經濟叢書

統計方法

金氏著

甯恩承譯

經濟叢書

統計方法

金氏原著
甯恩承譯

商務印書館發行

序

近來我們常常感覺到我們智識的饑荒,中國學術的饑荒,也常聽着大家呼喊着「學術饑荒,學術饑荒.」這種饑荒差不多是普遍的,各種學問都覺得饑荒.對於統計一方面的學問,我們更覺得饑荒.統計學在中國是一種新的科學,現在雖然大家都知道了統計學的重要,都知道我們需要統計方法,但是專門研究統計的人還很少,關於統計學和統計方法的書籍也很有限.要尋一本淺近詳明的統計方法課本真是尋鳳毛麟角了.

本書是一本淺近詳明的統計方法課本.敍述既然清楚簡明易於了解,引用的方法也很完備.在美國久已博得一般學校和人士們的贊許傳誦,在中國或者也是一本初步統計方法的善本.惟是「譒譯之難」我們人人都承認.統計學這種科學是必須精確的.譯的時候,一方面恐怕失了原意,不得不逐節逐句的推敲下去;他方面又須顧慮到中文的通順和明白清晰,要避去歐化中文的毛病.

更困難的，我們中國沒有確當的統計名詞．有些名詞意離音澀，一個人有一個譯法．有的名詞是從日本統計名詞生硬拉用來的．究竟是生澀．還有許多我們從來所沒有的名詞，必須創譯．創譯名詞是困難的．無論創譯得怎樣好，大家因爲和他不熟悉，便覺得生硬．有這種種困難再加上譯者的學識簡陋，要譯到十分完善的地步，自然是很難的事．不過投磚引玉希望因爲這本小書出版有更好的統計書籍出世．

原譯稿成於南大罷教風潮之中，譯者以輪迴教育一文鬧得風雨滿樓，全稿草草譯就後，未及覆按卽行付梓，也是譯者對於讀者一個抱歉的地方．

我的朋友林頌河君對於譯筆常有增進指正，我很感激．原稿中的圖表多由友人申郁文君幫忙製作，我也十分銘感．一並書此致謝．

甯恩承，

南開大學

一九二五，五，一．

原 序

本書的目的是要給與學生,經濟家,行政官員,著作家,及一切智識階級的一個簡易統計方法的課本.凡是打算求一點統計方法的普通智識而用以爲衆數張本的科學的研究,或分析,應用,衆數張本者,都不可不讀此書.本書的原意雖爲社會學,經濟學,行政管理之用,而此中所舉的一般的原則方法,亦可應用於各種統計.以作者自己之經驗,作者所教的學生並不全是專門的算學家,而大多數欲實用統計者更是如此.所以本書僅將所有應用於統計方法的緊要的算學定理原則,加以引用說明,其餘高深的算學原理,難懂的算學方法一概從略.

據作者所知在美國還沒有最新最近的完全統計方法出版.外國雖不乏善本,但是不是僅具統計的一部分,就是專合於生物學家,統計學家,或算學家.如本書這樣簡易完全者,還決無僅有.應這種時勢之要求,故本書敢出而問世.

本書列舉的參考書,僅是統計學中主要的作品.作者的意思並不是要讀者把每一章後的參考書完全參看,不過作爲學者參考的指南而已.學者若對於某項問題具有興趣欲求深造者,本書所舉參考書也可以足用了.

本書原稿經西克瑞博士 (Dr. Horace Secrist) 亢芒教授(Professor John R. Commons)俄弟教授(Professor T. K. Urdahl)及維斯康新大學各教授同學之品題,校閱,並給予許多有價值的指正,特致感謝.

本書所以能出而問世者大半皆由於余師亞當博士 (Dr. Thos. S. Adams) 指教之功,本書如有任何優點皆余師之力.

金 (Willford I. King)

維斯康新大學

(University of Wisconsin)

一九一一年九月.

統計方法目錄

統計方法

第一編　緒論

第一章　統計學沿革

1.　引言

在最近五十年一般人纔知道統計學的重要,統計學纔別立門戶自成一種科學. 實在, 現在的統計學也不過是樓房上最上一層的成功而已;他的基礎許多世紀以前就成立了. 這本書的目的是要講現代統計方法, 不是要講統計學的歷史,對於統計學的歷史自然不能作詳盡的講述,不過提綱挈要的將統計學演進的歷程從古到今約略的講述一下,使讀者對於現在的統計學更易於了解.

2.　古代的統計學

統計學的發達和國家組織的發達是同時的.部落時代無所謂統計，從部落成爲國家，各君主各王公都須要知道他個人所管區域內的各項事物,他就不得不蒐集一些消息了. 他要知道他的國裏有多少財富，以決定能收多少稅; 他要知道他有多少能打仗的兵，以決定他自己的戰鬬力大小; 有的時候發生一種特別的事，爲他自己的某一種目的也須要記載下來.凡此種種都是必須有一種統計.最早的統計是埃及建築金字塔時蒐集的統計張本(data).當時調查人民的數目,和財富的數目,爲是要分擔金字塔建築的工程. 這件事發生於紀元前 3050 年. 後來過了幾世紀,約在紀元前 1400 年的時候,羅馬斯第二(Romeses II)調查埃及全國的土地,他這調查的目的是要按他個人的意思把全國土地重新分配一下,使人民各得其所.

在聖經民數記(Book of Numbers) 上第一二兩章記載摩西(Moses)點數以色列(Israel)人的事迹.他的目的也是要知道戰鬬力的大小.後來,約在紀元前 1018 年, 大威(David) 又調查一次人口. 他的意思也是要知道他的戰

鬬力.就是在遠東,古時也有調查人口的事,約在紀元前1200年,中國大禹王分天下爲九州,記載河山人民,也就是古代的統計.

希羅道塔(Herodotus)說賴嘉克(Lykurgus)分拉剛尼亞地方爲三萬九千鄉與斯巴達九千鄉與蘭斯擔孟三萬鄉.

這是希臘調查的一個舉例.他的目的是要分配土地徵收租稅,分別人民,決定戰鬬力,等等事項.我們再看羅馬.羅馬自從他留斯(Servius Tullius)以後,有許多次戶口調查,調查的目的是徵收賦稅,調查人民的數目,各城中某一部分人民的生死應在某寺院註册.

到了中世紀封建時代,各諸侯各帝王常常調查他自己國內的人民財富.如夏理曼(Charlemagne),威廉(William the Conqueror),阿爾曼(Al-Mamum),德皇腓得利第二(Frederick II),英王愛德華第二(Edward II),都曾舉行調查,見於歷史.

由上邊這幾段事實看來,我們知道各君主都要調查戶口.他們所以要調查戶口的意思,無非要爲行政的一

個幫助;如徵稅,如分配土地,如徵兵,都要先知道人民的數目.他們的調查除羅馬而外,大都是臨時調查的,不按期的.

3. 重商時代 (Mercantilistic Period)

到了重商時代,西歐各國都採取重商主義.重商主義興起,而統計發達,那是當然的結果.因為信重商主義的人相信政府應當獎勵某一種實業,不惜用種種方法冀得土貨出口現金入口,而得有利的貿易均衡.若要知道某種法規對於管理出入口是不是急需,某種法規對於商業有什麼影響,自然非用統計不可.於是統計益為各國所重視了.更有此時各君王互爭雄長,政權日漸集中,政府的組織也漸漸複雜,需用統計之處較在中世紀更多了,更急了.用之急,考察的方法自然進步.各帝王平時卽測量自己國內的財源富力與敵國相比較,善自節用,以為戰爭時之用.各國因此成功者比比皆是,這又是統計見重於各帝王的原因.

1575年西班牙王腓力第二(Philip II)徵詢各教主縣令所轄區的情形.十七世紀之初徐來(Sully)曾為法王亨

利(Henry of Navarre)作一個很大的法國富源兵力表. 1665 年,克伯 (Colbert) 作商業統計. 1699 年路易十四 (Louis XIV) 限令各長官必須報告他所轄區的情形.此時各國已都知道統計學的重要了.

到了近代,統計學益形進步.首先用有組織的方法以蒐集統計張本者爲德國,德國不僅所用的方法日有進步,更是按期蒐集材料作成統計.1719年威廉第一起首蒐集每半年的報告,如人口報告,職業報告,住宅報告,不動產報告,賦稅報告,城市財政報告,都須按期呈報.後來這種報告改爲每三年蒐集一次,蒐集的結果卽作成表册.腓得力大帝(Frederick the Great)極信統計學的重要,於是他又擴大調查的範圍,舉凡人民的國籍,年齡,死亡,死亡的原因,農業,商業,製造業,都在調查之列.大帝個人對於統計的工作更富有興味,上有好者下必有甚焉,於是消息之精確,消息之完備,大見進步.自1747至1782年之間,完全的統計制度遂告成功了.

4. 近代的戶口調查

十年戶口調查先是美國的產物,美國憲法上規定:

『衆議院議員依人民數目比例選出之』憲法既有這種規定,戶口調查自然是必要之舉了.1790年,美國舉行第一次戶口調查.後來又過十幾年(在1801年)英國也舉行戶口調查.

1803年德國「關稅同盟」要免除各小國中間的二重稅僅徵出口稅,他規定關稅按人民多寡分擔.要使分擔平均,所以每三年舉行一次戶口調查.其他各國對於按期戶口調查也大爲贊成.各强國先後採用這按期調查的制度,最後中國亦於1911年舉行第一次正式戶口調查了.

物換星移,戶口調查也漸漸擴大他的範圍,近三四十年來戶口調查成爲一種極複雜的事了.1900年美國設立永久的戶口調查部.這部的職守,就是繼續的解決每期配口調查的問題,解釋每期戶口調查的統計.各强國差不多都設有統計局,專司用科學的估量作按期的統計.美國有中央統計局,各州有各州的統計局.

5. 比較統計學

現在各公共機關所蒐集的統計和私人所蒐集的統

計眞是指不勝曲，多極了．惟是各統計局各自爲政，不相合作，因而各局的統計各自享用，不相謀助．若把各地的統計拿來比較一下，那是一件很難的事．古時對於比較統計學不甚注意．迨後，重商主義盛行，歐洲各國互爭雄長，遂不能不調查對方的人民財富，以與自己的人民財富相比較，而知有所興廢．於是比較統計學便應時興起．

早在 1544 年，海底堡教授苗斯特(Sebastian Muenster) 發表一篇論文．論文裏邊說明古代各國的組織，財富，兵役，戰鬪力，商業，教會，法律，等項事．1562 年意大利人山收威奴(Francesco Sansovino)，1589年，包得樓(Giovanni Botero) 也發表同樣的論文．毛馬林的王(Seigneur de Montmarin)柏黎 (Pierre d'Avity) 作一部四函的大書．其中大半是比較統計．這部書的內容很完備，較之以前的比較統計也更精確．此後比較統計的論文發表日多，比較統計學遂大見進步．到了現在，更有統計彙編了"Statistical Dictionaries"．在這統計彙編中，差不多各項統計都有，所謂應有盡有，惟是各國的統計方法各自不同，所以仍欠精確．

6. 生命的統計與社會的統計

前節已經說過從前各國政府所以重視統計學的原因，是要測量他的國力，發達統計學作爲行政的一個幫助．十七世紀之初統計學又發現新用途了．在宗教革命之初，新教徒要減少教徒們的非法行爲，所以令教徒們必須在教會註册．關於生死婚喪都要註册，很多的英，德城市對於這件事辦理的很完備．1612 年斯崔堡大學教授歐瑞(Professor George Obrecht of Strasburg University)向政府建議說，政府應有一種完備的生命統計，和完備的罪犯統計．他作了一個統計的計劃，說明他的意思．他的這種宏規遠慮，是要用這種統計而知有所借鏡，對於改善人民的道德，建立生命保險，規定養老金的制度，都要有統計的根據．他的計劃可算深遠了．

1661 年倫敦葛蘭德(Capt. John Graunt)第一次作分析生命統計的研究．他研究的結論說，生產率和死亡率是常常不變的；生產男女的分配爲十四與十三之比；就是說十四個男孩子與十三個女孩子相比．每生一百個小孩子，次一年能死多少，他也算出來．所以若從精確的

生產調查表上,也能推算全國的人民共有多少.

1691 年布瑞斯盧 (Breslau) 的牧師牛門 (Caspor Neumann),從那城中的死亡註册官手中得了五千八百六十九個人的死亡記載.他從這個記載,證明兒童的死亡,並不是像平常一般人迷信七歲九歲是兒童的厄運,兒童們常在七歲九歲死亡.他這結論到了英國皇家學會,引起了一個著名的天文家科學家的注意.這天文家科學家就是海來(Edmund Halley),海氏利用牛門的數目算出第一個完全生命表.從這個表便能推度每年能死多少人.他這結論雖然不是為人壽保險而作,而人壽保險的科學的制度已見端倪了.

這個時候人們已經知道人壽保險了.人壽保險的起源,是因為船長常因船沉死亡,所以作船長的常和人家賭注.若是這船長不死,船主就要給那個人錢;若是船沉了他死了,那個人便要賠償船主的損失.還有一種和這相類的事也很風行一時,就是公司中重要人員的保險.因為這個重要人若死了,對於營業有極大的妨礙;所以公司也給他保險.但是保險費全是隨便亂定的,毫無科

學的精確，而且這種制度全為公司着想，若是一個要保險自己以便死後他的家中可以得着賠償，那是不行的。因此海氏的表作成近代人壽保險的基礎.1698 年倫敦成立第一個人壽保險機關，又過一年「孤寡保險會」"Society of Assurancy for Widows and Orphans"也成立了.

巴斯爾大學的教授邦納里(Jacques Bernyuolli)將算學應用於統計學上去. 他是 1705 年死的. 死後遺下或然數原理(the theory of probabilities).這個原理在近世統計學發達史裏邊佔極重要的地位.1741 年許斯密 (Johann Peter Süssmulch) 發表一篇論文，在這篇論文裏邊他用統計方法說明「自然律」"Natural Order". 於是統計學又進一步.許氏在他的論文裏邊說明結婚時期的男女數目，差不多是相等的，所以他絕對相信一夫一婦制. 他的意思是說一夫一婦制是萬世不易的神聖規律.同時他找出各選區人民的年齡，他又發現生產和死亡的比率是常常不變的. 他說城市中的死亡率比鄉村的死亡率高. 他的解釋，是因為城市中的奢侈淫佚，上干天怒了，上帝令城市中人民多死亡.

十九世紀之初著名科學家拉賴斯 (Laplace) 和否瑞 (Fourier) 繼許氏之後繼續研究其說. 其後又有柏里根 (Belgian) 地方一個著名天文家算學家快德來 (Lambert Adolphe Jacques Quetelet),把統計學應用於天文學和氣象學裏邊去. 他研究氣象的時候, 引起他研究草木生長的定期現象. 迨後他的研究範圍擴大,及於動物, 及於人類,舉凡社會的特質,道德的特質, 物質的特質,無一不在考究之列. 他得了一個很大的發現. 這個發現是什麼呢? 就是觀察各種現象, 他們的現象雖然千變萬化, 但是結果都是相彷的,每次考察裏邊都有一個常模(norm),大多數的項數都是和這常模相近的, 離常模愈遠, 項數愈少, 結果成爲算學上的常態遞減.他說假如若把考察的數目作成一個曲線圖, 必成一個規則的兩稱曲線 (binomial curve).這個曲線正與算學上的機遇定律(或然定律)(law of chance, or probability) 相合, 這種現象好似人類的行爲完全爲自然的定律所操縱, 各種行爲同是有規則的. 如犯罪,如自殺, 如意外事項的發生, 都表示一種比較不變的數目. 因此快氏相信人是環境的產物, 社會應爲個

人負責.他不承認個人還有一種勢力,他更反對定命論.

繼快氏而起者大有人在.各人爭相傳播快氏的學說.他們相信快氏學說是合於邏輯的,更爲推廣其意.如合胥爾爵士 (Sir F. W. Herschel) 於 1850 年宣稱,人類完全是環境的產物;無所謂自由意志.卽使有自由意志存在,這自由意志也是不可察覺的東西.歷史家布克爾(H. Thomas Buckle) 也有同樣的主張,相爲附和.

同時意大利學派對於這種學說大爲懷疑,一班德國統計家也隨着發生疑問.1871 年,斯克莫勒 (Gustav Schmoller) 另闢新說.他說人類行爲的規則,是因爲發生這種行爲的原因的規則.各事結果常常不同,這便是自由意志的表現.平常各事的結果常是不同的,常常和常態相違的,就使原因一致不變,結果也常不一致.此說一出,大爲一般人所歡迎,快氏的學說無形中就減價了.

7. 統計學對於經濟學的貢獻

很早一般人就知道統計學是有助於行政的.重商時代許多政府的政策,都是根據統計定的.十七世紀之末,金喬哥雷 (Gregory King) 用統計方法表明物價和供給

(supply)有一種一定的關係.到了十八世紀,無論那一個經濟學家多少都根據於統計以證明他的學說.歷史學派的經濟學家,更注重統計.因爲他們相信經濟學的原理定律不是憑空想出來的,必須有歷史的事實或具體的事實來證明.於是統計學在經濟學上遂成爲重要的科學了.富有經驗的政府統計家黑爾白蘭(Bruno Hildebrand)是力主是說的領袖.同時又有克尼(Karl Knies)作正確的統計應用方法.循序漸進,統計學上便應用日廣了.

8　統計方法

最早的分析統計張本的人,也就知道分析統計張本必須有一種方法.用這一定的方法去分析張本,所得的結果纔易於了解.統計學應用的範圍愈廣,張本愈增,統計方法問題也愈來愈複雜.精細的科學的考究並不是用粗糙的機器所能勝任.1741年安須生(Anchersen)作一統計表,比較歐洲各國的情況.1782年克樓木更利用幾何形比較各國情況.1824年毛恩(F. J. Mone)在他的統計學原理裏邊特別注重統計方法,注重用統計方法解決統計問題.1861年安琪爾(Ernst Engel)在他的人口調查的方法

裏邊也注重統計方法. 於是方法益形完備. 其後有盧密林 (Rumelun),克尼(Knies),衞諾(Adolf Wagner),布洛克(M. Block),又從而合之.在最近三四年中純粹的統計學理論更顯著的發達.生物學家如米振(August Meitzen),愛德臥斯(Francis Edgeworth),哥爾頓(Francis Galton),趙恩地克(E. L. Thorndike),克爾披生(Karl Pearson),又爾(G. Udny Yule),大文砲(C. B. Davenport).經濟界如柏替郞 (Jacques Bertillon), 寶來(A. L. Bowley),胡克 (K. H. Hooker),亞當 (Thos. S. Adams), 窪倫披生 (Warren Persons),對於統計學原理各有貢獻,統計學益形進步了.

9. 統計學的教授

先時學者教學統計學,不過是在教授的時候引一點政治經濟地理的數目作一個詳細說明罷了.統計學這個名詞,是完全指着這一類事說的.到了 1660 年,口林(Hermann Conring)在哈木斯德大學(University of Pelmstedt)作第一次統計講演. 口林是著名的醫生, 也是自然法的教授.十九世紀之初財政學家也視統計學爲他們的教授的一部.第一個人把這雜亂的知識作成合於邏輯整體的,

是馬堡大學 (University of Marburg) 教授阿沈衞爾 (Gottried Achenwall).(卽所謂統計學之父)阿氏從意大利文 "statista" 這個字造成「統計學」"statistics". 意文 statista 是政治家的意思. 阿氏以爲統計學是比較兩國的學問. 用比較纔能有一種政治行爲的正確指南. 他到西班牙,葡萄牙,法國,英國,丹麥,荷蘭,俄國,瑞典各國去演講.1746 年他起首講演了.

彼時教授統計學是包括在經濟學和地理的教授裏邊. 亞當斯密斯 (Adam Smith) 發表原富 "The Wealth of Nations" 第一次把經濟學作爲獨立的科學.後來又有斯梯臥(Stewart),馬爾薩斯 (Malthus),瑞克斗 (Ricardo),佘以 (Say), 沙頭利亞 (Sartorious), 教寇 (Jacob), 克拉斯 (Kraus)等繼續亞氏力爲傳授發表.

十九世紀之初瑞特(C. Ritter)發表一部地理與自然和歷史的關係."Science of the Earth in Relation to the Nature and History of Men". 這部書把地理也作爲獨立的科學了.而原來在統計學裏邊佔有極重要位置的人壽保險,現在也別具門戶,另立家業了.

在最近幾十年中，歐，美各大學纔把統計學獨立一科．但是研究的方法仍有些不同；有的把他作爲經濟學的附屬學科，有的作爲生物學的附屬學科．惟是統計學自爲一種科學，已毫無疑義了．

10. 統計學的派別

近代統計學要可分爲兩枝：一枝是統計方法，一枝是應用統計學．

統計方法是什麼呢？統計方法可以說是算學一部分的應用；以算學的一部分作成一定的方法規律，用這方法規律以解決各種事物的統計．這種規律有很多有些是能用於經濟的張本，也能用於生物的張本．有些祇能用於某一種張本，不能用於別種的張本．方法論學家並不是說限定某一種方法，是要因個個件事的特別考察而用一般的原理原則和各種方法．

應用統計學是什麼呢？從這「應用」的字意就知道應用統計學是把已經作成的統計規律，統計公式，應用到某一種具體的事實上去．統計方法與應用統計的關係，和論理統計學與應用統計學的關係是一樣的．方法論學

家大概是算學家.應用統計學者大概是戶口調查的專家,政府中的官吏,社會學家,經濟學家,保險公司的會計,或人類某種行爲的考察家.

應用統計學裏邊又可分爲兩枝:一枝是敍述 (descriptive)應用統計學;一枝是科學應用統計學(scientific).敍述統計學是敍述現在的,或過去的,數目記載.美國的戶口調查各種比較表,各部來源的表,無論是從前的,或是現在的,都是敍述統計學的好例.一般人對於這種統計最有興味.因爲每個人都要知道他的州裏有多少人口;這一州的人口比別一州是多呢,還是少;這州裏的財富是增加了呢,還是減少;這一州的外國人比別一州的外國人是多呢,還是少;凡此種種都是人民歡喜知道的事.

科學家對於統計學和平常人不同,科學家不僅是爲好奇心理而研究統計學,他們是更要用統計的張本作爲他們學說的根據.物理學家要蒐集過去的張本作爲物理上定律的根據.心理學家要用統計的根據,證明他的心理學說.生物學家要用統計證明遺傳定律,氣象學者要

用統計求出日斑和氣候的關係.經濟學者願意用統計證明「貨幣數量學說」.政治家要知道衞生對於關稅法有何影響.科學統計既採用統計方法的規律,也用敍述統計所蒐集的張本.近世統計學,最後這一步,便是科學統計.

11. 結論

我們已經把統計學的沿革,從最古到現在約略講過了.學者具有這個背景,便可進而研究下邊的統計方法了.

參考書

Meitzen August: History, Theory and Technique of Statistics, tr. by Roland P. Falkner, Amer. Acad. of Pol. and Soc. Science, Phila., 1891, Part I.

John, V.: Geschichte der Statistik Ferdinand Enke, Stuttgart, 1884.

Block, Maurice: Traité Theorique et Pratique de Statistique, Guillaumin et Cie, Paris, 1886, Chaps. I and II.

Bertillon, Jacques: Cours Élémentaire de Statistique, Société d' Editions Scientifiques, Paris, 1895, Première Partie.

Yule G. Udny: An Introduction to the Theory of Statistics, J. B.

Lippincott and Co., Phila., 1911, Introduction.

第二章 統計學的定義

12. 定義

前一章已經講過從前有很多的學問都叫做統計,現在統計的意義和從前不同了.從前統計學是「研究國家」(the study of the state) 的意思.現在不是這個意思了.我們所要的統計學的定義,是現代統計學的定義,不是從前的定義.現在統計學的定義,應當把他應包括的意思全搜羅無遺,不相關的意思一律剷除乾淨.我們先看看從前學者對於統計學所下的定義.韋伯斯特(Webster)說:『統計學是把一國內的事實按照人民的情境 (condition) 而分類的意思.所謂事實更是指那可用數目表明的事實,用數表能表明的事實,或是用別種方法可表列或分類之事實.』韋氏這個定義是按照字面作出來的.因爲「統計」(statist) 這個字從「國家」(state)這個字來的.這個定義在從前還可以說得下去;但是現在不行了.現在的統計學

包括天文，生物，等等一切的張本，不僅限於國家.現在的統計學的定義，應當包括新的意義纔是.

與韋氏的定義相彷而範圍較大者爲寶來 (Bowley) 的定義.寶來說:『統計學是測量社會有機體的科學，以社會爲全體測量其各部.』這個定義，祇限於一方面，就是人，和人的行爲.現代統計學包括一切生物，天文，物質，社會等等現象，這個定義仍嫌太狹.

更有人說統計學是『計數的科學』"the science of counting". 這個定義雖然免去了限於一面的毛病，但是又有別一種毛病.統計學不僅以計數卽算了事，他的大部分更是估度的.農務部蒐集產麥的統計，並不是實實在在計量出產的多少共出多少斗，不過只是比前一年的產麥和今年的產麥作一個估度罷了.用這估度的方法也能作很精確的報告.其實，並沒有實在去計量.更因實際上測量很多數目時，實在的計數常常是不可能的.例如，調查戶口的時候有許多人民是遺漏了，更有許多人民調查重複了，所以實在的計數也不可靠.

這定義還有一個缺點是祇能應用到蒐集統計張本，

不能應用到分析張本. 實在, 蒐集張本和分析張本二者在統計學裏邊是不能偏廢的,所以這個定義仍欠完善.

統計學一個主要目的是要給我們一個鳥瞰以俯察大羣的複雜現象,要把複雜的現象作成一個簡單的形狀,以便於研究而易於了解.什麼是代表衆數的簡單的形狀呢? 我們常用平均數. (averages) 爲衆數的代表. 因爲這樣,於是寶來(Bowley) 又說統計學是『平均數的科學』"Statistics may rightly be called the science of average", 但是現代統計學又不僅到平均數爲止,除平均數以外更有表明變動的曲線方法,如表明每年每期的溫度變動曲線.表明大小的形圖,常見者爲各國的人口比較,財富比較,出產比較,支出比較等,都是常用圖表明的.更有相關表(correlation table),相關係數(correlation coefficient),表明兩物的關係. 這幾項都不是平均數所能包辦的, 所以這個定義仍然太狹.

我們認爲適當的定義而範圍較寬者如下:統計學是一種用分析某種列舉, 或估度的事實之結果, 以判斷集合的自然或社會現象的方法.這個定義比以前的定義都

完備.自然仍有一些幻想的統計問題沒有包括在這定義之內;但是實用上,這個定義也可算爲寬廣了.

參考書

Bowley, Arthur L.: Elements of Statistics, Chas. Scribner's Sons., N. Y., 1907, Chap. I.

Meitzen, August: Statistics, pp. 89–109.

Yule, G. U.: Introduction to Statistics, Introduction.

Block, Maurice: Traité de Statistique, Chap. IV.

第三章 統計學的淵源特質及用途

13. 何以要用統計學

人類心意不能同時了解很多的複雜的東西,若不是把兩種複雜的事實作成一種簡單的形狀,這兩種事是不能相比的. 假如一個人拿着兩張兩村的人民財產表. 每村有一百個人, 每人名下寫着他所有財產的數目, 他從頭逐一念下去. 每念一個人的時候, 同是念他的財產數目. 若有人能把這兩村的人民財產聽後完全記住, 還能比較, 那真是特異的人了. 平常的人聽過了自然是茫無

頭緒,毫無所得的.兩個小小的村莊的人民財產尚且如此,若是比較兩個大國的人民財產更該如何困難呢!其他衆數複雜的現象,若沒有簡單的方法也是不能比較.一林中的樹,把每棵樹的高矮大小都逐一的說出,聽的人該有一種什麽印象呢?把兩個地方的氣象記載,一天一天的念出來,兩地的氣候能相比嗎?使這複雜的衆數事實變成簡單的形狀,此統計學所以有用哩.統計學能把複雜的事實作成總和,作成平均數,我們用總和,平均數就和用簡易數一樣了.或是把複雜的事實作成圖表,曲線,顯出一般的「趨勢」,使這事實合於我們的理解.

14 統計學的用途

把各種事實作成簡單的形狀以便於比較者這是統計學一個主要目的.我們要知道中國人口的多寡,不僅知道了便算了事,更要用現在的人口和過去幾十年間的人口相比較.由比較我們纔知道人口是增加了,或是減少了.和別國人口相比,纔知道那一國的人口多,那一國的人口少.用人口的增加和食物的供給相比,和製造業相比,和礦產相比,和財富相比,纔可以知道一國人民和

物產的情境如何. 相比的用途是要得一個比較的大小,不是要絕對的大小.

然而比較之作並不是僅爲滿足我們的好奇心理爲了事,他的主要用途是要解决一切政府中或經濟界的重要問題. 假如若沒有戶口調查, 選舉國會議員怎能分配公平呢? 我們要問染肺結核病的人是增加呢, 或是已經減少了? 這個問題是重要的, 無論是由國家理財方面說,防疫政策方面說,個人衛生方面說, 都是重要的, 若答這個問題自然非用統計不可. 鐵路的運費應當增加嗎? 要答這個問題之先, 必須要知道收入支出的可靠的統計. 總而言之現代的政府行爲,差不多沒有一件不根據統計的,就是現代的大公司的行動,也是常常根據於統計.

會計學的猛烈進步,和市政府統一賬目制度的要求,正是急需正確的統計.

近來政府中立了很多的委員會以考察管理各種私人和政府的行爲, 差不多各方面都考察到了. 他們的範圍的大小雖然不同, 但是同是根據於統計的報告. 所以委員會的報告要依照統計消息. 而其結果是不是可靠,

要看所用的統計方法是否正確,所搜集的材料是否精確為斷.

保險公司算他的保率,也必須根據統計.新式保險公司的日日進化,人類的生活狀況日日變遷,新的統計須要繼續的搜集,新的運算要繼續的作去.例如,現在各國很注意的工人失業保險,正是急待解決的問題,但是因為統計學還沒達到極完善的地步,雖有正確的方法和正當的預防,對於這種工人失業保險,仍是不可能的.

不僅是實際政治家和實業家視統計學為無價之寶,就是理論的經濟學家,研究自然現象的科學家,也都用統計為他們的學理臆說的證明.生物學家要用統計證明遺傳定律,經濟學家要用統計證明他的人口定律,工資定律,物價定律,或者用統計證明各現象的關係,如財政恐慌和工人失業是不是有關係,須用統計方可證明.社會學家要說明賣酒和犯罪的關係,須用統計來證明,說明賣酒和貧窮的關係,也須要用統計學來證明,賣酒和自殺的關係或和其他類似此類事物的關係,都須要用統計學來證明.

寶來 (Bowley) 說:『統計學正當之職務是增個人的經驗』.若是沒有統計學的知識,我們大部分的觀念都是不的確的.若把這不淸楚的觀念,用數碼表示出來,我們便能看出各事的關係了,也可以漸漸尋出一種規律以資遵守而管理他的運行變化.

