

漢譯世界名著

統計學原理

(下)

鮑萊著
李植泉譯

商務印書館發行

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

中華民國二十七年七月初版

(32072.5)

漢譯世界名著統計學原理二冊

Elements of Statistics

每冊實價國幣貳元

酌加運費匯費

原著者

Arthur

譯述者

林泉

發行人

王雲五

印刷所

長沙南正路商務印書館

發行所

各商務印書館

版權必究

★E三九五二

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	r 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	<u>逆機率</u>	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);

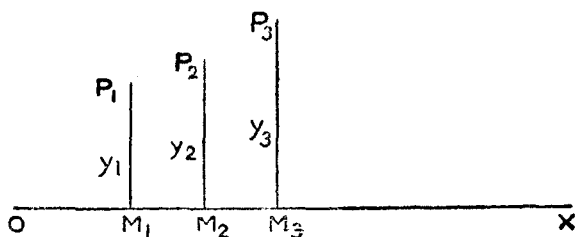
(2) 現象間相互關係之測算 (The measurement of relationship between phenomena);

(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

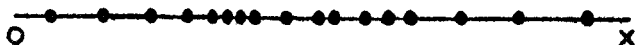
$$n = y_1 + y_2 + \dots$$

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

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

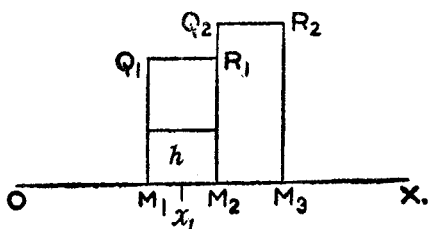
設 P_1, P_2, P_3, \dots 諸點，可視為在一連續曲線 (continuous curve) 上，則其軌跡 (locus) 即成頻數曲線 (Frequency curve)。

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，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，即爲離中差之一次，

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。

$$n = S y \dots \dots \dots (2)$$

$$m_1' = \bar{x} = S x y \div S y \dots \dots \dots (3)$$

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。

於是

$$nm_2 = S(x - \bar{x})^2 y = Sx^2 y - 2\bar{x}Sxy + \bar{x}^2 Sy = nm_2' - 2\bar{x} \cdot n\bar{y} + n\bar{x}^2$$

$$\therefore m_2 = m_2' - \bar{x}^2 \dots\dots\dots (5)$$

$$\sigma = \sqrt{m_2} \dots\dots\dots (6)$$

此之謂標準差，定義如上所述。

$$nm_3 = Sx^3 y - 3\bar{x}Sx^2 y + 3\bar{x}^2 Sxy - n\bar{x}^3$$

$$m_3 = m_3' - 3\bar{x}m_2' + 2\bar{x}^3 \dots\dots\dots (7)$$

m_3 ，在對稱曲線，其值為零。求算偏斜度，簡便之測量法，乃以橫坐標定為標準差之倍數，如此則測量所用具體單位之煩，可以免去也。

例如

$$\kappa = S \left\{ \left(\frac{x - \bar{x}}{\sigma} \right)^3 y \right\} \div n = \frac{m_3}{\sigma^3} \text{ (註二)} \dots\dots\dots (8)$$

即偏斜度之測量法。

同理

$$m_4 = m_4' - 4\bar{x}m_3' + 6\bar{x}^2 m_2' - 3\bar{x}^4 \dots\dots\dots (9)$$

為第四級動差，而

$$\kappa_2 = \frac{m_4}{\sigma^4} = \frac{m_4}{m_2^2} \dots\dots\dots (10)$$

為不用單位之測量法。

標準差既然求得， κ_2 之大小要看該類各份子自中心向外離散之程度而定，離散愈大， κ_2 值亦愈大。在差誤常態曲線之特種

情形下（見第二章第四節）， $k_2=3$ 。設 σ 不變，中心高度降低，而以外各部向外展開，則 $k_2>3$ 。

皮爾生教授用 $\beta_1 = \frac{m_3^2}{m_2^3}$ ，然則，如上所述， $\sqrt{\beta_1} = k$ 。皮氏及猶爾（Yule）氏，求偏斜度之公式，較為精審複雜。且彼用 u_i ，不用 m_i ，又以 β_2 代替 k_2 。至愛基華斯（Edgeworth）教授，拘於舊習，時用 $c = \sqrt{2m_2}$ （是謂模差（modulus）），作為差量（reduction）之單位，以代替 σ ，然則 $c = \sigma\sqrt{2}$ 。總而言之，用 c 原為避免公式上之繁複，然多用一 c 字母，似屬得不償失，蓋無論如何，標準差亦在所必需也。

愛基華斯氏復用 j 代 $\frac{m_3}{c^3}$ ，然則 $k = 2\sqrt{2j}$ ，又以 i 代表

$\frac{m_4}{c^4} - \frac{3}{4} = \frac{1}{4} \left(\frac{m_4}{\sigma^4} - 3 \right) = \frac{1}{4} (k_2 - 3)$ 。於是在差誤常態曲線中，
之值必為零。

第四節 動差算法舉例

\bar{x} , σ , k , k^2 四種基本測量之算法，舉例如下。

統計之敘述工作中，必須用四級以上之動差者，事極罕見，蓋自實際觀察求得較高級動差時，隨以俱來之差誤，致使更高級之動差失其功用也。

一、第一例——體格度量之同種羣類——年齡約略相同之兒童三千四百零四人之體重。此等兒童之身長（見第七章第五

例) 散佈若屬對稱, 其體重理宜表示正偏斜度, 實際上亦確為 $k = 0.643$ 。惟內中兒童一人, 體格迥異常人 (身長五英尺四吋, 體重十四石) (註三), 在計算動差時, 曾將其剔除。因 $k = 3$ 只餘 457 , 可知該曲線離常態不遠。

紐約市僱傭特許證上十四至十五歲兒童之體重

重量 (磅)	尺度 x	人數 y	乘積			
			xy	x^2y	x^3y	x^4y
65-	- 7	3	- 21	147	- 1,029	7,203
70-	- 6	9	- 54	324	- 1,944	11,664
75-	- 5	142	- 710	3,550	- 17,750	88,750
80-	- 4	301	- 1,204	4,816	- 19,264	77,056
85-	- 3	289	- 867	2,601	- 7,803	23,409
90-	- 2	380	- 760	1,520	- 3,040	6,080
95-	- 1	416	- 416	416	- 416	416
100-	0	404	—	0	—	0
105-	1	315	+ 315	315	+ 315	315
110-	2	320	+ 640	1,280	+ 2,560	5,120
115-	3	262	+ 786	2,358	+ 7,074	21,222
120-	4	221	+ 884	3,536	+ 14,144	56,576
125-	5	131	+ 655	3,275	+ 16,375	81,875
130-	6	76	+ 456	2,736	+ 16,416	98,496
135-	7	52	+ 364	2,548	+ 17,836	124,852
140-	8	20	+ 160	1,280	+ 10,240	81,920
145-	9	29	+ 261	2,349	+ 21,141	190,269
150-	10	14	+ 140	1,400	+ 14,000	140,000
155-	11	10	+ 110	1,210	+ 13,310	146,410
160-	12	2	+ 24	288	+ 3,456	41,472
165-	13	2	+ 26	338	+ 4,394	57,122
170-	14	5	+ 70	980	+ 13,720	192,080
175-	15	1	+ 15	225	+ 3,375	50,625
3,404			+ 4,906	37,492	+ 158,356	1,502,932
			- 4,032		- 51,246	
			+ 874		+ 107,110	

(第一表)

以102.5爲原點，單位爲五磅。

$$m_1' = \bar{x} = \frac{874}{3404} = .2568 \quad m_1 = 0. \text{平均數} 102.5 + .2568 \times 5 = 103.784 \text{磅}$$

$$m_2' = \frac{37492}{3404} = 11.014 \quad m_2 = m_2' + \bar{x}^2 = 10.948$$

$$m_3' = \frac{107110}{3404} = 31.466 \quad m_3 = m_3' - 3\bar{x}m_2' + 2\bar{x}^3 = 23.01$$

$$m_4' = \frac{1502932}{3404} = 441.519 \quad m_4 = m_4' - 4\bar{x}m_3' + 6\bar{x}^2m_2' - 3\bar{x}^4 = 413.542$$

$$m_2 = \text{修正後} = 10.948 - \frac{1}{12} = 10.865. \sigma = \sqrt{m_2} = 3.296, \text{即} 16.48 \text{磅}$$

$$m_4 \text{修正後 (註四)} = m_4 - \frac{1}{2}m_2 + \frac{7}{240} = 413.542 - 5.474 + .029 = 403.10$$

$$k = \frac{m_3}{\sigma^3} = .643 = \sqrt{\beta_1}, \quad k_2 = \frac{m_4}{m_2^2} = 3.457 = \beta_2$$

$$c = 4.661 \quad j = \frac{m_3}{\sigma^3} = .227 \quad i = .114。$$

二、孫巴克 (Sauerbeck) 之四十五種指數，爲測量各種商品價格之變動者，其平均數，則表示一般價格漲落之情況。此四十五種指數，對於一般運動之測量，乃不免於各個機差 (Chance deviation)，故可組成頻數羣類，而此羣類之標準差，又可用以測量平均數之確度。該類之偏態，尙不甚顯。因個體取材過狹，故無計算第四動差之必要。

孫巴克氏一九一六,四十五種商品之指數

指數	x	x^2	x^3	指數	x	x^2	x^3
68	- 68	4,624	- 314,432	138	+ 2	4	8
71	- 65	4,225	- 274,625	148	+ 12	144	1,728
84	- 52	2,704	- 140,608	148	+ 12	144	1,728
86	- 50	2,500	- 125,000	153	+ 17	289	4,913
93	- 43	1,849	- 79,507	154	+ 18	324	5,832
93	- 40	1,600	- 64,000	154	+ 18	324	5,832
100	- 36	1,296	- 46,656	157	+ 21	441	9,261
100	- 36	1,296	- 46,656	159	+ 23	529	12,167
101	- 35	1,225	- 42,875	159	+ 23	529	12,167
104	- 32	1,024	- 32,768	160	+ 24	576	13,824
104	- 32	1,024	- 32,768	161	+ 25	625	15,625
107	- 29	841	- 24,389	163	+ 27	729	19,683
114	- 22	484	- 10,648	163	+ 27	729	19,683
114	- 22	484	- 10,648	166	+ 30	900	27,000
119	- 17	289	- 4,913	168	+ 32	1,024	32,768
121	- 15	225	- 3,375	169	+ 33	1,089	35,937
125	- 11	121	- 1,331	172	+ 36	1,296	40,656
128	- 8	64	- 512	173	+ 37	1,369	50,653
128	- 8	64	- 512	174	+ 38	1,444	54,872
131	- 5	25	- 125	183	+ 47	2,209	103,823
132	- 4	16	- 64	197	+ 61	3,721	226,981
135	- 1	1	- 1	202	+ 68	4,356	287,496
135	- 1	1	- 1				
23	- 632	25,982	- 1,256,414	22	629	22,795	988,637
				23	- 632	25,982	- 1,256,414
				45	- 3	48,777	- 267,777

(第二表)

原點在 136。

$$\bar{x} = -\frac{3}{45} \text{ 平均數 } 136 - \frac{3}{45} = 135.93$$

$$m_2' = \frac{48777}{45} = 1083.933 \quad m_2 = m_2' - \bar{x}^2 = 1083.929 \quad \sigma^2 = \frac{267777}{45} = 5951 \quad \sigma = \sqrt{5951} = 77.14$$

$$m_3' = -\frac{267777}{45} = -5951 \quad m_3 = m_3' - 3\bar{x}m_2' + 2\bar{x}^3 = -5734$$

$$k = \frac{m_3}{\sigma^3} = .161$$

三、北極星赤經度之觀察(註五)。

與假定平均數之秒差	x	觀察頻數	xy	x^2y	x^3y	x^4y
+3.0	6	1	6	36	216	1,296
+2.5	5	5	25	125	625	3,125
+2.0	4	16	64	256	1,024	4,096
+1.5	3	38	114	342	1,026	3,078
+1.0	2	63	126	252	504	1,008
+0.5	1	72	72	72	72	72
0.0	0	82	—	0	—	0
-0.5	-1	73	-73	73	-73	73
-1.0	-2	61	-122	244	-488	976
-1.5	-3	36	-108	324	-972	2,916
-2.0	-4	21	-84	336	-1,344	5,376
-2.5	-5	12	-60	300	-1,500	7,500
-3.0	-6	6	-36	216	-1,296	7,776
-3.5	-7	1	-7	49	-343	2,401
		487	+407 -490	2,625	3,467 -6,016	39,693
			-83		-2,549	

$$\bar{x} = -.170$$

$$m_2 = 5.390 - .029 = 5.361, \sigma = 2.3$$

$$m_3 = -5.234 - 3(-.170) \times 5.390 + 2(-.170)^3 = -2.49 \quad \kappa = -.2$$

$$m_4 = 81.505 - 4(-.170)(-5.234) + 6(-.170)^2(5.390) - 3(-.170)^4 = 78.88, \kappa_2 = 2.7$$

此等觀察，於研究物體觀察用常態曲線表現之可能度時，時常用之。據觀察之結果，曲線尚能對稱，惟因 $k_2 < 3$ ，可見近於平均數處，集中過甚也。

四、下舉一例，示明機率表(Table of Chances)之可以作頻數羣類看待。所選出者為一不對稱之事實，即從十二骰子中擲出六點之機率是。例如，三個六點之機率為

$${}_{12}C_3 \left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^9$$

擲出六點次數	十二個骰子中之機率	
x	y	$\bar{x} = 2$
0.....	244,140,625 ÷ 6 ¹²	
1.....	585,937,500 ,,	$m_2 = 1\frac{1}{2}$
2.....	644,531,250 ,,	$m_3 = 1\frac{1}{3}$
3.....	429,687,500 ,,	$m_4 = 8\frac{1}{15}$
4.....	193,359,375 ,,	$\sigma = 1.29$
5.....	61,875,000 ,,	$\kappa = .516$
6.....	14,437,500 ,,	
7.....	2,475,000 ,,	$\kappa_2 = 3.1$
8.....	309,375 ,,	
9.....	27,500 ,,	
10.....	1,650 ,,	
11.....	60 ,,	
12.....	1 ,,	

2,178,782,336 ,,

五、於整數中，任意抽樣，其平均數，總在4.5左右。下表係從七位對數表之末一數字中，以二十五為一類，抽出四百類之結果。此種分佈，不甚對稱，且 $k_2 > 3$ 。

二十五個數字之和，已用五約過

與22.5之差	次 數	
Over9	1	以23為原點， $\bar{x} = -.2575$; 平均數, 22,7425 $m_2 = 8.8662$; 修正後, 8.783 $\sigma = 2.964$ $m_3 = 13.584$; $\kappa = .522$ $m_4 = 274.24$; 修正後, 269.8 $\kappa_2 = 3.50$
8 to 9	5	
7 ,, 8	9	
6 ,, 7	5	
5 ,, 6	12	
4 ,, 5	10	
3 ,, 4	15	
2 ,, 3	36	
1 ,, 2	48	
0 ,, 1	57	
0 ,, -1	62	
-1 ,, -2	58	
-2 ,, -3	39	
-3 ,, -4	17	
-4 ,, -5	13	
-5 ,, -6	10	
-6 ,, -7	2	
-7 ,, -8	1	
	400	

艾得頓 (Elderton) 氏 (註六), 曾介紹特別適宜用計算機 (Adding and multiplying machine) 計算動差之方法, 茲依本章所用標號, 錄之如下。

以 y_1, y_2, \dots, y_t 爲 $x=1, 2, \dots, t$ 時之次數

$${}_0S_1 = y_t, {}_0S_2 = y_t + y_{t-1}, \dots, {}_0S_t = y_t + y_{t-1} + \dots + y_1.$$

$${}_1S_2 = {}_0S_1 + {}_0S_2, {}_1S_3 = {}_0S_1 + {}_0S_2 + {}_0S_3, \dots, {}_1S_t = {}_0S_2 + {}_0S_3 + \dots + {}_0S_t$$

又 ${}_2S_2 = {}_1S_1 + {}_1S_2, \dots, {}_2S_t = {}_1S_1 + {}_1S_2 + \dots + {}_1S_t$, 等等

$${}_0S_t = \text{觀察數} = n$$

$${}_1S_t = ty_t + (t-1)y_{t-1} + \dots + 1 \cdot y_1 = n\bar{x},$$

式中 \bar{x} (爲平均數) $= nm_1'$ 。

$$\begin{aligned} {}_2S_t &= (1+2+\dots+t)y_t + (1+2+\dots+t-1)y_{t-1} + \dots \\ &+ (1+2)y_2 + y_1 = \frac{t(t+1)}{2}y_t + \frac{(t-1)t}{2}y_{t-1} + \dots + \frac{1 \cdot 2}{2}y_1 \\ &= \frac{n}{2}(m_2' + m_1'), \end{aligned}$$

式中 m_1', m_2', \dots 爲由原點計算之動差。

$$\begin{aligned} {}_3S_t &= \frac{1}{2}\{1 \cdot 2 + 2 \cdot 3 + \dots + t(t+1)\}y_t + \frac{1}{2}\{1 \cdot 2 + \dots + (t-1)t\}y_{t-1} \\ &- 1 + \dots = \frac{1}{6}\{t(t+1)(t+2)y_t + (t-1)t(t+1)y_{t-1} + \dots \\ &\dots + 1 \cdot 2 \cdot 3y_1\} \\ &= \frac{n}{6}(m_4' + 6m_3' + 11m_2' + 6m_1'). \end{aligned}$$

則依公式 5, 7, 9, 得

$$m_2 = \frac{2}{n} \cdot {}_2S_t - \bar{x}(1 + \bar{x})$$

$$m_3 = \frac{6}{n} S_t - 3m_2(1 + \bar{x}) - \bar{x}(1 + \bar{x})(2 + \bar{x})$$

$$m_4 = \frac{24}{n} \cdot {}_4S_t - 2m_3(3 + 2\bar{x}) - m_2(11 + 18\bar{x} + 6\bar{x}^2) - \bar{x}(1 + \bar{x})(2 + \bar{x})(3 + \bar{x}).$$

${}_1S_t, {}_2S_t, {}_3S_t, {}_4S_t$ 幾個數量，只須用重複加法，即可求出。用此法計算第五例所舉之動差，其程序可以充分顯示。

x 係以 14 為原點計算而來； y_x 為 x 之頻數。在 ${}_0S_x$ 直行中各項，乃於前一直行中在其左側之項與該項以上各項相加而得； ${}_1S_x$ 直行亦同樣由 ${}_0S_x$ 行算出，其他以此類推。 $t = 18$ 以 x' 代 $19 - x$ 。

數字之和 ÷ 5	x	xx'	${}_0S_x'$	${}_1S_x'$	${}_2S_x'$	${}_3S_x'$
31.5 以上	18	1	1	1	1	1
30.5	17	5	6	7	8	9
29.5	16	9	15	22	30	39
28.5	15	5	20	42	72	111
27.5	14	12	32	74	146	257
26.5	13	10	42	116	262	519
25.5	12	15	57	173	435	954
24.5	11	36	93	266	701	1,655
23.5	10	48	141	407	1,108	2,763
22.5	9	57	198	605	1,713	4,476
21.5	8	62	260	865	2,578	7,054
20.5	7	58	318	1,183	3,761	10,815
19.5	6	39	357	1,540	5,301	16,116
18.5	5	17	374	1,914	7,215	23,331
17.5	4	13	387	2,301	9,516	32,847
16.5	3	10	397	2,698	12,214	45,061
15.5	2	2	399	3,097	15,311	60,372
14.5	1	1	400	3,497	18,806	79,180
總計		400 = σ_{18}	3,497 = τ_{18}	18,808 = S_{18}	79,180 = s_{18}	235,560 = $4S_{18}$

$$\bar{x} = \frac{3497}{400} = 8.7425 \text{ 平均數} = 22.7425$$

$$m_2 = \frac{2}{400} \times 18808 - 8.7425 \times 9.7425 = 8.8602$$

$$m_3 = \frac{6}{400} \times 76180 - 3 \times 8.8662 \times 9.7425 - 8.7425 \times 9.7425 \times 10.7425 = 13.584$$

$$m_4 = \frac{24}{400} \times 285560 - 2 \times 13.584 \times 20.485 - 8.8662 \times 626.95 - 10744.15 = 274.24$$

- (註一) 爲初步研究用,可參閱 Whitworth 著: *Choice and Chance*。
- (註二) 本書爲便利計不用前此求偏斜度之標號,而用K。
- (註三) 譯者註:石(Stones),重量相當十四磅。
- (註四) 薛伯氏校正(Sheppard's Correction)參閱附錄五。上表中,其他計算亦同,各級人數(即各組次數),可視作恰當各該組之中心,此種解釋,設非分組得當,第二,四動差,顯有誇大之虞,惟如兩極端各級之人數極少,則第一,三動差,所受影響尙微。以 h 代各級之寬度,且不爲一,則經修正後之動差爲 $m_2 - \frac{1}{12}h^2$, 及 $m_4 - \frac{1}{2}h^2m_2 + \frac{7}{240}h^4$ 。
- (註五) 見 Quetelet 著: 概率論 (*Lettres sur la théorie des probabilités*) 第一百二十八頁。】
- (註六) 【類數曲線與相關】第十九至二十三頁。在一十三頁,艾得頓氏示吾人近於中心取用原點之法,以首數字計算之煩。可參考哈地氏著: 死亡率表編製原理 (Hardy, *The Theory of the Construction of Tables of Mortality*) 第五十九頁起。

第二章 代數機率與差誤常態曲線

第一節，初步原理

代數機率之方法與基本定理，可概述如下：

假如有交替之事件 N 個，其中任何一個實現之機會，與其他各個完全均等，設其中一個，已知業經實現，但不知究為何者；並設在 N 個事件中，有 M 個具有特殊性質，其餘 $(N - M)$ 個則無之；則具此特性之事件，出現機率 (Chance)，謂為 $\frac{M}{N}$ 。

例如，從一副五十二張之撲克牌中，抽取一張，則抽出「紅心」之機遇為 $\frac{13}{52} = \frac{1}{4}$ 。此五十二個事件，依吾人所知，每個均有均等機會，蓋承造紙牌者，在監製之下，必須力求紙牌重量相同摩擦力亦相同也。吾人無從指出何種情形會使一牌比他牌易於抽出，除非一個么點之牌，表面摩擦力較王牌者為小。在一理想體系中，如有 N 個事件，其中一個比別個易實現之情況，可謂絕無。公平賭博，所用之工具，必須求其均等，故此種賭博可供機率論之例證。

假如 $p = \frac{M}{N}$ ； $q = 1 - p = \frac{N - M}{N}$ 。 q 為具有特質之某事不出現之機率。如以此某事之出現謂之成功， p 乃即成功之機率， q

爲失敗之機率； p 乃對於 q 而成功， q 對於 p 而失敗。

第二節 機率乘法

假設以 $p_1 p_2$ 爲兩個各自獨立實驗中之成功機率，則 $p_1 \times p_2$ ，可如下文所示，爲二者均成功之機率。

設一實驗，有機會均等之交替事件 n_1 個，另一實驗，則有 n_2 個。則 $p_1 = \frac{m_1}{n_1}$ ， $p_2 = \frac{m_2}{n_2}$

此處所謂互相獨立者，乃指第一實驗所得之結果，對於第二實驗不生影響，故在 $n_1 \times n_2$ 個或有之重複事件中，其中各個全係機會均等。

在此 $n_1 \times n_2$ 個事件中，有 $m_1 \times m_2$ 個，爲二者均成功。

$m_1 \times (n_2 - m_2)$ ……前者成功後者失敗

$(n_1 - m_1) \times m_2$ ……前者失敗而後者成功

$(n_1 - m_1) \times (n_2 - m_2)$ ……二者均失敗

$n_1 n_2$ 個機會均等之事件內，只有 $m_1 m_2$ 個係二者均成功，餘則不能。故二者均成功之機率， p

$$= \frac{m_1 m_2}{n_1 n_2} = p_1 \times p_2$$

例如，擲二骰而俱得六點之機率爲 $\frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$

但如所作幾種實驗，並非互相獨立，而第一種既得結果足以影響第二種之機率，則公式便須變更，如下例所示。

假如從兩副紙牌中，每副各抽一張，則抽得兩個么點之機率爲 $\frac{4}{52} \times \frac{4}{52}$ ，蓋 $p_1 = \frac{4}{52} = p_2$ 也。

但如在一副牌內已抽出一張後，再抽第二張，則有下列幾種可能：

(1) 可供抽出之事件共有 51×52 。

(2) 如先抽出之牌，爲一A點，則在下餘 51 張中，尚有三個A點。

(3) 二者均成功之途徑，有 4×3 個。

(4) 前者成功，後者失敗之途徑，有 4×48 個。

(5) 前者失敗，後者成功之途徑，有 48×4 個。

(6) 二者均失敗之途徑，爲 48×47 個。

(7) 所以二者均成功之機率 $\frac{4}{52} \times \frac{3}{51} = \frac{1}{51}$ 。

此問題寫作下式亦可：

(1) 每副紙牌中，共可抽出 ${}_{52}C_2 = \frac{52 \cdot 51}{1 \cdot 2}$ 對。

(2) 在此 ${}_{52}C_2 = \frac{52 \cdot 51}{1 \cdot 2}$ 對中，抽得兩個么點，只佔 ${}_4C_2 = \frac{4 \cdot 3}{1 \cdot 2}$ 對。任向一對可被抽出之機會，彼此完全相同。所以，在抽兩個么點時，不論一齊抽出或挨次抽出，其抽得之機率，爲 $\frac{{}_4C_2}{{}_{52}C_2} = \frac{4 \cdot 3}{52 \cdot 51}$ 。

另外如欲從一副五十二張牌中，抓到十三張牌，內中要有紅心八張，他樣雜牌五張，則其實現之機率，爲

$$\frac{{}_{13}C_8 \times {}_{39}C_5}{{}_{52}C_{13}} = \frac{13 \cdot 12 \cdot 11 \cdot 10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 39 \cdot 38 \cdot 37 \cdot 36 \cdot 35 \cdot (13!)}{52 \cdot 51 \cdot 50 \cdot 49 \cdot 48 \cdot 47 \cdot 46 \cdot 45 \cdot 44 \cdot 43 \cdot 42 \cdot 41 \cdot 40 \cdot (8!)(5!)}$$

$$= \frac{105,857,037}{90,716,222,800} = \frac{1}{857} \text{ (約略)} = p;$$

因如以十三張爲一把，則機會均等可任得 ${}_{52}C_{13}$ 把 $=N$ ；任得紅心八張，機會均等可有 ${}_{13}C_8$ 種，其他花樣五張，機會均等可有 ${}_{39}C_5$ 種，

$$\therefore M = {}_{13}C_8 \times {}_{39}C_5; \text{ 如此則 } p = \frac{M}{N}.$$

第三節 機率加法

取二骰而擲之，點數之和，共得九點者，有下列各對：(3,6) (4,5) (5,4) (6,3) 換言之，在三十六種機會均等之事件中，得此結果者，只佔其四，故其機率爲 $\frac{4}{36} = \frac{1}{9}$ 。

但此結果由下亦可得之：擲出三點之機率爲 $\frac{1}{36}$ ，六點之機率亦爲 $\frac{1}{36}$ ，故擲出三與六點之機率，爲 $\frac{1}{36}$ 。同理，擲出(4,5) (5,4) 及 (6,3) 之機率，亦各爲 $\frac{1}{36}$ 。由此可知，整個實現之機率，爲各項交替雙重事件機率之總和。

概而論之，若一事件之成功，或由以 p_1 爲機率之機會而實現，與一以 p_1' 爲機率之機會而實現，或由幾個以 p_2, p_2', \dots 爲機率之機會而連續出現，則其成功之總機率，爲

$$p = p_1 p_1' + p_2 p_2' + \dots$$

第四節 差誤常態律之演繹

現在可以討論到與機率原理本身，及其在統計學之應用上，有同等重要性之一普遍定理。

設有一實驗（如擲骰，打牌，或猜數）於此，其成功之機率永為 p ，失敗之機率永為 q ，則 $p+q=1$ 。

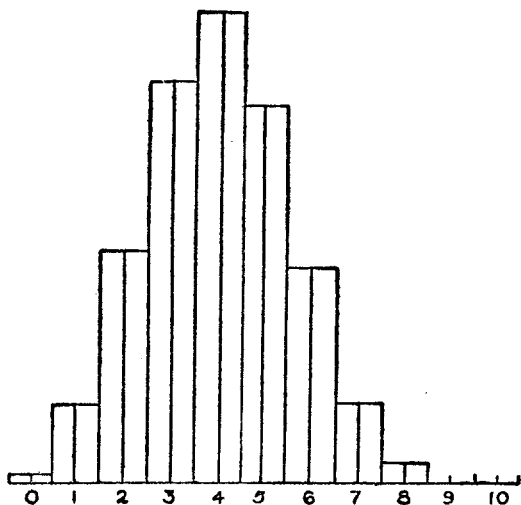
假如實驗重複舉行 n 次，然後考察其成功 r 次及失敗 $n-r$ 次之機率。依其排列之次序，須先有 r 次實驗為成功者，其餘均為失敗之次數，則機率乃為

$p \times p \times \dots$ 到 r 個因子 $\times q \times q \times \dots$ 到 $n-r$ 個因子 $= p^r \times q^{n-r}$ ；且不拘其他何種順序，其機率終屬相同。在一組之 n 次實驗中，選定任一 r 為成功次數，即可得該種順序，即成功有 nC_r 種方法是。是以，其全體機率乃係 $nC_r \cdot p^r q^{n-r}$ 。

成功 $0, 1, 2, \dots, n$ 次之機率乃自然形成二項式之連接項：
 $1 = (q+p)^n = q^n + n \cdot q^{n-1}p + \dots + nC_r \cdot q^{n-r}p^r + \dots + nqp^{n-1} + p^n$

例如， $p = \frac{2}{5}$ ， $q = \frac{3}{5}$ ，又 $n = 10$ 則得

r	nC_r	$p^r q^{n-r}$	$nC_r \cdot p^r q^{n-r}$
0	1	$3^{10} \div 5^{10}$	•006,046,617,6
1	10	2×3^9	•040,310,784,0
2	45	$2^2 \times 3^8$	•120,932,352,0
3	120	$2^3 \times 3^7$	•214,990,848,0
4	210	$2^4 \times 3^6$	•250,822,656,0
5	252	$2^5 \times 3^5$	•200,658,124,8
6	210	$2^6 \times 3^4$	•111,476,736,0
7	120	$2^7 \times 3^3$	•042,467,328,0
8	45	$2^8 \times 3^2$	•010,616,832,0
9	10	$2^9 \times 3^1$	•001,572,864,0
10	1	2^{10}	•000,104,857,6
			1,000,000,000,0



此縱標尺擴大百倍，故此圖面積為一百個單位的方格。

上圖例釋成功次數多寡不同之相對機率，並以之作成一『頻數羣』。

先就 p 及 n 之普通數值，將該羣類之動差求得，以上圖之橫標尺為 x 之尺度。

假設包括 n 個事件之實驗重複舉行 N 次，而 N 又係極大之數。則有 r 次成功之回數變為 $N \times {}_n C_r \cdot q^{n-r} p^r = y_r$

而且 $y_0 + y_1 + \dots + y_n = N(q+p)^n = N$ ，因 $p+q=1$ ，

$\bar{x} = m_1'$ ，由原點之第一動差

$$= (y_0 \times 0 + y_1 \times 1 + \dots + y_r \times r + \dots + y_n \times n) \div N$$

$$\begin{aligned}
&= n \cdot q^{n-1} p + n(n-1) / 2 \cdot q^{n-2} p^2 \times 2 + \dots + {}_n C_r q^{n-r} p^r \\
&\quad \times r + \dots + p^n \times n \\
&= n p (q+p)^{n-1} = n p \dots \dots \dots (11)
\end{aligned}$$

$$\begin{aligned}
m_2' &= (y_0 \times 0^2 + y_1 \times 1^2 + \dots + y_r \times r^2 + \dots + y_n \times n^2) \div N \\
&= \sum_0^r {}_n C_r \cdot q^{n-r} p^r = \sum \{r(r-1) + r\} \frac{(n)_r}{r!} q^{n-r} p^r \\
&= n(n-1) p^2 \sum \frac{(n-2)_{r-2}}{(r-2)!} q^{n-r} p^{r-2} + n p \sum \frac{(n-1)_{r-1}}{(r-1)!} \\
&\quad q^{n-r} p^{r-1} \\
&= n(n-1) p^2 (q+p)^{n-2} + n p (q+p)^{n-1} = n(n-1) p^2 + n p \\
&= n^2 p^2 + n p (1-p) = \bar{x}^2 + n p q \dots \dots \dots (12)
\end{aligned}$$

而以平均數為原點之第二動差, m_2

$$= m_2' - \bar{x}^2 = n p q = n p (1-p) \dots \dots \dots (13)$$

同樣

$$m_3' = \sum_0^n {}_n C_r \cdot r^3 \cdot q^{n-r} p^r = n(n-1)(n-2) p^3 + 3n(n-1) p^2 + n p \dots (14)$$

至以平均為原點之第三動差, m_3

$$\begin{aligned}
&= m_3' - 3m_2' \bar{x} + 2\bar{x}^3 \\
&= n(n-1)(n-2) p^3 + 3n(n-1) p^2 + n p - 3n^2 p^3 - 3n^2 p^2 \\
&\quad (1-p) + 2n^3 p^3 \\
&= n p (2p^2 - 3p + 1) = n p (1-p)(1+2p) = n p q (q-p) \dots \dots (15)
\end{aligned}$$

$m_4' = \sum_0^n {}_n C_r \cdot r^4 \cdot q^{n-r} p^r$, 而 m_4 亦可證明等於

$$3(pqn)^2 + pqn(1 - 6pq).$$

故依第一章第三節之公式, $\sigma = \sqrt{pqn}$, $\beta_1 = \frac{(q-p)^2}{pqn}$, $k_2 =$

$$\beta_2 = 3 + \frac{1-6pq}{pqn}, C = \sqrt{2pqn}, k = \frac{q-p}{\sqrt{pqn}}, i = \frac{1-6pq}{4pqn}.$$

標準差隨 \sqrt{n} 而變化。 \sqrt{n} 若大時, 偏斜度量數, k 及 $\sqrt{\beta_1}$ 必小。 n 若大時, $(k_2 - 3)$ 與 i 亦必低。

復次, 請就成功 r 次之機率, 及 n 增加時之圖形討論。

第一例: 當 $p = q = \frac{1}{2}$, n 為偶數 $= 2n'$ 時。

設 P_x 為 $n' + x$ 次成功之機率, 亦即 $(n' - x)$ 次失敗之機率。

$$P_x = {}_{2n'}C_{n'+x} \cdot \frac{1}{2^{n'+x}} \frac{1}{2^{n'-x}} = \frac{(2n')!}{(n'+x)!(n'-x)!} \cdot \frac{1}{2^{2n'}}$$

$$= \frac{(2n')!}{n'!n'!} \cdot \frac{1}{2^{2n'}} \cdot \frac{n'(n'-1)\cdots(n'-x+1)}{(n'+1)\cdots(n'+2)\cdots(n'+x)}$$

$$P_0 = \frac{(2n')!}{2^{2n'} \cdot n'!n'!} = \frac{1}{\sqrt{\pi n'}}, \text{根據瓦立斯氏 (Wallis) 定理, 核正}$$

至 $\frac{1}{n'}$ 止 (附錄一, 公式 132)。

$$\therefore P_x = \frac{1}{\sqrt{\pi n'}} \cdot \frac{\left(1 - \frac{1}{n'}\right)\left(1 - \frac{2}{n'}\right)\cdots\left(1 - \frac{x-1}{n'}\right)}{\left(1 + \frac{1}{n'}\right)\left(1 + \frac{2}{n'}\right)\cdots\left(1 + \frac{x}{n'}\right)}$$

$$\log(P_x \sqrt{\pi n'}) = \log\left(1 - \frac{1}{n'}\right) - \log\left(1 + \frac{1}{n'}\right) + \log\left(1 - \frac{2}{n'}\right) - \log$$

$$\left(1 + \frac{2}{n'}\right) + \cdots + \log\left(1 - \frac{x}{n'}\right) - \log\left(1 + \frac{x}{n'}\right) - \log\left(1 - \frac{x}{n'}\right)$$

$$\begin{aligned}
 &= -2\left(\frac{1}{n'} + \frac{1}{3n'^3} + \dots\right) - 2\left(\frac{2}{n'} + \frac{2^3}{3n'^3} + \dots\right) \dots\dots \\
 &\qquad - 2\left(\frac{x}{n'} + \frac{x^3}{3n'^3} + \dots\right) - \log\left(1 - \frac{x}{n'}\right) \\
 &= -\frac{2(1+2+\dots+x)}{n'} - \frac{2}{3} \cdot \frac{1^3+2^3+\dots+x^3}{n'^3} \dots \frac{2}{2t+1} \\
 &\qquad \cdot \frac{1^{2t+1}+2^{2t+1}+\dots+x^{2t+1}}{n'^{2t+1}} \dots - \log\left(1 - \frac{x}{n'}\right)
 \end{aligned}$$

t 爲任一整數

$$\begin{aligned}
 &= -\frac{x(x+1)}{n'} - \frac{2x^2(x+1)^2}{3 \cdot 4n'^3} - \dots \\
 &\qquad - \frac{2}{2t+1} \cdot \frac{x^{2t+2} + \dots + \dots}{(2t+2)n'^{2t+1}} \text{ (註一) } - \dots + \left(\frac{x}{n'} + \frac{x^2}{2n'^2} + \dots\right)
 \end{aligned}$$

以 $x = \tau\sqrt{n'} = \tau c$, 因爲根據第一章第三節,

$$c^2 = 2pqn = 2 \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot 2n' = n'.$$

$$\begin{aligned}
 \log(P_x \cdot c\sqrt{\pi}) &= -\tau^2 - \frac{\tau}{\sqrt{n'}} - \frac{2}{3n'} \frac{\tau^2}{4} \left(\tau + \frac{1}{\sqrt{n'}}\right)^2 - \dots \\
 &\qquad - \frac{2}{(2t+1)(2t+2)} \left(\frac{\tau^{2t+2}}{n'^t} + \dots\right) - \dots + \left(\frac{\tau}{\sqrt{n'}} + \frac{\tau^2}{2n'} + \dots\right) \\
 &= -\tau^2 + \text{包含 } \frac{1}{\sqrt{n'}} \text{ 各項}
 \end{aligned}$$

故如上述 P_0 之值, 將 $\frac{1}{\sqrt{n'}}$ 略去, 則

$$P_x = \frac{1}{c\sqrt{\pi}} e^{-\tau^2} = \frac{1}{c\sqrt{\pi}} e^{-\frac{x^2}{c^2}} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{\sigma^2}} \dots\dots\dots (16)$$

因爲標準差， $\sigma = c/\sqrt{2}$ 。

$$\text{又因 } c^2 = \frac{n}{2}, \therefore P_x = \frac{\sqrt{2}}{\sqrt{\pi n}} e^{-\frac{2c^2}{n}}$$

第二例：在 p 與 q 不相等時。

設 P_x 爲 $pn+x$ 個成功之機率（註二），亦即 $qn-x$ 失敗之機率。

$$\begin{aligned} P_x &= \frac{n!}{(pn+x)!(qn-x)!} \cdot p^{pn+x} q^{qn-x} \\ &= \frac{1}{(pn)!(qn)!} p^{pn} q^{qn} \cdot \frac{qn(qn-1)\cdots(qn-x+1)}{(pn+1)(pn+2)\cdots(pn+x)} \frac{p^x}{q^x} \\ &= \frac{\left(1 - \frac{1}{qn}\right)\left(1 - \frac{2}{qn}\right)\cdots\left(1 - \frac{x-1}{qn}\right)}{\left(1 + \frac{1}{pn}\right)\left(1 + \frac{2}{pn}\right)\cdots\left(1 + \frac{x}{pn}\right)} \\ \log(P_x/P_0) &= \sum_{s=1}^{s=x} \log\left(1 - \frac{s}{qn}\right) - \sum_{s=1}^{s=x} \log\left(1 + \frac{s}{pn}\right) - \log\left(1 - \frac{x}{qn}\right) \\ &= - \sum_1^x \left(\frac{s}{qn} + \frac{s}{pn} - \sum_1^x \frac{1}{2} \left(\frac{s^2}{q^2 n^2} - \frac{s^2}{p^2 n^2}\right) - \dots\right) \\ &\quad - \sum_1^x \frac{1}{t} \left(\frac{s^t}{q^t n^t} \pm \frac{s^t}{p^t n^t}\right) - \dots - \log\left(1 - \frac{x}{qn}\right) \\ &= - \frac{x(x+1)}{2} \cdot \frac{p+q}{pqn} - \frac{x(x+1)(2x+1)}{6} \cdot \frac{p^2 - q^2}{2p^2 q^2 n^2} \\ &\quad - \frac{x^2(x+1)^2}{4} \cdot \frac{p^3 + q^3}{3p^3 q^3 n^3} - \dots \\ &\quad - \frac{1}{t} \cdot \frac{x^{t+1} + \dots}{t+1} \cdot \frac{p^t \pm q^t}{p^t q^t n^t} - \dots + \left(\frac{x}{qn} + \frac{x^2}{2q^2 n^2} + \dots\right) \end{aligned}$$

以 $x = \tau c$, 而 $c^2 = 2pqn = 2\sigma^2$

$$\begin{aligned} \log(P_x/P_0) = & -\frac{\tau^2 c^2 + \tau c}{c^2} + \frac{2\tau^3 c^3 + 3\tau^2 c^2 + \tau c}{3c^4} (q-p) \\ & - \frac{\tau^4 c^4 + 2\tau^3 c^3 + \tau^2 c^2}{\frac{3}{2} \cdot c^6} \cdot (1-3pq) - \dots \\ & - \frac{\tau^t + 1}{t(t+1)2^{-\tau c^{2t}}} (p^t \pm q^t) - \dots + \frac{2\tau c p}{c^2} + \frac{2\tau^2 c^2 p^2}{c^4} + \dots, \end{aligned}$$

因 $P+q=1$,

$$\begin{aligned} = & -\tau^2 + \frac{\tau}{c} \left\{ -1 + \frac{2\tau^2}{3} (q+p) + 2p \right\} \\ & + \frac{\tau^2}{c^2} \left\{ (q-p) - \frac{2\tau^2}{3} (q-p) + 2p \right\} \\ & + \text{包含 } \frac{1}{c^3} \text{ 之項} \end{aligned}$$

以 τ 為有限數論, 換言之, 在 x 之乘數值中, 只就其可與 \sqrt{pqn} 比較者討論之。

如刪略 $\frac{1}{c}$ (即刪略 $\frac{1}{\sqrt{n}}$), 則得

$$P_x = P_0 e^{-\tau^2} = P_0 e^{-\frac{x^2}{c^2}} \dots \dots \dots (17)$$

如保留 $\frac{1}{c}$, 略去 $\frac{1}{c^2}$ (即刪略 $\frac{1}{n}$), 則得

$$P_x = P_0 e^{-\tau^2} \cdot e^{-\frac{\tau(q-p)(1-\frac{2}{3}\tau^2)}{c}}$$

$$= P_0 e^{-\tau^2} \left\{ 1 - \frac{q-p}{c} \left(\tau - \frac{2}{3}\tau^3 \right) \right\}, \text{ 因 } \frac{1}{c^2} \text{ 已略去}$$

$$\begin{aligned}
 &= P_0 e^{-\frac{x^2}{\sigma^2}} \left\{ 1 - \frac{q-p}{c} \left(\frac{x}{c} - \frac{2}{3} \frac{x^3}{c^3} \right) \right\} \\
 &= P_0 e^{-\frac{x^2}{2\sigma^2}} \left\{ 1 - \frac{k}{2} \left(\frac{x}{\sigma} - \frac{x^3}{3\sigma^3} \right) \right\}, \dots\dots\dots (18)
 \end{aligned}$$

因 $c = \sqrt{2}\sigma$, $\sigma = \sqrt{2pq/n}$, $k = \frac{q-p}{\sigma}$ 。

P_0 之值可用斯德令 (Stirling) 氏級乘 (factorial) 定理求之 (附錄三, 公式 134), 即 $m! = m^m \sqrt{2\pi m} \cdot e^{-m + \frac{1}{12m}}$, 此係已將 $\frac{1}{m^2}$ 刪略, 如再將 $\frac{1}{m}$ 刪除, 則 $m! = m^m \sqrt{2\pi m} e^{-m}$ 。

$$\begin{aligned}
 P_0 &= \frac{n!}{(pn)!(qn)!} p^n q^n \\
 &= \frac{n^n}{(pn)^n (qn)^n} \sqrt{\frac{2\pi n}{2\pi pn \cdot 2\pi qn}} \cdot e^{-n+pn+qn} \cdot p^n q^n \\
 & \qquad \qquad \qquad \text{略去 } \frac{1}{pn}, \frac{1}{qn}, \dots\dots\dots, \\
 &= \frac{1}{\sqrt{2\pi pqn}} \quad (\text{因 } p+q=1), = \frac{1}{c\sqrt{\pi}} = \frac{1}{\sigma\sqrt{2\pi}}.
 \end{aligned}$$

以 y 代 P_x , 則得方程式

$$y = \frac{1}{\sqrt{2\pi pqn}} e^{-\frac{x^2}{2pqn}} = \frac{1}{c\sqrt{\pi}} e^{-c^2} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} \dots\dots\dots (19)$$

此係已將 $\frac{1}{\sqrt{n}}$ 略去, 如保留 $\frac{1}{\sqrt{n}}$, 而刪去 $\frac{1}{n}$, 則

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} \left\{ 1 - \frac{k}{2} \left(\frac{x}{\sigma} - \frac{x^3}{3\sigma^3} \right) \right\} \dots\dots\dots (20)$$

如上所述, 當舉行包有 n 種事件之實驗時, 成功之次數總比

pn 多 x , (此處之 p 為單純試驗中之成功機率), 在此情形之下之機率, 即上列方程式所表示者也。

$y = \frac{1}{c\sqrt{\pi}} e^{-\frac{x^2}{\sigma^2}}$ 所代表之曲線, 謂之差誤常態曲線 (註三) (normal curve of error)。圖形附在本書之末者即是。

$\frac{1}{\sqrt{n}}$ 項之重要如以 $n=1000$, $P=\frac{1}{10}$, 代入之即可見之。若是則 $\sigma = \sqrt{90} = 9.5$, k 略等於 .084。當 x 大於 σ 時, 機率感受影響頗為靈敏。

n 若大, 任一成功次數之實際機率必小, 舉例言之, 假如 $P = \frac{1}{10}$, $n=1000$, 成功恰當 500 次 (最大概然數) 時之機率, 僅約達 $\frac{1}{4}$ 而已。吾人所需用者並非特定縱坐標之量數, 而在各數值全距之機率總和數量; 所謂各數值之全距者, 例如由 x_1 至 x_2 , 即是, 此處之 $x_2 - x_1$, 與 $\sigma (= \sqrt{pqn})$, 為同一方次。

藉一極著盛名之定理 (註四) 吾人可以從縱坐標之總和, 而達面積之積分, 成功之次數, 大小與 $pn + x_1$ 同, 而未必比 $pn + x_2$ 大, 則其全部機率必為 $\int_{x_1}^{x_2} y dx$, 此處之 $y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}$, 並已將含有 $\frac{1}{\sqrt{n}}$ 各項略去。

以 z 代 $\frac{x}{\sigma}$ 則得 $\int_{x_1}^{x_2} y dx = \frac{1}{\sqrt{2\pi}} \int_{z_1}^{z_2} e^{-\frac{1}{2}z^2} dz$, 此函數數值表, 容當附及 (即第三表)

上文所提及之函數, 其重要常數, 得出之法如下:

曲線面積 = $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz = (p+q)^n$ 之極限，當 n 漸近於無限大時，= 1。

$$\therefore \int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2} dz = \sqrt{2\pi}; \quad \int_{-\infty}^{\infty} e^{-n^2} \cdot du = \sqrt{\pi};$$

$$\int_{-\infty}^{\infty} e^{-au} du = \sqrt{\frac{\pi}{a}}; \quad \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} dx = 1$$

以 $m_s = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} \cdot x^s dx$ ，為對平均數之第 S 次動差，

此乃以平均數為原點，面積為 1。

此曲線經過原點之縱坐標成為對稱，且 $m_{2t+1} = 0$ ，不拘 t 為何值（註五）。

$$\begin{aligned} m_2 &= \frac{1}{2\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2\sigma^2}} dx \\ &= \left[-\frac{\sigma}{\sqrt{2\pi}} x e^{-\frac{x^2}{2\sigma^2}} \right]_{-\infty}^{\infty} + \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2\sigma^2}} dx \\ &= 0 + \sigma^2 = \sigma^2 \dots\dots\dots (21) \end{aligned}$$

此在公式 13 已論及矣。

$$\begin{aligned} m_{2t} &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x^{2t} e^{-\frac{x^2}{2\sigma^2}} dx \\ &= \left[-\frac{\sigma}{\sqrt{2\pi}} x^{2t-1} e^{-\frac{x^2}{2\sigma^2}} \right]_{-\infty}^{\infty} + \frac{(2t-1)\sigma^2}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x^{2t-2} e^{-\frac{x^2}{2\sigma^2}} dx \\ &= 0 + (2t-1) \sigma^2 m_{2t-2} \dots\dots\dots (22) \end{aligned}$$

$$\text{故 } m_4 = 3\sigma^2 m_2 = 3\sigma^4 = 3m_2^2,$$

$$k_2 = \beta_2 = \frac{m_4}{m_2^2} = 3, i = \beta_2 - 3 = 0$$

此數依公式 16 亦可求得，當 n 為無限時。

$$\begin{aligned} m_{2t} &= (2t-1)(2t-3)\cdots\cdots 3 \cdot 1 \sigma^{2t}, \text{ 用歸納法,} \\ &= \frac{(2t)!}{2^t t!} \sigma^{2t} \cdots\cdots\cdots (23) \end{aligned}$$

$$\text{例如 } m_6 = 15\sigma^6, m_8 = 105\sigma^8.$$

平均差， n ，因面積為一，故

$$= \frac{2}{2\sqrt{2\pi}} \int_0^{\infty} x e^{-\frac{x^2}{2\sigma^2}} dx = \left[-\frac{2\sigma}{\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} \right]_0^{\infty} = \sigma \cdot \sqrt{\frac{2}{\pi}} \cdots (24)$$

$$\therefore \frac{\sigma}{n!} = \sqrt{\frac{\pi}{2}}$$

『機誤』(見第一編第六章第一節)，可查 z (而此 z 須使 $\int_0^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz = \frac{1}{4}$) 之數值表而求得。據計算機誤之數值為 $x = z\sigma = .674490\sigma$

本書之末，附有曲線圖。其轉向點(points of inflection)係使 $Dx^2 = 0$ 而得，式中之

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}.$$

如是，則 $\log y + \text{常數} = -\frac{x^2}{2\sigma^2}$

$$\frac{1}{y} Dxy = -\frac{x}{\sigma^2}$$

$$\frac{1}{y} \cdot 1.2xy - \frac{1}{y^2} (D_x y)^2 = -\frac{1}{\sigma^2}$$

$$\therefore -\left(-\frac{x}{\sigma^2}\right)^2 = -\frac{1}{\sigma^2} \quad (\text{在轉向點}) \dots\dots\dots (25)$$

曲線自底線。至 σ 之一段，當等於表上數值之

$$\frac{1}{\sigma\sqrt{2\pi}} \int_0^{\sigma} e^{-\frac{1}{2}\frac{x^2}{\sigma^2}} dx = \frac{1}{\sqrt{2\pi}} \int_0^1 e^{-\frac{1}{2}z^2} dz = F(1) = .3413;$$

照樣查表，則可立得下列各近似值：

各段底線之曲線面積比例

底線	面積	底線	面積
0 - .2 σ	.07926	- .2 σ 至 + .2 σ	.1585
0 - .6 σ	.2257	+ .2 σ ,, + .6 σ	.1465
0 - 1.0 σ	.3413	+ .6 σ ,, + 1.0 σ	.1156
0 - 1.4 σ	.4192	+ 1.0 σ ,, + 1.4 σ	.0779
0 - 1.8 σ	.4641	+ 1.4 σ ,, + 1.8 σ	.0449
0 - 2.2 σ	.4861	+ 1.8 σ ,, + 2.2 σ	.0220
0 - 2.6 σ	.4953	+ 2.2 σ ,, + 2.6 σ	.0092
0 - 3.0 σ	.49865	+ 2.6 σ ,, + 3.0 σ	.0033

(註六)

所謂機誤為自中心位置向左向右佔有全觀察數量，恰當一半之距離。

據計算結果， $F(z) = \frac{1}{4}$ ，當 $z = .67449$ (約略數)。故該曲線之四分位數，乃在

$$\pm .67449\sigma \dots\dots\dots (26)$$

單個觀察在此全距之內或在此全距之外，機會恰各佔一半，此，.67449 σ ，即『機誤』，測算精度時常用之，以代替 σ 等。

第五節 代數機率與經驗

上文之分析，純爲抽象者，紙牌及骰子，種種例證，亦僅使『機會均等』一詞易於想像而已。吾人現應就成功之次數，依其代數機率之比例而出現一點，考察其究有何種證據。吾人雖必因（在簡單事例）成功次數之比例完全異樣——例如取百錢而擲之，表面在上者，竟佔其九十，或從全副牌中抽取一張，抽後混入重抽，如此遞抽五十次，所抽之牌，設全爲紅心——而驚奇，然此驚奇之感覺，不過使吾人認爲全宇宙中，機遇顯然事件，本有若干法則，存乎其中。於此舍經驗與試驗不爲功。以一般論之，據賭博者之經驗，事件之實現，無論如何大都合於代數之機率，如玩五十二張之牌，四人對局，每人每次派一張，以得大牌爲得分，則大牌之出現次數，即係根據此理，其勝利與否乃依所算之機率而決定也。人壽保險與意外保險，即基於此一信念而生，蓋個別事實雖屬變動無常，然大量事件，頗可預測於事先，且此信念業經事實不斷證明矣。已往爲考察事件實現次數與其應有之本來機率比較，曾作極多之實驗，結果顯屬成功。

然均等機率之原有條件，吾人不能斷其必然能以完全滿足，而且吾人所得亦不能期望超過一約略之驗證也。

粗率之實驗，用極簡單之用具，即可舉行。例如，取撲克牌若

干副，將其有圖像各張，統行剔出後，每次抽出出張，計其共得點數，然後重行插入再抽。如此進行，以至九十次為止。

設紙牌副數異常之多，足使每次抽出四張中之各色牌在抽取時，不受牌數之拘束，則共得 r 點之機率，乃為：

$\frac{1}{10^4}(x+x^2+\dots+x^{10})^4$ 即 $\frac{x^4}{10^4} \cdot \left(\frac{1-x^{10}}{1-x}\right)^4$ 一式內 x^r 之係數，此機率與逐次實驗之結果，可表列如下：——

	機率合計	×90=希望數	實驗得數
4 至 9	.0126	1.134	0
10 ,, 14	.0871	7.839	7
15 ,, 19	.2375	21.375	27
20 ,, 24	.3256	29.304	25
25 ,, 29	.2575	21.375	22
30 ,, 34	.0871	7.839	9
55 ,, 40	.0126	1.134	0
	1.0000	90.000	90

九十組（每四張為一組）所有各點總數為 1956，平均數每張合 5.43。各副所有各牌之平均數為 5.5。

試驗結果與希望數相合，至少亦相逼近。

第六節 白諾立氏定律

吾人次須研究理論頻數與所期頻數（即理論示吾人所期望之頻數），二者相互間之適應程度。為達此目的，差誤律(Law of error)可供測驗之用。

即就上文所舉實驗， $r=15$ 至 19 一類討論之。由此全距中，猜數之機率為 $.2375=p$ 。九十次實驗中，在此全距內，猜數 t 次

之機率為 $(q+p)^{90}$ 之第 $t+1$ 項。實現機會最多之數，非為 21 即為 22，而其各種可能的成功數之標準差為

$$\sqrt{pqn}, \text{ 其中 } n=90, \text{ 故在 } 4 \text{ 左右。}$$

在此重疊舉行多次之繁複實驗中，從 17 至 26 一類內，任何成功數之機率，用前表查出，大約等於 $\frac{2}{3}$ ；求得一大與 27 相等之數（如上試驗表中所列），其機率約為 $\frac{1}{8}$ 。與 21 有三個標準差之分歧，未必即有此事；換言之，實現次數多過 33，少於 9，殊屬罕見之事也。

此種方法，概而言之，已引到白諾立氏定律 (Bernoulli's law)，茲將此定律解釋如下。假如成功機率為 p 之一實驗，舉行 n 次，以 $p'n$ 代表成功次數，則若 n 增加， p' 將漸近於 p 。比 $p \sim p'$ 較大之離中差之實現機率，為

$$2 \int_z^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz,$$

但

$$z = \frac{p \sim p'}{\sqrt{\frac{p(1-p)}{n}}},$$

故當 \sqrt{n} 增加時，任何離中差之機率必隨之減少。 n 增加至相當程度，可使機率任意縮小（註七）。

基於一般之經驗及多次實驗之結果，白氏定律 確合於事實。

於是，設有原為機會均等之事件，吾人已知其條件如何，可以用數學機率方法，計算各種事件之機率，且吾人可以期望計算之結果，實際上可在差誤律決定範圍之內。

下文將述及種種實驗之結果。爲首三則係在示已得之分配與依差誤律而得者之比較，其餘各例，在用抽樣方法示於範圍甚廣之羣類內決定級距之程序。

第七節 例證

I. 假如任意抽取一單位數字，則抽得之數比五少（如 0, 1, 2, 3, 或 4）之機率爲 $\frac{1}{2}$ 。取一對數表，將其第七位小數，每次抽出五十個，將其少於五之數 r 個，記下。抽得 r 個此等數字之機率， $(\frac{1}{2} + \frac{1}{2})^{50}$ 展開式之第 $r+1$ 項即是。 $n=50$, $p=q=\frac{1}{2}$, $\sqrt{pqn} = 3.535 = \sigma$ 。

抽得機會最多之數， pn ，爲 25。得數不過於 $25+x$ 之機率，即前列常態機率表之 $F(z)$ ，此處之 $z = \frac{x}{\sigma} = \frac{x}{3.535}$ ，假設 $n=50$ 在對稱曲線， n 已甚大，足以應用常態曲線，庶可不用二項式級數。

每次抽取五十字之試驗，舉行三百次

實在頻數與理論頻數，二者之相合程度，果如理論所示（參閱第十章）。標準差理論上應爲 $\sqrt{pqn} = 3.535$ 。援用第一章第四節所得第二級動差之平方根，亦可求得觀察之實在標準差。其平均數爲 25.043。以 25 爲原點而得之第二動差，爲 $(1 \times 11^2 + 0 \times 10^2 + 3 \times 9^2 + \dots + 1 \times 10^2) \div 300 = 11.30$ ，而以平均數爲原點之第二動差，則爲 $11.300 - .043^2 = 11.293$ 。其平方根爲 3.361，與

z	$F(z)$	差額 (註八) $\times 300 =$ 實現次數 (註九)						
		期望數	實在數	在				
13.5	-3.2522	.4994	.0008	.2	0	1	在	14
14.5	-2.9694	.4986	.0020	.6	0或1	0	,,	15
15.5	-2.6866	.4966	.0047	1.4	1或2	3	,,	16
16.5	-2.4038	.4919	.0088	2.6	2或3	2	,,	17
17.5	-2.1210	.4831	.0161	4.8	5	3	,,	18
18.5	-1.8382	.4670	.0270	8.1	8	7	,,	19
19.5	-1.5554	.4400	.0416	12.5	12或13	9	,,	20
20.5	-1.2726	.3984	.0595	17.85	18	18	,,	21
21.5	- .9898	.3389	.0787	23.6	24	26	,,	22
22.5	- .7070	.2602	.0959	28.8	29	21	,,	23
23.5	- .4242	.1643	.1082	32.5	32或33	32	,,	24
24.5	- .1414	.0561	.1122	33.7	39	42	,,	25
25.5	+ .1414	.0561	.1082	32.5	32或33	36	,,	26
26.5	+ .4242	.1643	.0959	28.8	29	30	,,	27
27.5	+ .7070	.2602	.0787	23.6	24	28	,,	28
28.5	+ .9898	.3389	.0595	17.85	18	15	,,	29
29.5	1.2726	.3984	.0416	12.5	12或13	16	,,	30
30.5	1.5554	.4400	.0270	8.1	8	5	,,	31
31.5	1.8382	.4670	.0161	4.8	5	2	,,	32
32.5	2.1210	.4831	.0088	2.6	2或3	2	,,	33
33.5	2.4038	.4919	.0047	1.4	1或2	1	,,	34
34.5	2.6866	.4966	.0020	.6	0或0	1	,,	35
35.5	2.9694	.4986	.0008	.2	0	0	,,	36
36.5	3.2522	.4994						

299.6

理論之標準差，相差不過 .174，而此 .174 數，原亦難免之相差也（參閱公式120）。

II. 如不求各個數值之希望數，而照下列所示之法，仍可測驗其分佈情況。

設有書一本，每頁三十七行，如以一百頁為基準，數其各行首字之字母，其字母數為1, 2 而不過3者，共有若干次數。據試驗結果，在三千七百行中，行首字字母數不過3個者，共有一千三百一十七次。故此種各行首字字母不過3個者之機率，為

$$p = \frac{1317}{3700}, q = 1 - p = \frac{2383}{3700}.$$

每頁中各行首字，字母不過三個，共有 r 字之機率約與 $(q+p)^{37}$ 二項式之第 $r+1$ 項相當。

理論之標準差為 $\sqrt{pqn} = 2.913 = \sigma$ 。

其實現次數如下：

各行首字字母在 3 以下者之實現次數	出現此種字之頁數
7	1
8	2
9	9
10	6
11	8
12	17
13	15
14	12
15	13
16	5
17	4
18	2
19	3
20	2
21	0
22	1

