# 統　計　學

上　册

鄭堯柈著

商務印書館發行

# 統　計　學

## 第一編　緒論

### 第一章　緒言

多數學者將宇宙間分為自然界現象及人生社會現象二部,(人生社會現象大而言之亦得算為自然界現象)此二方面所涉複雜現象,非祇其本身複雜而已,且其相互間亦變化無窮,佛家所謂『諸行無常』即此之謂也。但吾人欲於此等變化無窮之現象中,以求得各種知識,實非易事。譬如無數昆蟲成為一團時,雖祇欲知其數目,亦覺甚難,何況於此瞬刻變化之現象中以求其理解,眞所謂難中之難矣。然吾人依多年學問上技術上研究之結果,漸次有說明此等現象之可能,而其中以數字說明一法為最上策,此數字說明學者名之曰統計,此統計範圍之所以廣,應用之所以宏也。依上說知統計是用數字以說明一切,可說是數字應用之一種,其特徵是在能簡單明瞭說明現象,使世人無插入疑問之餘地。故雖極小極簡單一册統計報告,亦能說明一切現象,並能使任一國人看了均毫無困難地得能明瞭其國家社會現象,實可稱為說明一切現象上最便利之國際語也。用數字以說明一切,對於數量上固甚適當,但對於質量上則見困難,故學者發明補救方法,將其內容詳細分類,以逐次明瞭其性質,

用數字及分類二法，於是近代統計乃趨於完全境界。

依上述知統計是由數字構造而成，是用數字以說明一切現象，但其構成方法如何，則有預行略述之必要。統計之構成，是由將大量之同一種類物，在一定空間與時間之下，一個也不少，一個也不重複，從頭至尾全體依預定目的舉行觀察後所得數字，始得稱為統計。將全體一個不遺漏一個不重複以行觀察一事，專門家稱之為大量觀察。(Aggregative observation; Massenbeobachtung)。

大量觀察之特徵是在不遺漏不重複之計量上，但世人對此每多誤會，多以為大量觀察就是大體觀察，差不多的數目就以為已大量觀察過了，此實大錯而特錯。現各機關辦理統計事務者，尚多作此想，豈非喪失統計之價值。須知差不多大約模等數值，實不得算為統計。然統計學內有所謂估計(Estimation; Schätzung)者則又為何物，對此又不能不預為說明幾句。估計是在事實上不能全體大量觀察時，定其內之一小部份，先依大量觀察方法實地調查，將此調查所得結果為計算之基礎，以估計其全體。此時所算出數字雖不免有粗雜想像等現象含在其內，但其基礎仍是正確之數字，不過於正確之統計上施行一種估計而已，此是統計之應用，與徒從臆斷推測亂下數字粉飾外形等似是而非之統計，則有天壤之別。此等單從個人主觀而來之數字，不能說為統計。統計最少有一部份非從大量觀察而來不可。大量觀察即是實地調查，故非從實地依大量觀察方法調查而來之數字，不得稱為統計之數字。

依上述知非實地調查而得之數字，不能稱為統計之數字，對此一點雖覺很費時間很費手續，或視為迂緩之事，然此處卻是統計之價值。凡

一般學問上沿革歸納學派之價值，是在將實際之事實爲基礎，以歸納求出該事實之原理原則。可是統計學亦是從實地調查所得之結果，以產生其原理原則。在學問上之性質，二者完全相一致。沿革歸納學派上所有之價值與利益，統計學上亦完全具備。沿革歸納學派所受詰難，統計學上亦所不免。但在實際上逐一調查所得之結果，於客觀上比此最澈底最可靠者，恐已屬於無，此處就是統計之價值。且對此結果無論何人均不能否認，其數字有無上之權力，此所以依世界之進步，而各國統計調查之日益重視也。

依現今經濟界之進步與其活動能力之旺盛，凡經營生產配給之業務者，均感有非立於正確之根據以定各種計劃，不足以戰勝一切。再加之現今社會上各階級間之利害關係愈趨於複雜化，各人爲保護自己之利益及緩和相互間之利害衝突起見，竭需正確之統計。依經濟之發展而要求統計之完備，同時依經濟之充實得滿足編製統計上所需要之經費。此二者相偕並進，於是統計隨經濟界之發展而日趨進步也。

現在民主主義盛行，各人之能力已無甚差異，故欲以一人之主觀思想去支配多人，已勢所不能。在此不能以個人主觀以支配衆人之世，執政者勢必一變其『民可使由之不可使知之』之態度，別想妥當方法，以便說服衆人。其方法爲何，則是依實際之事實及從來之成績，以證明此後設施，必須依此方向進行，而後始能歸服衆心。但欲知此等實際事實及經過成績，則必須依統計方法實地調查實際計量，使成爲無論何人不能挾入異論之數字，於是政治家得用此數字，以定種種方針。此所以依時代思想之進步，而統計之需要亦隨之日益重視也。

統計之需要已如上述,知爲萬政之本。旣爲萬政之本,就不能有個人之主觀的推測在內,非用實地調查所得數字不可,對此一點應當特別注意。現有多數從事統計工作人員,專依自己主觀以定數字,並於其外形稍事粉飾,就算一種統計。在上級機關者旣不能說明依如何方法調查,如何方法編製,只混亂地命令下級機關施行各種調查。而在下級奉到命令者,旣不能明白調查本旨,又無能力自定調查方法,見上級機關旣有此項命令分發下來,在尙有下級機關者,就將其所受命令,完全轉飭下屬。如已無下級機關可轉飭時,則將別處所得同樣調查之結果,依樣葫蘆,如法泡製,就說是由調查而來。如別處尙無此項調查報告時,那更不對,當事者就將自己腦海作爲調查區域,辦公桌上作爲整理場所,口銜香烟,亂定數字,粉飾外形,就算是一種統計,以搪塞報告之責。此等統計,寧有價値之可言。可是不明白內容者,一見就以爲此非甚整齊甚明瞭之統計乎,焉知其內容一點也不能信任,完全是杜撰假造,將此等杜撰假造之統計以爲施政之本,實屬貽害民衆,貽害統計之將來,其危孰甚。故在從事統計者,應感覺自己責任之重大,遇有以上諸缺點時,須根本改正,以期求得正確之統計,爲施政之本,並爲將來統計比較之本,庶乎其可也。

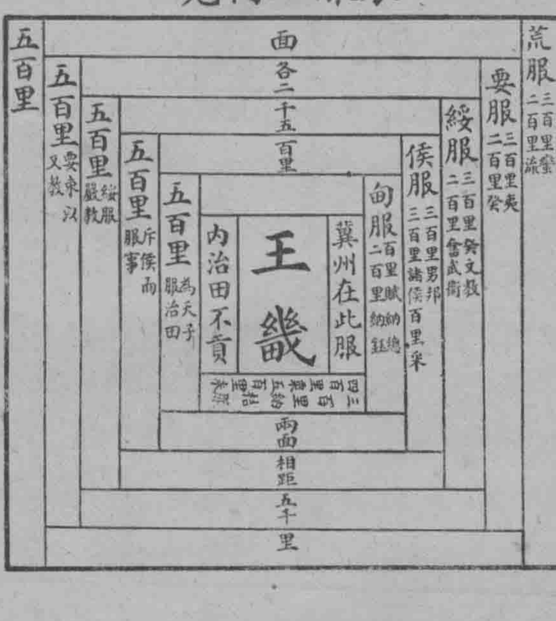
# 第二章　統計學之發達史及現代之統計學派

欲明統計學之概念，則以先知統計學之歷史，較為迅速。

統計調查事業由來已久，古時部落相結，乃成國家，國家之雛形既具，其國君為軍事上財賦上便利起見，乃感覺到有明瞭領土詳形之必要，於是統計調查之舉尚焉。我國統計調查事業倡行最古，堯之五服制（第1表詳情請參照周制）禹之禹貢篇（第2表）為其調查報告中之最古者，其

第　1　表

堯制五服表



第 2 表　禹貢九州表

| 州名 | 冀州 | 兗州 | 青州 | 徐州 | 揚州 | 荊州 | 豫州 | 梁州 | 雍州 |
|---|---|---|---|---|---|---|---|---|---|
| 境域 | 之東豫河之北 河兗河之西雍河 帝都之地三面距 | 濟河惟兗州 | 海岱惟青州 | 海岱及淮惟徐州 | 淮海惟揚州 | 荊及衡陽惟荊州 | 荊河惟豫州 | 華陽黑水惟梁州 | 黑水西河惟雍州 |
| 土質 | 厥土白壤 | 厥土黑墳 | 海濱廣斥 厥土白墳 | 墳 厥土赤埴 | 泥 厥土惟塗 | 泥 厥土惟塗 | 下土墳壚 厥土惟壤 | 厥土青黎 | 壤 厥土惟黃 |
| 田 | 中中 厥田惟 | 中下 厥田惟 | 上下 厥田惟 | 上中 厥田惟 | 下下 厥田惟 | 下中 厥田惟 | 中上 厥田惟 | 下上 厥田惟 | 上上 厥田惟 |
| 賦 | 錯 厥賦上上 | 九等 乃同賦第 十有三載 厥賦貞作 | 厥賦中上 | 厥賦中中 | 下錯 厥賦下上 | 厥賦上下 | 中 厥賦錯上 | 三錯 厥賦下中 | 厥賦中下 |
| 植物 | — | 厥木惟條 厥草惟繇 | — | 草木漸包 | 厥木惟喬 厥草惟夭 篠簜既敷 | — | — | — | — |
| 運輸 | 夾右碣石入于河 | 浮于濟漯達于河 | 浮于汶達于濟 | 浮于淮泗達于河 | 沿于江海達于淮泗 | 至于南河 浮于江沱潛漢逾于洛 | 浮于洛達于河 | 逾于沔入于渭亂于河 西傾因桓是來浮于潛 | 河會于渭汭 浮于石積至于瀧門西 |
| 產物 | 島夷皮服 | 厥貢漆絲牙篚織文 | 鉛松怪石萊夷作牧厥篚檿絲 厥貢鹽絺海物惟錯岱畎絲枲 | 厥篚玄纖縞 孤桐泗濱浮磬淮夷蠙珠暨魚 厥貢惟土五色羽畎夏翟嶧陽 | 橘柚錫貢 椎木島夷卉服厥篚織貝厥包 厥貢三品瑤琨篠簜齒革羽毛 | 璣組九江納錫大龜 底貢厥名包匭菁茅厥篚玄纁 栝柏礪砥砮丹惟箘簵楛三邦 厥貢羽毛齒革惟金三品杶幹 | 磬錯 厥貢漆枲絺紵厥篚纖纊錫貢 | 織皮 厥貢璆鐵銀鏤砮磬熊羆狐貍 | 厥貢惟球琳琅玕 |

後歷代亦時行調查，以期明瞭國內詳情。至歷來學者利用已有之統計材料，並依統計之概念以議論政治經濟及社會現象之策論，亦甚衆多。如齊管仲之富國策與鹽鐵論，魏李俚之物價策與地方民衆策，秦商鞅之田制論，漢桑弘羊之均輸說與耿壽昌之義倉說，後漢桓潭張林之商政策，隋長孫平之貯蓄策，唐張全義之農政策，宋王安石之財政策及元苗好謙之農政策等爲其中之最著者，是則史記戰國史等著作，亦未始非記述派的統計學，惜無系統的研究，不能成爲一種專門學問。以致近日所述統計學，完全須由西歐輸入，實堪浩嘆。

此處所說統計二字，並非吾人依字義上解說之合計總計總算共計等意義，乃是英文 Statistics 德文 die Statistik 法文 la Statistique 意文 Statistico 荷蘭文 Statistick 西班牙文Estadistica葡萄牙文 Estatistico 等之譯語。世人對此每多誤會，多以爲合計總計就是吾人所說之統計，此實大謬不然，對此亟應立期改正。至原語之由來，其說不一，但以由拉丁文 Status(狀態國家)，而來較爲正確。

