

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
李植泉 著
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

第二編	數理統計之部	361
第一章	頻數曲線	361
第一節	導言	361
第二節	頻數羣類及曲線	362
第三節	動差所用之標號	368
第四節	動差算法舉例	370
第二章	代數機率與差誤常態曲線	379
第一節	初步原理	379
第二節	機率乘法	380
第三節	機率加法	381
第四節	差誤常態律之演繹	382
第五節	代數機率與經驗	395
第六節	白諾立氏定律	396
第七節	例證	398

第八節	對於抽樣法之應用.....	402
第九節	抽樣方法舉例.....	405
第十節	範圍實非爲無限或選擇未能獨立之例.....	408
第十一節	小數律.....	411
第三章	大數律(普遍的差誤律).....	417
第一節	平均及總和之標準差及均立方差誤.....	417
第二節	差誤曲線之發生.....	420
第三節	用多項式定理證明之.....	422
第四節	愛基華斯氏之證明.....	427
第五節	普遍的差誤律或大數律之說明.....	431
第六節	範圍有限制之例.....	433
第七節	例證.....	436
第四章	差誤律之應用.....	449
第一節	平均數及總和數之精度.....	449
第二節	平均數之精度.....	451
第三節	平均數之常態分配.....	452
第四節	加權總和及加權平均之絕對差誤.....	454
第五節	相對差誤.....	457
第六節	例證(一).....	463
第七節	平均數之比較.....	469

第八節 例證 (二)	471
第九節 平均數與平均數間差額之重要	473
第十節 趨勢之存在	485
第十一節 週期性	488
第五章 經驗頻數方程	493
第一節 皮爾生氏曲線系	494
第二節 愛基華斯氏法	496
第三節 巴里多氏方程	497
第四節 梅克漢氏公式	500
第六章 相關論	503
第一節 導言	503
第二節 相關係數	507
第三節 之特性	510
第四節 相關面	512
第五節 愛基華斯氏法	514
第六節 常態相關面之性質	518
第七節 直線相關	521
第八節 相關率	524
第九節 未分級變量之相關	526
第十節 相聯	529

第十一節	相依	531
第十二節	時間數列之相關	535
第十三節	時間數列之圖式比較法	540
第七章	相關例證	543
	例一 例二 例三 例四 例五 例六 例七	
第八章	淨相關與複相關	565
第一節	淨相關	565
第二節	複相關	572
第九章	平均數動差及相關等測量之精度	581
第一節	逆機率	581
第二節	某類在範圍中所佔比例, P , 之精度	584
第三節	通用方法	588
第四節	算術平均數之精度	589
第五節	標準差之精度	590
第六節	平均數之標準差 (罔計及逆機率)	592
第七節	相關係數之標準差	598
第十章	資料與公式適應之測驗	605
第一節	測驗方法	605
第二節	例證	613

附錄 數學摘錄

一 求 π 值之瓦立斯氏定理.....	617
二 整數之乘冪總和.....	617
三 求 $m!$ 之斯德令公式.....	618
四 猶勒麥克老令定理——用求和表示積分.....	620
五 薛伯氏對於頻數曲數動差之修正.....	624
六 對於普遍的差誤曲線第二近似值之動差及常數....	627
七 未加權平均數之比率.....	633
八 加權平均數之比率.....	636
九 動差.....等中差誤之標準差之常態性.....	639
十 最小二乘法.....	643

第二編

插表目錄

第一表 兒童體重.....371

第二表 孫巴克氏四十五種商品之指數.....373

第三表 $F(z) = \frac{1}{\sqrt{2\pi^2}} \int_0^z e^{-\frac{1}{2}z^2} dz$ 之值 (常態機率表).....394

第四表 $F(z) = \frac{1}{6\sqrt{2\pi}} \left\{ 1 - (1 - z^2)e^{-\frac{1}{2}z^2} \right\}$ 之值

