相同,惟後者之變動較前者大爲劇烈.因此,潘蓀氏旣單以計量商情循環爲目的,乃用勃拉特斯脫利氏指數爲計量包括10種商品而較爲敏感之指數之基數.根據此基數以計算差離,與根據長期趨勢線而計算之法完全相同.

此法亦用於一條計算利率變動的基線之決定.十條主

(註 1) Review of Economic Statistics, Harvard Economic Service, April, 1923, 73—74, July, 1923, 192—194.

(註 2) 參考第六章.

要鐵路債劵之利息,形成變動較小之數列,因此用以爲基數,以測量與之同時變動而較之爲敏感之二個數列,即 60—90 天期票利率與4—6月期票利率.兩者爲使其"交點"(即代表每一數列之線與基線相交之點)能與其他一般商情指數之數列之"交點"相合,各須有若干修整,修整之法詳見章末參考書中.

[作爲分析工作的一步之緊縮]

經濟材料之許多數列以貨幣單位計值,如圓,鎊,法郎等,此種數列常因物價水平之變動而失其眞相.如美國二十七邦 1913 年建築工程之契約價值爲 85,800 萬美金; 1922 年契約價值在同一區域內者合計334,400萬美金.(註) 1922年之建築量眞能相當於 1913 年之四倍乎?其實不然.每年之建築契約價值,其決定不但依據於實際之建築量,且須依據建築材料與建築勞動之費用之高低;而自 1913—1922年間,此後者之價格增加甚多.若吾人單欲知建築量之變動,則上述之價值中必須除去因1913—1922年間建築費用增加而起之變動.此種校正之法名爲緊縮 (Deflation).

遇此種情形,中心問題爲適當之緊縮指數 (Deflating index) 之選定.在討論指數之一章中,貨幣工資依照生活費指數而緊縮,成爲實在工資變動之計量.吾人可依下列之方法,依照建築材料批發價格指數緊縮建築契約之價值:

(註) 根據 F. W. Dodge & Co. 編製之數字.

(1)	(2)	(3)	(4)
年份	建築契約價值	建築材料價格指數	建價契約緊縮價值
1913	858,000,000元	100	858,000,000元
1922	3,344,000,000元	168	1990,000,000元

最後一行之值,乃以(3)行之指數除(2)行中與之相當之實際價值再乘100而得.結果將實際價值化爲依照基期(在此例爲1913年)之物價而計算之價值.緊縮價值表示1922年實際建築量增至1913年建築量兩倍以上.

若將工資變動一併計算,緊縮之結果將更爲正確.美國電話電報公司(A. T. & T. Co.)作一建築費用指數,將建築材料價格之變動與建築工資之變動兼包在內,用如此之指數緊縮建築契約之價值,情形見於下表:

表七十四

建築契約之實際價值與緊縮價值

	建築契約價值(單位:百萬圓)	A.T.&T.公司建築費用指數(1914=100)	建築契約緊縮價值(單位:百萬圓)
1914			
正月	51.1	100	51.1
二月	39.1	101	38.7
三月	58.1	102	57.7
四月	79.7	101	78.9
五月	72.0	100	72.0
六月	81.8	100	81.8
七月	72.0	100	72.0
八月	77.3	100	77.3
九月	47.1	100	47.1
十月	53.4	99	53.9
十一月	45.5	99	46.0
十二月	42.3	98	43.2

表七十四(續上)

建築契約之實際價值與緊縮價值

	建築契約價值(單位：百萬圓)	A.T.&T.公司建築費用指數(1914=100)	建築契約緊縮價值(單位：百萬圓)
1915			
正　月	43.3	99	43.7
二　月	48.8	99	49.3
三　月	75.6	99	76.4
四　月	76.5	100	76.5
五　月	77.1	102	75.6
六　月	92.2	104	88.7
七　月	94.7	104	91.1
八　月	90.4	102	88.6
九　月	82.0	103	79.6
十　月	88.6	104	85.2
十一月	88.0	107	82.2
十二月	82.9	112	74.0
1916			
正　月	62.8	118	53.2
二　月	66.3	121	54.8
三　月	94.5	124	76.2
四　月	100.9	127	79.4
五　月	131.4	127	103.5
六　月	140.7	128	109.9
七　月	114.4	125	91.5
八　月	127.0	122	104.1
九　月	132.2	123	107.5
十　月	151.4	125	121.1
十一月	122.4	128	95.6
十二月	112.9	136	83.0
1917			
正　月	90.8	138	65.8
二　月	95.2	141	67.5
三　月	132.7	144	92.2
四　月	148.5	148	100.3

表七十四(續上)

建築契約之實際價值與緊縮價值

	建築契約價值(單位:百萬圓)	A.T.&T.公司建築費用指數(1914=100)	建築契約緊縮價值(單位:百萬圓)
1917			
五月	157.6	151	101.7
六月	206.5	160	129.1
七月	159.2	171	93.1
八月	165.6	170	97.4
九月	122.5	166	73.8
十月	154.5	154	100.3
十一月	94.3	155	60.8
十二月	90.8	155	58.6
1918			
正月	161.6	156	103.6
二月	137.3	161	85.3
三月	115.3	162	71.2
四月	128.9	164	78.6
五月	120.4	167	72.1
六月	248.2	171	145.1
七月	153.0	176	86.9
八月	146.4	180	81.3
九月	124.5	180	69.2
十月	166.1	183	90.8
十一月	130.3	183	71.2
十二月	57.3	186	30.8
1919			
正月	54.1	186	29.1
二月	98.7	186	53.1
三月	121.8	186	65.5
四月	188.8	178	106.1
五月	234.7	178	131.9
六月	285.4	180	158.6
七月	317.7	183	173.6
八月	295.1	194	152.1

表七十四(續上)

建築契約之實際價值與緊縮價值

	建築契約價值 (單位:百萬圓)	A.T.&T.公司 建築費用指數 (1914=100)	建築契約緊縮價值 (單位:百萬圓)
1919			
九　月	228.7	199	114.9
十　月	307.5	201	153.0
十一月	220.6	207	106.6
十二月	226.7	221	102.6
1920			
正　月	226.1	227	99.6
二　月	205.3	240	85.5
三　月	302.2	248	121.9
四　月	304.9	254	120.0
五　月	264.9	261	100.3
六　月	260.1	266	97.8
七　月	204.5	265	77.2
八　月	202.7	267	75.9
九　月	182.2	261	69.8
十　月	178.6	256	69.8
十一月	132.9	250	53.2
十二月	100.1	236	42.4
1921			
正　月	111.6	231	48.3
二　月	102.4	223	45.9
三　月	163.8	221	74.1
四　月	221.3	212	104.4
五　月	240.9	206	116.9
六　月	226.2	201	112.5
七　月	212.2	197	107.7
八　月	220.4	194	113.6
九　月	244.9	188	130.3
十　月	222.4	186	119.6
十一月	190.8	187	102.0
十二月	198.3	186	106.6

表七十四(續上)

建築契約之實際價值與緊縮價值

	建築契約價值(單位:百萬圓)	A. T. & T. 公司建築費用指數(1914=100)	建築契約緊縮價值(單位:百萬圓)
1922			
正月	165.3	185	89.9
二月	172.7	185	93.4
三月	293.4	187	155.9
四月	353.0	186	189.8
五月	361.8	188	192.4
六月	342.4	195	175.6
七月	350.1	199	175.9
八月	322.0	197	163.5
九月	271.5	201	135.1
十月	253.1	201	125.9
十一月	243.4	197	123.6
十二月	214.2	196	109.3
1923			
正月	218.7	202	108.3
二月	229.9	209	110.0
三月	333.2	212	159.5
四月	362.9	217	167.2
五月	373.9	217	172.3
六月	323.6	221	146.4
七月	274.2	219	125.2
八月	258.1	219	115.6
九月	253.5	217	116.8
十月	319.9	216	148.1
十一月	289.3	211	137.1
十二月	267.9	215	124.6

實際價值與緊縮價值皆繪於圖60.

價值數列(value series)大部分受價格變動之影響,在進一步分析以前,常宜先予此因子以校正.然每一數列代表一

新問題,無一指數可以普遍適用於一切數列之緊縮者.美國勞工統計局所編製之批發物價指數,曾廣泛應用於緊縮以金圓記值之經濟材料,惟在此應用中此指數對於許多場合全不適當.例如用批發物價指數以緊縮貨幣工資之値乃不合理之事.蓋所用之緊縮指數,其計量之價格變動,應爲與吾人所欲緊縮之數列有關者也.

茲舉一有趣之例;紐約聯邦準備銀行統計部對於紐約以外之銀行淸算額予以緊縮時所用之方法,爲取下列諸數

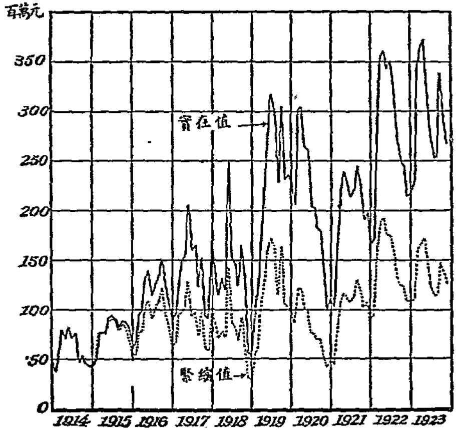


圖 60.——1914—1923 年建築契約逐月實在值與緊縮値之比較.

列加權而結合之,以爲緊縮指數;

地租指數……………………………(權數)1

批發物價指數……………………………2

工資指數……………………………$3\frac{1}{2}$

生活費指數……………………………$3\frac{1}{2}$

如是作成之緊縮指數,頗能除去物價變動與工資變動對於紐約以外的銀行清算額之影響.緊縮後所得之數列,常被取以代表美國貿易額之變動.

緊縮一價值數列,常爲研究該數列之第一步.用本章及以後數章中所說明之方法,乃得開闢一更進一步分析研究之途徑.

參考書

Bowley, A. L. Elements of Statistics (132—169)

Davies, G. R. Introduction to Economic Statistics (100—130).

Karsten, Karl. Chats and Graphs (477—481).

Lipka, Joseph. Graphical and Mechanical Computation.

Moore, H. L. Forecasting the Yield and the Price of Cotton (28—37).

Pearl, Raymond. Medical Biometry and Statistics (332—341).

Persons, W. M. Indices of Business Conditions. Review of Economic Statistics. Prel. Vol. I, 1919 (1—107).

Running, T. R. Empirical Formulas.

Steinmetz, C. P. Engineering Mathematics (209—255).

關於最小平方法之參考書見附錄一.

密勒氏統計方法論

下　冊

徐堅譯

商務印書館發行

密勒氏統計方法論

下　册

F. C. Mills 著
徐　堅譯

商務印書館發行

第八章

時間數列之分析:季節變動及循環變動之計量

分析時間數列時有許多問題,長期趨勢之計量僅屬其一.前曾指出,此種數列受季節性與循環性之週期變動所影響;而此種變動對於商情之作用,平常較之長期趨勢尤爲重要.下表中之數列,受季節影響極爲顯著,吾人可用以說明計量此種季節傾向之方法.

表七十五(註)

美國生產者所得之雞蛋平均價格

(1910–1921)

	1910	1911	1912	1913	1914	1915	1916	1917	1918	1919	1920	1921
正　月	30.5	30.4	29.5	26.8	30.7	31.6	30.6	37.7	45.3	57.2	64.8	61.1
二　月	28.9	22.1	29.1	22.8	26.4	29.2	26.8	35.8	49.4	48.3	56.9	49.6
三　月	22.9	16.5	24.5	19.4	24.2	21.3	21.2	33.8	40.4	33.1	46.6	29.2
四　月	18.6	14.9	17.8	16.4	17.6	16.6	17.9	25.9	31.2	34.3	38.8	20.4
五　月	18.6	14.7	17.1	16.1	16.8	17.1	18.1	30.0	31.0	36.8	37.4	20.2
六　月	18.3	14.5	16.7	16.9	17.3	16.6	19.0	31.1	29.8	38.6	37.0	19.4
七　月	18.2	14.2	16.7	17.0	17.6	16.8	19.7	28.3	30.7	36.8	36.7	22.0
八　月	17.6	[illegible]	17.4	17.2	18.2	17.0	20.7	29.8	34.4	39.3	40.0	26.6
九　月	19.4	17.4	19.1	19.5	21.0	18.7	23.3	33.2	36.4	41.0	44.2	30.4
十　月	22.4	20.0	22.0	23.4	23.5	22.3	28.1	37.4	41.6	44.7	50.1	34.2
十一月	25.3	23.5	25.9	27.4	25.3	26.3	32.2	30.4	47.2	54.0	58.9	44.2
十二月	29.0	26.7	29.7	33.0	29.7	30.6	38.1	43.3	55.0	61.9	65.0	51.1
平均數	22.5	19.4	22.1	21.3	22.5	22.0	24.9	33.8	39.5	43.8	47.9	34.0

(價格以每打值洋若干分計算,平均價格係月第一日之價格之平均數)

(註) 此材料係美國農業部所編製.

〔就每月份之各項求算術平均數之法〕

計量季節變動時,比較上爲簡單而直接之方法,爲就每月份之許多項求算術平均數.上表之材料包括十二年,故正月份,二月份以及其他諸月份各各有十二項數字.按月份而計十二項之平均數,則得表七十六之第(2)項.

表七十六

根據實際價格之算術平均數擬製季節變動指數之方法

(1) 月份	(2) 每月鷄蛋價格之算術平均數	(3) 除去長期趨勢之影響後之算術平均數	(4) 季節變動之指數
正月	每打洋 39.77 分	每打洋 40.73 分	138.8
二月	35.61	36.38	123.9
三月	27.76	28.34	96.5
四月	22.53	22.92	78.1
五月	22.82	23.01	78.4
六月	22.93	22.93	78.1
七月	22.89	22.70	77.3
八月	24.47	24.08	82.0
九月	26.97	26.39	89.9
十月	30.81	30.04	102.3
十一月	35.63	34.66	118.1
十二月	41.26	40.10	136.6
平均數	....	29.357	100.0

第(2)項之簡單算術平均數,對於鷄蛋價格季節變動之範圍可予少許指示.由正月至四月價格慘跌,四月以後逐漸上漲,經七月之微跌,至十二月達最高點.若非因長期趨勢對於此平均數之影響,則將此種數字化成百分數,即可用爲季

節變動之計量.自1910—1920,年雞蛋之價格大漲,此種漲價與季節變動同樣反映在各月之平均數中.單就長期趨勢而言,平均上每逢二月則價格超過比其前之一月正月,每逢三月則超過比其前一月之二月;其他各月情形亦同;故若欲單獨計量季節之變動,必須由各月之平均數內除去長期趨勢之影響.

將一直線配合於1910—1921年之十二個按年計算之雞蛋價格平均數時,可見在此一時期內,每打雞蛋之價格,平均每年增加2.32分.每月平均增加之數爲此數之十二分之一,即.193分.故每一月之平均價格皆受此數之影響.若長期趨勢之力量單獨發生作用,則二月份各項之平均數將大於正月份各項之平均數.193分,三月份平均數亦大於二月份平均數.193分,其餘各月情形相同.此因子之影響極易除去,祇須將任何一月爲基期,而將其餘各月之數字改正.如取正月爲基期,吾人可由二月份之平均數中減去.193分,在三月份之平均數中減去.386分,由四月份之平均數中減去.579分,其餘可類推.若取十二月爲基期,則吾人應於十一月份平均數上加.193分,於十月份平均數上加.386分,其餘可類推.在目前之研究上,最好取近於一年中心之月爲基期,故取六月爲基期,從七月份平均數減去.193分,從八月份平均數減去.386分,……而在五月份平均數上加上.193分,在四月份平均數上加上.386分,其餘類推.如此校正過後所得之平均數列於表七

十六第(3)行中.

此行諸平均數所代表者,單爲所研究之十二年中季節變動之影響.爲應用便利起見,最好化爲百分數形式,以各月份數字之平均數(29,357)爲基數,如此求得之百分數列於表七十六第(4)行中,此種指數在時間數列之分析中之用途將於後文討論之.

[環比法]

潘莎氏另創一法以計算季節變動.此法曾用於幾種重要之研究中,最著者卽潘莎氏商情指數之編製也.環比法(method of link relatives)之第一步爲計算環比,將所研究之時期內每月之數字,化爲其前一月之數字之百分數,如1910年正月之數字,化爲1909年十二月數字之百分數.1910年二月之數字,化爲1910年正月之數字之百分數.餘可類推.下表乃根據表七十五計算而得之環比.

表七十七

每月雞蛋價格之環比

月份	1910	1911	1912	1913	1914	1915
正月	107.4	104.8	102.8	97.2	93.0	105.4
二月	94.8	72.7	98.6	85.1	92.5	92.4
三月	79.2	74.7	84.2	85.1	85.2	72.9
四月	81.2	90.3	72.7	84.5	72.7	77.9
五月	100.0	95.7	96.1	98.2	95.5	103.0
六月	98.4	98.6	97.7	105.0	103.0	97.1
七月	99.5	97.9	100.0	101.0	101.7	101.2
八月	93.7	109.2	104.2	101.2	103.4	101.2
九月	110.2	112.3	109.8	113.4	115.4	110.0
十月	115.5	114.9	115.2	120.0	111.9	119.3
十一月	112.9	117.5	117.7	117.1	107.7	117.9
十二月	114.6	122.1	114.7	120.4	117.4	116.3

表七十七(續上)
每月雞蛋價格之環比

月　份	1916	1917	1918	1919	1920	1921
正　月	100.0	99.0	106.9	104.0	104.7	94.0
二　月	87.6	95.0	106.7	84.4	87.8	81.2
三　月	79.1	94.4	81.8	68.5	81.9	68.9
四　月	84.4	76.6	77.2	103.3	83.3	69.9
五　月	101.1	115.8	99.4	107.3	96.4	99.0
六　月	105.0	103.7	96.1	104.9	98.9	96.0
七　月	103.7	91.0	103.0	95.3	99.2	113.4
八　月	105.1	105.3	112.1	106.8	109.0	120.9
九　月	112.6	111.4	105.8	104.3	110.5	114.3
十　月	120.6	112.6	114.3	109.0	113.3	112.5
十一月	114.6	105.3	113.5	120.8	113.6	129.2
十二月	118.3	109.9	116.5	114.6	114.2	115.6

於是進而求正月份諸價格對於十二月份諸價格;二月份諸價格對於正月份諸價格等等月份與月份間之平均的關係.在前例中,此種關係用每月份諸價格之算術平均數化成百分數以代表之.潘蓀氏方法乃取諸環比之中位數代表之.即將正月份之許多環比順大小爲序排列而取其中位之値,其餘各月份亦同樣處置.如此求得之各月份之中位環比(median link relatives)列於表七十八第(2)行中:

表七十八
用環比法編製季節變動指數之法

(1) 月　份	(2) 每月雞蛋價格之中位環比	(3) 鏈比	(4) 改正後之鏈比	(5) 季節變動之指數
正　月	103.4	100.0	100.0	139.5
二　月	90.1	90.1	89.7	125.1
三　月	80.5	72.5	71.9	100.3
四　月	79.5	57.4	56.7	79.1
五　月	99.2	56.9	56.0	78.1
六　月	98.7	56.2	55.1	76.9
七　月	100.5	56.5	55.2	77.0
八　月	105.2	59.4	57.8	80.6
九　月	111.0	65.9	63.9	89.1
十　月	114.6	75.5	72.9	101.7
十一月	115.8	87.4	84.0	117.2
十二月	116.0	101.4	97.1	135.4
平均數	……	……	71.69	100.0

(2) 中位環比爲根據移動基期而計量逐月之變動,每月爲其後來之一月之基期.吾人必須進一步將逐月之變動表現爲對於一固定基期之百分數.因此以正月爲基期而計算鎖比(chain relatives);正月份之數字爲100;二月份爲90.1,與中位環比相等,因二月份之一切環比皆係根據正月份爲基期而作者也.但三月份之中位環比爲以二月份爲基期而求得之,80.5以正月份爲基期而求得之三月份之鎖比將爲80.5×90.1或72.5同樣,任何一月之鎖比,皆可以該月之中位環比乘前一月之鎖比而求得,如此求得各月之鎖比列於表七十八第(3)行中.

若材料中無向上或向下之長期趨勢存在,則以十二月份之鎖比乘正月份之中位環比,其積將爲100.實際上,則101.4與103.4(此後一數目爲正月份之環比)之乘積等於104.8.此差異之由來,因有一固定之長期趨勢存在.而影響於十二個環比之每一個也.吾人應注意此差異爲一累積的差誤之結果,當中位環比依次被乘於鎖比時,錯誤重複凡十二次.此種情形與本金按一定之複利率增殖之情形相同.此地之差誤(因長期趨勢而起之每月增加率)相當於利率.一定數目之本金(P_0)按固定利率複式增殖n年時其增殖之公式如下:(P_n代表n年末之本金):

$$P_n=P_0(1+r)^n$$

增加之率以下式代表之

$$r=\sqrt[n]{\frac{P_n}{P_0}}-1$$

可將此處之數值代入公式：(註)

$$104.8=100(1+r)^{12}$$

$$r=\sqrt[12]{\frac{104.8}{100}}-1$$

$$=1.004-1$$

$$=.004$$

故知因長期趨勢而起之增加，其率爲每月 .004. 欲消除此因子之影響，二月份之鏈比必須被除於1.004 卽 $(1+r)$，三月份之鏈比必須被除於1.008，卽 $(1+r)^2$，其餘類推；而十二月份之鏈比則被除於1.044，或 $(1+r)^{11}$. 用此法，則作爲季節變動之計量之鏈比可除去長期趨勢之影響. 改正後之鏈比列於表七十八第(4)行中.

表示季節變動之指數，與其根據正月份之數字計算，不

註 $P_n=104.8$，r 實爲 .00388 此地四捨五入作爲 .004，適與按單利公式計算者相等. 故原義之表現甚不明晰. 讀者幸注意及之. 後面 $(1+r)$，$(1+r)^2$，$(1+r)^{11}$ 等誤作單利計算，譯者以職責所限，未便校正.——(譯者).

如根據十二個月份之數字之平均數計算.蓋根據平均數則十二個月份之百分數之總和將爲1200,每一月份之數字在解釋與應用上可比較稱爲簡單.將前表各月份數字化爲其平均數(71.69)之百分數時,結果所得之指數如表七十八第(5)項中所列者.

[應用移動平均數計算之法]

計量季節變動之另一方法,爲應用移動平均數.此種變動既有一固定週期——十二個月,故移動平均數之應用,較之用於包含週期長短不同之循環變動者,較有把握.變動之量(即季節性擺動之振幅 Amplitude)不能爲原來一致,故用移動平均數作出之線不能完全免除季節之影響.惟可取各月實際數目對於移動平均數之關係而平均之,再根據如此求得之平均數而作季節變動之指數.

移動平均數代表之中點,當然必須與之相比之原來數字之日期相合.故須作二次平均.例如已知之鷄蛋價格爲每月之第一日者,1910年十二項按月價格之平均數所代表之中點爲六月十五號.自1910年二月至1911年正月之十二項數字之平均數所代表之中點七月十五號.欲求與七月一號之實際價格可互相比較之數字,必更取此兩平均數而平均之.即求得"12月移動平均數"後更須求此移動平均數之"2月移動平均數",始可與原來之數字相比.下表爲如此得來之平均數:

表七十九

1910—1921年鷄蛋價格之移動平均數

(經2月平均數調整後之12月平均數)

(價格以每打值洋若干分計)

	1910	1911	1912	1913	1914	1915
正 月	... ...	20.25	21.28	20.78	22.74	22.29
二 月	... ...	19.99	21.45	20.79	22.80	22.20
三 月	... ...	19.82	21.60	20.79	22.91	22.06
四 月	... ...	19.64	71.75	20.87	22.97	21.91
五 月	... ...	19.46	21.94	20.99	22.89	21.90
六 月	... ...	19.38	22.08	21.19	22.66	21.98
七 月	22.47	19.33	22.00	21.49	22.57	21.98
八 月	22.18	19.58	21.63	21.88	22.64	21.84
九 月	21.63	20.20	21.16	22.32	22.55	21.74
十 月	21.21	20.66	20.89	22.57	22.39	21.78
十一月	20.89	20.88	20.79	22.65	22.36	21.88
十二月	20.57	21.07	20.76	22.69	22.35	22.02

	1916	1917	1918	1919	1920	1921
正 月	22.24	30.08	36.72	41.39	46.60	40.47
二 月	22.51	30.80	37.01	41.86	46.63	39.30
三 月	22.86	31.59	37.33	42.25	46.80	38.16
四 月	23.29	32.39	37.64	42.57	47.15	36.93
五 月	23.77	33.08	38.14	42.98	47.50	35.74
六 月	24.33	33.59	38.96	43.56	47.75	34.63
七 月	24.94	34.17	39.90	44.17	47.72	... ...
八 月	25.60	35.10	40.31	44.84	47.26	... ...
九 月	26.51	35.93	39.96	45.76	46.24	... ...
十 月	27.36	36.43	39.79	46.50	44.74	... ...
十一月	28.20	36.69	40.17	46.71	43.26	... ...
十二月	29.20	36.68	40.77	46.67	41.81	... ...

將實際價格表現爲與之相當之移動平均數之百分數，結果如下表：

表八十

鷄蛋實際價格對於其12月移動平均數計

百分數關係

	1910	1911	1912	1913	1914	1915
正　月	··· ···	150.1	138.8	129.0	135.0	141.8
二　月	··· ···	110.6	135.7	109.7	124.6	131.5
三　月	··· ···	83.2	113.4	93.3	105.6	96.6
四　月	··· ···	75.9	81.8	78.6	76.6	75.8
五　月	··· ···	75.5	77.9	76.7	73.4	78.1
六　月	··· ···	74.8	75.6	79.8	76.3	75.5
七　月	81.0	73.5	75.9	79.1	78.0	76.4
八　月	79.4	79.2	80.4	78.6	80.4	77.8
九　月	89.7	86.1	90.3	87.4	93.1	88.0
十　月	105.6	96.8	105.3	103.7	105.1	102.4
十一月	121.1	112.5	124.6	121.0	113.1	120.2
十二月	141.0	136.2	143.1	145.4	132.9	139.0

	1916	1917	1918	1919	1920	1921
正　月	137.6	125.4	126.1	138.2	139.1	151.0
二　月	119.1	116.2	133.5	115.4	122.0	123.2
三　月	92.7	107.0	108.2	78.3	99.6	76.5
四　月	76.9	80.0	82.9	80.6	82.3	55.2
五　月	76.1	90.7	81.3	85.6	78.7	56.5
六　月	78.1	92.6	76.5	88.6	77.5	56.0
七　月	79.0	82.8	76.9	83.3	76.9	··· ···
八　月	80.9	84.9	85.3	87.6	84.6	··· ···
九　月	87.9	92.4	91.1	89.6	95.6	··· ···
十　月	102.7	102.7	104.5	91.1	112.0	··· ···
十一月	114.2	107.4	117.5	115.6	131.5	··· ···
十二月	130.5	118.1	134.9	182.6	155.5	··· ···

由此可見同一月份之實際價格，其對於移動平均數之百分數關係，各年相差甚大；正月份之數字，雖皆超過平均數，然小者爲平均數之125.4%，大者爲151%.欲求吾人所須之指

數,必須取每月份之十一個百分數而平均之.此處無論算術平均數或中位數皆可應用.用此兩法求得之結果列於下表.第(2),(3)兩項爲實得之算術中數及中位數.此種指數若加校正而使十二個月份之指數平均時等於 100，則將更爲有用.第(4),(5)兩項之指數乃經如此調整而得者.

表八十一

根據移動平均數計算而得之

雞蛋價格季節變動指數

(1) 月份	(2) 算術中數 (未校正時)	(3) 中位數 (未校正時)	(4) 算術中數 (已校正後)	(5) 中位數 (已校正後)
正月	137.4	138.2	137.9	138.7
二月	122.2	122.0	122.6	122.5
三月	95.9	96.6	96.2	97.0
四月	77.0	78.6	77.2	78.9
五月	77.3	77.9	77.6	78.2
六月	77.4	76.5	77.7	76.8
七月	78.4	78.0	78.7	78.3
八月	81.7	80.4	82.0	80.7
九月	89.9	89.7	90.2	90.1
十月	103.3	103.7	103.7	104.1
十一月	118.1	117.5	118.5	118.0
十二月	137.2	136.2	137.7	136.7
平均數	99.65	99.608	100.0	100.0

〔配線比例法〕

求季節指數之法,新近又發明一種,(註)具有若干顯著之優點,此法即配線比例法是也,應用此法時,首先配合一適當之直線或曲線以代表材料之長期趨勢,次求各月之實在數值對於長期趨勢線上與之相當之一點之數值(長期趨勢值或正則值)之百分比,於是更求每月份諸百分比之平均數.此法與上述應用移動平均數計算季節指數之法相類,惟實在數值化爲百分數時,從任何代表長期趨勢之函數所得出之正則值,皆可爲其計算之基數耳.欲瞭解季節變動之實際情形,以選定每月之平均數值,可利用一多項頻數表(multiple frequency table).藉此表之助,吾人可徵驗是否有確定之季節變動存在,並決定選擇何種形式之平均數最能作代表每一月份之標本數值.茲將此法用於前數例所用之材料.

對於表七十五中所列 1901—1921 年每年平均雞蛋價格,可配合一直線,其方程式如下:

$$y=29.45+2.32x,$$

原點爲 1916 年正月.用以前敍述之方法,每月之正則值甚易求得.材料包括十二年,故將實在值化爲正則值之百分比時,每月份可得十二個百分比,正月份之十二個百分比大至

(註) 此法之基本觀點爲法爾克氏(Helen D. Falker:"The Measurement of Seasonal Variations", Journal of the American Statistical Association, June, 1924. 167—179)與赫爾氏(Lincoln W. Hall:"Seasonal Variations as a Relative of Secular Trend" 同上書 June, 1924, 156—166)獨自發明者.

並 194.8 小至 102.5, 五月份之十二個百分比大至 113.7, 小至 48.2. 圖 61 之多項頻數表, 乃就每月份之各項分組而列爲頻數表形式者也.

從此表觀之,顯有季節變動之存在,冬季各月之百分數常較春夏爲高. 此顯著之季節變動必可用指數表示.

此表助吾人選擇一種平均數以計量季節變動, 中位數似不足爲代表,蓋除非每月之百分數之頻數分配十分集中,

	正月	二月	三月	四月	五月	六月	七月	八月	九月	十月	十一月	十二月
自170以上	/	/										
165–169.9	//											
160–164.9												//
155–159.9	/											/
150–154.9												/
145–149.9	//	/										
140–144.9		//	/								//	/
135–139.9	/										/	
130–134.9		/									/	//
125–129.9										/		/
120–124.9	/	//								/		/
115–119.9	///		///							//	///	/
110–114.9		//		/	/	/			/	/	/	
105–109.9		/					/		//			/
100–104.9	/	/	/					//	/	/	//	/
95–99.9			/	/	/		/	//	/	//		
90–94.9		/		/	//		/				//	
85–89.9			//	//	/		/	/	//	//		
80–84.9			/	/	/	/		//	/			
75–79.9			/	/		//	/		/	//		
70–74.9			//	/		/	//	/	//			
65–69.9				/	//	/	/	//	/			
60–64.9					/	/	/	/				
55–59.9				//	/	/	/	/				
50–54.9							/					
未滿50				/	/	/						

圖 61.——每月鷄蛋之實在價格對於與之相當之長期趨勢値之比之頻數分配.(根據 1910—1921 年之材料而作.)

一二項之增加與減少卽足以大改中位數之值也.另一方面,包括全體各項之算術中數,又不免受極大極小之項之過分之影響.皆不如取每月份近中心之若干項而求其算術平均數.若觀察多項頻數表而不能直接確定應以何種平均數形式爲最佳,可先爲每月作幾個指數,而後比較彼此之結果而決定之.玆就表七十五之材料,計算得各月實在値對於長期趨勢値之百分比,更就此許多百分比用六種不同方法爲每月份作成六個不同之平均數.此簡單平均數卽未校正指數,列於下表.取此種平均數而校正之,使每一組平均數之平均數等於100,卽得同表之校正指數.(此種平均數非由圖61之頻數分配求出,而爲由實在値對於長期趨勢値之諸百分數求出者.)

表八十二

雞蛋價格季節變動之未校正指數與校正指數

(根據實在値對於直線長期趨勢値之百分比而作)

月份	未校正指數(中間各項之平均數)					
	2	4	6	8	10	12
正月	140.5	138.2	138.0	139.2	139.8	141.4
二月	120.8	121.6	123.4	123.8	123.6	125.7
三月	92.8	95.0	96.9	97.1	96.7	98.3
四月	80.0	79.0	79.1	78.1	78.3	78.9
五月	79.8	79.3	79.1	78.8	78.8	79.2
六月	77.6	77.9	78.6	78.6	79.0	78.9
七月	76.1	77.3	77.3	77.5	77.7	78.1
八月	80.3	80.5	81.4	81.8	82.1	82.0
九月	88.3	89.2	89.7	90.1	90.2	90.0
十月	100.7	102.2	102.4	102.4	102.3	102.4
十一月	116.8	115.7	116.0	117.5	117.3	117.3
十二月	133.0	134.0	134.7	135.6	135.5	135.2

表八十二(續上)

雞蛋價格季節變動之未校正指數與校正指數

(根據實在值對於直線長期趨勢值之百分比而作)

月分	校正指數(中間各項之平均數)					
	2	4	6	8	10	12
正　月	142.1	139.4	138.4	139.1	139.6	140.5
二　月	122.2	122.6	123.8	123.7	123.5	124.9
三　月	93.9	95.8	97.2	97.1	96.6	97.7
四　月	80.9	79.7	79.3	78.1	78.2	78.4
五　月	80.7	80.0	79.3	78.8	78.7	78.7
六　月	78.5	78.6	78.8	78.6	78.9	78.4
七　月	77.0	77.9	77.5	77.5	77.6	77.6
八　月	81.2	81.2	81.6	81.8	82.0	81.5
九　月	89.3	89.9	90.0	90.1	90.1	89.5
十　月	101.8	103.1	102.7	102.4	102.2	101.8
十一月	117.9	116.7	116.3	117.5	117.2	116.6
十二月	134.5	135.1	135.1	135.6	135.4	134.4

在第一行中之未校正指數爲根據每月之頻數分配取其中心之兩項而求得之算術平均數;第二項中之數字乃中心四項之算術平均數,餘按此類推.

在此例中,季節變動在各年皆明顯而規則,故根據不同項數平均而得之結果間尙無顯著之差異.根據中心四項之平均數而作之各月份之指數在此處,或可謂最佳.一般言之,中心三項,四項,或五項之平均數,無論與中位數比較或與包括每月全體項數之平均數比較,皆較易穩定而有代表真相之資格.每月各項在頻數分配表中越爲集中,則指數所須根據之項數亦越可減少.

下表並列用上述各種不同之方法求得之雞蛋價格季節變動指數凡四組,以資比較.

表八十三

雞蛋價格季節變動之指數
(已經校正)

月份	移動平均數法*(中位數)	算術平均數法	環比法	配線比例法(取四項平均數)
正月	138.7	138.8	139.5	139.4
二月	122.5	123.9	125.1	122.6
三月	97.0	96.5	100.3	95.8
四月	78.9	78.1	79.1	79.7
五月	78.2	78.4	78.1	80.0
六月	76.8	78.1	76.9	78.6
七月	78.3	77.3	77.0	77.9
八月	80.7	82.0	80.6	81.2
九月	90.1	89.9	89.1	89.9
十月	104.1	102.3	101.7	103.1
十一月	118.0	118.1	117.2	116.7
十二月	136.7	136.6	135.4	135.1
平均數	100.0	100.0	100.0	100.0

[幾種季節變動指數之比較]

此四種指數所表示之雞蛋價格季節變動,彼此相當合符.二月與三月有少許重要之出入,三月之最大最小兩指數,相差至百分之四.此計量季節變動之四種方法中,究以何者爲最佳,無絕對之標準,惟可就大致上言之.實際價格之算術平均數計算固爲簡便,然遇有極大極小之項存在,或遇材料在所研究之時期內,其數值大有變更時,不宜應用.環比法計算之手續繁冗,用於分析性質單純而頗有規則之數列,結果

並不能比簡單之計算法更佳.(註)就實在值對於移動平均數之百分比,或就實在值對於爲一適當之長期趨勢線所決定之正則值之百分比,而求其平均數之法,或爲最普遍適用之法.兩者之計算皆甚簡單,當正則值依據一長期趨勢之數學曲線時尤其如此;且在理論上亦有正當之理由證明根據比率求平均數較之根據絕對值求平均數爲佳.

[每月正則值之決定]

欲決定循環變動,須先除去長期趨勢與季節變動之影響;吾人於進而討論此問題前,對於將每年之長期趨勢值或正則值化爲每月之長期趨勢值或正則值之方法,應先略加說明,此方法僅爲簡單之算術計算.

茲假定下列之數字以資說明:

年　份	x商品之 每年正則生產額
1910	186
1911	330
1912	474
1913	618

(註) 潘瑟氏指出環比法有下列之優點:(注意,前二點非爲環比法所獨有,卽移動平均數法及配線比例法亦莫不然也.)第一,環比之相對數分配,使吾人能判斷逐月變動或季節變動之規則性之程度如何,第二,應用中位數(或中心幾項之平均數爲指數)故巨大而非季節性的變動之影響,可以大爲減除.第三,可以應用幾個性質不同之統計數列以計量季節變動.例如在某一時期50個代表城市的銀行淸算額之環比,可以與另一時期100個代表城市的銀行淸算額之環比相結合.(Journal of American Statistical Assoceation, June, 1923—717)

此材料之圖見圖62,A.

此等數字爲連續之四年之正則生產額,此中長期趨勢可以下列方程式代表之:

$$y = 186 + 144x,$$

原點爲1910年.因每年之數字皆爲該年之總數,故可認爲此數集中於各年之中點(即七月一號),經過如此作成之各點可繪一直線,其斜度爲144,此數即在此時期內每一年增加之生產額.

化每年正則值爲每月正則值之法,可分兩個步驟.第一,先求一以每月平均數表示各年產量之方程式.此可以12除前一方程式之186及144兩數值而得之:

$$y = 15.5 + 12x,$$

原點爲1910年之中點.此方程式之圖見圖62,B.圖中狹條表示集中於各年中點(七月一日)之每月生產額.圖中直線之斜度爲12,代表某年某月生產額比較前一年同月之生產額增加之量.故以1910年七月一日爲中心之一月,其正則生產額爲15.5;以1911年七月一日爲中心之一月其正則生產額爲27.5;餘可類推.注意,在此方程式中x之單位爲一年.

第二,乃計算每月之正則值或長期趨勢值.若自1910年七月一日至1911年七月一日之間,每月正則生產量之增加爲12,則自1910年七月一日至同年八月一日之間,其增加應爲1.因此爲便利起見,吾人可將橫生標之單位由一年改爲

一月.用此單位,則方程式爲

$$y = 15.5 + x.$$

以1910年七月一日爲原點.此原點不便於用,因每月之實際材料皆假定爲集中於每月之中點者也.若1910年七月一日

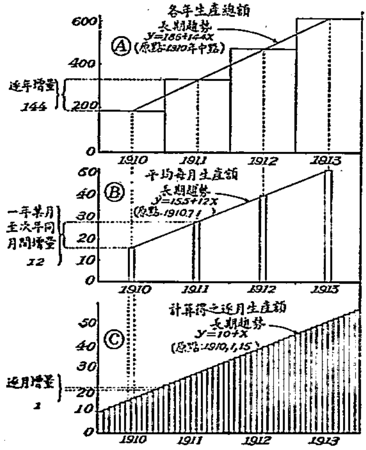


圖62——表示決定每月長期趨勢值之方法(X單位在Ⓐ及Ⓑ爲一年,在Ⓒ爲一月)

之縱坐標爲15.5,而每月之增量爲1,則同年正月十五日之縱坐標將爲10.將原點移至1910年正月十五日,則方程式如下:

$$y = 10 + x$$

根據此方程式則任何一月之正則生產額皆可立時求得.圖62,c,表示1910—1913年一時期內依次相繼各月正則生產額間之關係.

圖63表示1910年全年(正則)總生產額,平均每月(正

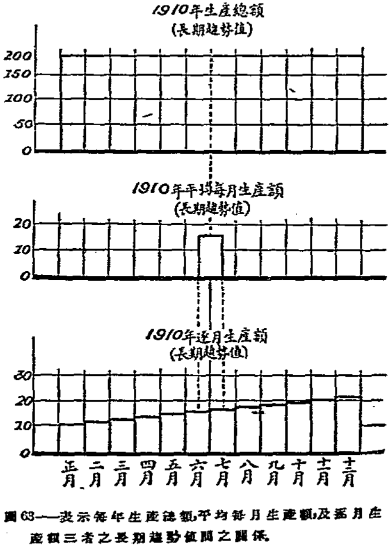


圖63——表示每年生產總額,平均每月生產額,及逐月生產額三者之長期趨勢值間之關係.

則)生產額與各月份正則生產額三者間之關係.(註)特別表示在將平月每月生產額化成各月份正則生產額時,必須校正半月之差.

在上例之數列中,因長期趨勢而起之每月增量,等於以 144 除每年增量.凡遇每年之數字爲十二個月數字之總和(關於生產額之材料皆如此)時,每月增量皆可如此求之.而在價格數列或任何其他數列中,其每年數字爲十二個月之數字之平均數者,則每月增量等於12除每年增量.凡遇此種情形,上述兩個步驟中之第一個即可省去.

[循環變動之計量]

欲計量一數列之循環變動,須除去該數列中長期趨勢及季節變動之影響.在大部分之經濟研究中,此種循環變動常爲主要興趣所在,而計量此種變動常爲分析時間數列之中心問題.茲取美國1901—1923年煙煤生產數字而分析之,以說明計算之手續.

對於此時期中每年生產額必須先作一長期趨勢線.選定之方程式爲

$$\log y=a+bx+cx^2,$$

用最小平方法求得常數之值.得曲線之方程式爲

$$\log y=2.646052+.014226x-.0010246x^2$$

(註) 圖62與63原爲達文浦(D.H.Davenport)氏所作.

用此方程式可求得各年之正則生產額.結果列於下表,並與實在數值相比較.

表八十四

1901—1923英國煤產額

(實在生產額與正則生產額之比較)

(1) 年份	(2) 實在生產額(單位:百萬噸)	(3) 長期趨勢之對數	(4) 正則生產額(單位:百萬噸)	(5) 實在額對正則額之百分比
1901	225.83	2.365589	232.05	97.3
1902	260.22	2.401332	251.96	103.2
1903	282.75	2.435025	272.29	103.8
1904	278.68	2.466670	292.87	95.1
1905	315.06	2.496265	313.52	100.5
1906	342.87	2.523810	334.05	102.6
1907	394.76	2.549307	354.25	111.4
1908	332.57	2.572754	373.90	88.9
1909	379.74	2.594153	392.78	96.7
1910	417.11	2.613502	410.68	101.6
1911	405.91	2.630801	427.36	95.0
1912	450.10	2.646052	442.64	101.7
1913	478.44	2.659253	456.30	104.9
1914	422.70	2.670405	468.17	90.3
1915	442.62	2.679509	478.09	92.6
1916	502.52	2.686562	485.92	103.4
1917	551.79	2.691567	491.55	112.3
1918	579.38	2.694522	494.91	117.1
1919	465.86	2.695429	495.94	93.9
1920	568.67	2.694286	294.64	115.0
1921	415.92	2.691093	491.01	84.8
1922	404.51	2.685852	485.12	83.4
1923	545.30	2.678561	477.05	114.3

煙煤之生產額之循環表現於每年之數字者，根據此表已可直接決定．惟現今之問題爲計量反映於每月之數字中之煤產額循環變動．

第一步手續爲決定所研究之時期中，每月在長期趨勢線上之縱坐標．此縱坐標之數値爲該月之"正則"生產額．然長期趨勢線爲配合於每年之材料者，故必須將其化爲每月之情形．

參加數列之每個數字，皆視爲其所指之時期的中點之代表，卽1913年總產額視爲集中於1913年七月一日；1913年正月之生產額視爲1913年正月十五日之產額．

1913年之長期趨勢値爲456.30百萬噸，此數視爲以1913年七月一日爲中點之一年之正則值．以12除之，得38.025百萬噸，此數視爲以1913年七月一日爲中點之一月之正則生產額．如此之一數字旣不能與六月又不能與七月之實際生產數字相比較，因彼此皆有十五日之差也．此點極易校正．以12除1914年之正則値，得39.014百萬噸，此數爲以1914年七月一日爲中點之一月之正則値．39.014與38.025間之差（.989），代表一年之某一月與前一年之同月之平均每月生產額間因長期趨勢而起之差．由一月至次月間之增量等於以12除.989，等於.0824此數等於自1913年七月一日至同年八月一日之增加量．自1913年七月一日至同年七月十五日之增加量爲其半，卽.0412．前已求得以1913年七月一日爲中點之一

月,其正則值爲 38.025 百萬噸,則以 1913 年七月十五日爲中點之一月,其正則值爲 38.025+.0412,卽 38.0662 百萬噸;以 1913 年八月十五日爲中點之一月其正則值爲 38.0662 + .0824, 卽 38.1486 百萬噸. 九月之長期趨勢值爲 38.1486+.0824, 卽 38.2310 百萬噸.1913 年七月起至 1914 年六月止,每月之正則生產額皆可用同法求得.

若數列之長期趨勢始終取直線形式,則全時期之每月正則值皆可以此法求之,每月之正則值加一固定之增量,卽得其下一月之正則值.但在此數列中,正則發展之率年年不同,(註)正則值之增量因此亦隨之而變.表八十五代表每年七月之正則值或長期趨勢值(各集中於七月十五日),並列舉自 1912—1922 年中各年正則生產額之每月增量.

表八十五

正則發展率之變動

1912—1922 美國煙煤生產額

日　期	正則生產額 (單位:百萬噸)	十二個月以內 正則生產額之每月增加量(單位:百萬噸)
1912 七月	36.9341	.0918
1913 同上	38.0662	.0824
1914 同上	39.0490	.0690
1915 同上	39.8650	.0543
1916 同上	40.5120	.0391
1917 同上	40.9742	.0233
1918 同上	41.2461	.0071
1919 同上	41.3238	—.0090
1920 同上	41.2074	—.0252
1921 同上	40.8070	—.0409
1922 同上	40.3957	—.0560

(註) 嚴格言之,正則發展之率乃逐月變動者,惟假定每十二月之時期內之長期趨勢爲直線形式而內推各月之正則值,並不致發生重大錯誤,而在計算上則較之用一較爲複雜之方程式容易甚多.

計算每月正則值時,此發展率之變動必須計及,前面計算1913年七月至1914年六月之每月正則值時所用之手續,對於每一包含十二個月之時期,必須重復一次.

吾人尙須求季節變動之指數.其法可於前述諸法中任取其一.此地所用者爲移動平均數法,根據1913-1921年之一時期中之每月生產額而求得,(註)結果列於表八十六第(5)行中.

計算煙煤生產額之循環變動之第一步手續,爲將每月之實在數化爲對於該月正則值或長期趨勢值之百分數,表八十六,就研究之全部時期中取二年爲代表以說明計算之方法,此兩年之百分數列於第(4)行中.在此種百分數中尙

表八十六

1913—1914美國煙煤生產額循環變動計算表

	(1) 月份	(2) 實在生產額 (單位:百萬純噸)	(3) 正則生產額 (單位:百萬純噸)	(4) 化實在值爲正則值之百分數 (2)÷(3)	(5) 季節變動指數	(6) 除去季節變動後對於正則值之百分數差離 (4)−(5)
一九一三	正月	42.274	37.503	112.7	107	+5.7
	二月	37.057	37.598	98.6	92	+6.6
	三月	37.538	31.693	99.6	102	−2.4
	四月	34.169	37.788	90.4	83	+7.4
	五月	37.205	37.852	98.2	92	+6.2
	六月	37.405	37.977	98.5	95	+3.5
	七月	38.858	38.066	102.1	97	+5.1
	八月	41.590	38.149	109.0	106	+3.0
	九月	41.424	38.231	108.4	106	+2.4
	十月	46.164	38.313	120.5	113	+7.5
	十一月	43.233	38.396	112.6	104	+8.6
	十二月	41.519	38.478	107.9	103	+4.9

(註) 此地所用美國煙煤生產額季節變動指數係紐約聯邦準備銀行統計部編製者.

表八十六(續上)

1913—1914美國煙煤生產額循環變動計算表

	(1) 月份	(2) 實在生產額 (單位:百萬純噸)	(3) 正則生產額 (單位:百萬純噸)	(4) 化實在值爲正則值之百分數 (2)÷(3)	(5) 季節變動指數	(6) 除去季節變動後對於正則值之百分數差離 (4)−(5)
一九一四	正月	40.191	38.581	104.2	107	−2.8
	二月	35.472	38.643	91.9	92	−.1
	三月	45.455	38.725	117.4	102	+15.4
	四月	23.609	38.808	60.8	83	−22.2
	五月	28.551	38.890	73.4	92	−18.6
	六月	31.412	38.973	80.6	95	−14.4
	七月	34.305	39.049	87.9	97	−9.1
	八月	37.751	39.118	96.5	106	−9.5
	九月	39.019	39.187	99.6	106	−6.4
	十月	37.685	39.256	96.0	113	−17.0
	十一月	33.392	39.325	84.9	104	−19.1
	十二月	35.862	39.394	91.0	103	−12.0

須除去季節變動之影響,季節變動之指數列於第(5)行;觀此顯然可見若第(4)行之百分數,比較與二相當之第(5)行之百分數爲小時,則實在值對於正則值之差離爲負數;若相反則差離爲正數.第(6)行之差離,乃由第(4)行各百分數減去季節指數而得;故第(6)行中各項爲除去長期趨勢與季節變動之影響後之數字,表示煙煤生產額之循環變動.此中意外變動與不規則變動之影響當然未曾除去.以第(6)行之指數作爲循環變動之計量,乃假定此種意外變動之影響,終究能互相抵消者也.(然在煤一方面,此種假定並不常爲合理).

詳細之計算僅舉兩年爲例;用同樣方法計算其他各年,所得結果列於下表.表中列示1913-1923年一時期每月之實際材料,以及除去季節影響以後其對於正則值之百分數差離.

表八十七

1913—1923年美國煙煤生產額

每月實在生產額及除去季節變動後對於正則值之校正差離

	1913		1914	
	實在生產額 (單位:百萬美噸)	對於正則值之校正差離 (百分數)	實在生產額 (單位:百萬美噸)	對於正則值之校正差離 (百分數)
正　月	42.274	+5.7	40.191	−2.8
二　月	37.057	+6.6	35.472	−.1
三　月	37.536	−2.4	45.455	+15.4
四　月	34.169	+7.4	23.609	−22.2
五　月	37.205	+6.2	28.551	−18.6
六　月	37.405	+3.5	31.412	−14.4
七　月	38.858	+5.1	34.305	−9.1
八　月	41.590	+3.0	37.751	−9.5
九　月	41.424	+2.4	39.019	−6.4
十　月	46.164	+7.5	37.685	−17.0
十一月	43.233	+8.6	33.392	−19.1
十二月	41.519	+4.9	35.862	−12.0
	1915		1916	
正　月	37.194	−12.7	46.593	+8.9
二　月	29.321	−17.8	45.187	+20.3
三　月	31.801	−21.7	43.829	+6.8
四　月	29.968	−7.5	33.628	+.3
五　月	30.938	−14.1	38.804	+4.0
六　月	33.957	−9.7	37.742	−1.7
七　月	35.578	−7.8	38.113	−3.0
八　月	38.161	−10.4	42.698	−.7
九　月	40.964	−3.5	42.098	−2.3
十　月	44.198	−2.6	44.807	−2.7
十一月	44.737	+7.6	44.927	+6.5
十二月	45.814	+11.1	44.098	+5.3

表八十七(續上)

1913—1923年美國烟煤生產額

每月實在生產額及除去季節變動後對於正則值之校正差離

	1917		1918	
	實在生產額(單位:百萬美噸)	對於正則值之校正差離(百分數)	實在生產額(單位:百萬美噸)	對於正則值之校正差離(百分數)
正　月	47.969	+10.7	42.227	−4.3
二　月	41.353	+9.4	43.777	+14.4
三　月	47.869	+15.3	48.113	+14.9
四　月	41.854	+19.4	46.041	+28.8
五　月	47.086	+23.1	50.443	+30.4
六　月	46.824	+19.4	51.138	+29.0
七　月	46.292	+16.0	54.971	+36.3
八　月	47.372	+9.5	55.114	+27.6
九　月	45.108	+4.0	51.183	+18.0
十　月	48.337	+4.8	52.300	+13.7
十一月	47.690	+12.1	43.895	+2.3
十二月	44.037	+4.2	40.184	+5.7
	1919		**1920**	
正　月	42.193	−4.8	49.748	+13.5
二　月	32.103	−14.3	41.055	+7.5
三　月	34.293	−19.0	47.850	+14.0
四　月	32.712	−3.8	38.784	+11.0
五　月	38.168	+.4	39.841	+4.0
六　月	37.685	−3.8	46.095	+16.8
七　月	43.425	+8.1	45.988	+14.6
八　月	34.613	−.4	49.974	+15.3
九　月	48.209	+10.7	50.421	+16.1
十　月	57.200	+25.5	53.278	+16.5
十一月	19.006	−58.0	55.276	+30.5
十二月	37.235	−12.8	53.275	+26.6

表八十七(續上)

1913—1923年美國煙煤生產額

每月實在生產額及除去季節變動後對於正則值之校正差離

	1921		1922	
	實在生產額(單位:百萬美噸)	對於正則值之校正差離(百分數)	實在生產額(單位:百萬美噸)	對於正則值之校正差離(百分數)
正月	41.148	−6.8	37.600	−14.5
二月	31.524	−15.2	40.951	+8.8
三月	31.055	−26.3	50.193	+21.7
四月	28.154	−14.3	15.760	−44.1
五月	34.057	−8.8	20.501	−41.4
六月	34.635	−10.4	22.309	−39.8
七月	31.047	−21.1	17.003	−54.9
八月	35.291	−19.6	22.328	−50.7
九月	35.893	−18.1	40.964	−4.3
十月	44.686	−3.4	45.173	−.7
十一月	36.805	−19.6	45.262	+8.7
十二月	31.627	−25.3	46.450	+12.8
	1923			
正月	50.123	+18.1		
二月	42.160	+13.4		
三月	46.807	+15.2		
四月	42.564	+23.7		
五月	46.076	+23.7		
六月	45.490	+19.3		
七月	45.644	+17.9		
八月	48.864	+17.2		
九月	46.216	+10.7		
十月	49.171	+11.3		
十一月	42.946	+4.7		
十二月	39.707	−2.3		

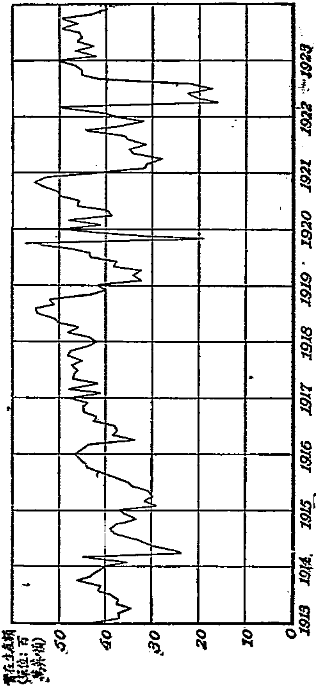


圖 64——1913—1923 英國每月煤炭生産額

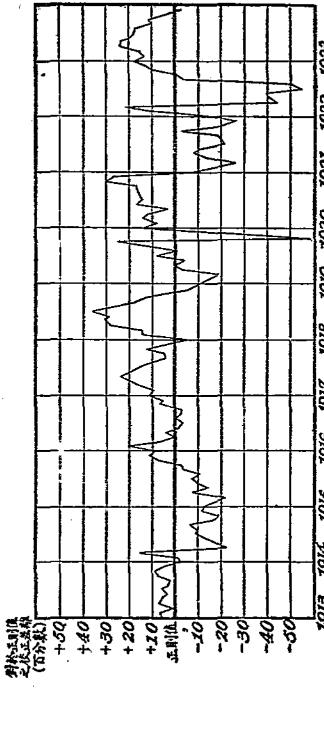


圖65——1913—1923 美國煙紙生產指數之循環變動與意外變動

每月實在生產額繪於圖64,循環變動(對於正則值之校正差離)繪於圖65.此等循環之反映一般之商情循環相當確切.然對於煙煤之材料,意外與不規則變動,所不能以上述方法消去者,較之對於大部份經濟數列更為重要.如1919年十一月與1922年之一部分,煙煤產額之小主要由於礦場之勞工糾紛.在分析任何統計數列時,此種不規則性乃不可不計及者也.

