裘爾倍氏

統計研究法

李仲珩譯

世界書局印行

原 序

自1911年G. U. Yule之統計學概論"An Introduction To The Theory Of Statistics"(1)出版,遂引起余編著是書之決心,Yule氏博覽羣書,經驗宏富,加以擇材精審,方法新穎,故其著作不僅適於統計問題之研究,且於此日形擴大之範圍中,使讀者對於統計方法得一鳥瞰,洵統計界之寶筏也!其書自1911年初版後,截至1920年,翻印至四次,而大戰期內(1915—1919)四年之間,售出三版,亦足見英人研究統計學之努力及興趣矣.至其主因大故,則F. Klein在歌廷根『應用數理協進會』廿週年(1918)紀念演講詞中(2),言之至審,固無庸贅言也.

統計目的,既在致用,本書編撰,亦即以此爲前提;故除取材於Yule之原著外,尤注意旁搜廣引,舉凡現今能應用統計之科學,莫不稱其實例,其尚未應用者,亦期化爲能用,總之能用統計之學科愈多愈妙,而所舉實例亦求切而且眞,斯則著者之初意云爾.至於本書所用算法,雖以求和法爲多,然其他諸法,亦莫不具備.

應用差誤論於統計研究,自不無可議之處,而尤以應用由概算推出之差誤論於農業森林計算上爲尤甚.

本書說理,不厭詳盡,惟求讀者,易於明瞭,否則以應用統計學科之衆,自不必人人能讀,又非著者之原意矣!

一九二〇年四月著者敍於奥國之鹽城(Salzburg).

---

(1) London, Charles Griffin and Company. 出版

(2) Jahresbericht der Deutschen Mathematiker—Vereinigung, 德國數學協會 27. Band, P. 217, 及 223.

後序

十八年春,余承乏中央大學,值吾師周君適先生有統計專班之設,命遂譯 E. Czuber 之「統計研究法,」以作講義.時國中統計應用,雖已萌芽,對於理論,則未注意.而先生獨取楚伯此著,以爲圭臬者,以其舉例精而備,論理博而洽,誦之,足窺統計學理之堂奧也.惜余以應留學試,倉卒返湘,譯途中輟.在湘謁吾師方彼川先生,先生亦以統計書籍缺乏,須亟編譯爲囑.抵德後,復得周師來書催促,余始於功課之暇,因巴陽國家圖書館之便,得取原著參考書籍擇條翻覽,其爲法意諸文者,則煩吾友 Erika Perron 詳爲轉譯,務期實事求是,譯毋違意;並取原著錯誤,是正二三十處,其意有未顯,詞嫌簡略者,則刪去原文,而改作之.歷時二載,始成此帙.雖然,心疏筆誤,失檢鑄錯,情理之常,聖賢難免,吾師 Oskar Perron 嘗謂「錯誤爲人類之通病,」以彼淵博,尚復云然,嗟余小子,其能無僭?海內碩宿,若發其誤,而舉以告者,則不佞將再拜而受其賜!

至於擬在國內所譯諸章,則以典籍缺乏,德文生疏,全賴

周師口講意授,發凡糾謬.他如譯名,則與徐國英女士亦多所商榷.特誌於斯,藉表謝意.

民國二十年四月譯者序於巴陽明與大學

目　次

導　言

第一編　定標論

第一章　標準之符號及配合

第二章　標準之性

第二章　中值

第五章　兩標相關性應用

第六章　相關係數之應用

第七章 多標相關性

導　言

1. 統計學者,專論事物之外相而不及其內容,與物理化學生理之研究方法適相反者也.其結果,可度,可數,可衡,可以數字表之,故有量的性質.

統計學中,發生最早之問題,爲求全體中能合一定標準或一羣標準事物之箇數,此必待數計而後知者,固有量的意義,然若論其是否合此標準,則同時亦含性質之意義矣.譬如僅稱量之大小輕重,而不定其確數,僅言色之深淺紅綠,而不窮其物性,則不獨含有性質意義,直謂之專言性質,亦無不可.

學者之研究一物也,或叩其爲彼爲此,或審其大小輕重,前者僅須觀察即能判然,後者必加度衡始生差異,蓋所謂量者,雖物體本身所固有,然必藉度衡始能顯示之.

職此之由,統計方法遂分二種:一曰同度統計[1],先立固定標準,次將材料之合此標準者,列爲一類,然後數計類元,加以統計如調查人口之分宗教性別國籍是.二曰異度統

(1) C. V. L. Charlier, Statistisk Tidskrift 1010 (統計期刊瑞典)

計,先立能縮能伸之標準,次定材料之度量,爲之分類而統計之,如殖民調查之注意年齡分佈是.

2. 至於所謂事物,雖係指同類者言,然其同樣度仍有差異,標準之符合者愈多,其同樣度亦愈大,故同樣度者,隨標準之多寡而殊,非有一定不移之性質也.取同樣度,惟求適合問題之研究即爲已足,過大或過小,均足使所得結果,失去科學意義.

欲研究人類體格,必須注意其因年齡性別而生之發育差異,若不分老少男女而統計之,其結論必至毫無意義.故欲得一科學上有價值之結果,則須取性別年齡血統相同之人羣,方能奏效.至於遺傳研究,則不獨須取同類之植物,且須取同母之植物,再依一定目標,加以選擇,其所需同樣度之大如是.

供統計研究,同類事物之集體,名曰集團,其分子曰個體,或簡稱元,亦曰樣元.個體之總數,曰集團之範圍.

組成集團之個體,最常見者爲天然物.而尤以實體爲最,如人類,鳥獸,植物,及其各種器官是.其他若藝術,實業,職業等,亦可作集團之研究.至於抽象事物之能組成集團者,則有自然現象,文化事業,及自物理,化學,生理,醫學,農業等試驗所得之結果,與夫同量之屢測,各時各地之貨價及工資等.總之已組成集團及能組成集團之事物,不勝枚舉.集團

意義之廣狹蓋無垠矣.

欲使一集團之統計結果,易於領悟,且有科學之價值,則必須加以說明,然後讀者對於統計之結果,方可一目瞭然.如研究體格,若標題僅為「體格」,則必嫌不足,如改為「成年男子之體格」,則較前為美,倘標明為「自二十歲至三十歲男子之體格」,則精審有加,若更能別其血統,定其年齡,分其性別,則尤善矣.

集團之意義愈廣,則應用統計之學科亦愈多,舉其犖犖大者,如人類學,動物學,植物學,生理學,遺傳學,物理學,醫學,化學,農業,森林,經濟,保險,及實驗心理等,皆須統計以整理其材料,擴大其範圍,故凡一切由經驗構成之科學,莫不有統計之致用.且其致用之範圍與日俱廣,前途發展,正未可限量也.

3. 考 Statistik (統計學) 一字,係由 Staat (Status 國家) 引申而得,蓋即大規模有組織人羣之標識,原為研究人民生存發展條件及方法之科學,亦即古代之政治學也.迄後屢加改作,至其結果,可以數字表示,始有今日之義

此種範圍不廣,組織嚴密之學問,至晚近而益形擴大,蓋宇宙現象,其原因或複雜至不可思議,或淆混至不可區析,或繁困至不可分究.而欲於渠茫之現象中,蒐羅事實之共相,固舍統計學莫由,此亦現今學子之所公認者也.然僅蒐

集事實,而不加以整理,則猶不能致用.故必須序列之,比較之,更從而推斷之.然此三者莫不含有數字之運用,是則統計研究,固有賴於數學者矣.

以廣義言,統計學者,就任一現象範圍中,蒐集某項事實,而整理其次序,推出其結論,以說明其現象,研究其原因之科學也.數學方法之致用於此者,曰「統計法」,亦曰「數學統計」,又稱「理論統計」.其論材料之審測,蒐集,及整理者曰「實用統計」.然實用統計,必須理論以引導其發展.而理論統計,亦賴前者以供給其研究之材料,而導出新穎之問題.是則二者蓋若輔車之相依,幾無分立之可能也.

現今科學,除本身固有之研究方法外,更用統計方法以助之者,固已數見不鮮.即如物理,化學,力學,星學,生理,遺傳,在昔學者或未發見其與統計之關係,或謂所學已臻精確之極境,或不知蒐集有統計價值之材料,而作有系統之觀察,近亦以其可以致用,而加意研究之矣.

第一編　定標論

論固定標準

第一章　標準之符號及配合

1. 本章所用之集團,其個體僅有非此即彼之性質,而無長短大小之度量,茲依固定標準,而研究其互變如下.

研究此種集團,步驟有二:先組類別,再計每類中樣元個數,所得結果,即為集團中標準分佈於個體之情狀.

雖然所謂集團,若非在特例,所有該項事物,同時聚於一處,則恆不能包括其全體,而為全體中之小部.實即所取之試樣也.故欲視此種部分研究,所得某標準分佈於其樣元之情形,即為彼在總體中分佈於其所有個體之真象.則非加以假說不可.然據經驗所得,此種假說,往往與事實不符,惟集團之範圍愈大,則其距事實亦愈近,斯為統計學中,積無數經驗,而推得之原理耳.

統計學中,集團組織,至為重要,蒐集材料,不可稍存偏見,至令標準之中,有所獨重.尤不可使此種偏見,成為事實,故

個體取出,不當擇選,號碼給與,宜爲任意,然後依一定法則,取每五,每十,或每若干個,分爲小羣而研究之,庶不至有所偏重,而貽乖誤之譏.

有時所取集團,卽爲該項事物之全體,如一國之人口統計等,則因所有個體均已在觀察範圍中,故僅須按所擇標準,以求其對於個體之分佈,卽可直接得其眞象,而不必有賴於假說.反之欲研究一鄉或一市髮色對於青年之分佈,則宜取區中學童作爲試樣,而認由此推斷之結果,卽爲標準分佈之眞象,是則有待乎假說者.

上述步驟中,惟類別組織,須特別加以申述,以其多係配合手續也.

2. 設僅定一標準,則集團中之個體,對之必有合有不合.試依對分法列合者爲一類,而以拉丁大楷字母如 A 者表之,其不合者爲一類,而以相當希臘小楷字母如 α 者表之.如是則此集團分爲正反二類,吾人若以 (A), (α) 表其相當範圍,N 表集團之範圍,則得

$$N=(A)+(\alpha). \qquad (1)$$

倘所設之標準有二,更按原義以 A, B, α, β, 表其正反諸類,且同時注意標準之配合,則又可得新類.

$$AB, \alpha B, A\beta, \alpha\beta, \qquad (2)$$

式中並列字母如 AB, 爲同時適合 A, B 二標準之組類.

αB則爲適合B而不適合A者.

上述各類,依其連合符號之個數,稱爲二級類,其級數在僅含二個標準之集團中爲所有類級之最高者.以其中除二級類外,僅有一級類.

$$A, B \quad \alpha, \beta \text{也.} \tag{3}$$

所謂A者,實包括一切有A性質之個體而言,而不論其同時適合B標與否,故若以括號表示類之範圍,則可得一重要之關係.即

$$(A)=(AB)+(A\beta) \quad (B)=(AB)+(\alpha B)$$

$$(\alpha)=(\alpha B)+(\alpha\beta) \quad (\beta)=(A\beta)+(\alpha\beta) \tag{4}$$

$$N=(AB)+(\alpha B)+(A\beta)+(\alpha\beta). \tag{5}$$

若所設之標準有三,則可按所注標準之同時爲三爲二,或僅爲一,以構成由三級至一級之一切類.而得其

三級類爲:

ABC	$A\beta\gamma$
αBC	$\alpha B\gamma$
$A\beta C$	$\alpha\beta C$
$AB\gamma$	$\alpha\beta\gamma$;

二級類爲:

AB	AC	BC
αB	αC	βC

$$A\beta \qquad A\gamma \qquad B\gamma$$
$$\alpha\beta \qquad \alpha\gamma \qquad \beta\gamma;$$

一級類為:

$$A \qquad B \qquad C$$
$$\alpha \qquad \beta \qquad \gamma$$

依上述符號之性質,一切低級類之範圍,均可以相鄰高級類之範圍表之.故最高級類或最末級類,在分類組織中至為重要.蓋由其範圍可以表示標準之各種分佈也.由是易知

$$(A)=(AB)+(A\beta)$$
$$(AB)=(ABC)+(AB\gamma)$$
$$(A\beta)=(A\beta C)+(A\beta\gamma),$$

結果可得

$$(A)=(ABC)+(AB\gamma)+(A\beta C)+(A\beta\gamma)$$ 諸式.

同理有

$$\alpha=(\alpha BC)+(\alpha B\gamma)+(\alpha\beta C)+(\alpha\beta\gamma),$$

由是得

$$(A)+(\alpha)=(ABC)+(\alpha BC)+(A\beta C)+(AB\gamma)+(A\beta\gamma)$$
$$+(\alpha B\gamma)+(\alpha\beta C)+(\alpha\beta\gamma)=N$$

及 $(B)+(\beta)=(C)+(\gamma)=N.$

3. 吾人若認一切適合標準之類咸為正類,而視集團

爲零級正類,且以N表其範圍,則由一切正類之範圍,可以推算其餘各類之範圍.故若所定標準僅有一個,則由N及(A)可得所求一級類爲

$$(\alpha)=N-(A).$$

若所定之標準有二,則由(AB),(A),(B)及N可推出其餘一切類之範圍.例如所求爲(αB)則由

$$(B)=(AB)+(\alpha B)$$

得　$$(\alpha B)=(B)-(AB).$$

所求爲$(\alpha\beta)$則由

$$N=(AB)+(\alpha B)+(A\beta)+(\alpha\beta)$$

$$=(AB)+(B)-(AB)+(A)-(AB)+(\alpha\beta)$$

得　$$(\alpha\beta)=N-(A)-(B)+(AB) \tag{6}$$

如所設之標準有三,且(ABC),(AB),(AC),(BC),(A),(B),(C),N,諸正類範圍咸爲已知,則可求得其餘諸類之範圍.例如自

$$(B)=(AB)+(\alpha B)$$

得　$$(\alpha B)=(B)-(AB);$$

自　$$(BC)=(ABC)+(\alpha BC)$$

得　$$(\alpha BC)=(BC)-(ABC)$$

自　$$(\alpha B)=(\alpha BC)+(\alpha B\gamma)=(BC)-(ABC)+(\alpha B\gamma)$$

得　$$(\alpha B\gamma)=(B)-(AB)-(BC)+(ABC) \tag{7}$$

又若

於 $(\alpha\gamma)=(\alpha B\gamma)+(\alpha\beta\gamma)$ 中

以依(6)式求得之 $(\alpha\gamma)$ 及(7)式所得之 $(\alpha B\gamma)$ 代入,則得

$$(\alpha\beta\gamma)=N-(A)-(B)-(C)+(AB)+(AC)+(BC)-(ABC) \tag{8}$$

故知其餘諸類,均可以正類表示.

4. 玆更述關於同級之正反類個數,及集團中所有正類之總數,以明其相互之關係,則因每一配合標準中,每一位置可以正反二種符號相替換(換言之,即可以拉丁字母及希臘字母相置換),可得

$2\binom{n}{1}$ 個一級類,

$2^2\binom{n}{2}$ 個二級類,

....

....

$2^n\binom{n}{n}$ 個 n 級類.

至於 x 級正類之個數,則在上列相當結果中消去其第一因子 2^x 即得.吾人若視集團爲零級正類,則得所有正類之總共個數爲

$$1+\binom{n}{1}+\binom{n}{2}+\cdots\cdots\cdots\cdots+\binom{n}{n}=(1+1)^n=2^n.$$

例如所設之標準有三,即 $n=3$,則可展得各級正類及所

有正類之個數如下:

零級類	1個
一級類	6個
二級類	12個
三級類	8個
所有正類總數	8個.

此處最宜留意者,即最末類之個數與正類之總數相符,當所設標準有 n 時,均等於 2^n,且此二組類羣,每組中之各類,均互為獨立.質言之,即任一最末類之壓數,(即範圍)不能以其他最末類之壓數表示,任一正類之壓數,亦不能以其他正類之壓數表示.反之,任一 m 級類之範圍,則可以同級及低級之一切正類範圍表示,故知同級諸類之範圍,則非互為獨立者.且知其中獨立類之個數,與該級及一切低級正類之總數相同,亦即謂當級數為 m 時,僅有

$$\binom{n}{m}+\binom{n}{m-1}+\binom{n}{m-2}+\cdots\cdots+1\text{ 個獨立類也.}$$

故當 $m=1$ 時,僅有 $n+1$ 個獨立類;

當 $m=2$ 時,僅有 $\frac{1}{2}(n^2+n+2)$ 個獨立類;

當 $m=3$ 時,僅有 $\frac{1}{6}(n^3+5n+6)$ 個獨立類;

當 $m=4$ 時,僅有 $\frac{1}{24}(n^4-2n^3+11n^2+14n+24)$ 個獨立類;

當$m=5$時,僅有$\frac{1}{120}(n^5-5n^4+25n^3+5n^2+94n+120)$個獨立類;

餘類推.

又若所設之標準有四,即$n=4$,則在

$2\binom{4}{1}=8$ 個一級類中,僅有 5 個互爲獨立;

$2^2\binom{4}{2}=24$ 個二級類中,僅有 11 個互爲獨立;

$2^3\binom{4}{3}=32$ 個三級類中,僅有 15 個互爲獨立;

至於四級類則有$2^4\binom{4}{4}=16$個,且均互爲獨立.

5. 例 新育嬰孩依三種標準之分類.

[註] 設所定之標準有三,則依算式可得二級類十二個,即

$AB, A\beta, \alpha B, \alpha\beta$; $BC, B\gamma, \beta C, \beta\gamma$; $AC, A\gamma, \alpha C, \alpha\gamma$; 是但設上語僅有七個互爲獨立以其餘五類均可以此七類表之也.如設

$(AB), (BC), (B\gamma), (\beta\gamma), (AC), (A\gamma), (\alpha C)$ 互爲獨立,則可得

$(\alpha B)=(B)-(AB)=(BC)+(B\gamma)-(AB)$,

$(\beta C)=(C)-(BC)=(AC)+(\alpha C)-(BC)$,

$(A\beta)=(A)-(AB)=(AC)+(A\gamma)-(AB)$,

$(\alpha\gamma)=(\gamma)-(A\gamma)=(B\gamma)+(\beta\gamma)-(A\gamma)$,

$(\alpha\beta)=(BC)+(B\gamma)+(BC)+(\beta\gamma)-\{(AB)+(A\beta)+(\alpha B)\}=(AB)-(BC)$
$+(\alpha C)-(A)+(\beta\gamma)$.

正標準:　A活，B法生，C男；

相當反標準：α死，β私生，γ女

設1913年奧國新生嬰孩紀錄,其最末類之屢數為:

$(ABC)=391718$　$(A\beta C)=52965$

$(\alpha BC)=10312$　$(\alpha\beta C)=2189$;

$(AB\gamma)=370200$　$(A\beta\gamma)=49880$

$(\alpha B\gamma)=7838$　$(\alpha\beta\gamma)=1686$.

由是得其總和

$$N=886788.$$

依算式

$$(AB)=(ABC)+(AB\gamma),\ (AC)=(ABC)+(A\beta C),$$

$$(BC)=(ABC)+(\alpha BC)$$

$$(A)=(ABC)+(A\beta C)+(AB\gamma)+(A\beta\gamma)$$

$$(B)=(ABC)+(\alpha BC)+(AB\gamma)+(\alpha B\gamma)$$

$$(C)=(ABC)+(\alpha BC)+(A\beta C)+(\alpha\beta C);$$

所得一切正類之屢數如下:

$(N)=886788$　新育嬰孩之總數

$(A)=864763$　活嬰孩

$(B)=780068$　法生嬰孩

$(C)=457184$　男孩

$(AB)=761918$　活法生

$(AC)=444683$ 活男孩

$(BC)=402030$ 法生男孩

$(ABC)=391718$ 活法生男孩

由是求得最末各類之屢數為

$(\alpha B\gamma)=(B)-(AB)-(BC)+(ABC)=7838$ 死法生女孩

$(\alpha\beta\gamma)=(N)-(A)-(B)-(C)+(AB)+(AC)+(BC)$

$-(ABC)=1686$ 死私生女孩.

6. 上述相異二組類屢數,知其一組集團中標準之分布,可以完全決定.所謂相異二組類屢數者,係指最末類組及正類組而言.其他當標準為 n 時亦合 2^n 類之混成組,或亦能顯示集團中標準之分布,惟實際上則僅用上述二組.至此二組是否果能顯示標準分布之眞象,則其解答恆視所與之値為最末類或正類而異.蓋類屢數事實上僅可為正為零,但絕對不能為負.(4),(5)二式所示正類屢數,恆可以最末類屢數和表之,故若所與最末類屢數組為正,則所得正類之屢數亦為正,而此種與組即能顯示標準分布之實情.

反之,最末類屢數,雖可以正類屢數表示,惟算式中加減互見,故須所與屢數能使所得結果無一為負,此種與組方能顯示標準分布之眞象.故若所與之正類屢數為正,而欲判其是否合理,則宜依法求出一切最末類,察其結果是否

爲正或零,然後推斷之.

在上述情形中,欲考驗其結果,則須加以核算,庶不至以事實上不可能之數系,顯示標準分布之情形.

例如所與三標分布之正類屢數爲:

$$N=1000 \qquad (AB)=42$$
$$(A)=525 \qquad (AC)=147$$
$$(B)=312 \qquad (BC)=86$$
$$(C)=470 \qquad (ABC)=25$$

則由此算出之最末類爲:

$$(ABC)=25$$
$$(\alpha BC)=(BC)-(ABC)=61$$
$$(A\beta C)=(AC)-(ABC)=122$$
$$(AB\gamma)=(AB)=(ABC)=17$$
$$(A\beta\gamma)=(A)-(AB)-(AC)+(ABC)=361$$
$$(\alpha B\gamma)=(B)-(AB)-(BC)+(ABC)=209$$
$$(\alpha\beta C)=(C)-(AC)-(BC)+(ABC)=262$$
$$(\alpha\beta\gamma)=N-(A)-(B)-(C)+(AB)+(AC)$$
$$+(BC)-(ABC)=-57$$

因上述結果中有負數出現,故所與正類組爲不合理,亦即不能表示標準分布之眞象也.

第二章　標準之倚性

7. 設有分布於集團個體上之二標準A, B,則依個體之性質,有時A, B,可以互相吸引,即個體之適合標準A者亦易適合於B.是則A之出現利於B而不利於β也反之,二標準中亦有相衝者,如A易與β同時出現,而不易與B同時出現是.欲詳其理宜先明不倚之意義.

不倚云者,二標準A, B中,A標對於適合B標個體之分布,與其對於適合β標個體之分布,二相對度數相同之謂其意義可以算式

$$\frac{(AB)}{(B)}=\frac{(A\beta)}{(\beta)} \tag{1}$$

表之.由是易知

$$\frac{(AB)}{(B)}>\frac{(A\beta)}{(\beta)} \tag{2}$$

可以表A, B二標相吸

$$\frac{(AB)}{(B)}<\frac{(A\beta)}{(\beta)} \tag{3}$$

可以表其相衝.

又吾人可視(1)式之標準互爲獨立,(2),(3)中之標準互爲相倚.再按$\frac{(AB)}{(B)}-\frac{(AB)}{(\beta)}$之爲正爲負,稱(2)式爲正倚(3)式爲負倚以區別之.

設標準A對於一雙正反標準B, β不倚之條件已完全成立,則B對於A, α; α對於B, β; β對於A, α;不倚之條件

亦應成立換言之,卽依(1)式應有

$$\frac{(BA)}{(A)}=\frac{(B\alpha)}{(\alpha)}$$

$$\frac{(\alpha B)}{(B)}=\frac{(\alpha\beta)}{(\beta)} \qquad (1a)$$

$$\frac{(\beta A)}{(A)}=\frac{(\beta\alpha)}{(\alpha)}$$

也.今自(1)式有

$$\frac{(B)-(AB)}{(B)}=\frac{(\beta)-(A\beta)}{(\beta)}$$

因得 $$\frac{(\alpha B)}{(B)}=\frac{(\alpha\beta)}{(\beta)}$$

而上列諸式中之第二式,遂獲證明.又由(1)式及上式應用加比之理可得

$$\frac{(AB)+(A\beta)}{(B)+(\beta)}=\frac{(A)}{N}=\frac{(AB)}{(B)}$$

$$\frac{(\alpha B)+(\alpha\beta)}{(B)+(\beta)}=\frac{(\alpha)}{N}=\frac{(\alpha B)}{(B)}$$

故有 $$\frac{(BA)}{(A)}=\frac{(B\alpha)}{(\alpha)}$$

是卽上列諸式中之第一式也.

由此又可推出第三式

$$\frac{(\beta A)}{(A)}=\frac{(\beta\alpha)}{(\alpha)} \qquad \left[\frac{(A)-(BA)}{(A)}=\frac{(\alpha)-(B\alpha)}{(\alpha)}\right]$$

又在 $$\frac{(AB)}{(B)}=\frac{(A)}{N}$$

中若以A,B互換,則得

$$\frac{(AB)}{(A)}=\frac{(B)}{N}$$

因有 $$(AB)=\frac{(A)(B)}{N}$$

或 $$\frac{(AB)}{N}=\frac{(A)}{N}\cdot\frac{(B)}{N} \tag{4}$$

此式與概算有關,學者宜特別注意,茲特述其意義如下:若A,B互為獨立則其聯標之相對屢數恆等於其單標之相對屢數積,此理對於標組中之一切標偶,咸為有效.

故 $$\left.\begin{aligned}\frac{(A\beta)}{N}&=\frac{(A)}{N}\cdot\frac{(\beta)}{N},\\ \frac{(\alpha\beta)}{N}&=\frac{(\alpha)}{N}\cdot\frac{(\beta)}{N},\\ \frac{(\alpha B)}{N}&=\frac{(\alpha)}{N}\cdot\frac{(B)}{N},\end{aligned}\right\}(4a)$$

三式亦能成立.

由是更得獨立標準之重要關係

$$(AB)(\alpha\beta)=(\alpha B)(A\beta)=\frac{(A)(B)(\alpha)(\beta)}{N^2} \tag{5}$$

8. 以上所得均係由(1)式推出之結果,故其關係可以等號表示.今若以(2)式或(3)式易(1)式,則凡一切依法推得之關係,皆無相等之性質,而須以大於或小於表明之.

由表A,B正倚之(2)式,所推得屢數之關係,咸與(2)式有相同之意義.故若改書(2)為

$$(AB)(\beta)>(A\beta)(B)$$

更取其與恆等式

$$(AB)(B)=(AB)(B)$$

之和,則得

$$\frac{(AB)}{(B)}>\frac{(A)}{N}\text{及}\frac{(BA)}{(A)}>\frac{(B)}{N}. \qquad (6)$$

更自恆等式

$$N(A)=N(A)$$

中減去不等式

$$N(AB)>(A)(B),$$

則有

$$\frac{(\beta A)}{(A)}<\frac{(\beta)}{N}\text{及}\frac{(A\beta)}{(\beta)}<\frac{(A)}{N}. \qquad (7)$$

同理自

$$N(B)=N(B) \text{ 減 } N(AB)>(A)(B),$$

得 $$\frac{(\alpha B)}{(B)}<\frac{(\alpha)}{N} \text{及} \frac{(B\alpha)}{(\alpha)}<\frac{(B)}{N}. \tag{8}$$

以(7)式為出發點,更依上法行之,又可得

$$\frac{(\alpha\beta)}{(\beta)}>\frac{(\alpha)}{N} \text{及} \frac{(\beta\alpha)}{(\alpha)}>\frac{(\beta)}{N}. \tag{9}$$

[自 $N(\beta)=N(\beta)$ 中減去 $N(\beta A)<(A)(\beta)$ 卽得]

改書(9)式為

$$N(\alpha\beta)>(\alpha)(\beta),$$

若以 $(\alpha B)+(\alpha\beta)$ 代其中之 (α), 則得

$$\frac{(\alpha\beta)}{(\beta)}>\frac{(\alpha B)}{(B)}.$$

若以 $(A\beta)+(\alpha\beta)$ 代其中之 β, 則得

$$\frac{(\beta\alpha)}{(\alpha)}>\frac{(A\beta)}{(A)}. \tag{10}$$

(2)及自(6)至(10)諸式,均係表 A, B 正倚時之關係,吾人僅須將其中之不等號反置,則可得若干表 A, B 負倚之關係式.卽

$$\frac{(AB)}{(B)}<\frac{(A\beta)}{(\beta)} \tag{3}$$

$$\frac{(AB)}{(B)}<\frac{(A)}{N}, \quad \frac{(BA)}{(A)}<\frac{(B)}{N}; \tag{6'}$$

$$\frac{(\beta A)}{(A)}>\frac{(\beta)}{N}, \quad \frac{(A\beta)}{(\beta)}>\frac{(A)}{N}; \tag{7'}$$

$$\frac{(\alpha B)}{(B)}>\frac{(\alpha)}{N},\quad \frac{(B\alpha)}{(\alpha)}<\frac{(B)}{N}; \tag{8'}$$

$$\frac{(\alpha\beta)}{(\beta)}<\frac{(\alpha)}{N},\quad \frac{(\beta\alpha)}{(\alpha)}<\frac{(\beta)}{N}; \tag{9'}$$

及
$$\frac{(\alpha\beta)}{(\beta)}<\frac{(\alpha B)}{(B)},\quad \frac{(\beta\alpha)}{(\alpha)}<\frac{(A\beta)}{(A)}. \tag{10'}$$

上列算式雖多,欲用以表示倚向,則仍須審察情形,擇其能解決問題,並能予吾人以直接解答者.然此種選擇,恆視所取材料而異.因時制宜,是在學者.固無一定不移之法則也.

等式所示不倚關係,不必恰與事實吻合.蓋此種關係,即實際存在,經意外之擾,亦不必如算式所示之純確.故驟得一表互倚之不等式,欲判其係經擾亂而成,抑係表眞實相倚,則須依統計目的,取異時異地之材料,來源互別之集團,加以研究.必須所得結果表示之關係均同,始能謂之互倚.若關係不一,倚向時變,則結果必與互倚無干,而陷於無定狀態.又互倚情形分顯著及可疑二種.然眞實互倚之倚度,有遞昇性,標準倚否,亦即以此判決之.茲先就下列諸例,證明其互倚之存在如次.

9. 例一　新育嬰孩之死活,及其性別之互倚.

以 A 表活,α 表死,B 表男,β 表女,依1913[1]年奧國統計報

(1) Österr. Statistik, Neue Folge, Heft 1, P. 33*

告,其類屡數爲:

$$(AB)=1059 \quad (A\beta)=1000$$

$$(\alpha B)=1313 \quad (\alpha\beta)=1000$$

由上列與每千女孩$(A\beta)$及$(\alpha\beta)$相當之男孩個數(AB)及(αB),得$N=4372$,及一次類

$$(A)=2059, \quad (B)=2372,$$

$$(\alpha)=2313, \quad (\beta)=2000.$$

由是求得

$$\frac{(BA)}{(A)}=\frac{1059}{2059}=0{\cdot}514,$$

$$\frac{(B\alpha)}{(\alpha)}=\frac{1313}{2313}=0{\cdot}567,$$

因$\frac{(BA)}{(A)}<\frac{(B\alpha)}{(\alpha)}$由(3)知男性及活生爲負倚,換言之,亦即活生中男孩之數較死亡中男孩之數爲少也.

別法由

$$\frac{(A\beta)}{(\beta)}=\frac{1000}{2000}=0{\cdot}5,$$

$$\frac{(A)}{N}=\frac{2059}{{\cdot}4372}=0{\cdot}462,$$

知$\frac{(A\beta)}{(\beta)}>\frac{(A)}{N}$.依(7)式復得男性與活生爲負倚.蓋因女孩中活生屡率旣較集團中活生屡率爲大,則男孩中活生屡

率自較小也.

又因

$$\frac{(\alpha\beta)}{(\beta)}=\frac{1000}{2000}=0\cdot5,$$

$$\frac{(\alpha B)}{(B)}=\frac{1313}{2372}=0\cdot553,$$

知$\frac{(\alpha\beta)}{(\beta)}<\frac{(\alpha B)}{(B)}$,按(10)式所得結果,與前相同.亦卽謂女孩中死亡之數,較男孩中死亡之數爲少.

上述三結果中以第一及第三爲較大,故其結果較爲可靠.亦卽謂由此結果,吾人較易定斷活生及性別爲負倚也.

10. 例二 活生及法生

設以A, B表活孩及法生.α, β表死孩及私生.則依1913[1]年奧國統計所得每萬人口中之類屬數爲

$$(AB)=262,\quad (\alpha B)=6,$$

$$(A\beta)=36,\quad (\alpha\beta)=1;$$

由是算出

$$N=305;\quad (A)=298,\quad (B)=268$$

$$(\alpha)=7,\quad (\beta)=37$$

因 $$\frac{(AB)}{(B)}=\frac{262}{268}=0\cdot977,$$

(1) 見 Österr. Statistik, Neue Folge, Bd. 14, Heft 1, P. 37.

$$\frac{(A\beta)}{(\beta)}=\frac{36}{37}=0\cdot973,$$

故知A, B爲正倚.然以其相差過小,不易判定,故須更依他法,考其結果.如取

$$\frac{(B\alpha)}{(\alpha)}=\frac{6}{7}=0\cdot857 \text{ 與}$$

$$\frac{(B)}{N}=\frac{268}{305}=0\cdot878$$

相較,則所得結果與(8)式同.且因其相差較大,故可斷α, B爲負倚.亦即謂A, B爲正倚也.

11. 例三 聾啞及癡鈍.

以A表癡鈍,B表聾啞.則由1901年英格蘭及威爾士之統計,得

人口總數爲 N =32,528,000

癡鈍之數爲(A)=48,882.

聾啞之數爲(B)=15,246

聾啞癡鈍之數爲(AB)=451

由上列數系得

$$\frac{(AB)}{(B)}=\frac{451}{15{,}246}=0\cdot0296,$$

$$\frac{(A)}{N}=\frac{48882}{32528000}=0\cdot0015,$$

$$\frac{(BA)}{(A)}=\frac{451}{48882}=0\cdot0092,$$

$$\frac{(B)}{N}=\frac{15246}{32528000}=0\cdot0005$$

因知

$$\frac{(AB)}{(B)}>\frac{(A)}{N}\quad\frac{(BA)}{(A)}>\frac{(B)}{N}.$$

而與(6)式所示情形吻合.又因二者之相差甚大,故易斷聾啞及癡鈍爲正倚.蓋癡鈍在聾啞中居千分之29·6.而在全人口中僅占千分之1·5.又聾啞在癡鈍中居千分之9·5.而在全人口中僅占千分之0·5也.

12. 例四 父子之瞳色.

分瞳色爲深(褐色及黑色),淺(藍色及灰色)二種.而以A表淺瞳色之父,α表深瞳色之父,B表淺瞳色之子,β表深瞳色之子.則由 F. Galton 精密統計所得

$$(AB)=471,\quad(A\beta)=151,$$

$$(\alpha B)=148,\quad(\alpha\beta)=230;$$

由是算出

$$N=1000;\quad(A)=622,\quad(B)=619,$$

$$(\alpha)=378,\quad(\beta)=381$$

及諸歷率之百分數

$$\frac{(BA)}{(A)}=\frac{47100}{622}=76\%$$

$$\frac{(B\alpha)}{(\alpha)}=\frac{14800}{378}=39\%.$$

而 $$\frac{(BA)}{(A)}>\frac{(B\alpha)}{(\alpha)}$$

即謂淺瞳色之父有76% 生淺瞳色之子.深瞳色之父則僅39% 生淺瞳色之子也.

又因

$$\frac{(AB)}{(B)}=\frac{47100}{619}=76\%,$$

$$\frac{(A\beta)}{(\beta)}=\frac{15100}{381}=40\%,$$

而 $$\frac{(AB)}{(B)}>\frac{(A\beta)}{(\beta)}$$

故可得同樣結果.即謂淺瞳色之子,有76% 有淺瞳色之父.而深瞳色之子則僅有40% 有淺瞳色之父.故知淺瞳色之父子爲正倚.然瞳色深淺父可傳諸其子,子不能傳諸其父,故首式所示較爲自然.

13. 上述表A,B互倚之算式雖多,然其較差均不一致.故宜求出一判決互倚之方法以爲準繩.其法先設二標爲不倚,由(4)及(4a)二式用一次類屢數積

$$\frac{(A)(B)}{N},\ \frac{(A)(\beta)}{N},\ \frac{(\alpha)(B)}{N},\ \frac{(\alpha)(\beta)}{N},$$

算出二次類AB, $A\beta$, αB, $\alpha\beta$,之屢數而以$[AB]$$[A\beta]$$[\alpha B]$$[\alpha\beta]$

表由此算得之結果.吾人若以(AB),$(A\beta)$,(αB),$(\alpha\beta)$表觀察所得之厖數,則僅當A,B之眞質,不倚成立時,相應諸值始能相等.否則$[AB]-(AB)$必生一乖差δ又因

$$(AB)-[AB]=\delta \tag{11}$$

之值量既能伸縮,符號又可變更,故可利用之以表倚度及倚向.由等式

$$[AB]+[A\beta]=(A),(AB)+(A\beta)=(A)$$

有　$$(AB)-[AB]+(A\beta)-[A\beta]=0$$

而得　$$(A\beta)-[A\beta]=-\delta. \tag{12}$$

同理由　$$(AB)+(\alpha B)=(B),[AB]+[\alpha B]=(B)$$

得　$$(\alpha B)-[\alpha B]=-\delta. \tag{13}$$

以之代入

$$(\alpha\beta)-[\alpha\beta]+(\alpha B)-[\alpha B]=0$$ 中,

又得　$$(\alpha\beta)-[\alpha\beta]=\delta. \tag{14}$$

亦卽謂α,β與A,B之倚向相同,$A\beta$與αB之倚向相同,而與前者相反.又當$\delta>0$時A,B爲正倚.$\delta<0$時A,B爲負倚.

以$[AB]$之值代入(11)式而以二次類厖數表其結果,則得

$$\begin{aligned}\delta&=(AB)-\frac{(A)(B)}{N}\\&=\frac{1}{N}\{[(AB)+(A\beta)+(\alpha B)+(\alpha\beta)](AB)\end{aligned}$$

$$
\left.\begin{aligned}
&-[(AB)+(A\beta)][(AB)+(\alpha B)]\Big\}\\
&=\frac{1}{N}\Big\{(AB)(\alpha\beta)-(A\beta)(\alpha B)\Big\}\\
&=\frac{1}{N}\begin{vmatrix}(AB) & (A\beta)\\ (\alpha B) & (\alpha\beta)\end{vmatrix};
\end{aligned}\right\}\quad(15)
$$

(15)式中行列式之號表倚向,其絕對值表倚度.

欲求 $(A\beta)-[A\beta]$之值,僅須於(15)式中得B與β互換,則原式變爲

$$
\begin{vmatrix}(A\beta) & (AB)\\ (\alpha\beta) & (\alpha B)\end{vmatrix}\quad(12a),
$$

由此式變號復得原式.由是知依此求得之 $(A\beta)-[A\beta]$ 其符號,與(12)式所示者相同.同理將(15)及(12a)中之列互易,則得$(\alpha B)-[\alpha B]$及$(\alpha\beta)-[\alpha\beta]$之值.

14. 因 δ 爲絕對量,故用之測量倚度,極感不便,以其不適於比較也.欲免其弊,宜依事實本性,求出一相對量,此量當A,B互爲獨立時,其值最好爲零,以便依其爲正爲負而決定A,B爲正倚或負倚,又因標準之極大正倚爲A每現一次,B亦現一次;或B每現一次,A亦現一次.或A,B同現,而$(A\beta)$及(αB)爲零.其最大負倚爲:A每現一次,B不現一次;或B每現一次,A不現一次;或A,B不同現,而(AB)及$(\alpha\beta)$爲零.故此量當A,B爲完全正倚及負倚時,宜有極限,其最便之值莫如當A,B完全正倚時爲$+1$,完全負倚時

爲 -1, 以介於 -1 及 $+1$ 間之正負分數已足表異度互倚也.

算式之能括示上述一切條件者,厥爲

$$q=\frac{(AB)(\alpha\beta)-(A\beta)(\alpha B)}{(AB)(\alpha\beta)+(A\beta)(\alpha B)}$$

蓋當不倚成立,則分子等於零,而得 $q=0$.完全正倚成立,則 $(AB)=0$, $(\alpha B)=0$, 而得 $q=+1$.完全負倚成立,則 $(AB)=0$, $(\alpha\beta)=0$, 而得 $q=-1$ 也.又按(15)式更得 q,δ 之關係式爲

$$q=\frac{N\delta}{(AB)(\alpha\beta)+(A\beta)(\alpha B)} \tag{16}$$

式中之 q 稱爲倚係數,亦曰尤氏倚率,以其創自 yule[1] 也.

15. 例(1)夫婦之瞳色.

以 A 表淺瞳色之夫, B 表淺瞳色之婦, α 表深瞳色之夫, β 表深瞳色之婦,則由 F. Francis 統計所得之分布如下:

$(AB)=309,\quad (A\beta)=214,$

$(\alpha B)=132,\quad (\alpha\beta)=119.$

由是得

$N=774;\quad (A)=523,\quad (B)=441,$

$(\alpha)=251,\quad (\beta)=333.$

又設 A,B 爲不倚,算出

$[AB]=298,\quad [A\beta]=225,$

(1) Trans. Roy. Soc., A, Vol. 194 (1900), §13.

$$[\alpha B]=143, \quad [\alpha\beta]=108.$$

而得觀察結果及計算終值之比較如下:

觀察結果	計算所得	差
309	298	11
214	225	−11
132	143	−11
119	108	11

由是知A,B二標為正倚,且知其倚係數為

$$q=\frac{8523}{65019}=0.131$$

此值雖小,然足表嫦婚者於瞳色有取同之傾向.蓋因差與觀察結果之比值,在此例中介於3·5%與9·2%之間,而其倚度為逕异也.

(2)死孩之性別及其父母之婚否.

以A表法生,α表私生,B表男,β表女,則由1913[1]年奧國嬰孩統計依婚姻,性別分類,所得以10·000為底之相當死亡為:

$$(AB)=256, \quad (A\beta)=207,$$
$$(\alpha B)=397, \quad (\alpha\beta)=327,$$

由是得

(1) Österr. Statistik, Neue Folge, Bd. 14, Heft 1. P. 35*

$N=1187;\quad (A)=463,\quad (B)=653$

$(\alpha)=724,\quad (\beta)=534;$

$[AB]=254.7,\quad [A\beta]=208.3,$

$[\alpha B]=398.3,\quad [\alpha\beta]=325.7.$

因依法求得之公差,僅爲1.3.其相當倚率

$$q=\frac{1533}{165891}=0.009$$

更形微小,故死孩性別與其父母婚否之倚性不能斷定而此種乖差,僅能視爲偶然現象.

(3)植物生長所受母種單純及雜配之影響,Darwin氏曾取多種植物,加以研究. yule[1] 氏特於其中擇其便於計算者三十八種,以作統計研究之材料.

吾人若以A表配種,α表純種,B表高,β表矮.則因第一分法爲絕對,第二分法爲相對.故後者宜取一算術中量爲判值,大於此值者爲B,小於此值者爲β.由是分得之類屬數爲:

$(AB)=395,\quad (\alpha B)=179,$

$(A\beta)=168,\quad (\alpha\beta)=372.$

依法求出

$\delta=106.4\quad q=0.66$

(1) Phil, Trans, Roy, Soc. A, Bd. 194 (1900), P, 293

於是配種與長度之正倚完全決定.以此處 δ 及 q 之值頗大也.

然此三十八種植物長度,所受配種之影響各不相同.且有數種爲負倚,數種爲正倚,玆取其爲正倚者,按倚度大小列舉三種如下:

1. 山胡椒 $(AB)=17, (\alpha B)=12, (A\beta)=17, (\alpha\beta)=22;$

2. 香木犀 $(AB)=39, (\alpha B)=25, (A\beta)=16, (\alpha\beta)=30;$

3. 紫英藤 $(AB)=63, (\alpha B)=18, (A\beta)=10, (\alpha\beta)=55;$

由是算出

1: $\delta=2.5,$ $q=0.29$

2: $\delta=7.0,$ $q=0.49$

3: $\delta=22.5,$ $q=0.90$

故配種對於植物生長之影響,恆視其種類而異,不必配合卽佳.

(4)夫婦之身材.

Darwin 之研究遺傳也,曾細考婚媾身材之互擇.其所得結果,可表列如下:

		妻		
		大	中	小
夫	大	18	28	14
	中	20	51	28
	小	12	25	9

取上列九格表以大小為標準而重列之,則得二四格表如下:

		妻	
		大	不大
夫	大 ………	18	42
	不大………	32	113

		妻	
		小	不小
夫	小 ………	9	37
	不小………	42	117

由第一表所得之倚係數為

$$q=\frac{690}{3378}=0.20,$$

由第二表所得之倚係數為

$$q=-\frac{501}{2607}=-0.19;$$

故知 q 之大小,與材料配置有關.至於標準分佈之百分數,則在第一表中 30% 大夫娶大婦,36% 大婦嫁大夫.在第二表中,19.5% 小夫娶小婦,17.6% 小婦嫁小夫.是則身體魁梧者之相吸,遠勝於身體矮小者之相吸.而其相吸之最甚者,則又當推大小適中之夫婦,以其在配合總數中獨占 24.9% 也.