平均數為 $13.17 = \bar{x}$ ；從此觀察算出之標準差為 2.922。然後計算，在各級標準差（平均數算出）中，所期之實現次數如下：

		差額 $\times 100$	實現數	
$\bar{x} - 3\sigma = 4.43$	$F(-3) = .499$	2.3	$7\frac{1}{2}$ 以下	1
$\bar{x} - 2\sigma = 7.34$	$F(-2) = .177$	13.6	$7\frac{1}{2}$ 至 $10\frac{1}{2}$	17
$\bar{x} - \sigma = 10.26$	$F(-1) = .341$	34.1	$10\frac{1}{2}$,, $13\frac{1}{2}$	40
$\bar{x} = 13.17$	$F(0) = .0$	34.1	$13\frac{1}{2}$,, $16\frac{1}{2}$	30
$\bar{x} + \sigma = 16.08$	$F(1) = .341$	13.6	$16\frac{1}{2}$,, $19\frac{1}{2}$	9
$\bar{x} + 2\sigma = 19.00$	$F(2) = .477$	2.3	$19\frac{1}{2}$,, $22\frac{1}{2}$	3
$\bar{x} + 3\sigma = 21.91$	$F(3) = .499$	100.0		100

觀察之量數，非為整數不可者，不易使各級成為 σ 之倍數。但

若觀察之分級狹小，或其量數係繼續不斷者，則此法——以標準差， σ 之相等的次倍數(submultiple)為準——甚為簡捷，且在測驗未應用之前，分級即可決定，故可供為一種良好而簡單測驗之用。

III. 以一商號題名錄，亦可作同樣之試驗。設此題名錄共有七十四頁，頁各四十家。為行政上目的，各商店之僱用女人者，皆逐一標明。結果得知商店僱用婦女者，佔有五分之一。故在任何一頁，查出 r 家商店之機率，為 $(q+p)^{40}$ 式中之第 $r+1$ 項。此處 $p = \frac{1}{5}$, $\sigma = \sqrt{(\frac{1}{5} \cdot \frac{4}{5} \cdot 40)} = 2.53$, $pn = 8$ 。

	期望數	實在數
$pn+2\sigma$ 至 $pn+3\sigma$	1.7	2或3
$+\frac{3}{2}\sigma$, $+2\sigma$	3.3	5, 6或7
$+\sigma$, $+\sigma$	6.8	4或5
$+\frac{1}{2}\sigma$, $+\sigma$	11.0	9, 10, 11
0, $+\frac{1}{2}\sigma$	14.2	13或14
$-\frac{1}{2}\sigma$, $+0$	14.2	15或16
$-\sigma$, $-\frac{1}{2}\sigma$	11.0	8
$-\frac{3}{2}\sigma$, $-\sigma$	6.8	7
-2σ , $-\frac{3}{2}\sigma$	3.3	2
-3σ , -2σ	1.7	5

最末一欄，有交替之數，乃由於不易使各款 (entries) 與前定各級適合。

在此情形之下，獨立性之先決條件，未能完全實現：查出被標出商號之機率，不應因同頁被標出商店之有無，而受其影響；惟實際上有時同一商店因設有支店關係，店名或屢見不一見，而所有支店均僱或均未僱有婦女也。

第八節 對於抽樣法之應用

此定理對於在若干次試驗中預料之成功次數，有一主要用途，即在藉抽樣方法，以研究範圍寬廣之羣類。茲將此定理簡述如下：

(1) 在一範圍 (universe) 之內，設有人或物 N 個，其中有 pn 個，為具有某種特性者， N 為已知數， p 為未知數。

(2) 由範圍之內，隨機抽出 n 個事物，並發覺其中有 $p'n$ 個，係含有某種特性者。

(3) 假如 $\frac{n}{N}$ 之值甚小(註一〇)並設在抽選過程中，各個事物各有其均等的被抽之機會，且抽取此一事物時，不至影響其他任何事物之被選機緣，則取得 $(p+x/n)$ 個事物之機率可由 $\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{x^2}{2\sigma^2}}$ 而求得，式中之 $\sigma = \sqrt{\frac{p(1-p)}{n}}$ ，可由前列常態機率表 (第三表) 求之。至於精度 (precision)，以 $\frac{1}{\sigma}$ 測算之，係隨 \sqrt{n} 而增加。

(4) 以下第九章第五節示知吾人，如以數值表示 σ ，則由抽樣觀察而得之 p' 值，可用以替代未知數 p 之真值 (true value)。

(5) 結果如下所述，該範圍中 p 之值，為 $p' \pm \sqrt{\frac{p'(1-p')}{n}}$ 。此式之含義，乃以 p' 為所得材料中之最或然值 (probable value)，對於 p' 變量之機率，可用第三表 (常態機率表) 求出，該表以標

準差爲單位，而標準差卽爲： $\sqrt{\frac{p'(1-p')}{n}}$ 也。須知此種數值只能用於特定之『範圍』，其中各個成員，可以有逐一調查之機會也。此種條件以及其他條件之重要性，可於下例表示之。

在理丁(Reading)地方，曾調查工人住宅六百零九家，據云其中有一百五十四家，每房間所住人數，在一與二人之間。 $n=609$, $p'n=154$, $p'=.253$, $\sqrt{p'q'/n}=.0176$ 。此等住宅所占比例，爲 $.253 \pm .0176$ 。

此處所謂之『範圍』，乃指全部住宅(約有12,000所)而言，此六百零九所，乃由全部住宅中選出。此羣類乃依地方住所姓名錄而規定，藉助所謂「重要住宅詳單」(“principal residents’)，將中等階級住宅，及大戶除開，並因得熟習地方情形者之助，將非住宅之房舍剔除。勞動階級調查之確實與否，要視地方住所姓名錄之詳確若何與剔除方法之是否適當而定。例如將卑屋陋室之貧民窟，屏出不計，則該『範圍』，必爲減縮多多；反之，若將住中等階級之街市誤計在內，苟非於調查進行中，業已發現錯誤，則該『範圍』亦將爲之擴大也。

在此情形之下，選擇時，在經修正過後之住所名錄，每二十擇一記下。如此選擇，自比用純粹之隨機抽樣法，較爲精確，參閱第四章第九節便知。一般保持隨機性之方法，乃擇『範圍』中之事物，而派定一自1至N之號碼目，然後用數字表查出，或選出

n 個號碼(註一一)。於此，務須注意，如不能得純粹隨機性之把握，則其他方法，以確有把握較隨機抽樣能得精確之結果者為佳。舉例言之，本章第九節所舉之實驗，隨機之決定，即非按頁抽選，或任意點出，所可達到；因各頁之各項，並非互相獨立也。各個事項，務須有被取用之同等機會，已成一規則，苟違背之，則結果所受影響，非常大也。

訪問（如每間房間所住之人數）之不確，固以避免為宜，但錯誤之來，果係由此，則人數之以多報少，或以少報多，機會各佔一半，結果不至有若何之影響也。

須知結果之正確與否，乃視 n ，即抽樣之多寡而定，與 N ，即『範圍』之全體個數，無關。『範圍』之廣狹，僅於 N 個事物為數衆多，散佈甚廣時對於問題有影響。蓋惟其如此，欲挨門逐戶，逐一檢點，以保持各個事物被選之機會均等，勢所難能；反或將與所查大部事物，性質根本不同之部分，因忽略其存在，而多所遺漏，斯亦於事難免。且也，若 p 小，則 pn 必略鉅，而 pn 為相對的小。假設 pn 小，與差誤曲線近似值（見本章第四節）相去必遠，而包括 $K\left(\frac{1-2p}{\sqrt{p(1-p)n}}\right)$ 之一項，乃不能略去，是則須用二項式 $(p+q)^n$ 之各該項，而不宜查積分表。據檢查若干實例之經驗所知，設 p 之值甚屬渺小，則具有該種性質者，往往一無所見。例如一有民宅一萬所之城市，其中居住人數衆多擁擠不堪者，只有三

十所，現抽查八百所，則查不到密聚雜居之住宅之機率為 $q^n p^0$ ；如以 $p = .003, q = .997, n = 800$ ，則其機率為.09。故以此樣本所得之材料為根據，作成之報告，對密聚雜居之情形，或鮮提及，除非言明並無該項證據存在也。但若 $p = .03, q^n p^0$ 僅達 $\frac{3}{10^{11}}$ ，則該現象必將得到證明。關於小數(small number)出現之機率，容再於本章十一節論之。

最後一點，吾人務須注意者，即當將應行計入之事物用名單抽點或其他方法決定時，必須儘量排除障礙，以免應行選入而未包羅於內也。設調查時房主拒絕報告，或寄售之貨物，內中一部，竟然失落，則該房屋或貨物情況之出乎常態，可想而知，故苟非將障礙排除，則該『範圍』之內，必有幾個部分，真象未能表明也。

第九節 抽樣方法舉例

根據一九一一年，英格蘭及威爾士之戶口普查，將一萬二千八百三十個自治區，各別列以號碼，然後按對數表之號次，選定二百五十區。下列一表，將各區樣本中之人口分配與全類之人口分佈狀況，作一比較（全類指戶口普查報告冊——第六二五八號，第四二八頁）

各區人數

	100以下	100- 200	200- 300	300- 400	400- 500	500- 1000	1000以上
250樣本中所估區數...	35	52	400	27	20	41	33
1000p'	140	208	168	108	80	164	132
1000 $\sqrt{\left(\frac{p'q'}{250}\right)}$	22	26	24	20	17	23	21
每1000中實在數.....	152	192	147	108	80	173	146

由此觀之——以上表第一行為例——人口在一百人以下者，二百五十區中佔三十五區

$$p' = \frac{35}{250} = .14$$

據此推測，每千區之數，必為1000之.14=140。

p'之標準差為 $\sqrt{\frac{p'(1-p')}{250}} = .022$ (見上節)，然則1000p'，即推測值，140，之標準差必為22。實際上英格蘭及威爾士各區人口在一百人以下者，為千分之一百五十二。推測與實數相差約半個標準差。(見統計學報 Statistical Journal 一九一二至一九一三年號第一八二頁)

(2)自三千八百七十八家公司股息率單中，擇出四百家，列表於下：

	股息百分率					
	三鎊以下	三鎊	四鎊	五鎊	六鎊	八鎊
樣本所取公司數.....	34	180	117	60	48	33
1000p'	85	270	292½	150	120	82½
1000 $\sqrt{\left(\frac{p'q'}{400}\right)}$	14	22	23	18	16	14
單中全數每1000中公司數	75	272	311	177	180	57

(見統計學報 Statistical Journal 第一九〇六年號第五二頁)

(3) 一地名索引共有地名三萬一千二百一十個，就中選擇五百個，並將緯度列爲一表。選擇時，爲確保隨機性(randomness)起見，乃將各行地名，號以數碼，然後依照數學表中之號碼抽選；用一英尺置在各行之旁，以與吋數相值（以各行第一位之經度第一單位數字爲準）者爲入選。此種精密方法，乃爲保持各項之獨立，所不可少。

南北緯度

	0°—10°	10°—20°	20°—30°	30°—40°	40°—50°	50°—60°	60°—70°	70°—80°	80°—90°
樣本所取地方數.....	22	56	104	103	93	112	9	1	0
1000p'.....	44	112	208	206	186	224	18	2	0
1000 $\sqrt{\left(\frac{p'q'}{500}\right)}$	9	14	18	18	17	19	6	?	?
全單中每1000中地方數	51	111	201	200	200	215	18	3.4	0.9

須知上表所包地方，凡在北緯(80°以北，南緯80°以南之地，均未選入。經用 $n=1000$ 之其他選擇方法，得知南北緯 80° 以上者佔千分之一云。

(4) 照一九一一年戶口普查之戶主表 (householder schedule)，依次將全卷每五十取出一個，後將家屬分爲幾類。

	有職業者		無職業者		—— 合計		
	男性		女性				
	二十歲以上	二十歲以下	十八歲以上	十八歲以下	十四歲以上	十四歲以下	
抽樣中所佔人數.....	538	112	310	74	386	718	2138
1000p'.....	251	52	145	35	181	336	1000
1000 $\sqrt{\left(\frac{p'q'}{213}\right)}$	9	5	8	4	8	10	—
戶口普查表冊之每千人中人數	258	55	144	33	185	325	1000

第十節 範圍實非為無限或選擇未能獨立之例

前於說明試驗時論及差誤常態律（見本章第四節，下同）曾假設每擲一次骰或抽一把牌，其實現之機率，完全相同（第二章第四節），而且假設每次試驗（Trial），不受已出現過者之影響。實則此項條件，完全見諸事實者，本極罕見，但吾人可用相同方法，證明差誤常態律，乃在較寬之假定下求得者也。

設一含有 N 個事物之範圍，在 N 個事物中，有某種屬性者有 pN 個，不含某種屬性者有 qN 個（如此則 $p+q=1$ ）。自 N 內，抽出 n 個來，抽選時，須用一種方法，使在該範圍內之每個事物，各有被抽之機遇。以 P_x 代表所抽事物中之 $(pn+x)$ 個能具有該種屬性者之機率（註一二）。舉例言之，假如所謂『範圍』，係一裝有白色球一百個連其他雜色球，共有一千個球之箱子。茲假如箱內事物業已攪和均勻，然後取出五十個球來，如是，則 $N=1000$ ， $p=\frac{1}{10}$ （白色為其屬性）， $n=50$ ， $pn=5$ ，而 P_x 即為抽取時白球佔有 $5+x$ 個之機率。

種種不同之抽取方法，共數有 ${}^N C_n$ 種。

抽得白球 $pn+x$ 個，雜色球 $(qn-x)$ 個之各種抽取方法，共有 $p_n C_{pn+x} \times q_n C_{qn-x}$ 種。

$$\text{故 } P_x = \frac{p_n C_{pn+x} \times q_n C_{qn-x}}{{}^N C_n}$$

$$= \frac{(pN)!(qN)!n!M!}{(pn+x)!(pM-x)!(qn-x)!(qM+x)!N!}$$

此處之 $M=N-n$ 。

援照斯德令氏 (Sterling) 定理, 應用到階乘 (Factorial)

上面, 將 $\frac{1}{pn}, \frac{1}{n}$ 及其他較小之數量省略。(見附錄三公式134)

$$P_0 = (pN)^{pN} (qN)^{qN} (n)^n M^M (pn)^{-pn} (pM)^{-pM} (qn)^{-qn} (qM)^{-qM} N^{-N} \cdot (2\pi)^{\frac{1}{2}-\frac{x}{2}}, e^0 \cdot \left(\frac{pNqNnM}{pnpMqnqMN} \right)^{\frac{1}{2}},$$

e 之指數為

$$pn + pM + qn + qM + N - pN - qN - n - M$$

$$= 0, \text{ 因 } p+q=1.$$

當將此諸指數集攏於一起時, 則得

$$P_0 = \left(\frac{N}{2\pi p q n M} \right)^{\frac{1}{2}} \dots \dots \dots (27)$$

$$\begin{aligned} \frac{P_x}{P_0} &= \frac{(pn)!(pM)!(qn)!(qM)!}{(pn+x)!(pM-x)!(qn-x)!(qM+x)!} \\ &= \frac{(pn)^{pn} (pM)^{pM} (qn)^{qn} (qM)^{qM} \cdot (2\pi)^0 \cdot e^0 \cdot (pn \cdot pM \cdot qn \cdot qM)^{\frac{1}{2}}}{(pn+x)^{pn+x+\frac{1}{2}} (pM-x)^{pM-x+\frac{1}{2}} (qn-x)^{qn-x+\frac{1}{2}} (qM+x)^{qM+x+\frac{1}{2}}} \\ \therefore \frac{P_x}{P_0} &= \left(1 + \frac{x}{pn}\right)^{pn+x+\frac{1}{2}} \cdot \left(1 - \frac{x}{pM}\right)^{pM-x+\frac{1}{2}} \cdot \left(1 - \frac{x}{qn}\right)^{qn-x+\frac{1}{2}} \\ &\quad \cdot \left(1 + \frac{x}{qM}\right)^{qM+x+\frac{1}{2}} \end{aligned}$$

$$\log P_x / P_0 = - \left(pn + x + \frac{1}{2} \right) \log \left(1 + \frac{x}{pn} \right) - \left(qn - x + \frac{1}{2} \right) \log \left(1 - \frac{x}{qn} \right)$$

$$\begin{aligned}
& -\left(pM-x+\frac{1}{2}\right)\log\left(1-\frac{x}{pM}\right)-\left(qM+x+\frac{1}{2}\right)\log\left(1+\frac{x}{qM}\right) \\
& = -\left(pn+x+\frac{1}{2}\right)\left(\frac{x}{pn}-\frac{x^2}{2p^2n^2}+\dots\right) \\
& \quad +\left(qn-x+\frac{1}{2}\right)\left(\frac{x}{qn}+\frac{x^2}{2q^2n^2}+\dots\right) \\
& \quad +\left(pM-x+\frac{1}{2}\right)\left(\frac{x}{pM}+\frac{x^2}{2p^2M^2}+\dots\right) \\
& \quad -\left(qM+x+\frac{1}{2}\right)\left(\frac{x}{qM}+\frac{x^2}{2q^2M^2}+\dots\right) \\
& = \frac{x}{2}\left(-\frac{1}{pn}+\frac{1}{qn}+\frac{1}{pM}+\frac{1}{qM}\right)-\frac{x^2}{2}\left(\frac{1}{pn}+\frac{1}{qn}+\frac{1}{pM}+\frac{1}{qM}\right) \\
& \quad +\frac{x^2}{4}\left(\frac{1}{p^2n^2}+\frac{1}{q^2n^2}+\frac{1}{p^2M^2}+\frac{1}{q^2M^2}\right)+\dots\dots\dots
\end{aligned}$$

n 當然小於 N , 且可認為小於 $\frac{1}{2}N$ (而不背乎通理 [generality])

故 n 必小於 M 。

設 pn 及 qn 皆相當的大, 則依 $\frac{1}{\sqrt{pn}}$ 之昇冪進行。

若以 $\frac{x^2}{n}$ 為可與整一 (unity) 相比敵之數量, 則如第四節例

二所言, $\frac{x}{n}$ 屬於 $\frac{1}{\sqrt{n}}$ 次 (order), $\frac{x^2}{n^2}$ 屬於 $\frac{1}{n}$ 次。

於是將 $\frac{1}{\sqrt{n}}$ 次或更高次各項省略, 則得

$$\log P^x P_0 - \frac{x^2}{2} \left(\frac{p+q}{pqn} + \frac{p+q}{pqM} \right) = -\frac{x(n+M)}{2pqnM} = -\frac{x^2 N}{2pqnM}$$

以 σ^2 代 $\frac{pqnM}{N}$ 。

$$P_0 = \frac{1}{\sigma\sqrt{2\pi}}$$

$$y = P_x = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}$$

此之謂差誤常態曲線 (normal curve of error), 而 σ (如上公式21所示) 即為其標準差。

$$\sigma^2 = pqn \cdot \frac{N-n}{N} = pqn \left(1 - \frac{n}{N}\right) \dots\dots\dots (28)$$

而且在第四節所言之情形下, 當較其值 (pqn) 為小, 但當 N 變為無限大時, 則當然漸與其值相等。

第十一節 小數律

自 $(p+q)^n$ 之展開式各項, 推算常態曲線時, 曾假定不只 n , 即 pqn , 亦甚大。但當 p 異常之小, 小至 pn 不復為大時, 則在此情形之下, q 幾等於 1, 此乃為有趣味之事實。

設 $u = pn$, 且為一極小之有限數。 $p = \frac{u}{n}$, $q = 1 - \frac{u}{n}$ 。

在 n 個獨立之實驗中 (independent experiments), 成功 r 次之機率, 為

$$P_x = \frac{n!}{(n-r)!r!} p^r q^{n-r}$$

$$= \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{r-1}{n}\right) \cdot \frac{u^r}{r!} \cdot \left(1 - \frac{u}{n}\right)^n q^{-r}$$

略去 $\frac{1}{n}$ ；則括弧中之 $r+1$ 個因數乘積，（係在 1 與 $1 - \frac{r(r-1)}{2n}$ 之間），可當作 1。

同時 $\left(1 - \frac{u}{n}\right)^n$ 亦漸等於 e^{-u} ，而 $q^{-r} = \left\{ \left(1 - \frac{u}{n}\right)^{-n} \right\}^{\frac{r}{n}}$ 漸等於 $(eu)^{\frac{r}{n}}$

且於 $\frac{r}{n}$ 達到 0 時，漸及於 1。

總之當 $\frac{1}{n}$ ， $\frac{r}{n}$ 及 $\frac{r^2}{n}$ 略去後，

$$P_r = e^{-u} \cdot \frac{u^r}{r!} \cdots \cdots \cdots (29)$$

$$\sigma^2 = pqn = u \left(1 - \frac{u}{n}\right) \therefore \sigma = \sqrt{u} \text{ (約略數)} \cdots \cdots \cdots (30)$$

$$k = \frac{q-p}{\sigma} = \left(1 - \frac{2u}{n}\right) / \sqrt{\left\{u \left(1 - \frac{u}{n}\right)\right\}} = \frac{1}{\sqrt{u}} \text{ (約略數)} \cdots \cdots (31)$$

整個曲線於是以 u 而決定，與 P 及 n 各別無關，因其平均數為 u ，標準差為 \sqrt{u} ，而其“ k ”為 $\frac{1}{\sqrt{u}}$ 也。由此觀之， p 與 n 之值，不易由觀察數量中而分別決定也。

當 u 為整數時，二項式之最大項為 $P_{pn} = \frac{e^{-u} u^u}{u!}$ (註一三)，

如是則，

$$P_r = P_{pn} \cdot \frac{u^{-r} \cdot r!}{r!} = \frac{P_{pn}}{\frac{r}{u} \left(\frac{r-1}{u} \right) \dots \left(\frac{r-r+1}{u} \right)},$$

於 $\frac{r}{u}$ 歷經各整數值時，則 p_r 以迅速之速率變小，例如假以 $u=6$ ，

而 $r=3u$ ，則 $P_{3u} = .00004$ 。

總而言之，各種觀察值，與其平均數，相差絕不至過大。小數之變動與現已討論之分配定律，頗能相應適合，吾人亦曾喚起注意，關於例證，則包吉維 (Bortkiewicz *Das Gesetz der kleinen Zahlen*, 1898) 及毛達拉 (Mortara: *Annali di Statistica*, Serie V, vol. 4, 1912) 二氏，均曾舉出。更有趣味者，由此理論引起所謂小數之恆性，(permanence of small numbers) 此應與以注意。在多數事物中，如有少數現某種特色者，則據一般經驗所知，此種小數，鮮有超過限度；而完全消滅，亦極罕見。舉凡意外事件，火災，及新聞紙用以充實篇幅之奇事與巧合莫不皆然。各種職業之專家，自專醫某種無名耳疾之醫生起，至古玩商人止，均恃此小數之恆性，而維其生活者也。

舉例言之；設每年由於各種原因之死亡人數為 500,000，而自一八七五年至一八九四年因脾臟扶斯致死之人數如下：

5, 4, 10, 14, 12, 18, 9, 15, 8, 18, 11, 11, 11, 12, 7,
4, 3, 6, 7, 10,

平均數 $9.75 = pn = u_0$ $e^{-u} = .00005842_0$

r	$e - u \cdot ur / r!$		預測數	實在數
0	.00006	$\times 20$	$= .001$	0
1 至 4	.0343		$= .7$	3
5 ,, 9	.4564		$= 9.1$	6
10 ,, 14	.4403		$= 8.8$	8
15 ,, 19	.0683		$= 1.4$	3
20 ,,	小		—	0

(註一) 見附錄二, 公式133。

(註二) 下文為簡便起見, 假定 pn 為整數, 故 P_0 為最大項; 而 n 既大, 則 $\frac{1}{n}$ 之乘方乃可於最後刪略之, 但與算式證明無干。(註三) 參閱 Edgeworth 氏所著 大英百科全書 第十二部之「機率」一文, 自第三百九十一頁起。

(註四) 附錄四。

$$(註五) \quad \text{因 } m^2t+1 = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2t+1 e^{-\frac{x^2}{2\sigma^2}} dx = \int_{-\infty}^{\infty} \phi(x) dx,$$

$$= \int_0^{\infty} \phi(x) dx + \int_{-\infty}^0 \phi(x) dx = \int_0^{\infty} \phi(x) dx - \int_0^{\infty} \phi(x') dx', (x = -x = 0)$$

(註六) 平均差及機誤定義見第一編第六章第一節。所謂平均差, 乃組成羣類之各項數量, 對中心數 (普通係算術平均數) 之差額, 不計符號之正負, 而得之平均。

(註七) 注意, $p \sim p'$ 為比例 (proportion) 之離差。實際結果, 離中差應為 $pn \sim p'n$, 故 $z = \frac{pn - p'n}{\sqrt{p(1-p)n}}$ 。如是, 則 \sqrt{n} 增加, 機率亦隨之增加。(註八) 例如 $r=13.5$, 及 14.5 , 則 $F(z) = .4994$ 及 $.4986$ 。此差額, $.0008 \times 300$, 即為 $r=14$ 時之希望數 (Expected number)。

(註九) 前欄最靠近之全數。

(註一〇) 必要之修正, 當 $\frac{n}{N}$ 不能刪略時, 本章第十節論及之。

- (註一一) 例如, $N=10,000$, $n=500$, 則可取七位數字表按頁取其各數目之末四位數字, 以至取得自 0 至 10,001 間之數, 五百個為止。然後再查明各號碼所指之事物, 而加以調查。此法於調查各區人數時用之。見下節。
- (註一二) 例如從一副五十二張之牌, 每抽取十三張為一把, 則每把中抽到三個 A 點之機率, 為 $P_2 = {}_4C_3 \times {}_{48}C_{10} \div {}_{52}C_{13}$, 由此看來 $N=52$, $n=13$, $p = \frac{1}{13}$, $x=2$, P_x 乃約略等於 .041。
- (註一三) 如 u 等於 10 時, 此式與 $\frac{1}{\sqrt{2\pi u}}$ 相差, 不過百分之一。

第三章 大數律(普遍的差誤律)

在上文中，吾人既已以差誤常態曲線，作為二項式 $(q+p)^n$ 之極限，並已表明其整數在 p 有確定意義時之應用。但現有同一之方程式 $y = \frac{1}{\sigma\sqrt{\pi}2} e^{-\frac{x^2}{2\sigma^2}}$ ，乃為十分廣泛之假說下之得數，本章之主要目的，即欲對此種種假說而加以發揮也。

在未論及通則(general law)之先，有若干重要之命題(proposition)不可不加以考慮，此種命題為何，即由一大羣類或多數羣類中所選衆多量數之平均數或總和數之標準差，與由各原來量數而算得之標準差，二者相互間之關係是也。此種命題(本章第一節)，僅受機率基本定理之支配，與極限值或略去微量(small quantities)方法無關。

第一節 平均及總和之標準差及均立方差誤

(1)以 $u_1, u_2, \dots, u_i, \dots, u_{m_1}$ 為一類數羣類之 m_1 個量數，而以 \bar{u} 為其平均數， σ_u 為其標準差。

假以 $u_i = \bar{u} + u'_i$ 。

則 $m_1\bar{u} \doteq \sum u_i \therefore \sum u'_i = 0$

$$\begin{aligned}\text{而 } m_1\sigma_u^2 &= Su_t'^2 = S(u_t - \bar{u})^2 = Su_t^2 - 2\bar{u}Su_t + m_1\bar{u}^2 \\ &= Su_t^2 - 2\bar{u} \cdot m_1\bar{u} + m_1\bar{u}^2\end{aligned}$$

$$\therefore Su_t^2 = m_1(\sigma_u^2 + \bar{u}^2) \dots \dots \dots (32)$$

(2) 設以 $v_1, v_2, \dots, v_t, \dots, v_{m_2}$ 爲第二頻數曲線之 m_2 個量數其平均線爲 \bar{v} , 標準差爲 σ_v 。

從每類中任意各抽出一物, 譬如抽得者爲 u_s 及 v_t 。今請就 $u_s + v_t$ 之所有或有數值 (每次抽選兩個, 各次抽選彼此須完全獨立不致互受牽掣) 所組成之羣類中, 求其平均數及標準差。茲設以 H_2 爲此一羣類之平均數, 以 S_2 爲其標準差。

吾人假設各次抽選彼此互不相涉, 抽選次數乃有無限之多, 則在此一新羣類中, $u_s + v_t$ 之 $m_1 \times m_2$ 個或有數值, 必以均等頻數而出現。

然則 $H_2 \times m_1 \times m_2$

$$\begin{aligned}&= (u_1 + v_1) + \dots + (u_1 + v_t) + \dots + (u_1 + v_{m_2}) \\ &\quad + (u_2 + v_1) + \dots + (u_2 + v_t) + \dots + (u_2 + v_{m_2}) \dots \dots \dots \\ &\quad \dots \dots \dots \text{共有 } m_1 \text{ 行, 每行各有 } m_2 \text{ 項。} \\ &\quad + (u_{m_1} + v_1) + \dots + (u_{m_1} + v_t) + \dots + (u_{m_1} + v_{m_2}) \\ &= m_2 \cdot Su_t + m_1 \cdot Sv_t = m_2 m_1 \bar{u} + m_1 m_2 \bar{v}\end{aligned}$$

$$\therefore H_2 = \bar{u} + \bar{v} \dots \dots \dots (33)$$

$$m_1 m_2 (S_2^2 + H_2^2) = S(u_s + v_t)^2$$

$$\begin{aligned}
 &= (u_1 + v_1)^2 + \dots + (u_1 + v_{m_2})^2 + (u_2 + v_1)^2 + \dots + (u_2 + v_{m_2})^2 + \dots \\
 &= m_2 S u^2 + m_1 S v^2 + 2 \cdot S u_1 \cdot S v_1 \\
 &= m_2 m_1 (\sigma_u^2 + \bar{u}^2) + m_1 m_2 (\sigma_v^2 + \bar{v}^2) + 2 m_1 \bar{u} \cdot m_2 \bar{v} \\
 &= m_1 m_2 \{ (\sigma_u^2 + \sigma_v^2) + (\bar{u} + \bar{v})^2 \}
 \end{aligned}$$

$$\therefore S_2^2 = \sigma_u^2 + \sigma_v^2 \dots \dots \dots (34)$$

(3) 如組成之羣類, 不用總和, 而用 $\overline{u_s - v_t}$ 之差, 則依同樣步驟, 可得 $H_2 = \bar{u} - \bar{v}$, 但 $S_2^2 = \sigma_u^2 + \sigma_v^2$, 與前相同。

(4) 復次以從三類中抽出之和 (或差), 另成一類, 其三類之平均數為 $\bar{u}, \bar{v}, \bar{w}$, 標準差為 $\sigma_u, \sigma_v, \sigma_w$, 而和或差為 H_3, S_3 。

$$\text{則 } H_3 = \bar{u} \pm \bar{v} \pm \bar{w} \dots \dots \dots (35)$$

其理顯而易見。

至於 S_3 之求得, 可設先使 u_3 與 v_t , 先行結合, 然後再加上 W , 並援用上已證明兩次之公式。

$$S_3^2 = S_2^2 + \sigma_w^2 = \sigma_u^2 + \sigma_v^2 + \sigma_w^2 \dots \dots \dots (36)$$

此公式, 用歸納方法, 可引申於任何多少羣類也。

(5) 有一重要事例, 不可不知者, 乃此數類之標準差, 完全相等, 如此, 則 $\sigma_u = \sigma_v = \dots \dots \sigma$ 。

如此總和乃由此等羣類 n 個所組成, 而其標準差為 S , 則

$$S^2 = S_n^2 = \sigma^2 + \sigma^2 + \dots \dots \text{直到 } n \text{ 項為止} = n\sigma^2$$

$$\therefore S = \sigma \cdot \sqrt{n} \dots \dots \dots (37)$$

(6) 復次設吾人不用 n 個量數之和，而用其平均數，則在此組合之羣類中，各項必以 n 除之，然則此平均數組成之羣類標準差，必為取用總和法所得羣類之標準差再被 n 除。

$$\therefore \sigma_a = \frac{S}{n} = \frac{\sigma}{\sqrt{n}} \dots \dots \dots (38)$$

(7) 最後，如此平均數係從 n 個項目而來，而均由此同一（無限大）之原始羣類獨立抽出，使從 n 個項目中抽取任何一個之機遇，不至因以前所抽出者而受影響，則仍為 $\sigma_a = \frac{\sigma}{\sqrt{n}}$ 。

下文假設各原始量數，均係由各羣類之平均數而來，且 $0 = \bar{u} = \bar{v} = \dots \dots$ ，且 $0 = Su = Sv = \dots \dots$ 。

$$\begin{aligned} u_s \text{ 與 } v_t \text{ 相加之均立方 (mean cube). 爲 } & \frac{1}{m_1 m_2} S (u_s + v_t)^3 \\ = \frac{1}{m_1 m_2} \{ m_2 S u^3 + m_1 S v^3 + 3 \cdot S v S u^2 + 3 \cdot S u S v^2 \} & = \frac{1}{m_1} S u^3 + \frac{1}{m_2} S v^3 \\ = u \mu_3 + v \mu_3, \text{ 即爲各類以平均數爲中心而得之第三動差之總和。} \end{aligned}$$

是故， m_3 —— n 個項目（均自一類選來）總和之第三動差——等於 $n \mu_3$ ，（ μ_3 爲該類之第三動差），如用總和則

$$K = \frac{m_3}{S^3} = \frac{n \mu_3}{n^2 \sigma^3} = \frac{K'}{\sqrt{n}} \dots \dots \dots (39)$$

（ K' 爲該類之“ K ”值，其定義已見公式(8)）

不論， n 個項目之平均數，抑或爲其總和， K 值終爲相同。

第二節 差誤曲線之發生

現研究到一種分析工作，由此分析即將入於應用差誤曲線之途。下舉一極簡單之例，可為本章前後兩部之樞紐。

設有多數之原始羣類，可以差誤常態曲線代表之，吾人可證明此等羣類之總和與平均，均屬於常態。

因如以 $x_t = u_t + v_t$ 則 u_t, v_t 各部數值同時實現之機率為

$$\begin{aligned} & \frac{1}{\sigma_u \sqrt{2\pi}} e^{-\frac{u^2 t}{2\sigma_u^2}} \times \frac{1}{\sigma_v \sqrt{2\pi}} e^{-\frac{v^2 t}{2\sigma_v^2}} \\ &= \frac{1}{2\pi \sigma_u \sigma_v} e^{-\frac{1}{2} \left(\frac{u^2 t}{\sigma_u^2} + \frac{(v_t + u_t)^2}{\sigma_v^2} \right)}. \end{aligned}$$

將上式對 u 值積分之，則 $x_t (+\delta x)$ 之整個機率等於

$$\begin{aligned} & \frac{1}{2\pi \sigma_u \sigma_v} \int_{-\infty}^{\infty} e^{-\frac{\sigma_u^2 + \sigma_v^2}{2\sigma_u \sigma_v^2} \left(u - \frac{\sigma_u^2 x_t}{\sigma_u^2 + \sigma_v^2} \right)^2} e^{-\frac{x_t^2}{2(\sigma_u^2 + \sigma_v^2)}} du \cdot \delta x \\ &= \left(\text{以 } u' \text{ 代 } u - \frac{\sigma_u^2 x_t}{\sigma_u^2 + \sigma_v^2} \right), \\ & \frac{1}{2\pi \sigma_u \sigma_v} \cdot e^{-\frac{x_t^2}{2(\sigma_u^2 + \sigma_v^2)}} \cdot \delta x \cdot \int_{-\infty}^{\infty} e^{-\frac{\sigma_u^2 + \sigma_v^2}{2\sigma_u^2 \sigma_v^2} du'^2}, \text{ 又依第二} \end{aligned}$$

章第四節公式 20 至 21

$$= \frac{1}{S_2 \sqrt{2\pi}} e^{-\frac{x_t^2}{2S_2^2} \delta x}, \text{ 式中 } S_2^2 = \sigma_u^2 + \sigma_v^2.$$

故 x 值之機率為 $\frac{1}{S_2 \sqrt{2\pi}} e^{-\frac{x^2}{2S_2^2}} \dots \dots \dots (40)$

以上如用歸納方法，得出通則甚為易易。從一以 σ 為標準差之常態曲線中，舉行 n 個獨立之抽選，則自所抽得之總和或平

均中，求得 x 之機率，分別為

$$\frac{1}{\sigma\sqrt{2\pi n}} e^{-\frac{x^2}{2n\sigma^2}} \text{ 及 } \frac{\sqrt{n}}{\sigma\sqrt{2\pi}} e^{-\frac{nx^2}{2\sigma^2}} \dots\dots\dots (41)$$

此原始曲線非為常態時，亦能求得上述得數，惟此得數只係一第一次近似值 (First appropriation)，然亦滿足此分析工作中所得之某種條件。此種得數，極為重要，其兩個證法，當另闢下節專論之。

第三節 用多項式定理證明之

在此證明中，吾人須引申上節，(公式33至39)所用方法得出之動差，無論方次 (order) 多少，皆與由具有相當標準差之差誤常態曲線所得者相同。

設有 n 個元始羣類，各有 m_1, m_2, \dots, m_n 個可量之事物；在任何一類中，譬如第 t 類，以 \bar{u}_t 為其平均數， σ_t 為其標準差， $\epsilon_{t2}, \epsilon_{t3}, \dots$ 為其以平均數為中心之動差，並設各項為 $\bar{u}_t + \epsilon_{t1}, \bar{u}_t + \epsilon_{t2}, \dots, \bar{u}_t + \epsilon_{ts}, \dots$

則， $\epsilon_{t1} + \epsilon_{t2} + \dots + \epsilon_{ts}, \dots = 0$ 。

從每一羣類任意抽選一項，而將抽得之 n 項相加；假設各自各類抽取一個，各個互為獨立，不相關涉，並設自一羣類中抽得某一量數之機率，亦不受以前所行抽選之影響。

在第 S 次抽選時，其相加之和為 $H + E_s$ ，(此處之 $H = \bar{u}_1$

+ $\bar{u}_2 + \dots + \bar{u}_n$, 而 $E_s = {}_1u_s + {}_2u_s + \dots + {}_nu_s$ 。

以 S, M_2, M_3, \dots 爲 E_s 之頻數曲線之標準差及動差, 所謂 E_s 之頻數曲線者, 換言之, 卽總和數之頻數曲線也。

$M_2 = S^2 = ({}_1u_s + {}_2u_s + \dots + {}_nu_s)$ 之諸可能值平均數。

此種數值有 $m_1 \times m_2 \times \dots \times m_n = N$ 個。於是將第一節前半節之程序, 歸納起來,

$$M_2 = S^2 = \frac{1}{N} \left\{ \frac{N}{m_1} S_1 u^3 + \dots + \frac{N}{m_2} S_2 u^3 + \dots \right\} + \frac{3}{N} \left\{ \frac{N}{m_1 m_2} S_1 u \cdot S_2 u + \dots \right\}$$

$$= \sigma_1^2 + \sigma_2^2 + \dots + \sigma_n^2, \text{ 因爲 } 0 = S_1 u = S_2 u = \dots \text{ 依同理,}$$

$$M_3 = \frac{1}{N} S({}_1u_s + {}_2u_s + \dots + {}_nu_s)^3$$

$$= \frac{1}{N} \left\{ \frac{N}{m_1} S_1 u^3 + \frac{N}{m_2} S_2 u^3 + \dots \right\} + \frac{3}{N} \left\{ \frac{N}{m_1 m_2} S_1 u S_2 u^2 + \dots \right\}$$

$$= {}_1\mu_3 + {}_2\mu_3 + \dots + {}_n\mu_3, \text{ 因爲 } 0 = S_1 u \dots \dots \dots (42)$$

$$\text{且 } M_4 = \frac{1}{N} \left\{ \frac{N}{m_1} \cdot S_1 u^4 + \dots \right\} + \frac{4}{N} \left\{ \frac{N}{m_1 m_2} S_1 u^3 S_2 u + \dots \right\}$$

$$+ \frac{12}{N} \left\{ \frac{N}{m_1 m_2 m_3} S_1 u^2 S_2 u S_3 u + \dots \right\}$$

$$+ \frac{6}{N} \left\{ \frac{N}{m_1 m_2} \cdot S_1 u^2 \cdot S_2 u^2 + \dots \right\}$$

$$+ \frac{24}{N} \left\{ \frac{N}{m_1 m_2 m_3 m_4} S_1 u S_2 u S_3 u S_4 u + \dots \right\}$$

$$= {}_1\mu_4 + {}_2\mu_4 + \dots + {}_n\mu_4 + 6(\sigma_1^2 \sigma_2^2 + \sigma_1^2 \sigma_3^2 + \dots + \dots) \dots \dots (43)$$

$$\therefore M_4 - 3S^4 = S({}_1\mu_4 + 6(\sigma_1^2 \sigma_2^2 + \dots) - 3(\sigma_1^2 + \sigma_2^2 + \dots)^2) = S(\mu_4 - 3\sigma_4^4)$$

若幾類元始曲線之標準差相同，動差亦一樣，如是， $\sigma_1 = \sigma_2 = \dots = \sigma$, $1\mu_3 = 2\mu_3 = \dots = \mu_3$ ，其他以此類推，則 S (總和) $= \sigma\sqrt{n}$ ，
 σ_a (平均數) $= \frac{\sigma}{\sqrt{n}}$ (44)

$$K \text{ (總和或平均)} = \frac{M_3}{S^3} = \frac{n\mu_3}{n^3\sigma^3} = \frac{K'}{\sqrt{n}} \text{ (45)}$$

(K' 為各元始曲線中之“ K ”)。

$$K_2 = \frac{M_4}{S^4} - 3 = \frac{n(\mu_4 - 3\sigma^4)}{n^2\sigma^4} = \frac{1}{n} \left(\frac{\mu_4}{\sigma^4} - 3 \right) \\ = \frac{\text{各元始曲線中之“}K\text{”值之和}}{n} \text{ (46)}$$

故當 \sqrt{n} 變大時， K 漸近於零，而於 $\frac{1}{n}$ 可以略去時， K_2 亦可作為零。

欲求更高級動差，須以數值表示 m_t 而以 t 為任何之整數；換言之，即為 $(1u + 2u + \dots + nu)^t$ 之平均值，至 $(1u + 2u + nu)^t$ 式 (依多項式定理 (註一))，在 $n_1 + n_2 + \dots = t$ 條件之下，所有各可能值加在一起時，乃即

$$S \frac{t!}{n_1!n_2! \dots} \cdot 1u^{n_1} \cdot 2u^{n_2} \dots$$

之平均值。

第一就 t 為偶數時之例言之。

$$M_{2t} = \frac{(2t)!}{n_1!n_2! \dots} \cdot 1u^{n_1} \cdot 2u^{n_2} \dots \text{各項平均數之和}$$

當 $n_1 + n_2 + \dots = 2t$

$$M_{2t} = \frac{(2t)!}{n_1! n_2!} \cdot 1^{n_1} n_1 \text{ 平均值} \times 2^{n_2} n_2 \text{ 平均值} \times \dots \text{各項之和}$$

因自各元素曲線中之抽取係獨立的，故每一 $1u$ 與每一 $2u \dots$ 之實現，各有均等之次數也。

$$\therefore M_{2t} = \frac{(2t)!}{n_1! n_2! \dots} \cdot 1^{\mu_{n_1}} \cdot 2^{\mu_{n_2}} \dots$$

其次再就一種情形——各元素曲線之標準差及動差均相等——而論；在此情形之下， $1\mu_{n_1} = 2\mu_{n_1} = \dots = \mu_{n_1}$ ，以此類推。

在任一抽出項中，設有 F 個因子 $1\mu_{n_1}, 2\mu_{n_1}, \dots$ 則此一項，以各種形式而出現，有 nC_f 次之多：

$$1\mu_{n_1} \times 2\mu_{n_2} \times 3\mu_{n_3}, \dots, 1\mu_{n_2} \times 2\mu_{n_1} \times 3\mu_{n_3}, \dots, 1\mu_{n_3} \times 2\mu_{n_2} \times 3\mu_{n_1}, \dots,$$

式中之每一項均與 $\mu_{n_1} \times \mu_{n_2} \times \mu_{n_3} \dots$ 相吻合。

故 $M_{2t} = nC_f \frac{2t!}{n_1! n_2! \dots} \mu_{n_1} \times \mu_{n_2} \times \dots$ 各項之和，而此各項均適合 $n_1 + n_2 + \dots = 2t$ 之條件。

由此，以 $S^2 = n\sigma^2$, $S^{2t} = n^t$ ，且 $nC_f = n(n-1)\dots(n-f+1)/f!$

$$\therefore \frac{M_{2t}}{S^{2t}} = \frac{(2t)!}{n_1! n_2! \dots} \cdot \frac{\left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \dots \left(1 - \frac{f-1}{n}\right)}{f!} \cdot \frac{n^f}{n^t} \cdot \frac{u_{n_1}}{\sigma^{n_1}} \cdot \frac{\mu_{n_2}}{\sigma^{n_2}}$$

\dots 各項之和。

再次，僅就一種元素曲線而論，設此一曲線必須滿足下列條件：無論 p 爲何值， $\frac{\mu^p}{\sigma^p}$ 終爲有限，換言之，即 $\left(\frac{u}{\sigma}\right)^p$ 之平均值爲有

限，或者該曲線之有效全距，可與其標準差相比較。吾人所欲討論者，諸可能值中何者為有限？及何者屬於 $\frac{1}{n}$ 次(order)或更高次？

因 μ_1 等於0， n_1, n_2, \dots 之每一項在不為零之各項，其值必不下於2；故因 f 項 n_1, n_2, \dots 等項之和等於 $2t$ ，則各項之最大可能數乃為 t ，而且 $f > t$ 。

如 $f < t$ ，則分數 $\frac{n^f}{n^t}$ 屬於 $\frac{1}{n}$ 以上之次。

如 $f = t$ ，則 $2 = n_1 = n_2, \dots$ ，且當略去 $\frac{1}{n}$ 時，則得僅有之項

$$\frac{(2t)!}{2^t} \cdot \frac{1}{f!} \cdot \frac{n^f}{n^t} \cdot \left(\frac{\mu_2}{\sigma^2}\right)^f = \frac{2t!}{2^t t!},$$

因為 $\mu_2 = \sigma^2$ 。且 $\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right) \dots \left(1 - \frac{f-1}{n}\right)$ 在1與 $1 - \frac{f(f-1)}{2n}$

之間。當略去含有 $\frac{1}{n}$ 之各項時，

$$\text{是以 } M_{2t} = S^{2t} \cdot \frac{(2t)!}{2^t t!} = 1 \cdot 3 \cdot 5 \cdot (2t-1) S^{2t} \dots \dots \dots (47)$$

依同理， $\frac{M_{2t+1}}{S^{2t+1}} = \frac{(2t+1)!}{n_1! n_2! \dots} \cdot \frac{1}{f!} \cdot \frac{n^f}{n^{t+\frac{1}{2}}} \left(\frac{\mu_{n1}}{\sigma^{n1}}\right) \left(\frac{\mu_{n2}}{\sigma}\right) \dots \dots$ 等項

之和。

由此觀之，每一項之分母，無不含有 n 次方，若欲求其最大項，只須證得在 n_1, n_2, \dots 等數量之中，有一個等於3，其他各值則均等於2；如是， $2t+1 = n_1 + n_2 + \dots = 2(f-1) + 3 = 2f+1$ ， $f=t$ ，如依次以3代替 n_1, n_2, \dots ，則可得 f 個相等項。

$$\begin{aligned} \text{然則 } \frac{M_{2t+1}}{S^{2t+1}} &= f \times \frac{(2t+1)!}{2^{t-1}3!} \cdot \frac{1}{t! \sqrt{n}} \cdot \frac{\mu_3^{2t-1} \cdot \mu_3}{\sigma^{2t+1}} \\ &= \frac{t}{3} \cdot 1 \cdot 3 \cdot 5 \cdots \frac{1}{2t+1} \cdot \frac{\mu_3}{\sqrt{n} \sigma^3} \cdots \cdots (48) \end{aligned}$$

如將包含 $\frac{1}{\sqrt{n}}$ 之項略去,

$$\therefore M_{2t+1} = 0,$$

$$\text{且 } M_{2t+1} = \frac{t}{3} \cdot 1 \cdot 2 \cdot 3 \cdots \frac{1}{2t+1} \cdot M_3 \cdot S^{2t-2} \cdots (49)$$

因如保留含有 $\frac{1}{\sqrt{n}}$ 之項, 而將含 $\frac{1}{n}$ 者刪除, $\frac{M_3}{S^3} = \frac{\mu_3}{\sqrt{n} \cdot \sigma^3}$,

如此得來之動差, 與 (參閱公式 23, 及附錄六) 由曲線 (已

$$\text{將含 } \frac{1}{\sqrt{n}} \text{ 項者刪除) } y = \frac{1}{S\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2s^2}},$$

得來者, 毫釐不差。再如將有 $\frac{1}{\sqrt{n}}$ 之項保留, 而刪除其 $\frac{1}{n}$ 項, 則

$$\text{可由曲線 } y = \frac{1}{S\sqrt{2\pi}} \left[1 - \frac{k}{2} \left(\frac{x}{s} - \frac{1}{3} \cdot \frac{x^3}{s^3} \right) \right] e^{-\frac{x^2}{2s^2}}$$

得來, 惟式中所用之 $K = \frac{M_3}{S^3}$ 。

是故, 如以標準差相合, 動差亦一致, 即認為曲線亦脗合。則上述之方程式, 便為該頻數曲線之第一, 二近似值。

第四節 愛基華斯氏之證明

愛基華斯教授(Professor Edgeworth 著 “Law of Error”, Camb. Phil. Trans., Vol. XX., Part I., 1904)所作證明,較爲簡捷普通,惟含有頗深奧之數學概念,此乃以上所分析(主要亦係以愛氏理論爲根據)力圖避免者也。

愛氏原著之公式,可求無數次之近似值,茲爲簡便起見,下列僅就第一,二兩次,闡明之。

援用上述之標號及條件,

設 $E_s = {}_1u_s + {}_2u_s + \dots + {}_n u_s$ 。

設 a 爲任一固定小數量,僅用以抽選同元(dimensions)之各項,

則 $e^{aE_s} = e^{a \cdot {}_1u_s} \cdot e^{a \cdot {}_2u_s} \cdot e^{a \cdot {}_2s} \cdot {}_3u_s \dots$ 相一致。

$e^{a \cdot {}_1u_s}$ 之平均值,換言之,即

$$\left(1 + a \cdot {}_1u + \frac{a^2}{2} \cdot {}_1u^2 + \frac{a^3}{3!} \cdot {}_1u^3 + \dots\right) \text{之平均值} = 1 + a \cdot {}_1\mu_1 + \frac{a^2}{2} \cdot {}_1\mu_2 + \frac{a^3}{3!} \cdot {}_1\mu_3 + \dots \quad (\text{此處 } {}_1\mu_1 = 0)$$

從各元素曲線之抽選既係獨立,則

$e^{a \cdot {}_1u_s} \times e^{a \cdot {}_2u_s} \times \dots$ 乘積之平均數 = 各該平均數之乘積。

$\therefore 1 + a \cdot M_1 + \frac{a^2}{2} M_2 + \frac{a^3}{3!} M_3 + \dots =$ 因子 n 個之乘積,例如

以下之因子: $\left(1 + \frac{a^2}{2} \cdot {}_1\mu_2 + \frac{a^3}{3!} \cdot {}_1\mu_3 + \dots\right)$

$$\begin{aligned}
& \therefore \log \left(1 - aM_1 + \frac{a^2}{2} M_2 + \dots \right) \\
&= \sum_{t=1}^{t=n} \log \left(1 + \frac{a^2}{2} \cdot t\mu_2 + \frac{a^3}{3!} \cdot t\mu_3 + \dots \right) \\
&= \frac{a^2}{2} St\mu_2 + \frac{a^3}{6} \cdot St\mu_3 + \frac{a^4}{24} St\mu_4 + \dots - \frac{1}{2} \left(\frac{a^2}{2} S(t\mu_2) + \dots \right)^2 \\
&\therefore 1 + aM_1 + \frac{a^2}{2} M_2 + \dots \\
&= e^{\frac{a^2}{2} \cdot St\mu_2} \cdot e^{\frac{a^3}{6} St\mu_3} \cdot e^{\frac{a^4}{24} (St\mu_4 - 3S(t\mu_2)^2)} \\
&= \left(1 + \frac{a^2}{2} St\mu_2 + \dots + \frac{1}{p!} \left(\frac{a^2}{2} St\mu_2 \right)^p + \dots \right) \cdot \left(1 + \frac{a^3}{6} St\mu_3 + \dots \right) \\
&\quad \cdot \left(1 + \frac{a^4}{24} \left\{ St\mu_4 - 3S(t\mu_2)^2 \right\} + \dots \right) \dots
\end{aligned}$$

以 a^4 爲準，令各係數相等。

$$M_1 = 0$$

$S^2 = M_2 = St\mu_2 = S\sigma_i^2 = n\sigma^2$ ，設 σ^2 爲 $\sigma_1^2, \sigma_2^2, \dots$ 之平均

$M_3 = St\mu_3 = n\mu_3$ ，設 μ_3 爲 $\mu_{13}, \mu_{23}, \dots$ 之平均

$$\frac{M_4}{24} = \frac{1}{2} \cdot \left(\frac{1}{2} \cdot St\mu_2 \right) + \frac{1}{24} \left((St\mu_4 - 3S(t\mu_2)^2) \right)$$

$$\therefore M_4 - 3S^2 = S(t\mu_4 - 3\sigma_i^4)$$

$$k_2 = \frac{M_4}{S^2} - 3 = \frac{1}{n^2 \sigma^4} S \left(\frac{t\mu_4}{\sigma_i^4} - 3 \right) \sigma_i^4 = \frac{1}{n} k_2'$$

$\left(\frac{t\mu_4}{\sigma_i^4} - 3 \right) \left(\frac{\sigma_i}{\sigma} \right)^4$ 之平均。

$$k = \frac{M_3}{s^3} = \frac{n\mu_3}{n^{\frac{3}{2}}\sigma^3} = \frac{k'}{\sqrt{n}}, \left(k' = \frac{\mu_3}{\sigma^3}\right).$$

$$\begin{aligned} &\therefore 1 + \frac{a^2}{2}S^2 + \frac{a^3}{3!}M_3 + \dots + \frac{a^t}{t!}M_t + \dots \\ &= e^{\frac{1}{2}a^2s^2} \left(1 + \frac{1}{6}a^3 \cdot s^3 \cdot \frac{1}{\sqrt{n}}k' + \dots\right) \left(1 + \frac{1}{24}a^4s^4 \cdot \frac{1}{n}k_2' + \dots\right) \dots \end{aligned}$$

上列各方程式之右方， a 之指數，均等於 μ 之附數，或乘羈之附數之和，或各 u 值之乘積。

現假設所有各元素曲線中，不拘 p 值爲何， $\frac{\mu_p^p}{\sigma^p}$ 終爲有限，其結果， a^p, s^p 之係數，必包有因子 $\frac{1}{n^{\frac{1}{2}(p-2)}}$ ，由以上推算以 a^4 爲準，所得結果可知。

略去 $\frac{1}{n^{\frac{1}{2}}}$ 及其以上之乘羈

$$1 + \frac{a^2}{2}s^2 + \dots + \frac{a^t}{t}M_t + \dots = 1 + \frac{1}{2}a^2s^2 + \dots + \frac{1}{t!}a^{2t} \cdot \frac{s^{2t}}{2^t} + \dots$$

\therefore 每奇數級動差， $M_{2t+1} = 0$

$$\text{至偶數級動差， } M_{2t} = (2t)! \frac{s^{2t}}{t! 2^t} = 1 \cdot 3 \cdot \dots \cdot (2t-1) \cdot s^{2t} \dots \quad (50)$$

如在差誤常態曲線（公式23）中。

現存其 $\frac{1}{\sqrt{n}}$ ，而刪其 $\frac{1}{n}$ 。

M_{2t} 如前述。

$$\frac{M_{2t+1}}{(2t+1)!} = \frac{1}{(t-1)! 2^{t-1}} s^{2t-2} \cdot \frac{1}{6} \cdot s^3 \cdot \frac{k'}{\sqrt{n}},$$

$$M_{2t+1} = \frac{(2t+1)!}{(t-1)! 2^{t-1}} \cdot \frac{s^{2t-2}}{6} \cdot M_3 = \frac{t}{3} \cdot 1 \cdot 3 \cdot 5 \cdots (2t+1) M_3 \cdot$$

s^{2t-2} 換言之，即下列曲線之第 $(2t+1)$ 級動差

$$\frac{1}{s\sqrt{2\pi}} \left\{ 1 - \frac{k}{2} \left(\frac{x}{s} - \frac{1}{3} \frac{x^3}{s^3} \right) \right\} e^{-\frac{x^2}{2s^2}} \cdots \cdots \cdots (51)$$

(參閱附錄六)

由此觀之，藉用動差相等之測驗，從 n 種抽選之結果，將其總和或平均組成之頻數曲線，在已知條件之下，如取其第一近似值，略去 $\frac{1}{\sqrt{n}}$ ，即得常態曲線；如取其第二近似值，即得偏態曲線如上述。

更進一步之近似值，愛氏雖均曾作出，然此乃純粹理論之研究，本書不取也。

第五節 普遍的差誤律或大數律之說明

上文所證之定理，茲可概述於下，其有效條件，仍敘及並為擴大焉。

設有元素羣類甚多(n)，每一羣類，各可以其頻數軌跡表示之，然則自一類抽取一數 U 之機率，必為 U 之函數。

從各類中，各取出一事物， n 個事物之總計為 H ，則自此類抽選，不至一一或至微一一影響其他各類之抽取；照此手續重抽

進行，乃得若干數值之H，如此每一H之作成，絕不受其他各次抽選之影響（註二）。

然則如此等元素羣類之頻數軌跡果能適合某種條件，則H之頻數軌跡，必具有一定形式，此形式之第一近似值為

$$y = \frac{1}{s\sqrt{2\pi}} e^{-\frac{x^2}{2s^2}}, \text{ 而第二近似值則為}$$

$$y = \frac{1}{s\sqrt{2\pi}} e^{-\frac{x^2}{2s^2}} \left[1 - \frac{k}{2} \left(\frac{x}{s} - \frac{1}{3} \frac{x^3}{s^3} \right) \right]$$

式中之 s^2 ，為軌距之第二動差， ks^3 為軌距之第三動差。

量數之平均數之頻數軌跡，與其總和之頻數軌跡，形式相同，而且二者之K亦係同值。設 Sa 係平均數所成羣類之標準差， $Sa \approx \frac{s}{n}$ （如 σ 可為元素曲線標準差之代表， $S = \sigma\sqrt{n}$ ，而 $Sa = \frac{\sigma}{\sqrt{n}}$ ）。在H之頻數方程式中，乃係屬於 $\frac{1}{\sqrt{n}}$ 次，與1相對照，且當 n 為極大，或當元素曲線係對稱時（在此情形之下， $k=0$ ），只須求其第一近似值即足。

元素曲線必須適合之條件，為：如 up 為第 p 次之動差， σ 為其中任一羣類之標準差，則不論 p 值若干， $\frac{up}{\sigma^p}$ 終為一極小有限數（註三）（此值如乘以 $\frac{1}{n^{\frac{p}{2}-1}}$ 則變為極微可以省略）。欲達

到此目的，必須頻數曲線之絕大部份，離散情形，在底線上，自其平均數算起，向左向右不超過二，三個標準差。此種條件，平常頻數羣類當 n 全然甚大時當可滿足之。

第一次及第二次之近似值，只在 $\frac{x}{s}$ 之值中常適度時，方能有效，因超過此第一及第二近似值，則更高次近似值之結果，出入甚大矣，只有如此產生之 H 之頻數曲線之中心部份，方可加以決定，而中心以外各部則缺乏一般之形式，而且只能假定其集合體 (aggregate volume) 甚小，而超出 $3s$ 之機率甚微可以忽略也。所謂中心部份之全距，全視 n 值而定。互相獨立之元素數目愈增加，則可以決定之形式之全距益擴張。就平常情形言之，如 n 增大至 100 時，則頻數曲線，即在原點兩側各有 $2s$ 之全距也。

由此觀之，吾人不能以極端數值之位置，並不與差誤律相一致為理由，而否認差誤律對於某種觀察應用上之可能性。

第六節 範圍有限制之例

在第一節吾人曾有假定，謂一個項目之選擇，不至影響其他抽選之機會。

又如在第二章第十節所論，吾人現請就『範圍』（即從此範圍中抽選）受有限制之情形加以研究。

假設一羣類含有 n 個事物，此羣類乃自含有 N 個事物之羣類

中抽選而來，而此 N 個事物之量數為 $\bar{u} + u_1, \bar{u} + u_2, \dots, \bar{u} + u_N$ ，在此式中 \bar{u} 為其平均數，而 $\sum_1^N u_t = 0$ 。茲以 $H + E$ 為所選出 n 個事物之總和，而 $H + n\bar{u}$ 。

E 之均等或然值，有 NC_n 個，例如

$$u_1 + u_2 + u_3 + \dots + u_n$$

$$u_1 + u_3 + u_4 + \dots + u_N$$

.....