為使此一數列之循環變動,可與其他數列之循環變動相比較,宜將此種數字化為可相比較之形式.此一數列之百分數差離(循環變差),其程度可較另一數列者大,若無一公分母比較時將甚困難.標準差可作為公分母,欲將不同數列之循環變動相比較時,須先將每月或每年之差離,用標準差為其計量單位以表示之.

分析之過程於是完結.關於一數列,吾人首先求得其長期趨勢線之方程式,根據此方程式可以決定任何日期之正則值或長期趨勢值.於是更計量季節變動,而算出此種變動之指數.最後,使循環變動(加上無從計算之意外而不規則之變動)隔離其他影響而單獨表示,用數學方法固猶可更進一步分析循環變動,以除去意外變動之影響,將循環變動分析為較簡單之形式;惟應用此種精密之方法於經濟材料時須萬分留意.關於此種方法之技術之討論不在本書範圍以內.

〔關於時間數列之分析上一般應加考慮之點〕

下列各點,乃在分析時間數列時一般上應加考慮者.其中一部分前已論及,而在此予以綜結.

1. 材料在被研究之時期內其性質必須前後一致,此爲基本要件;此點不僅限於指材料之外觀,其出處,其命定價格之方法等等而言,且及於潛在之經濟的與社會的情況.具體言之,此點直接與前文中兩個因子之研究有關:

a. 以單獨之一曲線代表長期趨勢之事,僅於影響此長期趨勢之種種情況在所研究之時期內,一成不變時,始爲合理.

b, 季節變動指數之計算,僅限於季節變動相當一致之時期,即非常之時期不應包括在內,並且潛在的情況有變化時,即應另作新指數,在編製此種指數之前首須對於材料及其附屬之因子作一精密研究,以徵驗爲季節變動之源之各種情況,在所研究之時期內是否未有變更.

2. 求一長期趨勢線之方程式時所取之時期,其包括之循環之數應爲一整數,爲時期之始末之年,就所研究之該數列言,應爲正常之年.

3. 求長期趨勢線與計算季節變動指數時,應儘可能取較長之時期.十年可謂最低限度,能更長則更佳.

4. 代表物量之時間數列而又以圓爲單位者,一般上

在進一步分析以前須先按貨幣購買力之變動加以校正,銀行清算額,出口價值,與建築執照價值等皆其例也.並須注意在校正時所用之物價指數是否適當.

參考書

Chaddock, R. E. Principles and Methods of Statistics (Chapter XIII).

Crum, W. L. The use of Median in Determining Seasonal Variation, Journal of the American Statistical Association, marck, 1923.

Faer Falkner Helen D. The Measurement of Seasonal Variation, Journal of the American Statistical Association, June, 1924.

Hall, Lincoln W. Seasonal Variation as a Relative of Secular Trend. Tournal of the American Statistial Ass ocration, June, 1924.

Hart, W. L. the Method of Monthly Means for Determination of a Seasonal Variation. Journal of theAmerican Statistical Association, Sept., 1922

King, W.I. An Improved Method for Measaring the Seasonal Factor Journal of the American Statistical Association, Sept., 1924, (金氏在此文中提出一種計量季節因子之逐年變動之法).

Moore, H. L. Economic Cycles: Their Law and Cause.

Moore, H. L. Generating Economic Cycles.

Persons, W. M. Correlation of Time Series, (載 H. L. Rietz 所編 Hand book of Mathematical Statistics, 151—158)

Persons, W. M. Indices of Business Conditions, Review of Economie Statistics, Prel, Vol. I, 1919.

第九章

物量指數

經濟學家以及實業家之明悉自身活動之更大意義者,早慨嘆美國生產量統計之缺乏,關於若干大宗原料,猶可得相當正確之生產量數字,惟關於無數之製造品竟無可靠之材料可得.最近政府各部及私人機關之活動,致欲矯正此種情況.在將來,關於生產之情形,必可有正確而包羅豐富之材料無疑.此種材料,不論為研究價格或研究商情循環,不論對於抱有特殊目的之實業家或對於興趣較為普遍之經濟學家,皆異常須要不能或缺者也.

當吾人所研究者不僅限於一二種商品之情況時,立即有綜合幾種生產量數字以作一平均數之必要.研究一般生產情形之變動者,其所遇之問題與研究一般物價變動者同,若欲決定一生產之長期總趨勢,若欲觀察生產中之季節變動與循環變動,無數獨立之數字必須化為一單獨之指數,蓋如此之指數,其意義易為人所瞭解也,在1919年美國戰時工業局價格部在米契爾氏領導之下從事此種指數之編製.此

指數包含九十種原料之生產量,包括1913—1918年各年,其後性質相同範圍較廣時期較長之指數尙有多種.玆取其中若干種形式而敍述之,以說明編製物量指數之方法.

〔戰時工業局之生產指數〕

戰時工業局之編製生產指數,爲研究價格變動時之附帶工作.所用之方法與編製物價指數之方法同.編製物價指數時,種種商品每月或每年之價格以1917年之生產量與輸入量之和乘之,諸乘積相加而得之總值更化爲物價指數.物價有變動,而物量則不變.編製物量指數時,各種商品逐年之生產量與輸入量之和以1917年之價格乘之,求得諸乘積之總值,更將各年總值化爲比值而成生產指數.(註)價格旣不變,每年總值僅表現生產量與輸入量之變動.

關於九十種原料,求得之總值共有三組:其一代表每年生產量與1917年價格之乘積,其一代表每年價格與1917年生產量之乘積,第三者代表每年之價格與每年生產量之乘積.根據第一組而得生產指數,根據第二組而得物價指數,根據第三組而得逐年實際價值變動之指數,此最後一指數乃代表生產量變動與物價變動之共同結果者也.此等總值與指數見下表.

(註)個別之商品亦曾各各加權,權數與各商品在製造過程中經歷之手續爲正比例.鑛產物原料之權數大於農產物原料,因後者中甚多未曾加工者也.所用權數未見有詳細說明.

表八十八

1913—1918年九十種原料之

生產指數,價格指數,每年價格指數(註)

	總值(單位:百萬元)			指數		
年份	每年生產量×1917年價格	每年價格×1917年生產量	每年價格×每年生產量	生產量	價格	價格×生產量
1913	$ 30,375	$ 19,973	$ 17,390	100	100	100
1914	30,207	19,224	16,694	99	96	96
1915	32,482	19,699	18,455	107	99	106
1916	33,700	27,383	22,785	111	117	131
1917	34,748	34,748	34,748	114	174	200
1918	35,169	38,251	39,153	116	192	225

第一種指數表示生產量之增加在1914年至1916年間爲最大,1917年與1918年之間則增加甚小.價格之增加則與其他指數所表示者相類,第三行中.生產量因子與價格因子皆爲變數,所列數字,與常被引用於表示戰時生產量之增加之數字相類似.此中數值之增加大部分爲價格變動之結果,代表生產量之實際增加者僅一小部分.

上面所用之方法可以下列公式表示之.以1913年爲基期而表示1914年生產量變動之指數,其式爲

(註) 見戰時工業局物價公報第一期第45頁"History of Prices During the War"一文,原表之指數分列五個原料集團,植物性農產物,動物性農產物,林產物,礦產物,漁產物.

$$\frac{\Sigma q1914\cdot p1917}{\Sigma q1913\cdot p1917}.$$

此式卽費暄氏在採求其"標準"指數時所試驗之幾種總值式指數公式之一,僅 p 與 q 之地位對換耳.費氏之標準指數公式及其供平常應用之簡單式對於計量物量變動與物價變動同樣可以應用.標準公式用於物量指數時其形式爲

$$\sqrt{\frac{\Sigma(q_1p_0)}{\Sigma(q_0p_0)}\times\frac{\Sigma\ q_1p_1)}{\Sigma(q_0p_1)}},$$

此中 q_0 與 p_0 代表各商品在基年之產量與價格,而 q_1 與 p_1 代表在計算年之產量與價格.計算此種指數之手續,與前述計算標準物價指數之手續相同,僅將價格與產量之地位對調而已.

[戴氏生產指數]

戴氏 (Edmund E. Day) 經常編製兩種生產量指數,極爲有用.兩者編製時之原則不同,編製之目的亦少異.其一稱爲美國基本原料生產量之"未校正指數"(Unadjusted index 玆先就此種指數之特質言之.

此種指數包括自 1899 年至現在之各年,自最初編製至今,其包含之數列經過數次變更,此地就編製 1909 1922 年指數時所用者而言之.

編製此種指數時,先將可以蒐集之生產量材料分爲四個集團,卽關於農產者,關於畜產者,關於林產者,及關於鑛產

者.每一集團亦如基本原料之全部門一樣,各作一個指數,所以如此分類者,因影響於此四個集團之情况,有顯著之不同也.

編製此未校正指數之法,與前文所述編製物價指數之法,實際相同.先以1919年爲基,將各個生產數目化爲比值,(農產物以1919—1920年兩年之平均數爲基數).每個小集團之指數爲此比值之加權幾何平均數,(林產物除外,因其指數所根據者僅美國材木採伐量之一個數列也).權數係視各商品在集團內之比重而定,而此比重乃以基年之總值爲標準者.

如此計算之四個集團之指數其根據之商品數目各各不同.1909—1922年之指數所根據之商品數目各如下:

農產物……………………16種
畜產物……………………8種
林產物……………………1種
鑛產物……………………13種

(此時期後面幾年中鑛產物包括16種)

最後一步手續爲將此等集團指數結合爲一單獨之基本原料生產指數.此最後之指數爲四個集團指數之加權幾何中數,權數各與各集團在基年之生產總價值爲比例.

戴氏用完全同樣之方法更作一製造量指數,(Index of volume of manufacture) 包括自1914至1922各年.編製此種指

數有若干問題須附加考慮者,特別關於加權問題. 由1914至1918各年之指數,爲分別以1919年與1914年爲權數之兩個指數之幾何平均數.其餘各年則以1919年爲權數.此指數所根據數列有三十一個,代表歷年由製造過程中所生產之商品量或消費之原料量.

茲列各種未校正指數之數值於下表:

表八十九

美國生產量之未校正指數(註)

1909—1922 (1919=100)

年份	農產物(收穫物)	林產物	畜產物	鑛產物	前四項之綜合	製造品
1909	86.0	128.8	82.8	74.3	83.4	......
1910	83.5	115.8	75.0	81.7	82.6	......
1911	83.8	107.1	87.0	81.0	84.4	......
1912	99.6	113.3	84.3	86.7	91.7	......
1913	88.4	111.1	83.9	93.0	87.8	......
1914	99.5	108.1	80.3	87.6	90.1	75
1915	104.9	90.4	88.1	93.3	95.3	86
1916	91.6	100.7	93.6	101.8	94.6	102
1917	97.5	96.0	93.2	112.2	98.1	105
1918	98.1	84.9	103.2	112.7	102.0	102
1919	100.0	100.0	100.0	100.0	100.0	100
1920	109.2	86.4	92.5	114.2	103.0	104
1921	90.5	78.1	94.9	91.6	92.0	80
1922	100.7	100.7	102.1	97.5	100.7	109

(註) 此表之數字取自 Review of Economic Statistics, July, 1922; July, 1923.

將以上各種指數互相比較之,可得有趣味之結果.各數列之長期趨勢顯然不同,各表示其特質;逐年生產量之變動亦同樣有顯著之差異.上面在討論物價指數已經指出,氣候一般上爲農產品之決定因子,而其他生產部門則比較直接的爲當時物價與市面情況所支配.

[生產量之校正指數]

前面分析時間數列時業已指出,此種數列之循環變動常成爲最有興味之對象,對於研究商情變動者尤爲如此.生產量與貿易量之變動爲商情循環中根本重要之形相,故在物量之研究中循環變動尤爲注意之中心.計量單獨一數列之循環變動所須之方法,已經前數章中說明.研究一般商情時,顯然須進一步結合若干數列之循環變動爲一單獨之指數.此種代表生產量在商情循環過程中所生變動之指數,其效用可不言自明.

當所用之材料以年爲單位時,此種循環變動指數之編製甚簡單.因此中無季節變動之問題存在,須計及者僅長期趨勢而已.編製之方法可有兩種.戴氏在編製其生產量之"校正"指數時,對於兩種方法皆曾加試驗.

第一種方法,先就所包含之每一數列作一長期趨勢線,於是將實在數字化爲與之相當之長期趨勢值之百分數.當每一數列皆已化爲百分數時,每一年之幾個百分數之加權算術平均數即爲該年之校正指數.在求平均數時所用之權

數,與計算未校正指數時所用者相同.結果所得之校正指數為比值之形式;惟此種比值非對於任何固定之基數而言,而為對於一假設之"正則"之百分數也.此卽所求之代表生產量循環變動之指數.

另一方法較為簡單.每一未校正指數有一長期趨勢,"此長期趨勢者,在未校正指數所根據之幾個原來數列中所包含之固定趨勢之合力也."然則為何不計量此一長期趨勢以代替個別計量原來數列之長期趨勢,而直接由未校正指數求出校正指數乎?

後一方法之是否可靠,可取兩法之結果而比較之.戴氏如此比較之結果證明實際上兩者之結果相同.因此在求校正指數時所應用者,為此較為簡單之方法.

關於農產與礦產之校正指數,相當於前表所列之未校正指數者,茲列於下表:

表九十

美國生產之校正指數(註)

1909—1922

年份	農產物(收穫物)	礦產物
1909	98	97
1910	99	102
1911	93	98
1912	109	101
1913	95	104
1914	106	95
1915	110	98

(註) 指數錄自 Review of Economic Statistics, July, 1922; July, 1923.

表九十(續上)

美國生產之校正指數
1909—1922

年份	農產物(收穫物)	礦產物
1916	95	107
1917	100	111
1918	99	108
1919	99	93
1920	107	104
1921	88	81
1922	96	84

[根據每月之數字而作之物量指數]

上面諸例僅及每年之指數.將上述之方法稍加改進.即可用之於生產量及貿易量數列之每月數值以求其指數.此種指數之包含內容最豐富者,美國聯邦準備局每月編製及出版之指數即其一也.

[聯邦準備局所作之基本工業生產量指數]

此指數包括之時期爲由1913年正月起直至現在,每月公佈於聯邦準備公報上.根據之統計數列有二十二個,此等數列乃經常可得每月之材料者.其中二十個數列列於表九十一.編製最後之指數時,對各數列各各加權,權數視1919年工業調查之結果中所表現之各數列之相對重要性而定之,求權數時所根據之材料,爲製造過程中所增加之價值與僱傭之人工之數字.

編製此指數時,並未作去除長期趨勢影響之嘗試.最後

之指數爲比值形式,以1919年平均每月生產量爲基數.惟此基數之爲"正則",與長期趨勢線之縱坐標之爲"正則"者意義迥乎不同耳.

求平均數之前,先將各數列以1919年爲基數化爲比值.季節影響則儘量消除之,故在計算比值以前,先以季節變動指數校正1919年之每月平均數.其法頗可注意.(註1)

季節變動之指數爲每月實在值對於12月移動平均數之差離之中位數,其計算之法前章已詳言之.所包括各數列之季節變動極有趣味,故將其季節指數綜列於下表:

表九十一

若干統計數列中之季節變動指數(註2)

	正月	二月	三月	四月	五月	六月
生鐵生產量	99	88	103	100	102	89
鋼塊生產量	99	89	102	100	102	99
棉花消費量	104	97	107	100	105	102
小麥生產量	112	89	89	82	84	97
糖融解量	69	93	128	125	121	123
猪屠宰量	142	119	99	86	99	106
牛屠宰量	102	81	88	84	86	92
小牛屠宰量	85	73	102	121	128	114
羊屠宰量	104	89	91	80	83	90
木材斫伐量	80	84	97	101	119	113
煙煤生產量	107	92	102	83	92	95
白煤生產量	97	84	100	96	105	105
銅生產量	100	91	105	103	104	102
鞋底皮生產量	102	90	100	103	108	109
報紙生產量	106	92	101	101	101	102
水門汀生產量	55	64	83	104	120	117
煤油生產量	98	90	102	100	102	101
雪茄煙生產量	89	88	100	96	100	105
捲煙生產量	95	69	99	91	99	106
加工煙葉生產量	94	94	107	100	102	102

(註1) 此法對於此問題之特別應用,乃美國國立經濟研究局麥柔萊(Frederik R. Macaulay)氏之力.

(註2) 錄自聯邦準備公報,1922年十二月號第1416頁.

表九十一(續上)

若干統計數列中之季節變動指數

	七月	八月	九月	十月	十一月	十二月
生鐵生產量	100	104	101	105	102	98
鋼塊生產量	100	104	101	107	102	95
棉花消費量	99	100	95	99	95	97
小麥生產量	77	106	113	124	124	121
棧駱解量	127	126	99	71	60	53
豬屠宰量	84	65	62	83	109	146
牛屠宰量	96	99	111	128	122	111
小牛屠宰量	109	96	99	100	92	81
羊屠宰量	100	110	120	117	110	103
木材斫伐量	107	113	110	109	92	75
煙煤生產量	97	106	106	113	104	103
白煤生產量	100	102	99	110	102	100
銅生產量	97	100	97	103	97	99
鞋底皮生產量	98	101	97	103	92	97
報紙生產量	98	101	95	103	98	102
水門汀生產量	109	119	117	121	105	86
煤油生產量	105	103	99	103	98	99
雪茄煙生產量	103	103	103	113	106	94
捲煙生產量	114	108	104	110	99	86
加工煙葉生產量	100	107	106	108	95	85

此等指數係根據1913—1919年一時期之材料而作,對於表現各個不同之工業集團之季節變動上,甚爲重要.

作每月比值之基數之數字,先經校正而除去季節變動之影響.茲可取聯邦準備公報上登載之棉花消費量數字,以說明其方法.表九十二所列者,爲1919年之實在數字,季節變動指數,以及除去季節變動後1919年每月之校正平均數.

1919年每月平均消費量爲4,933,(百包),若無季節變動之存在,而置長期趨勢與意外變動不計,此數即可視爲該年每月之實在消費量.惟編製季節變動指數時,發見正月份之數字平均上較一年之平均數大4%,二月份之數字較一年

表九十二

除去季節變動之校正法之應用

（棉花消費量）

(1) 月份	(2) 棉花消費量 1919 (單位：一百包)	(3) 季節指數	(4) 每月校正 平均數 1919
正　月	5,569	104	5,130
二　月	4,333	97	4,785
三　月	4,335	107	5,278
四　月	4,759	100	4,933
五　月	4,879	105	5,179
六　月	4,743	102	5,032
七　月	5,103	99	4,884
八　月	4,973	100	4,933
九　月	4,911	95	4,688
十　月	5,560	99	4,884
十一月	4,913	95	4,686
十二月	5,117	97	4,785
	59,195		59,195
每月平均	4,933		

之平均數小3%,……故以季節指數乘4,933,所得者爲各月份之每月校正平均數,表示若季節變動爲唯一之擾亂因子時,1919年全年之實在消費量在各月份應有之分配.所有實在數字化作比値時,皆以此每月校正平均數爲基數.

舉例言之,1921年正月實在棉花消費量爲3,665(百包).以1919年爲基數化之爲比値時,吾人以1919年正月份之校正平均數5,130除之,更乘以100.於是吾人得1921年正月之比値71.4,此數已除去季節變動之影響.欲求二月份之比値,則取1919年二月份之校正平均數爲基數.此法妙在將基數作一次校正以後,不須更作其他校正,而比値中季節變動之影響卽可除去也.

由代表各數列之此種校正比値,極易求得每月之最後之指數.每一比値與其權數相乘,諸乘積相加其總和除以權數之總和,卽得所求之指數.該局將校正比値與包括一切商品之最後指數一起發表,此法大增此編製工作之效用.

除此代表基本工業生產量之一般性指數外,聯邦準備局之研究分析部更作一套有用之指數,表示專門部分之活動,包括農業活動指數,鑛產指數,及製造業生產指數.編製此等專門指數時,並未除去季節變動及長期趨勢之影響.此種指數自1919年正月以後,逐月皆有.

[美國電話電報公司之"一般商情"指數]

有一少許不同之法,可以美國電話電報公司統計部所編製之一般商情之綜合指數爲例,此指數已追算至1887年,逐月編製,在此時期內其包含之成分曾變更若干次.

就1923年底時此種指數之結構言,所用之數列有十一個,此種數列乃認爲足以反映一般商情變動者.每一數列皆

按前二章所述之方法分析之.配合長期趨勢線,化實在數字爲正則值之百分數,更校正此等百分數而除去其所受季節變動之影響,惟在結合此十一個數列爲一之前,猶有一事須作者.若干數列之循環變動較之其他數列劇烈甚多,欲着手求諸數列之平均數,首須全部化爲可以互相比較之單位.不然則變動之範圍最大之數列將得過大之權.校正之法,卽將對於每數列正則之差離化爲該數列標準差單位也.

1923年十二月之指數,包括下列各數列:(附其權數)

數列	權數
生鐵生產量	20
水運貨噸哩數	10
車運載量	10
鋼塊生產量	10
羊毛機器之活動量	10
棉花消費量	10
紙生產量	10
木材生產量	5
皮革生產量	5
煙煤生產量	5
電力生產量	5
總計	100

玆將1877—1923年此種指數之數值列於表九十三,並作圖66以表示之.

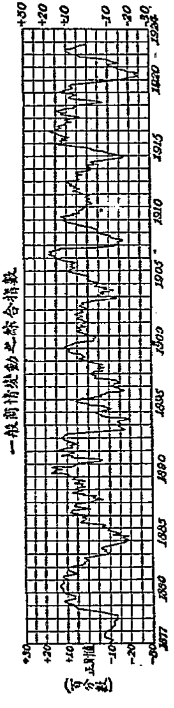


圖 66

(美國電話電報公司所編之綜合指數)

表九十三(一)

一般商情變動之綜合指數

1877—1923

英國電話電報公司編製

（表中數字爲對於正則之百分差離，右上角有*號者表示該月以後之曲線其成分與以前不同）

	1877	1878	1879	1880	1881	1882	1883	1884	1885	1886	1887	1888
正　月	−9	−8	−13	+7	+12	+11	+9	−10	−16*	−10	−2	−2
二　月	−9	−9	−13	+9	+12	+11	+7	−9	−19	−8	+3	−2
三　月	−10	−10	−13	+11	+12	+12	+4	−8	−17	−6	+9	−7
四　月	−11	−11	−13	+12	+12	+12	+2	−8	−16	−8	+8	−6
五　月	−10	−11	−11	+10	+12	+10	+2	−7	−19	−7	+8	−2
六　月	−8	−12	−9	+8	+13	+8	+1	−6	−14	+3	+6	−2
七　月	−7	−12	−7	+7	+13	+5	+1	−6	−11	+4	−2	−3
八　月	−7	−12	−3	+7	+13	+7	+1	−8	−15	+2	+2	+2
九　月	−7	−12	+2	+7	+12	+9	+2	−10	−14	+2	+7	+2
十　月	−7	−12	+6	+7	+12	+10	+2	−12	−12	+1	+5	+8
十一月	−7	−12	+6	+9	+12	+10	−2	−15	−11	+2	+7	+2
十二月	−8	−13	+7	+11	+11	+9	−6	−18	−9	+3	+1	+3

表九十三(二)

一般商情變動之綜合指數

1877—1923

美國電話電報公司編製

(表中數字係對於正則之百分差離,右上角有*號者表示該月以後之曲線其成分與以前不同.)

	1889	1890	1891	1892	1893	1894	1895	1896	1897	1898	1899	1900
正月	+6	+11	+5	+0*	+12	−14	−8	−1	−13	−8	+1	+10
二月	+5	+8	+1	+14	+12	−10	−9	−2	−11	−2	+1	+11
三月	+3	+10	−5	+11	+12	−14	−10	−7	−12	−3	+4	+10
四月	+1	+12	−6	+9	+11	−14	−11	−5	−12	−7	+1	+8
五月	+4	+17	−7	+7	+11	−14	−8	−8	−15	−7	+2	+6
六月	+2	+15	0	+9	+5	−20	−6	−8	−13	−4	+3	+6
七月	+7	+16	+8	+6	−2	−18	−1	−8	−14	−8	+3	+3
八月	+6	+12	+7	+6	−13	−10	0	−14	−10	−5	+6	0
九月	+1	+15	+13	+7	−17	−7	+1	−16	−8	−5	+8	−3
十月	+8	+17	+12	+8	−17	−7	+6	−17	−4	−7	+9	−3
十一月	+7	+8	+7	+10	−16	−6	+4	−16	−4	−4	+8	−4
十二月	+5	+5	+7	+12	−15	−4	+3	−13	−2	−2	+11	−4

表九十三(三)

一般商情變動之綜合指數

1877—1923

英國電話電報公司編製

(表中數字係對於正則之百分差離;右上角有*號者表示該月以後之曲線其成分與以前不同.)

	1901	1902	1903	1904	1905	1906	1907	1908	1909	1910	1911	1912
正月	−1	+2	+3*	−10	−5	+10	+17	−14	−5*	+12	0	−2
二月	0	+2	+1	−5	−5	+9	+14	−15	−5	+10	−1	+3
三月	+2	+2	+2	−8	−2	+7	+15	+15	−5	−12	0	+3
四月	+5	+5	+4	−6	−1	+5	+16	−14	−5	+11	−3	+6
五月	+4	+5	+3	−8	0	+8	+18	−17	−3	+6	−2	+5
六月	+3	+2	+2	−10	0	+7	+16	−17	−2	+7	−1	+3
七月	+4	+4	−3	−12	−1	+7	+18	−16	+1	+4	−2	+5
八月	+4	+3	−1	−12	0	+8	+16	−13	+2	+4	−1	+6
九月	0	+6	+2	−8	+2	+5	+12	−11	+5	+4	−1	+5
十月	+8	+5	−6	−9	+2	+10	+13	−11	+8	+4	−1	+8
十一月	+2	+2	−11	−4	+4	+11	−4	−9	+9	+2	−1	+7
十二月	+1	+3	−14	−5	+7	+13	−12	−8	+10	+1	−2	+7

表九十三(四)

一般商情變動之綜合指數

1877—1923

英國電話電報公司編製

(表中數字係對於正則之百分差離,右上角有*號者表示該月以後之曲線其成分與以前不同。)

	1913	1914	1915	1916	1917	1918	1919	1920	1921	1922	1923
正月	+10*	0	−10	+11	+17	+4	+4*	+13	−19*	−16	+8
二月	+8	−2	−13	+12	+18	+3	−3	+10	−18	−11	+8
三月	+3	0	−12	+13	+14	+10	−8	+12	−22	−10	+10
四月	+5	−1	−9	+9	+12	+10	−6	+9	−25	−16*	+9
五月	+5	−5	−8	+12	+14	+13	−2	+6	−25	−14	+10
六月	+4	−4	−4	+11	+12	+10	0	+8	−23	−11	+7
七月	+5	−2	−3	+8	+10	+14	+7	+8	−25	−12	+6
八月	+2	−8	−2	+12	+10	+15	+9	+7	−21	−10	+5
九月	+3	−10	+2	+13	+7	+11	+8	+5	−19	−6	+1
十月	+5	−13	+5	+15	+9	+10	+7	−1	−18	−2	+1*
十一月	−1	−17	+9	+17	+11	+5	+4	−6	−15	+4	0
十二月	−2	−18	+14	+16	+6	+8	+3	−11	−14	+4	−2

〔貿易量指數〕

然物量指數之包羅最爲豐富者,當推紐約聯邦準備銀行報告部所編製之貿易量指數.(註)觀其所包括之統計數列之表,可知其性質與範圍爲如何.表中並註明每數列之權數.

表九十四

貿易量指數中所包含之統計數列
(紐約聯邦準備銀行)

生產活動	權數	
1. 消費者之貨物	8	
2. 生產者之貨物	9	
3. 工廠雇傭人工	6	
4. 汽車與貨車	2	
5. 房屋建築	4	
	—	29%
原始分配		
6. 商品車運量	5	
7. 其他車運量	2	
8. 批發貿易	8	
9. 出口	3	
10. 進口	2	
11. 巴拿馬運河運輸	1	
12. 穀物出口	1	
	—	22%

(註) 關於此指數之詳細記述見施乃德(Carl Snyder)氏"A New Index of the Volume of Trade"一文. Journal of the American Statistical Association, Dec., 1923, 49—63.

表九十四(續上)

貿易量指數中所包含之統計數列
(紐約聯邦準備銀行)

生產活動	權數	
對於消費者之分配		
13. 百貨公司銷售量	8	
14. 連環商店銷售量	3	
15. 連環食品雜貨店銷售量	6	
16. 函信販賣店銷售量	3	
17. 新人壽保險額	2	
18. 娛樂收入額	2	
19. 廣告費	2	
	—	26%
一般業務活動		
20. 紐約城以外各地之負債	8	
21. 紐約城之負債	5	
22. 郵政收入	1	
23. 交通	1	
24. 電力生產	2	
	—	17%
金融活動		
25. 紐約股票交易所出售之股票	2	
26. 新組織之資金	2	
27. 穀類預售額(支加哥)	1	
28. 棉花預售額(紐約)紐奧爾闌	1	
	—	6%
		100%

此中若干數列爲由更基本之材料結合而成者,如消費者貨物之生產量數列乃結合十五項材料而成,代表生產者貨物之生產量數列亦包含同數之材料.

決定若干數列之權數時遇有若干專門問題,一般言之,權數乃根據製造中所增加之價值,生產物之價值,或交換之商品之價值而定者.

結合此等數列爲一單獨之指數時,其所用方法爲吾人所已熟知者.先配合一適當曲線代表每一數列之長期趨勢而定其正則.凡遇數列之受強烈季節影響者,則計算其季節指數.因包含甚多物價數列之故,緊縮之問題特爲嚴重.對於若干數列,適當之緊縮指數不難求得,然另有甚多之數列,緊縮指數必須結合幾個不同數列而編製之,例如緊縮對於紐約外個人帳目之負債時,所用緊縮指數乃由下列數列組合而成者:

數列	權數
批發商品價格	2
生活費指數	3
工人平均每週所得額指數	3
股票價格指數	1

每一數列經過以上一切校正以後,乃可加以最後之結合,求得之1919—1923年一時期指數之數值如下:

表九十五

英國貿易量指數,1919—1923

(紐約聯邦準備銀行編製)

	1919	1920	1921	1922	1923
正　月	97	109	90	93	109
二　月	97	104	92	96	110
三　月	95	108	91	102	113
四　月	101	105	92	99	109
五　月	106	104	91	103	110
六　月	108	102	92	103	107
七　月	110	101	91	97	100
八　月	107	99	95	99	102
九　月	107	96	94	103	100
十　月	109	95	92	103	106
十一月	105	94	92	104	105
十二月	105	93	93	105	100

前面幾頁中所述者,決非已將流行之一切貿易量及生產量指數盡數列舉.(註),其主要目的,專在舉例以說明編製之方法,雖無一標準之方法,然物量指數之編製不須用複雜方法一點已可明瞭.至於若干困難問題,——緊縮"正則"之決定,季節變動之計量等,固永遠存在,非經更進一步之試驗不能完全解決者也.惟用現今已有之方法,已可得若干甚有價

(註) 參考潘葉氏所編製之"英國貿易指數",(包括自1903年至今各月)在 Review of Economic Statistics, 1932 年四月號上有關於此種指數之紀述;並參考 Review of Economic Statistics 上發表之每月製造業及礦業指數;以及 Federal Reserve Bulletin 上每月發表之零售貿易及批發貿易之精密指數.

值之計量,代表實業活動之情況.不僅使吾人對於經濟制度之認識大為增進,且使吾人統制此制度之力量亦加強也.

參考書

Day, Edmund E. An Index of Physical Volumn of Production. Review of Economic Statistics, Sept., 1920, Jan., 1921.

Day, Edmund E. The Volume of Production of Basic Materials in the United States. Review of Economic Statistics, July, 1922.

Persons, W. M. An Index of Trade for the United States. Review of Economic Statistics, April, 1923.

Snyder, Carl. A New Index of the Volumn of Trade. Journal of the American Statistical Association, Dec., 1923.

Snyder, Carl. A New Clearings Index of Business for Fifty Years. Journal of American Statistical Association, Sept., 1924.

第十章

關係之計量：直線繫聯

討論平均數、離中趨勢量與偏態量時，吾人所研究者爲敍述一單獨之頻數分配之方法．藉此等計量之助，吾人能決定集中值之所在，並描述集中值周圍之分配之性質，單獨一變數之許多數值沿尺度而排列之狀態因而得其表現．分析時間數列時，所遇之問題略異．此時吾人所處理者，爲一變數因子在時間之過程中之諸變值，所欲決定者，爲長期趨勢、循環變動、季節變動及意外變動諸不同之力量所引起之數值變動，其程度爲如何．以前數章，即討論如何計量此每種力量(除意外變動外)對於某一數列之影響之種種方法者也．

在此種種方法之中，有幾種可以適用於現今所須討論之問題．處理時間數列時，吾人曾見時間因子與長期趨勢因子間，其關係可以一固定之數學方程式代表之．即在許多經濟數列中，長期之趨勢或發展似爲時間之一函數．若遇如此之關係，不問其爲確切或僅爲近似，求一數學式以代表之，總有顯著之利益．現今所須研究之問題與此相似，而其範圍則

較之猶大爲廣闊：此問題即當處理時間數列時，時間與已知數列各項正則值間之關係既可用一固定數學方程式代表，則研究別種變數間關係時，同樣方法可否應用乎？吾人可以數學的方法計量花生產量與棉花作格間，穀物之收成與雨量間，勞動之薪給與其生產力間之關係乎？如此事爲可能，則經濟學家可得一強有力之工具，使其所用方法之精確性差可與自然科學家之工作相比.

〔應課稅之個人收入額與汽車登記額間之關係〕

茲取美國各邦應課稅之個人收入額與汽車登記之數字間之關係，作爲標本的問題而研究之.兩者之數字列於表九十六第(2)、(3)兩行中.(註)

將此數字描繪於圖67，每一點代表每一邦中應課稅之個人收入額與登記之汽車輛數間之關係，如此之圖名爲散佈圖(scatter diagram).由此圖觀之，兩變數間顯有關係存在.一般上個人收入額大之各邦，即爲汽車登記數大之邦.惟其關係並不完全.應課稅之個人收入額相同之兩邦，其登記之汽車數可相差甚大.如懇塔啓邦與俄克拉何馬邦之應課稅個人收入額各爲69,000，但前者之汽車登記數爲126,000，後者爲221,000.若關係爲完全者，Y變數之任一數值只能與X變

(註)五個主要工業邦(紐約、本雪爾文尼亞、紐澤賽、伊里諾斯、馬薩諸塞五邦)未列在內.個人收入額與汽車所有權間之函數關係，在工業區與全體鄉村不同.故其除外乃合理之事也.

表九十六

1921年美國四十三邦之

應課稅之個人收入額與登記之汽車數

(1) 邦名	(2) 應課稅之個人收入額 1921 (單位:千) X	(3) 汽車登記數 1921 (單位:千) Y	(4) XY	(5) X^2	(6) Y^2
阿拉巴馬	43	82	3,526	1,849	6,724
亞黎桑那	18	35	630	324	1,225
亞堪薩斯	34	67	2,278	1,156	4,489
加利福尼亞	386	673	259,778	148,996	452,929
科羅拉多	70	145	10,150	4,900	21,025
康涅狄格	123	137	16,851	15,129	18,769
德拉瓦	16	21	336	256	441
佛羅里達	42	97	4,074	1,764	9,409
喬治亞	68	131	8,908	4,624	17,161
伊達何	23	51	1,173	529	2,601
印第安那	150	400	60,000	22,500	160,000
衣阿華	111	460	51,060	12,321	211,600
堪薩斯	89	291	25,899	7,921	84,681
肯塔基	69	126	8,694	4,761	15,876
路易亞阿那	68	80	5,440	4,624	6,400
緬因	44	77	3,388	1,936	5,929
瑪利蘭	113	140	15,820	12,769	19,600
密支干	250	477	119,250	62,500	227,529
明尼蘇達	125	328	41,000	15,625	107,584
密西西比	26	65	1,690	676	4,225
密蘇利	173	346	59,858	29,929	119,716
蒙他拿	37	58	2,146	1,369	3,364

表九十六(續上)

1921年美國四十三邦之

應課稅之個人收入額與登記之汽車數

(1) 邦名	(2) 應課稅之個人收入額 1921 (單位：千) X	(3) 汽車登記數 1921 (單位：千) Y	(4) XY	(5) X^2	(6) Y^2
奈勃拉斯加	72	238	17,136	5,184	56,644
奈瓦達	10	10	100	100	100
新罕布什爾	32	42	1,344	1,024	1,764
新墨西哥	12	24	288	144	576
北加羅林那	44	148	6,512	1,936	21,904
北達各塔	18	92	1,656	324	8,464
俄亥俄	367	720	264,240	134,689	518,400
俄克拉何馬	69	221	15,249	4,761	48,841
俄勒岡	63	118	7,434	3,969	13,924
路得島	48	54	2,592	2,304	2,916
南加羅林那	25	90	2,250	625	8,100
南達各塔	22	119	2,618	484	14,161
田納西	61	117	7,137	3,721	13,689
塔克薩斯	200	467	93,400	40,000	218,089
猶他	26	47	1,222	676	2,209
威爾滿	18	36	648	324	1,296
維基尼阿	76	141	10,716	5,776	19,881
華盛頓	116	185	21,460	13,456	34,225
西維基尼阿	75	93	6,975	5,625	8,649
威士康新	148	341	50,468	21,904	116,281
亞俄明	22	26	572	484	676
	3,602	7,616	1,215,966	603,968	2,612,065

數之一個數值相配.

吾人之第一個問題為求一方程式以代表此關係.此關係雖不完全,然明明存在者也.此地之關係可擬之為長期趨勢,且為一可以直線代表之長期趨勢.用最小平方法求得一配合於散佈圖上諸點之一直線之方程式,則此方程式可以在數學上表示兩變動間之平均關係 (Average relationship).

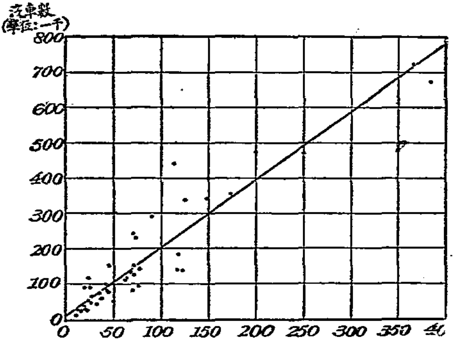


圖67.——散佈圖:表示美國各邦應課稅之個人收入額與汽車登記額間之關係,並其平均關係線.

此種直線固亦可憑觀察配給,而應用最小平方法可得更正確之結果.

用最小平方法時須解下列方程式:

$$\Sigma(Y)=na+b\Sigma(X)$$

$$\Sigma(XY)=a\Sigma(X)+b\Sigma(X^2)$$

解此兩方程式時所須之數値，可根據表九十六所列材料求之．代入方程式得

$$7,616=43\,a+3,602\,b$$

$$1,215,966=3,602\,a+603,968\,b$$

解之得

$$a=16.92$$

$$b=\ 1.91$$

所求之方程式爲

$$Y=16.92+1.91\,X\text{(註)}$$

此直線繪於圖67.

關於所研究之兩變數，卽各邦應課稅之個人收入額與各邦汽車登記數，其間關係吾人已得一數學式以代表之．在方程式中，前者爲自變數或X變數，後者爲應變數或Y變數．此方程式成爲此兩變數間函數關係之計量，惟僅爲平均關係之表現耳．然則此方程式效率如何？若爲一完全之關係，而繪成各點皆在描述此關係之線上，則可用方程式爲確切之工具，由一變數之値決定另一變數之値，吾人可確信其結果

*(註) 在討論繫聯之各章中，大寫字母(W,X,Y,等等)乃用以代表變量之原來數値，自實際數值尺度之零點計起者，小寫字母(w,x,y,等等)乃用以代表變數之數値，表現爲其算術中數之差離者.

爲正確.惟用一固定方程式代表之線亦可配合於與之相距甚遠而散漫分佈之諸點.遇此情形,方程式在表面上似爲描寫一確切之關係,惟離中趨勢過大,吾人不能用之而無疑.此與應用一平均數時之問題相同.吾人必須知此平均數之效率,知此平均數周圍之集中趨勢之程度爲如何,始能應用適當.因此代表兩變數間關係之方程式,若吾人不知其在實際經驗中有效之程度爲如何,實無甚意義.故對於所配繪之線之周圍之離中趨勢,吾人必須有一計量.

前曾指出,在敍述頻數分配時,標準差爲離中趨勢之最佳而普遍之計量.此種計量顯然爲決定一代表平均關係之方程式之可靠性時所須要者.此線上之標準差不僅一般地爲此方程式之效率指數,且使吾人能計量根據此方程式而作之估量(Estimation)之正確程度.

[標準誤之計算]

平均關係線上之標準差,作爲估量之正確性之計量者,可稱爲估量標準誤 (Standard error of Estimate) 或簡稱爲標準誤 (Standard error).'標準差'之一名詞,平常限於代表以算術中數爲標準而求得之諸差離平方之中數之根.標準誤之記號爲S.

計算S時,必須知與X之每一已知値相當之Y之正則値.將X之已知各値代入方程式

$$Y=16.92+1.91\,X,$$

Y之正則值即可求得.於是即可求Y之實在值對正則值之差離,所求之計量即此諸差離平方之中數之方根(標準差)也.計算之法說明於下表:

表九十七

(1) 邦名	(2) 汽車登記數 1921 (單位:千) Y實在值	(3) Y計算值	(4) δ (2)−(3)	(5) δ^2
阿拉巴馬	82	99	− 17	289
亞黎桑那	35	51	− 16	256
亞堪薩斯	67	82	− 15	225
加利福尼亞	673	755	− 82	6,724
科羅拉多	145	151	− 6	36
康涅狄格	137	252	−115	13,225
德拉瓦	21	48	− 27	729
佛羅里達	97	97	0	……
喬治亞	131	147	− 16	256
伊達阿	51	61	− 10	100
印第安那	400	304	+ 96	9,216
衣阿華	460	229	+231	53,361
堪薩斯	291	187	+104	10,816
肯塔基	126	149	− 23	529
路易亞阿那	80	147	− 67	4,489
麥因	77	101	− 24	576
瑪利蘭	140	233	− 93	8,649
密支干	477	495	− 18	324
明尼蘇達	328	256	+ 72	5,184
密西西比	65	67	− 2	4
密蘇里	346	348	− 2	4
蒙他拿	58	87	− 29	841

表九十七(續上)

(1) 州名	(2) 汽車登記數 1921 (單位:千) Y實在值	(3) Y計算值	(4) d (2)−(3)	(5) d^2
奈勃拉斯加	238	155	+83	6,889
奈瓦達	10	36	−26	676
新罕布什爾	42	78	−36	1,296
新墨西哥	24	40	−16	256
北加羅林那	148	101	+47	2,209
北達各塔	92	51	+41	1,681
俄亥俄	720	719	+1	1
俄克拉何馬	221	149	+72	5,184
俄勒岡	118	137	−19	361
路得島	54	109	−55	3,025
南加羅林那	90	65	+25	625
南達各塔	119	59	+60	3,600
田納西	117	134	−17	289
塔克薩斯	467	399	+68	4,624
猶太	47	67	−20	400
威爾蒙	36	51	−15	225
維基尼阿	141	162	−21	441
華盛頓	185	239	−54	2,916
西維基尼亞	93	160	−67	4,489
威士康新	341	300	+41	1,681
華俄明	26	59	−33	1,089
				157,790

$$S_y=\sqrt{\frac{157,790}{43}}$$

$$=60.6$$

計算之結果得S_y之值爲60.6 (S_y爲代表Y變數標準誤之記號.)解釋時,與以算術中數爲中心之標準差恰恰相同.假定各項在關係線周圍取近似於正態之分配,則全體例數百分六十八將在$\pm S$(在此地爲60.6)之範圍以內,百分之九十五將在$\pm 2S$(在此地爲121.2)以內,百分之九九・七將在$\pm 3S$(在此地爲181.8)以內.若無一點落於線(配合於代表X與Y之相當值之各點之線)外時,S將爲零,則根據X估量而得之Y值將絕對正確.線周圍之離中趨勢越小,S之值亦越小.因此S之值爲代表兩變數間關係之線之效率之指示者,請注意標準誤之單位與原來Y諸值所用之單位爲相同者.

[估量之方法]

且就此等結果而研究其意義.假定某一邦登記之汽車數爲吾人所不知,而必須與以估量,此時有二法可取,第一爲僅根據吾人關於Y變數之知識而估量.在所研究之四十三邦中,汽車總數爲7,616,000輛.以43除此總數,得平均數爲177,116輛.關於某一邦登記之汽車數無專門消息時,其他各邦數字之算術中數可以作爲該邦之最可能之數值.(一觀察數列之最可能之值,爲該數列之中數.)吾人有何法以判斷此估量之正確否乎?原來分配之標準差爲中數周圍離中趨勢程度之計量,故亦爲根據中數而作之估量之正確性之計量.分配若近於正態形式,則一百次中有六十八次之情形,該成爲問題之一邦,其實在值與中數之差不出標準差以上.表

九十六中所載各邦登記汽車數之標準差爲171.4.因此算術中數爲一種合理的估量之基礎,而標準差則對於作此估量時所含之機率予若干指示.(當然,根據人口或其他相似之因子而作估量之可能性,爲吾人此處所未論及者.)

估量某一邦登記汽車數時,若吾人知該邦應課稅之個人收入額,則可存另一方法.根據前數頁之研究結果,知汽車登記數與應課稅之個人收入額之間,其平均關係可以方程式代表之,其式如下:

$$Y=16.92+1.91\,X$$

(注意:每一變數,其單位皆爲1000)

若該邦應課稅之個人收入額爲200,000,則可由此方程式作估量,得該邦之汽車登記數爲399,000.此數乃根據平均關係方程式而決定之Y之最可能之值.此一估量較之以前以Y之中數作其最可能之值者是否較佳?吾人關於X,Y之間之平均關係之知識,對於根據X之一已知值而估量Y值是否有助?

此等問題,可以Y之標準誤以及Y之標準誤與Y之標準差間之關係答復之.Y之標準誤(即以平均關係線爲中心之標準差)爲60.6.標準差爲171.4.如此顯見由方程式所得之估量,較之根據Y之中數數値而作者爲正確.前一估量,則在一百次中有六十八次其錯誤不超過60.6,或以原來單位表示之,卽不超過60,600輛;後一估量,則在一百次中有六十八

次其錯誤不超171,400輛.(註)兩變數間之關係雖非固定與完全,然吾人對於此種關係之知識,足使吾人大大減少估量之錯誤

[繫聯係數]

今得兩種計量足以助吾人敘述變量間關係者.其一爲關係之基本方程式,表現兩個相伴之變數當其一發生某種變動時另一變數平均上所發生之變動之程度.其二爲標準誤,此爲在平均關係線周圍"分散"程度之計量.標準誤與標準差皆用絕對數表示,所用單位皆爲計量原來之 Y 諸值者,此其同點也.標準誤使吾人決定在某一例中根據關係方程式而作之估量落於某一定界限以內之機率.

ι. 在計量離中趨勢時,曾感有求一離中性之抽象計量之必要,抽象計量者,與原來問題中之絕對單位脫離關係者也.前曾指出,當取不同之分配互相比較時,此種計量特別須要.因此,爲計量離中程度之故,吾人曾用一離中係數.在目前之問題中,正須要一同樣之計量.吾人須有一表示兩變數間關係程度(degree of relationship)之計量,須有一抽象係數之與某一數列中所應用之特殊單位無關者,皮爾生(Karl Pearson)氏乃創一如此之指數.

(註) 在此例中,因有少數諸項與分配之正態過遠,故不能由 S_y 與 σ_y 決定確切之機率.特別因六個或八個特大之項使其他諸項相形見小者之存在,對於此兩種計量之影響甚大.在解釋中加此限制,則吾人可用 S_y 與 σ_y 爲離中趨勢之有用的計量.

此種計量可以前面所討論者說明之,前已指出根據關係方程式而作之估量,其效率可將Y之標準誤(關係線周圍分散狀態之計量)與Y之標準差相比較而決定.若標準誤與標準差相等,則關係方程式對於吾人毫無用處.但標準誤若小於標準差,則此方程式之應用可增進估量之正確性.故方程式之效率,可為標準誤與標準差間之關係所指示.但此兩者皆為絕對數項,故以此一除彼一,可得一抽象之計量.於是吾人可寫作

$$\text{繫聯之計量}=\frac{S_y}{\sigma_y}.$$

將此比寫成下式,更為有用:

$$\text{繫聯之計量}=\sqrt{1-\frac{S_y^2}{\sigma_y^2}}.$$

當此計量用於直線繫聯時,名為繫聯係數(coefficient of correlation),而以r為記號以表示之.

對此公式略加研究,可助吾人明瞭r之意義.若關係線周圍無離中趨勢之存在,S_y之值將為零;而方程式所敍述者為兩變數間一完全之關係.此時r之數值,觀公式即可知為1.

S_y之值,其最大限度為等於σ_y之值.在此種情形下,關係方程式不能改進吾人之估量,依照r之公式計之,其值將為0.此值表示在兩變數之間毫無關係存在;換言之,配合最佳

之直線將爲水平線,此線穿過代表Y諸值之中數之一點.此值表示Y之較大諸值與X之較大諸值相伴,或與其較小諸值相伴之傾向,並不存在.兩變數之變動彼此絕對無關.在此種情形下,每一點對於配合之線之差離,等於其對於中數之差離,而兩個"標準差",如前所述,彼此相等.

因此0與1爲r值之界限,實際工作中遇見之值,皆在此限界之內,關係程度高者則近於1.r之值越大,則越可信任關係方程式所表示之關係,與之接近之實例占甚大之百分數,就前舉處理汽車登記數與應課稅之個人收入額之一例言之,吾人得

$$
\begin{aligned}
r&=\sqrt{1-\frac{(60.6)^2}{(174.4)^2}}\\
&=.935.
\end{aligned}
$$

此值表示在所選各邦之此兩變數間有一固定而密切之關係存在.(註)

以關係方程式中常數b之正負號加於緊聯係數上,或可使其更有意義.此符號表示直線之傾斜爲正爲負,加於r之前,則吾人可知關係之性質爲正爲反.如現今一例中,一變數之大值與另一變數之大值相伴,緊聯爲正,係數應寫作

(註) 此地r值之所以若是其大者,一部分由於有六項或八項數值較主要之部分甚大.此種特例對於緊聯係數之影響,下文將詳細討論之.

+.935. 當取棉花生產量與棉花價格作成緊聯,則關係爲相反,蓋一變數之大值平常與另一變數之小值相伴者也.

當吾人已得上述三種計量時,則在一事例中之關係之計量業已完成.平均關係方程式爲連結兩變數之潛在法則(如可以假定有一法則存在)之表示.標準誤以絕對數單位計量關係線周圍之離中趨勢.緊聯係數爲一抽象計量,表示平均關係合於實際之程度.

[計算法詳解]

前節說明研究變量間關係時必須之各種計量時,故意避免對於計算法作詳細說明.茲可重行研究詳細之計算法,以及使計算手續減少至最小限度之若干方法.

前節說明中所用之手續,爲求三種計量時應用之合理手續,此法固可普遍應用,惟利用求 S_y 之一種簡法,尙可大減計算之繁煩.此法可先就上文引用之材料說明之.而目前之討論,暫限於變數間關係可以一直線描寫之事例.

第一步之問題爲求得關係方程式.用最小平方法配合

$$Y=a+bX$$

式之線.

第二步爲 S_y^2 之計算,卽標準誤平方之計算.在上文說明時,此數由計量各觀察値對於配合之線之差離而求諸差

離平方之中數以得之.然此數亦可由下列方程式求得:(註 1)

$$S_y^2=\frac{\Sigma(Y^2)-a\Sigma(Y)-b\Sigma(XY)}{N}.$$

a,b 兩值爲配合之直線之方程式中兩個常數,其他諸值則與原來之觀察值有關.將必須之各值代入此方程式,(各值由表九十六與(367 頁)求得)得:(註 2)

(註 1) 標準誤由下列公式計算之

$$S_y^2=\frac{\Sigma(d^2)}{N}$$

d 代表對於配合之線之一個差離,即在某一例中 Y 之實在值與計算值之差,計算值乃根據下列方程式求得者:

$$Y_c=a+bX.$$

(Y_c 代表 Y 之計算值)

若以 Y 代表實在值,則

$$d=Y_c-Y$$

或

$$d=a+bX-Y \qquad (1)$$

每有一點,即有一個此種形式之方程式.各以 d 乘之,相加,則得

$$\Sigma(d^2)=a\Sigma(d)+b\Sigma(dX)-\Sigma(dY). \qquad (2)$$

但線爲用最小平方法配合者,

$$\Sigma(d)=0$$

$$\Sigma(dX)=0$$

(證明見附錄一)

故

$$\Sigma(d^2)=-\Sigma(dY) \qquad (3)$$

回至方程式(1),各以 Y 乘之,相加得

$$\Sigma(dY)=a\Sigma(Y)+b\Sigma(XY)-\Sigma(Y^2) \qquad (4)$$

將 $\Sigma(dY)$ 之相當值代入方程式(3)則得

$$\Sigma(d^2)=\Sigma(Y^2)-a\Sigma(Y)-b\Sigma(XY) \qquad (5)$$

由此求出 S_y^2 之公式.