(5)童年各種殘廢之互倚

英國科學協會曾取十四歲之男女學童 50,000, 研究其身體及精神上之欠缺.

其注意點有四:

A. 不勻稱,不完全之身體發育.

B. 行動怪異,所表示之神經欠缺.

C. 身體瘦弱,顏色蒼白,所表現之營養不足.

D. 學業不良,所表現之經神衰萎.

於此四標中,每取二標相配,共得六種組合由計算結果,知各組標準,咸爲互倚.玆將其倚係數表列如下:

配合	q			
	男孩		女孩	
	1888–91	1892–94	1888–91	1892–94
AB	0.898	0.750	0.904	0.784
AC	0.903	0.848	0.952	0.916
AD	0.893	0.846	0.929	0.900
BC	0.862	0.783	0.914	0.814
BD	0.893	0.897	0.926	0.905
CD	0.791	0.823	0.863	0.835

表內除男孩中之*BD*及*CD*外,第二期各標之倚度,均較第一期相當標準之倚度爲小.至於其所以相差之原因及意義,則吾人不知此表如何作法,殊不能加以臆斷.惟女孩中各種殘缺之倚度,較男孩中相當殘缺之倚度爲大,則顯而易見耳.

第三章 間接互倚

16. 上章所述,其最要者,爲當A,B分佈於集團N時,若A,B之類屬數大於或小於A,B不倚時所應有者;且在

$$\frac{(AB)}{N}>\frac{(A)}{N}\cdot\frac{(B)}{N}\text{及}\frac{(AB)}{N}<\frac{(A)}{N}\cdot\frac{(B)}{N}$$

即 $$(AB)>\frac{(A)(B)}{N}\text{及}(AB)<\frac{(A)(B)}{N}$$

中各式左右相差甚鉅,則須斷定A,B實爲互倚,而不能謂此種差異爲與實不倚經擾亂而成者.倘二者相差過小,則須取相異公式,考其倚度之昇降而判定之.

標準互倚,可分直接及間接二種.前者起源於本身,後者有賴乎媒介,此其不同之要點也.至欲辨別互倚之爲彼爲此,則僅能依互倚情形,加以臆斷.惟此種臆斷,須經統計方法之考驗,始能決其是否合理.又因判定直接互倚,須對集團作全體觀察,考驗間接互倚,須取局部互相比較,故前者又名全體互倚,後者又名局部互倚.

例如取各校學生,依A發育不全,B神經昏亂,C精神衰弱諸標準作一統計,而研究B之存在,是否可爲A,C互倚之媒.亦即研究神經昏亂,是否能使腦力失常,以致一方直接影響身體之發育,他方直接影響精神之健全也.若然,則B之存在,實爲A,C互倚之主因.而A,C之互倚爲間接.考驗間接互倚之存在與否,宜取集團中適宜部分,加以比較,而不可爲全體觀察所限制.

例如人羣之中,有種牛痘亦有不種者.若統計天花結果所得未種痘者患天花之數,較天花與種痘無關時應有之數爲多,固因種痘有預防天花之效,而二者爲直接互倚.然若謂不種痘者,大抵皆貧賤之徒,其居處旣不衛生,其飲食尤缺營分,細菌流傳,迅若光電,天花作疾,患者必多.是則天花之流行也,視乎生活之衛生與否,而牛痘之佈種也,亦以生活狀况爲轉移.故若以A表未種痘者,B表天花,C表不衛生之生活,則據後說,A,B之互倚,實因A,C及B,C均爲直接互倚之故.亦即謂A,B爲間接互倚也.然試取集團中講求衛生未種牛痘之部分人羣,而研究其中A,B之互倚,所得結果與不講求衛生未種牛痘部分人羣中所有者相同,則第二說法根本無效.

又若某種性格,在祖孫間同時具有之數,較性情與遺傳無關時應有之數爲多,其爲互倚固已毫無疑義.然欲斷定互倚之種別,則須考驗此種性格,是否由祖直接傳諸其孫,抑係父子相承,祖傳諸父,父更傳諸其子.吾人若以A,B,C分表孫,父,祖具有某種性格者,則上述問題變爲討論A,C互倚是否因B媒介,或A,B;B,C直接互倚而致.欲判決之,宜於同具此性之祖孫中,依其父之具有是項性格與否,分爲二部,若各部中祖孫性格之互倚如故,則其爲直接互倚無疑.

7. 判決間接互倚之公式,與前章所述諸式相同,惟須取其中之總集團,以相當部分集團代之.按7節所述,在含N元之集團中,A, B不倚之條件為

$$(AB)=\frac{(A)(B)}{N}.$$

故集團C中$(A)_c$, $(B)_c$不倚時(ABC)之屢數應為

$$(AB)_c=\frac{(A)_c(B)_c}{(C)}$$

或 $$[ABC]=\frac{(AC)(BC)}{(C)}.$$

又因決定媒介互倚,須取全集團中與問題相宜之部分,加以研究.故欲判定A, B互倚是否因C媒之故,則須取觀察所得之(ABC)與計算所得之$[ABC]$相較,或取ABC之屢數,與分團(C)中(AC)及(BC)之屢率相較,換言之亦即取(ABC)與(C)中(AC)及(BC)不倚時應有之屢數相較也.若比較結果

$$(ABC)>\frac{(AC)(BC)}{(C)}, \qquad (17)$$

則A, B因C而成間接正倚之臆斷,遂獲證明.

(17)式所表關係為適於應用起見,又可化為一羣公式如下:

$$\frac{(ABC)}{(AC)}>\frac{(BC)}{(C)} \qquad \frac{(ABC)}{(BC)}>\frac{(AC)}{(C)}. \qquad (18)$$

在(17)式

$$(ABC)(C)>(AC)(BC)$$

中,以$(BC)+(\beta C)$代(C), $(ABC)+(A\beta C)$代(AC), 則得

$$\frac{(ABC)}{(BC)}>\frac{(A\beta C)}{(\beta C)}. \tag{19}$$

反之若以$(AC)+(\alpha C)$代(C), $(ABC)+(\alpha BC)$代(BC), 則上述關係,變爲

$$\frac{(ABC)}{(AC)}>\frac{(\alpha BC)}{(\alpha C)}. \tag{20}$$

上列諸式皆係視C爲媒介所得表A, B間接正倚之結果.吾人若改視B爲媒介,則所得表$A\ C$正倚之關係式爲

$$(ABC)>\frac{(AB)(BC)}{(B)}, \tag{17*}$$

$$\frac{(ABC)}{(AB)}>\frac{(BC)}{(B)}\quad \frac{(ABC)}{(BC)}>\frac{(AB)}{(B)}, \tag{18*}$$

$$\frac{(ABC)}{(BC)}>\frac{(AB\gamma)}{(B\gamma)}, \tag{19*}$$

$$\frac{(ABC)}{(AB)}>\frac{(\alpha BC)}{(\alpha B)}. \tag{20*}$$

18. 例(1)按英國統計$N=10000$男學童中A體力衰萎,B神經欠缺,及C神經不振者其正類之屢數如下:

$N=10000$

$(A)=877$　　$(B)=1086$　　$(C)=789$

$$(AB)=338 \qquad (AC)=338 \qquad (BC)=455.$$

$$(ABC)=153.$$

吾人欲研究 $A\ C$ 互倚,是否以 B 爲全因.可取體力衰萎諸學童中精神不振者之屢率$\frac{(AB)}{(A)}$,與全集團中精神不振者之屢率$\frac{(C)}{N}$相較以斷定 $A\ C$ 之互倚與否,然後再討論其互倚是否以 B 爲媒.則因二者之百分數爲

$$100\frac{(AC)}{(C)}=38.5 \qquad 100\frac{(C)}{N}=7.9$$

而 $\frac{(AC)}{(A)}>\frac{(C)}{N}$ 甚鉅,故按(6)式所示可斷定 A, C 正倚無疑.

又因

$$100\frac{(ABC)}{(AB)}=45.3, \qquad 100\frac{(BC)}{(B)}=41.9;$$

而 $$\frac{(A\beta C)}{(AB)}>\frac{(BC)}{(B)}$$

$$100\frac{(A\beta C)}{(A\beta)}=100\frac{(AC)-(ABC)}{(A)-(AB)}=34.3,$$

$$100\frac{(\beta C)}{(\beta)}=100\frac{(C)-(BC)}{N-(B)}=3.7$$

而 $\frac{(ABC)}{(\beta)}>\frac{(\beta C)}{(\beta)}$視 B 及 β 爲集團,則按(6)式無論神經欠缺與否,A, C 咸爲正倚.且因後差與前差之比爲 30.6 : 34, 故神經欠缺,不能爲體力衰萎精神不振之共因.換言之亦即

B不能爲A, C正倚之媒介也.

(2) Galton 之研究遺傳也曾取最少有六子者七十八家,而統計其三代之瞳色,若以A, B, C,分表淺瞳色之子,父,祖.α, β, γ表其爲深瞳色者,則其類屢數如下:——

$$(ABC)=1928,\quad (AB\gamma)=596,\quad (A\beta C)=552,$$
$$(A\beta\gamma)=508,\quad (\alpha BC)=303,\quad (\alpha B\gamma)=225,$$
$$(\alpha\beta C)=395,\quad (\alpha\beta\gamma)=501.$$

細察上列數系,可知三代相承,淺瞳色最多,深瞳色較少,至於淺瞳色之人,而有深瞳色之父及子者,爲數尤屬寥寥.吾人若進而研究祖孫淺瞳色之互倚,是否由父子遞傳而得.或淺瞳色人,不必有淺瞳色之子,卽可有淺瞳色之孫,則自上述記錄,因

(a) (BC)對於(C)及$(B\gamma)$對於(γ)之百分數爲

$$100\frac{(BC)}{(C)}=\frac{223100}{3178}=70.2 \text{ 及}$$

$$100\frac{(B\gamma)}{(\gamma)}=\frac{82100}{1830}=44.9;$$

而 $$\frac{(BC)}{(C)}>\frac{(B\gamma)}{(\gamma)}.$$

(b) (AB)對於(B)及$(A\beta)$對於(β)之百分數爲

$$100\frac{(AB)}{(B)}=\frac{252400}{3052}=82.7, \text{ 及}$$

$$100\frac{(A\beta)}{(\beta)}=54.2;$$

而　$$\frac{(AB)}{(B)}>\frac{(A\beta)}{(\beta)}.$$

(c) (AC)對於(C)及$(A\gamma)$對於(γ)之百分數爲

$$100\frac{(AC)}{(C)}=\frac{248000}{3178}=78.0,\text{及}$$

$$100\frac{(A\gamma)}{(\gamma)}=\frac{110400}{1830}=60.3;$$

而　$$\frac{(AC)}{(C)}>\frac{(A\gamma)}{(\gamma)}.$$

故極易斷定淺瞳色之祖孫父子均爲正倚.且因其差異之大小,可知相鄰二代之倚度相近,祖孫之倚度較弱.

以70.2−44.9=25.3與82.7−54.2=28.5相近而78.0−60.3=17.7較二者均爲小也.至此問題之最後判決,則須分別計算(d)以B爲媒,及(e)以β爲媒時,淺瞳色祖及深瞳色祖所生淺瞳色孫之百分槪率,即得

(d):　$$100\frac{(ABC)}{(BC)}=\frac{192800}{2231}=86.4$$

$$100\frac{(AB\gamma)}{(B\gamma)}=\frac{59600}{821}=72.9$$

(e):　$$100\frac{(A\beta C)}{(\beta C)}=\frac{55200}{947}=58.3$$

$$100\frac{(A\beta\gamma)}{(\beta\gamma)}=\frac{50800}{1009}=50.3$$

按(19*)式AC雖因B而爲正倚,然以其倚度86.4−72.6

$=13.8$ $58.3-50.3=8.0$較AC直接正倚時之倚度17.7爲小,故與其謂祖孫淺膚色爲藉媒於B之間接正倚,不如謂B爲促成A, C直接互倚之因子.

(3) 1891年英格蘭及威爾士人口統計中,所有A盲聾,B白痴,C聾啞,及各種殘缺配合之結果爲:

	I 所有男子總數	II 五十五歲以上男子總數
N	14,053,000	1,377,000
(A)	12,281	5,538
(B)	45,392	10,309
(C)	7,707	746
(AB)	183	65
(AC)	51	14
(BC)	299	47
(ABC)	11	3

吾人若依上表以求I,II每千人中(1)所有各種殘缺之屢率,(2)盲聾白痴之直接倚度,及(3)此二種殘缺以聾啞爲媒之間接倚度.

則須(1)千倍I,II之

$$\frac{(A)}{N},\ \frac{(B)}{N},\ \frac{(C)}{N},$$

(2)千倍I,II之

$$\frac{(AB)}{(B)}及\frac{(A\beta)}{(\beta)}相較,$$

(3)千倍I,II之

$$\frac{(ABC)}{(BC)}及\frac{(A\beta C)}{(\beta C)}相較,$$

所得結果表列如下:

殘缺之種類	I. 每1000男子中之個數	II.每1000五十五歲以上男子中之個數
盲聾	0.87	4.02
癡妄	3.23	7.49
聾啞	0.55	0.54
癡妄中之盲聾	4.0	6.3
常態中之盲聾	0.9	4.0
癡妄聾啞中之盲聾	36.8	63.8
常態聾啞中之盲聾	5.4	15.7

細察表中對立各項,易知三種殘缺之中,以白癡爲最甚,聾啞爲最少.盲癡則與年齡俱長,聾啞則與年齡無關.又在I, II中,盲聾白癡均爲正倚.而尤以老年人爲甚,同樣在聾啞中,二者亦爲正倚,亦以老年人爲甚,以此時I, II之倚度爲36.8對63.8也又聾啞中之盲聾,爲5.4對15.7, 亦以老年人爲最.

第四章　複分類

19. 以上所論之標準,僅有正反二面.換言之即個體之不合正標者,必合其反標,非此即彼,蓋分類法中之最簡單者也.

至於複分類,則其基本標準復各分爲若干子標,如所定基本標準爲 $A,B,\cdots\cdots K$ 則其子標爲

$$A_1,A_2,\cdots\cdots A_i\cdots\cdots A_m$$

$$B_1,B_2\cdots\cdots B_j\cdots\cdots B_n$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$K_1,K_2\cdots\cdots K_p\cdots\cdots K_s$$

是盡取

$$A_iB_j\cdots\cdots K_s$$

之可能組合,則得集團中所有之末類,且知其個數爲 $mn\cdots\cdots s$,分類愈詳,固可使結果愈臻精密,然若過於繁冗,則徒足使讀者頭昏目眩,而不能收提綱絜領之效.

此種分法之便於觀察及研究者,爲基標僅有二個,子標不過 m 及 n 之複分類,蓋其中各種關係及研究,均可由雙檢表顯出.按圖索驥,而無求端理紊之勞,且以其最爲常見,故詳論之.

雙檢表者,以 $A_1, A_2, A_i\ A_m$ 爲行,$B_1B_2\cdots\cdots B_i\cdots\cdots B_n$ 爲列,合

mn交場之表格也.其中每一交場,卽配示 mn 類中一類之屢數.如第 i 行第 j 列之交場所示,爲 (A_iB_j) 是.至於行和

$$\sum_{j=1}^{n}(A_iB_j)=(A_i)$$

則所以表 A_i 之總屢數,而不問 B 之分佈如何.列和

$$\sum_{i=1}^{m}(A_iB_j)=(B_j)$$

則所以表 B_j 之總屢數,而不問 A 之分佈如何.

$$\sum_{i=1}^{m}(A_i)=\sum_{j=1}^{n}(B_j)=N$$

則爲集團之範圍.茲將雙檢表之格式舉列如次:

	A_1	A_2	A_3	…	…	…	
B_1	(A_1B_1)	(A_2B_1)	(A_3B_1)	…	…	…	(B_1)
B_2	(A_1B_2)	(A_2B_2)	(A_3B_2)	…	…	…	(B_2)
B_3	(A_1B_3)	(A_2B_3)	(A_3B_3)	…	…	…	(B_3)
…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…
	(A_1)	(A_2)	(A_3)	…	…	…	N

欲得標準分佈之烏瞰,及同類材料之比較,則須將各類原有之屢率,化爲其在1,100,或1000中之屢數

$$\frac{(A_iB_j)}{N},\ 100\frac{(A_iB_j)}{N}\ \text{或}\ 1000\frac{(A_iB_j)}{N}.$$

按舊表式樣,以造出新表,然後依新表中標準之組合及關係,以解決一切疑難問題.

例如以A表致死之病源,B表死者之職業,則表相異病源之子標 $A_1, A_2 \cdots\cdots A_m$, 表不同職業之子標 $B_1\ B_2 \cdots\cdots B_n$ 依統計目的可隨時合併,如將數種病源合而為一,以求某幾種職工死於此病者之數在其死亡總數中所占之部分.及死於此病之某幾種職工,在死於此病之一切職工中所占之數目是.至欲解決此種問題,則須由原表取已擇定之病源及職業諸交場相加,而求出 (B_j) 或 (A_i) 之和,以計算各種分配之數值,例如死亡職工B_1, B_2分佈於病源A_1, A_2, A_3之百分數為

$$100\frac{(A_1B_1)+(A_2B_1)+(A_3B_1)+(A_1B_2)+(A_2B_2)+(A_3B_2)}{(B_1)+(B_2)}$$

A_1, A_2, A_3,分佈於B_1, B_2之百分數為

$$100\frac{(A_1B_1)+(A_2B_1)+(A_3B_1)+(A_1B_2)+(A_2B_2)+(A_3B_2)}{(A_1)+(A_2)+(A_3)}$$

是.

20. 吾人若進而討論A, B子標之互倚,則須取每二同行每二同列之四交場,依14節所述而研究之.

又設諸A_iB_j為不倚,由

$$[A_iB_j]=\frac{(A_i)(B_j)}{N} \tag{21}$$

算出與諸 (A_iB_j) 相當之一切不倚值 $[A_iB_j]$ 而表列之,則得四週與原表相符之同形表.

	A_1	A_2	A_3	…	…	…	
B_1	$[A_1B_1]$	$[A_2B_1]$	$[A_3B_1]$	…	…	…	(B_1)
B_2	$[A_1B_2]$	$[A_2B_2]$	$[A_3B_2]$	…	…	…	(B_2)
B_3	$[A_1B_3]$	$[A_2B_3]$	$[A_3B_3]$	…	…	…	(B_3)
…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…
	(A_1)	(A_2)	(A_3)	…	…	…	N

取二表而比較之,則標準之互倚與否可以立斷.若未經擾亂之眞實不倚完全成立,則二表中之相當屢數應一一吻合.

若眞實不倚完全成立,則四場行列式

$$\begin{vmatrix}[A_1B_1] & [A_2B_1]\\ [A_1B_2] & [A_2B_2]\end{vmatrix}$$

爲零,以由(21)式其分子

$$(A_1)(B_1)(A_2)(B_2)-(A_1)(B_2)(A_2)(B_1)$$

等於零也.故 A_1, B_1 之爲正倚或負倚可依相當行列式

$$\begin{vmatrix} (A_1B_1) & (A_2B_1) \\ (A_1B_2) & (A_2B_2) \end{vmatrix}$$

之爲正爲負而斷定之.至其倚度之大小,則依行列式之絕對值而定.

又因正倚條件可書作

$$\frac{(A_1B_1)}{(A_2B_1)}>\frac{(A_1B_2)}{(A_2B_2)} \tag{22}$$

$$\frac{(A_1B_1)}{(A_1B_2)}>\frac{(A_2B_1)}{(A_2B_2)}, \tag{23}$$

負倚條件可書作

$$\frac{(A_1B_1)}{(A_2B_1)}<\frac{(A_1B_2)}{(A_2B_2)} \tag{22'}$$

$$\frac{(A_1B_1)}{(A_1B_2)}<\frac{(A_2B_1)}{(A_2B_2)}, \tag{23'}$$

由(22)及(23)式,知正倚條件爲鄰列比及鄰行比上昇.由(22')及(23')知負倚條件,爲鄰列比及鄰行比下降.

若雙檢表中之鄰行比,及鄰列比均爲逐漸上昇,或依次下降,無有間斷,則各行及各列中按序間離,所取二數之比亦必如是,表之具有此種性質者曰均性表.雖然均性與否,與子類之排列次序有關,故表之有均性者,既可調換其子類而失去,其不具是項性質者,亦可改列其子類而化均.

雙檢表既可調換其子類而化爲均性,一切互倚問題,乃可迎刃而解,而個體性質,始得一目瞭然.

21. 複分類中研究互倚,比較局部固屬詳盡,總賅全體尤爲重要;蓋不如是,則不知整個關係,難收舉綱得目之效.科學家之實驗也,固可取所得結果兩兩比較,以求出個別之差異.尤當求出中量或概誤以表試驗之結果.故欲解答 AB 全體是否互倚之問題,則非局部比較所能奏效,職此之由.K. Pearson[1] 特創二種係數,以爲 A, B 互倚之總測茲述其一如下.Pearson 以爲若 A_i, B_j 完全不倚則應有

$$(A_iB_j)=[A_iB_j]$$

或 $$(A_iB_j)-[A_iB_j]=0;$$

而爲 $$(A_iB_j)-[A_iB_j]=\delta_{ij}; \qquad (24)$$

中 $\delta_{ij}=0$ 時所得之特例.若 $\delta_{ij}\neq 0$, 則自可依其冠號而判決正負互倚.是則 δ_{ij} 實括示 A_i, B_j 之三種可能互倚,故可利用之以創造新數.雖然 δ_{ij} 之爲數也,若雙檢表無有均性,則其冠號正負互見,而諸差 δ_{ij} 時有互相消去之虞,而消去部分之大小,則又因子類排列之方法而異,故宜乘 δ_{ij} 而用 δ_{ij}^2,又因 δ_{ij} 爲絕對量,同一數值恆因屢數之大小而變更其功效,普遍比較尤爲不便,故宜取

$$\frac{\delta^2_{ij}}{[A_iB_j]}$$

而捨其原值.由是 Pearson 氏始以

---

(1) Drapers' Company Research Memoirs, Biometric Series I, 1904.

$$x^2=\Sigma\frac{\delta^2{}_{ij}}{[A_iB_j]} \tag{25}$$

造出

$$C=\sqrt{\frac{x^2}{N+x^2}} \tag{26}$$

以測集團中A,B互倚之總結倚度,而名之曰概倚係數.當完全不倚成立時,則$\delta_{ij}=0,x^2=0$而以$C=0$爲不倚之條件.當$\delta^2{}_{ij}$爲極大,而x^2增至無窮時.則C趨近於一,而以一爲互倚之上限.

又若以δ_{ij}之值代入(25)式,而注意$\Sigma(A_iB_j)$及$\Sigma[A_iB_j]$均等於N,則得

$$x^2=\Sigma\frac{(A_iB_j)^2}{[A_iB_j]}-N,$$

以S表$\Sigma\frac{(A_iB_j)^2}{[A_iB_j]}$,則得

$$C=\sqrt{\frac{S-N}{S}} \tag{27}$$

應用(27)式不必計算δ_{ij},即可直接求出C值,故雖$(A_iB_j)>\delta_{ij}$,平方計算較爲困難,然若有平方表在,仍以應用(27)式爲佳.

在(27)式中,吾人若改書S爲

$$\Sigma(A_iB_j)\frac{(A_iB_j)}{[A_iB_j]}.$$

則當眞實不倚成立時$\frac{(A_iB_j)}{[A_iB_j]}=1, \Sigma(A_iB_j)=N$.故 $C=0$ 爲不倚之條件,至於C之上限則理論上僅當$S=\infty$時能達子類數目爲m及m之集團,其標準之最大互倚爲雙檢表中僅對角場上有數,其餘歷數一槪爲零.下表所示卽$m=6$時此種雙檢表之形狀.

	A_1	A_2	A_3	A_4	A_5	A_6	
B_1	(A_1B_1)						(B_1)
B_2		(A_2B_2)					(B_2)
B_3			(A_3B_3)				(B_3)
B_4				(A_4B_4)			(B_4)
B_5					(A_5B_5)		(B_5)
B_6						(A_6B_6)	(B_6)
	(A_1)	(A_2)	(A_3)	(A_4)	(A_5)	(A_6)	N

蓋依前例死亡職工與致死病源之極端互倚,爲一種職工僅有一種致死病源,亦卽一種病源僅能致死一種職工也

於是 S 變爲 $\frac{(A_iB_i)^2}{[A_iB_i]}$ 之和,而對於每 i 有 $(A_iB_i)=(A_i)$ $=(B_i)$ 及 $[A_iB_i]=\frac{(A_i)(B_i)}{N}$, 因之 $\frac{(A_iB_i)^2}{[A_iB_i]}=N$, $\Sigma\frac{(A_iB_i)^2}{[A_iB_i]}=$ mN 而

$$C=\sqrt{\frac{m-1}{m}}.$$

由是觀之,雖在極端互倚,亦須分類詳密,C値始與上限相近,亦即m須增至無窮,C值始近於一.故欲求粗略分類之互倚關係,概倚係數殊不能用,以其變化欠靈也.當$m=5$時$C=\sqrt{\frac{4}{5}}=0.8944$,吾人若謂$C$值須過此數始得謂之與一相近,則子標數目不能在5×5以下.

又概倚係數,全係理論創作,故其意義不易領略,此其又一缺點也.

22. 例(1) 髮色及瞳色.

下表爲德國Baden自由邦人口統計中淺褐(A_1),棕(A_2),黑(A_3),紅(A_4),四種髮色,及藍(B_1),綠(B_2),棕(B_3),三種瞳色,對於6800男子之分佈,其十二類之屢數如下.

髮色及瞳色之(A_iB_j)表

瞳色	髮色				
	A_1淺色	A_2棕色	A_3黑色	A_4紅色	
B_1藍色	1768	807	189	47	2811(B_1)
B_2灰色或綠色	946	1387	746	53	3132(B_2)
B_3棕色	115	438	288	16	857(B_3)
	(A_1)2829	(A_2)2632	(A_3)1223	(A_4)116	6800N

觀表易知髮色分佈,以淺褐色爲多,紅色最少,瞳色分佈以灰綠色爲衆,棕色最少,又因紅髮在棕瞳色中分佈之百

分數.

$$100\frac{(A_4B_3)}{(B_3)}=\frac{1600}{857}=1.9\%$$

大於其在B_1及B_2中分佈之百數

$$100\frac{(A_4B_1)+(A_4B_2)}{(B_1)+(B_2)}=\frac{10000}{5943}=1.7\%$$

故其分佈以在深瞳色中爲多然以之與A_1, A_2, A_3之分佈相較,則又勿逮遠甚.又黑髮對於三種瞳色分佈之百分數爲

$$100\frac{(A_3B_1)}{B_1}=\frac{18900}{2811}=6.7\%$$

$$100\frac{(A_3B_2)}{(B_2)}=\frac{74600}{3132}=23.8\%$$

$$100\frac{(A_3B_3)}{(B_3)}=\frac{28800}{857}=33.6\%$$

故知其分佈在淺瞳色中爲最少,棕瞳色中爲最多.

至欲研究互倚,則須依法求出各標不倚時之歷數,茲特表列如下:——

Badem男子髮色及瞳色之$[A_iB_j]$表

瞳色	髮色				
	淺色	棕色	黑色	紅色	
藍色	1169.46	1088.02	505.57	47.95	2811
灰色或綠色	1303.00	1212.27	563.30	63.43	3132
棕色	356.54	331.71	154.13	14.62	857
	2829	2632	1223	116	6800

表中小數,所以保存算術精度,並無其他意義.此種特殊屢數,與觀察結果乖差頗大,惟紅髮分佈則極爲相近,又此表所有四場行列式如

$$\begin{vmatrix} 1088.02 & 47.95 \\ 331.71 & 14.62 \end{vmatrix} = 15906.9 - 15905.5$$

在一定近似範圍內均等於零,是爲此表之特性.

取 $(A_iB_j)-[A_iB_j]$之差,則得其符號之正負如下.

+	−	+	−
−	+	+	−
−	+	+	+

由是易知各種髮色及瞳色互倚之正負.如淺髮與藍瞳爲相吸,淺髮與綠瞳及棕瞳爲相衝是.

於依 (22) 及 (23) 求出之商

(22)

2.19	4.28	4.02
0.68	1.86	14.07
0.20	1.52	18.00

(23)

1.87	0.58	0.25	0.89
8.23	3.17	2.59	3.31

中,取前者同行相鄰二數之差,後者同列相鄰二數之差,則得其結果之符號如下.

+	+	−
+	+	−

試將(23)之末行移置於首行之後,則上述之符號變遷完全失去,且其鄰列比均爲遞降,故若將髮色子標次序改爲淺褐,紅,棕,黑,取原表而順列之,則原表化爲均性,至欲決定概倚係數,則須求出4×3=12個$\frac{(A_iB_j)^2}{[A_iB_j]}$.

$\frac{(A_1B_1)^2}{[A_1B_1]}$	2672.7	$\frac{(A_1B_2)^2}{[A_1B_2]}$	686.8	$\frac{(A_1B_3)^2}{[A_1B_3]}$	37.1
$\frac{(A_2B_1)^2}{[A_2B_1]}$	598.5	$\frac{(A_2B_2)^2}{[A_2B_2]}$	1586.9	$\frac{(A_2B_3)^2}{[A_2B_3]}$	578.3
$\frac{(A_3B_1)^2}{[A_3B_1]}$	70.7	$\frac{(A_3B_2)^2}{[A_3B_2]}$	987.9	$\frac{(A_3B_3)^2}{[A_3B_3]}$	538.1
$\frac{(A_4B_1)^2}{[A_4B_1]}$	46.1	$\frac{(A_4B_2)^2}{[A_4B_2]}$	52.6	$\frac{(A_4B_3)^2}{[A_4B_3]}$	17.5
	3388.0		3314.2		1171.0

由是得$S=3388.0+3314.2+1171.0=7873.2$

$$C=\sqrt{\frac{1073.2}{7873.2}}=0.37.$$

(2)兄弟體格及姊妹性情.

下列統計結果係取自 Pearson[1] 氏,其中A, B意義雖同,傳主則異.因其分類僅有3×3=9,故概倚係數幾不能用,但爲示例起見,暫不顧及.

兄弟體格之分佈

(1) Biometrika Vol. III.

		兄			
		強	中	弱	
弟	強	906	20	140	1066
	中	20	76	9	105
	弱	140	9	370	519
		1066	105	519	1690

姊妹性情之分佈

		姊			
		熱 烈	温 良	粗 暴	
妹	熱烈	193	177	77	452
	温良	177	996	165	1338
	粗暴	77	165	120	362
		452	1338	362	2152

第一表中屢數,以身體俱強之兄弟(A_1B_1)爲多,第二表中屢數,以性情賢淑之姊妹(A_2B_2)爲夥,至二者之相當不倚屢數表則爲

體 格

	兄			
	強	中	弱	

弟	強 …………	672.4	66.2	327.4	1066
		(+233.6)	(−46.2)	(−187.4)	
	中 …………	66.2	6.5	32.2	105
		(− 46.2)	(+69.5)	(− 23.2)	
	弱 …………	327.4	32.2	159.4	519
		(−187.4)	(−23.2)	(+210.6)	
		1066	105	519	1690

性　情

		姊			
		熱烈	温良	粗暴	
妹	熱烈 ………	94.9	281.0	76.0	452
		(+103.1)	(−104.0)	(+ 1.0)	
	温良 ………	281.0	831.9	225.1	1338
		(−104.0)	(+164.1)	(− 60.1)	
	粗暴 ………	76.0	225.1	60.9	362
		(+ 1.0)	(− 60.1)	(+ 59.1)	
		452	1338	362	2152

表中括號內之數目,爲乖差δ_{ij},其冠號表標準之倚向,故知兄弟姊妹之同點爲正倚,異點爲負倚.

取諸δ_{ij}應用,(25) (26)式,或取(A_iB_j)及$[A_iB_j]$應用(27)式,均可

由第一表求出$C=0.68$.

由第二表求出$C=0.36$.

故兄弟體格之互倚較姊妹性情之互倚爲大.

第二編　變標論

第一章　集團中之分佈

23. 變標云者,對定標言,其大旨雖定,長短級度可隨個體變遷之標準也.個體對於定標或合或否,僅有兩途,對於變標則無有不合,惟其大小級度則因個體而殊,故變標純爲量的標準.集團之中,變標級度,皆個體所挾以俱來,不過借度衡以顯示其數目.譬諸長度,無木不有,本身尺寸,亦所夙具,學者欲詳其究竟,故度之而數示其結果焉.

今後所研究之材料,爲其中個體各現有變標定度之集團.至於變標數目之多寡,則依統計目的而決定.

變標級度,通常以X表之,各元所現之特値則書爲x.此x者,莫不藉度衡或數計而決定其總額,即所以供研究之資料者也.

變量X,學者稱爲集團之判量或級量.因其特値x,實造成集團之等級,故爲描畫集團算術性質之變量,而與集團有同一之範圍.

至X之算術性質,則可分爲相異二種情形,述之如下:(1)

X能經歷定限內所有各值而爲連續變量.如大小,輕重,面積,時間及光之強度等.(2) X僅能經歷一定數值,若整數,而爲斷續變量.如生物某種器官之數目增減是.

雖然,此種分別,大半得自理論,事實方面,連續變量,僅能爲之粗略分段取值.否則以度衡之不完善,知覺之欠精審,吾人實不能取X所能有之值,一一定之也.然此種分段,已足應科學上之需求,故研究連續及間續集團可取同一方法.

24. 數測集團,所得數系,名曰原表.以其雜亂無章,故不能直接察出標度分佈於個體之情狀.然若取此數系,依大小次序而整理之,則集團中判量之最大值及最小值,可以立見.而變幅界限,判量變域,亦因之而定.此種經整理而得之結果,學者特名之曰分佈初表,以示與原表有別.

取定值爲單位,劃分X爲等距諸值 X_1 X_2, X_3……X_n, 使之最少佈滿變幅,遇必要時,亦可伸出限外.所得X_n至X_{n+1}間之距離,即爲分類之類幅,亦曰類距.

集團之中,諸元判值,居同一類幅者,成爲一類,其個數即此類之屢數,諸屢數和與集團範圍相等.

雖然集團範圍,大小不一,分類幅距,長短各殊,欲求制宜,必先達變,下述各節當可與讀者以一臂助.

a. 選擇類幅,宜注意集團之範圍,判量之變域,及定類

之單位,研究之原因等.

分類要則,須使表之中段各類,屢數均不爲零,僅兩端諸類,可以空缺.此類幅與範圍之關係,而宜留意者也.

·原定單位及其倍數,均可以爲類幅,至於何者較爲適宜,則須視變域之大小,分類之詳略而定,又分類數目不可過多,亦不可過少,過多則必濺繁困而不易明瞭,過少則必感粗率,而不能應用.

已擇定之類格,遇必要時,亦可析之使精,併之使簡.前者於求精密結果時行之,後者於求簡明關係時致用,至於精略之標準,析併之方法,則須實地經驗,始能領悟,固非言語所能概示也.

在間績集團中,判量僅有整值,分類數目因之而定,類幅意義完全失去.學者欲分類方法之劃一也,特取各距中點,以爲劃定類限之節點,而謂二相鄰節點間之距離,爲類幅.

b. 類格位置,與節點及中點之記號數字有關.而以能用整類記示爲妙,否則亦須以整數配示二種列點之一,故在 1 *cm*. 之類幅,當以整個 *cm*. 表節點, $\frac{1}{2}$ *cm*. 表中點,或以整 *cm*.表中點, $\frac{1}{2}$*cm*. 表節點,在 2 *cm*.之類幅,當以偶數表節點,奇數表中點,或以偶數表中點,奇數表節點,而在判量爲年齡,類幅爲五歲之集團,則當以 0,5,10,

……爲節點,2.5,7.5,12.5……爲中點,或以0,5,10……爲中點,−2.5,2.5,7.5,12.5……爲節點,雖後者之第一類幅,含有負齡部分,然以其於計算有益,於理論無違,故存而不避,由是易知類幅雖定,位置可移,而屢數爲零之中段各類,亦可不變類幅,而化去矣.

c.類格之幅位既定,始可取合格樣元依次編入,以決定其屢數.至於編列手續,則視豫備工作而異.

最簡單者,爲已有分佈初表,而求各類屢數,以表中判值爲順次遞昇,其同值屢見者,亦復重列數示.故僅須劃定類幅之界限,數計限中所有判值之個數,即可得各類之屢數.

若夫僅有原表,而集團範圍又甚廣漠,則欲由此製出分佈初表,不獨費時曠日,亦且傷體勞神.故不如直就原表取同類樣元,加以圈定.爲計算便利計,亦可列同類樣元於一處,每取五元或豎四橫一而貫之,或邊四斜一而絡之.如

 及  是.

由此求得之屢數和,若與集團範圍完全相合,則圈定數計必無錯誤.否則,必係計算失檢,而又須重新圈數矣.

在範圍甚大之工作,如人口統計,死亡研究等,則須與個體以卡片,而數示此元所具各標判值於其上.取所有卡片依各類判盈而分裘之,更從而數計每包中所有卡片數目,卽得相當各類之屢數,至欲考驗此種結果,有無錯誤,則須取各包卡片逐一檢其脊否誤置,而後重新數計之,由是類元編製完全決定,而無須藉助於原表及初表矣.

編定類元之方法,雖已列述如上,然若遇有與節點同值之判盈,果用何法以處之乎?此不得不稍加討論者也.

若此判值爲計算比例所得之近似結果,則宜取原有比值重加計算,求出與節點號數相異之數值,以判定其類屬.如節點號碼爲2.5,$\frac{797}{320}$及$\frac{803}{320}$之近似值亦爲2.5,則因$\frac{797}{320}$之確值爲2.49……,$\frac{803}{320}$之確值爲2.509……故前者應屬於左類,後者應屬於右類.

選擇類格,固宜使集團樣元無一與節點相合,然若所定位置至爲妥善,不幸有一二樣元,落在節點,則移換類格雖可使此弊免去,然移換之後,節點及中點必至以碎數配示,而與計算以莫大障礙.故不如依左右二類屢數,比裂此元而分屬之.但比分結果,繁雜小數在所不免,故不如分爲二半左右分隸,於是類屢數中遂有小數.5.

*d.*經有效之整理及數計,得分佈表,表含兩欄,左示類

號右表屢數.

表類之法,則爲二種,或以類限,或用中點.

中點示類頗爲簡明,且能與吾人以計算便利,以其視中點判値爲類元之公共判量,並假定類元平均分佈於幅內,而不分計類元相異之判値也.故若以x_i表第i類中點之判量,z_i表其相當屢數,則依中點示類之意義,判量爲x_i者,其屢數爲z_i,亦即謂以x_i爲中點之類,其類元個數爲z_i也.又此種類格之節點判量,爲相鄰二中點判量之算術中數.至於節點示類,則詳細顯明,尤爲特出.如以12.5—13.5(糎,年及其他)表類,z表其屢數,則依節點示類之意義,知判量在12.5以上13.5以下之類元有z個.其判量恰爲12.5或13.5者,則劈爲二半而分屬之.又此類中點之判量13,爲相鄰二節點判値之算術中數.在間斷集團中,示類號碼,即個體實現之判値,但爲計算便利計,僅有採用類幅者.

25. 例(1)同一試驗場內[1]九齡松樹之高.

下表所示,爲測度所得125株松樹之高度,其超過n公分(cm)而不滿$n+1$公分者,以$n+1$公分入算.

---

(1) H. Voter: Die Ausgleich ungsrechnung bei Bodenkultur-Versuchen. Mitteilumgen aus der Königlich-Söchsir-Chen forstlichen Versuchsanst alt zu Tharadt, Bdi (1918), S. 8-12.

125株九齡松之高度原表[1]

234	227	148	147	194	159	196	126	188	132	175	179	209	171
233	146	119	222	219	195	209	150	120	197	181	214	176	173
159	190	183	206	186	102	180	166	162	146	201	175	158	195
192	184	177	167	246	129	181	209	139	193	162	153	236	201
194	170	170	173	169	156	180	169	198	201	216	113	169	247
61	126	137	149	117	159	219	169	184	202	142	129	200	172
206	178	139	166	166	185	188	157	70	194	172	159	216	164
149	150	195	204	187	153	182	198	114	107	147	172	202	254
192	143	177	270	181	139	221	176	176	223	203	237	262	

分佈初表

x	z	x	z	x	z	x	z	x	z	x	z
61	1	142	1	166	3	182	1	198	2	227	1
70	1	143	1	167	1	183	1	200	1	233	1
102	1	146	2	169	4	184	2	201	3	234	1
107	1	147	2	170	2	185	1	202	2	236	1
113	1	148	1	171	1	186	1	203	1	237	1
114	1	149	2	172	3	187	1	204	1	246	1
117	1	150	2	173	2	188	2	206	2	247	1
119	1	153	2	175	2	190	1	209	3	254	1
120	1	156	1	176	3	192	2	214	1	262	1
126	2	157	1	177	2	193	1	216	1	270	1
129	2	158	1	178	1	194	3	219	3		
132	1	159	4	179	1	195	3	221	1		
137	1	162	2	180	2	196	1	222	1		
139	3	164	1	181	3	197	1	223	1		

(1) 欲使所得結果,有實用及科學之價值,則標題務須詳盡.此表既未載出地點,復未標明土性,不無缺憾.

觀表易知判量變域,以61爲下限,270爲上限,長凡209公分大於集團範圍125一倍有半.

吾人若取分佈初表,將$x=61$至$x=270$間相差爲1公分之各數依序列入其屢數爲零者,則於z行內書0表之,則所得結果合類二百有十,繁冗異常,而屢數空白,兩端尤甚故宜取1公寸(dm)爲類幅,整公寸爲中點,5公分之倍數爲類限,而重列之.因得第一節點爲55,最末節點爲275,而首末兩類屢數均不爲零之分佈表.

以1公寸爲幅之分佈表

X	z	X	z
55— 65	1	165—175	17
65— 75	1	175—185	17·5
75— 85	0	185—195	13
85— 95	0	195—205	13·5
95—105	1	205—215	6
105—115	3	215—225	7
115—125	3	225—235	3
125—135	5	235—245	2
135—145	6	245—255	3
145—155	11	255—265	1
155—165	10	265—275	1
			125

然表中空類有二,屢數亦起落無常,以之代表分佈,實嫌

未足,故宜縮小類數,依統計目的,適當方法,取二類或數類併而一之,若所併爲 v 類,則其併法有 v 個,蓋新類可以第一第二或第 v 類爲起點.至兩端所餘類數之不滿 v 個者,則以屢數爲零之空類補足之,而各成一新類.

由是作成之表,曰化表.因其起點之不同,復分爲各種化位以別之.

在上表中併二類爲一類,則得二種類幅爲 2 公寸之化表如下:

分 佈 化 表

I,第一化位

X	z
55— 75	2
75— 95	0
95—115	4
115—135	8
135—155	17
155—175	27
175—195	30.5
195—215	19.5
215—235	10
235—255	5
255—275	2
	125

II,第二化位

X	z
45— 65	1
65— 85	1
85—105	1
105—125	6
125—145	11
145—165	21
165—185	34.5
185—205	26.5
205—225	13
225—245	5
245—265	4
265—285	1
	125

表中屢數昇極而降,至合常規,足表分佈現象.

(2)新育嬰孩之體重[1]

取1902/3及1904/5波落乜(Bologna意大利城名)產科實習醫院統計原表,以100 g 爲幅,重加整理,所得嬰孩體重之分佈如下:

以100 g 爲幅之嬰孩體重分佈表

X	x	z		X	x	z	
		男孩	女孩			男孩	女孩
1550—1650	1600	…	0.5	3150—3250	3200	30.5	27
1650—1750	1700	…	0.5	3250—3350	3300	28.5	17.5
1750—1850	1800	…	1	3350—3450	3400	23.5	16
1850—1950	1900	1		3450—3550	3500	18	19.5
1950—2050	2000	1.5	1	3550—3650	3600	18.5	13
2050—2150	2100	2.5	2	3650—3750	3700	13.5	11
2150—2250	2200	3	3	3750—3850	3800	13	8
2250—2350	2300	4.5	9	3850—3950	3900	6	3
2350—2450	2400	1.5	6.5	3950—4050	4000	4	2.5
2450—2550	2500	3	8	4050—4150	4100	5	1
2550—2650	2600	8	8.5	4150—4250	4200	6.5	1
2650—2750	2700	14	18	4250—4350	4300	3.5	…
2750—2850	2800	14.5	17	4350—4450	4400	2.5	…
2850—2950	2900	14	18	4450—4550	4500	0.5	…
2950—3050	3000	17	25	4550—4650	4600	3	1
3050—3150	3100	26	30.5	4650—4750	4700	1	
						288	269

(1) 見 C. Gini, Sulla Variabilita deidue Sessi Cagliari 1910, P. 22-43.

上表雖不規則,然各有其特徵,若以數學方法詳加整理,則雖以200g為類幅,其無常起落仍不能免.必須以300g為幅,始有下列各種化位分佈表之形狀.