15. 統計常態定律 (Law of Statistical Regularity)

近世統計學有一個最有價值的發現,就是能把一大羣極複雜的事物用比較很少的項數作代表,能使我們對於這件事物得有一個很淸楚的印象,他的結果還能比較的精確.譬如我們要知道美國工人的平均工資,自然不能把全美國的工人的工資都蒐集來一個不遺漏,然後再求他們的平均工資.只要有一組可爲全體代表的工人,我們把這一組工人的工資數目搜集來,用正當方法得出一個平均數,這個平均數雖然不能和從全體工人的工資之平均數絲毫不差,眞正相合.但是所差很小,實際上忽略去了也沒有什麼要緊.人類學者要硏究某一民族身體上的特質,只要由全民族中抽出一小部分人來謹愼的測量了,就可發現這族的特質,自無須把全族中個個都測

量到了，然後纔行．這是因爲自然界在算學上有一個或然定律．(law of possibilities)．這個定律說，若是從一大羣項數裏邊，任便取出一些項數來求他的平均數，這個平均數便可代表全體項數了；全體項數的特質也可以由這平均看出了．例如有一籃子，內盛一百萬個胡桃，把兩個人的眼睛用手巾蒙上，令這兩個人隨便從籃中往外拿胡桃，每一個人拿出三百個．他們兩個人所拿出來的胡桃，雖然大小不同，但是每個人所拿的平均重量差不多是相同的．再有用這全體一百萬求出的每個胡桃的平均重量，和從三百個胡桃中求出的平均重量，也是相等的．

這個定律也可以用擲骰子來證明．假如，取四個骰子，兩面的點數共爲二十八，平均每次應有十四點向上．(兩面爲二十八，一面自然是十四)若連擲五十次，應該有七百點向上．若真照此作一個實驗，用四個骰子連擲五十次，所得的點數必與七百極相近．一般賭徒依這個原理連擲不已，終不易得勝贏錢卽以此．罪犯數目的常常不變，和自殺的數目常常不變，都是依照這個定律．所以前章講過從前的人相信人是環境的產物，並無所謂自由

意志，就是根據這個定律．保險公司所以能知道每年死多少人，也是根據這個定律．這個定律叫做「統計常態定律」．

然而也不可過信這個定律說是無論有多少數只要用一小部分標樣 (sample)，所得的結果和自全體所得的結果是永久一樣的．作標樣的項數愈多，則錯誤的機遇愈減．若是僅用很少的幾項爲標樣，結果的錯誤自然很大．若是作標樣的項數增加，則錯誤減少，最後錯誤到了極小，我們便可置諸不理了．

16. 衆數的惰性(Inertia of Large Numbers)

衆數惰性定律是或然原理的一個支系，這個定律是由衆數的現象生出來的．在一大羣的數目的裏邊，有一部分數目的變化向這一方向，同時有一相同部分的數目向相反的方向變化．反正兩個變化相消，所以全體的變化就很小了．在某一小區域內一年一年的產麥量數，可以相差很多，但是全世界的產麥量幾十年也不變．某一城的這一年的火災損失，可以五十倍於前一年的火災損失，但是這一年全國的火災損失和前一年的損失還是一

樣.火險公司本着這種道理,纔可以預先算出每年的火險損失.統計學的大部分也根據於這個定律,和或然原理.

這種惰性定律,並不是說無論時間如何變化,也是永久不變的;只是說大多衆數中的變化比較少項數中的變化比較的有規則.美國全國的火災的失損,固然一年一年的不變,但是因爲舊式的木料房屋漸漸改爲石料的建築了,火災的損失也有漸減的趨向.全世界的產麥量數雖然常是不變的,但是因爲某處開墾了很多新麥田,麥的產量也可增加.

這種惰性有時也不甚顯著.有一種事物他的變化偏向某一方向,惰性便不顯了.例如某省各城的負債已經都到了法定限額了.假如有一城把他的債還清了,對於全省內的各城之負債總額定有影響的.決不能再有一個城同時增加他的負債,把這城的減少數目抵消了,使全體沒有變化.假如沒有負債的法定限制,這一城的淸償債務,對於全省各城的總負債,定然影響很少.若有法定限制,他的影響便大了.

17. 不信任統計學

不懂科學的人依照他們對於新發明新發現的態度,可分為兩種:一種人是不問是非,盲目的相信新的發明,無論什麼怪誕不經的奇說,他都是沒問題的承認.再有一種人是稍有智識,一知半解的人,他們對於什麼真理都懷疑.他們說科學真理都是憑空揑造的.不懂統計學的人對於統計學的態度也有這兩種.有的人絕對相信統計學.他們說:『數碼不會撒謊的.』"Figures won't lie"有的人完全不信任統計學.他們說統計學全是胡說亂道,完全是騙人的東西.這兩派人各有理由,各有證據,未能說誰是誰非.

統計有一個缺欠就是他不能把他的內容好壞顯然的表現出來.一個極粗率的表,和一個用幾個月的工夫作成的精確表,在表面看來,同是一個表,他的價值是一樣的.其實何嘗如此.若判斷統計表現的價值,這統計的作者的能力如何,他的統計是不是可靠,他是不是一個巧敏的統計家都要注意.一位謹愼的考察者對於表中內部的證據,對於他的好點,是有價值的導線.近來有一個社會主義的作家,因為考察的怳忽,作了一個很可笑的

失敗.他的論文根據於政府的報告,那知道政府的報告根本就不可靠,報告上的錯誤的顯著,凡有統計學識的人,都可以看得出來.

統計學固然可以證明一切的事物,也有的證明是非科學的,只是勉强想用數碼表明自己所要得的結果.有些數碼很難分析,他的意思也是有疑義的,也是可辯駁的.但是大多數情境若是用正當的科學方法,不偏不倚的態度,從事分析,常是得到清楚正確的結果.所以統計學是最有用的科學;惟須知道他的正當用途,統計學纔有大價值.

18. 統計學的進步精確

我們已經知道統計學的精確是依照數碼的精確,數碼精確並不是一件容易事.但是也不能因爲他不容易,使用可疑的精確爲統計表的根據,以其不容易爲藉口.實在這種表若能附一説明,説明張本的來源,錯誤或精確的大概,使讀者不致於誤解.這種表也有科學的價値.這種價値是因爲他可爲將來可作更精確考察的基礎.一開首的考察,只要把應當勝過的難點表明出來,把所用

方法的優點劣點指示出來，結果的大概能推測出來，便算行了.這種原始的考察(preliminary investigation)是要緊的.若沒有原始考察，則大多數的詳細的考察，都是不可能的.即是可能的,也極困難.光浪速率的測量,並不是一下子就到現在這樣子，乃是一次一次測量，一次一次的偪近精確,最後纔有現在這樣精確.現代的生命表(life table),也不是一次作成的,是由牛門 (Casper Neumann)的教會人口記載.漸漸的進步，漸漸的改良,纔有現在生命表.所以不精確的考察，若是照着最後的結果說，自然是不對的.但是原始考察中不精確,正是本來的面目，正無可爲罪.

19. 統計學的限度

統計學雖然對於各種科學的考察都是極有用的工具，他的應用也有一個限度；他也有他不能勝過的缺欠.統計學常用平均數,但是這平均數所代表的各項是極不一致的.有的項極大，有的極小,而平均數把這種大小不同的情形完全不顧.這自然是一個缺欠.本書後邊雖然也有種種方法免去這種缺點,然而現在還沒有一種方法

可以使大羣的事物一目瞭然,同時他的各項小的不規則也完全呈露出來.我們常常想這不規則的項數是無關重要的.但是這種假設,並不是永久對的.這些小項固然有時不重要,但有時也很重要.火柴公司的工人,依全國的工人言之,是極不重要的一小部分,而受燐毒的工人,更是小部分中的一小部分.若是依全體工人的總平均痛苦言之,這一小部分的不幸,自然沒有什麽影響,並不算重要.但是對於身受其害的工人,我們能說是這痛苦不重要嗎?我們能說這小部分工人的痛苦是不重要的,無須爲他們設法校正嗎?統計學的本質就是沒把每個項看作重要的事,統計學所重的是全體,是大多數,不是個項.若是個項十分重要的時候,只好用別的方法.

20. 統計消息的來源

統計學中搜集張本的方法有種種,第一個方法是個人親自去搜集張本.這種方法僅能從一大羣項中取出一些標樣來分析之,他不能把所有的項全考察遍了.這是他的缺點.惟是,個人親自考察身臨其境,所考察事的真實狀態可以知道,標樣的精確也完全操於個人之手,這

也是他的長處.這個方法常用於自然科學,在某限度內也可用於社會的考察,或經濟的考察.若是考察的範圍太寬,這種方法是不適用的.

這個方法的一個改正,就是僱用一些調查員(enumerators)通信員去調查,然後考察者把他們調查的結果蒐集起來.支加哥的作投機事業的人曾用這個方法搜集世界各國麥子的收穫狀況,以預先知道麥的產量.這種方法,顯然的要用很大的費用,科學的徵詢是不行的.

第二個方法是用他人已搜集的統計,比較他們的成績,得出我們自己的結果.這種方法很難應用,實際也不常用.因爲我們不知道這材料的來源是否正當,他所用的方法是否正確,這材料是不是精確.又兼個人考察的範圍不同,目的不同,更不易適用.

因爲有這種困難,所以大多數社會考察,經濟考察,都要求助於政府的報告.政府的報告是無偏無私的,考察的範圍又通及全國.有這兩個長點,我們自要取其長的.近幾十年來各國戶口調查,各國統計局所搜集的各種統計極其完備,所謂應有盡有,其範圍既這樣寬廣,所

以研究社會科學的人,就視這些統計爲取材料的礦源了.政府調查比個人調查還有一個長點,就是政府有强迫力以搜集消息,個人是不能如此的.固然這種强迫的答語不一定精確,也不一定真實,但是被詢者的惰性,和忽略,總可免去了.在私人考察中這被詢者的惰性和忽略,是最大的兩件阻礙物.

21. 統計學的各方面(Phases of Statistics)

在統計學沿革裏邊我們已經知道統計學有好幾方面.每一方面在某一時期各有一學派主持,從歷史上說他的演進如下:

I. 經驗的統計學

(A)要在用以幫助行政管理.

II. 比較的統計學

(A)用爲經濟學說的根據.

(B)用爲決定政府政策的根據.

III. 用科學方法分析的統計學

(A)用以證明經濟的社會的科學的各種學說各種臆說.

(B)用以為政府行為的指南.

現在的統計學顯然有兩個分枝

I. 統計方法(Statistical Method);

II. 統計消息(Statistical Information).

統計方法對於統計家是最要緊的知識.有了正確的統計方法,纔能得到正確的統計消息.但是平常一般人所有興味的不是統計方法,是統計消息.正如工程師造橋一般.一般人只知道注意這個橋造的怎麼樣,工程師造橋時所用的算學方法他們全不注意.其實,算學對於造橋是最重要的東西.作統計而沒有統計方法的知識,正如一個工程師打算造一個大鐵橋沒有三角的知識一樣.這本書專為討論初步的統計方法,以為研究統計學的初步.這種統計方法的知識是很重要的,研究社會科學更重要.

參考書

Meitzen, August: Statistics, pp. 55-110, 143-155, 207-219.

Bowley, A. L.: Elements of Statistics, Chap. I.

Block, Maurice: Traité de Statistique, Chap. IV.

Bertillon, Jacques: Cours Élémentaire de Statistique, Parts IV and V.

Bowley, Arthur L.: An Elementary Manual of Statistics, Macdonald & Evans, London, 1910, Chap. I.

第二編　搜集材料

第四章　解決問題

22. 決定問題

統計考察家開宗明義第一件事,便是要確定他要解決的問題的性質.若其問題的範圍或外形稍有變更,則所用的方法卽要全部變更 至少也要有一部分的變更.例如一個人要研究比較工資,以便作爲他經濟學說的根據.第一他必要決定是要貨幣的工資呢,或是真正工資呢?第二他要決定是要知道這工資是按照一定的勞力給與的呢?或是按照一定的出產品給與的?或按照一定的時間給與的?更要問這工資是工人個人每年的進款呢?或是工人全家每年的進款?這些問題各自不同,所用的解決問題的方法亦逈然不同.所以第一件事,便是把問題的性質,範圍,定得十分清楚.

23. 問題成分的選擇

差不多每一個統計問題都有比較,這種比較常用百分比率表之,例如我們要比較各城市中的死亡人口,我們常說一千人中死了幾人.論城市的發展,我們常用百分表明之,比較自殺的數目,常說十萬人中自殺的有幾人.這幾個例,都要有一個分母和一個分子,以分母除分子所得的商數,謂之係數.我們選定這分子分母的時候要極注意,假如一下子選錯了,那就不能比較了,結果也全錯了.錯誤之爲物千形百態,我們不易查覺,常爲錯誤所欺.嘗有兩事相比的結果確類真正的結果,我們便以爲滿足了.社會上一般人也信之不疑了.迨後錯誤發現,大家又反其相信之心,爲絕對的不相信,對於統計學全部亦有些懷疑了.

有一舉例很可以說明選擇問題的成分錯誤.當美國與西班牙戰爭的時候,有一家新聞紙討論戰爭的死亡.據那新聞紙上所說在戰爭期中,美國海軍的死亡率,僅爲千分之九.同時紐約城中的死亡率爲千分之十六.於是結論便可說,在戰爭期中充當海軍,比住在紐約城中

還要平安了．我們對於這種結論若略加以思索，卽知這報紙上的數目是比較錯了．其所以得這個比率的原因，因爲他以紐約城中死亡總數爲分子，以全城人口爲分母，就得了這千分之十六的死亡率．以海軍死亡總數爲分子，以海軍全體爲分母，又得千分之九海軍的死亡率．就事實言之，這種數目是全然不能相比的．我們都知道老年人和小孩子的死亡率是高的，强壯的少年的死亡率是很低的．充當海軍的人都是强壯少年，不僅是強壯少年，入伍時又經過嚴格的檢查，具有疾病或身體素弱者，都被淘汰去了，剩下的人自然全是很强健的少年．自不能以老弱具備的紐約居民與此等強健少年海軍相提並論．如若比較紐約居民的死亡率與海軍的死亡率，必須以紐約城中之强壯少年的死亡率與海軍的死亡率相比，所得結論纔能正確，纔可以推斷兩方的死亡機遇．同理，假如一個人要比較克郎底克 (Klondike) 的慘殺案與支加哥的慘殺案，以推論二城強暴特性，自然不能以各城人口總數目除各城的慘殺案數卽得正當的結論．因爲平常慘殺案，不是小孩子作的，也不是婦女或老年人作的．慘殺案

的比率，當以確實的案件爲分子，以十六歲至六十歲中間的男子總數爲分母，所得的結果纔能正確. 法國統計學家伯提郎 (Bertillon) 對於係數曾說:『總要以結果與生此結果真正的原因相比.』就是說選擇分子分母要謹愼,此分子分母所得的商數纔能相比.

參考書

Meitzen, August Statistics, pp. 168-170.

第五章　統計的單位(The Statistical Unit)

24.　單位之決定

統計學是敍述數目的科學,是敍述預先定準了單位的數目的科學. 遽然看來，决定單位這件事,是很簡單一件事. 其實並不然.在未得結論之前,必須先想我們所要的結果的性質. 最早的調查可算爲計數人口. 美國原是每十年調查戶口一次，此調查的原意，就是要知道每州的人口多寡，以分配衆議院議員的名額. 美國憲法修正十四第二節規定衆議院議員按照各州人口之多寡依比

例選出之，不納稅的印底安人除外．由這一條規定便生出疑問了．作這條憲法人的原意，就是如字面上那樣規定嗎？選舉人這個「人」怎麼解釋呢？一個在城市中作工的印底安人卽是憲法所指的人嗎？或是憲法僅指純印底安人呢？憲法上所謂印底安人也包括半血統的印底安人嗎？(半血統，是父母中有一人是白人，一人是印底安人.)假如半血統的人是包括在內的，那們一個人具有十六分之一的印底安血統，算不算是印底安人呢？一個法國旅行家，當調查戶口那一天他正在紐約城，這個人字的單位算不算他？假如這個「人」字只指本地的住民而言，那們大口塔地方一個農夫，當調查時他越境到加拿大去作買賣去了，這個人算不算作住民呢？假如這個農夫算作住民，他的兒子同一個工程師到巴拿馬去住幾個月，這該如何算法呢？由這幾個問題就可表明定單位的困難，簡單的單位如「人」字還有這們多的難題，其他單位可想而知了．美國戶口調查部每十年調查田地一次，「一塊田」這個名詞該怎麼定呢？五英畝一個市場公園，是不是一塊田地？美政府的數千畝的牧場算作一塊田嗎？設一個人有兩塊

地,每塊都是八十英畝,兩地相距約半英里.這個人他自己兼耕這兩段地,這算作一塊田呢,或是兩塊田呢?假如他僱人耕一塊,其餘一塊自耕,還是如前一樣算法或另算呢?假如他租出八十畝結果如何呢?假如他把這田租與十二家,這十二家都受他的監督,這該算幾塊田呢?

一個「人」字「一塊田」作爲單位,尙不易定得精確清楚,若以「罪犯」爲單位,其困難更當如何?必須重罪纔算爲罪犯嗎?假如一個人犯了殺人罪,他賄賂法官得了釋放出獄,他還是不是罪犯呢?這種問題我們便有不能決定的困難了.然而若打算要比較犯罪的統計,我們必須先定出罪犯的單位來.單位最後的決定,宜與名稱相對.卽使名稱的合宜有可疑之點,也必須把單位,精確的無誤的決定清楚了.相比的兩個時間或兩地位必須要用同一的單位,欲單位之正確,在考察以前須詳細審查,把所有的瑣屑小點全都找出來,凡能想到的問題都要想到.假如決定單位這個人並不是親自去調查,他要僱一些調查員去調查,單位的說明更須用明瞭的語句,單位的詳解要使調查員易懂易尋.調查員多爲平常智力之人,如僱用一

些調查員去調查稍有雙關之語句,便足引起很大的混淆.

25. 單位必具之特質

不僅是要把單位的定義定得十分清楚正確,而擇定的單位更須具有可以正確決定之性質者方可.假設一個人要比較兩個社會的教育情形,他擇定了以「有教育的人」爲單位,爲分子,這便錯了.因爲並不能有一個人合於這個名稱的,那種人是有教育的人呢?大學畢業嗎?小學畢業生呢?或識字者便是呢?單位須是用感覺能測量的東西,如「大學畢業生」,「得有高級中學文憑者」,「入學幾年者」,「能讀寫者」,「熟習某幾種事業者」,或他種具有特別證據可確定之事爲單位,較之以「有教育」三字爲單位之內心的特質爲單位要好得多了.總而言之,抽象的名稱,必須以其具體的證據測量之.

參考書

Bowley, A. L.: Elements of Statistics, Chap. III, Meitzen, August, Statistics, pp. 117-118

第六章　搜集張本的計劃

26. 初步計劃(Preliminary Plans)

在搜集材料之前,必須先詳細審查問題的各方面,免得空費精力再重作第二次徵詢. 如將問題謹慎研究, 則錯誤可減至最小度. 統計工作有一特質, 卽是要觀察事物於事前. 各種可能的錯誤的來源都要尋出來, 除去之. 大概的結論也可以事前約略測定, 舉凡問題, 問題的成分,單位,調查單,調查員,列表,進行方法,時間,費用, 都要有詳盡的說明. 統計工作是煩膩的, 卽使我們用盡種種預防方法, 錯誤與誤解仍是不免. 但是吾人萬不可因此自餒, 要知費一點鐘於事前之預備, 便省十點鐘於事後之尋錯校誤.

前邊已說過考察有種種方法, 大概分有兩種: 一爲原始考察,一爲繼起考察.

I. 繼起考察(Secondary Investigation)

27. 特質

這種考察需要原始的工作很少,差不多各種事都是依照已搜集的材料作去,而搜集材料的計劃到不甚容易.

II. 原始考察(Primary Investigation)

28. 一般的特質

原始考察因考察的境地不同，所用的方法也不同．只要考察者自信那一種方法最合宜，便用那一種方法．平常有四種可能的方法：(1) 個人親自考察；(2) 通信員的估計；(3) 被詢者所填之調查單；(4) 調查員所填之調查單．究竟要那一種方法爲最合宜，自然要依問題之性質，所希望之結果與所有的財力爲轉移．

29. 個人親自考察(Personal Investigation)

這種方法最適用於內涵的研究，用這種方法最好的例子，便是普來(Le Play)在歐洲的工人預算考察．他的方法是在一個工人的家裏住幾個月的工夫，躬自考察他的預算．這一家考察完了，再用同一的方法到第二家的工人去住．如此連續作去，一直用幾年的工夫便得了一個非常精確的統計．但是這種方法考察這些工人的家要用很多的時間，結果還是極少數的家數，仍不能據此以代表全體．卽使普來用盡畢生光陰於此等考察之上，而考察之工人家數，亦屬有限．然其考察之方法大有可注意之點．阿支楊(Arthur Young)用旅行方法，亦可謂內涵

研究考察方法之一種. 這種親自考察的方法, 雖然因爲個人親自徵詢所得的結果比較精確,而所考察的範圍太狹,恐不足代表全體. 兼之個人自信成見太深,也不免有不相宜之點. 蓋個人之偏見,與個人之慾望,每於不知不覺中混入結論裏邊去了,這自然是這種方法的缺點.

30.　通信員的估計方法(Estimates from Correspondents)

若是我們僅要知道一件事的大概結果,我們可用通信員的估計方法. 這種方法極容易舉辦,費用也極廉省. 例如搜集每年的收穫報告,便可用這種通信員估計的方法.他們只要估量這一年的收穫比前一年的收穫增加了百分之幾,或是減少了百分之幾便行了. 每個人的報告,自然不十分精確, 但是他們的錯誤到可互相補償, 有的報告的錯誤是比真正數大了,有的錯誤是比真正數小了,於是這兩種錯誤便可以補償抵消了.假如我們若能得到多數的報告, 則這種結果必與真正結果相差不遠. 通信員估計的方法之變相,卽是於全國各處設有代理人以搜集通信員的估計報告.

31. 被詢者所填之調查單(Schedules to Be Filled by Informants)

更有一種外延的考察,卽是被詢者所填的調查單的方法. 此方法與前法不同之點,是調查單上所問的問題,是被詢者所熟知的事項. 這種方法最大的缺點, 是調查的成功與否, 全是依被詢者對於此事之興味大小而定. 除非有政府的命令或政府所派的代表具有强迫實行之全權, 大多數的調查單都是寄不回來的, 卽是有寄回的調查單,這調查單填的也難完全, 常常有許多的錯誤. 若是問題十分簡單, 寄回來的正確調查單或者多一些, 平常的被詢者對於這種調查全然不懂, 卽使稍懂一點, 填表時也極不經心. 所以調查單的問題必須十分簡單, 調查單上必須寫明調查的目的, 負責任的人, 負責的機關, 否則被詢者之狐疑妄想與其惰性相合,他們便把這調查單置之不理了. 問題所問的事,更須是現在的事,若問過去的事,一方面旣少精確之記載,而填時更多煩擾. 此法之主要利點爲用很少的費用, 可得各地的消息. 若收回的調查單, 有一些填的十分完全, 可爲全體之標樣者,這

調查也可得到一個大致精確的結果.私人調查常用此法.政府調查也可用此法,如工資統計,失業統計,地方歲出統計,天氣統計等等,常常都是用這種方法.若是有一種法律規定,被詢者必須填答這調查單,否則加以懲戒,則所得的結果常是滿意的.有時自願的報告,如氣候報告等,亦常有可信的張本.如調查者能定期調查,被詢者亦按期報告,也常得有可信的張本.若偶然舉行一次調查,則結果便不甚可靠了.後邊所論各種填表製題種種原則,亦應用於此處.為便利起見,容後章備述.

32. 調查員所填之調查單(Schedules in Charge of Enumerators)

政府中重要的調查,常派調查員去調查,此法若施之於私人調查,則所費未免太大.這種派調查員,自然是外延調查法之最好方法,所用的調查單可以完備一些.比把調查單直接寄與被詢者定然完備,調查的範圍也無妨擴大一些.然而調查單的格式形狀也須要加意選擇,以便於調查員之攜帶運用.如調查單是一張很大的摺紙,用時既不方便,且易於摺壞,調查單上的行列地方須要

合宜,不至於使目力混淆,綱目題名須有正當的聯貫,彼此的關係,須要清楚,各題名綱名的字句須極清晰,使平常智力之人亦一目瞭然.各問題中的字句也須十分斟酌,以免去可能的疑誤,與衝突的解釋.我們所要的精確程度也要明白書出,有這種種預防,便可減少許多錯誤,免去許多不必須的混淆,調查員也省了許多時間.

下面的表是個職業與工資調查單一個樣張.這張調查單是要從工人方面調查,不是要從僱主方面訪問.調查員必有的工具是解釋的說明調查單的用途,及一張已經填好作爲樣本的調查單.

現在更有一種很通行調查單,即是硬片調查單,把每個人的消息填在一張硬片上,這些硬片全是分開的,一張硬片只填一個人的事.假如同一些材料,先後的排列分類欲其不同,最好用這種硬片的方法.例如第一表原來要按職業把工人分類,後來他又要依工資分類,再後又要依失業分類,如此則用硬片爲最便利.凡是繼續的擴張的記載,硬片制度爲唯一的好方法.然而硬片得用很多,要佔很大的地位,算總和時也不便利,也是他的

缺點. 美國調查戶口調查單與硬片合用, 調查員先將張本塡入調查單內, 然後再按類分寫在各相當硬片上, 再用電機將此等硬片表列穿在一起.總之此方法是否合宜應用,要依考察的問題之特質而定.

33. 選擇問題(The Choice of Questions)

選擇問題時先要看這調查單是由被詢者自願的塡寫呢,或是法律規定他必須塡寫.二者的情形不同,選擇的問題也自不同.若是這調查單是個人隨意自願塡寫的,我們的問題須要極簡單. 問題的數目要極少, 更要容易答的問題, 否則大部分的被詢者都不肯塡答了. 若是調查單是法律規定被詢人必須塡答,問題的數目不妨多一些,但是問題也務須簡單,以便答案正確有價值.若派調查員去調查,問題也可稍爲複雜,問題的數目也可加多,若調查員具有一種法律允許的强迫權更可如此.然問題亦不可太難, 使調查員用印出來說明書之助, 還不能得出正當的解釋, 如調查員澈底明瞭這問題, 即使被詢者對於這問題完全不清楚,調查員也能用相關的幾個問題得出正確答語來.

尤須注意者,增加一個問題,即是增加一分費用.國家戶口調查,只要一個詢問,便費幾萬元.所以問題的數目,不能不受財力之限制,所有的問題以極必需的爲限.不重要的問題,大可去掉不問.若是打算把調查的結果作成表册,最好的問題是用「是」「否」或數字能答的問題.假如一個問題問被詢者的教育如何,這問題的答語,定然分歧不一.大概是「好」「有一點」「很好」「高級中學」等等答語.此類答語,决不易塡入統計的表裏去.假如我們問他在學校讀書幾年,這種用數目的答語,自然可以塡入表裏邊去,結果也可確當了.

更有一件要注意的是問題不要引起被詢者厭惡發怒,也不要引起他的偏見.如被詢者一有反對之心,則正確的答語便不可得了.對於個人私德有礙的問題,或對於身心兩方之缺點所有的發問,最易引起被詢者之反感,最好不問這類的問題.如必須問這類問題的時候,我們也要另想方法得出我們所希望的結果來.例如一般婦女們總是喜歡把她年齡少報幾歲,若必須問她們年齡時,可用旁證的問題問之,不要問她現年幾何,要問她是那

一年生的.從她的生日算到現在,便可知她現在的年齡了.一個精明的調查員可用種種方法得到他要得的結果,此不過一例而已.

再有一項該注意的是要考究這問題是不是正是我們所要問之點.有沒有令人誤解生疑的地方.用他人調查之結果最大之難點,即是問題所問之點不能一致,問題稍不一致,結果便大不相同.例如一個人要調查某一鐵路的哩數,他先要決定算不算支路,算支路與不算支路,結果自然不同的.

總起來說,選擇問題,有下列幾個原則:

1. 問題的數目要比較的少.
2. 要用數目或「是」「否」能答的問題.
3. 問題要簡單易懂.
4. 問題不要引起人之反感或偏見.
5. 問題要必須的,不必要的問題即去之.
6. 問題愈具有旁證性質者愈好.
7. 問題所問之點要正是所要之點.

34. 規定調查的範圍

現在已將問題之決定,徵詢的方法,調查單的填法,問題的選擇等要略述過.現在再講考察的範圍.考察者先要決定他要考察的範圍.時間,地位,固然是無限制的,而其考察之範圍必要有一定.若出此範圍以外,卽不復問了.例如研究收入,也可是一城的收入,也可是數城的收入,一國的收入,或數國的收入;我們可以比較同時各地的收入,也可以比較同地異時的收入,究竟我們所要研究的收入是那一種,不得不預先定好.

35. 代表的張本(Representative Data)

私人考察的範圍,每不能如個人所願之廣.故有時用標樣的張本以濟其窮.譬如要考察工人家中的收入支出預算表,不必把全國的工人家庭都調查了,只要標樣的工人家庭便夠了.所謂標樣家庭是什麼呢?卽其可以代表全體工人家庭的家庭.若標樣正確,調查的件數不少,結果一定可滿意的.夾平教授(Professor Chapin)研究紐約工人狀況.雖其考察之家數遠不如美國戶口調查部調查之多,其結果則相合,此卽用標樣張本之一例.但是此法易滋流弊.若一下子將標樣選錯了,他卽不能

代表全體了.偶然發生之事我們選之爲標樣,或考察者預存成見,以其所希望的事選爲標樣,皆不能代表全體的真面目.例如一位社會學家因爲他要證明生活狀況日下,他就盡選些貧家調查,結果自然生活日非了.再有一位樂觀的社會學家他盡檢着富家調查,結果全是富人了.這兩種標樣都不能代表真正的生活狀況.正確的標樣(sampling),必從衆數中隨意(毫不加以個人私見)選一些數,再將這些分爲各組,由各組中選出之.

此法之變象,卽是將各項依其大小排列之,愈近愈妙.由組距相等處選出標樣來.此法固具科學上的精確,惟實際上難於應用.生物學上常用此法,蓋生物學所研究之事,每患項數太多,故用此以濟之耳.

36. 調查員之選擇

考察的結果好壞,要看調查員的好壞,這是毫無疑義的.調查員要好,結果自然是好.調查員要不好,結果一定不會好的.調查員必須很聰明,纔可以把被詢者的含混的答語免去,使其答語正合於載述.除聰明外更須誠實,勤勉.不謹愼的調查員,塡調查單時,隨便七亂八糟的

塡上,結果全行無用.有偏見的人不能爲調查員,因爲私心希望得某種結果,這種成見不知不覺中便雜入記載中去了.調查員更須雅緻和靄,對人要客氣,庶幾進行順利,正確之答案可得了.

第七章 搜集材料(The Collection of Material)

37. 繼起方法

我們若用已搜集之材料爲根據,是謂繼起方法.用繼起方法須有幾件應注意之事,萬不可據他人之材料以爲己有,毫不考究其來源便貿然採用.第一要知道原來搜集這材料的人是否可信,他的能力如何,他的材料是他個人在房裏憑空杜撰的也未可知.假如我們把他杜撰的材料作爲科學上的根據,豈不太愚.即是我們已知他的材料可信,我們還須注意下列各事:

1. 數碼是從何處得來的.
2. 單位的定義,與給予調查員之說明如何.
3. 搜集這材料的原意何在.

4. 他用什麼搜集方法.

5. 數碼的精確程度如何.

如果以上各點都有滿意的決定,於是考察者纔可用他人的數碼.

數碼中常有不相合者,蓋由於某部分之數碼忽略遺漏,誤置於他部分總和之中.學者以謹愼的推理,與精密的觀察,或可尋出錯誤的來源何在.

同一樣的事情,若是他的來源不同,數碼也便不相同了.若遇此種情形時,當極力考察這不相同的數目的來源,假如二數的來源都很可靠,即可用力詳爲考究.否則除了有實證正確之數目都可棄之不用,或僅留兩數相合之數碼,不相合者便棄之.

38. 元始方法(The Primary Method)

調查員或被詢者把調查單都已寄回來了,此時並不能立即製表,仍有許多事須作.每一張調查單都要察視一遍,視其是否有遺落之處,或錯誤之處.有時若調查單省略不完,可作第二次徵詢.惟第二次徵詢,仍須更多的工夫,仍須更多的費用,而所得者常是特別的少.平常的

空白並不是由於忽略,常由於問題有些難答.

若數碼的錯誤非常的顯然,卽可較正之,或竟棄之不用.有時全調查單太不完全,或太纏繞不清,我們也可以全部棄之.任要少而正確,不要多而錯誤.數少而正確之數還可校正,得出逼近的精確.數多而錯誤,則校正無術了.

參考書

Meitzen, August: Statistics, pp. 120–129, 155–185.