(註 2) 在此計算中,b 之數值之小數,比較前面所取者多取幾位.

$$S_y^2=\frac{2{,}612{,}066-(16.92\times 7{,}616)-(1.91238\times 1{,}215{,}966)}{43}$$

$$=\frac{157{,}814}{43}$$

$$=3{,}670$$

$$S_y=60.6$$

由此往前,可依照前已敍述之方法進行,r由下式計算之:

$$r=\sqrt{1-\frac{S_y^2}{\sigma_y^2}}$$

惟係數 r 可不經 S 而直接計算之.上列 r 之公式可化爲

$$r^2=\frac{a\Sigma(Y)+b\Sigma(XY)-Nc_y^2}{\Sigma(Y^2)-Nc_y^2}$$

式中c_y代表Y之中數與計算中應用之原點間之差.(註)若原

(註) 公式
$$r^2=1-\frac{S_y^2}{\sigma_y^2}$$

可寫作
$$r^2=1-\frac{\Sigma(d^2)}{\Sigma(y^2)}$$

式中y代表Y諸值與其算術中數之差離.但

$$\frac{\Sigma(y^2)}{N}=\frac{\Sigma(Y^2)}{N}-c_y^2$$

式Y代表對於一假定原點(在此地爲原來尺度上之零點)之差離,c_y代表此原點與Y諸值之中數間之差.

故
$$r^2=1-\frac{\Sigma(d^2)}{\Sigma(Y^2)-Nc_y^2}$$

將$\Sigma(d^2)$之相當值$(\Sigma(d^2)=\Sigma(Y^2)-a\Sigma(Y)-b\Sigma(XY))$代進此方程式,即得

$$r^2=1-\frac{\Sigma(Y^2)-a\Sigma(Y)-b\Sigma(XY)}{\Sigma(Y^2)-Nc_y^2}.$$

化簡,得
$$r^2=\frac{a\Sigma(Y)+b\Sigma(XY)-Nc_y^2}{\Sigma(Y^2)-Nc_y^2}.$$

點爲原來之Y尺度上之原點,c_y將等於Y諸值之算術平均數.

現今應用表九十六之材料,得

$$c_y=\frac{7,616}{43}=177.116.$$

其他諸值與上文計算S_y時所用者相同,代入公式,得

$$r^2=\frac{1,105,337.98}{1,263,152.7}$$

$$=.875$$

$$r=.935$$

因此實際上在用最小平方法配合一直線之後,求S及r(欲完全描寫兩變量間之關係,此兩者爲必須之計量)時所須之數值皆已齊備,須另求者僅$\Sigma(Y^2)$與c_y而已.

〔繫聯表之構成〕

上述之例僅包含四十三個觀察值,若觀察值爲數極多,欲保留各個原來數值而研究關係,實際上爲不可能之事.個別各項必須集爲重要之幾組,一切計算必須完全根據已分組之材料而進行.此僅謂吾人必須處理已彙爲頻數分配之材料也.吾人所處理者既爲兩個變數,過去之簡單頻數表必須改變以適合目前問題之需要.此種變形之頻數表,排列以助吾人研究關係時計算所須之數值者,名爲繫聯表(correlation table).

茲研究聯邦準備銀行對於一種標準形式之商業票據之貼現率與商業銀行對於同樣票據之貼現率間之關係,以爲製作繫聯表之標本問題.會員銀行接收此票據後,既可向聯邦準備銀行再貼現,故兩者之貼現率間有某種程度之關係存在,乃可以預料之事也.目前吾人之目的,爲予此關係以計量.

第一步爲將原來之觀察值列表,已有之材料,爲自 1920 年七月一日至 1923 年十二月一日止之42個月間十二個聯邦準備銀行所在之每一城市之每一變數之每月份數值.(註)

X—— 聯邦準備銀行貼現率

Y—— 商業銀行貼現率

	3.75 至 4.24%	4.25 至 4.74%	4.75 至 5.24%	5.25 至 5.74%	5.75 至 6.24%	6.25 至 6.74%	6.75 至 7.24%	共計
7.75 至 8.24%					/	/	//	4
7.25 至 7.74%		/			[illegible]	[illegible]	/	17
6.75 至 7.24%			[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	117
6.25 至 6.74%		[illegible]	[illegible]	[illegible]	[illegible]			47
5.75 至 6.24%	/	[illegible]	[illegible]	[illegible]	[illegible]			165
5.25 至 5.74%	[illegible]	[illegible]	[illegible]					126
4.75 至 5.24%	[illegible]	[illegible]						34
4.25 至 4.74%	//							3
共計	24	227	48	20	123	20	42	504

圖68——列表求聯繫表中各項之法

(註) 聯邦準備銀行之貼現率,乃關於90日滿期之已承受之商業票據者,商業銀行之貼現率,乃關於顧客之第一等商業票據以30日至60日滿期者(有爲30日至90日者),以每30日之時期內慣常之率代表該時期之中點.

製表時使聯邦準備銀行貼現率與同一城市之商業銀行貼現率相配合,圖68示製表之法.

列表既畢,乃可作一繫聯表,以便利以後之計算.表如下:

表九十八

繫聯表

聯邦準備銀行之貼現率及商業銀行之貼現率

X——聯邦準備銀行貼現率(%)

Y——商業銀行貼現率(%)

組距						3.75—4.24	4.25—4.74	4.74—5.24	5.25—5.74	5.75—6.24	6.25—6.74	6.75—7.24	共計
	中點					4.00	4.50	5.00	5.50	6.00	6.50	7.00	
		f				21	227	48	20	123	20	42	504
			d'			0	1	2	3	4	5	6	
				fd'		0	227	96	60	492	100	252	1227
					f(d')²	0	227	192	180	1968	500	1512	4579
7.75—8.24	8.00	4	8	32	256					1	1	2	
7.25—7.75	7.50	17	7	119	833					7	9	1	
6.75—7.24	7.00	117	6	702	4212			5	4	63	9	36	
6.25—6.74	6.50	47	5	235	1175		2	9	10	22	1	3	
5.75—6.24	6.00	155	4	924	2496	1	90	29	6	30			
5.25—5.74	5.50	126	3	378	1134	11	110	5					
4.75—5.24	5.00	34	2	68	136	10	24						
4.25—4.74	4.50	3	1	3	3	2	1						
共計		504		2161	10245								

注意在此表中每一變數皆有一假定原點,計算時所用之單位爲組距.吾人可用X'與Y'代表對於假定原點之差離(假定原點之位置爲原來尺度上4,4之一點.)

[關係量之計算]

配合一直線於材料時以及求 S, r 兩計量時所須之一切數值,無一不可由此繫聯表求得.計算 $\Sigma(X')$, $\Sigma(X')^2$, $\Sigma(Y')$ 及 $\Sigma(Y')^2$ 所用之法,爲讀者所已熟知者,配對之數值之乘積 $\Sigma(X'Y')$ 可直接由繫聯表計算之.惟爲初學者求簡單計,或宜將材料從新整理爲直行式如下表.表中之每一列代表原來繫聯表中記錄頻數之一格.

表九十九

聯邦準備銀行之貼現率與商業銀行之貼現率

(計算配合曲線時所需數值之方法)

(1) X'	(2) Y'	(3) f	(4) $f(X'Y')$
0	1	2	0
0	2	10	0
0	3	11	0
0	4	1	0
1	1	1	1
1	2	24	48
1	3	110	330
1	4	90	360
1	5	2	10
2	3	5	30
2	4	29	232
2	5	9	90
2	6	5	60
3	4	6	72
3	5	10	150
3	6	4	72
4	4	30	480
4	5	22	440
4	6	63	1,512
4	7	7	196
4	8	1	32
5	5	1	25
5	6	9	270
5	7	9	315
5	8	1	40
6	5	3	90
6	6	36	1,296
6	7	1	42
6	8	2	96
			6,289

配合一直線及計算標準誤與緊聯係數時所須之數值爲：

$$N = 504 \qquad \Sigma(X')^2 = 4{,}579$$

$$\Sigma(X') = 1{,}227 \qquad \Sigma(X'Y') = 6{,}289$$

$$\Sigma(Y') = 2{,}161 \qquad \Sigma(Y')^2 = 10{,}245$$

配合最佳之直線,其方程式爲：

$$Y' = 2.7155 + .6458\,X'.$$

代入公式

$$S_y^2 = \frac{\Sigma(Y')^2 - a\Sigma(Y') - b\Sigma(X'Y')}{N}$$

得

$$S_y^2 = \frac{10{,}245 - (2.7155 \times 2{,}161) - (.6458 \times 6{,}289)}{504}$$

$$= .6257$$

$$S_y = .791$$

欲決定緊聯係數之值時,僅需將適當數值代入方程式

$$r^2 = \frac{a\Sigma(Y') + b\Sigma(X'Y') - Nc_y^2}{\Sigma(Y')^2 - Nc_y^2}$$

代入後得

$$r^2 = \frac{(2.7155 \times 2{,}161) + (.6458 \times 6{,}289) - (504 \times 18.38437)}{10{,}245 - (504 \times 18.38437)}$$

$$= \frac{663.9}{979.27}$$

$$= .67795$$

$$r = +.82$$

以上一切計算皆以組距爲單位,以原來尺度上4.4之一點爲原點.r之值不因此而受影響,惟估量方程式與標準誤皆須改正.

S_y之值,以組距爲單位時,爲.791.因Y變數之組距爲.5%,故用原來之單位時,

$$S_y = .5\% \times .791$$
$$= .40\%$$

方程式亦可同樣改正.因X, Y兩者之組距皆爲.5%,故可寫作

$$2Y' = 2.7155 + .6458(2X')$$

或

$$Y' = 1.3577 + .6458X'$$

決定在所研究之時期內聯邦準備銀行貼現率與相當之商業銀行貼現率間之關係之數值今有三個,方程式爲平均關係之敍述,標準誤爲根據此方程式而作之估量之可靠性之計量,緊聯係數爲兩變數間之關係之程度之抽象計量.

標準誤S_y之意義,用圖69表示之.此散佈圖上畫一平均關係線,可稱爲"估量帶"(zone of estimate)者亦於此線之周圍劃出.包括等於$2S$之闊度而以配合之直線爲軸之估量帶,假定分配爲正態時,全體點數應有68%位於其內.闊度等於$6S$而以配合之直線爲軸之估量帶,在同一假定下,全體點數應有99.7%落於其內.S之值越小,帶亦越狹,因此根據平均

關係方程式而作之估量將超正確.

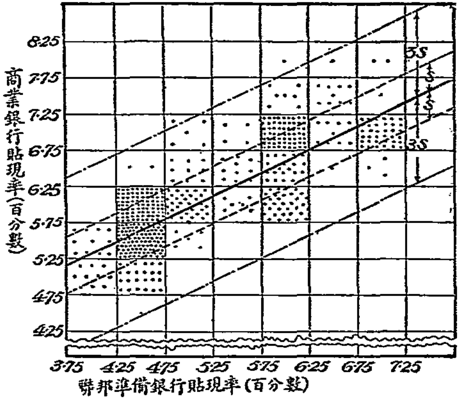


圖69.——散佈圖:表示聯邦準備銀行貼現率與商業銀行貼現率之關係,並其平均關係線及估量帶.

[繫聯係數之乘積率公式]

以前所舉諸例,繫聯係數皆由下列公式算出:

$$r^2=\frac{a\Sigma(Y)+b\Sigma(XY)-Nc_y^2}{\Sigma(Y^2)-Nc_y^2}$$

此公式乃從以最小平方法配合直線之方法直接化出者.平常應用之公式與此少有不同,學者宜熟知之.

當配合一直線於材料時,若以平均數之一點爲原點,則

兩個正則方程式原爲

$$\Sigma(Y)=na+b\Sigma(X)$$

$$\Sigma(XY)=a\Sigma(X)+b\Sigma(X^2)$$

者,化成

$$\Sigma(xy)=b\Sigma(x^2),$$

因

$$\Sigma(x)=0 \qquad 而 \qquad \Sigma(y)=0$$

也.此時所須求之常數僅爲代表斜度之b,其值可以下式之關係求之

$$b=\frac{\Sigma(xy)}{\Sigma(x^2)}.$$

在同樣情形之下,下列公式

$$r^2=\frac{a\Sigma(Y)+b\Sigma(XY)-Nc_y^2}{\Sigma(Y^2)-Nc_y^2}$$

化爲

$$r^2=\frac{b\Sigma(xy)}{\Sigma(y^2)},$$

因當以Y諸值之中數爲計算差離之原點時,$c_y=0$故也.以b之相當數代b,則得

$$r^2=\frac{\Sigma(xy)\cdot\Sigma(xy)}{\Sigma(y^2)\cdot\Sigma(x^2)}.$$

但 $\Sigma(y^2)=N\sigma_y^2, \Sigma(x^2)=N\sigma_x^2$.

故

$$r^2=\frac{\Sigma(xy)\cdot\Sigma(xy)}{N^2\sigma_y^2\sigma_x^2}$$

而

$$r=\frac{\Sigma(xy)}{N\sigma_x\sigma_y}$$

式中 x 與 y 代表以平均數之點爲原點之差離.

此公式可寫作

$$r=\frac{p}{\sigma_x\sigma_y}$$

式中

$$p=\frac{\Sigma(xy)}{N},$$

p 之值爲諸對 x 與 y 之乘積之算術中數(平均乘積 mean product).

依此公式計算緊聯係數時,所取之途徑與前述者少許不同.吾人知求算術中數與標準差時,可任取一假定原點以爲計算一切差離之中心,而後校正其因此假定原點而引起之錯誤.同樣,平均乘積 p 亦可用簡法求之,先用一假定中數而後校正之.

若以 x' 與 y' 代表對於任意選定之假定中數點之差離,而 p' 代表此種差離之平均中數,則

$$p'=\frac{\Sigma(x'y')}{N}.$$

計算p'非難事,因差離可以中點爲原點而計算,而用組距爲單位也.既得p'後,吾人可依下列公式求眞正之平均乘積,其式爲:

$$p=p'-c_xc_y$$

式中c_x與c_y各代表x諸值與y諸值之眞正中數與假定中數間之差.(註)

(註) 下面爲此種關係之證明:

x'=任何一點對於x諸值之假定中數之差離.

x=同點對於x諸值之眞正中數之差離.

c_x=x諸值之眞正中數與假定中數間之差.

y'=同點對於y諸值之假定中數之差離.

y=同點對於y諸值之眞正中數之差離.

c_y=y諸值之眞正中數與假定中數間之差.

$$x'=x+c_x$$

$$y'=y+c_y$$

$$x'y'=(x+c_x)(y+c_y)=xy+c_xy+c_yx+c_xc_y$$

求N個點各點之此種乘積之和,得

$$\Sigma(x'y')=\Sigma(xy)+c_x\Sigma(y)+c_y\Sigma(x)+Nc_xc_y.$$

但 $\Sigma(y)=0$ 與 $\Sigma(x)=0$

故 $$\Sigma(x'y')=\Sigma(xy)+Nc_xc_y$$

$$\frac{\Sigma(x'y')}{N}=\frac{\Sigma(xy)}{N}+c_xc_y$$

$$\frac{\Sigma(xy)}{N}=\frac{\Sigma(x'y')}{N}-c_xc_y$$

或

$$p=p'-c_xc_y$$

[乘積率法用於未分組材料之情形]

此法可先就用於未分組材料之情形，借前文中美國各邦個人收入額(X)與汽車登記數(Y)之數字以說明之。計算時所須諸值，如表九十六所示者，如下：

$$N=43$$

$$\Sigma(X)=3,602$$

$$\Sigma(Y)=7,616$$

$$\Sigma(X^2)=603,968$$

$$\Sigma(Y^2)=2,612,066$$

$$\Sigma(XY)=1,215,966$$

平均乘積可由下列方程式求之

$$p=\frac{\Sigma(xy)}{N}=\frac{\Sigma(x'y')}{N}-c_xc_y$$

吾人可取原來之兩尺度上之真正原點爲假定原點，得

$$p=\frac{\Sigma(XY)}{N}-c_xc_y.$$

求兩個標準差之公式爲

$$\sigma_x=\sqrt{\frac{\Sigma(X^2)}{N}-c_x^2}$$

$$\sigma_y=\sqrt{\frac{\Sigma(Y^2)}{N}-c_y^2}$$

(當假定原點在原來尺度上爲零點時，X與x'兩記號，Y

與y'兩記號各各相合,如上列公式之情形.)

此等計量可根據表九十六之數值極易求得之:

$$c_x=\frac{3,602}{43}=83.767 \qquad c_y=\frac{7,616}{43}=177.116$$

$$c_x^2=7,016.910 \qquad c_y^2=31,370.077$$

$$p=\frac{1,215,966}{43}-14,836.476$$

$$=13,441.803$$

$$\sigma_x=\sqrt{\frac{603,968}{43}-7,016.910}$$

$$=83.84$$

$$\sigma_y=\sqrt{\frac{2,612,066}{43}-31,370.077}$$

$$=171.41$$

代入公式解之,求繫聯係數,得

$$r=\frac{p}{\sigma_x\sigma_y}=\frac{13,441.803}{83.84\times171.41}=+.935$$

描寫X與Y間平均關係之直線之方程式,可由上面計算中所須之數值求得之.當以平均數之點為原點時,此方程式可寫作

$$y=r\frac{\sigma_y}{\sigma_x}x.$$

代入適當之值,得

$$y=+.935\cdot\frac{171.41}{83.84}\cdot x$$

$$=1.91\,x$$

注意此即用最小平方法求得之方程式.此中無代表 y 截數之常數項存在,因原點在平均數之一點上,而為最小平方線所必須通過者也.(註)

(註) 公式

$$y=r\frac{\sigma_y}{\sigma_x}x$$

之等於根據最小平方法而作之公式一點極易證明.當直線通過平均數點時,$Y=a+bX$ 之方程式變為 $y=bx$.但

$$b=\frac{\Sigma(xy)}{\Sigma(x^2)}.$$

故可寫作

$$y_c=\frac{\Sigma(xy)}{\Sigma(x^2)}x$$

此即等於

$$y_c=r\frac{\sigma_y}{\sigma_x}x$$

因此從一方程式可寫作

$$(1)\ y_c=\frac{\Sigma(xy)}{N\sigma_y\sigma_x}\cdot\frac{\sigma_y}{\sigma_x}x$$

$$(2)\ y_c=\frac{\Sigma(xy)}{N\sigma_x\cdot\sigma_x}x$$

$$(3)\ y_c=\frac{\Sigma(xy)}{N\sqrt{\frac{\Sigma(x^2)}{N}}\cdot\sqrt{\frac{\Sigma(x^2)}{N}}}x$$

$$(4)\ y_c=\frac{\Sigma(xy)}{\Sigma(x^2)}x$$

(y_c 一記號在此等方程式中代表 y 之計算值,以免與方程式右邊之 y 之實在值相混.)

當用乘積率法計算繫聯係數乃決定響應方程式(Equation of regression)時,標準誤S_y可將最初舉出之r公式略爲改變而求之.由公式

$$r^2=1-\frac{S_y^2}{\sigma_y^2}$$

可得公式

$$S_y=\sigma_y\sqrt{1-r^2}.$$

如吾人已得σ_y與r之值,即可就此公式計算S_y之值.此地即爲

$$S_y=171.41\sqrt{1-(.935)^2}$$
$$=60.6$$

[乘積率法用於已分組材料之情形]

乘積率法對於必須製作雙頻數表或繫聯表之材料亦可應用.表一百示其詳細手續.

此表除所選之假定原點不同外,其餘悉如前文舉以表示同一材料之表相同.

取5.5之一數値作爲X諸値之假定中數,取6.5之一數値爲Y諸値之假定中數.以組距爲單位由此原點計算差離.繫聯表之每一格內有三個數字,與計算$\Sigma(x'y')$之値相關連者.在中間之一數字爲落於該格之項數.如有七對數値其X之値在5.75—6.25(中點6.0)之間,而其Y之値在7.25—7.75(中點7.5)之間.此七對數値每對之x'(X諸値對於假定中數之

表一百

聯繫表

表示聯邦準備銀行貼現率與商業銀行貼現率間之關係並說明計算聯繫係數之法

Y—商業銀行貼現率(%) 組距	中點	f	d'	fd'	$f(d')^2$	X—聯邦準備銀行貼現率(%) 3.75—4.24	4.25—4.74	4.75—5.24	5.25—5.74	5.75—6.24	6.25—6.74	6.75—7.24	共計	$\Sigma(x'y')$
	中點					4.0	4.5	5.0	5.5	6.0	6.5	7.0		
		f				24	227	48	20	123	20	42	504	
			d'			−3	−2	−1	0	+1	+2	+3		
				fd'		−72	−454	−48	0	+123	+40	+126	−285	
					$f(d')^2$	216	908	48	0	123	80	378	1753	
7.75—8.24	8.0	4	+3	+12	36					+3 1 (+3)	+6 1 (+6)	+9 2 (+18)		+27
7.25—7.74	7.5	17	+2	+34	68					+2 7 (+14)	+4 9 (+36)	+6 1 (+6)		+56
6.75—7.24	7.0	117	+1	+117	117			−1 5 (−5)	0 4 (0)	+1 63 (+63)	+2 9 (+18)	+3 36 (+108)		+184
6.25—6.74	6.5	47	0	0	0		0 2 (0)	0 9 (0)	0 10 (0)	0 22 (0)	0 1 (0)	0 3 (0)		
5.75—6.24	6.0	156	−1	−156	156	+3 1 (+3)	+2 99 (+198)	+1 20 (+20)	0 6 (0)	−1 30 (−30)				+182
5.25—5.74	5.5	126	−2	−252	504	+6 11 (+66)	+4 110 (+440)	+2 5 (+10)						+516
4.75—5.24	5.0	31	−3	−102	306	+9 10 (+90)	+6 24 (+144)							+231
4.25—4.74	4.5	3	−4	−12	48	+12 2 (+24)	+8 1 (+8)							+32
共計		504		−359	1235									+1222

$A.M_x=5.5$

$c_x=\frac{-285}{504}$

$=-.565$

$c_x^2=.319$

$s_x^2=\frac{\Sigma f(d')^2}{N}$

$=\frac{1753}{504}$

$=3.478$

$\sigma_x^2=s^2-c^2$

$=3.478-.319$

$=3.159$

$\sigma_x=\sqrt{3.159}$

$=1.777$

$A.M_y=6.5$

$c_y=\frac{-359}{504}$

$=-.712$

$c_y^2=.507$

$s_y^2=\frac{1235}{504}$

$=2.450$

$\sigma_y^2=2.450-.507$

$=1.943$

$\sigma_y=1.394$

$p=\frac{\Sigma(x'y')}{N}-c_xc_y$

$=\frac{1225}{504}-.402$

$=2.4305-.402$

$=2.0285$

$r=\frac{p}{\sigma_x\sigma_y}$

$r=\frac{2.0285}{1.777\times1.394}$

$=\frac{2.0285}{2.477}$

$r=+.82$

$M_x=5.218$

$M_y=6.144$

(註)表內一切計算皆以組距爲單位.

差離),以組距爲單而計之皆爲 +1,而其 y' (Y 諸値對於假定中數之差離),以組距爲單位而計之,皆爲 +2.故每對之乘積爲 $x'y'=+2$. 此數寫在方格之上方.然在此方格內者共有七對如此之數字,故此一集團之 $x'y'$ 之總和爲 +14. 此數寫在方格下方之括弧內.欲求 $\Sigma(x'y')$, 必須將所有各格內如此得來之數値求其代數和,加法先就各列進行,每列之和寫在表之最右一行內,然後總加之得 $\Sigma(x'y')=+1225$,以組距爲單位.

計算標準差時求 c_x, c_y 之法爲吾人所熟知者,在此地以組距爲單位其數値爲

$$c_x=-.565 \qquad c_y=-.712$$

故得

$$\frac{\Sigma(xy)}{N}=\frac{\Sigma(x'y')}{N}-c_xc_y$$

$$=\frac{1225}{504}-.402$$

$$=2.0285$$

此爲以組距爲單位時平均乘積p之數值.更進,則

$$r=\frac{\Sigma(xy)}{N\sigma_x\sigma_y}$$

$$=\frac{p}{\sigma_x\sigma_y}$$

$$=\frac{2.0285}{1.777\times1.394}$$

$$=+.82$$

計算r時,最後一分數式中之分子分母(平均乘積與兩個標準差)皆以組距爲單位,故r可求得而不必先化該等數字爲原來之單位.故全部計算皆用簡單之組距單位進行.

求代表一描寫x與y間平均關係之直線之方程式時,應用

$$y=r\frac{\sigma_y}{\sigma_x}x,$$

則 σ_y 與 σ_x 應以原來尺度之單位表示.(註) 以組距乘現在之數值卽可化之爲如此之單位.

$$\sigma_x\,(\text{用原來單位}) = 1.777 \times .50 = .8885$$

$$\sigma_y\,(\text{用原來單位}) = 1.394 \times .50 = .697$$

將已知值代入公式,得

$$y = .82\frac{.697}{.8885}x$$

$$= .64x.$$

[響應線]

論述繫聯時平常所應用之若干名詞,在前面之討論中故意省去.惟其中有幾個應加說明者.

在前面之說明中,配合最佳之線之方程式,當平均數之點爲原點時,爲

$$y = .64x.$$

在此方程式中,y 表示爲 x 之函數;卽以 x 作爲自變數而 y 作爲應變數.此方程式表示在 x (聯邦準備銀行貼現率)發生一單位之變動時 y (商業銀行貼現率)所發生之平均變動.此關係線恰與長期趨勢線相當,蓋後者所表示者,爲時間上發生一單位之變動時在某一數列中所發生之平均變動也.如此

(註) 當組距相同有如現在之情形時,此種改變並非必要,蓋分母分子兩者係未能因之而變也.在實際上,在計算之此一階段中,兩個標準差均以用原來之單位表示爲佳.

描述兩變數間之平均關係之一線，名爲響應線 (line of regression). 其方程式名爲響應方程式 (regression equation)，而代表如此一線之斜度之 $r\frac{\sigma_y}{\sigma_x}$ 之值名爲響應係數 (coefficient of regression). 最初用此等名詞者爲加爾登 (Galton) 氏，在其關於父輩之身長與兒輩之身長間之關係之研究中，發見兒輩之身長對於種族平均身長較之父輩身長對於種族平均身長爲接近，不論父輩爲高於平均身長或低於平均身長，兒輩有回至或復歸(regress)於平均數之趨向. 故稱圖上表示此兩變數間關係之線爲 line of regression. 此名詞現仍普遍應用，雖其原來之意義於應用時大半已毫無關係矣.（譯者按中文譯名"響應線"或"回應線"本已不復充分表示原來之意義.）

任何之情形下皆可計出兩條響應線. 其一表示 Y 爲應變數 X 爲自變數時兩者間之平均關係；其一表示 X 爲應變數 Y 爲自變數時兩者間之平均關係. 兩者之意義可用圖表示之.

圖 70 爲直接由圖 69 之散佈圖化出者. 每行之小圈代表落於該行內之 Y 各項之中數數值，如第一行內包括二十四例，將 X 值位於 3.75% 與 4.25% 之間之各項全包在內. 二十四例之 X 值固同屬一組，而 Y 值則各異，列頻數分配表示之. 其他各行用同樣方法各得一中數，各中數與 Y 對 X 之響應線 (line of regression of Y on X) 皆繪於圖 70 上.

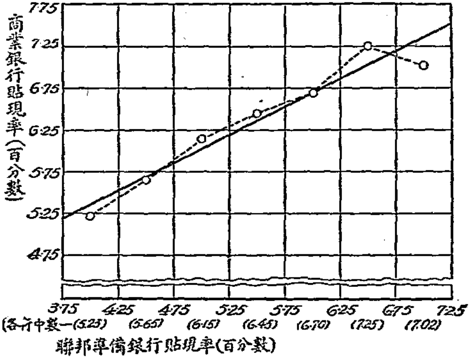


圖70.——表示商業銀行貼現率與聯邦準備銀行貼現率間之關係(折線聯結各行之中數,直線表示聯邦準備銀行貼現率發生一單位之變動時,商業銀行貼現率所發生之平均變動,即表示Y對X之影響.)

表一百另一

表一序列之算術中數之計算表

組距	中點 (m)	頻數 (f)	fm
5.75—6.24	6.0	1	6.0
5.25—5.74	5.5	11	60.5
4.75—5.24	5.0	10	50.0
4.25—4.74	4.5	2	9.0
		24	125.5

$$M=\frac{125.5}{24}=5.23$$

圖中 X 變數(聯邦準備銀行貼現率)爲自變數．當其由 4.0% 增至 4.5, 5.0, 5.5% 等等時，商業銀行貼現率之平均數亦隨之增高．商業銀行之平均貼現率 5.23% 與聯邦準備銀行貼現率 4.0% 相聯；商業銀行之平均貼現率 5.65%，與聯邦準備銀行平均貼現率 4.5% 相聯；諸如此類．直線——嚮應線或平均關係線——之斜度，表示聯邦準備銀行貼現率增加一單位時商業銀行貼現率與之相應之平均增加率．

兩變數間之關係更可由另一方面觀察之．如問題爲：已知商業銀行之某一貼現率時，與之相聯之聯邦準備銀行平均貼現率爲何？已知商業銀行貼現率之某一變動，聯邦準備銀行貼現率之與之相當之變動爲何？此時商業銀行之貼現率視作自變數，而聯邦準備銀行貼現率爲與之相聯之應變數．問題可以圖 71 答復之．以小圈表示而以一折線連結之諸點，代表每列間各項之算術中數．如第一列中三個 X 項之平均數爲 4.17%．此爲與商業銀行貼現率 4.5% 相聯之聯邦準備銀行平均貼現率．聯邦準備銀行平均貼現率之與商業銀行貼現率 5.0% 相聯者，爲 4.35%，諸如此類．配合於此諸點之直線表示兩貼現率間之關係，斜度表示商業銀行貼現率發生一單位之變動時聯邦準備銀行貼現率與之相聯而起之平均增加(或減少)量．

此線即 X 對 Y 之嚮應線 (line of regression of X on Y)．代表此直線之方程式，其普通公式如下：

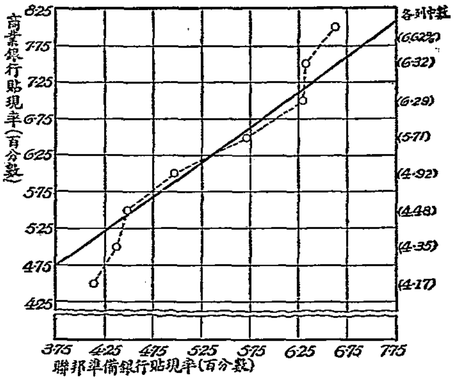


圖71.——表示商業銀行貼現率與聯邦準備銀行貼現率間之關係.(折線聯結各列中數,直線表示商業銀行貼現率發生一單位之變動時,聯邦準備銀行貼現率所發生之平均變動.即表示X對Y之響應.)

$$x=r\frac{\sigma_x}{\sigma_y}y$$

代入現在之數值,得

$$x=.82\frac{.8885}{.697}y$$

或

$$x=1.045y$$

吾人可見此方程式中之各因子,與Y對x之響應線之

公式中若無異.(註) 若 r 等於一,則兩線相合;若兩個標準差且相等,則響應線將平分坐標軸所成之角.在無論何種情形之下,各點若作於以 x 及 y 之標準差為單位分格之圖紙上,可得 $y=rx$,而響應線之斜度等於 r 之值.

響應係數以 b 為記號.在一簡單之繫聯問題中,皆有兩個此種係數,代表兩條響應線之斜度,其式如下:

$$b_{yx}=r\frac{\sigma_y}{\sigma_x}$$

$$b_{xy}=r\frac{\sigma_x}{\sigma_y}$$

(b 字後之兩個小型字母表示兩變數之關係,位於前者為應變數.)

兩公式中皆有係數 r 之存在.既如此,r 顯然可從響應

(註) 公式

$$x=r\frac{\sigma_x}{\sigma_y}y$$

可化為

$$x=\frac{\Sigma(xy)}{\Sigma(y^2)}y$$

此為配合於圖 71 上各點之一直線之方程式,以此直線為中心之水平差離 (horizontal deviation) 之平方之和為一最小數.

公式

$$y=\frac{\Sigma(xy)}{\Sigma(x^2)}x$$

為配合於圖70 上各點之一直線之方程式,以此直線為中心之垂直差離 (vertical deviation) 之平方之和為一最小數.明乎此則兩條響應線之差別可知矣.

係數算出.因

$$r=\sqrt{b_{yx}\cdot b_{xy}}=\sqrt{r\frac{\sigma_y}{\sigma_x}\cdot r\frac{\sigma_x}{\sigma_y}}=\sqrt{r^2}.$$

故吾人若已知兩條響應線之斜度,r 即可決定.在現今之例中,用已知之確切數值,得

$$r=\sqrt{.64248\times 1.045}=.819$$

[響應方程式之用法]

前文所舉之兩個響應方程式爲

$$y=.64x$$

及

$$x=1.045y,$$

其所表示者,爲X對於X諸值算術中數之差離,與Y對於Y諸值算術中數之差離間之關係;卽原點爲在平均數之一點也;應用此方程式時,吾人可用X與Y之原來數值,惟必須先化之爲對於其中數之差離.舉例以明之,如欲決定與聯邦準備銀行貼現率6% 相聯之商業銀行之正則貼現率,X變數(聯邦準備銀行貼現率)之中數數值爲 5.218%.6% 之率對於其中數之差離爲 +.782,將此值代入上列公式之第一個,得

$$y=.64(+.782)$$
$$=+.500$$

此數爲與 x 差離(+.782)相聯之 y 平均差離.欲求與聯邦準備銀行貼現率6% 相聯之商業銀行正則貼現率, +.500% 必

須加於商業銀行貼現率之中數 6.144% 之上,故吾人所求之值爲6.644%.

由於關係方程式之形式不便於此種計算,故計算時頗費周折.此方程式可改爲較便利之形式如下.

設

$$\overline{X}=X\text{ 諸值之算術中數}$$

$$\overline{Y}=Y\text{ 諸值之算術中數}$$

則

$$y=r\frac{\sigma_y}{\sigma_x}x$$

可寫作

$$Y-\overline{Y}=r\frac{\sigma_y}{\sigma_x}(X-\overline{X})$$

在此最後之一方程式中,X 與 Y 所代表者爲變數在原來尺度上之數値,而非對於各自中數之差離.以坐標圖之用語言之,如此之變更,表示將原點由平均數之一點移至與原來尺度上零點相當之一點.

此種形式之方程式效用較大,爲說明此點計,茲將

$$y=.64\,x$$

之方程式用上法改變之,得

$$Y-6.144=.64(X-5.218)$$

$$=.64\,X-3.340$$

$$Y=.64X+2.804$$

此方程式,即原點業經移動而變數之原來數值可以直接應用者也.欲求與聯邦準備銀行貼現率6.0%相聯之商業銀行平均貼現率,可以6.0%代入方才求得之方程式.

$$Y=(.64\times 6.0)+2.804$$
$$=6.644$$

求得之結果,與用另一種形式之方程式求得者完全相等,但在應用之許多方面、皆以能直接代入實在數值之方程式為較佳.

方程式

$$x=r\frac{\sigma_x}{\sigma_y}y$$

可同樣改為

$$X-\overline{X}=r\frac{\sigma_x}{\sigma_y}(Y-\overline{Y})$$

[結論]

計量兩變數間關係時,須有若干數值;前面許多頁中所討論者,為計算此等數值時所用之兩種方法,兩法頗不相同.茲可綜括敍述兩者之步驟如下.最小平方法.固為兩者所共同之基礎,惟最好用此名詞代表上述之第一法,因第一法中配合關係線一事,為其第一而且主要之步驟也.

I. 最小平方法.

A. 材料不須分組者.

1. 用最小平方法配合一直線於材料,將材料簡單列

爲幾行，所須之$\Sigma(X)$，$\Sigma(Y)$，$\Sigma(X^2)$，$\Sigma(Y^2)$，$\Sigma(XY)$諸值即極易求得．如此求得之方程式足以描述兩變數間之平均關係．

2. 根據下列公式求標準誤S_y之值：

$$S_y^2=\frac{\Sigma(Y^2)-a\Sigma(Y)-b\Sigma(XY)}{N}$$

S_y爲根據關係方程式所作估量之可靠性之計量，解釋時與解釋以算術中數爲中心之標準差時情形相同．

3. 根據下列公式求緊聯係數r之值：

$$r^2=\frac{a\Sigma(Y)+b\Sigma(XY)-Nc_y^2}{\Sigma(Y^2)-Nc_y^2}$$

將響應方程式中常數b之正負號加於r．在兩變數之關係可用一直線表示之範圍內，此係數爲此關係之程度之計量．

4. 若欲求一敍述X對Y之響應關係之方程式(X爲應變數)，可以適當之數值代入下列兩個正則方程式

$$\Sigma(X)=na+b\Sigma(Y)$$

$$\Sigma(XY)=a\Sigma(Y)+b\Sigma(Y^2)$$

所得之方程式將如下式：

$$X=a+bY$$

欲求S_x，可將S_y之公式加以適當之改變而計算之．r之值將與前面以Y爲應變數時之值相同．

B. 材料須分組者.

1. 選定一適當之組距,將各項列爲繫聯表形式.

2. 計算配合一直線於材料時所必須之數值.計算時可爲每一變數選定一假定原點,而一切數值皆以組距爲單位表示之.計算

$$\Sigma(X'Y')$$

之值時,爲便利起見可重列一直行式之表.

3. 依上面已舉出之公式計算標準誤.

4. 依上面已舉出之公式計算繫聯係數.

5. 若前面之計算用組距單位進行者,現今須將平均關係方程式及標準誤改用原來之計量單位表示.若曾用假定原點,則方程式應改正,使變數指對於眞正原點之差離而言.

II. 乘積率法

A. 材料不須分組者.

1. 將成對之觀察值列於平行之直行中,計算 $\Sigma(X)$, $\Sigma(Y)$, $\Sigma(X^2)$, $\Sigma(Y^2)$ $\Sigma(XY)$ 諸值.

2. 諸值各以 N 除之,最初之兩個商可以 c_x, c_y 之記號代表之.即

$$\frac{\Sigma(X)}{N}=c_x \quad 與 \quad \frac{\Sigma(Y)}{N}=c_y$$

3. 由下列公式計算平均乘積

$$p=\frac{\Sigma(XY)}{N}-c_xc_y$$

4. 由下列公式計算兩個標準差

$$\sigma_x=\sqrt{\frac{\Sigma(X^2)}{N}-c_x^2}$$

$$\sigma_y=\sqrt{\frac{\Sigma(Y^2)}{N}-c_y^2}$$

5. 由下列公式計算繫聯係數

$$r=\frac{p}{\sigma_x\sigma_y}$$

6. 將適當數值代入下列公式,求攀應方程式

$$y=r\frac{\sigma_y}{\sigma_x}x$$

$$x=r\frac{\sigma_x}{\sigma_y}y$$

(注意,兩方程式之原點皆在平均數點上.)

7. 必要時可將算術中數代入下列公式使原點移至兩個原來尺度之零點

$$Y-\overline{Y}=r\frac{\sigma_y}{\sigma_y}(X-\overline{X})$$

$$X-\overline{X}=r\frac{\sigma_x}{\sigma_y}(Y-\overline{Y})$$

8. 由下列公式計算兩個標準誤

$$S_y=\sigma_y\sqrt{1-r^2}$$

$$S_x=\sigma_x\sqrt{1-r^2}$$

B. 材料須分組者.

1. 照上面B.1.作一繫聯表.

2. 為每一變數選定一假定中數,各項對於假定中數之差離以組距為單位計算.

3. 計算c_x與c_y,以組距為單位.

4. 計算σ_x與σ_y,以組距為單位.

5. 計算繫聯表內每一格之$\Sigma(x'y')$,以組距為單位,將如此求得之數字相加,以求全表之$\Sigma(x'y')$.

6. 由下列公式計算以組距為單位之平均乘積之值

$$p=\frac{\Sigma(x'y')}{N}-c_xc_y$$

7. 由下列公式計算r

$$r=\frac{p}{\sigma_x\sigma_y}$$

8. 將σ_x與σ_y化為原來單位.

9. 將適當數值代入下列公式求響應方程式

$$y=r\frac{\sigma_y}{\sigma_x}x$$

與

$$x=r\frac{\sigma_x}{\sigma_y}y.$$

10. 必要時,用下列公式可將原點移至原來尺度上之零點,

$$Y-\overline{Y}=r\frac{\sigma_y}{\sigma_x}(X-\overline{X})$$

$$X-\overline{X}=r\frac{\sigma_x}{\sigma_y}(Y-\overline{Y})$$

11. 由下列公式計算兩標準誤

$$S_y=\sigma_y\sqrt{1-r^2}$$

$$S_x=\sigma_x\sqrt{1-r^2}.$$

在無論何種情形之下,作一散佈圖而繪響應線於其上,總爲有益.此種散佈圖,常能使吾人對於材料所包含之關係,對於應用之方法是否適當,較之單獨就數字研究時,得一更眞實之觀念.

[關係量之限制]

關於前面諸頁所討論之幾種關係量其普遍性爲如何,此問題自然發生.其應用僅限於某種形式之分配乎?抑可視爲十分普遍而到處適用之計量乎?

吾人已知標準差對於一正態之分配有確切而固定之意義.遇正態分配時,已知中數與標準差之值,則吾人可知在

全部觀察值之項數中,其值位於某一指定界限以內之項數所占之確切之百分數.若分配失去正態,標準差固仍有意義,惟在解釋時其意義不能如前者之確切耳.記住此點,可進而研究下列之公式

$$r^2=1-\frac{S_y^2}{\sigma_y^2}.$$

當應變數之原來諸值,在其中數周圍爲一正態分配,而在最小平方線周圍之分配亦爲正態時,S_y與σ_y皆有專門而確切之意義,根據此兩者之間之關係而計算一計量如r者,乃爲完全合理之事.兩種分配有一種失去正態時,S_y與σ_y之比較所表示之意義即因之減少.但吾人知雖當分配情形頗不合於正態之時,標準差所表示之意義不能如正態分配時所表示者同樣確切,然仍不失爲一有用之計量.同樣,雖當正態性之要件無一存在之時,標準誤與繫聯係數仍可計算與利用.惟遇此種情形在解釋上應加注意耳.吾人應充分認識此種計量,惟有在應變數之原來分配與最小平方線周圍之分配兩者皆爲正態或近於正態時,始表示完全之意義.

茲舉一簡單之例以說明一大異於正態分配之情形所加於繫聯係數數值之影響.下表選1919年美國工業調查關於紐約邦之若干數字.

表一百零二

1019年紐約邦十一個城市中之工廠工人數及生產總值

城名	工人數 (單位:千) (X)	生產總值 (單位:百萬元) (Y)
巴達維亞	2.2	9
皮棱	2.2	10
麻李	3.5	11
日尼瓦	2.5	10
格倫爾斯	2.8	12
依太卡	1.7	10
密特爾頓	2.2	10
拉克斯基勒	2.1	11
倫賽萊	1.4	10
土那宛達	1.8	16
紐約城	63.8	5,261

當僅取表上最前十個城市作爲一集團時,求得之數值如下:

$$\sigma_y=1.8682$$

$$S_y=1.8669$$

$$r=-.034$$

十個點與其響應線繪於圖 72.(上列繫聯係數無普遍之意義,諸城市乃專爲說明此問題而選者.)

當將紐約城加入此集團時,所求得之十一城市之數值爲:

$$\sigma_y=1509.3$$

$$S_y=7.53$$

$$r=+.999988$$

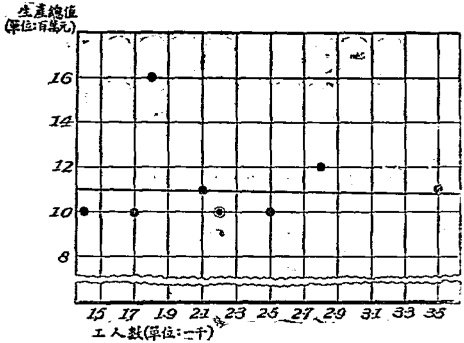


圖 72.——表示紐約邦中抽選之十個城市中工廠工人數與生產總值間之關係.

十一個點與響應線繪於圖 73.

結果之所以大相差異者,其理由甚爲明顯.當將一大城市加入十個小城市之集團時,兩變數之標準差之值皆因之大增,Y變數(生產物之價值)之標準差自 1,8682 增至 1509.3.惟計量配合線周圍之散佈狀態之S_y,其值未有如此突兀之變化.十個城市之標準誤爲 1.8669;十一個城市之標準誤爲 7,53.所以如此者,因一特例之存在,其大力足以使用最小平方法所作之直線,不得不通過或非常接近於代表此特例之一點.因此S所受一非常之特例之影響,常較σ所受者爲小,r之値

既係於下列之關係

$$r=\sqrt{1-\frac{S_y^2}{\sigma_y^2}},$$

故此種特例之存在,常能增加緊聯計量之價值.上述之例中,因一特例之加入,竟能使緊聯係數之數值簡直等於由零而變爲一.當然,此種結果實無意義.

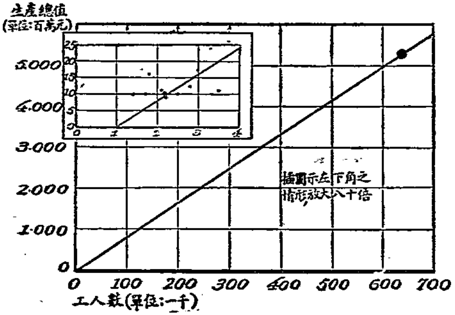


圖73.——表示紐約邦中揀選之十一個城市中工廠工人數與生產總值間之關係.

此例所代表之情形固爲極端,然無論何時凡分配非爲正態,則皆有此變相之存在,僅程度不同耳.在實際應用上各種關係計量不能限於完全正態之分配,惟若認爲有此種擾亂之影響存在時,解釋時須特別注意.

參考書

Bowley, Arthur L. Elements of Statistics (350—397).

Brunt, David. The Combination of Observations (148—170).

Chaddock, R. E. Principles and Methods of Statistics (Chap. XII).

Elderton, W. P. Frequency Curves and Correlation (106—124).

Galton, Francis, Correlations and their Measurement. Proceedings of Royal Society. Vol. XLV, 1888 (136—145).

Tones, D. C. A First Course in Statistics. (102—131).

Kelley, Truman L. Statistical Method (151—195).

Moore, H. L. Forecasting in the Yield and the Price of Cotton (12—51).

Pearl, Raymond. Medical Biometry and Statistics (292—318).

Pearson, Karl. Regression, Heredity and Panmissia. Phil. Transactions, Royal Society. Series A. Vol. CLXXXVII, 1896 (253—318).

Rietz, H. L. and Crathome, A. R. Simple Correlation (In Rietz, H. L. Handbook of Mathematical Statistics, 120—129).

Rugg, H. O. Statistical Methods applied to Education (233—307).

Whitaker, E. T. and Robinson G. The Calculus of Observations (317—336).

Yule, G. U. An Introduction to the Theory of Statistics (157—209).

第十一章

時間數列間關係之計量

前章所述計量繫聯諸方法,原爲分析非歷史性數列而作,卽非用於處理時間數列而用於處理頻數數列者也.計量時間數列間之繫聯時,有若干特殊之問題須予單獨論述者.

吾人已知此種數列爲種種力量所影響.此等力量可分爲長期趨勢,循環變動,季節變動與意外變動.前曾敍述分開此等力量的結果之方法.吾人藉之可以將整個之數列,分爲若干構成部分而予以各別之研究.欲決定時間數列間之繫聯,此種工作甚爲重要.苟非如此,所得結果將虛僞而易滋誤會.繫聯之問題,原爲求一數變量間關係程度之精確計量.但每一時間數列皆爲幾個變數結合而成.故將此種數列作成繫聯時,在可能範圍內,必須使各變數相互分開而各予以單獨之處理.

兩個時間數列間之關係(如利率與債劵價格間之關係),可就下列諸成分之一或全部而研究之:

a 長期趨勢

b 循環變動

c 季節變動

d 一時間單位與其次之一時間單位間之變動(如星期與星期間,月與月間,或年與年間之變動).

此種關係首先可將曲線互相比較而研究之.用此簡單之方法可獲得許多知識,長期趨勢之相似或不相似,循環變動間之一般關係,皆可由此決定.欲作更精確之比較,吾人可用繫聯係數;惟如此應用時特別重要者,即吾人對於如此應用之本質與所得結果之意義應有確切之認識也.

繫聯係數決不能用與長期趨勢之比較,兩數列間若僅有相同之長期趨勢存在,並不能表示兩者有互相依屬之關係.故根據長期趨勢值所作之繫聯係數將毫無意義.且欲比較長期趨勢,猶有較爲簡單之方法可用也.

根據同樣之理由,兩時間數列除非爲皆無長期趨勢存在之少見之特殊情形,否則繫聯係數不應以其原來之絕對數值爲根據.當吾人處理平常之統計數列時,計算r之法中,須計算兩變數之各項對於各自算術中數之差離,並須求配對之差離之乘積之總當和.相配者始終爲同正負號之差離時,r爲一正數;當相配者始終爲異正負號之差離時,r將爲一甚大之負數.時間數列中若有顯著之向上向下之長期趨勢在,即不能用此法求得r之有意義數值.例如計量1900—1920年間汽車生產量與醃肉價格之兩數列間之關係,兩數列之

長期趨勢皆急陡向上,當以兩數列之算術中數爲準而計算其每年數字之差離時,成對之項在前數年皆爲負數,而在後數年者皆爲正數.若在此基礎上進行計算,結果所得之 r 值將爲一甚大之正數.此數與人一十分錯誤之印象,蓋在此兩數列中實無何等關係可以想像也.此種緊聯係數所計量者固不過兩長期趨勢間之關係而已.

吾人固未嘗不可應用緊聯係數以決定兩數列季節變動之相似性.惟其效用如何尙成疑問;且亦有其他較爲簡單之方法可用也

故實際上緊聯法旣不宜用於計量長期趨勢間之關係,亦不宜用於計量季節變動間之關係.其應用之範圍,僅限於兩個或兩個以上數列之循環變動之比較與逐月逐年短期變動之比較.在作此種比較時,若欲求緊聯之有效的計量,凡不在研究範圍以內之其他力量之影響,必須儘可能予以淸除.未開始緊聯之正面工作以前,必須先經一番篩擠,除去一切不相關之材料.材料若未經如此之"蒸溜",則結果求得之緊聯係數,在解釋上將感困難.

[循環變動間緊聯之計量]

計量並除去作用於時間數列之諸因子中若干因子之影響之法,前面有一章業經討論.若計算觀察值之差離時不以兩數列之算術中數爲準而以兩條長期趨勢線爲準,則因長期趨勢之存在而生之僞緊聯即可以避免.若吾人以循環

變動爲注意之中心,則對於正則値之差離此種變動,實爲甚有意義之數值.若所用之材料爲每年之數字,即無清除季節變動之問題存在.

爲說明此計量時間數列間關係之方法起見,可以棉花產量與棉花價格之循環變動爲材料,而決定此兩者之間是否有何關係存在.所用數字爲1900—01年至1922—23年間一時期中各收穫年(crop year)之數字.

在着手研究繫聯以前,棉花價格尙須稍與校正作爲研

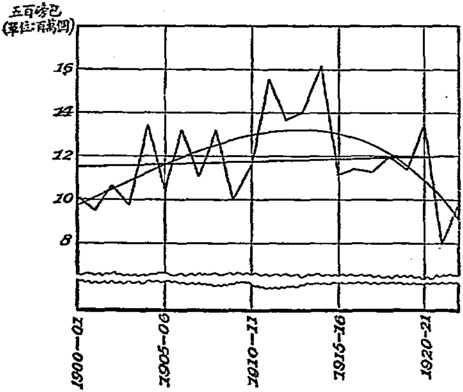


圖 74.— 1900—01 至1922—23 各收穫年之美國棉花生產量,及其長期趨勢線.

究出發點之原來數字,爲每一收穫年中由九月至五月中等內地棉在紐約市場上之平均現貨批發價格.但此種價格非但反映棉花市場上情況之變更,且反映一般物價水平之變動.爲消除此後一因子之影響計,先求每一收穫年中九月至五月一時期之勃拉特斯武利氏物價指數,用以緊縮原來之價格.勃拉特斯武利氏指數,以 1913—14 收穫年之平均數爲100,一律化成比値以應此須要.玆將原來兩數列之數字並緊縮以後之價格列於表一百零三.

表一百零三

1900—1923年棉花生產量及棉花價格

(1) 收穫年	(2) 美國棉花生產量 (單位:一千包)	(3) 每磅棉花價格 (單位:一分)	(4) 勃氏物價指數 (1913—14=100)	(5) 每磅棉花價格緊縮値 (單位:分)
1900—01	10,123	9.58	84.8	11.30
1901—02	9,510	8.64	86.2	10.02
1902—03	10,631	9.50	90.0	10.56
1903—04	9,851	13.20	88.6	14.90
1904—05	13,438	8.69	89.3	9.73
1905—06	10,575	11.40	92.3	12.35
1906—07	13,274	10.97	98.8	11.10
1907—08	11,107	11.41	93.2	12.24
1908—09	13,242	9.81	91.3	10.74
1909-10	10,005	14.62	100.6	14.53
1910—11	11,609	14.80	97.8	15.13
1911—12	15,693	10.34	100.0	10.34
1912—13	13,703	12.35	104.8	11.78
1913—14	14,156	13.40	100.0	13.40
1914—15	16,135	8.63	105.2	8.20
1915—16	11,192	12.04	121.2	9.93
1916—17	11,450	18.29	151.0	12.11
1917—18	11,302	29.96	197.9	15.14
1918—19	12,041	30.06	203.1	14.80
1919—20	11,421	38.63	226.3	17.07
1920-21	13,440	16.90	152.9	11.05
1921—22	7,954	18.67	127.2	14.68
1922—23	9,762	28.26	149.7	17.54

此材料繪成圖74及圖75.對於每一數列各配兩條長期趨勢線.

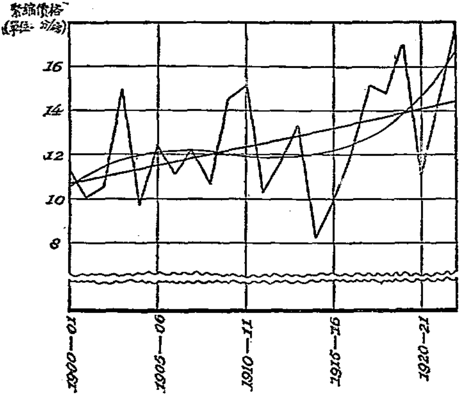


圖75.——1900-01至1922—23各收穫年中等內地棉在紐約市上之價格及其長期趨勢線.(數字代表各收穫年之每年平均價格經過用勃拉特斯武利氏批發物價指數平縮後之結果)

現今須以每一數列之長期趨勢爲準計算該數列每年數字之差離,而求兩數列之差離間之緊聯係數.茲先使對於兩條三次拋物線之差離相聯結.(若欲將兩數列相聯,必須其長期趨勢線爲同一形式).計算之法見於下表.