以300g為幅之分佈化表

男孩

X	x	z
1850—2150	2000	5
2150—2450	2300	9
2450—2750	2600	25
2750—3050	2900	45.5
3050—3350	3200	85
3350—3650	3500	60
3650—3950	3800	32.5
3950—4250	4100	15.5
4250—4550	4400	6.5
4550—4850	4700	4
		288

女孩

X	x	z
1450—1750	1600	1
1750—2050	1900	2
2050—2350	2200	14
2350—2650	2500	23
2650—2950	2800	53
2950—3250	3100	82.5
3250—3550	3400	53
3550—3850	3700	32
3850—4150	4000	6.5
4150—4450	4300	1
4450—4750	4600	1
		269

(3) Umbrien(在意大利中部)後備兵之頭蓋指數[1]

(1) *C. Gin.i Variabilità e mutabilità. Bologna, 1912.*

譯者按人種學中有所謂(α)頭蓋指數,其測法如下:自前額至後腦設想一垂直平分面,而測頭蓋在此面內最長之距以為蓋長.然後取右壁與此面垂直距離最遠之點與左壁稱點相連,而測其距是為頭蓋寬.百倍頭蓋寬以長除之,得頭蓋指數.

下表所示為1859至1863年入伍6209個 Umbrien 後備兵以上鄰整數入算之頭蓋指數.

Umbrien後備兵頭蓋指數分佈表

x	z	x	z	x	z
68	1	79	370	90	149
69	1	80	299	91	155
70	2	81	391	92	111
71	4	82	550	93	54
72	8	83	573	94	57
73	13	84	722	95	23
74	28	85	456	96	9
75	45	86	608	97	10
76	95	87	369	98	3
77	152	88	316		
78	188	89	447		6209

上表類限為 67·5—68·5, 68·5—69·5 等,其中部歷數起落無常,故須用化位法以除去之.如併三類為一,則以68為起點者應補二空類於末端,以69為起點者,應補二空類於首段,以70為起點者,首尾各應補一空類以足之,如是所得之分佈化表其狀如下:

三種異位分佈表

I		II		III	
x	z	x	z	x	z

69	4	67	1	68	2
72	25	70	7	71	14
75	168	73	49	74	86
78	710	76	292	77	435
81	1240	79	857	80	1060
84	1751	82	1514	83	1845
87	1293	85	1786	86	1433
90	751	88	1132	89	912
93	222	91	415	92	320
96	42	94	134	95	89
99	3	97	22	98	13
	6209		6209		6209

新表屢數昇極而降,至合常規,尤以第Ⅰ表中段自$x=78$至$x=90$處之對稱為ⅡⅢ所不及.

以中點判量為類元之判量公值,此種說法,固欠精密,似不如取類元判量算術中數之為近理,然所謂算術中值者,由原表求出以應用於分佈初表則可,至在化表則當以z為重,而不必斤斤於x之為算術中數與否.讀者細閱上例,當知此種精密計算之不能與新表以特殊影響也.

蓋依後法計算,則類元判量算術中數之在表Ⅰ類81中者為

$$\frac{299.80+391.81+550.82}{1240}=81.20,$$

在表II類85中者爲

$$\frac{722.84+456.85+608.86}{1786}=84.94,$$

在表III類80中者爲

$$\frac{370.79+299.80+391.81}{1060}=80.02$$

其值均與中點約量相差不遠,故以中點約量表之亦無大誤.至於末端中值則依後法計算應爲98,96.7及97.2然實補入空類所致.況此值與99,97,98尚云相近,且遠居末點,尤不足以影響中央分佈.

26. 統計學中類幅相等,爲普徧要求,蓋不如是,則不獨有礙關係之判斷,且足妨算法之致用.然通常所見,大都不等,尤以表末列約量在某數以上或某數以下者爲一類,而不予以終止界限者爲尤甚.如人口統計中以百歲及百歲以上者歸爲一類是.此種類幅不等之結果,可由屢數之驟起驟落察出,吾人若不知其原表,而欲化爲等幅,則亦無從着手,假定屢數爲均佈,而化大幅爲小幅,蓋萬不獲已之方法.實失精確之意義.

茲將類幅不等之1911[1]年奧國入款分佈表列舉於下.

(1) Mitteilung en des K. K. Finanzministerium Bd. XVIII (1912)

(2) 奧元昔名 Krone 今稱Schilling約當德幣 0.59 Mark.

表中類幅變更不下七次,首三次與第五次之變換,可自 z 行中,數值之起落察出,表中末項收入最大,確數不明,惟據他方報告,知此565人之收入總額,超過入款在 1800 至 2000 者134.979人之收入總額甚多.作收入統計時,若依一定類幅(如0.2千)求其分佈則所得結果旣異常長冗,而後部屢數必屢見空缺,至於按末欄方法求出平均屢數,則不獨與事實相違,亦且毫無意義.吾人試進察此表,則其所示者,旣極明晰,尤爲扼要,由是觀之參差類幅非但有時不能化爲一律,且

1911年奧國入款分佈表

收入分級(以千奧元[2]計) X	人數 Z	以0.2千爲幅所得之平均類屢數
1.2— 1.4	288.077	288.077
1.4— 1.6	204.950	204.950
1.6— 1.8	146.283	146.283
1.8— 2.0	134.979	134.979
2.0— 2.4	142.862	71.431
2.4— 2.8	94.042	71.431
2.8— 3.6	128.230	47.021
3.6— 4.4	69.419	47.021
4.4— 5.2	47.174	32.057
5.2— 7.2	60.534	32.057
7.2— 9.2	27.607	32.057
9.2—12.0	19.331	32.057
12—20	19.556	17.355
20—40	10.654	17.355
40—60	2.695	17.355
60—80	1.238	17.355
80—100	649	...
100—120	379	...
120—140	246	...
140—160	186	...
160—180	143	...
180—200	111	...
200 以上	565	...
	1,399.910	

可利用之以表分佈特徵.而類幅不等,遂爲天然不可免之事實,又如人口死亡,幼童最甚,若在死亡統計中無論老幼皆以一年爲類幅,則其結果之乖謬,可斷言矣.

關於此種問題,顯著之例爲白喉死亡對於年齡之分佈.下表所示,即1891—1900十年之間英格蘭及威爾士統計所得之結果.依等幅分類,則屢數至第四年增至極點,首五年中死亡總數爲49.979,而五齡以內之死於此病者,乃占犧牲總數61.3%.至於末欄所示,則自第六行起爲每齡中之死亡均數,其所示意義,較中行明瞭甚多,以五齡以下之屢數下降,較中行爲速故也.

1891—1900英格蘭及威爾士白喉死亡分佈表

年齡 X	死亡數 Z	每歲平均額
0—1	4.186	4.186
1—2	10.491	10.491
2—3	11.218	11.218
3—4	12.390	15.390
4—5	11.194	11.194
5—10	23.348	4.670
10—15	4.092	818
15—20	1.123	225
20—25	585	117
25—35	786	79

35—45	512	51
45—55	324	32
55—65	260	26
65—75	127	13
75歲以上	35	不定
	80.671	

27. 表列分佈,固係善法,然若所予數系過長,則欲對於全體作一鳥瞰殊非易事.而尤以範圍過大之集團爲甚.然吾人若以方法圖示其結果,則屢數變遷,分佈特點,可以一目瞭然.圖示方法分爲二種,其一先定一軸,取相宜單位劃定節點,然後於各中點豎立垂線,依各類屢數與以相當長度,所得端點順次以直線連之.則得一以首端空類中點爲起點,末端空類中點爲終點之屢數多邊形.圖1所示爲由25節例3第I化表所得表Umbrien人頭蓋指數之屢數多邊形.

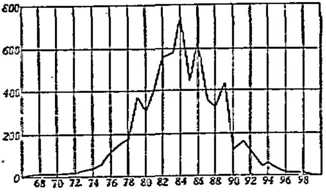


圖1 Umbrien 後備兵頭蓋指數之屢數多邊形

其二,以類幅爲底,於其上分作矩形,使其高與相當屢數

相等,而以各矩形之對稱軸表之.如是所得之圖形爲梯級式,名曰梯級圖.吾人若將矩形稱軸端點,順次以直線連接,即得屢數多邊形.是爲上述二種圖示法之關係.圖 2 所示爲由25節例(1)第 II 化位所得九齡松高度分佈之梯級圖.虛線所示即與此相當之屢數多邊形.讀者細考

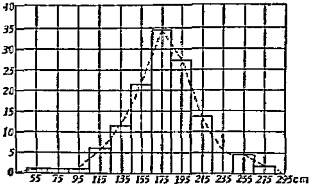


圖2 九齡松高之梯級圖及多邊形

二者之畫法,當知其相互之關係.梯級圖與多邊形,外觀雖異,其與基線所包之面積則等.以屢數多邊形各邊截梯級圖相當矩形所成三角片,兩兩相等故也.又因總面積之大小與屢數之總和有關,故可利用之表集團範圍.例如以廣8mm表類幅,高2mm表一樹,則在前例當以$8\times2=16mm^2$表個體,而梯級圖或多邊形之面積應爲$125\times16=2000mm^2$.至於二圖之顯著差異,則爲梯級圖中矩形面積可表相當屢數,而多邊形片段,則除其上限爲一直線,或相鄰三端點爲共線外,不能表示,而視角之凹入或凸出以判其大於或小於應有之屢數.又梯級圖中最大缺點,爲假定各類樣元勻佈幅內,致幅中分佈眞象,無由窺出.至於屢數多邊形,則除

最高點外,其餘均爲漸昇或漸降,故較切事實,然吾人爲計算便利計,以後仍假設其爲平均分佈者.

28. 前述之圖示方法,僅能表示一個集團之分佈情狀,而不能取範圍不同者互相比較,此種目的,誠不易達,然吾人若將集團範圍化爲定値,則雖原數不同者,亦可比較.通常恆用 1, 100, 1000 及 10之較高冪爲固定範圍,以其便於計算也.由是化得之屢數,對原有之絕對屢數,稱曰相對屢數.欲化原有屢數,成範圍爲一時之屢數,則須以N除此原値,欲化之成範圍爲 100 或 1000 時之屢數,則須以 $\frac{100}{N}$ 或 $\frac{1000}{N}$ 乘之.

吾人若以多邊形之全面積爲基量,而以1, 100或1000配示之,則其中各類之面積,可表其相對屢數,是即上述化法之發何意義也.

以上所言屢數,皆係一類所有,學者欲統計結果易於領悟,亦有取一定節點以前諸類之屢數和,以顯示分佈情形者.故舉凡判量在定限以內之類元,莫不盡數羅入,以成一類,然後將此定限歷經節點,所得諸値列成和表,附於分佈表尾,至於和表樣元之效域,則超過中點而以節點爲限.

下表所示爲25節(1)所舉類幅爲1dm之九齡松高度分佈表,與之相連者爲其相應和表,及範圍爲 1 之相對和表.至於末欄所示,則爲取前欄相鄰二數之差,所得範圍爲 1

九齡松高度分佈之和表

x_i	z	和表	相對和表	相對類度數
55— 65	1	1	0.008	0.008
65— 75	1	2	0.016	0.008
75— 85	0	2	0.016	0.000
85— 95	0	2	0.016	0.000
95—105	1	3	0.024	0.008
105—115	3	6	0.048	0.024
115—125	3	9	0.072	0.024
125—135	5	14	0.112	0.040
135—145	6	20	0.160	0.048
145—155	11	31	0.248	0.088
155—165	10	41	0.328	0.080
165—175	17	58	0.464	0.136
175—185	17.5	75.5	0.604	0.140
185—195	13	88.5	0.708	0.104
195—205	13.5	102	0.816	0.108
205—215	6	108	0.864	0.048
215—225	7	115	0.920	0.056
225—235	3	118	0.944	0.024
235—245	2	120	0.960	0.016
245—255	3	123	0.984	0.024
255—265	1	124	0.992	0.008
265—275	1	145	1.000	0.009
	125			1.000

之相對類度數表.相對類度數之計算,要以此法爲佳,以其總和爲1,有非時取近似值之直接計算所能及者.

上表第十行所示,謂九齡松高15 dm 者凡11株,亦卽謂此11株者,其高與15 dm 相差不過5 cm,又高度在155 cm 以內者凡31株,實佔全數24.8%,高度在145–155 cm 內者凡11株,實佔全數8.8%,此全表所示之分佈情形也.

又因和表各數,依次遞昇,將 z 行之起落完全免去,此其特別優點,而非分佈表所能及者.

自度數多邊形,或梯級圖,求和多邊形,須自各類終點分作垂線,使其長等於前列諸角點之經量和,或矩高和,然後依次以直線連其末點,卽成和多邊形.和多邊形之經量,在任一長條中所畫成之面積,以原定單位計,實等於判值在經量足點以內諸類之絕對度數和.

圖3所示爲上述和表之圖形,以其中矩形底邊(卽類幅)相等,故僅須依原定單位測量經線之高,卽得絕對度數,至於相對度數亦可以最末經長爲單位,而測

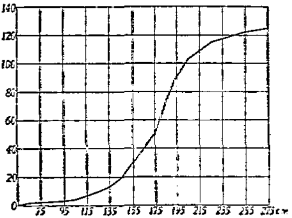


圖3 九齡松高度分布之和多邊形

出之.

29. 若集團範圍為連續增長,則雖縮小類幅,增加類數,其各類屢數,仍無空缺,如是可使屢數多邊形及梯級圖與平面縣曲線相近,此種綿續曲線,稱為屢數曲線,學者恆假定其與表研究材料特性之理想曲線多少相近,而以為此種材料之標幟.然據經驗所得,各種材料均有其相當曲線,若植物外形,礦石比重,實可藉資識別焉.夫以集團之廣,材料之衆,曲線形狀,宜若甚多.然若材料單純,分法適當,則此屢數曲線者,彙而別之,亦不過數大類而已.至於混成之集團,雜合之材料,則其本體曲線,形狀各異,亦無從而研究之矣.他若意外之擾亂,計算之舛誤,均足影響曲線之形狀,掩蔽分佈之真象.如是而欲區別園義之輕重,求出實際之情形,則非取原定材料,重加研究不可.通常學者過重曲線之局部,昧於關係之總絡,以為曲線之分類愈詳,結果愈美,然徵諸事實,則曠時費日,而所獲甚微,

學者之研究科學也,由簡而繁,化繁為簡,統計之道亦復如斯.先詳理想曲線之理論,次求複雜情形之化簡,如幾何學其根據之理想空間,雖不必全符事實,然據此以推實在,未有或謬者也.

30. 分佈之研究及判決,以模範屢數曲線

$$y=\frac{h}{\sqrt{\pi}}e^{-h^2(X-A)^2} \qquad (1)$$

爲出發點.此式表一理想分佈,當 $X=A$ 時 y 爲極大,且以 $X=A$ 爲稱軸.離此則向兩方等勢下降,而伸至無窮,其外形頗與樂鐘之縱截面相近,吾人若以 A 爲判量起點,或以 X 與 A 之差 x 表 $X-A$.

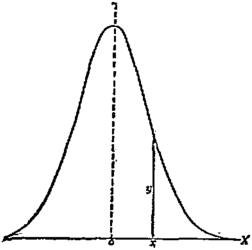


圖4 模範屢數曲線

則(1)式化爲

$$y=\frac{h}{\sqrt{\pi}}e^{-h^2x^2} \qquad (2)$$

(1)(2)之幾何圖形,完全一致,其最大差異,爲(1)式有二參數 A 及 h,(2)式僅有一參數 h,二參數中,A 值僅能影響曲線之位置,至於 h 則因 $x=X-A=0$ 時,y 值爲極大,而等於 $\frac{h}{\sqrt{\pi}}$,與 h 有關,故能直接影響曲線之形狀.

(1)(2)曲線,與橫軸所含面積,應等於相對屢數之總和,以(1),(2)式之定積分,$\frac{h}{\sqrt{\pi}}\int_{-\infty}^{\infty}e^{-h^2(X-A)^2}dX=\frac{h}{\sqrt{\pi}}\int_{-\infty}^{\infty}e^{-h^2x^2}dx=\frac{2}{\sqrt{\pi}}\int_{0}^{\infty}e^{-t^2}dt=1$ 也.

實際分佈,除自人種學及生物學中取出之材料,均不能充分適合模範曲線.即其對稱特性,經意外之擾,亦難純淨顯出,而曲線大形,遂失均稱.然若乖差甚微,則依材料本質,亦可斷其對稱,而模範曲線,遂爲判決均稱之關鍵圭臬.

抑有進者,實際分佈固不必如模範曲線所示者之合法.然其乖差之程度何如,則須以模範曲線所示者爲標準而判決之.是則模範曲線所示數量之關係,又不啻判決分佈之關鍵矣.

分佈之例,可依材料本質,判決其爲對稱者,如A. Quetelet所得1516名比國兵士胸圍統計之結果,其分佈表及屢數多邊形如下:

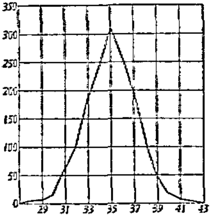


圖5 比國士兵之胸圍屢數多邊形

x 比利時寸	Z
28	2
29	4
30	17
31	55
32	102
33	180
34	242
35	310
36	251
37	131
38	103
39	42
40	19
41	6
42	2
	1516

體高及胸圍,分別測算,起於募兵時男子體格之檢定. A. Quetelet[1] 出始應用之以研究人種及其他科學,其後因研究之要求,測算之範圍益廣,遂之

(1) Recherches Sur la loi de la Croissance de L'homme (1831).

限於士兵者,進而施諸羣衆,向之限於男子者,進而施諸婦孺,而年齡之範圍,測算之部分,亦隨之而愈演愈進日趨日廣.邇來學者復發見死亡與體格之關係,此種研究之興趣,更有一日千里之勢而人種學遂爲保險,醫學,及保險術中之要科.保險事業最發達之國家,若美利堅等,則視保險之所值,而分研究之程度;雖其保險種類之不同,猶未見有能離人種學而獨立者.使保險事業愈昌,則吾知人種學之研究當益進矣.(1)

31. 集團分佈,其爲對稱,可云特例,通常所見不稱居多.其不稱度,亦復各異,故對稱曲線爲數有限,而不稱曲線之形狀無窮,惟不論曲線之對稱與否,均各有一頂點,且其左右二端,均向此點兩旁下降,是則其惟一公共性質耳.然下降速度既各線不同,而一線之中,復左右異勢,其爲龐雜概可想知.學者欲其便於研究也,特依其下降較速一端所在判域部分之爲小爲大,而分不稱曲線爲左傾右傾,以資識別.

在間續集團中,倘其範圍甚大,不稱程度亦著,則此種不稱,必係眞實情形,反之若集團範圍不廣,不稱程度亦微,則大半非眞實情形,而係擾亂所致.至於連續集團則除此二

(1) G. Bohlmann, Anthropometrie und Lebens-Versicherung. Zeitschrift für die gesamte Versicherungswissenschait, Bd. XIV (1914) P. 743—786

種情形外,尚可因分類方法生出輕微不稱,亦可改變類幅而消去之.故每遇輕微不稱,必須改變類幅,驗其是否可以消去.然後始能取範圍較廣之集團重復考驗,以判其是否係表材料特性之真實情形,抑為紊擾亂而生之變態結果.

Fechner氏曾作共頂異形之二相稱曲線,取甲之左枝,與乙之右枝於峯點連接,以為表不稱分佈之理想曲線.Pearson氏則以方程式表與對稱曲線同樣伸至無窮之不稱曲線.圖六所示,即表不稱分佈之右傾及左傾屢數曲線之理形也.

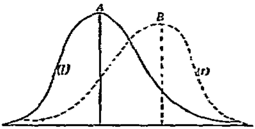


圖6 右傾(r)及左傾(l)屢數曲線

下列三例,均為分佈不稱者,特舉述之以資參考.

(1) K. Marbe[1]之研究生物也,曾親植數種豌豆,取配合所得新種之收穫.依特殊目的,相當方法而序列之.然後屏去蠹莢,而數計各莢中之粒數.所得結果,為自一至十且以十為最高,今將依粒數分類之豆莢, 60536 枚之分佈表,列舉如下:

右圖所示之屢數多邊形,有顯明之左傾,其左端終點 0,無論集團範圍若何增大,均不改變.其右端終點,則不然,倘

(1) K, Marbe: Die Gleichförmig Keit in der Welt (1919) P. 95

豆莢按粒數多寡之分佈

粒	莢
1	3792
2	8567
3	12150
4	12742
5	10388
6	7083
7	4225
8	1473
9	115
10	1
	60536

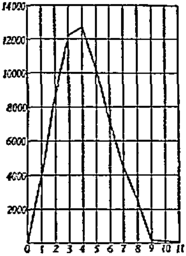


圖7 豆莢按粒數多寡之分佈

能尋出種子粒數在十以上之豆莢,即可向右移動.

此種屢數多邊形,可與一左端交軸於0,右端與軸無窮接近之不稱曲線套合.反之,6圖所示之曲線,亦必有一屢數多邊形與之套合. Marbe之行此實驗也所有特例悉數免除.而各項結果,均爲一致.故可視此種分佈,爲材料擺派之眞實情形.

間續集團之實例已略如上述,茲更將關於連續集團之二例列舉如下:

(2) W. Johannsen[1] 曾取完整紅豆558粒,測其長度分佈.其法凡長過 n 公釐者,則視爲長 $n+1$ 公釐,換言之即以釐爲

(1) W. Johannsen Elemente der exakten Erblichkeitslehre, 2 Aufl. (1913), P. 13.

類幅,而定其分佈也.下表所示卽其結果.8圖所示卽表此分佈之屢數多邊形.雖因範圍過小,分類太詳,起伏迭見,然其爲左傾則甚明瞭.

紅豆長度分佈表

公釐	
17—18	3
18—19	7
19—20	21
20—21	23
21 22	53
22—23	69
23—24	85
24—25	75
25—26	72
26—27	56
27-28	39
28—29	25
29—30	21
30—31	4
31—32	4
32—33	1
	558

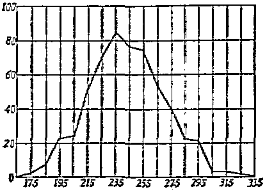


圖8 表紅豆長度分佈之屢數多邊形

(3)吾人試取下表及圖所示類似材料,互相比較,則其最可注意者,爲新生男孩[1]體重之分佈對稱,而成年男子[2]體重之分佈左傾是也.此種分佈,係按種族血源如英格蘭人,蘇格蘭人,威爾士人,愛爾蘭人等,分別統計,故其結果可表體重分佈之實際情形.

(1) 由25節(2)第一表化得

(2) G.u. Yule, Introduction etc., P. 95

新生男孩之體
重分佈表

千　斤	z
1.5—2.0	1
2.0—2.5	14
2.5—3.0	59
3.0—3.5	140
3.5—4.0	53
4.0—4.5	19
4.5—5.0	2
	288

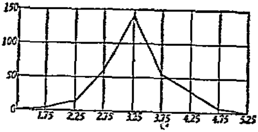


圖9　新生男孩之體重分佈圖

成年男子之體
重分佈表

英　磅	z
90—100	2
100—110	34
110—120	152
120—130	390
130—140	867
140—150	1623
150—160	1559
160—170	1326
170—180	787
180—190	476
190—200	263
200—210	107
210—220	85

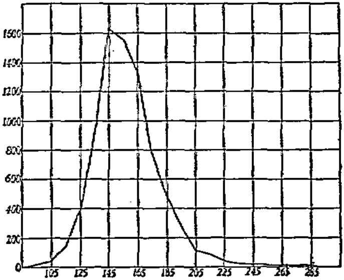


圖10成年男子之體重分佈圖

欲證明分佈不穩與分類方法有關,可取同一集團,依二種分法,所得分佈

220—230	41
230—240	16
240—250	11
250—260	8
260—270	1
270—280	
280—290	1
	7749

之結果,互相比較.下述之例,卽1522粒比國黑豆,依二種不同分類法,所得長度分佈之結果也.

細察二表,易知第一分法所得長度分佈,甚爲對稱.第二分法所得結果,則爲顯著右傾.觀11圖當益明瞭,圖中實

比國黑豆依第一分類方法所得之長度分佈表

耗	z
8.75— 9.75	2
9.75—10.75	43
10.75—11.75	314
11.75—12.75	309
12.75—13.75	316
13.75—14.75	30
14.75—15.75	6
15.75—16.75	2
	1522

比國黑豆依第二分類方法所得之長度分佈表

耗	z
9—10	7
10—11	67
11—12	466
12—13	760
13—14	201
14—15	15
15—16	5
16—17	1
	1522

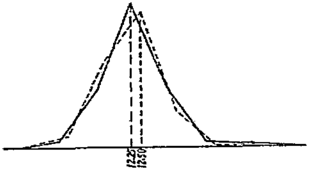


圖11比國黑豆依二種不同分類法所得長度分佈之屢數多邊形

線所示,為由第一方法所得之屢數多邊形,虛線所示為由第二方法所得者.

至於妥善分類之通則,則以屢數最密之判值所在位置能與類幅中點或節點疊合為佳.否則亦須設法移近,但移置之時,宜注意新位置之號碼,毋使數字過繁.

32. 不稱程度,以片面分佈為其極限.片面分佈者,屢數自一極大值,單向一方下降之分佈也.其例在經濟範圍中,至為常見.如收入分佈,收入最少者人數最多.愈豐者,人數愈少.又如都市房屋用價,最低者為數最多,價高則數少.此吾人所習見者.下表所示為1885—1886[1]年英格蘭及威爾士按課稅多寡,所得房屋用價之分佈.表中類幅大小不一,然其變遷則可自z行中數字起落,完全察出.故欲加以比較,則須由行中各值,求出以10£(鎊)為幅之平均屋數,始能如願.第12圖所示,即由平均數所得與正雙曲線枝相近之分佈梯級圖也.

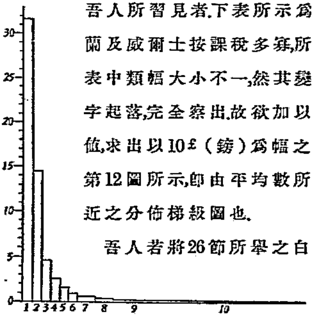


圖 12 房價分佈圖

吾人若將26節所舉之自喉死亡分佈表,以五年及十年為類幅,而改列之,則得一片面

(1) K. Pearson, Phil. Trans. Roy. Soc. (A), Bd. 186 (1896) P. 396.

1885—1886英威房價分佈表

類	X 鎊 £	Z	z 以10£爲幅之平均數
1	0— 10	3,174.806	3,174.806
2	10— 20	1,450.781	1,450.781
3	20— 30	441.595	441.595
4	30— 40	259.756	259.756
5	40— 50	150.968	150.968
6	50— 60	90.432	30.432
7	60— 80	104.128	52.064
8	80—100	47.326	23.663
9	100—150	58.871	11.774
10	150—300	37.988	2.532
11	300—500	8.781	439
12	500—1000	3.002	60
13	1000—1500	1.036	21

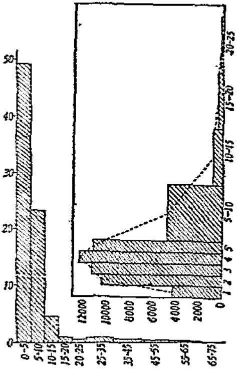
分佈表,及梯級圖13a.然若取原表第三行之年商而圖示之,則得圖13b所示之左傾雙方分佈,

又如取收入及房租分佈表之第一類之類幅,詳加分析,則所得結果,必與自喉死亡之分佈相同,以判值在首類末點以下之諸類,其屢數旣不爲零又非極大也.

33. 統計學中最常見者,雖爲異度不稱,以及對稱之單峯分佈,然此並非集團研究所得唯一之結果也.學者曾得

以五年及十年爲幅
之白喉死亡分佈表

X	Z
0— 5	49.479
5—10	23.348
10—15	4.092
15—20	1.123
20—25	585
25—35	786
35—45	512
45—55	324
55—65	260
65—75	127
75(以上)	35
	80.671



13 a　13 b

圖 13 白喉死亡分佈圖

數種分佈與此頗有出入,茲特示例於下.

栗木複葉,雖係成對分排,然葉梢亦有單葉出現,下表所示爲 8554[1] 片複葉之分佈,其第一列中之 X,爲子葉片數,第二列之 Z 爲複葉片數.

X:	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Z:	8	5	142	75	876	237	2674	527	2947	223	753	26	59	2

(1) K. Pearson, Phil. Trans. Roy. Soc. (A), Bd. 197 (1901)

表中偶 X(1) 與奇 X 之分佈,完全不同,若以幾何圖形表之,則得14圖所示近於對稱之二重分佈.

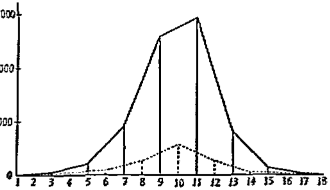


圖14依于葉數目所得栗木複葉分佈圖

Pearson(2) 氏曾得一與模範曲線完全相反之分佈圖形.即其在德國 Breslau 十年間逐日觀察所得形如 U 字之雲度分佈曲線是第15圖所示,即與此相當之度數多邊形也.

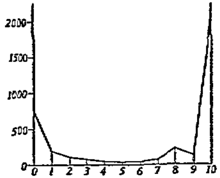


圖15相異雲度分佈圖

X:	0	1	2	3	4	5
Z:	751	179	107	69	46	9
X:	6	7	8	9	10	
Z:	21	71	194	117	2089	

(1) 此處係係未適度之複葉,據 Pearson 觀察所得,即係由末端子葉分裂而成者.

(2) Proc. Rod. Soc. (A) Bd. 197. P. 287.

【附註】Pearson及其弟子以為用分析函數,表各種分佈為生物統計學中最要之問題,遂求出一法,以定函數中之參數. Pearson氏又謂不稱分佈,可由多數均稱分佈疊置而得,蓋通常所用之材料,不必單純,故所得曲線不合規則.若細為區析,則一種純淨材料,必有一常規曲線與之相當.集團既係雜數種材料而成,其分佈曲線自可由多數模範曲線求得.

第二章　中值

34. 以分佈表及屢數曲線描寫集團,固甚詳盡,然尚欠簡明提要.盍取同類二集團而比較其分佈,則其不同之點,固可就曲線形狀,詳為敍述.然欲使全體關係一覽無餘,則實非圖表所能奏效.欲得括示全體關係之標幟,須由各集團中依一定界說,求出一量,以表分佈之特性.取二量相較,則分佈之異點立見.欲求此量,宜察分佈之顯著異點,一為各團判量之不同,二為各團判域之互異.由全團判值,可求出各種中數.所謂中數者,依某種界說,所求得介於最大及最小判值間之數值也.依數學方法,如此之值,多至無數.然吾人為適於應用計,僅取其與屢數曲數特性有關者論之.至區別判域,則須與上下限,其手續上之不便姑置勿論,即其須與兩數,已屬缺點.蓋此種標幟,愈簡愈妙,換言之,亦即以一數為佳也.此種表判域大小之量,曰散布量.

中值及散布量之理論,已如上述.然此二者非僅有理論上之根據,且具有實用之功能,以後當詳論之.

35. 數學上理想中值之多,既如上述,至吾人所欲論者,則爲適合下列條件之諸值.各中值用途之大小,亦卽以其適合條件之程度爲轉移.

適於統計應用之中值,其必具性質如次:

a. 易於領悟.蓋統計學科,重在致用,其所得結果,亦卽以大衆了解爲前提.故數學中定義抽象,觀念難明之中值,槪不適用.

b. 其定義須尖刻明瞭,而無二義.務使讀者能顧名思義,而不必須練習有素,經驗豐富者,始能了解.

c. 此值須與一切觀察結果有關,然後始能代表全體性質.

d. 須有便於數學上運用之優點.如當二個或更多集團之中值已個別算出時,則由此聯合之集團,其中值可由已知中值算出 *ε*.至其算法則愈簡愈妙.

e. 易於計算.中值應用既多,其計算自以簡單爲妙,若過重其他性質,而忽視計算之便利,勢必至繁困難勝,不能應用.此創造中值時,所不可不加之意者.

中值之具上述條件者,爲數寥寥,依其用途及意義之大小而序列之,不過算術中值,必值,密值,幾何中值,及調和中

値而已.

36. 算術中值. 設變量X之特値爲$X_1,X_2\cdots\cdots X_n$,則取諸量之和,以N除之,可得一新數,是爲諸値之算術中數,吾人若以M表之,則得其公式如下:

$$M=\frac{1}{N}\Sigma(X). \tag{1}$$

諸X中,或完全不同,或部分相等,均與定義無害.設其中之某X出見Z次,則此N數中有Z個相等,而上式可書爲

$$M=\frac{\Sigma(ZX)}{\Sigma(Z)}. \tag{2}$$

是爲就分佈初表求算術中値之公式.至於(1)式,則當由原表求算術中値時應用之.

此種中値之算術定義,可以概算觀念說明如次.若$X_1X_2\cdots\cdots X_n$之出現機會相等,則其出現概率各爲$\frac{1}{N}$,今N數中有Z個X,則X出現之概率當爲$\frac{Z}{N}$,故算術中値爲X及其概率乘積之總和.

在間續集團,因其判値皆爲整數,故範圍有限,而應用(2)式之計算亦無困難.但其結果多非整數,故倘須加以說明.

至於範圍甚大之連續集團,而欲應用初表以求M,則計算繁雜殊爲不便,故宜擇一相當分佈表,假定各類之判値,均與其類距中點所有者相同,而後計算之.即謂先化連續

集團爲間組,然後求其算術中值也.其所得結果,固非算術中數之眞值,而爲其近似.然若類幅愈小,則分佈於幅中之各值,愈相接近,而所得之結果亦愈精.吾人若以 x 表類距中點之判值,則得一與(2)式形式相同意義各異之公式

$$M=\frac{\Sigma(zx)}{\Sigma(z)}. \tag{3}$$

37. 應用(3)式以求算術中值,宜若甚爲簡便,然當 x 及 z 之位數過繁,則仍不免困難.故欲此式有應用功效,非將數字設法化簡不可,化簡之法:

a.取原有判值以類距爲單位表之.

b.棄原有判值,而取其與定值之差代之.

吾人若以整數表類距中點,而以附 .5之數,表其節點,則首條所示,不難達到.至於末條,則可於分佈表之中段擇一類距中點爲出發值,命之曰 U. 則其餘中點判值均可以 $U+\varepsilon$ 表示,而得

$$\Sigma(zx)=U\Sigma(z)+\Sigma(z\varepsilon) \tag{4}$$

因之 $$M=U+\frac{\Sigma(z\varepsilon)}{\Sigma(z)}. \tag{5}$$

因表中之 ε 其次序在 U 以上者爲負,以下者爲正,又可互相消去而使計算化簡.

例,右表爲由25節(1)所述之松高分佈以10公分爲幅重加整理所得之新表.表末所附之 M,即依上法求得者.

茲爲闡明化位對於算術中値之影響起見,特取以20公分爲類幅之化表,而舉其計算結果如下.

若以公寸表高度中値,且僅取一位小數,則由:

以10公分爲幅之分佈表,所得之 M=17.7公寸.

以20公分爲幅之第一化表,所得之 M=17.7公寸.

以20公分爲幅之第二化表,所得之 M=16.0公寸.

至於 Johannsen[1] 之所謂分佈化形,則除以正負號冠於屢

算術中値之計算

類距中點	屢數	乖差	積
60	1	−11	11
70	1	−10	10
80	0	− 9	0
90	0	− 8	0
100	1	− 7	7
110	3	− 6	18
120	3	− 5	15
130	5	− 4	20
140	6	− 3	18
150	11	− 2	22
160	10	− 1	10
170	17	0	−131
180	17.5	1	17.5
190	13	2	26
200	13.5	3	40.5
210	6	4	24
220	7	5	35
230	3	6	18
240	2	7	14
250	3	8	24
260	1	9	9
270	1	10	10
	125		+218

(1) Elemente der exakten Erblichkeitslehre, 2. Aufl. (1913), P. 34, 11.

由化表求算術中數

1

x	z	ε	$z\varepsilon$
65	2	−5	10
85	0	−4	0
105	4	−3	12
125	8	−2	16
145	17	−1	17
165	27	0	−55
185	30.5	1	30.5
205	19.5	2	39
225	10	3	30
245	5	4	20
265	2	5	10
	125		+129.5

2

x	z	ε	$z\varepsilon$
55	1	−6	6
75	1	−5	5
95	1	−4	4
115	6	−3	18
135	11	−2	22
155	21	−1	21
175	34.5	0	−76
195	26.5	1	26.5
215	13	2	26
235	5	3	15
255	4	4	16
275	1	5	5
	125		+88.5

數乘差前以外,别無二致,惟此法應用於範圍較小,數目不大之分佈,至爲便利,故特舉例說明之.

例. Skagen(丹麥北海漁場之一)附近,所捕之鰈,有取其尾鰭而數其刺數者,所得 703 尾中之刺數最少爲47,最多爲61,其分佈如下.

尾刺數 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61

鰈魚數 5 2 13 23 58 96 134 127 111 74 37 16 4 2 1

茲依樂氏方法,以53為定值,而求其算術中數.

x		53	54	55	56	57	58	59	60	61	(1)
			52	51	50	49	48	47			
z	+	134	127	111	74	37	16	4	2	1	(2)
	−		96	58	23	13	2	5			
ε		0	1	2	3	4	5	6	7	8	(3)
(2)同行差	+	134	31	53	51	24	14		2	1	(4)
	−							1			
(3)(4)同行積	+	0	31	106	153	96	70		14	8	(5)
	−							6			

取(5)式諸值之和 $478-6=472$ 得 $M=53+\frac{472}{703}=53\cdot67$.

若鰈魚尾刺數目相同,則最少者當得53枚,最多者不過54刺,而在703尾中應有具53刺者 232 尾,具54刺者 471 尾.斯卽 M 不為整數之意義也.

38. 上述算法,實為求和法之基礎.求和法者,對於統計研究應用最便之法也.以其純係加減手續故名.欲研究之,先列分佈表之通式如下,此表依計算目的,僅三欄已足.表中之 a 為首類前一空類中點之判值.

因此法為繼續求和,故在長分佈表稍數繁冗.欲免此弊,

可於表之中段取定值U爲出發點,然後自上下兩端向中間算去,則所得結果,異常簡單,玆舉其算法於下:

應用求和法之基表法式

類距中點	類屢數	對U之乘值
$a+1$	z_1	$-(k-1)$
$a+2$	z_2	$-(k-2)$
$a+3$	z_3	$-(k-3)$
…	…	…
…	…	…
…	…	…
$a+(k-2)$	z_{k-2}	-2
$a+(k-1)$	z_{k-1}	-1
$U=a+k$	z_k	0
$a+(k+1)$	z_{k+1}	1
$a+(k+2)$	z_{k+2}	2
…	…	…
…	…	…
…	…	…
$a+(n-2)$	z_{n-2}	$n-k-2$
$a+(n-1)$	z_{n-1}	$n-k-1$
$a+n$	z_n	$n-k$
	$N=\Sigma(z)$	

$$(6)\begin{cases} s_1=z_1 \\ s_2=z_1+z_2 \\ s_3=z_1+z_2+z_3 \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ s_{k-2}=z_1+z_2+z_3 \\ +\cdots+z_{k-2} \end{cases}$$

$$(7)\begin{cases} s_n=z_n \\ s_{n-1}=z_n+z_{n-1} \\ s_{n-2}=z_n+z_{n-1} \\ +z_{n-2} \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ s_{k+2}=z_n+z_{n-1} \\ +z_{n-2}+\cdots+z_{k+2} \end{cases}$$

$$(8)\quad s_{k-2}+z_{k-1}=S_0^-$$

$$(9)\quad s_{k+2}+z_{k+1}=S_0^+$$

由(6),(7)求和得

$$s_1+s_2+\cdots+s_{k-2}=(k-2)z_1+(k-3)z_2+\cdots+z_{k-2}$$
$$=(k-1-1)z_1+(k-2-1)z_2+\cdots+(2-1)z_{k-2}$$
$$=-\sum_1^{k-2}(z\varepsilon)-s_{k-2}=S_1^-; \tag{10}$$

$$s_n+s_{n-1}+s_{n-2}+\cdots+s_{k+2}=(n-k-1)z_n+(n-k-2)z_{n-1}+\cdots+z_{k+2}$$
$$=(n-k-1)z_n+(n-k-1-1)z_{n-1}+\cdots+(2-1)z_{k+2}$$
$$=\sum_n^{k+2}(z\varepsilon)-s_{k+2}+S_1^+. \tag{11}$$

令 $$\Sigma(z\varepsilon)=\sum_1^{k-2}(z\varepsilon)-z_{k-1}+z_{k+1}+\sum_{k+2}^{n}(z\varepsilon)$$
$$=-(S_1^-+s_{k-2})-z_{k-1}+z_{k+1}+S_1^++s_{k+2}$$
$$=S_1^+-S_1^-+S_0^+-S_0^-, \tag{12}$$

若命

$$S_0^+-S_0^-=\Delta_0$$
$$S_1^+-S_1^-=\Delta_1 \tag{13}$$

則得 $$\Sigma(z\varepsilon)=\Delta_1+\Delta_0 \tag{12a}$$

設 $$\frac{\Sigma(z\varepsilon)}{\Sigma(z)}=\eta$$

則由(5)式得算術中值

$$M=U+\eta. \tag{14}$$

若依(6)式進求 s_{k-1} 及 s_k，依(7)式進求 s_{k+1}，則因 $s_k+s_{k+1}=$

$\Sigma(z)$得計算結果之驗算法.

又因

$$s_{k-1}=S_0^- \qquad s_{k+1}=S_0^+$$

咸爲已知,故僅須求出S_1^-及S_1^+卽可計算算術中值.

39. 例(1).玆取25節例1所述之幼松高度依基表式樣作分佈表,並於主要數字之旁,註明計算符號,以醒眉目.惟此種符號,若稍加棘習,卽不難諳誦,故於後舉各列,一概從略.

依求和法求算術中數

公分	z	s	
60	1	1	
70	1	2	
80	0	2	
90	0	2	
100	1	3	
110	3	6	
120	3	9	
130	5	14	
140	6	20	
150	11	31	
160	10	41	(S_0^-) 90
			(S_1^-)
170	17	58	

成年男子體重之算術中數

英磅	z	s	
95	2	2	
105	34	36	
115	152	188	
125	390	578	
135	867	1445	
145	1623	3068	2249
155	1559	4627	
165	1326	3122	3921
175	787	1796	
185	476	1009	
195	263	533	

180	17.5	67 (S_0^+) 151
190	13	49.5　(S_1^+)
200	13.5	36.5
210	6	23
220	7	17
230	3	10
240	2	7
250	3	5
260	1	2
270	1	1
	125	

205	170	270
215	85	163
225	41	78
235	16	37
245	11	21
255	8	10
265	1	2
275		1
285	1	1
	7749	

驗算　$58+67=125$　　　驗算 $4627+3122=7749$

$\Delta_0=67-41=26$　　　$\Delta_0=3122-3068=54$

$\Delta_1=151-90=61$　　　$\Delta_1=3921-2249=1672$

$\eta=\frac{26+61}{125}\cdot 10=6.9$　　　$\eta=\frac{54+1672}{7749}\cdot 10=2.23$

$M=170+6.9=176.9$公分　$M=155+2.23=157.23$磅

每一英磅(Troy)$==373.2419g$.

η 原係比值故無單位可言,若欲以原定單位表之,則宜取其與類幅長度之積,如在上例,類幅爲10公分故以類幅單位計算,η 應有之値爲0.69×10公分.

(2)成年男子之體重(見31節例3),其中位計算可由上表察出,無庸具載.

(3)下述二例為關於間續集團者.(a)鰈之尾鰭刺數,(b)木藍莢中種子數[1].前者之範圍為703,後者為178,其算法以(a)之最小判值,(b)之最大判值為出發點,故(a)表僅具基表下部,(b)表僅具基表上部.

鰈尾鰭刺之算術中值

x	z	s	
47	5	703	
48	2	698	3992
49	13	696	
50	23	683	
51	58	660	
52	96	602	
53	134	506	
54	127	372	
55	111	245	
56	74	134	
57	37	60	
58	16	23	
59	4	7	
60	2	3	
61	1	1	
	703		

$\Delta_0=698,\ \Delta_1=3992$

$$\eta=\frac{698+3992}{703}=6.7$$

$M=47+6.7=53.7$

木藍每莢種子粒數之算術中值

x	z	s	
3	1	1	
4	2	3	
5	8	11	
6	13	24	
7	22	46	
8	45	91	
9	63	154	
10	23	177	330
11	1	178	
	178		

$\Delta_0=0-177,\ \Delta_1=0-330$

$$\eta=-\frac{177+330}{178}=-2.9$$

$M=11-2.9=8.1$

(1) F. Ludwig, Botanisches Zentralblatt, 卷 73 (1898), P. 343.

又 s 之驗算值,在(a)居該欄之首,在(b)居該欄之末,至於 M 不爲整數之解釋,則已於37節例 2 述之矣.

40. 算術中值對於中值應具之條件,均能完全滿足.故其在統計學中所居地位亦最優越.吾人若名之爲平均值,則其意義更易明瞭,如云每日之平均工資,易知其爲以日數除逐日工資總和所得之值是.由日數及平均工資,不難求得工資總和.