以上數值之和，必等於零，此乃顯而易見之事實，故 E 之平均值亦等於零。

設 s^2 為 E 之標準差。

則 $NC_n \cdot s^2 = NC_n$ 個平方之和，例如 $(u_1 + u_2 + \dots + u_n)^2$ 等，每個內包含 n 項即是。

在此總和中每一平方，例如 u_i^2 ，出現 $\frac{n}{N} \times NC_n$ 次，而每一乘積 $\cdot 2u_s u_t$ 出現 $\frac{1}{NC_2} \times \frac{n(n-1)}{2} \times NC_n$ 次，因共有 $n \times NC_n$ 個平方及 $\frac{n(n-1)}{2} \times NC_n$ 個乘積也。

$$\therefore NC_n \cdot S^2 = \frac{n}{N} \times NC_n \cdot S_1 N u_{t2} + 2 \cdot \frac{n(n-1)}{N(N-1)} \cdot NC_n \cdot S u_s u_t$$

$$S^2 = \frac{n}{N} \cdot N\sigma_2 + \frac{n(n-1)}{N(N-1)} \left\{ (S u_{st})^2 - S u_{st}^2 \right\},$$

(σ 為所從抽選之『範圍』之標準差)

$$= n\sigma^2 - \frac{n(n-1)}{N-1}\sigma^2, \quad \therefore \text{Sum} = 0$$

假如 $\frac{1}{N}$ 可以刪略,

$$= \sigma^2 \cdot n \frac{N-n}{N-1} = \sigma^2 n \cdot \left(1 - \frac{n}{N}\right) \dots \dots \dots (52)$$

設 σ_a 為抽選 n 次所得平均數 $\left(\frac{H}{n} + \frac{T}{n}\right)$ 之標準差。

$$\text{則 } \sigma_a = \frac{s}{n} = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\left(1 - \frac{n}{N}\right)} \dots \dots \dots (53)$$

假如 N 為無限大, 可依前文而得 $\frac{\sigma}{\sqrt{n}}$ (公式 38)。至 $\frac{n}{N}$ 一

值, 如果刪除之則標準差必有張大之虞。

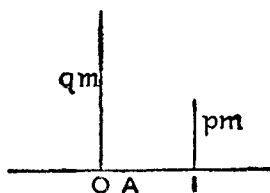
據證明 (Isserlis, 統計學報 Stat. Journal, 一九一八年號第七十五頁下同), 總和或平均所得之頻數, 酷近常態, 設 N 為極大 (事實上亦多如此)。

如以 $\frac{\sigma}{\sqrt{n}}$ 照查常態機率表 (第二章第四節第三表), 則必有將離中差超過實際數之危險。

附註——前論大數律係由 $(p+q)^n$ 之極限得來一言, 現可證明乃為一般分析之特殊情形。

設每一元素羣類, 各具有 qm 個零, pm 個單位, 而 $p+q=1$ 。

$$\text{此類之常數為 } \bar{x} = \frac{qm \times 0 + pm \times 1}{qm + pm} = p。$$



pm 與平均數, A , 之距離為 $+q$, qm 與 A 之距離為 $-p$ 。

$$\mu_2 = \frac{qm(-p)^2 + pm(q)^2}{(p+q)m} = pq, \sigma = \sqrt{pq},$$

$$k' = \frac{\mu_3}{\sigma^3} = \frac{q(-p^3) + p(q)^3}{(pq)^{\frac{3}{2}}} = \frac{q-p}{\sqrt{pq}}.$$

從每一 n 個曲線中各抽出一個數值, 而後相加, 所得之總數, 適合上述 H 構成之條件。

將此總數列成一頻數曲線, 其平均數為 pm , 標準差為 $\sigma\sqrt{n}$
 $= \sqrt{pqn}$, 而 $k = \frac{k'}{\sqrt{n}} = \frac{q-p}{\sqrt{pq}}$, 如上於第二章第四節所述。

第七節 例證

大數律, 如取其積分形式, 以第二近似值而論, 乃為

$$\begin{aligned} P(\pm x) &= \frac{1}{\sigma\sqrt{2\pi}} \int_0^x e^{-\frac{z^2}{2\sigma^2}} dz \mp \frac{k}{6\sqrt{2\pi}} \left\{ 1 - \left(1 - \frac{x^2}{\sigma^2}\right) e^{-\frac{x^2}{2\sigma^2}} \right\} \quad (\text{註四}) \\ &= \frac{1}{\sqrt{2\pi}} \int_0^z e^{-\frac{1}{2}z^2} dz \mp \frac{k}{6\sqrt{2\pi}} \left\{ 1 - (-z^2) e^{-\frac{1}{2}z^2} \right\} \\ &= F(z) \mp kf(z) \dots\dots\dots (54) \end{aligned}$$

式中 $P(x)$ 爲對平均數之正標準差不超過 x 之機率, $z = \frac{x}{\sigma} F(z)$

已列表見二章第四節第三表, 至 $f(z) = \frac{1}{\sigma\sqrt{2\pi}} \left\{ 1 - (1-z^2)e^{-\frac{1}{2}z^2} \right\}$

之表則列於次頁 (第四表)。

$k = \frac{u_3}{\sigma^3}$, u_3 爲該曲線之第三動差, σ 爲其標準差, 或依理論算定, 或由觀察計算而來, 結果均同。

下列例證八則, 解明以觀察配合曲線之方法。爲首二則 (字與磚), 以量數之原始, 示明與大數律之符合; 次二則 (頭蓋骨及鯨魚) 示在生物測量上之應用; 由再次一則 (年齡) 與心理現象乃有間接之關係; 最後三則 (速度, 食物消費量及價格) 其變化繁雜無常, 頻數曲線之形式無法預測。

以下各例, 僅將第一則全部作出, 其他則未也。

(1) 茲有冗長之書一本, (A) 依次取其 10,000 行之行首第一整字, 數各該字之字母若干; (B) 以十行之行首字爲一批, 各批字母相加, 然後計算 1000 批之字母總數若干; (C) 再以 100 行行首字爲一批, 各批字母相加, 便得 100 個總數。

A 之頻數曲線, 純由觀察而來, 故其形式如何, 不得而預知; B 之頻數曲線, 似能滿足大數律之條件, 惜『 n 』只爲 10, 故 A 苟非近於常態, 則僅能預測其中心部分; 至於 C, 以 n 爲 100, 第二近似值可與極大部分適合, 又如 A 恰爲對稱形式則第一近似值, 即

足應用矣。

A——10,000個字依其所含字母數之分佈情況

字	母	數	觀察數		xy	x^2y	x^3y	(註五) 差額		
			x	y				z	$F(z)$	$\times 10,000$
1	或 .5 至	1.5	-7	127	- 889	6,223	- 43561	-1.62	0.447	490
2	,, 1.5 ,,	2.5	-6	1,792	-1,752	64,512	-287072	-1.27	.396	770
3	,, 2.5 ,,	3.5	-5	1,984	- 9920	49,600	-248000	- .92	.321	1,020
4	,, 3.5 ,,	4.5	-4	1,240	- 4960	19,840	- 79360	- .58	.219	1,280
5	,, 4.5 ,,	5.5	-3	968	- 2904	8,712	- 26136	- .23	.091	1,390
6	,, 5.5 ,,	6.5	-2	812	- 1624	3,248	- 6496	+ .12	.048	1,330
7	,, 6.5 ,,	7.5	-1	893	- 893	893	- 893	+ .47	.181	1,130
8	,, 7.5 ,,	8.5	0	634	0	0	0	+ .82	.294	850
9	,, 8.5 ,,	9.5	1	602	+ 602	602	+ 602	+1.17	.379	570
10	,, 9.5 ,,	10.5	2	460	+ 920	1,840	+ 3680	+1.52	.436	330
11	,, 10.5 ,,	11.5	3	260	+ 780	2,340	+ 7020	+1.87	.469	180
12	,, 11.5 ,,	12.5	4	116	+ 464	1,856	+ 7424	+2.22	.487	80
13	,, 12.5 ,,	13.5	5	69	+ 345	1,725	+ 8625	+2.57	.495	30
14	,, 13.5 ,,	14.5	6	21	+ 126	756	+ 4536	+2.92	.498	10
15	,, 14.5 ,,	15.5	7	18	+ 126	882	+ 6174	+3.27	.499	10
16	,, 15.5 ,,	16.5	8	4	+ 32	256	+ 2048	+3.62	.500	0
10,000					-31942	163,285	-791518			
					+ 3395		40109			
					-28547		-751409			

$\bar{x} = -2.8547$. 平均數是 $8 - \bar{x} = 5.1453$

$\mu_2 = 16.5285 - x^2 = 8.1792, \sigma = 2.860$

$\mu_3 = -75.1409 - 3(-2.8547)(16.5285) + 2(-2.8547)^3 = 18.1704$.

$\kappa = \frac{\mu_3}{\mu_2^{3/2}} = .78$.

爲計算動差，即以 8 爲假定原點。

配合常態曲線， $z = \frac{x - \bar{x}}{\sigma}$ 。橫列第一行， $F(z) = .447$ 。表示平均數，與 .5 個字母 ($x = -7.5$) 間之部分。在 .5 個字母以下者，依常態曲線所示，共有 530 件；此外最末一縱列與觀察值 y ，並不甚近。可知原素曲線雖非常態，但其爲單峯的且爲連續的，乃無可置疑，又不論其偏斜度若何，其大部總在 $\bar{x} \pm 2\sigma$ 範圍之內。故如自曲線隨機抽選若干元素加在上面，則求得大數律之條件均已齊備矣。

B——十字爲一批，千批字母總和之分佈情況

字母數	$F(z)$	差額	觀察數	$F(z)$	差額
		$\times 1000$		千 $Kf(\varepsilon)$	$\times 1000$
				(註七)	
		4	0	.078	0
26.5	-2.650	.496	13	.091 (註六)	8
31.5	-2.119	.483	39	.095	37
36.5	-1.588	.444	89	.071	99
41.5	-1.057	.355	154	.025	173
46.5	-.526	.201	263	.026	213
51.5	+.005	.002	202	.072	191
56.5	+.536	.204	153	.095	135
61.5	+1.067	.357	88	.078	78
66.5	+1.598	.445	38	.091	40
71.5	+2.129	.483	13	.078	18
76.5	+2.660	.496	3	.039	7
81.5	+3.191	.499	1	.067	2
86.5	+3.722	.500	0		0

關於此 1000 批總和，平均數為 51.453, $\sigma=9.4155$, $k=.40$ 93。字母數之和，最大不過 87，最小亦在 26 之上。第七八九三欄關於 z , $F(z)$ 及差額之計算與 A 相同。此者對於常態曲線之配合情形，良好多多，而在其全距 31.5 至 76.5，即平均數士 2σ 之間，其為佳妙，可謂無以復加，但公式應用結果，在 31.5 以下過多，而在 76.5 上者甚少，此頗嫌美中不足耳。

第二近似值通盤配合甚接近，不過在 86.5 之上，有一項未能包括在內而已（參閱第十章配合之測驗）。

此等觀察之標準差即為 A 之標準差（2.860），而被 $\sqrt{10}$ 乘；此等觀察之 K 即為 K (.81) 而 $\frac{1}{\sqrt{10}}$ 乘（公式 37 及 39）。

但 $2.860 \times \sqrt{10} = 9.04$, $.81 \div \sqrt{10} = .25$ ，而吾人由 B 直接求得之標準差為 9.42, K 為 .41，相差如此之鉅，可知乃係每十字一批，每批字母攙集時，未能完全獨立之故；迨經分析，乃知連續所抽十字之間，果有相當關聯也。實際上，如照 C 式連續將百字字母相加，則得 $\sigma=33.311$ 非復 $2.86 \times \sqrt{100}$ 矣，然若將各批字母數次序變更重為排列，使以便將全書各部份之行首字盡行包括於每次一百字之各批中，則 σ 為 28.87，則與理論相合。

C——百字為一批，共百批字母總數之分佈

字母數		$F(z)$	差 額 ×100	觀察數
415	-3.001	.499	.7	1
435	-2.400	.492	2.8	2
455	-1.800	.464	7.9	7
475	-1.200	.385	16.0	19
495	-.599	.225	22.5	25
515	-.001	.000	22.6	18
535	+ .602	.226	15.9	18
555	+1.202	.385	7.9	6
575	+1.803	.464	2.8	3
595	+2.403	.492	.7	0
615	+3.003	.499	.1	1
635	+3.604	.500		

此表觀察數與公式融合程度頗密切 (參閱十章) 即用第二近似值, 亦難期有顯明之精進也。

此一實驗之舉行, 目的在解明大數律 (及相關面, 如公式102所論), 結果可謂極為滿意, 甚且將隨機抽樣之困難, 亦為例釋無遺。

(2) 一花園中, 行人路側有邊界用磚堆壘而成, (並非實砌), 兩端縱長相間連接, 乃經時日既久, 風吹雨打, 遊人雜踏, 歷經消磨, 迄今已非原狀。茲以四塊磚為一列, 量其 143 列之長, 量法務求精密, 以至一英寸之十六分之一為止。但其變量必須顧及, 究

其變量之原因，一則由於製造時之欠精，二則由於壘磚時，距離排列之不準，三則由於壘成後之推動，四則由於測量之困難，綜此數因，繁複已極，但各為獨立，而影響細微。此等影響，可視為差誤之總和，而數量分佈情形，或與常態，對稱相近似，亦未可知也。

四磚為一列，各列長度之分佈

長 度	觀察次數	依公式計算之結果 (常態曲線)
35	1	.7
$35\frac{1}{16}$	1	1.4
$35\frac{1}{8}$	3	2.7
$35\frac{1}{4}$	7	5.1
$35\frac{1}{2}$	11	8.0
$35\frac{3}{8}$	4	11.6
$35\frac{1}{2}$	21	15.0
$35\frac{7}{8}$	7	17.7
$35\frac{1}{2}$	30	18.3
$35\frac{9}{16}$	16	17.5
$35\frac{5}{8}$	13	14.9
$35\frac{11}{16}$	6	11.4
$35\frac{3}{4}$	11	7.9
$35\frac{13}{16}$	7	5.0
$35\frac{7}{8}$	4	2.7
$35\frac{15}{16}$	1	1.4
36	0	.6
	143	142.9

除非有量至一英寸之最近八分之一，而非十六分之一之顯明傾象，則配合結果，必甚良好；如更將趨勢修改，配合尤佳。

(3)茲就塞利格門教授所著『英埃蘇丹赫買堤族問題觀』(Professor C. G. Seligman's "Some aspects of the Hamitic Problem in the Anglo-Egyptian Sudan")一書中，採取下列量數，至其類數分類，乃循原著者之請，經本人分析而成。

頂可(Dinka)族頭蓋骨及體高之測量

與平均數 相距等級	F(ε)	頭部指數		鼻樑指數		體高	
		差額 ×148	觀察數 *	差額	觀察數	差額	觀察數
3σ以上	.4966	.2	2	.1	0	.1	0
2½σ-	.4938	.9	1	.5	1	.7	2
2σ-	.4772	2.3	2	1.3	0	1.8	1
1½σ-	.4332	13.6	4	3.7	4	5.1	1
σ-	.3413	22.2	14	7.8	6	10.7	6
½-	.1915	28.3	18	12.8	13	17.4	22
0-	0	28.3	30	16.3	27	22.2	24
½	.1915	22.2	30	16.3	12	22.2	25
1σ	.3413	13.6	25	12.8	8	17.4	23
1½σ	.4332	6.5	13	7.8	6	10.7	6
2σ	.4772	2.3	7	3.7	4	5.1	4
2½σ	.4938	.9	2	1.3	3	1.8	1
3σ以下	.4938	.2	0	.5	1	.7	1
	.4986		0	.1	0	.1	0
總計.....		148	148	85	85	116	116
平均數.....			72.7		91.6		178.6
標準差.....			3.70		13.0		9.66

除兩極端項註有*號者外，全距係為常態形，且以上取材之

例如此之少，與常態曲線相差並不如所預料之多。

(4) 北海漁業調查所，五百五十四條鯨魚之體長測量，結果如下：

長 度		F(σ)	差 額 ×554	觀察數
35.5	2.825	.4976	1.3	0
34.5	2.076	.4810	9.2	6
33.5	1.327	.4077	40.6	50
32.3	.578	.2183	104.9	105
31.5	.171	.0679	158.6	166
30.5	.920	.3212	140.3	145
29.5	1.669	.4524	72.7	61
28.5	2.418	.4922	22.0	10
27.5	3.167	.4992	3.9	7
26.5	3.916	.5	.4	3
25.5	4.665	.5	0	1
				554

平均數 31.778; $\sigma=1.335$.

兩極端各項未能與常態相緊合。

(5) 美國聖路易 (St. Louis, V. S. A.) 公立學校報告中，曾將第六級年齡不同之學童人數，調查如下。

下列之表，將所得資料與大數律之第一第二近似值作一比較。

用近似值求得之人數

年齡	學童人數	第一近似值	第二近似值
10	23	59	27
11	201	207	204
12	673	620	670
13	1,001	983	995
14	729	785	746
15	310	323	307
16	80	67	79
17	13	9	15
18	1	0	0

平均年齡, 13.665; $\sigma = 1.190$; $\kappa = .2059$ 。

第一近似值, 在平均數左右 2σ 以內, 尚能配合。

第二近似值, 則與觀察值密切配合 (參閱附錄六另圖)。

(6) 一百人行路之速度 (見 “Die Schwankungen der landwirtschaftlichen Reinertrage-Mitscherlich”) 茲依此一百人在兩點間, 所用之時間, 作一觀察, 計算如下。

平均速度每秒, 1.5846 公尺。 $\sigma = .2179$ 公尺。

速度	各級速度人數	
	計算數	實准數
平均數 + .50m 以上	1.1	2
+ .40 至 .50	2.3	2
+ .30 ,, .40	5.0	4

+ .20 ,, .30	9.6	11
+ .10 ,, .20	14.3	10
0 ,, .10	17.7	18
- .10 ,, 0	17.7	20
- .20 ,, - .10	14.3	15
- .30 ,, - .20	9.6	8
- .40 ,, - .30	5.0	7
- .50 ,, - .40	2.3	3
- .50 以下	1.1	0

(7) 採用 1918 年勞工生活費委員會所搜集之材料，將在城市之家庭九百七十家，每星期食物費，用『等成年』(equivalent adults 一即以兒童作為成年人之幾分之幾計算——)除之。平均數為 10.75 先令； $\sigma = 3.156$, $k = .84$ 。

每週每一單位 之食物費	家 數		
	實際數	用第二近似值 求得之數	由皮爾生氏第 三形態求得數 (註八)
5.5 先令以下	18	22	7
5.5 先令	107	123	122
7.5	255	233	252
9.5	245	248	250
11.5	173	168	172
13.5	101	89	95
15.5	38	51	45
17.5	17	22	19
19.5	9	11	7
21.5 以上	7	1	1

$$\beta_1 = .708, \beta_2 = 4.035.$$

(8) 美國 272 處之麵粉價格，已由調查得來。內有五個城市，每磅之價美金四分，此顯屬例外，故將其放棄。其餘 267 處，平均

每磅價 2.629 分, 而 $\sigma = .3334$ 。

依麵粉價格對各城市之分類

平均數	以第一近似值算出	實際數
+3 σ 以上	.4	2
+ $\frac{5}{2}\sigma$ 至 3 σ	1.3	3
+2 σ $\frac{5}{2}\sigma$	4.4	1
+ $\frac{3}{2}\sigma$ 2 σ	11.8	9
+ σ $\frac{3}{2}\sigma$	24.5	16
+ $\frac{1}{2}\sigma$ σ	40.0	43
0 $\frac{1}{2}\sigma$	51.1	67
- $\frac{1}{2}\sigma$ 0	51.1	47
- σ - $\frac{1}{2}\sigma$	40.0	37
- $\frac{3}{2}\sigma$ - σ	24.5	25
-2 σ - $\frac{3}{2}\sigma$	11.8	9
- $\frac{5}{2}\sigma$ -2 σ	4.4	4
- $\frac{3}{2}\sigma$ 以下	1.7	4

在平均數士 2 σ 全距之內, 與常態頗稱融洽, 而與第十章所論之測驗, 亦無不合。

(註一) 多項式定理乃為二項式定理之引申; 茲將其證明, 扼要述之如下:
 $(a_1+b_1+c_1+\dots)(a_2+b_2+c_2+\dots)(a_t+b_t+c_t+\dots)$ t 個因子之乘積
 = 各可能項, 諸如 $a_1 a_2 b_3 c_4 d_5 b_6 \dots k_t$ (每一附數 (suffix) 各出現一次) 之總和。 a 出現 n_1 次, b 出現 n_2 次……諸如此類之項數, 乃

即每次取 t 個事物而有 n_1 個相同, n_2 個相同……, 意即 $\frac{t!}{n_1! n_2! \dots}$,

之排列(permutation)數。以 $a_1 = a_2 = \dots = 1u, b_1 = b_2 = \dots = 2u, \dots$ 以此類推, 則得結果如正文。

- (註二) 設自一基本羣類抽選, 不能彼此獨立, 而各數量乃成批出現, 則須多得 H 之數值, 當所得數值甚多時, 方能求出其最終類數形式之已知近似值,
- (註三) 以準確之語言之, 此數非他, 即此比率與在被包含之常態曲線中之相當比率之相差也。
- (註四) 見附錄六, 積分部分。
- (註五) 見第二章第四節第三表。
- (註六) $+$, z 為負數時用之。
- (註七) 見第三章第七節 $f(z) = \frac{1}{\sqrt{2\pi}} \left\{ 1 - (1 - z^2)e^{-\frac{1}{2}z^2} \right\}$ 數值表
- (註八) 見第五章第一節。

第四章 差誤律之應用

第一節 平均數及總和數之精度

依前章所述，如 n 個可以度量之事物，乃由一大範圍中隨機抽選而來，而在此範圍中之大小，乃分佈成一十分連續之頻數羣類，同時在此羣類中，與其標準差 (σ) 相比，並無距離其平均數甚遠者，則此平均數，即屬於一以 $\frac{\sigma}{\sqrt{n}}$ 為標準差且形式幾為常態之頻數曲線。

σ 普通均須由觀察之本身得來，與大範圍之標準差或有不同，但二者相差之量，不過 $\frac{\sigma}{\sqrt{n}} \times \frac{1}{\sqrt{n}}$ 次而已（見本編第九章第五節）。

下文所舉第一例（每家住屋之人數），示吾人以十二個地方之情形，就此十二個地方，以樣本之平均數，與樣本所自來之大範圍中之平均數比較。

次舉二例（數字及緯度），表明若干平均數之分佈，與差誤常態曲線相合之情形。

在適用原理之情形下，不止平均數之標準差，可以指明，即平均數之差誤超過該標準差某種倍數之機率，亦可決定。

因範圍爲不可知之數，吾人無法考察，該範圍中之頻數羣類，與愛基華斯氏所舉之條件（第三章第四節），是否相合，不得而知。吾人測驗之法，有時只能借重樣本之本身。假如吾人於每 n' 項中抽出 k 個樣本，然後將歷次所抽得之平均數，組成頻數羣類。如此則在範圍中之條件苟能滿足時，此一頻數羣類必幾近常態，但如 n' 並不甚大時，則亦不完全近於常態。如此言屬實，則不妨另抽一較大樣本，使其包括 $n = n' \times k$ 個項目。如是則此一大樣本之平均數，乃即 $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_k$ 之平均數，而且此大樣本，既由一羣類中抽出 k 個事物而成（此羣類既近於常態，必能滿足所舉條件），則大樣本平均數中之差誤，必具有以 $\frac{\sigma}{\sqrt{n}}$ 爲標準差（ σ 爲由 n 個觀察值中計算而得）之常態頻數無疑。以 k 個數量 $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_k$ 之標準差，必與 $\frac{\sigma}{\sqrt{n'}}$ 相近也。

如下文第四節所舉關於緯度之例，以二〇〇〇項集成八〇個羣類之分配情形，本不得而知。惟 $k = 80, n' = 25$ 。八十個羣類之平均數之標準差，爲 1.628。由此推論，在『範圍』中之標準差，必約略爲 $1.628 \times \sqrt{25} = 8.14$ 。然則以二千項全部爲基礎之平均數之標準差，必爲 $\frac{1.628 \times \sqrt{25}}{\sqrt{2000}} = \frac{1.628}{\sqrt{80}}$ ，讀下文自知。

反之，吾人可對於經 n 次抽樣組成之頻數羣類，加以考察，視其是否與愛基華斯氏所列條件相合。果如相合，則彼平均數之差

誤，必係一常態之頻數。

第二節 平均數之精度

茲自若干區域之戶主調查表中抽取樣本（見第二章第九節第一例），並將十二區之每家住宅平均人數算出。

註冊區	五十抽一之抽樣			全部各區		標準差	
	住宅數	人數	每宅人數	住宅數	每家住宅人數		
白斯諾爾	東北區	277	1,224	4.42	13,850	4.35	.14
	西南區	278	1,261	4.54	13,505	4.60	.14
受爾地支	南區	187	792	4.24	9,331	4.26	.18
	西區	152	693	4.56	7,623	4.34	.19
	東北區	156	653	4.19	7,847	4.39	.19
斯皮特非爾		130	637	4.90	6,476	4.79	.21
懷德加白		117	519	4.44	5,914	4.72	.22
聖喬治		187	924	4.93	9,374	4.88	.18
色得外爾		95	387	4.07	4,800	4.37	.25
萊母好斯		133	611	4.59	6,655	4.54	.21
邁爾恩德	西南區	267	1,211	4.54	13,366	4.71	.15
	東北區	207	839	4.05	10,364	4.40	.17

吾人經將全部各區之調查表冊，加以探討後，得知家家住宅人數之標準差(σ)，全距乃在2.38至2.75之間。

上表第一項之標準差，如取最低及嚴格之 σ 值，乃為 $\frac{\sigma}{\sqrt{277}}$ 。其他標準差，計算方法相同。

樣本平均數與全體平均數之相差，少於算得之標準差者有六件，多於算得之標準差，而超過之額不及標準差之四分之一者有四件，相差超過標準差之額當標準差之百分之三十者有一件，又相差之額有標準差之兩倍者有一件。

第三節 平均數之常態分配

從七位數之數學用表中，依次將各數之末位數記下，集成十個數字，復行相加，如此累次進行，俟得至一千個總和為止。（註一）

此一千個總數，組成一羣類，其平均數為 45.014，標準差為 9.025，而隨機抽樣之範圍為無限大（假如 0 至 9 之數字分佈均勻）時，平均數為 45，標準差為 $\sqrt{82.5} = 9.083$ 。

下列一表，將一千個總數之分配與差誤常態曲線，二者作一比較。

在某種限度內1000個總和之次數

與平均數之距離	算出之次數	標準差	觀察數	差額
$\frac{7}{2}\sigma$ 以上	6	2.4	8	+ 2
2σ	17	4.1	17	+ 0
$\frac{3}{2}\sigma$	44	6.5	47	+ 3
σ	92	9.1	75	-17
$\frac{1}{2}\sigma$	150	11.3	157	+ 7
0 至 $\frac{1}{2}\sigma$	191	12.4	197	+ 6
0 至 $-\frac{1}{2}\sigma$	191	12.4	201	+10
$-\frac{1}{2}\sigma$	150	11.3	148	- 2
$-\sigma$	92	9.1	77	-15
$-\frac{3}{2}\sigma$	44	6.5	50	+ 6
-2σ	17	4.1	20	+ 3
$-\frac{7}{2}\sigma$ 以下	6	2.4	3	- 3

標準差係依公式 $\sqrt{p(1-p)n}$ （見公式 (13)）算來，在此， $n=1000$ ， p 為依常態定律落於某一級內之比例；例如在 σ 與 $\frac{3}{2}\sigma$ 之間，應有全部之 0.92，故 $p=0.092$ 。

將觀察值排列起來，以 9.205 作為 σ 。

理論數與觀察數之相差，少於標準差者有九件，多於一標準差而不及兩標準差者有三件。

由此可知，常態曲線實為該一羣類之適宜代表。

在此全部樣本中算得平均數為45.014，標準差則為 $\frac{9.205}{\sqrt{1000}} = 0.29$ ，且就一般觀之，竟與十個數字總和之平均數相近，實出意料之外。

取一地名索引，從其所有三萬一千二百一十個地名中，用粗略方法隨機抽出二十五個，其緯度只記到度數為止，分數及南北方位均刪略之。將二十五個之緯度，加以平均，如此重疊施行，以得到八十個平均數為止。

總平均數為 35.0° ，而此80個平均數所成之羣類，標準差為 1.628° 。

下列之表，依上例之計劃，將觀察之分配，以與差誤常態曲線比較。

與平均數之距離	算出之次數	標準差	觀察數	差額
$\frac{3}{2}\sigma$ 以上	.5	-	1	0至1
2σ	1.3	1.1	2	1
$\frac{3}{2}\sigma$	3.5	1.8	1	2至3
σ	7.4	2.6	10	3
$\frac{1}{2}\sigma$	12.0	3.2	12	0
0至 $\frac{1}{2}\sigma$	15.3	3.5	12	3
0至 $-\frac{1}{2}\sigma$	15.3	3.5	16	1
$-\frac{1}{2}\sigma$	12.0	3.2	15	3
$-\sigma$	7.4	2.6	7	0
$-\frac{3}{2}\sigma$	3.5	1.8	2	1至2
-2σ	1.3	1.1	1	0
$-\frac{3}{2}\sigma$ 以下	.5	-	1	0至1

八件之差額在一標準差以下 餘二件之差額則略超過之。

平均數 35.0° ，既從 $80 \times 25 = 2000$ 緯度樣本中得出，標準差爲 $\frac{1.628}{\sqrt{80}}$ 度 = .18 度，故平均數之第一位小數，並不能確切求得。

第四節 加權總和及加權平均之絕對差誤

前於第三章第一節曾論及，如 $H + E$ ，係從 n 個頻數羣類，獨立抽出 n 個數量之總和，（在此 n 個頻數羣類中，平均數爲 $1\bar{u}$, $2\bar{u}$, $\dots\dots n\bar{u}$ ，標準差爲 $\sigma_1, \sigma_2, \dots\dots \sigma_n$ ，而 $H = 1\bar{u} + 2\bar{u} + \dots\dots + n\bar{u}$ ），則 $H + E_t$ ——任一次抽選之總和——必等於 $1u_t + 2u_t + \dots\dots nu_t$ ，而其標準差爲 s ($s^2 = \sigma_1^2 + \sigma_2^2 + \dots\dots + \sigma_n^2$)。

將上文重爲排比，可知如用加權總和 (weighted sum) $H + E_t = W_1 \cdot 1u_t + W_2 \cdot 2u_t + \dots\dots + W_n \cdot nu_t$ ($W_1, W_2, \dots\dots$ 爲常數)，標準差變爲

$$s^2 = W_1^2 \sigma_1^2 + W_2^2 \sigma_2^2 + \dots\dots + W_n^2 \sigma_n^2 = S (W_t^2 \sigma_t^2) \dots (55)$$

至加權平均數 $\frac{H + E}{SW_t}$ 之標準差， s_a ，係從下式得來

$$s_a^2 = \frac{S(W_t^2 \sigma_t^2)}{(SW_t)^2} \dots\dots\dots (56)$$

n 值若大， $\frac{1}{\sqrt{n}}$ 值必微，故可略去，第三章第五節後半段所列之其他條件，乃亦滿足，總和及平均之頻數，必係常態，而第一章第四節之第三表，亦確可用來決定對平均值 H 之離中差之機率。

設 $\bar{\sigma}^2$ 爲 $\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2$ 之加權平均值，則

$$\bar{\sigma}^2 S(W_t^2) = S(W_t^2 \sigma_t^2)$$

$$s^2 = \bar{\sigma}^2 S(W_t^2)$$

$$s_a^2 = \bar{\sigma}^2 \frac{(W_t^2)}{(SW_t)^2}$$

然後再設 $SW_t = n\bar{w}$, $W_t = \bar{w} + w_t$, $n\sigma_w^2 = S(w_t^2)$, 則 \bar{w} 爲 w 羣類之平均數, σ_w 爲其標準差。 $S w_t = 0$ 。

於是

$$S(W_t^2) = S(\bar{w}^2 + 2\bar{w}w_t + w_t^2) = n\bar{w}^2 + 2\bar{w}S w_t + S w_t^2 = n(\bar{w}^2 + \sigma_w^2) \quad (58)$$

$$s^2 = n\bar{\sigma}^2(\bar{w}^2 + \sigma_w^2) \quad \dots\dots\dots (59)$$

$$s_a^2 = \bar{\sigma}^2 \cdot \frac{n(\bar{w}^2 + \sigma_w^2)}{(n\bar{w})^2}$$

$$\therefore s_a = \frac{\bar{\sigma}}{\sqrt{n}} \cdot \sqrt{\left(1 + \frac{\sigma_w^2}{\bar{w}^2}\right)} \quad \dots\dots\dots (60)$$

上列末一公式，在權數爲已知且無差誤時，用作求加權平均數之標準差，甚爲簡便。如此，原來項目之離中差，已按 $1:\sqrt{n}$ 之比率降低，且當 n 大時，離中差亦愈小，然因子 $\sqrt{\left(1 + \frac{\sigma_w^2}{\bar{w}^2}\right)}$ 罕大於 $\sqrt{2}$ ，因用以測量權數之標準差，對權數平均值之比率者，爲 $\frac{\sigma_w}{\bar{w}}$ ，而此比率通常均小於整一也。

如此平均數未曾加權，則

$$s_a = \frac{\bar{\sigma}}{\sqrt{n}} \quad \dots\dots\dots (61)$$

基本公式 $s^2 = S(Wt^2\sigma_t^2)$ ，爲英國小量收入協會之委員會 (Committee of the British Association on Small Incomes) 所用。(參閱 Statistical Journal, 1910, p. 62, 該學報中應用字母略異)。

茲共有三十一級，每級中不繳所得稅者，估計共有人數 N_t ，標準差爲 S_t ；此數等級之平均收入爲 I_t ，標準差爲 s'_t 。則此類總收入爲 $N_t I_t$ ，標準差爲 σ_t ，而

$$\begin{aligned}\sigma_t^2 &= \text{平均值}\{(N_t + e_t)(I_t + e'_t) - N_t I_t\}^2 \\ &= \text{平均值}\{N_t e'_t + I_t e_t\}^2\end{aligned}$$

如將 e 之乘積省略，則

$$= N_t^2 s'^2_t + I_t^2 s_t^2。$$

$N_t I_t$ 之總和之標準差必爲 s ，而

$$s^2 = S(N_t^2 s'^2_t + I_t^2 s_t^2)。$$

依此辦法，將各級之標準差分別估計，例如 s_1, s_2, \dots 及 s'_1, s'_2, \dots ，是也。

設各級確實人數，業已知曉，且僅所有各級中之平均收入，易有差誤，則吾人應可採用上述公式 $s^2 = (Wt^2\sigma_t^2)$ ，而在此情形之下，則 $s^2 = S(N_t^2 s'^2_t)$ ，此式吾人如以 $s_t = 0$ ，當然亦可求得之。至所有各級平均收入中之差誤標準差，則爲 $\frac{\sqrt{S(N_t^2 s'^2_t)}}{SN_t}$ 。

在調查中 $S(I_t^2 s_t^2) = 315 \times 10^5$ ， $S(N_t^2 s'^2_t) = 4 \times 10^8$ ，故 N 內

之差誤，不甚重要。 $S(N_t) = 4023$ ， $S(N_t I_t) = 284,700$ ，至於一九一一年不付所得稅者中，除工人外，其平均收入，係七十一鎊，其標準差為五鎊。

第五節 相對差誤

絕對差誤，離中差，及某一數值與觀察值對於平均數（或真值）之真正差額，上文業已述及，現請更進而討論相對差誤，及離中差（第一編第八章曾述及之）。

設 x 為某數量之觀察值，該數量之真值或平均數則為 x' ，且 $x = x'(1+e)$ ，於是 $e = \frac{x-x'}{x'}$ 即為相對差誤或離中差。（註二）

(1) 乘積及商數

假如 F_1, F_2 兩因子，互為獨立，但測量時，因有差誤，故成為 $F_1(1+e_1), F_2(1+e_2)$ 。結果 e 即為乘積 P 之相對差誤，因而 $P(1+e) = F_1(1+e_1) \cdot F_2(1+e_2)$ ，式中 $P = F_1 F_2 \dots \dots \dots$ (62)

如果 e 之乘積可以略去時， $e = e_1 + e_2 + e_1 e_2 = e_1 + e_2$ 。

是以，如 $\sigma, \sigma_1, \sigma_2$ 為 P, F_1, F_2 之標準差，依照公式(34)， $\sigma^2 = \sigma_1^2 + \sigma_2^2$ 。

此一得數，可以引申使因子數達於任何之有限數，於是

$$\sigma^2 = \sigma_1^2 + \sigma_2^2 + \sigma_3^2 + \dots \dots \dots (63)$$

x^n 之差誤（ n 為有限數），如用 $x^n(H_e) = \{x(1+e_1)\}^n$ 算來，

(至 e_1 , 乃為 x 之差誤),

$$\therefore e = ne_1 + \frac{n(n-1)}{2} e_1^2 + \dots = ne_1 \dots \dots \dots (64)$$

當刪除其平方時, 如以 σ 為 x 之標準差, x^n 之標準差為 $n\sigma$ 。

如 n 為分數, 此得數即為真實。例如, 立方根中之差誤, 即為該數量差誤之三分之一。如此, 假如 1006 一數, 誤書為 1000 (其相對差誤為 .006), 則立方根中之相對差誤, 即為 .002, 但此乃以 10 為根, 並非 $10.02 = 10(1+.002)$ (約略數) 也。

設 e 為 $Q = F_1/F_2$ 中之差誤; 且 F_1 與 F_2 彼此互相獨立,

$$Q(1+e) = \frac{F_1(1+e_1)}{F_2(1+e_2)} \dots \dots \dots (65)$$

$e = (1+e_1)(1+e_2)^{-1} - 1 = e_1 - e_2 +$ 平方及乘積。

$\sigma_q^2 = \sigma_1^2 + \sigma_2^2$, 此處 σ_q 乃為 e 之標準差 $\dots \dots \dots (66)$

設 e 為乘幕, a^x , 之差誤, 如 a 為已知, e_1 為 x 中之差誤,

$$a^x(1+e) = a^{x(1+e_1)}$$

$e = a^{xe_1} - 1$ 如將 e_1^2 略去, 則

$$= e_1 \cdot x \log a \dots \dots \dots (67)$$

就一般情形言之, e 為一函數, $f(x)$, 中之差誤, 則

$$f(x) \times (1+e) = f\{x(1+e_1)\} = f(x) + e_1 x f'(x) + \dots$$

$$e = x \frac{f'(x)}{f(x)} \cdot e_1 \dots \dots \dots (68)$$

(2) 平均數中之差誤

設 \bar{m} 為 n 個數量 $M_1, M_2, \dots, M_t, \dots, M_n$ 之未加權平均

數，並設 $M_t = \bar{m} + m_t$ ，於是 $\sum m_t = 0$ 。又設 $n\sigma_m^2 = \sum m_t^2$ 。

假設該種數量，因觀察有誤，致以 M_t 寫成 $M_t(1+e_t)$ ……等等，其他以此類推，茲以 e 為其平均數中之相對差誤。

$$\bar{m}(1+e) = \frac{1}{n} \sum \{M_t(1+e_t)\} = \bar{m} + \frac{1}{n} \sum (M_t e_t)。$$

$$\text{於是} \quad e = \frac{1}{n} \sum \left(\frac{M_t}{\bar{m}} e_t \right) \dots\dots\dots (69)$$

如 s_a, σ_t 為 e, e_t 之標準差，則以公式(55)，

$$s_a^2 = \sum \left(\frac{M_t}{n\bar{m}} \sigma_t \right)^2 = \sigma^2 \sum \left(\frac{M_t}{n\bar{m}} \right)^2，$$

又如 σ^2 為 $\sigma_1^2, \dots, \sigma_t^2, \dots$ 之加權平均數，或所有各標準差，均各相等

$$s_a^2 = \sigma^2 \cdot \frac{\sum (\bar{m} + m_t)^2}{n^2 \bar{m}^2} = \sigma^2 \cdot \frac{\bar{m}^2 + \sigma_m^2}{n \bar{m}^2}, (\because \sum m_t = 0)$$

$$\text{同時} \quad s_a = \frac{\sigma}{\sqrt{n}} \sqrt{\left(1 + \frac{\sigma_m^2}{\bar{m}^2}\right)} \dots\dots\dots (70)$$

此公式與前論(公式60)加權平均數之絕對差誤時所用公式，完全相同。

在此公式中， σ_m, \bar{m} 及 \sqrt{n} ，均為已知，故 $\frac{s_a}{\sigma}$ 之比，乃得確實算出。無論關於各個量數之已知情形為何，在此環境之下，均須將 σ 計算出來。

第三章第五節所舉條件，在計算平均數時，大概均可滿足，(設隨機抽樣之條件，已經妥為維持)，故如 n 大時，均可查常態次數表。至 n 大不過二十時，該表亦可約略應用也。

(3) 加權平均數中之相對差誤

[以統計學報(Statistical Journal), 1911-12號, 第八十一至八十八頁一文為根據]

設 $\bar{m}_w = \frac{S(W_t M_t)}{S W_t}$, 而 M_t (以及 \bar{m}, σ_m 等) 意義同前, $W_1, W_2, \dots, W_t, \dots, W_n$ 為權數。

設 $W_t = \bar{w} + w_t$, 此處之 $n\bar{w} = S W_t$, 而 $S w_t = 0$, 並設 $n\sigma_w^2 = S w_t^2$ 。

於是 $S(W_t M_t) = n\bar{w}\bar{m}_w$ 。

現假設吾人對於權數所知不確, 以致誤用 $W_t(1+\eta_t)$ 以代 W_t, \dots 。

設 M 中之差誤, 如故, 而以 e 為 \bar{m}_w 中之差誤。

$$\text{則 } \bar{m}_w(1+e) = \frac{S\{W_t(1+\eta_t)M_t(1+e_t)\}}{S\{W_t(1+\eta_t)\}}。$$

$$\begin{aligned} \therefore e &= \frac{S\{W_t(1+\eta_t)M_t(1+e_t)\} \cdot S W_t - S(W_t M_t) \cdot S W_t(1+\eta_t)}{S(W_t M_t) \cdot S\{W_t(1+\eta_t)\}} \\ &= \frac{S(W_t M_t e_t) \cdot S W_t + S(W_t M_t \eta_t) \cdot W_t - S(W_t \eta_t) \cdot S(W_t M_t)}{S(W_t M_t) \cdot S W_t} \end{aligned}$$

略去 $e\eta$ 及 η^2

$$\begin{aligned} &= \frac{S(W_t M_t e_t)}{S(W_t M_t)} + \frac{S\{(W_t M_t \cdot n\bar{w} - W_t \cdot n\bar{w}\bar{m}_w)\eta_t\}}{S(W_t M_t) n\bar{w}} \\ &= \frac{S(W_t M_t e_t)}{n\bar{w}\bar{m}_w} + \frac{S\{W_t(m + \bar{m} - \bar{m}_w)\eta_t\}}{n\bar{w}\bar{m}_w} \dots \dots \dots (71) \end{aligned}$$

$$\bar{m}_w = \frac{S\{(\bar{w} + w_t)(\bar{m} + m_t)\}}{n\bar{w}} = \frac{n\bar{w}\bar{m} + \bar{m}S w_t + \bar{w}S m_t + S w_t m_t}{n\bar{w}}$$

$$= \bar{m} \left\{ 1 + \frac{1}{n} S \left(\frac{w_t}{\bar{w}} \cdot \frac{m_t}{\bar{m}} \right) \right\}$$

$$\bar{m}w - \bar{m} = \frac{1}{n\bar{w}} S(w_t m_t) \dots \dots \dots (72)$$

$$e = \frac{S(W_t M_t e_t)}{n\bar{w}\bar{m}_w} + \frac{S(W_t m_t \eta_t)}{n\bar{w}\bar{m}_w}, \text{此爲約略之數, 且須將 } \bar{m} - \bar{m}_w$$

之差刪卻, 如將全式寫出, 則至第二項之分子爲 $S\{W_t(M_t - \bar{m}_w)\eta_t\}$ 。

設 $\bar{\sigma}$, σ_t , σ_t' 爲 e , e_t , η_t 之標準差。

$$\text{於是 } \bar{\sigma}^2 = \frac{1}{(n\bar{w}\bar{m}_w)} \{S(W_t M_t)^2 + S\{W_t(M_t - \bar{m}_w)\sigma_t'\}^2\} \quad (73)$$

設 $\sigma_1 = \sigma_2 = \dots = \sigma$, 且 $\sigma'_1 = \sigma'_2 = \dots = \sigma'$, 或設 σ^2 , σ'^2 , 爲加權平均數, 使 $\sigma^2 S(W_t M_t)^2 = S(W_t M_t \sigma_t)^2$, 且使 $\sigma'^2 S(W_t m_t)^2 = S\{W_t(M_t - \bar{m}_w)\sigma_t\}^2$ 。於是

$$\bar{\sigma}^2 = \frac{1}{(n\bar{w}\bar{m}_w)^2} \{ \sigma^2 \cdot S(W_t M_t)^2 + \sigma'^2 S\{W_t(M_t - \bar{m}_w)\}^2 \} \dots (74)$$

無論何種差誤, 凡在測量之環境下, 有發生或然差誤或可能差誤時, 均宜將其標準差估計算出來。

其餘有關數量, 可由觀察算得。在平常情形下, 求得此一得數之良好近似值, 乃爲

$$\bar{\sigma}^2 = \frac{\sigma^2}{n} \left(1 + \frac{\sigma_w^2}{\bar{w}^2} \right) \left(1 + \frac{\sigma_m^2}{\bar{m}^2} \right) + \frac{\sigma^2}{n} \left(1 + \frac{\sigma_w^2}{\bar{w}^2} \right) \frac{\sigma_m^2}{\bar{m}^2} \dots \dots (75)$$

此近似值, 求法如下:—

$$S(W_t M_t)^2 = S\{(\bar{w}^2 + 2\bar{w}w_t + w_t^2)(\bar{m}^2 + 2\bar{m}m_t + m_t^2)\}$$

$$= n\bar{w}^2\bar{m}^2 + n\bar{w}^2\sigma_m^2 + nm^2\sigma_w^2 + n\sigma_w^2\sigma_m^2$$

$$+ S w_t^2(m_t^2 - \sigma_m^2) + 4\bar{n}\bar{w}S w_t m_t + 2\bar{w}S w_t m_t^2 + 2\bar{m}S m_t w_t^2$$

$$\therefore \frac{S(W_t M_t)^2}{n(\bar{w}\bar{m})^2} = \left(1 + \frac{\sigma_w^2}{\bar{w}^2}\right) \left(1 + \frac{\sigma_m^2}{\bar{m}^2}\right) + \frac{\sigma_w^2 \sigma_m^2 R_{22}}{\bar{w}^2 \bar{m}^2} + \frac{4\sigma_w \sigma_m r}{\bar{w}\bar{m}} + \frac{2\sigma_w \sigma_m^2 r_{12}}{\bar{w}\bar{m}^2} + \frac{2\sigma_w \sigma_m r_{12}}{\bar{w}\bar{m}}$$

式中

$$r = \frac{Swm}{n\sigma_w \sigma_m}, r_{12} = \frac{Swm^2}{n\sigma_w \sigma_m^2}, r_{21} = \frac{Sw^2 m}{n\sigma_w \sigma_m}, R_{22} = \frac{Sw^2 m^2}{n\sigma_w \sigma_m^2} - 1 = \frac{Sw^2(m^2 - \sigma_m^2)}{n\sigma_w \sigma_m^2}$$

$$\begin{aligned} S\{W_t(M_t - \bar{m}_w)\}^2 &= S\{W_t(m_t + \bar{m} - \bar{m}_w)\}^2 \\ &= SW_t^2 m_t^2 + 2(\bar{m} - \bar{m}_w) SW_t^2 m_t + (\bar{m} - \bar{m}_w)^2 SW_t^2 \\ &= \bar{w}^2 \cdot n\sigma_m^2 + 2\bar{w} Sw_t m_t + Sw_t^2 m_t^2 - \frac{2}{\bar{w}} Sw_t m_t (2\bar{w} Sw_t m_t + Sw_t^2 m_t) \\ &\quad + \left(\frac{Sw_t m_t}{n\bar{w}}\right)^2 n(\bar{w}^2 + \bar{w}^2) \end{aligned}$$

$$\begin{aligned} \therefore \frac{S\{W_t(M_t - \bar{m}_w)\}^2}{n(\bar{w}\bar{m})^2} &= \frac{\sigma_m^2}{\bar{m}^2} + \frac{2\sigma_w \sigma_m^2}{\bar{w}^2 \bar{m}^2} \cdot r_{12} \\ &\quad + \frac{\sigma_w^2 \sigma_m^2}{\bar{w}^2 \bar{m}^2} (R_{22} + 1) - 4 \frac{\sigma_w^2 \sigma_m^2}{\bar{w}^2 \bar{m}^2} r^2 - 2 \frac{\sigma_w^3 \sigma_m^2}{\bar{w}^3 \bar{m}^2} r \cdot r_{21} + \frac{\sigma_w^2 \sigma_m^2}{\bar{w}^2 \bar{m}^2} \left(1 + \frac{\sigma_m^2}{\bar{w}^2}\right) r^2 \\ \therefore \bar{\sigma}^2 &= \frac{1}{n} \left(\frac{\bar{m}}{\bar{m}_w}\right)^2 (l_1^2 \sigma^2 + l_2^2 \sigma'^2) \end{aligned}$$

式中

$$\begin{aligned} l_1 &= \left(1 + \frac{\sigma_w^2}{\bar{w}^2}\right) \left(1 + \frac{\sigma_m^2}{\bar{m}^2}\right) + 4 \frac{\sigma_w}{\bar{w}} \cdot \frac{\sigma_m}{\bar{m}} r + 2 \cdot \frac{\sigma_w}{\bar{w}} \cdot \frac{\sigma_m^2}{\bar{m}^2} r_{12} + 2 \frac{\sigma_w^2}{\bar{w}^2} \cdot \frac{\sigma_m}{\bar{m}} r_{21} + \frac{\sigma_w^2}{\bar{w}^2} \cdot \frac{\sigma_m^2}{\bar{m}^2} R_{22} \\ 2^2 &= \frac{\sigma_m^2}{\bar{m}^2} \left\{ \left(1 + \frac{\sigma_w^2}{\bar{w}^2}\right) + \left(\frac{\sigma_w^4}{\bar{w}^4} - 3 \frac{\sigma_w^2}{\bar{w}^2}\right) r^2 + 2 \frac{\sigma_w}{\bar{w}} r_{12} - 2 \frac{\sigma_w^3}{\bar{w}^3} r \cdot r_{21} + \frac{\sigma_w^2}{\bar{w}^2} R_{22} \right\} \end{aligned}$$

如此， r ， r_{12} ， r_{21} ， R_{22} 等，各於其公式內包含 m_t ， w_t 或 $m_t^2 - \sigma_m^2$ 各因子，各因子之總和等於零，故除其他因子 (m_t^2 ， w_t^2) 之巨值，特別為正數或為負數外，此等因子乘積之和，其值必小，而且含此等數之項與其他各項相較，亦有變為甚小之傾向，並且 $\frac{\bar{m}\bar{w}}{\bar{m}} = 1 + r \frac{\sigma_w \sigma_m}{\bar{w}\bar{m}}$ 。

如將 r ， r_{12} ， r_{21} ， R_{22} 全行刪掉，則得近似值如上。

第六節 例證(一)

茲舉數例，詳細推敲，以明有關量數之相對量數 (relative magnitude)。

(1) 第一例計算工資，所加權數，頗為粗略，即以男女及各級年齡之工人人數作為權數，以求男工一種之平均工資，此種重大錯誤，在極不完善之調查中始有之。

量數觀察易生差誤之部分，如以 σ 為代表，在求近似值之公式中，有因子二，均常大於 1，而一般又小於 2；此二因子可由觀察中算出。

在另一方面，加權中易生差誤之部分，在公式(75)中，以 σ' 代表，在求近似值，或求完全值之兩種公式中，均含有因子 $\left(\frac{\sigma_m}{\bar{m}}\right)^2$ ，即數量之標準差，對其平均值之比率之平方也。論及加權平均數時，有一極為普通之情形，即此一比率甚小時，權數中差誤之影響尤小，有時較諸數量中同等差誤之影響，所小甚多。是故，在平常情況下，對數量之確度，應較對權數之確度，必多加注意方可。

最後，關於加權平均數，第一章第四節之第三表，可用以測量當 n 大時離中差大於 $\bar{\sigma}$, $2\bar{\sigma}$, $3\bar{\sigma}$, ………之機率，至 n 小至等於 20 時，該表亦可用查近似值也。

(工 程 及 造 船 業 除 外)

一 九 〇 六 年

門 類	僱 工 人 數	男 工 平 均 收 入	
	W 以 千 為 單 位	M	
		先 令	便 士
生 鐵 業	14	34	4
鋼 鐵 業	54	39	1
馬 口 鐵 業	11	42	0
鐵 道 車 輛 業	46	30	9
鑄 鐵 業	12	31	4
電 氣 器 材 業	15	34	7
金 屬 線 業	8	35	7
銅 業	8	31	9
金 銀 業	8	36	6
寶 石 業	3	38	0
利 刃 (刀 剪) 業	3	31	2
歛 鑄 業	8	31	5
自 轉 車 業	7	34	4
水 管 業	7	28	3
製 釘 業	5	31	0
牀 架 業	2	36	3
蹄 鐵 業	2	27	9
科 學 用 具 業	2	36	10
針 業	2	31	9
鏈 環 業	1	35	4
鎖 業	1	28	0
鐘 表 業	1	32	7
活 字 鑄 造 業	1	33	3
雜 類	45	32	5
總 計	266		

行業數， $n=24$ 。S. W. = 266。 $\bar{w} = \frac{1}{n}$ S. W. = $11\frac{1}{12}$ 。

二十四項收入之算術平均數， $\bar{m} = 33$ 先令。 $6\frac{1}{2}$ 便士 = 33.511先令。

$\sigma_m = 3.47$ 。 $\sigma_w = 14.74$ 。 m 及 w 為各項對 \bar{m} 及 \bar{w} 之離中差。

$$r = \frac{Swm}{n\sigma_w\sigma_m} = .150, \quad r_{12} = \frac{Swm^2}{n\sigma_w\sigma_m} = .095, \quad r_{21} = \frac{Sw^2m}{n\sigma_w^2\sigma_m} = .280,$$

$$R_{22} = \frac{Sm^2w^2}{n\sigma_m^2\sigma_w^2} - 1 = .264, \quad \left(\frac{\sigma_m}{\bar{m}}\right)^2 = .011, \quad \left(\frac{\sigma_w}{\bar{w}}\right)^2 = 1.77,$$

$$\frac{\sigma_m}{\bar{m}} = .104, \quad \frac{\sigma_w}{\bar{w}} = 1.33.$$

以各行業之人數為權數，則收入平均數， $\bar{m}_w = 34$ 先令 $2\frac{1}{2}$ 便士， $= \bar{m}\left(1 + \frac{r\sigma_m\sigma_w}{\bar{m}\bar{w}}\right)$ ； $\therefore \left(\frac{\bar{m}}{\bar{m}_w}\right)^2 = .959$ 。

現依前節求近似值（小體字）之標號，

$$\begin{aligned} \therefore I_1^2 &= 2.77 \times 1.011 + 4 \times 1.33 \times .104 \times .150 + 2 \times 1.33 \\ &\quad \times .011 \times .095 + 2 \times 1.77 \times .103 \times .280 + 1.77 \times .011 \\ &\quad \times .264 = 2.80 + .083 + .003 + .102 + .005 = 2.99. \end{aligned}$$

$I_1^2 \times \left(\frac{\bar{m}}{\bar{m}_w}\right)^2 = 2.87$ 。依求近似值公式，得2.80。

$$\begin{aligned} I_2^2 &= .011\{2.77 + (3.13 - 3.99) \times .0225 + 2 \times 1.33 \times .095 \\ &\quad - 2 \times 2.35 \times .150 \times .280 + 1.77 \times .264\} \\ &= .011\{2.77 - .020 + .253 - .198 + .467\} = .011(2.77 \\ &\quad + .50) = .036. \end{aligned}$$

$l_2^2 \times \left(\frac{\bar{m}}{\bar{m}_w}\right)^2 = .035$ 。依求近似值公式，得 $\cdot 031$ 。

$$\bar{\sigma}^2 = \frac{1}{24}(2.87\sigma^2 + .035\sigma^{12})。$$

各業男工平均收入，或許在33先令中有六便士之差誤，在此情形之下， $\sigma = \frac{1}{66}$ ， $\sigma^2 = .00023$ 。

至權數上之差誤，因既以男女工人總數為權數，而非男工人數，於理顯有不合，故必甚大。

其差誤根據原來報告計算，係為 $\cdot 23$ ，故 $\sigma^2 = .053$ 。

由此， $\bar{\sigma}^2 = .000027 + .000077 = .000104$ 。 $\bar{\sigma} = .01$ 。故平均數為 $\bar{m}_w(1 \pm \bar{\sigma})$ ，換言之，即34先令 $2\frac{1}{2}$ 便士 ± 4 便士。

在此極端情況下，各個權數中之差誤，雖當數量中差誤十五倍之多，但計算結果，差誤僅為 $\cdot 0088$ ，而數量差誤亦有 $\cdot 0052$ 。

(2) 加權平均最重要之用途，莫過於物價指數。

前編第九章嘗言，掉換基年，無異掉換權數；但今日，依本章所述原理，設必要條件不致變更，結果受影響固甚微也。

取孫巴克氏之一九〇〇及一九一一各年商品物價指數，再改用一九〇〇年為基期。例如英麥價格，以一八六七至一八七七年為一〇〇，則一九〇〇年為四九，一九一一年為五八。如以一九〇〇年為一〇〇，則一九一一年為一一八。數目共有四十五個，其算術平均數為107.82，但孫巴克所得平均指數，在一九〇〇為

75.07, 在一九一一年為79.69, 成100:106.16之比。

如用簡單之平均數, 當所有各數均為100時, 則對於比率之加權, 實際乃全相等, 惟斯氏辦法, 如一九〇〇年各項指數為 p_1, p_2, \dots , 一九一一年者為 p'_1, p'_2, \dots , 則總指數為 $I_1 = \frac{p_1 + p_2 + \dots}{45}$, $I_2 = \frac{p'_1 + p'_2 + \dots}{45}$, 而 $I = 100 \frac{I_2}{I_1}$, 得出一九〇〇至一九一一年之變動情形, 即100變到106.16。又

$$I = 100 \frac{\sum p'}{\sum p} = 100 \frac{\sum p \cdot p'}{\sum p^2}$$

詳言之, 即各項變動情形之比率, 用一九〇〇年各項指數作為權數也。

現請就孫氏指數, 研究其平均數之確度。以 p 寫作 w , 是為權數, 以 $\frac{p'}{p}$ 為 m , 是為量數。茲將有關數量列下:—

$$\bar{w} = 75.07, \quad \sigma_w = 20.67, \quad \frac{\sigma_w}{\bar{w}} = .275,$$

$$\bar{m} = 107.82, \quad \sigma_m = 20.03, \quad \frac{\sigma_m}{\bar{m}} = .186,$$

$$r = -.2944, \quad \bar{m}w = 106.2, \quad r_{12} = .506,$$

$$r_{21} = .347, \quad R_{22} = .936.$$

如依公式(75), 將 $r, r_{12}, r_{21}, R_{22}$ 略去, 則

$$\sigma^2 = \frac{\sigma^2}{45} (1.076)(1.035) + \frac{\sigma_1^2}{45} (1.076 \times (.186)^2) = \sigma^2 \times .025 + \sigma^2, \\ \times .00083.$$

如包括在內，則 $\bar{\sigma}^2 = \sigma^2 \times .024 + \sigma'^2 + .0012$ 。

二者之差，幾全由於 r_{12} ，亦即由於平均 w_m^2 ；自一九〇〇至一九一一年反常的增加（用 m 測量），可由自基期一八六七至一八七七年之反常變動（以 w 測量之）中看出；但此點雖有關係，然謂有絕大影響，則恐未必。

m 之差誤， σ ，幾全因用整數(round number)之故，其值漸有約達於 $\frac{1}{300}$ 之趨勢，故

$$\sigma^2 \times .024 = (.0005)^2,$$

如此之微，雖省略亦無傷也。

w 之差誤，如商品輕重，有一定標準，則亦可以算出。非然者，即假定本應加同等權數，如上交替演算之例是也。於是 $\frac{\sigma w}{w} = .275$ ，可將實在權數距假定真權數之離散度求出，而 $\sigma'^2 \times .0012 = (.275 \times .034)^2 = (.0093)^2$ 。

是以 $\sigma^2 = (.0005)^2 + (.0093)^2 = (.0093)^2$ (約略數)，至於指數可書如下式：

$$106.2(1 \pm .0093) = 106.2 \pm 1,$$

由此式可示吾人用指數時，應有之伸縮範圍。

實際上，根據兩假定算出指數之差，為

$$107.8 - 106.2 = 1.6。$$

第七節 平均數之比較

如兩種調查之差誤，彼此完全獨立，且其平均數成爲 $A_1(1 \pm \sigma_1)$ ， $A_2(1 \pm \sigma_2)$ 形式，則 A_1/A_2 之標準差，根據公式(66)，必爲 $\sqrt{(\sigma_1^2 + \sigma_2^2)}$ 。

但兩期各項之差誤，時常顯現同一意義（或全爲正或全爲負）；例如某級工人兩期之工資均低估者是。遇此情形，欲消滅差誤，則惟比較方法尙矣。

茲取一簡單商數 $Q = F_1 \div F_2$ 爲例。

設 e_1 及 e_2 爲 F_1 及 F_2 中之相對差誤，其標準差爲 σ_1, σ_2 ，則 Q 中差誤之標準差，依公式(66)之規定，當爲 $\sqrt{(\sigma_1^2 + \sigma_2^2)}$ 。

惟如 $d = e_1 - e_2$ ，則平均值 $d^2 =$ 平均值 $e_1^2 +$ 平均值 $e_2^2 - 2$ 平均值 e_1e_2 。苟 e_2 之各值，與 e_1 之各值，出現機會完全均等，上式之末一項始能消滅，如 e_1 及 e_2 或者以同一符號而出現，則末一項不能消滅。

例：如 e_2 永等於 $\frac{1}{2}e_1$ ，則 $\sigma_2^2 = \frac{1}{4}\sigma_1^2$ ，平均值 $e_1e_2 = \frac{1}{2}$ 平均值 $e_1^2 = \frac{1}{2}\sigma_1^2$ ，且平均值 $d^2 = \sigma_1^2 + \frac{1}{4}\sigma_1^2 - \sigma_1^2$ ，而此比率之標準差，必爲 $\frac{1}{2}\sigma_1$ 。

加權與不加權平均數比率之必要分析，可參閱附錄七及附錄八。

求近似值之公式如下，符號同上節。

如以 s_r 代表二不加權平均數比率之標準差，

$$s_r^2 = \frac{1}{n} \sigma_d^2 \left(1 + \frac{\sigma_m^2}{\bar{m}^2} \right) \dots\dots\dots (76)$$