表一〇四

棉花產量與棉花價格間聯繫係數之計算法

(1) 年份	(2) 棉花產量對於長期趨勢拋物線之差離（單位—千包） x	(3) 棉花躉售價格對於長期趨勢拋物線之差離（單位—千包） y	(4) x^2	(5) y^2	(6) xy
1900—01	+371	+.79	137,641	.6241	+293.09
1901—02	−603	−1.05	363,609	1.1025	+633.15
1902—03	+145	−.93	21,025	.8649	−134.85
1903—04	−1,012	+3.11	1,024,144	9.6721	−3,147.32
1904—05	+2,200	−2.26	4,840,000	5.1076	−4,972.00
1905—06	−1,027	+.24	1,054,729	.0576	−246.48
1906—07	+1,325	−1.06	1,755,625	1.1236	−1,404.50
1907—08	−1,164	+.08	1,354,896	.0064	−93.12
1908—09	+681	−1.38	463,761	1.9044	−939.78
1909—10	−2,806	+2.47	7,873,636	6.1009	−6,930.82
1910—11	−1,405	+3.14	1,974,025	9.8596	−4,411.70
1911—12	+2,530	−1.60	6,400,900	2.5600	−4,048.00
1912—13	+452	−.13	204,304	.0169	−58.76

表一〇四（續上）

棉花產量與棉花價格間關聯系數之計算法

(1) 年份	(2) 棉花產量對於長期趨勢地物線之差離（單位一千包） x	(3) 棉花原棉價格對於長期趨勢地物線之差離（單位一千包） y	(4) x^2	(5) y^2	(6) xy
1913—14	+887	+1.47	786,769	2.1609	+1,303.89
1914—15	+2,923	−3.81	8,543,929	14.5161	−11,136.63
1915—16	−1,878	−2.24	3,526,884	5.0176	+4,206.72
1916—17	−1,388	−.30	1,926,544	.0900	+416.40
1917—18	−1,206	+2.37	1,454,436	5.6169	−2,858.22
1918—19	−31	+1.56	961	2.4336	−48.36
1919—20	−102	+3.21	10,404	10.3041	−327.42
1920—21	+2,586	−3.59	6,687,396	12.8881	−9,283.74
1921—22	−2,104	−.91	4,426,816	.8281	+1,914.64
1922—23	+636	+.82	404,496	.6724	+521.52
共計	0	0	55,236,930	93.5284	−40,752.29

$$\sigma_x=\sqrt{\frac{55,236,930}{23}}=1,550 \qquad \sigma_y=\sqrt{\frac{93.5284}{23}}=2.017$$

$$r=\frac{\Sigma(xy)}{N\sigma_x\sigma_y}=\frac{-40,752.29}{23\times1,550\times2.017}=-.567$$

係數之値爲 $-.567$, 即表示在所研究之一時期內,美國棉花生產量對於其長期趨勢之差離.與相當之棉花在紐約市場上之價格對於其長期趨勢之差離.兩者之間有相當程度之負緊聯(negative correlation)存在.若將大戰時期幾年之變態情形除去,求得之係數其値自必較大,惟由於某種原因不宜於如此處理.

根據已計算得之數值,吾人可以推算一估量方程式,估量當生產量發生一種變動時棉花價格所發生之變動.此響應方程式,前已指出,其式爲

$$y=r\frac{\sigma_y}{\sigma_x}x$$

在此地,y 與 x 所指者爲對於長期趨勢拋物線之差離,代入適當數值,吾人得

$$y=-.567\frac{2.017}{1,550}x$$

$$y=-.0007378\,x$$

此方程式表示平均在棉花生產量之差離(x)爲高於長期趨勢一單位時,棉花價格之差離(y)爲低於長期趨勢00.07單位.生產量之單位爲1000包,緊縮價格之單位爲一分.爲解釋上簡便起見,可以 1,000,000 包爲 x 之單位,而響應方程式成爲

$$y=-.7378\,x.$$

於是棉花生產量若超過正則生產量一百萬包,則價格(此地皆指緊縮後之價格)方面所發生之變動,爲低於正則價

格每磅的3/4分之數.此爲1900—1923年間平均關係.此種情形並非有試必驗,觀r之値爲$-.567$可知.此種法則或與此相似之法則,若爲無試不驗者,r之値將爲-1矣.

計量響應線周圍分散狀態之S_y可由下列公式求之

$$S_y=\sigma_y\sqrt{1-r^2}$$

此例中,S_y之値爲1.66分.此數之意義已在前文中說明.

(應請特別注意者,欲根據前面之方程式而估量將來之數字,須視兩條長期趨勢線能否合理延長於材料之外爲斷).

前面分析中,差離皆用絕對單位計算,而所得結果亦惟有用絕對單位始能解釋,如棉花若干包,每磅値價若干分.然有時宜連結兩數列對於其長期趨勢之百分數差離,遇此種情形,則求得之標準誤與響應方程式亦以百分數表示.兩法中究取何種,平常須視結果將用於何種方面而決定.

上面討論之問題中,顯然有一任意決定之因子在內,而爲以前研究緊聯問題時所未有者.差離之計算,不以中數爲準,而以長期趨勢線爲準;而長期趨勢線乃任意選定者也.應用不同之長期趨勢線,可得頗不相同之結果.上面例中所用之長期趨勢線爲兩條三次拋物線,但吾人幾有同樣充分之理由,假定兩數列中潛在之長期趨勢最宜以兩條直線表示之.嘗作成如此之直線,而將兩數列對於各自長期趨勢直線之差離相聯繫時,結果得緊聯係數之値爲$-.603$,S_y之値爲1.72分.(此等結果表面上似與將兩數列對於各自長期趨

勢拋物線之差離相聯繫時所得之結果相矛盾,因假定長期趨勢爲直線形式時,無論S_y或r皆爲一較大之數值也.其故在長期趨勢爲直線形式時,σ_y亦較大;而r之數值當然爲S_y與σ_y兩者間之關係所決定者.)

[決定時間數列之繫聯之困難]

現今有兩個繫聯係數,其中究以何者爲兩數列的循環變動間之關係之正確計量乎?不幸此問題不能確定答復,此爲研究時間數列之繫聯時所遇主要困難之一.假如在配合於每一數列之兩種曲線中,吾人能確斷何者爲長期趨勢之最正確之計量,此問題卽可解決;惜乎無一客觀之標準可以作最後之試驗.由表面上觀之,三次拋物線在材料之範圍以內爲兩數列長期趨勢較佳之計量;惟此點無從確證.能得出最大繫聯係數之長期趨勢線,決非卽爲最佳之長期趨勢線,因一數值大之係數可僅僅表示兩長期趨勢線皆配合不當,而兩者之錯誤皆趨於同一方向也.(註)

在對於長期趨勢之差離之繫聯中,既有此任意決定之因素存在,致其結果之可靠性稍形減色.求兩數列之循環變動間之繫聯係數時,若選定若干不同之長期趨勢線,則所得結果可以大不相同.此地中心問題不在作繫聯時機械過程,而在如何選定每一數列之適當長期趨勢線.若經過觀察以

(註) 參考潘森氏: "Correlation of Time Series," Journal of the American Statistical Association, June, 1923, 725.

及與外界事實相比較後,證明吾人所選定之曲線似能確切代表繫聯中每一數列之長期趨勢時,求得之係數即可認爲有效.惟在解釋與運用此結果時,仍不應忘卻在最初之計算中有此個人判斷之因素存也.吾人研究之目的,若爲在兩數列之循環變動間建立一函數關係,若爲根據求得之結果作成一估量(響應)方程式,則此點尤關重要.

[繫聯係數與時間順序之計量]

前面將棉花生產量與棉花價值作成繫聯時,吾人之目的,爲對於棉花生產量之變動所加於棉花價格之影響儘可能地作一確切估量;求得一方程式,代表當以應用之特殊長期趨勢線爲準而計量差離時在此兩數列間之關係.吾人視棉花價格爲棉花生產量之函數,而以計量此函數關係爲研究之目的.在此種情形之下,吾人假定兩數列間有函數關係存在,而欲決定一數列之循環依從或反映另一相關數列之循環之程度.此問題本質上即開始討論繫聯時所提者,當研究任何形式之數列間之關係時,常爲主要問題.

研究時間數列時,有時遇另一種問題,與此稍有不同.在兩個時間數列中假定固定之循環存在,兩者之循環在時間上究相合乎?抑一數列之循環經常在另一數列之循環之前或在其後乎?在決定此種"抄前"或"落後"之程度上,繫聯係數甚有用處,此種問題僅欲決定時間關係Temporal relationship),與平常計量函數關係之問題不同.

[股票價格循環與一般商情循環間之關係]

爲說明解決時間關係問題之方法起見,以工業股票價格之循環變動與美國電話電報公司所編製之綜合指數上所表示之一般商業活動之循環變動爲材料,決定此兩種循環在時間上之關係,1877—1923年一時期之綜合指數.其逐月之數字見表九十三,股票價格之數字列於下表:

表一〇五

1903—1923年美國工業股票價格之循環(註)

(表內數字爲實在值與正則值之差,以標準差爲單位.×交易所停市)

	1903	1904	1905	1906	1907	1908
正月	-.2	-2.0	-.1	+2.3	+1.6	-1.3
二月	-.1	-2.1	+.2	+2.2	+1.4	-1.5
三月	-.3	-2.1	+.6	+1.9	+.6	-1.1
四月	-.5	-2.0	+.7	+1.7	+.6	-.8
五月	-.5	-2.1	+.2	+1.4	+.4	-.5
六月	-.9	-2.1	+.3	+1.5	+.2	-.5
七月	-1.4	-1.8	+.7	+1.3	+.3	-.2
八月	-1.7	-1.6	+.8	+1.7	-.3	+.3
九月	-1.9	-1.3	+.7	+1.7	-.5	+.1
十月	-2.3	-.9	+.8	+1.7	-1.3	+.2
十一月	-2.4	-.3	+1.1	+1.7	-1.9	+.5
十二月	-2.1	-.1	+1.8	+1.7	-1.6	+.5

(註) 此表中數字爲潘森氏分析之結果,詳哈佛經濟研究委員會出版之 Review of Economic Statistics. 1903年正月至1914年七月一時期,根據12種工業股票之平均價格;1914年七月以後,根據20種工業股票之平均價格(Dow-Jones 指數).

表一〇五（續上）

1903—1923年美國工業股票價格之循環

（表內數字爲實在值與正則值之差，以標準差爲單位，×交易所停市）

	1909	1910	1911	1912	1913	1914
正　月	+.5	+1.0	-.1	-.4	-.2	-.7
二　月	+.3	+.5	+.1	-.4	-.5	-.6
三　月	+.3	+.8	-.1	-.1	-.7	-.6
四　月	+.5	+.5	-.1	+.3	-.6	-.8
五　月	+.8	+.4	0	+.2	-.7	-.8
六　月	+.9	+.1	+.1	+.2	-1.1	-.7
七　月	+1.1	-.4	+.1	+.2	-.9	-1.1
八　月	+1.4	-.3	-.3	+.3	-.7	×
九　月	+1.4	-.3	-.7	+.4	-.5	×
十　月	+1.4	0	-.7	+.3	-.8	×
十一月	+1.4	+.1	-.5	+.2	-.9	×
十二月	+1.3	-.2	-.4	0	-.9	-2.54
	1915	1916	1917	1918	1919	1920
正　月	-2.42	+.42	+.61	-1.00	-.56	+1.16
二　月	-2.47	+.33	+.15	-.68	-.53	+.40
三　月	-2.30	+.31	+.41	-.80	-.13	+.78
四　月	-1.69	+.05	+.56	-.86	+.14	+.83
五　月	-1.72	+.08	+.32	-.61	+.77	+.17
六　月	-1.54	+.11	+.59	-.63	+1.22	+.17
七　月	-1.28	-.04	+.29	-.59	+1.56	+.11
八　月	-.74	+.15	-.03	-.54	+1.01	-.27
九　月	-.27	+.62	-.41	-.50	+1.38	-.16
十　月	+.25	+.97	-.75	-.22	+1.86	-.31
十一月	+.40	+1.40	-1.32	-.35	+1.62	-.80
十二月	+.58	+.70	-1.41	-.50	+1.24	-1.28

表一〇五(續上)

1903—1923年美國工業股票價格之循環

(表內數字爲實在值與正則值之差,以標準差爲單位.×交易所停市)

	1921	1922	1923			
正 月	−1.04	−.61	+.66			
二 月	−1.01	−.37	+.95			
三 月	−1.02	−.13	+1.14			
四 月	−.92	+.19	+.93			
五 月	−.88	+.34	+.53			
六 月	−1.49	+.32	+.35			
七 月	−1.54	+.45	+.01			
八 月	−1.66	+.69	+.13			
九 月	−1.41	+.81	+.09			
十 月	−1.32	+.84	−.11			
十一月	−.97	+.51	+.14			
十二月	−.69	+.64	+.37			

兩數列之材料繪於圖76.將圖中兩曲線比較之,顯然可見在兩數列之變動間有某種關係存在,惟僅僅根據如此之比較,不能得確定之結論.現今吾人之目的在決定兩數列之循環在時間上是否相合;如不合,則一數列之循環平均先於另一數列之循環之時間爲幾何.在分析商情循環時,此種研究之意義甚爲顯著者也.

玆取1903年正月至1914年六月之材料作研究之資,省去大戰時期之幾年,因該幾年中兩數列皆大受變態情形之影響也.

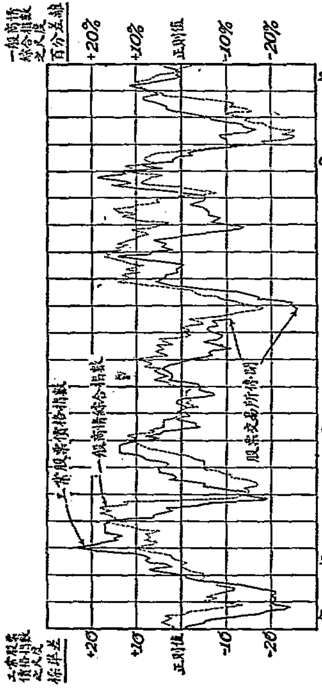


圖 70.——1903—23 年,工業股票價格指數與一般商情綜合指數之比較

先將兩數列中同時間各項彼此相聯求其繫聯係數,得值+.55.其次,將早一月之股票價格與各月之一般商情指數相聯,即以1903年正月之股票價格與1903年二月之一般商情指數相乘,以二月之股票價格與三月之一般商情指數相乘,諸如此類.自1903年正月至1914年六月之整個時期皆用此法處理.最先計算時,原有逐月數字138個,惟此地僅有137個,因1903年正月之商情指數與1914年六月之股票價格數字皆不在計算之列也.因此C_x與C_y之值(x與y之假定中數與實在中數之差)與兩個標準差將稍有不同.此種差異極易校正.結果計算得繫聯係數之值爲+.65.依照此法,再處理兩變數之其他配合形式.結果綜列之於下表,並給圖77示之.

表一〇六

工業股票價格與一般商情指數間繫聯係數
(根據1903—14年一時期之材料)

			繫聯係數
股票價格與商情指數同時			+.55
股票價格前於商情指數一月			+.65
同	上	二月	+.70
同	上	三月	+.73
同	上	四月	+.76
同	上	五月	+.76
同	上	六月	+.76
同	上	七月	+.74
同	上	八月	+.71
同	上	九月	+.67
同	上	十月	+.61
同	上	一月	+.54

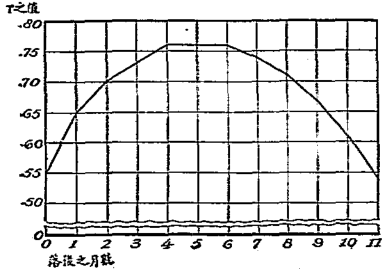


圖77.——工業股票價格指數與一般商情指數間緊聯係數(1903—14)，表示由不同的配對所得之結果.(各種配對中除兩者同時者外其餘皆爲一般商情指數落後於股票價格指數.)

緊聯係數之最大數値爲+.76,此數乃股票價格前於商情指數4,5與6月時所求得者.當股票價格抄前之時期爲3個月至6個月時,緊聯係數表現穩定之性質;此點卽表示在此兩個數列之循環變動之間,在此範圍以內時間上無確定之間隔.由所得之結果觀之,股票價格之循環變動前於一般商情指數之循環變動之平均時間似爲5月,但此非淸楚劃定如一固定之關係者.

在此處之硏究中,包括之時期太短,故不能得一最後之結論.結果表示一般商情指數之變動落後於股票價格之變動之時間在大戰以後較戰前爲短.股票價格先於商情指數三個月之時,緊聯係數最大.

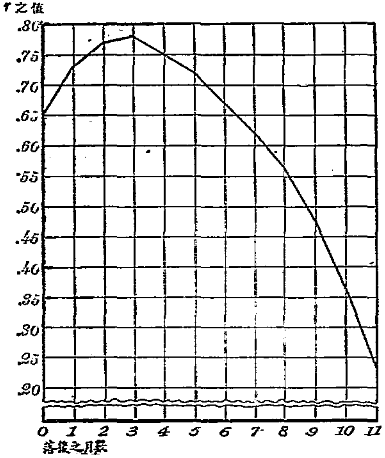


圖78.——工業股票價格指數與一般商情指數間聯繫係數(1919—23),表示由不同的配對所得之結果.(各種配對中除兩者同時者外其餘皆爲一般商情指數落後於股票價格指數.)

[用移動平均數聯繫時間數列之循環之法]

在前面討論中所處理之循環,僅限於從數學方法所配合之長期趨勢線計算而得者,惟吾人知計量長期趨勢之線亦可用移動平均數法求之.對於此種長期趨勢線之循環差離亦可作成聯繫,其法與處理從其他長期趨勢線所求得之

差離者完全相同.此種從移動平均數線計算而得之差離,其算術中數不一定爲零.此點爲與從最小平方法配合之曲線所求得之差離不同者;故在作成繫聯時,亦須作一相當之校正.

表一〇七

工業股票價格與一般商情指數間之繫聯係數

(根據1919—23年一時期之材料)

			繫聯係數
股票價格與商情指數同時			+.75
股票價格前於商情指數一月			+.83
同	上	二月	+.87
同	上	三月	+.88
同	上	四月	+.85
同	上	五月	+.82
同	上	六月	+.77
同	上	七月	+.72
同	上	八月	+.66
同	上	九月	+.57
同	上	十月	+.46
同	上	十一月	+.33

移動平均數之缺點與長期趨勢之數學曲線之缺點同.由根據移動平均數而作之長期趨勢線所求得之差離,不必僅僅表示循環變動之結果.每處之結果須視應用之移動平均數所包括之時期之長短而定;欲決定何者爲長期趨勢之最佳計量,並無完全之標準.循環差離由移動平均數計算時,固未嘗不可以求得有意義有用之係數,然吾人必須認清計算中有一任意決定之因素在,當應用其結果時,不得不加以相當保留.

[短期變動之繫聯]

前面敍述構成時間數列中各值之變動因子時,曾指出繫聯係數一般上不能用以比較兩數列之長期趨勢與季節變動.至於兩數列之循環變動間,其函數關係與時間關係,吾人若已儘可能除去其他變動之影響,則可以繫聯係數作其有效之表示.然繫聯係數及與之連帶應用之計量,對於處理時間數列上尚有一種用途,即用以計量兩時間數列短期變動間之關係,如逐年、逐月、甚至逐週、逐日之關係是也.此問題與上節所討論者相異,在解釋結果時,兩者不宜相混也.

表一〇八

1901—1923年棉花生產量及價格間繫聯係數計算表

(根據第一級差)

(1) 收穫年	(2) 某一年生產量與其前一年生產量之差(單位百萬包) X	(3) 某一年價格與其前一年價格之差(單位每磅一分,緊縮值) Y	(4) X^2	(5) Y^2	(6) XY
1901—02	−.613	−1.28	.375769	1,0384	+.78464
1902—03	+1.121	+.54	1,256641	.2916	+.60534
1903—04	−.780	+4.34	.608400	18,8356	−3,38520
1904—05	+3.587	−5.17	12,666569	26.7289	−18.54479
1905—06	−2.863	+2.62	8.196769	6.8644	−7.50106
1906—07	+2.699	−1.25	7.284601	1.5625	−3.37375
1907—08	−2.167	+1.14	4.695889	1.2996	−2.47038

表一〇八（續上）

1901—1923年棉花生產量及價格間緊聯係數計算表

（根據第一級差）

(1) 收穫年	(2) 某一年生產量與其前一年生產量之差（單位百萬包） X	(3) 某一年價格與其前一年價格之差（單位每磅一分，緊縮值）Y	(4) X^2	(5) Y^2	(6) XY
1908—09	+2.135	−1.50	4.558225	2.2500	−3.20250
1909—10	−3.237	+3.79	10.478169	14.3641	−12.26823
1910—11	+1.604	+.60	2.572816	.3600	+.96240
1911—12	+4.084	−4.79	16.679056	22.9441	−19.56236
1912—13	−1.990	+1.44	3.960100	2.0736	−2.86560
1913—14	+.453	+1.62	.205209	2.6244	+.73386
1914—15	+1.979	−5.20	3.916441	27.0400	−10.29080
1915—16	−4.943	+1.73	24.433249	2.9929	−8.55139
1916—17	+.258	+2.18	.066564	4.7524	+.56244
1917—18	−.148	+3.03	.021904	9.1809	−.44844
1918—19	+.739	−.34	.546121	.1156	−.25126
1919—20	−.620	+2.27	.384400	5.1529	−1.40740
1920—21	+2.019	−6.02	4.076361	36.2404	−12.15438
1921—22	−5.486	+3.63	30.096196	13.1769	−19.91418
1922—23	+1.808	+2.86	3.268864	8.1796	+5.17088
	−22.847	−25.55	140.548313	203.6683	−117.37216
	+22.486	+31.79			
	−.361	+6.24			

$$c_x=\frac{-.361}{22}=-.0164$$

$$c_x^2=.00026896$$

$$c_y=\frac{+6.24}{22}=+.284$$

$$c^2_y=.080656$$

$$\sigma_x=\sqrt{\frac{140.548313}{22}-.00027}$$

$$=\sqrt{6.38829}$$

$$\sigma_x=2.528$$

$$\sigma_y=\sqrt{\frac{208.6688}{22}-.08066}$$

$$\sigma_y=\sqrt{9.40429}$$

$$\sigma_y=3.067$$

$$p=\frac{\Sigma XY}{N}-c_xc_y$$

$$=\frac{-117.37216}{22}-(.284\times-.0164)$$

$$=-5.3304404$$

$$r=\frac{p}{\sigma_x\sigma_y}$$

$$r=\frac{-5.3304404}{7.753376}$$

$$=-.687$$

比較短期變動之法有數種.或將兩數列順次相繼各項間之差之絕對數作成繫聯,或用此差之百分數或比率作成繫聯.表一〇八說明計算棉花生產量與棉花價格之逐年絕對變動(第一級差)間之繫聯時所用之手續.(2),(3)兩行之數字乃從表一〇三中之原來數字求得者.

計算 r 之法,與前數例中用一假定原點以計算差離者相同,此處所用原點爲0,惟 X 與 Y 諸値之代數和皆非零.故有許多數値皆須經過校正.結果得 r 之値爲 $-.687$

用平常公式可計算得華應方程式與 S_y 之值如下:

$$y=.8335x$$

$$S_y=2.23\text{分}$$

將前列關於棉花諸例中求得之結果集而比較之,對於

總的繫聯問題可予若干有趣味之說明.事實上吾人所計量之繫聯屬於三種不同之事物,(即繫聯之存於對三次拋物線之差離間者,存於對長期趨勢直線之差離間者,以及存於棉花生產量與價格之逐年變動間者是也.)嚴格言之不能相比.惟吾人如欲估計與一定收穫量相伴之棉花價格,可在三種研究之結果中任取其一以爲根據.

	r	S_y
棉花生產量與價格之循環之繫聯（由三次拋物線計差離時）	−.567	1.66分
棉花生產量與價格之循環之繫聯（由長期趨勢直線計差離時）	−.603	1.72分
同一材料之逐年變動之繫聯	−.637	2.23分

最後一例之 r,其值大於前兩例者,而其標準誤亦較大.此顯然之矛盾,其理由前已言之.蓋逐年變動之標準差,亦較大於直線或拋物線長期趨勢周圍之標準差也.

初觀之,將對於三次拋物線之差離作繫聯,而以其結果爲根據,則估量之差誤似爲最小;以對於逐年變動之研究爲根據,則估量之差誤似爲最大,其實在第一例中暗中含有一假定,即不問代表生產量或代表價格之長期趨勢線,皆可以延至所研究之時期以外也.然此假定大非可靠者.故該例之估量標準誤未能將真正之機率表出.直線長期趨勢亦然.惟根據逐年變動而求得之計量中,則未含此種假定;雖然,當遇一特殊情形之年頭時,其困難將大於其他之二例.

參考書

Moore, H. L. Economic Cycles: Their Law and Cause.

Moore, H. L. Forecasting the yield and the Price of Cotton.

Moore, H. L. Generating Economic Cycles.

Persons, W. M. Correlation of Time Series. Journal of the American Statistical Association, June, 1923. (This article is also published in Rietz, H. L., Handbook of Mathematical Statistics 150—165).

Persons, W. M. Indices of Business Conditions. Review of Economic Statistics, Prel. Vol. I. 1919.

Persons, W. M.-The Variate Difference Correlation Method and Curve fitting. Quarterly Publications of the American Statistical Association, June, 1917.

Snow, E. C. Trade Forecasting and Prices (with discussion). Journal of the Royal Statistical Society, May, 1923 (332—398).

Stamp, J. C. The effect of Trade Fluctuations upon Profits (with discussion). Journal of the Royal Statistical Society, July, 1918(563—608).

Yule, G. U. On the Time Correlation Problem, with Especial

Reference to the Variate Difference Correlation Method (with discussion). Journal of the Royal Statistical Society, July, 1921 (497—537).

第十二章

關係之計量：非直線繫聯

在前兩章中所討論之事例，僅限於兩變數間關係之可用一直線描述者。繫聯係數 r，爲關係程度之計量，表示兩變數接近於一直線關係之程度。僅當直線能適當配合於代表 X 與 Y 各對數值之各點時，r 始有意義。

配合曲線於時間數列之問題，前文曾有說明。彼時曾遇許多事例，其長期趨勢非直線形式，而須以一高次曲線代表之者。在目前討論之問題中亦有同樣情形。兩變數間關係可爲直線所不能描述，可成爲一高次之繫聯。遇此種情形，則實際數字將與配合最佳之直線相差甚巨，而在其周圍散佈甚廣。r 之值將細小失實。若能配合一代表眞正關係之曲線，則散佈之範圍將大減，吾人乃可計量眞正之繫聯。表一〇九之材料即其一例也。

表一〇九

紫花苜蓿生產量與灌溉水量(註)

(1910—1915年加利福尼亞邦合縱斯地方研究之總結)

灌水深度(吋)	每畝產量噸數						平均數
	1910	1911	1912	1913	1914	1915	
0	3.85	5.94	5.52	2.75	2.89	2.35	3.88
12	4.78	7.52	6.51	4.31	5.83	4.84	5.63
18	×	×	7.02	5.69	8.02	6.46	6.80
24	6.00	8.38	8.32	6.89	9.96	7.96	7.92
30	7.53	9.54	9.43	7.97	11.06	8.32	8.98
36	7.58	9.33	9.38	8.22	12.48	8.63	9.27
48	8.45	9.52	8.63	8.83	10.62	8.05	9.02
60	×	×	10.17	7.25	10.70	5.55	8.42

×材料缺

圖79爲此材料之圖.

配合兩不同之曲線於圖中諸點,其一爲直線,其方程式爲

$$Y=5.038+.0886X$$

Y代表每畝產量噸數,X代表灌水之深度,以吋爲單位.兩變數之關係,如此線所描述者,其程度以緊聯係數r表出之,r之值爲+.68.

(註) 此表錄自貝開脫(S. H. Beckett)氏與洛布生(R. D. Robertson)氏"The Economical Irrigation of Alfalfa in Sacramento Valley"一文,載加利福尼亞大學農事試驗站公報280號.1917年五月.

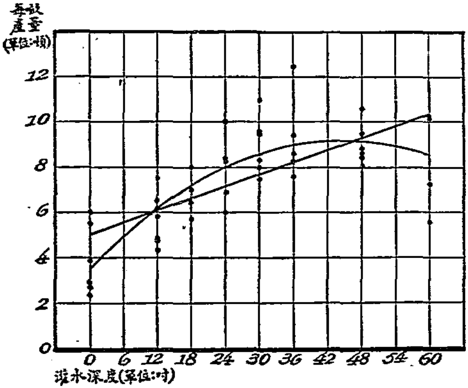


圖 79.——散布圖：表示紫花苜蓿產量及灌漑水量間之關係，及兩條迴歸線.

細觀此圖，可知此直線非最能適合於材料者.故 r 必非紫花苜蓿生產量與灌水深度間關係程度之眞確計量.

[拋物線式關係]

圖上另一曲線爲二次拋物線，乃用最小平方法求得者.此曲線之方程式爲

$$Y=3.55+.252X-.002816X^2$$

灌水量之增加對於紫花苜蓿產量之影響,用此二次拋物線表示時顯然較用直線表示時正確甚多.此種研究中最重要之結果,爲決定因灌水量過多而紫花苜蓿產量反見減少之一點,而關於此種減少,乃直線所不能作任何表示者也.

表一一〇

紫花苜蓿生產量之實在值與計算值之比較

(1) 灌水深度 X	(2) 實在產量 Y	(3) 由拋物線方程式計算之正則產量 Y_c	(4) 實在值與正則值之差 (2)—(3) d	(5) d^2
0	3.85	3.55	+.30	.0900
0	5.94	3.55	+2.39	5.7121
0	5.52	3.55	+1.97	3.8809
0	2.75	3.55	−.80	.6100
0	2.89	3.55	−.66	.4356
0	2.35	3.55	−1.20	1.4400
12	4.78	6.16	−1.38	1.9044
12	7.52	6.16	+1.36	1.8496
12	6.51	6.16	+.35	.1225
12	4.31	6.16	−1.85	3.4225
12	5.83	6.16	−.33	.1089
12	4.84	6.16	−1.32	1.7424
18	7.02	7.17	−.15	.0225
18	5.69	7.17	−1.48	2.1904
18	8.02	7.17	+.85	.7225
18	6.46	7.17	−.71	.5041
24	6.00	7.97	−1.97	3.8809
24	8.38	7.97	+.41	.1681
24	8.32	7.97	+.35	.1225
24	6.89	7.97	−1.03	1.1664

表一一〇(續上)

紫花苜蓿生產量之實在值與計算值之比較

(1) 灌水深度 X	(2) 實在產量 Y	(3) 由拋物線方程式計算之正則產量 Y_c	(4) 實在值與正則值之差 (2)—(3) d	(5) d^2
24	9.96	7.97	+1.99	3.9601
24	7.96	7.97	−.01	.0001
30	9.53	8.57	−1.04	1.0816
30	9.54	8.57	+.97	.9409
30	7.43	8.57	+.86	.7396
30	7.97	8.57	−.60	.3600
30	11.06	8.57	+2.49	6.2001
30	8.32	8.57	−.25	.0625
36	7.58	8.97	−1.39	1.9321
36	9.33	8.97	+.36	.1296
36	9.38	8.97	+.41	.1681
36	8.22	8.97	−.75	.5625
36	12.48	8.97	+3.51	12.3201
36	8.63	8.97	−.34	.1156
48	8.45	9.15	−.70	.4900
48	9.52	9.15	+.37	.1369
48	8.63	9.15	−.52	.2704
48	8.83	9.15	−.32	.1024
48	10.62	9.15	+1.47	2.1609
48	8.05	9.15	−1.10	1.2100
60	10.17	8.53	+1.64	2.6896
60	7.25	8.53	−1.28	1.6384
60	10.70	8.53	+2.17	4.7089
60	5.55	8.53	−2.98	8.8804
			+24.22	80.9871
			−24.21	

故吾人應取拋物線式者作爲關係方程式,而不取直線形式者.標準誤S_y,爲必須與方程式相伴之計量,計算時可先求各實在值對於計算值之差離,而將諸差離之平方之算術中數開方,求得之根卽爲S_y之值.其計算之步驟可於表一一〇見之.表中正則產量乃根據上列拋物線方程式計算而得者.

將表上求得之差離平方之總和代入公式

$$S_y=\sqrt{\frac{\Sigma d^2}{N}},\text{ 得}$$

$$S_y=\sqrt{\frac{80.9871}{44}}=1.36$$

[繫聯指數]

吾人現今尙須求一第三種之數値,卽關係程度之抽象計量是也.前章處理直線關係之事例時,知此種計量(繫聯係數,)可由S_y與σ_y之已知值求之.遇事例之爲非直線關係者,有如目下之問題,吾人亦可用同樣之方法求得一類似之計量.繫聯係數之一名詞與其記號r,旣專指直線響應之事例,吾人可稱此普遍之計量爲繫聯指數(Index of correlation)而以ρ(Rho)爲記號以代表之.(註)

(註) 斯比曼(Spearman)氏曾用此記號代表用等級差異法 (Method of Rank-differences)求得之繫聯係數,但此種用法在乎常經濟統計上無甚顯著之意義,此記號用於現今之目的不致引起誤會也.

繫聯指數之普通公式為:(註)

$$\rho^2_{yx}=1-\frac{S_y^2}{\sigma_y^2}$$

S_y之值前已求得,σ_y之值用業已熟知之方法求之,得2.27.代入ρ公式得:

$$\rho_{yx}=\sqrt{1-\frac{1.8406}{5.177}}$$
$$=.80$$

此值大於同材料之繫聯係數甚多,蓋r之值為+.68也.此差異之原因,由於二次拋物線對於材料比較直線大為適合,繫聯顯非直線形式,r乃其不適當之計量也.

[繫聯指數之意義]

吾人必須瞭解ρ之意義及其限度.ρ之值決定於配合曲線周圍之散佈狀態.與Y諸值算術中數周圍之散佈狀態間之關係.配合之線為直線時,ρ與r相合為一,r蓋ρ之特殊情形也.ρ之限度為0與1.其值為0時,表示兩變數間若有關係存在者,此關係不能以所應用之特殊方程式描述.其值為1

(註) 以X為應變數,則公式為

$$\rho^2_{xy}=1-\frac{S_x^2}{\sigma_x^2}$$

記號後之兩小字母,其一個經常指應變數,其第二個經常指自變數,此點必須註明,因ρ之值,當X為應變數時不一定與Y為應變數時相同也.在直線繫聯中,不論X或Y為應變數,r之值總相同,故不必有此種分別.

時,表示所應用之特殊方程式其描述之關係爲一完全者.關係爲高次曲線時,不應加正負號於ρ之前,因關係在全域內可有一段爲正,一段爲負,有如上面紫花苜蓿之例所示者也.

緊聯指數ρ所指之曲線形式,若不明白指出,則ρ將毫無意義.惟在此一方面,r之意義則經常明顯者,因其永指直線而言也.ρ則不然,若不特別指出曲線之形式,必致意義混亂.故緊聯指數可視爲以某一曲線形式描述兩變數間關係之適當性計量(Measure of adequacy).

當然,方程式中常數之個數若與點數相等,則不論有若干點數,吾人皆可作一曲線一一通過之.在此種情形下ρ之值必等於1,惟此値毫無意義.當常數之個數與點數相等時,數學函數之任何應用皆有此不合理性在,ρ之爲1,僅此不合理性之表現而已.當吾人應用此類指數時,配合曲線之平常原則應時留腦際.此等指數不應認作有絕對而獨立之意義,蓋其意義永爲相對者,永指所用之特殊函數而言也.緊聯之一切計量莫不如此,此事常爲人所忽略,結果致得謬誤失實之結論.

[計算緊聯指數之簡法]

在前例中,爲使吾人標準誤及緊聯指數之意義能得一確切之理解起見,故計算此兩個計量時所用之方法頗爲繁冗.惟此種計算之繁冗,若吾人利用在處理r時所發見之關係,即可使其大爲減少.當曲線爲$Y=a+bX+cX^2+dX^3\cdots$式之

定冪級數時，S_y之公式僅須將當關係方程式爲一直線時所用以計算S_y之公式稍加推廣卽得.曲線爲定冪級數式時S_y之普遍公式如下：

$$S_y=\frac{\Sigma(Y^2)-a\Sigma(Y)-b\Sigma(XY)-c\Sigma(X^2Y)-d\Sigma(X^3Y)-\cdots\cdots}{N}$$

同樣，以前求 r 之公式亦可推廣成求ρ 之普遍公式，而適合於此種形式之一切方程式.此公式爲(註)

$$\rho^2_{yx}=\frac{a\Sigma(Y)+b\Sigma(XY)+c\Sigma(X^2Y)+d\Sigma(X^3Y)+\cdots\cdots Nc_y^2}{\Sigma(y^2)}$$

吾人應注意S 與ρ兩公式之特質.當計算此兩計量時，所須之數值僅爲平均關係方程式中諸常數.配合曲線時已用過之若干數值.以及$\Sigma(Y^2)$ 與c^2_y.故S與ρ之數值可視爲配合曲線之直接副產；此兩數值補襯應方程式之不足；對於吾人所研究之兩變數間之關係，欲作一完全之敍述時，此兩計量乃不可或缺者也.襯應方程式之作用爲敍述平均關係，標準誤 S 爲計量根據此方程式而作之估量之可靠性，而在關係可以所用之特殊曲線描述之範圍內，ρ 爲關係程度之抽象指數.

以上幾個公式之應用，可以紫花苜蓿產量問題之材料說明之.下列諸值，或由表一〇九之材料中求得，或配合曲線之過程中求得者，合於現今之需要：

(註) 此公式之證法見附錄一.

$a=3.5468$　　　　$\Sigma(X^2Y)=407,448.00$

$b=.2520$　　　　$c_y^2=55.9504$

$c=-.0028162$

$\Sigma(Y)=329.03$　　　　$\Sigma(Y^2)=2688.3129$

$\Sigma(XY)=10,269.96$　　　　$N=44$

代入求二次拋物線標準誤之公式

$$S_y^2=\frac{\Sigma(Y^2)-a\Sigma(Y)-b\Sigma(XY)-c\Sigma(X^2Y)}{N}$$

則得

$$S_y^2=\frac{2688.3129-(3.5468\times329.03)-(.2520\times10,269.98)-(-.0028182\times407,448)}{44}$$

$$=\frac{80.7345}{44}$$

$$=1.8349$$

$$S_y=1.36$$

二次拋物線式繫聯指數可由下列方程式計算之

$$\rho^2_{yx}=\frac{a\Sigma(Y)+b\Sigma(XY)+c\Sigma(X^2Y)-Nc^2_y}{\Sigma(Y)^2-Nc_y^2}$$

代入適當之數值

$$\rho^2_{yz}=\frac{145.7608}{2688.3129-(44\times55.9504)}$$

$$\rho_{yz}=.80$$

此種求 S 與 ρ 之方法，適用之範圍甚廣，僅須將公式少加改動，以適應各處所應用之特殊方程式耳．下章對此點尚有說明，而此一般方法之說明詳見於附錄一．

[繫聯比]

吾人於此尚須敍述繫聯計量之第三種特殊形式，此即潘莾氏所首倡之繫聯比 (correlation ratio) 是也，其記號為 η (Eta)，此種計量可視為 ρ 之特殊情形，惟計算時所用之方法略異．

吾人知無論何處，兩變數間為一定形式曲線所描述之關係，其程度皆可以下列公式定之：

$$繫聯之計量=\sqrt{1-\frac{S_y^2}{\sigma_y^2}}$$

當 S_y 代表一直線周圍之標準差時，繫聯係數 r 恰為此種計量，繫聯指數 ρ 為同樣性質之計量，惟為一普遍之形式．繫聯比亦為完全同樣性質之計量，僅其 S_y 所代表之標準差，以通過繫聯表各行中數之一線為中心．因此代表配合曲線之方程式中，其常數之個數直增加至與繫聯表上之行數相等．若所有各行之中點皆在一直線上，繫聯比與繫聯係數將相等．若各行中點不在一直線上，繫聯比將大於繫聯係數．

因此在繫聯比之一觀念中,並未包括任何新原理被用於非直線響應;在兩變數間之關係可用一通過各行中數之曲線描述之範圍以內,繫聯比爲關係程度之計量.若爲一完全之關係,若已知數值無一離開如此作成之曲線者,η之值將爲1.若兩變數間無關係存在,若在曲線周圍散佈之程度與Y諸值之中數周圍散佈之程度相等,η之值將爲0.

平常用以求繫聯比之公式,與前列之公式少許不同.在通過各行中數之曲線周圍之標準差,不用S_y爲記號,而用σ_{cy}爲記號,σ_{cy}之意義,除經常指一繫聯表而言外,與以前所用之S_y之意義毫無出入.

公式可寫爲

$$\eta_{yx}=\sqrt{1-\frac{\sigma_{cy}^2}{\sigma_y^2}}$$

當Eta寫成(η_{yx})時,指Y對X之響應(Y爲應變數).當寫作(η_{xy})時,指X對Y之響應(X爲應變數),則其值爲在通過繫聯表各列中數之曲線周圍之離中趨勢所決定.η_{yx}與η_{xy},與r不同,r對於兩種響應其值不變,而η_{yx}與η_{xy},若響應非爲直線形式,則其值不同.

[繫聯比之計算]

下面之繫聯表,表示農事試驗中所下氮素肥料之量(每畝若干磅)與相當之小麥產量(每畝若干蒲)間之一般關係.各點繪於圖80.

表一一一(註)

表示每畝小麥產量及所下氮素肥料量間之關係之緊聯表

Y—每畝產小麥之蒲數 \ X—每畝施用氮素肥料磅數	0—19.9	20—39.9	40—59.9	60—79.9	80—99.9	100—119.9	120—139.9	140—159.9	160—179.9	共計	各列中數
32—35.9				5	16	12	4	5	2	44	107.27
28—31.9			1	20	21	8	4	1		55	88.91
24—27.9			16	19						35	60.86
20—23.9			13							13	50.0
16—19.9		12								12	30.0
12—15.9		18								8	30.0
8—11.9	3	5								8	22.50
4—7.9	10									10	1.00
0—3.9	8									8	1.00
共計	21	25	30	44	37	20	8	6	2	193	
各行中數	5.05	15.12	24.4	28.73	31.73	32.4	32.0	33.33	34.0		

(註) 本表根據達文波段(E. Davenport)氏所記述之實驗而作(錄自裴雷(Bailey)氏美國農業百科全書"比較農業"條,)表中實在數字係專為此地之說明而任意選定者。惟達文波段氏實驗中表示有與現今之假定相同之法則存在。

欲根據前述之公式計算 η_{yx},吾人須有 σ_y 及 σ_{ay} 之值,此後者即在通過各行中數之曲線周圍之標準差.前者之值其求法吾人已熟知之矣.而 σ_{ay} 之值,可用最初用於計算 S_y 之方法計算之,即求各點對於響應線之差離而平方之是也.惟在現今之情形中,描述關係之曲線為通過每一行之中數者,故各行中數可以代替由響應方程式計算而得之"正則"值.計算 σ_{ay} 時,先求各項對於其所屬之一行之中數之差離而

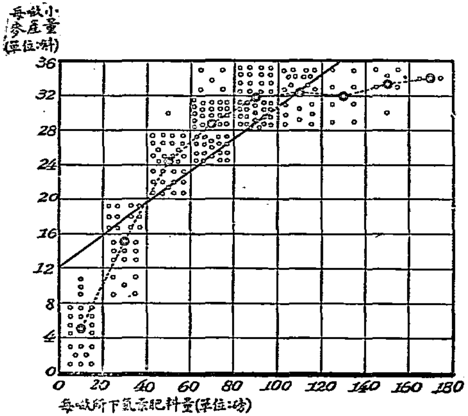


圖80.—散佈圖表示小麥產量及所下氮素肥料量間之關係,及直線式響應線與聯結各行中數之線.

平方之諸差離平方相加，求平方之中數，開方求方根，即得其値；其法正與計算標準差之法同。今用表一一一第一行中之材料，以說明計算手續之一部分。此行包括X値在0與20之間之各項。屬於此行之二十一項其Y値之中數爲50.5；差離乃以此爲中心而計算者。

表一一二

以一序列之中數爲準計算差離平方之法

組距			對本行中數(5.05.)之差離		
(每畝生產小麥蒲數)	m	f	d	d^2	fd^2
8—11.9	10	3	4.95	24.5025	73.5075
4—7.9	6	10	.95	.9025	9.0250
0—3.9	2	8	—3.05	9.3025	74.4200
共計					156.9525

其他每一行之差離平方總和亦用同樣之方法求得。結果得各行中數周圍之標準差，σ_{cy}，爲2.420。而σ_y之値爲9.188。

將已知値代入公式

$$\eta^2_{yx}=1-\frac{\sigma_{cy}^2}{\sigma_y^2}$$

$$\eta^2_{yx}=1-\frac{(2.42)^2}{(9.188)^2}$$

$$=1-.0694$$

$$=.9306$$

$$\eta_{yx}=.965$$

此值爲緊聯比之值,計量在通過各行中數之線周圍之散佈程度.其意義將於後文討論之.

前例所用之計算方法尙可大事減省.設σ_{my}代表各行中數對於全部Y值之算術中數之標準差.計算此值時每一行之中數以該行所包括之項數爲權數而加權之.吾人可以證明(註)

$$\sigma_{ay}^2=\sigma_y^2-\sigma_{my}^2$$

代入方程式

$$\eta^2_{yx}=1-\frac{\sigma_{ay}^2}{\sigma_y^2}$$

得

$$\eta^2_{yx}=1-\left(\frac{\sigma_y^2-\sigma_{my}^2}{\sigma_y^2}\right)$$

$$=\frac{\sigma_{my}^2}{\sigma_y^2}$$

$$\eta_{yx}=\frac{\sigma_{my}}{\sigma_y}$$

(註) 下列證明乃自游爾氏(Yule)書中轉錄者.

設有一數列,其中數爲M,此數列爲兩個小數列所合成,兩小數列之中數爲M_1及M_2.N爲觀察值之總項數,N_1及N_2各爲構成此數列之兩數列中所包含之觀察值之項數,$N_1+N_2=N$;則σ,σ_1及σ_2三者之間有何關係?設

$$M_1-M_2=c_1$$

第一個小數列中各觀察值對於原點M之差離之平方之中數，以S_1^2代表之，則

$$S_1^2=\sigma_1^2+c_1^2$$

同樣

$$S_2^2=\sigma_2^2+c_2^2$$

但$N_1S_1^2$等於第一數列中各項以M爲原點計算而得之差離之平方總和，$N_2S_2^2$等於第二小數列中各項以M爲原點計算而得之差離平方總和，故：

$$\sigma^2=\frac{N_1S_1^2+N_2S_2^2}{N}$$

而
$$N\sigma^2=N_1S_1^2+N_2S_2^2 \tag{1}$$

但
$$S_1^2=\sigma_1^2+c_1^2\ ,\quad S_2^2=\sigma_2^2+c_2^2$$

故
$$N\sigma^2=N_1(\sigma_1^2+c_1^2)+N(\sigma_2^2+c_2^2) \tag{2}$$

在現今之例中，大數列之中數以M_y爲代表，而構成此大數列之小數列(分列爲各行之諸項)之中數各以m_{y1}……等代表之。設以S_{cy}代表Y值之任何一行中各項數值對於該行中數之標準差，則各行之小數列其標準差爲S_{cy1}……等等，而其中數與全體Y值之中數之差爲M_y-m_{y1}等等。代入上面求得之方程式(2)，得

$$N\sigma_y^2=n_1[S_{cy1}^2+(M_y-m_{y1})^2]+n_2[S_{cy2}^2+(M_y-m_{y2})^2]+\cdots\cdots \tag{3}$$

$$N\sigma_y^2=\Sigma n[S_{cy}^2+(M_y-m_y)^2] \tag{4}$$

但
$$N\sigma_{cy}^2=\Sigma(n\cdot S_{cy}^2)$$

因每一行中
$$S_{cy}^2=\frac{\Sigma d^2}{n}$$

(d代表以該行中數爲原點之一個差離)。於是各行皆爲

$$\sigma_{cy}^2=\frac{\Sigma d^2}{N}=\frac{\Sigma(n\cdot S_{cy}^2)}{N}$$

將此式代入方程式(4)

$$N\sigma_y^2=N\sigma_{cy}^2+\Sigma n(M_y-m_y)^2 \tag{5}$$

由各行中數之標準差之定義，

$$\sigma_{my}^2=\frac{\Sigma n(M_y-m_y)^2}{N}$$

故由(5)式可得 $$\sigma_y^2=\sigma_{ay}^2+\sigma_{my}^2 \tag{6}$$

因決定σ_{my}較之決定σ_{ay}容易甚多，η之值平常皆由此式計算之.茲可取表一一一之材料以說明此方法.

表一一三

兼聯比之計算法

X各組中點(磅)	Y各項之算術中數(磅) m_y	對於全體Y值之中數(25.005)之差離 d	d^2	f	fd^2
10	5.05	−19.955	398.202	21	8,362.242
30	15.12	−9.885	97.713	25	2,442.825
50	24.40	−.605	.366	30	10.980
70	28.73	+3.725	13.876	44	610.544
90	31.73	+6.725	45.226	37	1,673.362
110	32.40	+7.395	54.686	20	1,093.720
130	32.00	+6.995	48.930	8	391.440
150	33.33	+8.325	69.306	6	415.836
170	34.00	+8.995	80.910	2	161.820
共計				193	15,162.769

$$\sigma_{my}=\sqrt{\frac{15,162.769}{193}}$$

$$=8.864$$

將已知值代入公式

$$\eta_{yx}=\frac{\sigma_{my}}{\sigma_y}$$

得

$$\eta_{yx}=\frac{8.864}{9.188}$$

$$=.965$$

計算繫聯比之步驟可以簡約綜述如之下:

1. 將所有各項列成繫聯表形式.

2. 就各行之Y數值而求其算術中數.

3. 計算Y數列全體之算術中數

4. 求各行算術中數對於數列全體之算術中數之差離,求各差離之平方,乘以各該行所包括之項數,而求其總和.

5. 求得之總和以總項數除之,再開平方卽得σ_{my}.

6. 計算σ_y.

7. 以σ_y除σ_{my},得η_{yx}.

用相同之方法,可求X對Y之繫聯比之值;將適當數值代入公式

$$\eta_{xy}=\frac{\sigma_{mx}}{\sigma_x}$$

σ_{mx}一記號代表各列之中數對於X數列全體之中數之標準差.X對Y之繫聯比之值,其大小爲各項對於通過各列中點之線之(水平的)差量所決定.其值平常皆與Y對X之繫

聯比之值不同.本例η_{xy}為.824,代表兩數列間關係之線越近於直線形式,則兩個繫聯比越接近於相等.

η與r同,其值決不能超過1,其值為1時,代表觀察值各項之各點無一不在通過各行(或各列)中點之線上.更由一

$$\eta_{yx}=\frac{\sigma_{my}}{\sigma_y}$$

公式,顯然可見當σ_{my}為0時,繫聯比之值亦為0.此乃每行之中數其值與Y數列全體之中數之值相等時之情形.若X變數之數值增減時Y變數之數值不起相當之變動,即有此種情形存在.此時觀察值各項在繫聯表各行中之分配與Y值全體之分配相同,而兩變數間顯無任何關係存在也.

繫聯比從無一負值.惟兩變數間之關係,其為正,為負,或為正負交錯,吾人一觀繫聯表即可知之.

以繫聯係數與繫聯比相比較時,前者有一特殊之優點,即當其值與兩個標準差之值業已決定時,響應線之方程式極易求得.η則不然,已計算得η之值時,再求表示兩變數間關係之"法則"之數學式,必須更作一番計算,以配合一曲線於各行或各列之中數.

[繫聯比之校正]

η僅能應用於材料之項數甚多而能列為一繫聯表之時.若如此排列之項數有限,而恰於每一行中分得一項,則σ_{my}及σ_y之值將相等,而η必然為1,根據之項數極少而應用

之組數極多,則所求得之緊聯比將毫無意義.

潘蓀氏創一校正法,校正初步求得之緊聯比,以特别消去因分組過多而生之錯誤.若以κ代表行(或列)之數,吾人所求之值可由下列公式計算而得:

$$\text{校正}\,\eta^2=\frac{\eta^2-\frac{(\kappa-3)}{N}}{1-\frac{(\kappa-3)}{N}}$$

將此式用於校正前例所得之結果,得

$$\text{校正}\,\eta^2=\frac{(.965)^2-\frac{(9-3)}{193}}{1-\frac{(9-3)}{193}}$$

$$=.929$$

$$\text{校正}\,\eta=.964$$

此例校正之數甚微.惟遇N小而κ甚大之時,此種校正將大減η之值.

[緊聯比與緊聯係數間之關係]

當兩變數間之關係爲絕對之直線式時,則通過各項中數點之線,當然與緊聯係數所根據之線相合,在此種情形下,η與r之值相等.當兩變數間之關係非直線形式時,求得之η值與r值不等,η經常大於r.蓋除非一直線能通過所有諸中數點,否則在通過各項中數點之一線周圍之差離,總較在配

合於諸點之直線周圍之差離爲小.而在代表平均關係之線之周圍,其差離愈小,則繫聯計量之値愈大.如在紫花苜蓿之問題中,r之値爲+.68,而根據二次拋物線求得之繫聯指數,其値爲.80.同材料之繫聯比爲.83.就表一一一之材料言,η_{yx}之値爲.964;r之値爲+.793,兩者相差甚大.所以相差之原因可由圖80見之,圖中一爲X對Y之響應直線,一爲通過各行中數點之線.響應關係大異於直線形式;響應直線周圍之差離,大於在通過各行中數點之線之周圍之差離甚多.

r與η間之關係,爲測定一例之響應關係之直線性(Linearity)之便利工具,因響應爲嚴格之直線式時,兩計量之値相等;而兩者相差之程度,隨響應異於直線式之程度而俱增.直線性之一般測驗式爲

$$\zeta=\eta^2-r^2$$

卽在直線式響應之例中,η與r之値亦可因偶然之變動而有差異.惟ζ(Zeta)若反映巨大之差異時,卽表示所研究之關係不能用一直線敍述,而r非一繫聯之適當計量.

上面之例中,η等於.963,r等於.795,ζ之値爲.300.觀乎此數之大,吾人已可充分確斷響應非直線形式.將來在第十六章中尚須說明如何決定ζ之一定數値之意義之法.

參考書

Bowley, A. L. Elements of Statistics (365—367).

Kelley, Truman L. Statistical Method (238—245).

Pearl, Raymond. Medical Biometry and Statistics (311—318).

Pearson, Karl. Mathematical Contributions to the Theory of Evolution (XIV).

On the General Theory of Skew Correlation and Non-linear Regression. Draper's Company Research Memoirs, Biometric Series II. 1905.

Notes on the History of Correlation. Biometrika, Vol. 13, 1920 (25—45).

On a Correction Needful in the Case of the Correlation Ratio. Biometrika, Vol. 8, 1911 (254—256).

On the Correction Necessary for the Correlation Ratio. Biometrika, Vol. 14, 1923 (412—417).

Rietz, H. L. (editor) Handbook of Mathematical Statistics (129—131).

Yule, G. U. An Introduction to the Theory of Statistics (204—207).