至算術中值之運算優點,更可述之如次.

a. 在(5)式中若 $U=M$, 則 $\Sigma(z\varepsilon)=0$, 亦卽謂一切判價,對於算術中值之乖差總和爲零也.

b. 設諸集團之範圍爲 N_1, N_2……則其聯合集團之範圍,爲 $N=N_1+N_2+$…….吾人若以 X_1, X_2, ……;及 X 表相當各集團之判值,則得

$$\Sigma X=\Sigma X_1+\Sigma X_2+\cdots\cdots.$$

或
$$N\frac{\Sigma X}{N}=N_1\frac{\Sigma X_1}{N_1}+N_2\frac{\Sigma X_2}{N_2}+\cdots\cdots$$

亦卽
$$NM=N_1M_1+N_2M_2+\cdots\cdots. \tag{15}$$

是爲算術中值之關係式.若式中 M_1 爲已知,則由此可直接推出總團之算術中值,而不必詳細計算.故在 M 爲已知之集團,欲求其擴充以後之算術中值,則僅須求出擴充部分之 M 代入(15)卽得.

c. 若變量 X 爲變量 X_1 及 X_2 之和(或差), 則 X 之算

術中值,等於X_1及X_2之算術中值和(或差).

設關於X_1, X_2之分佈表,其範圍爲N_1及N_2,則每取X_1之一值與X_2之一值配合,可得一值X,因二者之配合方法爲N_1N_2故X之值有N_1N_2個,求其總和有

$$\Sigma X = N_2\Sigma X_1 \pm N_1\Sigma X_2,$$

以N_1N_2除之得

$$M = M_1 \pm M_2. \qquad (16)$$

此式不難推及於多數變量之代數和.

有以上各式則實行統計時,範圍較大之集團,可分作數部研究,而能收分工合作之效.例如統計生物之集種器官,其集團亦可分作數部,依其分佈結果,求出個別之算術中值,以推出其總團之算術中數.

d.算術中值之顯著關係

$$\Sigma(z\varepsilon) = 0,$$

實示M在圖形中所取位置.蓋式中z表各類面積,ε爲各類中點垂線與算術中值垂線之距離.故$z\varepsilon$表各類面積對於算術中值垂線之靜力旋勢.$\Sigma(z\varepsilon)=0$者,卽靜力旋勢之總和爲零,亦卽謂算術中值垂線通過全面積之重心也.

41. 心值. 心值者,分曾經整理之範圍爲二等分之判值也.因判值在其上者之屢數和,與在其下者之屢數和相

等,故曰必值,通常以 C 表之.

取集團之判值,依序排列,其數次出現者,亦數次列之,則當集團範圍爲 $2n+1$ 時,序中第 $n+1$ 判數,即爲必值,且係一義確定.以其前後各有 n 判值,而位於 $n+1$ 之值,僅有一個也.若在範圍爲 $2n$ 之集團,則第 n 及 $n+1$ 判值間各數,均合必值條件,而 C 論爲不定,學者爲應用便利計,特取二點中心之值表之.

間續集團,C 值存在,僅當其範圍爲 $2n+1$ 時有可能性,若此條件能達,且居第 $n+1$ 位之判值其屢數爲 1,則此值方爲必值,以此時其上下各有 n 判值也.然此種特別條件,通常集團萬難達到,故依必值嚴格意義,此種中值,對於間續集團不能應用.

至於求連續集團之必值,則可自分佈表首端或末端數至屢數和與 $\frac{N}{2}$ 相等,或最近之節點.若此點上下之屢數和雙方相等,則該點即必值所在,否則 C 點必居次一類幅,而依屢數比例兩分其距.又行求和法所用之分佈表,極適於求 C 值之用.

例: 下述各例,均36節中已引用者,其分佈表,亦詳該節.

(1)鰈魚尾鰭刺數. 因集團範圍爲 703, 故必值上下之屢數和各爲 351, 今判值在54以上者其屢數和爲 245, 以下者爲 331, 故按 C 值定義,無存在可能.

(2)木藍莢中種子粒數. 集團範圍爲178,其半爲89然判值之在8以下者僅46,超過8者占87,故C值不能存在.

(3)幼松高度之C值算法. 判值爲175公分及在175公分以下之屢數總和爲58,範圍之半數爲62.5.故C值應在類幅175—185中,又因此類之屢數爲17.5,故屢數62.5−58=4.5應得之長度爲

$$\frac{62.5-58}{17.5}\cdot 10=2.57\text{公分},$$

由是得 $C-175+2.57=177.57$公分.

(4)成年男子之體重. 由表得範圍之半爲3874.5,至節點150之屢數和爲3068,故C值應在類幅150—160中,由計算得其與150之差爲

$$\frac{3874.5-3068}{1559}\cdot 10=5.17,$$

故 $C=150+5.17=155.17$英磅.

由表末起,計算C值,其法如次.至節點160之屢數和爲3122.故C應在160—150中.

又因 $$\frac{3874.6-3122}{1559}\cdot 10=4.83$$

故 $C=160-4.83=155.17$英磅.

42. 必值對於中值應具條件,不能完全滿足,其所以引用之者,純以其容易了解,不難計算之故.其定法雖與全體觀察有關,然僅重判值位置,而棄數量本身,故雖以下判值任意減小,以上判值任意增大,而C值可毫無變動.以其發

化欠靈之故,其應用價值頗爲有限.又算術中值,所有運算優點,心值皆付闕如.故心值爲已知之集團,其聯合後,所有心值既須重新計算而不能由原值推出,二變量和(或差)之心值與其個別所有者復乏相當關係,而不能以公式表明.至其最大缺點,則爲求 C 值時所在類幅,其屢數爲平均分佈之假設,此種過欠嚴密之計算,尤缺科學之意義.

心值與圖形之關係至爲顯著,以其係平分全面積之垂線足點,故當屢數曲線完全對稱時,與算術中值相合.若在不稱分佈則曲線不稱愈甚,二值距離愈遠,故 C 值有判决不稱之功用.

所謂不稱曲線者,與底軸平行,作諸線與之相割,所得割點之相當經線,其左右距離有同一不等方向之曲線也[1].由是可得算術中值與心值之大小關係.圖16中 C 爲心值,其經線分全面積爲二等分.設 σ' 及 σ'' 爲二部分面積之重心,則全面積之重心 Σ 平分直線 $\sigma'\sigma''$,然 Σ 之垂足卽 M 所在,故

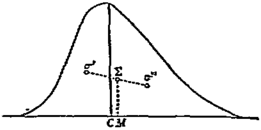


圖16 算術中值及心值之大小關係

當左傾成立,則 σ' 距 C 經較 σ'' 爲近,Σ 居 C 經之右,而 $M>C$.

(1) H. E. Timerding Analyse des Zufalls P. 80 (1915)

反之若右傾成立,則 $M<C$.

43. 密值 密值者,集團中出現次數最多之判值也.

因間續集團之類屢數,實示相當判值之出現次數,故知分佈表中,唯一最大屢數之判值,即爲密值.

若在連續集團,則分佈表中之最大類屢數,僅能示吾人以密值所在之類幅,欲進求此值,則須假定幅中堆積點之存在,堆積點者,屢數集中之點,即密值所在地也.若不與節點偶合,恆在屢數最多之類幅中.

就分佈表以求密值,其最簡者,爲表中僅有一屢數最大之類幅.若所與爲初分佈表,且集團之範圍甚大,則以其中屢數之起落無常,大值之出現不一,密值不易察出,故須以分類法,化去其起伏,使全表僅有一最大屢數,倘仍不能達到,則須以化位法濟之.然依前章所述材料不純,實可引起屢數起落,故多峯分佈,亦可謂由材料混雜所致.如取二種密值不同之材料,混而爲一,則所得分佈必至起落無常,由是可知行化位法,不可過甚,否則必失分佈眞象,而貽乖誤之譏.

由初分佈表求密值,若該表不能顯示此值所在,則除依上述方法行之外,亦可由別法達到,其法以某類及其前 v 類其後 v 類共 $2v+1$ 類之屢數和,或算術中值,代本類屢數,其首尾兩類則以前後 v 空類足之,以消去屢數之無常起

落,Th. Wittstein之統計表均消法,蓋取$2v+1$類之算術中値,以代本類屢數而消去表中起伏者.然此法多非$v=1,2$時所能奏效,且全係人爲法則,故不如前法之美.

吾人旣依前述方法,消去表中起伏,而得唯一最大屢數,即可進求密値所在位置,使其不獨與相鄰各類有關,且能表全體分佈特性.欲解決此項問題,可假定理想分佈曲線之存在,然後依觀察結果,以決定其解析方程式.Pearson氏曾依此種觀念,求出決定數種實際分佈曲線之方法.設曲線方程式爲已知,則可決定其極大値,以算出該値之緯量,而密値遂定.然一種分佈,旣有多種曲線與之套合,而所謂最精配合,復非一義確定.故所得結果,仍不能謂之精確無誤.

由是觀之,密値之易於了解,洵有之矣,至謂其一義確定則猶未也.

欲研究曲線之決定方法,宜先明密値本身之根本意義.所謂密値者,以事實言,爲集團棣元向此集中之特値,故得之可明材料之傾向.以其能顯示分佈狀態,故由此可察出材料之態度.以形式言,密値實分分佈曲線爲左右二枝,左枝漸昇,右枝迭降,故在Fechner統計學中此値最爲重要.以其可爲曲線各枝分別研究之公共出發値也.Fechner欲求密値之精確決定,費盡心血,曾得方法二種,一曰經驗法,二

曰依比法.然同一分佈,用此二法所得結果,既時相乖異,而孰是孰非,復無精密方法可以判定,故二法優劣無從決斷.

至吾人所欲論者,爲(一)僅有一最大屢數,(二)有二相鄰相等最大屢數.後者固可改變類距代爲簡單,然亦可保存舊有分類法而以他法取之.

44. 通常精確足用之密值定法:先擇表中類屢數爲最大之相連三類,以抛物線

$$y=\alpha x^2+\beta x+\gamma$$

表其屢數曲線,更取屢數最大之類幅中點爲原點,類幅爲單位,則得三類四節點之緯量爲 $-\frac{3}{2}, -\frac{1}{2}, \frac{1}{2}, \frac{3}{2}$(圖

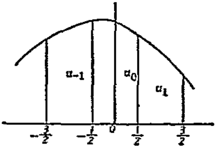


圖17密值定法

17).吾人若自左而右依次以 u_{-1}, u_0, u_1 表三類屢數,則因諸值與以類幅爲底,相當屢數爲高之諸面積相等,故有

$$u_{-1}=\int_{-\frac{3}{2}}^{-\frac{1}{2}} y\,dx=\frac{13\alpha}{12}-\beta+\gamma,$$

$$u_0=\int_{-\frac{1}{2}}^{\frac{1}{2}} y\,dx=\frac{\alpha}{12}+\gamma,$$

$$u_1=\int_{\frac{1}{2}}^{\frac{3}{2}} y\,dx=\frac{13\alpha}{12}+\beta+\gamma;$$

由是得原式之係數關係式為:

$$\Delta u_{-1}=u_0-u_{-1}=-\alpha+\beta$$

$$\Delta u_0=u_1-u_0=\alpha+\beta$$

$$\Delta^2 u_{-1}=\Delta u_0-\Delta u_{-1}=2\alpha.$$

又由原式求得 y 為極大之條件為

$$0=2\alpha x+\beta$$

或 $$x=-\frac{\beta}{2\alpha},$$

因得中值與中類節點 $-\frac{1}{2}$ 之距離為

$$x+\frac{1}{2}=\frac{\alpha-\beta}{2\alpha}=-\frac{\Delta u_{-1}}{\Delta^2 u_{-1}}. \tag{17}$$

上述結果係以類幅 l 為單位計算而得者,故欲以原有單位表之,則須以類幅乘此結果設 X 為節點 $-\frac{1}{2}$ 之料值,則得

$$D=X-\frac{\Delta u_{-1}}{\Delta^2 u_{-1}}l.$$

例 取25節例(1)所舉幼松高度分佈表,以10公分為幅而再列之,所得密值算式如下.

$u_{-1}=17$ $\Delta u_{-1}=0.5$ $\Delta^2 u_{-1}=-5.0$ $-\frac{\Delta u_{-1}}{\Delta^2 u_{-1}}\cdot 10=1$

$u_0=17.5$ $\Delta u_0=-4.5$

$u_1=13$ $D=175+1=17.6$ 公分.

實際計算,普通符號均可略去,僅書其所表之數已足,讀者自習數例當知其毫無困難也.

Leipzig學生體高之中值M,C,D.

x	z	s	
59.5	0.5	0.5	
60.5	0.5	1	
61.5	0	1	
62.5	0	1	
63.5	1	2	
64.5	8	10	
65.5	20	30	
66.5	415	71.5	
67.5	72	143.5	
68.5	137	280.5	
69.5	215.5	496	
70.5	271	767	1037
71.5	323.5	1090.5	
72.5	305	956.5	1376.5
73.5	274.5	651.5	
74.5	183.5	377	
75.5	101.5	193.5	
76.5	52	92	
77.5	27.5	40	
78.5	7	12.5	
79.5	3	5.5	
80.5	1.5	2.5	
81.5	0	1	
82.5	1	1	
	2047		

(2) 左表所表爲1843至1862[1]年德國Leipzig大學體格檢查所得廿歲新生2047名之高度,表中x係以Sachsen舊寸爲單位者,每寸等於23.6公釐,茲依前述各法求出其三種中值如下:

$$\begin{array}{rr} 956.5 & 1376.5 \\ -767 & -1037 \\ \hline 189.5 + & 339.5 \end{array}$$

$$529 : 2047 = 0.26$$

$$M = 71.5 + 0.26 = 71.76;$$

$$\begin{array}{r} 1023.5 \\ -767 \\ \hline 256.5 \end{array} : 323.5 = 0.79$$

$$C = 71 + 0.79 = 71.79;$$

$$\begin{array}{lll} 271 & 52.5 & -71.0 \\ 323.5 & -18.5 & \\ 305 & & \end{array}$$

(1) Th. Fechner Kollektivma Blehre P. 136 及 137.

$$52.5:71.0=0.74$$

$$D=71+0.74=71.74.$$

依計算結果,三種中種咸在一類,且甚爲接近,故可謂爲均稱分佈.

411名巴什旗人體高分佈[1]

公分	x	和	算術中數
148—151	1	7	2.7
151—154	6	39	13
154—157	32	110	36.7
157—160	72	182	60.7
160—163	78	220	73.3
163—166	70	226	75.3
166—169	78	187	62.7
169—172	39	136	45.3
172—175	19	68	22.7
175—178	10	33	11
178—181	4	16	5.3
181—184	2	6	2
	411		410.7

(3)求密值時,須注意之點已於43節述及,茲爲讀者易於明瞭起見,特示例如下.右表所示,爲粤國所俘411名巴什旗人[2]體高之分佈,表中峯值有二,吾人若假定其爲經意外擾亂之單峯分佈,則依前述方法,取相連三類屢數之和,或算術中值,以代中類屢數,而消去其起伏,表末二欄,即其均消結果也.欲求密值,二種結果,均可應用,惟算術平均數之總和,較集團範圍爲小,尙須留意.

(1) Baschkiren土耳其游牧民族現組巴什旗自由國都Ufa,爲赤俄聯邦之一.

(2) 見R. Pecch.奥國俘虜所人體測量記錄

441名巴什族人之體高分佈

東巴什族人		西巴什族人	
x	z	x	z
150—153	3	146—149	1
153—156	5	149—152	1
156—159	14	152—155	10
159—162	13	155—158	34
162—165	18	158—161	67
165—168	13	161—164	60
168—171	8	164—167	60
171—174	2	167—170	57
174—177	2	170—173	23
177—180	1	173—176	13
	79	176—179	6
		179—182	3
		182—185	1
			342

茲依末欄結果計算密值如下:

73.3　2　−14.6

75.3　−12.6

62.7

$$D=163+\frac{2}{14.6}$$

$$=163.14\text{公分}$$

上述方法,有掩蔽分佈真像可能.不能謂爲盡善,因按原表依法計算所得之二密值,一爲 160.43,一爲 166.17,相差甚大故也.探諸實情,巴什旗種,實分東西二族,細爲區析,可得上表,蓋表之前十類蓋除西巴什旗人外,更益以東巴什旗人者.由此二表可得相差頗大之二密值163.5及160.46.惟上表因範圍太小,致末欄數字變化尙缺常規.此種密值求法,是否可用,殊屬疑問.

45. 若分佈表中有二相鄰相等之最大屢數,則求密值時,上法不能適用,而須以他法易之.其法除屢數爲極大之

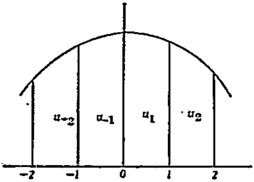


圖18密值之定法

二類外,更取其旁兩類之屢數而以三次拋線

$$y=\alpha x^3+\beta x^2+\gamma x+\delta$$

表其屢數曲線,則依18圖所示有

$$u_{-2}=\int_{-2}^{-1}ydx=-\frac{15}{4}\alpha+\frac{7}{3}\beta-\frac{3}{2}\gamma+\delta$$

$$u_{-1}=\int_{-1}^{0}ydx=-\frac{1}{4}\alpha+\frac{1}{3}\beta-\frac{1}{2}\gamma+\delta$$

$$u_1=\int_{0}^{1}ydx=\frac{1}{4}\alpha+\frac{1}{3}\beta+\frac{1}{2}\gamma+\delta$$

$$u_2=\int_{1}^{2}ydx=\frac{15}{4}\alpha+\frac{7}{3}\beta+\frac{3}{2}\gamma+\delta.$$

因 $u_{-1}=u_1$ 故得

$$\begin{aligned}
\alpha&=\frac{u_2-u_{-2}}{6}-\frac{u_1-u_{-1}}{2}=\frac{u_2-u_{-2}}{6}\\
\beta&=\frac{u_2+u_{-2}}{4}-\frac{u_1+u_{-1}}{4}=\frac{u_2+u_{-2}}{4}-\frac{u_1}{2}\qquad(19)\\
\gamma&=\frac{5(u_1-u_{-1})}{4}-\frac{u_2-u_{-2}}{12}=-\frac{u_2-u_{-2}}{12}\\
\delta&=\frac{7(u_1+u_{-1})}{12}-\frac{u_2+u_{-2}}{12}=\frac{7u_1}{6}-\frac{u_2+u_{-2}}{12}.
\end{aligned}$$

又 y 爲極大之條件爲

$$0=3\alpha x^2+2\beta x+\gamma,$$

因得

$$x=\frac{-\beta\pm\sqrt{\beta^2-3\alpha\gamma}}{3\alpha}. \tag{20}$$

欲判別 x 之二値,孰爲合理,可取由此算出之 D,察其是否在屢數爲極大之二類中,而後決斷.以依密値意義,D 須在此二類中也.

例 下列分佈爲 450 名歐洲男子之頭蓋直立弧距,[1] 蓋 Fechner 取 Welcker 測量結果整理而得者.其第一表分佈與上述情形吻合,故舉出而計算其密値.

歐洲男子之頭蓋直立弧距

(a)

X 公釐	Z
365.5—370.5	1
370.5—375.5	2
375.5—380.5	5
380.5—385.5	17
385.5—390.5	24

(b)

X 公釐	Z
367.5—372.5	3
372.5—377.5	1
377.5—382.5	7
382.5—387.5	22
387.5—392.5	30

(1) 譯者按人種學中,頭蓋測量,有所謂 (a) 直立弧距,(b) 平經周圍,及 (c) 橫拱弧距者,其定義如次:

(a) 自鼻根經頭頂而達於後腦小突起之弧距.

(b) 過額下隆起及後腦拱骨之頭蓋平立周圍.

(c) 自右耳根至左耳根橫過頭頂之弧距.

390.5—395.5	36	392.5—397.5	33
395.5—400.5	41	397.5—402.5	55
400.5—405.5	59	402.5—407.5	50
405.5—410.5	65	407.5—412.5	73
410.5—415.5	65	412.5—417.5	52
415.5—420.5	51	417.5—422.5	55
420.5—425.5	40	422.5—427.5	35
425.5—430.5	17	427.5—432.5	12
430.5—435.5	19	432.5—437.5	14
435.5—440.5	4	437.5—442.5	5
440.5—445.5	2	442.5—447.5	2
445.5—450.5	2	447.5—452.5	1
	450		450

由 $u_{-2}=59, u_{-1}=65, u_1=65, u_2=51$ 得

$$\alpha=-\frac{4}{3}, \beta=-5, \quad \gamma=\frac{2}{3},$$

$$x=-\frac{5\pm\sqrt{27\frac{2}{3}}}{4}=\begin{cases}-2.56\text{類幅}\\ 0.065\text{類幅}=0.33\text{公釐}.\end{cases}$$

上述二值,僅後者爲合理,以其在屢數最大之類中,而前者則在其外也.由是得

$$D=410.5+0.33=410.83\text{公釐}.$$

吾人若取(*a*)表移換節點而重列之,則得僅有一極大屢數之(*b*)表,由是算得之密值爲:

50	23	−44
73	−21	
52		

$$D=407.5+\frac{23}{44}\cdot 5=410.11\text{ 公釐.}$$

與由(a)表所得之D值相差頗大,蓋範圍太小,有以致之也

至在範圍較大之集團,則密值所受分類之影響甚微,今取44節所述 Leipzig 大學新生體高分佈,用四種不同分類法,重列而求其密值.下表所示即入算須用之三類屢數及密值算法.

I

沙寸	z		
70−71	371	52.5	−71.0
71−72	323.5	−18.5	
72−73	305		

$$D=71+\frac{52.5}{71}=71.74$$

II

沙寸	z		
70.25−71.25	280	47	−70
71.25−72.25	327	−23	
72.25−73.25	304		

$$D=71.25+\frac{47}{70}=71.92$$

III

沙寸	z		
70.5−71.5	290	40.5	−75.0
71.5−72.5	330.5	−34.5	

IV

沙寸	z		
70.75−71.75	309	9	−41.5
71.75−72.75	318	−32.5	

72.5−73.5　286　　　　72.75−73.75　285.5

$$D=71.5+\frac{40.5}{75}=72.04 \qquad D=71.75+\frac{9}{41.5}=71.70$$

上述四密值其最大相差不過 72.04−71.74=0.30 吋,由此可見密值與範圍之關係矣.

46. 算術中值及心值,在不稱分佈中之位次前巳言及.茲更就密值與二者之關係論之.觀19圖知左傾分佈之峯點,居面積之左半部,因之 $C>D$, 又由42節所述知此時 $M>C$, 故有

$$D<C<M,$$

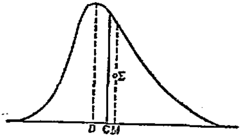


圖 19 M,C,D, 之位置及其關係

至在右傾分佈,則此種大小關係完全顛倒,而變爲

$$M<C<D.$$

例(1)右表所示,爲來英國某醫院就診之腸熱症患者 8689 人按年齡之分佈.其左傾至爲顯著,故適於說明第一情形之用.

腸熱病依年齡之分佈

X 年	z	s	
0−5	266	266	
5−10	1143	1409	266
10−15	2019	3428	
15−20	1955	5261	7656
20−25	1319	3306	
25−30	857	1987	
30−35	503	1130	

$$\begin{array}{r} 5261 \\ -1409 \\ \hline 3852 \end{array} + \begin{array}{r} 7656 \\ -266 \\ \hline 7390 \end{array} = 11242$$

$M=12.5=\frac{11242}{8689}\cdot 5=18.97$

$4344.5-3428=916.5$

$C=15+\frac{916.5}{1955}\cdot 5=17.34$

1143　　876　　−940

2019　　− 64

1955

$D=10+\frac{876}{940}\cdot 5=1464.$

35—40	299	627
40—45	163	328
45—50	98	165
50—55	40	67
55—60	14	27
60—65	8	13
65—70	4	5
70—75	1	1
	8689	

取三值比較知 $D<C<M$ 且相差頗大.

(2) 1911 年維陰（Wien 奥京）食畜市依活重售出之牲畜,按每百公斤 (kg) 價格分類得下列右傾分佈.

$M=125-\frac{227752+514670}{228403}\cdot 10=92.94$

$C=90+\frac{114201.5-84297}{57541}\cdot 10=95.20$

57541　2404　−36380

59945　−33976

25969

$D=100+\frac{3404}{36380}\cdot 10=100.66.$

依活重出售之畜數及每百公斤之價格

奥幣	z	s
30—40	425	425
40—50	2208	2633
50—60	6274	8907
60—70	17627	26534
70—80	21719	48253
80—90	36044	84297
90—100	57541	141838

100—110	59945	201783	
110—120	25969	227752	514670
120—130	651	228403	
	228403		

故$M<C<D$,其實際大小關係與理論結果全同.

M,C,D,三值之差$C-M$,$M-D$,$D-C$,固視全體分佈如何而異其值,然其與不稱程度之關係則尤為密切.又三差之比率,可表諸值之相對位置,以與全體分佈有關,故其正負大小與時俱異.然其變化通常已不若絕對差之甚,至在相似分佈,則尤與常數相近,Fechner曾假定不稱分佈曲線為二段模範曲線合成,而詳究三中值之位置關係,然其所得之位置定律,則實際意義殊少.

在不稱度不甚大之分佈,必值約分算術中值與密值間之距離為1∶2,因之$|M-C|$與$\frac{1}{2}|M-D|$(1)相近.其結果離此太遠者或以為不獨不能屬於常規分佈之列,且在不稱分佈範圍之外,故中值關係,可由此而評騭之.

因密值決定與分類有關,故其結果在三中值中為最不可靠,然若依上述觀念,按三者位置關係由算術中值,及必值以求之,似不失為較良之法.惟此種觀念與事實不侔,故為用殊少,讀者細閱下例當知其然也.

---

(1) 此種關係蓋Pearson依羃數曲線$y=y_0x^re^{-rx}$求得者,此曲線之x軸以M為原點. Phil. Trans. A, Bd.186 (1895), P. 375.

(1)以後將於52節詳述之美國新兵體高分佈,其中值爲:

$$M=66.701 \quad C=66.651 \quad D=66.422\text{英寸}$$

依中值大小關係爲左傾分佈,又

$$M-C=0.050, \quad M-D=0.279,$$

而$\frac{1}{3}(M-D)=0.093$故$M-C$介於$\frac{1}{6}(M-D)$與$\frac{1}{5}(M-D)$之間.

(2)由本節第一分佈表,得腸熱症分佈中值爲

$$M=18.97 \quad C=17.34 \quad D=14.64\text{齡}.$$

依其大小次序,亦爲左傾分佈,又因$M-C=1.63$ $M-D=4.33$其比值爲$1:2.66$與$1:3$相近,可與規則強合.

(3)由本節第二分佈表,得食畜活重價之中值爲:

$$M=92.49 \qquad C=95.20 \quad D=100.66\text{奧元},$$

又因 $C-M=2.71$ $D-M=877$而$3(C-M)=8.13$可謂與規則相近.

47, 幾何中值. 諸值$X_1, X_2, \cdots\cdots, X_n$之幾何中值,爲諸數乘積之$n$次純對根,若以$G$表之則有

$$G=(X_1X_2\cdots\cdots X_n)^{\frac{1}{n}} \tag{21}$$

在上式中若有值之一爲負或零,幾何中值即失去其意義.故G不必如前述三中值對於任何分佈均能應用.

除上述特別情形外,幾何中值皆爲一義確定,至其計算

則遇 n 大於 3 時,若用對數法卽 $n=2,3$ 亦以應用此法爲便.由是得幾何中值與算術中值之關係,諸 X 幾何中值之對數等於諸 X 對數和之算術中值.

$$\log G=\frac{1}{n}(\log X_1+\log X_2+\cdots\cdots+\log X_n).\quad(22)$$

因之幾何中值與算術中值有共同代數優點;如聯合集團之幾何中值,與分團幾何中值,有與40節(15)式相似之關係是

$$N\log G=N_1\log G_1+N_2\log G_2+\cdots\cdots.\qquad(23)$$

諸 X_i 之分佈,對算術中值爲對稱時, $\log X_i$ 對 $\log G$ 不必有同樣性質.

茲設 $\log X_i$ 之分佈,對 $\log G$ 爲對稱,而進考其相當 X_i 對於 G 之分佈.則因在反對方向與 $\log G$ 等距之 $\log X$,其二值 $\log X'$ $\log X''$ 有

$$\log X'-\log G=-(\log X''-\log G)$$

之關係.故 $\log X$ 二值之眞數與 $\frac{X}{G}$ 之二值相當.且互爲倒數.若 $\frac{X'}{G}=q,\frac{X''}{G}=\frac{1}{q}$, 則得相等經線與 G 經距離互爲倒數之不稱分佈,而以 G 爲其密值.此種分佈是否實際存在,則吾人不敢斷言,更無從細辨.

因幾何中值,純係抽象之數學中數,缺乏淺顯意義,加以計算繁冗,不易求得,故爲用殊少.

乘積元判值,與中值之差,而取其與某中值之比,以研究其分佈之變化,此幾何中值之所由與也.若不論此種中值為誰,而以G表之,則吾人之研究對象,幾為諸商

$$\frac{X_1}{G}, \frac{X_2}{G}, \ldots\ldots \frac{X_n}{G}$$

之分佈.因上列比值或大於一,或小於一,而諸值對數$\delta_1, \delta_2, \ldots\ldots \delta_n$遂有正有負,若於

$$\log X_1-\log G=\delta_1, \log X_2-\log G=\delta_2 \cdots \log X_n-\log G=\delta_n$$

中,擇定G值使$\Sigma\delta=0$,則得

$$\log X_1+\log X_2+\log X_3+\cdots\cdots+\log X_n=n \log G.$$

而有 $G=(X_1X_2X_3\cdots\cdots X_n)^{\frac{1}{n}}.$

範圍較小之集團,其變域自不若較大者之廣,故中值決定須與判值全體有關.以絕對值入算,其結果必無比較功能,故不若取比值入算之為美.二者固幾何中值所專具,而非其他中值所能有;然其意義難明,計算緊困,以視算術中值之衆美皆備,必值密值之易於了解者,則其應用價值有遜色焉.

幾何中值之弱點既多,故其實際應用除人口統計外,殊為罕見.吾人若假定人口蕃殖為按幾何級數與年俱增者,則在殖民期間其中葉之人口總數,應為原有人口總數及結果人口總數之幾何中值.設殖民期間之年限為n,以V

表任何時之人口總數,V_0表原有者,V_n表n年後之結果,r表蕃殖因子,則得

$$V_n=V_0r^n$$

$$V_0V_n=V_0^2r^n$$

$$\sqrt{V_0V_n}=V_0r^{\frac{n}{2}}=V_{\frac{n}{2}}$$

因人口依幾何級數蕃殖之說不甚可靠,由此求出之中數,亦不必與事實全符,期間愈長,相差愈甚,故幾何中值,僅可用於近似估計,不能視為精確結果.

據奧國中央統計委員會調查[1], 1912年終奧國人口總數為

28,748,850,

1913年終增至

28,946,103,

其幾何中值為

28,847,300.

但據實際調查1913年中人口總數為

28,847,478.

與上述結果相差不遠,故此處幾何中值尚可應用.

48. 對數法: 與幾何中值有關,而於集團研究另開途

(1) Neue Folge Bd. 14 (1918), P. 16

徑者,爲 Fechner 之對數法.此法與前述算術方法相倖,惰乘差取比,其理由前節已略舉一二,至在若干集團研究,以應用此法爲便.當最末判值超過算術中值過甚時,算術方法尤不若此法之美,見其又一原因也.

對數法者,乘判值本身,依對數分類以研究標準分佈之方法也.若所分各類爲等隔,則因與彼相當之眞數爲諸毫無規則之判值,故由此所得之分佈,與用算術方法求得者完全不同.

取對數分佈表,依前述方法而研究之,則得新舊中值之關係如下.

a. 對數分佈之算術中值M,爲前述幾何中值之對數,故 M 之眞數等於G.

b. 對數分佈之心值 C,爲前述心值之對數,以 C 等分對數判值爲二部,而 C 亦等分原有判值爲二部,故 C 在對數順序中之位次與 C 在算術順序中之位次相當.

c. 對數分佈之密值D,爲對數判值之集中點.後者與前者之差,相當於算術判值與D眞數之比.若以 D_c 表 D 之眞數,則因 D_c 在算術分佈中,爲諸判值對 G 比值之密集點,故可名之爲比率密值.

至於對數方法之應用,及其與算術方法之關係,則可舉例說明如下:

例:(1)美術品雖依人類意旨而別高下,然亦可按一定規則分類研究,Fechner 曾取明與(München)美術館,及 Darmstadt 圖畫室之實事畫[1],而研究其高度之分佈,依算術方法求得之三中值爲 M=55.26, C=44.30, D=35.40 公分.表中最小高度爲13,最大爲 265 公分.後者超過算術中值三倍有餘,故分佈結果,爲顯著左傾又曲線右枝伸出甚遠,且多空類.

事實畫之高度算術分佈表

X 公分	Z	X 公分	Z
10–20	13	140–150	5
20–30	41	150–160	…
30–40	54	160–170	1
40–50	43	170–180	…
50–60	22	180–190	…
60–70	20.5	190–200	1
70–80	15	200–210	…
80–90	10	210–220	…
90–100	8.5	220–230	…
100–110	5	230–240	1
110–120	3	240–250	…
120–130	6	250–260	…
130–140	3	260–270	1
			253

取分佈原表用對數方法而重列之,則得新表.表中對數類距爲0.08.對數類限之旁,並列相當算術類限,以示與相等對數類幅相當之算術類距爲遞昇者.以其由 2.5 增至 46.3 公分也.取對數類距中點真數與D_c之比則得 v 稱諸

(1) Kollektivma Biehro P. 342.

∘ 晉事畫高度之對數分佈表

Log X	x 公分	類幅 (算術值)	z	v	差
1.08–1.16	12.0– 14.5	2.5	4	0.38	
1.16–1.24	14.5– 17.4	2.9	5	0.46	0.08
1.24–1.32	17.4– 20.9	3.5	5	0.55	0.09
1.32–1.40	20.9– 25.1	4.2	19	0.66	0.11
1.40–1.48	25.1– 30.2	5.1	22	0.79	0.13
1.48–1.56	30.2– 36.3	6.1	38	0.96	0.17
1.56–1.64	36.3– 43.7	7.4	32	1.15	0.19
1.64–1.72	43.7– 52.5	8.8	31	1.38	0.23
1.72–1.80	52.5– 63.1	10.6	26	1.66	0.28
1.80–1.88	63.1– 75.9	12.8	18	2.00	0.34
1.88–1.96	75.9– 91.2	15.3	19	2.40	0.40
1.96–2.04	91.2–109.6	18.4	13	2.88	0.48
2.04–2.12	109.6–131.8	22.2	9	3.47	0.59
2.12–2.20	131.8–158.5	26.7	8	4.17	0.70
2.20–2.28	158.5–190.5	32.0	1	5.01	0.84
2.28–2.36	190.5–229.1	38.6	1	6.03	1.02
2.36–2.44	229.1–275.4	46.3	2	7.24	1.21
			253		

値,取 v 欄相鄰諸値之差,得末欄所示比値增量.至於 v 欄計算,則依對數法爲對數類距中點與 D 値諸差之眞數.亦即

$$1.04-D,\quad 1.12-D,\quad 1.20-D,\cdots\cdots$$

之眞數也.此值在第七類以前爲純小數以後則由中點之值漸大而變爲雜小數矣

上表之三中值爲 $M=1.669$, $C=1.644$, $D=1.538$

其相當眞數爲 $G=46.63$, $C=44.06$, $D_a=34.53$ 公分.

由此求得之 C 值,與用算術方法求得者相近,又舊表之幾何中值,小於其算術中值,其比率密值,亦在算術密值之下.

上述研究,用對數方法所收之效果爲:

a、減小類數,併26類爲17類,

b、縮小頻度,

c、加詳分佈表之首段,而約併其後段.

(2)有取耿湖(Genf 日內瓦)元月之雨量,而統計之者自1845—1892四十八年之間,得雨天 477 日.其算術分佈表如下:

1845—1892耿湖元月雨量之算術分佈表

X 公釐	Z	X 公釐	Z	X 公釐	Z
0—1	133	14—15	3	28—28	1
1—2	88	15—16	3	29—30	
2—3	43.5	16—17	2	30—31	1
3—4	28	17—18	5	31—32	
4—5	27	18—19	1	32—33	1

5− 6	28	19−20	3	33−34	…
6− 7	27.5	20−21		34−35	…
7− 8	14.5	21−22	3	35−36	…
8− 9	16	22−23		36−37	…
9−10	11.5	23−24	2	37−38	…
10−11	12	24−25		38−39	…
11−12	10	25−26		39−40	1
12−13	6.5	26−27			477
13−14	5.5	27−28			

由表得其三中值爲:

$$M=4.486,\quad C=2.403,\quad D=e\text{公釐},$$

因該表以 0 爲密值,故成爲絕對左傾,其判值伸張幾達於算術中值九倍.表尾空類尤多,故須以對數方法,細析其首段,約併其後段,吾人若取原表以 0.2 爲對數類幅而研究之,則得

耿湖元月雨量之對數分佈表

Log X	X 公釐	類幅 (算術值)	z	v	差
−1.5 bis−1.3	0.03− 0.05	0.02	8	0.006	0.004
−1.3 bis−1.1	0.05− 0.08	0.03	8	0.010	0.006
−1.1 bis−0.9	0.08− 0.13	0.05	9	0.016	0.009
−0.9 bis−0.7	0.13− 0.20	0.07	9	0.025	3.015
−0.7 bis−0.5	0.20− 0.32	0.12	28	0.040	0.023
−0.5 bis−0.3	0.32− 0.50	0.18	14	0.063	0.037

−0.3 bis −0.1	0.50− 0.79	0.29	34	0.100	0.04
−0.1 bis 0.1	0.79− 1.26	0.47	45	0.16	0.09
0.1 bis 0.3	1.26− 2.00	0.74	66	0.25	0.15
0.3 bis 0.5	2.00− 3.16	1.16	47	0.40	0.23
0.5 bis 0.7	3.16− 5.01	1.85	53	0.63	0.37
0.7 bis 0.9	5.01− 7.94	2.93	67	1.00	0.59
0.9 bis 1.1	7.94−12.59	4.65	53	1.59	0.92
1.1 bis 1.3	12.59−19.95	7.36	27	2.51	1.47
1.3 bis 1.5	19.95−31.62	11.67	7	3.98	2.33
1.5 bis 1.7	31.62−50.12	18.50	2	6.31	
			477		

上表之三中值爲

$$M=0.314,\quad C=0.374,\quad D=0.8.$$

其相當眞數爲 $G=2.061$, $C=2.366$, $D=6.310$ 公釐.

前後二表完全不同,前爲單方分佈,後係顯著右傾,前者空類累累,後者完全消併.又因比值 $\frac{X}{D_0}$ 自 0.006 變至 6.31, 故觀察結果之變化甚大.

(3) 46 節例 2 所述之食畜活重價分佈表,爲顯著右傾.茲依對數方法加以研究,則得下表:

表中 V 欄所示爲 $\frac{X}{D_0}$ 在對數類幅中之疆限, v 爲屬於對數類幅中點之比值,其對數類幅爲 0.05.

食畜活價及頭數

Log X	X 與 幣	類 幅 (算術値)	z	V	v
1.50−1.55	31.6− 35.5	3.9	87	0.31−0.35	0.33
1.55−1.60	35.5− 39.8	4.3	338	0.35−0.39	0.37
1.60−1.65	39.8− 44.7	4.9	637	0.39−0.44	0.42
1.65−1.70	44.7− 50.1	5.4	1571	0.44−0.50	0.47
1.70−1.75	50.1− 56.2	6.1	2853	0.50−0.56	0.53
1.75−1.80	56.2− 63.1	6.9	7027	0.56−0.63	0.59
1.80−1.85	63.1− 70.8	7.7	14633	0.63−0.70	0.66
1.85−1.90	70.8− 79.4	8.6	15918	0.70−0.79	0.74
1.90−1.95	79.4− 89.1	9.7	33386	0.79−0.89	0.84
1.95−2.00	89.1−100.0	10.9	65388	0.89−0.99	0.94
2.00−2.05	100.0−112.2	12.2	68124	0.99−1.11	0.05
2.05−2.10	112.2−125.9	13.7	18441	1.11−1.24	1.18
			228403		

三中値爲　　G=1.961, C=1.979, D=2.003

相當眞數爲　M=91.41, C=95.28, D_0=100.69與元

用算術方法得 C=95.20與上列 C 値殊爲相近.

上表右傾較前尤甚,其算術中値,及幾何中値,則以92.49−91.41＝1.08相差尙小.其價格則以在比率密値附近者爲多.

49. 調和中値, 倒數$\frac{1}{X}$與 Z 配合所得諸$\frac{Z}{X}$之算術中値倒數,曰 X 之調和中値.若以 H 表之則得

$$\frac{1}{H}=\frac{1}{N}\Sigma\left(\frac{Z}{X}\right). \tag{24}$$

統計術中,間有應用調和中值之趨勢,如就單位錢幣所購某物之件數,或重量,而求其每件或單位重之市價是.

有販雞蛋者,嘗取每月平均馬(Mark)所購蛋數,而統計之,期年而後,所得十二月之結果爲:

月份: I, II, III, IV, V, VI, VII, VIII, IX, X, XI, XII,

每馬蛋數: 9, 11, 15, 16, 15. 14, 13, 11, 10, 9, 6, 7.

欲求是年每蛋之平均價,其法有二:一先求全年每馬所購平均蛋數得 $\frac{136}{12}=11.33,$

再求每蛋之值得 $\frac{100}{11.33}=8.8$ 分尼.

二,以諸值代入(24)式

$$\frac{1}{H}=\frac{1}{12}\left(\frac{1}{6}+\frac{1}{7},+\frac{2}{9}+\frac{1}{10}+\frac{2}{11},+\frac{1}{13}+\frac{1}{14}+\frac{2}{15}+\frac{1}{16}\right)=0.0949$$

得全年平均蛋價爲 9.5 分尼.其所以較用前法求得之結果爲昂者,蓋每馬所購蛋數之調和中值 $H=10.54$,較其算術中值爲小也.

[$a,b(a\neq b)$ 之算術中值爲 $\frac{a+b}{2}$,其調和中值爲 $H=\frac{2ab}{a+b}$,因二者之差 $M-H=\frac{(a-b)^2}{a+b}$ 爲正,故 $M>H$]

又前法係由蛋數之算術中值入門,後法直就蛋價之算

術中值着手,此其根本異點也.

至調和中值之另一應用,則將於96節述之.

第三章 散布量

50. 樣元散布,意義有二,一爲判值之伸張,二爲歷數之堆積.

判值伸張,可以兩端界值表示,此大小界值之差,學者稱爲變域.

以理論言,堆積愈甚,變域愈小,故伸張與堆積之間,似有互相消長之關係.

然此種關係,即令成立,亦無用處,蓋範圍特別擴大,既可增加堆積而不改變域.樣元偶爾變動,復可陡換界值而延展伸張,是則堆積連續增加,變域不必連續擴大,變域特別擴大,堆積不必特別增加.二者之關係,固無數示之可能也.

由是觀之,僅知判值變域,無從決定散布,所謂已與變域云者,不過示人以判定界值而已.

當集團範圍無窮增長,實際分佈,既設理想曲線爲之極限.實際界值,獨不能自理想曲線求出理想界值,以爲之限乎? 斯說理論方面,似持之有故,事實方面,則勞而無功.何則,以曲線至微之變動,足引起界值甚大之遷移也.理想界值,既與事實相違,故其決定方法,可以不論.

度量之足表散布者,除須與判值及度數全體有關外,須具與中值條件相類之性質,換言之即應有簡明提要易於了解之優點也.至計算便利運用如意,尤足增高其應用價值.

散佈量之創造,以差誤論爲基礎,差誤論者,概算中根據幾遇觀念,所得之理論也.事實雖不相同,其公式大都不須改變,即可應用.散布量之時云變化,時云綿延,其意義亦與概算所有者,完全相同.

下節特舉散布量之要而切用者,述其用途,及意義,加以評隲,其算法亦反復說明,務期詳盡.

51. 均方差. 散布量之用途最廣者,厥爲均方差,均方差者,判值對於M乖差平方算術中值之方根也.

設X爲判值,Z爲度數,$X-M=\delta$,爲X對於算術中值之乖差,N爲集团之範圍,則得X對於M之均方差爲:

$$\mu=\sqrt{\frac{1}{N}\Sigma(Z\delta^2)}. \tag{1}$$

因均方差以算術中值爲出發點,故其優點特多,吾人若命未定之出發值爲U,則 $X-U=\varepsilon$ 之平方和,必隨U值而變,今定U值使$\Sigma(\varepsilon^2)$爲極小,則因

$$\Sigma(\varepsilon^2)=(X_1-U)^2+(X_2-U)^2+(X_3-U)^2+\cdots+(X_n-U)^2$$

(1) H.Bruns, Wahrscheinlich keits rechnung und Kollehtivma Blehre P.119

(2) Phil. Trans.Roy. Soc. A, Bd. 185 (1894) P. 80

之極小條件爲

$$\frac{d\Sigma(\varepsilon^2)}{dU}=-2[X_1+X_2+\cdots+X_N-NU]=0.$$

而 $U=\frac{1}{N}\Sigma(X)=M$. 故 M 爲諸 U 中,使 $\Sigma(\varepsilon^2)$ 爲極小之出發值.