Bowley, A. L.: Elements of Statistics, Chaps. II and III.

Yule, G. U.: Introduction to Statistics, Chaps. XIII and XIV.

Block, Maurice: Traité de Statistique, Chaps. IX and X.

Bertillon, Jacques: Cours Élémentaire de Statistique, Chaps. IV and V.

Bowley, A. L.: Elementary Manual of Statistics, Chap. XIII.

第八章 逼近與精確(Approximation and Accuracy)

39. 完全精確之難得

在第一章已經說過,統計學是估量的科學,並不是確實列舉的科學.譬如我們要計量美國各煤礦出產總額,

也不過知道一個大概的逼近數目,並不是每車煤都量得十分精確,絲毫也不差.各煤礦公司的報告,自然免不了錯誤,一定更有一些小煤礦公司沒有報告,他們的產量自然沒算在這總產量之內,若把這些事全算起來,則所謂煤產總量不知要與真正總量差幾萬噸,或幾百萬噸.所以精確只是比較的精確,並沒有絕對的精確.一九零九年美國產煤總量爲三萬九千七百萬噸,即差一百萬噸.亦不過爲百分之一的四分之一耳.作統計時有時即錯百分之四或百分之五,仍無關重要.此百分之一的四分之一,自是無關輕重之數.

凡能計量的東西,他的絕對精確,都是不可能的.用平常的呎可以把一個針量成糎,若再用精尺 (Vernier),可以量爲糎之十分之一,精確之度已行增加.如再用更精的方法,可以量成糎之千分之一,但是無論如何,針長之精確程度,是不能絕對精確.精細的方法,不過益行逼近精確而已,絕對的精確,是永遠得不到的.

40.　精確標準

在實質的科學上舉行測驗,還可得到很大的精確,

社會現象便不易測量了.他的精確更不易得,社會現象的錯誤,是常常有的,形形色色,紛至沓來,卽使極端謹愼,錯誤仍不能免.幸而社會現象中之小錯誤,常是無關宏旨的,對於一個問題之結果,並無偌大妨害,而欲得極大的精確,平常是妄費時間的.海關的每年收入,固然可以精確至於分,但是對於普通的統計,比較此種多餘的精確數碼,不但無益而且擾亂觀聽,蓋此種繁多之數,吾人不易領會,反將主要之數亦行忘卻了.

每一個統計問題,在未作之先,卽要決定一個精確的標準,所蒐集之材料,都要努力達到這個標準.惟此種精確標準,並不可謂爲精確之最高度,調查人的年齡,固然可以精確到一個人從生到現在共活多少日數,但是求出這日數後亦無何等用途.我們所要的數目,只要對於我們的問題具有相當的精確,卽算够了.

所以若量一個大湖的面積,用方哩爲單位,卽可爲精確了.量一個小水池,則須用立方呎或加倫爲單位.量鐵路之長以哩爲單位.量光浪則需用一糎之百萬之一爲單位.我們所要者乃比較的精確,並非絕對的精確.

41. 畧數(Round Number)

有時繁多的數碼,易於混淆觀聽,故用略數.卽使精確的數碼可以得到,亦不如用略數爲優.例如比較中國與美國的人口,最好是說中國人民四萬萬,美國人民九千萬,若把精確的戶口調查數目幾萬幾千幾百幾十幾人念給一般人聽,他們反到不得要領了.蓋數碼繁多,他們不易記住,反到不能比較了.

現在通行的書報雜誌上常用略數.卽科學上亦有用略數的時候.假如若知道美國過去三十年間之鋼鐵出產比較量,他的精確的數目固然可以得到,而下表亦可謂正確了.

年	鋼鐵產量以十萬噸爲單位
1880	12
1885	17
1890	43
1895	61
1900	102
1905	200

其餘的數碼,若是統計的不精確,即可無須寫出.如要更精確一些,亦可把全數寫出,他人或可據此數爲較爲精確的組合.教科書中比較之事項,常用略數示與讀者,在平常之論說著述中,亦常用略數以醒目.

42. 可能的精確

決定我們所要的精確標準,固然很容易.要使每項數目,都達到這定的精確之度,就不可能了.一個工程師測量地形,或者他要知道角度一秒十分之一,而他的測量器僅能量到秒爲止.工資調查者或者要知道每個工人每年收入的幾元錢,而零工如能得到十元或二十元,已不是易事了.地質學家或者要知道上次冰期到現在共有多少日數,其實如能知道共有多少萬年,也即算滿意了.立法機關或者要知道全國修道費的確數目,但是以報告的不完全和報告的淩亂不清,他們以爲最切近之數目,不知要與真正數目相差幾萬元.這樣看來,精確的標準,不是由統計家隨意定的,乃是由於外界情境定的,統計家要知道他的報告僅能表明實在能得的精確,並不能如他所期望那樣精確.精確的程度,也不能如他最正確張

本那樣精確.

43. 書寫或讀出數碼的精確

在表上讀出數碼的精確程度,應當說明在縱格上的題詞上,或用腳註說明之.若數碼出了我們預定精確範圍之外,除了第一數碼,所有不精確的數碼,都可去掉不要.

常有數碼估量近於正確數,仍以多留一數碼爲是.因爲若多去掉一個數碼,便是增加一點錯誤.例如樹葉之長記載是 2·96 生的米突,雖然實際上僅能讀到生的米突的十分之一,但最末數碼6仍以留存爲是.因爲2·96這個數,總要比3·0更近於真正葉長.若是最末數碼是0,也要把這 0 像別的數碼一樣塡上.譬如一個樹葉之長爲七生的米突,而讀出之精確能至米厘,於是寫時該寫「7·0生的米突」,不應只寫「7·生的米突」,卽算了事.後邊這數目是表明精確僅至生的米突的意思,換一句說,7生的米突,是包括6·5至7·5生的米突的意思.凡是6·5至7·5之數,都可歸入這7組內.若是精確之度至於糎,則 7·0組內僅包括6·95至7·05各數目,若精確之程至於糎的

百分之一，其能歸入7·00內者，僅爲6·995至7·005各數而已.

無論何時若必須去掉幾個數碼以致整齊一律，或因爲數碼的位數出了我們所預數位以外了，(如我們只要三位小數現在有五位是——譯者) 去的時候，必須要看看餘下的數碼是否正確. 例如下列各數，都改爲一位小數之數，結果如後.

原數	改爲一位小數後之數
27.25001	27.3
27.249987	27.2
18.20995	18.2
18.9478	18.9
18.95172	19.0
19.09162	19.1
24.05002	24.1
23.04997	23.0

凡是過於5的小數，便算爲一個整數進上一個1.不到5的小數，把不算了.若一個小數正是爲·5，進爲整數

也可以，去之不算也可以．去取之權操之作者．(按此即是我們所謂四捨五入法)

44. 各算法結果可能的精確

僅用一種運算方法得出一些假精確的數目，這是極大一個錯誤，例如張，王，李三個兒童．張為七歲，王為九歲，李為六歲，求三兒之平均年齡，常有人算成下式者．

$$7+6+9=22$$

$$22\div3=7.333333333\cdots\text{歲}$$

這種小數要寫多少位都有，把一張紙都寫滿了也寫不完．平常以為此數一定正確的，如稍加思索，則知其誤了．張兒僅為七歲，若這七歲是對的，依以前四捨五入的方法論之，張兒或為七歲不足五月，或為七歲零五月二十九日．這便是精確的數目．王李二童之年齡，也可依此類推．我們平常算年齡，以年計算便為精確，若把一年分成為千萬分小分，那就毫無意義了．上邊的累贅不盡的小數，是決不能增加精確的．所以無論乘，除，開方，解指數時，如結果若是小數或帶有小數，我們要注意這假精確．下邊的討論，或可為決定算法結果之精確的一助．

乘法的精確

設 m = 乘數

n = 被乘數

x = 乘數的可能錯誤

y = 被乘數的可能錯誤

於是 $(m+x)(n+y)=mn+my+nx+xy,$

$(m-x)(n-y)=mn-my-nx+xy.$

二式合在一起其積爲

$$mn+xy\pm(my+nx).$$

但 xy 與 $my+nx$ 相比爲值太小,故 xy 可略而不書,於是相乘之積僅爲

$$mn\pm(my+nx).$$

$my+nx$ 之值是能算出的,於是積之精確與積之數位都可決定了.

試以實數例明之

$$726\times10{,}200$$

726 的可能錯誤是 $0.5=x,$

10,200 的可能錯誤爲 $50=y,$

$$m=726$$

$$n=10{,}200.$$

將積代入 $mn+xy\pm(my+nx)$ 式中,得

$$7{,}405{,}200+25\pm(36{,}300+5{,}100)=7{,}405{,}225\pm41{,}400.$$

如以 xy 之值過小棄而不算,便得

$$7{,}405{,}200\pm41{,}400.$$

從 + 號爲 7,446,600,

從 − 號爲 7,363,800,

所以最大精確爲 7,400,000.

除法的精確

設　$a=$ 被除數

$d=$ 除數

$x=$ 被除數的可能錯誤

$y=$ 除數的可能錯誤

商數必在 $\frac{a+x}{d-y}$ 與 $\frac{a-x}{d+y}$ 之中間.

但 $\frac{a+x}{d-y}-\frac{a-x}{d+y}=\frac{(a+x)(d+y)-(a-x)(d-y)}{d^2-y^2}=\frac{2dx+2ay}{d^2-y^2}$,

可能錯誤等於$\frac{2dx+2ay}{d^2-y^2}$之一半即爲$\frac{dx+ay}{d^2-y^2}$,

∴商數差不多等於

$$\frac{a}{d}\pm\frac{dx+ay}{d^2-y^2},$$

可能錯誤既然定出,則絕對精確數目即可決定了。試以實數例之:

$$1,440\div.012$$

$$a=1,440$$

$$d=\quad .012$$

$$x=\quad 5$$

$$y=\quad .0005$$

商數爲

$$\frac{a}{d}\pm\frac{dx+ay}{d^2-y^2}=\frac{1,440}{.012}\pm\frac{.012\times5+1,440\times.0005}{.000144=.000,00025}$$

$$=120,000\pm\frac{.78}{.00014375}=120,000\pm5,426^+$$

$$=125,426^+\text{ 或 }114,574^-.$$

最精確之結果爲十萬,但以可能錯誤爲 5,426⁺ 或然錯誤定然很小,故此數之精確,亦具有到萬位之可能性。

方根的精確

設　　　　$n=$ 求方根之數

　　　　　$e=$ 此數的可能錯誤

真正方根必在 $\sqrt{n+e}$ 與 $\sqrt{n-e}$ 之中間,而可能錯誤必逼近 $\sqrt{n}-\sqrt{n-e}$.

以錯誤開方後大減,故此數僅爲 $\sqrt{e}$ 之分數.

例

$\sqrt{14,400}$ 之値必在 $\sqrt{14,450}$ 與 $\sqrt{14,350}$ 之間,可能錯誤爲 $\sqrt{14,400}-\sqrt{14,350}$.

卽等於 $120.0-119.79+=0.21-$.

但原數之可能錯誤爲 50,而 50 之方根爲 7^{+}.

所以方根的可能錯誤,較原數之可能錯誤之開方爲小.

平方的精確

設　　　　求 n 的平方

　　　　　e 爲可能錯誤

於是真正平方,必在 $(n+e)^2$ 與 $(n-e)^2$ 之間.

但　　　　$(n+e)^2=n^2+e^2+2ne$,

$$(n-e)^2=n^2+e^2-2ne,$$

卽 $$n^2+e^2\pm 2ne.$$

e^2之值太小,可以不算,結果僅爲

$$n^2\pm 2ne.$$

於是正確數碼之位數卽可得矣.

例

$$(1{,}200)^2=(1{,}200)^2\pm 2\times 1{,}200\times 50=1{,}440{,}000\pm 120{,}000,$$

結果等於 1,560,000 或 1,320,000.

精確之位數僅至百萬.

以上各例中,要注意可能錯誤,不一定與或然錯誤相合,具最大可能錯誤之機遇,比較的少如43節所述商積方根等,如需絕對的精確,則最末之數碼,總以加上爲是,蓋多加一個數碼,卽是減少或然之錯誤.

45. 補償錯誤與重積錯誤(Compensating vs. Cumulative Error)

一個統計的結果是不是精確,要依照他所含的錯誤是補償的呢,或是重積的而定.許多人估量一條直線,一定有一些人所估量的長比真正線長,還有一些人估量的

比真線短,長短互相補償,結果必逼近線之真長.再如兩個人量一條鐵索,一個人量得太緊,一個人量得太鬆,結果的錯誤便互相補償了.一千人估量今年的收穫與前一年收穫之比較,雖個人所估之數都不正確,但是若把這一千人的估量合起來,則去真正的結果不遠了.這就是統計學上之常態定律的一具體之應用.

然而如上例鐵索之銷環太短,那麼這鐵索愈長錯誤愈大.再如一些歲出報告表是錯的,若項目愈多,錯誤愈大,吾人卽無術使之補償平衡了.一般婦女都歡喜把年齡報小了,常把年齡報得很小.二十四歲的婦人她說是十九歲,若幾萬的婦女都是這樣,結果錯誤愈大,卽無術校正了.校正重積錯誤的理論,可以一故事說明之.有某小飯館掌櫃的說,他每賣一頓飯都損失一點,但是後來把損失都補上了,因爲他的顧客太多了.總一句說,若項數很多的補償錯誤而其錯誤又比較的小,這種錯誤可以不必過問.反過來說,若是重積錯誤,這錯誤對於總和,平均均有極大的影響.

46. 總和的精確

重索之力基於弱環，各分項若不正確，也不能得出正確的總和來.例如一百個鐵路公司報告他們的支出，精確之度皆至於分.有一家鐵路公司的報告，其精確僅至於千元.那麼這些報告的總和精確，僅能說至千元爲止，再不能比千元更精確了.假設這個公司的支出報告爲 2,673,000 元，我們僅知道這個數是在 2,672,500 與 2,673,500 二數之中間，只要是在這二數中間的數，無論什麼數，都可謂之正確數.設再有一個鐵路公司他的報告是 16,295,472.16 元，我們決不能把這個數加在 2,673,000 上，於是說這個總合的精確是近於元角分.因爲 2,673,000 的可能錯誤是 500，無論如何，這總合要錯 500 的，其正確之狀當如下.

$16,295,472.16	
2,673,000	逼近的
$18,968,000	正確逼近數

然而數目相加時，也不可矯枉過正，把出乎精確範外之數碼全行去掉，如下列各數之和，遽然看來，似是 46，其實並不是 46，乃是 48.

6.321
2.4926
21.4632
8.
7.3875
2.246
48.0903
48………正確逼近數

上邊的例，如小數位全去掉不算，總和便錯2，所以各項中的正確的數碼，都要保留，總和中之不正確之數始可去掉.

47. 平均數的精確

已知總和精確之度，決不能超於成此總合之各項的精確之度，現在再講算術平均數的絕對精確.設 $m_1m_2m_3\cdots$ 爲估量的各數，n 爲項數，$e_1\ e_2\ e_3\cdots\cdots$ 爲各項之錯誤.於是估計平均數爲 $\Sigma\, m/n$，最大可能的平均數爲

$$\frac{(m_1+e_1)+(m_2+e_2)+(m_3+e_3)\cdots\cdots\cdots\cdots+(m_n+e_n)}{n},$$

或爲 $\frac{\Sigma m+\Sigma e}{n}=\frac{\Sigma m}{n}+\frac{\Sigma e}{n}$.

所以平均數可以正確的寫爲$\Sigma m/n \pm \Sigma e/n$, 但$\Sigma e/n$爲一項的平均可能錯誤，所以算術平均的可能錯誤，等於組中各項的平均可能錯誤.

求這個可能錯誤時，我們假設錯誤都是同方向的，實際上若錯誤是補償的錯誤，而項數又很多，則結果的錯誤便互相補償了. 所以算術平均數的或然錯誤，僅爲可能錯誤之一小分. 設E爲算術平均數的可能錯誤，則此平均數之或然錯誤僅爲$E/\sqrt{n}$.

各項數目平均的或然錯誤比各項的或然錯誤小,這種原則對於科學上有極大的貢獻.物理學家常把同一個現象觀察許多次,將各次的結果相加再求出一個平均數來，測量家常把一件東西測量許多次，然後得出一個精確的角度. 社會學家研究一個現象常到各地考察，以便使某地的特別情形以他地的情形補償之,個人的偏見亦可用平均方法減消之.

然亦不可矯枉過正，謂平均數可以免除一切錯誤. 偏見的錯誤是不能互相校正的,其錯誤並不能用平均數的方法免除之. 即非偏見的錯誤有時錯誤太大，亦不能

用平均減免之，若項數極少錯誤的影響更大。

48.　決定小數點的位置

用算尺 (slide rule) 算出之數，僅是一些數碼，初學者對於決定小數點之位置，常感困難，宜熟自練習，俾一望即可決定小數點的位置何在.下邊決定小數點的經驗方法，可為學者一助.

1.　以乘數的第一個有效數碼為整數位，其餘的數碼為小數，被乘數亦如此.

2.　用心算算出此二數之積，視此積含一位整數或兩位整數.

3.　把這整數加在從前乘的時候沒算的整數上去，就得出共有的整數位的數目了.若被乘數乘數有一個是小數的數，或二者全是小數的時候，就把這小數位由那整數減下.減得的結果，就是積上的有效數碼之前的零(○) 的位數.(如最後得數為 3 就是有效數碼之前的零位該是3——譯者)結果若是正號的數，就是表明積中應有的整數位數.若是負號的數，就是表明積中有效數碼之前的零的位數.

例一

$$42{,}000\times .025 \text{ 用算尺得數 } 105$$

$$4.2\times 2.5=10+ \text{ 或是兩個數碼.}$$

被乘數中還有四位整數,乘數有一個小數點和一個零.

$$2+4-2=4\cdots\cdots \text{積中應有的整數位,}$$

所以其積爲 1050.

例二

$$.036\times .024 \text{ 用算尺得數 } 864,$$

$$3.6\times 2.4=8+ \text{ 或是一個數碼.}$$

乘數被乘數,各有兩個小數點和兩位零,即是四個數碼.

$$1-4=-3\cdots\cdots \text{積中有效數碼之前的零位.}$$

所以其積爲 .000864.

除法

用算尺作除法和乘法一樣,他的小數位平常一看就可以知道.但是他的小數位不能一看便知的時候,下邊所述的方法,未始不可爲學者之一助.

自小數點向右的數我們算他是負數,自小數向左的數我們算他是正數.將被除數中自小數點向第一個有效數碼的位數記下來,把除數中自小數點到第一個有效數碼也記下來.得出來的數由被除數所得的數依代數減法減去自除數所得的數.得出的結果,再和由表中相當格中的數目相加,就是所求的數了.

設　　D=被除數中第一有效數碼

　　　d=除數中第一有效數碼

除法決定小數點的方法

除數及被除數之性質		被除數大於或等於除數		被除數小於除數	
		$D=d$ $D>d$	$D<d$	$D=d$ $D>d$	$D<d$
被除數爲1或大於1	除數爲1或大於1	+1	0	0	−1
	除數爲小數	0	−1		
被除數爲小數	除數爲1或大於1			0	−1
	除數爲小數	+1	0	0	−1

第　二　表

例一

$$.002 \div .04$$

這個數我們得出被除數的第一有效數碼的位數是 -3, 除數的第一有效數碼的位數是 -2,

$$(-3)-(-2)=-1$$

被除數是小數,除數也是小數,而被除數小於除數, $D<d$, 由表中得出 (-1),再把這個 (-1) 加上

$$(-1)+(-1)=-2,$$

所以其商為 .05.

例二

$$1{,}000 \div .025$$

算得除數被除中之第一有效數碼的位數為 $4-(-2)=6$, 由表中得數 (-1),

$$6+(-1)=5$$

所以其商為 40,000.

平方

決定平方小數位的方法和乘法一樣.

平方根

把要算的數自小數點分成幅(periods)，每兩位爲一幅. 再從小數點向第一有效數碼計數(count)幅數，於是就得出方根的第一有效數碼的位數了.

例一

$$\sqrt{36'00}$$

向第一有效數碼計數幅數得 +2, 故其根爲60.

例二

$$\sqrt{00'00'81}$$

第一有效數碼之數爲 −3故其根爲 .009.

參考書

Bowley, A. L.: Elementary Manual of Statistics, Chaps. II and III, IV.

"Student:" The Probable Error of a Means, Biometrika, Vol. VI, p. 1, 1908.

Bowley, A. L.: Elements of Statistics, Chap. VIII.

第三編 分析材料

(Analysis of the Material Collected)

第九章 製表(Tabulation)

49. 通論(General Rules)

遽然一想，好似製表一事是世界上最簡單的事．但是若作一個複雜數目的表，並不簡單．製表在統計學中佔極重要的部分，如一旦將一科學的表製成，分析工作已作了大半了．故從前的統計家視製表爲最後的一步．

製表第一個問題是把數碼都放在一張表內呢，或是放在幾張表內呢？放在一張表內，固然有賅簡之便，張本亦易於接近，然而張幅太大，看的時候易致混淆，行列紛紜也不易一看便可瞭然．如使其清楚不亂可用粗細線，或大小地位分別之，然仍不如把這張大表分爲幾部分爲尤妙．

每一張表應自爲單位，把許多不同的事項都放在一

張表上,極不相宜.蓋各事有各事的相宜的表;目的不同,表自然也不同.若將一些不相關的事項都湊在一個表內,便混淆不清,不倫不類了.例如將工資統計與失業統計放在一個表上,便不相宜.因爲一是說明失業之性質,一爲說明工人賺錢能力,二者之性質截然不同,豈可混爲一談.

製表第二個問題是表上用絕對的數碼 (absolute figures) 呢,或僅用百分數便够了呢?或二者兼用呢?解答這個問題,宜先看情形再定.假如若比較世界各國麥的產量,有絕對的數碼便行了,無須再作成百分數.假如要比較城市與鄉村的瘋狂人數,僅從絕對數碼着想,是毫無意義的,如作成百分數便有意義了.在鄉村中每百人中有多少人是瘋的,在城市中每百人有多少人是瘋的,這自然可以相比了.原始的考察或參考的考察,絕對數碼與百分數宜兼具,在平常的表上絕對數碼,百分數,平均數,都是必要的項目,三者缺一不可.

製表第三個問題是行列標目的數目,須用多少標目爲最適宜呢?一方面標目分得愈多,精確之度愈大.在他

一方面,標目繁多,易擾亂正當的注意,統計上所示之趨向亦忽略了.一方面說標目愈多愈妙;他方面說愈少愈妙.究竟要分多少標目呢?解答這個問題,須視各張本的性質與情境.如能因勢制宜,便可謂盡統計家之能事了.平常的表,總是有幾個主要標目,每一主要標目之下,有幾個分標目.同時用許多平列的標目殊非得計,宜將總和百分等數填入總標目之下,以示主要之意.再將詳細說明填入分目中.若是此表張幅太大,各總和平均數相距太遠不易比較,最好把總和平均數等寫在另一個表上,作爲總表,則一般的結果,便可一目瞭然了.概略表中卽用此法,將結果用略數表之.因爲 42 節已經述過數碼過多,心意易於混亂,反致主要之意義失卻了.故表上寫"支出以萬爲單位"或"產量以百萬斛爲單位"便够了.這樣一來,可省卻許多數碼,然在原始表中,萬不可如此省略,因爲原始表是作爲參考用的,如太省略,則無從參考了.

50. 表的題目(The Title of the Table)

無論總表的題目或分表的題目,都要顯明易解.看

表的人,決不願費許多力,去尋查腳註說明,纔可懂得題目的意義.題目以明瞭爲上,更要正合於表內的張本.平常一個錯誤,是題目的範圍太狹,不足以包括表內的張本.這種錯誤,我們自要極力避免.此外更有一錯誤,是題目的字意不定,語涉雙關,有兩種意思.例如"各職業百分數以國籍分,"這句話可解爲某一職業內按國籍分之工人百分數,也可解爲各國某職業中之工人百分數.

縱格的標目,應書明所用的單位,如"英寸(高)""元"等是.

題目前之字碼,以羅馬字碼爲宜,阿拉伯字碼,不宜用於題目上.各個標目之地位大小,要於所表輕重相稱,平列的標目之地位大小,宜相似相同.若標目很複雜時,不可用打字機打,若有許多平列的標目的時候,須用有法則的方法排列之.例如各省可依地位順序排列之,依面積大小排列之,依人民之多寡排列之,或以其他合於論理的方法排列之.

51. 表的形狀(Form of the Table)

未作表之前，先要作一個完全的草稿，然後再劃格填數.先有草稿,纔可決定用紙的大小，算出縱格的寬度與排列標目的順序，百分數，平均數，總和的地位須留出來.相比的各項數目要放在一起，若能放在一縱格內，較同列在一橫格內爲尤佳.相比的縱格愈相近愈好.總和，百分數，平均數，宜相鄰.蓋一般讀者所常注意者，只是總和百分數平均數.此三數宜置於表頂上或左邊格中，以便易於察覺.習慣上是把這些數放在表之末尾，美國戶口調查部施用新法，把總和平均數放在表頂上.此大可注意的一個方法.而不重要的總和，只用以爲校正者，卽可置於表末.

表內的線，也是表明各支目的輕重的.重要的組，應用粗線或雙線.線的寬度，隨支目的級階的而下降，重要愈減，則線愈細.無論表之大小數碼，在表上應成橫條，每條有五行至八行之寬.各條之間有一狹的空白，於是尋一數目便非常容易了.

各種不同的張本，既難於同歸在一表內，故奇異無可歸之張本，卽置於“雜項”(或稱其他)格內.此法既省

地位，又免得表上有許多空白的地方.表上有許多空白處是不好的,因爲空白太多,看的時候容易混淆.例外之事,可用(米)星號誌於其上,於表下說明之.然而例外事項過多,或雜項格內太繁,表之價值常爲之減色,不可不注意.

52. 製表的精確

表中每項都要精確纔好，苟有一二錯誤發現，卽不免疑及工作之全體了.若使他精確，自然須有一種校正的方法.第一要審察每項的來源是否正確,若來源正確,餘事都爲機械的工作了.校正總和的方法，常可把縱格各數相加,再橫線各數相加,視其結果是否相等.如兩數相等,卽可斷定加法無誤,否則錯了.不能施行此法的時候,只好重加一次,以資校正.

校正百分數的方法較爲簡單,將各數相加視其和是否等於百分之百.如等於百分之百卽算正確了.校正平均數的方法，是用項數乘平均數，視其積是否等於原有之和.如等於原有之和,便對了.

乘除可作兩次，或二人同算，如此則錯誤便可減至

最小度了.

53. 分析結果

表上的結果,很不易一見卽能盡知其詳,最好表外再附一分析說明書,指出主要的結論,可能的錯誤,及其或然的原因.於是製表的價值益形顯著了.統計家必有的特質是,分析表的能力,解釋結果的正確,說明結論的簡賅淸析,三者缺一,便不足言統計家的能事.

參考書

Bowley, A. L.: Elements of Statistics, Chaps. IV and VI.

Bowley, A. L.: Elementary Manual of Statistics, Chap. VI.

Bertillon, Jacques: Cours Élémentaire de Statistique, Chap. I, Annexe I.

第十章 簡單圖(Simple Diagrams)

54. 圖之用途

用數碼比較兩件事是不易了解的,卽使了解了,也不能記住.若把一些數碼,念給一般聽衆,那更是毫無意

義.統計學中一個重要的題目,便是把複雜的數碼作成顯明的表現,使人一望便知,一望便懂,因此統計學中發明了許多方法,以說明表的意義.最普通的法子便是簡單圖.本章卽說明幾個簡單的製圖方法.

55. 統計地圖(Cartograms)

與地理上的地位有關涉的現象,最好用統計地圖.統計地圖有多種,各具各長,學者宜視情地之宜採用之.如只有一張地圖,而顏色印刷費用又不甚昂貴,最好把他印成各種顏色,或敷上各種陰影.美國戶口調查部發行的美國統計地圖(Statistical Atlas of the United States),大可爲我們借鏡.其中引人注意的圖,與易於解說的圖不在少數,頗足以效法.

如顏色印刷費用太大,可用各種橫線及斜線以表明密度的不同.報紙上常有的降雨地圖(rainfall maps),卽是此法的好例.

更有一種圖是點圖.點圖有時特別應用.例如一個人要表明中國麥的產量,他卽用一個點代表十萬石,按各省的產量的多寡定點的多少,產麥最多之地其點最密

於是一望即知各省產麥的比較如何了.美國維斯康新大學教授泰樂爾(Taylor)先生,曾用點圖的方法,研究農業經濟.

56 形圖(Pictograms)

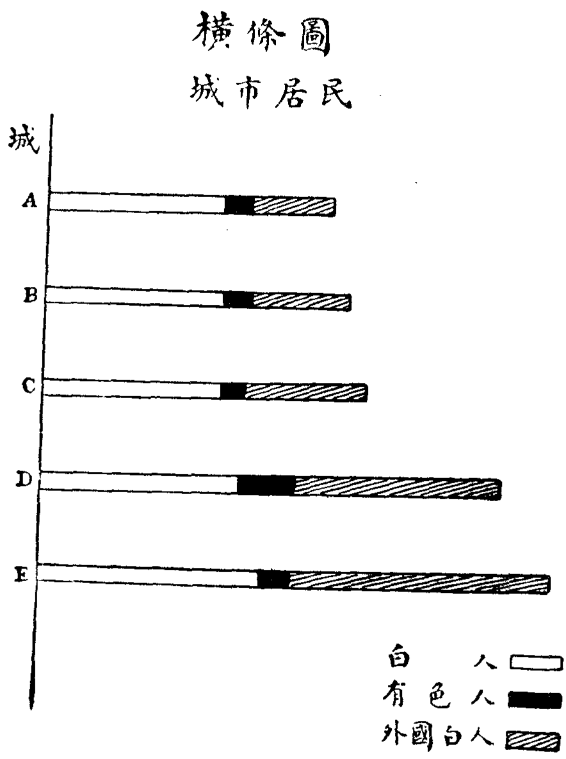
一個作麪包的人,他打算作一個廣告,表明他的麪包比別人的好,他的麪包比別人的重,他的麪包一塊重十六兩,別人的麪包僅重十二兩.他若僅用字句說明,這廣告的效力定然很小.若用下邊的圖,便立時可有反應,

比較圖

第一圖

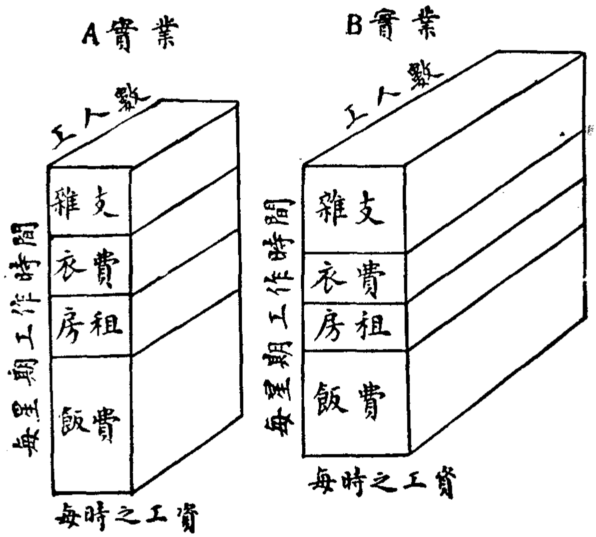
這個圖就是說明圖的重要. 統計學中的圖有多種,最普通最簡單者爲橫條圖 (for diagram),數目的大小,只用條線的長短表之. 但爲表明數目的成分起見,也常

於條線加以陰影以示區別.如第二圖卽是.此法之優點,在於將所有的東西都放在一圖上.惟各組難於比較,也是他的大缺點.除第一組外,其餘各組皆不在一直線上,故甚難比較,如將各組分開另成一圖,就易於比較了.