此式所用之 σ_d ，為 e_t ， e'_t 相減差額之標準差，至 e_t ， e'_t 乃代表在測量二個時期各別數量 M_t ， M'_t 中之差誤。

然若以 s_r 代表二加權平均數比率之標準差，則在某種條件下，其近似值為

$$s_r^2 = \frac{1}{n} \left(1 + \frac{\sigma_u^2}{\bar{a}^2} \right) \left\{ \sigma_d^2 + \left(\frac{\sigma_u}{1 + \bar{u}} \right)^2 (\sigma^2 + \sigma'^2) \right\} \dots\dots\dots (77)$$

式中 σ_d 仍舊， σ 為數量中差誤之標準差， σ' 係權數中差誤之標準差，又

$$M'_t = (1 + \bar{u} + ut) M_t$$

但 $\sum u_i = 0$ ，故 $1 + \bar{u}$ 足以測量數量平均增長率，而 σ_u 為 u 之標準差，可用以測量增長率之擴散度。

如此，假設數量中差誤，在兩期中均有漸至相同之傾向，則公式(77)括弧{ }中之第一項必小，又如數量幾以同率而增長，則其第二項必小。至於標準差，不拘何種情形，必隨 $\frac{1}{\sqrt{n}}$ 而漸減。

在普通多易適合之條件下，即使測量數量時，及實行加權時，原始之差誤甚巨，但亦能於加權平均數比率中求得極高之確度。惟須注意者，計算方法不可隨時變更，以便求得相似差誤及 σ_d 之

低值也。

第八節 例證(二)

英國產業工人平均每週工資變動情況

	1880		1900		M之增加率 $1+\frac{u}{W}$
	人數	工資	人數	工資	
	單位一千人 W	單位先令 M	單位千人 W'	單位先令 M'	
農業					
英格蘭及威爾士...	135	15	120	16.2	1.08
蘇格蘭.....	24	18	20	21.2	1.18
愛爾蘭.....	98	9	86	10.4	1.16
建築業.....	84	27	123	31.0	1.15
印刷業.....	8	31	13	32.9	1.06
造船業.....	7	28.5	13	34.8	1.22
工程業.....	72	25	106	30.5	1.22
煤礦業.....	44	23	75	34.3	1.49
泥水業.....	9	31	11	38.1	1.23
棉業.....	52	16	54	19.5	1.22
紡毛業.....	12	14	12	13.6	.97
毛線業.....	12	14	12	14.4	1.03
煤氣業.....	3	27	8	31.0	1.15
木器業.....	12	23	18	24.8	1.08
	<u>572</u>		<u>671</u>		

各業人數，係根據英格蘭威爾士戶口普查總報告，第三十五表。增加率乃自統計學報(Statistical Journal, 1909, p. 93) 務德(G. H. Wood) 先生一文中轉錄而來。平均工資數，為由各方

材料算得。至各比率之確度較 M 確度，尤當重視。

$$n = 14, \bar{m} = 21.54, \bar{m}' = 25.20, \bar{m}_w = 18.69, \bar{m}'_w = 24.09$$

$$\frac{\sigma_m}{\bar{m}} = .319, \frac{\sigma_m'}{\bar{m}'} = .351, \frac{\sigma_w}{\bar{w}} = 1.00, \frac{\sigma_w'}{\bar{w}'} = .90, \bar{u}' = .160, \sigma_u = .12$$

$$r = -.42, r_{21} = .44, r_{21}' = -.42, R_{22} = .25。$$

不加權平均數之比率爲 $\frac{\bar{m}'}{\bar{m}} = 1.170$ ；加權平均數之比率

$$= \frac{\bar{m}'_w}{\bar{m}_w} = 1.288。$$

務德氏得出之不加權與加權平均數（用各種權數）比率爲

$$\frac{100}{86} = 1.163, \frac{100}{82} = 1.219。$$

$$sr^2 = \frac{1}{14} \left(1 + 1.00^2 \right) \left\{ \sigma_a^2 + \left(\frac{.12}{1.160} \right)^2 (\sigma^2 + \sigma'^2) \right\}$$

$$= .143\sigma_a^2 + .0015(\sigma^2 + \sigma'^2),$$

根據求近似值公式(77)，暨求完全值公式(148)。得出約略數如下：

$$sr^2 = .145\sigma_a^2 + .022\sigma^2 + .0035\sigma'^2 + .016\sigma'a^2,$$

$\sigma'a$ 爲兩期權數中差誤相減餘額之標準差。

該期權數之影響，既已忽略不計，但權數之變動甚爲劇烈，故求近似值公式，對於數量中差誤，自然無效。

欲觀此種差誤之影響如何，設一八八〇年工資中差誤 (σ) 爲 $\frac{1}{20}$ ，權數中差誤 (σ') 爲 $\frac{1}{10}$ ，並設由此差誤相似性，乃使 $\sigma_a = \frac{1}{2}\sigma$ ，而 $\sigma'a = \frac{1}{2}\sigma'$ 。

然則 $sr^2 = .000091 + .000055 + .000035 + .000040 = .00022$ 。

$$sr = .015$$

平均數之比率，可寫作下式：

$$\frac{\bar{m}'w}{\bar{m}w} (1 \pm \sigma_r) = 1.288 \pm .020$$

詳言之，增加百分數，並非為二九，乃在二七至三一之間也。

實際上，基本差誤，或許較現所假定者為大。此數字乃為示演算方法之例，並說明各項之影響耳。如為試驗學理，則此處 $n = 14$ ，為數過狹，不能為密切應用原理之用，如欲正式研究工資一般變動情形，則非將產業部門範圍擴張，確實決定人數及平均工資不可也。

第九節 平均數與平均數間差額之重要

實用統計上時常發生一極端重要之問題，即決定兩相似羣類平均數之差額之問題也。此相差之額，或由於觀察上之差誤——尤其在隨機抽樣所包個體過狹時——抑完全歸因於特質之不同，皆應加以決定。例如，有兩類觀察得來之死亡率，一為千分之14.7，一為千分之14.3，吾人何以處之，將謂第一類之死亡率高耶，抑將認為實有.4相差，而逕將人口妄分兩類耶？

設觀察得來者，較在機會均等之抽選中所期望者，差額甚大，則知此一差額，甚為重要，換言之，現象間之真正差額，關係頗為

重要也。

一般分析方法，不外下述：設有兩類，一類有事物 n_1 個，一類有事物 n_2 個，平均數一為 \bar{x}_1 ，一為 \bar{x}_2 。自大範圍中任意抽出 n_1 及 n_2 個事物，將二者之平均數分別算出，即由此平均數之相減差額，組成頻數曲線，然後計算此曲線之標準差，設已算得之標準差為 σ 。

以 σ 與 $\bar{x}_1 - \bar{x}_2$ 相比。比率大於 3 之機率為 $\cdot 0027$ ，因 $\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz$ 自 3 至 ∞ ，及自 -3 至 $-\infty$ 之積分總和為 $2(\frac{1}{2} - F(3)) = 2(\cdot 5 - \cdot 49865) = \cdot 0027$ （見第三表，第一章第四節）

依同理，比率大於 2 之機率，為 $\cdot 0456$ ，大於 1 之機率為 $\cdot 3174$ ；至其是否大小與 $\cdot 674$ 相等，機會相同，尚在兩可之間。如此，設 $\bar{x}_1 - \bar{x}_2$ 不大於 $\cdot 674\sigma$ ，謂其有真實差額，必無證據可言，換言之，即差額乃由各羣類之性質而來，未可歸因於機率離差也。但當 $\bar{x}_1 - \bar{x}_2$ 大過於此數，則結果之反機率 (improbability) 必增加，至比率等於 2 時，則必以 21 對 1 ($\cdot 9544$ 對 $\cdot 0456$) 而失敗。在至 2σ 時，若非差額果係真實，則可謂無此機會。至 3σ ，則將以 370 對 1 而失敗。此種事件，機會愈益減少，如果有其事，則 $\bar{x}_1 - \bar{x}_2$ 之差，將名之為重要差。而達 4σ 時，則將以 15,000 對 1 而失敗，更無此機會矣。依此方法，欲求得確然不疑之點，至為困難。故非用機率尺度，以明重要 (significant) 一詞不可。茲將計算 σ 之規則，於

下文說明之，因上文所論條件，均已滿足，故在任何情形之下，差誤之頻數羣類，終為常態形勢，且在任何情形之下， σ 與實現機率之關係，終不外乎本段所言也。

例 A. —— 具有特質之事物在範圍中所佔之比例

(1) 設一範圍中共有 N 個事物，就中具有特質之事物為 P 個， p 及 N 為已知數。 $q=1-p$ 。

將 n 個隨機抽選出來，設其中 $p'n$ 個，為具有該種特質者。

$p' \sim p$ 之標準差為 $\sqrt{p' \left(\frac{1}{n} - \frac{1}{N} \right)}$ ，如可將 $\frac{n}{N}$ 略去，則等

於 $\sqrt{\frac{pq}{n}}$ 。

例如——取一骰連擲一千二百次，六點出現有一百八十次，

$$N = \infty, p = \frac{1}{6}, n = 1200, p' = \frac{3}{20}, \sigma = \sqrt{\frac{1}{6} \cdot \frac{5}{6} \cdot \frac{1}{1200}} = .0108,$$

$$\frac{p-p'}{\sigma} = \frac{.0167}{.0108} = 1.6,$$

於此不能不聲明者，骰子之六面，並無有欠一律之證明。

(2) 設從範圍中抽選兩種樣本 (n_1, p_1) (n_2, p_2) ，並將 $\frac{n_1}{N}, \frac{n_2}{N}$ 刪略。

$p_1 - p$ 及 $p_2 - p$ 之標準差為 $\sqrt{\frac{pq}{n_1}}$ 及 $\sqrt{\frac{pq}{n_2}}$ 。

故 $p_1 \sim p_2 = (p_1 - p) \sim (p_2 - p)$ 之標準差，乃各該標準差平方之和之平方根（見公式 34），且

$$= \sqrt{\left\{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right\}} \dots\dots\dots (78)$$

若 p 為不可知之數，但只有從抽樣中得來一途，則須將樣本合併始能得其最優值，如 $p(n_1 + n_2) = p_1n_1 + p_2n_2$ 是。

例如——從一小城市選出 1000 家，其中二百 (n_1) 家，家主為技師，八百 (n_2) 家，家主為勞動者。有學齡兒童者在第一類 ($p_1 = .4$) 中有八十家，在第二類 ($p_2 = .525$) 中有四百二十家 (家數係假定的)。

$$p \times 1000 = 80 + 420 \quad \therefore p = \frac{1}{2} = q$$

$$p_1 \sim p_2 \text{ 之標準差} = \sqrt{\left\{\frac{1}{4}\left(\frac{1}{200} + \frac{1}{800}\right)\right\}} = .04$$

$$\frac{p_2 - p_1}{\sigma} = \frac{.525 - .4}{.04} = 3 \text{ (約略數)}。$$

此差額即為重要。

(3) 從各種不同之未知範圍中，抽出樣本 (n_1p_1), (n_2p_2) 個，而 n_1 及 n_2 均甚大。

例如——從兩國共抽出一千人，其中有藍眼睛者，一國為三百人，一國為二百五十人。

此次抽選， $p_1 = \frac{3}{10}$ ，其 $\sigma = \sqrt{\left(\frac{3 \times 7}{10^5}\right)} = .014$ ，而該國 (設抽選時不問該國種族氣候如何) 全國碧睛人數，約近於 p_1 。同理，另一國家之數，約略為 $p_2 = \frac{1}{4}$ 。

$p_1 - p_2$ 之標準差，兩獨立羣類相減餘額之標準差，即：

$$\sqrt{\sigma_1^2 + \sigma_2^2} = \sqrt{\left(\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}\right)} = 0.02 \dots\dots\dots (79)$$

$$\frac{p_1 - p_2}{\sigma} = \frac{3 - 0.25}{0.02} = 2.5。$$

此法一般於比較兩類職業工人（如礦工及瓦匠）之死亡率時多用之。

設據觀察所知，於 n_1, n_2 中，每年死亡者，有 m_1 及 m_2 人，則死亡率 (r_1, r_2) 爲 $\frac{m_1}{n_1} \times 1000$ 及 $\frac{m_2}{n_2} \times 1000$ ，又因缺乏其他證據，只得假定兩類之危險性完全相同，故

$$p_1 = \frac{m_1}{n_1}, \quad p_2 = \frac{m_2}{n_2}。$$

此種礦工，假定係從所有礦工（之範圍中）任意抽樣而來，瓦匠亦然。則

$$r_1 - r_2 \text{ 之標準差} = 1000 \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}} = 1000 \sqrt{\left\{ \frac{m_1(n_1 - m_1)}{n_1^3} + \frac{m_2(n_2 - m_2)}{n_2^3} \right\}}。$$

此外，尚有較爲簡單之程序，即以各類與整個成年男子全數比較，以查礦工死亡率是否與一般職業者有別，此宜用例 (1) 方法。

以上均假定在一範圍中，終有一律之機率， p 。然一範圍中有各區各層 (strata) 之不同，因而機率亦不能一律，於此乃發生問題：全以該範圍爲整個對象，從中隨機抽樣乎？抑就宇宙之分區

分層，各就同一之比例，以抽樣本乎？關於此點，猶爾氏 (Theory of Statistics, 第二百八十一頁) 曾有一公式，茲介紹於下。

設一範圍在 t 層中有事物 n_1, n_2, \dots, n_t 個，並設具有某種特質者，在 t 層中，為 $p_1 n_1, p_2 n_2, \dots, p_t n_t$ 。

$$N = n_1 + n_2 + \dots + n_t, \text{ 並以 } p_1 n_1 + p_2 n_2 + \dots = P N.$$

設 t 層中，所檢查之事物為 $k n_1, k n_2, \dots, k n_t$ ，總之， $KN = n_0$ 。

以 $p_1 = P + d_1, p_2 = P + d_2, \dots$ ，而 $P = p_1 \frac{n_1}{N} + p_2 \frac{n_2}{N} + \dots$ ，
故 $S(nd) = 0$

樣本中 p_1 之標準差為 $\sqrt{\frac{p_1 q_1}{k n_1}}$ ， p_2 之標準差及其他均以此類推。

如此，設 σ 為樣本中 P 之標準差，則依公式 55，

$$\begin{aligned} \sigma^2 &= \left(\frac{n_1}{N}\right)^2 \frac{p_1 q_1}{k n_1} + \left(\frac{n_2}{N}\right)^2 \frac{p_2 q_2}{k n_2} + \dots = \frac{1}{k N^2} (n_1 p_1 q_1 + n_2 p_2 q_2 + \dots) \\ \therefore k N^2 \sigma^2 &= n_1 p_1 (1 - p_1) + \dots \\ &= S(np) - S(np^2) = NP - S\{n(P+d)^2\} \\ &= NP - NP^2 - 2P \cdot Snd - Snd^2 = NPQ - N\sigma_p^2 \\ \sigma_p^2 &= \frac{1}{N} (n_1 d_1^2 + n_2 d_2^2 + \dots) \\ \therefore \sigma^2 &= \frac{PQ}{n} - \frac{\sigma_p^2}{n} \dots \dots \dots (80) \end{aligned}$$

所謂 σ ，乃觀察結果之標準差， P 為在範圍中所佔之實在比

例，而 σ_p^2 則為各層(strata)之加權均方差(weighted mean square of deviations)。

如從範圍中任意抽取數目，則差誤之標準差，將為 σ_0 ，而

$$\sigma_0^2 = \frac{PQ}{n}。$$

故 $\sigma^2 = \sigma_0^2 - \frac{\sigma_p^2}{n}$ ，且如從各層依比例抽取，所有差誤之標準差，必消滅。

茲欲調查四城鎮之經濟狀況（見生計與貧乏一書），每街每二十戶中，提出一家，而不用將所有各戶排號任意從每二十號中各取其一之辦法。因吾人如此辦理，不致有某區未被代表之弊，而此弊則隨機抽樣乃不可免者也。且用此辦法，上列公式所提之優點，可以完全得到，因各街市之社會情形，多有某種相似之處也。茲設十個相等區域共有一萬六千戶，且設各區在某種標準之下之比例為 $\cdot 02, \cdot 06, \cdot 10 \dots \cdot 38$ 。於是 $N=16000$; $n_1=n_2=\dots=1600$; $p_1=\cdot 02, p_2=\cdot 06 \dots P=\cdot 2$; $d_1=\cdot 18, d_2=\cdot 14 \dots$;

$$\sigma_p^2 = \frac{1}{10}(\cdot 18^2 + \cdot 14^2 + \dots), \sigma_p = \cdot 115$$

茲設每區各擇八十戶而調查之， $n=800, k=\frac{1}{20}, \sigma^2 = \frac{\cdot 2 \times \cdot 8}{800} - \frac{\cdot 0132}{800}, \sigma = \cdot 0136$ ，於是得數乃為 $\cdot 20 \pm \cdot 0136$ ，或百分之 20 ± 1.36 。

以上乃就分層而言，如抽選方法，並不依各層排列抽選，則

應得 $\sigma = .0141$ 。確度之增加，為數無幾，然分層抽取方法，乃基於常識，苟非有何困難，當以用此法為是。

例 B. —— 一範圍中含有若干可量物件之例

(1) 設在一範圍中，有 N 個物件，物件量數之平均為 \bar{x} ，標準差為 S 。

從中任意抽出 n 個，其平均數為 \bar{x}_1 。

則 $\bar{x}_1 - \bar{x}$ 之標準差， σ_1 ，依公式 52，乃為 $S\sqrt{\left(\frac{1}{n} - \frac{1}{N}\right)}$ 。

實例——某城市有住宅 10,000 所，每所平均住人 4.5，標準差為 2。一千所工人住宅之平均數為 4.7。是則 $N = 10,000$ ， $n = 1000$ ， $\bar{x} = 4.5$ ， $\bar{x}_1 = 4.7$ ， $s = 2$ 。

$$\sigma = 2\sqrt{\left(\frac{1}{1000} - \frac{1}{10000}\right)} = .06, \quad \frac{\bar{x}_1 - \bar{x}}{\sigma} = \frac{.2}{.06} = 3.3.$$

(2) 在範圍間，只知其樣本為 n ， \bar{x} ， s 。又由樣本， n ，中抽出小樣本 n_1 ，由 n_1 而得 \bar{x}_1 ， σ_1 。

範圍中自抽出第一次樣本後，設第一次抽出者，並非抽自有某特種平均數之一類，而係隨機抽得，則所餘樣本仍可自一不可知之範圍中獨立隨機抽出，茲設所餘樣本為 n_2 ， \bar{x}_2 ， σ_2 。

於是 $n_1 + n_2 = n$ ， $n_1\bar{x}_1 + n_2\bar{x}_2 = n\bar{x}$ 。

$$\therefore \frac{\bar{x}_1 - \bar{x}_2}{n} = \frac{\bar{x}_1 - \bar{x}}{n_2} = \frac{\bar{x} - \bar{x}_2}{n_1}.$$

自原點算出之動差，則得

$$n(s^2 + \bar{x}^2) = n_1(\sigma_1^2 + \bar{x}_1^2) + n_2(\sigma_2^2 + \bar{x}_2^2)$$

$\bar{x} \sim \bar{x}_1$ 之標準差為 $\sqrt{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)}$ 。

但 $\bar{x} \sim \bar{x}_1$ 對 $\bar{x}_1 \sim \bar{x}_2$ 之比，乃為常數，即等於 $\frac{n_2}{n}$ 。

∴ $\bar{x} \sim \bar{x}_1$ 之標準差為 σ ，而

$$\sigma^2 = \frac{n_2^2}{n^2} \left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \right) = \frac{\sigma_1^2}{n_1} + \frac{s^2 - 2\sigma_1^2}{n} - \frac{n_1}{n n_2} (\bar{x} - \bar{x}_1)^2,$$

將 σ_2 及 x_2 消去，即得出。

· 設 n_1 小於 n_2 ，而 n 及 n^2 為甚大。

則 $(\bar{x} - \bar{x}_1)\sqrt{n}$ 屬於 $\sigma\sqrt{n_1}$ 次，即屬於 σ_1 次。故 $(\bar{x} - \bar{x}_1)^2$ 一項可以略去。

所以
$$\sigma = \sqrt{\left(\frac{\sigma_1^2}{n^2} + \frac{s^2 - 2\sigma_1^2}{n}\right)} \dots\dots\dots (81)$$

(見 Biometrika, 卷五, 第一百八十二頁)。

此法係蘇格蘭人口普查 (第七千一百六十三號第二百八十八頁) 比較各業男子家庭大小所用。

設 $\frac{n_1}{n}$ 小，如按例一所云， $\frac{n}{N}$ 可刪略時， $\sigma = \frac{\sigma_1}{\sqrt{n}}$ ，至 σ_1 係由觀察而得，用以代表不可知之 s 者。

如 $\sigma_1 = s$ (此種情形必須標準差不受分類之影響，而受影響者僅為平均數) 則依吾人所預料，

$$\sigma = \sigma_1 \sqrt{\left(\frac{1}{n_1} - \frac{1}{n}\right)}.$$

實例——蘇格蘭人口普查

結婚總數 $n = 133,960$ 。

\bar{x} , 每對夫婦所生子女數, = 5.82, 其 $s = 3.099$,

鍋匠之生育, $n_1 = 923$, $\bar{x}_1 = 6.00$, $\sigma_1 = 3.069$,

$$\sigma = \sqrt{\left(\frac{9 \cdot 24}{923} + \frac{9 \cdot 60 - 18 \cdot 46}{133,960}\right)} = \cdot 10.$$

$$\frac{\bar{x}_1 - \bar{x}}{\sigma} = \frac{\cdot 18}{\cdot 10} = 1.8, \text{ 差額顯然重要。}$$

實例——第三章第七節最後一例, 美國二百六十七處麵粉價格, 內有沿大西洋北部數州之地方一百四十二處。

	地方數	平均數	標準差
美國全國	$n = 267$	$\bar{x} = 2,625$	$s = \cdot 293$
北大西洋各州	$n_1 = 142$	$\bar{x}_1 = 2,748$	$\sigma_1 = \cdot 244$

$$\sigma = \sqrt{\left(\frac{\cdot 0595}{142} + \frac{\cdot 0858 - \cdot 01190}{267}\right)} = \cdot 017,$$

$$\frac{\bar{x}_1 - \bar{x}}{\sigma} = \frac{\cdot 123}{\cdot 018} = 7 \text{ (約略數)}, \text{ 可知北大西洋各州麵粉價格}$$

確比美國全國之平均為高。

(3) 從兩個未知之範圍 ($N_1\bar{x}'s_1$) 及 ($N_2\bar{x}''s_2$) 中取出兩個樣本 ($n_1\bar{x}_1\sigma_1$) 及 ($n_2\bar{x}_2\sigma_2$)。

$\bar{x}_1 \sim \bar{x}_2$ 之 σ 乃為

$$\sqrt{\left\{s_1^2\left(\frac{1}{n_1} - \frac{1}{N_1}\right) + s_2^2\left(\frac{1}{n_2} - \frac{1}{N_2}\right)\right\}} = \sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)} \text{ (約略值)},$$

此係兩個獨立觀察之差額, 此兩個獨立觀察各由例(1)而來。

若在範圍中, 非用抽樣方法, 則不可知, 且 $\frac{n_1}{N_1}, \frac{n_2}{N_2}$ 之值為小時, 必須用

$$\sigma^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \dots\dots\dots(82)$$

此外尚有他種變量(variant),亦可用此原則解決之。

實例——生活費指數中之食物費用,與第三章第七節所列者相似,在全部羣類中其 $\bar{x} = 10.3$ (先令), $s = 3.3$ 。

有技能工人五百六十六人之家庭, \bar{x}_1 為 10.9 , 無技能工人二百六十六人家庭之 \bar{x}_2 , 為 9.3 。

此兩類之標準差,並未算得,但大概與 s 相近似。

$$\bar{x}_1 - \bar{x}_2 \text{ 之 } \sigma \text{ 為 } 3.3 \sqrt{\left(\frac{1}{566} + \frac{1}{266}\right)} = .25.$$

$$\frac{\bar{x}_1 - \bar{x}_2}{\sigma} = \frac{1.6}{.25} = 6 \text{ (約略數)}, \text{ 差額重要可知。}$$

具有可量事物之範圍之成層排列(stratification),猶爾(Yule: Theory of Statistics 第345頁)氏,亦曾研究,茲介紹於下:

設在一範圍($N\bar{x}$ s)間,包括($n_1\bar{x}_1s_1$), ($n_2\bar{x}_2s_2$)……數羣類,並設 $kn_1, kn_2, \dots\dots$ 自各羣類抽選而來,平均數為 ($\bar{x}_1 + \delta$), ($\bar{x}_2 + \delta_2$)……,至 kN 之平均數為 $\bar{x} + D, kN = n$ 。

$$\text{於是 } N = n_1 + n_2 + \dots\dots; N\bar{x} = n_1\bar{x}_1 + n_2\bar{x}_2 + \dots\dots$$

$$\text{以 } \bar{x}_1 = \bar{x} + d_1, \bar{x}_2 = \bar{x} + d_2, \dots\dots;$$

$$\text{於是 } \sum nd = 0$$

$$N s^2 = n_1(s_1^2 + d_1^2) + n_2(s_2^2 + d_2^2) + \dots\dots$$

$\delta_1, \delta_2, \dots\dots$ 之標準差之平方為

$$\frac{s_1^2}{kn_1}, \frac{s_2^2}{kn_2}, \dots$$

以 σ 為 D 之標準差。

$$kN(\bar{x} + D) = kn_1(\bar{x}_1 + \delta_1) + kn_2(\bar{x}_2 + \delta_2) + \dots$$

$$\therefore D = \frac{n_1}{N}\delta_1 + \frac{n_2}{N}\delta_2 + \dots$$

$$\therefore \sigma^2 = \left(\frac{n_1}{N}\right)^2 \frac{s_1^2}{kn_1} + \left(\frac{n_2}{N}\right)^2 \frac{s_2^2}{kn_2} + \dots$$

依公式55,

$$= \frac{1}{kN^2} (n_1 s_1^2 + n_2 s_2^2 + \dots)$$

假為 n 個樣本，係從整個範圍中隨機抽選出來，以 σ_0 為平均數之標準差。

$$\text{於是 } \sigma_0^2 = \frac{s^2}{n} = \frac{1}{nN} \{n_1(s_1^2 + d_1^2) + n_2(s_2^2 + d_2^2) + \dots\}$$

$$\therefore \sigma^2 = \sigma_0^2 - \frac{1}{n} \cdot \frac{S(nd^2)}{N}$$

以 $N\sigma_m^2 = S(nd^2)$ ，如是則 σ_m^2 即為該層平均數之加權均方差。

$$\text{於是 } \sigma^2 = \sigma_0^2 - \frac{\sigma_m^2}{n} \dots \dots \dots (83)$$

平均數之確度，公式80所不能精確者，茲用成層排列法可以推進矣。

例如本章第二節之例，以 N 為最末七條街（自斯皮特菲爾街以上）之住宅總數。在此七條街每家住宅之平均住人數， \bar{x} ，在普查報告所得為 4.64， s 為 2,750 而住宅抽樣，係五十取一，

故 $k = \frac{1}{50}$; $N = 57000$, $n = 1140$ 。

	住宅數	每宅人數
		\bar{x}
斯皮特菲爾街	$n_1 = 6500$	$+ .15 = d_1$
懷德加白爾街	$n_2 = 59$	$+ .08 = d_2$
聖喬治街	$n_3 = 94$	$+ .24 = d_3$
色得外爾街	$n_4 = 48$	$- .27 = d_4$
萊母好斯街	$n_5 = 66$	$- .10 = d_5$
邁爾恩德街 西南區	$n_6 = 134$	$+ .01 = d_6$
邁爾恩德 東北區	$n_7 = 104$	$- .24 = d_7$
	570	

$$S(nd^2) = 1806 \quad \sigma m^2 = \frac{1806}{57000} = .0317$$

$$\sigma_0^2 = \frac{(2.75)^2}{1140} = .006634 \quad \sigma_0 = .08145$$

$$\sigma^2 = .006634 - \frac{.0317}{1140} = .006606 \quad \sigma = .08128$$

以各區爲層，從七層中，用抽樣方法所得之數，雖略有進步，但厥值甚微。

第十節 趨勢之存在

上述原理，如進一步應用之，可作關於時間數列之研究，藉以檢視變動及運動是否漫無趨向，換言之，即考查趨勢 (trend) 或週期 (periodicity) 之存在與否也。

此種方法及其難點，舉二例以明之：

(1) 一八五〇至一八九九年賽馬時間記錄如后

分,秒		分,秒		分,秒		分,秒		分,秒	
1850	2 56	1860	2 56	1870	2 52	1880	2 49	1890	2 40 $\frac{1}{2}$
1851	2 52	1861	2 44	1871	2 51	1881	2 46	1891	2 54 $\frac{3}{4}$
1852	3 0	1862	2 49	1872	2 52	1882	2 49	1892	2 43 $\frac{1}{2}$
1853	2 52	1863	2 54	1873	2 50 $\frac{3}{4}$	1883	2 53	1893	2 44 $\frac{1}{2}$
1854	3 0	1864	2 47	1874	2 43 $\frac{1}{2}$	1884	2 49	1894	2 50
1855	2 58	1865	2 51	1875	2 49 $\frac{1}{2}$	1885	2 43 $\frac{2}{3}$	1895	2 48 $\frac{2}{3}$
1856	3 4	1866	2 53	1876	2 50	1886	2 54 $\frac{2}{3}$	1896	2 45 $\frac{2}{3}$
1857	2 50	1867	2 54	1877	2 54 $\frac{1}{2}$	1887	2 50 $\frac{1}{3}$	1897	2 45
1858	2 53 $\frac{1}{2}$	1868	2 47 $\frac{1}{2}$	1878	2 54	1888	2 42 $\frac{2}{3}$	1898	2 45 $\frac{1}{3}$
1859	2 55	1869	2 59	1879	3 2	1889	2 45	1899	2 44
十年平均數2	56.05	2	51.45	252.395		2	48.22	2	46.26

以上數字恰配合常態曲線，其平均數為 2 分 50.87 秒，標準差為 5.20 秒。兩記錄差額之標準差，乃為 $5.2\sqrt{2} = 7.4$ 秒。超過此額者，數年間只有十一次，而差額絕無超出兩倍此數者；足見每兩次賽馬之間，並無猛烈變動之事。但先年與最末幾年，時間相差，有時卻達 20 秒。如以十年為一期，則兩期平均數相差額之標準差，為 $52\sqrt{\left(\frac{1}{10} + \frac{1}{10}\right)} = 2.33$ 秒。一八五〇至一八五九年之平均數及一八九〇至一八九九年之平均數，相差將及十秒之多，與一八五〇至一八五九年及一八八〇至一八八九年之平均數之差額相同，差額均甚為重要。至中間數年之相差，尚不為巨。由此可知自一八五〇至一八八〇年之間，必有幾種原因，乃使賽馬逐漸加快也。

(2) 英格蘭威爾士一八六〇至一九〇九年間之結婚率

1860	17.1	1870	16.1	1880	14.9	1890	15.5	1900	16.0
1861	16.3	1871	16.7	1881	15.1	1891	15.6	1901	15.9
1862	16.1	1872	17.4	1882	15.5	1892	15.4	1902	15.9
1863	16.8	1873	17.6	1883	15.5	1893	14.7	1903	15.7
1864	17.2	1874	17.0	1884	15.1	1894	15.0	1904	15.3
1865	17.5	1875	16.7	1885	14.5	1895	15.0	1905	15.3
1866	17.5	1876	16.5	1886	14.2	1896	15.7	1906	15.7
1867	16.5	1877	15.7	1887	14.4	1897	16.0	1907	15.9
1868	16.1	1878	15.2	1888	14.4	1898	16.2	1908	15.1
1869	15.9	1879	14.4	1889	15.0	1899	16.5	1909	14.7
十年平均數	16.70	16.33	14.86	15.56	15.55				

此五十年間平均數為15.80，五十個記錄之標準差為.89，且如不計次序，幾成常態形式。各年間，並無相差甚鉅之情形。取十年記錄，作一平均，兩平均數間差額之標準差為0.4，可知一八七〇至一八七九至一八八〇至一八八九年之下降，及一八八九後之上升，甚為重要。

。第一個二十五年，較第二個二十五年間之變動為劇，茲作更精密之測驗如下：

1860-1884 $\sigma = .894$ $\sigma\sqrt{\frac{1}{5}} = .566$		1885-1909 $\sigma = .613$ $\sigma\sqrt{\frac{1}{5}} = .388$	
	平均數		平均數
1860-4	16.70	1885-9	14.50
1865-9	16.70	1890-4	15.24
1870-4	16.96	1895-9	15.88
1875-9	15.70	1900-4	15.76
1880-4	15.22	1905-9	15.34

由此可見自一八七〇至一八七四至一八八五至一八八九年之低降，及自一八八五至一八八九至一八九五至一八九九年之上升，均頗有重要之含義。

論辯必須以一圖式解明之，蓋惟有圖式乃表明在何期間始可施用測驗也。

第十一節 週期性

關於一長度未定之週期是否存在之一般問題，乃係數學上之問題，即係倒數分析(harmonic analysis)上之問題，非現在所可研究；但如週期長度已知，則不妨從事測驗週期性之影響也。

取一既知距度(interval)為例，譬有按月之記錄，共一年，試查正月份二月份之相差，是否與隨機觀察出現（不計時間先後）者相同。茲設記錄每年十二次積至 t 年之久，則應共有記錄 $12 \times t$ 之多，又設其平均數為 \bar{x} ，如以 x 代表任何觀察，則標準差（依平均數算來） $\sigma = \sqrt{\left(\frac{\sum (x - \bar{x})^2}{12t}\right)}$ 。

如有兩類記錄，各有 t 個，均係隨機抽來，則各類平均數之相互差額之標準差，為

$$\sqrt{\left\{\frac{\sigma^2}{t} + \frac{\sigma^2}{t}\right\}} = \sigma \sqrt{\frac{2}{t}}$$

如此記錄作為一羣類而幾近常態形式，或與第三章第五節所舉條件相符，則超過上列標準差之機率，即可由第二章第四節（第三表）常態機率表查出。

反之，如不採用隨機之抽取，而以 t 年之正月份記錄之平均數，與 t 年二月份記錄之平均數相比，則見兩平均數相減之差額，

超過 $\sigma\sqrt{\frac{2}{t}}$ 有兩三倍之多，正二月比較如此，其他各月亦然；於是吾人可以證明，若非某月有極端反常之記錄，即所量數量受一年時間之影響也。

雖然，此一方法不易盡行包括所有之證據。例如查閱前編第七章第四節之一百八十個失業記錄，可知平均數為 4.269，標準差為 1.924。以 15 個作一平均，則兩平均數差額之標準差為 $1.924\sqrt{\frac{2}{15}} = .70$ 。以正月或十二月份之平均數，與四月；五月，六月，或七月之各該平均數比較，雖超過 .70，但差額尚不算大；即隨機抽樣得來者，亦不過如此。但此外尚有累積證據 (cumulative evidence)，甚難加以測量。例如，自十二月經由正月二月三月（如將一九一二年之反常記錄取消）四月以至五月之平均數之低落，以及自五月以後至十月，按月之平均數之上升是。此乃波浪形運動 (wave motion)，非此法所能測量者也。

另有一法，係比較某月對其下月之下降上升數，然欲求其精確，卻亦甚難。

	正月 至	二月 至	三月 至	四月 至	五月 至	六月 至	七月 至	八月 至	九月 至	十月 至	十一月 至	十二月 至
	二月	三月	四月	五月	六月	七月	八月	九月	十月	十一月	十二月	正月
低落數	12	13	10	10	5½	7	2	7½	8½	10	15	10
上升數	3	2	5	5	9½	8	13	7½	6½	5	0	4

由此觀之，二月數少於前一正月者，有十二年，多於前一正月者有三年。如數相同，則使兩行各佔二分之一。在此十五次試

驗中，各次之正負，機會一律均等，取得 10 以上而具同符號之機率約為 $\frac{1}{8}$ ，足見三月至四月，四月至五月，十月至十一月之運動，即在隨機抽樣中，亦不見得即無實現之機會。但 12 以上具有同號者，機率僅有 $\frac{1}{28}$ ，則自正月至二月，二月至三月，七月至八月，十一月至十二月，如不受季節影響，該種運動實現必甚難。

討論至此，似已得一結論：自十一月至三月，四月或五月，乃有累積下降勢，而在初夏之季，則有累積增高勢。

茲為求更確定之結果起見，再舉一例如下：十八年間捕獲鱈魚每月報告 (Granton, 北海漁業調查所)，單位為每月每船一英擔 (hundred weight) 數。記錄共有二百一十六件，其平均數為 172，標準差約為 108。

取一月之平均數，與全部之平均數相比，二者差額之標準差為 $108 \sqrt{\left(\frac{1}{18} + \frac{1}{216}\right)}$ ，即約為 26.5，而一月之平均數與另一月之平均數，相減差額之標準差，為 $108 \sqrt{\frac{2}{18}} = 36$ ，以上兩種試驗，均以隨機抽選及無季節影響為前提。

記錄上之平均數

正月	101	四月	83	七月	247	十月	227
二月	115	五月	145	八月	282	十一月	181
三月	125	六月	196	九月	267	十二月	101
						一年	172

上表正月二月四月及十二月，低於平均數，多於 27 之兩倍有餘，三月及五月低於平均數，只比一 27 為多，七月至十月每月大

於平均數約比兩倍 27 猶多，六月及十一月大於平均數只在—27 之內。於此得一結論；七月至十月一季，比十二月至四月之季為佳。

再看上下月間變動，在下列時節，均多於36，三月至四月，四月至五月，五月至六月，六月至七月，九月至十月，十月至十一月，及十一月至十二月。四月顯然為最惡劣之月份；但八月是否即為最佳，斯時尚屬疑問。

試從原始材料中，得

	一月至二月	二月至三月	三月至四月	四月至五月	五月至六月	六月至七月	七月至八月	八月至九月	九月至十月	十月至十一月	十一月至十二月	十二月至一月
下降數	5½	10	16	4	4	7	4	11	11	11	15	8
上升數	12½	8	2	14	14	11	14	7	7	7	3	9

由此可知多於11或少於7之數，頗為重要。

附註：——

(1)因以未知真實平均數為中心算得之離中差，與由觀察之平均數算出者不同，

故平均數之標準差時常寫作 $\frac{\sigma}{\sqrt{n-1}}$ ，而不用第三章第一節之 $\frac{\sigma}{\sqrt{n}}$ 。

設以 \bar{x}_0 為一羣類之真實平均數，其標準差為 σ_0 ，又設從中抽出 n 個事物，而得平均數 \bar{x} ，至 n 個事物各別分為 X_1, X_2, \dots 而其標準差為 σ 。

$$\text{以 } \bar{x} = \bar{x}_0 + d$$

X_1, X_2, \dots 對真實平均數之離中差，為 $X_1 - \bar{x}_0, X_2 - \bar{x}_0, \dots$ ，其標準差依假設當為 σ_0 。

$$\text{故 } \sigma_0^2 = \text{平均}(X_1 - \bar{x}_0)^2 = \text{平均}(X_1 - \bar{x} + d)^2 = \text{平均}(X_1 - \bar{x})^2 + d^2$$

因依公式 (38)，此平均數之標準差為 $\frac{\sigma_0}{\sqrt{n}}$ ，

$$\text{故} \quad = \sigma^2 + \frac{\sigma_0^2}{n}$$

$$\therefore \sigma_0^2 = \frac{n}{n-1} \sigma^2$$

$$\text{而且} \quad \frac{\sigma_0}{\sqrt{n}} = \frac{\sigma}{\sqrt{n-1}} = \sqrt{\left\{ \frac{\sum(X-\bar{X})^2}{n(n-1)} \right\}}$$

此處 σ 須以 $\sqrt{n-1}$ 除而非以 \sqrt{n} 除。

此種改正只有原理上的重要性，因其差額只於 n 為極小時方能察覺，而且 σ 與此差額，無論何時，均易陷于同次之差誤。

(2) 前論及時間數列時，曾言以 t 年之平均數為中心，測量時間數列之觀察離中差，必須注意平均時有無陷於差誤之危險（此差誤用 $\frac{\sigma}{\sqrt{t}}$ 測量之，而 σ 即為鄰近各年份之觀察數值之標準差）。觀察與此一平均數之差額之標準差，非為 σ ，乃係 $\sqrt{\left(\sigma^2 + \frac{\sigma^2}{t}\right)} = \sigma \sqrt{\frac{t+1}{t}}$ 。因 t 小，差誤易於察覺，前編第七章第三節，結婚率與平均每人擔負進出口貨值之圖，所示之離中差，估計頗欠完善，於此可知，本編第七章第一節第六實例，所算相關，亦欠精確。

(註一) 此種抽選，未能完全滿足獨立條件，參閱 *Statistical Journal*, 1912-13, p. 702 nixon。

(註二) 在第一編第八章，即以上文之 $\frac{\sigma' - \sigma}{x}$ 作差誤，此差誤知名之為 e_1 ，則 e 與 e_1 二者之關係為 $e_1 = -e + e^2 - e^3 \dots$ ，但普通大都假定， e^2 值甚微，不妨略去，故 $e_1 = -e$ 。

第五章 經驗頻數方程

所有頻數羣類，未必全能用大數律表現之。蓋方程式背後之特種繁複獨立原因，未可一概認為適用於所有之觀察羣類也。常態曲線之主要用途，為對於平均數及其他函數（其作成方法，業已知曉）之應用。至於對於人體測量的（anthropometrical）或生物測量的（Biometrical）羣類之應用，則須按量數類別，逐一加以檢證，而心理上及倫理上之性質，是否依常態形式而分佈，則須作專門之調查。雖然，吾人可得而認定者，在極多類別中，各類之中心部分（自中心計上下各一二標準差之間），亦頗合常態分配之形式，且觀察去平均數二標準差以上之機率，並不算大，故常態頻數表即在非常態情形下，亦足供吾人之參考。

為求羣類之完全敘述，或用一較有伸縮性之辦法，以包進比差誤曲線較寬之各類，或用基於經驗得出之方程式，以配合特別觀察羣類。茲於本章擇其適於達到此（及其他）目的之方程式，略一討論之。

一般所用方法，係用一含有二，三，或四個未知常數之數學方程式。擇定常數之時，須方程所代表之曲線，能與由觀察構成之圖相配合；圖上點數既比常數數目為多，於是方程式亦多於未

知數，乃不得不決定適當解法矣。普通解此難題方法，係用最小二乘法（見附錄九），惟遇觀察得來之頻數曲線時，一般多用皮爾生法；此法係將由曲線方程式中得出之動差，用數學推算方法，使與由觀察得來（本編第一章第四節）之動差相等。其實，前論由觀察材料之前三級動差，以求平均數，標準差，及偏斜度(\bar{x}, σ, k)時，即已曾用此法，下節所論曲線系 (system of curve)，尤惟此是賴。此外尙有其他方法，然亦不外二途：一求得可以滿足條件——觀察資料係從隨機抽樣而有最小反機率者得來——之常數；二，或從所選點——而為方程式確切相合之點——中取用少數幾點。

第一節 皮爾生氏曲線系

皮爾氏曲線系，因其符號為統計調查一般所應用，吾人頗有注意之必要，與現所論者，亦復有關係，故不惜一說明之。惟若為精詳討論，則不妨參閱艾得頓先生所著『頻數曲線與相關』（Elderton: Frequency Curves and correlation），及哈地先生著『死亡表編製原理』（Hardy: Theory of the Construction of Table of Mortality）。

$\mu_0, \mu_1, \dots, \mu_i, \dots$ 表示一頻數曲線之連續各級動差， μ_0 代表面積，係作整一視之。設曲線之重心點落於縱坐標上， μ_1 之值

等於零。在此情形下，標準差， σ ，即為 $\sqrt{\mu_2}$ 。 β_1 代表 $\frac{\mu_3^2}{\mu_2^3}$ ， β_2 代表

$$\frac{\mu_4}{\mu_2^2}$$

此方程式

$$D_x y = \frac{(x+a)y}{b_0 + b_1 x + b_2 x^2} \dots \dots \dots (84)$$

即為解析之基礎。

此方程式甚合一條件——當 $y=0$ 時，曲線必經過縱軸，並同時在另一位置成水平線，如在 $x=-a$ 時便是。換言之，該曲線只有一衆數。

實際上，分母即以 $b_2 x^2$ 為止，以下無須多求。

此方程式，經積分而得下列三個交替之一般形式：

$$y = y_0 \left(1 + \frac{x}{a_1}\right)^{va_1} \left(1 - \frac{x}{a_2}\right)^{va_2}, \quad y = y_0 \left(1 + \frac{x^2}{a^2}\right)^{-m} e^{-v \tan^{-1} \frac{x}{a}},$$

及
$$y = y_0 (x-a)^{q_2 x - q_1},$$

在上三式中， y_0 及各組之三常數 (v, a_1, a_2) , (m, a, v) , (a, q_2, q_1) 均可用動差在基本方程式中之 a, b_0, b_1, b_2 ，求出。哈地氏曾用較簡之符號告吾人以另一解析方法。

當 a, b_0, b_1, b_2 各有特定值，或相互間有特定之關係時，僅含有兩個（甚至僅有一個）常數之簡單方程式，不難求得。關於曲線，茲共分為主要七型，艾得頓氏曾示吾人，各型配合相當觀察類數羣類之法。惟所用代數及算術甚繁耳。此法應用於生活費之

食物費用之結果，在本編第三章第七節，可見一斑。

差誤常態曲線方程式，可書如下列形式：

$$D_{xy} = -\frac{xy}{\sigma^2} \dots \dots \dots (85)$$

此乃特型之一。

求得一般差誤曲線之第二近似值爲

$$D_{xy} = -\frac{\left(x + \frac{\kappa\sigma}{2}\right)y}{\sigma^2 + \frac{\kappa\sigma x}{2}} \dots \dots \dots (86)$$

此式已將 k^2 略去，亦一特例也。

據皮爾生氏及其同仁所知，單峯(unimodal)觀察頻數羣類，一般均可用此公式之一二變量表現之。故平均數， σ ， β_1 及 β_2 之計算，乃成爲用四個淺顯量數以表示羣類之一般而有效之方法，依此程序進行，普通即以一平均數代表一羣類。用此四數，代表羣類之曲線方程，即可成爲適當之形式，復可據以用插補法求相當任何 x 值之 y 值，不論觀察資料如何分級也。

現所欲討論者，並非此等方程式如何用於機率問題之問題，亦非此方程式如何由經驗得來之問題，更非欲討論與機率發生之假定有如何之關係也。

第二節 愛基華斯氏法

在一羣類偏斜度過大，不能適用普遍化之差誤律之第二近

似值時，則代表此羣類之差誤常態曲線，加以變形，即可得一公式，此為愛基華斯氏所採用。但為羣類之敘述，插補或其他用途上，便利至何程度，尙未有充分之試驗，現時自不能輕予決定。（參閱統計學報，一八九八年，十二月份及一九一六年七月份所發表二文）。

第三節 巴里多氏方程

設 $b_0=0$, $b_2=\frac{b_1}{a}=-\frac{1}{m}$ ，則由上述曲線系，可得一方程式

$D_x y = -\frac{m y}{x}$ 。此方程式所表示者，乃為當 m 為正數時，不拘 x 為何值，總歸向右偏斜之曲線。

如積分之，則方程式變為 $\log y = -m \log x + \text{常數}$ ，換言之，即為 $y = Cx^{-m}$ 。

曲線自 x 至 ∞ 之面積為

$$\begin{aligned} z &= \int_x^{\infty} Cx^{-m} dx = \left\{ \frac{Cx^{1-m}}{1-m} \right\}_x^{\infty} \\ &= \frac{C}{(m-1)x^{m-1}} \quad (\text{設 } m > 1) \end{aligned}$$

以 a 代 $m-1$, A 代 $\frac{C}{m-1}$ ，則得

$$y = \frac{Aa}{x^a - 1}, z = \frac{A}{x^a} \dots \dots \dots (87)$$

最末一方程式，係為所得問題所用巴里多氏律 (Pareto's

Law)之最簡式。A與 a 為常數, z 為所得在 x 磅(或 x 佛郎及其他等等)左右之總人數。

其他羣類與此一公式相合者甚多,例如年租等級不同之房屋,此為年租之變量數及次數全距可為極寬者,並可用雙對數尺繪製成圖者,即是。

仍以所得問題而論,所得總額自 $\pounds x_1$ 至 $\pounds x_2$ 者,為

$$\int_{x_1}^{x_2} xy dx = \frac{Aa}{a-1} \left(\frac{1}{x_1^{a-1}} - \frac{1}{x_2^{a-1}} \right).$$

全距中所得數為 $A \left(\frac{1}{x_1^a} - \frac{1}{x_2^a} \right)$ 。

此定律對於極低或極高之所得,一般均以為不便應用。但如用於最高所得,則得

$$\text{在 } \pounds x \text{ 或以上之所得總額} = \frac{Aa}{(a-1)x^{a-1}},$$

$$\text{在 } \pounds x \text{ 或以上所得數} = \frac{A}{x^a} = N,$$

$$\text{故 } \pounds x \text{ 或以上之平均所得} = \frac{a}{a-1}x,$$

用此數方程式,由原始資料中,即可求出 A, a 。

巴里多氏方程對於一九一一至一九一二年須繳所得附加稅之所得統計,在全距 $\pounds 5,000$ — $\pounds 55,000$ 之間,配合情形良佳,惟在 $\pounds 55,000$ 以上之所得數,算得者高於實在數耳。

$$a=1.5, \log A=9.618 \text{ 能切實配合。}$$

所得全距	所得次數	
	計算數	記錄數
£5 to £10.....	7,546	7,411
10 ,, 15.....	1,890	2,029
15 ,, 20.....	790	787
20 ,, 25.....	424	438
25 ,, 35.....	411	382
35 ,, 45.....	199	186
45 ,, 55.....	103	107
55 ,, 65.....	70	56
65 ,, 75.....	50	37
75 ,, 100.....	118	55
100 and over	83	66
Totals.....	11,700	11,554

所得總額：計算額 £166,000,000；記錄額 £145,000,000。

如繪一雙對數尺圖，全距——全距上劃一直線，則直線即為恰當之近似值——可以一目瞭然，而其斜度，即可取為 α 之試驗值。此值可選用 x 之兩值，例如 x_1 與 x_2 ，測驗出來，而 x_1 與 x_2 可由臨近經驗線之點，譬如 N_1 與 N_2 ，而得 N 之值。

$$\text{於是 } \alpha = \frac{\log N_1 - \log N_2}{\log x_2 - \log x_1}。$$

如上表，以 $x_1 = 5,000$ ， $x_2 = 45,000$ ，則得 $N_1 = 11,554$ ， $N_2 = 821$ ；故 $\alpha = 1.63$ 。

雖然，此法係假設最高所得之次數，乃與定律相合；但一般情形，則未必盡然；實際上仍以三值 x_1 ， x_2 ， x_3 為宜。

於是，方程式

$f(a) = \left(\frac{1}{x_1^a} - \frac{1}{x_2^a} \right) \div \left(\frac{1}{x_1^a} - \frac{1}{x_3^a} \right) = \frac{\text{自 } x_1 \text{ 至 } x_2 \text{ 所得次數}}{\text{自 } x_1 \text{ 至 } x_3 \text{ 所得次數}} = k$ (已知數) 足可得出 a 值。設 $a=1.60$ 為試驗值；依次以 $a=1.5, a=1.55, a=1.60, a=1.65$ ，以解算 $f(a)-k$ ，再用插補法求 a 之值，使 $f(a)=k$ 愈近愈佳。然後就此得數，測驗其他各值。 a 之值既已求得，求 A 不難矣。

此外尚有一法，應用資料或可較為完全，其法係用方程式 x_1 至 x_2 全距中之平均所得

$$= \frac{a}{a-1} \left(\frac{1}{x_1^{a-1}} - \frac{1}{x_2^{a-1}} \right) \div \left(\frac{1}{x_1^a} - \frac{1}{x_2^a} \right)。$$

關於此種公式之應用，可參閱英國下議院所得稅委員會之報告（一九〇六年三六五號，第二二〇至二三〇頁，第二四〇至二四一頁，及第二四五至二四六頁）。

第四節 梅克漢氏公式

由方程式 $-\frac{1}{y} \cdot \frac{dy}{dx} = a + bc^x$ ，可引出一實際應用上甚為重要之公式。

為簡便起見，用積分法，以 $a = -\log s$ ， $b = -\log c \times \log g$ ，則 $\log y = x \log s + c^x \cdot \log g + \text{常數}$

$$\therefore y = ks^x \cdot (g)^{c^x} \cdot (k, s, g, c, \text{均為常數}) \dots \dots \dots (88)$$

此即梅克漢 (Makeham) 氏公式。在此公式中， y 為某一時

代活至 x 歲之人數，並寫作 l_x 。

某一期間， δ_x ，死亡人數，對該期開始時生存人數之比，再用期間除之，當期間無限消滅時，則為

$$\frac{l_x - l_{x+\delta_x}}{l_x \cdot \delta_x} = -\frac{1}{l_x} \cdot \frac{dl_x}{dx} \dots \dots \dots (89)$$

此式多以 μ_x 代表之，所謂『死亡勢』（“force of mortality”）者即是。

用此公式之微分方程，則得

$$\mu_x = a + bc^x \dots \dots \dots (90)$$

此係以死亡勢為二數之和，二數者一為常數 a ，另一（即 $bc^x = \mu'_x$ ）則因 $\frac{D_x \mu'_x}{\mu'_x} = \log c$ ，而依一定之幾何級數而增加者也。

哈地氏（同前書第八十八頁）用 $\overline{a+u'_x}$ 代 a ，其式較為繁複，英國議會之管理國營保險法案一九一二至一九一三年報告，第一部第五百八十五頁亦如之。

哈地氏又用雙曲線方程式為分度(graduation)之用，

$$z = \log \frac{y}{N-y} = k + \frac{m}{a+x} + \frac{n}{b+x}$$

y 為年齡在 x 歲以下之丈夫人數， N 為丈夫總人數（見死亡表編製法第五十頁至第五十一頁及 CD. 6907. P. 595）。

第六章 相關論

第一節 導言

各現象相互間是否彼此了不相關，各自獨立？此為統計學重要問題之一。如答覆係為反面，則請測量其依賴性如何。

本章討論對象，以此問題為主要；提及此種問題，必聯帶論及二三變量，而各種變量之變動，又或有共同之原因。

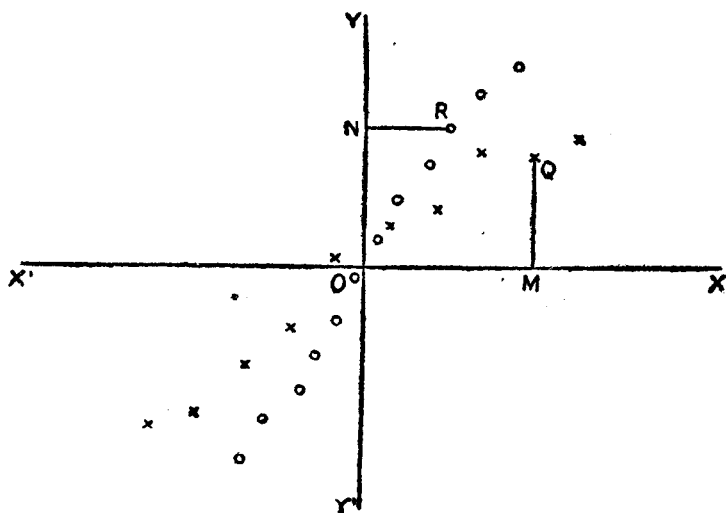
設有一對之觀察，如人之身長及指距，兩弟兄之身高，一家庭之所得及房租均是。假以各對量數為 $(X_1, Y_1), (X_2, Y_2) \dots$ ，並設 X, Y 各成一類數分類，平均數分別為 \bar{x}, \bar{y} 。如 X 與 Y 彼此完全獨立，則知 X 之值，未必即知 Y 之相對值；對於 \bar{y} 之某種離中差，有無一定之機率，乃視 Y 之類數曲線而定；但如 X, Y 變動，有共同之原因，存乎其間，則一知 X 值為何，相對之 Y ，離中差之機率，必得而預測矣。

X 與 Y 可在一方程內，發生密切之關係，此為當然之事實，例如 X 磅與 Y 公斤，乃為同一體重之不同說法，依此而論， $X = 2.204Y$ ， Y/X 必為常數也。然有多種事實，吾人今後欲加研究者，關係並不如此簡單；雖知 X 為若干， Y 值竟無從決定，在此情

形下，惟有藉重成列量數（如身高），以求隨此 X 而變之 Y 值。

如 Y 之各項，各與一 X 相聯，其頻數曲線平均數（或形式），與在未依 X 值而分類時之 Y 各值平均數（或形式）不同，則在 X, Y 兩數量之間，必有共同性質存在，此謂之 X 與 Y 相關。

解析之法，當以將 Y 之觀察值列為整列（array）為第一，每一整列同時將與 X 相同之各值全行列入，如列普通交叉表然。當 $X = X_t$ ， Y 整列之平均數 \bar{Y}_t ，如具有完全獨立性而觀察數甚多，則有漸近於所有各 Y 之平均數， \bar{y} ，之傾向；苟該平均數與 \bar{y} 相差，駕乎隨機抽樣所得之上，則可知 Y 值並未與 X 值脫離牽涉，故不能互為獨立也，



上圖，設O代表X，及Y二列之平均數。以 x_t, y_t 為 X_t, Y_t 超過平均數 \bar{x}, \bar{y} ，上之部分，並以OM為任一值， x_t ，MQ為整列中各 y 之平均數；如此，則 $MQ = \bar{Y}_t - \bar{y}$ ； \times 號表示Q之各處位置。

於是，如Y與X各為獨立，則 $\times \times \times$ 只於觀察次數不足以求得真實平均數時，距 $X X'$ 方為甚遠。如Y不能脫離X之影響，則Q必形成一定軌跡，而此軌跡即為經過各點隨手劃成之線。在此情形下，可以 $\bar{Y}_t - \bar{y} = f(X_t - \bar{x})$ ，故當 X_t 值已知， Y_t 之實在值，雖不得而知；但整列之平均數， \bar{Y}_t ，卻可見於疊次抽樣中，故亦可約略決定之。

同理，如取相當 y_t (ON) 任一值之X整列，則此等整列之平均數——0000標記之各R值即是——有與曲線 $\bar{X}_t - \bar{x} = f_1(Y_t - \bar{y})$ 相合之趨勢（此處， f_1 與 f 同，須注意）。

Q之軌跡，名曰Y對X之迴歸線，R之軌跡，亦名X對Y之迴歸線。

此線曲線多大略為直線，尤以鄰近O點時為最，故 $\bar{Y}_t - \bar{y}$ 約略等於 kx_t （當 x_t 甚小時， k 為常數）。

此線之斜度，以 k 表示之，約略與 $\frac{MQ}{OM}$ 相等（不論OM之值如何），求之方法，只須將 $\frac{MQ}{OM}$ 之各值，加以平均即可。以後當再論之（本章第三節及第七節）。

茲另從各方解決此問題：

設有 n 個量數 X_1, X_2, \dots , 並有 n 個量數 Y_1, Y_2, \dots 。

隨機抽出一 X , 再獨立抽得一 Y , 乃得一乘積 XY 。據經驗所知, 一特定 X 被抽選之後, Y 之各值將以相等之頻數而出現。且就一般而論, 每一 n^2 個乘積 $X_1 Y_1, X_1 Y_2, \dots, X_2 Y_1, \dots, X_n Y_n$ 必以相等之頻數而發生。

極大一數, N 個, 乘積之和 $= S(XY) = S(\bar{x} + x)(\bar{y} + y) = N\bar{x}\bar{y} + \bar{x} \cdot S y + \bar{y} S x + S x y$ 。

此處 $S y$ 漸近於 $\frac{N}{n}(y_1 + y_2 + \dots + y_n) = 0$, $S x$ 亦漸近於 0 。

$S(xy)$ 漸近於 $\frac{N}{n^2}(x_1 y_1 + x_1 y_2 + \dots + x_2 y_1 + \dots + x_n y_n) = \frac{N}{n^2} \cdot S x \cdot S y = 0$ 。

故 $S(XY)$ 漸近於 $N\bar{x}\bar{y}$, 乘積 XY 之平均值漸近於 X 之平均數與 Y 之平均數之乘積。

但如抽取 Y 時, 不能與 X 之抽取脫離開涉, 則 n^2 個乘積 $x_1 y_1, \dots$, 即不能以相同頻數而出現, 故, 平均 $XY = \bar{x}\bar{y} + \text{平均 } xy$ 。

又如 \bar{m} 為 $M_1 M_2, \dots$ 之未加權平均數, 依第四章第五節所用之符號, $M_t = \bar{m} + m_t$, 而 \bar{m}_w 為加權平均數, 至 w_t —— M_t 之權數 —— 則等於 $\bar{w} + w_t$, 又 \bar{w} 乃權數之平均數也。

$$\bar{m}_w = \frac{S(\bar{w} + w_t)(\bar{m} + m_t)}{n\bar{w}} = \frac{1}{n\bar{w}}(n\bar{w}\bar{m} + S w_t m_t)$$

$$= \bar{m} \left(1 + \frac{1}{n} S \frac{w_t}{\bar{w}} \cdot \frac{m_t}{\bar{m}} \right) \dots \dots \dots (91)$$

至於 $\bar{m}_w = \bar{m}$ 之條件有三：

(1) 只於 $S w_t m_t = 0$ 時，有之

(2) 只於 n 甚大時，有之

(3) 只於大權數與大數量比與小數量難於發現時，有之。反是，則不等。

依第三章第一節所論，如用上述符號， $(x+y)^2$ 之平均值為 $\sigma_x^2 + \sigma_y^2$ ，所設 σ_x 與 σ_y ，抽取 X 與 Y 時，在完全獨立情形下，乃為 X, Y 之標準差。

由此可見，如不能彼此獨立時，則上式應當變為

$$S^2 = \text{平均值}(x+y)^2 = \sigma_x^2 + \sigma_y^2 + 2 \cdot \text{平均值}xy \dots \dots \dots (92)$$