第十三章

關係之計量與估量問題

經濟學及實業方面之數量方法,謂其集中於估量問題,誠非過當.響應方程式、標準誤及繫聯係數,其所以重要者,大抵由於其與決定或然產量(probable production)、或然價格、或然商情變動等實際問題有關.此非謂估量問題僅僅與預測將來變動之嘗試有關係也.當吾人根據若干不同之觀察值以決定一最有可能之值時,或當吾人應用描述兩個或幾個變數間之關係之方程式時,吾人所作者即估量.統計技術之價值,大部分在於其能作估量之實際效用.

此種目的,在前數章中討論由一個變數之已知值估量另一變數之數值時,經常居主要之地位.此種方法之有效與否須視若干假定為斷,可於此略述之.

[估量時包含之假定]

前文曾言,一觀察值數列之最或然值(most probable value)為該數列之算術中數.諸觀察值在此中數周圍成正態分配時,則根據此中數而作之估量,其機率可以標準差確

切計量之.同樣若觀察值在響應線周圍成正態分配時,則根據此響應方程式而作之估量,其機率可以標準誤確切計量之.而響應方程式之意義與效率,可以該變數之估量標準誤與其標準差相比較,而決定之.

由此兩數值間之關係,且可求得一抽象之關係計量,卽繫聯係數或繫聯指數是也.此係數或指數,惟有在響應線周圍之分配以及在中數周圍之分配皆爲正態或近於正態時,始爲正當而確切之計量.與正態分配大有出入時,此等計量之意義將大爲減低.

前面之討論中始終用算術數值.言根據中數作估量時,吾人乃指算術中數而言.所謂中數周圍及響應線周圍之分配爲正則之假定,乃指用算術的方法計算差離而言.然吾人果能假定經濟分析中一切分配皆爲算術式乎?估量之作成與估量之錯誤之計量僅能用算術的名辭乎?若不然,則上文說明之方法能否適合於其他分配?此等問題最好作爲專門之問題以解答之.

[估量問題]

表一一四記載1881—1913年美國衛麥之產量與價格.兩數列各配合一適當之長期趨勢線,而算出每數列中各項實在值對於正則值之比率.

玆欲計量此兩變數間之關係,此關係似可以普通爲$Y=aX^b$式之雙曲線適當描述之.用最小平方法配合此曲線

時,此方程式必須改爲對數式

$$\log Y=\log a+b\log X$$

表一一四

美國雀麥產量與價格

年份	美國雀麥生產量（百萬蒲）	生產量之直線長期趨勢（註1）	實在生產量對長期趨勢值之比率	芝加哥每蒲雀麥價格	價格之直線長期趨勢（註2）	實在價格對長期趨勢值比率
1881	416	448	.929	47	36.0	1.30
1882	488	471	1.036	37	35.3	1.05
1883	571	494	1.156	31	34.6	.90
1884	583	517	1.128	29	34.0	.85
1885	629	540	1.165	28	33.2	.84
1886	624	563	1.108	25	32.5	.77
1887	659	586	1.124	30	31.2	.96
1888	701	609	1.151	24	30.5	.79
1889	751	632	1.188	24	29.8	.81
1890	523	655	.798	43	29.0	1.48
1891	738	678	1.088	31	28.3	1.10
1892	661	701	.943	30	27.5	1.09
1893	639	724	.882	31	26.8	1.16
1894	662	747	.886	28	26.1	1.07
1895	824	770	1.070	19	25.3	.75
1896	780	793	.983	18	23.6	.76
1897	790	816	.969	24	25.0	.96
1898	843	839	1.005	25	26.4	.95
1899	926	862	1.074	23	27.8	.83
1900	914	885	1.033	25	29.2	.86
1901	778	908	.857	42	30.6	1.37
1902	1053	931	1.131	33	32.0	1.03
1903	899	954	.911	38	33.4	1.14
1904	1009	977	1.033	30	34.8	.86
1905	1090	1000	1.090	31	36.2	.86
1906	1036	1023	1.013	39	37.6	1.04
1907	805	1046	.770	51	39.0	1.31
1908	851	1069	.796	52	40.4	1.29
1909	1068	1092	.978	43	41.8	1.03
1910	1186	1115	1.064	35	43.2	.81
1911	922	1138	.810	31	44.6	1.14
1912	1418	1161	1.221	37	46.0	.80
1913	1122	1184	.948	41	47.4	.87

（註1） 配合長期趨勢線時所用之材料,其包括之時期較此地所研究者爲長.

（註2） 配合長期趨勢線時全時期分爲1881—1895, 1896—1913之兩段.由基羅(H. B. Killough)氏就每段之材料各作一直線.

配合此種形式曲線所須之正則方程式爲

$$\text{I.}\quad \Sigma(\log Y)=N\log a+b\Sigma(\log X)$$

$$\text{II.}\quad \Sigma(\log X.\log Y)=\log a\Sigma(\log X)+b\Sigma(\log^2 X)$$

解此方程式所必須之數值可根據表一一五決定之.

表一一五

配合一曲線於美國雀麥生產量及價格之材料時
所須諸數值之計算表

第一例

(1) 年份	(2) 價格對長期趨勢之比 Y	(3) 產量對長期趨勢之比 X	(4) logY	(5) logX	(6) $\log^2 Y$	(7) $\log^2 X$	(8) logY·logX
1881	1.30	.929	.1139434	.9680157-1	.01298320	.001022995	−.0038444
1882	1.05	1.036	.0211893	.0153598	.00044899	.000235923	.0003255
1883	.90	1.156	.9542425-1	.0629578	.00209375	.003963685	−.0028808
1884	.85	1.128	.9294189-1	.0523091	.00498169	.002736242	−.0036920
1885	.84	1.165	.9242793-1	.0663259	.00573362	.004399125	−.0050222
1886	.77	1.108	.8864907-1	.0445398	.01288436	.001983794	−.0050557
1887	.96	1.124	.9822712-1	.0507653	.00031431	.002577217	−.0009000
1888	.79	1.151	.8976271-1	.0610753	.01048021	.003730192	−.0062524
1889	.81	1.188	.9084850-1	.0748164	.00837500	.005597494	−.0068468
1890	1.48	.798	.1702617	.9020029-1	.02898905	.009604432	−.0166852
1891	1.10	1.088	.0413927	.0366289	.00171336	.001341676	.0015162
1892	1.09	.943	.0374265	.9745117-1	.00140074	.000649053	−.0009539
1893	1.16	.882	.0644580	.9454686-1	.00415483	.002973674	−.0035150
1894	1.07	.886	.0293838	.9474337-1	.00086341	.002763216	−.0015448
1895	.75	1.070	.8750613-1	.0293838	.01560969	.000863408	−.0036712
1896	.76	.983	.8808136-1	.9925575-1	.01420540	.000055450	.0008895
1897	.96	.969	.9822712-1	.9863238-1	.00031431	.000187038	.0002425
1898	.95	1.005	.9777236-1	.0021661	.00049624	.000004692	−.0000483
1899	.83	1.074	.9190781-1	.0310043	.00654885	.000961267	−.0025089
1900	.86	1.073	.9344985-1	.0141003	.00429045	.000198818	−.0009236
1901	1.37	.857	.1367206	.9329808-1	.01725316	.004491573	−.0091629
1902	1.03	1.131	.0128372	.0534626	.00016479	.002858250	.0006863
1903	1.14	.911	.0569049	.9595184-1	.00323817	.001638760	−.0023036
1904	.86	1.033	.9344985-1	.0141003	.00429045	.000198818	−.0009236
1905	.86	1.090	.9344985-1	.0374265	.00429045	.001400743	−.0024515
1906	1.04	1.013	.0170333	.0056094	.00029013	.000031465	.0000955
1907	1.31	.770	.1172713	.8864907-1	.01375256	.014884361	−.0133113
1908	1.29	.796	.1105897	.9009131-1	.01223008	.009818214	−.0109580
1909	1.03	.978	.0128372	.9903389-1	.00016479	.000093337	−.0001240
1910	.81	1.064	.9084850-1	.0269416	.00837500	.000725850	−.0024656
1911	1.14	.810	.0569049	.9084850-1	.00323817	.008374812	−.0052076
1912	.80	1.221	.9030900-1	.0867157	.00939155	.007519613	−.0084036
1913	.87	.948	.9395193-1	.9768083-1	.00365792	.000537855	.0014027
共計	32.83	33.338	17.6715068-18	14.0375350-14	.21721807	.099422642	−.1194567
							+.0051562
							−.1143005

由此表得

$$N=33$$

$$\Sigma(\log Y)=-.32849 \qquad \Sigma(\log X.\log Y)=-.1143005$$

$$\Sigma(\log X)=.037535 \qquad \Sigma(\log^2 X)=.096423$$

代入正則方程式,得

$$-.32849=33\log a+.037535b$$

$$-.1143005=.037535\log a+.096423b$$

解之得

$$\log a=-.00861$$

$$b=-1.18206$$

而所求之方程式爲

$$\log Y=(9.99139-10)-1.18206\log X$$

或

$$Y=.9804X^{-1.18206}$$

此卽描述雀麥產量與價格間之平均關係之方程式(在兩數列之實在値皆表現爲對於各自長期趨勢線之比率時).與此方程式相當之曲線繪於圖82.

[用對數名辭表現之估量標準誤]

此方程式之可恃程度如何?欲答覆此問題,吾人必須計算標準誤S.在配合曲線時既全部用對數名辭進行,標準誤亦可用同樣之名辭計算之.仿照以前處理直線關係及定冪級數關係時所用之方法,對於剛才求得之對數曲線可得S

之公式如下:

$$S^2_{\log y}=\frac{\Sigma(\log^2 Y)-\log a\Sigma(\log Y)-b\Sigma(\log X.\log Y)}{N}$$

代入適當之數值,得

$$S^2\log=\frac{.21721807-(-.00861\times-.32849)-(-1.18206\times-.1143005)}{33}$$

$$=\frac{.07927928}{33}$$

$$S^2_{\log y}=.0024024$$

$$S_{\log y}=.04901$$

估量標準誤,其值取對數形式時,爲 .04901. 吾人如爲處理對數,則解釋此標準誤之方法,與解釋其他曲線之標準誤之法完全相同.假定對數值在描述平均關係之曲線周圍其分配之狀態爲正態,則估量值之對數與實在值之對數間,相差不超過 .04901 者,其機會爲 68%,不超過 .09802 者,其機會爲 95%,不超過 .14703 者,其機會爲 99.7%.

〔估量標準誤之解釋:估量帶〕

此種標準誤在實在數值上表示何種意義乎?彼僅表示吾人所處理者,始終爲比率而非絕對數也.兩數之對數之差,等於原來兩數之一對於另一數之比率.故欲知 S 在某處之絕對值爲何,須視吾人所處理之數值之大小爲斷,若欲將 S 化爲絕對值,必須指一定之估量而言.卽將 X 之某值代入平均關係方程式而估量一與之相當之 Y 值.所用者若爲對數

方程式,則估量之結果亦爲一對數.在此估量之對數上加$S_{\log y}$,求和之逆對數,卽得一估量帶之上限,其高於響應線之距離等於S.由此估量之對數中減去$S_{\log y}$之值,求差之逆對數,卽得一估量帶之下限,此下限低於響應線之距離亦等於S.該例之Y值在100次中將有68次歸於如此劃定之限度以內.估量帶之上下限其對於響應線之距離等於$2S$與$3S$者,其絕對值亦可以同法決定之.

如此劃定之關於對數曲線之估量帶,將與前面處理簡單之直線方程式時敍述之估量帶大不相同.簡單之估量帶,其兩邊各距估量曲線$1S$之距離者,自首至尾其絕對闊度不變,而且始終以響應線爲中心.但對數式估量帶,如以自然數計之,其闊狹有變動,且其在曲線兩方之闊度不等.然在此曲線兩方之比率乃經常相等者.卽低於計算值$1S$之數值對於計算值之比率,等於計算值對於較計算值大$1S$之數值之比率.而此曲線繪於對數尺度之格紙上時,以所作之曲線爲中心兩邊各距$1S$之長之估量帶,亦成對稱形式,與前面簡單估量方程式之估量帶所取之形式相同.慣於以比率思想及慣於運用對數格紙者,對於此種計量極易解釋.

[以比率表示之估量標準誤]

比率旣始終相等,故估量之標準誤可以比率表示之.在現今之例中,以S_r代表用比率表示之估量標準誤,吾人得

$$S_r=\text{anti-log}\,S_{\log y}=\text{anti-log}\,.04901=1.12$$

$S_{\log y}$ 在上面求得時為正數,故比率大於一.此為較大數對於較小數之比率.何謂也?卽實在値在 100 次中應有 68 次為如下之情形:若大於估量値其超過之數不多於12%;若小於估量値,則將在如此之一限度以內,卽估量値大於實在値之數將不超過實在値之 12%.此種形式始終以較小數為基數,而將較大數化為其比率;故不甚方便.若能一律以估量値為基數,當較為便利.欲達此目的,可將 $S_{\log y}$ 作成負數形式,求相當之自然數如下: $-.04901=9.95099-10$,相當之自然數為.8933.此比率表示較小數對於較大數之關係,合前後兩種比率形式而並記之,S_r 之意義可易於明瞭.

$$S_r=.89—1.12$$

此計量之解釋,卽當分配為正態時,實在値在 100 次中有 68 次不能小於估量値之 89%,不能大於估量値之 112%.此種用比率表示之標準誤,意義簡單而確定,於大部分實際用途上,比較用絕對數値表示之標準誤常更為重要.(註)

欲求 $2S$ 或 $3S$ 之値時,吾人不能簡單的將上面之百分數乘以 2 或 3.須將 $S_{\log y}$ 乘以 2 或 3 而將結果所得之數變為自然數.為應用便利起見,應如前例同時求正負兩値之逆對數計算之法極簡.

$$2S_{\log y}=.09802$$

(註) 1922 年達文波脫氏曾在一篇未發表之著作中指出百分數式之可靠性計量之重要性.此種計量曾應用於幾種研究中,惟尚未有計算此種計量之現成方法,而其可能性尚未盡數利用耳.

此值爲正時,其逆對數爲 1.25;爲負時,其逆對數爲 .80.

$$3S_{\log y}=.14703$$

正負值之逆對數各爲1.40與.71.茲將各標準誤綜列於下:

$$S_r=.89\text{至}1.12$$

$$2S_r=.80\text{至}1.25$$

$$3S_r=.71\text{至}1.40$$

實在值與估量值,位於 S_r 之值所指示之或然百分數限界 (Probable percentage limits) 之內者,100個中有68個.位於$2S_r$之值所指示之限界內者,100 個中有95個,位於$3S_r$之值所指示之限界內者,100個中有99.7個,始終以實在值之對數在配合曲線周圍成一正態分配爲假定.

[估量標準誤之應用]

茲可說明 $S_{\log y}$ 之應用.設雀麥生產量超過長期趨勢值 50%(即其對於長期趨勢之比率爲 1.50),則價格方面與此相當之最有或然性之比率爲何?並此種估量之正確程度爲何?

估量方程式爲

$$\log Y=(9.99139-10)-1.18206\log X.$$

將 .176091(1.50 之對數)一值代入此方程式,得 $\log Y$ 之值爲 9.78324—10.相當之自然數爲.607.此數表示生產量若爲正則(根據一定之長期趨勢線而計量所得之正則)之 150%,價格或爲正則(根據長期趨勢線計算而得)之 60.7.

欲知此計量之可靠性,必須求標準誤.用已經求得之S_r之數值,求得60.7之89%爲54, 60.7之1.12%爲68.此兩數值之意義,即在設定之生產狀況下,實在價格不小於長期趨勢值之54%而不大於長期趨勢值之68%之機會,100中占68.(註1)相當之$2S_r$與$3S_r$之數值可依照上述之方法決定之.

[根據對數值之緊聯指數]

吾人猶須計算第三種計量,即抽象之緊聯指數.(註2)關於下列形式之方程式

$$\log Y = \log a + b\log X,$$

ρ之公式變爲

$$\rho^2_{\log y \log x} = \frac{\log a\Sigma(\log Y) + b\Sigma(\log X \log Y) - Nc^2_{\log y}}{\Sigma(\log^2 Y) - Nc^2_{\log y}}$$

式中$c_{\log y}$代表Y諸值之對數之算術中數與原點(在此地爲對數尺度上零點)間之差,代入適當數值,得

(註1) 在現今之問題中如此決定估量之可靠性時,即發生所用之長期趨勢線適宜與否之問題.此問題另節詳論之.

(註2) 雖在對數式中關係成一直線形式,然關係計量之記號仍用ρ而不用r.因此種以對數表示之計量,不能用解釋尋常緊聯係數之完全相同之方法解釋之也.

$$\rho^2_{\log y \log x}$$

$$=\frac{(-.00861\times-.32849)+(-1.18206\times-.1143005)-(33\times.00009909)}{.21721807-(33\times.00009909)}$$

$$=\frac{.13466882}{.2139481}$$

$$=.629445$$

$$\rho_{\log y \log x}=.793$$

繫聯指數之值為 .793. 當吾人所處理者為對數有如目前之例時,此值應如何解釋乎?

吾人如從下列關係上觀察之,其意義自明:

$$\rho_{\log y \log x}=1-\frac{S^2_{\log y}}{\sigma^2_{\log y}}$$

在此例中,此等數值為

$$S_{\log y}=.04901$$

$$\sigma_{\log y}=.08052$$

將此等數值作成平方,代入上列公式,得

$$\rho^2_{\log y \log x}=1-\frac{.002402417}{.00648328}$$

而

$$\rho_{\log y \log x}=.793$$

此值所計量者何耶?吾人已知 r 與較普遍之指數 ρ 為兩變數間關係程度之計量,此關係乃為一定之函數式所描述者.一例之 ρ, 其值決定於配合曲線周圍之離中趨勢對

於Y諸值之中數周圍之離中趨勢兩者間之關係.若以響應方程式代替Y之中數作爲估量之根據時,估量之離中趨勢卽可因之大減者,吾人可認此方程式所代表之關係爲重要.故此ρ之值,決定於S_y與σ_y兩個數量間之關係.

前面各章所處理之各例中,此兩種離中趨勢之計量皆以絕對差離(Absolute deviations)表示之,而ρ之值則視此兩種絕對離中趨勢計量間之關係而決定.現今之例其唯一不同之處,卽在吾人所處理者,爲對數的或比率的離中趨勢(Logarithmic or ratio variability);差離之計算,不用自然數,而用對數而巳.

指數ρ必須根據此事實解釋.其值與平常一樣,爲S^2及σ^2兩離中趨勢計量間之關係所決定,惟在此例中此兩計量皆爲用對數表示者.簡言之,決定ρ之值者,爲在配合曲線周圍之比率離中趨勢與在Y之幾何中數(所以爲Y之幾何中數者,因此值相當於Y對數之算術中數也.)周圍之比率離中趨勢兩者間之關係也.

吾人於此得一組計量,其在比率之領域內所盡之任務,恰與S與ρ(在直線關係中爲r)在自然數之領域所盡者,完全相同.此等計量求得之方法亦與求S與ρ之方法相同,所異者僅其所根據之關係方程式中,應變數爲$\log Y$(或在相反之情形則爲$\log X$).計算此等數值之普通公式與處理自然數時所用者相同,僅凡在Y之處皆以$\log Y$代替之耳.其法

與用對數格紙代替自然數格紙時相類.

吾人宜注意者,Y若用對數表示,不論X是否用對數形式,求得之數值皆爲對數或比率之形式.前面所配合之曲線其式爲

$$\log Y=\log a+b\log X.$$

卽平常拋物線或雙曲線之對數式也.若曲線之式爲

$$\log Y=\log a+X\log b,$$

卽變冪曲線

$$Y=a(b^X)$$

之對數式,S與ρ之值將仍取對數形式.

上面兩個對數式皆爲直線形式,惟此非應用此兩計量時必要之條件.S與ρ之兩計量,不論處理者爲比率或自然數,不論函數關係爲直線式或非直線式,皆可普遍適用者也.

於此地最好將曾用以分別各種不同計量之記號綜述之.問題爲算術式關係時,吾人可用S_y, σ_y,與ρ三個記號;前二者爲用絕對數表示離中趨勢之計量.ρ表示用自然數時之關係程度.若用Y各值之對數,最好於記號後加註以示分別:用$S_{\log y}$與$\sigma_{\log y}$兩記號分別表示在配合曲線周圍之對數的離中趨勢與在Y各值對數之算術中數周圍之對數的離中趨勢.若$S_{\log y}$化爲比率形式,可寫爲S_r.因指數ρ在此地解釋時必然有少許不同之處,故此記號可寫爲$\rho_{\log y \log x}$或$\rho_{\log y x}$.

[倒數在計量關係時之應用]

尙有一種曲線亦可用以描述雀麥產量與價格間之關係,此種曲線之應用引吾人入繁聯之第三個領域,其中有若干新觀念加入,而解釋各種計量時必須另用一種方式.此曲線之形式爲

$$Y=\frac{1}{a+bX},$$

此式可以增加分母之項數而展開之:

$$Y=\frac{1}{a+bX+cX^2}$$

此種雙曲線式曾被用於幾種研究中,作爲多種商品之"須要"曲線(Demand curve)之近似曲線.

代表此種曲線之方程式可書爲

$$\frac{1}{Y}=a+bX$$

此式爲描述Y諸值之倒數與X之原來諸值間之關係之直線方程式.配合此種曲線時所須之正則方程式爲

$$\text{I. } \Sigma\left(\frac{1}{Y}\right)=Na+b\Sigma X$$

$$\text{II. } \Sigma\left(\frac{X}{Y}\right)=a\Sigma(X)+b\Sigma(X)^2$$

計算方程式中所必須之諸數值之方法,見下表:

表一一六

配合一曲線於雀麥產量與價格之材料時所須諸數值之計算表

第二例

(1) 年份	(2) 價格比率 Y	(3) 生產量比率 X	(4) $\frac{1}{Y}$	(5) $\frac{X}{Y}$	(6) $\left(\frac{1}{Y}\right)^2$	(7) X^2
1881	1.30	.929	.7692308	.7146154	.59171602	.863041
1882	1.05	1.036	.9523810	.9866667	.90702957	1.073296
1883	.90	1.156	1.1111111	1.2844444	1.23456788	1.336336
1884	.85	1.128	1.1764706	1.3270588	1.38048307	1.272384
1885	.84	1.165	1.1904762	1.3869048	1.41723358	1.357225
1886	.77	1.108	1.2987013	1.4389610	1.68662507	1.227664
1887	.96	1.124	1.0416667	1.1708334	1.08506951	1.233376
1888	.79	1.151	1.2658228	1.4569620	1.60230736	1.324801
1889	.81	1.188	1.2345679	1.4666667	1.52415790	1.411344
1890	1.48	.798	.6756757	.5391892	.45653765	.636804
1891	1.10	1.088	.9090909	.9890909	.82644628	1.183744
1892	1.09	.943	.9174312	.8651376	.84168001	.889249
1893	1.16	.882	.8620690	.7603449	.74316296	.777924
1894	1.07	.886	.9345794	.8280373	.87343865	.784996
1895	.75	1.070	1.3333333	1.4266666	1.77777769	1.144900
1896	.76	.983	1.3157895	1.2934211	1.73130201	.966289
1897	.96	.969	1.0116667	1.0093750	1.08506951	.938961
1898	.95	1.005	1.0526316	1.0578948	1.10803329	1.010025
1899	.83	1.074	1.2048193	1.2939759	1.45158955	1.153476
1900	.86	1.033	1.1627907	1.2011628	1.35208221	1.067089
1901	1.37	.857	.7299270	.6255480	.53279343	.734449
1902	1.03	1.131	.9708738	1.0980583	.94259504	1.279161
1903	1.14	.911	.8771930	.7991228	.76946756	.829921
1904	.86	1.033	1.1627907	1.2011628	1.35208221	1.067089
1905	.86	1.090	1.1627907	1.2674419	1.35208221	1.188100
1906	1.04	1.013	.9615385	.9740385	.92455629	1.026169
1907	1.31	.770	.7633588	.5877863	.58271666	.592900
1908	1.29	.796	.7751938	.6170543	.60092543	.633616
1909	1.03	.978	.9708738	.9495146	.94259591	.956484
1910	.81	1.064	1.2345679	1.3135802	1.52415790	1.132096
1911	1.14	.810	.8771930	.7105263	.76946758	.656100
1912	.80	1.221	1.2500000	1.5262500	1.56250003	1.490841
1913	.87	.948	1.1494253	1.0893552	1.32117852	.898704
	32.83	33.338	34.3360320	35.2571485	36.85702940	34.168554

將適當數值代入正則方程式,得

$$34.3360320=33a+33.338b$$

$$35.2571485=33.338a+34.168554b$$

解之,

$$a=-.1357$$

$$b=1.1643$$

故所求之方程式爲

$$\frac{1}{Y}=-.1357+1.1643X$$

[用倒數表示之標準誤與繫聯係數]

欲決定此方程式之效率,必須有標準誤及繫聯係數.求此兩計量時所必須之兩個公式,可照以前諸例求得之.以$\frac{1}{Y}$代表一個實在值之倒數.則每個實在值對於計算值之差離爲,

$$d=a+bX-\frac{1}{Y}$$

各乘以d而總加之,得

$$\Sigma(d^2)=a\Sigma(d)+b\Sigma(dx)-\Sigma\left(\frac{d}{Y}\right)$$

因

$$\Sigma(d)=0 \text{ 而 } \Sigma(dX)=0.$$

故

$$\Sigma(d^2)=-\Sigma\left(\frac{d}{Y}\right)$$

將前面差離方程式乘以 $\frac{1}{Y}$ 而總加之得

$$\Sigma\left(\frac{d}{Y}\right)=a\Sigma\left(\frac{1}{Y}\right)+b\Sigma\left(\frac{X}{Y}\right)-\Sigma\left(\frac{1}{Y}\right)^2$$

將 $\Sigma\left(\frac{d}{Y}\right)$ 之相當量代入前一方程式得

$$\Sigma(d^2)=\Sigma\left(\frac{1}{Y}\right)^2-a\Sigma\left(\frac{1}{Y}\right)-b\Sigma\left(\frac{X}{Y}\right)$$

而吾人得 $S_{\frac{1}{y}}^2$ 之公式爲

$$S_{\frac{1}{y}}^2=\frac{\Sigma\left(\frac{1}{Y}\right)^2-a\Sigma\left(\frac{1}{Y}\right)-b\Sigma\left(\frac{X}{Y}\right)}{N}$$

將此 $S_{\frac{1}{y}}^2$ 之公式代入聚聯指數之普通公式得

$$\rho^2=1-\frac{S_{\frac{1}{y}}^2}{\sigma_{\frac{1}{y}}^2}$$

化簡之得

$$\rho^2_{\frac{1}{y}x}=\frac{a\Sigma\left(\frac{1}{Y}\right)+b\Sigma\left(\frac{X}{Y}\right)-Nc^2_{\frac{1}{y}}}{\Sigma\left(\frac{1}{Y}\right)^2-Nc^2_{\frac{1}{y}}}$$

將適當數值代入兩方程式得

$$S_{\frac{1}{y}}=.1191$$

$$\rho_{\frac{1}{y}x}=.766$$

求原來之 Y 諸值以倒數表示時之標準差得

$$\sigma_{\frac{1}{y}}=.1851$$

（此等計量皆加註 $\frac{1}{y}$, 以別於根據自然數或對數而求得之同樣計量）.

〔用倒數表示之估量標準誤之解釋〕

吾人將如何解釋此等結果乎?與前面所有之此種問題相同,方程式為由 X 之已知值估量 Y 之值之工具.標準誤 $S_{\frac{1}{y}}$ 為此種估量之可靠性之計量,而 $\rho_{\frac{1}{y}x}$ 為兩變數間關係程度之抽象計量.惟現今所有此等計量皆用倒數表示.方程式使吾人能估量 Y 之倒數,標準誤亦僅以倒數形式表示其

意義,ρ之值爲兩計量($S_{\frac{1}{y}}^2$與$\sigma_{\frac{1}{y}}^2$)間關係所決定,而此兩計量乃皆以倒數表示者也.

茲舉例說明之若某年之雀麥產量爲正則產量之150,則最有或然性之價格當爲何者?以1.50作爲X之值代入下列方程式

$$\frac{1}{Y}=-.1357+1.1643X$$

得

$$\frac{1}{Y}=1.6108$$

而

$$Y=.621$$

即在如此之生產情形時,吾人可料雀麥之價格約爲正則價格之62%.吾人求得關於此估量之可靠性之計量爲

$$S_{\frac{1}{y}}=.1191$$

此數必須用於以倒數表示之估量值,如下

$$1.6108+.1191=1.7299$$

$$1.6108-.1191=1.4917$$

將此兩倒數變爲自然數,得.578與.670,此即吾人所求之值也.於是知最有或然性之價格爲正則值之62.1%,假定在曲線周圍倒數之分配近似於正態,則100次機會中有68次之情形爲價格位於正則值之57.8%—67.0%之間.$2S$與$3S$之限界亦可同樣求之:先在用倒數表示之估量值上加上與減去

兩倍與三倍於.1191之數,更將結果化爲自然數卽得.兩倒數間一定量之差,其情形正與兩對數間一定量之差相同,其値以絕對量計算時,隨其在Y値範圍內之地點不同而有變化.因此根據$S_{\frac{1}{y}}$而決定之可靠性之限界,僅在已經作一特殊之估量後,始能用自然數表示.

[由算術式.幾何式及倒數式計量得出之關係計量之比較]

解釋ρ時發生同樣之問題.吾人已知繫聯係數之數値爲曲線周圍離中趨勢程度對於應變數原來數列之平均數周圍離中趨勢程度之比較所決定.處理自然數時,曲線周圍之離中趨勢與應變數算術中數周圍之離中趨勢相比,兩者皆以自然數表示(卽S_y與σ_y相比也.)處理對數時,配合曲線周圍之離中趨勢與應變數之對數之算術中數周圍之離中趨勢相比,兩者皆以對數計算,但吾人知對數差離可用比率解釋.對於曲線之對數差離,代表實在値對於計算値之比,對於原來數列的對數之算術中數之對數差離,代表應變數列之實在値對於其幾何中數之比. $\rho_{\log y}$之値隨此兩種差離間之關係(卽$S_{\log y}$與$\sigma_{\log y}$相比)而定.

用應變數之倒數配合一曲線時,曲線周圍之離中趨勢與原來數列之離中趨勢皆以倒數計算,卽$\sigma_{\frac{1}{y}}$根據諸實在値之倒數與所有倒數之算術中數間之差計算而得.然此種

倒數之算術中數,爲原來數列之倒數中數之倒數.故簡言之.緊聯指數$\rho_{\frac{1}{y}}$之値,隨配合曲線周圍之離中趨勢與應變數列倒數中數周圍之離中趨勢兩者間之關係(即$S_{\frac{1}{y}}$與$\sigma_{\frac{1}{y}}$相比)而定,兩種離中趨勢皆以倒數計算.

故描述兩變數間關係之曲線有三大族,

1. 曲線之用應變數之自然數値配合者.凡屬此族之一切曲線其方程式皆爲下式

$$Y=f(X)$$

2. 曲線之用應變數之對數配合者.凡屬此族之一切曲線其方程式將爲下式

$$\log Y=f(X)$$

3. 曲線之用應變數之倒數配合者.此種曲線之方程式將爲下式

$$\frac{1}{Y}=f(X)$$

三族中任何一族,其方程式皆可爲直線式或爲非直線式.僅就解釋之問題言,對於應取X之何種函數一事並無限制.(惟用前述諸法計算S與ρ時涉及若干限制,此點當於他處討論之.)

第一族曲線之估量標準誤,求得時其數値爲用應變數之原來計算單位表示者,故其意義簡單而直接.此種曲線之

繫聯係數,計量以配合曲線代替算術中數爲計算差離之標準時,應變數之絕對離中趨勢可因此而減少之程度.

第二族曲線之估量標準誤,求得時其數值爲用對數表示者,平常將其解釋爲比率較爲便利.繫聯係數 $\rho_{\log y}$ 所計量者,爲應用配合曲線代替幾何中數作爲計算差離(或比率)標準之時,應變數之對數的或比率的離中趨勢可因而減少之程度.

第三族曲線之估量標準誤,求法與其他兩種相同,惟求得時係用倒數表示者.繫聯係數 $\rho_{\frac{1}{y}}$ 所計量者,即用配合曲線代替倒數中數作爲標準而計算倒數差離時,以倒數表示之應變數其離中趨勢可因而減少之程度.

[抉擇關係計量之要素]

由此可知抉擇一種形式之曲線以敍述某一關係時,所須考慮之基本之點,爲應以何種形式之平均數作該數列集中趨勢之最適當之計量也.與此連帶之問題,即平均數周圍之離中趨勢,在自然數,對數與倒數三者中究應用何者計算,始能最接近於正態分配也.當選擇一曲線與應用 S 與 ρ 兩計量時,關於以上諸點輒常有一暗中之假定存在.

當重要者爲絕對數值,而應變數稍於算術尺度上時其離中趨勢近於正態者,則以算術式關係計量爲適用,但當吾人處理之數列其最重要之點不在變動之絕對量而在變動

之率 rate),而離中趨勢似依照一幾何法則時,算術中數與其他算術式之計量顯然不適用,遇此種情形,對數曲綫似較算術曲線爲佳,而計量估量可靠性與關係程度之數字,根據比率求得者,似較根據絕對數值求得者爲適宜.

倒數中數之應用,不若前兩種平均數之廣,應用於目前討論之一類問題上時,有若干原則須稍加注意者.一般言之,此種倒數式之計量,與算術式計量具有相同之缺點,僅其錯誤之方向相反耳.幾何式計量較之兩者似較適宜於普遍之應用但有一經濟學家所關心之特殊領域,爲倒數中數所特別適用者,如前舉之一例若應用倒數中數,即似爲正當者也.

倒數中數之應用,以倒數成正態分配爲假定;即假定在自然數上表示高於平均數之離中趨勢超過低於平均數者甚巨.應用$\frac{1}{Y}=a+bx$式之曲線時,即假定在X與Y之關係中,亦有同樣之情形存在.X增加一定數量時,Y之值減低若干,惟X之值減低同一數量時,Y值增加之數,將大於彼時所減低之數.許多商品之生產量與其價格間,當將價格作爲應變數時,其關係卽爲此種情形.當商品之生產量增加某一數量時,將使價格減低若干;惟當生產量減低同一之數量時,引起價格之高漲,高漲之數且大於彼時所減低之數.並且,求此種商品在一個時期中之平均價格時,倒數中數比較其他任何種

平均數更能作爲一標本之數值.(註) 遇此種情形,應用倒數曲線並用倒數關係表示一切估量之精確程度,自有鞏固之理由,可不待言也.

[算術式、幾何式、倒數式計量之比較]

欲將此各種不同之方法作一對照,惟有配合三種形式之曲線於同一材料取其結果互相比較之一法最爲有效.關於雀麥生產量與價格之材料(表一一四),配合幾何式倒數

(註) "買賣馬鈴薯者,所擬定之價格,常不能與基本經濟狀況適合.此種錯誤在秋季尤爲普通,或致價格過於抬高.在季節之初期價格既爲過高,馬鈴薯消費速度遲緩,不及出去所有之供給.農人商賈見其所持存貨不能悉以目前之價格出售.因馬鈴薯不能留至明年,故在上述情形之下,其價格必須充分壓低,使在季節終了以前全部供給盡可出售.若起首即有一適宜之價格,則除棧租之累積外,整個時期將維持原狀,而消費之量前後頗可爲一致.惟季節初期之變態高價使消費量減少,則其餘時期必須以變態之低價補償之,否則收穫之物不能全部售盡.

同樣,若季節之初價格異常低,則供給量將瞬息售盡,彼尚有餘存之馬鈴薯在手者,將見在季節之其餘時期能得一異常高昂之價格."

然則在一季節中有部分時期之價格異常高昂或異常低賤時,吾人將如何計算一平均價格,其與精確估量之供需情況相合者乎?既然"季節中一部分時期之低價必須以其餘時期不相稱之高價爲補償",則全時期價格之算術中數,"將較季節一開始即有一適當之價格時最後所得之平均數爲高.此種困難,若取逐月份價格之倒數中數,即可解除."(見渥金(oHlbrook Working)氏書)

式兩種曲線時之一切計算,前已說明.茲更配合一直線(算術式)於同一材料,並計算一切必須之相連之計量.三套結果列於下表.

表一一七

1881—1913年番麥生產量與價格之關係
配合不同形式曲線所得結果之比較
(各以價格為應變數)

	方程式	估量標準誤	聚聯係數
A	$Y=2.24-1.236X$	$S_y=.12$	$r=.783$
B	$\frac{1}{Y}=-.1357+1.1643X$	$S_{\frac{1}{y}}=.1191$	$\rho_{\frac{1}{y}}=.766$
C	$\log Y=-.00861-1.18206\log X$	$S_{\log y}=.04901$	$\rho_{\log y}=.793$

三種估量標準誤照現在情形不能相比,因僅有第一種所用者爲原來之計量單位(實在產量對於正則之比率)也.下表所列者,爲根據每一方程式所作之估量,表示與五種生產情形相伴之或然性最大之價格(作爲對於正則之比率表示).每一估量之後隨以一列數值表示估量標準誤所設定之界限.估量值加上與減去 S, $2S$ 及 $3S$ 後所得之值到處列出,以示實在數值在估量值周圍之或然離中趨勢.此外,更計量估量值與界限值間實際差數,以示三關係曲線周圍各各不同之離中趨勢量.此差數列於 A 之一行中.表中一切數值業已全部化爲原來之單位(價格對正則之比率),故可以互相比較.

表一一八

根據棉花生產量及其價格之三個關係方程式
所作之價格估量及估量標準誤之比較

(1) X之值（出產量對正則之比率）	(2) Y之估量值（價格對正則之比率）根據算術方程式(A)	(3) 算術式估量之界限	(4) Δ	(5) Y之估量值，根據倒數方程式(B)	(6) 倒數式估量之界限	(7) Δ	(8) Y之估量值，根據對數方程式(C)	(9) 對數式估量之界限	(10) Δ
.5	1.622	+3S=1.982	+.36	2.240	+3S=11.223	+8.983	2.224	+3S=3.114	+.890
		+2S=1.862	+.24		+2S= 4.803	+2.563		+2S=2.780	+.556
		+S=1.742	+.12		+S= 3.055	+.815		+S=2.491	+.267
		−S=1.502	−.12		−S= 1.708	−.472		−S=1.979	−.245
		+2S=1.382	−.24		−2S= 1.461	−.779		−2S=1.779	−.445
		−3S=1.262	−.36		−3S= 1.244	−.996		−3S=1.579	−.645
.8	1.251	+3S=1.[illegible]11	+.36	1.257	+3S= 2.281	+1.024	1.276	+3S=1.783	+.510
		+2S=1.491	+.24		+2S= 1.794	+.537		+2S=1.595	+.310
		+S=1.371	+.12		+S= 1.478	+.221		+S=1.429	+.153
		−S=1.131	−.12		−S= 1.093	−.164		−S=1.136	−.140
		−2S=1.011	−.24		−2S= .967	−.290		−2S=1.021	−.225
		−3S= .891	−.36		−3S= .867	−.390		−3S= .906	−.370

表一一八（續）

(1) X之值（生產量對正則之比率）	(2) Y之估計值（價格對正則之比率）根據算術方程式(A)	(3) 算術式估量之界限	(4) Δ	(5) Y之估量值，根據倒數方程式(B)	(6) 倒數式估量之界限	(7) Δ	(8) Y之估量值，根據對數方程式(C)	(9) 對數式估量之界限	(10) Δ
1.0	1.004	$+3S=1.364$ $+2S=1.244$ $+S=1.124$ $-S=.884$ $-2S=.704$ $-3S=.644$	$+.36$ $+.24$ $+.12$ $-.12$ $-.21$ $-.36$	.972	$+3S=1.490$ $+2S=1.265$ $+S=1.100$ $-S=.870$ $-2S=.789$ $-3S=.722$	$+.518$ $+.293$ $+.128$ $-.101$ $-.183$ $-.250$	.980	$+3S=1.372$ $+2S=1.226$ $+S=1.098$ $-S=.872$ $-2S=.781$ $-3S=.698$	$+.392$ $+.246$ $+.118$ $-.108$ $-.199$ $-.284$
1.2	.757	$+3S=1.117$ $+2S=.997$ $+S=.877$ $-S=.637$ $-2S=.517$ $-3S=.397$	$+.36$ $+.24$ $+.12$ $-.12$ $-.24$ $-.36$	.703	$+3S=1.106$ $+2S=.977$ $+S=.8.5$ $-S=.721$ $-2S=.667$ $-3S=.618$	$+.313$ $+.184$ $+.082$ $-.069$ $-.126$ $-.176$	.790	$+3S=1.100$ $+2S=.987$ $+S=.885$ $-S=.703$ $-2S=.632$ $-3S=.561$	$+.310$ $+.197$ $+.095$ $-.087$ $-.158$ $-.220$
1.5	.383	$+3S=.740$ $+2S=.620$ $+S=.500$ $-S=.260$ $-2S=.140$ $+2S=.020$	$+.36$ $+.24$ $+.12$ $-.12$ $-.24$ $-.36$	.621	$+3S=.708$ $+2S=.728$ $+S=.670$ $-S=.578$ $-2S=.541$ $-3S=.508$	$+.177$ $+.107$ $+.049$ $-.043$ $-.080$ $-.113$	.607	$+3S=.852$ $+2S=.761$ $+S=.680$ $-S=.542$ $-2S=.484$ $-3S=.443$	$+.245$ $+.154$ $+.073$ $-.065$ $-.123$ $-.174$

[估量帶及其意義]

如對此表作一仔細之研究,則根據此地三種方程式而作之估量,其性質即可以明瞭.基本差別之點,與其謂在估量之實在數值,無寧謂在計量此等估量之可靠性及指示實在值大致位於何種界限以內之標準誤.換言之,差別之點在對於曲線周圍離中趨勢之性質所作之假定.

算術式曲線之標準誤 S_y,表示估量值不論爲大爲小,其差誤皆有一致之絕對範圍.故此中假定曲線周圍之離中趨勢爲算術式者.高於估量值一 S_y (或 S_y 之任何倍數)之數值與低於估量值同數之數值,其算術中數等於估量值.此種情形可於圖 81 上見之.圖上所繪者,爲原來材料之各點,一條算術式之關係直線,以及以此線爲中心闊度等於 $2S$, $4S$ 與 $6S$ 之估量帶.

幾何式曲線之標準誤$S_{\log y}$,表示估量值不論爲大爲小,其差誤皆有一致之相對的或百分比的範圍(Relative or percentage range).即表示實在值所分佈之絕對範圍,當估量值小時較之當估量值大時狹小甚多.故此中假定曲線周圍之離中趨勢爲幾何式者.在此地,高於估量值一 S_y (或 S_y 之任何倍數)之數值與低於估量值同數之數值,其幾何中數等於估量值,圖 82 爲此種關係之圖示.圖上繪有代表原來材料之各點,並給一以

$$Y = .9804X^{-1.18206}$$

爲方程式之曲線,並劃出在幾何的意義上以關係線爲中心而其闊各等於 $2S_r$, $4S_r$ 及 $6S_r$ 之估量帶之界限.將圖81與圖82相比較,可使吾人對於根據算術式分配之假定而作之估量與根據幾何式分配之假定而作之估量兩者間之差異,得一較清楚之了解.

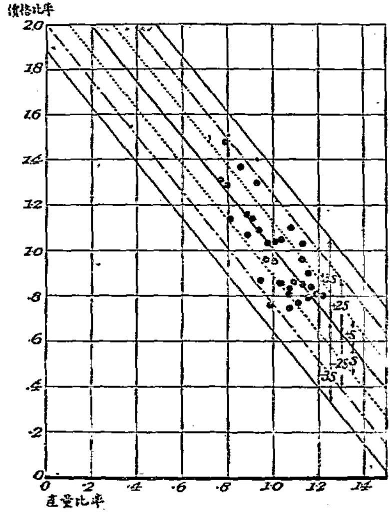


圖81.——雀麥產量與價格間之關係;表示算術式變直方程式與算術式估量帶之應用.

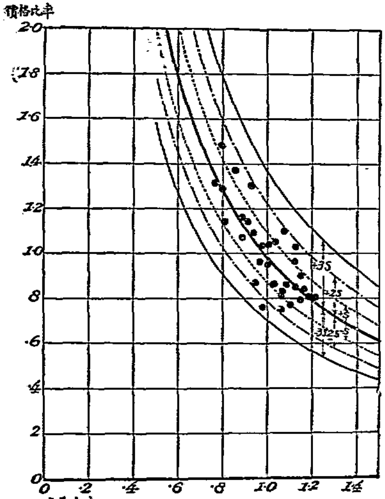


圖82.——雀麥產量與價格間之關係;表示對數式雙曲方程式與幾何式估量帶之應用.

圖82上之各點各線繪於對數格紙上,即成圖83之形式.在此種格紙上,關係曲線變成直線,估量帶變爲對稱,而其闊度亦前後一致.材料繪於對數格紙上時所發生之此種變化,使吾人知根據對數值作估量時所涉及之假定,根本上原極簡單也.

用倒數式標準誤$S_{\frac{1}{y}}$時,吾人更將價格愈高則曲線周圍離中趨勢亦愈大之一假定更推進一步.此種標準誤表示估量值低時,估量之差誤範圍極爲狹小,估量值高時,估量之差誤範圍極大.計算值,或估量值,永爲高於此值一$S_{\frac{1}{y}}$(或$S_{\frac{1}{y}}$之任何倍數)之數值與低於該值同數之數值兩者間之倒數中數.

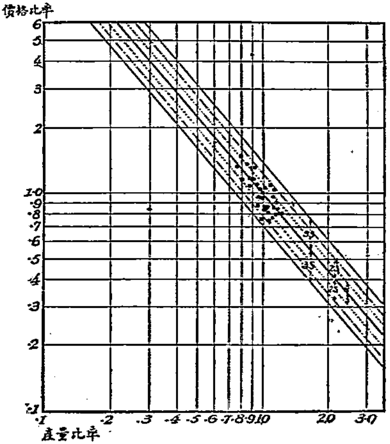


圖83.——雀麥產量與價格間之關係;表示對數式擬合方程式與幾何式估量帶之應用.(繪於雙對數格紙上)

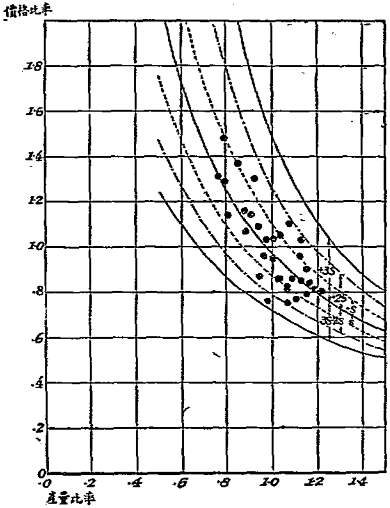


圖84.——雀麥產量與價格之關係;表示倒數式雙曲方程式與倒數式估量帶之應用.

在圖84中繪有以

$$\frac{1}{Y}=-.1357+1.1643X$$

爲方程式之曲線,並原來材料之觀察值,以及在倒數之意義上以配合曲線爲中心而具有$2S_{\frac{1}{y}}$, $4S_{\frac{1}{y}}$, $6S_{\frac{1}{y}}$之闊度之估

量帶.此圖與前二者之差別異常顯著,其估量帶尤然.假定關係曲線周圍之倒數成正態分配時,最外之一估量帶(闊度等於 $6S$ 者),表示 99.7% 之點數應在其範圍之內;最內之一估量帶(闊度等於 $2S$),表示所有點數中 68% 應在其範圍之內.若不繪自然數而繪倒數,此顯非正態之分配可化爲對稱形式,一如將幾何數值繪於對數格紙上所得者.

對於估量值之高者與低者,幾何式計量之 $S_{\log y}$ 位於算術式計量 S_y 與倒數式計量 $S_{\frac{1}{y}}$ 兩者之間.此後兩者固自有其功用,適合於某等情況;然將此等方法用於經濟分析時,幾何式之一族計量較之彼其他兩族者更有普遍之用途,諒爲事實.此僅表示比率常較絕對差數更爲重要,故根據

$$\operatorname{Log} Y=f(X)$$

式方程式作估量,而用 $S_{\log y}$ 以比率計量此等估量之可靠性,似爲合理之事.在此種情形中,前已指出,吾人以 $\rho_{\log y}$ 計量聯繫;而 $\rho_{\log y}$ 之值乃視曲線周圍之比率離中趨勢(Ratio variability)對於幾何中數周圍之比率離中趨勢之關係而定.(註)

(註) 華爾施(C. M. Walsh)氏"估量問題"(The problem of Estimation)一書中所論者,對於目前之問題特別適用.華氏引伽利略(Galileo)之言而擁護應用幾何中數以求估量值之平均數曰:"故一差誤之用以計量許多差誤者,必爲一估量與眞量間之比率,而非具體之量自身也.吾人不能以磅、尺、克郎計差誤;而必須以錯誤估量中之磅、尺、克郎之數對於估量事物之磅、尺、克郎之數之比例以計量之."此言更力證前文關於估量時應用對數函數及應用對數式估量差誤計量所言者之不誣.

參考書

Killough, H. B. A Statistical Analysis of Oat Prices 爲美國農業經濟局所作之稿。

Moore H. L. Elasticity of Demand and Flexibility of Prices. Journal of the American Statistical Association March, 1922.

Empirical Laws of Demand and Supply and the Flexibility of Prices. Political Science Quarterly, Dec., 1919.

Walsh, C. M. The Problem of Estimation (1—67)

下列二書中包含許多本章所討論之關係曲線之例

Schultz, Henry. The Statistical Measurement of the Elasticity of Demand for Beef. Journal of Farm Economics. July, 1924 (254—278)

Working, Holbook. Fastors Determining the Price of Potatoes in St. Paul and Minneapolis. Techuical Bulletin 10, University of Mimessta Agricuthual Experiment Station.

第十四章

關係之計量：複繫聯與偏繫聯

前數章中所處理之繫聯問題,僅包含兩個變數,一爲應變數,一爲單獨之自變數.在若干例中吾人曾見所研究之兩變數間,有頗爲高度之繫聯存在.惟一般言之,經濟現象不止受一個因子所影響,一個變數之變動可爲許多力量相互作用之結果;此乃顯然之事也.單獨處理兩個變數時其他一切因子皆置不問,此時常假定此單獨之自變數爲應變數變動之最重要之原因.(註) 如前舉苜蓿收成之例,吾人僅就一單獨之因子——灌溉——而察其加於產量之影響,惟所研究之各年中,其雨量與氣溫之變化必皆作用於產量也.同樣,實際上經濟分析所處理之一切因子,其變動皆有一個以上之原因.若欲使吾人之分析爲完全,必須應用種種方法,使能同時處理兩個以上之變數.吾人須要種種工具,藉以計量若干因子加於一單獨之變數之總影響.此種工具可將前述之方法

(註) 不應將繫聯係數誤解爲計量或確證因果關係之數字.請參考第十六章關於此點之論述.

少加展開而獲得之.

下表列 1890—1922 年堪薩斯邦每畝玉蜀黍產量與各年六月七月八月之平均氣溫.曾有一長期趨勢直線配合於生產數字，第(4)行中表示實在產量對於長期趨勢線上正則值之百分數.

表一一九

1890—1922 年玉蜀黍產量與氣溫(堪薩斯邦)(註1)

(1) 年份	(2) 實在產量，每畝蒲式數	(3) 長期趨勢值，產量，每畝蒲式數(註2)	(4) 實在值對長期趨勢值之百分比 X_1	(5) 六月平均氣溫 X_2	(6) 七月平均氣溫 X_3	(7) 八月平均氣溫 X_4
1890	15.6	22.4	69.6	77.6	83.1	76.1
1891	26.7	22.2	120.3	70.7	74.0	75.1
1892	24.5	22.1	110.9	73.4	77.5	76.5
1893	21.3	21.9	97.3	74.7	79.5	73.8
1894	11.2	21.8	51.4	74.2	77.8	78.0
1895	24.3	21.6	112.5	71.7	74.9	76.0
1896	28.0	21.5	130.2	74.1	78.1	78.7
1897	18.0	21.3	84.5	76.6	80.2	76.0
1898	16.0	21.2	75.5	75.0	77.7	78.2
1899	27.0	21.0	128.6	73.9	76.2	80.6
1900	19.0	20.9	90.9	74.9	77.9	81.0

(註1) 此玉蜀黍生產量材料錄自美國農業部公報 515 期，與該部之年報.氣溫之材料取自聯邦氣象局之報告.

(註2) 長期趨勢線之方程式為 $Y=19.907-1.502X$，以 1906 年為原點

表一一九(續上)

1890—1932年玉蜀黍產量與氣溫(堪薩斯邦)

(1) 年份	(2) 實在產量，每畝蒲數	(3) 長期趨勢值，產量，每畝蒲數(註 2)	(4) 實在值對長期趨勢值之百分比 X_1	(5) 六月平均氣溫 X_2	(6) 七月平均氣溫 X_3	(7) 八月平均氣溫 X_4
1901	7.8	20.7	37.7	77.3	85.0	79.1
1902	29.9	20.6	145.1	70.9	76.8	78.2
1903	25.6	20.4	125.5	67.2	78.3	75.3
1904	20.9	20.3	103.0	70.4	75.6	74.6
1905	27.7	20.1	137.8	75.5	74.5	78.7
1906	28.9	20.0	144.5	71.8	73.8	76.3
1907	22.1	19.8	111.6	72.0	78.4	78.1
1908	22.0	19.7	111.7	72.1	75.8	76.2
1909	19.9	19.5	102.1	73.1	78.1	80.1
1910	19.0	19.4	97.9	72.2	79.5	75.7
1911	14.5	19.2	75.5	80.5	78.6	76.4
1912	23.0	19.1	120.4	69.3	79.9	77.4
1913	3.2	18.9	16.9	74.2	82.1	84.2
1914	18.5	18.8	98.4	78.2	79.9	78.2
1915	31.0	18.6	166.7	69.2	74.0	70.1
1916	10.0	18.5	54.1	70.3	81.2	79.6
1917	13.0	18.3	71.0	72.8	80.8	73.4
1918	7.1	18.2	39.0	78.4	78.3	82.3
1919	15.2	18.0	84.4	72.3	80.2	78.3
1920	26.5	17.9	148.0	72.8	77.6	72.9
1921	22.2	17.7	125.4	74.4	79.2	78.6
1922	19.3	17.6	109.7	75.2	77.0	80.1

〔玉蜀黍產量與氣溫間之關係：初步分析〕

吾人知玉蜀黍之產量，受其生長季節氣溫所影響，現今研究之目的，為決定產量與所指出之三個月中每月氣溫之確切關係，以便得一基礎，以根據已知之氣溫估量其產量之大小。生長季節中各月之重要程度不同，故氣溫與產量之關係先宜分開研究，將三月中每月之關係，單獨決定之。

描述每畝產量與六月氣溫之關係之方程式如下

$$X_1=a+b_{12}X_2$$

描述每畝產量與七月氣溫之關係之方程式如下

$$X_1=a+b_{13}X_3$$

（兩式中之 X_1 皆代表實在產量對於長期趨勢產量之百分比關係，而 X_2, X_3……等代表絕對氣溫，以華氏表度數為單位。）以前諸例皆用 Y 與 X 代表變數，此地則不然，改為 X_1, X_2, X_3 等，以 X_1 代表應變數。第一式中代表斜度（響應係數）之常數，其記號為 b_{12}，右下角之1,2兩小字示此常數所指之變數，第一個小字經常指應變數（上面第一例中指 X_1），第二個小字指自變數（此地指 X_2），當問題中包含幾個變數之時，為分別不同之常數起見，此種小字乃屬必要。常數意義與以前諸例者完全相同，惟彼時所處理之變數僅有兩個，無須用小字分別之耳。

解適當之正則方程式求描述每畝產量（實在值對長期趨勢值之百分關係）與六月氣溫間平均關係之方程式中

之常數,則得

$$X_1=522.31-5.743X_2$$

S_{12}之值可由下列公式決定之

$$S_{12}^2=\frac{\Sigma(X_1^2)-a\Sigma(X_1)-b_{12}\Sigma(X_1X_2)}{N}$$

(S右下角之小字,與下面r右下角之小字,與前節所述者意義相同.)代入已知值,解之得

$$S_{12}^2=913\ 178$$

$$S_{12}=30.22$$

標準誤S,此計量根據關係方程式而作之估量之可靠性之數字,其意義前已詳言之矣.欲判斷關係方程式價值之大小,S_{12}須與σ_1(X_1之標準差)相比;此σ_1者,可視爲根據變數X_1之算術中數而作之估量之可靠性計量者也.茲求得其值爲

$$\sigma_1=34.477$$

由此觀之,根據關係方程式所作之估量,顯然比較根據算術中數而作者爲可靠.繫聯係數r用抽象之形式表示此關係.其值可由下列方程式決定之.

$$r_{12}^2=\frac{a\Sigma(X_1)+b_{12}\Sigma(X_1X_2)-Nc_1^2}{\Sigma(X_1^2)-Nc_1^2}$$

解方程式求r,並將b_{12}之正負號加於其前,得

$$r_{12}=-.4814$$

此等數值表示堪薩斯邦玉蜀黍每畝產量與六月氣溫間有一負繫聯存在,其程度不高.今試改用七月氣溫之數字,觀由此所作估量能否較為正確.