第二圖

說明複雜的算學關係僅用橫條圖是不足的,如遇此種情形的時候,最好畫兩三個體積圖.第三圖卽是比較每點鐘的工資,比較每日工作時間,比較工人消費的分配,比較A,B,兩實業中的工人數目.須注意者,比較的項數愈加,精確之度卽愈減,第三圖中工人數目的比較,每點鐘的工資比較,每星期工作時間的比較,是容易比較的.而收入支出的數目與百分數,卽不易比了.因爲很多



第 三 圖

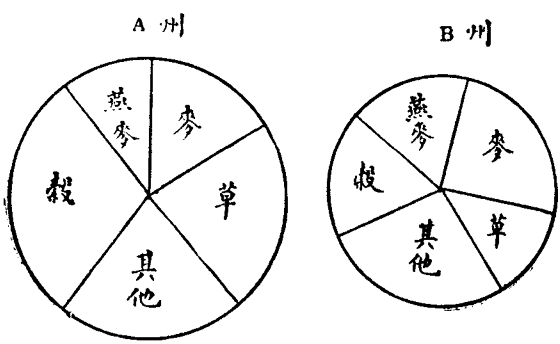
的矩形,平行四邊形,極不易分辨那個大於那個了.

再有一種形圖,卽是於圓形內分爲許多扇形,如第四圖.將各扇形加以陰影或顏色,以表明各個間之關係.第四圖的圓是代表全國已墾植的田,圓中各扇形代表各穀物的比較收穫量.

比較各體積,平常亦用立方或圓球以表之,也有卽以所要比較的物件的實形,畫成圖案以比較者.如常有畫軍艦以比較各國的海軍,畫商船以比較各國海外貿易,

圓圖

各種穀物所佔地之面積



第　四　圖

或於船上畫一些貨包以表明棉花的產量. 須注意者, 面積的圖以平方計算, 其所代表之物宜與平方相合. 體積圖以立方計算, 其所代表之物宜與立方相合. 若不注意及此,結果就錯了.

參考書

The Statistical Atlas of the United States (Census of 1900.)

Bailey, Wm. Bacon: Modern Social Conditions, Century Co., N. Y., 1906, pp. 54–56.

Bowley, A. L.: Elementary Manual of Statistics, Chap. V.

Bertillon, Jacques: Cours Élémentaire de Statistique, Chap. XI.

Block, Maurice: Traité de Statistique, Chap. XIII.

第十一章 次數表與次數圖 (Frequency Tables and Graphs)

57. 次數表的用途

自然現象中萬象雜呈, 極不一致, 各項具有各項的

特質.某一棵榆樹是高的,別一棵卽或是矮的;某人是富的,別一個人便是貧的;某一個城是大的,別一個城或者是小的.情態不同的事物,叫做個體 (variables).以上所舉的人,樹,城市,各自不同.卽同類之物,仍各不同.更有進者,同一事物,因時間的變遷,其特質亦隨之而變遷不同.同是這一棵樹,因爲時間變遷,長高了.同是這一個人,過幾年便老了.一個小城過些年會變大了.人,城,樹,仍是從前的人,城,樹.其前後之特質,便不可同日而語了.這種因時間變遷而起變化的事物,叫做歷史變遷 (historicalvariation).容第十五章中再詳加討論.

研究同類的事物,爲便利起見,可把各項分成組.(classes)最簡單的分組方法,是把所有的東西分成兩組.具某種特質的爲一組,不具這種特質的爲一組.如人可分爲瘋人與非瘋人兩組.工人可分爲在職的工人與失業的工人兩組.花可分爲白的與有色的.人可分爲高人矮人.這種兩組分法 (division by dichotomy),有時很適用,結果亦很滿意.然有時不易分清界限.如分人的高矮幾

尺高爲高,幾尺高爲矮,這是極不易定的.只好立一個强訂的界限,以分高矮.設强訂爲五尺,過五尺者爲高人,不及五尺爲矮人.然而此種强訂分法,與其只分兩組,不如多分幾組爲更好.實際的分類法,是把全體分爲寬度相同的組,組的界限是謂組限(class-limits).兩組限之距離叫做組距(class-interval).如樹林中最高的樹爲39尺,最矮的樹爲16尺,卽可把全體分成五組,各組的界限用略數寫爲15,20,25,30,35,40.這例中組距爲5尺,把全體分爲和諧的組,將應塡入某組的項數卽塡入該組,於是作成一表,此表叫做次數表.每組中的項數組成組形(size of class),或稱爲組之次數.

譬如某處有一百個人共有財產一百萬元,每人平均有一萬元.是這一百人的財產都相等呢?或是只有一人有990,100元,其餘九十九人各有一百元呢?研究此等財產分配問題,卽可用次數表.次數的作法,依財產的多寡將人分爲各組,然後再比較其次數分配.設有一個社會與前一個社會相彷,有一百個人,若是他們的財產分配很勻稱,大約可有下表的結果.

財產分配簡易次數表
（小組距）

每人之財產 m	次數或各組之人數 f
\$ 0 \$1,000	5
1,000－ 2,000	8
2,000－ 3,000	10
3,000－ 4,000	12
4,000－ 5,000	14
5,000－ 6,000	10
6,000－ 7,000	9
7,000－ 8,000	10
8,000－ 9,000	6
9,000－10,000	2
10,000－11,000	3
11,000－12,000	1
12,000－13,000	2
13,000－14,000	0
14,000－15,000	2
15,000－16,000	1
16,000－17,000	1
17,000－18,000	1
18,000 以上	3
	n＝100

第　三　表

這卽是說這社會中有千元以下的人有五人，一千元

以上二千元以下的人有八人,二千元以上三千元以下的人有十人,照此類推,至一萬八千以上者三人.

此表內各組的人數(卽次數)漸次增加,直抵最大組$4,000－$5,000(最大組名曰範數 mode)後各組的人數又減少了.其減少之狀很不規則.假如把組距增大了,看他有什麽影響.

財產分配簡易次數表

(大組距)

每人的財產 m	各組的人數 f
$ 0－$ 3,000	23
3 000－ 6,000	36
6 000－ 9,000	25
9,000－ 12,000	6
12,000－ 15,000	4
15,000－ 18,000	3
18,000 以上	3
	n＝100

第 四 表

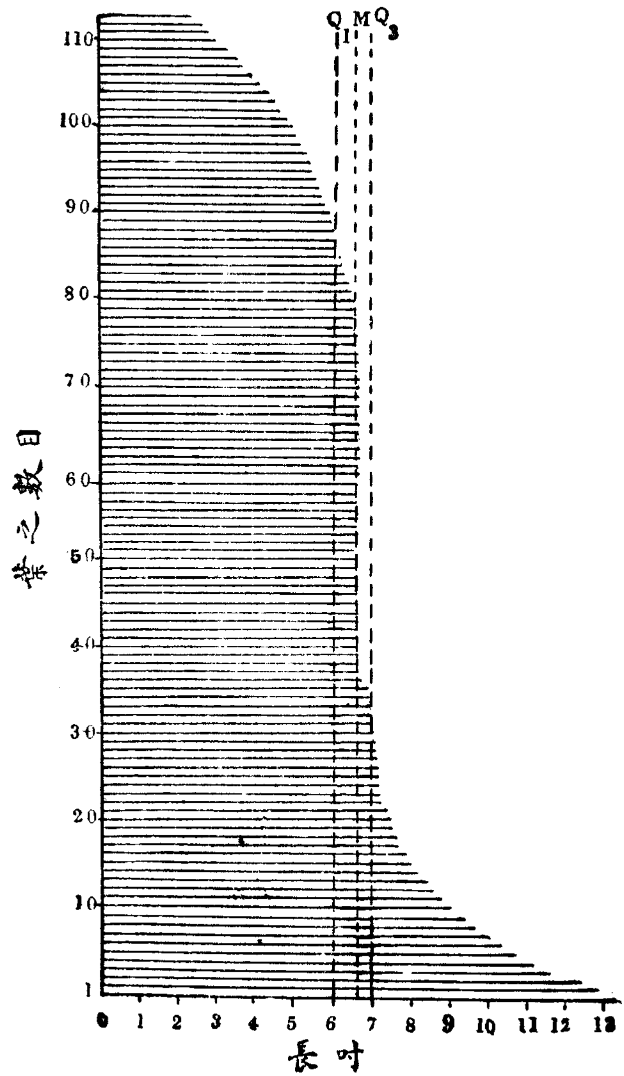
這表中一件極顯然的現象卽是組距增大,則第二格內數目起落的常態 (regularity) 也增加了.就是說把兩限的距離從$1,000增到$3,000,第二格內數目的漲落便有規則了.此種有規則的現象就叫做「常態」.但是欲得

此種常態，對於表中的詳細的表現，不得不為一度之犧牲.此即綱要的整齊,與詳細的精確之老爭論了.綱要之長點為整齊，惟不精確；詳細之長點為精確，但乏整齊.究竟要分多少組，要依統計家自己的決定.平常的組數多至於不犧牲逼近常態而止.

接着我們便要問上表的常態現象,是我們强訂臆定的一種財產分配呢，或是各種事物都有這種常態呢?稍加考察,便知這起落的常態,各事物都是如此，並不是我們杜撰臆斷的.

從一棵樹上,隨便（絕對的隨便,毫不加意的）採取一百十三個葉子，依葉子的長短順序排列之，若拿每一條線代表每葉之長，則其排列之狀當如第五圖.在這個圖我們看，接近特別長的葉與特別短的葉，其長度之增減很快.自第四十葉至六十葉中間的葉子，其長度差不多是一致的.這就是說假如以一生的米突為組距而分組，三生的至五生的的組中所有的葉數很少;七生的至十三生的組中的葉數也很少;惟六生的至七生的組內葉數極多，一大半葉子都在這六至七的組中.這種組形的起落，

113個樹葉之長的排列圖



第五圖

在自然現象中如此,在經濟現象中也是如此.

有一個實驗,將三個骰子連擲一百九十六次,其結果如下表.

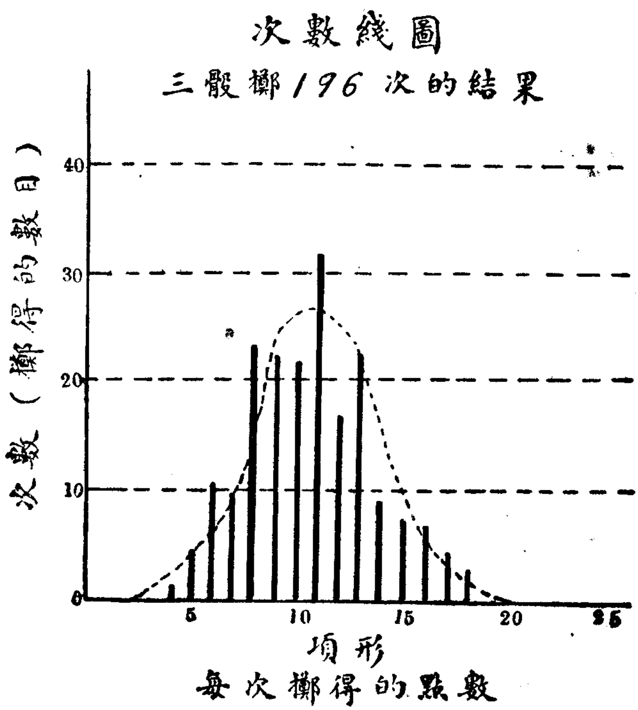
三個骰子擲得之次數表

點數(項形) m	發現的次數 f
4	1
5	4
6	11
7	10
8	24
9	22
10	23
11	32
12	17
13	23
14	9
15	7
16	7
17	4
18	3
	n=196

第　五　表

從上表可看出純粹機遇的事物的次數起落,的確和自然現象的次數起落相同,最多的次數近於十一.

諸如此類的例子很多,從以上幾個例即可得出下列的規律了.自然現象與經濟現象的常模(norm)叫做範數(mode),大多數項數近於範數;離範數漸遠的組,其項數漸少.自然現象中離範數最大的離差,可以約略決定一個界限,出此界限,就不能再有數目了.這個定律是快得雷(Quetelet)發明的,統計學的大部分即根據於此.



第六圖

在第五圖中範數在於第五十六葉上,從這一點劃的直線與向右邊的直線幾垂直了.這直線卽是表示大多數的葉子,於是我們就可以說,此樹的葉子,大多數的長,是6.7 生的米突.

讀者自然要問,若項數增加了,他的範數的位置變不變呢?就是說把一百十三個葉增爲五百個,範數的位置該如何呢?據實驗的結果,若是所選的項數的確可爲標樣,增加項數不過增加各組形的常態而已,項數愈多,則得正確標樣的機會愈多,有正確之標樣,卽可代表全體了.

58. 次數表的分類

大組距小組距之利弊,已要略說過了,還有幾個點應該牢記.組距必須相等,就是說各組的寬度必要一律.所有的組限都要記入相當的格內,否則結果一定錯的.各組之形要用單數碼表之,如 3 cm. 是.若組限是 3 cm.這組包含 2.5 cm. 至 3.5 cm. 所有的各項;若組距大於 1 可作爲 3–7 cm., 8–12 cm. 等等. 3–7 cm. 這一組包含 2.5–7.5 一切的項數.以原始計量中常用 1 爲組距,故前法僅用於標樣制度,若計量的目的專爲次數表之用,各組亦可

寫爲 3–7 cm., 7–11 cm., 11–15 cm., 等等. 現在這組距是偶數6.99, 這一項應在 3–7 cm. 組內.

分組時若能使組之中點爲偶數時, 最好用偶數, 不要用分數. 中點爲偶數時, 乘時及其他一切算學手續全感便利, 運算時我們以全體爲一致. 但是宜知自然現象之各項並不是整數, 實際計量時所用之數, 多少總有一點人造的强訂的界限, 自然的本體, 並無顯然的分界, 吾人所用的分類, 不過爲便利起見强訂一種界限而已.

59. 連續級數與分立級數(Continuous and Discrete Series)

自然界各物的形態與重量, 在某限度內可任在某點上, 此種形態重量, 僅能測出逼近的數目, 並無算學的精確. 這種測量的記載, 叫做連續級數(continuous series). 反過來說, 擲骰子僅能得出整確的點數, 這種項數的記載叫做分立級數(discrete series). 一張工資記載, 是分立級數. 因爲工資的算法, 常是每星期多少元, 很少用小於半元之數爲單位. (按此爲美國工人情形——譯者) 並沒有每星期賺 $14,39 的工人. 總而言之, 以金錢爲單位之

數常爲分立級數,因爲金錢的最小單位也是可分的.

60. 分立級數次數圖

我們已經知道圖可以表明各種統計,使讀者一目瞭然.圖對於次數表也有同一的功用.分立級數最簡單的圖爲線次數圖.此圖最適用於項數很少各項可用線代之事物.57節內第六圖表明擲骰子的次數,正可說明此意,因爲沒有小數的項數,所以劃緩和曲線(smooth curve)時,這緩和曲線不可釋爲是表明整數之間更有次數存在,不過各整數點之一種表現而已.鐘形曲線(bell-shaped curve),是表明擲的次數無限的各項相連,其高卽成一鐘形曲線.此鐘形曲線叫常態次數曲線(normal frequency curve).或稱錯誤的常態曲線(normal curve of error).在那擲三個骰子實驗中每次擲得十個點的共有二十二次,在曲線中二者仍爲同一的項數其常數爲二十六.此實驗中得十八點者有三次,三點者一次沒有.但並不可說三點者不會發現,三點者也可一樣的發現.每一個骰都有一面是一個點的,若是三個骰同是一點向上,便成了三點了.三點向上的機會與十八點向上的機會是均等的,不

過在這個實驗裏六點特別多，一點特別少而已．緩和曲線的目的，在於減少意外的變遷，樹建一般的趨向罷了．

61. 緩和矩形次數圖或直方圖

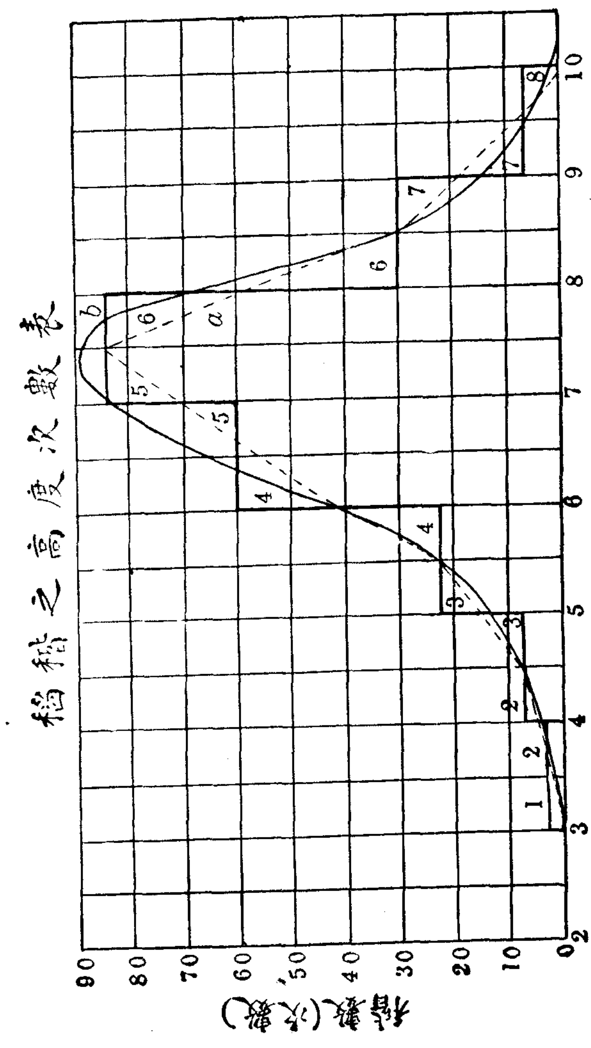
無論連續級數或分立級數最大項之形與最小項之形的中間必有一個大差．各項數目卽在此兩極端之中．57節所述之樹葉測驗，每一項都代表無限相仿的項數，然則自每項自不可僅劃一線代表之便算了事，須將張本分爲强訂界限之各組以每羣爲全體．假如我們到田中去量穀稭，卽知稭之長不出某限度之外，若量一組代表的穀稭，其結果或如下．

穀稭高度次數表

高尺（項形） m	稭之數目（次數） f
3－ 4	3
4－ 5	7
5－ 6	22
6－ 7	60
7－ 8	85
8－ 9	32
9－10	8
	n＝217

第 六 表

此表可用第七圖矩形圖說明之.



第七圖

這些矩形,的確可代表各組的相對的大小,惟以界限是強訂的,故組之大小不同,或排列不同,則結果大異.若量的項數增加,同時組距減小,則階級的形狀卽漸漸減小,最後可成爲一緩和曲線.代表217 個實在量的穀稭矩形,直方圖優於曲線圖,若代表全田中的穀稭,自然以曲線圖爲宜.曲線圖的意思,也就是代表全田中之穀稭.我們所要知道的,也是全田中之穀稭.所以我們的目的,就是設法把矩形圖作曲線圖,以便代表所有的穀稭.

普通逼近曲線的法子,把圖中各項先作成矩形,再從底的兩極端經過各矩形之上底中點,聯成線卽成第七圖之點線 a,此線大約逼近正確,但有一二錯誤.

第一,新畫的面積與舊矩形直方的面積,雖然應當相等,在線外的三角與在線內的三角形應兩兩相對,而舊圖中的第一與第八兩個三角形沒有相對的,所以新圖之面積,較舊圖之面積爲小.

第二,假如中組的組限不是 7 與 8,是 7.4 與 7.6,其線或較現在之圖爲高.

第三,設有一稭,其長出於我們所定範圍之外,卽是

說這穀稭特別矮,不足三尺,或特別的長過於十尺,這種消息,只好尋之於表外了.

將此三點牢記在心,卽可畫緩和曲線了.(或稱緩和直方圖)緩和曲線畫出,卽可代表全田的穀稭.圖如連線b這條曲線與擲骰子鐘形曲線極相似.

實際上直方圖簡直就不畫,僅將次數於各組之中點作成次數多邊形a,這是最容易的方法,惟不易緩和得適當.若爲精確起見,仍用前法爲宜,假如有幾個次數圖相比,不用次數多邊形,卽用緩和直方圖,萬不要用矩形直方圖,因爲矩形直方圖,須用許多許多的矩形,易於混淆,並且各直線常常相倂,更不淸楚.

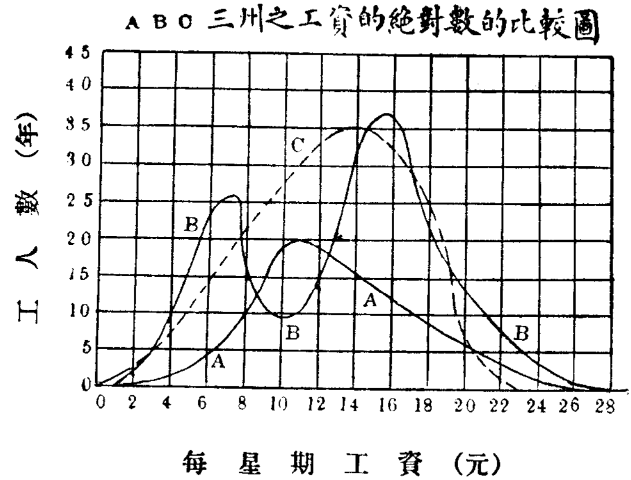
若緩和次數多邊形,要多邊形是由矩形直方圖來的緩和時應照矩形直方圖的緩和方法緩和之.就是說,曲線的頂常出於多邊形最高點之外,組大時更是如此.多邊形之棱角能免除者卽行免除之,曲線的方向能不變更卽不變更.如此,則大多數的棱角皆可免除了,至於緩和應到那種程度,各張本的情境不同,程度亦不同.自然現象或機遇的現象,常逼近一個規則的錯誤曲線,緩和時

卽可自由緩和.若社會現象經濟現象常態曲線上常有棱角者,緩和時只能將小的棱角免去,大的棱角仍然存在.更有注意者,緩和的功用在於顯出一般的趨向 (tendencies);然而趨向顯出,則實際的事實轉爲隱暗了.

緩和曲線常是代表連續級數的.連續級數的項數,可自無窮小起首項數漸次增大,但過某點後,項的數目又漸次減少,直至零點爲止.故畫曲線應底線起至底線止,關於曲線的起止,雖無絕對的規律,平常總是起於極端的項數止於極端項數.

62. 比較的直方圖

直方圖可用之以比較二組或二組以上的張本.第七表與第八圖卽是比較A,B,C,三州的工資和工人數目的.A州中的工人比較的少,大多數每人每星期之工資在八元與十六元之間.B州中工人顯然分爲兩組,或可說B州的婦女兒童之工資甚小,男人的工資很大.在C州得大工資的工人較AB兩州全少,但C州中一般人的工資,平均計之,則較AB兩州高.



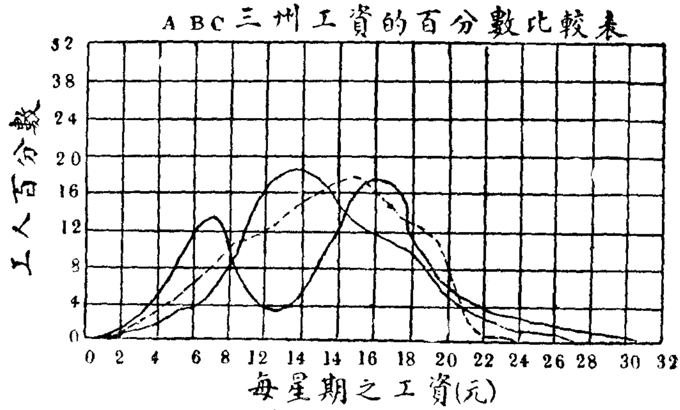
第　八　圖

第八圖中的工資，比較有一部分不甚清楚，因爲一州中的工人比其餘州中工人多的原故.因爲這一州的工人比那一州工人多，曲線的高底便不清楚了.欲矯此弊，可將次數作爲百分數，以各項總和除各項，就得出百分數了.下表就是把次數作爲百分數的表.例如，A 州中的工人總數是 98,000，卽用這個總和除各項，卽得出第三格內的百分數，百分之和自然爲 100.

A.B.C.三州工資之絕對數與百分數次數表

每星期之工資	A州		B州		C州	
	工人數目	工人百分數	工人數目	工人百分數	工人數目	工人百分數
$ 0— 1.99	25	0.0	.210	0.1	1.114	0.5
2.00— 3.99	1.460	1.5	4.630	2.3	4.986	2.3
4.00— 5.99	3.984	3.9	16.424	8.1	10.102	4.8
6.00— 7.99	5.025	5.1	24.898	12.3	17.170	8.1
8.00— 9.99	13.200	13.5	12.122	6.0	22.054	10.4
10.00—11.99	17.420	17.7	8.964	4.4	28.402	13.3
12,00—13.99	16.142	16.5	17.220	8.5	33.960	15.9
14.00—15.99	13.240	13.5	35.116	17.3	34.817	16.3
16.00—17.99	10.940	11.2	34.963	17.2	31.460	14.8
18.00—19.99	7.964	8.1	17.842	8.8	24.972	11.7
20.00—21.99	4.982	5.1	12.240	6.1	3.417	1.6
22.00—23.99	2.786	2.8	7.963	3.9	546	.3
24.00—25.99	962	1.0	6.241	3.0		
26.00—27.99	70	.1	3.196	1.5		
28.00—29.99			971	.5		
總和	98.000	100.0	203.000	100.0	213.000	100.0

第　七　表



第　九　圖

第九圖卽是用百分數作成的圖,此法是把各州不問大小,都視爲同等,研究時就便利了.平常比較兩件事物用百分數直方圖,比用實在數的直方圖爲好,其比較的結果,也比較的滿意.

總起來說,緩和直方圖,爲比較張本的最簡單最好的方法,其比較之應用可施之各方面.

63. 堆積次數表(Cumulative Frequency Tables)

把簡單次數表裏的次數相加,便可作成堆積次數表.其加法先將簡單次數中的第一組與第二組相加,卽作堆積次數的第二組.將此數再與簡單次數表的第三組相加,

卽得第三組堆積次數.如此繼續相加,便得一堆積次數表.假如我們還用穀稭的高度來說明堆積次數表,結果可如下.

穀稭高度堆積次數表

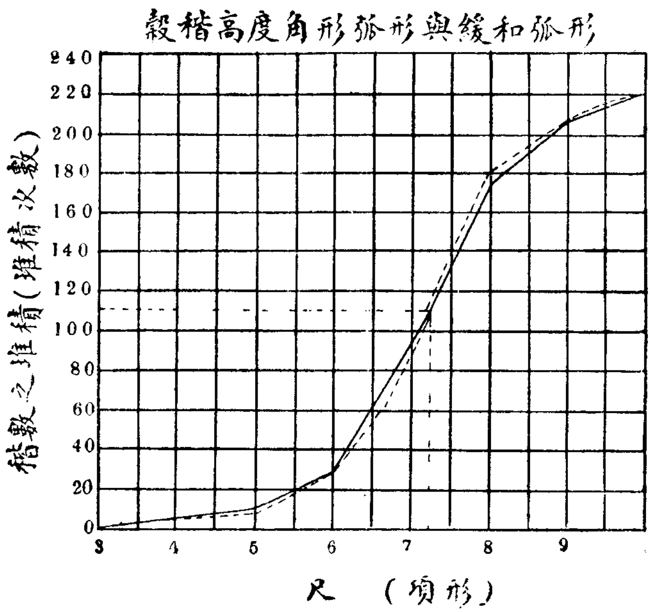
高度尺 m	稭之數目 f	堆積次數
3– 4	3	3
4– 5	7	10
5– 6	22	32
6– 7	60	92
7– 8	85	177
8– 9	32	209
9–10	8	217
	n=217	

第　八　表

64. 弧形(The Ogive)

把上邊的表中最後的格裏的數目,作爲一圖,就是堆積次數圖,或稱爲弧形.弧形的作法與次數多邊形不同.作次數多邊形時用各組的中點連線;作弧形時則用各組的上界限(upper limit).第十圖卽是用上表的數目所作成的弧形,弧形的優點在於一目瞭然,第十圖遽然

一看便十分清晰.這一點就是弧形勝於直方圖的地方.



第　十　圖

弧形可以用以比較事物,此點並不遜於直方圖.將次數作爲百分數其結果更好,此點又與直方圖相同.惟是一般人對於弧形不若對於直方圖之易於了解.弧形的原來用途是決定中數,四分數,百分數等,其意容後章再講.

65. 作圖的通律

1. 軸(axes)用重黑線.

2. 選定的格式(scale)應可容下各項,又合宜紙幅的大小.

3. 排列格式,應當用略數寫在方格紙的重線上,若方格紙是按十計分法(decimal plan);可寫5,10,15….;10,20,30……或寫其他易分的數目萬不可用3,6,9….一類的數目.因爲這些分數不能與格線相對.亦不要格式上的數碼僅與表上的數目相合.

4. 圖應畫在紙的重要部分,其格式的大小,應大至於可容下我們所要的詳細事項,小至於一看便全看得清楚.

5. 圖要便於攜取,也要精確.如能表出我們所要的主要各點,則更大的精確也就不必要了.

參考書

Elderton, W. Palin and Ethel M.: Primer of Statistics, Adam and Chas. Black, London, 1916, Chaps. I and II.

Yule, G. U.: Introduction to Statistics, Chap. VI.

Thorndike, E. L.: Are Introduction to the Theory of Mental and Social Measurements, Science Press, N. Y., 1904, Chaps. III and IV.

Bowley, A. L.: Elements of Statistics, Chap. VII.

Bowley, A. L.: The Measurement of Groups and Series, Chas. and Edwins Layton, London, 1903, 1st Lecture.

Edgeworth, F. Y.: The Law of Error. Ency. Britannica.

第十二章　平均數(Averages)

66. 平均數的用途

1. 平均數第一個用途,是,把一羣很多的數目用簡單數目淸淸楚楚表現出來. 例如樹林中的樹, 若把各個樹的高矮逐一告訴我們, 我們倒不十分淸楚了. 若說這林中的樹的平均是幾丈高, 便易於了解了. 也有一個一定的意義了.

2. 平均數第二個用途,是,用這簡易的平均數目比較各組事物. 這是第一項用途的支系. 無論比較什麼東

西,必須我們心中先有一個一定的清楚觀念,纔可比較.兩個樹林是不能比較的.若用平均數就可比較了.

3. 平均數第三個用途,是,僅以簡易的張本便可得出全體張本的現形來.要知一種民族的身高,並用不着把這民族中的人每個人都量到,只要量幾百人便够了.用這幾百人的身高作爲標樣,求其平均數,此平均數便可代表全族的身高了.

4. 平均數第四個用途,是,用平均數表明各組關係的算學觀念.我們固然可以說,這林中的樹比那林中的樹高,但是我們要知道一個一定的高度比率,必須用平均數.

1 範數(Mode)

67. 範數的定義

平均數中的最有用的一種便是範數.範數的定義,各統計稍異其詞.或謂範數是最常有的項;或謂範數是項數密積的地方;或謂範數爲緩和直方圖中豎線最高的地方.三者都有道理,三者同是一事.惟其說法不同而已我們說平均人,平均收入等等,我們的意思是說大多數

人,大多數的收入,我們說範數的(modal) 工人住宅是五間,對於某教堂的範數的捐款是五分,這意思就是說大多數工人住宅都是五間,大多數的捐款都是五分.也就是說工人五間住宅,人士五分捐款,是很普通的,最常有的,最通行的.

68. 決定範數的方法

在一個正確的緩和直方圖範數的位置可一望而知,只要尋出與豎線最大者相對的項卽是範數.也就是曲線最高的部分相對的項,便是範數.尋出曲線的最高部分,那是極容易的事.但是有時直方圖有好幾個範數,那該如何決定呢?決定每個範數的方法,仍與決定一個範數的方法相同.