第二節 相關係數

由此觀之， X 與 Y 未能彼此獨立時，『平均值 xy 』一數量，算式應用者甚多，且由此一數量，可藉明相關之存在及其總量。然其數值乃以測量 x, y 之單位為單位，並無相當天然尺度以名之，故下述一數量，實為最優之名稱。

設有 $X_1 Y_1, X_2 Y_2, \dots, X_t Y_t, \dots, X_n Y_n$ 若干對量數， X 及 Y 之平均數及標準差為 $\bar{x}, \sigma_x, \bar{y}, \sigma_y$ ，則 X, Y 間之相關係數，乃為

r_{xy} , r_{xy} 之求法, 如下:

$$r_{xy} = \frac{S\{(X_t - \bar{x})(Y_t - \bar{y})\}}{n\sigma_x\sigma_y} = \frac{1}{n} \left(\frac{x_t}{\sigma_x} \cdot \frac{y_t}{\sigma_y} \right)$$

(以 $X_t = \bar{x} + x_t$, $Y_t = \bar{y} + y_t$)

$$\begin{aligned} &= \frac{1}{n\sigma_x\sigma_y} \left\{ S(X_t Y_t) - \bar{x} S Y_t - \bar{y} S X_t + n\bar{x}\bar{y} \right\} \\ &= \frac{1}{n\sigma_x\sigma_y} \left\{ S(X_t Y_t) - n\bar{x}\bar{y} \right\} \dots\dots\dots (93) \end{aligned}$$

(因 $S Y_t = n\bar{y}$, $S X_t = n\bar{x}$)

依上例

$$\text{平均值 } XY = \bar{x}\bar{y} + r_{xy}\sigma_x\sigma_y$$

$$\bar{m}_w = \bar{m} \left(1 + r_{mw} \frac{\sigma_m}{\bar{m}} \cdot \frac{\sigma_w}{\bar{x}} \right)$$

$$s^2 = \sigma_x^2 + \sigma_y^2 + 2r_{xy}\sigma_x\sigma_y$$

以 r 代 r_{xy} , 可知 r 永不 > 1 , 或 < -1 。

因爲 $n^2 r^2 \sigma_x^2 \sigma_y^2 = (S x_t y_t)^2$;

$$\text{但 } n^2 \sigma_x^2 \sigma_y^2 - (S x_t y_t)^2 = (x_1^2 + x_2^2 + \dots) (y_1^2 + y_2^2 + \dots) - (x_1 y_1 + x_2 y_2 + \dots)^2$$

(因爲 $n\sigma_x^2 = x_1^2 + x_2^2 + \dots$, 且 $n\sigma_y^2 = y_1^2 + y_2^2 + \dots$)

$$\begin{aligned} &= (x_1 y_2 - x_2 y_1)^2 + (x_1 y_3 - x_3 y_1)^2 + \dots + (x_2 y_3 - x_3 y_2)^2 + \dots \\ &\quad + (x_n y_{n-1} - x_{n-1} y_n)^2 \end{aligned}$$

$n^2 \sigma_x^2 \sigma_y^2 > 0$ (除非 $x_1 y_2 - x_2 y_1 = 0 = x_1 y_3 - x_3 y_1 = \dots$)

$$\therefore \frac{x_1}{y_1} = \frac{x_2}{y_2} = \frac{x_3}{y_3} = \dots = \frac{x_n}{y_n} = \pm \frac{\sigma_x}{\sigma_y}$$

在此情形下，該式等於零，於是 $r = \pm I$ 。

$$\therefore r^2 = \frac{(\sum x_i y_i)^2}{n^2 \sigma_x^2 \sigma_y^2} < I,$$

$\therefore I > r > -I$ (y 隨 x 順變時爲例外)

於是 $r = +I$ 或 $-I$ (94)

r 之性質，可歸納於下：

(1) r 爲一隨所有觀察而起變化之數量，

(2) 當抽選 X, Y 各值完全獨立，而平均值 $xy = 0$ 時， r 之值爲零，

(3) 與測量 X, Y 所用之單位無關，

(4) 凡遇一正 x_i 與正 y_i 同時出現，或遇一負 x_i 與負 y_i 同來時， r 之值必增加，

(5) 但 r 增加僅能達到 $+I$ ，而不能超過 $+I$ ，當 x, y 之關係爲方程式『 $y = x \times \text{常數}$ 』時。

(6) 如正 x ，與負 y 同來，或負 x 與正 y 同出時， r 則由 0 漸變直至 $-I$ 爲止。

故 r 測算相關量(amount of corretation)甚爲敏捷。

爲與他種量數區別起見，此相關量數，謂之積和相關係數(product-sum coefficient of correlation)。

如以各對分爲整列，例如 $x_s y_s; x_{s2} y_s \dots$ ，而以 y_s 爲 n_i 個數

量 y_1, y_2, \dots 之平均數，則 $Sxy = Sx_s \cdot n_s \bar{y}_s$ ，而 $\frac{\sigma_y}{\sigma_x} \cdot r = S \left(\frac{n_s x_s^2 \cdot \bar{y}_s}{n \sigma_x^2} \right)$ ，(因 $S n_s x_s^2 = n \sigma_x^2$)；故 $\frac{\sigma_y}{\sigma_x} r$ 及為本章第一節 $\frac{MQ}{OM}$ 各值之加權平均數，而 $y = r \frac{\sigma_y}{\sigma_x} \cdot x$ 為 Q 之軌跡之近似值。

茲於下節討論(1)檢察 r 得各種數值之原因，(2)研究 X, Y 依各種假定之分佈，(3)求迴歸線方程式。

第三節 r 之特性

以 X, Y 為二個變量，隨其他變量 U, V, W ，而起變動，其變動之情形如下：

$$X_t = {}_1U_t + {}_2U_t + \dots + {}_pU_t + {}_1V_t + {}_2V_t + \dots + {}_qV_t,$$

$$Y_t = {}_1U_t + {}_2U_t + \dots + {}_pU_t + {}_1W_t + {}_2W_t + \dots + {}_qW_t,$$

${}_1U_t$ 係由一以 ${}_1\bar{u}$ 為平均數， ${}_1\sigma_u$ 為標準差之任何形式之類數羣類隨機抽來， ${}_2U_t$ 又另從其他羣類獨立抽來，其餘 U, V, W 各類亦然。 p 與 q 為任何整數。

以 ${}_1U_t = {}_1\bar{u} + {}_1u_t, \dots$ ，以 $X_t = \bar{x} + x_t, Y_t = \bar{y} + y_t$ ， \bar{x} 與 \bar{y} 為 X, Y 各可能值之平均數。以 σ_x, σ_y 為 X, Y 之標準差。於是

$$x_t = {}_1u_t + {}_2u_t + \dots + {}_pu_t + {}_1V_t + {}_2V_t + \dots + {}_qV_t,$$

$$y_t = {}_1u_t + {}_2u_t + \dots + {}_pu_t + {}_1W_t + {}_2W_t + \dots + {}_qW_t,$$

$$\sigma_x^2 = {}_1\sigma_u^2 + \dots + {}_p\sigma_u^2 + {}_1\sigma_v^2 + \dots + {}_q\sigma_v^2,$$

$$\sigma_y^2 = 1\sigma_u^2 + \dots + p\sigma_v^2 + 1\sigma_w^2 + \dots + q\sigma_w^2,$$

以上各 u, v, w 項, 依假說之規定, 彼此各係獨立 (見第三章第一節)。

如 $1\sigma_u = 2\sigma_u = \dots = \sigma_u, 1\sigma_v = 2\sigma_v = \dots = \sigma_v, \text{且 } 1\sigma_w = 2\sigma_w = \dots = \sigma_w$, 或以 $\sigma_u^2, \sigma_v^2, \sigma_w^2$ 爲 U, V, W 各標準差之平均值, 則

$$\sigma_x^2 = p\sigma_u^2 + q\sigma_v^2, \quad \sigma_y^2 = p\sigma_u^2 + q\sigma_v^2.$$

又 平均值 $x_t y_t =$ 平均值 $1u_t^2 +$ 平均值 $2u_t^2 + \dots +$ 平均值,

$$1u_t \cdot 1w_t + \dots + \text{平均值 } 1v_t \cdot u_t + \dots = p\sigma_u^2,$$

因依假說之規定, U, V, W 各項之抽取, 彼此完全獨立, 則在長久經驗之後, $1u_t \cdot 1w_t$ 等各項當等於零。

$$\text{故 } r = \text{平均值 } x_t y_t / \sigma_x \cdot \sigma_y = \frac{p\sigma_u^2}{\sqrt{\{(p\sigma_u^2 + q\sigma_v^2)(p\sigma_u^2 + q\sigma_v^2)\}}} \quad (95)$$

又在特別情形之下, 如 $\sigma_u = \sigma_v = \sigma_w$, 則

$$r = \frac{p}{p+q} \dots \dots \dots (96)$$

此爲 r 數值之最簡單概念; 如以文字說明之, 此式之意義, 乃指: 相關係數漸近於兩變量產生時之共同原因數, 對各變量所依之獨立原因總數之比。

如加常數 a, b, c, d , 使

$$\left. \begin{aligned} x_t &= a_1 \cdot 1u_t + \dots + ap \cdot pu_t + b_1 \cdot 1v_t + \dots + bq \cdot qu_t \\ y_t &= c_1 \cdot 1u_t + \dots + cp \cdot pu_t + d_1 \cdot 1w_t + \dots + dq \cdot qw_t \end{aligned} \right\} \dots (97)$$

於是 $\sigma_x^2 = S \cdot a^2 \sigma_u^2 + S \cdot b^2 \sigma_v^2, \sigma_y^2 = S \cdot c^2 \sigma_u^2 + S \cdot d^2 \sigma_w^2,$

平均值 $x_t y_t = \rho \sigma_u \sigma_v \sigma_w^2$,

求 r 之方程式，於此可得矣。

第四節 相關面

試就 U, V , 及 W 各類數曲線係常態之情形而論，先請考查當 $x_t = u_t + v_t, y_t = u_t + w_t$ 時 X, Y 之分類。

抽得 u_t, v_t, w_t 之機率為

$$\frac{1}{\sigma_u \sqrt{2\pi}} e^{-\frac{u_t^2}{2\sigma_u^2}} \times \frac{1}{\sigma_v \sqrt{2\pi}} e^{-\frac{v_t^2}{2\sigma_v^2}} \times \frac{1}{\sigma_w \sqrt{2\pi}} e^{-\frac{w_t^2}{2\sigma_w^2}}$$

消去 v_t 及 w_t 。

於從 U 類中抽得 u_t (以至 $u_t + \delta u$ 為止) 時，取得

$$x_t (\text{至 } x_t + \delta x), y_t (\text{至 } y_t + \delta y)$$

之機率為：

$$\frac{1}{\sigma_u \sigma_v \sigma_w (2\pi)^{\frac{3}{2}}} e^{-\frac{1}{2} \left\{ \frac{u_t^2}{\sigma_u^2} + \frac{(u_t - x_t)^2}{\sigma_v^2} + \frac{(u_t - y_t)^2}{\sigma_w^2} \right\}} \cdot \delta x \delta y \delta u$$

$$= \frac{1}{\sigma_u \sigma_v \sigma_w (2\pi)^{\frac{3}{2}}} e^{-\frac{1}{2} k \{ u_t - l x_t - m y_t \}^2 - \frac{1}{2} (a x_t^2 + 2h x_t y_t + b y_t^2)} \cdot \delta x \delta y \delta u$$

式中 $k = \frac{1}{\sigma_u^2} + \frac{1}{\sigma_v^2} + \frac{1}{\sigma_w^2}, kl = \frac{1}{\sigma_v^2}, km = \frac{1}{\sigma_w^2}$

$$a = \frac{1}{\sigma_u^2} - kl^2, b = \frac{1}{\sigma_w^2} - km^2, h = -klm.$$

從 u 之任何值，求得 x_t, y_t 之機率 (譬如 P_{xy})，將從 u 之指定值求得 x_t, y_t 之機率，相加即得。

故，以 x, y 代 x_t, y_t ,

$$P_{xy} = \frac{1}{\sigma_u \sigma_v \sigma_w \cdot 2\pi} e^{-\frac{1}{2}(ax^2 + 2hxy + by^2)} \cdot \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}k(u_t - lx_t - my_t)^2} \cdot du$$

$$= \frac{1}{\sigma_u \sigma_v \sigma_w 2\pi \sqrt{k}} e^{-\frac{1}{2}(ax^2 + 2hxy + by^2)}$$

但 $\sigma_x^2 = \sigma_u^2 + \sigma_v^2, \sigma_y^2 = \sigma_u^2 + \sigma_w^2, r\sigma_v\sigma_w = \sigma_u^2$

$$\therefore k = \frac{(\sigma_u^2 + \sigma_v^2)(\sigma_u^2 + \sigma_w^2) - \sigma_u^4}{\sigma_u^2 \sigma_v^2 \sigma_w^2} = \frac{\sigma_x^2 \sigma_y^2 (1 - r^2)}{\sigma_u^2 \sigma_v^2 \sigma_w^2}$$

$$a = kl - kl^2 = l(k - kl) = l\left(\frac{1}{\sigma_u^2} + \frac{1}{\sigma_w^2}\right) = \frac{1}{k\sigma_v^2} \cdot \frac{\sigma_y^2}{\sigma_u^2 \sigma_w^2}$$

$$= \frac{1}{\sigma_x^2 (1 - r^2)}$$

同理 $b = \frac{1}{\sigma_y^2 (1 - r^2)}$

$$-h = \frac{1}{k} \cdot kl \cdot km = \frac{1}{k\sigma_v^2 \sigma_w^2} = \frac{\sigma_u^2}{\sigma_x^2 \sigma_y^2 (1 - r^2)} = \frac{r}{\sigma_x \sigma_y (1 - r^2)}$$

$$\therefore P_{xy} = \frac{1}{2\pi \sigma_x \sigma_y \sqrt{1 - r^2}} e^{-\frac{1}{2(1 - r^2)} \left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} - \frac{2rxy}{\sigma_x \sigma_y} \right)} \dots \dots \dots (98)$$

將此法擴張，可知（見艾得頓著：『頻數曲線』第一百零九頁，該文係根據 Pearson，在皇家學會報，第一百八十七卷（一八九六年），A 175號之文）如 x_t, y_t 係由若干均為常態頻數（如公式98所示）之變量加權總和（weighted sum）而成，則 P_{xy} 即成 $ke^{-(ax^2 + 2hxy + by^2)}$ 形式，上文已述及；又當以曲線係整體為條件時，標準差 σ_x, σ_y 及 $r\sigma_x\sigma_y$ 之平均乘積用積分法表示，則 k, a, h, b 之值，與已論過之單純情形所得者相同。

此法頗有意趣，且此爲皮爾生氏用積和公式(the product-sum formula) 求算相關，在現代統計學中之一貢獻，但此乃基於元素頻數曲線係爲常態之假定，由此假定，積和方法之價值大爲消失，蓋觀察資料未必盡合常態也。茲爲免除此種假定起見，爰就愛基華斯氏『大數律』一文(該本上已介紹)，重爲分析推論於下。

第五節 愛基華斯氏法

$$\text{假以 } x_t = 1u_t + 2v_t + \dots + p u_t \dots + n v_t$$

$$y_t = 1v_t + 2v_t + \dots + p v_t \dots + n v_t$$

$1u_t$ 爲自平均數之離中差，換言之，即自以 $1\sigma_n$ 爲標準差之頻數曲線中抽得一數量之平均數而來之離中差， $2u_t \dots n u_t, 1v_t \dots n v_t$ 亦如之。設抽取各個 u_t 值，彼此完全獨立，則平均值 $1u_t \cdot 2u_t$ 漸近於零， v_t 各值亦爲獨立，則亦同然；但設有數 v_t 值不能離 u_t 值而獨立，則平均值 $1u_t \cdot 1v_t$ 平均值 $2u_t \cdot 2v_t \dots$ 平均值 $n u_t \cdot n v_t$ ，並不全漸近於零。至平均值 $1u_t \cdot 2v_t$ 一數量，則視爲近於零。

設 n 甚大， $\frac{1}{\sqrt{n}}$ 可以略去，且與第三章第五節所舉之其他條件相合，則 x, y 之各別頻數曲線必爲差誤常態曲線。

$$\text{於是 } \sigma_x^2 = S(p\sigma_u^2), \sigma_y^2 = S(p\sigma_v^2),$$

$$\text{而 } \text{平均值 } xy = S(\text{平均值 } p u_t, p v_t) \dots \dots \dots (99)$$

式中 p , 爲 1 與 n 間之一整數。

再將測量 x 與 y 之軸轉換, 經過 θ 角, 使 $2\theta = \frac{2r\sigma_x\sigma_y}{\sigma_x^2 - \sigma_y^2}$,

$r = \frac{\text{平均值 } xy}{\sigma_x\sigma_y}$ 即爲 x 與 y 間之相關係數。

試以 $X_t = x_t \cos\theta + y_t \sin\theta$,

$Y_t = x_t \sin\theta - y_t \cos\theta$,

再以 ${}_pU_t = {}_p u_t \cos\theta + {}_p v_t \sin\theta$,

${}_p v_t = {}_p u_t \sin\theta - {}_p v_t \cos\theta$ (註一)

於是 $X_t = S_p U_t$,

且 $Y_t = S_p V_t$

對 $\tan 2\theta$ 之值而言,

平均值 $X_t Y_t = \sin\theta \cos\theta (\sigma_x^2 - \sigma_y^2) - \cos 2\theta \cdot (\text{平均值 } xy) = 0$.

$\sigma_x^2 = \cos^2\theta \cdot \sigma_x^2 + \sin^2\theta \cdot \sigma_y^2 + \sin 2\theta \cdot r\sigma_x\sigma_y = S(\text{平均值 } {}_p U_t^2)$

$\sigma_y^2 = \sin^2\theta \cdot \sigma_x^2 + \cos^2\theta \cdot \sigma_y^2 - \sin 2\theta \cdot r\sigma_x\sigma_y = S(\text{平均值 } {}_p V_t^2)$

$\therefore \sigma_x^2 + \sigma_y^2 = \sigma_x^2 + \sigma_y^2$.

$\sigma_x^2 - \sigma_y^2 = (\sigma_x^2 - \sigma_y^2) \cos 2\theta + 2r\sigma_x\sigma_y \sin 2\theta = (\sigma_x^2 - \sigma_y^2) \sec 2\theta$

$= \sqrt{\{(\sigma_x^2 - \sigma_y^2)^2 + 4r^2\sigma_x^2\sigma_y^2\}}$

$4\sigma_x^2\sigma_y^2 = 4\sigma_x^2\sigma_y^2(1 - r^2)$

平均值 ${}_p U_t \cdot {}_p V_t = \sin\theta \cos\theta ({}_p \sigma_u^2 - {}_p \sigma_v^2) - \cos 2\theta$

(平均值 ${}_p u_t \cdot {}_p v_t$).

$\therefore S(\text{平均值 } {}_p V_t \cdot {}_p V_t) = \text{平均值 } X_t Y_t = 0 \dots\dots\dots (100)$

現依第三章第四節所用計算 X, Y 之動差方法。

設 α, β 爲任何小之常數, 茲用以收集相似之項。

$$e^{\alpha X_t + \beta Y_t} = e^{\alpha_1 U_t + \beta_1 V_t} \times e^{\alpha_2 U_t + \beta_2 V_t} \times \dots$$

將此指數函數展開, 得出 t 個可能值, 而取其平均值, 吾人依前述原理, 已知一和之平均值, 即爲各項之平均總和, 又知獨立因子 (如上式右方之因子) 積數之平均值, 即爲各因子之平均值積數, 且知一乘羈之平均, 乃爲零。

如以 k, l 爲任何之整數,

$$\left\{ 1 + \frac{\alpha^2}{2} \sigma_x^2 + \dots + \frac{\alpha^k}{k!} (\text{平均數 } X^k) \dots \right\} \\ \times \left\{ 1 + \frac{\beta^2}{2} \sigma_y^2 + \dots + \frac{\beta^l}{l!} (\text{平均值 } Y^l) \dots \right\}$$

= n 個因子之乘積

$$\left\{ 1 + \frac{1}{2} \alpha^2 (\text{平均值}_p U_t^2) + \frac{1}{2} \beta^2 (\text{平均值}_p V_t^2) + \alpha\beta \right. \\ \left. (\text{平均值}_p U_t V_t) + \dots \right\}$$

$$\therefore \log \left\{ 1 + \dots + \frac{\alpha^k \beta^l}{k! l!} (\text{平均值 } X^k Y^l) + \dots \right\},$$

(其 k 或 l 或可等於零),

$$= S \left[\log \left\{ 1 + \frac{1}{2} \alpha^2 (\text{平均值}_p U_t^2) + \frac{1}{2} \beta^2 (\text{平均值}_p V_t^2) + \alpha\beta \right. \right. \\ \left. \left. (\text{平均值}_p U_t V_t) + \dots \right\} \right]$$

= (用對數級數將其展開, 並歸併各項)

$$\frac{1}{2} \alpha^2 S (\text{平均值}_p U_t^2) + \frac{1}{2} \beta^2 S (\text{平均值}_p V_t^2) + \alpha\beta S \\ (\text{平均值}_p U_t V_t) + \dots$$

$= \frac{1}{2}a^2 \cdot \sigma_X^2 + \frac{1}{2}\beta^2 \cdot \sigma_Y^2 + a\beta \times 0$ (註二) + 含有 $a^3, a^2\beta, \dots$ 之項。

依第三章第四節末段之理論，有 a^3 之項係屬 $\frac{1}{\sqrt{n}}$ 次， a^2 項則

不然，故如將 $\frac{1}{\sqrt{n}}$ 略去，則

$$1 + \dots + \frac{a^k \beta^l}{k! l!} (\text{平均值 } X^k Y^l) + \dots = e^{\frac{1}{2}a^2 \sigma_X^2} \times e^{\frac{1}{2}\beta^2 \sigma_Y^2}$$

$$= \left\{ 1 + \frac{1}{2}a^2 \sigma_X^2 + \dots + \frac{1}{k!} \left(\frac{1}{2}a^2 \sigma_X^2 \right)^k \dots \right\}$$

$$\left\{ 1 + \frac{1}{2}\beta^2 \sigma_Y^2 + \dots + \frac{1}{l!} \left(\frac{1}{2}\beta^2 \sigma_Y^2 \right)^l \dots \right\}$$

取此方程式兩方各項係數令其相等，則在 $l=0$ 時，

$$\text{平均值 } X^{2k+1} = 0, \text{ 平均值 } X^{2k} = \frac{(2k)!}{2^k k!} = \sigma_X^{2k}$$

與在常態曲線同，當 Y 為零時， Y 亦同然。

凡與 X 或 Y 之奇數乘鬆 (odd power) 有關各平均值均等於零。

$$\text{平均值 } (X^{2k} Y^{2l}) = \frac{(2k)!}{2^k k!} \sigma_X^{2k} \cdot \frac{(2l)!}{2^l l!} \sigma_Y^{2l} \dots \dots \dots (101)$$

此等恰為將下列之平面積分，求得之平均乘鬆 (mean power)

$$z = \frac{1}{2\pi \sigma_X \sigma_Y} e^{-\frac{1}{2} \left(\frac{X^2}{\sigma_X^2} + \frac{Y^2}{\sigma_Y^2} \right)} = \frac{1}{\sigma_X \sqrt{2\pi}} e^{-\frac{1}{2} \frac{X^2}{\sigma_X^2}} \cdot \frac{1}{\sigma_Y \sqrt{2\pi}} e^{-\frac{1}{2} \frac{Y^2}{\sigma_Y^2}},$$

其 X 係與 Y 獨立。故如第三章第四節末段之理論，可以此方程式，用以求 X, Y 之頻數。

現當仍轉換為原軸。

前已示明 $\sigma_X \sigma_Y = \sigma_x \sigma_y \sqrt{1-r^2}$,

$$\begin{aligned} X^2 \sigma_Y^2 + Y^2 \sigma_X^2 &= (x \cos \theta + y \sin \theta)^2 \sigma_Y^2 + (x \sin \theta - y \cos \theta)^2 \sigma_X^2 \\ &= x^2 (\cos^2 \theta \sigma_Y^2 + \sin^2 \theta \sigma_X^2) + y^2 (\sin^2 \theta \sigma_Y^2 + \cos^2 \theta \sigma_X^2) \\ &\quad - x y \sin 2\theta (\sigma_X^2 - \sigma_Y^2) \end{aligned}$$

於是 x^2 及 y^2 係數之和 $= \sigma_X^2 + \sigma_Y^2 = \sigma_X^2 + \sigma_Y^2$

$$x^2 \text{ 及 } y^2 \text{ 係數之差 } = \cos 2\theta (\sigma_Y^2 - \sigma_X^2) = \sigma_y^2 - \sigma_x^2$$

故其係數為 σ_y^2 及 σ_x^2

$$\text{且 } \sin 2\theta (\sigma_X^2 - \sigma_Y^2) = \tan 2\theta (\sigma_x^2 - \sigma_y^2) = 2r \sigma_x \sigma_y$$

$$\text{故 } \frac{X^2}{\sigma_X^2} + \frac{Y^2}{\sigma_Y^2} = \frac{1}{\sigma_x^2 \sigma_y^2 (1-r^2)} \{x^2 \sigma_y^2 + y^2 \sigma_x^2 - 2r \sigma_x \sigma_y \cdot xy\}$$

而平面之方程式為

$$z = \frac{1}{2\pi \sigma_x \sigma_y \sqrt{(1-r^2)}} \cdot e^{-\frac{1}{2(1-r^2)} \left\{ \frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} - \frac{2rxy}{\sigma_x \sigma_y} \right\}} \dots \dots (102)$$

如上(公式98)所言,在假定——元素羣類係常態曲線——之下,理應如是。

此方程式 $\sigma_x^2 = S\sigma_u^2$, $\sigma_y^2 = S\sigma_v^2$, $r\sigma_x\sigma_y = S(r_p \cdot p\sigma_u \cdot p\sigma_v)$, 內, 係從公式99而來,其 rp 即為 p_u 與 p_v 之相關係數。

如將常數加進原來方程式內,則 $x = {}_1a_1u + {}_2a_2u + \dots$ 且 $y = {}_1b_1v + {}_2b_2v + \dots$ 而與分析程序無妨。

第六節 常態相關面之性質

以 x 與 y 變數之平均數為中心。

$$\begin{aligned} \text{容積} &= \int \int z dx dy \\ &= \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-r^2}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{1}{2(1-r^2)}\left(\frac{x}{\sigma_x} - r\frac{y}{\sigma_y}\right)^2} \cdot e^{-\frac{1}{2}\frac{y^2}{\sigma_y^2}} \cdot dx dy = 1 \end{aligned} \dots\dots\dots (103)$$

$$\begin{aligned} y\text{之第二動差} &= \int \int zy^2 dx dy \\ &= \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-r^2}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{1}{2(1-r^2)}\frac{x'^2}{\sigma_x^2}} \cdot y^2 \cdot e^{-\frac{1}{2}\frac{y^2}{\sigma_y^2}} dx' dy, \end{aligned}$$

其 $x' = x - r\frac{\sigma_x}{\sigma_y}y$; 然後對 x' 而積分, 則該式 (即 y 之第二動差)

$$= \frac{1}{\sigma_y\sqrt{2\pi}} \int_{-\infty}^{\infty} y^2 e^{-\frac{1}{2}\frac{y^2}{\sigma_y^2}} dy = \sigma_y^2 \text{ (依公式21之規定) } \dots\dots (104)$$

同理 x 之第二動差為 σ_x^2 。

$$\begin{aligned} xy\text{之平均積數} &= \int \int 2xy dx dy \text{ (積分之極限為 } \pm\infty \text{)} \\ &= \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-r^2}} \int \int \left(x' + r\frac{\sigma_x}{\sigma_y}y\right) e^{-\frac{1}{2(1-r^2)}\frac{x'^2}{\sigma_x^2}} \cdot y e^{-\frac{1}{2}\frac{y^2}{\sigma_y^2}} dx' dy \\ &= \frac{1}{\sigma_y\sqrt{2\pi}} \int r\frac{\sigma_x}{\sigma_y} y^2 e^{-\frac{1}{2}\frac{y^2}{\sigma_y^2}} dy = r\sigma_x\sigma_y \dots\dots\dots (105) \end{aligned}$$

$$\int \int zy x^3 dx dy = 3r\sigma_x^3\sigma_y, \int \int zx^2 y^2 dx dy = (2r^2 + 1)\sigma_x^2\sigma_y^2 \dots\dots (106)$$

由每一與 XOZ , YOZ 平行之平面, 所構成之部分, 係一常態

曲線。例如，以 $x = x_1$,

$$z = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-r^2}} \cdot e^{-\frac{1}{2(1-r^2)\sigma_y^2}\left(y - r\frac{\sigma_y}{\sigma_x}x_1\right)^2} \cdot e^{-\frac{1}{2}\frac{x_1^2}{\sigma_x^2}} \dots\dots\dots(107)$$

此即為一常態曲線，中心在 $y = r\frac{\sigma_y}{\sigma_x}x_1$ ，標準差為 $\sigma_y\sqrt{1-r^2}$ ，又最大縱坐標為

$$\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-r^2}} e^{-\frac{1}{2}\frac{x_1^2}{\sigma_x^2}}$$

相當一指定 x 值，各個 y 值組成之頻數羣類，自為常態形式，其標準差亦脫離 x 之關係。羣類之平均數（及其衆數與中位數），即在 $y = r\frac{\sigma_y}{\sigma_x}x$ 線上，（不論 x 值為何），此線即謂之迴歸線（本章第一節）。

$r\frac{\sigma_y}{\sigma_x}$ 即為 y 對於 x 之迴歸係數 (coefficient of regression)

同理，對一已知之 y 值， x_1 之頻數羣類亦屬常態，其標準差為 $\sigma_x\sqrt{1-r^2}$ ，其平均數係在 $x = r\frac{\sigma_x}{\sigma_y}y$ 線上，例如 $y = x \tan \phi_2$ 。

$r\frac{\sigma_x}{\sigma_y}$ 為 x 對於 y 之迴歸係數。

兩個迴歸係數之幾何平均為 r 。

橫面部分均為相似橢圓形。例如以 $z = z_1$,

$$\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} - 2r\frac{xy}{\sigma_x\sigma_y} = -2(1-r^2)\log(z_1 2\pi\sigma_x\sigma_y\sqrt{1-r^2}) \dots\dots\dots(108)$$

此等橢圓（以 $\sigma_x > \sigma_y$ ），任一之長軸，均與 x 軸，成 $-\theta$ 角，而 $\tan 2\theta$

$$= \frac{2r\sigma_x\sigma_y}{\sigma_x^2 - 2\sigma_y^2} \circ$$

$$\text{則 } \tan 2\theta = \frac{2r\sigma_x\sigma_y}{\sigma_x^2 - r^2\sigma_y^2}, \text{ 又 } \tan 2\phi_2 = \frac{2r\sigma_x\sigma_y}{r^2\sigma_x^2 - \sigma_y^2}。$$

如 $r = \pm 1$, 則 $\theta = \phi_1 = \phi_2$, 該面乃降為平面 $\frac{y}{\sigma_y} = \frac{x}{\sigma_x}$ 。反之, 如 $|r| < 1$, 則 $\phi_2 > \theta > \phi_1$, 而迴歸線乃落於該橢圓長軸平面兩邊之任一邊。如 $\sigma_x = \sigma_y$, 則 $\theta = \frac{\pi}{4}$, 而兩迴歸線便與包含數橢圓形主軸之平面成爲均等之角。

吾人於此應加注意, 一面純視五數量而定, 五數量者何? 兩平均數, 兩標準差, 及一相關係數也。

第七節 直線相關

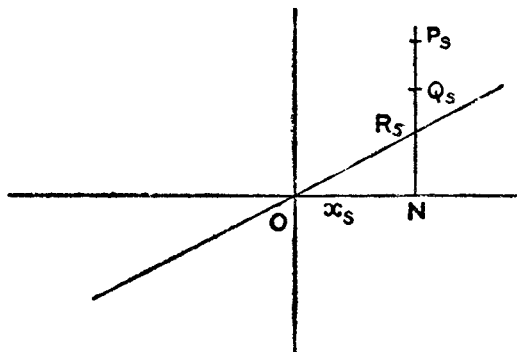
在某種情形下, 性質單純而具複雜原因之迴歸線——即對一變量(x)之已知值, 另一變量(y)之各列平均數之軌跡——係爲一經過多平均數所代表位置之直線, 關於此點, 前已論及。

但如條件未能嚴格滿足, 卻只能大略適應, 則須假定迴歸之直線性 (retilinearity), 亦能大略實現也。

然即使變量 x, y 並非常態分配, 迴歸依然大略合於直線; 雖分佈平面決不能用 r 值而決定, 而方程式 $y = r \frac{\sigma_x}{\sigma_y} x$ 仍可爲迴歸方程。

設相當 x 之一值 x_s , 整列中共有 n_s 個 y 值, 並設其平均數爲 \bar{y}_s 。以 $m_s = \frac{\bar{y}_s}{x_s}$, 則 m_s 乃爲僅由一羣類決定之迴歸線之斜度。

依本章第二節所論， $\tan\phi_1 = r \frac{\sigma_y}{\sigma_x}$ ，爲 $m_1, m_2, \dots, m_s, \dots$ 之加權平均數，至其權數則爲 $n_s x_s^2, \dots$ 也。



設 $ON = x_s$, NP_s 爲隨 x_s 而來之任何 y 值，又以 $NQ_s = \bar{y}_s$ 爲 n_s 此等數值之平均數。設一條線 $y = ax + b$ 與 NP_s 相遇於 R_s 點。

然則據猶爾氏（統計學報第一八九七年號第八百一十七至八百一十八頁）證明，經過 x_s 之所有各值後， $(R_s P_s)^2$ ——即 $S[y_s - (ax_s + b)]^2$ ，——所有各值之和，當於 $b = 0$ ，及 $a = r \cdot \frac{\sigma_y}{\sigma_x}$ 時，其值爲最小。此法根據最小二乘法，可參照附錄十。

此 $y = r \cdot \frac{\sigma_y}{\sigma_x} x$ 線經過各觀察值，經過時確使由各觀察值平行於 y 軸到該線之距離之平方之和成爲最小值。如是，無論分佈情況如何，該線始終爲迴歸之唯一良好代表。

請再進一步討論，試假定任一第 s 個整列內， y 之離散度，求爲 σ_2 且與 x_s 無關。因如各平均數有在一條直線之傾向，除在觀察

量不充足，以致平均數不能落在直線上爲例外外，則對平均數之離中差，如 $R_s Q_s$ ，必形成頻數曲線 $K e^{-\frac{(R_s Q_s)^2}{2\sigma^2}} \left(\sigma^2 = \frac{\sigma_2^2}{n_s} \right)$ 。

故離中差($R_1 Q_1, R_2 Q_2, \dots$ 之連帶機率(joint probability)爲 $K' e^{-\frac{1}{2\sigma^2} S n_s (R_s Q_s)^2}$ ，在 $S n_s (R_s Q_s)^2$ 爲最小時，此機率則爲最大。

$$\begin{aligned} S n_s (R_s Q_s)^2 &= S \{ n_s \{ \bar{y}_s - (a x_s + b) \}^2 \} \\ &= S (n_s \bar{y}_s^2) + a^2 S (n_s x_s^2) \\ &\quad + N b^2 - 2a S (n_s \bar{y}_s x_s) - 2b S n_s \bar{y}_s + 2ab S n_s x_s \\ (N &= n_1 + \dots + n_s + \dots) \end{aligned}$$

但 $S n_s \bar{y}_s = 0 = S n_s x_s$ ，因原點爲雙重平均數。

$$S (n_s x_s^2) = S x^2 = N \sigma_x^2, \quad S (n_s \bar{y}_s x_s) = S x y = N r \sigma_x \sigma_y$$

於是該式等於 $S n_s \bar{y}_s^2 + N (a \sigma_x - r \sigma_y)^2 + N b^2 - N r^2 \sigma_y^2$ 。

當 $b=0$ ， $a = r \frac{\sigma_y}{\sigma_x}$ 與前同，則此爲最小。

如假定一整列中之離中差，與 x 之相當值間，保持有直線性及獨立性，而距直線之離中差係由觀察太少而來，則此 $y = r \frac{\sigma_y}{\sigma_x} x$ 線，爲迴歸線之最或能之軌跡。

從此線計算得來之 $S n_s (R_s Q_s)^2$ 值，爲 $S n_s \bar{y}_s^2 - N r^2 \sigma_y^2$ 。

雖然，如測量之開始，或其結局，不能證明直線性之假定，則 r 爲分析中 useful 之參考雖有餘，但爲原因公同性度量之標準則不足。

第八節 相關率

爲不受關於觀察值分佈狀況假定之拘束，皮爾生氏制定一種量數，謂之相關率 (correlation ratio) —— (見 Draper's Company Research Memoirs, Biometric Series, II. 1905)。

假以 σ_y 爲第 s 個整列之標準差，使 $n_s \cdot s \sigma_y^2 = S(y_s - \bar{y}_s)^2$ ，又以 σ_a^2 爲 $1\sigma_y^2, 2\sigma_y^2 \dots$ 之加權平均值，使 $N\sigma_a^2 = S(n_s \cdot s \sigma_y^2) = SS(y_s - \bar{y}_s)^2$ ，裏面求和表示一整列內之各離中差相加，外面求和表示所有各整列之總和。

以 $N\sigma_m^2 = S(n_s \bar{y}_s^2)$ ，使 σ_m^2 爲各整列平均數之加權平均二次方。

於是 $N\sigma_y^2 = S(y^2)$ ，其求和普及於 y 之所有值，而且 $N\sigma_a^2 = S\{S(y_s^2) - 2\bar{y}_s S y_s + n_s \bar{y}_s^2\} = S(S y_s^2 - n_s \bar{y}_s^2) = S(y^2) - S(n_s \bar{y}_s^2)$ 。

$\therefore \sigma_y^2 = \sigma_m^2 + \sigma_a^2$ (如就別方面而論，其理甚明)。

於是 $\eta = \frac{\sigma_m}{\sigma_y} = \sqrt{\left(1 - \frac{\sigma_a^2}{\sigma_y^2}\right)} \dots \dots \dots (109)$

η 名曰相關率。所謂相關率，乃即若干整列之平均數散佈度，對於未編成整列之羣類散佈度之比率。

若 $\sigma_m = 0$ ，因而每一 $\bar{y}_s = 0$ 時，則 $\eta = 0$ 。換言之，即於每一整列之平均數，與該羣類之總平均數相合時，則其相關爲 0。

若 $\sigma_a = 0$ ，則 $\eta = 1$ 。換言之，如每一 $s \sigma_y = 0$ ，且每一整列之各項，

均集中於一點， \bar{y}_s ，時，則相關率為正一。

反是，則 η 小於1，而大於零， $1 > \eta > 0$ 。

在常態相關之下，每一 $s\sigma_y = \sigma_y\sqrt{(1-r^2)}$ ，如公式(107)，則
 $\sigma_a^2 = \sigma_y^2(1-r^2)$ ，而 $\eta^2 = r^2$ 。

又 $S n_s (R_s Q_s)^2 = N \sigma_m^2 - N r^2 \sigma_y^2 = N (\eta^2 - r) \sigma_y^2$

$$\eta^2 = r^2 + \frac{S n_s (R_s Q_s)^2}{N \sigma_y^2}, \dots\dots\dots (110)$$

且若非每一 $R_s Q_s$ 均為0，而各整列之各該平均數，均在 $y = r \frac{\sigma_y}{\sigma_x} x$ 線上時，則 $|\eta| > |r|$ 。

相關問題，討論至此，已可作一結論：

如 (x, y) 為一對測量——自平均數量起——兩個變量(關於時間，空間，或關於事物，或關於有機體)之量數，且當 x 已知為正(或負)數時，則假定 y 為正(或負)，或假定 y 為負(或正)；於是該兩個變量，乃謂之相關。在此情形下，當 n 增加時， $\frac{1}{n} S xy$ 不漸趨於零，而達一極限 $r \sigma_x \sigma_y$ 。所謂 $r=0, r=1, r=-1$ 者，各含有特定之意義，而 r 對 x 與 y 間，所有各種關係，感應均甚敏銳。依普通情形觀察， σ_a (各整列內平均散佈度)愈小，則 r 亦愈大。如 x 與 y 各為 $(p+q)$ 個獨立元素——在此若干元素中，只有 p 個為 x 與 y 所共有——之總和，則在各元素之標準差相等時， r 等於 $p/(p+q)$ 。假如 x 與 y 由許多複雜獨立原因——內有若干為 x, y 所

共有——作成直線，則(一)乃說明 r 各對之全部頻數分配，(二)迴歸線爲直線，(三)迴歸方程式，爲 $y = r \frac{\sigma_y}{\sigma_x} \cdot x$ 及 $x = r \frac{\sigma_x}{\sigma_y} \cdot y$ 。如常態頻數曲線，不能決定，但迴歸線卻爲直線，則此同一方程，亦足爲迴歸之經驗的良好說明。但如關於 x 及 y 之分配，或各整列之各該平均數，並無若何規定，則 r 數值，必毫無意義可言(r 在不等於0或1時，即係如此)。雖然，就一般而論， r 未嘗不可測量 x, y 共同原因系統之數額也。

第九節 未分級變量之相關

(Correlation between Ungraded Variables)

依照上述方法，測算相關，必須觀察有適當之詳密。未達到此詳細程度時，必發生有絕大興趣之問題。

頭 髮 顏 色

父 母

		淺	深	總 計	
兒	深	a	b	n_1	
	淺	c	d	n_2	
童		總計	m_1	m_2	N

以毛髮之深淺，將父母兒童各加區分；淺色髮之父母，所生 m_1 個兒童，髮爲淺色者 c 人，爲深色者有 a 人；同時深色髮之父母，所生兒童共有 m_2 個，其髮爲淺色者有 d 人，深色者有 b 人。設 $a + b$

$= n_1, c+d = n_2, \text{又 } n_1+n_2 = N = m_1+m_2。$

見此資料，吾人須審定，父母兒童之間，髮色深淺有無相關存乎其中，如果有相關，則請測算之。

在此情形下，如變量（譬如色質數額）果為常態分配，並有常態相關，則此問題即可確定。蓋 m_2 對於 N 之比， n_1 對於 N 之比，即得（用第二章常態機率表，反查即得）色質尺度之橫坐標，而此色質尺度，又依髮色之深淺，而劃分為二部；不論 r 為何值，由此等橫坐所成平面之一部分相關面，既為已知，且 b/N 對於此者之方程式，亦即反求 r 之方程式也。

皮爾生氏 (Phil. Trans. A. Vol CXCIV, 第一頁起)，艾得頓氏 (『頻數曲線』第七章)，均曾作應有之分析，而終局得一求 r 之方程式，由此方程式 r 可約略求出也。

如其資料吾人可得而支配，並能將兩變量由中位數各分為兩部，則可得一簡易之解法。

設學算術及學代數之智力，係依常態而分配。將一大羣之學生，依其學算術之智力（以分數為準，或用其他方法）為次第，列為一類，復依學代數之智力，另分列為一類；假以 b 為二者均在中位數以上之學生， c 為均在中位數以下者，其學算術在中位數之上，學代數在中位數之下者，有 d 人，學代數在中位數上，學算術在中位數下者，有 a 人。然此並非謂智力確經測量，不過僅排列

其次第而已。

$$\text{於是 } a+b = \frac{N}{2} = c+d = a+c = b+d,$$

$$\therefore a=d = \left(\frac{1}{4}-q\right)N, b=c = \left(\frac{1}{4}+q\right)N.$$

$$r = \sin 2\pi q$$

其理容下再述。定每個尺度之標準差爲一。設所求之平面爲

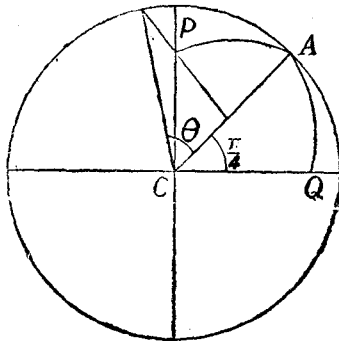
$$\frac{1}{2\pi\sqrt{1-r^2}} e^{-\frac{1}{2(1-r^2)}(x^2+y^2-2rxy)}$$

。則該平面之主軸，與 x 軸形成

$\frac{\pi}{4}(\frac{1}{4}+q)$ 等於 x 與 y 均爲正數之象限內之容積，該象限乃由平面

$y=0, x=0$ 所成，至此等平面又將相似橢圓形之水平部分，各切

去其面積之 $(\frac{1}{4}+q)$ 。茲有橢圓形 $x^2+y^2-2rxy=1$ 。



在上圖，橢圓形面積 $CPQ =$ 橢圓形面積之 $(\frac{1}{4}+q)$

以 θ 爲 P 之偏角 (eccentric angle)。

以其主軸而言之橢圓形爲

$$(1-r)x^2 + (1-r)y^2 = 1.$$

$$\frac{\tan\theta}{\tan\frac{\pi}{4}} = \frac{\text{長軸}}{\text{短軸}} = \sqrt{\frac{1+r}{1-r}}.$$

$$\therefore r = -\cos 2\theta$$

$$\frac{1}{4} + q = \frac{\text{CPA之兩倍面積}}{\text{橢圓形面積}} = \frac{2\theta}{2\pi}$$

$$\therefore 2\pi q = 2\theta - \frac{\pi}{2}$$

$$\sin 2\pi q = -\cos 2\theta = r$$

例如兒童為百分之四十，在兩者中均在中位數以上，

$$\frac{1}{4} + q = 4, q = \cdot 15, r = \sin \frac{3}{10}\pi = \cdot 81.$$

如 $q=0, r=0$ ，如 $q=\frac{1}{4}, r=1$ 。

據布魯昂(W. Brown)之實驗(見 *Biometrika*, Vol. VII, p. 366)，八十三名兒童中，代數智力高於中位數，而算術智力低於中位數者，佔一十一名。然則 $q = \frac{1}{4} - \frac{1}{8} = \cdot 12$ ，故 $r = \cdot 68$ 。布氏用全列次第——不只用中位數——計算，得知為 $\cdot 65$ ，又用分數得數 $\cdot 79$ 。至於兒童在校中之地位及年齡亦有關係，不可不加以修正，布氏亦曾注意及之。

但如此兩種品質，分配不能形成常態，則相關量之測算問題，即不能得以確定，歷經用多數方法，亦不能達到目的也。

第十節 相聯

如另無因果關係存乎其間，深色髮之父而有深色髮之子，其父親人數當為 $\frac{m_2}{N} \times \frac{n_1}{N} \times N = \beta$ ，即從 N 個事件中，深色髮者，父當有 m_2 人子有 n_1 人。

前於第九節所用符號仍舊，並以 a, γ, δ 代表 a, c, d 各區分內或然性最大之人數，則

$$a + b = a + \beta = n_1, a + c = a + \gamma = m_1 \dots \dots \text{餘類推。}$$

$$\begin{aligned} a - a &= b - \beta = c - \gamma = \delta - d = b - \frac{m_2 m_1}{N} \\ &= \frac{b(a + b + c + d) - (b + d)(a + b)}{N} = \frac{bc - ad}{N} = qN. \end{aligned}$$

q 即為相聯 (association) 之量數，其意義除在特別情形之外不能確定。

猶爾氏 不用 q 而用 $Q = \frac{bc - ad}{bc + ad}$ ，即相聯係數 (The “coefficient of association”) 或 $w = \frac{\sqrt{bc} - \sqrt{ad}}{\sqrt{bc} + \sqrt{ad}}$ ，即相連係數 (The “coefficient of colligation”) 為量數。(參閱 Introduction to Theory of Statistics, p. 37, 及 Statistical Journal, 1912, p. 593)。當 $bc = ad$, $q = 0$ 時, $Q = w = 0$ ，此為不相聯之例；如 a 或 d 為零, $Q = w = 1$ ； b 或 c 為零, $Q = w = -1$ ，此為相聯之最高量。

此等係數，在極端情形下，其意義乃可確定，但如以 $Q = \frac{2}{3}$ ，其命義如何，非檢視若干試驗不能知，且最後，不能以 Q 之大小而斷定相聯量之大小，因相聯一詞並無確定之量數意義也。

第十一節 相依

如吾人不以求得相聯量爲止，更進而求其存在之證據，換言之，即在品質彼此獨立時，求其觀察是否仍能發生，則理論基礎當愈爲鞏固矣。

如 $p:1-p$ ，爲兒童深色髮者對淺色髮者之比率，則 $p = \frac{n_1}{N}$ 爲自觀察中所得之最佳值。

故如將 N 名兒童武斷分爲兩類（例如以其基督教名自 A 至 K 爲一類，自 L 至 Z 爲一類），使分別佔有 m_1 及 m_2 人，則在第一類找到 a 之機率（已知第二章第十節所述），乃爲

$$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}, \left(\frac{1}{\sqrt{m_1}} \text{業已略去} \right)$$

但 $x = a - pm_1 = a - \frac{n_1 m_1}{N} = a - a = qN$ （註三）

$$\sigma^2 = p(1-p)m_1 \left(1 - \frac{m_1}{N} \right) = \frac{n_1}{N} \cdot \frac{n_2}{N} m_1 \cdot \frac{m_2}{N}。$$

離中差，或正或負，其等於 $(a-a)$ 時，出現之機率，爲

$$2 \int_{z\sqrt{2\pi}}^{\infty} \frac{1}{z\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz, \left(z = \frac{x}{\sigma} \right)。$$

$$\begin{aligned} \text{請注意 } \frac{x^2}{\sigma^2} &= \frac{q^2 N^5}{n_1 n_2 m_1 m_2} = q^2 N^3 \cdot \frac{(n_1 + n_2)(m_1 + m_2)}{n_1 n_2 m_1 m_2} \\ &= q^2 N^3 \left(\frac{1}{n_1 m_1} + \frac{1}{n_1 m_2} + \frac{1}{n_2 m_1} + \frac{1}{n_2 m_2} \right) \end{aligned}$$

$$= \frac{(a+\alpha)^2}{\alpha} + \frac{(b-\beta)^2}{\beta} + \frac{(c-\gamma)^2}{\gamma} + \frac{(d-\delta)^2}{\delta} = X^2.$$

例如，在分配 $\frac{65}{35} \frac{235}{165}$ 內

$$n_1 = 300, n_2 = 200, m_1 = 100, m_2 = 400, N = 500,$$

$$\sigma^2 = 19 \cdot 2, \alpha = 60, x = qN = 5,$$

$$\frac{x}{\sigma} = 1 \cdot 14, F(1 \cdot 14) = \cdot 373 \text{ (常態機率表)}, 2\{\frac{1}{2} - F(1 \cdot 14)\} = \cdot 254.$$

從500中抽出100個，出現者在65以上或55以下者之機率，為·746對·254之比，即約為三對一之比。

n_1, m_1, N 及 a 為已知，其餘 b, c, d, n_2, m_2 亦已知，則所得之機率，恰為均等之機率，足使 b, c, d 任何一數之出現，彼此完全獨立。此不能謂之為全部分配之機率；欲求此機率，吾人一則須知，機率 p 係由較廣之宇宙間而來，二則須知非如 $\frac{n_1}{N}$ 於有限之觀察個數即可決定者，三則須知 $\frac{m_1}{N}$ 即為對總機率之近似值。

為例釋此難點，請就上已論過之問題，再考慮之。

	未 種 牛 痘	已 種 牛 痘	總 計
痊 愈 者	a	b	n_1
死 亡 者	c	d	n_2
總 計	m_1	m_2	N

天花流行病患者，共有 N 人，就中種痘者為 m_2 人，死亡者有

n_2 人，其他範疇俱見上表。

痊愈率，如該項全部統計所述，乃為 $\frac{n_1}{N}$ ，又如種牛痘（無論直接影響，或其他連帶關係）與病之痊愈無關，則種痘及未種痘患者痊愈之機率，必為 $\frac{n_1}{N}$ ，而在 m_2 個種痘患者中， b 人業已痊愈之機率為

$$\int_{b-\beta}^{\infty} \frac{N^{\frac{3}{2}}}{\sqrt{2\pi n_1 n_2 m_1 m_2}} \cdot e^{-\frac{1}{2} \cdot \frac{x^2 N^3}{n_1 n_2 m_1 m_2}} \cdot dx,$$

($qN = b - \beta = x$)；如此數甚低，則種痘（或環境使然）與痊愈果有關係，乃有證明。

然 $\frac{n_1}{N}$ 一比率，須受 $\sqrt{\frac{n_1 n_2}{N}}$ 之標準差之影響，如此數不能略除，則其及於計算結果之影響，必須加以測驗。

如欲測量種痘（不問其中數種影響之存在有無證據）之利益（或弊點），即可比較 $\frac{b}{m_2}$ 及 $\frac{a}{m_1}$ 二痊愈率而得，但此法必須具有此等痊愈率之說明及其標準差，舍此二者，無他法可以直接求得。

相聯之存在及其測算之問題，如每種品質之交替數，並不如是之簡單，則變為複雜之問題。在複雜之情形下，吾人乃有若干不同之種類，如父子髮之色質有多種類別是。

皮爾生教授，為此種事實之測算，特設『相依係數』(coefficient of contingency)。(參閱 Drapers' Company Research

Memoir, Biometric Series I, 1904; 及艾得頓書之第十章)。

觀 察 次 數				
第 一 品 質 分 類				
類	a_1	a_2	n_1
分	b_1	b_2	n_2
品	c_1	c_2	n_3
質
第
m_1 m_2				N

設 n_1, n_2, \dots 為各行 (line) 之總數，以 m_1, m_2, \dots 為各列 (column) 之總數，並以 N 為觀察總次數，如上表所示。

如是，設 $n_1/N, m_1/N$ 為每一品質之第一類對總數之確實比例，則在無『相聯』時， a_1 處出現之最或然數，將為 $a_1 = \frac{n_1}{N} \times \frac{m_1}{N} \times N$ 。同樣， $a_2, \dots, \beta_1, \beta_2, \dots, \gamma_1, \gamma_2, \dots$ ，亦可為其他地位算出。

關於 $a_1 - \alpha_1, a_2 - \alpha_2, \dots, b_1 - \beta_1, \dots$ 各等項分歧，可供為相聯之測算。因超過之數，與不足之數，二者機會均等，故以用二次方， $(a_1 - \alpha_1)^2, \dots$ 較直線數量為便。

以上論四項範疇之例為準，得函數如下：

$$\chi^2 = \frac{(a_1 - \alpha_1)^2}{\alpha} + \frac{(a_2 - \alpha_2)^2}{\alpha_2} + \dots + \frac{(b_1 - \beta_1)^2}{\beta_1} + \dots + \dots \quad (111)$$

此函數用於測算，公式代表已知觀察之適切度，亦頗佔相當地位。

相依係數 (coefficient of contingency) 之定義，乃如下述

$$C = \frac{1}{\sqrt{\left(1 + \frac{N}{X^2}\right)}} \dots \dots \dots (112)$$

$X=0$ 時，則無相聯 (association)，而 $C=0$ 。

$$\frac{X^2}{N} = \frac{(a_1 - \alpha_1)^2}{n_1 m_1} + \frac{(a_2 - \alpha_2)^2}{n_1 m_2} + \dots \dots \dots,$$

只隨比率而變，與所測之全數無關。

據艾得頓氏 (Elderton 書第一百四十七頁) 證明，如 a_1, a_2 b_1 等數，出現於常態相關面之相當區分內一項亦不下於一小整數，則 C 將漸近於 r ，即相關係數。此種關係乃形成 X^2 之函數，而 C 之定義乃由 X^2 而出。

所得之 C 值，因所用之區分數而有不同，惟其如此，故其可用為測算之效用大減；幸現已有新法，足以克服此困難矣（見 *Biometrika*, Vol. IX pp. 116-139）。

至於得出某一 C 值，其意義若何，非具有極多試驗之經驗，不能領略也。

抑有進者， C 以及前節所論之法，羣類有限制之事例而無可以測量之性質者，用之甚佳。

第十二節 時間數列之相關

前文所論之相關，均為關於同時之兩統計羣類；關於兩個數

列，以 $x_1, y_1, \dots, x_p, y_p, \dots, x_t, y_t$ 各對為連續時間中數量之測量，吾人不得不測驗兩數列關係之法。但現所提出者，乃為 x 之一值，與其以往及後來之數本有關涉，不能互相獨立，而 x 與 y 之關係，僅能在時間上影響及於一般定期之進度，並無密切之聯系。時間上所有數列，幾均有一趨勢(trend)，此等趨勢，無論其依等速進行與否，抑或方向如何，均可有高度之相關係數，甚至數量非彼此獨立時亦有之。例如，介乎

$$1, 2, 3, \dots, 20$$

$$100, 98, 96, \dots, 62$$

($x_t = t, y_t = 100 - 2(t - 1)$)二者之相關係數為1。

求離中差之相關，必須將時間因素消除。

此外尚有一法(註四)，此法，一則可將每一種數量之修勻線求出；二則可計算每年之觀察值，與修勻線所得數值之差額；三則即以此種差額為據以測算相關之數量；換言之，即測量此種數量與圖式(見上編第七章第三節)所代表數值之相關。

如該數列特別有週期性，則其結果必至得出屬於週期性之相關，而此以用倒數分析法(harmonic analysis)為佳。又如此種數列係為『補償的』(compensated)，則於一正離中差之後，必隨一副離中差，其兆徵即可用相關反應出來。

惟在反覆擺動無一定標準之時，則一年之量數，並無一定之

趨勢，且與臨近各年之量數無關，於是從修勻之線，算出離中差間之相關係數，所測量之『關係』，與前論各羣類間之相關者，完全相同。

以 x_p, x_p ，為第 p 年為其中心， m 為奇數。則欲求之相關係數，乃在表示 $x_p - \bar{x}_p$ 與 $y_p - \bar{y}_p$ 之相關。

如線在修勻時，係用斐孫方法 (Professor Persons, The Review of Economic Statistics, No. I, 1919, Harvard University Press), 此法亦可應用。在此法之下，須將 m 年之平均數， \bar{y} ，求出 (m 年之趨勢，須在同一方向，而斜度須無顯著之變動，且如 t 為距期間中心之年數，修勻之線須成為 $y - \bar{y} = kt$ 之形式)。至 k 之決定，係由於一條件——對此線之離差平方之和必為一最小值，換言之，即 $S\{y_t - (\bar{y} + kt)\}^2$ 為一最低數 (y_t 為自中心 t 年之觀察) ——而得。故 $k = \frac{Sty_2}{St^2}$ 。

$$\text{如 } k = 2n + 1, St^2 = 2(1^2 + \dots + n^2) = \frac{n(n+1)(2n+1)}{3}。$$

此種方法，當趨勢繼續凹下（或凸出），而平均數常大（或小）於觀察值時，用於以移動平均法求出之線上，亦可戰勝一部困難。

更有一法，係克孚女士 (F. E. Cave, 文見 Royal Society's Proceedings, Vol LXXIV, p. 407, 1904) 及虎克 (Hooker), 見 Statistical Journal, 1905, 自六百九十六頁起) 所採用，近

更經皮爾生教授，克孚女士及其他等人，加以發揮。此法並非依觀察值求其相關，而係以相接觀察之差額以求相關。有 $m+1$ 年之時期於此，各年觀察值分別為 x_0, x_1, \dots, x_m 及 y_0, y_1, \dots, y_m ，相關係數即由以下各對： $x_1 - x_0, y_1 - y_0, x_2 - x_1, y_2 - y_1, \dots, x_m - x_{m-1}, y_m - y_{m-1}$ ，算出

因 $x_1 - x_0, x_2 - x_1, \dots$ 各等數量之平均數，為 $\frac{1}{m}(x_m - x_0)$ ；求 r 之公式，必須用對平均數之離中差；而此離中差，即為某年增值超過平均增值之部分，換言之，即對平均數之每年變量。如 x 與 y 之修勻線，顯著為凹形（或凸形），相關必由此兆徵而受支配，但如觀察值對一直線反覆擺動無一定標準時，則吾人可得測量相關，而不受時間因素之影響。

為免除凸凹形之困難，另有一較為詳密之方法，此即皮爾生教授所謂變差相關（variate difference correlation）者是也（註五）。此法之基礎，係假定 x_p 可示為 $x_p = X_p + bt_p + ct_p^2 + \dots$ ，（ X_p 不受時間影響），而時間影響，以拋物線函數表示之，同理 $y_p = Y_p + b't_p + \dots$

$$x_{p+1} - x_p = X_{p+1} - X_p + b + c(2t_p + 1) + \dots$$

虎克氏方法，刪略 c 及其他常數。

用第二級差額，則得

$$x_{p+1} - 2x_p + x_{p-1} = X_{p+1} - 2X_p + X_{p-1} + 2c + 6 + dt_p + \dots$$

用時復略卻 d 以及其他常數，使成一嚴格之拋物線形。

由此可知，如將時間因素消除，任何差額間之相關即等於 X_p 與 Y_p 之相關。當進而更高級差額時，相關係數不至受其影響，此種程序便告完成。

雖然，差額往往因觀察時之精度及小數之有效數字關係，除為首三級尚可用外，其他差額應用此法，不免招致絕大困難。此中關係可由 $2.6, 2.7, \dots$ 等數之二次方，仍取其至小數點後一位時知之。

6.8	Δ	Δ^2
7.3	.5	0
7.8	.5	.1
8.4	.6	0
9.0	.6	0
9.6	.6	0
10.2	.6	.1
10.9	.7	

第二級差額，如書全式，當全為 $.02$ 。實則此法未免過於詳密，普通統計之觀察用之，似不適宜。茲將各種方法之差別，示之如下：

x 量既已發生相關，如取用第四級差額，當為

$$6\{x_0 + \frac{1}{6}(x_2 + x_{-2}) - \frac{2}{3}(x_1 + x_{-1})\},$$

式中附數表示自中心向左向右之距離。此極端項，將該式增大許多。

以第五項爲基礎，用移動平均數法，則量數當爲

$$x_0 - \frac{1}{5}(x_{-2} + x_{-1} + x_0 + x_1 + x_2) = \frac{4}{5}\left\{x_0 - \frac{1}{4}(x_2 + x_{-2}) - \frac{1}{4}(x_1 + x_{-1})\right\}$$