此種研究所須之諸數值,皆可由表一一九之材料計算而得.解而求標應方程式中之常數,得關係方程式如下.

$$X_1=827.64-9.302X_3$$

並求得標準誤為

$$S_{13}=24.73$$

更求得繫聯指數

$$r_{13}=-.6963$$

此地求得之關係較為密切,而得一較佳之根據以供估量,勝於前者應用六月氣溫之時.

將同樣方法施諸每畝生產量與八月氣溫,結果得

$$X_1=517.86-6.098X_4$$

$$S_{14}=29.98$$

$$r_{14}=-.4937$$

當然,八月之氣溫對於堪薩斯邦玉蜀黍產量亦有影響,氣溫低可使產量高於正則產量.關係固不若七月氣溫之大,惟仍為重要,現今所須者為結合此三因子之方法,俾可知此等因子加於玉蜀黍產量之總影響,以為估量之根據.吾人不能將三月之氣溫相加或平均,因七月氣溫之影響顯然較其他兩月為重要也.完成此目的所用之方法,其原則固為簡單.

[從三個自變數估計玉蜀黍之產量]

在在現之情形中,估量方程式或響應方程式將包含一單獨之應變數(玉蜀黍產量)與三個自變數.其形式爲

$$X_1 = a + b_{12\cdot34}X_2 + b_{13\cdot24}X_3 + b_{14\cdot23}X_4$$

若吾人能決定四個常數之數值,吾人可代入 X_2, X_3, X_4 之已知值於此方程式,而得 X_1 之一個估量值,其法正如處理兩個變數時所用者.所須之常數值可用最小平方法求之.

關係記號須加若干說明.方程式原極簡單,僅因記號右下角之許多小字,與人一繁冗之印象. b_{12} 之一記號其意義前曾說明.所代表者爲當僅有 X_1 X_2 兩個變數時 X_1 對 X_2 之響應係數(即 X_1 爲應變數時描述兩者間關係之線之斜度. $b_{12\cdot34}$ 一記號所代表者爲 X_1 對 X_2 之純響應係數(Coefficient of net regression). 更於小點後面加註3,4兩字者,僅表示 X_3, X_4 兩變數亦加入研究範圍之內,就此一常數 ($b_{12\cdot34}$) 言,彼兩變數之影響業已除去.此常數表示在根據 X_2, X_3, X_4 三個變數而作 X_1 之估量時必須加於 X_2 之權數.當然此數不能與 b_{12} 相同,因 b_{12} 所表示者爲單獨根據 X_2 而作 X_1 之估量時應加於 X_2 之權數也.同樣,常數 $b_{13\cdot24}$ 爲 X_1 對 X_3 之純響應係數,表示 X_2 及 X_4 同在研究範圍之內時加於 X_3 之權數.每一係數所代表者皆爲一簡單而單獨之常數,爲使此常數之確切意義明瞭起見,故必須加註小字耳.在小點前之小字稱爲主註(Primary subscript),在其後之小字稱爲副註(Secondary subscript),

[正則方程式之構成與解法]

第一步工作,(註1)爲求得必須之正則方程式,解之而求上列估量方程式中常數之值,用平常之方法,(註2)得正則方程式如下:

$$\text{I}\quad \Sigma(X_1)=Na+b_{12\cdot34}\Sigma(X_2)+b_{13\cdot24}\Sigma(X_3)+b_{14\cdot23}\Sigma(X_4)$$

$$\text{II}\quad \Sigma(X_1X_2)=a\Sigma(X_2)+b_{12\cdot34}\Sigma(X_2^2)+b_{13\cdot24}\Sigma(X_2X_3)+b_{14\cdot23}\Sigma(X_2X_4)$$

$$\text{III}\quad \Sigma(X_1X_3)=a\Sigma(X_3)+b_{12\cdot34}\Sigma(X_2X_3)+b_{13\cdot24}\Sigma(X_3^2)+b_{14\cdot23}\Sigma(X_3X_4)$$

$$\text{IV}\quad \Sigma(X_1X_4)=a\Sigma(X_4)+b_{12\cdot34}\Sigma(X_2X_4)+b_{13\cdot24}\Sigma(X_3X_4)+b_{14\cdot23}\Sigma(X_4^2)$$

本可將已知值代入此等聯立方程式,直接解之而求四常數之值.惟尚有法減去正則方程式之一,因而大減計算之繁.此法以每一變數對其算術中數之差離代替其絕對數值,因此除去原來方程式中之一常數項a.

令A_1, A_2, A_3等代表各變數之算術中數,x_1, x_2, x_3等代表各變數對於中數之差離,吾人乃可以$(x_1+A_1), (x_2+A_2), (x_3+A_3)$等代替$X_1, X_2, X_3$等絕對值.如此代入正則方程式後,可化簡至消去正則方程式中第一式,其餘各式變爲下式:

(註 1) 此地所用處理複繫聯問題之方法,係依因託萊氏(H. R. Tolley)及愛茲基耳氏(M. J. B. Ezekiel)氏所用者.(見"A Method of Handline Multiple Correlation Probrems," Journal of the American Statistical Assocition, Dec, 1923, 993—1003).兩位先生所創作之簡單而有效方法,使統計者受惠不淺.關於正則方程式之化簡一事,此地不詳說,學者欲知詳細,可觀讀該原文.

(註 2) 參觀附錄一,關於此方法之討論.

$$\frac{\Sigma(x_1x_2)}{N}=\frac{\Sigma(x_2^2)}{N}b_{12\cdot34}+\frac{\Sigma(x_2x_3)}{N}b_{13\cdot24}+\frac{\Sigma(x_2x_4)}{N}b_{14\cdot23}$$

$$\frac{\Sigma(x_1x_3)}{N}=\frac{\Sigma(x_2x_3)}{N}b_{12\cdot34}+\frac{\Sigma(x_3^2)}{N}b_{13\cdot24}+\frac{\Sigma(x_3x_4)}{N}b_{14\cdot23}$$

$$\frac{\Sigma(x_1x_4)}{N}=\frac{\Sigma(x_2x_4)}{N}b_{12\cdot34}+\frac{\Sigma(x_3x_4)}{N}b_{13\cdot24}+\frac{\Sigma(x_4^2)}{N}b_{14\cdot23}$$

上列各方程式中一切變數,皆指原來變數對於各自算術中數之差離言.故$\frac{\Sigma(x_1x_2)}{N}$僅爲x_1與x_2兩變數之平均乘積(mean product),$\frac{\Sigma(x_2^2)}{N}$爲σ_2^2,諸如此類,如以p_{12},p_{13}等等記號代表各平均乘積,並將標準差平方之記號代入,則正則方程式爲:

$$p_{12}=\sigma_2^2b_{12\cdot34}+p_{23}b_{13\cdot24}+p_{24}b_{14\cdot23}$$

$$p_{13}=p_{23}b_{12\cdot34}+\sigma_3^2b_{13\cdot24}+p_{34}b_{14\cdot23}$$

$$p_{14}=p_{24}b_{12\cdot34}+p_{34}b_{13\cdot24}+\sigma_4^2b_{14\cdot23}$$

此爲正則方程式之最易解之形式

由表一一九中排列之材料,可求得下列數值:

$\Sigma(X_1)=3298.1$ $\Sigma(X_1^2)=368846.67$

$\Sigma(X_2)=2426.9$ $\Sigma(X_2^2)=178755.75$

$\Sigma(X_3)=2581.5$ $\Sigma(X_3^2)=202163.79$

$\Sigma(X_4)=2553.8$ $\Sigma(X_4^2)=197890.32$

$$\Sigma(X_1X_2)=240967.22$$

$$\Sigma(X_1X_3)=255954.11$$

$$\Sigma(X_1X_4)=253664.85$$

$$\Sigma(X_2X_3)=189941.83$$

$$\Sigma(X_2X_4)=187909.38$$

$$\Sigma(X_3X_4)=199845.00$$

$$c_1=\frac{\Sigma(X_1)}{N}$$

$$=99.9424 \qquad c_1^2=9988.4833$$

$$c_2=73.5424 \qquad c_2^2=5408.4846$$

$$c_3=78.2273 \qquad c_3^2=6119.5105$$

$$c_4=77.3879 \qquad c_4^2=5988.8871$$

解正則方程式所須數量,根據上列數值極易求出.茲將此等數量列於下面.

$$\sigma_1^2=\frac{\Sigma(X_1^2)}{N}-c_1^2$$

$$=\frac{368,848.67}{33}-9988.4833=1188.688$$

$$\sigma_2^2=\frac{178,755.75}{33}-5408.4846=8.3564$$

$$\sigma_3^2=\frac{202,163.79}{33}-6119.5105=6.6645$$

$$\sigma_4^2=\frac{197,890.32}{33}-5988.8871=7.7893$$

$$p_{12}=\frac{\Sigma(X_1X_2)}{N}-c_1c_2$$

$$=\frac{240,967.22}{33}-(99.9424\times73.5424)=-47.967$$

$$p_{13}=\frac{255,954.11}{33}-(99.9424\times78.2273)=-62.039$$

$$p_{14}=\frac{253,664.85}{33}-(99.9424\times77.3879)=-47.519$$

$$p_{23}=\frac{189,941.83}{33}-(73.5424\times78.2273)=2.790$$

$$p_{24}=\frac{187,909.38}{33}-(73.5424\times77.3879)=2.932$$

$$p_{34}=\frac{199,845.00}{33}-(78.2273\times77.3879)=2.063$$

代入正則方程式得

$$-47.967=8.3564b_{12\cdot34}+2.790\ b_{13\cdot24}+2.932\ b_{14\cdot23}$$

$$-62.039=2.790\ b_{12\cdot34}+6.6645b_{13\cdot24}+2.063\ b_{14\cdot23}$$

$$-47.519=2.932\ b_{12\cdot34}+2.063\ b_{13\cdot24}+7.7893b_{14\cdot23}$$

解上列聯列方程式(註)得下列諸常數值:

(註) 任何解法皆可應用,惟其中最便易者當有三個或三個以上之方程式時,或爲杜里脫留氏法(Doolittle method)詳見附錄一.

$b_{12\cdot34}=-2.095 \quad b_{13\cdot24}=-7.394 \quad b_{14\cdot23}=-3.354$

故所求之方程式爲

$$x_1=-2.095x_2-7.394x_3-3.354x_4$$

此爲x_1對$x_2\,x_3\,x_4$之純響應之方程式代入三個自變數(六月氣温,七月氣温,八月氣温)之任何已知數值,即可决定應變數(每畝玉蜀黍之產量)之或然性最大之數值.吾人須注意在現今之方程式中,一切變數皆以某對於各自算術中數之差離表示,在實際應用上最好有一以原來數值表示之方程式.換言之,即最好將原點由平均數之點移至原來尺度之零點.欲如此,必須重新將常數項a加入.

a之數值可由下列方程式决定之.

$$A_1=a+A_2b_{12\cdot34}+A_3b_{13\cdot24}+A_4b_{14\cdot23}$$

其中大A表示各變數之算術中數.(註1)代入適當數值,得

$$99.9424=a+(73.5424\times-2.094816)$$
$$+(78.2273\times-7.39374)+(77.3879\times-3.353796)\text{(註2)}$$

(註1) 此方程式由第491頁上第一個方程式化出,其式原爲 $\Sigma(X_1)=Na+b_{12\cdot34}\Sigma(X_2)+b_{13\cdot24}\Sigma(X_3+b_{14\cdot23}\Sigma(X_4)X_1\,X_2$等絕對數,用與之相當之$x_1+A_1,x_2+A_2,$等代替之,得

$$\Sigma(x_1)+NA_1=Na+b_{12\cdot34}[\Sigma(x_2)+NA_2]+b_{13\cdot24}[\Sigma(x_3)+NA_3]$$
$$+b_{14\cdot23}[\Sigma(x_4)+NA_4]$$

因$\Sigma(x_1)=0\,\Sigma(x_2)=0$,諸如此類,故此等變數可消去,全體各項各除以$N$,得公式如前.

(註2) 因假定原點各在原來尺度上之0點,故$A_1=c_1A_2=c_2$等等.欲使a之值更爲精確,故係數$b_{12\cdot34}b_{13\cdot24}$等之小數較在響應方式中多取兩位.

解之得

$$a=1091.94$$

故用原來數值表示之響應方程式爲,

$$X_1=1091.94-2.095X_2-7.394X_3-3.354X_4$$

[估量標準誤之計算]

根據此方程式所作之估量,是否較根據以前僅包括一個自變數之方程式所作之估量爲可靠?欲答復此問題,須計算標準誤之值.此地之標準誤須以$S_{1\cdot234}$之記號表示.小字指一單獨之應變數(X_1)與三個自變數.此標準誤之值可由下列公式計算而得(註)

$$S^2_{1\cdot234}=\sigma_1{}^2-b_{12\cdot34}p_{12}-b_{13\cdot24}p_{13}-b_{14\cdot23}p_{14}$$

(註) 此公式之來由如下.已知一方程式爲下式

$$x_1=b_{12\cdot34}x_2+b_{13\cdot24}x_3+b_{14\cdot23}x_4$$

(式中變數指對於中數之差離),則每一剩餘量(residual)可由下列方程式計算之

$$d=b_{12\cdot34}x_2+b_{13\cdot24}x_3+b_{14\cdot23}x_4-x_1 \cdots\cdots\cdots\cdots(1)$$

全部乘以d而綜加之,得

$$\Sigma(d^2)=b_{12\cdot34}\Sigma(dx_2)+b_{13\cdot24}\Sigma(dx_3)+b_{14\cdot23}\Sigma(dx_4)-\Sigma(dx_1)$$

但根據配合之方法,

$$\Sigma(dx_2)=0$$

$$\Sigma(dx_3)=0$$

$$\Sigma(dx_4)=0$$

故 $\Sigma(d^2)=-\Sigma(dx_1) \cdots\cdots\cdots\cdots(2)$

以x_1乘每一個餘量方程式(residual equation)(1),綜加之得

$$\Sigma(dx_1)=b_{12\cdot34}\Sigma(x_1x_2)+b_{13\cdot24}\Sigma(x_1x_3)+b_{14\cdot23}\Sigma(x_1x_4)-\Sigma(x_1{}^2)$$

將$\Sigma(dx_1)$之相當值代入方程式(2),得

$$\Sigma(d^2)=\Sigma(x_1^2)-b_{12\cdot34}\Sigma(x_1x_2)-b_{13\cdot24}\Sigma(x_1x_3)-b_{14\cdot23}\Sigma(x_1x_4)$$

$$S^2_{1\cdot234}=\frac{\Sigma(d^2)}{N}=\frac{\Sigma(x_1^2)}{N}-b_{12\cdot34}\frac{\Sigma(x_1x_2)}{N}-b_{13\cdot24}\frac{\Sigma(x_1x_3)}{N}-b_{14\cdot23}\frac{\Sigma(x_1x_4)}{N}$$

因變數指數對於中數之差離而言,得

$$S^2_{1\cdot234}=\sigma_1^2-b_{12\cdot34}p_{12}-b_{13\cdot24}p_{13}-b_{14\cdot23}p_{14}$$

標準誤亦可由下列普通式之方程式求之,以原來尺度之0點為原點,

$S_{1\cdot234\cdots\cdots n}$

$$=\frac{\Sigma(X_1^2)-a\Sigma(X_1)-b_{12\cdot34\cdots n}\Sigma(X_1X_2)-b_{13\cdot24\cdots n}\Sigma(X_1X_3)-b_{14\cdot23\cdots n}\Sigma(X_1X_4)-\cdots}{N}$$

此式之證明見附錄一.

代入適當數值,得

$$S^2_{1\cdot234}=1188.688-100.482-458.700-159.369$$

$$=470.137$$

$$S_{1\cdot234}=21.68$$

此數之解釋,與以前各例標準誤之解釋相同.根據X_1之算術中數而作之估量,其可靠性以σ_1計量之,其值為344.77.根據純響應方程式而作之估量,當產量視為六、七、八三月氣溫之函數時,其可靠性以$S_{1\cdot234}$計量之,其值為21.68. 由此可見根據方程式而作之估量,較之單根據X_1之材料而作者大為可靠.吾人固尚未將引起玉蜀黍產量變動之一切因子盡加計及,然已就影響堪薩斯邦玉蜀黍每畝產量之三個因子之結果而計量之,且以確切之言辭表示之矣.

[複繫聯係數]

現今吾人須一第三種之計量,即抽象之繫聯係數是也.此係數之值,如吾人前此所述,視 S 與 σ 之關係而定.現今可由下列公式計算之.

$$R^2_{1\cdot 234}=1-\frac{S^2_{1\cdot 234}}{\sigma_1^2}$$

當所研究者爲一個單獨之應變數與幾個自變數之間之關係時,此計量稱爲複繫聯係數(Coefficient of multiple correlation)而用 R 爲記號以代表之.R 右下角之小註,在小點左者指應變數,在小點右者指自變數,將 $S^2_{1\cdot 234}$ 之相當值代入此公式,得

$$R^2_{1\cdot 234}=1-\frac{\sigma_1^2-b_{12\cdot 34}p_{12}-b_{13\cdot 24}p_{13}-b_{14\cdot 23}p_{14}}{\sigma_1^2}$$

此式可化爲(註)

$$R^2_{1\cdot 234}=\frac{b_{12\cdot 34}p_{12}+b_{13\cdot 24}p_{13}+b_{14\cdot 23}p_{14}}{\sigma_1^2}$$

(註) 複繫聯係數亦可由以原來尺度上零點爲原點之普通公式求之,其式如下.

$$R_{1\cdot 234\cdots n}=\frac{a\Sigma(X_1)+b_{12\cdot 34\cdots n}\Sigma(X_1X)_2+b_{13\cdot 24\cdots n}\Sigma(X_1X_3)+b_{14\cdot 23\cdots n}\Sigma(X_1X_4)+\cdots-Nc_1^2}{\Sigma(X_1)^2-Nc_1^2}$$

代入適當之數值，得

$$R^2_{1\cdot234}=\frac{100.482+458.700+159.369}{1188.688}$$

$$R_{1\cdot234}=.778$$

複繫聯係數者，表示一個單獨之應變數與若干自變數之總體間之關係程度之指數也。彼計量應變數之變動依於其他因子之共同作用之程度爲如何。吾人若視一切自變數合成一個單獨之自變數列，則此係數之意義自易明瞭。於是係數乃成爲應變數與自變數列間關係之計量，恰如前此在較簡之事例中成爲兩變數間之繫聯計量也。在複繫聯中，自變數列有幾個構成因子，惟此事不能改變係數之基本意義。因注意者，R之前未加正號或負號。在現今之一例中，所有自變數皆與玉蜀黍產量成負繫聯，固可以在R前加一負號，惟有時對若干自變數爲正繫聯，對其他若干自變數又爲負繫聯，因此R經常不標正負號。欲知自變數中何者與應變數成正繫聯，何者成負繫聯，可於純響應方程式中常數之正負號見之。

[各種關係計量之比較]

綜合諸因子而作之研究，使吾人關係玉蜀黍產量變動之原因之知識，以及吾人所作估量之可靠性，究竟增進至若何程度乎？若集由此分析過程中所得之各種關係計量列於一表，則此問題將較易答復。

表一二〇

關於堪薩斯邦玉蜀黍產量之若干計量之比較

估量之根據	估量之可靠性之計量	繫聯係數
X_1之算術中數	$\sigma_1=34.477$	
$X_1=522.31-5.743X_2$	$S_{12}=30.22$	$r_{12}=-.4814$
$X_1=827.64-9.302X_3$	$S_{13}=24.73$	$r_{13}=-.6968$
$X_1=517.86-6.098X_4$	$S_{14}=29.98$	$r_{14}=-.4937$
$X_1=1091.94-2.095X_2-7.394X_3-8.854X_4$	$S_{1\cdot234}=21.68$	$R_{1\cdot234}=.778$

若分析中更加入其他因子,如生長季節之雨量等,則S之值猶可減小,而R之值因而增大.以上所說明之方法儘可擴充至包含任何個數之變數,每增加一變數,聯立方程式中亦增加一個方程式.

[複繫聯法惟在直線關係始爲有效]

前面討論過程中,有一重要之條件未曾特別說明,即此複繫聯法之有效無效,須視每一對變數間是否成直線關係爲斷,如問題之包括四變數者,可有六對變數(即須求六個平均乘積也).若在此六個關係中,有任何一個與直線形式大有出入,所得之結果,其價值即減低,在此種情形下應用所求得之方程式,亦未可謂謬誤,惟以此方程式作爲估量之根據時,不能如方程式之能計及真正關係者之爲佳耳.遇此種情形,S與R之數值,將表示估量之根據直線關係之假定而作

者,不十分可靠(註)

[複繫聯法應用之例]

茲舉例說明估量方程式之應用,1922年堪薩斯邦六月之平均氣溫爲華氏75.2°,七月平均氣溫爲77°,八月平均氣溫爲81°,則玉蜀黍每畝之或然產量爲何?將此等數值代

$$X_1=1091.94-2.095X_2-7.394X_3-3.354X_4$$

方程式中之X_2, X_3, X_4,卽得

$$X_1=1091.94-(2.095\times75.2)-(7.394\times77.0)-(3.354\times80.1)$$

$$X_1=96.4$$

此結果將實在值作爲對於正則之百分數而表示.玉蜀黍產量之長期趨勢方程式爲

$$Y=19.967-.1502X$$

(式中Y代表產量,X代表同時間因子,以1906年爲原點).由此方程式求得1922年正則產量爲每畝17.6斛.估量之產量爲此數之96.4%,或每畝17斛.

吾人將預期實在值位於此估量值周圍之何種界限以內乎? $S_{1.234}$ 之值爲21.68.卽謂100次機會中有68次將爲實在產量比率位於(96.4+21.68)與(96.4−21.68)之間,卽118.1與74.7之間也.此等數值皆指實在產量對正則產量之百分

(註) 在附屬數列間之關係非直線形式時處理複繫聯問題之方法,愛茲基耳氏曾有說明載於"A Method of Handling Curvilinear Correlation for any Number of Variables"一文.見 Journal of the American Statistical Association Vol. XIX. N. S. NO. 148, 1924.

數而言.既知正則產量爲17.6,此等界限因而可化作用斛爲單位之數量.如是得界限之値爲20.8斛與13.1斛,100次機會中有68次實在產量應在此界限之內.1922年之實在產量爲每畝19.3斛.

在此說明中,所用之年(1922)本在研究範圍之內,同樣之方法亦可用於作將來若干年之估量,惟此時原有之不確定性質,猶將因長期趨勢線之延長而多加一重耳.

[偏繫聯或純繫聯之意義]

前節將三個自變數綜合處理,以決定六月七月八月三個月之氣溫影響堪薩斯邦玉蜀黍產量之程度,吾人以計量三者合力作用於產量之結果爲目的.尚有一相關之問題,在許多研究中可占重要地位者,即決定一應變數與一單獨之自變數當將其他一切因子皆視爲固定不變時兩者間之關係是也.具體言之,若七月與八月之氣溫以及其他一切影響產量之因子皆可視爲固定不變,則七月氣溫之變動加於生產量之影響爲如何乎?此爲純繫聯或偏繫聯問題(Problem of net or partial correlation)

若能創一方法,使一個單獨之因子離其他因子而獨立以便於分別研究,則經濟學家與一般社會科學家之分析力量必大受其惠,此乃顯然之事也.此方面之學者,將因此而具有一種力量,例如化學家所具者,以消除不相關之影響而集中注意於一單獨之因子.當化學家研究一原素加於另一原

素之影響時,先設法除去其他一切原素,而其分析之效率,大部分視可能使當前之目的物如此孤立之程度而定.

在經濟分析中,欲在引起某一數列變動之諸因子中僅取其一而消除其他一切因子,並非常爲可能之事,一個經濟現象之直接與間接的原因,爲數過多,相互作用過於複雜,故經濟學家在簡化自己之問題爲兩個變數之工作上,決不能與化學家爭勝.但在一定限度以內,統計家亦能運用自然科學家所用之方法,使若干因子固定不變,而研究其他一因子變動之結果.達此目的之方法,乃研究社會科學者手中最有力工具之一也.

偏繫聯法可就堪薩斯邦玉蜀黍產量問題說明之.現今之目的在決定產量與三個月中每一個月之氣温間之純繫聯,該三月之平均氣温爲吾人所已知者.

[偏繫聯與簡單繫聯之分別]

偏繫聯之問題與平常計算兩變數間之關係時所遇之問題必須分別清楚.吾人前面已求得一描述玉蜀黍產量與七月氣温間平均關係之方程式以及標準誤及繫聯係數:

$$X_1=827.64-9.302X_3$$

$$S_{13}=24.73$$

$$r_{13}=-.697$$

當一切其他因子皆置而不論時,此等計量即所以描述問題中之關係者.其他一切因子未被計及,祇爲略去.宛如化學家

研究一原素對另一原素之反應時,試驗管中有各種雜質存在,而彼不加以清除也.經濟學家平常不能指明問題中一切"雜質"之所在而與以清除,惟其所求得之計量爲出於此種未校正之材料一事,乃彼等所應認識者也.

[偏繫聯法]

決定玉蜀黍產量及七月氣溫間之偏繫聯,即求一在其他一切因子若爲不變時兩者間之繫聯計量.吾人對曾加研究之其他因子絲毫不苟,惟欲求一關於玉蜀黍產量之計量,能專門反映七月氣溫之變動者耳.

欲達此目的,或可提出一種可能之方法:即吾人若能得,時期極長之材料時,可選取六月八月之氣溫不變之若干年.如吾人能得三十年之材料,各年之六月氣溫平均皆爲 74°,八月氣溫平均皆爲 78°,而在此各年中,玉蜀黍產量與七月氣溫各各不同,於是可求七月氣溫與玉蜀黍產量間之關係,蓋吾人可確知所得結果未受六月氣溫與八月氣溫變動之影響也.不幸此種使若干因子固定不變之方法不能應用.一般上材料過於有限而過於不一致,使吾人無從由其中選出適於目的之此種數字.惟尚有其他方法可以達到此目的者.

[變應係數與繫聯係數間之關係]

未談此方法以前,請先暫時回至簡單繫聯之問題上.吾人已知描述兩變數間關係之方程式可以有兩種形式表示,一以 y 爲應變數,一以 x 爲應變數.此等方程式,將變數用其

對各自算術中數之差離表現時,其式如下

$$y=r\frac{\sigma_y}{\sigma_x}x$$

與

$$x=r\frac{\sigma_x}{\sigma_y}y$$

在此等方程式中,$r\frac{\sigma_y}{\sigma_x}$與$r\frac{\sigma_x}{\sigma_y}$兩值稱爲響應係數 (Coefficient of regression),代表兩條響應線之斜度.繫聯係數等於此兩響應係數相乘之積之方根,即

$$r=\sqrt{r\frac{\sigma_y}{\sigma_x}\cdot r\frac{\sigma_x}{\sigma_y}}$$

用平常之記號表示之,得

$$r_{yx}=\sqrt{b_{yx}\cdot b_{xy}}$$

式中b_{yx}爲y對x之響應係數(y爲應變數),b_{xy}爲x對y之響應係數(x爲應變數),若用x_1與x_2代替y與x,加註之小字亦作相當之改變,則得下式

$$r_{12}=\sqrt{b_{12}\cdot b_{21}}$$

注意,r之值不論x_1或x_2爲應變數,爲數不變,即$r_{12}=r_{21}$也.

此引吾人立即入於目前問題之解決.吾人所求者爲其他因子固定不變時兩變數間關係之計量,即兩變數之偏繫聯係數是也.惟在研究複繫聯時吾人已求得若干偏繫聯係

數.響應方程式之描述一單獨之應變數與若干自變數間之關係者,其式如下:

$$X_1=a+b_{12.34}X_2+b_{13.24}X_3+b_{14.23}X_4$$

$b_{12.34}$, $b_{13.24}$, 及 $b_{14.23}$ 諸值,皆偏響應係數也.此等數值所指者爲同時根據三個自變數而作估量時每一自變數所應得之權數.如係數 $b_{13.24}$ 爲同時計及 X_2, X_3 與 X_4 三變數而作 X_1 之估量時應加於 X_3 之權數.此三係數之數值,可用解聯立方程式之簡單方法求得之,吾人前已見之矣.

惟用此法以求現今所求之偏繫聯係數時,尚須其他數值.係數 r_{yx} 爲 $b_{yx}.b_{xy}$ 之方根,係數 r_{13} 爲 $b_{13}.b_{31}$ 之方根.同樣,現今所求之係數,所以計量在 X_2(六月氣溫)與 X_4(八月氣溫)固定不變時 X_1(玉蜀黍產量)與 X_3(七月氣溫)間關係程度者,可由下列公式求之:

$$r_{13\cdot24}=\sqrt{b_{13\cdot24}.b_{31\cdot24}}$$

此方程式中 $b_{13\cdot24}$ 爲 X_1 對 X_3 之偏繫聯係數,而 $b_{31\cdot24}$ 爲 X_3 對 X_1 之偏繫聯係數.

此等記號之意義,在討論之過程中已經說明矣.小字所指者爲各數值所代表之特殊關係爲何,點左之兩小字(主註)指該值所專指之兩變數,點右之兩小字(副註)指爲專門比較前兩變數起見視爲不變之其他諸變數,在此地此種變數有二,然其個數固可爲一或任何其他數目也.於是求 X_1 與 X_3 兩變數間偏繫聯係數之普遍公式可寫爲

$$r_{13\cdot2456\cdots n}=\sqrt{b_{13\cdot2456\cdots n}\cdot b_{31\cdot2456\cdots n}}$$

主註中兩小字,前者指應變數,後者指自變數,

在目前問題中,欲求$r_{13\cdot24}$,僅須再求一個數值.係數$b_{13\cdot24}$之值在前節說明中已得之矣.而$b_{31\cdot24}$此響應係數則爲下列方程式中之常數之一:

$$X_3=a+b_{31\cdot24}X_1+b_{32\cdot14}X_2+b_{34\cdot12}X_4$$

此方程式者,以七月氣溫爲應變數,以六月氣溫八月氣溫及玉蜀黍產量爲自變數,而表示其間之平均關係者也.若已知X_1, X_2與X_4爲某某諸數值,吾人即可用此公式求得與此等數值相伴之或然性最大之七月氣溫.(當關係爲如此表示時,其所表示者寧爲伴聯Association而非依從之關係 Relation of dependence)

〔解正則方程式〕

所須之數值,其求法與前例完全相同.照舊用四個正則方程式,而四個又可化爲三個,其式如下

$$p_{13}=\sigma_1^2b_{31\cdot24}+p_{12}b_{32\cdot14}+p_{14}b_{34\cdot21}$$

$$p_{23}=p_{12}b_{31\cdot24}+\sigma_2^2b_{32\cdot14}+p_{24}b_{34\cdot21}$$

$$p_{34}=p_{14}b_{31\cdot24}+p_{24}b_{32\cdot14}+\sigma_4^2b_{34\cdot21}$$

式中平均乘積與標準差之值,除σ_1^2以外皆爲前次解方程式時所用者.將各值代入上列方程式,得

$$-62.039=1188.688b_{31\cdot24}-47.967b_{32\cdot14}-47.519b_{34\cdot21}$$

$$2.790=-47.967b_{31\cdot24}+8.3564b_{32\cdot14}+2.932b_{34\cdot21}$$

$$2.063=-47.519b_{31\cdot24}+2.932b_{32\cdot14}+7.7893b_{34\cdot21}$$

解方程得下列諸値

$$b_{31\cdot24}=-.0531$$

$$b_{32\cdot14}=+.0574$$

$$b_{34\cdot21}=-.0807$$

[偏繫聯係數之計算]

由以前之一例得

$$b_{13\cdot24}=-7.394$$

代入公式

$$r_{13\cdot24}=\sqrt{b_{13\cdot24}\times b_{31\cdot24}}$$

得

$$r_{13\cdot24}=\sqrt{-7.394\times-.0531}$$

$$=-.6266$$

(在單獨處理兩個變數之簡單之例中,r以擧應係數之正號或負號爲記號;此地情形亦同若兩個擧應係數皆爲正,r亦爲正;若兩者皆負,r亦爲負.)

此數卽堪薩斯邦玉蜀黍產量與七月氣溫間之偏繫聯係數.所計量者爲當六月氣溫與八月氣溫不變時此兩變數間之關係程度.

[計算偏繫聯係數之另一種方法]

若處理之變數有許多個數,則吾人可作成許多此種係數,成一整個數列,當計算許多個此種係數時,可用一與前述者少許不同之方法,其法對於系統排列上有若干便利之處.

僅關係於兩個變數之簡單繫聯係數,稱為零次係數(Coefficient of zero order).此種係數以r_{12}, r_{24}等等形式之記號代表之.關係於兩個變數之偏繫聯係數,其外另有一單獨之變數視為固定不變者,稱為一次係數(Coefficient of first order)以$r_{12\cdot3}$, $r_{24\cdot3}$等記號代表之.同樣尚有二次,三次,四次直至n次係數;當計量一單獨之應變數與一單獨之自變數間之關係時,視為不變之變數尚有若干,求得之係數即稱為若干次係數.

每一偏繫聯係數皆可由較之低一次之係數求得.如一次係數可由下列關係式求得之

$$r_{12\cdot3}=\frac{r_{12}-r_{13}\cdot r_{23}}{(1-r^2_{13})^{\frac{1}{2}}(1-r^2_{23})^{\frac{1}{2}}},$$

求二次係數之式為

$$r_{12\cdot34}=\frac{r_{12\cdot3}-r_{14\cdot3}\cdot r_{24\cdot3}}{(1-r^2_{14\cdot3})^{\frac{1}{2}}(1-r^2_{24\cdot3})^{\frac{1}{2}}}$$

故任何次偏繫聯係數之普通方程式為

$$r_{12\cdot345\cdots n}=\frac{r_{12\cdot345\cdots(n-1)}-r_{1n\cdot345\cdots(n-1)}\,r_{2n\cdot345\cdots(n-1)}}{(1-r^2_{1n\cdot345\cdots(n-1)})^{\frac{1}{2}}(1-r^2_{2n\cdot345\cdots(n-1)})^{\frac{1}{2}}}$$

故由零次繫聯係數出發,可逐級求其較高一次之係數.若純用算術方法推算,工作將甚繁冗,惟有若干現成之表可使計算之繁減少至最小限度.(註) 此法可用前一問題之材料說明之.

現今吾人須求三個次係數,即 $r_{12\cdot34}$, $r_{13\cdot24}$, 及 $r_{14\cdot23}$ 是也.此三係數將爲玉蜀黍產量與三個重要月份中每月氣溫間之偏繫聯計量.計算第一個係數之公式已列於上,求第二個係數者,爲

$$r_{13\cdot24}=\frac{r_{13\cdot2}-r_{14\cdot2}r_{34\cdot2}}{(1-r^2_{14\cdot2})^{\frac{1}{2}}(1-r^2_{34\cdot2})^{\frac{1}{2}}}$$

而求第三個係數之公式爲:

$$r_{14\cdot23}=\frac{r_{14\cdot2}-r_{13\cdot2}r_{43\cdot2}}{(1-r^2_{13\cdot2})^{\frac{1}{2}}(1-r^2_{43\cdot2})^{\frac{1}{2}}}$$

但三個係數亦可由另一套公式求之.其中一次係數之結合與此少異.其式如下:

$$r_{12\cdot34}=\frac{r_{12\cdot4}-r_{13\cdot4}r_{23\cdot4}}{(1-r^2_{13\cdot4})^{\frac{1}{2}}(1-r^2_{23\cdot4})^{\frac{1}{2}}}$$

$$r_{13\cdot24}=\frac{r_{13\cdot4}-r_{12\cdot4}r_{32\cdot4}}{(1-r^2_{12\cdot4})^{\frac{1}{2}}(1-r^2_{32\cdot4})^{\frac{1}{2}}}$$

$$r_{14\cdot23}=\frac{r_{14\cdot3}-r_{12\cdot3}r_{42\cdot3}}{(1-r^2_{12\cdot3})^{\frac{1}{2}}(1-r^2_{42\cdot3})^{\frac{1}{2}}}$$

(註) 參考 J. R. Miner, Tables of $\sqrt{1-r^2}$ and $1-r^2$ for use in Partial Cotion and in Trigonometry. Johns Hopkins Press, Baltimore, Md. 1922.

計算二次係數時每個若皆用兩種方法，則所得結果可彼此比較，以覆驗計算中有無錯誤。

[一次係數之計算]

二次係數須在必要之一切一次係數皆已求得後始可計算，求一次係數時所必須之方程式有如下式者

$$r_{12\cdot3}=\frac{r_{12}-r_{13}\cdot r_{23}}{(1-r^2_{13})^{\frac{1}{2}}\,(1-r^2_{23})^{\frac{1}{2}}},$$

可依照偏繫聯係數之普遍公式作成。須求之一次繫數既有多個，吾人宜用一有系統之方法計算之。下頁所載表式使計算便捷並適於利用前面言及之現成表格。

計算每個一次係數之手續甚簡單。每一計算中所必須之零次係數有三個，此三數須排列表中，其先後之次序依照其在所求之一次係數之公式中作爲分子時之次序。求此分數式中分子之法，爲由第一個零次係數減去第二，三兩個零次係數之乘積。此乘積項之數值在表上列爲一行。分數式中之分母爲$\sqrt{1-r^2}$式之兩項之乘積，此兩項乃由第二與第三個零次係數所化出者，表式之排列可使計算之進行系統化。

表一二一

示一次偏繫聯係數之計算法

堪薩斯邦玉蜀黍產量與氣溫

〇次之 r						一次之 r	
符號下加註之小字	繫聯係數	$(1-r^2)^{\frac{1}{2}}$	分子中之乘積項	分子全部	分母	符號下加註之小字	繫聯係數
12	−.4814		−.2604	−.2210	.6653	12.3	−.3322
13	−.6968	.7173					
23	+.3737	.9275					
14	−.4937		−.1994	−.2943	.6873	14.3	−.4282
13	−.6968	.7173					
43	+.2862	.9582					
24	+.3633		+.1070	+.2563	.8887	24.3	+.2884
23	+.3737	.9275					
43	+.2862	.9582					
13	−.6968		−.1799	−.5169	.8130	13.2	−.6358
12	−.4814	.8765					
32	+.3737	.9275					
14	−.4937		−.1749	−.3188	.8166	14.2	−.3904
12	−.4814	.8765					
42	+.3633	.9317					
34	+.2862		+.1358	+.1504	.8642	34.2	+.1740
32	+.3737	.9275					
42	+.3633	.9317					
12	−.4814		−.1794	−.3020	.8102	12.4	−.3727
14	−.4937	.8696					
24	+.3633	.9317					
13	−.6968		−.1413	−.5555	.8333	13.4	−.6666
14	−.4937	.8696					
34	+.2865	.9582					
23	+.3737		+.1040	+.2697	.8928	23.4	+.3021
24	+.3633	.9317					
34	+.2862	.9582					
14	−.4937		−.1994	−.2943	.6873	14.3	−.4282
13	−.6968	.7173					
43	+.2862	.9582					
12	−.4814		−.2604	−.2210	.6653	12.3	−.3322
13	−.6968	.7173					
23	+.3737	.9275					
42	+.3633		+.1070	+.2563	.8887	42.3	+.2884
43	+.2862	.9582					
43	+.3737	.9275					

當然，係數 $r_{23\cdot4}$ 與 $r_{32\cdot4}$ 相同，$r_{34\cdot2}$ 與 $r_{43\cdot2}$ 相同，餘照此類推，故關於此數等值可不必重複計算.

[二次偏繫聯係數之計算]

由此等一次偏繫聯係數，可計算所須之三個二次偏繫聯係數，其法與上面所用者相同.計算程序列於下表，對於所求之每一個係數皆用兩種不同之一次係數結合形式計算之，俾測驗計算之正確與否也.

表一二二

示二次偏繫聯係數之計算法

堪薩斯邦玉蜀黍產量與氣溫

一次之 r						二次之 r	
符號下加註之小字	繫聯係數	$(1-r^2)^{\frac{1}{2}}$	分子中之乘積項	全部分子	分母	符號下加註之小字	繫聯係數
12.3	−.3522		−.1235	−.2087	.8653	12.34	−.2412
14.3	−.4282	.9037					
24.3	+.2884	.9575					
13.2	−.6358		−.0679	−.5679	.9035	13.24	−.6265
14.2	−.3904	.9206					
34.2	+.1740	.9847					
14.2	−.3904		−.1106	−.2798	.7601	14.23	−.3681
13.1	−.6358	.7719					
43.2	+.1740	.9847					
12.4	−.3727		−.2014	−.1713	.7100	12.34	−.2411
13.4	−.6066	.7454					
23.4	+.3021	.9533					
13.4	−.6666		−.1126	−.5540	.8847	13.24	−.6262
12.4	−.3727	.9280					
32.4	+.3021	.9533					
14.3	−.4282		−.0958	−.3324	.9031	14.23	−.3681
12.3	−.3522	.9432					
42.3	+.2884	.9575					

注意,此地求得之$r_{13\cdot24}$之値,與從兩個偏響應係數所求得者相同.

此種係數之意義,在前面討論此問題之一節中已詳細說明.上面將以上所得各種結果並列比較,以示吾人知識因上面之分析而獲得之增益.

$$r_{12}=-.4814 \qquad r_{12\cdot34}=-.2412$$

$$r_{13}=-.6968 \qquad r_{13\cdot24}=-.6265$$

$$r_{14}=-.4937 \qquad r_{14\cdot23}=-.3681$$

六月氣溫所加於玉蜀黍產量之純結果,較諸簡單繫聯係數所指示者顯然爲小,所以然者,蓋六月氣溫與七、八兩月之氣溫間有一正繫聯存在;粗糙之簡單繫聯,僅計及兩個變數,故其指示六月氣溫之重要性也過於其實.由於同一之理由,故所有之偏繫聯係數無不小於簡單係數.雖然,七月氣溫所加於玉蜀黍產量之影響,其大於六月及八月氣溫之影響一點,猶顯而易見者也.

偏繫聯係數又名爲純繫聯係數,其所謂純(net),當然僅指變數之業已計及而視作固定者言.此外尚有其他因子,如六、七、八三個月之雨量亦爲影響玉蜀黍產量者,且與該三月之氣溫成繫聯;若將此等因子一併包含在研究範圍之內時,各純繫聯係數之値將較現今之値或大或小.

[離中趨勢之計量]

吾人在求得此等偏繫聯係數後,猶可計算一種計量,此

種計量相當重要.此即在若干相關之變數視作固定時,剩餘之離中趨勢之計量是也.問題爲若能使堪薩斯邦六月七八月之氣溫固定不變,則玉蜀黍產量之離中趨勢又爲如何乎?換言之:即吾人若能在玉蜀黍產量中除去其因氣溫變動而發生之變動,則玉蜀黍產量中所餘留之變動爲何?此種離中趨勢之計量用$\sigma_{1\cdot234\cdots n}$表示之.此計量稱爲$n$次標準差.

此數可由下列之普遍方程式求得之

$$\sigma^2_{1\cdot23\cdots n}=\sigma_1^2(1-r^2_{12})(1-r^2_{13\cdot2})(1-r^2_{14\cdot23})(1-r^2_{1n\cdot23\cdots n-1})$$

將此公式應用於前面研究玉蜀黍產量所得之結果,得

$$\sigma^2_{1\cdot234}=1188.688\,[1-(.4814)^2][1-(-.6358)^2][1-(-.3681)^2]$$

$$\sigma^2_{1\cdot234}=470.3364$$

$$\sigma_{1\cdot234}=21.69$$

回觀前面關於此問題之討論,可見$\sigma_{1\cdot234}$之值與$S_{1\cdot234}$之值相等.(上面實際上求得之值相差.01).即當X_2 X_3,與X_4作爲固定時,變數X_1之標準差僅爲響應線周圍之離中趨勢計量,即當根據X_2 X_3與X_4作估量時之估量標準誤也.原來數列之離中趨勢,減小至據關係方程式而作之估量值接近於實在值之程度.餘留之離中趨勢乃因此等估量值與實在值間之差數而來.而此等差數者,僅爲響應線周圍之差離,S即由此計算者也.明乎此兩計量——$\sigma_{1\cdot234}$與$S_{1\cdot234}$——之相同.則兩者之意義可愈爲明顯.

$\sigma_{1\cdot234}$與$S_{1\cdot234}$既相同,則複繫聯係數$R_{1\cdot234}$可由下列之方

程式計算之.

$$R^2_{1.23\cdots n}=1-\frac{\sigma^2_{1.23\cdots n}}{\sigma_1^2}$$

或將$\sigma^2_{1.234\cdots n}$之公式插入,此方程式可改爲下式

$$1-R^2_{1.23\cdots n}=(1-r^2_{12})(1-r^2_{13.2})(1-r^2_{14.23})\cdots(1-r^2_{1n.23\cdots(n-1)})$$

但第一法直接根據最小平方法而作,接近問題較爲簡單,平常較後者爲佳.

[複繫聯及偏繫聯法之限制]

處理複繫聯及偏繫聯問題時吾人所敍述之種種計量,惟在諸變數間之關係在無論何處皆爲直線形式時始爲有效,如有四個變數,可得六種不同之配對.若欲用上述之方法研究許多變數綜合之影響,或其中每一變數之純影響,此六種關係必須全爲直線響應,若自然數之響應非直線形式,則應用對數或倒數或可得一直線關係,如X_1以對數形式表示時可與其他每一個變數之自然數間成一直線關係,則吾人可得下列之估量方程式

$$\operatorname{Log}X_1=a+b_{12.34}X_2+b_{13.24}X_3+b_{14.23}X_4$$

於是與之相當之S與R之值將表現爲比率,如前章諸例所示者.

另一重要之限制須注意者,根據大量之變數而求得之複繫聯及純繫聯係數,非在觀察值亦爲大量時,無甚意義.當研究所涉及之變數甚多而個數有限時,係數之值將甚高,然

實爲錯誤,不足代表眞相,複繫聯與偏繫聯方法在此等限制所設定之範圍以內,在經濟分析上,爲非常有用之工具.

參考書

Bowley Arthur L. Elements of Statistics (398—408)

Edgeworth, F. Y. On Correlated Averages. Phil. May.5th series, Vol. XXXIV, 1892 (194).

Ezekiel, M. J. B. A Method of Handling Curvilinear Correlation for any Number of Variables. Journal of the Amierican Statistical Association, Vol. XIX, N. S. 148, 1924.

Haas, G. C. Sale Prices as a Basis for Farm Land Appraisal. Technical Bulletin No. 9. University of Minnesota Agricultural Experiment Station.

Kelley, Truman L. Statistical Method (279—310).

Kelley, Truman L. Partial and Multiple Correlation (載 Rietz, H. L. Handbook of Mathematical Statistics, 139—149).

Miner, J. R. Tables of $\sqrt{1-r^2}$ and $1-r^2$ for use in Partial Correlation and Trigonometry.

Pearl, Raymond. Medical Biometry and Statistics(319—331)

Pearson, Karl Regression Heredity and Panmixia. Phil. Transactions Royal Society, Series A. Vol. CLXXXVII. 1896(253—318)

Smith, Bradford B. The Use of Punched Card Tabulating Equipment in Multiple Correlation Problems. (供美國農業

部農業經濟局統計員之用）

Tolley H. R. and Ezekiel, M. J. B. A Method of Handling Multiple Correlation Problems. Journal of the American Statistical Association Dec. 1923.

Yule, G. U. An Introduction to the Theory of Statistics. (229—253).(包括多個變數之繫聯其記法爲Yule氏所創本書有其說明.)

第十五章

機率理論之初步及差誤正態曲線

在第四章中，曾指出數量材料雖來自大不相同之領域者，其頻數分配中仍有同族相似之點存在．並曾舉出一種基本形式，卽圖示時以名爲"正態曲線"(Normal curve)或"差誤正態曲線"(Normal curve of error)之對稱鐘形曲線所代表者．此曲線從前曾被認爲一基本法則之代表者，足以敍述一切數量材料之分配狀態．由今視之，此實爲一錯誤之觀念．蓋現今認爲正態曲線者，不過可用以敍述頻數分配之多種曲線之一耳．雖然，此種形式之重要性遠勝其他諸形式，而其特性爲統計家所不得不認識者也．

[機率之基本定理]

欲求此種認識，莫如對於機率論之若干基本原理略加研究．詳細說明機率理論將大越本書之範圍．下文所述僅作爲該問題之楔子，期以簡單數字之例，說明機率之原理與正態差誤曲線間之關係耳．

在此討論中，吾人可用下列之標準記號．若一事件可以

n 種形態出現, a 爲成, b 爲敗;則代表結果成之機率 p 可寫作

$$p=\frac{a}{n}$$

而代表結果敗之機率 q 可寫作

$$q=\frac{b}{n}$$

成敗結果之和旣等於事件之總數,故

$$a+b=n$$

以 n 除之

$$\frac{a}{n}+\frac{b}{n}=1$$

故

$$p+q=1$$

卽成敗機率之和等於一必然(Certainty)也.

因此機率可寫作比率.代表此比率之分數式,其分子代表成(或敗)之結果之次數,其分母代表可能之結果之總數.

[簡單機率之例]

擲幣得面向上則爲成,則成之機率爲

$$p=\frac{1}{2}$$

敗之機率

$$q=\frac{1}{2}$$

投一骰子,六點爲成,則

$$p=\frac{1}{6}$$

而

$$q=\frac{5}{6}$$

一副紙牌52張,抽得鏟花一點(Ace of spades)之機會爲 $\frac{1}{52}$,未得該牌之機會爲 $\frac{51}{52}$

[機率之相加]

由52張紙牌中抽牌一次,能得鏟花一點或鏟花兩點之牌之機會爲如何?遇此種情形,即得幾種結果之任何一種皆作爲成時,則成之機率爲幾種結果之每一種之機率相加之總和.如現今之例中即爲

$$p=\frac{1}{52}+\frac{1}{52}=\frac{1}{26}$$

由一套紙牌抽得心花(heart)牌或鏟花牌之機會爲

$$p=\frac{13}{52}+\frac{13}{52}=\frac{1}{2}$$

[機率之相乘]

兩事件之結果彼此不相影響,則此兩事件稱爲獨立(Independent).如骰子之第一擲諒不能對於其後之第二擲有何影響也.一繁複事件(Compound event)(即彼此獨立之兩事件一同出現者)之機率,爲每一事件之機率之乘積.故投骰子時前後相繼之兩擲中,先得一點繼以兩點之機會,爲

$$p=\frac{1}{6}\times\frac{1}{6}=\frac{1}{36}$$

當計算某一結果之機率時,有時須將機率旣相乘而復相加.例如求同時擲兩骰子時得點子總數爲5之機會即是也;先將兩骰子標以"甲""乙"二字以示分別,下列各種之結合皆得點子總數爲5.

骰子甲	骰子乙
1	4
2	3
3	2
4	1

擲骰子甲得一點之機會爲$\frac{1}{6}$,擲骰子乙得四點之機會爲$\frac{1}{6}$,兩者相遇同時出現之機會爲$\frac{1}{36}$.同樣,其他結合出現之機會各爲$\frac{1}{36}$.但四種結合莫不得點子總數5,即其結果皆作爲成,故

$$p=\frac{1}{36}+\frac{1}{36}+\frac{1}{36}+\frac{1}{36}=\frac{1}{9}$$

此例已答復下列之問題:卽同擲兩骰子時,恰得5點之機率爲何是也.吾人更可將問題改爲同擲兩骰子時至少得5點之機率爲何?此時點子總數爲5或5以上者將視爲成.茲列每種如此之結果之機率於下:

兩骰子點數和爲	12	之機率	$=\frac{1}{36}$
同　上	11	同上	$=\frac{2}{36}$
同　上	10	同上	$=\frac{3}{36}$
同　上	9	同上	$=\frac{4}{36}$
同　上	8	同上	$=\frac{5}{36}$
同　上	7	同上	$=\frac{6}{36}$
同　上	6	同上	$=\frac{5}{36}$
同　上	5	同上	$=\frac{4}{36}$
		機率總和	$=\frac{30}{36}$

故同時擲兩骰子得點數至少在5以上之機會爲$\frac{30}{36}$或$\frac{5}{6}$.

[二項展開式與機率之計算]

此等事實可以一普遍式表示之.茲先舉一簡單之例,以示此普遍式之由來.

同時取兩幣而擲之,可能之結果有四:

甲乙　甲乙　甲乙　甲乙
背背　背面　面背　面面

（此地兩幣各以甲乙表示之）.兩幣皆爲背,一幣爲背一幣爲面,及兩幣皆爲面之結果,其機會各爲 $\frac{1}{4}$, $\frac{1}{2}$, 及 $\frac{1}{4}$. 若同時擲甲乙丙三幣,則有八種可能之結果,卽

甲乙丙　甲乙丙　甲乙丙　甲乙丙
背背背　背背面　背面面　背面背

甲乙丙　甲乙丙　甲乙丙　甲乙丙
面背背　面背面　面面背　面面面

結果無一面出現,有一面出現,有兩面出現,及有三面出現之機會各爲 $\frac{1}{8}$, $\frac{3}{8}$, $\frac{3}{8}$, $\frac{1}{8}$.