假如次數表上只有一個很確定的範數,含此範數的組立時便可決定.若表上有許多的不規則的數目,雖然只有一個範數存在,而含範數的組也不易選擇.若有這種情形的時候,最好用羣集方法 (a process of grouping) 以求範數.其進行手續如下:

第一步將每兩組相加,得出一列新數目.於是將上

限移下一組,(即是從第二組起)再兩組相加,又得一列新數.再三組相加,得出一列新數,再將上限移下一組,再三組相加,再從第三組起三組相加.如仍須加時,可每四組相加,一直尋出常態(regularity),而最大次數的位置,不因上限變更而變更時爲止,範數即在新數中的最大數中.在某限度內,用此方法即可求出範數.下邊的表,即說明此種方法.表中之數,僅是接近於範數的各項,沒有極端數,因爲極端數對於羣集是無用的.

羣集方法求範數

項形 m	次數 f					
5	48	100				
6	52		108	156		
7	56	116			168	
8	60		122			
9	62	122		182		178
10	60		118		180	
11	58	114				
12	56		119	177		174
13	63	123			179	
14	60		108			
15	48	88		148		171
16	40		72		120	
17	32					

第 九 表

第一列的數目,最大數是123,故範數在13或14組內.其餘列中最大數,總在9組內,迨後組限雖變,而最大數常在9組中,故可決定範數逼近於9.

若次數表的各組組距很大,應在組限內決定範數.第七圖矩形直方圖內6-7組大於8-9,故範數的真正位置,距7較距8近.用範數組相鄰組之較量(weights)決定範數的位置,由經驗中得有一種方法,亦殊可用.其方法可用下邊的公式表之.

設　l=組的低限.

c=組距.

f_1=相鄰低組的項數.

f_2=相鄰高組的項數.

Z=範數.

於是

$$Z=l+\frac{f_2c}{f_2+f_1},$$

若將61節內之穀稭之高次數表的數目代入上邊的公式內,可得如下:

$$Z=7+\frac{32\times1}{32+60}$$

$$=7+\frac{32}{92}$$

$$=7.347+\cdots$$ 所求範數

若次數極不規則時,可用範數組 (modal class) 相鄰組兩組或兩組以上爲較量,但一圖上有許多範數時,此法不應用.

範數也可在弧形上約略決定.曲線上幾成直線之點卽爲範數,亦可用算學方法決定之.用尺切於弧上,沿弧向前移動,視其與尺成相反方向時,卽爲範數.惟是平常用弧形決定之範數,常是無價値的.

69. 範數的優點

1. 要免除極端的差別範數是有用的.幾棵特別高的樹,或特別矮的樹,對於全林中的範數樹是毫無影響的.教會捐款中有一張一千元的支票,對於範數捐款也是無影響的;若對於算術平均數就有影響了.

2. 決定範數時,除非項數極少時,可不必知道極端項數.要知道一國範數富力,不必知道百萬富翁的數目與其財產;也無須知道貧無立錐者的數目,卽可得範數

來.

3. 如標樣張本選得合宜,可決定得相當的精確.

4. 平常人想範數是羣數中的最好的代表.我們說某社會中的工人範數工資是每日兩元.大家都容易懂.若說平均工資是$2.17,聽者便不十分清楚.因爲實際上並沒有每日賺$2.17的人.

70. 範數的劣點

1. 有許多時並沒有一個單純的範數,市城並無範數的城.並沒有一個人知道範數城市是什麼意思.而在工資統計中又同時有許多的範數,各級工人有各級的範數,並不是有一個單純的範數.

2. 要表明極端項數時,範數是完全無用的.假如一個人要知道每個人的財富力大小,那們範數的富力還有什麼用呢.

3. 範數不能用算學方法決定,有時用任何方法也很難精確決定.

4. 範數乘項數之積,得不出正確的總和.這是他不如算術平均數之一點.

5. 用一些少數齊一的項數,決定大多數不同的項,這是一個大錯. 社會上的財富本是千變萬化的不同,只因爲一張財富表上有三個人同是 $992,餘人各自不同,所以就有人說這社會的每人範數財富是 $992. 其實並不是如此.(此種困難可用分組方法免除之.)

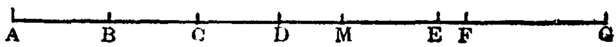
II. 中數(Median)

71. 中數的定義與決定中數的方法

把一些相似的東西,依其大小順次排列之,叫做行列(arrayed).第五圖卽是代表113個樹葉的行列.行列中的正中間的一項叫做中數項.第五圖中的行列中數卽是第五十七項.因爲這一項距兩端同遠,同是距五十六項.中數即是6.7 cm.若全體的項數是偶數,有兩項中項,我們即以爲中數在這兩項之間.假如純任便的取很多的樹葉作一個實驗,(取 113 個葉子)即可知道他們的中數常是不變,於是可知中數也可以代表一羣的數目.

中數的定義,可以說是行列中項形與他項最相近的項.用算學的話來說,設各離差(deviations) 全爲正號,自中數各離差之和爲最小.

自中數的離差



第　十　一　圖

說明

設 AB, AC, AD, AM, AE, AF, AG 爲行列之七項, AM 爲中數.

自 M 的離差之和 = BM + CM + DM + ME + MF + MG

設任取另一項 AE, 自 AE 的離差之和等於

(BM + ME) + (CM + ME) + (DM + ME) + ME + (MF − ME) + (MG − ME)

= BM + CM + DM + MF + MG + 2ME

此和比前和多一個 ME

所以自中數的離差(正號的)之和爲最小.

若第五圖各項,都已行列整齊,中數極易求得,只須從一端數起直數到正中間的一項便是中數了.然而實際上並不是這樣簡單,常是一個次數表上有很多的組,每

組有很多的項,平常求中數的方法,是不易實行的.有一簡單求中數的方法,就是先將各數作成弧形,並緩和之,自豎線一邊的中點向弧作一橫線,令與弧相交,自此交點向底邊作垂線,此線與底邊相交之點卽是所求的中數.如第十圖共有 217 項,橫線應從中點 108.5 項畫起,於是遂求得中數 7.15 尺.

我們常用次數表直接在組中決定中數.此可用插入法(interpolation)行之,設各項形的變移是一致的,卽可用下變的公式.

設 M=中數,

c=含中數組的組距.

l=低組限.

f=組中的項數.

i=自低組限向上數到中數之項數.

於是 $$M=l+\frac{c(2i-1)}{2f}$$

在 63 節內的次數表共有 217 項,中項爲第 109 項.但第 109 項爲 7-8 組內之 85 項的第 17 項,所以

$l=7,$

$c=1,$

$f=85,$

$i=17.$

代入公式中,即得

$$M=7+\frac{1(2\times17=1)}{2\times85}=7.19\text{ 尺}.$$

此數與從圖上所得的中數差不多一樣.

72. 中數的優點與劣點

中數之優點可列舉如下:

1. 決定中數較決定範數更能精確,當範數不能正確決定時,此優點益顯.

2. 中數受極端數的影響極小.從這一點說,中數很像範數不像算術平均數.教會中捐款一張一千元的支票,對於範數是毫無影響,對於中數之影響與別一項一樣,並無特別的影響.

3. 只要知道極端數的項數,即可求得中數各極端數的大小可不過問.例如已知有 \$100,000 以上的人有

多少,最窮的人有多少,於是中數的富力便可求得,無須知極貧與極富者的個人財產的多寡.

4. 範數的位置有時依照很少的幾項,中數決無此弊.

5. 中數可用於不能用一定單位量的東西.例如兒童的心意是不能用一種單位量的,但是很可以把兒童們依其智力大小排列之,中數智力的兒童便可決定了.此時算術平均數是毫無意義的,要比兩組兒童的智力也是無用的,中數則可比較.

中數的劣點可述如下:

1. 與範數同病,中數不能用簡單的算學方法決定,不如算術平均數可用簡單的算法可算得出來.

2. 和範數一樣,中數乘項數之積不與總和相符.

3. 與範數同病,若對於極端數與以較量時,中數是無用的.

4. 中數常爲實際上所沒有的數,此弊正與算術平均數相同.從這點說,中數遠不如範數了.中數的工資有時每日可爲 \$2.37½,實際上每日得此等帶小數的工資

的工人恐怕太少了.

5. 分立級數中,各組的差異很小,大多數的項數都在範數組中,有許多項大小相同都像中數,此時中數就不一定了.此時大於中數之項數與小於中數之項數,可以相差很多.如果相差太多,則中數代表全體之資格就喪失了.

自全體言之,爲實用起見,中數是極有價值的一種平均數.如研究工資問題,研究財富分配問題等,中數實遠優於範數及算術平均數.

III 算術平均數(The Arithmetic Average or Mean)

A簡易算術平均數

73. 算術平均數的定義

把各項合在一起全加起來,叫做總和.以項數除總和,除得之商,叫做算術平均數.

用次數表求平均數,自然先以各項乘次數,然後再相加,若各項的大小知道的不甚精確,可用各組的中點爲各項的代表,結果仍不至於大錯.組距小的時候,用中點更爲合宜.由前邊的算術平均數的定義,我們知道他

有一種特質，就是各項之離差依其正負號相加，其和等於零.可證明如下：

說明

自算術平均數之離差

第 十 二 圖

設 KM_1 KM_2 KM_3 爲一組項數中的一部分項數算術平均數爲 KA 由算術平均數的定義得

$$\frac{KM_1+KM_2+KM_3\cdots\cdots KM_n}{n}=KA$$

代入

$$\frac{(KA-AM_1)+(KA-AM_2)+(KA+AM_3)\cdots\cdots+(KA+AM_n)}{n}=KA$$

$$\therefore\quad KA-AM_1+KA-AM_2+KA-AM_3\cdots\cdots+KA+AM_n=n\overline{KA}$$

$$n\overline{KA}-AM_1-AM_2-AM_3+\cdots\cdots AM_n=nKA$$

所以 $-AM_1-AM_2+AM_3+\cdots\cdots AM_n=O$

所以說自算術平均數所得各離差之和等於零.

74. 求算術平均數的簡便法

若一組中有許多項，而各項形又差不多一樣，用下

法求其平均數較爲簡便．先任取一數爲平均數．將各項與此假設的平均數的離差，依其正負相加，加得之和，以項數除之，除得之商，加於假設的平均數(assumed average)上，其結果便是眞正的平均數．試舉例說明之：

求算術平均數的簡便方法

項	假設的平均數	自平均數之差異
747	740	+7
742	740	+2
735	740	−5
738	740	−2
730	740	−10
786	740	−4
		和　−12

第　十　表

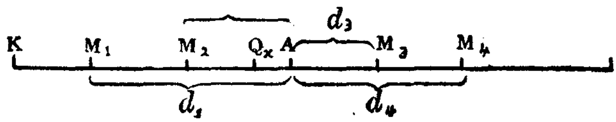
項數爲6

$$-12 \div 6 = -2$$

$$740 + (-2) = 738 = \text{眞正平均數}$$

由次數中用此簡便方法求平均數，自然先用各組之項數乘離差然後再加．

簡便方法的代數證明



第 十 三 圖

設 KM_1 KM_2 KM_3 KM_4 爲 n 項中的四項，KA 爲算術平均數.假設 KQ 爲平均數，d_1 d_2 $d_3 \cdots\cdots d_n$ 爲各項，自真正平均數 KA之離差.設 $QA=X$

求證 $$\frac{\Sigma \text{自 KQ 之離差}}{n}+KQ=KA.$$

證 $$\frac{\Sigma \text{自 KQ 之離差}}{n}=\frac{-(d_1-x)-(d_2-x)+(d_3+x)+(d_4+x)\cdots+(d_n+x)}{n}$$

$$=\frac{-d_1-d_2+d_3+d_4+\cdots\cdots d_n+nx}{n}$$

$$=\frac{nx}{n}=x$$

因爲自算術平均數之離差等於 0 而 $X+KQ=KA$

$$\therefore \quad \frac{\Sigma \text{自 KQ 之離差}}{n}+KQ=KA$$

Q. E. D.

75. 算術平均數之優點

1. 用簡單的加法除法卽可求得算術平均數,不像求範數中數必須畫圖或排各項成一行列,纔能求出.

2. 有時要與極端項數以較量,惟算術平均數能之.

3. 算術平均數每受項之影響,其位置決不能像範數那樣用很少的幾項卽能決定.

4. 算術平均數,人人皆知,無須費解,範數中數則不能如此.

5. 只要知道總和與項數卽可求得算術平均數,各項大小如何,可完全不必要.例如我們已經知道全國產糖總量爲若干,並知由外國進口糖之總量爲若干,以全國人民總數除糖的總量,便得出平均數.每人每年平均用多少糖便知道了.若求範數中數是不能的.

76. 算術平均數之劣點

1. 設如我們已經作好了一個次數圖,算術平均數不能用圖求出.

2. 若極端數有錯誤,平均數就不能精確決定了.從這一點看來,算術平均數不如範數與中數.

3. 算術平均數注重極端數,但是大多數的情形,我們並不要注重極端數.

4. 不能量的東西,不能用算術平均數,從這一點說,他不如中數了.

5. 算術平均數常是實際上所沒有的數目.例如用算術平均數每家人口可為 5.41 人,實際上這個數目,是絕對不可能的.

B. 較量算術平均數 (The Weighted Arithmetic Average)

77. 較量平均數的定義

各項先以各項之較量 (weights) 相乘,然後相加,加得之和,以各較量之和除之,除得之平均數,叫做較量平均數.所用之較量,可代表實在或估量的項數.誠若此,則較量平均數與簡易平均數無大出入了.例如我們知道某行中各鋪一些人的工資為若干,要知道這行中的每人平均工資,必須先把各鋪的平均工資以各鋪工人數乘之,然後相加,加得之和,再以此行中的工人總數除之,方可得正確的平均工資.若是僅用簡易平均數,除非各鋪中

的工人數目都是相同，那自然是不精確的．用較量平均數的方法求平均數，可當作此行中每個人的工資都加起來，以總人數除之．照此說來，其算法與簡易平均數的算法，並沒有偌大的差別．

再一方面說，較量並不代表項數，而代表比較的重要(importance)．例如教員算一個學生一學期的分數，他可用不同的數目代表各强訂的重要．如以 3 分爲班上的臨時記分之較量，2 爲筆試之較量，4 爲學期末之大考的分數的較量，各學生所得積之和以 9 除之．此時較量平均數與簡易平均數，便不甚相類了．

78.　較量的影響

若是較量的數目很少，較量的大小對於平均數有很大的影響．若較量的數目很多，互相補償的機遇亦多．則較量平均數與簡易平均數就相差不多了．然而也要看較量的大小與項形有沒有關係，若較量大，項形也大；或較量小，項形亦小，較量稍錯或稍有忽略，對於平均數很有影響．例如研究工資時，我們知道得小工資的人很多，得大工資的人很少．若僅用工資的簡易平均數，以小工人

的工資與公司總辦的薪金各爲一種工資,加在一起,以2除之,得出一種簡易平均工資;這就以總辦一人的薪金與千萬個小工人同一較量,其平均數就太大了.此時較量萬不可忽略的.然較量錯誤仍較原項形錯誤爲小,原項形的錯誤,並不能用較量的方法改正了,故有下邊的普通規律.

各項愈正準愈好,較量也要逼近的精確.但是若費很多的事求出較量大小的精確就不必須了.

IV 幾何平均數(The Geometric Average)

79. 幾何平均數的定義

把 n 項相乘,乘得之積開 n 次方,開得之結果叫做幾何平均數.算幾何平均數時,須用對數.戒安 (Jevons) 研究物價時,常用幾何平均數.普通一般統計家,尚不贊成此法.

80. 幾何平均數的特質

幾何平均數永久比算術平均數小一點,對於極端數可稍與以較量.所以幾何平均數位於算術平均數與中數之間,算幾何平均數時非常費工夫,這是他的缺點.更有

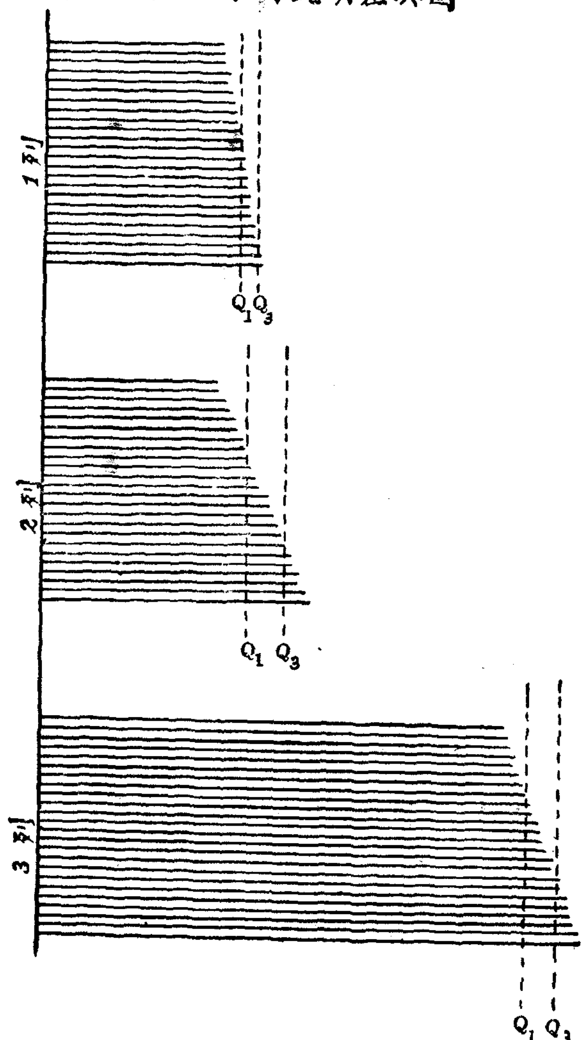
平常的人不懂，無算學腦筋的人也不容易懂，也是最大的缺點.

第十三章　差異(Dispersion)

81.　差異的意義

差異這名詞，是說一羣中各項的大小相差的意思.換一句話說，就是各項大小不整齊.我們說差異很小，卽是說這項與平均數的絕對大小相差很少.若說差異很大，就是說與平均數相差很大.一隊兵的身高，都在68吋至70吋之間，我們就可以說他們的身高很整齊，差異很小.若最矮的兵是62吋，最高的兵是74吋，我們就可以說他們的身高有一些差異了.再舉一例，上古時財富的分配很均，每個人的財產都差不多.現在與古時大不相同了，富者腰橫千萬，貧者窮無立錐.上古的財富分配可以說差異很小，現在的財富分配可以說差異很大.第十四圖中第一列的差異是一吋的十六分之五，第二列是一吋的十六分之十.

樹葉長之行列說明差異圖



第 十 四 圖

差異可用極端數之大小的差量之.換一句話說,卽是用限域(range)量之.亦可用自平均數之普通離差量之.而實際上計量差異並不用限域,因爲他太不一定了.例如某社會中最矮的人爲五呎,最高的人爲六呎一吋,這個限域自然是十三吋.假如只有一個最矮的人是三呎六吋,這限域就遽然增爲三十一吋了,而平均的身高並沒有大的變動.由此看來,因爲一個極端項,限域便受絕大的影響,這方法實際上決無何等價值.故計量差異須自平均數的離差算出,不可用限域,若必須用限域時,也要設法減少分數的極端數.

計量差異,分絕對的與比較的兩種.用絕對計量時,各項的平均大小是沒關係的;用比較計量時,他成爲根本重要了.一羣人的身高相差一吋,這並不算是什麼大差.若是他們的鼻子的長相差一吋,這就是一個大大的差了.第十四圖第二列與第三列之差異的絕對限域是相同的,而比較的差異第二列爲第三列的二倍,因爲第三列的平均組形爲第二列的二倍.

若比較各羣的比較差異(relaeti vdispersion),須求出

各羣的差異係數.他的求法,是以可代表各項形之數除差異的絕對計量,即得差異係數,此係數即代表一羣張本中的普通發生之差異分數.

82. 機(Moment)

差異的計量,平常用自算術平均數中數或範數之平均差 (average deviation). 計量這種離差常用各種機(moments).第一機是平均差.何為平均差呢?平均差就是用項數除離差的總和的意思.設 $m_1\ m_2\ m_3 \cdots\cdots m_n$ 為各項,n 為項數 $d_1\ d_2\ d_3 \cdots\cdots d_n$ 為各項自平均數的離差;可用下邊的公式表明各機.

第一機: $\Sigma d\ n$

第二機: $\Sigma d_2/n$

第三機: $\Sigma d_3/n$

1. 差異的計量與其係數(Measures and Coefficients of Dispersion)

A 第一羣,根據第一機.

83. 平均差與其差異係數

平均差可用中數求之,可用範數求之,也可用算術

平均數求之.求平均差時,我們以爲各項都是正號的數.

設 $m_1\ m_2\ m_3 \cdots\cdots m_n$ 爲各項;爲項數;$d_1\ d_2\ d_3 \cdots\cdots d_n$ 爲各項自平均數的離差;δ 爲平均差;a 爲算術平均數;M 爲中數;Z 爲範數平均差;可任便用下邊三個公式中之一求得之.

用算術平均數求平均差,

$$\delta = \frac{\Sigma(m-a)}{n} \text{或} \frac{\Sigma d}{n}.$$

用中數求平均差,

$$\delta_M = \frac{\Sigma(m-M)}{n} \text{或} \frac{\Sigma d_M}{n}.$$

用範數求平均差,

$$\delta = \frac{\Sigma(m-Z)}{n} \text{或} \frac{\Sigma d_z}{n}.$$

這些差異計量,都可用相當的平均的數除之,得出係數來.根據算術平均數之差異係數爲

$$= \frac{\Sigma(m-a)}{na} \text{或} \frac{\Sigma d}{na} \text{或} \frac{\delta}{a}.$$

根據中數之差異係數

$$= \frac{\Sigma(m-M)}{nM} \text{或} \frac{\Sigma d_M}{nM} \text{或} \frac{\delta_M}{M}.$$

根據範數之差異係數

$$=\frac{\Sigma(m-Z)}{nZ}\text{或}\frac{\Sigma d_z}{nZ}\text{或}\frac{\delta_z}{Z}.$$

下邊的表，說明自次數表中求差異係數的方法，以自中數之離差爲根據。

求平均差的算法

項 形 m	次 數 f	自中數之離差 d_M	fd_M
4	2	3	6
5	3	2	6
6	5	1	5
7	8	0	0
8	6	1	6
9	4	2	8
10	2	3	6
11	1	4	4
	$n=31$		$\Sigma d_M=41$

第 十 一 表

中數 $=M=7$

$$\delta=\frac{41}{31}=1.32+$$

差異係數

$$=\frac{\delta_M}{M}=\frac{1.32}{7}=0.19-.$$

上例我們以中數爲整數,僅是級數爲分立級數時中數纔能爲整數.若是連續級數時,須從第四組中求出正確的中數.還有一件事應注意的,是上例的離差,我們全拿他當作正號(positive)的離差.

這種差異係數的特質是.

1. 容易算也容易懂.

2. 每一項都算在內.

3. 依離差之大小與以較量,極端離差所得的較量比小離差所得者爲大,然並不以此而失掉了他們的比率,各較量還是成比例.

這種係數,用之以研究經濟問題,最爲相宜.如推算某國的個人財富分配用這種係數最好,因爲最富者與最貧者都注意到了.至於用那種平均數爲根據,是無關重要的,用那一種都可以.不過財富分配問題,用中數更爲相宜罷了.

B　第二羣,根據第二機

84.　均方差與其係數(The Standard Deviation and

Coefficient)

這羣中最常用的差異要算均方差．均方差的求法，雖然也可用範數或中數求其離差，但是平常皆用算術平均數求之．其公式如下：

$$\sigma = \text{均方差}$$

其餘各字代表之數，一如83節．於是

$$\sigma=\sqrt{\text{第二機}}=\sqrt{\frac{\Sigma d^2}{n}}=\sqrt{\frac{\Sigma(m-a)^2}{n}}.$$

下邊用普通的次數表，說明求均方差之直接方法．

由次數表中求均方差之直接方法

項形種 m	次數 f	mf	離差 d	d^2	fd^2
8	2	16	−3	9	18
9	4	36	−2	4	16
10	6	60	−1	1	6
11	9	99	0	0	0
12	6	72	+1	1	6
13	4	52	+2	4	16
14	2	28	+3	9	18
	$n=33$	$\Sigma m=363$ $a=11$			$\Sigma d^2=80$

第 十 二 表

$$\sigma=\sqrt{\frac{\Sigma d^2}{n}}=\sqrt{\frac{80}{33}}=\sqrt{2.4242}=1.56-.$$

既求出均方差，更須求其差異係數，卽用算術平均數除此均方差卽得，所以差異係數

$$=\frac{\sigma}{a}=\frac{1.56}{11}=0.14+.$$

85. 求均方差的簡便法

上邊的求均方差的方法，若算術平均數爲偶數時，算時極爲方便.若是小數時，則平方，乘，除，等等手續，就太麻煩了.故不得不用簡便法求之.其法如下.取近於算術平均數的一個整數爲假設的平均數.從這假設的平均數求出各項的離差，將各離差平方起來，再相加.從相加之和中減去假設的平均數與眞正平均數之差的平方之 n 倍，然後以 n 除之，除得之商再開方，卽得.

此法的代數公式爲

設 x = 假設的平均數，

a = 眞正的平均數，

於是　$$\sigma=\sqrt{\frac{\Sigma(m-x)^2-n(a-x)^2}{n}}.$$

試以實例說明之.

假設的平均數$=x=9$

求均方差之簡便法

項形 m	次數 f	mf	m−x或 d_x	$(m-x)^2$或 d_x^2	fd_x^2
6	2	12	−3	9	18
7	4	28	−2	4	16
8	5	40	−1	1	5
9	7	63	0	0	0
10	4	40	+1	1	4
11	3	33	+2	4	12
12	1	12	+3	9	9
	$n=26$	$\Sigma m=228$			$\Sigma d_x^2 64$

第 十 三 表

$$a=\frac{228}{26}=8.77-,$$

$$a-x=0.23+,$$

$$(a-x)^2=0.053,$$

$$n(a-x)^2=1.375+,$$

$$\sigma=\sqrt{\frac{\Sigma d_x^2-n(a-x)^2}{n}}$$

$$=\sqrt{\frac{64-1.375}{26}}=\sqrt{2.4086}=1.55.$$

故差異均方係數等於 σ/a

$$\frac{\sigma}{a}=\frac{1.55}{8.77}=0.177.$$

這簡便方法,根據自算術平均數之離差的平方之和爲最小的定理,此定理可證明如下.

設　$a=$ 真正的平均數.

$x=$ 假設的平均數.

求證　$\Sigma(m-x)^2>\Sigma(m-a)^2$

證　$(m-x)^2=m^2-2mx+x^2$

$$\Sigma(m-x)^2=\Sigma m^2-2x\Sigma m+nx^2$$

但　$\Sigma m=$ 總和 $=na$

$$\therefore\Sigma(m-x)^2=\Sigma m^2-2xna+nx^2$$

$$=\Sigma m^2+n(x^2-2ax)$$

$$=\Sigma m^2+n(x^2-2ax+a^2)-na^2$$

$$=\Sigma m^2-na^2+n(x-a)$$

$$\therefore\Sigma(m-x)^2>\Sigma(m-a)^2 \qquad \text{Q. E. D.}$$

由上式卽得

$$\Sigma(m-x)^2-n(x-a)^2=\Sigma(m-a)^2$$

$$\therefore\sqrt{\frac{\Sigma(m-x)^2-n(x-a)^2}{n}}=\sqrt{\frac{\Sigma(m-a)^2}{n}}$$

但

$$\delta=\sqrt{\frac{\Sigma(m-a)^2}{n}}$$

$$\therefore\quad\delta=\sqrt{\frac{\Sigma(m-x)^2-n(x-a)^2}{n}}.$$

此公式正是均方差簡便法的公式,所以證明其爲正確.

86. 均方差與其係數之特質與用途

先時生物學家多用均方差,經濟學家用均方差者尙少.算均方差時的離差平方多與極端數以較量,此種特質,有時也極有價値;但是以之研究經濟問題則不可.通常研究經濟者,多用平均差,惟算克爾披生的相關係數又爲例外.克爾披生的相關係數(Karl Pearson's Coefficient of Correlation),容後章再爲詳述.求均方差時之離差平方,可將負號免掉,運算時頗覺方便.研究高等整齊次數分配時,此特質更爲有用,所以生物學家樂用均方差.再一方面說,算均方差須經平方開方等手續,不如平

差容易算,所以非有特別原因,經濟家總是用平均差.

更有一種與均方差相似的差異計量是謂率(modus).率也是根據第二機,常用C代表之.其公式如下:

$$C=\sqrt{\frac{2\Sigma(m-a)^2}{n}}\text{或}\sqrt{\frac{2\Sigma d^2}{n}}$$

在初步的統計學中率並不重要,故本書從略.

C　第三羣根據四分數

87.　四分數十分數等(Quartile, Deciles, etc.)

我們說中數是一行列中的正中間一項.同一的說法,們可說四分數是把行列分爲四份的項.十分數是把行分爲十份的項.百分數是把一行列分爲一百份的項.二「四分數」,第五「十分數」,與中數顯然的是一項.中爲$\frac{n+1}{2}$項,第一「四分數」爲$\frac{n+1}{4}$項,第三「四分數」爲$\frac{(n+1)}{4}$項,第一「十分數」爲$\frac{n+1}{10}$項,第七「十分數」爲$\frac{(n+1)}{10}$項,第二十四「百分數」爲$\frac{24(n+1)}{100}$,餘此類推.第五中中數是第五十七項,第一「四分數」爲第二十九項,三「四分數」爲八十五項,用弧形決定四分數可以四除.決定十分數可用十除之.決定百分數可用百除之.在

次數表求此等數,可用求中數之公式.

88. 四分數計量與差異係數

以上所講的差異計量,都是把每項的離差都算,現在要講得出大概的情形就行.不必逐項逐字那樣瑣碎.一列數中,第一四分數與第三四分數中間之數,正為全體的一半,若是這一半的差異卽能代表全體的差異,則計量的方法就簡便了.第五圖第一「四分數」為 6.3 cm.,第三「四分數」為 7.1 cm.,限域的漲落為 0.9 cm..假如若從四分數的中間計量差異,則僅為 0.9 之半,卽是 0.45 cm..第十四圖第一列第二列卽知一羣中的四分數距離變更,則差異量卽變更.十四圖內極端數的距離增加一倍,四分數中間的距離差不多也增加一倍.然則四分差 (quartile deviation) 或為計量一行列的差異最簡單的方法了.其公式如下.

設 Q_1 = 第一四分數

Q_3 = 第三四分數

於是

$$\text{四分差} = \frac{Q_3 - Q_1}{2}.$$

第十四圖中的第二列第三列的四分差是相等的,惟第三列內的平均項形爲第二列的二倍,故須各用各形除之.此數可用四分數的平均長表之郎行.

$$\frac{Q_3+Q_1}{2}$$

於是差異的四分係數爲

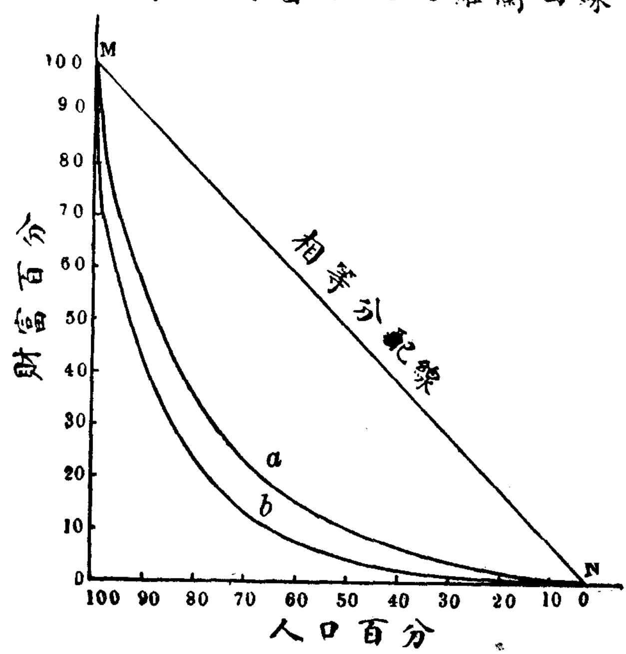
$$\frac{\frac{Q_3-Q_1}{2}}{\frac{Q_3+Q_1}{2}}=\frac{Q_3-Q_1}{Q_3+Q_1}.$$

四分差與其係數之長點,在於簡單易算.假如欲僅知一列數的大體,不及其極端雜項,四分差極爲勝任愉快.第五圖一看便知道,長於第三四分數的葉,或短於第一四分數之葉,對於四分差與其係數均沒有什麼影響.但是一半數目卻出於此四分數之外,此一半數目中未必沒有影響.所以欲以極端數以較量,四分差就完全無用了.從這一點說,四分差與均方差正是相反的,而平均差處於二者之中間.所以平均差常據優越的地位.

