漢譯世界名著

統計學原理

（下）

鮑萊 著
李植泉 譯

商務印書館

Arthur L. Bowley 著

李植泉 譯

漢譯世界名著

統計學原理

下

商務印書館發行

第二編

目錄

附錄　數學摘錄

第二編

插表目錄

第二編

插圖目錄

第二編　數理統計之部

第一章　頻數曲線

第一節　導言

數學方法，在統計領域中，極多部分，均甚重要。本書第一編，曾用代數法，將算術結果歸納於通則，較爲簡單之插補法，亦多用之。但非借重高深數理不能解決之問題，厥類甚繁，本編之作，卽欲對此方面，略加討論者也。然統計學中，應用數理之範圍，本至廣擴，非本編篇幅，所能盡行包羅，茲僅擇其方法之重要者，及與經濟學或其相關科學有直接關係之問題，討論之。實則醫學，生物學及其他科學中之統計問題，原亦須用同一方法，此事於相當之期刊見之。茲爲討論方便起見，本編通例，以認定由經濟社會調查而生之問題爲討論對象，應用例證，亦儘量於此限定範圍內取材。

數學方法之應用，有種種不同，然大要分之，不外三類：

(1)各羣類之有系統的敍述(The systematic description of groups)；

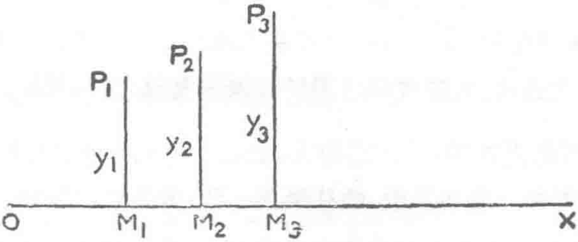
(2)現象間相互關係之測算(The measurement of relationship between phenomena);

(3)抽樣方法所得結果精度之測算(The measurement of the precision of results obtained by a process of sampling)。

各種分析之基礎,乃為機率原理(Theory of chance)(註一)其高深僅若干數學專家造詣及之,然吾人殊不能假定讀者之所通曉已超過代數機率之較淺理論;而欲介紹參考書籍,英文中又乏熟識教本;不得已,只得佔用許多篇幅,專論純粹數學原理;惟對於曾受數學訓練而非精於此道者,仍力求顧到,以期其能以領會。故於可能範圍內,能不用微積分而能到到證明者,必不用之;各項問題之結論,儘量用文字明白闡述,用算術例題證明;最簡單之例題,首先加以討論,以說明各種程序及結果,而較為普通之研究,則舉其概略,而以曾有透澈研究之書文,列為參考。至不諳數學之讀者,如將書中小體字,略去不讀,亦未為不可。書末附有附錄,凡他處不易得到證明之定理,擇要搜羅之,而分析過於繁難之部分,在教本中不便多所發揮者,亦歸併附錄之內。

第二節　頻數羣類及曲線(Frequency groups and curves)

以下專就頻數羣類之有系統的測量討論



設有任何羣類之數量於此，Ox軸卽代表此數量，在此軸上，分列尺度，具有數量 x_1者有 y_1 個，x_2 有 y_2個，其他以此類推；如此，圖上$OM_1=x_1$，$M_1P_1=y_1$其他以此類推，則該羣類，乃如上圖所示。

M_1M_2，M_2M_3，……各級(grade)不必相等。

假如該類之總數爲n，使

$$n=y_1+y_2+\cdots$$

如此，在x_1，x_2…等處觀察之『頻數』(frequency)爲

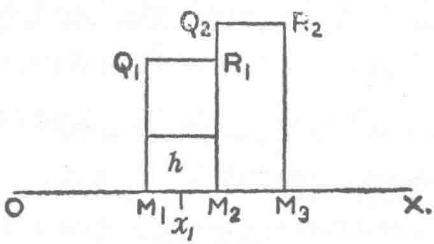
$$\frac{y_1}{n},\frac{y_2}{n}\cdots$$

設P_1，P_2，P_3……諸點，可視爲在一連續曲線(continuous curve)上，則其軌跡(locus)卽成頻數曲線(Frequency curve)。

又設量數並不分級，亦不按特定數值而分爲小羣，而每一次觀察，卽得一個量數，則此一羣類，卽成爲一滿佈星點之横軸，一點卽是一項目，

以後所論之公式，大部可以適用於此滿佈星點之綫，及頻數曲綫亦同。

一個羣類之各個量數，往往聚集一起，分別形成等級（譬如20–25，25–30……歲）或者在着手時，即已採用最狹之單位（譬如 55–56， 56–57……吋）。在此情形之下，各級個數，約略與長方形相等（譬如在 M_2M_3 一級之 $M_2M_3R_2Q_2$）。



以h代表各級之寬，x_1x_2……爲各級中點之橫坐標，y_1, y_2……爲長方形之高，而y_1h, y_2h……爲各級之個數。則

$$n=y_1h+y_2h\cdots\cdots,$$

如以h作單位，則

$$n=y_1+y_2+\cdots\cdots。$$

各級之頻數爲

$$\frac{y_1h}{n},\frac{y_2h}{n}\cdots\cdots$$

如果一個連續曲線之條件可以決定，然後構成圖式，而使建在 M_1M_2, M_2M_3 ……… 上面之諸部分曲綫面積，與 y_1h, y_2h ……… 成比例，則此曲線，卽爲該類之頻數曲綫。