惟此等結果,不必詳細計算個別之機率,亦可求得.前曾用 p 與 q 分別代表每一事件成敗之機率,若獨立之事件有二,兩者之機率相同,則繁複事件之機率可以下式

$$(p+q)^2$$

之展開式表示之.此地 p（擲得一面之機率）$=q=\frac{1}{2}$, 而各種結果之機率可以下式決定之

$$\left(\frac{1}{2}+\frac{1}{2}\right)^2=\frac{1}{4}+\frac{1}{2}+\frac{1}{4}$$

此卽第一例所得結果也.若獨立事件有三,則各種結果之機率爲

$$\left(\frac{1}{2}+\frac{1}{2}\right)^3=\frac{1}{8}+\frac{3}{8}+\frac{3}{8}+\frac{1}{8}$$

此卽第二例所得結果也.

若所求者非各種結果個別之機率,而爲若干次試驗中各種不同結果出現之或然頻數(Probable frequencies),則可由下式

$$N(p+q)^n$$

求之,式中N代表試驗次數,n爲獨立事件之數.如試驗次數爲200,獨立事件有2,或然頻數可由下式計算之,

$$200(p+q)^2=200(p^2+2pq+q^2)$$

若$p=q=\frac{1}{2}$,則得

$$200\left(\frac{1}{4}\right)+200\left(\frac{1}{2}\right)+200\left(\frac{1}{4}\right)=50+100+50$$

分別表示兩者皆實現,一者實現一者不實現,及兩者皆不實現之或然頻數.

若獨立事件有三,N次試驗中之或然頻數可由

$$N(p+q)^3$$

之二項展開式決定之.若N等於200,則得

$$200(p^3+3p^2q+3pq^2+q^3)$$

若p等於$\frac{1}{2}$,則得

$$200\left(\frac{1}{8}\right)+200\left(\frac{3}{8}\right)+200\left(\frac{3}{8}\right)+200\left(\frac{1}{8}\right)$$

式中各項依次表示三者皆實現,三者中有兩者實現,三者中有一者實現,及三者中無一實現之或然頻數.將各項乘得之頻數相加,其總和將等於試驗之次數;因一切可能之結果皆包括於此展開式中也.

因此吾人若預知機率,(註)則成或敗之次數不論幾何,成敗次數之或然頻數皆可以決定.此事於統計理論之發展上甚爲重要.

[在純偶然的領域中實在頻數與理論頻數之比較]

吾人可取若干試驗之結果與從二項展開式得出之理論頻數(Theoretical frequency)相比較,以說明若干重要之點.設將十二個骰子同時投擲若干次;凡四點,五點或六點中有任何一種出現時皆作爲成;凡一點,二點或三點中有任何一種出現皆作爲敗.(如以一擲爲標本,或得向上之點數爲3,1,5,1,2,4,4,6,3,2,3,5.此中有5個成,結果即如此記下.)根

(註) 平常機率分爲兩種:一爲先天的機率(A priori probabilities),一爲經驗的機率(Empirical probabilities),後者由觀察與經驗得來.美國人民在35歲以上,者其中能再活十年者,其機率爲74173/81822;此機率即後者之例.因此數之根據爲美國之死亡率調查表.表中載在35歲之人81822個中,十年後尙活者共有74173個,故得此機率.第十六章中將論及關於應用與整理試驗結果之若干問題.

據此成敗之條件,維爾頓氏曾作一模範之試驗,(註)將十二個骰子如此投擲 4096 次而紀錄其結果.茲將其結果載入表一二三第(2)行中,圖 85 爲其分配之圖示.根據計算結果,此分配之算術中數爲 6.139,標準差爲 1.712.

試將此結果與理論上推測之結果相比較,每次擲骰子十二個,故吾人爲處理12個獨立事件,試驗次數爲 4096.既以四點或五點或六點之出現爲成,故 $p=q=\frac{1}{2}$

二項展開式之各項爲

$$(a+b)^n=a^n+na^{n-1}b+\frac{n(n-1)}{1.2}a^{n-2}b^2+\frac{n(n-1)(n-2)}{1.2.3}a^{n-3}b^3+\ldots$$

今吾人有二項式

$$4096\left(\frac{1}{2}+\frac{1}{2}\right)^{12}$$

展開之

$$4096\left(\frac{1}{4096}+\frac{12}{4096}+\frac{66}{4096}+\frac{220}{4096}+\frac{495}{4096}+\frac{792}{4096}+\frac{924}{4096}+\frac{792}{4096}\right.$$
$$\left.+\frac{495}{4096}+\frac{220}{4096}+\frac{66}{4096}+\frac{12}{4096}+\frac{1}{4096}\right)$$

括弧外之 4096 與括弧內之各項分母相消,所得者爲將十二

(註) 游爾氏(G. U. Yule)在 An Introduction to the Theory of Statistics, 5th Ed. P.258. 及愛吉華斯 (F. Y. Edgeworth) 氏在大英百科全書十一版 Vol. XXII, 394 中曾引及此例.

個骰子擲 4096 次時各種可能的成功之理論頻數.茲將其列於下表第(3)項.

表一二三

擲骰試驗中實在頻數與理論頻數之比較

(1) 得四五六各點之骰子數	(2) 實在頻數	(3) 理論頻數
0	0	1
1	7	12
2	60	66
3	198	220
4	430	495
5	731	792
6	948	924
7	847	792
8	536	495
9	257	220
10	71	66
11	11	12
12	0	1
	4096	4096

此理論頻數之分配繪於下圖

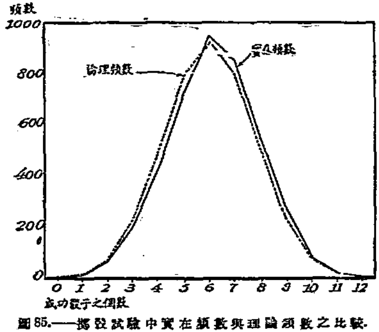


圖 85.——擲骰試驗中實在頻數與理論頻數之比較.

吾人能求得一標準式之分配以與試驗中得來之實在分配相比較時,所得利益實多.首先吾人能藉以決定頻數間之差異其由於偶然抽樣變動(Chance fluctuation of sampling)(或骰子所含有之偏誤)所引起之程度.假定試驗之次數增加至無限時所能得到之眞正分配狀態,亦可藉以決定.試驗之結果其獨立時之意義僅能限於所紀錄之擲骰情形,惟理論頻數則無此種限制,凡內在機率(Initial probability)相同之處咸可適用.理論頻數最主要之優點,在其供一推理(Generalization)之基礎,此乃試驗結果單獨所不可能者也.

若已知內在機率有如此例,卽可進而求理論頻數的分配之算術中數與標準差.當獨立事件之數目及成功之機率爲已知數時,成功數之算術中數(Mean number of successes)之普遍式爲

$$M=np$$

以本例之數值代入,得

$$M=12\times\frac{1}{2}=6$$

由實在頻數計算得之算術中數爲 6.139.

在同一情形之下,標準差之普遍式爲

$$\sigma=\sqrt{npq}$$

以本例之數值代入得

$$\sigma=\sqrt{12\times\frac{1}{2}\times\frac{1}{2}}=\sqrt{3}$$

由實在頻數計算而得之標準差爲1.712.

[差誤正態曲線]

今回至代表擲骰試驗之理論頻數之曲線(圖85)而研究之.此曲線成一完全對稱之十二邊多邊形,邊數(除底邊外)與所研究之問題中獨立事件之數相當.若獨立事件有六,則得一六邊形,若獨立事件之數爲二十,則得一二十邊形,餘可類推.顯然當 n 之數增大時,多邊形之邊數相應增加,其數與前者相當,而代表 $(p+q)^n$ 二項式之展開式之曲線,於是逐漸接近於光滑之曲線. n 若增至無限大,卽可得一完全光滑之曲線.此卽繪於圖87之差誤正態曲線也.

此正態曲線,或稱高斯式曲線(Gaussian curve).因對於此種曲線之特點之詳細的認識,使現代統計技術獲得大部分之力量.理論頻數之決定,可利用此種曲線之現成的積分表,其法比較根據二項展開式推算之法簡捷甚多.下文略舉數例以說明此種表之用途.

代表此曲線之方程式,可寫成數種形式,普通之形式爲

$$y=y_0e^{-\frac{x^2}{2\sigma^2}}$$

在此方程式中, y_0 代表最大之縱坐標,爲一常數; e 爲一常數(納皮爾對數 Napierian logarithms 之基數),其值爲2.7182818

而σ則代表標準差,(註1)最大縱坐標之值可由下列關係式求得,

$$y_0=\frac{N}{\sigma\sqrt{2\pi}}$$

故代表正則曲線之方程式可寫爲

$$y=\frac{N}{\sigma\sqrt{2\pi}}e^{-\frac{x^2}{2\sigma^2}}$$

此曲線之各種用途,與根據此曲線之積分而作之機率積分表之種種用途,太繁太多,此地不能盡述.茲可舉一例作爲此問題之小引.

〔應用於經濟之一例〕

美國電話電報公司統計部對於勃發羅(Buffalo)地方裝置共同四支電話線(Four-party line)之住戶之每年通話次數曾作一研究.將995個住戶中每戶每年通話次數列表分類.(註2)下表表示其結果與若干計算.

(註1) 高斯氏(Gauss)求差誤方程式(Error equation)之法可於論述最小平方理論之一切標準著作中見之.參觀附錄一後列參考書目.

(註2) 見 "Introduction to Frequency Curves and Averages." Statistical Bulletin, Statistical Methods Series, No. 1. 美國電話電報公司首席統計員發表.

表一二四

995個電話裝戶每年通話次數
示計算一類數分配之轉矩之法

(1) 每年通話次數組距	(2) 中點	(3) 類數	(4) 對於假定原點之差離（單位組距）	(5)	(6)	(7)	(8)
	m	f	x'	fx'	$f(x')^2$	$f(x')^3$	$f(x')^4$
0— 50	25	0	−10	0	0	0	0
50—100	75	1	−9	−9	81	−729	6561
100—150	125	9	−8	−72	576	−4608	36864
150—200	175	19	−7	−133	931	−6517	45619
200—250	225	38	−6	−228	1368	−8208	49248
250—300	275	50	−5	−250	1250	−6250	31250
300—350	325	90	−4	−360	1520	−6080	24320
350—400	375	85	−3	−255	765	−2295	6885
400—450	425	115	−2	−230	460	−920	1840
450—500	475	132	−1	−132	132	−132	132
500—550	525	144	0	0	0	0	0
550—600	575	116	1	116	116	116	116
600—650	625	79	2	158	316	632	1264
650—700	675	54	3	162	486	1458	4374
700—750	725	31	4	124	496	1984	7936
750—800	775	11	5	55	275	1375	6875
800—850	825	5	6	30	180	1080	6480
850—900	875	6	7	42	294	2058	14406
900—950	925	2	8	16	128	1024	8192
950—1000	975	1	9	9	81	729	6561
1000—1050	1025	1	10	10	100	1000	10000
1050—1100	1075	1	11	11	121	1331	14641
		995		−956	9676	−22952	283534

(註) 此種分組之法，以50一項歸入以50為上限之一組，其他適在組限上之各項皆同樣處理。

〔頻數分配之轉矩〕

有若干之名辭與記號爲以前所未用者可於茲介紹之：

$\nu_1=\dfrac{\Sigma f(x')}{N}$＝分配對假定原點之第一轉矩

$\nu_2=\dfrac{\Sigma f(x')^2}{N}$＝分配對假定原點之第二轉矩

$\nu_3=\dfrac{\Sigma f(x')^3}{N}$＝分配對假定原點之第三轉矩

$\nu_4=\dfrac{\Sigma f(x')^4}{N}$＝分配對假定原點之第四轉矩

"轉矩"(Moment)原爲一常用之力學名詞，指力之旋轉趨勢之計量.決定此種趨勢之大小者，自爲用力點離原點之距離.此名詞用於統計學上時，意義亦相類.各組之頻數即爲問題中之力，組中點與原點之距離則爲此方面之一重要因子.一個頻數分配，其對於任何原點之轉矩，可以用下法求之：以每組之距離(組中點與原點在X軸上之距離)之某次方乘本組之頻數，將各組之乘積相加，復以總項數除之.若求第一轉矩，用距離之一次方；若求第四轉矩，用距離之四次方，其餘按此類推.記號右下角之小字乃分別其代表之轉矩之次數者.

統計上最重要之轉矩，爲轉矩之用算術中數爲原點者.此種轉矩以π代表之，其關係如下：

對於算術中數之第一轉矩$=\pi_1=0$

對於算術中數之第二轉矩 $=\pi_2=\nu_2-\nu_1^2$

對於算術中數之第三轉矩 $=\pi_3=\nu_3-3\nu_1\nu_2+2\nu_1^3$

對於算術中數之第四轉矩 $=\pi_4=\nu_4-4\nu_1\nu_3+6\nu_1^2\nu_2-3\nu_1^4$

此等轉矩由分組之材料計算時,乃假定每組內各項可視爲集中於組中點處而處理之者.惟事實證明在某種條件之下根據此種假定而作之計算,犯一經常之錯誤.特別當由分組之材料計算時,第二第四兩轉矩之值與由未分組之材料計算而得者不同.

希巴特(W. F. Shepard)氏(註)創校正此偏誤之法,在下面兩條件存在時可施用之:

(1)分配爲關於連續變數者.

(2)頻數曲線以"高度接切"(High contact)爲特色者,即頻數曲線向兩方逐漸坦下者.

吾人用 μ 爲記號以代表對於算術中數之校正轉矩.應用希巴特校正法最後得下列公式:

$$\mu_1=0$$

$$\mu_2=\pi_2-\frac{1}{12}$$

$$\mu_3=\pi_3$$

$$\mu_4=\pi_4-\frac{1}{2}\pi_2+\frac{7}{240}$$

(註) 參考 Proceedings of London Mathematical Society, Vol. XXIX, 353—380.

(實施校正時$\frac{1}{12}$與$\frac{7}{240}$常用其相當之小數.083333與.029167.)

在作此種校正時,乃假定在計算對於算術中數之差離時所用之單位爲組距者也.

於此可附帶提起,標準差等於對於算術中數之第二轉矩之方根,在轉矩之未校正值,

$$\sigma=\sqrt{\pi_2}$$

在經過希巴特氏法校正後,

$$\sigma=\sqrt{\mu_2}$$

電話用戶頻數分配之轉矩之計算如下.此中應用希巴特氏校正法.因曲線表示相當高度之接切也.固非一連續變數之分配,然其單位(1)與全距相形極小,故可作爲連續變數處理之.

$$\nu_1=\frac{-956}{995}=-.960804$$

$$\nu_2=\frac{9676}{995}=9.724623$$

$$\nu_3=\frac{-22952}{995}=-23.067337$$

$$\nu_4=\frac{283564}{995}=284.988945$$

$$\pi_1=0$$

$\pi_2=\nu_2-\nu_1^2=9.724623-.923144=8.801479$

$\pi_3=\nu_3-3\nu_1\nu_2+2\nu_1^3=-23.067337+28.030370-1.773922$

$=3.189111$

$\pi_4=\nu_4-4\nu_1\nu_3+6\nu_1^2\nu_2-3\nu_1^4$

$=284.988945-88.652760+53.863384-2.556586$

$=247.642983$

$\mu_1=0$

$\mu_2=\pi_2-\dfrac{1}{12}=8.801479-.083333=8.718146$

$\mu_3=\pi_3=3.189111$

$\mu_4=\pi_4-\dfrac{1}{2}\pi_2+\dfrac{7}{240}=247.642983-4.400739+.029167$

$=243.271411$

[曲線型判準]

既得以上諸值,吾人乃可回至主要問題之研究,卽吾人關於正態曲線的知識之效用是也.有若干判準(criteria)可使吾人立卽決定某一分配之是否可以正態曲線描述.此種判準可自該分配之校正轉矩求得,

$$\beta_1=\frac{\mu_3^2}{\mu_2^3}=\frac{10.170429}{662.632015}=.01534853$$

$$\beta_2=\frac{\mu_4}{\mu_2^2}=\frac{243.271411}{76.006070}=3.200683$$

$$\kappa_2=\frac{\beta_1(\beta_2+3)^2}{4(4\beta_2-3\beta_1)(2\beta_2-3\beta_1-6)}$$

$$=\frac{.01534853\times 38.448470}{4(12.756686)\ (.355320)}=\frac{.5901275}{18.130823}$$

$$\kappa_2=.032548$$

此等判準在正態曲線其值爲

$$\beta_1=0$$

$$\beta_2=3$$

$$\kappa_2=0$$

在試驗的意義上，吾人可認爲該分配可以用正態曲線描述之.(註1)

〔應用縱坐標表配合正態曲線之法〕

縱坐標表(Table of ordinates)將正態曲線之縱坐標化爲最大縱坐標(y_0)之分數而表示之，此現成之表，使配合正態曲線之手續簡單化.下列短表雖太簡略不能用於實際之計算，然可用以表示手續之本質.(註2)

(註1) 若用皮爾生式第四種曲線(Pearsonian Type IV curve)描述之當更爲適合，但稽核可見正態曲線適合之程度亦相當完全，而其簡單一點則爲前者所不及之優點也.

(註2) 在潘爾氏 Tables for Statisticians and Biometricians 一書中有一詳細之表，其中正態曲線之縱坐標各以橫坐標表示之.

表一二五

正態曲線之縱坐標

(表示爲最大坐標之分數)

x/σ	y/y_0	x/σ	y/y_0
0.0	1.00000	2.5	.04394
0.1	.99501	2.6	.03405
0.2	.98020	2.7	.02612
0.3	.95600	2.8	.01984
0.4	.92312	2.9	.01492
0.5	.88250	3.0	.01111
0.6	.83527	3.1	.00819
0.7	.78270	3.2	.00598
0.8	.72615	3.3	.00432
0.9	.66689	3.4	.00309
1.0	.60653	3.5	.00219
1.1	.54307	3.6	.00153
1.2	.48675	3.7	.00105
.3	.42956	3.8	.00073
1.4	.37531	3.9	.00050
1.5	.32465	4.0	.00034
1.6	.27804	4.1	.00020
1.7	.23575	4.2	.00015
1.8	.19790	4.3	.00010
1.9	.16448	4.4	.00006
2.0	.13534	4.5	.00004
2.1	.11025	4.6	.00003
2.2	.08892	4.7	.00002
2.3	.07100	4.8	.00001
2.4	.05614	4.9	.00001
		5.0	.00000

在 x/σ 一行中,對於算術中數之差離各以標準差爲單位而表示.在 y/y_0 一行,離算術中數一定距離處正態曲線之縱坐標,各表示爲最大縱坐標之分數.如與算術中數相隔之距離等於一標準差處之縱坐標等於.60653y_0,而與算術中數相隔之距離等於2.8σ處之縱坐標等於.01984y_0.故配合一正態曲線於某一分配時,第一須將對於算術中數之差離化成

標準差單位,第二須決定最大縱坐標之值,最後須用一表如上列者以計算其他一切必須之縱坐標.最大縱坐標之值由下式求之.

$$y_0=\frac{N}{\sigma\sqrt{2\pi}}$$

或將$\sqrt{2\pi}$之值代入則成

$$y_0=\frac{N}{2.506628\sigma}$$

在本例

$$N=995$$

而

$$\begin{aligned}\sigma&=\sqrt{\mu_2}\\&=\sqrt{8.718146}\\&=2.953\end{aligned}$$

故

$$\begin{aligned}y_0&=\frac{995}{2.506628\times 2.953}\\&=134.42\end{aligned}$$

算術中數之值爲 476.96

以後之計算手續見於下表.

應用 y_0 公式時,以及作下表之一切計算時,標準差皆須用組距爲單位以表示之,蓋必如此而後計算得之理論頻數

與分組之實在頻數可以相比也.

表一二六

計算配合於電話用戶頻數分配之正態曲線之縱坐標

通話次數組中點 m	對於算術中數之差離(以組距爲單位) x	$\frac{x}{\sigma}$	$\frac{y}{y_0}$	計算得之頻數(縱坐標) y_c
25	−9.0392	−3.06	.00926	1.24
75	−8.0392	−2.72	.02474	3.33
125	−7.0392	−2.38	.05888	7.91
175	−6.0392	−2.05	.12230	16.44
225	−5.0392	−1.71	.23176	31.15
275	−4.0392	−1.37	.39123	52.59
325	−3.0392	−1.03	.58834	79.08
375	−2.0392	−.69	.78817	105.95
425	−1.0392	−.35	.94055	126.43
475	−.0392	−.01	.99995	134.41
525	.9608	.32	.94856	127.71
575	1.9608	.66	.80429	108.11
625	2.9608	1.00	.60653	81.53
675	3.9608	1.34	.40747	54.77
725	4.9608	1.68	.24385	32.78
775	5.9608	2.02	.13000	17.47
825	6.9608	2.36	.06174	8.30
875	7.9608	2.70	.02612	3.51
925	8.9608	3.03	.01015	1.36
975	9.9608	3.37	.00342	.46
1025	10.9608	3.71	.00103	.14
1075	11.9608	4.05	.00027	.04
				994.71

[配合適度之測驗]

圖 86 所給者爲實在分配情形與由上面計算而得之正

態曲線.就觀察言之,正態曲線對於材料雖在幾點上顯有出入,然大體上頗為適合.此地自然發生一問題,卽正態曲線何以不能與材料處處適合是也.可能之答復有二,不能適合之故,或由僅於任何範樣中莫不具有之偶然變動.在電話用戶之分配中眞有一潛在之定律完全合於差誤正態定律(Normal law of error);所選之特殊範樣固可有若干不規則性,然若能包括極多之例項,則此種不規則性自可消滅者也.另一種可能的答復,卽差異之發生,或由於此種分配根本上不合於差誤正態定律,而此種定律不能用於電話用戶按通話次數分組之分配,若遇此種情形,卽不能應用正態曲線.

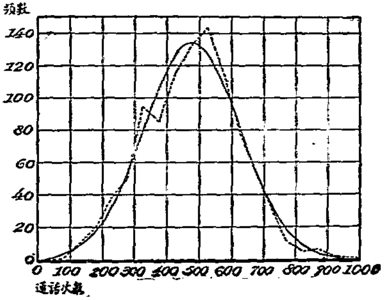


圖88.——表示配合一正態曲線於電話用戶頻數分配（按通話次數分組）之情形

此兩種解釋究以何者爲然乎?答復此問題之法有二:吾人或研究實在頻數與理論頻數相差最大之幾點,以判斷此等差異是否有可能爲任機抽樣 (Random sampling) 中偶然變動所引起之可能?或就整個曲線研究之,判斷若分配在本質上係合於差誤正態定律時,此觀察頻數與計算頻數(實在頻數與理論頻數)間之總差異是否有可能爲任機抽樣所引起,此兩法皆可應用.

但在應用此兩種方法之前,首須計算更精確之理論頻數.在組距中點之縱坐標,僅在該組距所包括之面積其頂邊之一段曲線成爲直線時,始能確切代表該面積.然在正態曲線之下,每組距所包括之面積其頂邊並非直線,故用正態曲線之縱坐標計算理論頻數,猶嫌粗略.欲精確決定理論頻數,必須計算面積以代替計算縱坐標.

[理論頻數之決定]

頻數曲線下之全面積,代表頻數之總和.如已知一片段之面積對於全面積之比率,此片段中所包含之頻數即不難求得.關於正態曲線,現有現成之機率積分表,由此表吾人可得所須之面積比率,眞幸事也.表一二七爲此種表之一例,用於精確之計算者,須較此表大爲詳細.(註)

(註) 希巴特博士所作機率積分之詳表,流傳甚廣.潘森氏 Tables for Statisticians and Biometricians 中包含一極便於應用之此種表式.

表一二七

差誤正態曲線下與一定橫坐標距離相當之面積表

(距離爲各縱坐標與最大縱坐標 y_0 間之距離,面積爲占全面積之部分)

x/σ	a	x/σ	a
0.0	.00000	2.0	.47725
0.1	.03983	2.1	.48214
0.2	.07926	2.2	.48610
0.3	.11791	2.3	.48928
0.4	.15542	2.4	.49180
0.5	.19146	2.5	.49379
0.6	.22575	2.6	.49534
0.7	.25804	2.7	.49653
0.8	.28814	2.8	.49744
0.9	.31594	2.9	.49813
1.0	.34134	3.0	.49865
1.1	.36433	3.1	.49903
1.2	.38493	3.2	.49931
1.3	.40320	3.3	.49952
1.4	.41924	3.4	.49966
1.5	.43319	3.5	.49977
1.6	.44520	3.6	.49984
1.7	.45543	3.7	.49989
1.8	.46407	3.8	.49993
1.9	.47128	3.9	.49995
		4.0	.49997

因正態曲線爲以最大縱坐標爲軸之一對稱形,上表所列諸值,對於中數之正負兩方皆可適用.

應用此表時,先將對於中數之差離化爲標準差單位,於是任何兩縱坐標線間之面積占全面積之部分極易決定之.例如:在最大縱坐標與離開中數 $+\sigma$ 之縱坐標間所包括之

項數,究竟占一正態分配全部項數之幾許部分?在表上 x/σ 行內找得 1.0,與之相對之值爲.34134.此卽全部項數中位於上述限界以內之部分,用比率形式表示者也.將此比率作成百分數,答數爲百分之 34.134.

圖 87 中以黑色部分 A 代表此部分,以表示此部分與曲線下全面積之關係.

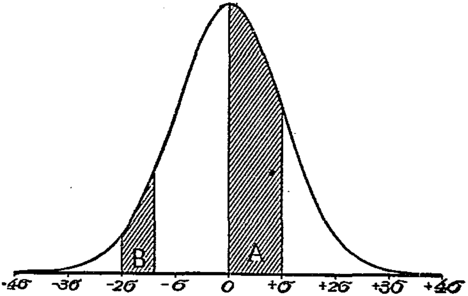


圖 87.——表示正態曲線下面積之計量

一正態頻數分配之全部項數中,位於距算術中數 -1.4σ 與 -2σ 處兩縱坐標線之間者,占幾何部分?由表上得知位於 y_0 與 -1.4σ 處縱坐標間之面積,占全部之 41.92%;位於 y_0 與 -2σ 處縱坐標間者,占全部 47.73%.兩百分數之差爲 5.81%,此卽位於 -1.4σ 與 -2σ 兩縱坐標間面積所占之部分也.將此百分數乘該分配之總項數,卽得真正之頻數.圖 87

之黑色部分B代表上述界限內之面積.

表一二八

由面積表計算理論頻數

(1) 組限	(2) 對於算術中數之差離 x/σ	(3) 各縱坐標線與 y_0 間之面積(全面積=1)	(4) 各縱坐標線與 y_0 間之頻數	(5) 各組之理論頻數	
0	−3.23	.4973810	496.88		(註)
50	−2.89	.4980738	495.58	0—50	1.92
100	−2.55	.4946129	492.14	50−100	3.44
150	−2.22	.4867906	484.36	100—150	7.78
200	−1.88	.4699460	467.60	150—200	16.76
250	−1.54	.4382198	436.03	200—250	31.57
300	−1.20	.3849303	383.01	250—300	53.02
350	−.86	.3051055	303.58	300—350	78.63
400	−.52	.1984682	194.48	350—400	106.90
450	−.18	.0714237	71.07	400—450	123.41
500	+.18	.0635595	63.24	450—500	134.31
550	+.495	.1896931	188.74	500—550	125.50
600	+.83	.2967303	295.25	550—600	106.51
650	+1.17	.3789995	377.10	600—650	81.85
700	+1.51	.4344783	432.31	650—700	55.21
750	+1.85	.4678432	465.50	700—750	33.19
800	+2.19	.4857379	483.31	750—800	17.81
850	+2.53	.4942969	491.83	800—850	8.52
900	+2.87	.4979476	495.46	850—900	3.63
950	+3.20	.4993129	496.82	900—950	1.36
1000	+3.54	.4997993	497.30	950—1000	.48
1050	+3.88	.4999478	497.45	1000—1050	.15
1100	+4.22	.4999876	497.49	大於 1050	.05
					995.00

(註) 理論分配表示在 -3.23σ 以下有 .62 項.此在本例將毫無意義,故加入 0—50 間之理論頻數中.

觀上述之例,根據此種面積表計算理論頻數之手續已可明瞭.蓋僅須決定各組限處縱坐標與最大縱坐標間之面積.用簡單之減法,卽可求得每組之面積.因而求得各組之理論頻數.表一二八應用電話裝戶之分配材料以示此手續之進行.

[測驗曲線適合程度之抽樣標準誤]

玆既得一套比較精確之理論頻數,請進而測驗配合曲線所得之結果以斷定正態曲線對於材料是否適合.吾人可先就觀察頻數與理論頻數間相差最大之數處,決定此種差異之意義.前曾言及,若已知機率,則可由下式求標準差

$$\sigma=\sqrt{npq}$$

式中 p 指成功之機率,q 指失敗之機率.處理頻數分配時,吾人可以$\frac{f}{N}$代替 p (f 代表 x 尺度上某一點之理論頻數,N 為總項數) 如此則 q 等於$\frac{N-f}{N}$,將其代入 p,q 之地位,則求標準差之普遍公式為

$$\sigma_s=\sqrt{N\times\frac{f}{N}\times\left(\frac{N-f}{N}\right)}$$

$$=\sqrt{\frac{f(N-f)}{N}}$$

此數名為抽樣之標準誤(Standard error of sampling).

觀察頻數與理論頻數間相差最大諸點如下:（f, 理論頻數, f_o, 觀察頻數

m	f_o	f	f_o-f
325	95	78.63	+16.37
375	85	106.90	−21.90
525	144	125.50	+18.50
775	11	17.81	−6.81

求第一組(中點325)抽樣標準誤,

$$\sigma_s=\sqrt{\frac{78.63(995-78.63)}{995}}=8.51$$

並求其他三組之標準誤,得下列諸值:

以325爲中點之組 = 8.51

以375爲中點之組 = 9.73

以525爲中點之組 =10.47

以775爲中點之組 = 4.18

關於電話通話次數之材料,若其分配與依差誤正態定律,觀察頻數與理論頻數間果有發生此種差異之可能乎?對此問題今可作確定之答復.在以325爲中點之一組處,差異爲+16.37;此組之σ_s爲8.51.差離爲σ_s之+1.9倍.若潛在之分配實爲正態時,在實際上此差離出現之機會爲如何?

吾人可參照機率積分表(表一二七)以答復此問.表上正對$\frac{x}{\sigma}=1.9$之數爲.47128.此即表示全部例數中約有47%位於最大縱坐標與距算術中數1.9σ處縱坐標間.離開算術中數$\pm1.9\sigma$之兩縱坐標線間,包有全面積94%.故一數値離開算術中數1.9σ以上之機會,約爲百中之六.故此一特殊之差離,可僅以抽樣之偶然變動解釋之,不足以證明正態曲線之不適合於現今之分配也.

對其他各組之差異可作同樣測驗.差異最大之組爲以375爲中點者,其差數-21.90相當於$2.2\sigma_s$.參照機率積分表可知100次機會中,有等於或大於此數量之負差離出現之次數約爲1.5.將正負雙方之差離合計之,則觀察頻數與理論頻數間如此之差異,在每百次試驗中約出現3次.於是吾人可得結論曰,本例中實際上所見之差異中,雖最大者亦不能證明正態曲線之不適於描寫材料(以通戶次數爲標準之住宅電話裝戶之分配.)

〔測驗適合程度之X^2測驗法〕

若不僅限於極端之差異,而計及所有之各組與所有之差異時,吾人可得更確定可靠之結果,乃顯然之事也.潘蓀氏曾發明一種如此之計量X^2(讀如Chi)使吾人能將全部材料盡數計及而決定某一形式頻數曲線之適合與否.已有理論頻數與觀察頻數時,計算此計量極易.舉例說明如下:

表一二九

χ^2之計算

測驗以差誤正態曲線配合於電話裝戶分配之適合程度

(1)	(2)	(3)	(4)	(5)
組限	觀察頻數 f_o	理論頻數 f	(f_o-f)	$\frac{(f_o-f)^2}{f}$
自150以下	10	13.14	−3.14	.75
150—200	19	16.76	+2.24	.30
200—250	38	31.57	+6.43	1.31
250—300	50	53.02	−3.02	.17
300—350	95	78.63	+16.37	3.41
350—400	85	106.90	−21.90	4.49
400—450	115	126.41	−11.41	1.03
450—500	132	134.31	−2.31	.04
500—550	144	125.50	+18.50	2.73
550—600	116	106.51	+9.49	.85
600—650	79	81.85	−2.85	.10
650—700	54	55.21	−1.21	.03
700—750	31	33.19	−2.19	.14
750—800	11	17.81	−6.81	2.60
800以上	16	14.19	+1.81	.23
	995	995.00	15組	$\chi^2=18.18$

注意,在作此表時曾將分配之兩端若干組合併爲一,在數值低之一端合併者有三組,數值高之一端合併者有六組,所以免在觀察頻數分配末端與理論頻數分配末端之間之微細差異過於放大也.

χ^2之值卽吾人所須之計量,所以表示適合程度者.其解

释必須用詳細之表.(註)下面由此種之表中摘錄一段以說明其手續:

由 X^2 及 n' 之數值求 P 之數值

X^2	$n'=14$	$n'=15$	$n'=16$
16	.249129	.313374	.382051
17	.199304	.256173	.318864
18	.175520	.206781	.232663
19	.123104	.164949	.213734
20	.095210	.130141	.171932

表中 X^2 代表 $\Sigma\left(\frac{(f_o-f)^2}{f}\right)$, n' 代表組數,本例

$$X^2=18.18$$

$$n'=15$$

$n'=15$, $X^2=18$ 時, P 之值爲 .206781; $n'=15$, $X^2=19$ 時, P 之值爲 .164949;內推得本例之 P 之數值爲 .199.

P 爲機率,卽實在分配取正態曲線形式時吾人所得之配合將等於此或更劣於此配合之機率.卽假定潛在之分配爲合乎正態時,在 1000 次試驗中有 199 次,或大概在 5 次中有 1 次,吾人所得之配合與今所得者同劣或較之更劣.此數決非過分,故吾人可以斷言正態曲線足以爲描述該分配之工具,因其對於材料相當適合也.

(註) 此種之表爲 W. Palin Elderton 所製,見潘蓋氏 Tables for Statisticians and Biometricians, 26-30

既得此結論,則所有一切知識,舉凡關於依照差誤正態定律之分配者,皆可以直接應用於該分配之研究.原來之頻數表僅許吾人研究該表所列舉之各組,今則可進一步,x 尺度上任何兩數值間所包括之項數莫不可以決定,且可計算一例位於 x 尺度上任何一點,或高於任何一數值,或低於任何一數值之機率.不僅吾人對於詳細之點之知識因而增進;如範樣確能爲代表,吾人且得一普遍推論之張本.此種張本,在將特殊分配獨立處理時,不論研究之範圍如何廣,皆無從獲得者也.(註)

[附記:關於頻數分配之敍述]

藉本章所說明之諸判準之助,吾人敍述一頻數分配時,在若干方面,可較早先各章中所用諸計量所能敍述者,稍爲精確.此問題之討論不在本書之範圍以內,惟對於此等後來加上之各種計量的本質,似宜略加指示也.

β_2之值爲計量某一曲線之頂端之平坦程度者.若$\beta_2=3$,如正態曲線之情形,則稱爲 Mesokurtic (常態峯),若 $\beta_2<3$,則較正態曲線爲平坦,稱爲 Platykurtic (平闊峯),若 $\beta_2>3$,如上述之例,則較正態曲線爲尖削,稱爲 Leptokurtic (高狹峯).

(註) 前已指出,正態曲線亦僅爲頻數曲線之一種形式,惟此種形式有根本之重要性耳.潘森氏求得若干標準曲線形式之方程式,並作詳細之描寫,蔚爲一頻數曲線之廣博系統.本書不能對於其他基本形式詳加討論,學者可參考本章參考書目中 A. Fisher 氏兩書.

由此等計量可算得一種偏態計量,較第五章所述者精確.皮爾生氏證明下式之值(註)

$$X=\frac{\sqrt{\beta_1}(\beta_2+3)}{2(5\beta_2-6\beta_1-9)}$$

爲一曲線之不對稱程度之計量.若將上文 β_1, β_2 兩值代入,則根據通話次數所作住宅裝戶之分配,其偏態爲

$$X=-.05558$$

(遇算術中數大於中位數時,X 爲正數;遇算術中數小於中位數時,X 爲負數.本例算術中數之值爲 476.96,而中位數之值爲 482.39,故偏態爲負數).

最後,算術中數與衆數間距離 d 可由下式定之

$$d=X\times\sigma$$

上述關於電話通話之分配中,其 σ 以原來單位表示時爲 147.65, 故

$$d=-.05558\times 147.65=-8.21$$

因

$$M_o=M-d$$

故

$$M_o=476.96+8.21=485.17$$

此數較之用第五章中所舉諸法求得之數更能接近於衆數值.

(註) 偏態計量 X 與測驗曲線配合之適合程度時所用之 X^2 兩者完全不同,切勿相混.

參考書

Bowley, Arthur L. Elements of Statistics (259—286)

Brunt, David The Combination of Observations (11—26)

Carver, H. C. Frequency Curves (載 Rietz, H. L. Handbook of Mathematical Statistics, 92—119)

Elderton, W. P. Frequency Curves and Correlation

Fisher Arne. An Elementary Treatise on Frequency Curves.

The Mathematical Theory of Probabilities.

Kelley, Truman, L. Statistical Method (94—108)(附 Kelleywood Table of the Normal Probability Integral.)

Pearl, Raymond. Medical Biometry and Statistics (220—263).

Pearson, Karl. Tables for Statistician and Biometricians. (表前導言特為可貴)

Shappard, W. F. On the Calculation of the Most Probable Values of Frequency Constants for Data Arranged According to Equi-distant Divisions of a Scale. (倫敦數學社會議錄第二十九期, 1898.)

The Calculation of the Moments of a Frequency Distribution. Biometrica, Vol. V.

Yule, G. U. An Introduction to the Theory of Statistics (291—316)

第十六章

統計歸納與抽樣問題

前文中所說明者,爲統計分析所應用之各種工具,並曾舉例表示此等工具在若干專門問題上之應用;惟自始至終側重於技術方面.吾人於此宜退開幾步,擴大眼界,對於與此等工具運用上有關之若干總問題,一加研究,此等工具在經濟與實業之研究中其適當之地位何在?其應用時所包含之假定與所受之限制如何?此等問題,本書雖不能詳加論述,然決不能略而不談者也.

〔統計敍述與統計歸納〕

討論此問題時,首宜將統計敍述(Statistical description)與統計歸納(Statistical induction)辨自清楚.吾人已知大量之數量材料可用統計方法作成簡明之敍述.數百個數千個單獨之事例,可以分類而作成頻數分配.更可將此分配之本質練成四種計量——集中趨勢量,離中趨勢量,偏態量及峯態量是也.吾人能如此以少數之計量,以代替錯綜複雜之無數個別事例,敍述其整個集團之幾種特質,裨益實非淺鮮.此種

工具,使吾人之有限之觀察力,能理解大量材料之意義.此外,兩個變量間之關係,亦藉統計方法之助而得以敍述.用數學方法配合曲線於材料後,由此曲線之方程式,可決定一個或多個變量發生變化時另一變量平均上發生幾多變化,除此關係方程式外,尚有指示配合曲線周圍離中趨勢之計量,最後尚有一抽象計量,指示應變數與一個或一個以上之自變數間之聯繫程度.

當結果之應用以實際上研究之材料爲限時,此各種之統計計量,僅爲敍述一個頻數分配之某幾種形相或敍述某種關係之方法.在此範圍以內,吾人可十分確信此等計量爲各種特性之確切敍述.惟當吾人欲推廣此結果,使結論普遍化,用之於原來材料所未包括之事例時,則遇另一列問題存在.

在邏輯上,由特殊事例之硏究求普遍之結論,謂之歸納;以與演繹(Deduction)相對,後者乃由若干普遍之原則求出特殊結論也.所謂統計歸納者,即假定由一定材料所求得之某一統計計量可適用於以外較廣之材料而將統計之結果普遍化(Generalization)也.此種方法乃吾人在實際統計工作中所不斷應用者,惟對於此法所暗中根據之假定與此法所受之限制,則非能常有清楚之認識,玆可就此種假定與限制而簡略研究之.

〔統計結果之普遍化〕

目前之問題爲研究統計歸納之有效性,可如下式提出:一種統計計量——一平均數,一頻數比率,一緊聯係數,——爲由一羣體(Population)中抽取若干材料研究而得者.(此地"羣體" Population 一詞指一集團之事物或現象,在意想上其中應有某種共同特性存在者.)若由同一羣體內另取其他範樣,所得之各種計量其數值亦能與前者相等乎?若不能相等,吾人果有法決定後來諸範樣所求得之計量大致與最初之範樣所得之結果之差異在何種範圍以內乎?此問題顯然爲非常重要者,潘蓀氏稱之爲"實用統計學之基本問題."若吾人不能確知後來諸範樣將得與最初之範樣相似之結果,——卽吾人若不能確知後來諸範樣之結果有某程度之穩定性(Stability),則由有限之事例研究所得之結果,推至其他事例時將爲無效,如是則任何研究,除非能盡包組成該羣體之一切事物或現象者,必將毫無意義.但盡包一切之研究,在經濟現象之研究中實不可能.物價指數,工資指數,生活費指數,描寫某商品生產量與價格間關係之方程式,氣温與農作物收成間之緊聯係數——凡此種種皆不得不以範樣之研究爲根據者也.

[自然一致性之假定]

由後來諸範樣求得之統計計量,若欲視爲穩定,吾人須先能承認下列兩假定爲確實有效:其一,以平常之語出之,卽自然一致性(Uniformity in nature)之假定是也;卽假定在自

然界內"個別事件變化差異之量,常有限制."當處理數量材料時,此自然一致之性表現爲大量惰性(Stability of large number),生產率或死亡率等現象所表現之奇特之有規則性,可爲此種惰性之代表.換言之自然界並非毫無秩序者,在自然界之一切過程中,皆表示有規則性,有秩序,有穩定性之原則存在,而此等原則在吾人處理大量之數量材料時乃甚爲明顯者也.因此當吾人將批發物價指數等計量普遍化時,即爲假定在吾人所計量之特性或關係之一點上,在此計量所應用之較大羣體內有共同一致之性存在.因有此共同一致之性存在,故若在此羣體內另抽許多範樣,則求得之統計計量與第一範樣統計計量之差,將在一定範圍之內,而此範圍乃吾人事先可以大致決定者.

此種假定,當然不能以純粹統計方法完全證明.故統計歸納中總有一先在之因素存在.統計上之結論,從不能完全獨立,必須藉理性與判斷之助始能令人確信.如取一香蕉輸入量與自殺率之範樣而研究之,見其中有高度之正聯繫係數,此點並不能使人確信在此兩變數之間,有何因果關係或臨時關係存在也.在取得此範樣之較大現象界中,吾人毫無理由假定在此種關係上能有共同一致之性也.

[範樣必須有代表性]

由羣體中抽取一範樣,取其統計計量用諸全羣體,此即普遍推理之法也.其有效與否尚受第二個條件所限制,此條

件在前面之討論中已暗中存在,即吾人所根據以求得第一批結果之範樣,必須足以完全代表此結果應用所及之整個羣體.可靠之統計歸納工作,以覓得此種有代表性之範樣爲其首要之條件.

惟此種範樣將如何獲得乎?公認爲主要之點者,即範樣所包括之事例,應爲全羣體中之任機事例(Random members),即選擇範樣時,對於一大集團之事例中,不應對於其中若干之取舍存好惡之成見也.全羣體之每一事例,應皆有同樣之機會參加其內.此總要件,誠如開恩斯氏所指示,應作如下之解釋,即關於普遍推理之一點上,範樣所包含者應爲全羣體中之任機事例.作純粹任機選擇時常須煞費苦心.若任取最易得之事例,表面似合任機選擇之條件,其實決非如是.若不於選擇事件時嚴密注意,無意中含有偏誤,則將求得之結論用之於事實之全領域時,可因而抹殺若干重要之因素.

[簡單抽樣之條件]

游爾氏在統計理論導言一書中(259—261),曾另舉若干條件,爲引申屬於抽樣過程之公式時所暗中假定者."簡單抽樣"(Simple sample)法所根據之假定爲:

其一,爲假定組成此範樣之個別事件,彼此完全獨立.如當一商品之價格變動,受另一種商品之價格變動所影響時,此兩種價格變動即非彼此完全獨立,若將兩者共納於一個範樣中,即違犯簡單抽樣之條件.

其二,爲假定範樣中所包括之許多例項,其所由收集之各地間無重要差別存在;且在觀察值所涉及之時期內,潛在的情況未有重要變更.

其三,爲假定吾人所觀察之特性,支配其出現之條件,不僅對每一範樣,且對範樣中每一個別事例,莫不一致.

所有此等抽樣之條件,如已全部遵照,則各種統計計量之由同一羣體諸不同範樣求得者,其數值變動之範圍,吾人可事先確定之.從一個範樣求出之統計計量,吾人將其應用於全羣體時,並非信其能完全適合也,惟將結論所應用之範圍如此推廣時,吾人對於其發生之錯誤之界限,已有相當確切之認識矣,必須之條件如能遵守,統計歸納固一有效之方法也.

[可靠性計量之意義]

吾人所求之可靠性計量,其本質如何,必須有清楚認識.當吾人以對於若干統計數量之知識爲根據,而作普遍之推理時,卽將由一個範樣計算得之中數,標準差,或緊聯係數等值,用於範樣以外之事例集團.然吾人若由同一羣體抽取若干不同之範樣而求此等數值時,結果之相差將在何種範圍以內乎?計量此範圍之數值,卽最初之結果推廣其應用之範圍時之可靠性計量也.

用一笨重之方法,卽研究無數不同之範樣,有如擲骰至4096次者,亦可求得此計量.例如有一平均數代表某種工人

每週工資額者,欲求一數以驗其可靠性.如此平均數爲根據第一個範樣求出,其中包含250個記錄,吾人可更取499個範樣,每個,亦包含250個記錄,而計算出499個平均數,平均數總共500個.彼此之值不等,惟列爲一表時,將成一非常接近於正態形式之頻數分配.更據此分配求得500個平均數之平均數(算術中數),以及其標準差.此標準差乃每週工資平均數(由相繼諸範樣計算而得者)之離中趨勢量也.

然繼續抽取四五百個範樣以驗某一計量之可靠性,實際上常爲不可能之事,當合於上述簡單抽樣諸條件時,此種可靠性計量,吾人有較直接之法計算之.

前章曾舉一例,將十二個骰子擲4096次,得五點四點六點皆視爲成,試驗之結果曾列一表,由結果求得標準差1.712.但吾人知已知成敗之機率與每次擲骰之個數時,可由下列公式

$$\sigma=\sqrt{npq}$$

計算得理論的標準差爲1.732.

用類似之方法,可計算得同樣性質之數量,在關於抽樣變動之範圍以內,可作爲統計結果之穩定性(可靠性)之計量.此各種之可靠性計量或標準誤,可列舉於下,惟不欲指示其所由求得之方法.欲知詳細之方法請參閱後列參考書目.

[諸主要統計計量之標準誤]

由某一範例求得之算術中數,其可靠性視原來材料之標準差之數值與範樣中所包括之事例個數而定,其確定之關係爲

$$\sigma_M=\frac{\sigma}{\sqrt{N}}$$

(σ右下角不註小字者指該範樣之標準差).此爲以標準誤表示該結果之可靠性者.惟用機差表示者更爲普通.吾人前已提及正態分配之標準誤常爲 $.6745\sigma$. 故得

$$P.E._{Mean}=.6745\frac{\sigma}{\sqrt{N}}$$

茲可舉例以說明此式之意義.假定吾人欲求某城市中十萬個工業勞動者每週之平均工資額,如欲包括全體勞動者而研究之,爲事勢所不能,故抽取一包括 900 個工人之代表範樣,得每週平均工資額爲 \$27.50. 此數可謂正確乎?若更取若干範樣而求平均數,其數值與此之差將如何?假定所研究之材料其標準差爲\$2.00,則得

$$P.E._{Mean}=.6745\frac{\$2.00}{\sqrt{900}}$$
$$=\$.045$$

當報告原來之研究結果時,每週平均工資將爲\$.27.50±.045. 即吾人若另取一大小相同之範樣,結果與 \$.27.50 相差至\$.045以上之機會,則占一半.

其他統計計量之可靠性,亦可同樣求一數值以計量之.下列諸公式乃求可靠性計量中最重要者之數值時所應用者也.(每處之機誤皆可以標準誤乘.6745求得之.)

$$\sigma_{Md}=1.25331\frac{\sigma}{\sqrt{N}} \qquad P.E._{Md}=.84535\frac{\sigma}{\sqrt{N}}$$

$$\sigma_{Q1}=1.36263\frac{\sigma}{\sqrt{N}} \qquad P.E._{Md}=.91908\frac{\sigma}{\sqrt{N}}$$

(Q_3與Q_1之可靠性計量數值相同.)

$$\sigma_{\sigma}=\frac{\sigma}{\sqrt{2N}}$$

此爲求標準差之標準誤之公式,僅於範樣中觀察值其分配爲正態時,始爲確切適用.惟下列之公式,可以適用於任何形式之分配,以決定 σ 之可靠性:

$$\sigma_{\sigma}=\sqrt{\frac{\mu_4-\mu_2^2}{4\mu_2\cdot N}}$$

正態分配中繫聯係數之可靠性可以下式計之:

$$\sigma_r=\frac{1-r^2}{\sqrt{N}}$$

此公式亦可用於複繫聯係數與偏繫聯係數,並可近似的用於繫聯比.

響應係數之標準誤,若遇正態分配,可由下式決定之:

$$\sigma_{b12}=\frac{\sigma_1\sqrt{1-r^2}}{\sigma_2\sqrt{N}}$$

此式可適用於任何次之響應係數.

前述測驗直線性的公式為

$$\zeta=\eta^2-r^2$$

惟吾人欲知某一例中 η^2 與 r^2 間之差為由於偶然之抽樣變動,抑由於其關係根本與直線性大有出入,勃萊克曼(Blakeman)氏提議以下式計算 ζ 之標準誤.

$$\sigma_\zeta=2\sqrt{\frac{\zeta}{N}}\sqrt{(1-\eta^2)^2-(1-r^2)^2+1}$$

前面曾有一章中舉一小麥產量之問題,今可用以說明此計量之用法,吾人已求得小麥產量與施用之氮素肥料量間之關係為

$$r=+.793$$

$$\eta=.964$$

故

$$\zeta=\eta^2-r^2=.300$$

將已知值代入

$$\sigma_\zeta=2\sqrt{\frac{\zeta}{N}}\sqrt{(1-\eta^2)^2-(1-r^2)^2+1}$$

解之得

$$\sigma_\zeta=.074$$

此地 ζ 之值爲.300,約爲其標準誤(.074)之4.05倍,關係之爲非直線形式,可無問題.而 η^2 與 r^2 間之差殆不可能爲偶然之抽樣變動所引起者也.

在正態曲線,X 之標準誤(即偏態計量之標準誤),由下列公式求得之:

$$\sigma_X=\sqrt{\frac{3}{2N}}$$

將此式用於前章所舉電話通戶次數之分配,得:

$$\sigma_X=\sqrt{\frac{3}{1990}}=.03883$$

前知 X 之值爲 $-.05558$,約爲其標準誤之1.43倍.故 X 之此值,大可爲抽樣變動所引起者,並不能證明眞正之分配爲不合於正態也.

計量算術中數與衆數間距離之數量 d,當遇正態曲線時,其標準誤可由下列公式決定之:

$$\sigma_d=\sqrt{\frac{3}{2N}}\sigma$$

將此式用於電話通話分配,得

$$\sigma_d=\sqrt{\frac{3}{1990}}147.65$$

$$=5.733$$

因 d 之值爲8.21,不及其標準誤之一倍半,故吾人可得結論,謂此處算術中數與衆數間之差,大可由於偶然之抽樣變動所引起.若 d 之值大於其標準誤三倍以上,則算術中數與衆數間之差或非單由於偶然之抽樣變動,而爲代表該分配之形式與正態形式有根本出入之處在也.

當將兩個平均數互相比較時,常遇一相當重要之實際問題,若取兩個範樣,各求其平均數,結果不同;則此兩範樣間之差異究竟僅由於抽樣變動乎?抑指示在此兩範例所由抽取之兩大事例集團間有眞正之差異在乎?欲決定此兩平均數間之差之標準誤,用下列公式:

$$\sigma_D=\sqrt{\frac{\sigma_1^2}{N_1}+\frac{\sigma_2^2}{N_2}}$$

式中 σ_1 與 σ_2 代表兩範樣之標準誤,N_1 與 N_2 代表兩範樣各各包括之例數.應用此計量時,兩平均數間實在差數與此差數之標準誤相比較,若實在差數大於標準誤之三倍,則謂此差異爲起於抽樣變動,殆不可能.遇此種情形,即證明此兩範樣所由抽取之兩大事例集團間有根本差異存在也.

用以比較理論頻數與實在頻數之抽樣標準差,其公式

$$\sigma_s=\sqrt{\frac{f(N-f)}{N}}$$

其用法已於前章說明.

上述各種可靠性計量,多半以標準誤表示而不用機誤表示.從標準誤求機誤,其法甚易,前已言之矣.惟在一般之應用上,似以標準誤較爲適宜.最精細之機率積分表,横坐標之距離用標準差爲單位.任何一種標準誤,用機率表示時其意義若何,根據此種之表,甚易決定.

[可靠性計量所受之限制]

從上列諸公式求得之各種可靠性計量,當吾人應用時,應充分了解其所受之限制.第一,此等計量須於範樣包括有充分大量之事例時,始有意義.N若低於15,即不宜應用上列求標準誤之各種公式;吾人對於所得之結果,若認爲非常重要;則範樣中包括之事例,其數尙大須增加.在繫聯係數,則以25個事例爲最低限度.蓋在解釋此等標準誤時,實有一假定存在,卽假定繼續抽取範樣時求得之統計計量,將形成一合於差誤正態定律之分配是也.如包括之事例爲數極多,卽使原來之材料不成如此之分配,此假定亦可與眞相近似.然事例若極少.此假定卽可十分不確,而標準誤卽不能用機率表示其確切之意義.

尙有最須注意之一點,卽吾人假定標準誤所計量之差誤,僅以簡單抽樣變動所引起者爲限也.此假定使任何統計歸納法之應用皆帶有若干可疑之成份.第一,如前面所舉之簡單抽樣之條件,是否已完全遵照,吾人無從確證.在處理經濟材料之時,能完全遵照此等條件者,雖或有之亦必甚少.第

二,引起後來諸範樣之變動者,除簡單抽樣之原因外,是否尚有其他原因存在?各種標準誤之如上述者,對於此點皆不能作何指示.當用此法以決定或然的可靠性時,凡由於偏見,由於抽樣過程中之非任機選擇,由於其他任何種類之經常的錯誤所引起之變動,全非吾人之所能察覺.故統計研究之結果,如欲與以適當之解釋與應用,則此等注意之點不可不時留腦際也.