89. 羅蘭曲線(The Lorenz Curve)

我們已經知道用直方圖表明次數分配是極有用的.再有一種表明一羣中的差異曲線，這曲線是羅蘭博士(Dr. Lorenz)發明的,所以叫做羅蘭曲線.用之以研究財富分配問題最合於用.這曲線並不像差異係數，予我們

表明財富分配之羅蘭曲線



第十五圖

一個數目的財富分配表現.從這點說，這曲線是不如差

異係數，惟其能表明各部分人民財富，不僅以一個平均數籠統代表全體，這也是他的長處．第十五圖說明作羅蘭曲線的方法．假如人民的財富是相等的分配，可畫一直線MN 表之．從兩極端聯成線卽成 MN 線，而實際的財富分配常爲 a b 線．若 a b 線離相等分配直線愈近，則財富的分配愈平均．若離直線愈遠，則貧人愈多，而財富盡聚於少數富翁手中了．

比較各時期財富的時候，曲線的末端往往相併，最好把曲線的極端部分另畫在一張紙上．於是便清楚了．上邊的極端數目可用橫格，右邊的極端項可用縱格．如此，則曲線淸分，極端項可以分晰了．

羅蘭曲線，也可用以研究土地分配，工資分配，收入分配等等問題．自全體言之，研究此等問題，多用此種曲線．所以他也是差異係數的一助．

參考書

Yule, G. U.: Introduction to Statistics, Chap. VIII.

Elderton, W. P. and E. M.: Primer of Statistics, Chap. IV.

Bowley, A. L.: Measurement of Groups and Series, Lects. II to IV.

第十四章 失稱(Skewness)

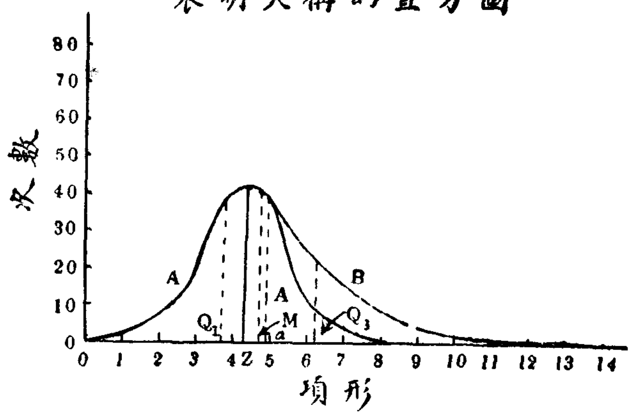
90. 失稱的解說

失稱用於次數分配的時候,是說與均稱相反的意思.就是說各數的分配不相稱,換一句話說,卽是範數以上的次數與範數以下的次數不相等,便叫做失稱.例如把一國內的所有田地的產麥作成一個次數表.假如全國的田地比較的一律,每畝地差不多都產十五斗麥,則田地的分配每是兩組兩兩成對,自範數至兩端,其距離定然相等.多於十五斗之組中的數目,與少於十五斗之組中的數目,其比率一定相等的,於是遂得一常態,均稱次數分配.正如我們已講過擲骰的次數一般.若反過來說,假如國內有一小部分田地極其磽薄,仍以之種麥,於是這種田地的產量,定然較肥田所產者距範數量爲大,此時的差異就不均稱了.這種分配,就叫做失稱向下的分配.

用直方圖來說明失稱的意思更易於了解.若直方圖有失稱存在,曲線就不是鐘形曲線,範數的兩邊就不均

.底邊的一邊,定比他一邊長.如第十六圖之曲線B,其部分較之常態線A傾向右方.

表明失稱的直方圖



第　十　六　圖

社會現象成完全均稱直方圖實是例外,往往都有高的失稱.但是應用失稱者僅是生物學家,其實在某種形之下研究經濟問題,失稱也是有用的.

91.　失稱對於平均數的影響

十六圖中的均稱曲線A上邊,範數,中數,算術平均,三數相合.曲線B三數就不相合了.右邊的大的項將術平均數移至右方.蓋以其常在重心處,故極端項對

之有極大的影響.據居直方圖中間的中數,也向右移其位置.但是各項的大小,對於中數是沒有影響的,所能影響中數者僅是項數.所以中數的位置雖也向右移動,究不若算術平均數移動那樣遠.若是曲線離均稱曲線不遠,中數所移之距離,恆爲算術平均數者之三分之二.故

$$M = Z + \frac{2}{3}(a - Z).$$

範數是仍舊不變的,不受新項的任何影響.總起來說,失稱曲線對於算術平均數影響最大,中數次之,範數絕不受影響.二數移動的方向,卽在曲線失稱的一方面.

92. 失稱的計量與其係數

假如我們要比較兩條曲線的失稱,必須用數字表之纔可以相比.失稱須求出係數,和差異求係數的道理相同,惟是失稱不能用項的平均數大小除之.因爲現在的問題,並不是說曲線的失稱對於形的大小成什麼比例;乃是要知道失稱這一邊的項數比那一邊的項數多若干項.所以他的分母,必是各項的差異.將此意牢記在心,卽可進行討論失稱的計量與他的係數了.

93. 失稱的第一計量與係數

失稱曲線將算術平均數從範數移動,此移動的距離,成爲最簡單一種計量失稱的方法.

設　　a =算術平均數

M =中數

δ_m =自中數的平均差

δ =自算術平均數的平均差

Z =範數

δ_z =自範數的平均差

j =失稱係數

最簡單的失稱計量,可以 $a-Z$ 代表之,最簡單的係數,可以下式代表之.

$$\frac{a-Z}{\delta_z}, \qquad \therefore \quad j=\frac{a-Z}{\delta_z}.$$

只要前後用的數相同,用 δ_z 或用 δ 是無關重要的.

上邊的公式,是一個理想的失稱計量,蓋以範數往往不能十分確定,故不得不用中數與算術平均數之差.但是我們已經知道這個差,僅爲算術平均數與範數之差的三分之一(見 91 節).若失稱很小數目就不清楚了,此

為其大缺點.用中數求失稱係數之公式如下:

$$j=\frac{a-M}{\delta_M}$$

94. 失稱第二計量與係數

這種計量根據失稱曲線的中數不在四分數的正中間了.蓋近於失稱一邊的四分數移向失稱的方向的距離,較對方的四分數的移動者為大.在失稱方面者趨向疏境.換一句話說,就是趨向次數少的一方,移動快;對方的四分數漸近最密處,移動慢.中數近於高次數之範數故遲遲移動,結果四分數自中數緩緩分離成比較的距離,用此兩距離之差以計量失稱.於是失稱第二計量的公式為:

$$(Q_3-M)-(M-Q_1) \text{ 或 } Q_3+Q_1-2M$$

以四分差除上式即得失稱係數.

$$j=\frac{Q_3+Q_4-2M}{Q_3-Q_1}$$

這種係數與差異四分係數有同一的缺點,二者都沒把極端項大小算上.用這種算法求得之美國財富分配曲線的差異與失稱富人之有 \$100,000者,或\$100,000,000者,都沒有影響的.二者都在上四分數之上,對四分數

中數毫無關係，這是四分係數的缺點，而其長處在於簡易易算．爲實際的應用亦具有相當的精確．然僅以極端項非根本重要者爲限．

95. 失稱的第三計量與係數

此係數根據第三機並根據數的立方其正負號不變的道理．第 73 節已經說過．自算術平均數之離差總和(依正負號相加)等於零，但是自算術平均數之離差的立方之和並不等於零．因爲立方時已將極端項加上較量了．在失稱曲線上，這是極重要的一件事，故第三機爲計量失稱的滿意的根據．從這一點說，這種係數是和四分係數相反，此係數注重極端數，四分數完全不注意極端數．此失稱計量的公式爲：

$$\sqrt[3]{\frac{\Sigma(m-a)^3}{n}}=\sqrt[3]{\frac{\Sigma d3}{n}},$$

將此失稱計量作爲係數，有許多數可爲分母，惟以其注意極端項仍以用注意極端的均方差爲分母爲最相宜，其公式如下：

設　　j=失稱係數　　則

$$j=\frac{\sqrt[3]{\frac{\Sigma d^3}{n}}}{\sigma}.$$

若用平均差爲分母其式如下：

$$j=\frac{\sqrt[3]{\frac{\Sigma d^3}{n}}}{\delta}$$

此爲最好的失稱係數，惟算時須費很多手續，這也是他的缺點.

此外計量失稱及其係數的方法尙多,然大多都是極其複雜艱深,難於實用,初步的統計學,尙談不到那些方法.

參考書

Yule, G. U.: Introduction to Statistics, Chap. VI.

Bowley, A. L.: Measurement of Groups and Series, Lecture II.

第十五章 歷史統計(Historical Statistics)

96 通論

前幾章所討論的張本，都沒把時間算在裏邊，其實比較各時間的張本，也是統計學中的一重要部分，比較

方法有種種,常用者爲:

1. 絕對數碼表;
2. 絕對歷史圖;
3. 對數表;
4. 對數歷史圖;
5. 指數;
6. 指數歷史圖.

以下逐一說明,惟第一項人人皆知,毋庸贅述了.

97. 絕對歷史圖或普通歷史圖(Absolute or Ordinary Historigrams) ——緩和——流動平均數——趨向

事物之變遷,經過一些時間的記載,叫做歷史級數istorical series).把這級數作成圖,以個體的大小爲縱(ordinates),以時間的久暫爲橫距(abscissa),叫做歷史.讀者不要以此與直方圖相混.歷史圖最重要的元素時間.直方圖則與時間是毫不相干的.

歷史級數的精確,與歷史圖的精確,都依記載時間距離爲標準.每時一記的氣候表,比每天一記者定然

精確;每天一記者,較一星期一記者尤爲精確;依此類推,時間距離愈短者愈精確.

作歷史圖時,先依歷史的張本作成各點,再將各點連成直線成爲多邊形.平常總要將此多邊形緩和成曲線,其緩和的方法,與前述的緩和直方圖的方法相仿.惟用手隨意劃曲線時,要常保持曲線最大可能的半徑,除非的確知道某處有裂口存在,不要使曲線上有尖的棱角.如人口調查,氣候報告等一類的統計,很少遽然的變動,故曲線上普通是沒有棱角的。

緩和歷史圖一個最好的方法,是用活動平均數(moving average)得出一個趨向(a trend).活動平均數僅用於循環歷史圖上,其目的在於免除該圖的起伏.但是多麼長的時間可爲算活動平均的一循環呢?這要依情形的不同而不同,要以其時間正爲一循環者爲宜.決定求活動平均數時所用的時長,最好的方法,先把張本作成歷史曲線.再察看這曲線的起落凹凸,看他第一個凸處與第二個相距是多少時間,平均是多少時間.再視其凹處.於是即可得出一個大概的波長(wave-length).用下表的張

作成第十七圖.從這個圖上我們知道他的波長是從六到八天.活動平均數所用的項數,最好要爲奇數,以便平均數放在正中間的一項旁邊.如上例卽以用七天爲宜的期長.

求活動平均數的第一步,是前七項加起來,以七除,所得的平均數,放在第四項相對的地方.再從第二項第八項相加,以七除之,所得的平均數放在第五項相的地方.依此類推,直到最後一項爲止.其結果當如下.如項數很多,求活動平均數就麻煩了,有一簡便方法,之可省工不少.其法將每次去掉那一項與每次加上那項的差,加在前一個總和上,或從總和內減去,然後再,而得平均數.

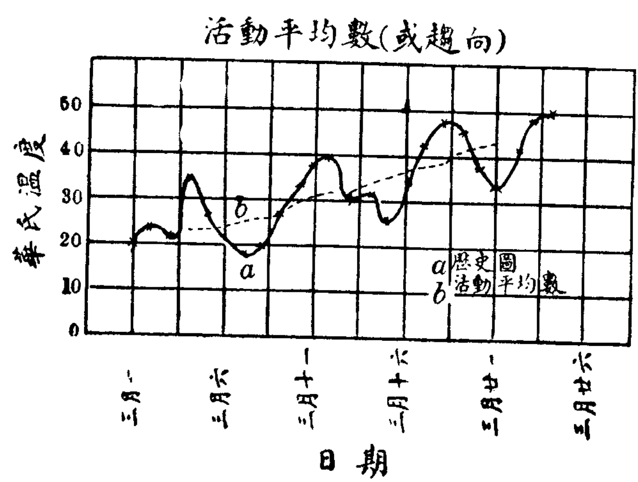
例如下表,從第一項到第七項的平均數爲24.0°,從二項到第八項的平均數也是24.0°.因爲第一項爲.0°第八項也是20.0°,去一個20.0°加一個20.0°,還是於無加減,自第三項至第九項加上的數爲28°,減去的爲25°,我們將總和上加上一個「3」便行了,於是其平數爲24.4°,這種手續比較的便利,幸學者用之.

決定趨向說明表

日 期	平均溫度(華氏度)	活動平均數七日爲一組
三 月 1	20	
2	25	
3	22	
4	35	24.0
5	26	24.0
6	22	24.4
7	18	26.1
8	20	26.7
9	28	28.7
10	34	29.9
11	39	31.9
12	40	32.7
13	30	33.6
14	32	34.9
15	26	36.1
16	34	37.1
17	43	38.4
18	48	38.9
19	47	41.1
20	39	43.3
21	35	44.3
22	42	
23	49	
24	50	

第 十 四 表

求趨向(trend)至於最末一位是不可能的,即使用一種通融辦法,將最後一項重疊相加,其結果亦不甚精確.其實順手隨便劃至極端項也未嘗不可.如上表將最末一項 50 連着重疊相加三次,也可得出一組新數,然而這種新數,不過是一個逼近數罷了.同是一個逼近數目,倒不如隨手劃下去反倒簡便.



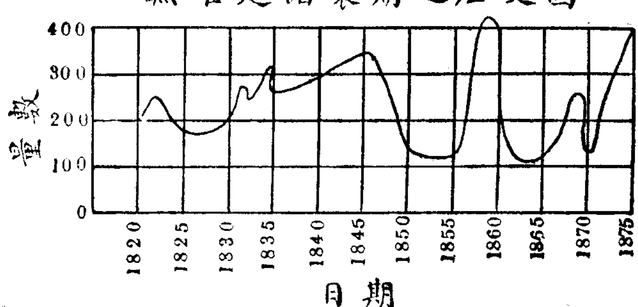
第　十　七　圖

第十七圖所示的活動平均數,所有的不規則的起伏,全都消滅了,僅得有一個全期氣候上昇的一個趨向.若研究短期的東西,須用原來的歷史曲線,僅示一班的趨

向線,是不能用的.

活動平均數不是萬應膏,並不是用之於所有歷史圖而皆準的妙法. 如第十八圖沒有一個有規則的循環期,故不能用活動平均數.若必欲一試,只有用很長的期間,或三十年爲一循環期,亦不爲長,而結果仍不過是一個全期的普通趨向,至於全期中何年變動最大一概抹殺了.

無確定循環期之歷史圖



第　十　八　圖

無論那種歷史圖,他的縱格都要適當選定,若格式太大了,各項形微有變化,便似有一個巨大起伏一般.反之,若佔地位太小,則畫成之曲線,便似全體一律,毫無變化一般.二十一圖的麥價歷史圖,正可說明此意,當其變爲指數縱格增加時,便可見其變化了.

98. 比較的或比例的變化(Relative or Proportional Changes)

假如某城的人口記載如下:

1890…………………100,000

1900…………………150,000

1908…………………200,000

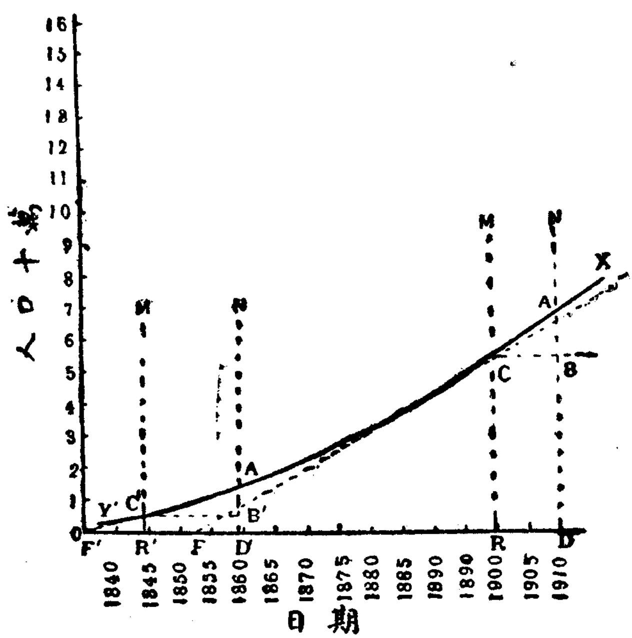
顯然的兩次人口增加的數目,其絕對數是一樣的,每次都是增加50,000.但是他的增加比例是不同的,第一次增加原數的百分之五十,第二次增加其原數的百分之三十三.每年增加人數,第一期爲每年平均爲50,000人,第二期每年平均爲6,250人.第二期增加率自然比第一期者大.惟是第二期之比率,係根據的人口數目較大的數,假如各以原數爲根據,則第一期人口增加比例率爲5,000\100,000=5%,第二期人口增加比例率爲6,250\15,000=4.17%,後邊這變化的比例率,纔是我們要研究的.

假如說紐約的人口近十年增加一百萬,六十年前每十年增加十萬,這種說法我們得不到一個兩期比較變化

的概念.馬續教授(Professor Afred Marshall)有見於此,發明一種圖,以便各時期的變化,比例率說明見十九圖.

XY 是表明某城中自 1840 至 1910 的人口歷史圖,現在要知道 1845 至 1860 與 1900 至 1910 兩期中,那一期的增加比例率大.設 M'R', D'N', MR, DN, 為兩期的相當

用馬續方法表明人口增加比率



第十九圖

縱距.與 XY 線交於 C′, A′, C, A 四點.作 C′B′, C B 線與底邊平行.作 A′ C′, A C 線並延長之與底邊交於 F′,F 兩點.

於是A′B′ AB 即代表兩期的絕對增加,A′B′/B′C′ AB/BC 爲其增加率,其增加比例率爲

$$\frac{\frac{A'B'}{B'C'}}{C'R}\quad 與\quad \frac{\frac{AB}{BC}}{CR}.$$

但　　AB/BC=CR/FR (因兩個相似三角形之對邊)

$$同理\ \frac{A'B'}{B'C'}=\frac{C'R'}{F'R'}$$

$$\therefore\quad \frac{\frac{AB}{BC}}{CR}=\frac{\frac{CR}{FR}}{CR}=\frac{1}{FR}=1900-1910\ 的增加比例率,$$

$$\frac{\frac{A'B'}{B'C'}}{C'R'}=\frac{\frac{C'R'}{F'R'}}{C'R'}=\frac{1}{F'R'}=1845-1860\ 的增加比例率.$$

所以增加比例率與 1/FR 成正比例,與 FR 成反比例.十九圖中的 FR 差不多等於 5 F′R′,所以 1900－1910 的增加比例率,差不多爲 1845－1860 的五分之一.

這個方法的長點,在於應用時極爲簡易,只要劃 AF,

CR 等線,計量 FR 線就算完事了.

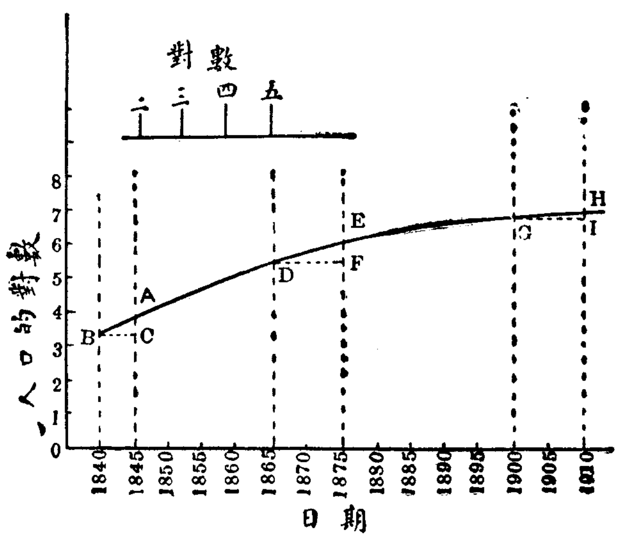
99. 對數歷史圖(Logarithmic Historigrams)

對數歷史圖之發明,純爲表明全期中各部分的比例變化,其理係據一個數的相等加便是以相等數相乘,也就是相等的比例變化.第二十圖就是一個對數歷史圖,自 1840 至 1845 的人口比例增加以 AC 代表之.自 1865 至 1875 間者以 EF 線代表之,自 1900 至 1910 間者以 HI 代表之,因爲 AC 差不多等於 EF.所以第一二兩期的比例變化,也差不多是相等的.但是變化的比例率便不同了.後邊這一期的線幾爲前期者之二倍,所以其變化比例率,當爲前期二分之一,最後 1900－1910 這一期的比例變化與變化的比例率都很小.

變化的比例率以曲線的陡峻表明之,其算法可以說以三角形之底除其高.如 AC/BC EF/DF 等,是這種結果,是不十分正確的,而其解說,適足以示出對數曲線的弱點.對數增加一倍時,而對數根並不一定加一倍.設 AC＝2HI 並不能說 1840 至 1845 之比例增加,定二倍於 1900 與 1910 的比例,以 3 的對數(0.477)增加某數的對數的三

倍.但某數以 6 乘之,其對數的增加,並不是 $2\times 0.477=0.954$, 乃是 0.778.於是可說比例變化,加倍小於對數歷史圖的縱格運行.寶來 (Bowley) 欲矯此弊,乃用縱線代表 2,3,4 等等的對數,用這些線之長以比較對數曲線的縱線變化,即可決定人口之增加為二倍三倍或四倍了.其格式見第二十圖.

對數歷史圖



第　二　十　圖

對數歷史圖,對於各時期之事物,固然極有價值,對於同時的各事物的形之比較就不行了．對此既然無用,故常將底線分配將各曲線縱移直至易比其比例變化時爲止．譬如比較十年中的鋼鐵價格與木料價格的趨向,我們並不管這兩種東西每一單位的價格如何,所要知者僅兩種東西價格的比較變遷而已.對數曲線對此頗能勝任裕如.惟有一缺點,卽是一般人對於對數曲線不甚熟悉,不易領會,而得一堅固的概念.因此對數曲線,多用於專門科學上,平常的統計多不用之.

100. 指數通論 (Index Number-General Characteristics)

把歷史統計作爲指數有兩種意思:第一是用他比較二時期或二時期以上的個體的比較變遷:第二是用他算出一組平均指數級數來.

簡單歷史圖上用數目,對於常模 (norms) 的起伏大小以資比較者,常是不能比的.當着二曲線代表的東西不同,或二者的數目相差太遠的時候,更不能相比.第二十一圖第一部分,就是表明這種不能相比的困難情形.

用絕對數目相比的時候，好似鋼的價格漲落很大，而麥的價格始終不變一般．其實並不是如此．其所以有這種現象者，完全由於兩種東西的價格相差太遠，選擇的單位不能相比的原故．假如不以每斗麥價爲單位，以每三十斗之價爲單位，則由其曲線的起伏便和鋼線的起伏相差不多了．

若在這圖的豎格上變更一種，對於麥價這是一個很大的變化，對於鋼價那是很小的變化，若打算把兩組數作成能比的數，免去此種困難，最好把他作成指數．其作法以某年的價格强訂爲根據，以之除物價級數中的各項，除得之數卽爲指數．或以全體的平均數爲根據，以之除各項也是可以的．最常用的指數還是用某年的價格爲根據，但是以一年的價格爲根據，不如以幾年價格的平均數好．因爲平均數是一個代表的數目，受偶然的現象的影響，機遇比較的少．故用全體物價平均數爲根據，是最好的根據．統計學中多用之．下邊表中的指數，卽是用這種方法算出來的．

指數離差

年	每噸鋼之價	每斗麥之價	鋼價指數	麥價指數
1890	$30	$1.05	120	117
1891	27	.90	108	107
1892	24	.94	96	104
1893	22	.83	88	92
1894	24	.88	96	88
1895	26	.92	104	102
1896	22	.72	88	80
	平均 25	平均$0.90	平均100	平均100

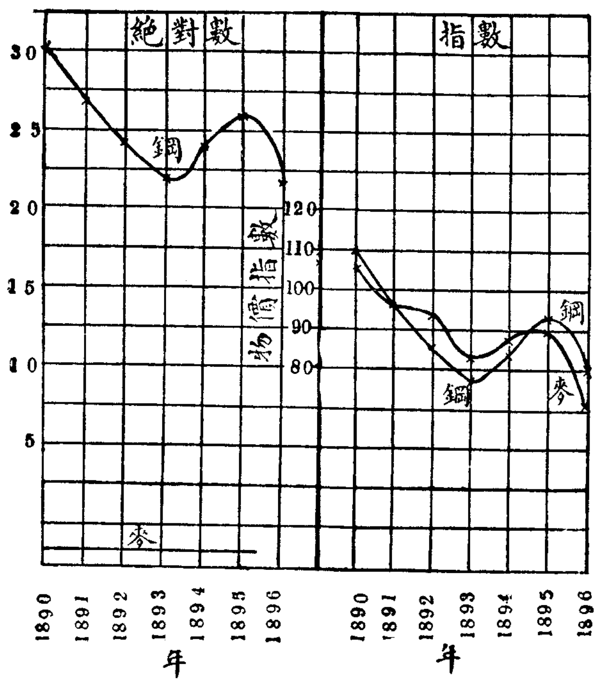
第十五表

以指數作成曲線圖最顯然的一件事,就是鋼價的漲落和麥價的漲落很相彷.現在若在豎格變更一種,對於麥鋼二價,都是代表相同的比例變更.所以這兩條曲線,纔可正確相比較.

於是可知將歷史張本作爲指數,對於比較同時期的各個體很爲方便.惟對於比較各時期中的同一物件,仍毫無進益.指數僅能比較各個體在某時期中變遷,至於

比較幾個時期中的比較的變遷,那便是對數歷史圖所有的事了.

鋼價麥價變遷圖



第二十一圖

101. 平均指數(Average Indices)

往往我們要把各種東西都合在一起找出一個普遍的趨向,故須每期當求出一個平均指數來.美國勞工部

每年所刊行的物價指數與工資指數,就是平均指數的好例.平均指數,是由很多物件重要不同的物件算出來的.其算法的第一步,是把每一種東西全時期的指數級數求出來,用同一時間爲根據,將所有的東西的指數級數都算出來.美國勞工部所用的時間根據是1890-1899.這十年下邊的表,是平均指數表一個縮影,說明求平均指數一個方法.此表僅舉七種東西爲例.實際上所用者,常爲數百種,此不過爲說明的方便而已.

指數的中數

年	物價指數							中數指數
	麥	棉花	鋼	木料	穀	羊毛	皮革	
1880	101	120	104	108	103	92	104	104
1881	97	90	102	103	97	99	102	99
1882	95	82	94	100	96	110	96	96
1883	102	108	99	90	100	100	98	100
1884	105	100	101	99	104	99	100	100

第 十 六 表

接着我們便要問用那一種平均數能得最好的結果呢?範數呢,中數呢,算術平均數呢?這要依問題的特別

性質而定.假如若研究生金或貨幣之量對於物價的影響,所用的平均數與研究生活費所用的平均數自然不同.研究金量對於物價影響的時候,無論那一種貨物價的變遷,都可以作爲標準,並不計較那一種貨物是重要,那一種次要,那一種不重要.假設再沒有別的東西能影響物價,則金量對於物價的漲落,無論那一種貨物全是一樣.如有特別大漲落,(如1880年的棉花與1882年之羊毛)雖初想以爲他一定有特別的影響,而算物價普遍趨向的時候,並不把他算入.不受極端數的影響的平均數,自然是中數,故用中數爲最宜.如將1880年的指數,按其數的大小順序排列之,卽得出中數爲104.這就那年的指數.1881年的指數爲99.依此類推,直到最末爲止,每年的平均指數,既然求出物價,普遍的趨向便可表明了.

假如我們要研究某時期內的生活費變遷如何.所用的平均數與上例所用的便不同了.因爲肉價昂漲,雖然和胡椒價的跌落,是同一範圍消費者,並不能因胡椒落價而補償抵消肉的漲價.因爲,他所用的肉量和所用的胡椒量是不同的.所以計算消費的指數(consumers' index)

時,須按照各物消費量與以較量.然則較量平均數(weighted arithmetic average) 是最好了.各種貨物的較量,不是永久一致的,平常所用者,由一羣消費者的預算表爲標準,以表中各物所費用的百分數爲較量.以之乘各物之指數,所得之積再相加,加得之和以較量之和除之,卽得消費者平均指數.下表說明此方法.

指數的較量平均

較量	40		16		14		6		24		消費指數
物品	食品		房租		衣服		電燈煤火		雜支		
年	指數	積	指數	積	指數	積	指數	積	指數	積	
1891	108	4,320	102	1,632	110	1,540	96	576	104	2,496	106
1892	99	3,960	100	1,600	101	1,414	101	606	101	2,424	100
1893	93	3,720	98	1,586	89	1,246	103	618	95	2,280	94

上表小數從略

第　十　七　表

1891 年的各積之和爲 10,564,較量之和爲 100,以較量之和除各積之和,其商數約爲 106,這就是那年的消費者指數.

平常這些原始指數，便是許多小指數的平均數．例如上表裏邊的食物指數，便是用許多食物的指數得出一個較量平均數而得來的．其所以用此方法的原因，而不用每種物件的直接較量平均數者，蓋由於每年每一種貨物的價格，完全得到常是不可能之事．用以上的方法，便可進行無阻了．

參考書

Mayo-Smith: Richmond, Statistics and Economics, Macmillan Co., N. Y., 1896, pp. 196-233.

Bowley, A. L.: Elements of Statistics, Chap. VII.

Fisher, Irving: The Purchasing Power of Money, Macmillan Co., N. Y., 1911, Chap. X.

Meitzen, August: Statistics, pp. 195-204.

Bertillon, Jacques: Cours Elementaire de Statistique, Chap. XI.

Edgeworth, F. Y.: Index Numbers, Palgrave's Dictionary of Pol. Econ.

Adams, Thos. Sewall: Index Numbers and the Standard of Value, Journal of Pol. Econ., Dec., 1901 and Jan., 1902.

Fountain, H.: The Construction of Index Numbers of Prices, Board of Trade Report on Wholesale and Retail Prices in the United Kingdom, 1903.

第四編　個體的比較

(Comparison of Variables)

第十六章　比較的各種方法

102.　比較的目的與其價值

前邊已經說過比較是研究統計學的最後目的地.用比較可以把一件事的時地關係表現得十分清楚,更可決定這現象是關聯的呢,或是獨立的.並可尋出原因結果的關係來.

我們要研究的事可有下列各事.

1. 單一個體的變遷.
2. 各羣的構造.
3. 兩個個體或兩個個體以上的變遷.

前一章已經講過用歷史表,簡易歷史圖,對數歷史圖,說明單一個體的變遷.各羣張本的構造比較方法也約略說一點.現在要把各羣的比較作一個總結,或者也

是學者的一助.

103. 兩羣或兩羣以上張本次數分配的比較

1. 用簡易次數表或直方圖.

若比較的目的在於表明各組的絕對的或比較的大小,這種次數表或直方圖就很可以用.例如各公司僱用的各級工人比較數目,及各地的工資分配情形,都可用這種方法.