變動參差爲自然界之通律，人事變幻，多所不免；頻數羣類，於是乃由大規模觀察而生。羣類可分爲四種：

（a）一羣類中各個分子，皆予測查，如某業成年男工工資是；

（b）從羣類中選樣觀察，如由一五萬戶之小城市，抽樣一千戶，而查其兒童人數若干，或測量某一種樹之樹葉，均屬此類；

（c）物體數量之重複測量（如重複測量某星位之赤緯），此種數量之變量，乃由於工具差誤(Instrumental "error")；

（d）各種數目實現之數學機率(mathematical probability)，（如擲錢五十個，擲得一，二，三……個表面向上之機率是）；或因未知之複雜原因而決定之事件次數。

無論現象屬於何類，普通所用方法，完全相同。在此法之下，須選出x值及y值之某種代數函數，並須以數值代表x各值及y各值之代數函數，以表示該羣類。其實，羣類之敍述乃係：

（1）決定其中心位置，

（2）測量觀察對此中心之離散度，

（3）測量其去中心之偏斜度，

（4）根據代表一羣類之圖形，作其他之測量。

中心位置之決定，可用算術平均數，中位數，或衆數，有時用幾何平均數亦可。算術平均數，其他計算上需用最廣，故應以之爲一切之慣常出發點。中位數無助於一般代數上之演算；精確數值又不可常得，如非別有用途，無計算之必要。至於衆數，普通甚難從觀察中確切決定，欲介紹其近似值，當此初步計算時又非所宜；然如有一代表羣類之確定代數公式，衆數可確實得出，則亦未嘗不可重視。

測量離散度，可以採用『概誤』(probable error)，卽四分位差(Half-intequartile range)，或用平均差(mean deviation)或均方差 (deviation of mean square) 在此數者之中，概誤一者與中位數同，只能求得其近似值，對於進一步之測量，甚難爲有系統之運用。平均差，其據以測算之原點位置，含混不淸，姑置不論，且因其根本量數，無視符號之正負，將來應用，必感極大困難。惟均方差，此種困難，可以完全避免。所謂均方差，乃各量數對平均數離差平方之平均數方根。故不僅易用代數推演，且爲多種演算所必需。此之謂標準差，數理統計上普遍應用者卽此是也（見第一編第六章）。

曲線不對稱，中位數，衆數及算術平均數，數者必不能相合於一點，而上下四分位數對中位數之距離，亦必不能相等。若羣類爲對稱時，此等數量，必等於零；故此種數量，可爲測量之根據；

但中位數，衆數及四分位數，只能得其近似值，作成之測量，必受觀察不充分之影響而難期完善，且變更觀察量，亦難得若何效果，除非將其轉移而逾過中位數或四分位數也。

測量必須與每一數值之位置感應銳敏。用在平均數下數值與平均數上數值兩個平均差之餘數，固無不可，惟此餘額，不易列成公式，以與其他有系統的測量，相互爲用。故爲免除困難起見，特採用平均立方差（各數值與平均數離中差三次方之平均數，符號或正或負，仍依其原來狀態），且立方差對偏態（skewness）之感應，甚爲靈敏。在度量差量時，自然而慣常之方法，須用名數表示之，如吋若干，磅幾何之類，標準差，均方差，及機誤之表示亦然，惟偏斜度之測量，苦無明確名數單位，故不得不使測量，脫離所用單位；而此方法，卽以離中差作爲標準差之倍數；例如，以x代表某種測量，$\bar{x}$代表平均數，σ爲標準差，則平均後之離中差，爲$\left(\frac{x-\bar{x}}{\sigma}\right)^3$，而偏斜度，乃可由$\frac{1}{n}\left\{所有\left(\frac{x-\bar{x}}{\sigma}\right)^3各值之和\right\}$而得。由此測量之結果，感應最爲敏銳，只惜缺乏明鮮單位耳。至其意義，是否可以明瞭，須視對曲線形狀，及實際測得之偏斜度，有無經驗而定也。

此外尙有用平均數之四次，五次乃至更高次方，以資測量者。皮爾生（Karl Pearson）教授，於其動差（moment）論，曾推論之。第一，第二，第三乃至更高級之動差，卽爲離中差之一次，

二次，三次乃至更高次方之平均。離中差可隨任何一點起算，而測得之動差，即爲對該點而言，惟一般應用，均以平均數爲測量之中心，而由其他各點得來之動差，不過圖演算之利便計耳。

第一編第五章，曾論平均數，謂其爲敍述一羣類之捷徑，在比較兩類材料時，此言尤信。然現時此概念應行擴大，現須另尋敍述主要表徵數之有系統的方法，而此方法，即用三數種符號，以測量平均數，標準差，偏斜度及其他類似數量也。若此等測量之意義及尺度，一經辨識淸楚，原始材料，便無效用（除非留作參考或製圖）；羣類之表示，完全用扼要方法；計算各羣類間之相互關係，即以此種數量爲根據，並恃作數理研究之基礎。

第三節　動差所用之標號

玆將動差所用之標號及術語，述之如下：——

$$m'_t=\frac{1}{n}(x_1{}^t y_1+x_2{}^t y_2+\cdots\cdots)=\mathrm{S}(x^t y)\div n\cdots\cdots\cdots\cdots(1)$$

此之謂一羣類對其原點而言之第 t 級動差。

$$n=\mathrm{S}y\cdots\cdots\cdots\cdots\cdots\cdots(2)$$

$$m_1'=\bar{x}=\mathrm{S}xy\div \mathrm{S}y\cdots\cdots\cdots\cdots\cdots\cdots(3)$$

乃爲該羣類之平均數。

$$m_t=\mathrm{S}(x-\bar{x})^t y\div n\cdots\cdots\cdots\cdots\cdots\cdots(4)$$

爲對平均數而言之第 t 級動差。