平常之機誤計量,用於經濟材料時,其所受限制甚嚴,故此種計量之可靠性,似常宜以實際之統計測驗之:續取諸範樣而研究之,分割一範樣爲若干小組而測驗其附屬之因子,此種之研究方法,關於某一計量之可靠性與其普遍應用之可能性,可供獻甚多之知識;比較將機誤之平常數學公式不加思索不加批判接受而應用之者,裨益實多矣.

參考書

Bowley, Arthur L. Elements of Statistics (312—342)

Broad, C. D. On the Relation between Induction and Probability. Mind. N. S. Vol. 27, 1928, 及 Vol. 29, 1920.

Editorial. On the Probable Errors Frequency Constants Bismetrika. Vol. 2 (273--281)

Elderton, W. P. Frequency Curves and Correlation (131—138).

Fisher, Arne. The Mathematical Theory of Probabilities.

Kelley, Truman L. Statistical Method (94—109).

Keynes, J. M. A Treatise on Probability.

Mills, Frederick C. On Measurement in Economics (載 Tugwell R. G. 所編 The Trend of Economics, 37—70)

Pearl, Raymond. Medical Biometry and Statistics (209—219)

Pearson, Karl. The Fundamental Problem of Practical Statistics. Biometrika. Vol. 13

Rietz, H. L. Random Sampling (載 Rietz H. L. 編 Handbook of Mathematical Statistics, 71—81)

Whitaker, E. T. and Robinson, G. The Calculus of Observations (164—208)

Yule, G. U. An Introduction to the Theory of Statistics (250—290, 335—356)

附　錄

一　應用於統計上之最小平方法

最小平方法用於處理一個未知量時,僅爲從若干獨立之觀察值求該未知量之最或然值之法,最或然值爲如此之值,卽諸觀察值對於此值之差離(或餘量 Residuals),作成平方而相加時,其總和爲一最小數者也.此值卽諸觀察值之算術中數.

若觀察值非直接指一個未知量,而指若干未知量之函數時,問題稍有不同,在前種情形中,每一觀察皆形成一個單獨之量,在此地,則每一觀察形成一觀察方程式.在此方程式中,包含結合一體之諸變數之觀察值,此地之未知量,爲表示諸變數間的函數關係之若干常數.吾人之問題爲求此等常數之最或然值,其實在之值乃吾人所不知者也.

此地之最或然值,與前種較簡單之情形相同,其餘量之平方之總和亦爲一最小數,惟此地之餘量,非對於一個單獨的量之差離,如最或然量爲算術中數時也.此地之餘量,乃對於一條表示最或然的函數關係之曲線而言.換言之,餘量爲

應變數之計算值與實在值(觀察值)間之差.

[正則方程式之由來]

茲以 Y 代表應變數之一個觀察值,以 Y_c 代表相當之計算值,以 v 代表餘量,或 Y 與 Y_c 間之差,以 W_1, W_2, W_3 與 W_4 代表各個自變數(或一個自變數之幾個不同的函數).則

$$Y_c = f(W_1, W_2, W_3, W_4)$$

$$v = Y_c - Y$$

$$= f(W_1, W_2, W_3, W_4) - Y$$

$$\Sigma(v^2) = \Sigma[f(W_1, W_2, W_3, W_4) - Y]^2$$

若某一例之函數爲下式

$$Y_c = aW_1 + bW_2 + cW_3 + dW_4$$

則

$$\Sigma(v^2) = \Sigma[(aW_1 + bW_2 + cW_3 + dW_4) - Y]^2$$

吾人之問題,爲決定諸常數之最或然值,此等常數乃確定函數關係者也.此等常數在本例以 a, b, c 與 d 代表之.(吾人須知一有觀察方程式,諸 W 之值卽爲已知者.在平常之情形中,W 爲一個自變數之幾個不同的函數,但此非必要之事.)假定觀察差誤之分配合於差誤正態曲線,吾人可以證明上列方程式中 a, b, c, d 之最或然值,爲使 $\Sigma(v^2)$ 變成一最小數之數值.

$$\Sigma[aW_1 + bW_2 + cW_3 + dW_4) - Y]^2 = \text{一個最小數} \qquad \text{(A)}$$

求常數時所必須之正則方程式,可用下法求得之.卽將

(A)式對未知量 a,b,c,d 之偏微分係數使等於 0.換言之,先將 b,c,d 作爲常數求上列函數對於 a 之微分係數,再將 a,c,d 作爲常數,求上列函數對於 b 之微分係數;再將 a,b,d 作爲常數,求上列函數對於 c 之微分係數;再將 a,b,c 作爲常數,求上列函數對於 d 之微分係數.如是進行時,求得方程式(A)對於 a 之偏微分係數爲

$$\frac{\partial}{\partial a}\Sigma[(aW_1+bW_2+cW_3+dW_4)-Y]^2=0$$

或

$$\text{I}\quad \Sigma W_1[(aW_1+bW_2+cW_3+dW_4)-Y]^2=0$$

求得方程式(A)對於 b 之偏微分係數爲

$$\frac{\partial}{\partial a}\Sigma[(aW_1+bW_2+cW_3+dW_4)-Y]^2=0$$

或

$$\text{II}\quad \Sigma W_2[(aW_1+bW_2+cW_3+dW_4)-Y]^2=0$$

求得方程式(A)對於 c 之偏微分係數爲

$$\frac{\partial}{\partial a}\Sigma[(aW_1+bW_2+cW_3+dW_4)-Y]^2=0$$

或

$$\text{III}\quad \Sigma W_3[(aW_1+bW_2+cW_3+dW_4)-Y]^2=0$$

求得方程式(A)對於 d 之偏微分係數爲

$$\frac{\partial}{\partial a}\Sigma[(aW_1+bW_2+cW_3+dW_4)-Y]^2=0$$

或

$$\text{IV}\quad \Sigma W_4[(aW_1+bW_2+cW_3+dW_4)-Y]^2=0$$

聯立 I, II, III, IV 四方程式而解之卽得 a, b, c, d 之最或然値.

[正則方程式之求法]

若觀察方程式皆爲一次方程式(卽式中之未知量 a, b, c, 等皆爲一次),正則方程式可以下法求之.

1. 寫一方程式代表假定中之關係.將諸變數之相聯的觀察値代入,得各個觀察方程式.

2. 將每一觀察方程式中第一個未知數之係數乘所屬之觀察方程式;將乘得之許多方程式相加,總和卽成爲第一個正則方程式.

3. 將每一觀察方程式中第二個未知數的係數乘所屬之觀察方程式;將乘得之許多方程式相加,總和卽成爲第二個正則方程式.

繼續用此法求其他正則方程式,直至其個數與未知數之個數相等.

配合曲線時之正則方程式,其實際上之求法尙可簡單化,避免如以前曾表明之情形,將觀察方程式一一寫出.下列各條,可以作爲求正則方程式之通則

1.將所欲配合的曲線之正程式寫出.爲說明此法起見,可用一普遍式如下

$$Y=aW_1+bW_2+cW_3+dW_4+\cdots\cdots \qquad (1)$$

式中Y代表應變數,$a,b,c,d,\cdots\cdots$代表問題中之常數(在此地爲未知量),而$W_1,W_2,W_3,W_4,\cdots\cdots$代表此等未知量之係數.此等係數假定爲變數,以變數稱之.此爲方程式(1).

2.將方程式(1)之第一個未知量的係數(即W_1)乘(1)中每項,加上總和記號Σ於每一變數之前.此爲第一個正則方程式(I).

3.將方程式(1)之第二個未知量的係數(即W_2)乘(1)中每項,置總和記號於每一變數之前.此爲第二個正則方程式(II).

4.將方程式(1)中第三個未知量的係數(即W_3)乘(1)中每項,置總和記號於每一變數之前,此爲第三個正則方程式(III).

5.將方程式(1)中第四個未知量的係數(即W_4)乘(1)中每項,置總和記號於每一變數之前,此爲第四個正則方程式(IV).

繼續進行,直至求得之正則方程式,其個數等於未知量之個數爲止.

[一組標準的正則方程式]

用上法可求得一組普遍的正則方程式,對於任何方程

式之可以作成

$$Y=aW_1+bW_2+cW_3+dW_4+\cdots\cdots$$

之形式者,皆可適用.玆列如下:

I　$\Sigma(W_1Y)$

$$=a\Sigma(W_1^2)+b\Sigma(W_1W_2)+c\Sigma(W_1W_3)+d\Sigma(W_1W_4)+\cdots\cdots$$

II　$\Sigma(W_2Y)$

$$=a\Sigma(W_1W_2)+b\Sigma(W_2^2)+c\Sigma(W_2W_3)+d\Sigma(W_2W_4)+\cdots\cdots$$

III　$\Sigma(W_3Y)$

$$=a\Sigma(W_1W_3)+b\Sigma(W_2W_3)+c\Sigma(W_3^2)+d\Sigma(W_3W_4)+\cdots\cdots$$

IV　$\Sigma(W_4Y)$

$$=a\Sigma(W_1W_4)+b\Sigma(W_2W_4)+c\Sigma(W_3W_4)+d\Sigma(W_4)+\cdots\cdots$$

對於任何形式之曲線,凡可以應用最小平方法配縮者,此等方程式皆可以適合,僅需將該處所用之特殊函數代入W_1, W_2, W_3, $W_4\cdots\cdots$等也.如欲配合一曲線,其方程式爲

$$Y=a+bX+cX^2+dX^3$$

者,代入上列標準正則方程式時,所根據之關係如下:

$$W_1=1$$

$$W_2=X$$

$$W_3=X^2$$

$$W_4=X^3$$

正則方程式應有之改變極明顯.$\Sigma(W_1Y)$變爲$\Sigma(Y)$;$\Sigma(W_1^2)$等於$\Sigma(1^2)$,等於N(觀察之總數)第一個正則方程式變爲

$$\Sigma(Y)=Na+b\Sigma(X)+c\Sigma(X^2)+d\Sigma(X^3)$$

其他正則方程式亦同樣改變.

上舉之例中,係數爲一個單獨之自變數X之一切不同的函數.此種情形,並非最小平方法所必要者.係數W_1,W_2,W_3等亦可代表若干自變數,如複繫聯之情形即爲如是.

應用最小平方法時,須記住此法所受之限制.此法在直接應用時,僅限於吾人所欲配合之曲線其方程式就其中常數言爲直線式者;換言之,一切觀察方程式就其未知量a,b,c等而言必須爲直線式.(當然非謂配合曲線之方程式必須爲直線式也.)舉例明之,如以$y=ab^{cx}$爲方程式之曲線即不能直接用最小平方法配給.觀察方程式若非直線式,在許多情形中,常可用對數先化之爲直線形式,然後再應用最小平方法.

[估量標準誤之公式之由來]

在本書正文中曾指出估量標準誤可作爲最小平方法之副產.此法可於此詳細證明之.

當使下式

$$\Sigma[(aW_1+bW_2+cW_3+dW_4)-Y]^2=\text{最小數}$$

對於第一未知量a之偏微分係數(Partical derivative)等於0時,

$$\Sigma W_1[(aW_1+bW_2+cW_3+dW_4)-Y]=0$$

因

$$aW_1+bW_2+cW_3+dW_4-Y=v$$

故配合曲線之必須條件爲

$$\Sigma(vW_1)=0$$

當使同式對於第二未知量 b 之偏微分係數等於0時,

$$\Sigma W_2[(aW_1+bW_2+cW_3+dW_4)-Y]=0$$

照前例代入 v,得

$$\Sigma(vW_2)=0$$

用同樣方法施之 c 與 d,得

$$\Sigma(vW_3)=0$$

與

$$\Sigma(vW_4)=0$$

結論:當應用最小平方法決定若干未知量之最或然值時,此等未知量若以 W_1, W_2, W_3, W_4 爲已知之係數,則下列諸關係爲最小平方法之必須條件:

$$\Sigma(vW_1)=0$$

$$\Sigma(vW_2)=0$$

$$\Sigma(vW_3)=0$$

$$\Sigma(vW_4)=0$$

既知此種關係,吾人乃得一種方法,極易求得 $\Sigma(v^2)$ 之值與估量標準誤之值.假定下列方程式

$$Y_e=aW_1+bW_2+cW_3+dW_4$$

中之常數吾人已用最小平方法求得.則每一個餘量之關係可以下式示之:

$$v=aW_1+bW_2+cW_3+dW_4-Y \tag{1}$$

用 v 乘全式,並將所有如此乘過後之餘量方程式總加得

$$\Sigma(v^2)=a\Sigma(vW_1)+b\Sigma(vW_2)+c\Sigma(vW_3)+d\Sigma(vW_4)-\Sigma(Yv) \tag{2}$$

但

$$\Sigma(vW_1)=0$$

$$\Sigma(vW_2)=0$$

$$\Sigma(vW_3)=0$$

$$\Sigma(vW_4)=0$$

故

$$\Sigma(v^2)=-\Sigma(Yv) \tag{3}$$

每一方程式(1)皆以 Y 乘之,相加,得總和爲

$$\Sigma(Yv)=a\Sigma(W_1Y)+b\Sigma(W_2Y)+c\Sigma(W_3Y)+d\Sigma(W_4Y)-\Sigma(Y^2) \tag{4}$$

將 $\Sigma(Yv)$ 之相當值代入(3),得

$$\Sigma(v^2)=\Sigma(Y)^2-a\Sigma(W_1Y)-b\Sigma(W_2Y)-c\Sigma(W_3Y)-d\Sigma(W_4Y)-\cdots \tag{5}$$

根據此式求 $\Sigma(v^2)$ 之値,可不必計算個別之餘量.凡所欲配合之曲線,其方程式爲下式,或用對數、倒數或其他方法可以化爲下式

$$Y=aW_1+bW_2+cW_3+dW_4$$

時,此法即可應用.將其用於一特殊之例時,僅須將原來方程式中未知量的係數所代表之函數代替此地 W_1, W_2, W_3, W_4.

等之地位.如配合一曲線,其方程式爲

$$Y=a+bX+cX^2+X^3$$

時,前已知

$$W_1=1$$

$$W_2=X$$

$$W_3=X^2$$

$$W_4=X^3$$

代入前列之方程式(5),得

$$\Sigma(v^2)=\Sigma(Y^2)-a\Sigma(Y)-b\Sigma(XY)-c\Sigma(X^2Y)-d(X^3Y) \quad (6)$$

標準誤 S_y 乃由下列方程式求得者.

$$S_y=\frac{\Sigma(d^2)}{N}^{(\text{註})}$$

式中 d 代表對於配合曲線之一個差離,於是差離 d 不過爲餘量 v 之另一名稱,故當 W_1, W_2, W_3, W_4 爲自變數時,Y 之標準誤之普遍式爲

$$S_y^2=\frac{\Sigma Y^2-a\Sigma(W_1Y)-b\Sigma(W_2Y)-c\Sigma(W_3Y)-d\Sigma(W_4Y)}{N}$$

(註) 茲以計量配合曲線周圍實際"分散"程度爲目的,故不用 $\frac{\Sigma(d^2)}{N-Nc}$,而專用 $\frac{\Sigma(d^2)}{N}$ 爲公式.(第一式中 N 爲觀察個數,Nc 爲配合曲線之方程式中常數之個數).依照最小平方法之理論若吾人欲決定一個觀察值或一個觀察方程式之平方中數差誤(Mean square error)時,應用第一式.

應用於特殊之例時,將原來方程式中未知量之係數代替 W_1, W_2, W_3, W_4 等,與以前之例同.

[繫聯指數之公式之由來]

假定有單獨之一個應變數 Y,單獨之一自變數 X,爲計量此兩變數間之關係程度起見,吾人曾以 ρ(Rho) 爲指數.此 ρ 乃根據下式求得者:

$$\rho^2_{yx}=1-\frac{S_y^2}{\sigma_y^2} \tag{8}$$

當單獨一個應變數 Y 與若干自變數 W_1, W_2, W_3, W_4 相聯繫時,繫聯指數之公式寫作

$$\rho^2_{y\cdot w1w2w3w4}=1-\frac{S_y^2}{\sigma_y^2} \tag{9}$$

若代表變數用之記號改變,指數後加註之小字亦須隨之改變.上列之式等於

$$\rho^2_{y\cdot w1w2w3w4}=1-\frac{\Sigma(d^2)}{\Sigma(y^2)}$$

其中 y 代表當以 Y 諸值之中數爲原點時每一 Y 值對於此原點之一個差離,但

$$\frac{\Sigma(y^2)}{N}=\frac{\Sigma(Y^2)}{N}-c_y^2$$

其中 Y 代表 Y 變數原來諸值,c_y 代表原來之原點(即0點)與 Y 諸值的算術中數之差.(c_y 與 c_x 兩記號切勿與方程式中常

數之一之 c 相混.)

故得

$$\rho^2_{y\cdot w_1w_2w_3w_4}=1-\frac{\Sigma(d^2)}{\Sigma(Y^2)-Nc_y^2} \tag{10}$$

前面吾人已求得一 $\Sigma(v^2)$ 之方程式($\Sigma(v^2)$ 與 $\Sigma(d^2)$ 相同),凡曲線爲用最小平方法配繪之處皆可適用者.茲先研究普遍之情形,設有若干個自變數,將與 $\Sigma(d^2)$ 相當之式代入上列方程式,則

$$\rho^2_{y\cdot w_1w_2w_3w_4}=1-\frac{\Sigma(Y^2)-a\Sigma(W_1Y)-b\Sigma(W_2Y)-c\Sigma(W_3Y)-d\Sigma(W_4Y)-\cdots}{\Sigma(Y^2)-Nc_y^2}$$

化簡之得繫聯係數之普遍式如下:

$$\rho^2_{y\cdot w_1w_2w_3w_4\cdots\cdots}=\frac{a\Sigma(W_1Y)+b\Sigma(W_2Y)+c\Sigma(W_3Y)+d\Sigma(W_4Y)+\cdots-Nc_y^2}{\Sigma(Y^2)-Nc_y^2} \tag{11}$$

當此式用於特殊之例時,須將原來方程式中未知量的係數所表示之函數代替式內 W_1, W_2, W_3, W_4 等.若所有此等係數爲一個單獨之自變數之幾個函數.(平常之情形皆爲如此),則繫聯指數可用 ρ_{yx} 代表之.

〔若干特殊情形〕

複繫聯中,不論自變數或應變數皆以 X_1, X_2, X_3, X_4 等等記號代表,繫聯計量用 R 代表,用數字註其後,此點在本書正文中已有敍述.

兩變數間關係爲直線式時,以 r 代替 ρ 作記號. r 所代表者爲平常之繫聯係數.其普遍式如下:

$$r=\frac{a\Sigma(Y)+b\Sigma(XY)-Nc_y^2}{\Sigma(Y^2)-Nc_y^2}$$

在兩種特殊情形中,此公式可以化簡.當以 X 諸值之算術中數爲原點時

$$a=c_y=\frac{\Sigma Y}{N}$$

$$a^2=c_y^2=\frac{a\Sigma Y}{N}$$

$$Nc_y^2=a\Sigma Y$$

而 r 之公式化爲

$$r^2=\frac{b\Sigma(xY)}{\Sigma(Y^2)-Nc_y^2}.$$

若以 Y 諸值之算術中數爲原點(不必同時在 X 諸值之算術中數上.)

$$\Sigma(y)=0,\ c_y=0$$

而繫聯係數之公式變爲

$$r^2=\frac{b\Sigma(Xy)}{\Sigma(y^2)}$$

在此後面的一種情形中,ρ 之普遍式亦可以簡單化,因其中 $a\Sigma(y)$ 與 Nc_y^2 兩項可以消去也.

[對於正則方程式的構成之覆驗]

當構成與分解一組正則方程式時,計算上發生錯誤之機會極多,故無論何處能用覆驗時卽應覆驗,直至構成正則方程式爲止之一切計算,有一簡便之法可以覆驗之,其法卽在每一觀察方程式中加入一項 s, 其値等於每一方程式中之一切已知量之和.例如有 1,3; 2,4; 3,6; 4,5; 5,10; 6,9; 7,10; 8,12; 9,11; 九對數値,形成九點,欲配合一線,對於其九個觀察方程式之 s 値如下:

	s
$3=a+1b$	5
$4=a+2b$	7
$6=a+3b$	10
$5=a+4b$	10
$10=a+5b$	16
$9=a+6b$	16
$10=a+7b$	18
$12=a+8b$	21
$11=a+9b$	21

(各式中 a 之係數皆爲 1,此數亦加在其他已知値也.)

配合之曲線其方程式爲下式之一類者,

$$Y=aW_1+bW_2+cW_3+dW_4$$

s 與其他計算値之間有下列之關係存在.在每一個觀察方程式,其關係爲

$$Y+W_1+W_2+W_3+W_4=s$$

在正則方程式,其關係爲

$$\Sigma(W_1Y)+\Sigma(W_1^2)+\Sigma(W_1W_2)+\Sigma(W_1W_3)+\Sigma(W_1W_4)=\Sigma(W_1s)$$
$$\Sigma(W_2Y)+\Sigma(W_1W_2)+\Sigma(W_2^2)+\Sigma(W_2W_3)+\Sigma(W_2W_4)=\Sigma(W_2s)$$
$$\Sigma(W_3Y)+\Sigma(W_1W_3)+\Sigma(W_2W_3)+\Sigma(W_3^2)+\Sigma(W_3W_4)=\Sigma(W_3s)$$
$$\Sigma(W_4Y)+\Sigma(W_1W_4)+\Sigma(W_2W_4)+\Sigma(W_3W_4)+\Sigma(W_4^2)=\Sigma(W_4s)$$

此式對於任何專門之問題皆可適用. s 方程式各按問題中之正則方程式用完全相同之方法化成.

應用此等覆驗式時,工作表中須添列幾行.惟每階段覆驗工作之機會,尚不能抵償外加之煩勞.下列一工作表表示在配合一曲線於 1,2; 2,6; 3,7; 4,8; 5,10; 6,11; 7,11; 8,9; 9,9 之九點時覆驗法之應用.曲線爲

$$Y=a+bX+cX^2$$

之二次拋物線.

表 A

表示覆驗正則方程式構成之法

Y	X	X^2	XY	X^2Y	s	Xs	X^2s
2	1	1	2	2	5	5	5
6	2	4	12	24	13	26	52
7	3	9	21	63	20	60	180
8	4	16	32	128	29	116	464
10	5	25	50	[illegible]	41	205	1025
11	6	36	66	396	54	324	1944
11	7	49	77	539	68	476	3332
10	8	64	80	640	83	664	5312
9	9	81	81	729	100	900	8100
74	45	285	421	2771	413	2776	20414

(X^3 與 X^4 兩行省去,因 $\Sigma(X^3)$ 與 $\Sigma(X^4)$ 之值可由現成之表求得也.)

s 一行中各值,各由其相當之觀察方程式求得,如由第一個觀察方程式

$$2=1a+1b+1c$$

得 s 之值為5(2加三個常數之係數).從表上求之,只須將在同一列之 Y,X 與 X^2 三行之數字相加再加上常數項 a 之係數1即得.

各行之總和求得後,即可用覆驗式如下徵驗計算之正誤:

$$\Sigma(Y)+N+\Sigma(X)+\Sigma(X^2)=\Sigma(s)$$
$$74+9+45+285=413$$
$$\Sigma(XY)+\Sigma(X)+\Sigma(X^2)+\Sigma(X^3)=\Sigma(Xs)$$
$$421+45+285+2,025=2,776$$
$$\Sigma(X^2Y)+\Sigma(X^2)+\Sigma(X^3)+\Sigma(X^4)=\Sigma(X^2s)$$
$$2,771+285+2,025+15,333=20414$$

後文討論正則方程式之解法時,尚須討論此種覆驗之其他應用.

[其他測驗]

前文曾提及用其他方法測驗計算之正誤之可能性.如以 W_1, W_2, W_3, W_4 代表配合曲線方程式中常數之係數,吾人知

$$\Sigma(vW_1)=0$$
$$\Sigma(vW_2)=0$$
$$\Sigma(vW_3)=0$$
$$\Sigma(vW_4)=0$$

若配合之曲線,其方程式爲下列之形式

$$Y=a+bX+cX^2+dX^3$$

則

$$\Sigma(v)=0$$

$$\Sigma(vX)=0$$

$$\Sigma(vX^2)=0$$

$$\Sigma(vX^3)=0$$

依照此種關係而覆驗之,可以測見工作之正確與否.

最後,吾人可用兩種不同的方法計算估量標準誤,以測驗工作之正確性.吾人可計算應變數之計算值與實在值間之差,如是得個別之餘量,更由此等餘量求S.再用上面求得之標準誤之普通公式求S,以兩者之結相比較.用表A之材料配合二次拋物線時,求得方程式爲

$$Y=-.92860+3.52316X-.267316X^2$$

用計算個別餘量之法求得之標準誤爲

$$S_y=.4941$$

用公式

$$S_y=\frac{\Sigma(Y^2)-a\Sigma(Y)-b\Sigma(XY)-c\Sigma(X^2Y)}{N}$$

得

$$S_y=.4947$$

此爲對於計算之正確性之最後的覆驗.

[正則方程式之解法]

解正則方程式之工作,一般上非經濟統計家之難事.未知數若僅有兩個或三個,則用簡單之代數方法即可解與之爲數相當之正則方程式.但即使正則方程式僅有三個,最好亦用一有系統之方法;若方程式超過三個以上,此種辦法更爲必要.此種系統的方法,用於解最小平方法中所遇之聯立方程式者,爲高斯氏及杜立脫爾氏所創.後者對於一般應用上或更爲便利,茲說明如下.

正則方程式中未知量之係數永以主要對角線爲軸兩方對稱.如求方程式

$$Y=aW_1+bW_2+cW_3+dW_4$$

中諸常數之最或然值時,正則方程式有四:

$$a\Sigma(W_1^2)+b\Sigma(W_1W_2)+c\Sigma(W_1W_3)+d\Sigma(W_1W_4)-\Sigma(W_1Y)=0$$

$$a\Sigma(W_1W_2)+b\Sigma(W_2^2)+c\Sigma(W_2W_3)+d\Sigma(W_2W_4)-\Sigma(W_2Y)=0$$

$$a\Sigma(W_1W_3)+b\Sigma(W_2W_3)+c\Sigma(W_3^2)+d\Sigma(W_3W_4)-\Sigma(W_3Y)=0$$

$$a\Sigma(W_1W_4)+b\Sigma(W_2W_4)+c\Sigma(W_3W_4)+d\Sigma(W_4)-\Sigma(W_4Y)=0$$

將Y諸項略去,其餘各項,顯然在對角線兩方成對稱之排列.由主要對角線上任何一項出發,其上面之一項與其左面之一項爲同一個係數.如對角線上係數爲$\Sigma(W_3^2)$之一項,向上觀之,有$\Sigma(W_2W_3)$與$\Sigma(W_1W_3)$兩係數,向左觀之,與之在同一橫線上者,亦爲同樣之兩係數,故在對角線上各項以左之項皆可省去,而將正則方程式寫作下式:

$$a\Sigma(W_1^2)+b\Sigma(W_1W_2)+c\Sigma(W_1W_3)+d\Sigma(W_1W_4)-\Sigma(W_1Y)=0$$

$$+b\Sigma(W_2^2)+c\Sigma(W_2W_3)+d\Sigma(W_2W_4)-\Sigma(W_2Y)=0$$

$$+c\Sigma(W_3^2)+d\Sigma(W_3W_4)-\Sigma(W_3Y)=0$$

$$+d\Sigma(W_4^2)-\Sigma(W_4Y)=0$$

[杜立脫爾氏法]

茲取正文第P.494頁上之正則方程式爲例,以說明杜立脫爾氏之解法.將諸式用上述之簡略式寫出時,得

$$8.3564b_{12\cdot34}+2.790b_{13\cdot24}+2.932b_{14\cdot23}+47.967=0$$

$$+6.6645b_{13\cdot24}+2.063b_{14\cdot23}+62.039=0$$

$$+7.7893b_{14\cdot23}+47.519=0$$

吾人欲解此等方程式求常數.爲便利起見,用 A 代表 $b_{12\cdot34}$, B 代表 $b_{13\cdot24}$, C 代表 $b_{14\cdot23}$. 一切計算工作以及必須之覆驗,皆列於下表.

表　B

用杜立脫爾氏法解正則方程式

列	(1) 倒數	(2) A	(3) B	(4) C	(5)	(6) s	
I		8.3564	2.790	2.932	47.967	62.0454	
II			6.6645	2.063	62.039	73.5565	
III				7.7893	47.519	60.3033	
1		8.3564	2.790	2.932	47.967	62.0454	
2	−.11966876	−1.000000	−.333876	−.350869	−5.740151	−7.424898	覆驗

表 B（續）

用杜立脫爾氏法解正則方程式

列	(1) 倒數	(2) A	(3) B	(4) C	(5)	(6) s
3			6.6645	2.063	62.039	73.5565
4			−.931514	−.978924	−16.015030	−20.715470
5			5.732968	1.084076	46.023970	52.841030
6	−.17442917		−1.000000	−.183094	−8.027923	−9.217017覆驗
7				7.7893	47.519	60.3033
8				−1.028748	−16.830133	−21.769807
9				−.204992	−8.702857	−9.991922
10				6.555560	21.986010	28.541571
11	−.15254227			−1.000000	−3.353796	−4.353796覆驗

C	B	A
−3.353796	−8.027923	−5.740151
−3.353796	+.634183	+2.468592
	−7.393740	+1.176743
		−2.094816

$$A = b_{12\cdot34} = -2.094816$$
$$B = b_{13\cdot24} = -7.393740$$
$$C = b_{14\cdot23} = -3.353796$$

覆驗：

方程式 I：　$8.3564 b_{12\cdot34} + 2.790 b_{13\cdot24} + 2.932 b_{14\cdot23} = -47.967$

代入已知值

$$8.3564(-2.094816) + 2.790(-7.393740) + 2.932(-3.353796) = -47.966985$$

說明——未知量 A, B, C 之係數列入表上指定之三行內.每一正則方程式之已知項列於第(5)行,(注意,已知項在表內之正負號,爲當其所屬之方程式等於0時所帶者.) s 行作覆驗之用.該行在 I, II, III 三列中之數值,爲各正則方程式中已知量之代數和.在求此代數和時,對角線以左現今省去之各項之係數必須包括在內.

下列各條爲解正則方程式之方法之摘要.

1.在第(1)列內記入正則方程式 I.

2.在第(1)行第(2)列內記入第(2)行第(1)列中數值之倒數,改變其符號.(此爲 A 之係數之倒數).以此倒數乘第(1)列內各項,將乘積記入第(2)列之相當行內.(第(2)列中第(2)(3)(4)(5)各行內之各項,其代數和應與同列第(6)行之值相等.)如此則已消去未知量 A,而以 $B\ C$ 表示之.(第(2)行第(2)列中之 -1 僅記入以便利覆驗者.第(6)(11)兩行中之 -1 亦然.)在第(2)列下可劃一粗線橫貫全表.

3.在第(3)列內記入正則方程式 II.

4.以第(2)列中 B 之係數(即 $-.333876$)乘第(1)列第(3)(4)(5)(6)各行中各項,將乘積記入第(4)列相當行內.

5.將第(3)列與第(4)列相加之和記入第(5)列,(第(5)列第(3)(4)(5)各行中各項之代數和應等於同列第(6)行之數值.)

6.將第(3)行第(5)列中數值之倒數,改變符號,記入第(1)行第(6)列,將此倒數乘第(5)列內各項,乘積記入第(6)列.(第(6)

列第(3)(4)(5)行中各項之代數和,應等於同列第(6)行之數值.)如此則已消去未知量B,而以C表示之.在第(6)列下可劃一粗線橫貫全表.

7.在第(7)列內記入正則方程式III.

8.將第(2)列內C之係數(即-.350869)乘第(1)列第(4)(5)(6)各行中各項.乘積記入第(8)列相當行中.

9.將第(6)列內C之係數(即-.189094)乘第(5)列第(4)(5)(6)各行中各項.乘積記入第(9)列相當行中.

10.將第(7)(8)(9)三列相加,其和記入第(10)列中.(第(10)列第(4)(5)兩行中兩數之和,應與同列第(6)行之數相等.

11.在第(1)行第(11)列中,將第(4)行第(10)列中數值之對數改變符號記入.將此倒數乘第(10)列中各項.乘積記入第(11)列內.(第(11)列第(4)(5)兩行之數值之代數和,應與同列第(6)行之數相等).如此則在第(5)行第(11)列內得C之數值.在第(11)列下,可劃一粗線橫貫全表.

若尚有其他未知量,如D與E時,則上面之運算最後所得之結果,僅爲將C作爲D與E之函數.故尚須繼續用此法前進,重複以上經過之步驟.吾人首須將方程式IV移下,記入第(12)列.於是第(2)(6)(11)各列中D之係數,又須用以與第(1)(5)(10)各列中必須之各項相乘,而記其乘積於第(16)列內;並用該列s行之數作覆驗.將第(16)列中D之係數之倒數,改變符號,乘該列各項,則D之值表現爲E之函數.至於E

之值可以同樣方法求之.

對於各步計算之覆驗,可於表上見之.計算時逐步隨時覆驗求得之結果,如此使錯誤之可能減至最小.

從表上求三個常數之值不難.從第(11)列得

$$C=-3.353796$$

從第(6)列得

$$B=-.189094C-8.027923$$

從第(3)列得

$$A=-.333376B-.350369C-5.740151$$

(第(6)行內各項,僅列入作覆驗之用.而第(2)(6)(11)各列中 −1.000000 一數,乃所以助覆驗之進行者.)

計算式可由表末見之.

最後的覆驗爲將此法求得之值代入正則方程式之一以觀其合符與否.表下之覆驗卽取方程式 I 時之情形.

參考書

Adams, Oscar S. Application of the Theory of Least Squares to the Adjustment of Triangulation. Special Bulletin No. 28, U. S. Coast and Geodetic Survey, 1915.

Brunt, David. The Combination of Observations.

Huntington E. V. Curve Fitting by the Method of Least Squares and The Method of Moments (載 Rietz, H. L.

Handbook of Mathematical Statistics, 62—70)

Merriman Mausfield. The Method of Least Squares.

Smith, Bradford B. The Use of Punched Card. Tabulating Epuipment in Multiple Correlation Problems. Washington, Burean of Agricultural Economics, 1923.

Weld, L. D. Theory of Errors and Least Squares.

Whitaker, E. T. and Robinson, G, The Calculus Observations (209—259)

Wright and Hayford. Adjustment of Observations.

二 統計符號

下列爲上文應用之最重要之各種統計符號,有時同一符號作數種用法,故其確切之意義須由其前後文決定之.

(一)一般代表變數與常數之符號:

x: 一變量(或變數)

y: 一變量(或變數)

一般言之,最後之幾個字母皆可用以代表一變量.幾個不同之變量亦可以一單獨之符號代表之,用小字註於右下方以示分別,如 X_1, X_2, X_3 或 $W_1 W_2 W_3$. [代表變量之字母,在後面大寫與小寫之間亦有分別:請參閱用於關係計量之符號].

a: 一常數(即其數值在討論之問題中不生變化之量是也.)一般以起首之幾個字母代表之.

(二)分析與敍述頻數分配時所用之符號:

m: 一個別觀察之數值;一組中點之數值.(有時用 a_1, a_2, a_3 等代表一數列中諸不同觀察之數值.)

f: 某一組內觀察值之個數;某一組之頻數.

i: 組距

l: 一組之下限

N: 某一數列或頻數分配中所包括之事例總數.

d: 一已知之觀察值對於一平均數之差離;常爲對於算術中數之差離.當寫作d_x或d_y時,爲x變數或y變數對於算術中數之差離.d有時亦用以代表算術中數與衆數間之差.

d': 一已知之觀察值對於一假定原點或假定算術中數之差離.

c: 一假定原點或假定算術中數與眞正算術中數間之差.$(c=M-M')$

Σ(Sigma): 總和之符號.如Σd爲全體差離之總和.

w_1,w_2,w_3: 求平時數時加於一數列中各項之權數.(不能與用以代表不同諸變量之相同之符號相混.)

y_o: 頻數曲線之最大縱坐標.

各種平均數,四分位數等之符號:

M: 算術中數,

Md:中位數.

M_o: 衆數.

M_g: 幾何中數.

H: 倒數中數.

M':假定算術中數之值.

Q_1: 第一四分位數或低四分位數.

Q_2: 第二四分位數或中位數.

Q_3: 第三四分位數或高四分位數

K: 在第一第三兩四分位數之中點上之數值.

D_3: 第三十分位數

各種離中趨勢與偏態計量之符號

$M.D.$: 平均差離(諸差離之算術中數)

σ: 標準差;即對於算術中數諸差離之平方之算術中數之方根.

s: 以算術中數以外之一點爲原點之標準差.

$P.E.$: 機誤

$Q.D.$: 四分位差

q_1: 中位數與低四分位數間之差.($Md.-Q_1$)

q_2: 高四分位數與中位數間之差.($Q_3-Md.$)

V: 離中趨勢係數.

sk: 偏態之計量.

X(Chi):根據判準 β_1 與 β_2 而作之偏態計量

(三)關於指數之符號

p_0': 某一商品在"0"時之價格(基期價格)

q_0': 同一商品在"0"時之數量(基期物量)

p_1': 同一商品在"1"時價格(計算期"1"之價格)

q_1': 同一商品在"1"時之數量(計算期"1"時之物量).

p_0'': 第二種商品在"0"時之價格.

q_0'': 第二種商品在"0"時之數量.

p_1'': 第二種商品在"1"時之價格.

q_1'': 第二種商品在"1"時之數量.

$\frac{p_1'}{p_0'}$: 一個價比(某一商品在"1"時之價格對於同一商品在"0"時之價格之關係.)

$\frac{q_1'}{q_0'}$: 一個量比(某一商品在"1"時之數量對於同一商品在'0"時之數量之關係.)

P_0: "0"時之價格水平

P_1: "1"時之價格水平

(四)計量關係時所用之符號

X: 一個變量之一個觀察值

Y: 一個變量之一個觀察值.(不同諸變量之觀察值亦可以用 X_1, X_2, X_3 或 W_1, W_2, W_3 等符號代表之.)

$\overline{X}$: X變量之若干觀察值之算術中數.(前文論複繫聯時曾在一證明中用 A_1, A_2, A_3……爲記號以代表 X_1, X_2, X_3……諸變數之算術中數,有時亦用 M_x 與 M_y 代表不同諸變數之算術中數.)

x: 一變量之值表示爲對於全體觀察值之算術中數之一差離者.(卽 X 數列之各項與其算術中數之差.) y 與 x_1, x_2, x_3……等記號之用法亦相同.至於原來之各觀察值,則以 Y, X_1, X_2, X_3……等代表之.

x': 一變量之值表示爲對於一假定原點之差離者.(即X數列之各項與其假定算術中數之差.) y'之意義相同.

Y_c: 一個變量之計算值或估量值,由平均關係方程式所決定者; y_c可用以代表一此種計算值之表示爲其對於算術中數之差離者.

p: 當兩變量皆表示爲對於各自算術中數之差離時之平均乘積,即 $p=\frac{\Sigma(xy)}{N}$. 當寫成 p_{12}等帶有小註之符號時,小註即指該平均乘積所關之兩變數,如 x_1 與 x_2.

p': 當兩變量表示爲對於假定算術中數之差離時之平均乘積,即 $p'=\frac{\Sigma(x'y')}{N}$.

r: 皮爾生式緊聯係數.當右下角註小字時,小字指此係數所關之變量.如 r_{yx}指 y 與 x 兩變量;而 r_{12} 指 x_1 與 x_2 兩變量.

ρ(Rho):普遍之緊聯指數(切勿與斯碧蒙氏(Spearman)之等級緊聯係數相混).右下角應註小字以指示與此計量相關之諸變量,如 ρ_{yx}, ρ_{xy}, $\rho_{\log x\cdot x}$, $\rho_{\log y\cdot\log x}$, $\rho_{\frac{1}{y}\cdot x}$, 等等.(第一個小字皆指應變量)

$\bar{d}$: 各觀察值對於一配合曲線之差離;即一變數之一觀察值與一相當之計算值之差.

v: 差數;意義與上述之$\bar{d}$相同.

S: 一配合曲線周圍之標準差;即估量標準誤.此符號右下角須加註小字以指示此計量所指之變數爲何,如S_y, S_x, $S_{\log y}$(以對數表示之估量標準誤),S_r(以比率表示之估量標準誤),$S_{\frac{1}{y}}$(以倒數表示之估量標準誤).

η(Eta): 繫聯比.符號右下角應加註小字以指示此計量所關之變數,如η_{yx}或η_{xy}.每處之第一個小字皆指應變數.

σ_{sy}:在通過繫聯表各行中數之一線周圍之標準差;y數列中各項對於各行算術中數之標準差.σ_{sx}對於繫聯表各列其意義與前者同;即x數列中各項對於各列算術中數之標準差是也.

σ_{my}:繫聯表上各行之算術中數對於y數列所有各項之算術中數之標準差,每一行之算術中數乃用各該行中所包括之項數加權者.σ_{mx}對於各列之算術中數其意義與前者同.

ζ(Zeta):響應之直線性之測驗($\zeta=\eta^2-r^2$)

κ: 計算一繫聯比時所用之行數或列數.

b: 響應係數;響應線之斜度,右下角加註小字時,小字指此係數所關之變數,如b_{yx}, b_{12}(指變數x_1, x_2).每處之第一個小字皆指應變數;b_{yx}爲y對x之響應係數;b_{xy}爲x對y之響應係數.

$R_{1.234}$: 一應變數x_1與自變數x_2x_3及x_4之集體間之複繫

聯係數,小字之次序可變,惟第一個小字常指應變數.

$r_{12\cdot34}$: 當 x_3 與 x_4 兩變數視爲固定不變時 x_1 與 x_2 兩變數間之偏繫聯或純繫聯係數.右下角所註小字,其次序因變數之結合不同而變化,小點前之兩小字經常指所計量之純繫聯存在於某某變數之間.

$b_{12\cdot34}$: 在 x_3 與 x_4 兩變數假定不變時 x_1 對 x_2 之偏響應係數,即同時根據 x_3 及 x_4 之數值作 x_1 之估量時所加於 x_2 之權數,所註之小字其次序因變數之結合不同而變更.

$\sigma_{1\cdot234}$: 四次標準差;與 $S_{1\cdot234}$ 相同.

(以上諸計量加註小字之個數與一研究中所包括之變數之個數相當,爲簡單起見僅假定有四個變數存在.)

(五)計量差誤時所用之符號

σ_M:算術中數之標準誤

σ_σ:標準差之標準誤,同樣,σ 之符號加註任何小字,卽爲該小字所指之計量之標準誤.

$P.E._M$:算術中數之機誤($P.E._M=.67449\sigma_M$).$P.E.$ 之符號加註任何小字卽爲該小字所指之計量之機誤.

(六)其他符號.

p: 一事件中成功之機率.

q: 一事件中失敗之機率.

s: 一觀察方程式中之已知量之和;用於覆驗正則方程

式之形成及解法者.

Δ(Delta):一般用以代表兩數量間之差或一變量之依次相繼諸值間之差,其所指之變量可以加註表明之,如Δ_y,第一級差寫作Δ^1,第二級差寫作Δ^2,諸如此類.

ν_1, ν_2, ν_3,等:一頻數分配對於一假定原點之各轉矩.

π_1, π_2, π_3等:一頻數分配對於算術中數之未校正轉矩.

μ_1, μ_2, μ_3等:經過用希巴特氏較正法較正後一頻數分配對於算術中數之各轉矩.(若不必用希巴特氏校正法時,μ可用作對於算術中數之未校正轉矩之符號.)

β_1: $\dfrac{\mu_3^2}{\mu_2^3}$

β_2: $\dfrac{\mu_4}{\mu_2^2}$

κ_2: 根據β_1與β_2求出之測驗曲線形式之判準.

χ^2: 用於測驗曲線配合之適合程度之一個數量.

參考書總目

Adams, Oscar S. Application of the Theory of Least Squares to the Adjustment of Triangulation. Special Publication No. 28. U. S. Coast and Geodetic Survey 1915.

Barlow. Tables of Squares, Cubes, Square Roots, Cube Roots, Reciprocals. Spon and Chamberlain. New York. 1919.

Beckett S. H. and Robertson, R. D. the Economical Irrigation of Alfalfa in the Sacraments Valley. Bulletin 280, Agricultural Experiment Station. University of California, May, 1917.

Bowley, A. L. Elements of Statistics. P. S. King and son, London. 1920.

Brinton, W. C. Graphic Methods for Presenting Facts. The Engineering Magazine Co., New York. 1914.

Broad, C. D. On the Relation between Induction and Probability. Mind N. S. Vol. 27, 1918, Vol. 29, 1920.

Brunt, David. The Combination of Observations. Cambridge University Press. 1917.

Chaddock, R. E. Principles and Methods of Statistics. Houghton Mifflin, Boston. (In preparation.)

Clark Wallace. The Gantt Chart. Ronald Press. New York. 1922.

Crum, W. S. The Use of the Median in Determining Seasonal Variation. Journal of the American Statistical Association March, 1923.

Davenport. E. Comparative Agriculture. In Boileys Cyclopedia of American Agriculture.

Davies. G. R. Introduction to Economic Statistics. Century. New York. 1922.

Day, Edmund E. An Index of the Physical Volume of Production. Review of Economic Statistics. Sept., 1920 Jan., 1921. Standardization of the Coustruction of Statistical Tables. Quarterly Publications of the American Statistical Association March, 1920. The volume of Production of Basic Raw Materials in the United States, Review of Economics. Statistics July. 1922.

Edgeworth, F. Y. On Correlated Averages. Phil. May., 5th series. Vol. 34, 1892.

Editorial. On the Probable Errors of Frequency Constants. Biometrika, Vol. 2 (273—281).

Elderton, W. Palin. Frequency Curves and Correlation. Layton London, 1906.

Ezekiel, M. J. B. A Method of Handling Curvilinear Correlation for Any Number of Variables, Journal of the American Statistical Association, Dec., 1924.

Faulkuer, Helen D. The Measurement of Seasonal Variation. Journal of the American Statistical Association, June, 1924.

Field, J. H. Some Advantages of the Logarithmic Scale in Statistical Diagrams. Journal of Political Economy Oct., 1917.

Fisher, Arne. An Elementary. Treatise on Frequency Curves Macmillan, New York. 1922. The Mathematical Theory of Probabilities Macmillan, New York. 1922.

Fisher, Irving A Weekly Index Number of Wholesale Prices Journal of the American Statistical Association. Sept., 1923. The Making of Index Numbers. Honghton Mifflin, Boston 1922. The "Ratis" Chart Quarterly Publicatoins of the American Statistical Association. June 1917. Revision of the Weekly Index Number. Journal of the American Statistical Association. Sept. 1924.

Flux, A. W. The Measurement of Price Changes. Journal of the Royal Statistical Society. March, 1921.

Galton, Francis Correlations and Their Measurement. Proceedings of the Royal Society. Vol. 45, 1888.

Griffin, F. L. Introduction To Mathematical Analysis. Honghton Mifflin, Boston. 1922.

Haas, G. C. Sale Prices as a Basis for Farm Land Appraisal.

Technical Bulletin, No. 9. University of Minnesota Agricultural Experiment Station Nov. 1922.

Hall. Lincoln W. Seasonal Variation as a Relative of Secular Trend. Journal of the American Statistical Association. June, 1924.

Hart, W. L. The Method of Monthly Means for Determination of a Seasonal Variation. Journal of the American Statistical Association. Sept. 1922.

Haskell, A. C. Graphic Charts in Business. Codex Book Co., New York, 1922. How to Make and Use Graphic Charts Codex Book Co., New York 1919.

Jones, D. C. A First Course in Statistics. Bell, London, 1921.

Karsten, Karl. Charts and Graphs Prentice Hall, New York. 1923.

Kelley, Truman L. Statistical Method. Macmillan New York. 1923.

Keynes, J. M. A Treatise on Probability. Macmillan, New York. 1921.

Killough, H. B. A Statistical Analysis of Oat Prices. Bureau of Agricultural Economics.

King, W. I. Elements of Statistical Method. Macmillan, New York. 1912. An Improved Method for Measuring the Seasonal Factor. Journal of the American Statistical Association, Sept., 1924.

Knibbs, G. H. The Theory and Justification of Curve Smoothing. In H. Secrist, Readings and Problems in Statistical Methods. Macmillan, New York. 1920.

Kurtz, Edwin. Replacement Insurance. Administration, July, 1921.

Lipka, Joseph. Graphical and Mechanical Computation. Wiley, New York. 1918.

Mellor, J. W. Higher Mathematics for Students of Chemistry and Physics. Longmans, London. 1922.

Merriman Mausfield. The Method of Least Squares. Wiley New York. 1897.

Mills, Frederick C. On Measurement in Economics (in Tugwell R. G. ed., The Trend of Economics. Knopf, New York 1924).

Miner, J. R. Tables of $\sqrt{1-r^2}$ and $1-r^2$ for Use in Partial Correlation and Trigonometry. Johns, Hopkins Press Baltimore. 1922.

Mitchell, W. C. History of Prices During the War. Price Bulletin, No. 1, War Industries Board 1919. The Making and Using of Index Numbers. Part I. Bulletin No. 284. U. S. Bureau of Labor Statistics. Oct., 1921.

Moore, H. L. Economic Cycles: Their Law and Cause. Macmillan, New York. 1914. Elasticity of Demand and Flexibility of Prices. Journal of the American Statistical Associatino March, 1922. Empirieal Laws of Demand and Supply and the Flexibility of Prices. Political Science Quarterly. Dec. 1919. Forecasting the Yield and the Price of Cotton. Macmillan. New York. 1917. Generating Economic Cycles. Macmillan. New York. 1923.

National Bureau of Economic Research. Income in the United States. (Edited by W. C. Mitchell.) Harcourt Brace and Co., New York. 1921.

North Dakota Agricultural College. Cost of Production and Farm organization on 126 Farms in North Dakota. Bulletin No. 165, Agricultural Experiment. Station. 1922.

Ogburn, W. F., and Thomas. Dorothy. Influence of the Business Cycle on Certain Social Conditions. Quarterly Publications of the American Statistical Association. Sept., 1922.

Peake, E. G. An Academic Study of Some Money Market and other Statistics. P. S. King, London. 1923.

Pearl, R., and Reed. L. J. Predicted Growth of Population of New York and Its Environs. Committee of Plan of New York. 1923.

Pearl, Raymond. Medical Biometry and Statistics, Saunders, Philadelphia. 1923.

Pearson, Karl. Mathematical Contrebutions to the Theory of Evolution. On the General Theory of Skew Correlation and Non-linear Regression. Draper's Company Research Memoirs. Cambridge University Press. 1905. Notes on the History of Correction, Biometrika, Vol. 13. On a Correction Needful in the case of the Correlation Ratio. Biometrika, Vol. 8. On the Correction Necessary for the Correlation Ratis. Biometrika, Vol. 14. Regression, Heredity and Panmixia. Phil. transactions, Royal Society, Series A, Vol. 187. 1896. Tables for Statisticians. and Biometricians. Cambridge University Press. 1914. The Fundamental Problem of Practical Statistics, Biometrika, Vol. 13.

Persons, Warren M., and Coyle, Eunice. A Commodity Price Index of Business cycles. Review of Economic Statistics. Prel. Vol. 3.

Persons, Warren M. An Index of Trade for the United States. Review of Economic Statistics. April, 1923. Correlation of Time, Series. Journal of the American Statisrical Association. June. 1923. Fisher's Formula for Index Numbers. Review of Economic Statistics. Prel. Vol. 3. Indices of Business Conditions. Review of Economic Statistics. Prel. Vol. 1. 1919. The Variate Difference Correlation Method and Curve Fitting. Quarterly

Publications of the Ameircan Statistical Association. June, 1917.

Prescott, Raymond. Law of Growth in Forecasting Demand Journal of the American Statistical Association. Dec., 1922.

Rietz, H. L. ed. Handbook of Mathematical Statistics. Honghton Mifflin, Boston.1924.

Rietz, H. L., and Crathorne, A. R. College Algebra. Holt, New York. 1917.

Rugg, H. O. Statistical Methods Applied to Education. Honghton, Mifflin, Boston. 1917.

Running, T. R. Empirical Formulas. Wiley, New York. 1917.

Schultze, Arthur. Graphic Algebra. Macmillan, New York. 1918.

Schultze, Henry. The Statistical Measurement of the Elasticity of Demand for Beef. Journal of Farm Economics. July, 1924.

Secrist, Horace. Introduction to Statistical Methods. Macmillan New York. 1917. Readings and Problems in Statistical Methods. Macmillin, New York. 1920.

Sheppard, W. F. On the Calculation of the most Probable Values of Frequency Constants for Data Arranged According to Equi-distant Divisions of a Scale. Proceedings of the London. Mathematical Society, Vol. 29, 1898. The calculation of the Moments of a Frequency Distribution, Biometrika, Vol. 5.

Smith, Bradford B. The use of Punched card Tabulating Equipment in Multiple correlation Problems. (Prepared *for the* use of statisticians of *the Bureau* of Agricultural Economics, U. S. Dept. *of* Agriculture.) 1923.

Snyder, Carl A New Index of the Volume of Trade. Journal of the American Statistical Association. Dec., 1923.

Snow, E. C. Trade Forecasting and Prices. Journal of the Royal Statistical Society. May, 1923.

Stamp, J. C. The Effect of Trade Fluctuations upon Profits. Journal of the Royal Statistical Society. July, 1918.

Steinmeltz, C. P. Engneering Mathematics. Mc Graw Hill, New York. 1917.

Stewart, Ethelbert. Labor Efficiency and Productiveness in Saw-mills. Monthly Labor Review. Jan., 1923.

Tolley, H. R. and Ezekiel. M. J. B. A Method of Handling Multiple Correlation Problems. Journal of the American Statistical Association. Dec., 1923.

Walsh, C. M. The Measurement of General Exchange Value. Macmillan, New York. 1901. The Problem of Estimation P. S. King and Son, London. 1921.

Weld, L. D. Theory of Errors and Least Squares. Macmillan, New York. 1916.

West. Carl J. Introduction to Mathematical Statistics. Adams, Columbus. 1918.

Whipple, G. C. Vital Statistics. Wiley, New York. 1919.

Whitaker, E. T., and Robinson, G. The Calculus of Observations. Blackie and Son, London, 1924.

Whitehead, A. N. An Introduction to Mathematics. Holt, New York, 1911.

Working, Holbrook. Factors Determining the Price of Potatoes in St. Paul and Minneapolis. Technical Bulletin, No. 10, University of Minnesota Experiment Station. Oct., 1922.

Wright, T. W. and Hayford, J. F. Adjustment of Observations Van Nostrand, New York. 1906.

Young, Allyn A. The Measurement of Changes in the General Price Level. Quarterly Journal of Economics. Aug., 1921.

Yule, G. Udney. An Introduction to the Theory of Statistics. Griffin London. 1919. On the Time Correlation Problem, with Especial Reference to the Variate Difference Correlation Method. Journal of the Royal Statistical Society. July, 1921.

Zizek, Franz. Statistical Averages. Holt, New York. 1913,

12

中華民國三十年四月初版

(合)密勒氏統計方法論二册 (83477)

上海 實價新法幣貳拾貳元

原著者 F. C. Mills

譯述者 徐坚

發行人 王雲五 長沙南正路

印刷所 商務印書館

發行所 商務印書館 各埠

(本書校對者張鷹天)

旦九九一上