統計調查事業倡行雖古，但將其發達成爲專門一學科，則爲近世之事。當中世之末近世之初，意大利皇帝欲知他國國況，命駐外使臣調查報告其駐在國之狀態，當時以調查方法之不良，不能得充分之效果，於是學者羣起研究，漸入於科學境域。依探究國家狀況所出最古著書爲海底堡大學教授苗斯特(Sebastian Münster)之萬國誌(Cosmographia 1544 年出版）此書是將各國之地理歷史風俗習慣政治宗教兵力等依次記述，與吾國之戰國史略同。但將此學科最初在大學講授者爲德之昆靈(Hermann Conring 1606-1681)，於 1660 年 11 月 20 日在𡁠爾母斯

旦脫 (Helmstadt) 大學，用「關於國家之重要事項」(Notitiarerum Politicarum nostri oevi Celebrrimarum)的題目以講授歐洲各國之現狀，憲法、行政、人口、經濟等事項，此爲舊派統計學之起源。其次爲斯買再(M. Schmeitzen 1679-1747) 用政治統計學講義 (Collegium "Politicostatisticum") 之題目，在 Jena(1723-1731) 及Halle(1731-1741) 二處講授一切，此(Statisticum)一語，實爲 (Statistik)一語發生之直接近因。至十八世紀中葉德之格丁根(Gottingen)大學教授阿痕發爾(Gottfried Achenwall 1719-1772) 始於講義上附以 Statistik 之題目，並倡言凡可以左右國家繁榮諸事項，稱爲國家之顯著事項，(Staatsmerkwündigkeit)而 Statistik 則以此爲其研究之目的等學說。其講述範圍甚廣，現在所稱之憲法、行政法、經濟學、地理學等，皆包涵在內，彼時以其爲應世之必需品，各國多翻譯之。於是阿氏之關於統計之見解流行各國，後世遂尊稱其爲統計之父(Vater der Statiitik)。承其後者則有休累湊爾(A. L. Von Schlözer 1735-1809)，氏主張「統計爲靜止之歷史，歷史爲流動之統計」。(Statistik als stille Stehende Geschichte, die Geschichte als ferlaufende)之學說。以上諸學者學說，均使用文字以記述各國狀況，數字不過作一種補助品而已，至統計表則完全不使用，又以其研究者皆爲大學教授，故稱該學派爲大學統計學派(Universitatsstatistiker)，又因其完全以記述爲主，故或稱之爲記述學派 (Descriptive school)。當時德國適爲腓特烈二世(Frederick II)對於統計調查事業十分注重，其調查方法亦隨之日趨進步。於是官府統計(Official statistics)之體裁，得以逐次趨入於整頓界域。對於官府統計上有巨大之供獻者爲德之皮

興(Atlon Friedrick Büsching 1724-1793),氏用比較方法,著有國家學之攝要一書,書內富於計算之材料,與阿痕發爾所述完全不同,由各國自身之比較一轉而爲各國相互間事實之比較,以期發見社會現象中所存在之統計法則。(統計法則意義請參照第四編)其研究是專以行政官廳所有之實際材料爲基礎,大增統計之正確程度,並促進官廳統計之發達,故後世稱之爲官廳統計之鼻祖。

與阿痕發爾同時在丹麥有安歇爾生(Anchersen 1700-1765),改革從來記述之統計,專用數字表以比較一切,稱爲表記統計學派(Tabellenstatistiker)。此派與大學統計學派旣在同一時期,勢不免互相排斥,甚至大學派之學者,詆表記派之學者爲表奴(Tabellenknechte)。以上各派總稱之爲舊派統計學,專以闡明各國國情爲其研究之目的,編製政治年鑑或統計辭典式之書籍。

該時歐洲大陸法國適在拿破崙一世之期,拿破崙認統計爲治世之必需品,特設統計局,並於1801年施行大規模之人口調查。普魯士亦於1805年設立統計局,依直接調查以作成正確之統計。該時普法二國對於官府統計,一若成爲相互競爭之勢。至理論一方面則大學派與實際統計之進步完全不相連絡仍以言文爲主記述爲旨,研究其獨有之學問,遂致統計學理陷於停滯狀態。

與大學統計學派同時興起者在英國有政治算術學派(Political arithmetic school),此學派至後世實奪舊派統計學之地位,爲現代統計學之起源。其始祖爲英人葛蘭溫脫(Capt. John Graunt 1620-1674),氏與昆靈爲同時期之人,以倫敦市民之出生死亡情形爲基礎,於1662年

發表一篇「用倫敦市民之死亡表爲根據以舉行自然與政治的觀察」,(Natural and Political Observation upon the Bills of Mortality of the City of London)。其論文內容,是說明倫敦人口中男女人數相等,雖依流行病以減少人口,但此減少數目以後即能由外部加以補充,每一婚姻平均約生小孩四人,其出生小孩中男女比例約爲 14 與 13,由出生起至第六年終爲止期內死亡人數最多,約占出生之$\frac{35}{100}$,並依年齡組別以設立各種死亡率。其論文完全以數字爲基礎,用歸納方法編出數字的法則,與舊派完全不同。葛氏之後有配第(William Petty 1623-1687),繼續葛氏之研究,與政治算術學派,完全依數字以推測各國政治上之事項。其次則有發見慧星之倫敦天文學者哈利(Edmund Halley 1656-1742)以德人紐曼(Kasper Neumann 1648-1715)於1691年所蒐集之德國北勃斯勞(Breslan)市人口統計材料爲根柢,作成死亡表(Halley's mortality tables),以研究生命年金之計算,定生命保險(Life assurance or insurance)之基礎。此等學者接踵出世後,此派之研究漸次成爲人口現象上實用的研究,其中以關於生命保險之數理,最爲發展。且此派學者非只限於英國在歐洲大陸亦甚衆多,如德人墨塞爾(Moser 著壽命之法則 1839年),法人得帕爾修(Deparcieux 1703-1768)荷蘭人刻舍蒲(Kersseboom 1691-1776)及瑞典人瓦格頓(Wargentin 1717-1783)等是。

依上述吾人得知政治算術學派與舊派統計學之異同處,政治算術學派是用數字的材料爲基礎以進行研究之步驟,其研究對象非只國家情勢而已,完全是依數字的材料以研究社會現象,並依此以檢討發見支

配法則之方法。

政治算術學派至德教士緒司密爾 (Johonn Peter Süssimilch 1707-1767)成爲新意義之統計學,緒氏說社會現象是依種種原因所作成之結果,以變動一切,其作用原因大別之可分爲恆常的神爲原因,與變動的人爲原因二種,在社會現象上如觀察單位少數時則人爲原因易於表現,如觀察單位多數時則人爲原因互相消滅而恆常的神爲原因得能明白表現等語。伊熱心政治算術學之研究,以各寺院所存在之婚姻洗禮死亡等之登記材料爲基礎,依大量觀察方法發見出生者比死亡者爲多男女出生之比例雖爲 21 與 20,但男子死亡數比女子爲多,故全社會男女人數,仍能相等,並論及兒童死亡率非常高大等關於人口現象上之常例。於 1741 年發表「人間體性變化上之神的秩序」(Gattliche Ordnung in den Veränderungen des menschlichen Geschlechte)一文,於是緒氏學說,遂爲世所公認。至緒氏對於統計學上之功績,則有:

(一)想定社會現象是爲有原因作用之物;

(二)依觀察之結果,知社會上有常態法則之存在;

(三)表明社會現象上只依大量觀察始能發見恆常的法則,將方法學的政治算術派,一變而爲獨立的科學。

以上三種功績實萬世不朽,只緒氏本人爲宗教家,故將社會秩序完全歸之於神意,其論法又多不依歸納法而依演繹法,且多武斷處,未始非白壁之瑕。

其他如英人馬爾薩斯 (T. R. Malthus 1766-1834)對於人口統計學上亦有特殊之貢獻。

至十九世紀中葉比人概特雷(Lambert Adolphe Jacques Quételet 1796-1874)出,解決新舊二派之爭執,打倒舊派,改良政治算術學為現時之統計學。概氏關於斯學之著作甚衆,其中最出色者爲次之四種:

(一)關於人間及其能力發達上之社會科學論(Sur l'homme et le developpement de ses facultés, ou essai de physique sociale, 1835, Bruxelles, 1869)

(二)道德學及政治學上幾率論之應用(Sur la théorie des probabilités appliquées aux sciences morales et politiques, 1849)

(三)社會組織論(Du systéme social et des lois qui le regissent, 1848)

(四)人間論(Anthropometrie, ou mesure des differentes facultés de l'homme, 1870)

概氏本爲天文學家,以眺天之眼來眺地面,以觀天之方法來觀察人間社會,預期發見人間社會上所存在之法則,因此蒐集數字的材料,用歸納方法以進行其研究步驟。其研究範圍非只人口現象而已,凡關於人類學及人體學的考察,皆加入在內。並兼用社會科學與自然科學雙方之觀察方法,以期完成關於人間之研究。彼依此計劃所得研究之收穫如次:

(一)人口統計方面只增加少數新材料而已,並無特殊之發見;

(二)創造道德統計(Moralstatistik)概氏以法國犯罪統計爲基礎,說明犯罪是每年依一定程序循環發生,其發生與氣候季節性別年齡教育職業等皆有密切之關係。並說國家之有犯罪預算,與有財政預算相

等,其犯罪之發生,完全受環境之支配等語;

(三)測定人之身長體重腕力呼吸脈搏數等以完成人類學之研究;

(四)綜合關於人間上各方面之研究,發明平均人(L'homme moyen)。此平均人是由蒐集關於人間肉體精神性質等各種觀察之平均數以構成之抽象人物,詳言之則爲綜合全國國民之身長體重脈搏數結婚年齡生存年齡犯罪傾向等之平均數所構成之理想人物以爲觀察人類之標準物。

概氏雖有濫行擴大人物研究之範圍,以致常起利用不正確材料之缺點,但常能下新鮮之結論,注入統計學以新生命。至其關於統計理論上之最大功績,則爲應用柏努利(J. Bernoulli 1654-1705)拉普拉斯(P. S. Laplace 1749-1827)傅立葉(A. Fourier 1768-1830)等學者所研究之幾率論(Theory of probability, Whascheinlichkeitsrechung)於統計學上,使確認大數法則(Law of large numbers, Gesetze der grossen Zahlen)之原理,賦大量觀察以生命,確立統計學之理論。其實際上功績則爲其就比利時統計行政當局時,貢獻官府統計之發達,並於1846年舉行近代式戶口大調查,爲現代各國戶口調查之模範。更於1853年發起萬國統計會議,以圖國際間之統一。概氏實得稱爲統計界之第一人,現在統計學皆由此發展。

在當時爲世人所注目者,尙有克尼斯(Karl G. A. Knies 1821-1898)之「獨立科學之統計學」(Die Statistik als selbständige Wissenschaft 1850)一論文。將統計學與從來記述國家狀況之學派相分離,主張其爲用數值以解說大量現象之一種專門學問。對於近代統計學之確立

上,實有重大之關係。

從十九世紀中葉以來,舊有之統計學均一變爲現代統計學,除上述諸學者外,尙多知名之士。現將其人名列舉如次:

a. 十九世紀有名人物

人口統計方面:——

Hoffmann, Dieterici, Engel, Boeckh, Wappäus, Hermann, Rümelin, Mayr, Newshome, Farr 等。

生物測定方面:——

Galton。

經濟統計方面:——

Hoffmann, Dieterici, Engel, Jevones, Block, Mayo-Smith, Wagner 等。

道德統計方面:——

Guerry。

b. 最近統計學界有名人物

人口統計方面:——

Kniffs, Letis, Newshalme, Knopp 等。

生物測定方面:——

Pearson。

經濟統計方面:——

Bowley, Zizik, Fisher, Walsh, Edgeworth, Moore, Secrist, Davies. Conrad, Bertillon, Mitchell, Persons, Adams, Crum, Patton, Pareto, Tyszka, Walff, Meerwarth, Jerome, Mills, Play, Bucher, Schmoller, Aftalion, Simiond 等。

方法論方面(一般統計學)