又依算術中值定義,有

$$\Sigma(\delta)=(X_1-M)+(X_2-M)+\cdots+(X_N-M)=\Sigma(X)-NM=0,(2)$$

故諸 X 對於 M 之乖差總和爲零.

算術中值與均方差之間,有一重要關係,即二者之平方和,等於判值平方之算術中數是蓋由

$$X+M=\delta$$

有 $X^2=M^2+2M\delta+\delta^2$

因 $\Sigma(\delta)=0$

而 $\Sigma X^2=NM^2+\Sigma(\delta^2)$,

故 $M^2+\mu^2=\frac{1}{N}\Sigma X^2.$ (A)

若集團範圍不大,則 μ 值,不難依法求出 M,取其與各 X 乖差平方對於 N 之平均值,開方計算之.然若集團範圍較大,M 位數過繁,則直接計算,殊爲不便.

較簡之法,先定各判值對任一 U 值之乖差 ε, 以進求其對 M 之差 δ, 其關係如下:

$$\varepsilon-\delta=X-U-(X-M)=M-U=\frac{\Sigma(ZX)}{N}-U$$

$$=\frac{1}{N}[\Sigma(ZX)-NU]=\frac{1}{N}\Sigma[Z(X-U)]=\frac{1}{N}\Sigma(z\varepsilon)=u,$$

由(2)得

$$\Sigma(\varepsilon^2)=\Sigma(\delta^2)+2u\Sigma(\delta)+Nu^2=\Sigma(\delta^2)+Nu^2,$$

以m^2表ε^2之平均值,則上式變為

$$m^2=\mu^2+u^2, \tag{3}$$

或 $$\mu^2=m^2-u^2. \tag{4}$$

式中m^2及u^2求得之後,μ^2即不難算出,但m^2及u^2均與ε有關,故吾人若定U值,使ε極便計算,則μ之求得,易如反掌.

在間續集團,判值咸為整數,所取U值,自以能使ε為整數者為佳.U值既定,祇須求出M,即可計算u值而求μ去.

茲取33節所示栗葉分佈,依上述目的,列為詳表,並舉非散布量算法,以為規範.

此種分佈現象,實可以下述數語概之,而不必有賴於M及μ,且較為顯切.

子葉為單數之複葉占全體……87.8%,

子葉為偶數之複葉占全體……12.8%,

子葉數目為 9 及11之複葉占全體……65.7%,

其中為 9 者占47.6%,

爲11者占52.4%.

子葉爲11之複葉占全體……34.4%.

$U=9$;

$u=\frac{10600-2855}{8554}=0.905$;

$M=9.905$

$m^2=\frac{36343}{8554}=4.2486$

$\mu^2=4.2486-0.8190$

$=3.4296 \quad \mu=1.85$

μ 之算法

X	Z	ε	Zε	Zε²
3	8	−6	48	288
4	5	−5	25	125
5	142	−4	568	2272
6	75	−3	225	675
7	876	−2	1752	3504
8	237	−1	237	237
			−2855	
9	2674	0		
10	527	1	527	527
11	2947	2	5894	11788
12	223	3	669	2007
13	753	4	3012	12048
14	26	5	130	650
15	59	6	354	2124
16	2	7	14	98
	8554		+10600	36343

52. 連續集團,一經分類,則變爲間續,故 μ 之計算,與上法無異,惟二者意義不同,學者宜注意及之.

間續集團之判值,爲各樣元所固有,至分類之連續集團,則各類判值以中點所有者表之,故所得 μ 值,爲由原表求出者之近似.所取之類幅愈小,則二值相距愈近,又因集團之範圍有限,故類幅不能小至無窮.

研究連續集團所取 U 值,以類格中點爲妙.若以類幅爲

單位,表判值對U之乖差ε,則ε必為整數,由此求出之u,m,μ亦然.取所得結果以類距乘之,即得以原有單位表示之諸量.

至於U值位置,則在長分佈表,以與算術中值同類為佳.故計算μ值,能先臆測M所在,而以該類中點為出發值,可收事半功倍之效.若在範圍較小之分佈表,則μ之計算不甚繁冗,所取U值即居極端,亦無大害.讀者細玩下例,當知U之位置與μ值計算之關係矣.

例:(1)取25878名,美國新兵之體高分佈表[1]而求其M及μ.

美國新兵體高之算術中值及均方差

X 英寸	z	ε	$z\varepsilon$	$z\varepsilon^2$
51−52	1	−15	15	225
52−53	1	−14	14	196
53−54	2	−13	26	338
54−55	1	−12	12	144
55−56	3	−11	33	363
56−57	7	−10	70	700
57−58	6	−9	54	486
58−59	10	−8	80	640
59−60	15	−7	105	735

(1) Phil. Trans. Roy. Soc., A, Bd. 186 (1895), P. 385.

60—61	50	−6	300	1800
61—62	526	−5	2630	13150
62—63	1237	−4	4948	19792
63—64	1947	−3	5841	17523
64—65	3019	−2	6038	12076
65—66	3475	−1	3475	3495
			−23641	
66—67	4054	0		
67—68	3631	1	3631	3631
68—69	3133	2	6266	12532
69—70	2075	3	6225	18675
70—71	1485	4	5940	23760
71—72	680	5	3400	17000
72—73	343	6	2058	12348
73—74	118	7	826	5782
74—75	42	8	336	2688
75—76	9	9	81	729
76—77	6	10	60	600
77—78	2	11	22	242
	25878		+28845	169630

$U=66.5;\ u=\dfrac{28845-23641}{25875}=0.2011;$

$M=66.5+0.2011=66.7011$ 英寸，

$m^2=\dfrac{169630}{25878}=6.5325$

$$\mu^2=6.5325-0.0404=6.4921$$

$$\mu=2.5479\text{英寸}$$

其精確足用之算術中值,及均方差爲:

$$M=66.70'' \quad \mu=2.55''.$$

例:(2)德國農人之夏季工資[1].表中X爲每日平均工價,

德國農人之夏季工資

X 馬克	z	ε	$z\varepsilon$	$z\varepsilon^2$
0.60−0.80	7	0	0	0
0.80−1.00	30	1	30	30
1.00−1.20	45	2	90	180
1.20−1.40	104	3	312	936
1.40−1.60	239	4	956	3824
1.60−1.80	249	5	1235	6175
1.80−2.00	257	6	1542	9252
2.00−2.20	50	7	350	2450
2.20−2.40	79	8	632	5056
2.40−2.60	84	9	756	6804
2.60−2.80	19	10	190	1900
2.80−3.00	34	11	374	4114
3.00−3.20	0	12	0	0
3.20−3.40	3	13	39	507
3.40−3.60	1	14	14	196
3.60−3.80	1	15	15	225
	1200		6535	41649

(1) A. Mitscherlich, Zeitschr. f. d. ges. Staatswissensch., VIII. 副刊 1903.

Z爲人數,其工價相差頗大,蓋取各地農人夏日工資,而統計之者.

$$U=0.70;\ u=\frac{6535}{1200}=5.4478\text{類幅}=1.0892\,M.\ (\text{馬克})$$

$$M=0.70+10.9=19\,M.$$

$$m^2=\frac{41649}{1200}=34.7075,$$

$$\mu^2=34.7075-29.6567=5.0508$$

$$\mu=2.2474\text{類幅}=0.45\,M.$$

53. 38節所述決定算術中值之求和法,亦可用之以計算均方差,以其全盤計算,僅用加法,故較前節所用方法爲美.茲除38節已述之諸和外,更取其餘諸和之求法,述之如次:

第一和表

x	z	s
$a+1$	z_1	s_1
$a+2$	z_2	s_2
$a+3$	z_3	s_3
...	...	...
...	...	...
...	...	...
...	...	...
$a+k-2$	z_{k-2}	s_{k-2}

$a+k-1$	z_{k-1}	S_1
$U=a+k$	z_k	
$a+k+1$	z_{k+1}	S_1^{+}
$a+k+2$	z_{k+2}	s_{k+2}
...	...	...
...	...	...
...	...	...
...	...	...
$a+n-2$	z_{n-2}	s_{n-2}
$a+n-1$	z_{n-1}	s_{n-1}
$a+n$	z_n	s_n
	$N=\Sigma(z)$	

由表之上段繼續求s_v之和得

$$\left.\begin{aligned}
&s_1'=s_1 && =z_1\\
&s_2'=s_1+s_2 && =2z_1+z_2\\
&s_3'=s_1+s_2+s_3 && =3z_1+2z_2+z_3\\
&\cdots\cdots\cdots\cdots && \cdots\cdots\cdots\cdots\\
&s_{k-3}'=s_1+s_2+s_3+\cdots\cdots+s_{k-3}\\
&\quad=(k-3)z_1+(k-4)z_2+(k-5)z_3+\cdots\cdots+z_{k-3}
\end{aligned}\right\}\quad(5)$$

取諸s_v'相加得

$$s_1'+s_2'+s_3'+\cdots\cdots+s_{k-3}'=$$

$$=\frac{(k-3)(k-2)}{2}z_1+\frac{(k-4)(k-3)}{2}z_2+\frac{(k-5)(k-4)}{2}z_3+\cdots+z_{k-3}=S_2^-.$$

因

$$(k-3)(k-2)=(k-1-2)(k-1-1)=(k-1)^2-3(k-1)+2$$

$$(k-4)(k-3)=(k-2-2)(k-2-1)=(k-2)^2-3(k-2)+2$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

故

$$2S_2^-=\sum_1^{k-3}(2\varepsilon^2)+3\sum_1^{k-3}(z\varepsilon)+2\sum_1^{k-3}(z). \tag{6}$$

由表之下端,同樣有

$$\left.\begin{aligned}
s_n'&=s_n &&=z_n\\
s_{n-1}'&=s_n+s_{n-1} &&=2z_n+z_{n-1}\\
s_{n-2}'&=s_n+s_{n-1}+s_{n-2} &&=3z_n+2z_{n-1}+z_{n-2}\\
&\cdots\cdots\cdots\cdots\cdots\cdots &&=\cdots\cdots\cdots\cdots\cdots\\
s_{k+3}'&=s_n+s_{n-1}+s_{n-2}+\cdots\cdots+s_{k+3}\\
&=(n-k-2)z_n+(n-k-3)z_{n-1}+(n-k-4)z_{n-2}+\cdots+z_{k+3}
\end{aligned}\right\} \tag{7}$$

求和得

$$s_n'+s_{n-1}'+s_{n-2}'+\cdots\cdots+s_{k+3}'=\frac{(n-k-2)(n-k-1)}{2}z_n$$

$$+\frac{(n-k-3)(n-k-2)}{2}z_{n-1}+\frac{(n-k-4)(n-k-3)}{2}z_{n-2}$$

$$+\cdots\cdots+z_{k+3}=S_2^+;$$

又諸分子之展式爲

$$(n-k-2)(n-k-1)=(n-k)^2-3(n-k)+2$$

$$(n-k-3)(n-k-2)=(n-k-1)^2-3(n-k-1)+2$$

故有

$$2S_2^+=\sum_{k+3}^{n}(z\varepsilon^2)-3\sum_{k+3}^{n}(z\varepsilon)+2\sum_{k+3}^{n}(z). \tag{8}$$

欲計算m^2須先求

$$\sum_{1}^{n}(z\varepsilon^2);$$

今

$$\sum_{1}^{n}(z\varepsilon^2)=\sum_{1}^{k-3}(z\varepsilon^2)+\sum_{k+3}^{n}(z\varepsilon^2)+4z_{k-2}+z_{k-1}+z_{k+1}+4z_{k+2};$$

以(6),(8)結果代入得

$$\left.\begin{aligned}\sum_{1}^{n}(z\varepsilon^2)&=2(S_2^++S_2^-)+3\left(-\sum_{1}^{k-3}(z\varepsilon+2z_{k-2}+\sum_{k+3}^{n}(z\varepsilon)+2z_{k+2}\right)\\&\quad-2\left(\sum_{1}^{k-3}(z)+z_{k-2}+\sum_{k+3}^{n}(z)+z_{k+2}\right)+z_{k-1}+z_{k+1}\\&=2(S_2^++S_2^-)+3\left(-\sum_{1}^{k-2}(z\varepsilon)+\sum_{k+2}^{n}(z\varepsilon)\right)-2(s_{k-2}+s_{k+2})+z_{k-1}+z_{k+1}\\&=2(S_2^++S_2^-)+3(S_1^++S^-)+(S_0^++S_0^-)=2\Sigma_2+3\Sigma_1+\Sigma_0\end{aligned}\right\}\tag{9}$$

式中　$\Sigma_v=S_v^++S_v^-$　(10)

由是得計算均方差之分佈表.

求算術中值及均方差之和表格式

x	z	s	s'
$a+1$	z_1	s_1	s'_1
$a+2$	z_2	s_2	s'_2
…	…	…	…
…	…	…	…
…	…	…	…
…	…	…	…
$a+k-3$	z_{k-3}	s_{k-3}	s'_{k-3}
$a+k-2$	z_{k-2}	s_{k-2}	S_2^-
$a+k-1$	z_{k-1}	$S_0^-=s_{k-1}$. S_1^-	
$U=a+k$	z_k	s_k	
$a+k+1$	z_{k+1}	$S_0^+=s_{k+1}$. S_1^+	
$a+k+2$	z_{k+2}	s_{k+2}	S_2^+
$a+k+3$	z_{k+3}	s_{k+3}	s'_{k+3}
…	…	…	…
…	…	…	…
…	…	…	…
…	…	…	…
$a+n-2$	z_{n-2}	s_{n-2}	s'_{n-2}
$a+n-1$	z_{n-1}	s_{n-1}	s'_{n-1}
$a+n$	z_n	s_n	s'_n
	$N=\Sigma(z)$		

至於選擇U值,應注意各點,已於前節述及.在特例,若以首類中點爲出發值,則表中僅係$S_0^+=\Sigma_0$, $S_1^+=\Sigma_1$, $S_2^+=\Sigma_2$;以末類中點爲出發值,則僅係$S_0^-=\Sigma_0$, $S_1^-=\Sigma_1$, $S_2^-=\Sigma_2$.

又因

$$s'_{k-3}=(k-3)z_1+(k-4)z_2+(k-5)z_3+\cdots\cdots+z_{k-3}$$

$$s_{k-2}=\quad z_1+\quad z_2+\quad z_3+\cdots\cdots+z_{k-2}$$

求和得

$$s'_{k-3}+s_{k-2}=(k-2)z_1+(k-3)z_2+(k-4)z_3+\cdots+z_{k-3}$$

亦即

$$s'_{k-3}+s_{k-2}=S_1^-; \tag{11}$$

同理自表之下段可得

$$s'_{k+3}+s_{k+2}=S_1^+. \tag{12}$$

54. 例:(1)茲取52節例2所述之農人夏作工資,依求和法而求其M及μ.表中除結算須用之數值外,並註普通運算之符號,以醒眉目.

用求和法計算M及μ

x 馬克	z	s	s'
0.70	7	7	7

0.90	30	37	44
1.10	45	82	126
1.30	104	186	312
1.50	239	425	489(S_2^-)
1.70	247	(S_0^-)672. 737(S_1^-)	
1.90	257	929	
2.10	50	(S_0^+)271. 473(S_1^+)	
2.30	79	221	439(S_2^+)
2.50	84	142	252
2.70	19	58	110
2.90	34	39	52
3.10	...	5	13
3.30	3	5	8
3.50	1	2	3
3.70	1	1	1
	1200		

計算

$929+271=1200;\ 425+312=737;\ 221+252=473.$

$U=1.90;\ \Delta_0=271-672=-401,\ \Delta_1=473-737=-264$

$u=-\dfrac{401+264}{1200}=-0.554$ 類幅 $=-0.1118$ 馬,

$M=1.90-0\cdot1118=1.79$ 馬

$\Sigma_0=271+672=943,\quad \Sigma_1=473+737=1210,$

$\Sigma_2=439+489=928,\quad 2\Sigma_2+3\Sigma_1+\Sigma_0=6429$

$m^2=\dfrac{6429}{1200}=5.3575$

$\mu^2=5.3575-0.3069=5.0506$

$\mu=2.2473$ 類幅 $=0\cdot49$ 馬

(2)又52節例(1)所述美國新兵體高之M及μ,可用求和法算之如下:

美國新兵體高之均方差

x 英寸	z	s		s'
51.5	1	1		1
52.5	1	2		3
53.5	2	4		7
54.5	1	5		12
55.5	3	8		20
56.5	7	15		35
57.5	6	21		56
58.5	10	31		87
59.5	15	46		133
60.5	50	96		229
61.5	526	622		851
62.5	1237	1859		2710
63.5	1947	3806		6516
64.5	3019	6825		10660
65.5	3475	10300	13341	

66.5	4054	14354	
67.5	3631	11524, 17321	
68.5	3133	7893	17250
69.5	2075	4760	9428
70.5	1485	2685	4668
71.5	680	1200	1983
72.5	343	520	783
73.5	118	177	263
74.5	42	59	86
75.5	9	17	27
76.5	6	8	10
77.5	2	2	2
	25878		

計算

$14354+11524=25878;\ 6825+6516=13341,\ 7893+9428=17321.$

$U=66.5;\ \Delta_0=11524-10300=1224,\ \Delta_1=17321-13341=3980$

$u=\frac{5204}{25878}=0.2011,\ M=66.5+0.2011=66.7011$ 英寸

$\Sigma_0=21824\quad \Sigma_1=30662,\quad \Sigma_2=27910$

$2\Sigma_2+3\Sigma_1+\Sigma_0=169630$

$m^2=\frac{169630}{25878}=6.5325$

$\mu^2=6.5325-0.0404=6.4921$

$\mu=2.5479$ 英寸

(3)取維陰天文台1775至1847[1]測候所得七十三年中元旦上午八時之氣壓結果,而求其 M 及 μ.

1775—1847維陰元旦上午八時之氣壓

x 奧寸	z	s	s'
27.0	1	73	
27.2	3	72　320.5	
27.4	4	69	735
27.6	7.5	65	251.5
27.8	12.5	57.5	186.5
28.0	11	45	129
28.2	8	34	84
28.4	10	26	50
28.6	10	10	24
28.8	4	6	8
29.0	2	2	2
	73		

計算:　$69+251.5=320.5$

$U=27.0$　$\varDelta_0=72$, $\varDelta_1=320.5$; $u=5.3767$ 類幅 $=1.0753$ 維寸[2]

(1)維陰(奧京)皇家天文台1775—1847氣象報告五卷.——自1848年起其觀察時間變更.故不錄.

(2) 1維寸=26.34 mm.

$$M=27.0+1.0753=28.0753\text{維寸}$$

$$\Sigma_0=72, \quad \Sigma_1=320.5 \quad \Sigma_2=735$$

$$2\Sigma_2+3\Sigma_1+\Sigma_0=2503.5$$

$$m^2=\frac{2503.5}{73}=34.2945$$

$$\mu^2=34.2945-28.9089=5.3856$$

$$\mu=2.321\text{類幅}=0.464\text{維寸},$$

此處算術中值,爲七十三年元旦之氣壓中數,可表是日氣象特徵,至均方差則爲是日氣壓變化之均值,若其餘各日之μ已知,則由此可察出全年氣壓之變化.

55. 分類連續集團之均方差,係假定各類判值與中點判值相同,而計算得之者.其結果必有差誤,故須依某種假設,估計其量,而除去之.如假定諸X在類格中爲平均分佈,以梯級形表全體分佈,而計算其差誤是.

以X表判量,x表中點判值,$X-x=\varepsilon$表此類判值對類中點之乖差,z表此類之屢數,取恆等式

$$M-x=M-X+\varepsilon,$$

平方之,有

$$(M-x)^2=(M-X)^2+2(M-X)\varepsilon+\varepsilon^2,$$

求其在類中總和得

$$Z(M-x)^2=\Sigma(M-X)^2+2\Sigma(M\varepsilon)-2(X\varepsilon)+\Sigma(\varepsilon^2),$$

因 $2\Sigma(M\varepsilon)=2M\Sigma(\varepsilon)=0, \; 2\Sigma(X\varepsilon)=2\Sigma(X)\Sigma(\varepsilon)=0,$

故　$Z(M-x)^2=\Sigma(M-X)^2+\Sigma(\varepsilon^2)=\Sigma(M-X)^2+z\mu_\varepsilon^2$.

式中μ_ε為此類之均方差.由第20圖,

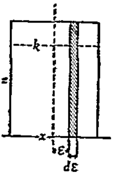


圖20 Sheppard 公式之求法

因　$d\varepsilon=\frac{X_2-X_1}{Z}$, $\mu_\varepsilon^2=\frac{\Sigma(\varepsilon^2)}{Z}$

$$=\frac{\Sigma(\varepsilon^2)d\varepsilon}{X_2-X_1}=\frac{\Sigma(\varepsilon^2)d\varepsilon}{\Sigma d\varepsilon}$$,類幅為k,

而　$\Sigma_\varepsilon^2=\frac{\int_{-\frac{k}{2}}^{\frac{k}{2}}\varepsilon^2 d\varepsilon}{\int_{-\frac{k}{2}}^{\frac{k}{2}}d\varepsilon}=\frac{k^2}{12}$

為常數故依上式求各類總和,有

$$\Sigma z(M-x)^2=\Sigma(M-X)^2+\mu_\varepsilon^2\Sigma(z).$$

式中X遍歷全圖判值.以$\Sigma(z)$除上式兩端,則得

$$\frac{\Sigma z(M-x)^2}{\Sigma(z)}=\frac{\Sigma(M-X)^2}{\Sigma(z)}+\mu_\varepsilon^2.$$

上式左端為51—54節用舊法求得之均差平方,左端首項為依假定算出之正確均差平方,若以μ_1表之,則有

$$\mu^2=\mu_1^2+\mu_\varepsilon^2,$$

或　$$\mu_1^2=\mu^2-\frac{k^2}{12}. \tag{13}$$

是為Sheppard公式[1]蓋 W. F. Sheppard 所求得者.由是知均

(1) 皇家統計學會叢刊 (Journ. Roy. Statist. Soc). 卷 IX(1897), P. 638.

差平方之差誤量爲$\frac{1}{12}$類幅平方,若以類幅計算,則爲$\frac{1}{12}$.故依原有觀念,以求均方差對於計算結果,影響甚微.若類幅甚小,則此種更正無關輕重,若其甚大,則計算均方差自以應用 Sheppard 公式爲佳.

49節例(1)所述美國新兵體高之μ值,以類幅計算有

$$\mu^2=6.4921,$$

由此減去 $\frac{1}{12}=0.0833\cdots\cdots$,則得

$$\mu_1^2=6.4088,$$

開方得 $\mu_1=2.5315$與原值$\mu=2.5479$相差爲0.0164.

56. 均淨差 均淨差者,將值對於某中值絕對偏差之算術中數,爲散佈論中之第二有用度量.至所謂中值,則或取M,或用心值,均無不可.後者依下述理論雖較前者爲優,然通常多有棄C而用算術中值者.

心值有使均淨差爲最小之特性,其事頗與51節所述M使μ爲最小之性質相近,茲以同樣觀念,證之如下:選定U爲出發值,便將值居其上者爲m個,居其下者爲$n-m$個,設Θ爲此時之均淨差,將U向右移動,但不改變其上下原有將值數目.若所移之距離爲u,則其偏差總和上段減小mu,下段增加$(n-m)u$,故U值移動以後所得均淨差爲

$$\Theta+\frac{(n-m)u-mu}{n}=\Theta+\frac{u}{n}(n-2m).$$

同理向左移動所得均淨差爲

$$\Theta+\frac{mu-(n-m)u}{n}=\Theta+\frac{u}{n}(2m-n).$$

前者當$m<\frac{n}{2}$,後者當$m>\frac{n}{2}$,咸有增量.故$m=\frac{n}{2}$時Θ爲最小,亦即謂居U上下判值個數相等或$U=C$時Θ爲極小也.

此後研究,將以θ表均淨差,而聲明所用中值爲C抑M,其未加註定者,則以算術中值爲依歸,至應用分佈表之計算方法,則可取52節例(1)所述美國新兵軀高求其均淨差以爲規範.由52節例(1),分佈表$U=66.5$,其上段乖差之和爲28845,下段絕對差和爲23641,故其總和爲52486.

今進求以$M=66.7011$爲出發值時,此和之變化,由54節例(2),知居M上段樣元個數爲11524,居其下者爲14354,由是得總和增量爲

$$(14354-11524)(66.7011-66.5)=569.113,$$

因之

$$\theta=\frac{52486+569.113}{25878}=\frac{53055.113}{25878}=2.0502\text{英寸}.$$

若自C值出發,則因$C=66.6509$,而得乖差總和之變化爲

$$(14354-11524)(66.6509-66.5)=427.047;$$

因之

$$\vartheta=\frac{52486+427.047}{25878}=\frac{52913.047}{25878}=2.0447\text{英寸.}$$

$2.0502-2.0447=0.0055$,故以C爲出發值所得之均淨差.較以M爲出發值所得者爲小.

應用求和法,計算均淨差:

按38節公式有

$$\sum_1^{k-2}z|\varepsilon|=-\sum_1^{k-2}z\varepsilon=S_1^-+s_{k-2}$$

$$\sum_1^{k+2}(z\varepsilon)=\qquad S_1^++s_{k+2}$$

而 $$\Sigma z|\varepsilon|=\sum_1^{k-2}z|\varepsilon|+z_{k-1}+z_{k+1}+\sum_1^{k+2}(z\varepsilon)=S_1^-+s_{k-2}+z_{k-1}$$

$$+z_{k+1}+s_{k+2}+S_1^+=S_1^++S_1^-+S_0^++S_0^-=\Sigma_0+\Sigma_1.$$

因得對U之平均乖差爲

$$\frac{\Sigma_0+\Sigma_1}{n},$$

由此依所用中值之爲M或C,求出乖差總和之變化均值而計算之,卽得均淨差.

重取54節例(2)美國新兵體高依求和法,而計算其均淨差,則因

$$\Sigma_0=21824\qquad \Sigma_1=30662$$

而$\Sigma_0+\Sigma_1=52486$,與由前法所得對U乖差總和吻合.再以

11524及14354二值入算,不難求出均淨差.

57. 四分值,十分值及百分值 在昔人種學中之散布量,至今尚能應用者,爲四分值,其定義如下:

取集團樣元,依序排列,再等分之爲四部,而以Q_1表第一區分點之判值,使一部居其下,三部居其上.Q_3表第三區分點之判值,使一部居其上,三部居其下.至第二區分點,則分樣元全體爲相等二部,故與心值相合,因得等分樣元全體爲四部三區分點之判值爲Q_1, C, Q_3.

若所論分佈完全對稱,則

$$C-Q_1=Q_3-C.$$

反之若爲左傾,則

$$C-Q_1<Q_3-C.$$

若爲右傾,則

$$C-Q_1>Q_3-C.$$

但完全對稱,至爲罕見,故Q_3+Q_1之半,不常與心值相重,學者爲劃一度量起見,特取$C-Q_1, Q_3-C$和距中點,所有判值爲散布量

$$Q=\frac{Q_3-Q_1}{2}. \tag{14}$$

名之曰四分值,式中Q_1爲第一四分值,Q_3爲第三四分值.Q爲二四分點之距離中心.四分值,除在完全對稱與C及M相重外,通常不必與特殊判值叠合,故在普通情形,Q值僅

示吾人以二四分點中心所在.換言之,亦即僅示吾人以二區分值爲界之類距中點也.

四分值之決定方法,與心值決定方法相同,惟計算Q_1須使下端樣元個數爲$\frac{N}{4}$,計算Q_3須使上端樣元個數爲$\frac{N}{4}$,以其計算手續,與決定心值相同,故仍以求和法爲妙.

在間續集團,四分值之定法,與心值之定法相同.

就四分值之意義,加以推廣,則得十分值,百分值等,吾人若分樣元爲十部,則得九個十分值,其第五值與C值相重.若分樣元爲百部,則得九十九個百分值,其第五十值與C值相重.所謂四分,十分,百分者無他,蓋便於比較起見,所行之特別分類法耳.

今更取美國新兵體高,依54節例(2)分佈表,而求其四分值如下.

按表 $N=25878$ $\frac{N}{4}=6469.5$其上段屢數和至$X=64$爲3806,故

$$Q_1=64+\frac{6496.5-3806}{3019}=64.882\text{英寸}.$$

其下段屢數和至$X=69$爲4760,故

$$Q_3=69-\frac{6469.5-4760}{3133}=68.454\text{英寸}.$$

又$C=66.651$英寸, $C-Q_1=1.769$ $Q_3-C=1.803$,

因之 $(Q_3-C)-(C-Q_1)=0.034$ 英寸,而分佈爲微度左傾.取二者之和折半,得

$$Q=\frac{68.454-64.882}{2}=1.786\text{英寸}.$$

是爲分佈之四分值.

計算四分值,頗幅極爲重要,其理與計算心值同,今取54節例1所述德國農人夏作工資,求其 Q 值,以爲規範.

$$N=1200 \quad \frac{N}{4}=300$$

$$Q_1=1.40+\frac{300-186}{239}\cdot 0.2=1.495\text{馬克}$$

$$Q_3=2.00-\frac{300-271}{257}\cdot 0.2=1.978\text{馬克}$$

$$C=1.742,\quad C-Q_1=0.247\quad Q_3-C=0.236$$

$$Q=\frac{1.978-1.495}{2}=0.242\text{馬克}.$$

此種分佈之缺乏常規,可由其屢數陡降察出.按前述規則,在左傾分佈,應有 $C-Q<Q_3-C$,今分佈既爲左傾,而 $C-Q_1>Q_3-C$ 者,亦缺乏常規有以使然也.

十分值之應用,恆見於死亡統計;死亡表,可視爲死亡對年齡之分佈,其生存者之數目,自表之下端算起,每加入一類死亡,即爲該類原有人數.至三中值 M,C,D 則依序表平均壽命,概約壽命,及Lexis標準年齡. W. Lexis者,德國著名

1905~1910
奧國死亡表之十分值

	男	女
d_1	11.5 月	3.68 月
d_2	11.06 月	1.40 年
d_3	56.62 年	13.98 年
d_4	32.25 年	35.32 年
d_5	48.62 年	52.02 年
d_6	58.56 年	61.78 年
d_7	65.74 年	68 12 年
d_8	71.82 年	73.61 年
d_9	78.08 年	79.27 年

國民經濟及人口統計家也.

旁表所示為1906至1910年,奧國死亡表之十分值 $d_1, d_2 \cdots\cdots d_9$. 男女並具.由此可見兩性死亡率之不同,女性各齡死亡,均較男性爲少,而尤以孩提至少壯爲尤甚.換言之,亦卽女子較男子爲適於生存也

茲爲節省篇幅起見,特取男性死亡表之小部,而求其十分值

年齡	死亡數	原有生存人數
32–33	455	60115
33–34	469	59660.

原表範圍,係以100000爲底,故自表之下端算起,與第四個十分值相當之生存人數應爲 $6\cdot\frac{N}{10}=60000$. 由是得

$$d_4=33-\frac{60000-59660}{455}=32.25\text{歲}.$$

58. 均方差均淨差及四分值等,旣皆爲散布度量則比

較二種分佈,無論以何者出發,其相對關係,應爲一致.取分佈互相比較,自以由類似材料求出者爲佳,若其屢數曲線,可用分析方法决定,則三散布量,是否同效問題,可藉理論解决.否則必須依統計結果,實行計算,又因吾人曾假定散布與散布量成正比,故各種分佈之散布量,概須求出,然後取各分佈中二散布量求其比值,若所得結果相近,則三值之爲同效甚明.

茲就後者,舉例論之.

右表爲巴什旗人,韃靼人之體高[1]分佈,除計算其散布量外,並求其必值以便比較.

巴什旗人及韃靼人體格散布量之比較

X 公分	巴什旗人	韃靼人
148—151	1	
151—154	6	6
154—157	32	15
157—160	72	38
160—163	78	56
163—166	70	49
166—169	78	32
169—172	39	23
172—175	19	20
175—178	10	6
178—181	4	1
181—184	2	
	411	246

[註] Tatareu爲土耳其民族之一,支別甚多,Kasan (Wolga)韃靼,Thrazicu韃靼(居希臘土耳其交界處),保加利亞韃靼,皆混血種.至純粹韃靼,則僅餘西伯利亞巴拉巴

(1) 見 R. Poech 奧國俘虜所人體測量記錄.

(Baraba) 草原以遊牧爲生之四萬人.此處並指俄羅斯韃靼共和國之居民而言,亦即 Kasan 韃靼也.

	巴什旗人		韃靼人
$\mu =$ 5.74公分	Q_3=167.89公分	$\mu =$ 5.61公分	Q_3=167.92公分
$\vartheta =$ 4.72公分	$Q =$ 4.12公分	$\vartheta =$ 4.57公分	$Q =$ 3.89公分
Q_1=159.65公分	$C-Q_1=$ 4.06公分	Q_1=160.13公分	$C-Q_1=$ 3.36公分
C =163.71公分	$Q_3-C=$ 4.18公分	C =163.49公分	$Q_3-C=$ 4.43公分.

由三散布量,所得之判斷全同,亦卽巴什旗人之體高散佈,較韃靼人爲大.因之其體格變化亦較強.

又二者之散布量比值,爲

巴什旗人	韃靼人
$\frac{\vartheta}{\mu}=0.82$	$\frac{\vartheta}{\mu}=0.81$
$\frac{Q}{\mu}=0.72$	$\frac{Q}{\mu}=0.69.$

因集團之範圍太小,故相當比值相差尙遠,今再取範圍較大英格蘭蘇格蘭成年男子之體高分佈論之.

英格蘭人及蘇格蘭人體高散布量之比較

X 英寸	英格蘭人	蘇格蘭人
57—58	1	
58—59	3	1
59—60	12	

60—61	39	2
61—62	70	2
62—63	128	9
63—64	320	19
64—65	524	47
65—66	740	109
66—67	881	139
67—68	918	210
68—69	886	210
69—70	753	218
70—71	473	115
71—72	254	102
72—73	117	69
73—74	48	26
74—75	16	15
75—76	9	6
76—77	1	4
77—78	1	1
	6194	1304

由表得二者之散布量爲

英格蘭人		蘇格蘭人	
μ = 2.60英寸	Q_3=69.16英寸	μ = 2.50英寸	Q_3=70.10英寸
ϑ = 2.05英寸	Q = 1.78英寸	ϑ = 1.96英寸	Q = 1.56英寸
Q_1=65.61英寸	$C-Q_1$= 1.80英寸	Q_1=66.99英寸	$C-Q_1$= 1.55英寸
C =67.41英寸	Q_3-C = 1.75英寸	C =68.54英寸	Q_3-C= 1.56英寸.

其相當各量均有同樣關係,而以英人之體格變化爲大,求散布量之比值得:

英格蘭人	蘇格蘭人
$\frac{\vartheta}{\mu}=0.79$	$\frac{\vartheta}{\mu}=0.78$
$\frac{Q}{\mu}=0.68$	$\frac{Q}{\mu}=0.63$

因知$\frac{\vartheta}{\mu}$大於$\frac{Q}{\mu}$而相當比值,相差不遠,且理論上,可視爲常數.故三散佈量有同等效力.

59. 因上述之散布量皆爲絕對度量,且與單位有關,故於分佈比較,殊爲不便.欲就支配樣元之散布,判斷材料之變化,則須注意散佈,乖差,中值三者之關係.蓋乖差愈大,則散布愈廣,乖差愈小,則散布愈削,而乖差大小,又以中值度量爲轉移,鼠類體格變化之不如人類,灌木高度變遷之不及喬木,皆以中值影響乖差,乖差關涉散佈也.然自中值大小,以斷定均方差則可,僅就均方差以斷定分佈變化則不可.何則,蓋就各種器官,以研究相異人種,及獸類,則其體格變化,各有定規,固不因所取器官不同,乖差迥異,而易其關係也. F. Galton 氏曾取男女相同之器官,作遺傳研究,所得相當中值之比例,與 13:12 至爲相近.故通常取女子某中值之$\frac{13}{12}$,即得男子應有之數.然若僅就中值之絕對量觀之,

則又大小互異,關係罯然矣.

故欲研究結果,便於比較,變化大小一目瞭然,必須依散布意義,乖差性質,求出一相對散布量,以顯示其結果.

變化係數者,K. Pearson(1) 倡用之相對散布量也.此值為均方差與算術中值比率之百分數,若以 V 表之,則有

$$V=100\frac{\mu}{M}. \tag{15}$$

Pearson 氏所以選定此值者,蓋因變度大小,一方以均方差為轉移,他方與算術中值成反比也.

58節曾取四種成年男子,而研究其體高分佈,茲為計算變化係數起見,特求其算術中值如下.

巴什旗人	164.05公分	英格蘭人	67.37英寸
韃靼人	164.09公分	蘇格蘭人	86.39英寸

前二者之平均體高,雖相差甚小,而後二者則因單位關係,相差頗大.故在前節,依絕對量評成之變化,實嫌不確.據計算所得,其相當之變化係數為:

巴什旗人	3.50	英格蘭人	3.86
韃靼人	3.42	蘇格蘭人	3.66.

至童年與成人之變化係數,是否相同,則可取美國 St. Louis 城八齡女生之體高研究,與以解答.

---

(1) Phil. Traus. Roy. Soc., A 卷 187 (1897) P. 277.

八齡女生之體高

x 公分	z	x 公分	z	x 公分	z
99.5	1	115.5	297	131.5	21
101.5	2	117.5	321	133 5	10
103.5	8	119.5	342	135.5	5
105.5	27	121.5	243	137.5	1
107.5	42	123.5	183	139.5	0
109.5	84	125.5	138	141.5	1
111.5	137	127.5	79		
113.5	222	129.5	28		2192

由表得 $M=118.271Cm$ $\mu=5.492Cm$ $V=4.64$ 故童年身體變化,較成人爲大.

60. 不稱程度,爲支配分佈形狀最要元素之一.完全對稱,實際已屬特例,且多係省略計算所得之結果.蓋在完全對稱,中值 M,C,D 應絕對吻合,若有差離,即生不稱.然此種乖差,有時僅可由甚微之小數位上察出,以測算之精確度言,已無計較必要.故雖非嚴密吻合,亦可謂之對稱分佈.

不稱程度愈強,中值相差愈大,由46節所述中值之相差最大者,爲算術中值及密值,故自其相差,可得不稱程度之絕對量.因絕對量不便比較,Pearson[1] 氏特依前節所述關於散布量之思維,造一相對度量,以測不稱,而以分佈傾度

(1) 哲學叢刊, A 卷 186 (1895) P. 370.

名之.此值爲自算術中值減去密値,再以均方差除得之結果.換言之,亦即

$$傾度=\frac{M-D}{\mu}也. \qquad (16)$$

若所取爲絕對量,則在相似分佈,伸張愈大,M 與 D 相差愈遠,伸張愈小,相差愈微.反之,若 M 與 D 之相差固定,則伸張愈大,不稱愈小,伸張愈小,不稱反大.是則 $M-D$ 之效力,實視材料之伸張爲轉移,故欲得可靠度量,非以散布量除之不可.

因在左傾分佈,$M>D$,而傾度爲正,右傾分佈 $M<D$ 而傾度爲負,故亦有稱二種不稱爲正向負向以別之者.然通常所得,其淨值恆爲小數.

玆取前述諸例,求其傾度,及與傾度有關諸值.

1. 美國新兵之體高（52節例1）:

$M=66.701$ 英寸, $D=66.422$ 英寸, $\mu=2.548$ 英寸;

傾度$=+0.109$.

2. 美國八齡女生之體高（59節）:

$M=118.271$ 公分, $D=118.850$ 公分, $\mu=5.492$ 公分;

傾度$=-0.105$.

3. 英格蘭成年男子之體高（58節）:

$M=67.373$ 英寸, $D=67.536$ 英寸, $\mu=2.595$ 英寸;

傾度$=-0.063$.

4. 蘇格蘭成年男子之體高（58節）:

$M=$ 68.392英寸, $D=$ 69.072英寸, μ=2.496英寸;

傾度=−0.272.

5. 年齡對於腸熱症之分佈（46節例1）:

$M=$ 18.97, $D=$ 14.64, μ=9.88 齡;

傾度=+0.438.

6. 維陰市場之食畜活價（46節例2）:

$M=$ 92.49, D=100.66, μ=16.32奥元;

傾度=−0.501.

61. 中值及散布量之用法及意義,前已述之詳矣.茲特取二例計算其一切數值,以示求和法之優點,且使讀者對於前述各論,得一連貫概念焉.

(1)劍橋每日之氣壓.是表爲13年間每日上午九時觀察之結果,共凡13·365+3=4748日,其類幅爲0.1英寸.

劍橋每日氣壓水銀柱高

X 英寸	z	s	s'
28.3	1	1	1
28.4	...	1	2
28.5	...	1	3
28.6	1	2	5
28.7	2	4	9

28.8	6.5	10.5		19.5
28.9	10.5	21		40.5
29.0	23	44		84.5
29.1	24	68		152.5
29.2	63.5	131.5		284
29.3	81	212.5		496.5
29.4	127	339.5		836
29.5	213	552.5		1388.5
29.6	289	841.5		2230
29.7	388	1229.5		3459.5
29.8	479.5	1709		9011.5
29.9	537.5	2246.5	5168.5	
30.0	586	2832.5		
30.1	550	1915.5	3241.5	
30.2	488	1365.5		3614
30.3	350.5	877.5		1876
30.4	246	527		998.5
30.5	150	281		471.5
30.6	85.5	131		190.5
30.7	35	45.5		59.5
30.8	7.5	10.5		14
30.9	2.5	3		3.5
31.0	0.5	0.5		0.5
	4748			

計算

1. $2832.5+1915.5=4748$; $1709+3459.5=5.1685$;

$1365.5+1876=3241.5$

2.

1915.5	3241.5	3614
2246.5	5168.5	9011.5
$\Sigma_0=4162$	$\Sigma_1=8410$	$\Sigma_2=12625.5$
$\Delta=-331$	$\Delta_1=-1927$	

3. $\frac{\Delta_0+\Delta_1}{N}=-0.4756$ 類幅$=-0.0476$英寸

$M=30-0.0476=29.9524$英寸 ($=760.7909mm$)

4. $\frac{N}{2}=2374$; $C=29.95+\frac{2374-2246.5}{586}\times0.1=29.9718$

($=761.2837mm$)

5.

537.5	48.5	-84.5
586	-36	
550		

$D=29.95+\frac{48.5}{84.5}\times0.1=30.0074$英寸($=762.1880mm$)

6. $D-M=0.0550$, $C-M=0.0194$, $3(C-M)=0.0582$

7. $\Sigma_0+3\Sigma_1+2\Sigma_2=54643$, $m^2=11.5086$

$\mu=0.1\times\sqrt{11.5086-(0.4756)^2}=0.336$英寸

8. $\Sigma_0+\Sigma_1=12572$; $\vartheta=\frac{12572+12.138}{4748}\times0.1=0.265$英寸

9. $\frac{N}{4}=1187;\ Q_1=29.65+\frac{1187-841.5}{388}\times 0.1=29.739$英寸

$$Q_3=30.25-\frac{1187-877.5}{488}\times 0.1=30.187\text{英寸}$$

$$Q=0.224\text{英寸}$$

$$C-Q_1=0.233\text{英寸} \qquad Q_3-C=0.215\text{英寸}$$

10. $\frac{\vartheta}{\mu}=0.789 \qquad \frac{Q}{\mu}=0.667$

11. $V=100\times\frac{0.336}{29.952}=1.122$

12. 傾度$=-\frac{0.055}{0.336}=-0.1637.$

其分佈情形,讀者可依上述結果自加評定.

又分佈之主要結果爲:

$M=29.9524$英寸 $\mu=0.336$英寸 $V=1.122$

$C=29.9718$英寸 $\vartheta=0.265$英寸 傾斜$=-0.1637.$

$D=30.0074$英寸 $Q=0.224$英寸

是例取自Pearson(1)氏『氣壓壓數分佈.』Pearson氏除英倫三島二十測候所外,更取 Dover Dungeness, London Cambridge 三所氣象報告.加以研究,其觀察時間少至五年,多至十三年,所得重要結果,爲氣壓分佈右傾,中值變化甚小.如自23處觀察結果,求得之密值其變限爲

(1) 哲學叢刊 A, 卷 190 (1898) P. 423.

29.9232及30.0493.最大相差不過0.1261英寸,必値變限爲29.8457及29.9834.其最大相差不過0.1377.至Pearson作此研究之主要目的,蓋欲如首章所云,求理想曲線與觀察結果互相配合耳.

(2)下述材料,與前論各例所屬範圍不同,是爲二十四至二十五歲男子所娶新婦之年齡分佈.