而其極端項又將該式減少許多。

用第八級差額，則得

$$x_0 - \frac{4}{5}(x_1 + x_{-1}) + \frac{2}{5}(x_2 + x_{-2}) - \frac{4}{35}(x_3 + x_{-3}) + \frac{1}{70}(x_4 + x_{-4})$$

用移動平均數，則爲

$$x_0 - \frac{1}{8}(x_1 + x_{-1}) - \frac{1}{8}(x_2 + x_{-2}) - \frac{1}{8}(x_3 + x_{-3}) - \frac{1}{8}(x_4 + x_{-4})$$

但第二級差額爲

$$x_0 - \frac{1}{3}(x_1 + x_0 + x_{-1})$$

此與僅用三項之移動平均法所得相同，可知不免於極大之機率差誤(chance error)。

各種方法必須加以詳細研究，並須多用以長經驗。據吾人所知，移動平均法對極端各項，頗有看輕之勢，而差額法則每因觀察之粗率，感受嚴重之影響。總而言之，兩種方法，相關之測量，均以假設爲基礎，非如測量羣類相關時之意義鮮明也。

第十三節 時間數列之圖式比較法

除決定相關量數外，尙有一問題，卽如何以圖式表示其關係。方法可示如下：

以 $x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_n$ 爲對移動平均數之離中差（如

第七章第六例之表)，或使——在無趨勢時——代表實在量數，復設 \bar{x} 與 \bar{y} 為其平均數。

構造一圖，表示 y 值，尺度之決定——以便利為主，時間作為橫軸。於是 x 值在此圖上，以任何尺度，任何原點，均無不可。

以 b 為 x 之原點，以一單位 x 相當 c 個單位 y 。決定 b 與 c 時，使距代表 x_1, y_1 ，各對之點，垂直距離平方之總和，須為最小為便。（見附錄十）。

換言之， $S\{c(x+b)-y\}^2$ ，其值為最小。

對 b 與 c 而微分之，得 $c = \frac{S(x-\bar{x})(y-\bar{y})}{n\sigma_1^2}$ ，又 $c(\bar{x}+b) = \bar{y}$ ，（ σ_1 為 x 之標準差）。

各離中差之平均數，必在縱標尺上之同一點，且與 x 之平均數之差，必乘以 $\frac{r\sigma_2}{\sigma_1}$ 然後在 y 軸上以平均數為中心，向上向下分別測量（ σ_2 為 y 之標準差， r 為相關係數）。

關於此法，1912年出版之統計學報（Statistical Journal）第799-800頁，曾用一例，可以參考。

（註一） X, Y, U, V 之意義，自與本章第三節所用者不同。

（註二） 見公式100。

（註三） q 之意義與前數節所用者相同，並非指第二章之 $1-p$ 也。

（註四） 請參閱統計學報一九〇一年號，第四百八十五頁起，虎克（Hoeker）一文。

（註五） 見 *Biometrika*, Vol. X. 自第一百九十七起，及自三百四十頁起。

第七章 相關例證

茲爲例證上述理論並示明測算步驟起見，特就數種實驗及觀察之結果，一一縷列於下。

在未開始舉例之前，有一先決問題，必須釋明者，即 r 之理論值，僅可於無量數觀察值中得之。前論機誤時，曾言 r 由 n 對中算來，去其真值，必有若干相差，此相差數額之標準差，如自差誤常態尺度 (normal scale of error) 算出，當爲 $\frac{1-r^2}{\sqrt{n}}$ 。如在例一，相關係數爲 $\cdot 6$ ；其變量共有二十四對，則其去真值之差額當在 $\cdot 6$ 之 $\frac{1-6^2}{\sqrt{24}} = \cdot 13$ 以內；但若差額達 $\cdot 13$ 之三倍，乃幾爲不可能之事實。反之，如吾人不知真實係數爲若干，務須以計算值 $\pm \frac{1-r^2}{\sqrt{n}}$ 爲宜。

下舉例證數則之目的，僅在示知從觀察中，求 r 之計算方法。其他各例，則釋明當觀察量數甚多時，求得各整列之平均數，並與方程式 $y - \bar{y} = r \frac{\sigma_2}{\sigma_1} (x - \bar{x})$ (在迴歸線爲直線時，此爲衆平均報之軌跡) 比較之法。至於最後一則，以 1000 對之分配，逐一與理論相關面之分配比照。一般言之， x 與 y 之計算，並非由其平均數，而係自一武斷之原點，故 $r = \left(\frac{S_{xy}}{n} - \bar{x}\bar{y} \right) \div \sigma_1 \sigma_2$ ，見公式 93。

例一：爲求得一實例，以解相關起見，可在所有環境均爲已知，而相關係數可用由果推因方法以事推出時，用一數學表從中任意抽出若干數字。以五個數字之和作爲 x_i ，以任一五個數字之和作爲 y_i ，但後一五個數字之中，有三個業已被包括於前一五個數字之內，其餘二個數字，則爲不同，如此共湊成二十四對 $(x_i y_i)$ …… $(x_i y_i)$ ……。此等各對之相關係數，爲 $\frac{2}{3}$ (公式96)。在一僅取二十四對之實例中，相關係數爲 0.537；相關係數 $\frac{2}{3}$ 之標準差爲 $\frac{1 - .36}{\sqrt{24}} = .13$ ，故對若是其小之數目，所差並不甚巨。

計算方法，可以下列一表示明之。

x	y	x^2	y^2	xy
22	32	484	1,024	704
27	27	729	729	729
12	19	144	361	228
21	30	441	900	630
21	26	441	676	546
27	26	729	676	702
23	25	529	625	575
17	22	289	484	374
25	23	625	529	575
11	9	121	81	99
16	24	256	576	384
20	28	400	784	560
37	29	1,369	841	1,073
33	25	1,089	625	825
18	20	324	400	360
24	26	576	676	624
22	17	484	289	374
17	16	289	256	272
32	27	1,024	729	864
29	29	841	841	841
26	17	676	289	442
27	20	729	400	540
26	26	676	676	676
21	17	441	289	357

$$n = 24 \quad \bar{x} = 23\frac{1}{2} \quad \bar{y} = 23\frac{1}{2}$$

$$24\sigma_1^2 = 13706 - 24(23\frac{1}{2})^2$$

$$\sigma_1 = 6.18$$

$$\sigma_2 = 5.56$$

$$r = \frac{13354 - 24\bar{x}\bar{y}}{24\sigma_1\sigma_2}$$

$$= .537$$

554 560 13,706 13,756 13,354

如 x 與 y 各以 23 為原點而計算，其計算手續必較簡易。

例二：吾人如有稀少而散見之觀察值，則將全部完全計算出來亦頗簡易。舉例言之，茲將二十六個城市之嬰兒死亡率，列如次表，以與此二十六個城市之人口（以一千為準）比較。然則 r 僅當其標準差之兩倍，因而其準確數值亦欠一定，但吾人可得而證明者，城市愈大者，死亡率亦愈高。如欲切實研究嬰兒死亡原因問題，必須多舉事例，並考慮除人口略數以外之其他種因素。

二十六個城市之人口與嬰兒死亡率

人口 x	死亡率 y	xy
000.		
55	162	8,910
39	201	7,839
36	241	8,676
35	182	5,670
31	179	5,549
30	174	5,220
27	176	4,752
24	208	4,992
24	168	3,912
23	206	4,738
22	172	3,784
20	200	4,000
19	218	4,142
19	198	3,762
19	132	2,508
16	155	2,480
15	148	2,220
15	220	3,300
15	141	2,115
12	169	2,028
7	155	1,085
6	129	774
6	167	1,002
5	150	750
5	171	855
4	161	644
總計	5,529	4,558
		95,707

平均數 20.354 175.31 -

$\sigma_1 = 12.1$ $\sigma_2 = 27.9$

$$r = \frac{Sxy - n\bar{x}\bar{y}}{26\sigma_1\sigma_2}$$

$$= \frac{95707 - 26 \times 20.354 \times 175.31}{26 \times 12.1 \times 27.9} = .34$$

$$r \text{ 之標準差} = \frac{1 - .34^2}{\sqrt{26}} = .17$$

例三：北海漁業調查所(North Sea Fisheries Investigation)，之青魚統計，可為一良好例證。該種青魚身上具有圈輪 (ring)，每年生一輪，故可為年齡之表明，此項統計即以魚之體量與圈輪之關係為題材。

原始曲線雖有斜度；但各整列之平均數，均甚臨近理論上之迴歸直線。

表上， y 軸代表魚之體量，原點為 31 cm. 單位為 1 cm. 圈輪數以 x 軸為代表，以七輪為原點，以一輪為單位，出現次數則填於方格表內。

n_2 為 y 某一種值之次數，最末二列記 n_2y 與 n_2y^2 及其總和，以便計算圈輪之平均數及標準差。同樣，以 n_1 為某一 x 整列之總次數，由 n_1x 與 n_1x^2 之各該總和，算出青魚體量之平均數及標準差。

最後一行，列明每一整列之平均數，此乃在每一 x 整列中，用其出現次數乘相當之 y 值而得。

在每一次數之下，在括弧中，註明相當之 $x \times y$ 值；例如，在 $x = -1$ 列下， $y = 3$ 一排 (row) 共有四件，而 $xy = -3$ ，故四件與 xy 之總值，相乘而得 $4 \times -3 = -12$ 。如此乘得之各項，分別彙類下列四象限內。

至於原點之決定，以使 $\sum xy$ 之各零項，愈多愈妙（請參照猶

爾氏著：統計原理概論第一百八十三頁)。

青魚之圈輪數及體量 (長度單位為公分)

圈輪數 x														總計		
	4	5	6	7	8	9	10	11	12	13	6	n_2	n_2y	n_2y^2		
Size. cm. y																
35 4	—	—	1	—	1	2	2	—	—	—	—	6	24	96		
			(-4)		(4)	(8)	(12)									
34 3	—	1	4	4	15	14	7	3	1	1	50	150	450			
		(-6)	(-3)		(3)	(6)	(9)	(12)	(15)	(18)						
33 2	1	1	11	26	26	22	11	3	3	1	105	210	420			
	(-6)	(-4)	(-2)		(2)	(4)	(6)	(8)	(10)	(12)						
32 1	1	24	49	53	26	7	5	1	—	—	166	166	166			
	(-3)	(-2)	(-1)		(1)	(2)	(3)	(4)	—	—						
31 0	—	28	43	45	21	6	2	—	—	—	145	0	0			
30 -1	1	15	21	16	7	1	—	—	—	—	61	-61	61			
	(3)	(2)	(1)		(-1)	(-2)										
29 -2	2	3	5	—	—	—	—	—	—	—	10	-20	40			
	(6)	(4)	(2)													
28 -3	1	1	3	2	—	—	—	—	—	—	7	-21	63			
	(9)	(6)	(3)													
27 -4	—	3	—	—	—	—	—	—	—	—	3	-12	48			
		(8)														
26 -5	—	1	—	—	—	—	—	—	—	—	1	-5	25			
		(10)														
總計 n_1	6	77	137	146	96	52	27	7	4	2	554	431	1,369			
n_1x	-18	-154	-137	0	96	104	81	28	20	12	$\Sigma n_1x = 32$					
n_1x^2	54	308	137	0	96	208	243	112	100	72	$\Sigma n_1x^2 = 1330$					

各整列之

平均數 30.17 30.85 31.34 31.65 32.25 32.92 33.07 33.3 — —

$\bar{x} = 32 \div 554 = 0.0578$ 。平均數, 7.0578 輪。

$\sigma_1^2 = 1330 \div 554 - 0.0578^2 = 0.083$ (註一) $\sigma_1 = 1.521$ 。

$\bar{y} = 431 \div 554 = 0.778$ 。平均數, 31.778 公分。

$\sigma_2^2 = 1369 \div 554 - 0.778^2 = 0.83$ (註二) $\sigma_2 = 1.335$ 。

S_{xy}	++	--	+-	-+	
4	52	3	7	4	
16	88	30	2	6	
24	66	21	—	12	$S_{xy} = 636 + 146 - 9 - 154 = 619$
45	24	12	9	6	$r = \frac{S_{xy} - 554xy}{554\sigma_1\sigma_2} = .528$
84	30	12		4	r 之標準差 = .026
63	12	10		22	
36	26	9		3	
15	14	6		48	
16	15	9		49	
	4	24			
		10			
	636	146		154	

$$\frac{\text{體長} - 31.778 \text{公分}}{1.335 \text{公分}} = .528 \frac{\text{輪數} - 7.0578}{1.521}$$

輪	數	從方程式推算之體長 公分	整列之平均數(公分)
	4	30.36	30.17
	5	30.82	30.85
	6	31.29	31.34
	7	31.75	31.65
	8	32.21	32.25
	9	32.68	32.92
	10	33.14	33.07
	11	33.60	33.3

例四：下舉一例，釋明相關甚小時，求 r 所得之值。以七位對數表，2500-2549，及 2600-2649，二組中每一數之末位字為 x ；以比 x 大 50 之對數末一位字為 y ，換言之，即以 2550-2599 及 2650-2699 之每數末一字為 y 。求得 r 為 .086，此值在一百對中乃小於其標準差。

茲又採用一算術計算法，此法較之本章之其他方法均為簡便，且可即時算出相關比率。

成對數字之出現數

x	0	1	2	3	y 4	5	6	7	8	9	n_x	S_y	xS_y	$ns\bar{y}^2$
0	—	—	1	—	1	1	—	—	—	4	8	53	0	351
1	3	—	—	1	1	—	—	1	1	1	8	31	31	120
2	—	2	1	—	3	1	—	—	—	1	8	30	60	112
3	2	—	—	1	—	1	—	2	2	3	11	65	195	334
4	1	2	2	—	—	2	1	3	1	—	12	51	204	217
5	1	1	—	—	1	1	4	1	—	—	9	41	205	187
6	1	1	1	2	—	1	—	2	1	—	9	36	216	144
7	3	3	2	1	1	3	1	—	3	1	18	68	476	257
8	—	2	—	3	2	1	—	—	2	1	11	49	392	218
9	—	—	—	1	—	—	—	1	1	3	6	45	405	338
	11	11	7	9	9	11	7	10	11	14	100	469	2,184	2,328

$$\bar{x}=4.72, \sigma_x=2.69, \bar{y}=4.69, \sigma_y=3.03, n=100,$$

$$S(x-\bar{x})(y-\bar{y})=Sxy-100\bar{x}\bar{y}=S(xS_y)-100\bar{x}\bar{y}=2184-2114=70$$

$$r=\frac{70}{100\sigma_x\sigma_y}=.086. \quad r \text{ 之標準差爲 } .1.$$

茲以 n_x 爲各個 x 數 0, 1, …… 出現之次數。 S_y 爲相當之 y 值總和；例如，第一行爲 $2+4+5+6+9 \times 4=53$ 。 \bar{y}_x 爲對某一 x 值之 y 值平均數，等於 $S_y \div n_x$ ，

$$\text{又} \quad ns\bar{y}_x^2 = (S_y)^2 \div n_x.$$

相關比率可由最末一列算出（見第六章第八節）。

$$\begin{aligned} 100\sigma_m^2 &= Sn_s(\bar{v}_s - \bar{y})^2 \text{ (註三)} = Sn_s\bar{y}_s^2 - 2ySn_s\bar{y}_s + n\bar{y}^2 \\ &= Sn_s\bar{y}_s^2 - n\bar{y}^2 = 2828 - 2200 = 128. \end{aligned}$$

$$\eta = \frac{\sigma_m}{\sigma_y} = \frac{1.13}{3.03} = .37, \text{ 相關係數之值雖小，而相關比率卻甚大。}$$

例五：下表資料，係根據『紐約市兒童之身長體重報告』一文，共有兒童 3,406 人，年齡在 14-15 之間。

身長	人數	報告中之平均體重	-	-	由方程式算出來之體重	-
原點 01 時 x	ns	原點 100 磅 \bar{y}_s	$ns\bar{y}$	$x.ns\bar{y}_s$	原點 100 磅 -	$ns\bar{y}_s^2$
-12	1	-12	- 12	144	-49.6	144
- 9	1	-20	- 20	180	-36.6	400
- 7	13	-18	- 234	1,638	-27.9	4,212
- 6	59	-19	-1,121	6,726	-23.6	21,299
- 5	96	-17	-1,632	8,160	-19.2	27,744
- 4	190	-14	-2,660	10,640	-14.9	37,240
- 3	283	-12	-3,396	10,188	-10.5	40,752
- 2	349	- 8	-2,792	5,584	- 6.2	22,336
- 1	440	- 3	-1,320	1,320	- 1.9	3,960
0	434	+ 2	+ 868	0	+ 2.5	1,736
+ 1	400	+ 7	+2,800	2,800	+ 6.8	19,600
+ 2	355	+11	+3,905	7,810	+11.2	42,955
+ 3	307	+17	+5,219	15,657	+15.5	88,723
+ 4	200	+20	+4,000	16,000	+20.0	80,000
+ 5	137	+24	+3,288	16,440	+24.2	78,912
+ 6	78	+30	+2,340	14,040	+28.6	70,200
+ 7	34	+35	+1,190	8,330	+32.9	41,650
+ 8	15	+34	+ 510	4,080	+37.2	17,340
+ 9	6	+42	+ 252	2,268	+41.6	10,584
+10	7	+42	+ 294	2,940	+45.9	12,348
總計	3,405	-	11,479	134,945	-	622,135

$\bar{x} = 2270, \sigma_1 = 2.99$ 。 平均數 61.227 吋。

$\bar{y} = 3.371, \sigma_2 = 16.3$ (註四) 平均數 103.37 磅。

$$r = \frac{Sxy - n\bar{x}\bar{y}}{nc_{12}} = \left(\frac{134945}{3405} - 0.7652 \right) \div \sigma_1\sigma_2 = 0.797$$

迴歸方程式為 $\frac{\text{體重} - 103.37}{16.3} = 0.797 \times \frac{\text{身長} - 61.227}{2.99}$

體重 = $103.37 + 4.345$ (身長 - 61.23)。

由此方程式求得之體重，列於上表之第六列，須用以與相當第三列各項身長之平均體重對照。公式所得之值與實在相合甚近，約自 57 吋至 70 吋；但在 57 吋以下，實在體重並不如公式之急轉直下。迴歸線實際上對低等身材者，並非為一直線。

$$3405\sigma_m^2 = S n_s \bar{y}_s^2 - 3405 \bar{y}^2,$$

$$\sigma_m = 13.1$$

$$\eta = \sigma_m \div \sigma_2 = .81.$$

此處相關率與相關係數，實係一數。

例六：第六章第十二節所論之時間數列相關，茲以英國聯合王國之按年平均每人進口貨值，與英格蘭與威爾士之各年結婚率為例以明之。

以 x 為各年超過五年平均數之值，每五年以平均數為中心。由結婚率，計算 y 值亦然。

$$\bar{x} = -.62, \quad \bar{y} = .3, \quad \sigma_1 = 36.9, \quad \sigma_2 = 3.61, \quad Sxy = 4309,$$

$$n = 50, \quad r = \frac{4309 \div 50 - .3 \times .62}{36.9 \times 3.61} = .65, \text{ 其標準差為}$$

$$\frac{1 - .65^2}{\sqrt{50}} = .09.$$

此處相關之計算，即係用移動平均法。

為比較差額以求相關起見，茲將最初五年差額列表於下，其他各行刻從略。

		進口貨值			結婚率			
X		DX	D ² X	Y	DY	D·Y	DX·DY	D ² X·D ² Y
1845	330	-15	-	172	0	-	0	-
1846	315	+6	+21	172	-14	-14	-84	-294
1847	321	-20	-36	158	+1	+15	-30	-540
1848	291		+91	159		+2		+182

	DX	D ² X	DY	D ² Y
平均數	15.74	1	-0.19	.04
標準差	57.13	80	5.3	6.78

DX·DY 之和 = 8902。D²X·D²Y 之和 = 12076。

故從第一級差額得 r 為 .60，從第二級差額得 r 為 .45。

例七：依第三章第七節第一例所舉之實驗，以十字字母之數相加，共有 1,000 個總和。茲以前五字之字總和為 A，B 為次五字字母之和，於是 $x = A + B$ 。在每十字取出之後，更取出五字，將五字字母之和定為 C；於是 y 乃為 $B + C$ 。如此共得 1,000 對，此千對之相關係數必為 $\frac{1}{2}$ ，其標準差為 .024。

但實際上相關係數乃為 .553，超過所期值有兩標準差以上。此中原因，或可以歸於缺乏完全獨立性為解釋。以 250 總和為一類（其標準差為 .047），則四類之相關係數為 .56, .50, .58, .59。

迴歸線在中心部分自 $x = 40$ 至 $x = 61$ ，幾為一條直線；此數之外，每一 x 值出現不下於 20 次；而一整列平均數之標準差大於 2，故比較方法，不無相當價值。 y 值之所有標準差，為自 1.3

至 2.0。

x 值	相當之 y 值平均數	y 之 四 分 位 差	由方程式 算出之 \bar{y}_s
40	48.3	10½	45.3
41	44.8	10	45.8
42	49.1	11	46.4
43	47.4	11	46.9
44	47.1	9	47.5
45	46.4	9½	48.0
46	48.9	12	48.5
47	46.1	10	49.1
48	50.2	11	49.6
49	51.1	12	50.2
50	51.5	11½	50.7
51	49.6	7	51.3
52	53.6	13	51.8
53	53.6	16	52.3
54	51.1	10	52.9
55	51.9	6	53.4
56	53.4	14	54.0
57	52.7	12	54.5
58	52.7	11	55.1
59	60.2	8	55.6
60	56	11	56.1
61	57.2	11	56.7

四分位差，從理論得來 ($2\sigma\sqrt{1-r^2}$ 之 .67)，依公式 26 及 107，其值為 105；而觀察得來之四分位差，平均為 10.75。此似與 x 值無關，果如理論（公式 107）預期。

此等數之適應性，圖上表示無遺，至圖上之迴歸方程式乃為

$$\frac{y-51.50}{9.24} = .553 \frac{x-51.46}{9.43} \circ$$

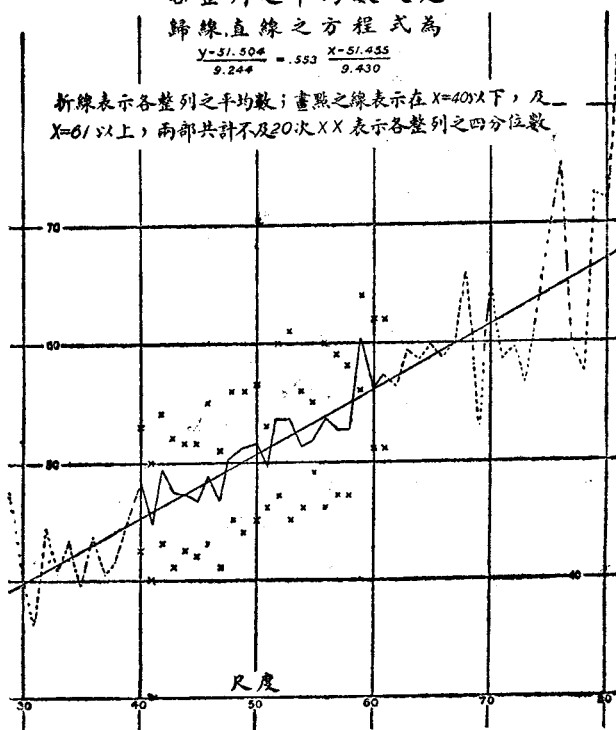
字母數實驗

各整列之平均數及迴

歸線，直線之方程式為

$$\frac{y-51.504}{9.244} = .553 \frac{x-51.435}{9.430}$$

折線表示各整列之平均數；畫點之線表示在 $X=40$ 以下，及 $X=61$ 以上，兩部共計不及 20 次 XX 表示各整列之四分位數



(第一圖)

如分級時，將羣類分為較寬之級，以便將由抽樣而來之差誤

消滅。則迴歸之數字表示，可以較為明鮮。如：

y 之分級	次 數	y 之平均數	相當之 x 平均數	方程式得來之 x
30-39	85	36.3	43.6	42.9
40-49	348	44.7	47.7	47.6
50-59	360	54.3	52.7	52.0
60-69	173	63.0	57.7	58.0
70-79	30	72.7	64.1	63.4

此處係指 x 對 y 之迴歸而言，而上圖則為 y 對 x 之迴歸。

在 30 以下及 80 以上尚有二數未計。

比較觀察分配，與常態相關面之分配，其方法甚多，茲擇其單簡者介紹如下：

設真 $r = \frac{1}{2}$ ， x 與 y 標準差之平均值為 $\sigma = 9.32$ 。則相關方程式為

$$z = \frac{1}{2\pi\sigma^2\sqrt{1-r^2}} e^{-\frac{1}{2(1-r^2)\sigma^2}(x^2+y^2-xy)}$$

$$\text{但 } x = \frac{X-Y}{\sqrt{2}}, y = \frac{X+Y}{\sqrt{2}}.$$

$$\text{方程式變為 } z = \frac{1}{\pi\sigma^2\sqrt{3}} e^{-\frac{X^2}{3\sigma^2} - \frac{Y^2}{\sigma^2}},$$

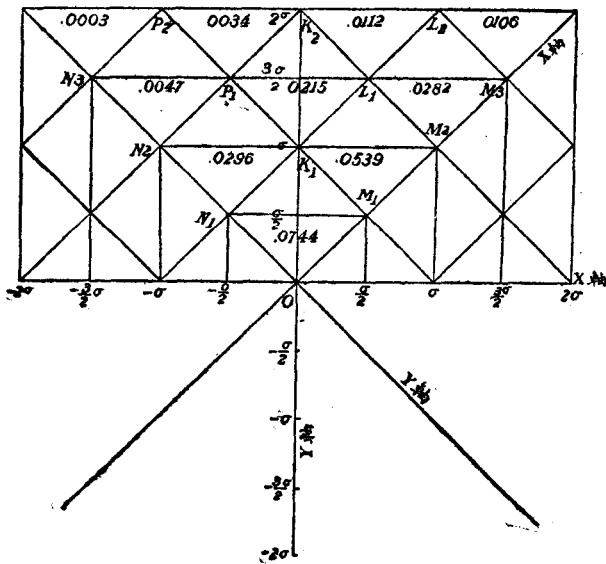
此代表對於主軸之平面，對中心軸斜角有 45° 。

以 $X = X_1$, $X = X_2$, $Y = Y_2$, $Y = Y_1$ 為限界之面積上，其

容積爲

$$\frac{1}{\sqrt{2\pi} \cdot \sigma\sqrt{\frac{3}{2}}} \int_{X_1}^{X_2} e^{-\frac{X^2}{2(\sigma\sqrt{\frac{3}{2}})^2}} dx \cdot \frac{1}{\sqrt{2\pi} \cdot \sigma\sqrt{\frac{1}{2}}} \int_{Y_1}^{Y_2} e^{-\frac{Y^2}{2(\frac{\sigma}{\sqrt{2}})^2}} dY,$$

如用第二章之常態機率表，不難立即求得。



圖上在 X 軸上，將 OM_1, M_1M_2, \dots 等距離劃開，使其各等於 $\frac{\sigma}{\sqrt{2}}$ 。以 σ_1 爲 X 之標準差， σ_2 爲 Y 之標準差。

$$\sigma_1 = \sigma\sqrt{\frac{3}{2}}, \quad \sigma_2 = \sigma/\sqrt{2}.$$

於是 $OM_1 = M_1M_2 = \dots = \sigma_1/\sqrt{3} = 0.577\sigma_1$ 。

對 X 軸垂直之平面，構成一立體形，此立體形之各部分容積，

經 OM_1 爲 $F(.577)$, 經 M_1M_2 爲 $F(1.155) - F(.577)$, 其他以此類推, 此在常態機率表中查出, 乃爲 $.2180, .1580$。

次復將 Y 軸劃爲 ON_1, N_1N_2等距離, 使各等於 $\frac{\sigma}{\sqrt{2}}$, 即使等於 σ_2 也。

以對 Y 軸垂直之平面爲限界, 所構成之容積, 在 ON_1 爲 $F(1)$, 在 ON_2 爲 $F(2) - F(1)$, 餘類推, 即等於 $.3413, .1359$。

因方程式中, X 與 Y 之積分, 各自獨立, 垂直於 X 軸之各部, 必在 N_1, N_2等處, 被平面割爲相等部分。故組成下表, 表明在第一線之平方 $ON_1K_1M_1, M_1K_1L_1M_2, M_2L_1L_2M_3, \dots$ 上, 第二線平方 $N_1N_2P_1K_1, K_1P_1K_2L_1$ (其他各線以此類推), 常態平面之容積各部分。

常態頻數平面之分配

X/σ_1		.577	1.155	1.732	2.31	2.89	3.46
$F(X/\sigma_1)$ (差額)		.2180	.1580	.0824	.0311	.0085	.0017
Y/σ_2	$F(Y/\sigma_2)$ 差額	差 額 乘 積					
1	.3413	.0744	.0539	.0282	.0106	.0028	.0006
2	.1359	.0296	.0215	.0112	.0042	.0012	.0002
3	.0214	.0047	.0034	.0017	.0007	.0002	.0000
4	.0014	.0003	.0002	.0001	.0001	.0000	.0000

OX 與 OY 二軸所成之四象限，各個分配狀況全同。表中之小數以 1000 乘之，即得 1000 對數之理論上的分配，假如置偏斜度於不論。

用方格紙將觀察值逐一記下，並將在 X, Y 各平方內出現次數註明。結果書如次表，表中各排之第一行，重記已知之理論數，第三行為觀察數。

理論數與觀察數，在中心左右三平方內，中心上下兩平方內，頗能吻合，即在 $\pm 1.7\sigma_1$ 與 $\pm \sigma_2$ 之內也。隨機抽樣中分歧若此之機率，約為 $\frac{2}{3}$ （見第十章例釋）。

自此方格再向左，情形乃全不同（觀察數為 31，而期望數為 41），向右則又多出許多（觀察數 54，期望數 41）。但十二個方格自中心向左，略有盈餘，向右則有不足。此恰與吾人對於原始曲線（見第三章第七節第一例）之偏斜度所預料之現象相仿。茲將偏斜度之影響，另於本章附註中討論之，而修正後之結果，即列入下表各排之第二行。改進之效果甚為顯著；例如最末三列向之期望數，現為 $33\frac{1}{2}$ （觀察值為 3·1），最末三行向右，則為 49（觀察值為 54）。

1000 對字母之和

觀察值分配與常態及與偏態頻數之比較。

中心之縱橫直線，並非坐標軸，實成 45° 斜角之對稱軸也。

第一行 常態分配..... (例如 29.6)

第二行 第二近似值..... (例如 28.7)

第三行 觀察值..... (例如 35)

0	0	.1	.1	.2	.3	.3	.2	.1	.1	0	0
0	0	?	?	0	1	0	1	?	?	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	.2	.7	1.7	3.4	4.7	4.7	3.4	1.7	.7	.2	0
0	0	0	1	1.9	4.3	5.1	4.9	3.2	1.7	?	0
0	0	0	1	1	1	5	6	4	0	0	0
.2	1.2	4.2	11.2	21.5	29.6	29.6	21.5	11.2	4.2	1.2	.2
?	.2	2.8	10.8	22.3	30.5	28.7	20.7	11.6	5.6	2.2	?
0	0	5	11	20	38	35	20	11	5	0	0
.6	2.8	10.6	28.2	53.9	74.4	74.4	53.9	28.2	10.6	2.8	.6
0	2.7	11.1	33.3	63.0	79.2	69.6	44.8	23.1	10.1	2.9	1.6
0	2	12	34	47	77	72	61	24	8	5	1
.6	2.8	10.6	28.2	53.9	74.4	74.4	53.9	28.2	10.6	2.8	.6
0	2.7	11.1	33.3	63.0	79.2	69.6	44.8	23.1	10.1	2.9	1.6
0	1	9	36	75	78	64	48	26	15	5	4
.2	1.2	4.2	11.2	21.5	29.6	29.6	21.5	11.2	4.2	1.2	.2
?	.2	2.8	10.8	22.3	30.5	28.7	20.7	11.6	5.6	2.2	?
0	1	1	8	21	32	28	19	9	8	0	2
0	.2	.7	1.7	3.4	4.7	4.7	3.4	1.7	.7	.2	0
0	?	?	.2	1.9	4.3	5.1	4.9	3.2	1.7	?	0
0	0	0	0	0	3	2	0	3	1	0	0
0	0	.1	.1	.2	.3	.3	.2	.1	.1	0	0
0	0	?	?	0	1	0	1	?	?	0	0
0	0	0	0	0	0	0	0	0	0	0	0

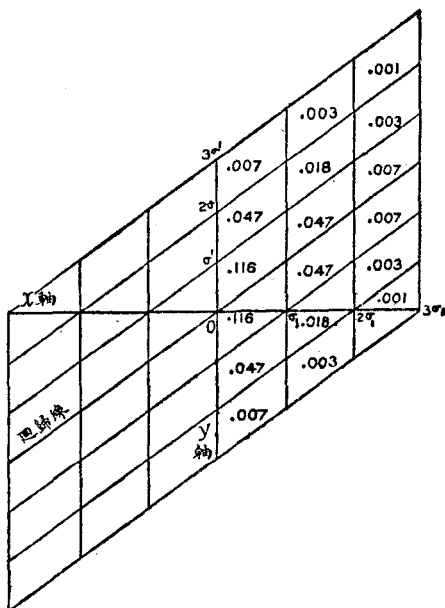
與整個期望值分歧之機率，據測驗結果（見第十年例釋），約等於 $\frac{1}{5}$ ；換言之，如此之實驗，五次只有二次，可望其密切相合（註五），反之，如分配果為常態及對稱，向左右所得如是之分歧，幾全為不可能之事；為補充此理論，必須求第二近似值也。

關於觀察值之分配，與常態平面下之分配，是否能以一致之問題，簡單之測驗方法，可由分析 x - 整列之分配而得，無須如上段之須變為對稱軸也。

$$\iint \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} e^{-\frac{1}{2(1-r^2)}\left(\frac{x^2}{\sigma_1^2} + \frac{y^2}{\sigma_2^2} - \frac{2rxy}{\sigma_1\sigma_2}\right)} dx dy$$

$$= \int \frac{1}{\sigma_1'\sqrt{2\pi}} e^{-\frac{y'^2}{2\sigma'^2}} dy' \times \int \frac{1}{\sigma_1\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma_1^2}} dx,$$

式中所用之 $\sigma' = \sigma_2 \sqrt{1-r^2}$, $y' = y - \frac{r\sigma_2}{\sigma_1}x$, 換言之, y' 係自迴歸線上與 y 軸平行, 如公式 107 所論。



$F(z)$

0-1 1-2 2-3
 .341 .136 .022

$F(z)$	乘 積			
0-1	.341	.116	.047	.007
1-2	.136	.047	.018	.003
2-3	.021	.007	.003	.001

以標準差作區分之單位，示如上表及圖。就此為基準，將字母實驗之結果，列表如下：

	-3σ	-2σ	$-\sigma$	0	σ	2	3σ	
	0	0	0	0	0	0	0	0
$3\sigma'$	0	0	0	0	0	0	0	0
	0	1	3	7	7	3	1	0
	0	0	0	4	8	2	0	0
$2\sigma'$								
	0	3	18	47	47	18	3	0
	0	1	17	63	56	18	4	1
σ								
	1	7	47	116	116	47	7	1
0	0	1	57	131	114	42	6	0
	1	7	47	116	116	47	7	1
	0	7	49	98	103	48	9	1
$-\sigma'$								
	0	3	18	47	47	18	3	0
	1	2	13	47	50	16	0	0
$-2\sigma'$								
	0	1	3	7	7	3	1	0
	0	2	3	10	7	3	0	0
$-3\sigma'$								
	0	0	0	0	0	0	0	0
	0	0	1	0	3	1	0	1

迴歸線

縱列示 x - 整列，可與本章例七之詳細表對照參看。每一間隔，上面一數為計算得來之數值，而下面之數乃係觀察之值。

對於偏斜度之修正，可以在同一方向改進適應程度，與以前之表同。

附註：——求相關面之第二近似值

當 $\frac{1}{\sqrt{n}}$ 次之各項，依一般大數律，加以保留時，一含有差誤之平均立方 (mean cube of error) 之項，必現諸方程式上 (見第三章第五節)。同理，據愛基華斯教授之證明 (註六) 相關面之方程式，在同樣條件下，應書如

$$z = z_0 - \frac{1}{6} \left(k_{30} \frac{\partial^3}{\partial x^3} z_0 + 3k_{21} \frac{\partial^3}{\partial x^2 \partial y} z_0 + 3k_{12} \frac{\partial^3}{\partial y \partial x^2} z_0 + k_{03} \frac{\partial^3}{\partial y^3} z_0 \right) \dots (113)$$

此處
$$z_0 = \frac{1}{2\pi\sqrt{1-r^2}} e^{-\frac{1}{2(1-r^2)}(x^2+y^2-2rxy)}$$

w 與 y 為觀察值與平均數（被其數標準差除後）之差額。

$$k_{10} = \text{平均值 } x^3, k_{21} = \text{平均值 } x^2y,$$

$$k_{12} = \text{平均值 } xy^2, k_{03} = \text{平均值 } y^3.$$

如本章七例, $x\sigma = A+B$, $y\sigma = B+C$ ($\sigma = 9.44$ (約略數))。

平均值 $xy\sigma = \text{平均值 } B^2 = \frac{1}{2}\sigma^2$, 又 $r = \frac{1}{2}$ (在實驗中得 $r = .55$)。

平均值 $x^3\sigma^3 = k_{30} = k_{03} = 'k'$ (見第一章第三節) = .409。

平均值 $x^2y\sigma^3 = k_{21} = \text{平均值 } B^3 = \frac{1}{2} \text{平均值 } (A+B)\sigma = \frac{1}{2}'k' = k_{12}$ 。

如將此等數值微分時, 則得

$$z = \frac{1}{\pi\sqrt{3}} e^{-\frac{2}{3}(x^2+y^2-xy)} \left\{ 1 - .01(x+y)(18+11xy-8x^2-8y^2) \right\} = z_0(1-w).$$

z_0w 一式不易加以積分, 較為簡易之步驟, 係對適當之面積上, 將 z_0 積分, 然後約略加以修正。西普生氏定則 (Simpson's rule) 所用之方法, 如

$$z_0, k, z_0, -k, z_0, 0, z-h, 0,$$

為長方形上平面四角之縱坐標, 而長方形之對角線為

$$2h, 2k,$$

z_{00} 為長方形中心之縱坐標; 則平均縱坐標為

$$\frac{1}{3}(2z_{00} + z_0k + z_0 - k + z_0h + z - h_0).$$

故如將 z_0w 算出以求前節常態類數面平方之分配表各容積之中心及四角, 則須用一量數 z_0w' (w' 為中心值兩倍及四角度之平均數) 使每一容積復原。如此通盤施行, 而將所得之值, 加上或從常態曲線得來之數減去, 以計算前段之修正值。

吾人須加注意者, 即當平面係指其主軸而言, 因而以 $x+y = \sqrt{2}X$, $-x+y = \sqrt{2}Y$, 則 w 在 Y 軸上變為對稱, 但非在 X 軸。

(註一) 薛伯氏校正, 見附錄五。

(註二) 第六章第八節, y 各值由其平均數計算, 此處須通盤減去 y 。

(註四) 計算根據之資料, 此處未曾查明。

-
- (註五) 表中重線, 僅爲表示應用第十章測驗方法之區分。
- (註六) 差誤律 (Camb. Phil. Trans., Vol. XX, 1905, Part II, § 6), 及統計學報 (Statistical Journal, 1917, 第 268 頁起。愛基華基氏, 在教本中不用模差(modulus), 而用標準差作單位。

第八章 淨相關與複相關

第一節 淨相關

依第七章所論之調查，可知一量數之變化，足以影響另一量數之變化。然情況尙不止此而已，一量數之變動，往往波及於其他多數之量，因亦隨之而變動。故頻數分配不能再以三元(dimension)之平面爲代表矣。關於此者，須求一類似之函數，前論者不過爲其簡單之情形耳。

討論至此，前言之迴歸方程，已非直線或曲線之方程式可比，此時之方程，乃一變量，與其他若干變量發生關係者也。斯時可將現有之許多變量，用類似『局部微分』(partial differentiation)方法，剔除其中之一，使其不至再生影響。如此一一提出，然後求其所餘兩個變量之相互關係，此即極爲重要之淨相關 (partial correlation)法也。

下文討論三個變量之例，詳細解剖，而將其較爲普遍之解法，提綱舉要，加以說明。

設有 x, y, z ，三個變量於此，各由其平均數算來，並設其彼此互有相關。假如三者關係如此： $z = ax + by + c$ ，此爲一理論之

平面，由此平面所得 z 之平均值，與 x, y 一對之值相應。

現求出 a, b, c ，求時以使 z 之觀察值 (observed value) 對由方程式求得值之觀察離中差，具有最低之反機率 (improbability) 為妙。

設 \bar{z}_s 為 k_s 個觀察之平均數，每種觀察， x 有 x_s (到 $x_s + \delta x$) 個成員， y 有 y_s (到 $y_s + \delta y$) 個成員。

以 η_s 代 $\bar{z}_s - (ax_s + by_s + c)$ ，此即第 s 個觀察羣類平均數對其理想值 (ideal value) 之離中差。

以 $\sigma_x, \sigma_y, \sigma_z$ 為 x, y, z 各頻數羣類之標準差。

於是如在長久經驗之後， z 羣類之標準差，與 x, y 之值無關，則 η_s 之標準差必為 $\frac{\sigma_z}{\sqrt{k_s}}$ (公式38)，而 η_s (到 $\eta_s + \delta\eta$) 出現之機率，為 $ke^{-\frac{k_s \eta_s^2}{2\sigma_z^2}} \cdot \delta\eta$

設有 n 對數值，如 x_s, y_s ，而共有觀察 N 個，則

$$N = k_1 + k_2 + \dots + k_n$$

$\eta_1 \dots \eta_s \dots \eta_n$ 出現之機率，為 $Ce^{-\frac{1}{2\sigma_z^2} \phi}$ ，式中 $\phi = k_1 \eta_1^2 + \dots + k_s \eta_s^2 + \dots + k_n \eta_n^2$ ，而 C 為常數。

$\phi = S_1^n k_s [z_s - (ax_s + by_s + c)]^2$ 為最小時，機率必最大，茲求出 a, b, c ，以達此目的。

$$\phi = S(k_s \bar{z}_s^2) + a^2 S(k_s x_s^2) + b^2 S(k_s y_s^2) + c^2 S k_s - 2a S(k_s x_s \bar{z}_s) - 2b S(k_s y_s \bar{z}_s) - 2c S(k_s \bar{z}_s) + 2ab S(k_s x_s y_s) + 2ac S k_s x_s + 2bc S y_s$$

但 $\sum x_s = 0 = \sum y_s$ 。 $\sum k_s \bar{z}_s = z$ 所有各值之和 = 0。

$\sum k_s x_s^2 = N\sigma_x^2$ ， x_2 在全羣類重現 k_s 次，又 $\sum k_s y_s^2 = N\sigma_y^2$ 同。

$\sum k_s x_s \bar{z}_s = \sum xz$ ， $\therefore k_s \bar{z}_s =$ 在第 S 羣類中 z 所有值之和。

又 $\sum k_s y_s \bar{z}_s = \sum yz$ 。 又 $\sum k_s x_s y_s = \sum xy$ 。

$$\therefore \phi = S(k_s \bar{z}_s^2) + Na^2\sigma_x^2 + Nb^2\sigma_y^2 - 2a\sum xz - 2b\sum yz + 2ab\sum xy + Nc^2。$$

以 $\sum xy = Nr_{xy}\sigma_x\sigma_y$ ， $\sum xz = Nr_{xz}\sigma_x\sigma_z$ ， $\sum yz = Nr_{yz}\sigma_y\sigma_z$ 。

然則 ϕ 必為最小，如 $\frac{\partial \phi}{\partial a}$ ， $\frac{\partial \phi}{\partial b}$ ， $\frac{\partial \phi}{\partial c}$ 各為零（註一），換言之，如

$$0 = \frac{\partial \phi}{\partial c} = 2Nc \quad \therefore c = 0,$$

$$0 = \frac{\partial \phi}{\partial a} = 2N(a\sigma_x^2 + b\sigma_x\sigma_y r_{xy} - \sigma_x\sigma_z r_{xz}) = 0,$$

$$0 = \frac{\partial \phi}{\partial b} = 2N(a\sigma_x\sigma_y r_{xy} + b\sigma_y^2 - \sigma_y\sigma_z r_{yz}) = 0。$$

故 $a\sigma_x + \sigma_y r_{xy} = \sigma_z r_{xz}$

$$a\sigma_x r_{xy} + b\sigma_y = \sigma_z r_{yz}。$$

$$\therefore \frac{a\sigma_x}{r_{xz} - r_{xy}r_{yz}} = \frac{b\sigma_y}{r_{yz} - r_{xz}r_{xy}} = \frac{\sigma_z}{1 - r_{xy}^2} \dots\dots (114)$$

方程式 $z = ax + by + c$ 變為

$$\frac{z}{\sigma_z} = R_x \cdot \frac{x}{\sigma_x} + R_y \cdot \frac{y}{\sigma_y} \dots\dots\dots (115)$$

式中， $R_x = \frac{r_{xz} - r_{xy}r_{yz}}{1 - r_{xy}^2}$ ， $R_y = \frac{r_{yz} - r_{xz}r_{xy}}{1 - r_{xy}^2}$ 。

R_x ， R_y 為 z ， x 及 z ， y 間之淨迴歸係數 (partial regression

coefficient); 對某一 y 值, $z = R_x \frac{\sigma_z}{\sigma_x} x + \text{常數}$, 對某一 x 值, $z = R_y \frac{\sigma_z}{\sigma_y} \cdot y + \text{常數}$, 與 $y = r \frac{\sigma_y}{\sigma_x} \cdot x$ 相對照之公式, 已詳第六章第五節。

類似此式之方程, 當 x 以 y, z 表示, 或當 y 以 z, x 表示時, 自亦可照樣解出。

x 與 z (y 為不變) 淨相關係數 (partial correlation coefficient) 之定義, 以僅有兩變量之例為比喻, 可解為當 z 用 x 與 y 表示 (如公式115), 及 x 用 z 與 y 表示時, 分別求得淨相關係數之幾何平均數。故淨相關係數為

$$\frac{r_{xz} - r_{xy}r_{yz}}{\sqrt{1-r_{xy}^2} \sqrt{1-r_{yz}^2}}$$

以上之研究, 均根據猶爾 (Yule) 先生之論文 (統計學報 1897 年號, 自八百三十一頁起) 及其所著之書, 蓋此科目大部均為猶爾 先生之貢獻也。惟此地所論者, 與猶爾 氏略有不同, 本書重要之考慮, 一則係由於差誤律之通用 (見第三章第五節), 一則係假定 z 之標準差與 x 及 y 之各值無關, 但此不合通例之事也。至猶爾 先生, 則不用此假定, 而用最小二乘法 (least square), 此法因其原理根基之難點, 本書, 除極少為例外外, 不取之。

z 對於 x 及 y 之方程式, 與猶爾 先生所用者相同, 而與由常態複相關原理所得者, 亦無少異。

例一: 生活費調查委員會, 於一九一八年, 用家計調查法, 搜

集多數工人家庭之每週食物費用資料（參閱第三章第七節第七例）。就中有三百九十家，係有技能工人。今將此 390 家資料，依各家人數，分為十四歲以上以下二類（參閱統計學報1919年號，第三百六十頁）。

茲將其標號及數量列下：

	食 物 費 用	十四歲以上人數	十四歲以下人數
平均數	51先令	2.48	3.56
對平均差	$z \times 5$ 先令	x	y
標準差	$\sigma z = 5.03 \times 5$ 先令	$\sigma x = .836$	$\sigma y = 1.40$
$r_{xy} = -.0525, r_{zx} = .504, r_{xy} = .315。$			

由此得方程式 $z/\sigma z = .52x/\sigma x + .35y/\sigma y$ ，更由此式得公式：一
食物費用（先令） = $14.5 \times 9.4 \times$ 十四歲以上人數 + $3.7 \times$ 十四
歲以下人數，並組成下表：

家庭食物費用

十四歲以 上之人數	照 公 式 算 得				實 在 數 之 平 均 數			
	兒 童 數 額				兒 童 數 額			
	2	3	4	5	2	3	4	5
2	40.7	44.4	48.1	51.8	40.5(74)	45.2(74)	47.1(53)	52.9(25)
3	50.1	53.8	57.5	61.2	54.8(21)	51.2(17)	58.2(16)	64.9(17)
4	59.5	63.2	66.9	70.6	58.0(10)	60.2(10)	78.1(6)	— (0)

括弧中數目係平均實在數

就標準差大而事件次數小者而論，經驗所得與公式所得，二者適應情形甚佳。

由此實例觀之，淨相關法與平常所用之交叉表法，至為接近；但一經用此公式，即將此兩得數，使生始終一貫之關係。現得一結論：（一）平均計算，每添一成年人（家庭收入普通均恃成年人）家庭食物費用，應增九先令五便士，而添一兒童應加三先令八便士；（二）兒童人數愈多生活程度愈低，因一兒童之養育費約當成年人三分之二也（此處所謂成年人，係指在十四歲以上者而言）。

例二：下列資料係自倫敦郡之一九一一年人口普查中摘錄而來：

$z+3.7$ 為一住宅之房間數

$x+4.15$ 為一家庭之人數

$y+.86$ 為一家庭十歲以下之兒童人數

3.7, 4.15, .86, 為倫敦郡之各該平均數。

$$r_{xy} = .57, r_{xz} = .44, r_{yz} = -.03, \sigma_z = 2.59, \sigma_x = 2.32, \sigma_y = 1.24,$$

$$R_x = .676, R_y = .402。$$

關於 1,023,951 家之數字，求至三位有效數字，堪稱確實。

$$z = x \times \frac{\sigma_z}{\sigma_x} \times .676 - y \times \frac{\sigma_z}{\sigma_y} \times .402 = x \times .754 - y \times .840$$

$$\text{或 (房間數} - 3.7) = .754(\text{人數} - 4.15) - .84(\text{兒童數} - .86)。$$

兒童人數增加時，某一定等級之家庭，房間數必急劇變少。

以下式記之亦可：

$$\begin{aligned} \text{房間數} &= 1.29 + .68 \text{ 人數} - .84 \text{ 兒童} \\ &= 1.29 + .68 \text{ 成人 (十歲以上)} - .16 \text{ 兒童 (十歲以下)} \end{aligned}$$

例三：據『生計與貧乏』(Livelihood and Poverty) 一書之記載，五百六十六家之社會研究，曾將收入，房租及家庭組織逐一調查。收入增加與有酬報人數多者，房租必較多，但若收入數與有酬報之人數不變，則兒童人數愈多，則可供為房租之用者愈少。

房租： $z + 6.075$ 先令，式中 6.075 先令為平均數。

相等成年人數： $x + 3.287$ ，式中 3.287 為平均數。

收入： $y + 31.712$ 先令，式中 31.712 先令為平均數。

相等成年人數，係依成人與兒童需用之房屋為根據，將成人，兒童照武斷之標準，加以分類；五歲下兒童，作為成人之四分之一，五歲至十四歲，作為二分之一，男童十四至十八歲，女童十四至十六歲，均為四分之三，自此以上則作為一。

房租對房間數之相關甚近，故房租可用以推算房間數。

$$\sigma_z = 1.33, \sigma_x = 1.22, \sigma_y = 13.0, r_{xy} = .543,$$

$$r_{yz} = .456, R_x = -.136, R_y = .532.$$

$$\text{故} \quad \frac{z}{\sigma_z} = -\cdot 136 \frac{x}{\sigma_x} + \cdot 532 \frac{y}{\sigma_y}$$

$$\text{或} \quad z = -\cdot 148x + \cdot 0544y_0$$

於是，如收入不變更，家庭變大則房間數減少。

以上三例，證明有兒童之家庭，比無兒童者，漸有減少每人食物費用及房間數之傾向，且於相當程度內測量其損失。

吾人仍須研究三變量之三原理論上分配，以與具有兩變量之常態相關面相呼應。下文特就簡單情形將分析及結論示明之，吾人將見所有證明方法，與具有兩變量之例，並無不同也。

節二第 複相關

下述分析，須假設各元素具有常態頻數，方能成立。

以 X, Y, Z 為因變量， U, V_1, V_2, V_3 ，為自變量，二種變量之關係為 $X = U + V_1, Y = U + V_2, Z = U + V_3$ 。

設 U, V_1, V_2, V_3 ，係從常態羣類中隨機抽出，該羣類之平均數為 $\bar{u}, \bar{v}_1, \dots$ ，標準差為 $\sigma_u, \sigma_{v_1}, \dots$ 。

以 $\bar{X}, \bar{Y}, \bar{Z}$ ，為 X, Y, Z 之平均數； $\sigma_x, \sigma_y, \sigma_z$ 為其標準差，並以 $X = \bar{X} + x, \dots, U = \bar{U} + u, \dots$

則至終局， $\bar{X} = \bar{U} + \bar{V}$ 等等諸如此類，且

$$\therefore \quad x = u + v, \quad y = u + v_2, \quad z = u + v_3_0$$

設 u, v_1, v_2, v_3 , 各個彼此完全獨立, 則

平均值 $uv_1 = 0 =$ 平均值 $v_1 v_2$ 其他以此類推。

茲以 r_{xy} 爲 x 對 y 之相關係數。

則 $\sigma_x^2 = \sigma_u^2 + \sigma_v^2 \dots \dots$,

$$\begin{aligned} \text{又 } \sigma_x \sigma_y r_{xy} &= \text{平均值}(u+v_1)(u+v_2) = \sigma_u^2 \\ &= \sigma_y \sigma_z r_{yz} = \sigma_z \sigma_x r_{zx} \end{aligned}$$

x_1, y_1, z_1 , 抽出之值對特定值 u, v_1, v_2, v_3 , 而生之連帶機率爲

$$\frac{1}{\sigma_u \cdot \sqrt{2\pi}} e^{-\frac{1}{2} \frac{u^2}{\sigma_u^2}} \times \frac{1}{\sigma_1 \sqrt{2\pi}} e^{-\frac{1}{2} \frac{v_1^2}{\sigma_1^2}} \times \dots = P_u,$$

受下列條件之支配: $x_1 = u + v_1, y_1 = u + v_2, z_1 = u + v_3$, 消去 v_1, v_2, v_3 。

x_1, y_1, z_1 , (因某一 u 值而生) 之連機率, 可以下

式得之: $-2 \log(P_u \cdot 4\pi^2 \sigma_u \sigma_{v_1} \sigma_{v_2} \sigma_{v_3})$

$$= \frac{u^2}{\sigma_u^2} + \frac{(u+x_1)^2}{\sigma_{v_1}^2} + \frac{(u+y_1)^2}{\sigma_{v_2}^2} + \frac{(u+z_1)^2}{\sigma_{v_3}^2} = a(u-b)^2 + c,$$

$$\text{式中 } a = \frac{1}{\sigma_u^2} + \frac{1}{\sigma_{v_1}^2} + \frac{1}{\sigma_{v_2}^2} + \frac{1}{\sigma_{v_3}^2}$$

$$ab = \frac{x_1}{\sigma_{v_1}^2} + \frac{y_1}{\sigma_{v_2}^2} + \frac{z_1}{\sigma_{v_3}^2}$$

$$c = \frac{x_1^2}{\sigma_{v_1}^2} + \frac{y_1^2}{\sigma_{v_2}^2} + \frac{z_1^2}{\sigma_{v_3}^2} - ab^2.$$

對某一 u 值而抽來之 x_1, y_1, z_1 , 其全部機率, 爲

$$P = \int_{-\infty}^{\infty} P u du,$$

如再以 x_1, y_1, z_1 爲常數 則

$$= \frac{e^{-\frac{c}{2}}}{(2\pi)^{\frac{3}{2}} \sigma_u \sigma_{v1} \sigma_{v2} \sigma_{v3}} \int_{-\infty}^{\infty} \frac{e^{-\frac{a}{2}(u-b)^2}}{\sqrt{2\pi}} du = \frac{1}{(2\pi)^{\frac{3}{2}} \sigma_u \sigma_{v1} \sigma_{v2} \sigma_{v3}} \cdot \frac{1}{\sqrt{a}} e^{-\frac{c}{2}}.$$

以 x, y, z 代 x_1, y_1, z_1 ,

$$ac = \left(\frac{1}{\sigma_u^2} + \frac{1}{\sigma_{v1}^2} + \dots \right) \left(\frac{x^2}{\sigma^2} + \dots \right) - \left(\frac{x}{\sigma} + \dots \right)^2.$$

但 $\frac{1}{\sigma_u^2} + \frac{1}{\sigma_{v2}^2} + \frac{1}{\sigma_{v3}^2} = \frac{1}{\sigma_u^2} + \frac{1}{\sigma_y^2 - \sigma_u^2} + \frac{1}{\sigma_z^2 - \sigma_u^2} = \frac{\sigma_y^2 \sigma_z^2 - \sigma_u^4}{\sigma_u^2 \sigma_{v2}^2 \sigma_{v3}^2}$

且 $a = \frac{1}{\sigma_u^2} + \frac{1}{\sigma_x^2 - \sigma_u^2} + \frac{1}{\sigma_y^2 - \sigma_u^2} + \frac{1}{\sigma_z^2 - \sigma_u^2}$

$$= \frac{\sigma_x^2 \sigma_y^2 \sigma_z^2 - \sigma_u^4 (\sigma_x^2 + \sigma_y^2 + \sigma_z^2) + 2\sigma_u^6}{\sigma_u^2 \sigma_{v1}^2 \sigma_{v2}^2 \sigma_{v3}^2}$$

$$\therefore e \{ \sigma_x^2 \sigma_y^2 \sigma_z^2 - \sigma_u^4 (\sigma_x^2 + \sigma_y^2 + \sigma_z^2) + 2\sigma_u^6 \}$$

$$= x^2 (\sigma_y^2 \sigma_z^2 - \sigma_u^4) + \dots - 2xy \sigma_u^2 \sigma_{v3}^2 - \dots$$

以 $R \sigma_x^2 \sigma_y^2 \sigma_z^2$ 代 $\sigma_x^2 \sigma_y^2 \sigma_z^2 - \sigma_u^4 (\sigma_x^2 + \dots) + 2\sigma_u^6$,

故 $R = 1 + 2r_{xy} r_{yz} r_{zx} - r_{xy}^2 - r_{yz}^2 - r_{zx}^2$.