Lexis, Bowley, Zizik, King, Yule, Pearson, Thorndike, Secrist, Kaufmann, Meitzen, Elderton, Kelley, Jones, Forcher, Bortikiewicz, Tschuprow, Czuber, Schott, Maeller Keynes, Westergaard, Walsh, Julin, Jordan, Rietz, Jerome, Mills, Chaddock, Sutchiffe March, Darmois 等。

至現代統計學界,從德人馬約 (George Von Mayer 1841-1925) 主張統計學非只方法論而已,實爲由其研究之對象及其結果以成立之一種獨立的社會科學後,起多數學者之共鳴,故近日統計學界仍分爲二派。一爲社會統計學派,主張統計學是以社會爲其研究對象,依特殊研究方法以研究其構造組織及其他關於一切社會現象的一種學問,將統計學定作爲一種社會科學,馬約等屬於此派。他爲統計方法論派,主張統計學是爲於可以適用大量觀察的各種學問範圍內,攻究統計調查方法及研究方法的一種方法,英美學者多屬於此派。

# 第三章　統計學之定義

由上章知統計學是依時代之變遷以更改其內容，故欲定確立不變之定義一事，實爲勢所不能。依此關於統計學之定義由來甚衆，法拉提(Johannes Fallate 1609-1655)列舉古來學者所下不同之定義凡53種，摩爾(Rebert Von Mohl 1799-1875)列舉63種，至1869年，在荷蘭開第七次國際統計會議(Congries International des Statistique)時，概特雷提出歷年學者關於統計之定義達180種之多。統計學之定義在概特雷時期已有如此之衆，何況在六十年後之今日其定義之多可想而知。現綜合最近各學者定義，略述統計之定義如次：

受有多數原因影響之事實於特立目的之下，先依組織的蒐集方法與正確的數計方法以施行一切後，再將其測得之結果依預定程序排列之，使成爲數字的集團。如此所得之數字集團，吾人稱之爲統計。

受有多數原因影響之事實，是爲社會現象，但又非只限於此，其他各方面所有同種之大量現象（自然界一切大量現象）亦得包含在內，於特定目的之下依組織地以行蒐集者，是說其必須在確立範圍內，依一定之目的組織與順序以行蒐集材料，如只漠然而無一定之目的則其所蒐集材料，於嚴格意義上不得稱爲統計資料，其由此整理所得結果，亦只成爲一種數字的集團，毫無統計價值。至所蒐集材料不依適當方法以行數計以行測定時，則所得結果難保其爲正確。又將所得結果不依預定秩

序以行排列時則難以應用。其依秩序排列者，是將其數字依預定方式以行排列，預定階級以行分類之謂也。

至統計學之定義如次：

於上述統計上，研究其說明記述之方法及由此以發見各現象之眞理原則等之學問，吾人稱之爲統計學。

此說明記述之方法，吾人稱之爲統計方法 (Statistical methods, statistische Methode)。

# 第二編　統計材料之蒐集

統計學之意義，是施行特殊觀察方法於自然界及社會界之現象上，以蒐集關於此二種現象內之數字材料，並於此數字材料上講究特殊的數字處理方法，以發見關於現象上之眞理原則。所欲蒐集之數字材料，稱爲統計資料，爲編製統計之基礎，其蒐集方法之巧拙，關係於統計之價值甚大，故對於統計資料之蒐集方法實有特殊研究之必要。現將其蒐集方法逐次說明如次：

## 第一章　統計材料之分類

統計資料依蒐集方法之不同，約可分次之三項以行說明：

(1) 單位與品質；

(2) 直接統計與間接統計；

(3) 靜態統計與動態統計。

### 第一節　單位與品質

蒐集統計資料時，對於目的物之單位 (Unit; Erhebungseinheit) 與品質 (Attribute or characteristics; Erhebungsmerkmal)，有預先明白了解之必要。單位則是成立觀察大量之個體，此個體得用一、二、

三、等數字以計數之。例如人口調查時之住民，交通調查時之通行者，住宅調查時之住宅，家畜調查時之家畜等是。品質是爲關於個體之性質及機能上之事項，例如人口調查時之人的年齡體性職業教育等是，觀此單位與品質之區別。一若甚明瞭然。但當實地調查時，常易於混亂難分。例如調查某校內之學生數時，其學生當然爲其單位。可是遇調查全國某種學校數並附屬調查其內容時，學生數變爲表示各學校之特質，於此時之學生只能算爲品質，不能算爲單位。依此知單位有時亦得變爲品質，對此宜特別注意。

單位之種類約可分爲人的單位與實物單位（非人的單位）二種，依此統計調查亦可分爲人的統計調查與實物統計調查二種，人的單位如學生數通行者數等是，實物單位如住宅數家畜數等是。

決定單位一事，一見若甚簡單容易，但實際上常遇可爲單位之標準甚衆，使辦理者易入迷路。如人口調查之主要目的，是調查特定區域及特定時間內所有人口確數，但對此人口又有原籍人口，常住人口，現在人口，市人口之分，其中之誰爲調查之目的一事，在未熟悉統計原理者實難以決定。此處非只對於決定單位而言，凡關於統計研究上皆得使用次之口號，『目的是支配方法』。依此知決定單位之標準雖有多種，但只須觀察其目的之所在，以定其最適宜之單位。於必要時得將在常識上算爲一單位者，分割爲數單位或逐出於計算之外。例如戶口調查時將一幢房屋內住民分爲數戶，房屋調查時多將廢頹腐朽之房屋不算入在內。

品質約可分爲時間的品質空間的品質及實質的品質三種，時間的品質是指示調查事項之於時間上的關係，空間的品質是指示調查事項

之於空間上的關係，宇宙間現象皆限止於時間與空間之關係上，故調查事項當然也不能逃出時間與空間之範圍外。至實質的品質更得細分爲次之數種：

(a)一般的品質與限定的品質

一般的品質，是指調查現象全部所具有之屬性而言，例如人之年齡體性等事項。限定的品質是指調查現象內只有一部分具有之屬性而言，例如人之所屬職業薪俸等事項。

(b)不變的品質與可變的品質

不變的品質是說其屬性依時間之進行而不更改，例如人之體性。至年齡職業等依時間之經過而不同，故稱之爲可變的品質。

(c)於客觀上易於認識之品質與難於認識之品質

如體性等於外觀上易下正確的判斷，是屬於前者。如職業年齡等於客觀上難下正確的判斷，是屬於後者。凡觀察屬於後者時，則俟其本人之報告較爲確實。

(d)自自然的品質與社會的品質

自然的品質是不俟人爲而由天然存在，例如年齡及體性等事項。社會的品質是由人類構造而成，例如職業國籍族籍等事項。

(e)單位上爲單一之品質與得合羣之品質

如體性年齡等各單位只能有一個，是屬於前者。至職業等品質則一人同時可以從事數種，是屬於後者。

(f)質的品質與量的品質

體性婚姻宗教死因職業等是屬於前者，年齡工資耕地面積等是屬

於後者。量的品質中又可分爲計量與計數二種，計量的品質，是爲年齡及身長等。計數的品質，是爲學生數家畜數房間數等。前者得觀察至一以下之小數，而後者則不能小至一以下。

品質之區別已如上述，至其決定上之注意事項，則有次之數種：

(1)遇決定之品質有多種時，則依決定單位時所說之方法，先省察統計調査目的之所在，以選定其目的上最適合之品質。並同時要使用同性質之品質於各次同種類之調査上，以期比較上之便利。

(2)於一單位上得觀察多數品質時，則宜充分權衡各品質之輕重，以重者先行，輕者後辦。如人之品質雖甚衆多，但對於戶口調査等大規模調査時，其調査事項雖只增加一項，亦須增加巨大之經費，故各國對於戶口調査之事項，均有一定之限制，擇國內最切要者先行舉辦。

## 第二節　直接統計與間接統計

專爲得統計材料，特於大量現象上施行大量觀察後所得之統計，稱爲直接統計 (Primary statistics, Primarestatistik)。

於已爲別項目的所存在之材料內，收集統計資料，以作成之統計，稱爲間接統計(Secondary statistics, Sekundärestatistik)。

依戶口調査所得之人口統計，是專爲得人口統計上之材料起見，特於人口現象上施行戶口調査之大量觀察，故得稱爲直接統計。反之貿易統計，是於已爲別項目的每月每日所報告之貿易狀況內，蒐集其所需要之材料，故得稱爲間接統計。

直接統計對於調査事項之選定雖覺自由，但當舉行時太費金錢，爲

貧弱國家所不樂用。至間接統計對於調查事項上雖被限制，但有節省費用之便利，不過其所得結果太不正確，故為現代統計學界所漸次排斥。

## 第三節　靜態統計與動態統計

宇宙間之大量現象，其構成分子雖時時刻刻新陳代謝變化無窮，但如取極短之瞬間，則比較的可稱為靜止的狀態。此靜止狀態時所有宇宙間之大量現象，稱為靜大量（Bestandsmassen）。依時間之經過所連續發生現象或行為之總體，稱為動大量（Bewegungsmassen）。人口工場等總體於瞬間內可說為靜止，故稱為靜大量。出生死亡貿易等之總體依時間而不同，故稱為動大量。

靜大量上所施行之大量觀察名曰靜態調查，由此所得統計稱為靜態統計（Bestandsstatistik）。在動大量上施行大量觀察後所得統計，稱為動態統計（Bewegungsstatistik）。無論何種大量現象，不是靜大量就是動大量，依此知各種調查必為靜態調查與動態調查中之一種，故無論何種統計不是靜態就是動態。

靜態統計之原則是在計數上，動態統計之原則是在登記上。計數是在一定瞬間內檢查數計該大量現象上各單位。例如戶口調查之定某瞬間期內調查全國或全省之戶口上各狀況。登記是將逐次發生之現象逐次記錄於登記簿上，至一定期間後，觀察其全體變化情形。例如人事登記之逐次將出生死亡婚姻等現象登記於人事登記簿內，而後依每年每月以觀察其人口上變動狀況，

在同一現象之靜態統計與動態統計間，實多有密切之關係。且各種

現象上對其靜態與動態非雙方同時調查後，實難以明瞭其詳細情形。例如人口之出生現象，若只知其出生數其利益不甚巨大，必須與人口數相比較後，始能完全明瞭其出生現象。但在同一現象內欲同時知其靜動二方詳情一事，有時爲事實所不能。蓋因靜態調查與動態調查其施行手續完全雙方無關，故只能一方依統計的方法以舉行調查者實甚衆多。現將此種情形略述二種如次；

(1)有動態調查而無靜態調查者。

爲行政上目的起見，於動大量上繼續施行登記，對此其記錄雖皆存在，但其靜態調查有時不能舉行。例如吾人雖能依人事登記得知某種疾病之死亡數，但於特定瞬間調查該疾病之患者人數一事，爲事實所不能。

(2)有靜態調查而無動態調查者。

例如職業在戶口調查時得詳細查明各職業所屬人數，但其就職轉職等職業上之變動狀況實無從調查。

# 第二章　統計材料蒐集上各要項

## 第一節　統計材料之蒐集者

十九世紀末葉以來，對於統計材料之質與量二方面，皆有長足之進步。其進步之唯一原因，在統計材料蒐集機關之被社會所注重，及其自身之改善上。在今日各國所有蒐集統計材料之機關，大別之可分爲公的機關與私的機關兩種。公的機關是指國家及地方行政團體而言，其中最重要者爲國家之行政關機。私的機關在統計思想進步並認識統計的研究爲必要之國家較爲衆多，如美國之報館雜誌社等所蒐集之統計材料實甚富足，此等私的機關皆熱心蒐集各國有益之統計材料，能補充政府統計之不足，以滿足各界之需要，其統計之發表亦甚迅速，故材料甚覺新鮮。

## 第二節　統計蒐集機關之組織

現代科學的職務組織皆如金字塔，凡決定事業全體之根本方針與分發重要命令之首要者，實據金字塔之上端，其下依次分段，人數亦依次增加。

統計蒐集機關之組織亦宜如金字塔，有一指導者據其頂上以定一切方針。此指導者之責任甚大，故最少須備次之數項條件：

(1)對於數字的研究上須有特殊之興味；

(2)對於現代統計學之理論及實務上須有充分之研究;