2. 用百分次數表與直方圖.

這種百分表與百分曲線對於各組的實際的大小,一點不能表現.其主要目的,在於使各地的高低羣的比較,分配清清楚楚表現出來.這種表圖,並不能表明各公司實際上僱用多少工人,但是各公司的比較工資,卻能清清楚楚表現出來.

3. 用堆積表與弧形.

弧形最大的用處在於決定中數,四分數,十分數等等.但是也可代替簡易次數表與直方圖.

4. 用百分堆積表與弧形.

這種圖表,若用之以比較某種東西,較絕對堆積表

更好,但是用以決定中數就不行了.

5.　用羅蘭表與羅蘭曲線.

若表明財富分配或收入分配等等,這種方法最好.

6.　用差異係數.

用差異係數可以計量自某種平均數的差異的數目的量數.

7.　用失稱係數.

用失稱係數的幫助,可將各項的欠集中偏向某一方的數目計量表出來.

8.　用相關係數.

容後章再講.

104.　兩個體或兩個體以上的比較變遷方法

各種比較方法已經約略述過了,現在不過總束於此.其主要方法為:

1.　用絕對歷史表或歷史圖.

表明各個體實際上的變遷,這是最好的方法.若用這種方法將各國麥之收穫張本作成曲線,這期中的各國產量的變遷與某時期的比較出品,都可一目瞭然了.

2. 用指數與指數圖.

這種方法是比較兩級數的變遷,並不是比較他們的實際數目大小.所有的曲線都以同一根據爲標準,若比較同時期內的各個體對於根據的比例變遷,那是很容易的.如紐約與吳明(Wyoming)兩城,在1900至1910期中的人口是否增加,或是那一城增加率大,一看便知道了.紐約城的人口的實際增加,也可表明他的比較的比例變遷.但是仍須記着指數曲線並不能表明十年與前一個十年相比的比例變遷.

3. 用對數的指數圖.

同一時期的各個體相比較,用簡易對數圖的竪的運動,便知其與前期的比較的比例變遷.其自横線的斜度,卽是表明各個體在那時期的變遷變的比例率.前邊已經說過用對數圖比較個體的利益,在於曲線的竪移直至互相逼近爲止.比的時候用目察其變移而比其趨向,若是將原來的張本作成指數,並用指數的對數作圖.則相比的法子是算學的,並不是機械的.對數曲線既然不能表明各個體的實際上的大小,所以作成指數或曲線的動移,

都是毫無損失的.

4. 用相關係數.

容下章詳述.

105. 比較曲線的作法

兩條曲線或兩條以上的曲線相比,最好是把他們都用同一格式同一軸線畫在一張紙上. 如怕他混亂, 可用各種顏色表明之. 若是將此圖付印的時候, 也可用更好的方法以代替顏色. 如用輕重線,圓點,斜線等等,都是很好的法子. 相比的線固然要放在一張紙上, 但是也不要一張紙畫的曲線太多,致各線都擠在一起,混亂不淸了. 若有五六條線相比, 最好用一條作爲根據, 每張紙上都把這條線畫上,再兩兩相比,這條作爲比較根據的線,應畫得粗一些,以示與餘線的區別.

106. 長短期的起伏

我們知道所謂長期短期,不過是一種比較對待的名詞.同是一個時期,對於這種事是長期,對於別一種事或者就是極短的期.惟是大多數的歷史級數兩種或多於兩種的起伏常是同時發生的.例如研究十九世紀的結婚率,

全世紀有一些下落的趨向;每五年十年有一次財政的恐慌;更有每年六月有一次起伏.這三種變化,每一個原因各自不同;但是這三種原因都是同時活動.研究氣候的變化也是一樣道理,每五六天就有一個循環.這是因爲有一種旋風(Cyclones)按期經過美國.每一年氣候有一個循環,是因爲地球循軌道繞太陽旋轉.每十五年有一循環的原因,現在還不知道.要考究每一循環的自身,惟有用物理學家的方法,把非原因的元素都免除.可惜統計學不能像物理學家那樣能自如管理他的試驗情境.統計家能作第二件事.就是能將顯然的額外的原因的影響免除了.例如研究失業的長期變遷,若能把季候的起伏(Seasonal fluctuations)減除之就更好了.用97節所講的活動平均數最好,這浪波的長度自然是一年.所以所用的活動平均之長自然也是一年.

107. 長期變遷的免除

上節講的免除短期變遷方法,我們已領會了.但是往往我們所要研究的是短期變遷,所以現在要講免除長期變遷的方法.例如研究失業的季候變遷,常有因經濟

恐慌以致成特別起伏者．求算季候的一個簡易方法，就是從幾年的季候失業算出一個季候平均數.假如僅能得到按月的記載，其運算手續可如下面的表．用這個平均數,失業的季候趨向就十分顯然了．

更有一種在歷史圖上減免長期變遷的方法,其法如下：把張本作成曲線選出相當的期間，求出活動平均數的線原張本與此線的各離差卽得出來了.再把各離異作成表．最後再以之於同一底邊上作成圖．於是長期變遷就可減免了．

失業的百分數

月	年						
	1900	1901	1902	1903	1904	1905	平均
一　月	4.1	5.0	4.7	12.6	7.4	5.1	6.5
二　月	3.6	4.8	5.2	12.1	8.1	4.6	6.4
三　月	3.2	4.1	5.4	9.2	5.2	4.3	5.2
四　月	2.1	2.7	3.8	7.4	6.1	4.2	4.2

第　十　八　表

108.　兩歷史圖上長短期變遷的比較

在比較統計學裏邊這是兩個普通最重要的方法.比

某物的供給價格短期起伏的比較

年	供給			物價		
	供給的指數	指數的活動平均數	自活動平均數的離差	物價的指數	指數的活動平均數	自活動平均數的離差
1880	80			146		
1	82			140		
2	86	84	+2	130	133	− 3
3	91	85	+6	117	129	−12
4	83	87	−4	133	124	+ 9
5	85	89	−4	127	117	+10
6	89	89	0	115	114	+ 1
7	96	91	+5	95	109	−14
8	93	92	+1	100	104	− 4
9	90	93	−3	106	100	+ 6
1890	91	94	−3	103	96	+ 7
1	94	96	−2	94	89	+ 5
2	100	98	+2	75	83	− 8
3	105	99	+6	66	80	−14
4	102	100	+2	75	79	− 4
5	96	101	−5	91	80	+11
6	98	103	−5	87	82	+ 5
7	106	105	+1	81	83	− 2
8	114	108	+6	76	83	− 7
9	112	109	+3	82	86	− 4
1900	109	111	−2	91	88	+ 3
1	106	112	−6	100	88	+12
2	112	113	−1	89	88	+ 1
3	120	114	+6	76	89	−13
4	118	114	+4	82	91	− 9
5	112	113	−1	100	96	+ 4
6	110	112	−2	106	101	+ 5
7	107			114		
8	113			103		
	平均=100			平均=100		

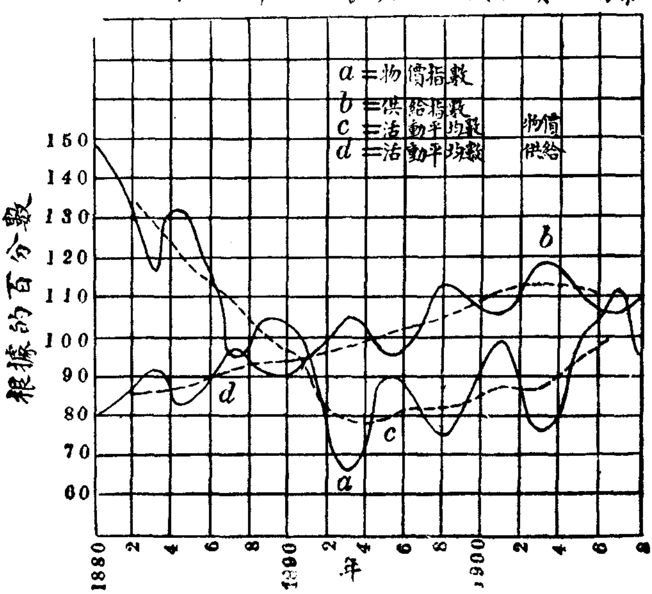
第 十 九 表

註——上表中的數目精確僅至於整數

較短期離差的方法，是把各項都作成指數，用方纔講過的方法把兩組最後的指數在同一底邊上作成曲線.其方法可用下邊假設的表和二十二二十三兩圖說明之.

二十二圖表明以活動平均數免除短期變遷的方法.曲線d表明某種物品的供給漸漸上昇，直到1904年纔微微下降.曲線c是表明物價的趨向.供給增加，物價下落，

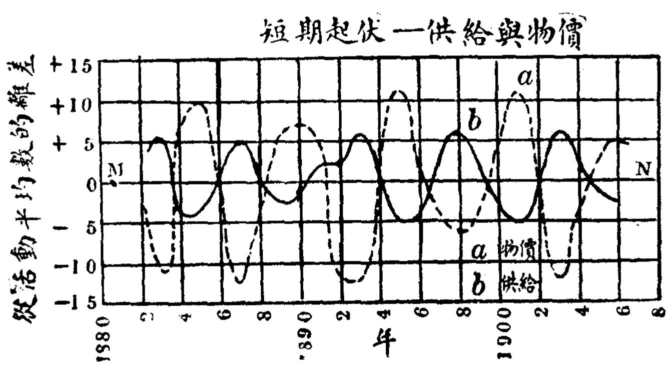
指數圖活動平均數表明供給與物價之關係



第二十二圖

正是符合經濟學的原理.惟至 1894 年物價不循常軌,他不管供給的上昇,他也上昇了.這或者由於別國這種物品供給減少的原故,也或者由於本國內對於這物品需求增大了.如人口的增加,財富的增加,或人民習慣的變更,都可爲需求增加的原因.

二十三圖表明的方法,正是和二十二圖相反.二十三圖是表明免除長期變遷的方法.設想把二十二圖裏邊的活動平均數的線都拉直了,放在二十三圖 MN 線上,而原來的曲線仍然如故,使之仍保持與活動平均數的線的關係,不要使他和這平均線離開,於是便成了二十三圖了.



第 二 十 三 圖

這樣一個圖,我們看出這兩條曲線有一個顯然的關係. a 若上昇, b 就下降; b 若上昇, a 就下降.其相反的趨勢極爲顯然. 但是在長期變遷中這種關係就不清楚了. 平常短期變遷比長期變遷更有用,所以短期變遷也更重要.

參考書

Yale, G. U.: Introduction to Statistics Chaps. X and XI.

Bowley, A. L.: Elements of Statistics, Chap. VII.

Hooker, R. H.: Correlation of Successive Observations, Jour. of the Royal Statistics Society, LXVIII, 696-703.

Norton, J. P.: Statistical Studies in the N. Y., Money Market, Macmillan and Co., N. Y., 1902.

第十七章　相關(Correlation)

109. 相關的定義

前一章研究各種比較張本的方法.當着研究比較方法的時候,已與相關接近了.相關是什麼呢?相關是說兩羣張本的中間存立一種原因結果的關係. 譬如飛雞鳥

(Fiji islands) 的椰子收穫增加了,同時美國的貨幣供給也增加了.除非我們能證明這一件事是那一件事的原因,或是這兩件事有一另外共同的原因.我們就不能說他這兩件事是相關的. 沒有相關的事物可用 $a\propto b$ 表之, 有相關存在時,則 a 的變化必與 b 的變化有一定的關係.

110. 相關的種類

假如非原因的分子都能完全免除了,則無論什麼時候 a 對 b 都有一種的確的算學的關係.例如爲 $a\propto b$, $a\propto 1/b$, $a\propto\sqrt{b}$, $a\propto(b+x)$ 等等.但是實際上的外界影響特別的多,也特別的複雜,要把我們所樂意要的影響完全去掉了,是不可能的.平常所能免除者,不過一二最重要的影響而已.不和諧的小分子仍紛然存在.要打算在二事之間求出一種精確算學關係,真是一件難事.若能知道這一件事減少,同時那一件事,有上昇或下降的趨向就算滿意了.假如有兩件事的起伏常常趨於同一方向,或是相反的方向.我們便能決定這兩件事有一種關係,這種關係便叫做相關.若是兩條曲線的起伏常常向同一的方向這叫做正相關 (Direct correlation).若是常常向相反的

方向這叫做反相關 (Inverse correlation).二十一圖中麥價曲線和鋼價曲線的起伏常是同一方向,所以這個相關是正相關.二十三圖中的物價曲線與供給曲線的起伏正與二十一圖相反.物價上昇時,同時供給減少;供給增加時物價下落.按照短期變遷說這相關是反相關.這兩件事在全期中的變化,總有一種極近的關聯.我們便說他們有一個高度的相關.第一舉例中的麥價鋼價都是依照別一種元素爲他們的原因.如貨幣數量的增減,或普通商業情形的變更,皆可爲他們的共同原因.第二舉例中的供給變化,就是物價漲落的原因,這兩件事的變化,是直接成因果的.

111. 相關的應用

我們可以研究兩種歷史事實的相關;也可以研究任便兩種現象不含有時間元素的事物的相關.第一種的舉例,如生金運動與出口的相關,結婚率與麥價的相關,失業與銀行清算的相關等等都是.第二種舉例是很多很多的.如樹葉的長與其寬的關係就是一例.長的葉一定也寬嗎?就是說,若樹葉之長增加了,他的寬同時也增加`

呢,還是仍舊不變呢?再如高格的父親,他的兒子也必定比平常人高嗎,或是和平常人一樣高呢?假如高父親一定有高兒子,這就有一種相關.相關的求法常用指數歷史圖.因爲他能把我們不欲要的起伏減免了.若求兩個體的相關,用次數圖也可得出一個大概情形.要求精確的相關這次數圖是不夠用的.因爲圖只能表明一個大概情形,不能與我們一個數目的計量.相關度有多們大,用圖是表不出來的.須用相關係數(Correlation coefficient).什麼叫相關係數呢?相關係數就是主體對體間之相關度的數目計量.但是主體(Subject)和對體(Relative)又是什麼呢?主體的意思是說作爲標準的那件事物.對體就是與主體相比的事物,或以主體爲標準計量的事物.

係數 +1 表明完全的正相關.若係數是〇表明絕對沒有相關.若係數爲 -1,就是表明完全反相關.

112. 克爾披生的相關係數(Karl Peason's coefficient of correlation)

若是一件事的兩特質相比,如樹葉之長與其寬,或是研究一對相關聯的事,如父子的身高,夫妻的年齡,或

是研究兩種歷史個體的長期變遷的關係等等的問題,最好的相關係數,就是大生物學家克爾披生(Karl Peason)發明的相關係數.這種係數,就叫做克爾披生相關係數.

設 x_1, x_2, x_3……爲主體的各項自算術平均數的各離差. y_1, y_2, y_3……爲對體各項的相當離差. 設 σ_1 爲主體的均方差, σ_2 爲對體的均方差, n 爲成對項的項數,以 n 代表克爾披生的相關係數.於是得式如下:

$$r=\frac{\Sigma(xy)}{n\delta_1\delta_2}.$$

算這個係數的方法表明在下表裏邊,把相當的夫與妻的年齡都放上一直線上,這表裏邊的數目,是大多數項數中可爲標樣者的代表.

$$n=\frac{\Sigma(xy)}{n\delta_1\delta_2}=\frac{287}{20\times4.02\times4.17}=\frac{287}{335.27}=+.856$$

若把下面的表一加研究,就知道分子的大小可以決定是係數的大小. 假如主體對體全比他的平均數大,或全比平均數小,則 xy 必是正號的數. 若主體大於他的平均數對體小於平均數,則 xy 必是負號的數.若 xy 的值大多數是負數,則係數必小.若大多數全是正數,則其值必

用克爾披生相關係數求算夫妻之年齡

主體			對體			
h 夫年	x 自平均數之離差	x^2	w 妻年	y 自平均數之離差	y^2	xy
22	−8	64	18	−8	64	+64
24	−6	36	20	−6	36	+36
26	−4	16	20	−6	36	+24
26	−4	16	24	−2	4	+ 8
27	−3	9	22	−4	16	+12
27	−3	9	24	−2	4	+ 6
28	−2	4	27	+1	1	− 2
28	−2	4	24	−2	4	+ 4
29	−1	1	21	−5	25	+ 5
30	0		25	−1	1	6
30	0		29	−3	9	0
30	0		32	+6	36	0
31	+1	1	27	+1	1	+ 1
32	+2	4	27	+1	1	+ 2
33	+3	9	30	+4	16	+12
34	+4	16	27	+1	1	+ 4
35	+5	25	30	+4	16	+20
35	+5	25	31	+5	25	+25
36	+6	36	30	+4	16	+24
37	+7	49	32	+6	36	+42
平均30	$\Sigma x^2=$	324	平均26		$\Sigma y^2=348$	$\Sigma(xy)=+287$

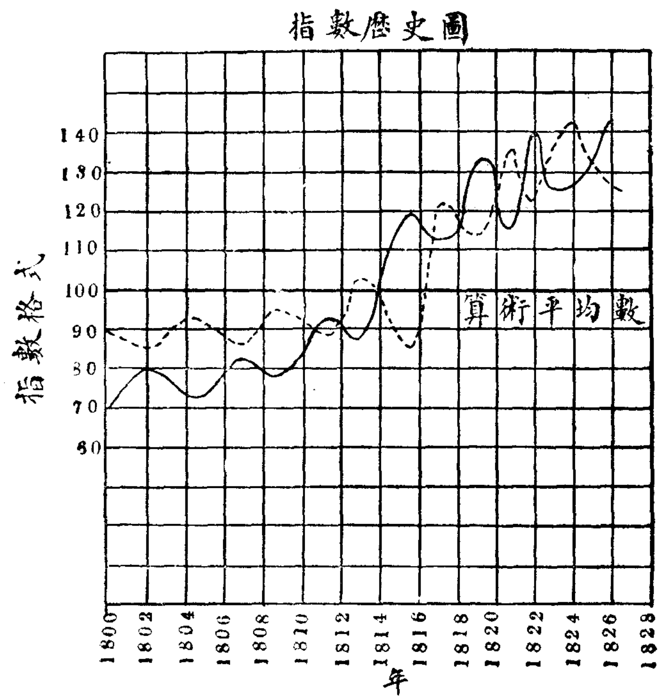
第　二　十　表

大.假如 xy 是十分一致的負數,這個係數就是代表一種高度的反相關.所以各對項對於平均數的比較,地位是要緊的,而與平均數的距離則不甚重要.

113. 克爾披生的係數對於歷史個體的長期變遷的應用

這種係數可用於某特別時候的一對事物上,也可用於歷史張本上.這係數的原來形狀,僅能用於長期變遷.例如二十二圖曲線 a 與曲線 b 都趨向與算術平均線相反的方向.所以他的係數是個負數.這係數的大小雖然也受短期起伏的影響,但是影響很小.因爲通過平均線的起伏很少的原故,而此短期之起伏,一見便知其爲反相關.

二十四圖表明一個圖其長期變遷是同一方向的,而短期起伏是反對的方向.用披生求相關係數的方法,得出的係數是一個很大正數.爲長期變遷自然很好,但是對於短期起伏的反對的關係毫不顧及了.求長期變遷的係數,即按照 112 節所述的方法,將同日期的項數與離差兩兩放在一起,然後按法求之.



第 二 十 四 圖

114. 克爾披生的係數對於短期起伏的應用

前邊已經講過披生的相關係數對於長期變化的應用；我們也可用同一的方法求出短期起伏的相關係數。例如，二十三圖裏邊的長期變遷已經完全減除了，所用的離差是從趨向算出來的，不是自算術平均數算出來的

用披生係數方法說明短期起伏的相關

年	供給				物價				y^2	xy
	供給指數	活動平均指數	自活動平均數之離差 x	x^2	物價指數	活動平均數	自活動平均數之離差 y			
1880	80				146					
1	82				140					
2	86	84	+2	4	130	133	− 3	9	− 6
3	91	85	+6	36	117	129	−12	144	−72
4	83	87	−4	16	133	124	+ 9	81	−36
5	85	89	−4	16	127	117	+10	100	−40
6	89	89	0	0	115	114	+ 1	1	0
7	96	91	+5	25	95	109	−14	196	−70
8	93	92	+1	1	100	104	− 4	16	− 4
9	90	93	−3	9	106	100	+ 6	36	−18
1890	91	94	−3	9	103	96	+ 7	49	−21
1	94	96	−2	4	94	89	+ 5	25	−10
2	100	98	+2	4	75	83	− 8	64	−16
3	105	99	+6	36	66	80	−14	196	−84
4	102	100	+2	4	75	79	− 4	16	− 8
5	96	101	−5	25	91	80	+11	121	−55
6	98	103	−5	25	87	82	+ 5	25	−25
7	106	105	+1	1	81	83	− 2	4	− 2
8	114	108	+6	36	76	83	− 7	49	−42
9	112	109	+3	9	82	86	− 4	16	−12
1900	109	111	−2	4	91	88	+ 3	9	− 6
1	106	112	−6	36	100	88	+12	144	−72
2	112	113	−1	1	89	88	+ 1	1	− 1
3	120	114	+6	36	76	89	−13	169	−78
4	118	114	+4	16	82	91	− 9	81	−36
5	112	113	−1	1	100	96	+ 4	16	− 4
6	110	112	−2	4	106	101	+ 5	25	−10
7	107				114					
8	113				103					
	平均100		總和	358	平均100		總和	1,593	−728

表中各數止於整數

第二十一表

就用這離差得 y, x 格中的數，再將這些離差平方之，就得出 $\sigma_1 \sigma_2$. 因此，第十九表卽變成第二十一表.用第二十一表中的數目便得

$$\sigma_1=\sqrt{\frac{\Sigma x^2}{n}}=\sqrt{\frac{358}{25}}=\sqrt{14.32}=3.78$$

$$\sigma_2=\sqrt{\frac{\Sigma y^2}{n}}=\sqrt{\frac{1593}{25}}=\sqrt{63.72}=7.98$$

$$n=\sqrt{\frac{\Sigma(xy)}{n\sigma_1\sigma_2}}=\frac{-728}{25\times3.78\times7.98}=-.965$$

前表自 1882 至 1906 共為二十五年，所以 n =25. 卽用之以求係數.指數，活動平均數，離差等數僅止於整數，要欲更精確，則不可不算出小數位來.上邊的 -.965 相關係數表明物價供給有很高度的反相關.就是說若是供給增加，則物價同時必然下落；若是供給減少，同時物價必然上昇.

115. 同流離差的係數(The Coefficient of Concurrent Deviation)

如以上所說將披生係數稍為變易，固然是一個比較短期起伏的極好的方法.但是，算的時候太麻煩，須費很多的時間和精力，纔能算得出來這個係數.現在有一種

極簡單的求相關係數的方法，而且這種方法，平常大多數的短期起伏，都可用以得出一滿意的結果來。可惜對於長期變遷是完全無用的。因為長期變遷是要一個普遍的趨向。而這種方法所得的結果，一點也不能表出普遍的趨向，所能表明者僅短期起伏而已。

比較兩條歷史曲線的時候，假如這兩條曲線同時趨向同一方向，我們就說這短期起伏之中有一個正相關。假如這兩條線同時趨向相反的方向，我們就說他有一個反相關。兩線同時趨向同方向就是離差的同流。同時趨向相反的方向就是離差的分流。算這種離差的時候，不是自算術平均數算出，也不是自活動平均數算出，僅從他的前一項算出。其算法也與從前大不相同，只算離差的方向；不算離差的大小。若是這一項比前一項大，就是正方。若是這一項比前項小，就是負方。這種算法有一種毛病，就是不問兩數的相差大小全是一律看待。有時前一年僅僅有一個小小的變動，而變動的原因也不是主要的原因，但是假如若是相同的方向（同為正或同為負），我們算的時候，就不管他相差的大小，都是一律看待。只

要他是這一個方向，大差是這個方向，小差也是這個方向，大小是沒有分別的.實在，我們所要的，是主要原因生出的離差，若把不重要的原因所生的離差也照樣來算，那就不對了.但是若項數極多，錯誤互相補償的機會大增，這種錯誤也不十分利害了.

下邊有一個由經驗中得來的公式，由這公式中得出來的結果之解釋和前邊的公式相仿.就是也用 $+1$ 表示完全的正相關；-1 表示完全的反相關；〇表示毫不相關.

設　　$r=$相關係數

　　　$n=$成對項的項數

　　　$c=$同流離差的數

於是

$$r=\pm\sqrt{\pm\frac{2c-n}{n}}$$

此處用的正負號要稍爲解釋一下子.假如 $\frac{2c-n}{n}$ 等於負數就用(—)號.根號以外也用負號.這是因爲要他開得方根以後，他的符號仍然一致不變.

用同流離差求供給物價之短期起伏的相關

年	供給		物價		xy積
	供給指數	與前年的離差(x)	物價指數	與前年的離差(y)	
1880	80		146		
1	82	+	149	−	−
2	86	+	130	−	−
3	91	+	117	−	−
4	83	−	133	+	−
5	85	+	127	−	−
6	89	+	115	−	−
7	96	+	95	−	−
8	93	−	100	+	−
9	90	−	106	+	−
1890	91	+	103	−	−
1	94	+	94	−	−
2	100	+	75	−	−
3	105	+	66	−	−
4	102	−	75	+	−
5	96	−	91	+	−
6	98	+	87	−	−
7	106	+	81	−	−
8	114	+	76	−	−
9	112	−	82	+	−
1900	109	−	91	+	−
1	106	−	100	+	−
2	112	+	89	−	−
3	120	+	76	−	−
4	118	−	82	+	−
5	112	−	100	+	−
6	110	−	106	+	−
7	107	−	114	+	−
8	113	+	103	−	−

第二十二表

這種係數的離差可用第二十一表中的數目作成第二十二表,在二十二表中.

$$n=28 \qquad c=o$$

於是

$$r=\pm\sqrt{\pm\frac{2c-n}{n}}$$

$$r=\pm\sqrt{\pm\frac{0-28}{28}}$$

$$r=-\sqrt{-(-1)}$$

$$r=-\sqrt{1}$$

$$r=-1$$

所以爲完全的反相關.

我們再假設一例.譬如說有 48 個時間.於是 n 等於 47.有 16 對離差是同流的.於是其式便爲

$$r=\pm\sqrt{\pm\frac{2c-n}{n}}=\pm\sqrt{\pm\frac{32-47}{47}}$$

$$=-\sqrt{-\frac{-15}{47}}$$

$$=-\sqrt{.3191}$$

$$=-.56.$$

這個結果表明反相關僅有一中等的度數。

用同流離差求係數的方法，除了簡易而外，還有一個優點，就是他能用於起伏極不規則的曲線.圖上.如十八圖他的不規則的狀態，幾不能用活動平均數，求出一個緩和狀態來，用這種方法便行了.

前邊表中的數目，我們所用的是指數.這不過爲說明方便起見罷了.其實，求係數時，無論是用披生方法，或是同流離差，把張本作成指數，是全然無用的，而且引出一點算學的錯誤來.

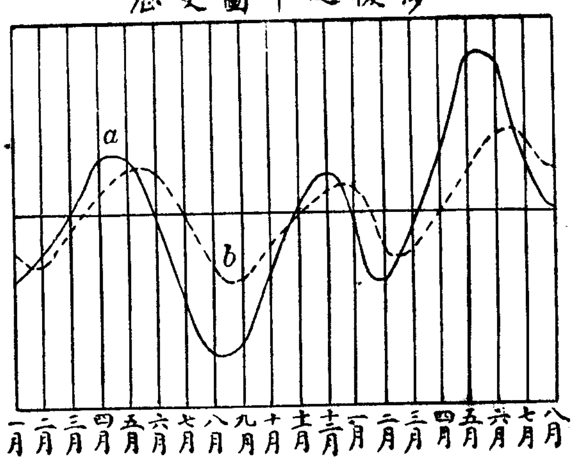
116.　前移的用途(The use of the lag)

比較歷史圖，我們常常看見兩件事中間，顯然有一種相關，但是這兩條曲線的起伏，並不相對.這種現象，大概由於從原因到結果，須經過相當的時間.例如失業爲貧困的原因，貧困又爲極貧的原因.但是由失業到極貧，必須經過一個相當的時間.

假如我們的確知道某一現象是別一現象的原因，更知道這原因結果之間，必須經過一個相當的時間.但是兩條曲線不相對.於是最好將代表結果那條曲線充分的

向前移一下子,好使兩曲線的波曲相對.這就是說,相對的日期不必定是同日期.但是須前移多少呢?自然要看他的性質而定.試以例說明之.二十五圖裏邊顯然有一種相關,但是曲線b比曲線a約落後一個月,若算相關係數時,須將a線的五月與b線的六月作成一對,a線的六月與b線的七月作成一對,七月對八月,八月對九月,依此類推,便可知這兩件事的關係了.

歷史圖中之後移



第　二　十　五　圖

但是萬不可僅看圖上的起伏表現便隨便移動,將因果倒置了.譬如把b線往前移動得出一個高度的正相關

若把 a 線向後移動，卻得出一個顯著的反相關．所以我們必須由圖外找消息,以決定普通的相關時間是多們長,而用這曲線不過得出那相當時間的長罷了.前移能用於作比較歷史圖，也可用於算相關係數．惟以必要時爲限；不必要時不可濫用．如二十五圖二十二表卽可用前移了．

117. 或然的錯誤(The Probable Error)

假如兩條曲線只有兩三處起伏相應,決不能就說他們有一種相關．若是我們冒然說他有相關，那就和說擲一對骰子每次是雙六，連續三次，於是便說這兩個骰子有一定的關係．這是同一樣的不真確．這種雙六完全是偶然發生機遇的現象，毫不足爲憑．假如若雙六連成十次，我們便可推想這裏除機遇而外，還有一種力量在那裏動作.所以若是兩條曲線的起伏，在許多許多的時期，都有一種關係,我們就可相信他們大概有一種真正的關係．

例如 100 對離差中有 55 對是同流的，45 對是分流的，我們就想這不等的狀況完全是機遇現象．若是有 70

對是同流的; 30 對是分流的,這就不盡是機遇了.

所以相關係數的或然錯誤與對項項數成反比例,與係數的大小也成反比例.算學家已將或然的錯誤律(the law of probable error) 審慎的定出.下邊公式的證明,似出本書範圍之外,茲從略.僅書其式.

設 r=相關係數

n=對項項數

$$\text{或然的錯誤}=\frac{.67(1-r^2)}{\sqrt{n}}$$

這個公式,是假定純粹機遇張本,僅用於小係數的錯誤或項數很少的數目,至於組的大小或項形之規則與否,那是毫無關係的.就是說相關係數須是:

$$r=\pm\frac{.67(1-r^2)}{\sqrt{n}}$$

而 r 之值必在

$$r+\frac{.67(1-r^2)}{\sqrt{n}}\text{與}r-\frac{.67(1-r^2)}{\sqrt{n}}\text{之間.}$$

118. 相關係數的解釋

下邊各規律,是按照相關係數與或然錯誤的關係,以解釋相關的程度.

1. 如 r 小於或然錯誤,則絕無相關.

2. 如 r 大於或然錯誤六倍, 則可決然決定有相關存在.

下邊兩條是說明或然錯誤很小的時候的情形.

1. 若 r 小於 0.30 時,不能斷定有顯然的相關.

2. 若 r 大於 0.50 時,定然有相關.

參考書

Elderton, W. P. and E. M.: Primer of Statistics, Chaps. V and VI.

Pearson, Karl: The Grammar of Science, London, 1900, pp. 381–92.

Bowley, A. L.: Elements of Statistics, Part II, esp. pp. 315–334.

Yule, J. V.: Introduction to Statistics, Chaps. IX–XII and XVI.

Bowley, A. L.: Measurement of Groups and Series, Lectures V and VI.

"Student": On the Probable Error of a Correlation Co-efficient, Biometrika, 1908, VI, 302.