第二編

插圖目錄

第一圖	各整列之平均數及迴歸線.....	554
第二圖	差誤偏態曲線.....	630
第三圖	差誤常態曲線.....	646

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，較為簡單之插補法，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，即欲對此方面，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。茲為討論方便起見，本編通例，以認定由經濟社會調查而生之問題為討論對象，應用例證，亦盡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) 各羣類之有系統的敘述(The systematic description of groups)；

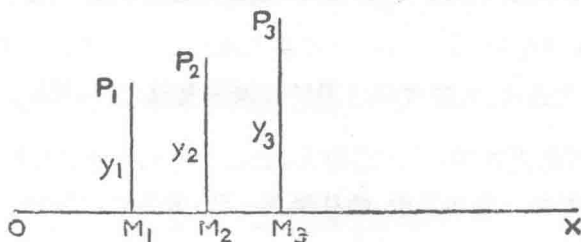
(2) 現象間相互關係之測算 (The measurement of relationship between phenomena);

(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軸即代表此數量，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。

假如該類之總數為 n ，使

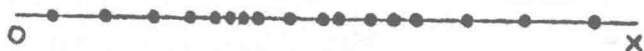
$$n = y_1 + y_2 + \dots$$

如此，在 $x_1, x_2, \dots$ 等處觀察之『頻數』 (frequency) 為

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

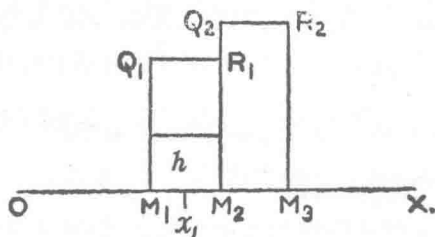
設 $P_1, P_2, P_3, \dots$ 諸點，可視為在一連續曲線 (continuous curve) 上，則其軌跡 (locus) 即成頻數曲線 (Frequency curve)。

又設量數並不分級，亦不按特定數值而分為小羣，而每一次觀察，即得一個量數，則此一羣類，即成為一滿佈星點之橫軸，一點即是一項目，



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， $x_1, x_2, \dots$ 為各級中點之橫坐標， $y_1, y_2, \dots$ 為長方形之高，而 $y_1h, y_2h, \dots$ 為各級之個數。則

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

如以 h 作單位，則

$$n = y_1 + y_2 + \dots。$$

各級之頻數為

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

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 $M_1M_2, M_2M_3, \dots$ 上面之諸部分曲線面積，與 $y_1h, y_2h, \dots$ 成比例，則此曲線，即為該類之類數曲線。

變動參差為自然界之通律，人事變幻，多所不免；類數羣類，於是乃由大規模觀察而生。羣類可分為四種：

- (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，羣類之敘述乃係：

- (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『機誤』(probable error)，即四分位差(Half-intequartile range)，或用平均差(mean deviation)或均方差(deviation of mean square)在此數者之中，機誤一者與中位數同，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，含混不清，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即此是也（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，即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 $\frac{1}{n}\left\{\text{所有}\left(\frac{x-\bar{x}}{\sigma}\right)^3\text{各值之和}\right\}$ 而得。由此測量之結果，感應最爲敏銳，只可惜缺乏明鮮單位耳。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，即為對該點而言，惟一般應用，均以平均數為測量之中心，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，謂其為敘述一羣類之捷徑，在比較兩類材料時，此言尤信。然現時此概念應行擴大，現須另尋敘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，一經辨識清楚，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，即以此種數量為根據，並特作數理研究之基礎。

第三節 動差所用之標號

茲將動差所用之標號及術語，述之如下：——

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

此之謂一羣類對其原點而言之第 t 級動差。

$$n = Sy \dots \dots \dots (2)$$

$$m_1' = \bar{x} = Sxy \div Sy \dots \dots \dots (3)$$

乃為該羣類之平均數。

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

為對平均數而言之第 t 級動差。