(3)非只模仿而已,要有創造的能力;

(4)須有努力鼓吹統計事業發達之意思力;

(5)尊重統計事業並確信其爲畢生事業,全力工作。對此一層則以優待統計家爲先決條件。

至最下層工作人員則爲調查員,對此等人員,事前宜正確規定其業務之分掌,明白其責任之歸屬,使全體工作人員成爲有機體的一團,而後工作始能整個進行於無阻。各國戶口調查於此意義上,特設立調查區以定調查員之工作範圍。當規劃此等調查區時,對於其境界線宜與行政上地理上系統上之境界線互相一致,對於其面積則宜考察調查事項之性質與調查員之人數以定其廣狹。

使用調查員時,對於調查員之選擇宜十分注意,須擇其對於事務上有十分興味熱心勤勉身體強健溫厚篤實信望昭著能嚴守祕密熟悉地方情形及精通調查事項之人士。如警察官吏稅務關係者新聞記者或通訊員等易爲被調查者所誤解,非於不得已時萬不能使用此等人士爲調查員。

## 第三節　調查事項

當決定單位及品質等,調查事項時,宜注意下列諸點:

(A)須選擇適合於研究目的之調查事項。

當蒐集統計材料時,對於調查之目的實有預行確定之必要,並宜依此目的以定種種統計調查之手段。

(B)前後宜相同。

同一種類之調查,雖在不同時間與空間上舉行時,對於調查事項仍須前後相同,以便於比較。

當比較時爲明瞭眞相計,有充分考察該現象之前發條件之必要。例如將學校內之死亡率與全國死亡率相較時,學校內之死亡率當然比全國爲低,對此並非是一入學校就可以減少死亡,實因全國人口內有老弱及幼孩等死亡率高大者混合在內,以致死亡率比學校爲大。由此知在調查各大量現象時,對於前發現象(例如死亡率之前發現象爲健康狀態年齡職業等)是有充分研究之必要。若前發現象有重要變化時,應將其狀況詳細算入於統計資料內。

(C)調查事項不宜過於太多。

同時調查多數事項而結果反不得正確者,實甚衆多。故對於調查項目應預先十分精選,抱少問多得主義,求其適用而已。

(D)調查上之三不可。

調查中有三不可事,第一關於智識道德等不能調查事項。蓋智識道德屬於精神範圍,是無形之物,不易從調查以得正確之數字,故當蒐集統計材料之前,須預先考究其欲調查之事項,能否依確實數量以行表示一事。第二關於個人之內行私事等不許調查事項。蓋調查個人祕密實爲統計上所大忌,若強欲調查,則反有害其他正當之調查事項,故必須嚴禁。但已得被調查者之同意時,不在此限內。第三關於事實上雖可調查而實際上勞多功少等不利調查事項。例如調查全國鞋數,事雖可能,但問有何利益。此三種不能,不許,不利事項,當決定調查項目時宜留意切避。

(E)對於調查事項宜附以正確明瞭之說明。

大量觀察是於大量現象上精查各單位，故對於調查事項有正確明瞭規定之必要。但當規定說明時宜注意次之諸點:

(1)注意調查員及被調查者之智識程度，使無論何人不能錯誤然以給與詳細之說明。

(2)對於調查單位之說明，應預先給以周密之研究，並指示明確範圍，使不致發生疑問及誤解等事。至發生疑問及誤解時之對策，亦宜先行研究妥當。

(3)對於說明之用語宜使用普通應用文語或口語。

(4)對於調查事項之定義宜有理論與實際之根據，使任何人觀了，皆能達到同一之結論。

(5)對於調查事項，宜切避其預設分類以詢問一切。

(6)宜用具體的確實事項，例如問教育程度時，若問以「有教育或無」「教育程度優良可或不可」等語，則雖能得到答案，但此等答案有主觀的助力在內，決不能說是正確的回答，故必須依「如何程度學校畢業」「何種試驗合格」「在某種學校受過教育幾年」等形式以詢問後，始能得到正確之答案。

## 第四節 調查時期

對此時期得分為靜動二態以行說明。

關於計數靜態之時期上所起問題，第一是在何年調查，第二是在何月何日調查，第三是在一日中選擇何時何分何秒調查，欲解決此等問

題，則有明瞭以下各點之必要。

(A)鑑察現象之性質

如人口調查時選擇人類比較靜止狀態之深夜，所謂夜間調查。如職業調查時採取失業者之最少，經濟活動能力之最旺盛時期。總之依現象之性質以決定調查時期可也。

(B)須圖比較上之便利

例如 1872 年在俄都開催萬國統計會議時，議決在西曆年號過零的一年年末各國一齊舉行戶口調查，其用意完全爲圖國際間比較上便利起見。

(C)須圖調查上之便利

對此一層可分爲被調查者之便利，調查者之便利與調查機關之便利三種，以行說明。一般依社會上習慣在年末節期，被調查者是甚繁忙，對此時期爲圖被調查者之便利起見，宜求切避。梅雨期降雪期及大寒大熱等時期是給障礙於調查員之活動能力，爲圖調查員之便利起見，亦宜切避此等時期。至調查機關在官署宜切避在選舉納稅等時期，在私的機關宜切避在決算期或營業最盛等多忙時期。關於此等各點，當決定調查時期時，應預先除去，以期得良好的結果。

對於動態現象之時期可分爲在事實發生時卽行調查登記及經過一定時間後始行調查載錄其發生事象二種。直接統計是常將事象發生與調查在同一時期，間接統計在一般上是調查時期較事象發生時期爲遲。

對於統計調查上不可有只依記憶以報告之惡弊，蓋因只依記憶實易起誤謬。故知事象發生時期與調查時期之差異，務宜求其短縮。

## 第五節 調查地方

關於調查地方得分直接調查(調查者與被調查者在同一地方)與間接調查(於已依別項目的所存在之資料內搜集所需要之統計材料)二種,以行說明。

對於直接調查之地方約可分為:

(1)調查者施行調查於被調查者所在地(如戶口調查),

(2)將被調查者招至(或自然到來)調查者所在地,(如犯罪統計則屬於被招至一類,貿易統計屬於自然到來一類)。

以行調查二種。但靜態現象之數計多在現象所在地,動態現象之數計則多在調查者所在地以施行一切。

對於間接調查之地方,約可分為:

(1)在原調查地使原調查機關自當其衝,

(2)仍在上述所在地由中央機關或地方機關派員專司其事,

(3)將欲調查之記錄蒐集於中央機關或地方機關,由中央機關或地方機關由此材料內抄錄謄寫關於統計調查上所必要之事項,等三種。

## 第六節 調查票式樣

調查票之式樣得分表式法與小票法二種。表式法是當調查時用統計表之式樣表,將編製統計之分類明示於調查票上使收到者只填入各欄應填數字已足,本法在行政統計調查上時常使用,第3表為江蘇省民政廳行政統計中之一種表式調查票。用表式以蒐集材料時,實難期其能

得正確之統計。蓋因票內只須填入數字，易使報告者起亂填數字搪塞報告之責等弊。

第 3 表

違警罪發生及拘到件數報告表（　年　月份）　縣

| 罪名 | 發生件數 | | | 拘到件數 | | | |
|---|---|---|---|---|---|---|---|
| | 本月份事件 | 本月以前事件 | 計 | 管內事件 | | 管外事件 | 計 |
| | | | | 本月事件 | 本月以前事件 | | |
| 妨害安寧罪 | | | | | | | |
| 妨害秩序罪 | | | | | | | |
| 妨害公務罪 | | | | | | | |
| 誣告偽證及湮沒證據罪 | | | | | | | |
| 妨害交通罪 | | | | | | | |
| 妨害風俗罪 | | | | | | | |
| 妨害衛生罪 | | | | | | | |
| 妨害他人身體財產罪 | | | | | | | |
| 合計 | | | | | | | |

一、發生件數欄內之「本月以前事件」係記入發生事實在本月以前而報告延至本月內者

二、管內事件係管轄範圍內之拘到事件

三、管外事件即在管轄範圍以外發生之事件而在管轄區域內拘到者

四、將每月份件數於次月份十五日以前報告到廳

責任記入者　　印

年　月　日

查上無誤

公安局長　　印

年　月　日

小票法是當單位調查時，用實際詢問方法，將被調查者情形依實填入調查票上，而後將其小票蒐集至整理機關，使整理成爲整個的數字。其調查式樣得分爲單記票 (Individualfragebogen) 與列記票 (Kollektivfragebogen) 二種。

單記票是每一調查單位，使用一張調查票。第 4 表是爲拙著江蘇省人口動態調查施行預定計劃(登在明日之江蘇第五期)中所使用調查票之一種，因每一人使用一張，故稱爲單記票。

第 4 表　江蘇省人口動態調查票

出生票

民國　　年　　月份

生字第　　號（　　）

| 江蘇省　縣/市　區　鎮/鄉公所 | | |
|---|---|---|
| (一)姓名及男女別 | 男　女　姓名 | |
| (二)出生年月日 | 年　月　日 | |
| (三)出生地名 | 省　縣/市　區　鎮/鄉 | |
| (四)母腹之第幾胎 | | |
| (五)處理者 | 醫師　產婆　收生婆　其他 | |
| (六)父母別 | 父 | 母 |
| (七)姓名別 | | |
| (八)出生年月日 | 年　月　日 | 年　月　日 |
| (九)職業 | 業名　地位 | 業名　地位 |

列記票是將數個調查單位同時記入在一張調查票上。第 5 表爲拙著中國戶口調查方法之商榷 (中國統計學社第一次年會論文載在立法院統計月報三卷三期)內所使用之戶口調查票，是將戶內全部人名及各人品質列記在一張調查票上，故稱爲列記票。

# 戶口調查

戶口調查是明瞭國民之生活社會之實況以作善政之基礎因此全國人民應當努力與政府合作以完成此文明的國家大事業對於各項調查應當誠實報告！詳細記入！

## 第5表

## 戶口調查票

中華民國二十四年十二月三十一日

### 記入須知

一、記入前必須先熟讀記入須知

二、如無事實可記各欄必須劃一斜線（一）

三、須用墨筆明白記入

| | |
|---|---|
| 調查區號碼 | 第　號 |
| 省　縣 | 市　區　鄉　鎮 |
| 坊　街 | 第　號 |
| 特種戶之種類名稱 | |
| 調查票之號碼 | 第　號　共　張內之第　號 |

| 項目 | (甲) 現在人員（記入各戶內之在二十四年十二月三十一日之夜半十二點鐘所有現在人員） | (乙) 一時的或偶然的不在者（記入各戶之在二四年十二月三一日夜半十二點鐘之一時或偶然的不在者） |
|---|---|---|
| 號碼 | 1 2 3 4 5 6 7 8 9 10 ／ 現在人員總計 現在人員合計 | 1 2 3 |
| 姓名 | | |
| 與戶主之關係 | | |
| 男女之別　男 | | |
| 男女之別　女 | | |
| 出生之年月日 | | |
| 婚姻狀況 | | |
| 職業及職業上之地位　本業 | | |
| 職業及職業上之地位　副業 | | |
| 失業 | | |
| 識字與否 | | |
| 廢疾 | | |
| 所屬宗教名稱 | | |
| 出生地 | | |

(丙) 將甲欄內所有一時的或偶然的現在者記入之（例如來客等）

| | 姓名 | 常住所 |
|---|---|---|
| (1)(2) | | |
| (3)(4) | | |
| (5)(6) | | |

戶口調查員　姓名　蓋章

查右無誤

主管人　姓名　蓋章

此項只適用於公共處所

臨時戶口調查處

此二種票式實各有優劣點，現在評論優劣之前，實有預先說明當整理調查票時所用單獨法則之必要。其法則是當整理統計材料時，先將每一單位作成一張小票，而後由此小票以舉行整理。此方法實能使整理迅速正確，爲統計界所樂用。單記票當整理時，卽能將其調查票移用爲小票。反之列記票非於整理之前，先將各觀察單位謄移於小票上不能進行其整理工作。在理論上單記票比列記票爲優，但在事實上則又未能盡然。蓋因調查單位較爲少數及塡記慣熟時，單記票尙感便捷。若在戶口調查等以多數民衆爲對象時，使各人塡入一張實覺太繁，且各人塡入方法優劣不等，實不能將其調查票卽移用爲小票，故各國在戶口調查等大規模調查時，皆使用列記票。