24—25歲男子所娶新婦年齡之分佈

x 年齡	z	s		s'
15.5	367	367		367
16.5	717	1084		1451
17.5	294	2378		3829
18.5	2121	4499		5647
19.5	3156	7655	8328	
20.5	4009	11664		
21.5	3593	16790	47415	
22.5	3604	13197		130279
23.5	3060	9593		34218
24.5	1774	6533		2465
25.5	1353	4759		18092
26.5	936	3406		13333
27.5	663	2470		9927
28.5	468	1807		7457

29.5	319	1339	5650
30.5	256	1020	4311
31.5	164	764	3291
32.5	134	600	2527
33.5	94	466	1927
34.5	77	372	1461
35.5	68	295	1089
36.5	59	227	794
37.5	33	168	567
38.5	40	135	399
39.5	27	95	264
40.5	18	68	169
41.5	21	50	101
42.5	11	29	51
43.5	14	18	22
44.5	4	4	4
	28454		

計算

1. $11664+16790=28454$; $4499+3829=8328$;

$13197+34248=47415$

2.

16790	47415	130279
7655	8328	5647
$\Sigma_0=24445$	$\Sigma_1=55743$	$\Sigma_2=135926$
$\Delta_0=9135$	$\Delta_1=39087$	

3. $\frac{\Delta_0+\Delta_1}{N}=1.695$, $M=20.5+1.695=22.195$ 齡

4. $\frac{N}{2}=14227$; $C=21+\frac{14227-11664}{3593}=21.714$ 齡

5. 3156 853 −1269

4009 −416

3593

$$D=20+\frac{853}{1269}=20.672 \text{ 齡}$$

6. $M-D=1.523$, $M-C=0.481$, $3(M-C)=1.443$.

7. $\Sigma_0+3\Sigma_1+2\Sigma_2=463526$, $m^2=16.2904$

$$\mu=\sqrt{16.2904-(1.695)^2}=3.663 \text{ 齡}$$

8. $\Sigma_0+\Sigma_1=80188$; $\vartheta=\frac{80188-8688.57}{28454}=2.512$ 齡

9. $\frac{N}{4}=7113.5$; $Q_1=19+\frac{7113.5-4499}{3156}=19.828$ 齡

$Q_3=24-\frac{7113.5-6533}{3060}=23.811$ 齡

$Q=1.991$ 齡

$C-Q_1=1.886$, $Q_3-C=2.097$

10. $V=100\times\frac{3.663}{22.195}=16.5$

11. 偏度$=\frac{1.523}{3.663}=0.416$.

其最堪注意者,為其變化係數及傾度特別強大.表中自 19.828至28.811 歲,3.983年間出嫁之新婦竟占全體半數,而其餘之新婦則支配在相差為26歲年限之間.此種特異情形,及其變化係數,與傾度之強大,皆為前述各例所未有.

結婚遲早,與戀愛進行,民族習慣,經濟狀況,三者有關.相異民族之婚期遲早,有表民族特性之功能.同一民族之婚期變遷,多蒙經濟情形之影響.斯時婚期改變,各有主因,至其他因子,則關係甚小矣.然無論情形奚似,此種分佈,咸為左傾,實可豫斷.以兩性結婚期間,大都以少年為衆也.

下表為1910年[1],瑞典及英國新婚夫婦之年齡分佈.自此可察出其相當關係及民族特性.

1910瑞典及英國結婚者之年齡

X 年齡	瑞典		英國	
	新郎	新婦	新郎	新婦
15—20	65	3501	3349	17546
20—25	8230	13144	93209	118542
25—30	12337	9099	91176	70411

(1) 見 Arkiv för Matematik, Astronomish Fysik 卷 XII 中 S. D. Wicksell 氏論文.

30—35	5705	3488	31485	20241
35—40	2199	1151	10274	5873
40—45	831	415	3471	1706
45—50	381	153	1387	636
50—55	163	44	558	171
55—60	73	17	194	64
60—65	22	3	76	28
65—70	4	—	40	23
70—75	—	—	23	11
75—80	—	—	10	—
	30015	30015	235252	235252

由其相當主要結果

	瑞典		英格蘭	
	男	女	男	女
M	28.698	25.808	26.921	25.114
C	27.719	24.758	26.155	24.221
D	26.911	23.623	24.889	23.386
μ	5.800	5.270	5.050	4.555
傾度	0.308	0.414	0.404	0.379,

可知瑞典婚期兩性均較英國爲遲,而尤以男子爲甚.

第四章　兩標相關性理論

62. 以上所論,概爲一標分佈於個體之情形及描寫此種情形之方法.如以中值及散布量描寫分佈之結果,多邊形及曲線表示分佈之大勢是.

若所定標準有二,且此二標準互相聯絡,同時出現,則吾人研究之問題,頗爲雙變數之變化.蓋由測算所得之結果,二標級度兩兩相當,成爲値偶,而分佈於集團中之個體,與雙變數値偶配合甚相類似也.

値偶云者,二標同時出現於一體之級度或一標對於二體之分佈.前者如研究葉之長廣,後者如統計夫婦之年齡.欲窮其變化之理,必須取此二値互相配置,然後加以討論.故値偶配合,對於二標相互關係之研究,甚爲重要.配合之法,約分二種:一以個體爲主,而後配置級度.二先取標度排列,然後求其屢數.後者如取甲標X出現於個體之級度,依其大小或爲漸升或爲遞降加以排列.次取乙標Y之級度,依法序之,然後作雙檢表,而計其屢數是.若二標級度升降之方向相同,則此種排列爲同向.否則爲逆向.然按諸實際,標度混淆,至爲參差,純爲同向或逆向者,蓋寥寥也.故所謂完全同向,或完全逆向之增長,實特例中之特例.

取甲乙二標分別列其級度,計其屢數,對於二標分佈變化之比較,至爲適宜.以吾人可就此二列中分別求其絕對散布量,加以比較,則二標變度之大小,即可一目瞭然也.然

若二標誌度單位各殊,則宜分別求其相對散布量,如變率等而比較之[1].

有生之物,稟性自然,身體發達,雖各有特徵.而同類之中,則多一致.如人體發育,大抵相類,其他如同種植物,共族禽獸,莫不如是.此盡師所深知,非必待精測而後喻者.故研究同種材料,即可假定一致性之存在,一致性既經成立,則二標相依,至爲明顯.

所謂相依,非必如函數論中,自變數及倚數之密且切也.蓋標準互倚,在統計最常見者,爲相關互倚.至函數互倚,則不過理論上之名詞,從未見有出現者.

標度分佈,既非單義決定,一一對應,則對於一定之X,必有若干不同之Y與之相當,對於一定之Y,必有若干X與之呼應.然此種配合,苟非任意雜湊,則依一致性之理論,必有常規.此常規者,即諸Y分佈依X而定.諸X分佈,視Y而決,換言之,亦即X與Y之分佈,Y與X之分佈,有相互關係也.學者爲研究便利計,特取Y分佈之中值,或散布量,與一定之X相當,或取X之分佈,與一定之Y相應,而窮其相互之關係焉.

標準互倚,既與宇宙之一致性有關,故其探求,對於自然

---

(1) C. Gini: Delle relazioni tra le intensita cograduate di due Caratteri Atti del Reale Ist Veneto, 1916—1917, t. LXXVI, P. 1147—1185

界之認識及支配含有重大意義.而統計乃爲治科學者所必須.世之不用統計方法,惟取一二標本以定生物之形態大小者,其去自然豈可以道里計哉.至於進化論及遺傳學之有賴於統計者尤多,若一旦棄統計而尙空論,則不爲盲人捫象者幾希.

取X,Y之諸值及其相當屢數,加以排列,得雙檢表,表有交場,羅成網狀.此諸場者,以幾何言,爲二組垂直平行線割成之矩形.吾人若以類幅爲單位,則諸矩形化爲正方形,學者爲研究便利計,特名此表爲相關表.茲述其填注方法如下.

(一)若判值X,Y咸爲整數(如統計生物之器官),則屢數應填於中線交點之上,以示値偶之分佈.

(二)若X,Y爲連續變數(如測量器官之大小,統計人口之年齡),則各類屢數應注入交場之中,以示集團中個體屬於此類者之個數.

相關表有行有列,通常均排X於橫列,置Y於直行,而書Y之級度於左行,X之級度於眉列.至X應示何標,則無有定規,而可從心所好,又相關表與解析幾何中之經緯網類似,以X與橫軸相當Y,與縱軸相應也.

63. 取已成表格,注明級度,即可依原表,將集團個體按其X,Y之所在,分別用點或符號記入相當交場.然後逐場

計其點數而數示之,若集團之範圍甚大,則此種方法,易致遺誤.欲免其弊,可以卡片配示個體而書其相當判值於其上,先取 X 在同類者別爲一包,再將各包卡片依 Y 級度分爲小包.如是所得之包數,自應與實有之類數相同,一再檢驗,即可填入各場矣.

在標準爲連續變數時,若遇判值與類限或角點相合之個體,則在前者可將此元劈半,後者可將此元四裂,分隸於交界各場.於是類屢中,途有 .5, .25, .75 諸小數.此種小數,改換類限,自可除去一部,然欲完全除去,則殊非易事.

相關表註入屢數後,更分行計其總數,填入下列,則得 X 之分佈.分列計其總數,填入右行,則得 Y 之分佈.取各行和或各列和相加,則得集團之範圍.

64. 例(1)關於生物之統計.(a)衿取草花梗及花辯之數目,(b)衿取草花辯之數目與最長花辯之長,是例材料爲 C. V. L. Charbier[1] 於1912年七月八日至二十三日在瑞典Lund近郊湖畔採集而得.表 I 中二標級度均爲整數.表II 中一爲整數一爲絡數.

衿取草器官之互關.

I. 花梗及花辯之數目

(1) C. V. L. Charlier: A, Statistical Description of Trientalis europea Arkiv för Botanik 卷 XII (1913)

		花梗數			
		1	2	3	
花瓣數	5	119	6	—	125
	6	103	51	1	155
	7	10	16	2	28
	8	1	5	5	11
	9	—	—	2	2
		233	78	10	321

II. 花瓣數目及最長花瓣之長

		花瓣數					和
		5	6	7	8	9	
最長花瓣長	10	1	—	—	—	—	1
	15	4	1	—	—	—	5
	20	13	4	—	—	—	17
	25	24	11	1	—	—	36
	30	30	21	2	1	—	54
	35	21	26	4	1	—	52
	40	17	30	—	3	—	50
	45	6	18	6	2	—	32
	50	2	15	5	1	1	24
	55	1	3	3	—	—	7
	60	—	2	2	1	—	5
	65	—	2	1	1	1	5
	70	—	—	1	1	—	2
		119	133	25	11	2	290

二表意義可略述之如下:

(*a*)花梗數目最常見者為1,其相對屢數為 0.726.花瓣數目最常見者為6,其相對屢數為 0.483.又花瓣數目與花梗數目為同向增長,以三行之算術中值5.5,6.3,7.8遞次上昇也.

(*b*)最長花瓣長為27.5–42.5者占全體53.8%,此長與花瓣數目亦為同向增長.以由五行所得之算術中值,為30.9,38.2,46.6,47.3,57.5公釐依次遞昇也.

(2)父子之蕃殖力.是例標度咸為整數,其詳細情形如下表:

III.父子蕃殖力之相關

		父之子女數																	和
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
子之子女數	0	5	8	7	14	18	2	2	3	8	3	4	4		…	…	…	…	78
	1	3	3	6	5	8	8	6	5	4		2		1	…	…	…	…	51
	2	7	5	6	13	12	12	12	6	5	4	2	1	1	…	…	…	…	86
	3	5	10	13	11	17	13	13	12	10	4	1	2	1	…	…	…	…	112
	4	4	16	18	24	23	5	18	10	7	8	1	5	2	1	…	…	1	143
	5	9	8	11	14	16	12	16	12	2	5	8	6	3	…	2	…	…	124
	6	3	4	10	16	13	11	11	10	10	1	2	2	1	…	…	…	…	94
	7	5	6	8	7	10	14	11	12	4	7	1	1	…	…	…	…	…	86
	8	3	5	4	15	19	7	10	8	4	2	2	1	…	…	1	…	…	81
	9	1	6	5	9	5	5	8	5	4	3	3	1	…	…	…	…	…	55

10	2	3	9	3	2	5	4	6	4	3	1	1	1	...	...	...	...	44
11	1	1	1	4	2	1	1	2	1	1	...	...	...	...	...	...	...	15
12		...	2	2	2	...	1	1	1	2	...	1	...	...	...	...	...	12
13	1	...	...	1	3	...	1	1	1	1	...	1	...	1	...	...	...	11
14	...	1	...	...	...	1	...	...	...	...	1	...	...	...	...	...	...	3
15	...	...	1	...	...	...	...	1	...	...	1	...		...	...	...	...	3
16	.	...	...	...	...	...	...	...	1	1	...	...	...	...	...	...	...	2
和	49	76	101	138	150	96	114	94	66	45	29	26	10	2	3		1	1000

是表凡子已結婚十五年以上,或夫妻間有一去世不能再行生殖者,方行收入.父則不受此種限制,又每家僅取一子,上表材料卽取自英國Peer族者(1).

此集團之個體,爲父及子.父生兒數,最常見者爲五,子則爲四,卒易 5/4 實爲屡數積集之區.至其餘各處,則以集團範圍太小之故,屡數增減頗不規則,又此表屡數最大之場爲4/4.

依上表結果,父之蕃殖力似較子之蕃殖力爲大,一若輾轉相傳,此族有絕種之慮者.其實子之生殖不必小於其父,蓋吾人所取材料,其中之父莫不有子,而子則不必有子也.

(3)下列二例,亦係關於矜取草之統計.蓋就 C.V.L.Charlier 之材料,而研究其他器官互關之理者.

---

(1) K. Pearson; Phil. Trans. Roy. Soc., A, Vol. 192 (1899) P. 287 (iii) 321.

IV. 衿取草最長花瓣之厚及長

公釐		幹厚											和
		0.425	0.525	0.625	0.725	0.825	0.925	1.025	1.125	1.225	13.25	1.425	
最長花瓣之長	10.5	1	...	...	...	...	...	...	...	...	...	...	1
	16.5	1	4	1	1	...	...	...	...	...	...	...	7
	22.5	1	9	16	3	1	...	...	...	...	...	...	30
	28.5	..	2	9	22	9	2	1	...	...	...	...	45
	34.5	..	..	8	19	20	4	1	...	...	...	...	52
	40.5	1	...	...	7	18	12	6	4	...	...	...	48
	46.5	...	...	...	1	8	9	3	2	1	...	...	24
	52.5	...	...	...	...	...	3	6	4	1	...	...	14
	58.5	...	...	...	...	...	...	2	2	1	2	...	7
	64.5	...	..	...	...	...	...	...	...	1	3	...	4
	70.5	..	...	...	...	...	...	...	...	1	...	1	2
		4	15	34	53	56	30	19	12	5	5	1	234

V. 衿取草最長花瓣之長及其廣

公釐		寬												和
		5.5	7.5	9.5	11.5	13.5	15.5	17.5	19 5	21 5	23.5	25.5	27.5	
長	10	1	...	...	...	...	...	...	...	...	...	...	...	1
	15	1	3	...	...	...	...	...	...	...	...	...	...	4
	20	1	2	7	...	...	...	...	...	...	...	...	...	10
	25	...	4	14	6	...	...	...	...	...	...	...	...	24
	30	...	...	2	17	3	3	1	...	...	...	...	...	26
	35	...	...	...	2	11	9	1	...	...	...	...	...	23

40	···	···	···	···	10	8	7	1	··	···	···	···	26
45	···	···	···	···	1	8	5	2	1	1	1	···	19
50	···	···	···	···	···	2	1	5	1	2	···	···	11
55	···	···	···	···	···	···	···		1	···	···	··	1
60	···	···	···	···	···	···	···	3	···		1	··	4
65	···	···	···	···	···	···	···	···	···	··	···	1	1
70	···	··	···	···	···	···	···	···	···	1	1	···	2
和	3	9	23	25	25	30	15	11	3	4	3	1	152

表中X,Y為類中點,又二表所示之分佈,與以前所見者不同,以諸質類咸在直線上,而此直線又係相關表之對角線也.

(4)下述範圍較大之連續集團,亦係自植物界取出,以示常春藤葉長廣之互關者,表中所用單位為$\frac{1}{8}$英寸,類幅倍之.

VI.常春藤葉長廣之關係

	長 X													和
	2.95	4.95	6.95	8.95	10.95	12.95	14.95	16.95	18.95	20.95	22.95	24.95	26.95	
2.95–4.95	7	13	···	···	···	··	···			···	···	··		20
4.95–6.95	26	97	18	1	···	···	···	···	···		···			142
6.95–8.95	3	101	106	37	4	···	···				···			251

Y 寬	10.95	3	33	190	152	31	4	2	…	…	…	…	…	415
	12.95	…	7	88	227	98	16	2	…	1	…	…	…	439
	14.95	…	1	26	137	216	66	7	…	…	…			453
	16.95	…	…	…	55	136	104	22	6	1	…	…	…	324
	18.95		…	3	11	50	89	49	9	2	…	…	·	213
	20.95	…	…	…	4	17	43	31	11	5	1	…	…	112
	22.95	·	·	…	…	4	9	21	17	4	1	1	…	57
	24.95	…	…	…	…	1	…	3	9	6	1	…	1	21
	26.95	…		…	…	…	··	4	11	7	5	1	··	28
	28.95	··	…	…	…		…	…	4	2	4	2	…	12
	30.95	…	…	…	…	·	…	…	1	2	1	3	…	7
	32.95	…	…	…	…	…	·	…		1	2	…	…	3
	34.95	…	…	…	…	…	…	…	…	…	…	2	··	2
	36.95	…	…	…	··	…	…	…	…	…	…	1	…	1
和		39	252	431	624	557	331	141	68	31	15	10	1	2500

集團範圍爲2500葉,爲自100株分產於二地之常春藤,各採25葉而得者,所謂葉廣,則係依葉長方向所作平行切線間之距離.

表中葉長,則以長9.95者爲多,葉廣則以廣13.95爲衆.至

交場中之屢數最大者,則為 9.95/11.95 其餘屢數咸向此處集中.

65. 雙標分佈,既有相關表說明其情形,更可作圖形括示其大勢.其法與前述關於單標分佈者相當.惟此處不能以平面圖形表示,而須用空間觀念耳.就相關表之諸場中點,各豎表面垂線,使其長與該類屢數相當,則垂線端點之起伏,卽表分佈之情狀.若所有標度咸爲整數,則此圖所示卽實際分佈.若標度爲綿續變數,則當範圍繼續增大,交場無限縮小,此諸端點漸與曲面接近,而成屢數曲面.取此面所包之體積全部爲單位,以測其中任一柱體之體積,卽得以該柱底爲類域之相對類屢數.爲觀察便利計,吾人更可依圖義,作成分佈模型.其法取已成木板,列相關表於其上,然後於各場中點,樹長與屢數相當之竿,各插小球於其端.觀諸球之起伏,卽知分佈之大概.

表示分布,除屢數曲面之外,亦有用稜柱體者,其法以各場爲底,相當屢數爲高,分立稜體.所得結果,卽示集團分佈之情形.惟此種方法,僅利於模示,而不適於圖表,以稜柱長短參差,不能全體繪出也.

前篇關於梯級圖多邊形及屢數曲線諸理論,對於屢數曲面及稜柱體,尤爲適用.如云統計圖形,必須範圍變更,始終如一者,方能表示材料之特性.此理對於屢數曲線爲眞,

對於屢數曲面則尤異也.

模範屢數曲面者,與模範屢數曲線相當,便於分佈比較之屢數曲面也.此面爲模範屢數曲線繞稱軸旋轉而成,吾人若以稱軸爲Z軸,横軸爲X軸,則得此面之方程式爲

$$Z=\frac{h^2}{\pi}e^{-h^2(x^2+y^2)}. \tag{1}$$

一切與Z軸平行之平面,不必通過此軸,即可割此面成模範屢數曲線,且諸曲線之參數咸同.讀者試取與zx面平行之平面割之,則得參數爲C之模範曲線

$$Z=Ce^{-h^2x^2}, \tag{2}$$

式中$C=\frac{h^2}{\pi}e^{-h^2y^2}$, y爲常數,舉一反三,當知其然矣.

單標分布之適合模範曲線者既鮮,雙標分佈之適合模範曲面者尤屬寥寥,惟近似言之,則頗不乏其例耳.

66. 雙關表之意義既甚重要,今請就其算法更進論之.計算之法,先取X及Y之分佈,逐行逐列依研究單標分佈之方法,分別求其中值及散布量,然後以此法行之於行和及列和,茲爲計算便利計,僅求其算術中值及均方差已足,由是

(1)對於一定之X必有一$M_Y^{(X)}$及一$\mu_Y^{(X)}$與之相當,此$M_Y^{(X)}$或$\mu_Y^{(X)}$之個數與所有行數相等.

(2)對於一定之Y,必有一$M_X^{(Y)}$及一$\mu_X^{(Y)}$與之相當,此$M_X^{(Y)}$或$\mu_X^{(Y)}$之個數,與所有列數相同.

(3)M_X表一切X之算術中值,μ_X表一切X之均方差.

(4)M_Y表一切Y之算術中值,μ_Y表一切Y之均方差.

茲取與64節表III相似之例,依上述方法以研究母女生殖力之關係.此例材料[1]取自英國 Barnet 及 Peer 二貴族.所取母女最少須結婚十五年.家有數女者,則任取其一人算.據此表作成之模型,極不規則,以其交遇皆非正方形也.

觀$M_Y^{(X)}$之起伏,知母女之生殖力為同向增長.惟至育11子之母,忽然發更,自是以下,起伏無常,以其範圍大小,結果遂不可用.

由$M_X^{(Y)}$之數列,亦可知女之生殖力大者,母之生殖力亦大,惟自身育九子之女以下,中值起落無常,其結果亦不可靠.

由是可知女性之生殖力,有遺傳性.

又$M_Y^{(X)}$及$M_X^{(Y)}$之差異,源於材料擇取之欠均,以諸女有無子者,而諸母則莫不有子也.

由諸行和及諸列和得X及Y之全體算術中值及均方差為:

$$M_Y=4.34 \qquad M_X=5.90$$

$$\mu_Y=2.97 \qquad \mu_X=2.83.$$

與前得諸值相較,可知均方差成相差不遠,惟在子數甚多之母女,則成例外.

VII.母女之生殖力

	母之子女數 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	和	$M\binom{y}{x}$	$\mu\binom{y}{x}$
女之子女數 0	5	9	11	18	21	15	8	9	6	3	2	3	…	…	…	…	110	5.39	2.57
1	12	5	14	15	10	13	9	8	5	3	2	2	…	…	…	…	98	5.10	2.77
2	9	9	10	15	18	15	9	3	2	4	2	…	…	…	…	1	91	4.89	2.74
3	5	10	16	11	9	14	13	10	4	8	2	…	…	…	…	…	105	5.63	2.83
4	5	5	19	17	21	15	18	10	14	2	1	5	1	…	…	…	133	5.80	2.64
5	7	6	7	17	23	9	12	13	14	8	3	2	2	…	…	…	123	6.13	2.83
6	4	5	6	11	15	12	15	14	7	5	3	3	1	…	…	…	103	6.22	2.70
7	5	4	3	8	4	13	9	8	5	10	2	1	1	…	…	…	73	6.48	2.86
8	1	2	4	12	9	9	8	5	12	3	4	1	2	1	…	…	73	6.77	2.84
9	…	…	4	3	3	4	7	5	3	2	2	1	…	…	…	…	34	4.85	2.41
10	…	…	1	2	1	3	4	6	3	2	…	1	…	1	…	…	24	5.63	2.43
11	…	…	2	1	1	1	…	…	1	2	…	…	…	…	…	…	8	4.25	2.82
12	…	2	1	2	3	…	1	1	…	…	1	…	1	…	1	…	13	5.46	4.01
13	…	…	…	…	2	1	…	…	…	…	1	…	2	…	…	…	6	4.83	3.58
和	53	57	100	192	140	124	113	92	76	52	25	22	10	2	1	1	1000		
$M\binom{x}{y}$	3.15	3.53	3.74	402	4.09	4.15	4.69	485	5.11	5.13	5.52	423	8.10	9.00	12.00	200			
$\mu\binom{x}{y}$	2.28	2.79	2.79	2.94	3.03	2.96	2.77	2.95	2.74	2.84	3.43	2.70	3.24	1.00	0	0			

(1) K. Pearson, Phil. Trans. Roy, Soc., A vol. 192 (1899),P. 285,319.

取64節表III而計算其算術中值,及均方差得:

X	$M_Y^{(X)}$	$\mu_Y^{(X)}$	Y	$M_X^{(Y)}$	$\mu_X^{(Y)}$
1	4.55	3.21	0	5.54	3.10
2	4.76	3.05	1	5.65	2.62
3	5.12	3.16	2	5.62	2.73
4	4.98	3.09	3	5.68	2.69
5	4.70	3.14	4	5.65	3.08
6	5.11	2.88	5	6.22	3.37
7	5.19	2.72	6	5.66	2.96
8	5.62	3.05	7	5.87	2.63
9	5.00	3.59	8	5.69	2.58
10	5.98	3.59	9	5.56	3.33
11	5.31	3.85	10	6.02	3.07
12	4.97	3.33	11	5.33	2.52
13	3.50	2.33	12	6.67	2.97
14	8.5	4.5	13	7.27	3.62
15	6.5	1.5	14	6.33	6.38
16			15	7.33	5.71
17	4	0	16	9.5	0.5

$M_Y=5.07$ $M_X=5.83$

$\mu_Y=3.14$ $\mu_X=2.91.$

此種結果,頗乏常規,且以其均方差過大故不可用.惟吾人若取相鄰之$M_Y^{(X)}$,每三數相加則得,

14.43 14.79 15.81 16.26 18.60

顯示Y之中值逐漸增加,然以此施於$M_X^{(Y)}$則否,以

16.81 17.55 17.22 16.91

非單向增長也.

總觀上列結果,男性生殖力遺傳與否,頗難決定,以其相互關係太不明顯也.

67. 欲進論相互關係,須應用幾何圖形,今擇要述之如下:就雙關表爲經緯制,點示各中值於其上,則得二點列,一表行中值,一示列中值.

若所論分佈爲理想分佈,則其行中值適在一與X軸平行之直線上,列中值適在一與Y軸平行之直線上.此二線之交點,得爲相關中點,亦名雙關表中點.此點與屢數曲面之峯點相當,其經緯XY有最大之屢數.斯時二標成爲獨立,並無相互關係,以對於一切X,Y之中值不變,對於一切Y,X之中值不變也.

在事實上,此種情形絕少出現,即最合理想之分佈其結

果亦多少與此相差.至就一般論,則中值點列成二曲線.吾人若求出曲線之解析方程式,使各與相當點列套合,則屬某X之行中值,或屬某Y之列中值,咸可由相當方程式直接算出.然曲線方程,求出不易,故此種說法,僅有理論上之價值而已.

在多種情形,中值位置,雖不必與理想分佈吻合,然若其散布得法,則吾人亦可以直線套合之.茲特論之如次.

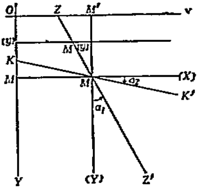


圖21 線性互關之基線

設行中值確在一直線KK'上,列中值確在一直線ZZ'上,則其交點M為全分佈之中值.過M與軸平行作二直線,與軸交於M', M''.茲設M'與一切Y之中值M_Y, M''與一切X之中值M_X相當(圖21)以x表X對M''之偏差,以y表Y對M'之偏差,b_1表ZZ'對$M(Y)$夾角之正切,b_2表KK'對$M(X)$夾角之正切則$M_X^{(Y)}$為直線 $y)$ 上一切X之中值,$M_Y^{(X)}$為直線(x)上一切Y之中值,而有

$$\Sigma^{(y)}(x)=mb_1y;\quad \Sigma^{(x)}(y)=nb_2x.$$

式中m為X沿(y),n為Y沿(x)之歷數和,取首式右端一切y,左端一切x之和,得

$\Sigma(nx)=b_1\Sigma(my)$,同理有 $\Sigma(my)=b_2\Sigma(nx)$,

由是得 $\Sigma(nx)$, $\Sigma(my)$ 之齊次線性方程式爲:

$$\Sigma(nx)-b_1\Sigma(my)=0$$

$$b_2\Sigma(nx)-\Sigma(my)=0.$$

依假定二直線相交於 M,即 $K\,K'$, $Z\,Z'$ 不相疊.合而 $1-b_1\,b_2$不等於零,換言之即二式之係數行列式不爲零,而

$$\Sigma(nx)=0 \qquad \Sigma(my)=0$$

也.故M'表一切Y之中值M_Y,M''表一切X之中值M_X.

設中值M_X,M_Y爲已知,則M之位置亦定,而b_1,b_2之值即可由是求出,茲述其求法如下:設p爲值偶xy之乘積中數.N爲集團之範圍,則

$$\Sigma(xy)=Np. \tag{3}$$

(3)式左端求法有二,一先依固定之y求x之和.二先依固定之x求y之和,然後再求總和,由是得

$$\Sigma\Sigma^{(y)}(xy)=\Sigma(y\Sigma^{(y)}(x))=b_1\Sigma(my^2)=Nb_1\mu_Y^2$$

$$\Sigma\Sigma^{(x)}(xy)=\Sigma(x\Sigma^{(x)}(y))=b_2\Sigma(nx^2)=Nb_2\mu_X^2.$$

式中μ_Y,μ_X爲和列及和行之均方差,代入(3)式得

$$b_1=\frac{p}{\mu_Y^2} \qquad b_2=\frac{p}{\mu_X^2}. \tag{4}$$

應用積和$\Sigma(xy)$以定$K\,K'$及$Z\,Z'$之法,爲 A. Bravais 所創,其後 F. Galton 復依其理造相關率[1]

(1) A. Bravais: Analyse mathématique sur les probabilités des erreurs de sitnation d'un point. Mem. presentés par divers savants II e ser., t. IX (1846) P. 255. F. Galton: Corretations und their measurement.

$$r=\frac{p}{\mu_X\mu_Y}, \tag{5}$$

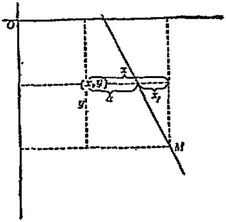
因均方差係絕對值,故r,p之符號相同,由是b_1,b_2又可以r表之如下:

$$b_1=r\frac{\mu_X}{\mu_Y} \qquad b_2=r\frac{\mu_Y}{\mu_X}. \tag{6}$$

以M爲原點,過此與舊軸平行,作直線爲新軸,則得ZZ'及KK'之方程式爲

$$\xi=r\frac{\mu_X}{\mu_Y}\eta \qquad \eta=r\frac{\mu_Y}{\mu_X}\xi. \tag{7}$$

上述b_1,b_2求法,係先假定中值確在直線上者.然嚴格論之,此種情形絕少出現.即在最合理想之分佈,其位置亦多少與直線相差.故b_1,b_2之決定,似宜自他點出發.設諸行(列)中值不完全在一直線上,吾人欲定一直線,使之與中值位置最爲相近.則此種條件,當各元對於此線偏差平方之總和爲最小時,自能滿足.至所謂偏差,則對ZZ'者爲順列向測得之距離.對KK'者爲依行向測算之結果.



$$a=x-x_1=x-b_1y$$

依前段所述.取M爲原點,過此所作之二直線其方程

式可書爲 $\xi=b_1\eta$, $\eta=b_2\xi$. 今點 (x,y) 對直線 $\xi=b_1\eta$ 之偏差爲 $x-b_1y$, 對直線 $\eta=b_2\xi$ 之偏差爲 $y-b_2x$. (x,y) 之既數爲 Z, 故二直線應滿足之條件爲:

$$\Sigma[z(x-b_1y)^2]=\text{極小}$$

$$\Sigma[z(y-b_2x)^2]=\text{極小}.$$

取首式對 b_1 求其微分,得

$$\Sigma[z(x-b_1y)y]=\Sigma(xy)-b_1\Sigma(xy^2)=Np-b_1N\mu_Y^2,$$

命之爲零,得 $b_1=\dfrac{p}{\mu_Y^2}$.同理有 $b_2=\dfrac{p}{\mu_X^2}$.

故無論分佈之情形若何,(4)式及(6)式咸決定通過 M 與中值位置最爲接近之二直線.至中值位置何時可以直線配合,配合結果,何時方能應用.則須作者審情度勢,自行決定,固無一定之規律.惟統計材料,千差萬別,中值位置,不必盡可以直線套合.此又學者宜特留意者也.至其他表示中值之方法,以後當更述之.偏差極小量,足示近似之程度,欲求其值,可取首式展開即得

$$\Sigma(x^2)-2b_1\Sigma(xy)+b_1^2\Sigma(y^2)=N\mu_X^2-2b_1Np+b_1^2N\mu_Y^2,$$

以(4)式代入,有

$$N\left[\mu_X^2-2\frac{p^2}{\mu_Y^2}+\frac{p^2}{\mu_Y^2}\right]=N\left[\mu_X^2-\frac{p^2}{\mu_Y^2}\right],$$

Proc. Roy. Soc. XIV (1888), P. 135. 繼續研究者有 F. Y. Edgeworth, K. Pearson 及 G. U. Yule 諸人.

由(5)式得

$$\Sigma[z(x-b_1y)^2]\text{之極小值}=N\mu_X^2(1-r^2); \qquad (8)$$

同理

$$\Sigma[z(y-b_2x)^2]\text{之極小值}=N\mu_Y^2(1-r^2). \qquad (9)$$

式中 z 及平方數咸爲正數,其積之和亦爲正數,因之 r^2 恆小於 1 而以 1 爲上限.

r 大小旣與 b_1, b_2 有關,而 b_1, b_2 又爲 ZZ', KK' 之方位正切,故甄別相關之種類及級度,莫如利用 ZZ' 及 KK' 之夾角 Θ, 由圖 22

$$\tan\Theta=\frac{\frac{1}{b_2}-b_1}{1+\frac{b_1}{b_2}}=\frac{1-b_1b_2}{b_1+b_2}=\frac{1-r_2}{\left(\frac{\mu_X}{\mu_Y}+\frac{\mu_Y}{\mu_X}\right)r}. \qquad (10)$$

式中 $1-r^2$ 及 $\frac{\mu_X}{\mu_Y}+\frac{\mu_Y}{\mu_X}$ 咸爲正數,故 $\tan\Theta$ 之符號,與 r 所

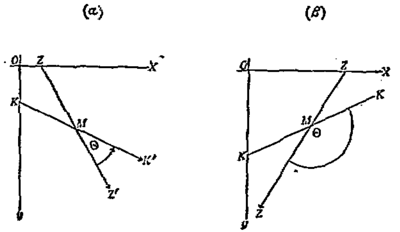


圖 22 正互圖及負互圖

有者相同.當 r 爲正則 Θ 爲銳角（圖22α),斯時甲變數之中值與乙變數爲同向增長.此種互關,卽名爲正互關.當 r 爲負,則 Θ 成鈍角（圖22β),斯時二變數爲逆向增長.此種互關,卽名爲負互關.

在特例若 r 爲零,則中值軸與經緯軸平行,而 Θ 成一直角.斯時或稱二標互爲獨立,其實此種論斷,頗無道理,蓋所謂 $r=0$ 者,僅就一定之計算精確度而言,是時中值軸與經緯軸之夾角雖小,未必眞等於零,惟學者爲稱呼便利計,特謂之爲無關耳.

又 $r=+1$ 或 -1 之特例,在事實上亦少出現,是時二中值軸疊合爲一,學者名之爲完全互關.當 $r=+1$ 時,二標爲同向增長,（圖23α),當 $r=-1$ 時一長一消（圖23β).

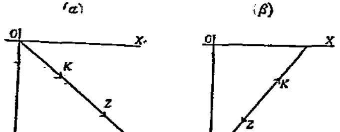


圖23完全正互關及完全負互關

68. 相關率 r 之意義及名稱,已如上述.茲更就(5)式取其數量性質論之.依定義,p 爲二度,$\mu_X\mu_Y$ 亦爲二度,因之

其比值 r 爲絕對量,故在同一材料,無論所定之單位如何,倘 x y, μ_X 及 μ_Y 均以同一單位計算,則所得之 r 恆同.

至於 b_1, b_2 之意義,則前者表各列中值 $M_X^{(Y)}$ 對公共中值 M_X 之偏差,後者表各行中值 $M_Y^{(X)}$ 對公共中值 M_Y 之偏差. b_1, b_2 二者最少有一須爲小數,以(6)式之中, r 既小於 1,而 $\frac{\mu_X}{\mu_Y}, \frac{\mu_Y}{\mu_X}$ 互爲倒數,倘不同時爲 1,最少有一須小於 1 也.又 b_1, b_2 僅當 X, Y 爲同度時爲絕對值,至其符號則恆與 r 所有者相同.

b_1, b_2 Galton 稱爲返祖係數. Galton 以爲體格遺傳有反祖現象,譬如父親高於常人一寸,則其子必退而與常人齊高.設 X 爲父之體高, Y 爲子之體高, x 爲 X 對 M_X 之偏差, y 爲 Y 對 M_Y 之偏差.則 $y=b_2x, x=b_1y$ 然 $b_1 \cdot b_2$ 通常不等於 1,故已知之值,與計算結果,大小各殊.且此種臆說,後經證明,與事實不合.故 b_1, b_2 之名稱,亦嫌失當,又 b_1, b_2 所定之直線, Galton 稱爲反祖直線.

G. U. Yule 以Galton所定名稱未能盡善,因改名 ZZ', KK' 爲特性線,(7)式爲特性方程.譯者依西名未當,則中名另定之原則,特名 b_1, b_2 爲消長係數,(7)式爲消長方程, ZZ', KK' 爲消長直線.

至(7)式之主要功用,則爲已知甲標之數量,即可估計乙標之平均數量.至此種估計之精確度,則以直線配合之精

確度爲轉移,換言之亦卽以(8),(9)所示極小量之大小爲轉移也.以N除(8),(9)二式,再開平方,而以m_X及m_Y表其結果,則得精密度之估計量爲

$$m_X=\mu_X\sqrt{1-r^2} \qquad m_Y=\mu_Y\sqrt{1-r^2}. \tag{11}$$

與 Galton 之觀念不同,而對於相關性,別具見解者,爲W. Wirth.氏曾於 E. Abderhalden: 之『生物研究法 (Handbuch der biolgischen Arbeitsmethoden, Berlin-Wien, 1920)』中發表其『體質心理學上的專門測算方法(Spezielle psychophysische Mabmlthoden),』其中最要之問題,爲求二相關量之變化比值,換言之,亦卽求以直線套合觀察點X/Y,而定其方向正切也.

如用消長直線之理論,以解決此問題,則其缺點有三:

1. 同爲一物,解答有二.

2. 每次運算,須先設一標爲固定,然後計算他標之偏差,同爲標準,待遇懸殊.

3. 所得結果,與經緯有關,若更換新軸,其結果勢將變易.

凡此種種,皆與問題本性相違.蓋吾人所求者,單義確定之解答也.

此種解答,W. Wirth 曾以他法得之.氏以爲當各點與一線之距離平方和爲極小時,則此直線與觀察點最爲接近

吾人若能求出此線而以平均直線名之,則其方位正切,即為二相關量之平均變率.至平均直線之位置,則恆在二消長直線之間.

統計材料之有平均變率者,其觀察點咸向一直線集中.斯時二消長直線漸相接近,其傾斜度亦相差甚微,反之則否.Wirth由是察出相關率與解答歧異之關係.相關率 r 愈大,則消長線相差愈小,而歧異愈微.當 $r=1$ 時,則歧異完全消失.而消長線與平均線合而為一,是時二標不僅有相互之關係,且成為線性之函數互倚[1].

第五章　兩標相關性應用

69. 前述理論之應用,與問題性質,範圍大小有關,如僅求兩標間是否有相關性,及此相關性之種類,則當集團之範圍甚小時,可行之如次:取 X 依昇向排列,各配以相當之 Y,若 Y 值亦為漸昇,則二標間之相關為正.若 Y 值依序下降,則此相關為負.若 Y 值昇降無定,則相關種類,非經計算不能決定.

依 Charlier 研究衿取草器官之結果,其中11株,根莖之

(1) Writh見解與「標誌相關」之關係,詳見以後120節.又學者如欲對此問題,及相關率構造度消長線,平均線之方向正切,再加研究,可參看(Wirth原文第§16 c).及 E. Czuber:「線性互關之理論」(Zur Theorie der Linearen Korrelation, Arkiv für Psychologie, Bd. XLI, 1921).

長(X)厚(Y)為:

X (公釐)	Y (公釐)
3	2.0
5	2.3
5	2.2
6	2.0
7	1.4
7	1.6
7	3.0
8.5	1.9
9	2.0
10	3.0
10	2.6

因Y值昇降不一,故二標之互關不能直接察出.

其中花13枝之雄蕊長(X),及雌蕊長(Y)為:

X (公釐)	Y (公釐)	
3.3	4.0	
4.0	4.4	15.3
4.1	3.5	

4.2	3.4	
4.2	4.8	
4.2	4.8	19.3
4.3	5.1	
4.4	4.6	
4.6	5.4	
4.9	3.8	19.8
5.0	5.1	
5.0	5 5	
5.2	5.2	

數序Y中雖時有起伏,然吾人若取相鄰四數歸爲一組,則易知二標之相關爲正.

欲求r之近似值,可先定$M_X^{(Y)}$,$M_Y^{(X)}$,及M_X,M_Y,過M引絲線以之繞M而轉,使與點列$M_X^{(Y)}$ $M_Y^{(X)}$套合,務期各線兩側之乖差可以互相消去.俟此情形達到,即依絲線位置,畫定直線,然後測其夾角α_1,α_2(圖21),以定方向正切b_1,b_2.由(6)式有

$b_1 b_2=r^2$,將b_1,b_2之值代入此式,則得

$r=\sqrt{b_1b_2}$.至r之符號,則當α_1,α_2爲正時爲正,爲負時爲負.

70. 至欲實行計算r,則首須求定M_X, M_Y, μ_X, μ_Y.蓋若M_X, M_Y爲已知,則相關表之中點亦定,由此即可計算p值

惟p之計算,最爲繁難.其直接算法,爲先定類X/Y對於中點M之經緯量$x=X-M_X, y=Y-M_Y$.再依符號求其乘積,然後以相當屢數Z倍之,以N除由此所得諸結果之和,即得乘積中值p.

但此種運算極易致誤,且嫌繁冗.故應用此法,頗多不便,茲述一較爲簡便之間接求法如下:

取M所在之類中點$\mathfrak{m}$(1)爲原點,類幅爲單位,以定各點對$\mathfrak{m}$之經緯.則如是所得之$\mathfrak{x}, \mathfrak{y}$咸爲整數,乘積$\mathfrak{x}.\mathfrak{y}$與屢數$Z$相乘之結果即可直接書出.以$N$除諸積之和,則得$p=\frac{1}{N}\Sigma(z\mathfrak{x}\mathfrak{y})$.設$\xi, \eta$爲$\mathfrak{m}$對$M$之相當經緯,$x y$爲點$\mathfrak{x}/\mathfrak{y}(X/Y)$對$M$之經緯,則得

$$\mathfrak{x}=x-\xi, \qquad \mathfrak{y}=y-\eta$$

$$\mathfrak{x}\mathfrak{y}=xy-\eta x-\xi y+\xi\eta,$$

$$\Sigma(z\mathfrak{x}\mathfrak{y})=\Sigma(zxy)-\eta\Sigma(zx)-\xi\Sigma(zy)+\xi\eta\Sigma(z)$$

但M爲表之中點,有$\Sigma(zx)=0, \Sigma(zy)=0$,故

$$\Sigma(z\mathfrak{x}\mathfrak{y})=\Sigma(zxy)+N\xi\eta,$$

以N除之,得

$$P-p+\xi\eta$$

或 $$p=P-\xi\eta. \tag{12}$$

式中各值咸以類幅爲單位.

(1) 若所論爲間積集團,則$\mathfrak{m}$爲M類之值點.

由是可知欲求以M爲出發點之乘積中值p,可先求以m爲出發點之乘積中值P,再自此值減去M對m之相對經緯乘積$\frac{P}{n}$卽得,惟運算時,宜特別注意各數之符號.

71. 茲取種類不同級度各異之相關略舉數例如下:

相關表IV 相關率之定法

公篩		幹厚											和
		0.425	0.525	0.625	0.725	0.825	0.925	1.025	1.125	1.225	1.325	1.425	
花瓣之長	10.5	${}_{1}16$	…	…	…	…	…	…	…	…	…	…	1
	16.5	${}_{1}12$	${}_{4}9$	${}_{1}6$	${}_{1}3$	…	…	…	…	…	…	…	7
	22.5	${}_{1}8$	${}_{5}6$	${}_{16}4$	${}_{7}2$	${}_{1}0$	…	…	…	…	…	…	30
	28.5	…	${}_{6}3$	${}_{5}2$	${}_{22}1$	${}_{5}0$	${}_{6}1$	${}_{1}2$	…	…	…	…	45
	34.5	…	…	${}_{12}0$	${}_{15}0$	${}_{20}0$	${}_{4}0$	${}_{1}0$	…	…	…	…	52
	40.5	${}_{1}4$	…	…	${}_{7}1$	${}_{15}0$	${}_{15}1$	${}_{6}2$	${}_{4}3$	…	…	…	48
	46.5	…	…	…	${}_{1}2$	${}_{15}0$	${}_{2}2$	${}_{3}4$	${}_{2}6$	${}_{1}8$	…	…	24
	52.5	…	…	…	…	…	${}_{3}3$	${}_{6}6$	${}_{4}4$	${}_{1}12$	…	…	14
	58.5	…	…	…	…	…	…	${}_{2}8$	${}_{2}12$	${}_{1}16$	${}_{2}20$	…	7
	64.5	…	…	…	…	…	…	…	…	${}_{1}20$	${}_{3}25$	…	4
	70.5	…	…	…	…	…	…	…	…	${}_{1}24$	…	${}_{1}36$	2
		4	15	34	53	56	30	19	12	5	5	1	234

例(1)取第64節(3)表IV所述衿取草之莖厚X,與其最長花瓣之長Y,而研究其相互之關係.