於是 x, y, z 同時而起之機率爲

$$P = \frac{1}{(2\pi)^{\frac{3}{2}} \sigma_x \sigma_y \sigma_z R^{\frac{1}{2}}} \cdot e^{-\frac{1}{2R} \left\{ \frac{x^2}{\sigma^2} (1 - r_{yz}^2) + \dots - \frac{2xy}{\sigma_x \sigma_y} (r_{xy} - r_{xz} r_{yz}) - \dots \right\}}$$

$$\begin{aligned} \text{因 } \frac{r_{xy} - r_{xz}r_{yz}}{\sigma_x\sigma_y} &= \frac{\sigma_x\sigma_y r_{xy}\sigma_z^2 - \sigma_x\sigma_z \cdot r_{xy}\sigma_y\sigma_z \cdot r_{yz}}{\sigma_x^2\sigma_y^2\sigma_z^2} \\ &= \frac{\sigma_u^2\sigma_z^2 - \sigma_u^2 \cdot \sigma_u^2}{\sigma_x^2\sigma_y^2\sigma_z^2} = \frac{\sigma_u^2\sigma_v^2}{\sigma_x^2\sigma_y^2\sigma_z^2} \end{aligned}$$

在特別情形之下

$$\sigma_u = \sigma_{v1} = \sigma_{v2} = \sigma_{v3}, \quad \sigma_x^2 = 2\sigma_u^2 = \sigma^2,$$

$$r_{xy} = \frac{1}{2} = r_{yz} = r_{zx}, \quad R = \frac{1}{2}.$$

則機率為

$$\frac{1}{2\sigma^3\pi^{\frac{3}{2}}} e^{-\frac{1}{4\sigma^2}\{3(x^2+y^2+z^2) - 2(xy+yz+zy)\}}.$$

對 x, y 之某值, z 之最或然值, 用 $\frac{\partial P}{\partial z} = 0$ 計算, 乃為 (如公式 115)

$$\frac{z}{\sigma_z}(1 - r_{xy}^2) = \frac{x}{\sigma_x}(r_{xz} - r_{xy}r_{yz}) + \frac{y}{\sigma_y}(r_{yz} - r_{xy}r_{xz}).$$

在特別情形下, 此又變為 $z = \frac{1}{2}(x+y)$ 。

艾得頓 (Elderton) 氏——追隨皮爾生氏——之證明, 如 x, y 與 z , 為任何有限變量數 (如上文之 $u, v \dots$) 之和, 全係常態類數, 並且 $(x, y), (y, z)$ 或 (z, x) 各對, 均有共同性, 餘則僅有一數時, 則 P 之形式, 乃

$$\text{爲 } K e^{-(ax^2+by^2+cz^2+2fyz+2gzx+2hxy)}$$

式中 a, b, c, f, g, h 均為待決之常數。

以機率之總數為一。

$$\text{設 } A, B, C, F, G, H \text{ 爲行列式 } \Delta = \begin{vmatrix} a & h & g \\ h & b & f \\ g & f & c \end{vmatrix}$$

之子行列式, 則 $A = bc - f^2, F = hg - af, \dots$,

$$BC - F^2 = a\Delta \dots\dots$$

$$\text{於是 } -\log \frac{P}{K} = a \left(x + \frac{h}{a}y + \frac{g}{a}z \right)^2 + \frac{c}{a} \left(y - \frac{F}{C}z \right)^2 + \frac{\Delta}{C}z^2$$

$$1 = \iiint P \cdot dx dy dz = (\sqrt{\pi})^3 \cdot K \frac{1}{\sqrt{a}} \cdot \sqrt{\frac{a}{C}} \sqrt{\frac{C}{\Delta}}, \text{ 而 } K\pi^{\frac{3}{2}} = \Delta^{\frac{1}{2}}$$

$$\begin{aligned} \sigma z^2 &= \iiint P z^2 dx dy dz = K\pi \cdot \frac{1}{\sqrt{a}} \cdot \sqrt{\frac{a}{C}} \int z^2 e^{-\frac{\Delta}{C}z^2} dz = K\pi^{\frac{3}{2}} \frac{1}{\sqrt{C}} \cdot \frac{1}{2} \left(\frac{C}{\Delta} \right)^{\frac{3}{2}} \\ &= \frac{C}{2\Delta} \end{aligned}$$

$$\text{同樣 } \sigma y^2 = \frac{B}{2\Delta}, \sigma x^2 = \frac{A}{2\Delta}$$

$$\begin{aligned} \sigma y \sigma z r_{yz} &= \iiint P z y dx dy dz = K\sqrt{\pi a} - \frac{1}{2} \int \int z y e^{-\frac{c}{a} \left(y - \frac{F}{C}z \right)^2 - \frac{\Delta}{C} z^2} dy dz \\ &= K\sqrt{\pi a} - \frac{1}{2} \int \int \left(y' + \frac{F}{C}z \right) e^{-\frac{C}{a} y'^2 - \frac{\Delta}{C} z^2} dy' dz, \end{aligned}$$

式中, $y' = y - \frac{F}{C}z$, 又積分之極限為 $\pm\infty$

$$= K\pi \cdot a^{-\frac{1}{2}} \sqrt{\frac{a}{C}} \cdot \frac{F}{C} \int z^2 e^{-\frac{\Delta}{C} z^2} dz = \frac{1}{2} \frac{F}{\Delta}.$$

$$\text{同理 } \sigma x \sigma y r_{xy} = \frac{1}{2} \frac{H}{\Delta}, \text{ 又 } \sigma x \sigma z r_{xz} = \frac{1}{2} \frac{G}{\Delta}.$$

$$\therefore a\Delta = BC - F^2 = 4\sigma y^2 \sigma z^2 (1 - r_{yz}^2) \Delta^2$$

$$f\Delta = GH - AF^2 = 4\sigma x^2 \sigma y \sigma z (r_{xy} r_{xz} - r_{yz}) \Delta^2$$

$$\Delta^2 = ABC + 2FGH - AF^2 - BG^2 - CH^2$$

$$= 8\Delta^3 \sigma x^2 \sigma y^2 \sigma z^2 (1 + 2r_{xy} r_{yz} r_{xz} - r_{yz}^2 - r_{xz}^2 - r_{xy}^2).$$

以 R 代括弧內之數量。R = $\begin{vmatrix} 1 & r_{xy} & r_{xz} \\ r_{xy} & 1 & r_{yz} \\ r_{xz} & r_{yz} & 1 \end{vmatrix}$

$$\text{於是 } \Delta = \frac{1}{8R\sigma_x^2\sigma_y^2\sigma_z^2}, a = \frac{1-ryz^2}{2R\sigma_x^2}, f = \frac{rxyrxz-ryz}{2\sigma_y\sigma_z R}.$$

$$\text{故 } P = \frac{1}{\sqrt{R}(2\pi)^{\frac{3}{2}}\sigma_x\sigma_y\sigma_z} e^{-\frac{1}{2R}\left\{\frac{x^2}{\sigma_x^2}(1-ryz^2)+\frac{2xyz}{\sigma_y\sigma_z}(ryz-rxyrxz)-\right\}}$$

與上述之特別情形相同。

如不用 x, y, z , 而用 n 個數量 x^1, x^2, \dots , 同一方法 (應歸功於皮爾生教授) 之更概括證明, 可得

$$P = \frac{1}{\sqrt{R}(2\pi)^{\frac{n}{2}}\sigma_1\sigma_2\cdots\sigma_n} e^{-\frac{1}{2R}\left\{\frac{x_1^2}{\sigma_1^2}R_{11}+\cdots+\frac{x_n^2}{\sigma_n^2}R_{nn}+\cdots\right\}}$$

$$\text{式中 } R = \begin{vmatrix} 1 & r_{12} & r_{13}\cdots r_{1n} \\ r_{21} & 1 & r_{23}\cdots r_{2n} \\ \dots\dots\dots\dots\dots\dots\dots \\ r_{n1} & r_{n2} & r_{n3}\cdots r_{nn} \end{vmatrix}$$

r_{st} 與 r_{ts} 數量相同, 而 R_{st} 為自第 s 列第 t 排斜交之子行列式, 原來正負符號仍舊。

證明之步驟如下:

如前述, $P = Ke^{-\phi}$, 式中 $\phi = a_{11}x_1^2 + \cdots + 2a_{12}x_1x_2 + \cdots, a_{11}\cdots, a_{12}\cdots$ 為常數。

$$\text{設 } \Delta n = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \dots\dots\dots\dots\dots\dots\dots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}$$

式中, $a_{st} = a_{ts}$, 並以 A_{st} 為第 s 列第 t 排交叉之子行列式

$$A_{nn} = \Delta_n - 1.$$

然後可將 ϕ 引申, 使 x_1 僅出在第一項內, x_2 僅理於第二項內, 其他以此類推, 則

$$\phi = a_{11}\left(x_1 + \frac{a_{12}}{a_{11}}x_2 + \cdots\right)^2 + \frac{a_{22}(a_{11}-a_{11}^2)}{a_{11}}\left(x_2 + \cdots\right)^2 + \cdots + \frac{\Delta_n - 1}{\Delta_n - 2}\left(x_n - 1 - \frac{\Delta_n \cdot n - 1}{\Delta_n - 1}\right)^2$$

$$+ \frac{\Delta_n}{\Delta_n - 1} x n^2$$

此可用歸法納證明，但手續極為繁雜。

$$1 = \iint \dots P \cdot dx_1 dx_2 \dots dx_n = K \pi \frac{n}{2} \cdot \left(\frac{1}{a^{11}}\right)^2 \left(\frac{\Delta_1}{\Delta_2}\right)^{\frac{1}{2}} \dots \left(\frac{\Delta_{n-1}}{\Delta_n}\right)^{\frac{1}{2}} = \frac{K \pi^{\frac{n}{2}}}{(\Delta_n)^{\frac{1}{2}}}$$

$$\sigma_n^2 = \iint \dots P x n^2 dx_1 dx_2 \dots dx_n = 1 \cdot \frac{1}{2} \frac{\Delta_n - 1}{\Delta_n} = \frac{A n n}{2 \Delta_n}$$

同樣，將次數改變， $\sigma_t^2 = \frac{A t t}{2 \Delta_n}$

$$\begin{aligned} \sigma_n \cdot \sigma_{n-1} \cdot \sigma_{n-1} &= \iint \dots P x n x_{n-1} dx_1 \dots dx_n \\ &= \left(\frac{\Delta_n}{\Delta_n - 2}\right)^{\frac{1}{2}} \iint x n x_{n-1} e^{-\frac{\Delta_n - 1}{\Delta_n - 2} \left(x n + 1 - \frac{A_{n, n-1}}{\Delta_n - 1}\right)^2 - \frac{\Delta_n}{\Delta_n - 1} x n^2} dx n - \\ &\quad 1 dx_{n-1} = \frac{A_{n, n-1}}{2 \Delta_n} \end{aligned}$$

同樣，將次數改變，

$$\sigma_s \sigma_t \sigma_{st} = \frac{A s t}{2 \Delta_n}$$

在求 R 之行列式中，以此等數值代 r_{st} ，……，則得

$$R = \frac{1}{(2 \Delta_n)^n \sigma_1^2 \dots \sigma_n^2} \begin{vmatrix} A_{11} & A_{12} & \dots & A_{1n} \\ A_{21} & A_{22} & \dots & A_{2n} \\ \dots & \dots & \dots & \dots \\ A_{n1} & A_{n2} & \dots & A_{nn} \end{vmatrix} = \frac{\Delta_n^n - 1}{(2 \Delta_n)^n \sigma_1^2 \dots \sigma_n^2}$$

此係根據一著名之行列式定理。

$$\frac{1}{K} = (2\pi)^{\frac{n}{2}} \sigma_1 \dots \sigma_n \sqrt{R}.$$

$$R_{ii} = \frac{1}{(2 \Delta_n)^{n-1} \sigma_2^2 \dots \sigma_n^2} \begin{vmatrix} A_{22} & \dots & A_{2n} \\ \dots & \dots & \dots \\ A_{n2} & \dots & A_{nn} \end{vmatrix} = \frac{a_{11} \Delta_n^n - 2}{(2 \Delta_n)^{n-1} \sigma_2^2 \dots \sigma_n^2}$$

此又係一著名之定理。

$$\therefore a_{11} = \frac{R_{11}}{2R\sigma_1^2}$$

同樣, $a_{12} = \frac{R_{12}}{2R\sigma_1\sigma_2}$, 又將次數改變, $a_{st} = \frac{R_{st}}{2R\sigma_s\sigma_t}$,

對於 x_2, x_3, \dots, x_n 之已知值, x 之最或然值, 乃可於下列之式求出:

$$\frac{x_1}{\sigma_1} R_{11} = -\frac{x_2}{\sigma_2} R_{12} - \frac{x_3}{\sigma_3} R_{13} \dots - \frac{x_n}{\sigma_n} R_{1n}$$

(註一) a, b, c 各值, 如以 ϕ 當作各平方之和不用微分, 亦可求得。

第九章 平均數動差及相關等測量之精度(註一)

第一節 逆機率

以前數章，討論抽樣方法所生差誤問題，均以宇宙——樣本所自取之大範圍——為觀點，並非樣本之本身；換言之，以前之問題，乃係樣本究能如何代表其範圍，代表至何程度之問題。但實際問題，乃適得其反：吾人由將何以樣本以推斷範圍？與言及此，乃不能不提出繁難費解之逆機率 (inverse probability) 學說矣。逆機率為何？其形式如是：原已抽出樣本之各範圍，何者究係產生該樣本者耶？

為求議論清楚起見，似以少論逆機率原理為宜；故舉數例於下，以解釋問題並求解答焉。

袋中有金鎊一枚，先令二枚，今遺失其一。迨將所餘二幣，取出一枚，視之乃為先令。請問金鎊遺失之機率，為何？

金鎊遺失之原來機率，為 $p'_1 = \frac{1}{3}$ ，先令遺失之機率，原為 $p'_2 = \frac{2}{3}$ ，假設一種錢幣遺失之機會，與另一錢幣之遺失機會，完全均等。

如金鎊果已遺失，則取出先令之機率，為 $p_1 = 1$ ，因袋中除先令外已無他物也。

如所遺失者為金鎊，所取出者為先令，其原來機率，為 $p'_1 p_1 = \frac{1}{3}$ 。

如所遺失者為先令，所取者亦為先令，則其原來機率，為 $p'_2 p_2 = \frac{2}{3} \times \frac{1}{2} = \frac{1}{3}$ 。

假定所給，只言此等機會均等之繁複事件之一，業已出現，但所遺失者為何，則不得而知。

故袋中所餘之第三個錢幣，其為金鎊，抑為先令，機會正復相等。

茲將此假設，照下列方法，概而論之：設有各種可能事件，其實現之機率，分別為 $p'_1, p'_2, \dots, p'_t, \dots$ ，於此等事件中，只知其中之一，業已出現。吾人可另外得一結果，如事件之第一，第二……第 t ……已經實現，則其各該機率，當分別為 p_1, p_2, p_t, \dots 。

第 t 個事件實現並發生此結果之原來機率，為 $p'_t \times p_t$ 。

第一序列之事件，實現並產生此結果之原來機率，當為下列比率：

$$p'_1 p_1 : p'_2 p_2 : \dots : p'_t p_t \dots = P_1 : P_2 : \dots$$

但吾人現知此等之任一個，確已出現，且雖多得此項消息，亦並不影響相對量數 P_1, P_2, \dots ，不過使其總數提升成為如此比

率, K , 因而使其等於一, 一即代表代數機率尺度之一定 (certainty)。故 $K \cdot S P_t = 1$, 而實現者果為第一序列第 t 個事件之機率, 為

$$K P_t = \frac{p_t' \cdot P_t}{p_1' p_2' \dots + \dots + p_t' p_t + \dots}$$

一囊中, 有類似黑白二種球共六個。茲取出一個白球。吾人將何以推斷囊中之白色原有若干耶?

欲答覆此問題, 須視關於黑白球分配之原有機率之假定為何。如原來, 一球之為黑為白, 須有均等之機會, 而 p_t' 為 t 個球白色之機率, 則 $p_t' = \frac{1}{2^t} \cdot {}_6C_t$ 。

不論假定云何, $p_t = \frac{t}{6}$ 。

$$P_0 : P_1 : \dots : P_6 = \frac{1}{6 \times 2^6} (0 : 6 : 30 : 60 : 60 : 30 : 6)。$$

$$S P_t = \frac{1}{2} \text{ 又 } K = 2。$$

囊中白球有 0, 1, 2, 3, 4, 5, 6 個之機率, 當分別為 0, $\frac{1}{32}$, $\frac{6}{32}$, $\frac{15}{32}$, $\frac{20}{32}$, $\frac{15}{32}$, $\frac{6}{32}$, $\frac{1}{32}$ 。

但如囊中白球數額, 係用擲骰而決定, 骰子上面之點即用以定白球之數, 則

$$p_1' = p_2' = \dots = p_6' = \frac{1}{6}; \quad P_t = \frac{1}{6}, \frac{t}{6};$$

$$S P_t = \frac{21}{36}, \quad K = \frac{12}{7}, \text{ 又 } K P_t = \frac{t}{21}。$$

就一般而論，如囊中共有球 n 個。就中白球若干，數額轉螺旋盤決定之，在圓盤之圓周，以相等之空間，記 $0, 1, \dots, n$ 之數字，螺旋盤特立軸旋轉，俟其停止，察其圓周對一固定針最近之數，即以此數作為白球之數，則

$$P'_t = \frac{1}{n+1}, P_t = \frac{t}{n}, SP_t = \frac{1}{2}, KP_t = \frac{2t}{n(n+1)}.$$

原始白球數額，在自 t 個以下之總和機率，為

$$K(P_0 + P_1 + P_2 + \dots + P_t) = \frac{t(t+1)}{n(n+1)} = f(t).$$

如 $f(t) = \frac{1}{2}$ ，則囊中有白色 t 個，或無白球 t 個，各有一半之機會；又如 n 甚大，則 $t = \frac{n}{\sqrt{2}}$ 一式約略可以適合上方程式。

故當 n 甚大時，則白球對全體球佔 $\frac{1}{\sqrt{2}} = .707$ 之比例，機會各佔一半。比例在 $\frac{1}{2}$ 與 $\frac{\sqrt{3}}{2}$ 之間，機率約為 $\frac{1}{2}$ 。

此一實例，至為重要，一則示吾人所有結果，乃全繫對於未知事件（即吾人所抽取者）相對原來機率之假定如何，二則指示吾人，如將各項機率總集起來，較之各個計算，可得較易理解之結果。

第二節 某類在範圍中所佔比例， P ，之精度

茲請應用逆機率原理，以樣本決定其全範圍。設有 N 個事物

之範圍，具有某種特質者為 p ，範圍中具有該種特質者，有 pN 個，茲由範圍中抽取 n 個事物，經查具有該項特質者為 $p'n$ 個。

由已知之 p 中，發現 $p'n$ 個之機率，為 $nC_{p'n} p^{p'n} q^{q'n}$ (見第二章第四節) 式中 $q=1-p$, $q'=1-p'$ 。

如 p 自 0 至 1 所有各值，原來即係機會均等，則自 p 各值中，從 p' 至 x ，發現 $p'n$ 之機率，為由各特定值而生各機率之總和，即 $= nC_{p'n} \int_{p'}^x x^{p'n} (1-x)^{q'n} dx$ ，於是依前節第一例之定理， p 原始各值，自 p' 至 x ，之機率，為

$$P_x = \frac{nC_{p'n} \int_{p'}^x x^{p'n} (1-x)^{q'n} dx}{nC_{p'n} \int_0^1 x^{p'n} (1-x)^{q'n} dx},$$

而此式如將 $\frac{1}{\sqrt{n}}$ 略去，可照下述，化為差誤常態曲線之形式。

以 $x = p' + z\sigma$ ，其 $\sigma^2_n = p'q'$ ，而 $1-x = q' - z\sigma$ 。至 σ 則與 $\frac{1}{\sqrt{n}}$ 同次 (order)。

如是

$$P_x = \frac{\int_0^z \left(1 + \frac{z\sigma}{p'}\right)^{p'n} \left(1 - \frac{z\sigma}{q'}\right)^{q'n} dz}{\int_{-\infty}^{\infty} \left(1 + \frac{z\sigma}{p'}\right)^{p'n} \left(1 - \frac{z\sigma}{q'}\right)^{q'n} dz} = \frac{\int_0^z f(z) dz}{\int_{-\infty}^{\infty} f(z) dz},$$

因為，如 $x=1$ ，則 $z = \sqrt{\frac{nq'}{p'}}$ ，又如 $x=0$ ，則 $z = -\sqrt{\frac{np'}{q'}}$ ，至此則於 n 甚大時而漸至於 $\pm \infty$ 。

於是

$$\log f(z) = p'n \log \left(1 + \frac{z\sigma}{p'}\right) + q'n \log \left(1 - \frac{z\sigma}{q'}\right) = z \times 0 - \frac{z^2 \sigma^2 n}{2} \left(\frac{1}{p'} + \frac{1}{q'}\right) \\ + \text{含有 } \sigma^3 n \text{ 之各項,}$$

即 $+\frac{1}{\sqrt{n}}$ 次各項,

$$= -\frac{1}{2}z^2, \text{ 當 } \frac{1}{\sqrt{n}} \text{ 略去時, 因 } p' + q' = 1 \text{ 也。}$$

$$\therefore P_x = \frac{\int_0^z e^{-\frac{1}{2}z^2} dz}{\int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2} dz} = \int_0^z \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz \dots\dots\dots (117)$$

故一範圍中, 成 p' 與 $p' + p_1$ 之比例, 則對宇宙之觀察之機

率, 爲 (以 $\frac{p_1}{\sigma}$ 代 z) $\int_0^{p_1} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{u^2}{2\sigma^2}}$ 式中 $\sigma = \sqrt{\frac{p'(1-p')}{n}}$ 。

以上分析均係根據投韓德 (Todhunter) 氏之機率學史 (History of the Theory of probability), 第五五四頁起。

從一已知之範圍中, 抽出之樣本內, 自 p 至 $p + p_1$ 中間抽得

一值之機率爲 $\int_0^{p_1} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{u^2}{2\sigma^2}} du$, (式中, $\sigma = \sqrt{\frac{p(1-p)}{n}}$), 此一

定理之反面 (converse), 即逆機率也。

上文分析之難點, 乃在一假定——自 0 至 1 所有 p 之各值, 原始即有均等之機會。茲將此假定說明於下:

設 $n = 100$, $p' = .1$ 。

如觀察係從一以 $p = .07$ 之範圍得來，則 $\sigma^2 = \frac{.07 \times .93}{100}$, $\sigma =$

$.0255$, $\frac{p' - p}{\sigma} = 1.18 = z$ 。自 $p = .06$ 至 $p = .08$ 之總和機 約略等於

$p = .07$ 之機率，此可以縱坐標 $\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$ ，以 $.02$ 為橫坐標（以標

準差之倍數為單位）求得之，換言之， $.02 \div .0255 = .78$ ，該總和

機率等於 $\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(1.18)^2}$ 之 $.78 = .157$ 。如此繼續逐一計算，乃得

出下表：

p -值	求得 $p' = .1$ 之約略機率
•00—•02	•000
•02—•04	•002
•04—•06	•029
•06—•08	•157
•08—•10	•262
•10—•12	•242
•12—•14	•159
•14—•16	•084
•16—•18	•039
•18—•20	•014
•20—•22	•005
•22—•24	•001
•24—•26	•000
	•994

由觀察看來， p 可以等於 $.1$ ，標準差為 $\sqrt{\frac{.1 \times 9}{100}} = .03$ ，且正

數偏斜度頗大。

p 之值去 $.1$ 多於兩個標準差，不論 p 值原來是否有均等之機會，其機率終為甚小。

所舉此例已釋明該項定理：除中心值附近爲例外外， p 值原始機率之分配，如何假定，並無關係。在雖小而甚重要之中心區域，無論實在定律如何，一假定——在某一區域之 p 之原始機率，與該區別成比例——頗可作一適當之第一近似值（參閱：愛基華斯氏在統計學報一九〇八年號第三八七頁一文，該文並列有參考書）

吾人既假定，所求數量重要數值，於影響於此分析之數值在全距之任一點上，原始即有均等之機會，則由此以上項或類似之調查，吾人可免陷於重大之差誤。

現可將用樣本求 p 之結果，總述於下。範圍中之最或然值，即爲觀察而來之 p' 。對觀察值之離中差——大小相當於 p_1 ——之機率，可由具有標準差

$$\sqrt{\frac{p_1(1-p_1)}{n}}, \text{ 或 } \sqrt{\frac{p(1-p)}{n} \left(1 - \frac{n}{N}\right)}$$

（式中 N 爲範圍中之事物總數，且 $\frac{n}{N}$ 不可略去）之常態差誤函數，求出其近似值。

一量數之精度，可用差誤——量數所易陷入之差誤——之標準差之倒數，測算之。

第三節 通用方法

茲更以概括之語論之，設未知範圍之函數，為 X ； n 個樣本，由範圍中隨機抽選而來，設 X' 為樣本之任何已知函數；並設 $X = X' + x$ 。

吾人如能證明，當範圍中之值為 X 時，求得 X' 值之機率為 $P_x = P_0 e^{-\frac{x^2}{\sigma^2}}$ （式中， P_0 為最大機率，係當 $X = X'$ 時所得，又 σ 為常數， $x = X - X'$ ），則可有理由，確定：樣本所示之證據——一則問題中函數之最或然值乃為 X' ，二則 X' 離中差之機率，可由具有標準差 σ 之常態函數求得之。上例中， p' 為 X' ， p 為 X ， x 為 $x\sigma$ ，而 $n\sigma^2 = p(1-p')$ 。較為普通之情形，倒轉之法，並不若是之直接（註二）。

欲決定以樣本為基礎之量數之精度，其法不外三種步驟：

- (1) 求真值與觀察值相差之標準差，
- (2) 求某一指定離中差發生之機率，
- (3) 應用逆機率原理。

第四節 算術平均數之精度

在第三章，已證明，如 n 個量數，由一頻數羣類（須與某種條件相合）中隨機而彼此獨立抽選而來， n 個量數之平均數，與範圍各值之平均數，相差僅有 x 之大，則此種機率為

$$2 \int_x^{\infty} \frac{1}{\sigma a \sqrt{2\pi}} e^{-\frac{x^2}{2\sigma a^2}} dx,$$

式中, $\sigma_a = \frac{\sigma}{\sqrt{n}}$, σ 為範圍中之標準差。

由此可知, 從樣本中觀察之標準差, σ' , 與 σ 相差, 可與 $\frac{\sigma}{\sqrt{n}}$ 等量齊觀, 故如 n 甚大, σ' 可以作為與 σ 相等也。

現可完成前此之議論, 如有 n 個事物之樣本中, 而此樣本又係從一離中差不至多過兩個標準差之羣類抽選而來, 樣本之平均數為 \bar{x} , 標準差為 σ , 則範圍中平均數與 \bar{x} 相差等於 x 之機率, 當 n 甚大時, 為

$$2 \int_x^{\infty} \frac{\sqrt{n}}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x}{\frac{\sigma}{\sqrt{n}}} \right)^2} dx, \dots\dots\dots, (118)$$

第五節 標準差之精度

次復引申此定理, 以測從樣本決定之標準差及第二級動差之精度。

設 \bar{x} , σ , μ_2 , $\dots\dots$ 為宇宙之未知常數, 而以 $\bar{x} + \bar{x}'\sigma'$, μ_2 , $\dots\dots$ 為由樣本中算得之相應值。

設以 $\bar{x} + x_t$ 為任何觀察, 並以 $x'_t = x_t - \bar{x}'$ 。

x_t^2 之頻數曲線, 以 μ_2 為其平均數, 而其標準差, 則可由 $\sigma_{x^2} =$ 平均值 $(x_t^2 - \mu_2)^2 = \mu_4 - \mu_2^2$ 求來。

其第四級動差為, 平均值 $(x_t^2 - \mu_2)^2 = \mu_8 - 4\mu_6\mu_2 + 6\mu_4\mu_2^2 -$

$$3\mu_2^4 = M_4.$$

在範圍間， $\frac{\mu_s}{\sigma^s}$ 依假定為有限數，不論 s 何值，故 $\frac{M_4}{\sigma d^4} = \left(\frac{\mu_8}{\sigma^8} - 4\right)$

$\frac{\mu_8}{\sigma^8} + 6\frac{\mu_4}{\sigma^4} - 3 \div \left(\frac{\mu_4}{\sigma^4} - 1\right)^2$ 為有限數。

同理，求 x_i^2 之任何級動差， $M_s \div \sigma d^2$ 為有限數。

故依第四章第一節及本章所論之定理，樣本中各項數量，

x_i^2 ，之平均數，與 μ_2 相差，有一具有常態頻數及標準差

$$\frac{\sigma d}{\sqrt{n}} = \sqrt{\frac{\mu_4 - \mu_2^2}{n}} \text{ 之「差誤」}。$$

$$\text{此平均數, } m_2 = \frac{Sx_i^2}{n} = \frac{1}{n} S(x'_i + \bar{x}') = \frac{1}{n} Sx_i'^2 + \bar{x}'^2$$

$$\text{因 } Sx'_i = 0, \text{ 故 } = \mu'_2 + \bar{x}'^2.$$

但 \bar{x}'^2 屬於 $\frac{1}{n}$ 次 (見公式 118)，而 m_2 適已證明屬於 $\frac{1}{\sqrt{n}}$ 次。故

\bar{x}'^2 可以略去，而以 μ'_2 代 m_2 。

由此可見，觀察得來之 μ'_2 ，與在範圍間之 μ_2 ，相差有一『差誤』，此差誤乃具有常態頻數，其標準差為

$$\sqrt{\frac{\mu_4 - \mu_2^2}{n}} \dots \dots \dots (119)$$

惟 $\sigma^2 = \mu_2$ ， $\therefore \delta\sigma = \frac{\delta\mu_2}{2\sqrt{\mu_2}}$ ；故 σ' 與 σ 相差，有一具有常態頻數，

而以

$$\sqrt{\frac{\mu_4 - \mu_2^2}{4n\mu_2}} \dots\dots\dots(120)$$

爲標準差之『差誤』。

如範圍爲常態， $\mu_4 = 3\mu_2^2$ ，觀察得來標準差及第二級動差之標準差，分別變爲

$$\frac{\sigma}{\sqrt{2n}} \text{ 及 } \sigma\sqrt{\frac{2}{n}} \dots\dots\dots(121)$$

用同樣方法， $m_4 = \frac{Sx_i^4}{n}$ 之標準差爲 $\sqrt{\frac{\mu_8 - \mu_4^2}{n}}$ ，

式中， $m_4 = \frac{1}{n}S(x'_i - \bar{x}')^4 = \mu'_4 - 4\bar{x}'\mu_3 + 6\bar{x}'^2\mu'_2 - 3\bar{x}'^4$ 。

故 μ'_4 ，差誤，其方次之低，有如 m_4 之差誤，換言之，即屬於 $\frac{1}{\sqrt{n}}$ 次也。

是以計算 μ_2 與 σ 之標準差時，吾人可以已知之 μ'_4 與 μ'_2 ，代替未知之 μ_4 與 μ_2 ，並於計算平均數之標準差時，吾人可以 σ' 代 σ ，以調正之。

高級動差或相關係數之標準差及差誤頻數曲線，無論如何，不能用此法（註三）算出，茲特於下文將全部基礎重新建立，將一切論辯改弦更張，至所根據之書籍，已介紹於本章之首矣。

第六節 平均數之標準差（罔計及逆機率）

設一範圍共有可量之物件 N 個，對量數 x_1 共有 $N \times y_1$ 個，對量數 x_2 共有 $N \times y_2$ 個，其他以此類推。並設隨機抽出 n 個物件， $\frac{n}{N}$ 甚小，故求一 x 值之機率，不至因前所抽過而受影響。

$$N = N \times y_1 + N \times y_2 + \dots, \therefore y_1 + y_2 + \dots = 1.$$

以 \bar{x}' , d' , μ'_2 為樣本之平均數，標準差，及第二級動差。茲請決定此等數值之精度，以為範圍中平均數，標準差，第二動差之代表。

設 x_1, x_2, \dots 從範圍之（未知）平均數算出，故

$$\mu_1 = S(x_i y_i) = 0.$$

以 μ_2 為範圍之第二動差，故 $\mu_2 = S(x_i^2 y_i)$ ，復以 $\mu_2 = \sigma^2$ 。

樣本不能於 x_1 處有 $x \times y_1$ 個，於 x_2 處有 $N \times y_2$ 個，（餘類推）。

設實際求得者，在 x_1 處有 $(y_1 + e_1)$ 個，在 x_i 處有 $(y_i + e_i)$ 個。

於是 $e_1 + e_2 + \dots + e_i + \dots = 0$ 。

x_1, x_2, \dots 本為常數。

y_i 既為在 x_i 求一物件之機率，且實驗舉行 n 次之多， e_i 自為常態頻數，而標準差為

$$\sqrt{\left\{ \frac{y_i(1-y_i)}{n} \right\}} \quad (\text{見第二章第八節}).$$

所以 e_i^2 所有各值之平均為 $\frac{y_i(1-y_i)}{n}$ ，而 e_i 係屬於 $\frac{1}{\sqrt{n}}$ 一

次。

假以 E 爲除第 s 及第 t 區分外其餘所有之總和差誤。以 Y 代 $1 - y_s - y_t$ 。

於是 $e_s + e_t + E = 0$

$$\therefore 2e_s e_t = E^2 - e_s^2 - e_t^2$$

$$\begin{aligned} \therefore \text{平均值 } e_s e_t &= \frac{1}{2} \text{平均值 } E^2 - \frac{1}{2} \text{平均值 } e_s^2 - \frac{1}{2} \text{平均值 } e_t^2 \\ &= \frac{1}{2n} \{ Y(1-Y) - y_s(1-y_s) - y_t(1-y_t) \} \\ &= \frac{1}{2n} \{ (1-y_s-y_t)(y_s+y_t) - y_s(1-y_s) - y_t(1-y_t) \} \\ &= -\frac{y_s y_t}{n} \end{aligned}$$

現以 F 爲 y_1, y_2, \dots 之任何直線函數，則

$$F = a_1 y_1 + a_2 y_2 + \dots$$

式中， a_1, a_2, \dots 爲已知常數。

$$\text{以 } F + f = a_1(y_1 + e_1) + \dots + a_t(y_t + e_t) + \dots$$

$$\text{則 } f = a_1 e_1 + \dots + a_t e_t + \dots$$

$$f^2 = S a_i^2 e_i^2 + 2 S a_s a_t e_s e_t$$

推算至此，當 e_1, e_2, \dots 之一切可能值，已查知確爲適當之比例，以 σ^2 代 f^2 之平均值，則

$$\sigma_f^2 = S a_i^2 (\text{平均值 } e_i^2) + 2 S a_s a_t (\text{平均值 } e_s e_t)$$

$$\begin{aligned}
&= \frac{1}{n} \{S a_i^2 y_i (1 - y_i) - 2S a_i a_i y_i y_i\} \\
&= \frac{1}{n} \{S a_i^2 y_i - F^2\} \dots \dots \dots (122)
\end{aligned}$$

以 $a_1 = x_1 \dots \dots, a_t = x_t \dots \dots$

$$F = S x_i y_i = 0$$

$$\bar{x}' = F + f = S x_i e_i$$

$$\therefore \sigma_{\bar{x}'}^2 (\text{之平均值}) = \frac{1}{n} S x_i^2 y_i \quad (\text{依公式122}) = \frac{\mu_2}{n} \dots (123)$$

故 \bar{x}' 屬於與 $\frac{1}{\sqrt{n}}$ 同次。

復次，以 $a_1 = x_1^2 \dots \dots a_t = x_t^2$

$$\mu_2 = F = S x_i^2 y_i, \quad \mu_2' = S (x_i - \bar{x}')^2 (y_i + e_i)$$

$\mu_2' - \mu_2 = f = S x_i^2 e_i - 2\bar{x}' S x_i y_i + \text{包含 } \bar{x}' e_i \text{ 及 } \bar{x}'^2 \text{ 之各項，而}$

此等乃屬於 $\frac{1}{\sqrt{n}}$ 一次。

$\therefore \mu_2' = \mu_2 = S x_i^2 e_i$ ，因為 $S x_i y_i = 0$ 當屬於 $\frac{1}{n}$ 次各項略去時。

\therefore 依公式122 $\sigma \mu_2^2 = (\mu_2' - \mu_2)^2$ 之平均值

$$= \frac{1}{n} \{S (x_i^4 y_i) - \mu_2^2\} = \frac{\mu_4 - \mu_2^2}{n} \dots (124)$$

現在 $\sigma^2 = \mu_2$

所以增加數 $\delta\sigma$ ， σ 之 $\delta\mu$ ，及 μ_2 ，用下方程式，連結起來

$$2\sigma\delta\sigma = \delta\mu_2, \text{ 或 } \delta\sigma = \frac{\delta\mu_2}{2\sqrt{\mu_2}} \dots \dots \dots (125)$$

$$\text{故} \quad \sigma^2 = \text{平均值}(\delta\sigma)^2 = \text{平均值} \frac{(\delta\mu_2)^2}{4\mu_2}$$

$$\therefore \quad \sigma_{\sigma}^2 = \frac{\mu_4 - \mu_2^2}{4\mu_2 n} \quad (\text{依公式124})$$

類似之分析，得出概括之結果

$$\sigma_{\mu p}^2 = \frac{1}{n} (\mu_{2p} - 2p\mu_p + 1\mu_{p-1} + p^2\mu_{p-1}\mu_2 - \mu_p^2) \dots (126)$$

故 \bar{x} , σ 與所有各級動差之標準差，均雜有 $\frac{1}{\sqrt{n}}$ 一項，且如 n 甚大時，明顯量數與真實量數之相差，必屬於 $\frac{1}{\sqrt{n}}$ 一次，於含有此等相差之公式中，均不妨刪略之。所以，以數值表示公式 123 至 126 時，動差 $\mu'_2 \dots$ 等之計算值，可用以代替未可知之真值也。

請注意，各級動差之標準差，隨兩倍於該次之動差而定，而此較高級動差，當方次增高時，亦立即隨之而變大。實際上，據吾人所知， n 之值如為普通常見者，高過四級之動差，精度必難保持，即此故也。

如範圍之頻數曲線，係為常態，則 $\mu_4 = 3\mu_2^2$ ，且

$$\begin{aligned} \sigma &= \frac{\sigma}{\sqrt{n}}, \quad \sigma_{\mu_2} = \sigma^2 \sqrt{\frac{2}{n}}, \quad \sigma_{\sigma} = \frac{\sigma}{\sqrt{2n}}, \\ \sigma_{\mu_3} &= \sigma^3 \sqrt{\frac{6}{n}}, \quad \sigma_{\mu_4} = \sqrt{\frac{6}{n}} \dots \dots \dots (127) \end{aligned}$$

為首兩個得數，如在常態曲線下，可用下法，直接得出。設有一羣類，其類數曲線

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-x_0)^2}{2\sigma^2}},$$

(式中, x_0 與 σ 為未知數), 今由此羣類內, 隨機抽出事物 n 個, 茲以 X_1, X_2, \dots, X_n 代表之。

此等特定之 n 個事物被抽選之機率, P_x , 為

$$\frac{1}{\sigma\sqrt{\pi}} e^{-\frac{(X_1-x_0)^2}{2\sigma^2}} \times \frac{1}{\sigma\sqrt{\pi}} e^{-\frac{(X_2-x_0)^2}{2\sigma^2}} \times \dots = \frac{1}{\sigma^n \pi^{\frac{n}{2}}} e^{-s \frac{(X_1-x_0)^2}{2\sigma^2}}.$$

設 \bar{x} 為 X 各值之平均數, $X_t = \bar{x} + x_t$, 則 $\sum x_t = 0$ 。

$$\begin{aligned} \log P_x &= -n \log \sigma - \frac{n}{2} \log \pi - \frac{1}{2\sigma^2} S \{ \bar{x} - x_0 + x_t \}^2 \\ &= -n \log \sigma - \frac{n}{2} \log \pi - \frac{1}{2\sigma^2} \{ n(\bar{x} - x_0)^2 + ns^2 \}, \end{aligned}$$

式中 s 為 X 各值之標準差。

\bar{x} 及 s 為已知數, 據以決定 σ 及 x_0 。

$\bar{x} - x_0$ 為最小時, 不論 σ 之值如何, P_x 必為最大。

設 x_0 為 \bar{x} 。

於是 P_x 為最大, 當 $\frac{dP_x}{d\sigma}$ 為零, 即當

$$0 = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \cdot ns^2, \text{ 又 } \sigma = s.$$

以 $\sigma = s + \gamma$, $x_0 = \bar{x} + \delta$, 並以 P_0 為 P_x 於 $\gamma = 0 = \delta$ 時之值。則

$$\begin{aligned} \log P_x - \log P_0 &= -n \log \frac{\sigma}{s} - \frac{n}{2\sigma^2} (\delta^2 + s^2) + \frac{n}{2s^2} S^2 \\ \frac{1}{n} \log \frac{P_x}{P_0} &= -\log \left(1 + \frac{\gamma}{s} \right) - \frac{1}{2s^2} \left(1 + \frac{\gamma}{s} \right)^{-2} (\delta^2 + s^2) + \frac{1}{2} \\ &= -\frac{\gamma}{s} + \frac{\gamma^2}{2s^2} - \dots - \frac{\delta^2}{2s^2} + \frac{2\gamma}{2s^3} \cdot s^2 - \frac{3\gamma^2}{2s^4} \cdot s^2, \end{aligned}$$

刪略 $\gamma^3, \gamma\sigma^2, \dots$,

$$= -\frac{\gamma^2}{s^2} - \frac{\delta^2}{2s^2}.$$

$$\therefore P_{\Delta} = P_0 e^{-\frac{1}{2} \cdot \left(\frac{\delta}{\frac{s}{\sqrt{n}}}\right)^2} \cdot e^{-\frac{1}{2} \left(\frac{\gamma^2}{\left(\frac{s}{\sqrt{2n}}\right)^2}\right)^2} \dots \dots \dots (128)$$

故 x 及 s 之差誤，彼此互不相干，且此等差誤均成常態之類數，而其標準差分別為 $\frac{s}{\sqrt{n}}$ ， $\frac{s}{\sqrt{2n}}$ ，（ s 為樣本之標準差，此可以範圍之標準差， σ ，代表之，當 n 甚大時，上文解答已得，不待詞費矣）。

第七節 相關係數之標準差

相關係數每易發生之差誤之標準差，可以下法得之（下文以薛伯博士（Dr. Sheppard）之法為依據）。

設有數值二對，係從平均數測算之離中差，如 x_i, y_i 是。該兩平均數之標準差為 σ_1, σ_2 ，第二動差為 λ, μ 。設共有對數 N 個，並設 $x_i N$ 之坐標，係在 $(\bar{x} + x_i, \bar{y} + y_i)$ 處，然則 $z_1 + \dots + z_i \dots = 1$ 。

$$\text{同時，} S_{z_i x_i} = 0 = S_{z_i y_i}.$$

經由各對取其 $x^s y^l$ ，而以 M_{sl} 代表 $x^s y^l$ 之平均值。於是 $M_{s1} = S_{z_i x^s y^1}$ 。復以 M 代 M_{11} 。

如是，如 r 為 N 對之相關係數，

$$r = \frac{M}{\sigma_1 \sigma_2}, \log r = \log M - \frac{1}{2} \log \lambda - \frac{1}{2} \log \mu.$$

現設隨機抽出 n 對，並設於 x_i, y_i 處，選得之數，為 $(z_i + e_i) n$ ，

假如 x', y' 為求得之平均數

$$x' = S(z_i + e_i) x_i = S_{z_i e_i}, \text{ 又 } y' = S_{y_i e_i}.$$

樣本各值與範圍（有 r, M, λ, μ ）各值之離差，將方程式對 $\log r$ 微分，可成下列聯系之條件。

$$\frac{\delta r}{r} = \frac{\delta M}{M} - \frac{\delta \lambda}{2\lambda} - \frac{\delta \mu}{2\mu}$$

當微量如 e_t, x', y' （均屬於 $\frac{1}{\sqrt{n}}$ 次）任何兩數之乘積被省略時，

$$\delta M = S(z_t + e_t)(x_t - x')(y_t - y') - S z_t x_t y_t = S x_t y_t e_t - x' \cdot S z_t y_t - y' \cdot S z_t x_t$$

$$\therefore \delta M = S x_t y_t e_t$$

如上（第六節公式122-126）所述，

$$\delta \lambda = S x_t^2 e_t \text{ 而 } \delta \mu = S y_t^2 e_t$$

$$\therefore \frac{\delta r}{r} = S \left(\frac{x_t y_t}{M} - \frac{x_t^2}{2\lambda} - \frac{y_t^2}{2\mu} \right) e_t$$

故依概括公式 122，如 σ_r 為 r 之差誤之標準差，

$$\begin{aligned} \sigma_r^2 &= \frac{r^2}{n} \left[S \left(\frac{x_t y_t}{M} - \frac{x_t^2}{2\lambda} - \frac{y_t^2}{2\mu} \right) z_t - S \left\{ \left(\frac{x_t y_t}{M} - \frac{x_t^2}{2\lambda} - \frac{y_t^2}{2\mu} \right) z_t \right\}^2 \right] \\ &= \frac{r^2}{n} \left\{ \frac{M_{22}}{M^2} + \frac{\lambda_4}{2\lambda^2} + \frac{\mu_4}{4\mu^2} - \frac{M_{31}}{\lambda M} + \frac{M_{13}}{\mu M} - \frac{M_{22}}{2\lambda\mu} \right\} \end{aligned}$$

$$\text{因 } S \left(\frac{x_t y_t}{M} - \frac{x_t^2}{2\lambda} - \frac{y_t^2}{2\mu} \right) z_t = 1 - \frac{1}{2} - \frac{1}{2} = 0,$$

式中， λ_4 與 μ_4 為 X

如原始分配，與常態相關面相合， $\lambda_4 = 3\lambda^2, \mu_4 = 3\mu^2, M = r\sigma_1$

$\sigma_2, M_{22} = (1+2r^2) \sigma_1^2 \sigma_2^2, M_{31} = 3r \sigma_1^3 \sigma_2, M_{13} = 3r \sigma_1 \sigma_2^3,$ (見公式 106), 又

$$\sigma r^2 = \frac{r^2}{n} \left\{ \frac{1+2r^2}{r^2} + \frac{3}{4} + \frac{3}{4} - 3 - 3 + \frac{1+2r^2}{2} \right\} = \frac{(1-r^2)^2}{n},$$

而 $\sigma_r = \frac{1-r^2}{\sqrt{n}} \dots \dots \dots (129)$

此為一般通用者，但係以分配近於常態為假設。

迴歸係數， y 對 x 而言，為 $r \frac{\sigma_y}{\sigma_x} = \rho$ 。用此符號， $\rho = \frac{M}{\lambda}$ ，以與上述類似之法，仍可得之，

$$\sigma \rho^2 = \frac{\rho^2}{n} \left\{ \frac{M_{22}}{M^2} + \frac{\lambda^4}{\lambda^2} \frac{2M_{31}}{\lambda M} \right\}$$

在任何種分配情形下。

在常態分配下

$$\sigma \rho^2 = \frac{\rho^2}{n} \left\{ \frac{1+2r^2}{r^2} + 3 - 6 \right\} = \frac{\rho^2}{n} \cdot \frac{1-r^2}{r^2}$$

$$\therefore \sigma \rho = \frac{1}{\sqrt{n}} \cdot \frac{\sigma_y}{\sigma_x} \cdot \sqrt{1-r^2}.$$

在常態分配下，用下列方法（茲根據皮爾生教授方法）亦可得之：

假設 $x_1, y_1, \dots, x_n, y_n$ 各對，從一平面抽選而來，該平面之未知中心為 x_0, y_0 ，標準差為 σ_1, σ_2 ，平均積為 $r \sigma_1 \sigma_2$ 。

設 $\bar{x}, \bar{y}, s_1, s_2, r'$ ，係由樣本中，算出。

n 對實現之機率為

$$P = \frac{1}{(2\pi\sigma_1\sigma_2\sqrt{1-r^2})^n} e^{-\frac{1}{2(1-r^2)} \left\{ \frac{(xt-x_0)^2}{\sigma_1^2} + \frac{(yt-y_0)^2}{\sigma_2^2} - \frac{2r(xt-x_0)(yt-y_0)}{\sigma_1\sigma_2} \right\}}.$$

$$\begin{aligned} \therefore \log Pz &= -n \log 2\pi \sigma_1 \sigma_2 - \frac{n}{2} \log(1-r^2) \\ &\quad - \frac{n}{2(1-r^2)} \left\{ \frac{s_1^2 + d_1^2}{\sigma_1^2} + \frac{s_2^2 + d_2^2}{\sigma_2^2} - \frac{2r'(r's_1s_2 + d_1d_2)}{\sigma_1\sigma_2} \right\} \end{aligned}$$

式中, $d_1 = x_0 - \bar{x}$, $d_2 = y_0 - \bar{y}$ 。

現在 $r, \sigma_1, \sigma_2, d_1, d_2$ 爲未知, r', s_1, s_2 爲已知。

$$\text{述明 } \frac{\partial P}{\partial d_1} = 0 = \frac{\partial P}{\partial d_2} = \frac{\partial P}{\partial \sigma_1} = \frac{\partial P}{\partial \sigma_2} = \frac{\partial P}{\partial r},$$

等幾種條件後,吾人乃得此五個未知數,蓋此五個未知數,乃使 Pz 爲最大也。

$$\frac{d_1}{\sigma_1^2} - \frac{rd_2}{\sigma_1\sigma_2} = 0 = \frac{d_2}{\sigma_2^2} - \frac{rd_1}{\sigma_1\sigma_2},$$

由此 $d_1 = d_2 = 0$, 除非 $r^2 = 1$ 時則反是。

如此,以 d_1 及 d_2 作零,

$$\frac{1}{\sigma_1} - \frac{s_1^2}{(1-r^2)\sigma_1^3} + \frac{rr's_1s_2}{(1-r^2)\sigma_1^2\sigma_2} = 0 = \frac{1}{\sigma_2} - \frac{s_2^2}{(1-r^2)\sigma_2^3} + \frac{rr's_1s_2}{(1-r^2)\sigma_1\sigma_2^2}$$

$$\therefore (1-r^2)\sigma_1^2\sigma_2 = s_1^2\sigma_2 - rr's_1s_2\sigma$$

並且

$$(1-r^2)\sigma_2^2 = s_2^2\sigma_1 - rr's_1s_2\sigma_2,$$

由此

$$\frac{s_1}{\sigma_1} = \frac{s_2}{\sigma_2} k, \text{ 又 } 1-r^2 = k^2(1-rr')$$

$$\text{又 } \frac{r}{1-r^2} - \frac{r}{(1-r^2)^2} \left\{ \frac{s_1^2}{\sigma_1^2} + \frac{s_2^2}{\sigma_2^2} - \frac{2rr's_1s_2}{\sigma_1\sigma_2} \right\} + \frac{r's_1s_2}{(1-r^2)\sigma_1\sigma_2} = 0.$$

$$\therefore r(1-r^2) - 2rk^2(1-rr') + r'(1-r^2)k^2 = 0.$$

故 $r = r'$, 又 $k = 1$

Pz 故爲最大值, 當樣本中各值, 作爲平面中各值時, 卽以求得之 Pz , 作爲 P_0 。

現以 $\sigma_1 = s_1 + \gamma_1$, $\sigma_2 = s_2 + \gamma_2$, 並 $r = r' + \rho$, 次將有微量 $d_1, d_2, \gamma_1, \gamma_2, \rho$ 之乘方

之一切函數展開, 而刪略去其第三乘方。則得

$$\begin{aligned} \frac{1}{n} \log \frac{P_2}{P_0} &= -\frac{1}{2(1-r'^2)} \left(\frac{d_1^2}{s_1^2} + \frac{d_2^2}{s_2^2} - \frac{2r'd_1d_2}{s_1s_2} \right) + \frac{r}{2-r'^2} \left(\frac{\rho\gamma_1}{s_1} + \frac{\rho\gamma_2}{s_2} \right) \\ &\quad + \frac{r'^2}{1-r'^2} \cdot \frac{\gamma_1\gamma_2}{s_1s_2} - \frac{2-r'^2}{2(1-r'^2)} \left(\frac{\gamma_1^2}{s_1^2} + \frac{\gamma_2^2}{s_2^2} \right) - \frac{1+r'^2}{2(1-r'^2)} \rho^2 \\ &= -\frac{1}{2(1-r'^2)} \left(\frac{d_1}{s_1} - \frac{r'd_2}{s_2} \right)^2 - \frac{d_2^2}{2s_2^2} - \frac{2-r'^2}{2(1-r'^2)} \left(\frac{\gamma_1}{s_1} - \frac{r'^2}{2-\gamma'^2} \cdot \frac{\gamma_2}{s_2} - \frac{r'}{2-r'^2} \rho \right)^2 \\ &\quad - \frac{2}{2-r'^2} \left(\frac{\gamma_2}{s_2} - \frac{r'\rho}{2(1-r'^2)} \right)^2 - \frac{\rho^2}{2(1-r'^2)^2} \end{aligned}$$

對兩極限依次積分

對 d_1 積分, 以 $d_2, \gamma_1, \gamma_2, \rho$ 作常數,

對 d_2 積分, 以 γ_1, γ_2, ρ 作常數,

對 r_1 積分, 以 γ_2, ρ 作常數,

對 r_2 積分, 以 ρ 作常數。

於是求得: 不論 $x_0, y_0, \sigma_1, \sigma_2$ 之值為何, 凡從 $r' + \rho$ 一值而來之觀察, 其全部機率為

$$P = Ke^{-\frac{1}{2} \frac{\rho^2}{\left(\frac{1-r'^2}{\sqrt{n}}\right)^2}}$$

換言之, 此分配係為常態, 其 r' 之標準差為 $\frac{1-r'^2}{\sqrt{n}}$ 。

第六節末及本節之末, 用小體字之推算, 指示吾人, 如樣本所自產生之頻數羣類為常態, 則求得樣本中之 \bar{x}, σ, μ 及 r 內各種差誤之機率, 可由常態機率函數 (the normal probability function) 得來; 反之, 範圍各數量對樣本相當各數量之離差, 其機率亦可用之求得。惟於他種條件下, 亦可求得同一解答, 茲尙

待加以證明。

無論何種情形，該種數量，係列如下式：

$$F + f = a_1(y_1 + e_1) + a_2(y_2 + e_2) + \dots,$$

式中， $e_1 + e_2 + \dots = 0$ ，又於 $0 = e_1 = e_2 = \dots$ 時， $f = 0$ 。

同理 $y_1 + y_2 + \dots = 1$ 。

樣本所取個數， n ，為甚大時（本章第六節）， e_1, e_2, \dots 之頻數曲線，必為常態。

如 e_1, e_2, \dots 彼此互不相干，或 x_1, x_2, \dots 各值，數目之多堪供吾人認為果係彼此獨立，則吾人可立即用第三章第四節之定理，而決定 f 之頻數係為常態。

如將全部分析（見附錄九）作完，吾人可得一結論：如一範圍（樣本所自出）之條件相同，則在此條件之下，吾人可認定為常態，由此常態性，即達到平均數之常態性（the normality of average）。換言之，該範圍所受之限制，為無論 t 值為何， $\frac{f}{\sigma^t}$ 終為有限數也。

（註一）參閱統計學報一九〇八年號第三百八十一頁起愛基華氏一文；猶爾著統計原理概論之最後一章；皇家學會會報一八九八年，第一九一卷（A. 220）皮爾生及費那（Félon）一文；第一九二卷（A. 229）薛伯（Sheppard）一文；（Biometrika）第二卷，第三部，第二八〇頁。

（註二）如 X 之所有各值，原即有均等之機會，則從範圍觀察（當 X 之值在 X'

士 x 極限內)之機率為 $2 \int_0^x P_x \cdot dx$, 如 x 甚小時; 又用逆機率, 宇宙中之值在此等極限內之機率, 為

$$2 \int_0^x P_x dx \div \int_{-\infty}^{\infty} P_x dx = \frac{2}{\sigma \sqrt{2\pi}} \int_0^x e^{-\frac{1}{2} \frac{x^2}{\sigma^2}} \cdot dx.$$

(註三) 此法係根據愛基華斯給着者之通信。

第十章 資料與公式適應之測驗

第一節 測驗方法

關於以數學公式，表示觀察資料，其通用方法，前已屢見不鮮，然公式究竟恰當與否，此為必有之疑問，為解答此疑問，於是測驗乃不可少矣。以常用語言之，此即配合完好與否之測驗問題也。

茲舉例論之，前述（第三章第七節第七例）九百七十家每『單位』每週食物費用，列表於下。

食物費用	m' 實在件數	m 計算件數	$e = m \sim m'$ 相差額	標準差	$\frac{e^2}{m}$
5.5先令以下	18	22	4	4.6	.7
5.5.....	107	123	16	10.4	2.1
7.5.....	255	234	21	13.3	1.9
9.5.....	245	249	4	13.6	.1
11.5.....	173	168	5	11.8	.1
13.5.....	101	89	12	9.0	1.6
15.5.....	38	51	13	7.0	3.3
17.5.....	17	22	5	4.6	1.1
19.5.....	9	11	2	3.3	.4
21.5以上	7	1	6	?	36.0
總計	970	970	88	---	47.3

計算得來之件數，係由大數律之第二近似值而得。以前所用之粗率方法，係將計算數與各區分觀察數之差額相加，不計符號之正負，而以其總計，作為總件數之百分數。如此算得之不相配合之百分數(percentage misfit)，即為 $88 \div 9.70 = 9.1\%$ 也。

此法缺點，在不能用任何之機率量數，令人一見不知其是否完善。綜觀二種公式，當視其求出之不相配合百分數，以最低者為較完善；又當有數組之類似觀察時，吾人可用此法，查出對公式之配合為最近，且可知何組觀察之分配最為整齊也。

但如將區分拼合，則不相配合之百分數，大致可以消滅也。

為測驗各個區分(compartment)之內容，尚有一簡法。如觀察共有 N 個， m_t 為一區分內之計算數，則於此區分內，求出觀察 $m_t + e_t$ 個之機率，為

$$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2} \cdot \frac{e_t^2}{\sigma^2}} \quad (\text{公式19}),$$

式中 $\sigma^2 = \frac{m_t}{N} \left(1 - \frac{m_t}{N}\right) N$ ，至於超過標準差某一倍數或次倍數(sub-multiple)之機率，則已見於常態機率表(第二章第四節)矣。上例每級，除最末一級外，其標準差，已述如前，由此可知，九個差額中，少於標準差， σ ，者，佔其四，在 σ 與 $\frac{3\sigma}{2}$ 之間者，只有其二，下餘三者，則無過於 2σ 。各量數均有其或然之機會，故全部羣類，可以認為有或然性，至最後一項，在 21.5 先令以上有七件，

乃其例外耳。

極端各級與中間各級之必不能連貫，乃一般觀察多類所不免者。

雖然，離中差並非為獨立，蓋其總數必為零；故即使某一區分內之離中差，以其本身觀之，為數之大，幾令人不能相信，但就全盤區分而論，則亦並非不可能之事。適於此種變更之測量，皮爾生教授曾為之制定，茲將此種之一部分分析，用簡單化之形式，並將其計算得數，列成簡表後，述及其應用方法，於下文討論之（請參閱 the Philosophical Magazine, 第 302 號 157-175 頁，一九〇〇年七月出版）。

設有一公式，假設其果能代表觀察之分配；茲由此公式，得出 m_1, m_2, \dots, m_n 個觀察，分成 n 級或區分，而觀察總數， $N = m_1 + m_2 + \dots + m_n$ 。

作一實驗，或一類觀察，設各區分為 $(m_1 + e_1) \dots (m_t + e_t) \dots (m_n + e_n)$ ，使 $e_1 + \dots + e_t + \dots + e_n = 0$ 。

$$\text{以 } p_1 = \frac{m_1}{N} \dots p_t = \frac{m_t}{N} \dots$$

設一羣類頗能適合公式，並從此羣類內作一觀察，則此一觀察係落在第 t 級內之機率為 p_t 。

從無限大之範圍中，隨機抽取 N 個，則 $m_t + e_t$ 落於此級之機率，為

$$\frac{1}{\sigma_t \sqrt{2\pi}} e^{-\frac{et^2}{2\sigma_t^2}},$$

式中, $\sigma_t^2 = p_t(1-p_t)N = p_t q_t N$, 而 $q_t = 1 - p_t$.