依上述知小票法旣能統一進行，又能澈底明瞭現象狀況，其所得統計之正確與完全，更非表式法所能比擬。現將其優點列舉如次。

(1)易於實現數計。

(2)得爲各種調查。

(3)調查正確，

(4)下級機關得免去製表之煩。

(5)小票記入之當事人得依機械式以行塡寫。

(6)便於統計材料之整理。

(7)得保持單位觀察之統一。

(8)得能不變動其單位調查狀況以保存之。

至其劣點亦有次之數項：

(1)增加整理機關之經費。

(2)因各小票上易生錯誤,故多費上下機關間往復訂正之手續。

(3)票數太多,難於保存。

(4)小票之整理上需要多數業務人員。

觀察上述各優劣點知優多劣少。輓近世界各國統計事業之所以有長足之進步,其原因雖甚衆多,但使用小票法一項,實爲其主要原因之一。故我國在能力應許範圍內,以使用小票法爲上票。

## 第七節　調查主義

調查主義得分爲自記主義與他記主義二種。自記主義是使被調查者本身或其代表者負擔記入數計之責。他記主義是由調查員自己負擔記入數計責任。

在施行自記主義之由被調查者本身負擔記入時,對於沒有利害關係之調查事項,其結果實較他記主義爲正確,但欲實行此種主義則以國民教育之普及爲第一條件。

當記入調查票或調查簿時,有所謂馬約 (Mayr)之三則:(1)檢討記入之正確;(2) 期記入之完全無缺;(3) 使依順序整頓記入,此三種原則當記入調查票時宜確守之。

## 第八節　調查障礙

各種統計材料雖均由單位調查而來,但在實際上常不能得十分正確之結果。今先就親自調查以言(需要者親自調查其目的物)。第一對於各人之感官及精神作用實有十分正確之必要,對此不幸依醫學上及生

理上之關係,其感官及精神作用不免時起錯誤。又當親自調査之用機械以舉行時,對於機械構造上難期其有理想的正確度。再加在使用中易受外界(振動溫度濕度等)之影響,生伸縮及其他故障於機器上,以致愈不能得十分正確之結果。

當所欲調査之目的物在數量上非常衆多或廣大的分布(不集積羅列在一處所)時,如仍欲親自調査,實勢所不能。當此時需要者必須使用別人爲其間之報告者,在此等使用中間媒介物之調査,實多主觀的見解含在其中,故除上述障礙外,尙須增加數種別項障礙,西諺云『數無僞人僞之』(Figures will not lie, but liars will figure) 卽此之謂也。當使用他人爲報告者時所生障礙,約可分爲次之四種:

(1)從無智而生障礙。本障礙是由報告者對於調査之目的性質不十分明白了解,以致發生種種錯誤。例如調査今天晚上十二時正之人口時,在此瞬間前數秒間死亡人數及數秒後出生人數均不能算入在內,但在不十分明瞭調査性質者,仍將此等人數算入在內,以致起不正確結果。

(2)從惡意而生障礙。本障礙是由報告者對於調査之目的性質雖能十分了解,但爲便於減輕租稅或其他目的起見,特作不實或過少之報告,有時爲驅於羞恥之念故意隱蔽事實,有時因感情關係故意作虛僞之報告。

(3)從誇張而起障礙。本障礙是由報告者對於調査之目的性質雖能十分了解,但爲宣傳起見故意作誇大之報告。如出版界所發表發行部數,私人機關所受商人報告資料等,常常有此傾向。

(4)從不信而起障礙。本障礙是由報告者對於調查之目的性質不十分明瞭，或對於調查機關不十分信用，以致襲於疑心暗鬼，糊塗事實，作錯誤之報告。在此種障礙中之最顯著者，為由課稅恐怖所起之不實報告。

欲除去上述四種障礙，則須舉行下列數項手續。

(1)在舉行實地調查之先，舉辦擴大宣傳，使被調查者明白調查本旨及其內容。

(2)對於調查事項須附以詳細平易之說明，及多數實例於調查票之背面，使減少由無智而來之錯誤。

(3)實地調查時調查員之言語舉動須十分謹慎，使被調查者不致起惡劣感情。

(4)對於故意隱蔽事實或為虛偽報告者及造謠惑衆妨害調查者，須在事前確定法律的制裁，並須公告人民，使有所儆懼。

(5)普及統計思想，使被調查者明白了解自己所繳報告書只使用於統計上，在統計上之發表只有總體上之狀況並無各別詳情，依其發表實不能明瞭各調查單位之所屬情形。再保證其不移用於他處，嚴守其祕密，使報告者有無記名投票時為選舉者之心理狀況。

(6)於事前舉辦與他們生活真有利益之實際工作，使被調查者對於調查機關發生信仰。

## 第九節　統計材料之檢查與訂正

當調查多數單位時，各單位既依一空格式以記入確定後，其次之問

題是爲收集問題。但當收集時對於檢查一項，務須十分留意。蓋因調查員只將自己擔任區域內所配佈之調查票全部收集後，其調查材料尙不能算爲十分完備，必須施行次述二項檢查手續，而後始可。

(1)檢查其所欲調查之單位有無遺漏或重覆，若有時須立卽施行訂正。其訂正手續，切忌有主觀的行爲。

(2)檢查各單位應塡事項有無遺漏或錯誤，若有時須依實際狀況以行訂正。

## 第十節 統計材料之保管

調查票之檢查訂正工作終了後所收集之材料，始能使用於整理上，此等材料稱爲統計原料。

當舉行大規模統計調查時，調查票是先由主管機關印刷妥當，於所定調查時期前分發屬下。當分發時對於分發張數須有十分餘裕，而後始不致有臨時缺乏之虞。至調查終結後其調查票之收集上，大多數是須經過許多機關而後始能蒐集至整理機關。在直接統計上，其大量現象之移動，非常迅速，若在輸送中一有失誤，則不能再得當時實情，故對於輸送上須十分注意。

統計調查之原材料，在靜態調查是於一定時期內將全部調查票由地方收集至中央。在動態調查是將一定期間內所連續發生之事實，依期由地方報告至中央。此等調查在大規模舉行時，調查票之容積，有意想外之巨大，故必須在事前預定計劃，使得依一定次序以行排列，而後在整理時始能節省巨大勞力。又在直接統計，其所收集材料，欲在期後再得

同樣事實一事,爲勢所不能。故對於避免紛失損毀之保管上,須十分注意,對於防火設備,尤須十分完備。

將各地方送到之材料收入保管以前,亦須施行前節所述二次檢查,以確認其材料是否完整。如檢查結果發見錯誤時,當卽發還地方機關,使加以訂正。但遇檢查之結果,其材料確爲完整時,地方機關之責任已盡,其後如有紛失與錯誤,須中央機關負完全責任,故在收管前之檢查,須嚴密施行。

當調製統計表時,實屢次重複使用調查票,以致多次移動其保存位置。故事前必須將調查票之所在處,排列清楚,而後始能節省時間與勞力。又因屢次使用調查票易起紛失與毀損,故當貯藏時,必須每一區域所有調查票之二面各附以厚紙板,並於其外面以布帶固繫之。

當整理列記式調查票時,必須先將各調查單位由列記票移入於卡片 (Tabulation card, Zählkart) 上而後始可。在此時其直接用以調製統計表者爲此項卡片,故此項卡片與調查票有同等之重要性,其保管上亦須施以同樣之處置。但此項卡片如有紛失或毀損時,仍得由調查票以重製之。又此項卡片之使用次數,實比調查票爲多,故必須保存於出入便利之處。

## 第十一節 報告期之厲行

統計之生命分爲長短二種,其得爲學問上之研究資料者,其生命甚長,故與其不完全而早事發表,不若延遲詳細發表之爲是。至政府公司等之年鑑或報告等,其目的是爲表現一年間之成績故以早日報告爲

上策。現今吾國各機關對於統計材料之提出期日，每多延遲致不能迅速發表新鮮之統計，以應社會之需要，對此實應力加矯正，務使能依期報告。

# 第三章　統計類似調查方法

各種大量現象能一一在實地調查，當然爲吾人所希望。但實地調查一事，非只消費巨大勞力與金錢而已，且有時爲財力人力所不能，故學者特發明一種間接或估計方法，以求達到調查之目的。現將此項類似調查方法略述如次：

## 第一節　估計

此處所說估計 (Estimation, Schatzung) 並非指臆測，亂斷，完全是利用既存之統計資料，以求出近於事實之結果。次之數種爲其中之最普通者。

(A) 將過去之統計事實爲基礎，以計算同一大量現象之現在或將來之統計。

例如戶口調查是依一定週期循環舉行，現先利用相連二回所查得人口爲計算之基礎，以求出平均每年之人口增加率，而後再由此以求出二回間之尚未舉行調查各年及將來之人口狀況。其數式如次：

$P_A$ 爲 A 年年末全國人口總數，

$P_B$ 爲 B 年年末全國人口總數，

r 爲平均每年人口增加率，

$y=B-A$ 爲調查週期，

其 (A+1) 年年末人口爲

$$P_{(A+1)}=A\text{ 年年末人口}+(A+1)\text{ 年增加人口}$$

$$=P_A+P_A\cdot r=P_A(1+r);$$

(A+2) 年年末人口爲

$$P_{A+2}=P_{A+1}+P_{A+1}\cdot r=P_{A+1}(1+r)=P_A(1+r)^2。$$

將如此算法逐次進行,至 A+y=B 年年末人口之數式爲

$$P_{A+y}=P_B=P_A(1+r)^y。\cdots\cdots\cdots\cdots\cdots\cdots(1)$$

施對數於(1)式二邊得

$$\log.P_B=\log.P_A+y\log.(1+r),$$

$$\text{即}\log.(1+r)=\frac{\log.P_B-\log.P_A}{y}。\cdots\cdots\cdots\cdots\cdots\cdots(2)$$

由 (2) 式求得 r 後,其週期中間未舉行調查各年份及將來之人口得依次式以求得之。

$$\text{及}\left.\begin{array}{l}P_{A+a}=P_A(1+r)^a;\ a<y\\ P_{B+b}=P_B(1+r)^b。\end{array}\right\}(\text{公式}1)$$

(例)某市於 19 0 年年末調查得全市人口爲一百萬,至 1920 年年末再行調查得全市人口爲一百五十萬,問 1912 年及 1922 年人口爲幾何?