由計算結果,得運算之基本量爲:

$M_X=0.814$公釐 $\mu_X=1.894$類幅$=0.189$公釐.

$M_Y=36.244$公釐 $\mu_Y=1.825$類幅$=10.95$公釐.

吾人若自m, 0.825/34.5出發,以粗線劃定通過m之行列,更於各場之中,分別以大字計其偏差乘積,小字計其相當度數,則粗線以內各場之乘積,概等於零.至各積之符號,則在十字右腋象限(I)及左肩象限(III)內者爲正,在左腋象限(II)右肩象限(IV)內者爲負.

依各場所在象限,以次取其大小數字之積,而求其代數和,則有

I		III		II	IV		
12	12	22	8	7	2		
12	16	18	3	4	2		
12	24	6	6	2	−4		
18	16	6	36	−13		I	+430
12	40	64	12			III	+251
12	20	54	16				+681
8	75		+251			II	− 13
9	24					IV	− 4
36	36						+664
36	+430						

由是得

$P=664:234=2.8376$

又

$$\xi=\frac{0.814-0.825}{0.1}=-0.11,\ \eta=\frac{36.244-34.5}{6}=0.2906\text{類幅}$$

$$\xi\eta=-0.0320,$$

故 $p=2.8376+0.0320=2.8696.$

$$r=\frac{2.8696}{1.894\cdot1.825}=0.830.$$

由是知二標為強大之正關,又此種結論,亦可自表IV用69節方法直接求出.

(2)就表V所示關於衿取草之統計,加以運算,則得其主要結果為:

$M_X=14.02$ 公釐, $\mu_X=2.197$ 類幅$=4.39$ 公釐.

$M_Y=35.53$ 公釐, $\mu_Y=2.222$ 類幅$=11.11$ 公釐.

$p=4.3668$ $r=0.894.$

由是知衿取草最長花瓣之長廣,為強度之正互關.

(3)取64節例(2)表III,父子蕃殖力,而細行研究,則二標之類幅為一,不須變換單位,即可直行計算.由66節得其主要結果:

相關表III 相關率之決定

		父之子女數 X																
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
子女之子女數 Y	0	$\overset{5}{25}$	$\overset{8}{20}$	$\overset{7}{15}$	$\overset{14}{10}$	$\overset{18}{5}$	$\overset{2}{0}$	$\overset{2}{5}$	$\overset{3}{10}$	$\overset{8}{15}$	$\overset{3}{20}$	$\overset{4}{25}$	$\overset{4}{30}$	…	…	…	…	…
	1	$\overset{3}{20}$	$\overset{2}{16}$	$\overset{6}{12}$	$\overset{5}{8}$	$\overset{8}{4}$	$\overset{8}{0}$	$\overset{6}{4}$	$\overset{5}{8}$	$\overset{4}{12}$	…	$\overset{2}{20}$	…	$\overset{1}{28}$	…	…	…	…
	2	$\overset{7}{15}$	$\overset{5}{12}$	$\overset{6}{9}$	$\overset{13}{6}$	$\overset{12}{3}$	$\overset{12}{0}$	$\overset{12}{3}$	$\overset{6}{6}$	$\overset{5}{9}$	$\overset{4}{12}$	$\overset{2}{15}$	$\overset{1}{18}$	$\overset{1}{21}$	…	…	…	…
	3	$\overset{5}{10}$	$\overset{10}{8}$	$\overset{13}{6}$	$\overset{21}{4}$	$\overset{17}{2}$	$\overset{13}{0}$	$\overset{13}{2}$	$\overset{12}{4}$	$\overset{10}{6}$	$\overset{4}{8}$	$\overset{1}{10}$	$\overset{2}{12}$	$\overset{1}{14}$	…	…	…	…
	4	$\overset{4}{5}$	$\overset{10}{4}$	$\overset{18}{3}$	$\overset{24}{2}$	$\overset{23}{1}$	$\overset{5}{0}$	$\overset{15}{1}$	$\overset{10}{2}$	$\overset{7}{3}$	$\overset{8}{4}$	$\overset{1}{5}$	$\overset{5}{6}$	$\overset{2}{7}$	$\overset{1}{8}$	…	…	$\overset{1}{11}$
	5	$\overset{9}{0}$	$\overset{8}{0}$	$\overset{11}{0}$	$\overset{14}{0}$	$\overset{16}{0}$	$\overset{12}{0}$	$\overset{16}{0}$	$\overset{17}{0}$	$\overset{2}{0}$	$\overset{5}{0}$	$\overset{8}{0}$	$\overset{6}{0}$	$\overset{3}{0}$	…	$\overset{2}{0}$	…	…
	6	$\overset{3}{5}$	$\overset{4}{4}$	$\overset{10}{3}$	$\overset{16}{2}$	$\overset{13}{1}$	$\overset{11}{0}$	$\overset{11}{1}$	$\overset{10}{2}$	$\overset{10}{3}$	$\overset{1}{4}$	$\overset{2}{5}$	$\overset{2}{6}$	$\overset{1}{7}$	…	…	…	…
	7	$\overset{5}{10}$	$\overset{6}{8}$	$\overset{8}{6}$	$\overset{7}{4}$	$\overset{10}{2}$	$\overset{14}{0}$	$\overset{11}{2}$	$\overset{12}{4}$	$\overset{4}{6}$	$\overset{7}{8}$	$\overset{1}{10}$	$\overset{1}{12}$	…	…	…	…	…
	8	$\overset{3}{15}$	$\overset{5}{12}$	$\overset{4}{9}$	$\overset{15}{6}$	$\overset{12}{3}$	$\overset{8}{0}$	$\overset{10}{3}$	$\overset{8}{6}$	$\overset{4}{9}$	$\overset{3}{12}$	$\overset{2}{15}$	$\overset{1}{18}$	…	…	$\overset{1}{27}$	…	…
	9	$\overset{1}{20}$	$\overset{6}{16}$	$\overset{5}{12}$	$\overset{9}{8}$	$\overset{5}{4}$	$\overset{5}{0}$	$\overset{8}{4}$	$\overset{5}{8}$	$\overset{4}{12}$	$\overset{4}{16}$	$\overset{3}{20}$	$\overset{1}{24}$	…	…	…	…	…
	10	$\overset{2}{25}$	$\overset{3}{20}$	$\overset{9}{15}$	$\overset{3}{10}$	$\overset{2}{5}$	$\overset{5}{0}$	$\overset{4}{5}$	$\overset{6}{10}$	$\overset{4}{15}$	$\overset{4}{20}$	$\overset{1}{25}$	$\overset{1}{30}$	$\overset{1}{35}$	…	…	…	…
	11	$\overset{1}{30}$	$\overset{1}{24}$	$\overset{1}{18}$	$\overset{4}{12}$	$\overset{2}{6}$	$\overset{1}{0}$	$\overset{1}{6}$	$\overset{2}{12}$	$\overset{1}{18}$	$\overset{1}{24}$	…	…	…	…	…	…	…
	12	…	…	$\overset{2}{21}$	$\overset{2}{14}$	$\overset{2}{7}$	…	$\overset{1}{7}$	$\overset{1}{14}$	$\overset{1}{21}$	$\overset{2}{26}$	…	$\overset{1}{42}$	…	…	…	…	…
	13	$\overset{1}{40}$	…	…	$\overset{1}{16}$	$\overset{3}{8}$	…	$\overset{1}{8}$	$\overset{1}{16}$	$\overset{1}{24}$	$\overset{1}{32}$	…	$\overset{1}{48}$	…	$\overset{1}{64}$	…	…	…
	14	…	$\overset{1}{36}$	…	…	…	$\overset{1}{0}$	…	…	…	…	$\overset{1}{45}$	…	…	…	…	…	…
	15	…	…	$\overset{1}{30}$	…	…	…	…	$\overset{1}{20}$	…	…	$\overset{1}{50}$	…	…	…	…	…	…
	16	…	…	…	…	…	…	…	…	$\overset{1}{33}$	$\overset{1}{44}$	…	…	…	…	…	…	…

$M_X=5.83$ $\qquad$ $M_Y=5\ 07$

$\mu_X=2.91$ $\qquad$ $\mu_Y=3.14$

又此處係間藉集團,故中點 m 爲 M 所居交場之中點 6/5. 其對 M 之經緯爲 $\xi=0.17$, $\eta=-0.07$, 由是得計算乘積中值 p 之表格.

及由此求出之乘積表

I				II			III		IV		
11	14	24	12	13	28	60	23	105	18	120	14
22	16	48	12	20	16	24	34	64	26	32	14
30	20	60	18	57	30	36	36	80	36	32	21
32	30	24	24	20	48	15	32	60	24	48	28
20	24	56	30	10	36	50	90	48	10	60	8
6	36	32	42	12	60	45	48	160	20	5	11
7	48	44	48	14	135	20	44	20	48	10	1227
8	60	10	7	24	18	50	78	50	36	30	−
20	18	10	35	32	42	30	40	105	40	40	
48	21	30	64	28	30	40	140	60	30	100	
48	24	60	27	90	16	1503	54	125	21	30	
40	33	25	1617	72	48	−	78	1700	60	24	
60	4	45	+	30	60		54	+	45	18	
24	56	50		48	96		72		48	120	

依前述方法再加計算得

$$P=0.587,\quad \xi\eta=-0.012\quad p=0.599$$

$$r=0.0656 \quad b_1=0.0608 \quad b_2=0.0708$$

$$\tan\Theta=\frac{0.9960}{2.0057\cdot 0.0656} \qquad \Theta=82°28'$$

及特性線之方程式為

$$\xi=0.0608\eta \quad \eta=0.0708\xi,$$

其對經緯軸之夾角為

$$\alpha_1=3°29' \quad \alpha_2=4°3'.$$

此種結果之意義如下:

若父之子數增1則子之子數平均增0.0708若子之子數增1則父之子數平均增0.0608.

故父子蕃殖力為極弱之正互關,圖24所示,

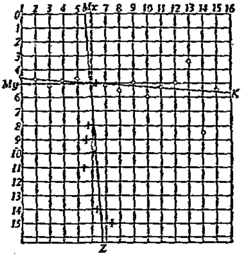


圖24 相關表III之幾何圖

即其幾何關係.圖中小圈,表行中值,十字表列中值,又K, Z 對中值點列之配合近似度為

$$m_Y=3.13 \qquad m_X=2.90,$$

足證直線Z對中值點列之配合較K為密近.

(4)同樣取63節表VII,母女之生殖力而研究其關係,得其相當結果為:

$$M_X=5.90 \qquad M_Y=4.34 \qquad \mu_X=2.83 \qquad \mu_Y=2.97$$

$$m(6/4) \qquad \xi=0.10 \qquad \eta=-0.34$$

$P=1.754$ $\xi\eta=-0.034$ $p=1.788$

$r=0.2130$ $b_1=0.2031$ $b_2=0.2234$

$\tan\Theta=\frac{0.9546}{2.0023\cdot 0.213}$ $\Theta=65^\circ 56'$

$\xi=0.2031\eta$ $\eta=0.2234\xi$

$m_X=2.76$ $m_Y=2.90$

與前例比較,可知母女之生殖力爲較強之正互關.又由圖25知Z,K之配合精密度亦相差甚微.

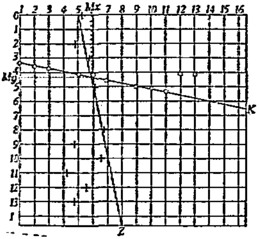


圖25表VII之幾何圖

(5)取61節(4)表VI,關於常春藤之例,注意其類幅依法計算得其結果爲:

$M_X=10.89$

$M_Y=13.23$

$\mu_X=1.72$類幅

$=3.44\cdot\frac{1}{8}$英寸

$\mu_Y=2.28$類幅

$=4.56\cdot\frac{1}{8}$英寸

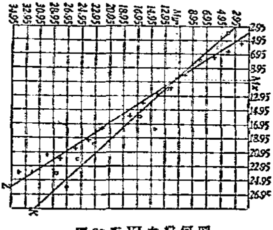


圖26表VI之幾何圖

m(9.95/13.95)

$\xi=-0.47$ $\eta=0.36$類擂

$P=3.1800$, $\xi\eta=-0.1692$

$p=3.3492$, $r=0.8540$

$b_1=0.6375$ $b_2=1.1201$

$$\tan\Theta=\frac{0.2707}{2.0800\cdot 0.8540}\qquad \Theta=8^\circ 40'$$

$\xi=0.6375\eta$ $\eta=1.1201\xi$

$m_X=0.8947$ $m_Y=1.1861$.

此種強度正互關,足證常春藤葉長度之比值,有保持常度之趨勢,圖26所示之結果,亦與前例各圖所示者不同.

72. 前述各節,係假定中值在直線上而求其線性互關.此種假定,稍嫌勉強.然事實上,則64節諸例俱足以證明線性互關之存在.惟中值點列之可以直線配合者固多,其顯呈曲線狀者亦夥.當此之時,而仍欲以直線配合,固無不可,然吾人於配合之外,尤當注意中值點列為曲線之事實,而另求一量（73節）以表明研究之結果,否則中值點列旣成曲線,相互關係自非線性,研究結果,必不能與其他線性互關比較矣.

茲為學者易於明瞭起見,特取中值點列呈曲線狀之例先加考察.

瑞典 Lund[1] 產科醫院,曾取新生男孩及胎盤而權其輕重,以研究其間之關係.下表所示為1223男孩與胎盤之重量及中值點列之散布.

VIII. 新育男孩與胎盤重量之相關表

		嬰孩之體重 X											和	
		1850	2150	2450	2750	3050	3350	3650	3950	4250	4550	4850		
胎衣重 Y	300	1	...	2	...	1	...	...	...	...	...	...	4	2450
	380	2	10	8	25	21	14	2	...	...	...	...	82	2702
	460	2	2	12	34	94	77	37	10	2	...	...	270	3180
	540	...	1	4	17	55	111	84	37	10	...	1	320	3433
	620	...	1	3	6	24	51	92	78	22	3	...	280	3647
	700	...	2	1	2	4	15	40	51	24	8	4	151	3834
	780	...	...	...	...	2	8	13	26	20	11	3	83	4008
	860	...	...	...	...	...	4	1	8	3	6	4	26	4158
	940	...	...	...	...	...	1	1	3	...	...	2	7	4079
和		5	16	30	84	201	281	270	213	81	28	14	1223	
		456	455	463	469	496	617	593	651	678	757	786		

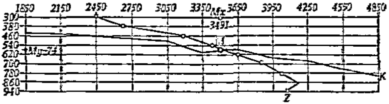


圖27 新育男孩及胎盤重量之中值圖

(1) S. D. Wicksell, Meddelande fran Lunds Astronomiska Observatorium (瑯德天文台報告) Nr. 60. 1917.

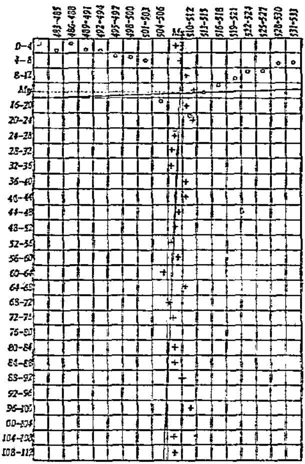


圖 28 相關表 IX 之幾何圖

觀表 VIII 數字之分配,已足斷定正關之存在,及中值之曲性.其彎曲狀態經圖示而益形明瞭 (圖27). 吾人若能增加集團之範圍,則 Z 及 K 之曲線形狀,當更顯著. 至圖中之縱橫線,則係通過交場中點之經緯. 下例中值位置,尤爲特別,學者宜細玩之.

取1881—1890十年間英格蘭及威爾士 625 註冊區所載每千嬰孩中男孩數目[1],而研究此種比率與嬰孩總數是否有關.表中某類之屢數,爲區域個數,此諸區域者,所有之嬰孩總數及男孩千分率均同,又此例與前例不同之點,爲

(1) Yule, An Introduction etc. 5. Edition, P. 163.

二標單位各殊,性質亦異.

表中凡與 r 計算有關之值,大都具備,今更取和行,和列之算術中值及均方差附列於下:

$M_X=509.24$　$\mu_X=2.505$類幅$=7.52$千分率.

$M_Y=14.10$　$\mu_Y=4.542$類幅$=18.17$千分

以 m (508/14) 為原點,得

$P=-0.1741,$

又因　$\xi=\frac{508-509.24}{3}=-0.413,\ \eta=\frac{14-14.10}{4}=-0.025,$

而　$p=-0.1844,$　$r=-0.016,$

$b_1=-0.016\frac{2.505}{4.542}=-0.0088,$

$b_2=-0.016\frac{4.542}{2.505}=-0.029,$

$\tan\Theta=-\frac{0.9997}{2.3647\cdot 0.016},$　$\Theta=92°10',$

$\xi=-0.0088\eta,$　$\eta=-0.029\xi,$

$m_X=2.506$　$m_Y=4.541.$

此種結果其最堪注目之處,約有數事:一曰證明大數定律(93節)為合理,當嬰孩總數增加,則千分率之變限減小,二曰顯示中值點列,可成曲線.表內行中值點列,雖可以直線套合,列中值點列則呈曲線狀態.因之 m_X 及 m_Y 相差甚鉅.三曰二標變化為弱度負互關.前節各例,無論級度大小,均為正關,此例則與之迥異.至此種顯呈曲線形狀之中

値點列,是否尙可以直線套合,則又統計學中一饒有趣味之問題也.

圖28爲表IX中値分佈之結果.

73. 中値點列旣不必盡成直線,表示結果,自以與中値位置無關之量爲佳.Pearson氏之相關比率,卽其類也.自相關比率,減去相關係數,卽得中値點列與套合直線之相差程度.

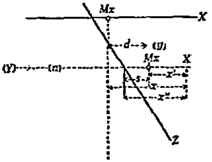


圖29 相關比率之求法

就某定列 (Y) 注意其中値 $M_X^{(Y)}$ 及 X 之公共中値 M_X,與平列中値之套合直線 Z (圖29), 而以

x 表該列中某 X 對 M_X 之偏差,

x' 表其 對 $M_X^{(Y)}$ 之偏差,

x'' 表其 對 Z 之偏差,

d 表 $M_X^{(Y)}$ 對 M_X 之偏差,

δ 表 $M_X^{(Y)}$ 對 Z 之偏差,

n 表 (Y) 列之度數,

N 表集團之範圍,

則得 (Y) 列均方差之平方爲

$$\mu_X^{(Y)2}=\frac{1}{n}\Sigma(x'^2). \tag{13}$$

公共均方差之平方爲

$$\mu_X^2=\frac{1}{N}\Sigma(x^2). \tag{14}$$

更注意各列之屢數,以求諸列均差平方之算術中值 $\mu_X^{(x)^2}$,則得

$$\mu_X^{(x)^2}=\frac{1}{N}\Sigma\left(n\mu_X^{(Y)^2}\right). \tag{15}$$

與(11)式同理,命

$$\mu_X^{(x)^2}=\mu_X^2(1-\rho_{XY}^2), \tag{16}$$

而名 ρ_{XY} 爲 X 對 Y 之相關比率,則得

$$\rho_{XY}^2=1-\frac{\mu_X^{(x)^2}}{\mu_X^2}. \tag{17}$$

同理自一行出發可得 Y 對 X 之相關比率 ρ_{YX}

$$\rho_{YX}^2=1-\frac{\mu_Y^{(B)^2}}{\mu_Y^2}. \tag{17*}$$

式中 $\mu_Y^{(B)}$ 爲各行均差平方之算術中值

若相關表之中值點列,與直線最爲接近,則 ρ_{XY} ρ_{YX} 相差甚微,而與 r 相等.若二者相差甚大,則與 r 相距亦遠,而中值點列呈曲線狀態.反之中值點列與直線乖離益甚,則 ρ_{XY} 與 r, ρ_{YX} 與 r 相差益鉅.

又 $\mu_X^{(x)^2}$ 及 $\mu_Y^{(B)^2}$ 不必依(15)式即可自中值出發,直接計算.

由圖 $\quad x=x'+d;$

平方此式而求其沿 (Y) 列之和,則因 $\Sigma x'=0$,而有

$$\Sigma(x^2)=\Sigma(x'^2)+nd^2=n(\mu_X^{(Y)^2}+d^2);$$

更求此式沿一切列之和而以N除其結果,則得

$$\mu_X^2=\mu_X^{(z)^2}+\mu_\delta^{(X)^2};$$

以μ_X^2除上式由(17)有

$$\rho_{XY}^2=\frac{\mu_d^{(X)^2}}{\mu_X^2}. \qquad (18)$$

由是得求相關比率之方法如下:

先定各列中值對M_X之偏差注意屢數而求其平方中值更以X之均差平方除之,卽得X對Y之相關比率.同理自各行出發,可得Y對X之相關比率.

又由圖

$$x''=x'+\delta;$$

自此入手可得

$$m_X^2=\mu_X^{(z)^2}+\mu_\delta^{(X)^2};$$

以(11) (16)二式代入,則有

$$\mu_X^2(1-r^2)=\mu_X^2(1-\rho_{XY}^2)+\mu_\delta^{(X)^2},$$

因之 $$\mu_\delta^{(X)^2}=\mu_X^2(\rho_{XY}^2-r^2). \qquad (19)$$

由是可知ρ_{XY}^2恆大於r^2,以$\mu_\delta^{(X)^2}$及μ_X^2咸爲正數也.在特例,當$\rho_{XY}^2=r^2$,則$\mu_\delta^{(X)^2}=0$,而列中值完全在直線Z上.反之若ρ_{XY}^2與r^2相差愈大,則$\mu_\delta^{(X)^2}$亦隨之增長,而列中值點列與直線相差愈遠.

74. 吾人已知表IX中,行中值點列與直線相差甚大,據

上述理論,ρ_{YX} 與 r 之差應極顯著.茲取此例再加研究,以

相關表IX　相關比率 ρ_{YX} 之計算

每一新育中之男孩 X	相應行之中値 $M_Y^{(x)}$	對公共中値之乖差 $M_Y=14.1$	乖差之平方	行匯數 n	末二行之積
480—482	2	+12.1	146.41	2	292.82
483—485	4	+10.1	102.01	2	204.02
486—488	2	+12.1	146.41	2	292.82
489—491	4	+10.1	102.01	4	408.04
492—494	3.3	+10.8	116.64	6	699.84
495—497	4.4	+ 9.7	94.09	15	1411.35
498—500	5.6	+ 8.5	75.25	18	1300.50
501—503	5.9	+ 8.2	67.24	46	3093.04
504—506	16.3	− 2.2	4.84	98	474.32
507—509	24.8	−10.7	114.79	125	14311.25
510—512	16.3	− 2.2	4.84	129	634.36
513—515	10.3	+ 3.8	14.44	88	1270.72
516—518	8.5	+ 5.6	31.36	47	1473.92
519—521	6.6	+ 7.5	56.25	26	1462.50
522—524	5.3	+ 8.8	77.44	11	851.84
525—527	5.0	+ 9.1	82.81	4	331.24
528—530	2	+12.1	146.41	1	146.41
531—533	2	+12.1	146.41	1	146.41
534—536	3.3	+10.8	116.64	3	349.92
				628	29245.32

證所述理論與事實相符.

上表所示,為其詳細結果及計算手續.至表 IX 首段自 645—467 至 477—479 及末段自 537—539 至 543—545 諸行,則以其屢數太小且多空類之故,從簡删去.此種删削,對 $M_Y=14.10$ 及 $\mu_Y=18.17$ 均無甚影響.

自表中主要結果得

$$\mu_Y^{(E)^2}=\frac{29245.32}{628}=46.5689,\ \mu_Y^{(E)}=6.82$$

$$\rho_{YX}=\frac{68.2}{18.17}=0.38;$$

此數與72節求出之 $r=-0.016$ 相差甚大,足證理論與事實相符.又由表可得 $\rho_{XY}=0.13$ 較 ρ_{YX} 為小,可知列中值點列與直線較為接近.

第六章 相關係數之用法

75. 相關係數之意義,已如上述,其應用當於87節更詳論之.茲擇須應用相關係數,方能解決之問題,述之如次:

問題一:已知變數 X_1 之判值數序為 $X_1^{(1)}, X_1^{(2)}, X_1^{(3)}, \cdots\cdots$; X_2 之判值數序為 $X_2^{(1)}, X_2^{(2)}, X_2^{(3)}\cdots\cdots$; $\cdots\cdots$; X_n 之判值數序為 $X_n^{(1)}, X_n^{(2)}, X_n^{(3)}\cdots\cdots$; 試求諸和 $X_1+X_2+X_3+\cdots\cdots+X_n$ 對其算術中值之均方差.

此問題可分二種情形研究,(a)各變數互相連屬,有一 X_1

卽有一X_2,一X_3……與之相連.(*b*)各變數互爲獨立,諸判値可任意配置.

欲解決此種問題,僅須取二變數硏究已足.

(*a*)若X_1,X_2互相連屬,則X_1,X_2與$X=X_1+X_2$有相同之範圍N,設M_1,M_2爲X_1,X_2之算術中値,x_1,x_2爲判値對M_1,M_2之乖差,則X_1,X_2之判値,可書作M_1+x_1,M_2+x_2,X之判値可書作$M_1+M_2+x_1+x_2$,但M_1+M_2爲X之算術中値,x爲X對此中値之乖差,故有

$$x^2=x_1^2+x_2^2+2x_1x_2$$

$$\Sigma(x^2)=\Sigma(x_1^2)+\Sigma(x_2^2)+2\Sigma(x_1x_2),$$

以N除此式,得

$$\mu^2=\mu_1^2+\mu_2^2+2\frac{\Sigma(x_1x_2)}{N}.$$

或 $$\mu^2=\mu_1^2+\mu_2^2+2r_{12}\mu_1\mu_2; \qquad (1)$$

式中 $$r_{12}=\frac{\Sigma(x_1x_2)}{N\mu_1\mu_2}$$

爲X_1,X_2之相關係數.

若所求爲X_1-X_2之均方差,則上式中除末項變爲負數外,其餘一概仍舊,由是得

$$\mu^2=\mu_1^2+\mu_1^2-2r_{12}\mu_1\mu_2. \qquad (2)$$

在特例,若變數無相互關係,則$r_{11}=0$,而

$$\mu^2=\mu_1^2+\mu_1^2 \qquad (3)$$

若完全互關成立,則當$r_{12}=+1$時,

變數和之$\mu=\mu_1+\mu_2$,

變數差之$\mu=|\mu_1-\mu_2|$;

當$r_{12}=-1$時,

變數和之$\mu=|\mu_1-\mu_2|$,

變數差之$\mu=\mu_1+\mu_2$;

至在其他情形,則

$$|\mu_1-\mu_2|<\mu<\mu_1+\mu_2.$$

上述結果,不難推及於多個變數,設變數之個數爲n,則 $\mu^2=\mu_1^2+\mu_2^2+\cdots\cdots+\mu_n^2+2r_{12}\mu_1\mu_2+2r_{13}\mu_1\mu_3+\cdots\cdots$
$+2r_{23}\mu_2\mu_3+\cdots\cdots$ (4)

此式右端諸平方項無論變數之聯結號爲加爲減,咸爲正數.至其後倍積,則視相當變數前符號之同否,而別其正負焉.

(b)若X_1, X_2不相連屬,則X_1+X_2之範圍$N=N_1N_2$,式中N_1爲X_1,N_2爲X_2之範圍,此N個x可排列如下:

$$\begin{array}{ll}
x_1^{(1)}+x_2^{(1)} & x_1^{(1)}+x_2^{(2)}\cdots\cdots x_1^{(1)}+x_2^{(N_2)} \\
x_1^{(2)}+x_2^{(1)} & x_1^{(2)}+x_2^{(2)}\cdots\cdots x_1^{(2)}+x_2^{(N_2)} \\
\cdots\cdots\cdots\cdots & \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
x_1^{(N_1)}+x_2^{(1)} & x_1^{(N_1)}+x_2^{(2)}\cdots\cdots x_1^{(N_1)}+x_2^{(N_2)};
\end{array}$$

由是得

$$\Sigma(x^2)=N_2\Sigma(x_1^2)+N_1\Sigma(x_2^2)+2\Sigma(x_1x_2).$$

因 X_1, X_2 不相連屬,故 $\Sigma(x_1x_2)=\Sigma(x_1)\Sigma(x_2)=0$

以 N 除上式得

$$\frac{\Sigma(x^2)}{N}=\frac{\Sigma(x_1^2)}{N_1}+\frac{\Sigma(x_2^2)}{N_2} \text{ 亦即 } \mu^2=\mu_1^2+\mu_2^2. \quad (5)$$

此式僅含平方項,故亦可應用之以求 X_1-X_2 之均方差.若所論變數有 n 個,則無論其聯結號爲加爲減,

$$\mu^2=\mu_1^2+\mu_2^2+\cdots\cdots+\mu_n^2 \quad (6)$$

吾人若就某種生物,取若干樣元,分别測其度量,計算各度之散布,再由各個散布以求全體之散布量,則須應用(a)款.若就不相連屬諸量之觀察結果,而求其代數和,則須應用(b)款.

76. 算術中值之均方差.設 $X_1, X_2, \cdots\cdots X_n$ 爲量 X 之觀察結果,M 爲其算術中值 $x_1, x_2, \cdots\cdots x_n$, x 爲 $X_1, X_2\cdots\cdots X_n$; M 之實際差誤,則

$$M+x=\frac{X_1+x_1+X_2+x_2+\cdots\cdots+X_n+x_n}{n};$$

由是得

$$nx=x_1+x_2+\cdots\cdots+x_n.$$

設 N_i 爲 x_i 出現之次數,則當觀察互爲獨立時,$N=N_1N_2\cdots\cdots N_n$ 爲 x 出現之次數.平方上式而求其和,則有

$$n^2\Sigma(x^2)=\frac{N}{N_1}\Sigma(x_1^2)+\frac{N}{N_2}\Sigma(x_2^2)+\cdots+\frac{N}{N_n}\Sigma(x_n^2)$$

$$+2\frac{N}{N_1N_2}\Sigma(x_1x_2)+2\frac{N}{N_1N_3}\Sigma(x_1x_3)+\cdots\ ;$$

以配合時 x_i 出現之次數,與其餘諸值組合之個數相同故也,又因觀察互為獨立,而

$$\Sigma(x_ix_k)=\Sigma(x_i)\Sigma(x_k);$$

若諸誤之正負量,能互相消去,則 $\Sigma(x_i)=0,\Sigma(x_ix_k)=0$ 以 N 除其餘諸項,則得

$$n^2\mu_M^2=\mu_1^2+\mu_2^2+\mu_3^2+\cdots\cdots+\mu_n^2. \qquad (7)$$

若 μ_i 咸為已知,則 μ_M 即可由此算出,由是得:

取個別觀察均誤平方之和,開平方而以觀察個數除其結果,即得算術中值之均方差誤.

若觀察之精確度相同,則諸均方差相等,($\mu_1=\mu_2=\cdots\cdots=\mu_n=\mu$) 而(7)式變為

$$n\mu_M^2=\mu^2.$$

由是得

$$\mu_M=\frac{\mu}{\sqrt{n}}. \qquad (8)$$

其意義如下:

若 n 個觀察之精密度相同,則其算術中值之均方差誤,等於任一觀察之均方差誤的 $\frac{1}{\sqrt{n}}$.

上述結果,可應用之以研究集團,吾人若命樣元之算術

中值爲M,均方差爲μ,集團之範圍爲n,則由(8)式決定之μ_M,爲集團算術中值之均方差.μ_M者,與μ之意義相當,表中值變化之數量也.集團之範圍愈大,則中值之變化愈小,而μ_M愈微.

77. 茲舉二例以說明(3)式及(8)式之應用.

例(1)由58節所述6194名英格蘭人及1304名蘇格蘭人之體高分佈表,得

$M_1=67.37$　　　$\mu_1=2.60$英寸

$M_2=68.39$　　　$\mu_2=2.50$英寸

試求此二民族體高之差異.

若依自然趨勢,就二者之算術中值,判定體高差異,則蘇格蘭人平均高於英人$d=1.02$英寸.惟此種差別,不必盡人如是,其可靠與否,與中值之變化有關.換言之亦卽以中值均方差之大小爲轉移,由是得d之均方差

$$\mu_d=\sqrt{\mu_{M_1}^2+\mu_{M_2}^2}=\sqrt{\frac{2.60^2}{6194}+\frac{2.50^2}{1304}}=.0077\text{英寸},$$

小於d值甚遠.按112節定律,若差別本身大於其均方差至三倍以上,則此種差別名爲『可靠』.今d值大於μ_d十三倍有奇,故吾人可謂蘇格蘭人普通高於英格蘭人1.02英寸.

(2)自25節例(2)新生嬰孩體重分布表,得

$N_1=288$;　　$M_1=3278.1$,　　$\mu_1=535$公分(g)

$$N_2=269;\qquad M_2=3092.6,\qquad \mu_2=447\text{公分}.$$

由是算出

$$\mu_{M_1}=31.5\qquad \mu_{M_2}=27.3\text{公分},$$

及 $d=M_1-M_2=185.5$之均方差

$$\mu_d=\sqrt{992.25+745.29}=41.7\text{公分}.$$

因 d 大於 μ_d 4.4 倍有奇,故新生女孩之體重較新生男孩爲輕,而平均値爲男孩體重 94.3%; 又據59節 Galton 定律,女孩體重應爲男孩體重92.3%.

78. 茲更取含有 r 之普通問題,求其解答.

問題二:設X_1, X_2, X_3……爲互相連屬之變數,x_1, x_2, x_3……爲其對相當中値M_1, M_2, M_3……之偏差,$V=f(X_1,X_2,X_3\cdots)$爲其任意函數,若變數値有N組爲已知,則函數V有N値爲已定.茲假定x_1, x_2, x_3……均較其相當之M_1, M_2, M_3,……爲甚小,試求此函數N値之算術中數及均方差.

命M爲所求之算術中値,μ爲所求之均方差.

依假定x_i爲甚小,故在

$$V=f(M_1+x_1, M_2+x_2, M_3+x_3,\cdots\cdots)$$

之展式中,x_i之高次項可以略去,吾人若以f表$f(M_1,M_2,M_3,\cdots\cdots)$, f_1,f_2,f_3……表f對M_1, M_2, M_3……之引數,f_{ki}爲f先對M_k次對M_i之引數,則得

$$V=f+f_1x_1+f_2x_2+\cdots\cdots+\frac{1}{2}(f_{11}x_1^2+f_{22}x_2^2+\cdots\cdots)+$$

$$f_{12}x_1x_2+f_{13}x_1x_3+\cdots\cdots;$$

取此式求和而以 N 除其結果,則有

$$M=\frac{1}{N}\left[Nf+f_1\Sigma(x_1)+f_2\Sigma(x_2)+\cdots+\frac{1}{2}(f_{11}\Sigma(x_1^2)\right.$$

$$\left.+f_{22}\Sigma(x_2^2)+\cdots)+f_{12}\Sigma(x_1x_2)+f_{13}\Sigma(x_1x_3)+\cdots\right].$$

又因 $\Sigma(x_1)=0$, $\Sigma(x_2)=0$, ……, 而

$$M=f+\frac{1}{2}(f_{11}\mu_1^2+f_{22}\mu_2^2+\cdots)+f_{12}\mu_1\mu_2r_{12}$$

$$+f_{13}\mu_1\mu_3r_{13}+\cdots\cdots \tag{9}$$

是爲函數 N 值之算術中值.

欲求函數諸值之均方差,可應用51節 (A) 式,因有

$$M^2+\mu^2=\frac{1}{N}\Sigma(V^2). \tag{10}$$

依假定 x_i 之高次方可以略去,故在一定精確度內

$$V^2=f^2+2f(f_1x_1+f_2x_2+\cdots)+f_1^2x_1^2+f_2^2x_2^2+\cdots+$$

$$+2(f_1f_2x_1x_2+f_1f_3x_1x_3+\cdots)+f(f_{11}x_1^2+f_{22}x_2^2+\cdots)$$

$$+2f(f_{12}x_1x_2+f_{13}x_1x_3+\cdots),$$

由是得

$$\Sigma(V^2)=Nf^2+f_1^2\Sigma(x_1^2)+f_2^2\Sigma(x_2^2)+\cdots+f(f_{11}\Sigma(x_1^2)$$

$$+f_{22}\Sigma(x_2^2)+\cdots)+2(f_1f_2\Sigma(x_1x_2)+f_1f_3\Sigma(x_1x_3)+$$

$$\cdots)+2f(f_{12}\Sigma(x_1x_2)+f_{13}\Sigma(x_1x_3)+\cdots)$$

以 N 除之,則有

$$\frac{1}{N}\Sigma(V^2)=f^2+f_1^2\mu_1^2+f_2^2\mu_2^2+\cdots+f(f_{11}\mu_1^2+f_{22}\mu_2^2$$

$$+\cdots)+2(f_1f_2\mu_1\mu_2r_{12}+f_1f_3\mu_1\mu_3r_{13}+\cdots)+$$

$$+2f(f_{12}\mu_1\mu_2r_{12}+f_{13}\mu_1\mu_3r_{13}+\cdots);$$

以此式及(9)代入(10),則得

$$\mu^2=f_1^2\mu_1^2+f_2^2\mu_2^2+\cdots+2(f_1f_2\mu_1\mu_2r_{12}+f_1f_3\mu_1\mu_3r_{13}+\cdots).\quad(11)$$

若變數無相互關係,則 $r_{12}=0$, $r_{13}=0$……而(9),(11)二式變為

$$\left.\begin{aligned}M&=f+\frac{1}{2}(f_{11}\mu_1^2+f_{22}\mu_2^2+\cdots\cdots)\\ \mu^2&=f_1^2\mu_1^2+f_2^2\mu_2^2+\cdots\cdots\end{aligned}\right\}.\quad(12)$$

吾人若將(9),(11)二式應用於函數(a) $V=X_1X_2$, (b) $V=\frac{X_1}{X_2}$, 則得其結果如下:

(a) 自 $f=M_1M_2$ 得 $f_1=M_2, f_2=M_1, f_{11}=0, f_{22}=0, f_{12}=1$ 代入(9)式及(11)式,得

$$M=M_1M_2+\mu_1\mu_2r_{12}$$

$$\mu^2=M_2^2\mu_1^2+M_1^2\mu_2^2+2M_1M_2\mu_1\mu_2r_{12};$$

命 $\frac{\mu_1}{M_1}=\alpha_1$, $\frac{\mu_2}{M_2}=\alpha_2$ 則得應用較便之公式

$$\left.\begin{aligned}M&=M_1M_2(1+\alpha_1\alpha_2r_{12})\\ \mu^2&=M_1^2M_2^2(\alpha_1^2+\alpha_2^2+2\alpha_1\alpha_2r_{12})\end{aligned}\right\},\quad(13)$$

式中 α_1, α_2 依假定咸為小數.

若變數無相互關係則$r_{12}=0$,而上式變爲

$$M=M_1M_2,$$

$$\mu^2=M_1^2M_2^2(\alpha_1^2+\alpha_2^2).$$

(b) 自$f=\frac{M_1}{M_2}$得$f_1=\frac{1}{M_2}$, $f_2=-\frac{M_1}{M_2^2}$, $f_{11}=0$, $f_{22}=\frac{2M_1}{M_2^3}$,

$f_{12}=-\frac{1}{M_2^2}$;代入(9)式及(11)式得

$$M=\frac{M_1}{M_2}+\frac{M_1}{M_2^3}\mu_2^2-\frac{\mu_1\mu_2}{M_2^2}r_{12}$$

$$\mu^2=\frac{\mu_1^2}{M_2^2}+\frac{M_1^2}{M_2^4}\mu_2^2-2\frac{M_1}{M_2^3}\mu_1\mu_2r_{12}.$$

命　$\frac{\mu_1}{M_1}=\alpha_1$, $\frac{\mu_2}{M_2}=\alpha_2$,則有

$$\left.\begin{aligned}M&=\frac{M_1}{M_2}(1+\alpha_2^2-\alpha_1\alpha_2r_{12})\\ \mu^2&=\frac{M_1^2}{M_2^2}(\alpha_1^2+\alpha_2^2-2\alpha_1\alpha_2r_{12});\end{aligned}\right\}\qquad(14)$$

若分子分母無相互關係則$r_{12}=0$,而

$$M=\frac{M_1}{M_2}(1+\alpha_2^2)\ M>\frac{M_1}{M_2}$$

$$\mu^2=\frac{M_1^2}{M_2^2}(\alpha_1^2+\alpha_2^2).$$

例:(1)就66節表VII求母女子數之比值.

由71節例4知

$$M_1=5.90\qquad \mu_1=2.83\qquad r_{12}=02.13$$

$M_2=4.34$ $\mu_2=2.97$

由此算出

$\frac{M_1}{M_2}=1.3594$ $\alpha_1=0.4796$ $\alpha_2=0.6843$

代入公式得

$M=1.901$ $\mu=1.016.$

因知母女子數之平均比值M,約為1.9,其均方差$\mu_M=\frac{\mu}{\sqrt{n}}$為0.032.

2)就72節表VIII求新生男孩及胎盤重量之比值.

由表算得

$M_1=3490.55$ $\mu_1=512.46$公分 $r_{12}=0.6284$

$M_2=574.27$ $\mu_2=117.35$公分

$\frac{M_1}{M_2}=6.0782$ $\alpha_1=0.1468$ $\alpha_2=0.2043$

$M=6.217$ $\mu=0.9725$

因知新生男孩及胎盤重量之平均比值M為6.217, M之均方差μ_M為0.028.

79. 統計術中,或因分工關係析集團為部分,分別求其結果,然後研究共相.或依材料性質,先有部分集團及結果,再求其整個關係.凡此二者,皆須由部分集團之算術中值,或其他數量,求總集團之相當中值或數量.

此種問題之簡單解法,為取部分中值,求其算術中數,然

由此所得之結果,通常不能應用,以此種解法,僅當各個中值之分兩相同時爲合理也.故欲求圓滿之解答,須先得詳細之記載,以決定各個中值之身價.

例如已知某物在國內各市場之平均價,欲由此求出該物在全國之平均價,則非知該物在各市之售出量不可,此售出量,卽表此物在該市平均價之分兩.依部分中值之身價,而求出之總中值,卽爲該物在全國之平均價.若部分記載不甚詳盡,售出額量無從稽考,則僅能取部分中值求其中數,此種中數,通常與依法求出之中數不同.

又如已知國內各區嬰兒與人口之比率,則全國嬰兒之百分數爲以各區人口爲權衡求出之算術中值.至若不注意各區人口,而求得之算術中值,則在統計學中毫無意義.

問題三:設 g 爲變數 X 之身價,M 爲部分中值之算術中值,M_1 爲注意各值分兩而求得之算術中數,x 爲 X 對 M 之偏差,r 爲 X 之平均分兩,μ_X 爲 X 對 M 之均方差,μ_g 爲 g 對 r 之均方差,N 爲集團之範圍,r 爲 X 與 g 之相關係數,試求 M, M_1, 及 r 之關係.

依假定

$$M=\frac{1}{N}\Sigma(X),\ M_1=\frac{\Sigma(gX)}{\Sigma(g)},\ r=\frac{1}{N}\Sigma(g).$$

$$\Sigma(gX)=\Sigma g(M+x)=M\Sigma(g)+\Sigma(gx)=NMr$$

$$+\Sigma(gx)=NMr+Nr\mu_X\mu_g,$$

以$\Sigma(g)=Nr$除上式,則得

$$M_1=M+r\mu_X\frac{\mu_g}{r};\qquad(11)$$

式中μ_X, μ_g, r均為正數,故當X與g之相關係數為正時,M_1大於M.為負時M_1小於M,又因M_1及M之差,與$\frac{\mu_g}{r}$成正比,故中值分兩變化愈大,則二者相差愈遠.

又(11)式可書為

$$r=\frac{M_1-M}{\mu_X}\cdot\frac{r}{\mu_g}.$$

此式亦可應用之以求相關係數,茲舉例說明如下:

例:就1891年元旦英格蘭及威爾士632行政區之貧民百分率,以研究英國貧民之分佈;至此處入算之貧民,則係依法律規定,應由政府救濟者.