至於所指各種差誤之連帶機率(joint chance), 爲

$$K e^{-\frac{1}{2}X^2}, \text{ 式中 } x^2 = S \cdot \frac{e_t^2}{m_t}, \text{ 又 } S e_t = 0,$$

而 K 則爲一常數。

因爲, 如共有兩個區分, $e_1 + e_2 = 0$, 且其連帶機率, 等於每個之機率。

$$\text{於是 } p = \frac{m_1}{N}, q = \frac{m_2}{N}, m_1 + m_2 = N.$$

機率爲

$$\frac{N^{\frac{1}{2}}}{\sqrt{2\pi m_1 m_2}} e^{-\frac{1}{2} \left(\frac{e_1^2}{m_1} + \frac{e_2^2}{m_2} \right)}, \text{ 因 } \frac{e_1^2 N}{m_1 m_2} = \frac{e_1^2 (m_2 + m_1)}{m_1 m_2}, \text{ 且 } e_1^2 = e_2^2.$$

如共有三個區分, $e_1 + e_2 + e_3 = 0$,

$$m_1 + m_2 + m_3 = N, \sigma_1^2 = \frac{m_1}{N} \cdot \frac{m_2 + m_3}{N} \cdot N,$$

對 σ_2^2, σ_3^2 , 亦然。

$$2e_1 e_2 = e_3^2 - e_1^2 - e_2^2,$$

$$r\sigma_1\sigma_2 = \text{平均值 } e_1 e_2 = \frac{1}{2}(\sigma_3^2 - \sigma_1^2 - \sigma_2^2)$$

$$= \frac{1}{2N} \{ m_3(m_1 + m_2) - m_1(m_2 + m_3) - m_2(m_1 + m_3) \}$$

$$= -\frac{m_1 m_2}{N}. \text{ (請與第九章第六節比照)}$$

e_1 與 e_2 —— 亦即 e_3 —— 併發之機率，依常態相關面，乃為

$$\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} e^{-\frac{1}{2(1-r^2)}\left(\frac{e_1^2}{\sigma_1^2} + \frac{e_2^2}{\sigma_2^2} - \frac{2re_1e_2}{\sigma_1\sigma_2}\right)}$$

現在

$$\begin{aligned}\sigma_1^2\sigma_2^2(1-r^2) &= \frac{m_1m_2(m_2+m_3)(m_1+m_2)}{N^2} - \frac{m_1^2m_2^2}{N^2} \\ &= \frac{m_1m_2m_3}{N}\end{aligned}$$

因 $m_1+m_2+m_3=N$ 也。

故 e 之指數為

$$\begin{aligned}& -\frac{N}{2m_1m_2m_3}(e_1^2\sigma_2^2+e_2^2\sigma_1^2-2r\sigma_1\sigma_2e_1e_2) \\ &= -\frac{N}{2m_1m_2m_3}\left\{\frac{e_1^2m_2(m_1+m_2)}{N} + \frac{e_2^2m_1(m_2+m_3)}{N} + \frac{2e_1e_2m_1m_2}{N}\right\} \\ &= -\frac{1}{2m_1m_2m_3}\{(e_1+e_2)^2m_1m_2+e_1^2m_2m_3+e_2^2m_1m_3\} \\ &= -\frac{1}{2}\left(\frac{e_1^2}{m_1} + \frac{e_2^2}{m_2} + \frac{e_3^2}{m_3}\right), \text{ 因 } e_1+e_2=-e_3.\end{aligned}$$

現設第二第三個區分，業已合併為一，因而內中包 $M+E$ 個觀察（其 $M=m_2+m_3$, $E=e_2+e_3$ ），則機率為

$$K_1 \cdot e^{-\frac{1}{2}\left(\frac{e_1^2}{m_1} + \frac{E^2}{M}\right)},$$

式中 K_1 為一常數。

由此觀之，只除第二區分，而不牽動第一區分，其效果，使常數變換，且在指數以 $\frac{e_2^2}{m_2} + \frac{e_3^2}{m_3}$ 代替 $\frac{E^2}{M}$ 。

同理，只有二區分，僅除其第三，而不動其第一第二，其效果必變換常數並在指數中以 $\frac{e_3^2}{m_3} + \frac{e_4^2}{m_4}$ 代替 $\frac{e_3^2}{m_3}$ 。其他以此類推。

故對於 n 個區分，差誤 e_1, e_2, \dots, e_n 機率， P ，為

$$Ke^{-\frac{1}{2}X^2} \text{ 式中, } x^2 = \frac{e_1^2}{m_1} + \frac{e_2^2}{m_2} + \dots + \frac{e_n^2}{m_n},$$

且 $e_1 + e_2 + \dots + e_n = 0 \dots \dots \dots (130)$

於此請注意， x^2 之式，與用以求相依係數者相同。

[皮爾生氏，曾用複相關方程——而不用上述歸納法——證明此公式。]

如各個區分之抽取，能以獨立，並不受 $e_1 + e_2 + \dots = 0$ 一條件之拘束，則機率乃為

$$Ke^{-\frac{1}{2}X^2} \times e^{-\frac{1}{2}\left(\frac{e_1^2}{N-m} + \frac{e_2^2}{N-m} + \dots\right)}$$

而其指數，乃為

$$-\frac{1}{2}\left(\frac{e_1^2 N}{m_1(N-m_1)} + \dots\right) = -\frac{1}{2}\left(\frac{e_1^2}{m_1} + \frac{e_1^2}{N-m_1} + \dots\right)。$$

設有區分甚多，最大之分數 $\frac{m_i}{N}$ 卻仍甚小，則指數之後段，與前段比較，相形之下，大可刪略，於是兩式漸至相等，且相關之效

驗甚弱也。

如中間並無相關，則最末一因子——如不刪略——既小於1實現之機率，必較有相關者為小。（進一步之推算中須將常數消除）。故不相關機率之總和，雖比現用方法為簡，但對公式之適切性，頗予吾人以不當不利之認識也。

差誤之每一系，凡得出 x^2 之某一值者，其機率必相同。論到對常態頻數平均數之離中差之機率，慣常之測算法，向左或向右如其大之離中差，發現之機率，為

$$2 \int_{\varepsilon}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz_0$$

依此同理，於測算此差誤系——或一或然性較小之系——發生之機率，只須將下式代以數值即得：

$$2 \int \int \dots \dots K e^{-\frac{1}{2}X^2} dx,$$

式中 dx 係代表 $de_1 de_2 \dots de_{n-1}$ ，積分有 $n-1$ 次，自 x 展至 ∞ ，其條件為 $e_1 + e_2 + \dots + e_n = 0$ ，至 K 如何選擇，須視 $\int_{-\infty}^{\infty} K e^{-\frac{1}{2}X^2} d = 1$ 而定。

因有此條件，致積分甚為繁雜，故仍以皮爾生氏推算之原本為參考為便。

至其得數為（當 n 為偶數時）

$$P = \sqrt{\frac{2}{\pi}} \int_{\varepsilon}^{\infty} e^{-\frac{1}{2}X^2} \cdot dx + \sqrt{\frac{2}{\pi}} e^{-\frac{1}{2}X^2} \left(\frac{x}{1} + \frac{x_3}{1.3} + \dots + \frac{x^{n-3}}{1.3 \cdot 5 \cdot \dots \cdot n-3} \right)$$

當 n 爲奇數時，爲

$$P = e^{-\frac{1}{2}X^2} \left(1 + \frac{X^2}{2} + \dots + \frac{X^{n-3}}{2 \cdot 4 \dots (n-3)} \right) \dots \dots \dots (131)$$

x^2 與 n 之各值，求出之 P 值，於第一卷，第 155 頁起之 *Biometrika*，曾列一表。茲爲決定一公式，是否能以充分代表一觀察之羣類，以簡短形式，而建立演算法則。吾人方法，只須取出 x^2 之值，至 x^2 則對某一 n 值，使 $P = \frac{1}{2}$ 強，或照查常態機率表，使 $P = .0455$ 弱，如此，乃相當於常態曲線中之兩個標準差。

n .	X .	P .	X^2	P .
3	1	.61	6	.050
4	2	.57	8	.046
5	3	.56	10	.040
6	4	.55	12	.035
7	5	.54	13	.043
8	6	.54	15	.036
9	7	.54	16	.042
10	8	.53	18	.035
11	9	.53	19	.040
12	10	.53	20	.045
13	11	.53	22	.038
14	12	.53	23	.042
15	13	.53	24	.046
16	14	.526	26	.038
17	15	.525	27	.041
18	16	.524	28	.045
19	17	.523	30	.037
20	18	.522		
25	23	.520		
30	28	.518		

如 $x^2 < n - 2$, 則觀察是否現於公式所代表之羣類, 其得失成敗之數乃相等。

如 $x^2 > 2n$, 則反機率甚大。

嚴格言之, 測驗應用時, 應以觀察原有之區分為區分, 因區分之歸併, 足以波及結果之 P 值。然事實上, 降格而用未分級之觀察, 時常遭遇困難, 且在連續變量之例, 如身長, 原始之分級, 乃隨測量之可能, 而細密至無以復加也。

但此尚為較小之難點, 其更為嚴重者, 在任一區分內, 觀察之 $m_i + e_i$, 必須為整數, 而一般 m_i 則不能, 且 e_i 之值, 有幾個乃於代表最完好之分配中, 方能出現。故其結果, 數目最少之區分內, 期望之數額, 亦應有合理的大, 否則對於 x^2 , 將不免有虛偽之分攤。此在事實上規定詳細之極端區分; 且不論其或拒絕或融合, 武斷之元素, 終不能免, 而詳密之測算, 亦不可能矣。

反之, 根據愛基華斯氏假說 (第三章第五節), 測驗差誤常態曲線, 或大數律之應用時, 在橫坐標於標準差小量倍數——對於測量之獨立元素, 數目愈低, 標準差亦愈少——之外而希望其密切配合, 絕無此理, 故測驗僅可用於數目甚多之中間區分; 然專顧及施用測量之範圍, 又將失去此法之詳密性矣。

故用此法之結果, 僅可得一寬泛, 但常能充分確定之得數。

第二節 例證

如吾人將第三章第七節第七例之極端一級刪去，則 $x^2=11.3$, $n=9$, $P=.18$ ，且即用公式第二近似值已足。

如用皮爾生氏公式（在同頁） $x^2=21.4$, $n=9$, $P=.006$ ，但如將最低最高極端級除外，則 $x^2=4.1$, $n=8$, $P=.77$ ；故此公式，除該類之極端級外，對於中間八級之表現，實優為之。

如僅分別計算各級之標準差，則所得結論亦同。

又第三章第七節第五例之學童年齡表， $n=8$ 。由常態曲線得： $x^2=16.7$, $P=.02$ ，竟不能適合。惟用第二近似值， $x^2=.47$ ，則 P 與 1 乃不能分開矣。

關於字母數之實驗（見第三章第七節第一例），十字之和，以 5 個字母為一級，得 $n=13$ ，而由常態曲線，得 $x^2=33$, $P=.001$ ，如取消二最高最低兩極端級， $n=10$, $x^2=6.1$, $P=.73$ 。惟第二近似值，雖盡包所有各級，仍得 $x^2=8.4$, $P=.74$ 。

100 字之和，以 20 字母為一級，由常態曲線得 $n=10$, $x^2=2.96$, $P=.975$ ，已不能再用更高級近似值。

茲為證實抽樣法則，或上述方法之適當起見，特舉一各種不同之例，以抽樣所成之分配，與樣本所自來之羣類比較。

以各種定率付股息之公司數

	樣本中數額 m'	所有各公司之相對數額 m	標準差	$\frac{e^2}{m}$
百分之三以下	34	30	5.3	.50
百分之三	108	108.8	8.9	0
百分之四	117	124.4	9.3	.44
百分之五	60	70.8	7.4	1.65
百分之六至百分之八	48	43.2	6.2	.53
百分之八	33	22.8	4.6	4.57
	400	400		7.72

由此, $n=6$, $\chi^2=7.72$, $P=.185$ 。結果良佳, 可惜最高級多有不合耳。

此種測驗, 應用於二次元(dimension)之分配, 如第七章之字每總和一千對之例, 即是。

從中心向左右各三, 向上下各二, 在理論上原有十一以上之觀察者, 茲共得二十四塊方格, 分別作為區分。外面方格, 用武斷方法, 以重線分九部, 以使各格貼近, 而期在第二近似值, 至少可得九個期望之觀察數。茲將結果列下:

	常態曲線		第二近似值	
	χ^2	P.	χ^2 .	P.
24個中間方格	20.8	.59	17.5	.79
9個外部方格	27.8		10.1	
共三十個區域	48.6	.035	27.6	.59

用第二近似值, 外部之改進, 甚為顯著。

附 錄

數 學 摘 錄

一、求 π 值之瓦立斯(Wallis)氏定理

經用簡單作圖之研究，吾人可見：當 n 爲一正整數時

$$\int_0^{\frac{\pi}{2}} \sin^{2n+1}x \cdot dx < \int_0^{\frac{\pi}{2}} \sin^{2n}x \cdot dx < \int_0^{\frac{\pi}{2}} \sin^{2n-1}x \cdot dx$$
$$\therefore \frac{2 \cdot 4 \cdot 6 \cdots 2n}{3 \cdot 5 \cdot 7 \cdots (2n+1)} < \frac{1 \cdot 3 \cdot 5 \cdots (2n-1)}{2 \cdot 4 \cdot 6 \cdots 2n} \cdot \frac{\pi}{2} < \frac{2 \cdot 4 \cdot 6 \cdots (2n-2)}{3 \cdot 5 \cdot 7 \cdots (2n-1)}$$
$$\therefore \frac{2^{2n}(n!)^2}{(2n+1)!} < \frac{(2n)!}{2^{2n}(n!)^2} \cdot \frac{\pi}{2} < \frac{2^{2n}(n!)^2}{(2n)!} \cdot \frac{1}{2n}$$
$$\therefore \frac{2^{2n}(n!)^2}{(2n)! \sqrt{2n+1}} < \sqrt{\frac{\pi}{2}} < \frac{2^{2n}(n!)^2}{(2n)! \sqrt{2n}}$$
$$\therefore \sqrt{\frac{\pi}{2}} = \frac{2^{2n}(n!)^2}{(2n)! \sqrt{2n}}, \text{ 至 } \frac{1}{n} \text{ 爲有效} \cdots \cdots \cdots (132)$$

請參閱，吉卜生著：微積分論 (Gibson: Treatise on the Calculus, 1806, EX. XXVI. 22)

二、整數之乘羈總和

如假定

$$S_m = \sum_{t=1}^{t=m} tr = am^{r+1} + bm^r + cm^{r-1} + \dots, ,$$

用歸納法可求 a, b, c 。

$$\begin{aligned} \text{因 } (m+1)^r - S_{m+1} - S_m &= a\{(m+1)^{r+1} - m^{r+1}\} \\ &+ b\{(m+1)^r - m^r\} + c\{(m+1)^{r-1} - m^{r-1}\} \dots \end{aligned}$$

令 m^r , 及 m^{r-1} 之係數相等, 則得

$$1 = a(r+1)$$

$$r = a \frac{(r+1)r}{2} + br, \text{ 又 } b = \frac{1}{2}$$

$$\frac{r(r-1)}{2} = a \frac{(r+1) \cdot (r-1)r}{6} + b \frac{r(r-1)}{2} + c(r-1), c = \frac{r}{12} \text{ 等等。}$$

$$\therefore \frac{1r + 2r + \dots + m^r}{m^{r+1}} = \frac{1}{r+1} + \frac{1}{2m} + \frac{r}{12m^2} +$$

$$\text{如將 } \frac{1}{m} \text{ 略去,} \quad = \frac{1}{r+1},$$

$$\text{如將 } \frac{1}{m^2} \text{ 略去,} \quad = \frac{1}{r+1} + \frac{1}{2m}, \dots \dots \dots (133)$$

三、求 $m!$ 之斯德令(Stirling)公式

用瓦立斯氏定理, 可求得此公式之第一近似值。

以

$$z = \frac{(2m)!}{(2m)^{m+1}(m-1)!} = (2m-1)(2m-2)\cdots(2m-m) \\ \div (2m)^m.$$

$$\log z = \sum_{t=1}^{t=m} \log \left(1 - \frac{t}{2m}\right)$$

$$-\log z = \frac{1+2+\cdots+m}{2m} + \cdots + \frac{1+2^r+\cdots+m^r}{(2m)^r} + \cdots \\ = \sum_{r=1}^{r=\infty} \left\{ \frac{1}{r \cdot 2^r} \cdot \left(\frac{m}{r+1} + \frac{1}{2} + \frac{r}{12m} \right) \right\},$$

此係根據附錄二，當刪除 $\frac{1}{m}$ 之較高級乘幂時。

$$\text{但 } \sum_{r, r+1} \frac{2^{-r}}{r \cdot r+1} = \sum \frac{2^{-r}}{r} - 2 \sum \frac{2^{-r-1}}{r+1} \\ = -\log(1 - \frac{1}{2}) + 2\{\log(1 + \frac{1}{2}) + \frac{1}{2}\} = 1 - \log 2 = \log \left(\frac{e}{2}\right).$$

$$\text{故 } -\log z = m \log \left(\frac{e}{2}\right) - \frac{1}{2} \log(1 - \frac{1}{2}) + \frac{1}{12m}$$

$$z = \left(\frac{2}{e}\right)^m \cdot 2^{-\frac{1}{2}} \cdot e^{-\frac{1}{12m}} = 2^{m-\frac{1}{2}} \cdot e^{-m} \cdot \left(1 - \frac{1}{12m} + \cdots\right)$$

$$\text{將 } \frac{1}{m} \text{ 略去 } = 2^{m-\frac{1}{2}} \cdot e^{-m},$$

惟依瓦立斯氏定理，

$$m! = \frac{(2m)! (2m)^{\frac{1}{2}}}{2^{2m} \cdot m!} \cdot \left(\frac{\pi}{2}\right)^{\frac{1}{2}}, \text{ 至 } \frac{1}{m} \text{ 爲有效,}$$

$$= z \times \frac{(m\pi)^{\frac{1}{2}} \cdot m^m}{2^{m-1}}$$

$$\therefore m! = m^m \cdot \sqrt{2\pi m} \cdot e^{-m} \dots\dots\dots (134)$$

此公式求出對於 $10!$ 值在百分之一以內之差誤，且如 m 增加，則確度亦增高。

如書其全式，乃爲

$$m! = m^m \cdot \sqrt{2\pi m} \cdot e^{-m} + \frac{1}{12m} - \frac{1}{360m^3} + \dots$$

參閱，克里斯圖氏(Chrystal)『代數』第三十章。

四、猶勒麥克老令定理(Euler-Maclaurin) —— 用求和表示積分

設以 $f(a), f(a+h), \dots, f(a+mh)$ 爲對於 $m+1$ 個 x 連續值之 $f(x)$ 值。

於是依泰樂爾(Taylor)氏展開法

$$f(a+h) = f(a) + hf'(a) + \frac{h^2}{2}f''(a) + \dots\dots$$

$$f(a+2h) = f(a+h) + hf'(a+h) + \frac{h^2}{2}f''(a+h) + \dots\dots$$

.....
.....

$$f(a+mh) = f(a + \overline{m-1h}) + hf'(a + \overline{m-1h}) + \frac{h^2}{2}f''(a + \overline{m-1h}) + \dots$$

以 $d = a + \overline{m+1}h$, $b = a + mh$, 並加之。

$$f(b) - f(a) = h \sum_a^d F(x) + \frac{h^2}{2} \cdot \sum_a^d F'(x) + \frac{h^3}{3!} \sum_a^d F''(x) + \dots,$$

式中, $F(x) = f'(x)$, 又 $\int F(x) \cdot dx = f(x) + \text{常數}$ 。

$$\therefore h \sum_a^d F(x) = \int_a^b F(x) \cdot dx - \frac{h^2}{2} \sum_a^d F'(x) - \frac{h^3}{3!} \sum_a^d F''(x) -$$

同理

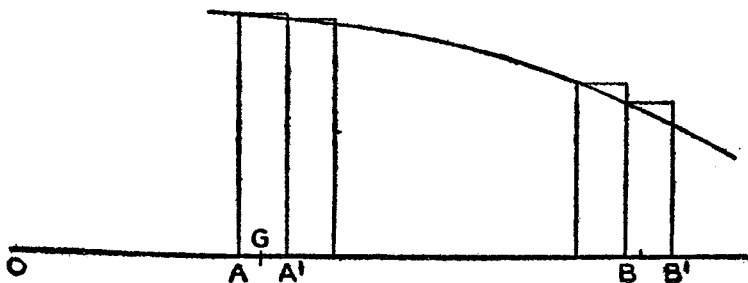
$$h \sum_a^d F'(x) = \int_a^b F'(x) dx - \frac{h^2}{2} \sum_a^d F''(x) - \dots$$

$$h \sum_a^d F''(x) = \int_a^b F''(x) dx - \dots$$

將上數方程式歸併, 則得

$$h \sum_a^d F(x) = \int_a^b F(x) \cdot dx - \frac{h}{2} \{F(b) - F(a)\} + \frac{h^2}{12} \{F'(b) - F'(a)\} \\ + \text{含有 } h^4 \text{ 之各項} \dots\dots (135)$$

$$\therefore h \sum_a^b F(x) = \int_a^b F(x) \cdot dx + \frac{h}{2} \{F(b) + F(a)\} + \frac{h^2}{12} \{F'(b) - F'(a)\} \\ + \text{含有 } h^4 \text{ 之各項} \dots\dots (135)$$



上圖，以 OA 代表 a ， OB 代 b ， AA' 及 BB' 代 h 。

$$AB = mh.$$

$hE(a)$ ， $hF(a)$ 為 AA' ， BB' 上之長方形面積。 $h \sum_a^b F(x)$ 為 AB 上各長方形面積之總和。

$$\int_a^b F(x) \cdot dx \text{ 為 } AB \text{ 上之曲線面積；} \frac{h}{2}[F(a) - F(b)] \text{ 一項，}$$

即為曲線面積與直線面積缺陷部分之第一近似值。

此定理應用於差誤曲線，頗有幾許困難。

$$\text{茲 } F(x) = \frac{1}{2\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}, \text{ 且當 } x \text{ 以 } \sigma \text{ 為比尚不甚大時，必為}$$

一有限垂直之長度。

$$\frac{1}{\sigma} = \frac{1}{\sqrt{pqn}}, \text{ 又 } \therefore \frac{1}{\sqrt{n}}, \text{ 必為垂直之有限數。}$$

橫距離 $AB = mb$ 必須為有限。據第二章第四節例二之分析，得知在 pn 上下之成功數，必屬於 \sqrt{n} 次；故 m 屬於 \sqrt{n} 次，而 h （單位級）屬於 $\frac{1}{\sqrt{n}}$ 次。換言之，造圖時，長方形必作為甚狹，以使許多長方形之和，以與 \sqrt{n} 比，終不得超過一有限之寬度。

依方程式 (135)， h 屬於 $\frac{1}{\sqrt{n}}$ 次， $\sum_a^b F(x)$ 每一有限數均含有 \sqrt{n} 項，故 $h \sum_a^b F(x)$ 屬於 $F(x)$ 次，因此等於 $\int_a^b F(x) \cdot dx$ 。

右邊各項，依次屬於 $\frac{1}{\sqrt{n}}, \frac{1}{n}, \dots$ 各次。(F'(b) 自爲一簡單數字比率)。

現以 h 之值作一，則對於差誤常態曲線（其中屬於 $\frac{1}{\sqrt{n}}$ 次者均刪之），自 $pn+x_1$ 至 $pn+x_2$ 之總成功機率爲

$$\int_{x_1}^{x_2} F(x) \cdot dx \dots\dots\dots (136)$$

更求進一級近似值，得

$$P_x = P_0 \cdot e^{-\frac{x^2}{2\sigma^2}} \cdot \left\{ 1 - \frac{k}{2} \left(\frac{x}{\sigma} - \frac{x^3}{3\sigma^3} \right) \right\},$$

式中， $\frac{1}{\sqrt{n}}$ 各項保留之， $\frac{1}{n}$ 各項刪略之。公式 135 之 h 項亦須保留之。

以自 pn 至 $pn+x$ 成功機率之總合，求出得數，最爲簡便；故於上圖以 A 爲由 G 向左之半個單位，而 $OG = pn$ ，至 G 乃爲曲線重心之橫坐標。然後以 $GB = x$ 。

自 G 至 B 機率總和 = 自 A 至 B $-\frac{1}{2} \cdot P_0$ 之總和

$$= \int_0^x P_x \cdot dx + \frac{1}{2}(P_2 + P_0) - \frac{1}{2}P_0.$$

以 $x = z\sigma$ 。

故自 0 至 $z\sigma$ 機率總和，當將 $\frac{1}{\sigma^2}$ 刪略時，等於

$$\frac{1}{\sqrt{2\pi}} \int_0^z e^{-\frac{1}{2}z^2} \left[1 - \frac{k}{2} \left(z - \frac{1}{3}z^3 \right) \right] dz + \frac{1}{2\sigma\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \dots (137)$$

$$\text{但 } \sigma = \sqrt{pqn}, k = \frac{q-p}{\sqrt{p/n}}.$$

參閱投韓德 (Todhunter) 著, 機率學史, 第九百九十三節。

五、薛伯 (Sheppard) 氏對於頻數曲數動差之修正

(參閱 *Biometrika* 第三卷, 第三百零八頁起)。

設以 $y=f(x)$ 爲一連續頻數曲線之方程式, 曲線面積爲一。

設 A_p 爲以 $x_p \pm \frac{h}{2}$ 爲基線之面積, p 爲整數, 並設不論 p 爲何值, A_p 之一切值, 由觀察均可知之。

由曲線方程算其第 t 次動差, 如 m_t , 乃等於 $\int_a^b x^t \cdot f(x) dx$, 式中 a 與 b 爲 x 之極端值。

如每個面積認爲均集中於各級之中心, 則從觀察求出第 t 級動差, 爲 μ_t , 等於 $\sum_a^b x^t \cdot A_p$ 。

但對於 μ_t 應加以如何之修正, 始能得 m_t , 茲請求之如下:

$$\begin{aligned} A_p &= \int_{x_p - \frac{h}{2}}^{x_p + \frac{h}{2}} f(x) \cdot dx = \int_{-\frac{h}{2}}^{\frac{h}{2}} f(x_p + x) \cdot dx \\ &= \int_{-\frac{h}{2}}^{\frac{h}{2}} \left\{ f(x_p) + x f'(x_p) + \frac{x^2}{2!} f''(x_p) + \dots \right\} \cdot dx \\ &= h f(x_p) + \frac{h^3}{24} f''(x_p) + \frac{h^5}{1920} f^{(4)}(x_p) + \dots \end{aligned}$$

所以

$$\mu_t = \sum_a^b h x_p^t f(x_p) + \sum_a^b \frac{h^3}{24} x_p^t f^2(x_p) + \sum_a^b \frac{h^5}{1920} x_p^t f^4(x_p) + \dots$$

依猶勒麥克老令定理 (公式135)

$$\begin{aligned} \int_a^b x^t f(x) \cdot dx &= \int_a^b x^t f(x) \cdot dx + \frac{h}{2} \{b^t f(b) + a^t f(a)\} + \frac{h^2}{12} [D(x^t f(x))]_a^b \\ &\quad - \frac{h^4}{720} [D^2(x^t f(x))]_a^b + \frac{h^2}{24} \int_a^b x^t f^2(x) \cdot dx + \frac{h^3}{48} \{b^t f^2(b) \\ &\quad + a^t f^2(a)\} + \frac{h^4}{288} [D(x^t f^2(x))]_a^b + \frac{h^4}{1920} \int_a^b x^t f^4(x) \cdot dx + \end{aligned}$$

含有 h^5 之諸項。

茲有一調查，其曲線為零而兩極端與坐標軸接觸，

$$\text{則 } f(a) = 0 = f(b) = f'(a) = f'(b),$$

並設接觸緊密，以致

$$h^2 f^2(a) = h^2 f^2(b) = 0, \text{ 且 } h^4 f^3(a) = h^4 f^3(b) = 0,$$

在此等情形下，又設乘數如 a^t, b^t ，與零並無重要之相差。

$$\begin{aligned} \text{該式變為 } \mu_t &= \int_a^b x^t f(x) \cdot dx + \frac{h^2}{24} \int_a^b x^t f^2(x) dx + \frac{h^4}{1920} \\ &\quad \int_a^b x^t f^4(x) dx + \dots \text{ 含有 } h^5 \text{ 之諸項。} \end{aligned}$$

於是

$$\int_a^b x^t f^2(x) dx = [x^t f'(x)]_a^b - t \int_a^b x^{t-1} f'(x) dx = t(t-1)m_{t-2},$$

用連續局部積分方法，並根據前述條件，則

$$\int_a^b x^t f^4(x) dx = t(t-1)(t-2)(t-3)m_{t-4},$$

因與動差相比 h 普通均為甚小，故涉及 h^3 之項，均可刪略。

$$\therefore \mu_t = m_t \frac{h^2}{24} t(t-1)m_{t-2} + \frac{h^4}{1920} t(t-1)(t-2)(t-3)m_{t-4}$$

此為近似之數。

依次以 t 之值為 1, 2, 3, 4, 則

$$\mu_0 = m_0 = \text{曲線面積} = 1.$$

$$\mu_1 = m_1 = 0 \text{ 如方程式平均數經過垂直線時。}$$

$$\mu_2 = m_2 + \frac{h^2}{12}$$

$$\mu_3 = m_3 + \frac{h^2}{4} m_1 = m_3 \text{ 如 } m_1 = 0.$$

$$\mu_4 = m_4 + \frac{h^2}{2} m_2 + \frac{h^4}{80} m_0 = m_4 + \frac{h^2}{2} \left(\mu_2 - \frac{h^2}{12} \right) + \frac{h^4}{80}$$

$$\therefore m_2 = \mu_2 - \frac{h^2}{12} \dots \dots \dots (138)$$

$$m_4 = \mu_4 - \frac{h^2}{2} \mu_2 + \frac{7h^4}{240} \dots \dots \dots (139)$$

至動差由平均數經過垂直線時， $m_1 = \mu_1$ ， $m_3 = \mu_3$ 。

六、對於普遍的差誤曲線第二近似值之動差及
常數曲線方程爲

$$y = \frac{1}{s\sqrt{2\pi}} e^{-\frac{x^2}{2s^2}} \left\{ 1 - \frac{\kappa}{2} \left(\frac{x}{s} - \frac{1}{3} \cdot \frac{x^3}{s^3} \right) \right\}.$$

以 m_p 代 $\int_{-\infty}^{\infty} \frac{1}{s\sqrt{2\pi}} e^{-\frac{x^2}{2s^2}} \cdot x^p dx$ 。

於是 $m_0 = 1, m_1 = m_3 = \dots = m_{2p+1} = \dots = 0, m_2 = s^2,$
 $m_4 = s^2, m_{2p} = 1 \cdot 3 \cdot 5 \dots (2p-1) \cdot s^{2p}$ (見公式23)。

以 M_p 代第二近似值之第 p 級動。則

於是

$$\begin{aligned} M_p &= \int_{-\infty}^{\infty} \frac{1}{s\sqrt{2\pi}} e^{-\frac{x^2}{2s^2}} \cdot x^p dx - \frac{\kappa}{2s} \int_{-\infty}^{\infty} \frac{1}{s\sqrt{2\pi}} e^{-\frac{x^2}{2s^2}} \cdot x^{p+1} dx \\ &\quad + \frac{\kappa}{6s^3} \int_{-\infty}^{\infty} \frac{1}{s\sqrt{2\pi}} e^{-\frac{x^2}{2s^2}} \cdot x^{p+3} dx \\ &= m_p - \frac{\kappa}{2s} m_{p+1} + \frac{\kappa}{6s^3} m_{p+3}. \end{aligned}$$

$\therefore M_{2p} = m_{2p}$, 因 $m_{2p+1} = 0 = m_{2p+3}$,

故偶數級動差不至因併入 κ 項, 而受影響。

$$M_2 = s^2 \dots \dots \dots (140)$$

$$M_{2p+1} = -\frac{\kappa}{2s} m_{2p+2} + \frac{\kappa}{6s^3} m_{2+4}, \text{ 因 } m_{2p+1} = 0$$

$$M_1 = -\frac{\kappa}{2s} \left(m_2 - \frac{1}{3s^2} m_4 \right) = -\frac{\kappa}{2s} \left(s_2 - \frac{1}{3s^2} \cdot 3s^4 \right) = 0$$

$$M_3 = -\frac{\kappa}{2s} \left(3s^4 - \frac{1}{3s^2} \cdot 15s^6 \right) = \kappa \cdot s^3 \dots \dots \dots (141)$$

$$\begin{aligned} M_{2p+1} &= -\frac{\kappa}{2s} \cdot 1 \cdot 3 \cdot 5 \dots (2p+1) \cdot s^{2p+2} \left\{ 1 - \frac{1}{3s^2} (2p+3)s^2 \right\} \\ &= \frac{p}{3} \cdot 1 \cdot 3 \cdot 5 \dots (2p+1) \cdot s^{2p-2} \cdot M_3 \dots \dots \dots (142) \end{aligned}$$

因 $M_1=0$ ，故原點為曲線之平均數。

欲求衆數 (mode)，須令 $\frac{dy}{dx}$ 等於零。

因 κ^2 為屬於 $\frac{1}{n}$ 次，且在第三章第三節之分析，業將其刪略，

故

$$\begin{aligned} \log(y s \sqrt{2\pi}) &= -\frac{x^2}{2s^2} + \log \left\{ 1 - \frac{\kappa}{2} \left(\frac{x}{s} - \frac{x^3}{3s^3} \right) \right\} = -\frac{x^2}{2s^2} \\ &\quad - \frac{\kappa}{2} \left(\frac{x}{s} - \frac{x^3}{3s^3} \right), \end{aligned}$$

以 $x = -\frac{1}{2}\kappa s$ ，刪略 κ^3 ，故

$$\therefore \frac{\text{衆數向右距平均數之距離}}{s} = \frac{1}{2}\kappa \dots \dots \dots (144)$$

於是在底線 ON 上，曲線之面積 (ON = $x = zs$)，如下得之：

$$Y_{z_0} = \frac{1}{\sqrt{2\pi}} \int_0^z e^{-\frac{1}{2}z^2} \left\{ 1 - \frac{\kappa}{2} \left(z - \frac{z^3}{3} \right) \right\} dz$$

$$= F(z) - \frac{\kappa}{6\sqrt{2\pi}} \{1 - (1-z^2)e^{-\frac{1}{2}z^2}\}$$

$$= F(z) - \kappa f(z),$$

式中 $F(z) = \frac{1}{\sqrt{2\pi}} \int_0^z e^{-\frac{1}{2}z^2} dz$

$$f(z) = \frac{1}{6\sqrt{2\pi}} \{1 - (1-z^2)e^{-\frac{1}{2}z^2}\}。$$

此等函數，在第二章第四節，及第三章第七節所列之表即是。

$$Y^0_{-z} = F(z) + \kappa f(z)$$

且自 $-z$ 至 $+z$ ，即 Y^z_{-z} ，之全部機率，為 $2F(z)$ ，如在常態曲線者然（註一）。

如 M 為中位數位置，則

$$\frac{1}{3} + MO \text{ 上面積} = Y^0_{-\infty}, \frac{1}{3} - MO \text{ 上面積} = Y^{\infty}_0$$

$$\therefore MO \text{ 上面積} 2 \text{ 倍} = Y^{\infty}_0 - Y^0_{-\infty} = \kappa \{f(-\infty) + f(\infty)\} = \frac{\kappa}{3\sqrt{2\pi}}。$$

在底線 OM 上之縱坐標，與 0 之縱坐標，即 $\frac{1}{8\sqrt{2\pi}}$ ，相差只

為含有 κ 之項。

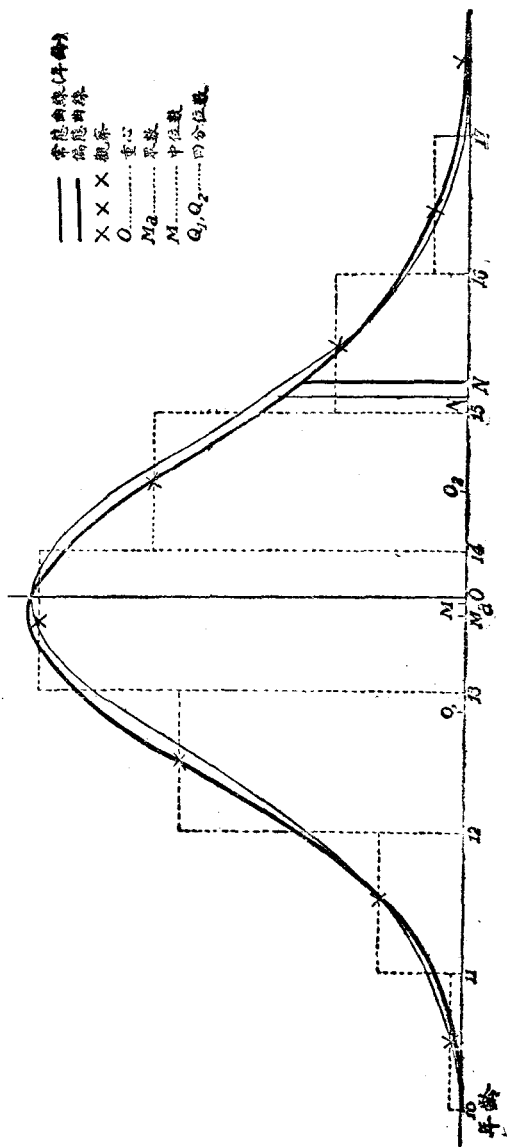
$$\therefore 2MO \times \frac{1}{8\sqrt{2\pi}} = \frac{\kappa}{3\sqrt{2\pi}}, \text{ 當 } \kappa^2 \text{ 可以刪略時。}$$

又 $MO = \frac{1}{6}\kappa s = \frac{1}{3}$ (衆數與平均數之距離) (145)

設底線 MN 上之面積—— M 為中位數， N 為任何一點——

差異偏態曲線舉例

聖路易公立學校第六級學生依年齡之分類



- 常態曲線(假)
- 偏態曲線
- X X 觀察
- O 重心
- Ma 眾數
- M 中位數
- Q₁, Q₂ 四分位數

(第二圖)

等於 ON_1 上常態曲線之面積，即 $F\left(\frac{x_1}{s}\right)$ ，式中 $x_1 = ON_1$ ，並設 $NN_1 = v$ (v 值甚小，且屬於 κ 次，於下式可以看出)。

$$ON = x_1 - v。$$

則 $F\left(\frac{x_1}{s}\right) = MO$ 上面積 + $(x_1 - v)$ 上面積

$$\begin{aligned} &= \frac{\kappa}{6\sqrt{2\pi}} + \frac{1}{s\sqrt{2\pi}} \int_0^{x_1 - v} e^{-\frac{x^2}{2s^2}} dx \\ &\quad - \frac{\kappa}{6\sqrt{2\pi}} \left\{ 1 - \left(1 - \frac{(x_1 - v)^2}{s^2} \right) e^{-\frac{(x_1 - v)^2}{2s^2}} \right\} \\ &= \frac{\kappa}{6\sqrt{2\pi}} + F\left(\frac{x_1}{s}\right) - \frac{v}{s\sqrt{2\pi}} e^{-\frac{x_1^2}{2s^2}} - \frac{\kappa}{6\sqrt{2\pi}} + \frac{\kappa}{6\sqrt{2\pi}} \left(1 - \frac{x_1^2}{s^2} \right) e^{-\frac{x_1^2}{2s^2}}, \end{aligned}$$

式中，具有 v^2 及 $v\kappa$ 各項，業已刪略矣。

$$\therefore v = \frac{\kappa s}{6} \left(1 - \frac{x_1^2}{s^2} \right)。$$

如吾人能得知，橫軸上從最低至三種位置之相對觀察數，且能斷定頻數曲線方程式，即為上文所論之式時，則平均數及 s 與 κ ，不難求得。此法可以數字例釋之。第三章第七節第五例之觀察相對數如下：

年齡極限	學童人數
0 至 13 歲	3044 之 .296
0 至 15 歲	3044 之 .867

0 至16歲

3044之·969

設 m 為年齡之中位數， s 歲為標準差， $\kappa s^3 =$ 第三動差：全為未知數。

在上列數字表中，設 M 代表其年齡中位數， N 代表十五歲。

則 MN 上面積，為 $.867 - .500 = F(1.112)$ ，（見第二章常態機率表）。

$$\text{故 } ON_1 = 1.112 = \frac{x_1}{s} = z_1$$

$$15 - m = MN = MO + ON_1 - NN_1 = \frac{1}{6}\kappa s + x_1 - \frac{1}{6}\kappa s \left(1 - \frac{x_1^2}{s^2}\right)$$

$$15 - m = z_1 s + \frac{1}{6}\kappa s z_1^2 \quad (\text{式中 } z_1 = 1.112)。$$

同理

$$16 - m = z_2 s + \frac{1}{6}\kappa s z_2^2, \quad (\text{式中 } z_2 = 1.866)$$

$$\text{又 } m - 13 = z_3 s - \frac{1}{6}\kappa s z_3^2, \quad (\text{式中 } F(z_3) = .204, \text{ 而 } z_3 = .536)。$$

N 如在 M 之左方時，必須取作負號，略經思索，便知其當然。

欲決定 m ， s ，及 κ 之值，現共有三方程式。

$$\frac{1}{6}\kappa s(z_1 - z_3) + s = \frac{2}{z_1 + z_3}$$

$$\frac{1}{6}\kappa s(z_2 - z_3) + s = \frac{3}{z_2 + z_3}$$

$$\therefore \kappa s = .278 \quad s = 1.187 \quad \kappa = .234 \quad m = 13.623$$

$$\text{平均數} = m + \frac{1}{6}\kappa s = 13.669。$$

(請參閱統計學報一九〇二年號,自339至348頁)

由九級全部算得之動差上,可求出 $s=1.190, =.206$, 且平均數 $=13.665$ 。

在已知平均數或中位數,或曲線已知為常態,而 $\kappa=0$ 時,兩種觀察值已足決定其他各值矣。

惟吾人須注意,代表第一近似值之曲線,與代表第二近似值者,必在 $x=s\sqrt{3}$ 處交叉。

MN上偏態曲線之面積 = ON 上當 x 等於 $\pm s$ 時,常態曲線之面積。

偏態曲線,在 ON 之右,超過常態曲線之距離 = 在 ON 之左,所短缺之距離。

七、未加權平均數之比率

設 $M_1, M_2 \dots M_n$ 為某一時間內 n 個數量之真值,而以 $M_1', M_2' \dots$ 為另外一時間內之值。

設 $n\bar{m} = \sum M_t, M_t = \bar{m} + m_t, \sum m_t = 0, n\bar{m}' = \sum M_t', M_t' = \bar{m}' + m_t', \sum m_t' = 0, n\sigma_m^2 = \sum m_t^2, n\sigma_{m'}^2 = \sum m_t'^2$ 。

設 $\bar{m}' = m(1 + \rho), M_t' = (1 + \bar{u} + u_t)M_t$ 式中 $\sum u_t = 0$,

$$\therefore \rho = \bar{u} + \frac{\sum m_t u_t}{n\bar{m}}.$$

\bar{u} 爲上述數量增加率之平均值, ρ 爲數量平均值之增加率。假如二者之較大者, 增加率不大, 或較小之量, 增加率較另一數量爲速, 則 \bar{u} , 與 ρ 漸趨相等。

假設此等數量因有錯誤, 致成 $M_t(1+e_t)$ 及 $M'_t(1+e'_t)$ ……等數。則依公式70之規定, m 及 m' 中差誤之標準差, 爲 $\frac{\sigma}{\sqrt{n}}\sqrt{\left(1+\frac{\sigma m^2}{\bar{m}^2}\right)}$ 及 $\frac{\sigma_1}{\sqrt{n}}\sqrt{\left(1+\frac{\sigma m'^2}{\bar{m}'^2}\right)}$ ——式中之 σ 與 σ_1 , 分別表示 e_t 與 e'_t 。

如兩組中之差誤, 彼此係屬獨立, 則依公式63之規定,

$$S_r^2 = \frac{1}{n} \left\{ \sigma^2 \left(1 + \frac{\sigma m^2}{\bar{m}^2} \right) + \sigma_1^2 \left(1 + \frac{\sigma m'^2}{\bar{m}'^2} \right) \right\},$$

式中之 S_r , 爲 $\frac{\bar{m}'}{\bar{m}}$ 即 $1+\rho$ 中差誤之標準差。

但 M'_t 各量數之差誤 e'_t , 與 M_t 各量數之差誤 e_t , 相差常屬無幾, 且正負號相同。

$$\text{以 } d_t = e'_t - e_t.$$

則由平均數之比率所生之差誤, 爲

$$\begin{aligned} \frac{\bar{m}'}{\bar{m}}(1+e) &= \frac{S\{M'_t(1+e'_t)\}}{S\{M_t(1+e_t)\}} \\ &= \frac{S\{M'_t(1+e'_t)\} \cdot SM_t - S\{M_t(1+e_t)\} \cdot SM'_t}{S\{M_t(1+e_t)\} \cdot SM'_t} \\ &= \frac{\bar{m}S(M'_te'_t) - \bar{m}' \cdot S(M_t e_t)}{\bar{m}' S\{M_t(1+e_t)\}} \quad \text{將 } e_t^2 \text{ 與 } e_t e'_t \text{ 略去,} \end{aligned}$$

$$\text{則} \quad = \frac{\bar{m}S(M_t'd_t) + S\{(\bar{m}M_t' - \bar{m}'M_t)e_t\}}{n\bar{m}\bar{m}'}$$

如將 $\bar{u} - \rho$ 刪略，

$$= \frac{S(M_t'd_t)}{n\bar{m}'} + \frac{S\{(u_t + \bar{u} - \rho)M_t'e_t\}}{n\bar{m}'} = \frac{S(M_t'd_t)}{n\bar{m}'} + \frac{SM_t'u_t e_t}{n\bar{m}'},$$

是故，如 S_r 為 e 之標準差， σ_d 為 d_t 之標準差， σ 為 e_t 之標準差，或為於非由一致之頻數曲線而來此等數值之加權標準差。

$$S_r^2 = \frac{1}{n^2\bar{m}'^2} \sigma_d^2 \cdot S(M_t'^2) + \frac{1}{n^2\bar{m}'^2} \cdot \sigma^2 \cdot S(M_t'^2 u_t^2), \text{ (見公式55)}$$

$$S(M_t')^2 = S(\bar{m}' + m_t')^2 = n(\bar{m}'^2 + \sigma_m'^2),$$

$$S(M_t'^2 u_t^2) = n\bar{m}'^2 \sigma_u^2 + n\sigma_m'^2 \sigma_u^2 + S u_t^2 (m_t'^2 - \sigma_m'^2) + 2\bar{m}'$$

$$S(m_t' u_t^2),$$

式中 $n\sigma_u^2 = S u_t^2$

$$= n\sigma_u^2 (\bar{m}'^2 + \sigma_m'^2) + \text{可以刪略之諸項。}$$

$$\therefore S_r^2 = \frac{1}{n} \sigma_d^2 \left(1 + \frac{\sigma_m'^2}{\bar{m}'^2}\right) + \frac{1}{n} \cdot \sigma^2 \cdot \left(1 + \frac{\sigma_m'^2}{\bar{m}'^2}\right) \cdot \sigma_u^2 \cdot \frac{1}{(1-\rho)^2}$$

..... (146)

如 e_t 與 e_t' 彼此係互為獨立， σ_d^2 必等於 $\sigma^2 + \sigma_1^2$ ，但如 $e_t = e_t'$ ，
..... 餘類推， σ_d 將為零。故 σ_d 可謂為在 0 與 $\sigma\sqrt{2}$ 之間。

第二項之大小，視 σ_u 而定，蓋 σ_u 表示各種數量增加率之變動且可由觀察而求得也。

故兩時期之觀察數量，如各具有類似之差誤，而此兩種數量之增加率，又相差無幾，則所算平均數之比率，其差誤必然甚小—— n 若大時，則尤小矣。

八、加權平均數之比率

論至加權平均數，其公式愈趨複雜。

設 $W_t = \bar{w}_t + w_t$, $W_t' = \bar{w}' + w_t'$ 為兩時期之兩個權數，而 \bar{w} , \bar{w}' 為權數之平均數。以 $n\sigma_w^2 = S w_t^2$, $n\sigma_w'^2 = S w_t'^2$ 。

設以 $W_t' = W_t(1 + \bar{v} + v_t)$ ，式中 $S v_t = 0$ ，並以 $n\sigma_v^2 = S v_t^2$ 。

假設 W_t, W_t' 之值，訛為 $W_t(1 + \eta_t)$ 及 $W_t'(1 + \eta_t')$ ，再以 σ' 為 η_t 之標準差。其他字母代表，與附錄七同。

$$\text{設 } \bar{m}_w = \frac{S(W_t M_t)}{S W_t}, \bar{m}'_w = \frac{S(W_t' M_t')}{S W_t'}$$

求 $\frac{\bar{m}'_w}{\bar{m}_w}$ 之差誤， e 。

$$\frac{\bar{m}'_w}{\bar{m}_w}(1 + e) = \frac{S\{W_t'(1 + \eta_t')M_t'(1 + e_t')\}}{S\{W_t(1 + \eta_t)M_t(1 + e_t)\}} \cdot \frac{S\{W_t(1 + \eta_t)\}}{S\{W_t'(1 + \eta_t')\}}$$

由此觀之，如再加以推演，而將 η_t, e_t 之乘積及平方取銷，則

$$e = \frac{S(W_t' M_t' e_t')}{n \bar{w}' \bar{m}'_w} - \frac{S(W_t M_t e_t)}{n \bar{w} \bar{m}_w} + \frac{S\{W_t(\bar{m}_w - M_t)\eta_t\}}{n \bar{w} \bar{m}_w} - \frac{S\{W_t'(\bar{m}'_w - M_t')\eta_t'\}}{n \bar{w}' \bar{m}'_w} \dots \dots \dots (147)$$

欲求近似值，只須將各乘積所有之總和取消，蓋某一因子之總和，其值為零也。由此乃得 $\bar{m}_w = m$, $\bar{m}_w' = \bar{m}'$, $\bar{u} = \rho$, $\bar{u}' = (1 + \bar{v})$ ，隨推演之進步，愈可化簡矣。

以 $d_t' = \eta_t' - \eta_t$ ，以 σ_d' 為其標準差。

$$\begin{aligned} \text{於是 } e &= \frac{S(W_t' M_t' d_t)}{n \bar{w}' \bar{m}'} + \frac{S(W_t' M_t' \eta_t)}{n \bar{w}' \bar{m}'} e_t + \frac{S(W_t M_t v_t)}{n \bar{w}' \bar{m}} e_t \\ &\quad + \frac{S(W_t' m_t' d_t')}{n \bar{w}' \bar{m}} + \frac{S(W_t' M_t' \eta_t)}{n \bar{w}' \bar{m}'} \eta_t + \frac{S(W_t m_t v_t)}{n \bar{w}' \bar{m}} \eta_t. \end{aligned}$$

所以，如用約略數，

$$\begin{aligned} S_r^2 &= \frac{1}{n} \left(1 + \frac{\sigma_w'^2}{\bar{u}'^2}\right) \left(1 + \frac{\sigma_m'^2}{\bar{m}'^2}\right) \sigma_d^2 + \frac{1}{n} \left(1 + \frac{\sigma_w'^2}{\bar{u}'^2}\right) \left(\frac{\sigma_m'^2}{\bar{m}'^2}\right) \sigma_d^2 \\ &\quad + \frac{1}{n} \left(\frac{\sigma_u}{1 + \bar{u}} + 1\right)^2 \left(\frac{\sigma_w'^2}{\bar{u}'^2}\right) \left(1 + \frac{\sigma_m^2}{\bar{m}^2}\right) (\sigma^2 + \sigma'^2) \\ &\quad + \frac{1}{n} \left(\frac{\sigma_v}{1 + \bar{v}}\right)^2 \left(1 + \frac{\sigma_w'^2}{\bar{u}'^2}\right) \left\{ \left(1 + \frac{\sigma_m^2}{\bar{m}^2}\right) \sigma^2 + \frac{\sigma_m^2}{\bar{m}^2} \sigma'^2 \right\} \quad (148) \end{aligned}$$

涉及 σ^2, σ_d^2 —— 測量 M_t 數量之差誤者 —— 之項，與平均數未加權者相似，但有一因子（大於 1 而普通小於 2）涉及權數，且有一項涉及一小因子 σ_v^2 （測量權數變動之離散度者），則為例外也。

於涉及 $\sigma'^2, \sigma_d'^2$ （測量權數之差誤者）之三項中，第一項包含因子 $\left(\frac{\sigma_m'}{\bar{m}'}\right)^2$ ，而第三項包括因子 $\left(\frac{\sigma_m}{\bar{m}}\right)^2$ ，此兩因子隨 M 數量之

離散度而變，離散輕者，該兩因子必小；至第二項，則含有 $\left(\frac{\sigma_u}{1+u}\right)^2$ ，而此因子，當數量增加率幾近相等時，其值必小。

σ , σ' , σ_d , σ_d' 係數之實在值，可於觀察中得之，而其相對重要性 (relative importance) 亦於以查出；但於離散不劇之數量，增加率與相等不至過甚時，則權數之差誤，與數量之等量差誤相比，必不甚大。

在此情形之下，第一近似值乃得

$$S_r = \frac{\sigma_d}{\sqrt{n}} \sqrt{1 + \left(\frac{\sigma_w}{w}\right)^2} \dots\dots\dots (149)$$

惟若 $\frac{\sigma_u}{1+u}$ 不為甚小，可得更近之近似值

$$S_r = \frac{1}{\sqrt{n}} \cdot \sqrt{\left[1 + \left(\frac{\sigma_w}{w}\right)^2\right] \left\{ \sigma_d^2 + \left(\frac{\sigma_u}{1+u}\right)^2 (\sigma^2 + \sigma'^2) \right\}} \dots (150)$$

σ_d ——測量數差誤相差量者——雖然多半小於 $\sqrt{2}\sigma$ ，然與差誤之一相比，其值已甚大。

故最好須於未刪除某項時，即由觀察中概略測驗其係數；並於看出所刪略之乘積並不甚小時，或任一差誤，或許特別甚大時，應以用公式147之完全式為佳。

(參閱統計學報 1911-12 年，第八一至八八頁，『平均數確度之測量』一文)。

九、動差……等中差誤之標準差之常態性

[此則係根據薛伯氏著：差誤理論之應用 (Application of the theory of Error) 一文，見皇家學會會報，第一九二卷，一八九八年版，A二二九，第一一七至一二八頁；但標號及論述，則已加變更]。

在一『宇宙』中，共有事物 N 個， x_1 處有 p_1N 個， x_2 處有 p_2N 個，其他以此類推， $p_1 + p_2 + \dots = 1$ ， $F = a_1p_1 + a_2p_2 + \dots$ 而 a_1, a_2, \dots 乃為常數。

茲就中抽出 n 個事物來， x_1 處有 n_1 個， x_2 處有 n_2 個，餘類推， $n_1 + n_2 + \dots = n$ 。

$$\text{以 } F + f = a_1 \frac{n_1}{n} + a_2 \frac{n_2}{n} + \dots$$

$$f = a_1 \frac{n_1}{n} + a_2 \frac{n_2}{n} + \dots - F \left(\frac{n_1}{n} + \frac{n_2}{n} + \dots \right) = b_1 \frac{n_1}{n} + b_2 \frac{n_2}{n} + \dots,$$

式中 $b_1 = a_1 - F$ ，其他以此類推。

$$\text{於是 } Sb_i p_i = Sa_i p_i - F \cdot Sp_i = F - F = 0$$

$$Sb_i^2 p_i = Sa_i^2 p_i - 2FSa_i p_i + F^2 \cdot Sp_i = Sa_i^2 p_i - F^2。$$

現請求 M_s (= 平均值 f^s)，並證明其對於 M_2 (= 平均值 f^2) 之關係，乃即於差誤常態曲線中所發生之觀象。

$$\text{將此式 } E = \left(p_1 e^{b_1 \frac{a}{n}}\right)^{n_1} \left(p_2 e^{b_2 \frac{a}{n}}\right)^{n_2} \dots\dots,$$

用多項式定理展開，則得各項之和：

$$\frac{n!}{n_1! n_2! \dots\dots} \left(p_1 e^{b_1 \frac{a}{n}}\right)^{n_1} \left(p_2 e^{b_2 \frac{a}{n}}\right)^{n_2} = P.e.f.a \text{ 各項之和,}$$

但此係以下為條件： $n_1 + n_2 + \dots = n$ ，且式中

$$f = b_1 \frac{n_1}{n} + b_2 \frac{n_2}{n} + \dots\dots,$$

$$P = \frac{n!}{n_1! n_2! \dots\dots} p_1^{n_1} p_2^{n_2} \dots\dots$$

P 乃為於 x_1 抽出 n_1 ，於 x_2 抽出 n_2 ……之全部機率，將下多項式展開即明：

$$(p_1 + p_2 + \dots\dots)^n$$

$$\therefore E = P \left(1 + af + \frac{a^2}{2} f^2 + \dots\dots + \frac{a^s}{s!} f^s + \dots\dots\right) \text{之和}$$

$$= M_0 + aM_1 + \frac{a^2}{2} M_2 + \dots\dots + \frac{a^s}{s!} M^s + \dots\dots$$

$$\text{又 } E = \left(Sp_t + \frac{a}{n} Sb_t p_t + \frac{a^2}{2n^2} Sb_t^2 p_t + \frac{a^3}{6n^3} C_3 + \frac{a^4}{24n^4} C_4 + \dots\dots\right)^n$$

將 $e^{b_i \frac{a}{n}}$ ……各項展開必如此，但

$$C_3 = Sb_t^3 p_t, C_4 = Sb_t^4 p_t \dots\dots, Sp_t = 1, Sb_t p_t = 0。$$

$$\therefore E = \left(1 + \frac{a^2}{2n^2} Sb_t^2 p_t + \dots\dots\right)^n。$$

求 E 之算式中，令爲首三個係數相等。

$$M_0 = 1, M_1 = 0, M_2 = \frac{1}{n} S b_t^2 p_t.$$

$$\therefore 1 + \frac{a^2}{2} M_2 + \dots + \frac{a^s}{s!} M_s + \dots = \left(1 + \frac{a^2}{2n} M_2 + \frac{a^3}{6n^3} C_3 + \dots \right)$$

現知， n 若大， $M_2, \frac{C_3}{n^{\frac{3}{2}}}, \frac{C_4}{n^2}, \frac{C_5}{n^{\frac{5}{2}}}, \dots$ 必爲有限數，故如將 $\frac{1}{\sqrt{n}}$

刪略，則得

$$\begin{aligned} 1 + \frac{a^2}{2} M_2 + \dots + \frac{a^s}{s!} M_s + \dots &= \left(1 + \frac{a^2}{2n} M_2 \right)^n = e^{\frac{a^2}{2} M_2} \\ &= 1 + \frac{a^2}{2} M_2 + \dots + \frac{a^{2t}}{t! 2^t} (M_2)^t + \dots \end{aligned}$$

故，在此情形下， s 若爲奇數， $M_s = 0$ ， $M_{2t} = \frac{(2t)!}{t! 2^t} (M_2)^t$ ，如

在差誤常態曲線中然。

此條件—— $\frac{C_3}{n^{\frac{3}{2}}}, \dots$ 爲有限數——與第三章第四節愛基華斯

氏之證明，原理相同，但又不可一概而論，該項定理應用，應就每一情形，而加以討論。

例如第九章第六節所論， $f = x_1^2 e_1 + x_2^2 e_2 + \dots$ (式中之 e_1 爲 $\frac{n_1}{n} - p_1$)， $F = \mu_2$ (樣本所自抽出之『宇宙』中第二動差)，及

$$b_t = x_t^2 - \mu_2.$$

$$M_2 = \frac{1}{n} S(x_i^2 - \mu_2)^2 p_i = \frac{1}{n} (\mu_4 - \mu_2^2)$$

$$C_3 = S(x_i^3 - \mu_3)^3 p_i = (\mu_6 - 3\mu_4\mu_2 + 2\mu_2^3)$$

$$\frac{C_3}{n^{\frac{3}{2}}} = M_2^{\frac{3}{2}} \cdot \frac{\mu_6 - 3\mu_4\mu_2 + 2\mu_2^3}{(\mu_4 - \mu_2^2)^{\frac{3}{2}}}$$

同理

$$\frac{C}{n^2} = (M_2)^2 \frac{\mu_8 - 4\mu_6\mu_2 + 6\mu_4\mu_2^2 - 3\mu_2^4}{(\mu_4 - \mu_2^2)^2} \dots\dots$$

由此推論， $\frac{\mu_4}{\sigma^4}, \frac{\mu_6}{\sigma^6}, \frac{\mu_8}{\sigma^8} \dots\dots$ 各比率若為有限數（而 $\sigma^2 = \mu_2$ ），

則 $\frac{C_3}{n^{\frac{3}{2}}}, \frac{C_4}{n^2} \dots\dots$ 為有限數，果不出所料。

是以，『宇宙』之頻數曲線，若與此等條件適合，在平均數左右果有合理之集中現象，而在超過 σ 之小量倍數以外，又無偏重情形，則對於第二動差——亦即標準差——之差誤，頻數曲線，必為常態形狀。

類似而較為簡短之證明，更可示吾人以平均數之差誤，亦具有常態之頻數（ $f = x_1 e_1 \dots\dots, F = 0, b_i = x_i$ ）。

就相關係數（見第九章第七節）之分析而論，

$$b_i = r \left(\frac{x_i y_i}{M} - \frac{x_i^2}{2\lambda} - \frac{y_i^2}{2\mu} - 1 + \frac{1}{2} + \frac{1}{2} \right), p_i = z_i$$

$$M_2 = \frac{1}{n} S r^2 \left(\frac{x_i y_i}{M} - \frac{x_i^2}{2\lambda} - \frac{y_i^2}{2\mu} \right)^2 z_i$$

$$= \frac{r^2}{n} \left(\frac{M_{22}}{M^2} + \frac{\lambda_4}{4\lambda} + \frac{\mu_4}{4\mu^2} - \frac{M_{13}}{M\lambda} - \frac{M_{13}}{M\mu} + \frac{M_{22}}{4\lambda\mu} \right)$$

$$C_s = S_r^3 \left(\frac{x_i y_t}{M} - \frac{x_i^2}{2\lambda} - \frac{y_t^2}{2\mu} \right) z_t = r^3 \cdot \left(\frac{M_{33}}{M^3} + \frac{\lambda_6}{8\lambda^3} + \dots \right).$$

以 $\lambda = \sigma_1^2, \mu = \sigma_2^2$, 則 $\frac{\lambda t}{\sigma_1^2}, \frac{\mu t}{\sigma_2^2}, \frac{M_{st}}{\sigma_1^s \sigma_2^t}$ 若為有限數, 不論 s 及 t 為

任何值, $\frac{C_s}{n^{\frac{s}{2}}}$ 必等於 $M_2^{\frac{s}{2}} \times$ 有限數, 且依此辦法, 更高級之項, 亦可

同樣處置也。

故如二次元 (two-dimensional) 頻數分配之動差及乘積, 果與前述之條件相合, 則相關係數之差誤曲線, 必為常態形式。

十、最小二乘法

此法之由來甚久, 於若干不確切之量數中, 指示所抽出之多個數值時, 用之。

設一數量 z , 因方程式 $z = u_1 x_1 + u_2 x_2 + \dots + u_k x_k$, (式中, u_1, u_2, \dots 為可以由觀察得來之數量) 而與 k 個未知之常數, x_1, x_2, \dots, x_k , 有關; 並設由 n 組觀察可得

$$1^s = 1u_1 x_1 + 1u_2 x_2 + \dots + 1u_k x_k$$

.....

$$n^s = nu_1 x_1 + nu_2 x_2 + \dots + nu_k x_k$$

式中之 z 及 u , 均為已知數。

n 若等於 k , x 各值即可確實求出。若 $n < k$, 則解答之式將有無限之多, 而方程式乃無從決定。

若 $n > k$, 則方程式必有矛盾, 於是問題乃變為如何決定 x_1, x_2, \dots 各值之問題, 蓋必決定 x_1, x_2, \dots 之值, 以使矛盾性縮為最小, 至矛盾之來, 則假定其由於測量 u 時, 量數之不完全也。

以 d_1, d_2, \dots 代表 $1z, 2z, \dots$ 與由 x_1, x_2, \dots 真值所得之值, 即 X_1, X_2, \dots , 二者間之相差額。

$$\text{則 } 1u_1X_1 + 1u_2X_2 + \dots + 1u_kX_k - 1z = d_1$$

$$2u_1X_1 + 2u_2X_2 + \dots + 2u_kX_k - 2z = d_2$$

.....

假定 d_1, d_2, \dots 即為差誤, 而此差誤之機率, 可由一常態曲

線 $P = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{d^2}{2\sigma^2}}$ 得來。此假定則多係根據理證 (demonstrations)

而得, 蓋在某一關於偶然差誤 (accidental error) 之假定下, 此常態形式可應運而出也。不論對於物體或測地 (geodetical) 之量數, 假定之效度如何, 均不便承認此等假定對於統計或生物測量能以適用, 至為對平均數之離中差量, 抑或由於抽樣方法之差誤, 則概所不問也。

欲得解答, 須先求 X_1, X_2, \dots 之各值, 由此各值可使 d_1, d_2, \dots 全體出現之機率為最大, 換言之, 即使

$$d_1^2 + d_2^2 + \dots$$

之和為最小。茲以 $f(d_1, d_2, \dots)$ 代表此總和。

成為最小之條件，為 $\frac{\partial f}{\partial X_1} = 0 = \frac{\partial f}{\partial X_2} = \dots$

由此得

$$\frac{1}{2} \cdot \frac{\partial f}{\partial X_1} = \sum_{t=1}^{t=n} u_1 (t u_1 X_1 + t u_2 X_2 + \dots - t z) = 0$$

$$\frac{1}{2} \cdot \frac{\partial f}{\partial X_2} = \sum_{t=1}^{t=n} u_2 (t u_1 X_2 + t u_2 X_2 + \dots - t z) = 0$$

.....

書如下式亦可：

$$X_1 \cdot \sum u_1^2 + X_2 \cdot \sum u_1 u_2 + \dots = \sum u_1 z$$

$$X_1 \sum u_1 u_2 + X_2 \cdot \sum u_2^2 + \dots = \sum u_2 z$$

.....

$$X_1 \sum u_1 u_k + X_2 \sum u_2 u_k + \dots = \sum u_k z,$$

以上共有 k 個方程式，由此可求出 k 個量數 X_1, X_2, \dots

此種方法於實用上，頗能求出大致甚佳之未知數之經驗值。其簡單形式，於第一編第十章第二節最後一法，曾一用之，而此法之效度，可以測驗者，本編第六章第七節曾採用之。

〔參閱麥理曼(Merriman)著：最小二乘法(Method of Least

Squares), 及外爾德(Weld)著: 差誤理論及最小二乘方(Theory of Errors and Least Squares)。

(註一) 應用猶勒麥克老定理, 依照 p, q, n , 之假設, 其相當公式, 爲

$2F(z) + \frac{e^{-\frac{1}{2}z^2}}{S\sqrt{2\pi}}$; 但若於資料爲連續時, 其最末一項, 必消滅。