依題知

$$P_A=P_{1910}=1,000,000,$$

$$P_B=P_{1920}=1,500,000,$$

$$y=B-A=1920-1910=10,$$

依（2）式得

$$\log.(1+r)=\frac{\log.1,500,000-\log.1,000,000}{10}$$

$$=\frac{6\cdot17609-6\cdot00000}{10}$$

$$=0\cdot017609,$$

依此知

$$1+r=1\cdot04138,$$

依公式（1）得

$$P_{1912}=P_{1910}(1+r)^2$$

$$=1,000,000,\times(1.04138)^2$$

$$=1,084,472;$$

$$P_{1922}=P_{1920}(1+r)^2$$

$$=1,500,000\times(1.04138)^2$$

$$=1,626,708。$$

關於人口增加法則之數理的研究請參考 G. H. Knibbs, The Mathematical Theory of Population; E. Czuber, Mathematische Bevölkerungstheorie。

(B)就與所求某大量有密切關係之已知正確統計事實上，施行一種估計方法以求出某大量。

例如欲知全國一年間米之消費量時，得由依戶口調查所得各年齡階級別人口數上，乘以由家計調查所得各年齡階級別人員每人每年間

米之消費量後，即得。

在此種估計上，其爲基本之假定適當與否，關係於估計之價值甚大，故對此應特別注意。

## 第二節 部分調査

部分調査(Teilerhebung) 是在不能調査某大量之全體，只能依特殊方法調査其中之一部分，但其結果仍能與調査全部有同一効果時，舉行之。本方法與前節之估計不同，其目的並非是爲明瞭現象之總體，完全是爲明瞭該大量現象之內部構成狀態以舉行一切。

由此方法所得結果之正確度上，實大有研究之必要。若其正確度能相當信賴時，其方法之舉行，實比實地全部調査爲節省勞力與費用。現將其方法列舉如次：

### (A)選樣調査

於調査之現象中選定其可爲樣本單位者一部分，就此一部分之樣本單位上，舉行詳密之調査以求其結論。例如欲知勞工之生計狀況時，多用此方法，擇勞工中之得爲標準單位者一部分上以實地詳密調査其家庭收支狀況。本法爲法人勒普來(Le Play 1806-1882)所首創，現在各國之家計調査，皆應用此方法。至其選爲樣本之單位，大多數是擇其含有普遍性質者，但欲認識此普遍性質實須由大量觀察而後始能明瞭，現在不由大量觀察以定其普遍性。於理論上不免有調査者之主觀思想在內，其結論不免有假設處，但與觀察全體相較，則有能詳密觀察各單位之長處。

(B)任意抽出調查

本調查並非於總體內選擇樣本部份以舉行調查，是隨意抽出其中之一部份以觀察一切。其抽出手段完全委之於自然，無主觀之介入，其抽出數目能相當衆多時，則各種單位皆能包含在內，其總體上所表現之狀態亦能於相似比例以表現於其單位間，其總體之狀態亦得依此窺見。此任意抽出調查既可節省勞力與費用，其結果亦爲一般所認可，故多樂用之。但當舉行時，萬不能缺少次之二條件：

(1)其抽出單位須相當多數。

(2)其結果須在數學理論上所允許之誤差範圍內，檢查此實際上所起誤差是否在理論上所預定範圍內之方法，普通是以其標準差爲標準。(詳情參照第七編內大數法則論)

(C)不完全調查

前述之抽出調查，其最初目的是由任意抽出之一部份材料所得調查結果以求其全體狀態，但現在所說之不完全調查，是爲本欲進行全體調查，後因有特別事故只能終結其中一部份之調查。雖同爲部份調查，其意義完全相異。

對此調查其洩漏部份內若含有特殊性質單位時，其調查之結論與總體之實況有不能近似之虞，對此宜加以注意。

## 第三節　委員調查

對於欲研究之大量現象有特殊利害關係及有專門學識者，發文書或口頭詢問以期搜集能解決該問題之材料，此方法稱爲委員調查(Enquête)。

本調查方法爲英國所最初使用，於該國之議會內組織某種委員會，常發質問於委員外之專門學者及有關係者，以求其回答，現今英法德各國之官廳議會，及各種團體機關常使用此法，以求解決問題之資料。

# 第四章　中央集查與地方分查

將所有調查票完全集中於中央，由中央機關依一定格式以舉行分類表章等工作者，稱爲中央集查(Zentralisierte Ausbeutung)，反之由地方的調查機關擔任分類表章等工作以作成統計表後，只將此統計表呈繳至中央機關，由中央機關總計各區統計表以作成總統計表者，稱爲地方分查(Dezentralisierte Ausbeutung)。

此二種方法之得失依調查之性質與規模而不同，但一般是中央集查勝於地方分查，現將中央集查之優點列擧如次：

(1)得大量生產上之利益。

(2)使事務紛繁之地方機關得免去製表之勞。

(3)中央機關多統計專門家。

(4)得依一定規則以進行一切事務。

(5)得使用便利之機器。

(6)得減少製表上之錯誤。

(7)得調製行政上學術上所必需之精密的統計表。

(8)經費與時間均比地方分查爲少。對此依德之恩格爾(E. Engel 1821-1896) 於 1870 年在德國統計時報上所發表之實例，即能明白了解。

恩格爾實例

| | 事務日數 | | 經費(單位法郎) | |
|---|---|---|---|---|
| | 中央集查 | 地方分查 | 中央集查 | 地方分查 |
| (甲) | 104,496 | 191,100 | 557,370 | 746,530 |
| (乙) | 82,977 | 140,600 | 429,870 | 559,031 |
| (丙) | 48,632 | 116,100 | 230,651 | 465,281 |

備考: (甲)爲於列記票內劃線以檢查事實之工作。

(乙)爲使用單記票以移轉列記票內各項事實之工作。

(丙)爲整理單記票之工作。

依上述知中央集查實大比地方分查爲優,故在能力應許範圍內,以實行中央集查爲上策。

# 第三編　統計材料之整理

調查票旣已由大量觀察方法蒐集在一處後，其次之問題是爲整理所蒐集之材料以編製統計表統計圖及各種統計值。對此無論其爲直接統計或間接統計，靜態統計或動態統計，於原則上實無差異。茲將其方法依次說明：

## 第一章　分類

### 第一節　分類之意義

在大量調査票內擇其關於某品質之全然一致或互相類似之各個體聚爲一集團，此方法名之曰分類。其目的是在用簡單少數之集團，以表示複雜之大量現象。

對此分類中之最簡單者是將某大量現象之全體，依一定之特性大別之爲二類。例如人口之分男女，勞工之分有業與失業等是。此外則依性質之不同，其分類亦隨之而異。例如調查職工之每月工資時，因其所要目的之不同，得依五元爲間隔，分爲「五元未滿」「五元以上（含五元在內以下同）十元未滿」「十元以上十五元未滿」……等，或依一元爲間隔以區分一切，若更欲詳細時得以二角爲間隔以區分之。此等男女，有業者失業者，及五元未滿，五元以上十元未滿等區分，稱之爲組

(class)。其間隔之五元一元等稱之爲組距,(class interval)。在此等一定組內所存在之單位個數,稱之爲次數 (frequence)。將此等單位之全體依一定組別順序以區劃一切，並於此各組內記入其所屬之次數後之總表,稱爲次數分配表(frequency table, Haufigkeitstafel)。例如第6表爲十六名學生之某科考試成績，現依其分數最少者順次數計,知 50 分者只 F 一人，55 分者有 G, I 二人，60 分者有 J, N 二人,……如此依次進行後卽得第7表。依第7表吾人一見卽能明瞭十六名學生間之分數分配狀況,故稱此表爲全體學生分數之次數分配表。其50分 55 分……等爲所說之組,一人二人……等爲所說之次數。

第 6 表

| 學生 | 分數 | 學生 | 分數 |
|---|---|---|---|
| A | 65 | I | 55 |
| B | 70 | J | 60 |
| C | 85 | K | 75 |
| D | 90 | L | 65 |
| E | 75 | M | 70 |
| F | 50 | N | 60 |
| G | 55 | O | 70 |
| H | 70 | P | 65 |

第 7 表

| 分數 | 學生數 |
|---|---|
| 50 | 1 |
| 55 | 2 |
| 60 | 2 |
| 65 | 3 |
| 70 | 4 |
| 75 | 2 |
| 80 | 0 |
| 85 | 1 |
| 90 | 1 |
| 計 | 16 |

對此分類吾人得分數量上之分類,質量上之分類,地理上之分類及時間上之分類等四種,以後依次說明。

## 第二節 數量上之分類

數量統計上,對於決定各組之境界及組之個數等較質量爲任意。例如依十歲組別以區分人口時，其組別有依 10 歲以上 20 歲未滿，20 歲

以上 30 歲未滿以行分配者。有依 15 歲以上 25 歲未滿，25 歲以上 35 歲未滿以行分配者。甚至依 10.5 歲以上 20.5 歲未滿，20.5 歲以上 30.5 歲未滿以行分配，亦未始不可。其境界只依研究上之便利，以決定之可也。至組之個數，則依組距之大小而不同。如將 0 歲以上 100 歲未滿之人口，依一歲組別，五歲組別或十歲組別以行分配時，其組數得分爲 100，20 或 10 等。在組數稀少時，各組所屬次數變爲多數，易於消失次數間所存在之差異性質，有害統計調查之目的，若組數過多時，則又使大量之概觀趨於困難，在實際上又多費金錢與勞力。故對於組數，亦應取其適宜。

又爲便於計算數理統計上所使用各數值（比例，平均數及其他各種係數）及易於認識特質起見，對於組距以相等爲宜。但爲便於各種問題之研究起見，亦得互相不同。例如爲表示人口之城市集中現象起見，其城市人口得使用二萬未滿，二萬以上五萬未滿，五萬以上十萬未滿等不同組距。有時又得將全範圍仍各爲相等之組距 只於其中一部分施行細分。例如依五歲組別以區分人口時，爲便於詳細觀察幼年時期之狀態起見，特將 0 歲至 4 歲一組人口分爲各歲別。

研計表內有時得不明示組之最小值或最大值而開放其兩端或一端者。例如對於組之最大端或最小端只書某某以上或某某以下等文字。

## 第三節　質量上之分類

關於質量之品質多數是易於分類，如男女別及婚姻狀態等之分類，實甚簡單。只職業產業及死因等品質，因其本身之內容本甚複雜，以致

組之決定方法依統計家而不同。故當分類此等品質時，宜預先研究其性質，以便決定適當之組別。玆將各國之職業與產業上之大分類及我國衞生部頒布之暫行死因分類列表如第 8 第 9 第 10 以資參考在第 8 第 9 二表內，各國所有類別文字之左上角附有之數字號碼，則爲各國本來之排列順序，但表內則依拍提永氏在國際統計會議所提出順序以行排列之。

## 第四節 地域上之分類

依各單位之空間的品質以行分類者，實甚衆多。其分類方法約可分爲次之三種：

(1)組之順序，是依面積之大小以行分類。

(2)依行政上之區域以行分類。例如各省之依縣級以排列縣別統計。

(3)依地理上之區域以行分類。例如將山脉河川等之自然的地勢爲基礎以行分類。本分類法對於重視地理的關係之生產及經濟統計上，有重大之使命。

## 第五節 時間上之分類

依動態現象之發生期間以行分類，此期間之組別得分爲數年一年六個月三個月一個月一星期或一日等。對此分類法上宜將各組之期間互相一致，以便於比較。但在實際上依月別以分類時，各月之日數，有多少之不同，此點宜加以注意；

## 第六節 多數品質之分類

上述各節是依一個品質以行分類之方法。但一般是每次舉行調查必同時調查數種之品質，因此起同時分類二個以上品質之問題。對此宜先依一個品質以行分類，而後由此所得各組上所屬之單位，再依第二品質以行分類，將此方法繼續進行，遂得第三第四等品質之分類。其分類情形如次：

總數
- 第一品質所分
- 各組之次數
  - 第二品質所分
  - 各組之次數
    - 第三品質所分
    - 各組之次數

至品質次序之決定，則依統計問題及製表者而不同。至數理的分類方法請參考下列二書。

Yule, G. U. An Introduction to the Theory of Statistics, PP. 7-74. 1924.