百分數	區鎮之個數
0.75—1.25	18
1.25—1.75	48
1.75—2.25	72
2.25—2.75	89
2.75—3.25	100
3.25—3.75	90
3.75—4.25	75
4.25—4.75	60
4.75—5.25	40
5.25—5.75	21
5.75—6.25	11
6.25—6.75	5
6.75—7.25	1
7.25—7.75	1
7.75—8.25	0
8.25—8.75	1
	632

由表可得$M=3.29$, $\mu_X=1.24\%$,至M_1, r,及μ_g則因右表未載各區人數,無從計算.但據該年英國人口調查,此632區中人口最少為二千,最多為五十萬,其平均人數,為每區45900,均方差為56400,

全國貧民之百分率爲$M_1=2.69$.由是估定

$$r=\frac{2.69-3.29}{1.24}\cdot\frac{45900}{56400}=-0.39,$$

因知是年英國人口與貧民百分率,爲顯著之負互關.

第七章 多標之相關性

80. 以上所論僅限於雙標分布,及二變數之互關.如父子,母女之生殖力,及樹葉長度之相關性等.

然宇宙事物,至爲繁頤,變數出現,同時多至三個以上者,往往而是.此種連屬變數,固可成對取出,以研究每兩相互之關係.然欲得其整個結果.則非此法所能奏效,例如五穀收穫,不獨受雨量之影響,亦且以温度爲轉移.而雨量多寡,又足以左右温度高低,故温度雨量二者,互爲因果,與收穫有關之變數也.吾人若僅依雙標理論,研究個別對收穫之影響,以爲統計結果,必至離眞而遠事情無意義矣!至於收穫與雨量之關係何如?温度對收穫之影響奚似?雨量與温度之牽涉大小.則又急待解決之問題,茲章所論之宏旨也!

又如遺傳學中,已證明祖孫性質有相互關係,欲進求父子之遺傳影響,則其變數亦有三個.

欲解決此種問題,必須取所有變標同時研究,以其相當影響,互相牽涉,不能分立也.

吾人之研究統計也,先取單標詳其理論,次定雙標窮其變化,今則更進一層而討論一組變標之相互影響.此種步驟,皆係依科學方法,循序漸進,故前者每為後者之預備,後者每為前者之結果.曩昔既藉單標分佈以論雙標互關,今將更依雙標互關,以窮多標之變化焉.

81. 討論雙標互關,所據基本觀念,為以直線方程連絡出現變數.然後依差誤平方和須為極小之條件,決定方程式之係數.並由此求出 r 表示相關之性質及級度而達到自甲標變化推出乙標平均變化之目的.今擬將此種觀念,推及多個變數,以研究其相互關係.

設 $X_1,X_2\cdots\cdots X_n$ 為某現象範圍中同時出現之 n 變數 $x_1,x_2\cdots\cdots x_n$ 為一組觀察值,設 X_1 可以其他變數之線式表示

$$X_1=a+b_2X_2+b_3X_3+\cdots\cdots+b_nX_n. \qquad (1)$$

若係數 $a,b_2,b_3,\cdots\cdots b_n$ 為已知,則(1)式之配合近似度,以

$$x_1-a-b_2x_2-b_3x_3-\cdots\cdots-b_nx_n$$

之大小為轉移,蓋此種線性關係,若能完全成立,則其結果應等於零也.吾人將以 $x_{1.23\ldots n}$ 表上式結果,而名之為(1)式之差誤,至 x 之足指數,則與諸變數之足指數相同,而以主要變數所有者列於第一.

以上所述為(1)式之意義及來由,茲更依科學原理,進論係數之定法.據最小二乘式之理論,線式配合最精之條件

為 $\Sigma(x^2_{1.23\cdots n})$

為極小.設 N 為集團之範圍,則此和有 N 項,以 $X_1X_2\cdots X_n$ 之值有 N 組也.

至於 b_i 之意義,則可述之如下:b_i 云者 X_1 對 X_i 之消長係數,其符號表 X_i 對 X_1 所生影響之性質,其大小表此種影響之強弱.若 b_i 為正,則 X_1 隨 X_i 而增,若 b_i 為負,則 X_1 逆 X_i 而減.

至欲得諸變數之整個關係,則須依(1)式順次取各變數以其餘變數表之,如以 $X_1, X_3, \cdots\cdots X_n$ 之線性函數表 X_2 是.惟此種線式,數繁項夥,對於運算諸多不便,茲為免除淆混計,特引用符號以別之.

82. 以

$$b_{12.34\cdots n}$$

表 $X_3, X_4, \cdots\cdots X_n$,為常數時,$X_1$ 對 X_2 之消長係數,而名之為偏消長係數[1],以示與全消長係數

$$b_{12}$$

有別,b_{12} 者僅有 X_1, X_2 而無其餘變數之消長係數也.至於 $b_{1.234\cdots n}$ 之指數,則第一批為相關變數之指數,第二批為輔

(1) 此量 Pearson 稱之為偏相關.其最先以數學方法研究之者為 G. U. Yule. On the Significance of Bravais' Formulae of Regression etc. Proc. Roy. Soc., vol. 60 (1897) P. 477.

助變數之指數,而以主要變數之指數書於首批第一.由是易知$b_{12.34\cdots n}$與$b_{21.34\cdots n}$有別,以前者爲以$X_2, X_3, \cdots\cdots X_n$表$X_1$時$X_2$之係數,後者爲以$X_1X_3X_4\cdots\cdots X_n$表$X_2$時$X_1$之係數也.至於第二批指數則與次序無關,學者恆依數字順序列之.

設$x_1, x_2, \cdots\cdots x_n$爲$X_1, X_2\cdots\cdots X_n$之一組觀察值,以$x_{1.23\cdots n}$表以觀察值代入(1)式後所得之差誤,則得

$$x_{1.23\cdots n}=x_1-a_1-b_{12.34\cdots n}x_2-b_{13.245\cdots n}x_3-\cdots$$
$$-b_{1n.23\cdots(n-1)}x_n \qquad (2)$$

此種方程式之個數,與觀察值之組數N相同,其組數則與變數之數目n相等,以此種方程式除(2)式所示之一組外,尚有

$$x_{2.13\cdots n}=x_2-a_2-b_{21.3\cdots n}x_1-b_{23.14\cdots n}x_3-\cdots\cdots$$
$$-b_{2n.13\cdots(n-1)}x_n \qquad (3)$$

等$n-1$組也.

83. 決定方程式(2)諸消長係數之條件,爲

$$\Sigma(x^2_{1.23\cdots n})=\Sigma(x_1-a_1-b_{12.34\cdots n}x_2-b_{13.24\cdots n}x_3-\cdots$$
$$-b_{1n.23\cdots(n-1)}x_n)^2$$

有極小值.吾人若取此式,對a_1及b分別求其微分,則得

$$\left.\begin{array}{l}\Sigma(x_1-a_1-b_{12.34\cdots n}x_2-b_{13.24\cdots n}x_3-\cdots-b_{1n.23\cdots(n-1)}x_n)=0\\ \Sigma x_2(x_1-a_1-b_{12.34\cdots n}x_2-b_{13.24\cdots n}x_3-\cdots-b_{1n.23\cdots(n-1)}x_n)=0\end{array}\right\}(4)$$

$$\left.\begin{array}{l}\Sigma x_3(x_1-a_1-b_{12.34\cdots n}x_2-b_{13.24\cdots n}x_3-\cdots-b_{1n.23\cdots(n-1)}x_n)=0\\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\ \Sigma x_n(x_1-a_1-b_{12.34\cdots n}x_2-b_{13.24\cdots n}x_3-\cdots-b_{1n.23\cdots(n-1)}x_n)=0\end{array}\right\}$$

爲決定a_1,b之條件式此式又可簡書爲

$$\left.\begin{array}{r}\Sigma(x_{1.23\cdots n})=0\\ \Sigma(x_2\cdot x_{1.23\cdots n})=0\\ \Sigma(x_3\cdot x_{1.23\cdots n})=0\\ \cdots\cdots\cdots\cdots\cdots\cdots\\ \Sigma(x_n\cdot x_{1.23\cdots n})=0\end{array}\right\}(4^*)$$

是爲最小二乘式中之標準方程,其個數 n 與變數之數目相等.

至於數計 X 所用之出發點,則通常爲算術中值,如 $x_1,x_2\cdots\cdots x_n$ 爲 $X_1, X_2\cdots\cdots X_n$ 對相當中值 $M_1, M_2\cdots\cdots M_n$ 之偏差是,式中

$$M_i=\frac{1}{N}\Sigma(X_i). \qquad i=1,2,3,\cdots\cdots n \qquad (5)$$

又因所取出發點算術中值,而

$$\Sigma(x_i)=0, \qquad i=1,2\cdots\cdots n \qquad (6)$$

由標準方程第一式可得

$$a_1=0, \qquad (7)$$

亦即(1)式之常數項爲零,而

$$x_1=b_{12.34\cdots n}x_2+b_{13.24\cdots n}x_3+\cdots\cdots+b_{1n.23\cdots(n-1)}x_n. \qquad (1^*)$$

至欲自其餘諸式直接求出 b 值,則須用積和 $\Sigma(x_1x_2)$, $\Sigma(x_1x_3)$, …… 其計算至爲麻煩,今特採間接方法求之.

84. 欲研究間接方法,須先明相關意義.在雙標分佈,若 X_1, X_2 爲所論變數,則其消長方程

$$x_1=b_{12}x_2 \qquad x_2=b_{21}x_1$$

僅有二個,由關係式

$$b_{12}=r_{12}\frac{\mu_{X_1}}{\mu_{X_2}} \qquad b_{21}=r_{12}\frac{\mu_{X_2}}{\mu_{X_1}}$$

卽得

$$r_{21}^2=b_{12}\cdot b_{21}.$$

是爲相關係數與消長係數之關係.今擬將此種關係,推及於多個變數,惟使與原有觀念,不致衝突.

設 $r_{12.34\ldots n}$ 爲同時注意其餘變數所得 X_1, X_2 之相關係數,

命　$$r_{12.34\ldots n}^2=b_{12.34\ldots n}\,b_{21.34\ldots n}. \tag{8}$$

則因右端首批指數互換,其值不變,與此種相關爲雙方交互之事實完全契合.而相關係數之首批指數,遂可任意排列.

吾人若以 $\mu_{1.23\ldots n}$ 表(1*)式之均方差,則與雙標論中之 μ_X 相當,有

$$\mu_{1.23\ldots n}^2=\frac{1}{N}\Sigma(x_{1.23\ldots n}^2). \tag{9}$$

因右端諸項咸自(1*)式導出,而(1*)式係數又由標準方程決定,故其平方和之値爲極小.

消長係數,相關係數,及均方差誤之輔助指數之個數,稱爲該量之級,如

$$b_{12.34\cdots n},\qquad r_{12.34\cdots n}\qquad \mu_{1.23\cdots n}$$

依次爲

$$n-2\qquad n-2\qquad n-1$$

級.是又

$$b_{12}\qquad r_{12}\qquad \mu_1$$

之級爲零.

85. 依上述定義,取方程式(4*)細加觀察,可得下述定理:若零級偏差之指數爲高級偏差輔助指數之一,則二差之乘積總和爲零.

此定理在普通情形固應成立,卽在一級偏差,亦以

$$x_{1.2}=x_1-b_{12}x_2\qquad x_{2.1}=x_2-b_{21}x_1$$

而

$$\Sigma(x_2\cdot x_{1.2})=\Sigma(x_2x_2)-b_{12}\Sigma(x_2^2)\quad \Sigma(x_1\cdot x_{2.1})=\Sigma(x_1x_2)-b_{21}\Sigma(x_1^2),$$

按67節有

$$\Sigma(x_2\cdot x_{1.2})=0\qquad \Sigma(x_1\cdot x_{2.1})=0.$$

在積和

$$\Sigma(x_{1.34\cdots n}x_{2.34\cdots n})$$

中,首以

$$x_1-b_{13.4\cdots n}x_3-b_{14.35\cdots n}x_4-\cdots-b_{1n.34\cdots(n-1)}x_n$$

代其第一因子,次以

$$x_2-b_{23.4\cdots n}x_3-b_{24.35\cdots n}x_4-\cdots-b_{2n34\cdots(n-1)}x_n$$

代其第二因子,分別應用上述定理而求其值,則得

$$\Sigma(x_{1.34\cdots n}x_{2.34\cdots n})=\Sigma(x_{1.34\cdots n}\cdot x_2)=\Sigma(x_1\cdot x_{2.34\cdots n}).$$

其意義可述之如下:

若二偏差之輔助指數相同,則在其積和中,一因子之輔助係數,可以略去.反之取甲因子之輔助係數,加於乙因子上,其結果亦復相同.

由是(4*)式中之第二式亦可書爲

$$\Sigma(x_{2.34\cdots n}\cdot x_{1.234\cdots n})=0.$$

吾人若以(2)式代其第二因子而展開之,則依前述定理有

$$\Sigma(x_1 x_{2.34\cdots n})-b_{12.34\cdots n}\Sigma(x_2\cdot x_{2.34\cdots n})=0$$

此式又可書爲

$$\Sigma(x_{1.34\cdots n}x_{2.34\cdots n})-b_{12.34\cdots n}\Sigma(x_{2.34\cdots n}\cdot x_{2.34\cdots n})=0.$$

由是得

$$b_{12.34\cdots n}=\frac{\Sigma(x_{1.34\cdots n}x_{2.34\cdots n})}{\Sigma(x^2_{2\cdot 34\cdots n})}. \qquad (a)$$

依67節(3)(5)二式,應有

$$r_{12.34\cdots n}=\frac{\Sigma(x_{1.34\cdots n}\cdot x_{2.34\cdots n})}{N\mu_{1.23\cdots n}\mu_{2.34\cdots n}}.$$

式中μ爲自(9)式求出之均方差.

以$\Sigma(x^2_{2.34\cdots n})=N\mu^2_{2.34\cdots n}$代入($\alpha$),而以上式除之,則得

$$b_{12.34\cdots n}=r_{12.34\cdots n}\frac{\mu_{1.34\cdots n}}{\mu_{2.34\cdots n}}. \tag{11}$$

因r之主要指數可以任意交換,又得

$$b_{21.34\cdots n}=r_{12.34\cdots n}\frac{\mu_{2.34\cdots n}}{\mu_{1.34\cdots n}}, \tag{11*}$$

取(11)與(11*)相乘即得(8)式所示之關係.

86. 據(11)式,消長係數已可求出.但若式中各量之級數甚高,則其計算仍多不便.然按積和特性,高級諸量,可化爲級數較前低一之量,故高級消長係數之算法,亦可因此化簡.今特證明如下:

按85節定理

$$\begin{aligned}\Sigma(x^2_{1.23\cdots n})&=\Sigma(x_{1.23\cdots n}\cdot x_{1.23\cdots n})=\Sigma(x_{1.23\cdots(n-1)}(x_1-b_{1n.23\cdots(n-1)}x_n\\&\quad-\cdots))=\Sigma(x_{1.23\cdots(n-1)}\cdot x_1)-b_{1n.23\cdots(n-1)}\Sigma(x_{1.23\cdots(n-1)}\cdot x_n)\\&=\Sigma(x_{1.23\cdots(n-1)}x_{1.23\cdots(n-1)})-b_{1n.23\cdots(n-1)}\Sigma(x_{1.23\cdots(n-1)}x_{n.23\cdots(n-1)})\\&=\Sigma(x^2_{1.23\cdots(n-1)})-b_{1n.23\cdots(n-1)}\Sigma(x_{1.23\cdots(n-1)}\cdot x_{n.23\cdots(n-1)}).\end{aligned}$$

據(9)及(α),上式三和依次等於

$$N\mu^2_{1.23\cdots n}\qquad N\mu^2_{1.23\cdots(n-1)}\qquad N\mu^2_{1.23\cdots(n-1)}b_{n1.23\cdots(n-1)},$$

因得

$$\mu^2_{1.23\cdots n}=\mu^2_{1.23\cdots(n-1)}(1-b_{1n.23\cdots(n-1)}b_{n1.23\cdots(n-1)}), \tag{β}$$

或

$$\mu^2_{1.23\cdots n}=\mu^2_{1.23\cdots(n-1)}(1-r^2_{1n.23\cdots(n-1)}). \tag{12}$$

由是易知高級消長係數恆小於 1.吾人若取式中之 n,以 $n-1$ 代之,則得循環公式

$$\mu^2_{1.23\cdots(n-1)}=\mu^2_{1\cdot23\cdots(n-2)}(1-r^2_{1(n-1)\cdot23\cdots(n-2)})$$

$$\mu^2_{1.23\cdots(n-2)}=\mu^2_{1\cdot23\cdots(n-3)}(1-r^2_{1(n-3)\cdot23\cdots(n-3)})$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\mu^2_{1.2} \qquad =\mu^2_1(1-r^2_{12}).$$

取諸式及(12)相乘則有

$$\mu^2_{1.23\cdots n}=\mu^2_1(1-r^2_{12})(1-r^2_{13.2})(1-r^2_{14.23})\cdots(1$$

$$-r^2_{1n.23\cdots(n-1)}), \tag{13}$$

是爲由低級相關係數及均方差計算高級均方差之公式.

同理

$$\Sigma(x_{1.34\cdots n}\cdot x_{2.34\cdots n})=\Sigma x_{1.34\cdots(n-1)}(x_2-b_{2n.34\cdots(n-1)}x_n-\cdots)$$

$$=\Sigma(x_{1.34\cdots(n-1)}\cdot x_2)-b_{2n.34\cdots(n-1)}\Sigma(x_{1.34\cdots(n-1)}\cdot x_n)$$

$$=\Sigma(x_{1.34\cdots(n-1)}\cdot x_{2.34\cdots(n-1)})$$

$$-b_{2n.34\cdots(n-1)}\Sigma(x_{1.34\cdots(n-1)}\cdot x_{n.34\cdots(n-1)});$$

據(9)及(α)上列三和依次等於

$$N\mu^2_{2.34\cdots n}b_{12.34\cdots n},\ N\mu^2_{2.34\cdots(n-1)}b_{12.34\cdots(n-1)},$$

$$N\mu^2_{n.23\cdots(n-1)}b_{1n.23\cdots(n-1)};$$

以此代入即得

$$\mu^2_{2.34\cdots n}b_{12.34\cdots n}=\mu^2_{2.34\cdots(n-1)}b_{12.34\cdots(n-1)}$$
$$-\mu^2_{n.34\cdots(n-1)}b_{1n.34\cdots(n-1)}b_{2n.34\cdots(n-1)}.$$

又由(11)及(11*)得

$$\frac{b_{2n.34\cdots(n-1)}}{b_{n2.34\cdots(n-1)}}=\frac{\mu^2_{2.34\cdots(n-1)}}{\mu^2_{n.34\cdots(n-1)}},$$

以式中 $b_{2n.34\cdots(n-1)}$ 之值代入前式,再依(β)消去μ,則得計算消長係數之循環公式爲:

$$b_{12.34\cdots n}=\frac{b_{12.34\cdots(n-1)}-b_{1n.34\cdots(n-1)}b_{n2.34\cdots(n-1)}}{1-b_{2n.34\cdots(n-1)}b_{n2.34\cdots(n-1)}}. \quad (14)$$

以(11)式代分子 b 值,(8)式代分母 b 值,則上式變爲

$$b_{12.34\cdots n}=\frac{r_{12.34\cdots(n-1)}-r_{1n.34\cdots(n-1)}r_{2n.34\cdots(n-1)}}{1-r^2_{2n.34\cdots(n-1)}}\ \frac{\mu_{1.34\cdots(n-1)}}{\mu_{2.34\cdots(n-1)}};$$

更交換其主要指數,復有

$$b_{21.34\cdots n}=\frac{r_{12.34\cdots(n-1)}-r_{1n.34\cdots(n-1)}r_{2n.34\cdots(n-1)}}{1-r^2_{1n.34\cdots(n-1)}}\ \frac{\mu_{2.34\cdots(n-1)}}{\mu_{1.34\cdots(n-1)}};$$

取二式乘積開方,則得計算相關係數之循環公式,

$$r_{12.34\cdots n}=\frac{r_{12.34\cdots(n-1)}-r_{1n.34\cdots(n-1)}r_{2n.34\cdots(n-1)}}{(1-r^2_{1n.34\cdots(n-1)})^{\frac{1}{2}}(1-r^2_{2n.34\cdots(n-1)})^{\frac{1}{2}}}. \quad (15)$$

命 $n=3$ 則得一級相關係數與零級相關係數之關係式爲:

$$r_{12.3}=\frac{r_{12}-r_{13}r_{23}}{(1-r^2_{13})^{\frac{1}{2}}(1-r^2_{23})^{\frac{1}{2}}}. \quad (16)$$

87. 茲將實地計算之步驟詳述如下:

先就值組 $X_1, X_2, X_3,\cdots$ 求出算術中值 $M_1, M_2, M_3,\cdots\cdots$ 均

方差μ_1, μ_2, μ_3……及每對相關係數 $r_{12}, r_{13}, \cdots r_{23}$……爲基本量.

I 然後自(16)式入手依(15)以求高級相關係數.

II. 更將此諸 r 代入(13)以求高級均方差.

III. 以所得結果代入(11),卽得所求之消長係數.

在特例若出現變數僅有三個,則其算法如下:

基本量: $M_1, M_2, M_3;\ \mu_1, \mu_2, \mu_3;\ r_{12}, r_{13}, r_{23}$.

I.

$$r_{12\cdot3}=\frac{r_{12}-r_{13}r_{23}}{(1-r_{13}^2)^{\frac{1}{2}}(1-r_{23}^2)^{\frac{1}{2}}},\quad r_{13.2}=\frac{r_{13}-r_{12}r_{23}}{(1-r_{12}^2)^{\frac{1}{2}}(1-r_{23}^2)^{\frac{1}{2}}},$$

$$r_{23.1}=\frac{r_{23}-r_{12}r_{13}}{(1-r_{12}^2)^{\frac{1}{2}}(1-r_{13}^2)^{\frac{1}{2}}}.$$

II.

$$\mu_{1.23}=\mu_1(1-r_{12}^2)^{\frac{1}{2}}(1-r_{13.2}^2)^{\frac{1}{2}},\quad \mu_{2.13}=\mu_2(1-r_{12}^2)^{\frac{1}{2}}(1-r_{13\cdot2}^2)^{\frac{1}{2}},$$

$$\mu_{3.12}=\mu_3(1-r_{13}^2)^{\frac{1}{2}}(1-r_{23.1}^2)^{\frac{1}{2}}.$$

因輔助指數可以任意交換,上列三式又可以他法求之,如

$$\mu_{1.23}=\mu_1(1-r_{13}^2)^{\frac{1}{2}}(1-r_{12.3}^2)^{\frac{1}{2}},$$

此式可供驗算之用.

III.

應用公式 $b_{12.3}=r_{12.3}\frac{\mu_{1.3}}{\mu_{2.3}}$ 須先算一級偏差,但此值除在

斯式外乏無用處,故其計算麻煩以能免去爲妙.據(13)式當 $n=3$ 有

$$\mu_{1.23}=\mu_1(1-r_{13}^2)^{\frac{1}{2}}(1-r_{12.3}^2)^{\frac{1}{2}}$$

$$\mu_{2.13}=\mu_2(1-r_{23}^2)^{\frac{1}{2}}(1-r_{12.3}^2)^{\frac{1}{2}},$$

相除得

$$\frac{\mu_{1.23}}{\mu_{2.13}}=\frac{\mu_1}{\mu_2}\left(\frac{1-r_{13}^2}{1-r_{23}^2}\right)^{\frac{1}{2}}.$$

又依(12)之式

$$\mu_{1.3}=\mu_1(1-r_{13}^2)^{\frac{1}{2}} \qquad \mu_{2.3}=\mu_2(1-r_{23}^2)^{\frac{1}{2}},$$

而

$$\frac{\mu_{1.3}}{\mu_{2.3}}=\frac{\mu_1}{\mu_2}\left(\frac{1-r_{23}^2}{1-r_{13}^2}\right)^{\frac{1}{2}};$$

與上式比較有

$$\frac{\mu_{1.3}}{\mu_{2.3}}=\frac{\mu_{1.23}}{\mu_{2.13}}.$$

因得

$$b_{12.3}=r_{12.3}\frac{\mu_{1.23}}{\mu_{2.13}} \qquad b_{13.2}=r_{13.2}\frac{\mu_{1.23}}{\mu_{3.12}}$$

$$b_{21.3}=r_{12.3}\frac{\mu_{2.13}}{\mu_{1.23}} \qquad b_{23.1}=r_{23.1}\frac{\mu_{2.13}}{\mu_{3.12}}$$

$$b_{31.2}=r_{13.2}\frac{\mu_{3.12}}{\mu_{1.23}} \qquad b_{32.1}=r_{23.1}\frac{\mu_{3.21}}{\mu_{2.13}}.$$

而三消長方程式爲:

$$x_1=b_{12.3}x_2+b_{13.2}x_3$$

$$x_2=b_{21.3}x_1+b_{23.1}x_3$$

$$x_3=b_{31.2}x_1+b_{32.1}x_2.$$

將計算出發點移於原點之上,則因

$$x_1=X_1-M_1,\ x_2=X_2-M_2,\ x_3=X_3-M_3$$

而得以原變數表示之消長方程式為

$$X_1=K_1+b_{12.3}X_2+b_{13.2}X_3$$

$$X_2=K_2+b_{21.3}X_1+b_{23.1}X_3$$

$$X_3=K_3+b_{31.2}X_1+b_{32.1}X_2.$$

此三式依幾何意義,為通過空間一點$M_1|M_2|M_3$之三平面.此三平面與二度相關論中之消長直線相當,故可名之為消長平面.

至三方程式中何者較為圓滿,則須視$\mu_{1.23}$, $\mu_{2.13}$, $\mu_{3.12}$之大小而定.以理論言,自以差誤最小者為妙.惟應用方面,則每因研究目的,僅取所須之式而討論之,其他計算一概從略.

至欲應用對數方法計算消長係數,則當式中平方數為已知時,除(1)式分子須先行計算外,其他均可直接行之.

88. 以上所述為相關係數之理論及算法,茲更就其應用範圍,加以討論,並附舉實例,依法計算,俾學者不獨於相關理論,能澈底了解,即實際運用亦有所遵循.

多變數相關理論之應用,以在生物學及遺傳學中為多.

其最先運用之以研究此項學科者,則爲相關論之創始人Galton 及Pearson 二氏.然此僅爲科學方面之研究,至其實際應用,則又多在醫學,畜牧,及園藝矣.

至於研究社會問題國民經濟[1],則須根據統計結果推出相互關係,故其應用相關理論之處爲數亦多.如結婚與生育,通商,失業,麥價之關係,肺癆與衞生情形之變化,疾病受分院治療之影響,貧民數目因家境變遷,救濟厚薄,年齡衰老,人口多寡之增減,銀行存款與借出,收入,折扣之關係,皆須應用相關原理以研究之.

他如物產統計中,收穫與氣象之關係,氣象相互之影響.亦相關論中最有趣味之問題也.

取異時同類之材料,分期加以研究,則各標觀察值之變化,可分二種,一爲慢性持久之變遷,一爲時起時落之昇降茲特就後者論其研究方法,以實際出現,每以後者爲多也

設吾人所究爲某年某種材料之互關,則可就此年前後各取若干年之同樣材料,分別求出各標觀察值之算術中數.取此年觀察值與相當中值之差,以爲觀察結果,則由此求出之互關必較與事實相近.

又材料之中,諸變數之相互關係,每不能同時實現.以當

---

(1) 1909年巴黎世界統計協會開會Yule曾於其所作報告之末列舉參考書名目.

甲標對於乙標之影響達於極大時,乙標對於丙標或且無甚影響,或其影響必須某時方能察出.如結婚對生育數之影響,以結婚後二年爲最大是.

總上觀之,相關論實爲研究原因,推測結果之最良工具.其對社會科學,及自然科學之應用,實無涯涘也[1].

69. 例(1)有取英國某區二十年間之秣草收穫X_1,雨量X_2,及在華氏42°[2],以上之春温總和X_3,而研究其相互關係者,其所得之基本量爲:

$$M_1=28.02\text{百磅（每畝）}$$

$$M_2=4.91\text{英寸}$$

$$M_3=594^{\circ}$$

$$\mu_1=4.42\text{百磅}$$

$$\mu_2=1.10\text{英寸}$$

$$\mu_3=85^{\circ}$$

$$r_{12}=+0.80$$

$$r_{13}=-0.40$$

$$r_{23}=-0.56.$$

試求變數之消長方程式.

---

(1) 最近亦有應用之以研究證實心理及死亡率者.習者可參看 Wilhelm Wirth "Spezielle Psychophysische Ma Bmethoden",Leipzig 1920, 及 Fredrik Esscher "Uber die Stenblichkeit in Schweden 1886—1914" Lund 1920. (2) 見252頁

取右表所示依 I 式求出之詳細結果,代入第 II 組公式中,可得諸一級均方差爲:

$\log \mu_{1.23}=0.42154 \quad \mu_{1.23}=\ 2.64$

$\log \mu_{2.13}=\bar{1}.77352 \quad \mu_{2.13}=\ 0.59$

$\log \mu_{3.12}=1.83654 \quad \mu_{3.12}=68.65.$

此值更可以下述二法驗正,

$$\log \mu_1=0.64542$$
$$\log(1-r_{12}^2)^{\frac{1}{2}}=\bar{1}.77815$$
$$\log(1-r_{13.2}^2)^{\frac{1}{2}}=\bar{1}.99797$$
$$\log_{1.23}=0.42154$$

$$\log \mu_1=0.64542$$
$$\log(1-r_{13}^2)^{\frac{1}{2}}=\bar{1}.96214$$
$$\log(1-r_{12.3}^2)^{\frac{1}{2}}=\bar{1}.81399$$
$$\log_{1.23}=0.42155.$$

應用 III 式,卽得

$\log b_{12.3}=0.52789 \quad b_{12.3}=\ 3.37$

$\log b_{13.2}=\bar{3}.56968 \quad b_{13.2}=\ 0.0037$

$\log b_{21.3}=0.23211 \quad b_{21.3}=\ 1.71$

$\log|b_{23.1}|=\bar{3}.57680 \quad b_{23.1}=-0.0038$

$\log b_{31.2}=0.39994 \quad b_{31.2}=\ 2.51$

零級相關係數	$\log(1-r^2_{\alpha\beta})^{\frac{1}{2}}$	分子積	分子	分子之對數	分母之對數	一級相關係數		
						$\log r_{\alpha\beta.\gamma}$	$r_{\alpha\beta.\gamma}$	$\log(1-r^2_{\alpha\beta.\gamma})^{\frac{1}{2}}$
$r_{12}=+0.80$	$\bar{1}.77815$	$+0.224$	$+0.573$	$\bar{1}.76042$	$\bar{1}.88042$	$\bar{1}.88000$	$r_{12.3}=+0.76$	$\bar{1}.81399$
$r_{13}=-0.40$	$\bar{1}.96214$	-0.448	-0.048	$\bar{2}.68124$	$\bar{1}.69643$	$\bar{2}.98481$	$r_{13.2}=+0.10$	$\bar{1}.99797$
$r_{23}=-0.56$	$\bar{1}.91829$	-0.320	-0.240	$\bar{1}.38021$	$\bar{1}.74029$	$\bar{1}.63992$	$r_{23.1}=-0.44$	$\bar{1}.95411$

$\log b_{32.1} = \bar{1}.70294 \quad b_{32.1} = -0.50.$

由是得以$M_1 M_2 M_3$爲出發點之消長方程式爲

$$x_1 = 3.37x_2 + 0.0037x_3$$

$$x_2 = 1.71x_1 - 0.0038x_3$$

$$x_3 = 2.51x_2 - 0.50x_2,$$

將出發值移至舊有原點,則得X_1, X_2, X_3之消長方程式爲:

$$X_1 = 9.28 + 3.37X_2 + 0.0037X_3$$

$$X_2 = -40.74 + 1.71X_1 - 0.0038X_3$$

$$X_3 = 526.13 + 2.51X_1 - 0.50X_2.$$

此三方程式之關係可說明之如下:

收穫與雨量之間,有一甚強之正互關,以其消長係數爲3.37表二者爲顯著之同向增長也.

收穫與春溫之間,有一甚弱之正互關,以其相關係數爲$r_{13.2}=0.10$,其消長係數爲0.0037,均爲數甚小也.

至雨量與春溫則爲弱小之負互關,蓋其$r_{23.1}=-0.44$, $b_{23.1}=-0.0038$, $b_{32.1}=-0.50$,亦卽當雨量增加則春溫下降,春溫上昇則雨量減少,惟其關係微小,不甚顯著耳!

取零級及二級之均方差對照排列,

---

(2) 其所以取此界限温度者蓋以較低温度對於牧草生長毫無影響之故.(註250頁[註2])

$\mu_1=4.42$　　$\mu_{12.3}=2.64$

$\mu_2=1.10$　　$\mu_{21.3}=0.59$

$\mu_3=85$　　$\mu_{32.1}=68.65$,

易知三標聯立,則μ值減小,其理由可自(12),(13)直接察出,毋庸再加說明.

就學理言,方程式中以第二式爲最圓滿,以其均方差誤爲最小.惟實用方面,則以第一式爲優,以吾人目的爲研究收穫所受雨量及春溫之影響也.

又收穫與春溫原爲負互關,因受雨量牽制,反變爲正,此亦吾人所宜留意者.

90. 例(2)英國法律,凡適合一定條件之貧民,得受政府之救濟.今取此種貧民之增率,而研究其與環境之關係.此處出現之變數爲:

X_1　1881—1891年,由政府救濟之貧民百分增率.

X_2　院外領取恤金,及居居貧民院者之比值增率.

X_3　年齡過65歲之貧民,與人口總數之比值增率.

X_4　人口之百分增率.

所得32市區之結果如下:

		$r_{\alpha\beta}$	$\log(1-r^2_{\alpha\beta})^{\frac{1}{2}}$
$M_1=104.7$	$\mu_1=29.2$	$r_{12}=+0.52$	$\bar{1}.93153$
$M_2=90.6$	$\mu_2=41.7$	$r_{13}=+0.41$	$\bar{1}.96004$

$M_3=107.7$ $\mu_3=5\cdot5$ $r_{14}=-0.14$ $\bar{1}.99570$

$M_4=111.3$ $\mu_4=23\cdot8$ $r_{23}=+0.49$ 1.94038

$r_{24}=+0.23$ 1.98820

$r_{34}=+0.25$ $\bar{1}.98598.$

茲就上述結果說明其意義:

以1881年之結果爲100.則十年之中,貧民增加4.7%,院外及院內之比值減少9.4%,65歲老者增加7.7%,人口增加11.3%,其中變化最大者爲X_2,最小者爲X_3,以μ_2大而μ_3小也.又貧民增率與院外院內比值爲最大正互關,貧民增率與人口增率爲惟一負互關.

欲計算一級及二級相關係數,宜依(15)先計算其相當分子,如由r_{12}, r_{13}, r_{23}計算$r_{12.3}$, $r_{13.2}$, $r_{23.1}$之分子以求一級相關係數.再由$r_{12.3}$, $r_{14.3}$, $r_{24.3}$計算$r_{12.34}$, $r_{13.24}$ $r_{23.14}$之分子,以求二級相關係數是.

$r_{\alpha\beta\cdot\gamma}$計算表

$r_{\alpha\beta}$		分子中乘積	分子	$r_{\alpha\beta\cdot\gamma}$		$\log(1-r^2_{\alpha\beta\cdot\gamma})^{\frac{1}{2}}$
12	+0.52	0.2009	0.3191	12.3	0.4013	$\bar{1}.96186$
13	+0.41	0.2548	0.1552	13.2	0.2085	$\bar{1}.99035$
23	+0.49	0.2132	0.2768	23.1	0.3553	$\bar{1}.97070$
12	+0.52	−0.0322	0.5522	12.4	0.5731	$\bar{1}.91355$

14	−0.14	0.1196	−0.2596	14.2	−0.3132	$\bar{1}.97772$
24	+0.23	−0.0728	0.3028	24.1	0.3580	$\bar{1}.97021$
13	+0.41	−0.0350	0.4450	13.4	0.4642	$\bar{1}.94730$
14	−0.14	1.1025	−0.2425	14.3	−0.2745	$\bar{1}.98298$
34	+0.25	−0.0574	0.3074	34.1	0.3404	$\bar{1}.97325$
23	+0.49	0.0575	0.4325	23.4	0.4590	$\bar{1}.94862$
24	+0.23	0.1225	0.1075	24.3	0.1274	$\bar{1}.99644$
34	+0.25	0.1127	0.1373	34.2	0.1619	$\bar{1}.99423$

$r_{\alpha\beta\cdot\gamma\delta}$計算表

$r_{\alpha\beta\cdot\gamma}$		分子中乘積	分子	$r_{\alpha\beta\cdot\gamma\delta}$		$\log(1-r^2_{\alpha\beta\cdot\gamma\delta})^{\frac{1}{2}}$
12.4	0.5731	0.2130	0.3601	12.34	0.458	$\bar{1}.94898$
13.4	0.4642	0.2630	0.2012	13.24	0.276	1.98275
23.4	0.4590	0.2660	0.1930	23.14	0.266	1.98408
12.3	0.4013	−0.0350	0.4363	12.34	0.457	…
14.3	−0.2746	0.0511	−0.3257	14.23	−0.359	1.97013
24.3	0.1274	−0.1102	0.2376	24.13	0.270	1.98359
13.2	0.2085	−0.0505	0.2590	13.24	0.276	…
14.2	−0.3123	0.0338	−0.3461	14.23	−0.359	…
34.2	0.1619	−0.0652	0.2271	34.12	0.244	1.98662
23.1	0.3553	0.1218	0.2335	23.14	0.266	…
24.1	0.3580	0.1209	0.2371	24.13	0.270	…
34.1	0.3404	0.1272	0.2132	24.12	0.244	…

因二級 r 之計算方法不一,吾人更可利用他法,取上述結果逐步驗正.

決定消長係數,似宜計算二級均差,以 $b_{12.34}=r_{12.34}\frac{\mu_{1.23}}{\mu_{2.34}}$ 中含有二級均差也.然此式亦可以三級均差表之,蓋

$$\mu_{1.234}=\mu_1(1-r_{14}^2)^{\frac{1}{2}}(1-r_{13.4}^2)^{\frac{1}{2}}(1-r_{12.34}^2)^{\frac{1}{2}}$$

$$\mu_{2.134}=\mu_2(1-r_{24}^2)^{\frac{1}{2}}(1-r_{23.4}^2)^{\frac{1}{2}}(1-r_{12.34}^2)^{\frac{1}{2}},$$

又因

$$\mu_{1.34}=\mu_1(1-r_{14}^2)^{\frac{1}{2}}(1-r_{13\cdot4}^2)^{\frac{1}{2}}$$

$$\mu_{2.34}=\mu_2(1-r_{24}^2)^{\frac{1}{2}}(1-r_{23\cdot4}^2)^{\frac{1}{2}}$$

而

$$\frac{\mu_{1.34}}{\mu_{2.34}}=\frac{\mu_{1.234}}{\mu_{2.134}}=\frac{\mu_1(1-r_{14}^2)^{\frac{1}{2}}(1-r_{13\cdot4}^2)^{\frac{1}{2}}}{\mu_2(1-r_{23}^2)^{\frac{1}{2}}(1-r_{23\cdot4}^2)^{\frac{1}{2}}},$$

因得

$$b_{12.34}=r_{12\ 34}\frac{\mu_{1.234}}{\mu_{2.134}}.$$

同理可得其餘諸式.

至於三級之 μ,則可依(13)應用前表諸量直接計算即得:

$$\log\mu_{1.234}=1.35739 \qquad \mu_{1.234}=22.8$$

$$\log\mu_{2.134}=1.50596 \qquad \mu_{2.134}=32.1$$

$$\log\mu_{3.124}=0.65772 \qquad \mu_{3.124}=4.55$$

$$\log\mu_{4\ 123}=1.32911 \qquad \mu_{4.123}=21.3.$$

r 及 μ 既經求出,即可按(12)式計算消長係數此數共有

十二.惟依統計目的,決定主要標準後,僅求三個已足.茲特完全計算,以便列出諸消長方程,加以討論.吾人若以$M_1M_2M_3M_4$爲原點,則得含x_1, x_2, x_3, x_4之消長方程式:

$$x_1 = 0.325x_2 + 1.384x_3 - 0.383x_4$$

$$x_2 = 0.644x_1 + 1.875x_3 + 0.405x_4$$

$$x_3 = 0.055x_1 + 0.038x_2 + 0.052x_4$$

$$x_4 = -0.336x_1 + 0.180x_2 + 1.174x_3.$$

將變數移至舊有原點,則得含X_1, X_2, X_3, X_4之消長方程:

$$X_1 = -31.174 + 0.325X_2 + 1.384X_3 - 0.383X_4$$

$$X_2 = -223.841 + 0.644X_1 + 1.875X_3 + 0.405X_4$$

$$X_3 = 92.711 + 0.055X_1 + 0.038X_2 + 0.052X_4$$

$$X_4 = 3.731 - 0.336X_1 + 0.180X_2 + 1.174X_4.$$

四方程式中,以第三式之乖差爲最小.惟依應用目的,則以首式爲佳.以吾人研究者,爲貧民增率所受撫恤分配,及年齡衰老,人口增加之影響也.

至第一式之意義,則可說明之如下:

當比值X_2, X_3, X_4各增百分之一,則因

$$0.325 + 1.384 - 0.383 = 1.326,$$

而X_1平均增加1.326%.

又自此式消長係數,可知貧民增率所受年齡衰老之影響最大;院外院內比值增加之影響甚小;至其與人口增率

則成反向增長,亦即人口增加貧民減少也.

取全相關係數及偏相關係數而對列之:

$r_{12}=+0.52$	$r_{12.34}=+0.46$
$r_{13}=+0.41$	$r_{13.24}=+0.28$
$r_{14}=-0.14$	$r_{14.23}=-0.36$
$r_{23}=+0.49$	$r_{23.14}=+0.27$
$r_{24}=+0.23$	$r_{24.13}=+0.27$
$r_{34}=+0.25$	$r_{34.12}=+0.24.$

則得貧民增率與人口增率之負互關更爲增大.

[附註]依二變數 X_1, X_2 間全相關之關係式

$$\mu_{1.2}^2=\mu_1^2(1-r_{12}^2),$$

可得多變數 $X_1, X_2, X_3 \cdots X_n$ 總共相關係數之關係式爲:

$$\mu_{1.23\cdots n}^2=\mu_1^2(1-R_{1.23\cdots n}^2),$$

式中 $R_{1.23\cdots n}$ 表 X_1 所受 $X_2, X_3, X_4 \cdots X_n$ 之影響.

取此式及(13)式相較,即得 R 及 r 之關係爲

$$1-R_{1.23\cdots n}^2=(1-r_{12}^2)(1-r_{13.2}^2)(1-r_{14.23}^2)\cdots(1-r_{1n.23\cdots(n-1)}^2).$$

由是易知 $R_{1.23\cdots n}$ 較各 r 爲大,在特例 R 大於 r_{12},但仍爲小數.

同理可得 $R_{2.134\cdots n}$ 等之關係式.

在例(1)收穫所受雨量及溫度之影響其總共相關係數爲

$$R_{1.23}=0.80,\quad R_{2.13}=0.84,\quad R_{3.12}=0.57.$$

又在例(2)

$$R_{1.234}=0.63,\quad F_{2.134}=0.64,\quad R_{3.124}=0.56,\quad R_{4.123}=0.44;$$

取諸值與 r 相較,易知 R 大於 r.

本書引用之統計材料

（數字所示爲節目次第）

食畜活重價之中值. 46. 48.

歐洲男子之頭蓋直立弧距. 45.

人口統計中幾何中數之一例. 47.

耿湖 (Genf) 之雨量（應用對數方法）. 48.

賈非盡之高度（應用對數方法）. 48.

食畜之價格（應用對數方法）. 46. 48. 60.

農人之夏作工資. 52. 54.

美國後備兵之體高. 52. 54. 60. 112. 118.

氣壓計之水銀柱高. 54.

奥國死亡表. 57.

巴什旗人及韃靼人之體高. 44. 58.

八齡女生之體高. 59. 60.

父子之體高. 121.

劍橋逐日之氣壓. 61.

男子在24至25歲結婚時其新婦之年齡. 61.

瑞典及英國人民結婚之年齡. 61.

其有相互關係者:

I. 花梗與花瓣之數目. 64.

II. 衿取草 Trientalis europaea 花瓣之數目與其中最長者之長. 64.

III. 父子間之蕃殖力. 64. 71.

統計學ABC

世界書局出版

統計學，是從大量的數字事實中間把顯著的真實表現出來的技術。研究統計學者，不可不備一本統計學ABC。

統計學是切實的學問，本書認清時代的需要，除討論統計學發達史，統計資料的來源和集合，統計表之編製，統計圖之種類和作法，以及平均數，變量，相互關係等以外，對於必要之公式等，亦儘量介紹；業經廣東中山大學採作教本，可見本書內容之忠實了。

蔡毓聰著　精裝一冊定價六角　平裝一冊定價五角

32.152.—18.9.12.

中華民國二十二年七月印刷
中華民國二十二年七月出版

統計研究法（全一册）

（每册定價銀一元一角）
（外埠酌加郵費匯費）

原著者　裴倍爾
譯者　李仲珩
發行人　沈知方
出版印刷者彙　世界書局　上海大連灣路
發行所　世界書局　上海四馬路 暨各省各埠

（本書責任校對者楊　威）

統計研究法　價洋一元一角