第十八章　變動比率(The ratio of variation)

119. 變動比率的定義

在 110 節裏邊已經說過,假如若能把額外的影響與非原因的分子一律完全免除,則原因結果永久有一種算學上的關係.我們也說過,若把額外的影響完全免除,那是不可能的,所能免除者,不過其主要者,所能得的算學上的關係,也不過是逼近精確.可能的關係有種種,吾人所用者僅一而已.

在第九章,我們已經知道某物的項數,常趨於某一特別形狀,也就是說趨於範數,如範數的樹葉,範數的身高,平常的麥價.範數的位置常近於算術平均數,但算術平均數比範數有定,故平常多用算術平均數爲根據,若範數能有一定時,也可以用範數.

兩曲線往往起伏相應,並不能就說他們的自平均數的離差,絕對的或比例的相同.換一句話說,就是起伏的波長可以相等,而其起伏的高度未必一致.譬如鐘擺的搖動,兩擺搖動,所用的時間雖然相等,而其搖動的弧度可以大不相等.一個搖成 5° 的弧,那一個或爲 20° 的弧.第二十三圖就可以說明這個意思.a b 二線的波長雖然

一樣,而其高度則不等.代表物價的 a 線,比代表供給的 b 線差不多高兩倍.這種物價離差,最好是代表沒有彈性的東西的價格,但是我們還要知道兩曲線,自其平均數的比例離差的平均比率,究竟是多少.一個買賣糧米投機的人,極願意知道,假如平常的收穫若減少百分之十,那們穀價該漲多少呢?社會改造家,或者願意知道,假如賣酒的鋪子若減少一半,該少賣出多少酒呢?社會學家一定很喜歡知道,假如把父母結婚的年齡增加了,對於生產率該有多們大的影響呢?這種關係叫做變動比率(the ratio of variation).因為變動比率是相關係數的一個系,所以他們很有關係.也常和相關係數相混,其實二者的性質是大不相同的.

二十三圖裏邊因為供給的比例起伏平均起來,差不多是物價起伏的一半,所以變動比率差不多是 0.50.

120. 變動比率的算法

前節已經說過,我們要求對體之自平均數的離差的平均率,以便與主體者相比較.第一步先要決定那一件是主體?那一件是對體?在生物學中孰為主體,孰為對體,

本沒有大關係的，任便選那一件都可以.研究社會的科學則取平均比例離差較大者爲主體，小者爲對體.我們所以要這樣選擇的原故，是要這比率小於一.本着這個道理，那們二十三圖中的物價曲線 a，自然定爲主體，供給曲線 b 爲對體.

歷史級數怎麼算法呢?歷史級數的變動率，是先把主體對體的各期的自平均數離差都求出來,再以主體的各離差,除對體的各離差,除得的商相加,再以商的項數除之.所得的結果就是平均率.譬如在某豎線上主體的離差是百分之 16，對體的離差是百分之 4，其商當是0.25.若主體的離差爲－10,對體的離差是＋2,其商當是－0.20.此結果不但不加,而且倒是須從總和中減了.所以若是曲線的起伏，是完全的規則的起伏，這種方法還可應用.否則,須用更好的方法較爲妥適.實際上這個方法常是不好用，所以須用別種方法.別種方法的的最好者就是葛爾頓圖 (Galton graph).葛爾頓圖名的由來，因爲創製這種圖者是葛爾頓教授 (Professor Francis Galton)發明的.因是得名.

121. 葛爾頓圖(The Galton graph)

用這種圖於兩種歷史個體的時候,第一步須把他們作成指數.若是研究長期的變遷,或者一個圖沒有顯明的趨向的時候,就用算術平均數除各項就得出指數了.得出指數後,於是以主體爲縱距,以對體爲橫距.將指數兩兩成對的作成圖.試以例說明之.

下邊的表,是美國1880年至1896年的銀行清算和移民的統計.二十六圖就是把這表上的數目應用於葛爾頓圖的情形.銀行清算是一種財產的測量,但是商業對於移民的影響須發生在次一年.所以須用116節所講的前移(Lag)方法,將移民張本向前移一年.使1881年的移民與1880年的銀行清算相比較,因爲移民指數的起伏,大於銀行清算的起伏,所以定移民爲主體,銀行清算爲對體.

122. 變動比率

在葛爾頓圖中,最好是把主體劃在縱線上,把對體劃在橫線上.用下表的數目按其主體對體縱橫劃上.於是得出很多的點.這些點有些相距很近,有些相距很遠,

用葛爾頓圖表明銀行清算與移民的變動比率之張本

主體			對體		
年	移民(萬)	移民指數	年	銀行清算(億)	清算指數
1881	67	136	1880	37	106
1882	79	161	1881	49	140
1883	60	122	1882	47	134
1884	52	106	1883	40	114
1885	40	81	1884	34	97
1886	33	67	1885	25	71
1887	49	100	1886	33	94
1888	55	112	1887	35	100
1889	44	90	7888	31	89
1890	46	94	1889	35	100
1891	56	114	1890	38	109
1892	62	126	1891	34	97
1893	50	102	1892	36	103
1894	31	63	1893	34	97
1895	28	57	1894	24	69
1896	34	69	1895	28	80
	平均49.1			平均35.0	

第二十三表

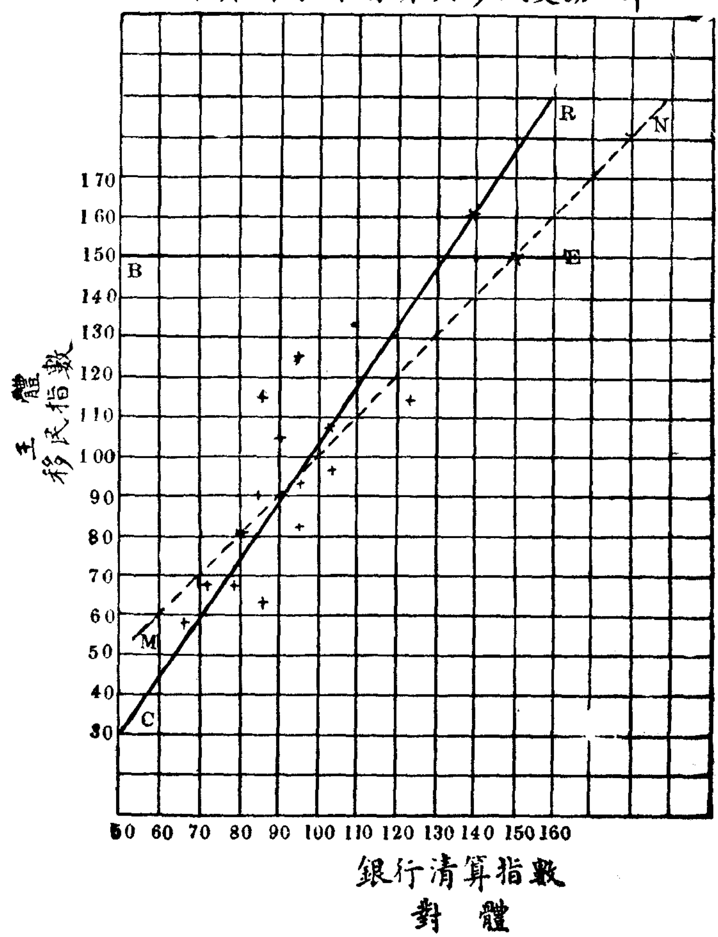
普通大體上說來衆點成隊的走向左下方.第二步手續便是按照這些點的普遍趨向劃一直線.劃的時候用目觀察.要使線的兩邊的點數相等，於是劃出的線，便是正確的方向了.

二十六圖中的 RC 差不多和這結果相近.假如他的相關是完全的，每一點必落在這直線上，或落在別一個極一定的曲線上.平常主體與對體的存在關係常是如此.若是他能正確，一定這個圖常是曲線，不是直線.但是實際上，各點既然若晨星一樣的散在各處，與直線既不接近，與曲線也不接近.也不能說曲線定然比直線好多少.所以通常都是劃直線，不劃曲線.更因爲依葛爾頓算變動率的方法只是用直線纔行.

二十六圖中的各點，既然那樣分散，故潛伏相關中的外力影響，正確變動率一樣都是顯然可見的.

下邊各規律應用於一般的葛爾頓圖.若點線向左下邊斜下，他的相關是正相關.若是點線向右下方斜下，他的相關是反相關.若是各點極其分散，亂列各處，沒有一個一定的趨向.不能劃一直線，也不能劃一個有規則的

用萬爾頓圖表明1880至1896年美國銀行清算與移民變動比率



第二十六圖

曲線.這是沒有相關.若是各點愈近於直線,則相關係數愈大.

若是主體增加百分之一,對體也增加(或減少)百分之一.主體減少百分之一,對體也減少(或增加)百分之一.他的變動率顯然的就是一.若變動率是一的時候,各點將形成一個45°的斜面,如第二十六圖中的MN,便是這相等比例變動線.但是對體變動比主體變動小,則其與縱線所成之角必小於45°.有這種情形的時候,這條線叫做退歸線(regression line).退歸線得名的由來,是因為研究生物學上的遺傳的時候,知道父母不能把他所有的特質,完全無遺的傳給他的子嗣.遺傳力不能全量的遺傳.例如父母比平常人高一吋,他的兒女的身高大概不能再比平常人高一吋,必要小於一吋.換一句話說,就是要退歸和大多數人一樣高.所以退歸線因此得名.約略言之,若是退歸的程度愈大,相關的程度愈小.就是說RC若離MN愈遠,則相關愈小.若是RC幾乎成為直線了,那就是說對體幾乎不受主體的影響了.例如研究高父親所生的兒子是不是也高,結果或者高一點.其最大的或

然數，是大概這種現象是完全偶然的.若平均起來，高父親的兒子並不比平常人的兒子高.

我們更要記着退歸線是由很少幾件事得出來的,並不十分精確.若或然錯誤很大的時候，實際上求變動率便沒有用了.

設 BE 爲任一横線與退歸線交於 A 點，顯然的對體平均變動與主體平均變動之比爲 AB/BC,或 ACB 切角.二十六圖 AB=82,BC=122,所以他的變動率等於 82/122=0.67.這就是說，若主體有百分之一的變動，對體就有百分之一的 67/100 的變動.相差那 0.33,叫做退歸率.

123. 長期變動的免除

如求相關係一樣，算短期起伏的變動率，必須先把長期變遷的趨向免除了,然後纔可以研究短期起伏的關係，其原理和求相關係數之免除長期變遷是一樣的.惟求法略有不同.第一步先要決定是不是有一種可用活動平均數表出一個趨向的性質.若是沒有這種性質，那們只好用算術平均數求離差了.(見 121 節)若是有一個一定的趨向，短期起伏的波長便決定了，第二步便是算出

活動平均線．第三步就是每項用那年活動平均數除之，則得出兩組指數級數．於是按照122節所用的方法兩數成對作圖．試以下邊假設的表說明之．

用葛爾頓圖表明銀行準備金與支票流轉短期變化的變動率之張本

年	主體			對體		
	B 銀行準備金平均數	b 活動平均數	$\frac{B}{b}$ 指數	C 支票流轉平均數	c 活動平均數	$\frac{C}{c}$ 指數
1880	8.2			27		
1881	8.0			26		
1882	7.2	7.7	94	22	26	85
1883	7.4	7.6	97	25	27	93
1884	7.7	7.5	103	30	28	107
1885	7.7	7.5	103	32	29	110
1886	7.5	7.4	101	31	30	103
1887	7.2	7.3	99	27	31	87
1888	6.9	7.3	95	30	32	94
1889	7.2	7.3	99	35	33	106
1890	7.7	7.2	107	37	35	106
1891	7.5	7.1	106	36	37	97
1892	6.7	7.0	96	37	38	97
1893	6.4	6.9	93	40	39	103
1894	6.7	6.7	100	40	40	100
1895	7.2	6.6	109	42	40	105
1896	6.5	6.5	100	41	39	105
1897	6.2			37		
1898	5.9			35		

第二十四表

上邊表中，每人平均銀行公積金一致的趨於減少，同時每人平均支票的流轉一致的增加.但是當着平常支票的流轉每人平均爲$26時,每變化一元並不代表相同的比例. 當着平常流轉增至$39 時,就代表相同的比例了. 所以若得正確的比例變動,須先得出每級數的活動平均數. 於是每點都有了正當的根據,若是原來的項用活動平均數除之,則此指數卽能表明由短期起伏所生的比較增加或減少了. 每對的指數,卽可用二十六圖所示的方法作圖.

124. 相關表(The correlation table)

當着主體對體的項數都是很多的時候,若把每項都作成指數,自然是極麻煩的事. 所以有相關表以補救這種困難. 相關表對於生物學功用很大. 因爲生物學上所有的數目項數常是極多的,若沒有相關表那就太麻煩了. 相關表的目的,是要把主體的項數作成次數表,再求出對體中的相當項數的某種平均數,以與主體的中數相比. 這種方法簡直就是用平均數代替各項罷了,其法較爲簡單. 下邊是一個最簡單的相關表以說明其用法. 這表是

E便取一羣樹葉,用樹葉之長爲主體,其寬爲對體,“主體”格中的前兩格是葉長的簡易次數表,右邊在“對體”格中的與各組相對的葉寬,就是那一組的葉寬.從後邊第二格裏邊是那組的所有的樹葉的平均寬,就用這個平均寬和那組的中數長得出變動率.要作成葛爾頓圖,自然要把各張本先作成指數.

作相關表有須注意之處現說明如下.

主體的各組距必須相等,對體的組距也是如此.自最小數直到最大數中間的各組,全要寫在表上,卽是沒有次數的組,也要照樣寫上.通常起草相關表,是先按照葉長順次排列,看這樣的長的葉他的寬各如何,卽在相當的格內劃點以代表之.如寬爲15糎,卽在15糎格中對着他的長劃一個點.若是16糎,卽放入16糎的格裏.然後數這點有多少,卽是他的次數填入正式表裏的相當的格中右邊的總和格裏的數目,不過只作爲與左邊第一格裏的數目一種對照罷了.

變動率的理想中最好的平均數是範數;因爲範數是離差的眞正中心.但是範數常常不甚準確,所以範數爲

求算葉之長寬相關表

主體			對體											
葉數	葉長種	以平均長除各組中數長	葉寬種 13—15	16—18	19—21	22—25	26—28	29—31	32—34	35—37	38—40	總數	平均寬	以平均寬除各組的平均寬
3	30—36	51.3	2	1								3	15.0	56.2
5	37—43	62.3	1	2	2							5	17.6	66.0
11	44—50	73.1		1	4	4	1	1				11	22.7	35.1
9	51—57	84.0			3	4	2					9	23.3	87.3
16	58—64	94.9		2	2	4	4	3	1			16	25.1	94.1
26	65—71	105.7		1	2	5	8	4	3	2	1	20	27.8	104.2
14	72—78	116.6					4	7	2	1		14	30.0	112.4
10	79—85	127.5				1	1	2	3	2	1	10	32.1	120.3
4	86—91	138.4							1	2	1	4	30.0	134.9
1	92—98	147.7									1	1	39.0	146.1
99		總和	3	7	13	18	20	17	10	7	4	99		
平均	64・31											平均	26.67	

第二十五表

根據，倒不如以算術平均數爲根據．算術平均數永是準確的，範數則不甚準確．項數很少的時候，更不準確．實際上應用多用算術平均數，上邊的表就是用算術平均數爲根據．

寶來(Bowley)說，用中數爲根據，是極其方便也很精確的方法．若是項數很多的時候，中數固然是最好的根據．若是項數分散凌亂極不一致，中數便不能用了．若是組距很大，須用71節所述的公式按法求出中項．若組距很小，卽用該組的中點爲中數．用中點爲中數，雖然微有錯誤，但是平常也够精確了．

用算術平均數時(如二十五表)各組內的樹葉長寬，都是以中點作爲他的長寬．例如44－50這一組我們就以中點47粴爲各葉之長．就是說我們把這組裏邊的葉都當爲47粴．第三格裏的指數，就是用全體葉的平均長除各組的中點得出來的．用一樣的法子，用全體葉的平均寬除各組寬的中點，就得出寬的指數了．

要把這表稍加研究，不但能用他決定變動率，也可知道主動被動間相關的大概情形．假如若是有相關的，

那們長的樹葉也該寬.於是使範數寬一致的向右,範數在表上作成45°的角,就是表示有相關存在.上邊的表範數斜向右下方,這就是表明有正相關的意思.若是他斜向左下方,那就是表明有反相關的意思.範數若愈能組成直線形,各項愈貼近範數,則相關度愈大.如有相關,則表底邊的和,必現出一個準確的範數,各和相加的總和,正可爲表中數目的一個對照.

幻想相關表,是一塊平面地.放入被動方格中的項數,按他點的多少,代表地面上的高底.於是便有一個亂山圖.表的中間是山頂.這山的表面叫做相關面(correlation surface).若是相關是完全的,山底必成卵形,山的中心在表的中心,山的表面是緩和的,向各方的斜面都是規則一律的.無論自那一點,將這山縱剖開都成一個鐘形,常態次數曲線.

相關表作完後的第一步,是作成葛爾頓圖.用主動每組的指數爲縱距,被動每組的指數爲橫距.再畫退歸線.最後按照122節所述的算變動率的方法求出變動率.

125. 結論

本書除了略述一點統計學的歷史外,所有統計方法的主要各點,已經講過了.關於衆數的蒐集,衆數的分析,衆數的比較,都講過了.我們相信這些事是重要的,對於社會科學更是重要.所以都約略的講過了.至於統計學的詳盡的討論,算學原理的演證,以及統計學真實的應用,要請讀者百尺竿頭更進一步.

參考書

Galton, Francis: Correlations and their Measurement, Proceedings of the Royal Society, 1888, XIV, 135.

Bowley, A. L.: Elements of Statistics, pp. 322-326.

Yale, G. U.: Introduction to Statistics, Chap. IX, and pp. 203-205.

Persons, Warren M.: The Correlation of Economic Statistics, Quar. Publications of the Amer. Statistical Association.

附錄一
第二十五表
對數表

No.	0	1	2	3	4	5	6	7	8	9
10	0000	0043	0086	0128	0170	0212	0253	0294	0334	0374
11	0414	0453	0492	0531	0569	0607	0645	0682	0719	0755
12	0792	0828	0864	0899	0934	0969	1004	1038	1072	1106
13	1139	1173	1206	1239	1271	1303	1335	1367	1399	1430
14	1461	1492	1523	1553	1584	1614	1644	1673	1703	1732
15	1761	1790	1818	1847	1875	1903	1931	1959	1987	2014
16	2041	2068	2095	2122	2148	2175	2201	2227	2253	2279
17	2304	2330	2355	2380	2405	2430	2455	2480	2504	2529
18	2553	2577	2601	2625	2648	2672	2695	2718	2742	2765
19	2788	2810	2833	2856	2878	2900	2923	2945	2967	2989
20	3010	3032	3054	3075	3096	3118	3139	3160	3181	3201
21	3222	3243	3263	3284	3304	3324	3345	3365	3385	3404
22	3424	3444	3464	3483	3502	3522	3541	3560	3579	3598
23	3617	3636	3655	3674	3692	3711	3729	3747	3766	3784
24	3802	3820	3838	3856	3874	3892	3909	3927	3945	3962
25	3979	3997	4014	4031	4048	4065	4082	4099	4116	4133
26	4150	4166	4183	4200	4216	4232	4249	4265	4281	4298
27	4314	4330	4346	4362	4378	4393	4409	4425	4440	4456
28	4472	4487	4502	4518	4533	4548	4564	4579	4594	4609
29	4624	4639	4654	4669	4683	4698	4713	4728	4742	4757

30	4771	4786	4800	4814	4829	4843	4857	4871	4886	4900
31	4914	4928	4942	4955	4969	4983	4997	5011	5024	5038
32	5051	5065	5079	5092	5105	5119	5132	5145	5159	5172
33	5185	5198	5211	5224	5237	5250	5263	5276	5289	5302
34	5315	5328	5340	5353	5366	5378	5391	5403	5416	5428
35	5441	5453	5465	5478	5490	5502	5514	5527	5539	5551
36	5563	5575	5587	5599	5611	5623	5635	5647	5658	5670
37	5682	5694	5705	5717	5729	5740	5752	5763	5775	5786
38	5798	5809	5821	5832	5843	5855	5866	5877	5888	5899
39	5911	5922	5933	5944	5955	5966	5977	5988	5999	6010
40	6021	6031	6042	6053	6064	6075	6085	6096	6107	6117
41	6128	6138	6149	6160	6170	6180	6191	6201	6212	6222
42	6232	6243	6253	6263	6274	6284	6294	6304	6314	6325
43	6335	6345	6355	6365	6375	6385	6395	6405	6415	6425
44	6435	6444	6454	6464	6474	6484	6493	6503	6513	6522
45	6532	6542	6551	6561	6571	6580	6590	6599	6609	6618
46	6628	6637	6646	6656	6665	6675	6684	6693	6702	6712
47	6721	6730	6739	6749	6758	6767	6776	6785	6794	6803
48	6812	6821	6830	6839	6848	6857	6866	6875	6884	6893
49	6902	6911	6920	6928	6937	6946	6955	6964	6972	6981
50	6990	6998	7007	7016	7024	7033	7042	7050	7059	7067
51	7076	7084	7093	7101	7110	7118	7126	7135	7143	7152
52	7160	7168	7177	7185	7193	7202	7210	7218	7226	7235
53	7243	7251	7259	7267	7275	7284	7292	7300	7308	7316
54	7324	7332	7340	7348	7356	7364	7372	7380	7388	7396

No.	0	1	2	3	4	5	6	7	8	9
55	7404	7412	7419	7427	7435	7443	7451	7459	7466	7474
56	7482	7490	7497	7505	7513	7520	7528	7536	7543	7551
57	7559	7566	7574	7582	7589	7597	7604	7612	7619	7627
58	7634	7642	7649	7657	7664	7672	7679	7686	7694	7701
59	7709	7716	7723	7731	7738	7745	7752	7760	7767	7774
60	7782	7789	7796	7803	7810	7818	7825	7832	7839	7846
61	7853	7860	7868	7875	7882	7889	7896	7903	7910	7917
62	7924	7931	7938	7945	7952	7959	7966	7973	7980	7987
63	7993	8000	8007	8014	8021	8028	8035	8041	8048	8055
64	8062	8069	8075	8082	8089	8096	8102	8109	8116	8122
65	8129	8136	8142	8149	8156	8162	8169	8176	8182	8189
66	8195	8202	8209	8215	8222	8228	8235	8241	8248	8254
67	8261	8267	8274	8280	8287	8293	8299	8306	8312	8319
68	8325	8331	8338	8344	8351	8357	8363	8370	8376	8382
69	8388	8395	8401	8407	8414	8420	8426	8432	8439	8445
70	8451	8457	8463	8470	8476	8482	8488	8494	8500	8506
71	8513	8519	8525	8531	8537	8543	8549	8555	8561	8567
72	8573	8579	8585	8591	8597	8603	8609	8615	8621	8627
73	8633	8639	8645	8651	8657	8663	8669	8675	8681	8686
74	8692	8698	8704	8710	8716	8722	8727	8733	8739	8745

75	8751	8756	8762	8768	8774	8779	8785	8791	8797	8802
76	8808	8814	8820	8825	8831	8837	8842	8848	8854	8859
77	8865	8871	8876	8882	8887	8893	8899	8904	8910	8915
78	8921	8927	8932	8938	8943	8949	8954	8960	8965	8971
79	8976	8982	8987	8993	8998	9004	9009	9015	9020	9025
80	9031	9036	9042	9047	9053	9058	9063	9069	9074	9079
81	9085	9090	9096	9101	9106	9112	9117	9122	9128	9133
82	9138	9143	9149	9154	9159	9165	9170	9175	9180	9186
83	9191	9196	9201	9206	9212	9217	9222	9227	9232	9238
84	9243	9248	9253	9258	9263	9269	9274	9279	9284	9289
85	9294	9299	9304	9309	9315	9320	9325	9330	9335	9340
86	9345	9350	9355	9360	9365	9370	9375	9380	9385	9390
87	9395	9400	9405	9410	9415	9420	9425	9430	9435	9440
88	9445	9450	9455	9460	9465	9469	9474	9479	9484	9489
89	9494	9499	9504	9509	9513	9518	9523	9528	9533	9538
90	9542	9547	9552	9557	9562	9566	9571	9576	9581	9586
91	9590	9595	9600	9605	9609	9614	9619	9624	9628	9633
92	9638	9643	9647	9652	9657	9661	9666	9671	9675	9680
93	9685	9689	9694	9699	9703	9708	9713	9717	9722	9727
94	9731	9736	9741	9745	9750	9754	9759	9763	9768	9773
95	9777	9782	9786	9791	9795	9800	9805	9809	9814	9818
96	9823	9827	9832	9836	9841	9845	9850	9854	9859	9863
97	9868	9872	9877	9881	9886	9890	9894	9899	9903	9908
98	9912	9917	9921	9926	9930	9934	9939	9943	9948	9952
99	9956	9961	9965	9969	9974	9978	9983	9987	9991	9996

附錄二

第二十六表

開方表

No.	0	1	2	3	4	5	6	7	8	9
10	10000	10201	10404	10609	10816	11025	11236	11449	11664	11881
11	12100	12321	12544	12769	12996	13225	13456	13689	13924	14161
12	14400	14641	14884	15129	15376	15625	15876	16129	16384	16641
13	16900	17161	17424	17689	17956	18225	18496	18769	19044	19321
14	19600	19881	20164	20449	20736	21025	21316	21609	21904	22201
15	22500	22801	23104	23409	23716	24025	24336	24649	24964	25281
16	25600	25921	26244	26569	26896	27225	27556	27889	28224	28561
17	28900	29241	29584	29929	30276	30625	30976	31329	31684	32041
18	32400	32761	33124	33489	33856	34225	34596	34969	35344	35721
19	36100	36481	36864	37249	37636	38025	38416	38809	39204	39601
20	40000	40401	40804	41209	41616	42025	42436	42849	43264	43681
21	44100	44521	44944	45369	45796	46225	46656	47089	47524	47961
22	48400	48841	49284	49729	50176	50625	51076	51529	51984	52441
23	52900	53361	53824	54289	54756	55225	55696	56169	56644	57121
24	57600	58081	58564	59049	59536	60025	60516	61009	61504	62001
25	62500	63001	63504	64009	64516	65025	65536	66049	66564	67081
26	67600	68121	68644	69169	69696	70225	70756	71289	71824	72361
27	72900	73441	73984	74529	75076	75625	76176	76729	77284	77841
28	78400	78961	79524	80089	80656	81225	81796	82369	82944	83521
29	84100	84681	85264	85849	86436	87025	87616	88209	88804	89401

30	90000	90601	91204	91809	92416	93025	93636	94249	94864	95481
31	96100	96721	97344	97969	98596	99225	99856	100489	101124	101761
32	102400	103041	103684	104329	104976	105625	106276	106929	107584	108241
33	108900	109561	110224	110889	111556	112225	112896	113569	114244	114921
34	115600	116281	116964	117649	118336	119025	119716	120409	121104	121801
35	122500	123201	123904	124609	125316	126025	126736	127449	128164	128881
36	129600	130321	131044	131769	132496	133225	133956	134689	135424	136161
37	136900	137641	138384	139129	139876	140625	141376	142129	142884	143641
38	144400	145161	145924	146689	147456	148225	148996	149769	150544	151321
39	152100	152881	153664	154449	155236	156025	156816	157609	158404	159201
40	160000	160801	161604	162409	163216	164025	164836	165649	166464	167281
41	168100	168921	169744	170569	171396	172225	173056	173889	174724	175561
42	176400	177241	178084	178929	179776	180625	181476	182329	183184	184041
43	184900	185761	186624	187489	188356	189225	190096	190969	191844	192721
44	193600	194481	195364	196249	197136	198025	198916	199809	200704	201601
45	202500	203401	204304	205209	206116	207025	207936	208849	209764	210681
46	211600	212521	213444	214369	215296	216225	217156	218089	219024	219961
47	220900	221841	222784	223729	224676	225625	226576	227529	228484	229441
48	230400	231361	232324	233289	234256	235225	236196	237169	238144	239121
49	240100	241081	242064	243049	244036	245025	246016	247009	248004	249001
50	250000	251001	252004	258009	254016	255025	256036	257049	258064	259081
51	260100	261121	262144	263169	264196	265225	266256	267289	268324	269361
52	270400	271441	272484	273529	274576	275625	276676	277729	278784	279841
53	280900	281961	283024	284089	285156	286225	287296	288369	289444	290521
54	291600	292681	293764	294849	295936	297025	298116	299209	300304	301401

No.	0	1	2	3	4	5	6	7	8	9
55	302500	303601	304704	305809	306916	308025	309136	310249	311364	312481
56	313600	314721	315844	316969	318096	319225	320356	321489	322624	323761
57	324900	326041	327184	328329	329476	330625	331776	332929	334084	335241
58	336400	337561	338724	339889	341056	342225	343396	344569	345744	346921
59	348100	349281	350464	351649	352836	354025	355216	356409	357604	358801
60	360000	361201	362404	363609	364816	366025	367236	368449	369664	370881
61	372100	373321	374544	375769	376996	378225	379456	380689	381924	383161
62	384400	385641	386884	388129	389376	390625	391876	393129	394384	395641
63	396900	398161	399424	400689	401956	403225	404496	405769	407044	408321
64	409600	410881	412164	413449	414736	416025	417316	418609	419904	421201
65	422500	423801	425104	426409	427716	429025	430336	431649	432964	434281
66	435600	436921	438244	439569	440896	442225	443556	444889	446224	447561
67	448900	450241	451584	452929	454276	455625	456976	458329	459684	461041
68	462400	463761	465124	466489	467856	469225	470596	471969	473344	474721
69	476100	477481	478864	480249	481636	483025	484416	485809	487204	488601
70	490000	491401	492804	494209	495616	497025	498436	499849	501264	502681
71	504100	505521	506944	508369	509796	511225	512656	514089	515524	516961
72	518400	519841	521284	522729	524176	525625	527076	528529	529984	531441
73	532900	534361	535824	537289	538756	540225	541696	543169	544644	546121
74	547600	549081	550564	552049	553536	555025	556516	558009	559504	561001

75	562500	564001	565504	567009	568516	570025	571536	573049	574564	576081
76	577600	579121	580644	582169	583696	585225	586756	588289	589824	591361
77	592900	594441	595984	597529	599076	600625	602176	603729	605284	606841
78	608400	609961	611524	613089	614656	616225	617796	619369	620944	622521
79	624100	625681	627264	628849	630436	632025	633616	635209	636804	638401
80	640000	641601	643204	644809	646416	648025	649636	651249	652864	654481
81	656100	657721	659344	660969	662596	664225	665856	667489	669124	670761
82	672400	674041	675684	677329	678976	680625	682276	683929	685584	687241
83	688900	690561	692224	693889	695556	697225	698896	700569	702244	703921
84	705600	707281	708964	710649	712336	714025	715716	717409	719104	720801
85	722500	724201	725904	727609	729316	731025	732736	734449	736164	737881
86	739600	741321	743044	744769	746496	748225	749956	751689	753424	755161
87	756900	758641	760384	762129	763876	765625	767376	769129	770884	772641
88	774400	776161	777924	779689	781456	783225	784996	786769	788544	790321
89	792100	793881	795664	797449	799236	801025	802816	804609	806404	808201
90	810000	811801	813604	815409	817216	819025	820836	822649	824464	826281
91	828100	829921	831744	833569	835396	837225	839056	840889	842724	844561
92	846400	848241	850084	851929	853776	855625	857476	859329	861184	863041
93	864900	866761	868624	870489	872356	874225	876096	877969	879844	881721
94	883600	885481	887364	889249	891136	893025	894916	896809	898704	900601
95	902500	904401	906304	908209	910116	912025	913936	915849	917764	919681
96	921600	923521	925444	927369	929296	931225	933156	935089	937024	938961
97	940900	942841	944784	946729	948676	950625	952576	954529	956484	958441
98	960400	962361	964324	966289	968256	970225	972196	974169	976144	978121
99	980100	982081	984064	986049	988036	990025	992016	994009	996004	998001