Czuber, E. Die Statistischen Forschungsmethoden, SS. 4-34. 1927 (此書已由李仲珩君譯爲中文但其內容只有原書之最初二編)。

# 第二章　製表(Tabulation)

## 第一節　製表之意義

社會現象之複雜程度與日俱增,故雖將統計材料正確完全查得後,如仍不將其混亂存在之單位,整頓於確定組內並用表形以現出時,多數人士仍難以理解。此種表形吾人稱之爲統計表(Statistical tables)。

統計表是爲最易認識統計材料上各特質之表式,能使觀者一目瞭然卽能求得所需要之數值。其所以使用統計表者,實因統計上所得結果,皆爲數字。如將其數字依甲市之戶數爲二十五萬六千八百二十戶,人口爲一百十七萬三千一百六十二人。乙縣之戶數爲七千六百六十五戶,人口爲三萬八千九百五十三人。丙縣之戶數爲五萬二千四百零五戶,人口爲二十五萬三千八百七十五人。丁縣之戶數爲五萬九千零九十二戶,人口爲二十八萬七千三百八十二人。戊縣之戶數爲一萬二千四百九十五戶,人口爲二萬零七百七十七人等式樣以書寫時。對於看者非但不能將正整之順序映入目中,且其一市四縣間之戶口相差程度與其數字間有無錯誤等情,均難以明瞭。現將上項一市四縣之戶口狀況依下列式樣以書寫時,卽能一目瞭然,明白其全體狀況。知甲市嶄然特大,乙縣非常稀少,丙縣與丁縣相差無幾,戊縣之戶數有一萬二千餘戶而人口只有二萬零七百七十七人,每戶人數平均不到二人,未免太少,其二項數

字中必有一方錯誤,現就原有材料檢查結查,知人口之二萬實爲六萬之誤,此等錯誤處在前述式樣書寫內,實難以發現。

| | 戶數 | 人口 |
|---|---|---|
| 甲市 | 256820 | 1173162 |
| 乙縣 | 7665 | 38953 |
| 丙縣 | 52405 | 253875 |
| 丁縣 | 59092 | 287382 |
| 戊縣 | 12495 | 20777 |

以上所述雖爲極簡單一例,但將其數字只依普通文字記述的方法書寫時,其內容何等混亂糊塗。若能依統計表體裁書寫時,則又何等簡單明瞭。故知欲簡單明瞭以表現大量現象之眞相時,實非使用統計表不可。

## 第二節 製表之方法

統計表之格式甚衆,其最簡單者爲於一個品質上給與各組所屬單位數,或於名義上爲二個以上之品質但表內各品質間全無關係只於一方記入各品質之組別他方給與各組所屬單位數而已,此等統計表實難發現各品質間之關係。例如將甲市戶口調查所得人口上之男女,婚姻,年齡等狀況列爲第 11 表。

依第 11 表雖能明瞭甲市人口對於上述各品質之構成狀況,但欲知男女二方面各自之婚姻別或各年齡組別之人口數,爲事實所不能,故須加以改正。現約略改之如第 12 表。

依第 12 表雖能知各年齡組別與男女間之婚姻關係,但尙不能稱爲完善。例如不要年齡別之婚姻關係只要男女年齡別人數,或不要年齡別只要男女總體上之婚姻關係時,在上表內不能卽刻明瞭,非再加以計算

不可。故必須求出男女二方面之總數，男女各方之總數，如第13表後，始能稱一完善之統計表。

第 11 表

| | | |
|---|---|---|
| 性別 | 男 | 650,780 |
| | 女 | 636,367 |
| 婚姻狀態 | 有配偶 | 497,529 |
| | 未婚 | 672,034 |
| | 鰥寡 | 100,405 |
| | 離婚 | 17,179 |
| 年齡 | 0—9 | 285,435 |
| | 10—19 | 279,126 |
| | 20—29 | 212,446 |
| | 30—39 | 166,680 |
| | 40—49 | 143,310 |
| | 50—59 | 94,570 |
| | 60—69 | 67,519 |
| | 70—79 | 29,718 |
| | 80—89 | 5,032 |
| | 90—99 | 304 |
| | 100— | 7 |

第 12 表

| 年齡 | 男 | | | | 女 | | | |
|---|---|---|---|---|---|---|---|---|
| | 有配偶 | 未婚 | 鰥夫 | 離婚 | 有配偶 | 未婚 | 寡婦 | 離婚 |
| 0—9 | —— | 144,246 | —— | —— | —— | 141,189 | —— | —— |
| 10—19 | 770 | 141,728 | 10 | 49 | 7,636 | 128,537 | 62 | 334 |
| 20—29 | 40,649 | 70,691 | 841 | 1,307 | 70,538 | 27,288 | 1,577 | 2,561 |
| 30—39 | 47,120 | 7,014 | 2,688 | 1,925 | 70,007 | 3,753 | 4,744 | 2,429 |
| 40—49 | 65,089 | 2,144 | 4,713 | 1,790 | 55,155 | 1,767 | 10,376 | 2,282 |
| 50—59 | 39,263 | 976 | 5,743 | 1,209 | 28,959 | 926 | 16,143 | 1,351 |
| 60—69 | 22,263 | 607 | 7,346 | 743 | 13,203 | 605 | 22,037 | 715 |
| 70—79 | 6,397 | 195 | 4,509 | 201 | 2,743 | 279 | 15,161 | 233 |
| 80—89 | 574 | 27 | 875 | 19 | 139 | 54 | 3,317 | 27 |
| 90—99 | 23 | 1 | 40 | —— | 9 | 4 | 223 | 4 |
| 100— | 1 | —— | —— | —— | —— | —— | 6 | —— |

# 第 13 表

| 年齡 | 全縣 | | | | | 男 | | | | | 女 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 總數 | 有配偶 | 未婚 | 鰥寡 | 離婚 | 總數 | 有配偶 | 未婚 | 鰥夫 | 離婚 | 總數 | 有配偶 | 未嫁 | 寡婦 | 離婚 |
| 全縣 | **1,287,147** | 497,529 | 672,034 | 100,405 | 17,179 | **650,780** | 294,140 | 367,632 | 26,765 | 7,243 | **636,367** | 248 389 | 304,402 | 73,640 | 9,936 |
| 0- 9 | 285,435 | — | 285,435 | — | — | 144,246 | — | 144,246 | — | — | 141,189 | — | 141,189 | — | — |
| 10-19 | 279,126 | 8.406 | 270,265 | 72 | 383 | 142,557 | 770 | 141,728 | 10 | 49 | 136,569 | 7,636 | 128,537 | 62 | 334 |
| 20-29 | 212,446 | 111,178 | 97,982 | 2,418 | 3.868 | 113,482 | 40,640 | 70,594 | 841 | 1,307 | 101,964 | 70,538 | 27,288 | 1,577 | 2,561 |
| 30-39 | 166,680 | 144,127 | 10,767 | 7,432 | 4,354 | 85,747 | 74,120 | 7,014 | 2,688 | 1,925 | 80,933 | 70,007 | 3,753 | 4,744 | 2,429 |
| 40-49 | 143,310 | 120,244 | 3.911 | 15,083 | 4,072 | 73,736 | 65,089 | 2,144 | 4,713 | 1,790 | 69,574 | 55,155 | 1,767 | 10,370 | 2,282 |
| 50-59 | 94,570 | 68,222 | 1,902 | 21,886 | 2,560 | 47,191 | 39,293 | 976 | 5,743 | 1,209 | 47,379 | 28,959 | 926 | 16,143 | 1,351 |
| 60-69 | 67,519 | 35,466 | 1,212 | 29,363 | 1,458 | 30,959 | 22,263 | 607 | 7,346 | 743 | 36,560 | 13,203 | 605 | 22,037 | 715 |
| 70-79 | 29,718 | 9,140 | 475 | 19,670 | 434 | 11,303 | 6,397 | 195 | 4,509 | 201 | 18,416 | 2,743 | 279 | 15 161 | 233 |
| 80-89 | 5,032 | 713 | 81 | 4,192 | 46 | 1,495 | 974 | 27 | 875 | 19 | 3,537 | 139 | 54 | 3,317 | 27 |
| 90-99 | 304 | 32 | 5 | 293 | 4 | 64 | 23 | 1 | 40 | — | 240 | 9 | 4 | 223 | 4 |
| 100- | **7** | 1 | — | 6 | — | 1 | 1 | — | — | — | 6 | — | — | 6 | — |

統計調查普通是在每一單位上調查三個以上之品質，此等品質實不能將其全部在一頁統計表上表現出來，故必須用多種統計表以表現之。普通每一統計表之組合上最多使用四個品質。蓋因使用多數品質以詳細分類時，各組所屬單位數太少，實有妨礙大數法則之適用，及損害統計價值之虞，對此宜注意之。

統計表之製作情形已如上述，必須將各種關係盡述於表上，使閱者能一目瞭然即能求得所要之數字。但此等關係是否能完全列入，其表內之排列是否能一目瞭然等之根據，完全在製表者之技術上。吾國各機關所發表之統計表，尙多不妥處，其原因實由各製表者對於製表技術，尙未十分熟練，以致所作成之統計表，多不合原理，使人難於理解。至製表技術中之最重要者，爲表形之製作與表內文字之書寫，對此二點能十分熟練時，其他各問題自能依次解決。現今吾國各機關之統計表內數字，多使用阿剌伯字，故現將表形之作成及文字之書寫，依橫書式之統計表逐次說明，將下列多數實例明白了解後，其記入中國數字之縱書式統計表，亦易於製作矣。

第14表　歷年生產死產男女別

| 年次 | 生產 | | | | | | | 死產 | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 總數 | | | 親生 | | 私生 | | 總數 | | | | 親生 | | | 私生 | | |
| | 總數 | 男 | 女 | 男 | 女 | 男 | 女 | 總數 | 男 | 女 | 未詳 | 男 | 女 | 未詳 | 男 | 女 | 未詳 |
| 民國　年 | | | | | | | | | | | | | | | | | |
| 民國　年 | | | | | | | | | | | | | | | | | |
| 民國　年 | | | | | | | | | | | | | | | | | |

第15表　歷年生產死產男女別

| 種別 年次 | 生產 | | | | | | | 死產 | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 總數 | | | 親生 | | 私生 | | 總數 | | | | 親生 | | | 私生 | | |
| | 總數 | 男 | 女 | 男 | 女 | 男 | 女 | 總數 | 男 | 女 | 未詳 | 男 | 女 | 未詳 | 男 | 女 | 未詳 |
| 民國　年 | | | | | | | | | | | | | | | | | |
| 民國　年 | | | | | | | | | | | | | | | | | |
| 民國　年 | | | | | | | | | | | | | | | | | |

現先將表內各部分名稱依實例以行說明，第 14 表上部之歷年生產死產男女別等文字稱爲表題 (title)，年次欄縱行下處稱爲表側 (Row heading)，其民國元年民國二年等文字稱爲表側文字，年次欄右側稱爲表頭(Column heading)，其生產死產親生私生總數男女等文字稱爲表頭文字，表側之右與表頭之下所有空白處，是記入應塡之數字，此處稱爲表身。此等表題表側表頭等之文字，旣須簡單明瞭，又須完全能明示統計表之內容。

統計表之組合上所用線紋，種類雖甚衆多。但一般是於表身與表頭表側相隔處，引以最粗之直線「——」。在二個以上之品質相並由此再行分類時，其品質間宜引以二重線「══」，如第 14 表內之親生與私生，第 15 表內之生產與死產均須由此再行分類者，故其間均引以二重線。其餘各種品質內之各項分別處，宜引以細直線「——」，如第 14 表第 15 表內之於男女分別處引以細直線。若表頭內容較爲複雜時，仍得使用數等之粗細線以分隔之，至表之外側近來統計報告書上漸趨於不劃直線，但在手製之統計表上，仍多使用粗細合成之二重線「══」，及其他各種粗細

線以包圍之。

第16表 主要農產物收穫額

| —— | 米 | | | | 麥 | | | 大豆 | 小豆 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 粳米 | 糯米 | | | 大麥 | 小麥 | | | | | | |
| 全國 | | | | | | | | | | | | |
| 民國 年 | | | | | | | | | | | | |
| 省別(民國 年) | | | | | | | | | | | | |
| 江蘇省 | | | | | | | | | | | | |
| 浙江省 | | | | | | | | | | | | |

第17表 貨幣鑄造額及發行額

| —— | 總數 | 銀元 | 銀角 | 銅幣 |
|---|---|---|---|---|
| 鑄造額(元) | | | | |
| 民國A年 | 10,767,190 | 7,596,524 | 2,240,168 | 930,498 |
| 民國B年 | | | | |
| 發行額(元) | | | | |
| 民國A年 | 10,766,400 | 7,596,000 | 2,240,000 | 930,400 |

第14表爲普通統計表之模形，其組合線紋爲「══」「══」「──」「──」等四種。

第15表爲第14表之變形，其組合線紋爲「══」「══」「──」「──」「～～」，及斜線等六種，與第14表相較多括弧斜線二種，此括弧不過代替直線而已無甚特別意義，隨製表者之所欲，直線或括弧均可。至左角欄內之斜線，在吾國之由中文數字直書式統計表上，實多使用之。此斜線在特殊情形上，或得說爲有意義之線紋，但普通是無甚理由與用處，蓋因吾人雖不使用此斜線及其兩邊上之文字，亦能明瞭表頭表側內之意義，故在外國統計表上，此斜線絕對不使用，只將左角欄內剩爲空白或引以橫細直線或書入能區別一切之文字而已，至吾國歷來統計表上之所以使用此等斜線者，大約在最初輸入統計學時，將統計學解作爲簿記學之一種，因簿記表上使用斜線，故將其應用至統計表上，此斜線使用與否並非重大問題，但吾國人士對於統計表上，多有非用此斜線不可之觀念，故特略加以解說。此斜線能不用時，當然可省去不少手續。

第16表爲全體與部分並記之例，其組合線紋爲「══」「──」「─」等三種，左角欄內引入一橫短細直線，表身處劃分爲全國與省別二部。

第18表　物價

| | 米（每石） | 柴（每擔） | 醬油（每斤） | 鹽（每斤） | | |
|---|---|---|---|---|---|---|
| 民國十八年 | 16元4角 | 1元5角4分 | 1角64 | 4分6厘 | | |
| 民國十九年 | 18元2角 | 2元1角6分 | 1角87 | 5分3厘 | | |

第19表 原籍人口各性各歲階級別

| 年齡 | 性別 | 民國二年末 | 民國七年末 |
|---|---|---|---|
| 1- 5 | 總數 | 7,116,516 | 7,309,749 |
| | 男 | 3,544,163 | 3,647,052 |
| | 女 | 3,517,363 | 3,612,697 |
| 6—10 | 總數 | 5,918,864 | 6,746,639 |
| | 男 | 2,998,351 | 3,409,666 |
| | 女 | 2,920,513 | 3,336,973 |
| 11–15 | 總數 | | |
| | 男 | | |
| | 女 | | |

第20表 某省漁業戶數

| 年次 | 本業 | 副業 |
|---|---|---|
| 民國元年 | 12,345 | |
| 民國二年 | 12,567 | |
| 民國三年 | 13,412 | |
| 民國四年 | 8,789 | 5,671 |
| 民國五年 | 8,951 | 6,127 |

以下各表是對於組合與書寫相並說明。

第 17 表為表示統計表內數字單位之例，其組合線紋為「——」「——」二種，其格式與第 16 表相類似，只於鑄造額及發行額之上下二

側引以直線，其幣額單位不記入於各數字之右側上，而只記入一「元」字於鑄造額及發行額等文字之右旁。

第 18 表爲記入統計表之數字單位及必要文字於相當欄內之例，其組合線紋爲「══」「━━」「──」等三種，其表頭處除記入品質名稱外，尙須記入數值單位，其表身處除應塡數字外，尙須記入物價之單位。

第 19 表爲明白統計表內數列間隔之例，其組合線紋爲「══」「━━」「──」「〰」等四種，並於每一年齡階級間加一空行，以期鮮明數列之區別。

第 20 表爲記入調查情形相異時各數字之例，其組合線紋爲「━━」「──」二種，表內之自民國元年至三年是只調查漁業總戶數，不分爲本業與副業，故將其戶數記入於本業與副業之中間，民國四年以後本業與副業分別調查，故得將其戶數分別記入。

第 21 表爲表示單位之例，其組合線紋爲「══」「══」「━━」「──」等四種，其與第 17 表不同處爲左角欄內不引横短線及鑄造額與發行額之間各年度欄與總數之間引以二重線而已。

第 22 表爲說明統計表內符號之例，其組合線紋爲「══」「══」「━━」「──」等四種，表內之數字遇有須特別說明之必要時得附以適宜之符號，並將其符號之原因說明於備考內，由此得知備考注意等項，於必要時亦得記入之。

第 23 表爲分割統計表之例，其組合線紋爲「══」「━━」「──」等三種，當統計表過於狹長時，得適宜分割爲數段，本表是只分割爲二段。又因其右表之上部須繼續於左表之下部，故於左表下端不劃横線表

第21表 貨幣鑄造額及發行額

| | 總數 | 銀元 | 銀角 | 銅幣 |
|---|---|---|---|---|
| | 鑄造額(元) | | | |
| 民國 年 | 10,767,190 | 7,596,524 | 2,240,168 | 930,498 |
| 民國 年 | | | | |
| | 發行額(元) | | | |
| 民國 年 | 10,766,400 | 7,596,000 | 2,240,000 | 930,400 |

第22表 保險公司資本金及公積金

| 年次 | 公司數目 | 基本金(元) | | 公積金(元) | |
|---|---|---|---|---|---|
| | | 總額 | 已付額 | 準備金 | 其他 |
| 民國甲年度 | 36 | 14,588,000 | 4,387,325 | | |
| | | 800,000* | 425,000* | | |

備考: 表內附有*者表示其兼營別業之金額

第23表　進口貨國別價值(二月份)

| 種類及國名 | 價值(關平兩) | |
|---|---|---|
| | 民國十七年 | 民國十八年 |
| 本色市布 英國 | 457,393 | 144,957 |
| 日本 | 457,611 | 542,830 |
| 其他各國 | 33,409 | 323,385 |
| 、 | | |
| 、 | | |

| 種類及國名 | 價值(關平兩) | |
|---|---|---|
| | 民國十七年 | 民國十八年 |
| 棉紗 英國 | — | 952 |
| 香港 | — | 1,440,072 |
| 日本 | 105,307 | 50,876 |
| 其他各國 | 192,522 | 1,406,278 |
| 、 | | |
| 、 | | |

第24表　國外貿易狀況(單位千兩)

| 年次 | 總數 | 輸入額 | 輸出額 |
|---|---|---|---|
| 民國十年 | 1,507,378 | 906.122 | 601,256 |
| 民國十一年 | 1,599,942 | 945,050 | 654,892 |
| 民國十二年 | 1,676,320 | 923,403 | 752,917 |
| 民國十三年 | 1,789,995 | 1,018,211 | 771,784 |
| 、 | | | |
| 、 | | | |

示其尙須繼續。

第 24 表是表示統計表上各欄幅度得廣狹不同之例，組合線紋爲「━━」「──」二種，當表內各欄數字之行數有多少不同時，得將各記入欄之幅度，依行數之情形變爲廣狹之不同，本表之總數欄內數字之行數，較他欄爲多，故其幅度亦得較他欄爲廣。

第 25 表爲表示統計表內之數字每五行空一行之例，組合線紋爲「══」「━━」二種線紋，「══」在外國統計表上時常使用。本表內之所以每五行空一行者，實爲努力使統計表上各數字之顯現起見，對於橫線除於品質與數字分界處劃線外，其餘概不使用，只於其表側各品質(本表爲年度)內，每五行空一行，以便於區別而已。

第 26 表爲折轉縱短橫長統計表之例，組合線紋爲「══」「━━」「──」等三種，遇縱短橫長之統計表時，得將橫之部分適宜縱斷以折轉之。此折轉方法旣能使表之體裁長短適度，又可減少印刷頁數。

第 27 表是爲附與號碼於統計表內及記入符號於無數字可塡處之例，組合線紋爲「══」「━━」「──」等三種，當橫之組別多數時，於表上甚難求得其屬於同組之數字，故須於表之左右二側記入分類各組名稱或附以順序之號碼，(本表是附以順序之號碼)以期易於找尋。又在表身內之無數字可塡處，必須劃一橫短線「──」或圓點「・」以期形式之整齊。

以上各表內總數之位置皆放在最初項內，與從來之放在最末一行有不同處，蓋因統計表上之總數，其目的並非是算術上之合計，是用以表示總體之狀況，故必須將此放在最易注目之第一行內，但當此時應將

## 第25表 國外貿易總價額

| 年次 | 總價額 | 年次 | 總價額 |
|---|---|---|---|
| | 千兩 | | 千兩 |
| 民國元年 | 843,617 | 民國十一年 | 1,599,942 |
| 民國二年 | 978,469 | 民國十二年 | 1,676,320 |
| 民國三年 | 825,468 | 民國十三年 | 1,789,995 |
| 民國四年 | 873,337 | 民國十四年 | 1,724,218 |
| 民國五年 | 998,204 | 民國十五年 | 1,988,516 |
| 民國六年 | 1,012,451 | 民國十六年 | 1,931,552 |
| 民國七年 | 1,040,776 | 民國十七年 | 2,035,388 |
| 民國八年 | 1,277,807 | ,, | |
| 民國九年 | 1,303,881 | ,, | |
| 民國十年 | 1,507,378 | | |

## 第26表 各種工業戶數

| 年次 | 造絲 | 機業 | 革類 | 造紙 | 油類 | |
|---|---|---|---|---|---|---|
| 民國甲年 | 346,279 | 427,636 | 20 | 15 | 5,716 | |

| 年次 | 酒類 | 醬油 | 肥皂 | 洋火 | 漆器 | |
|---|---|---|---|---|---|---|
| 民國　年 | | | | | | |

| 年次 | 製茶 | 製粉 | 罐頭 | 釦 | | |
|---|---|---|---|---|---|---|
| 民國　年 | | | | | | |

第 27 表

各國工業生產額

| 地方 | 製絲 | 織物 | 鐘表 | 革類 | 紙類 | ,, | ,, | , | ,, | 地方號碼 |
|---|---|---|---|---|---|---|---|---|---|---|
| 全國 | 546,542,789 | 1,189,271,099 | 3,950,468 | 34,713,274 | 103,087,117 | | | | | |
| 1 甲省 | 24,655 | 1,029,155 | — | 647,006 | 10,548,055 | | | | | 1 |
| 2 乙省 | 31,211 | 438,807 | — | 13,445 | — | | | | | 2 |
| 3 丙省 | 3,448,288 | 274,221 | — | 1,335 | — | | | | | 3 |
| 4 丁省 | 611,543 | 399,539 | — | 36,287 | — | | | | | 4 |
| 5 戊省 | 12,610,481 | 17,586,027 | — | 55,098 | — | | | | | 5 |
| 6 己省 | 13,131,791 | 1,903,199 | — | 2,937 | — | | | | | 6 |
| 7 庚省 | 3,334,515 | 40,562,922 | 506 | 7,500 | 80,000 | | | | | 7 |
| 8 辛省 | 4,417,915 | 53,428,668 | — | — | 609,000 | | | | | 8 |
| 、、 | | | | | | | | | | |

總計合計等名稱，改爲總數或其他表示全體意義之文字。

對於數字之記入除上列各表內所示各注意點外，尙須注意下述各點：

(1)明示小數點與數位。

a.「·」是爲小數點，例如 135.5。

b.「，」是用於數位，例如 12,345.5 對此數位吾國古法是每四位加一點，但在今日大多數是使用每三位加一點，求與各國相一致。

(2)小數點下之位數。

整數有二位以上時，其小數點下之位數有一位已足。如整數只有一位時，其小數點下之位數須有二位以上，

(3)數字書寫上之注意。

a. 數位須排列淸楚，字須正寫。

b. 雖爲同一數字而行欄不同時，須全部重寫，切忌寫「同」或「，，」等字樣。

c. 各欄內數字之前方，必須有一行之空白處，例如 | 12345 | 23456 | 之連續相並，則有失一目瞭然之精神。

d. 中國數字是用一二三四五等字而不用壹貳叁肆伍等字，用中國數字直書時，一一與二，一二與三，一三與二二等字，易於看錯，故於書寫時應特別注意。

關於製表技術上，除上述線紋之組合與數字之書寫外，尙須注意次之各點：

(1)對於表之頁數須預先決定其能否合爲一頁或分割爲數頁。

(2)對於表內之記入數字,須預先決定其是否記入實際上之數字(絕對數)或比例數。例如表示各國之農業產額時,記入絕對數或比例數均可。如表示城鄉疾病人數比較時,則以記入比例數(每千人)較爲便利。

(3)應依適宜程度以整理欄之數目,遇必要時得設立「其他」「雜」及「不詳」等項目。

(4)記載表頭處宜給與適當之寬度。

(5)記入工作完了後,必須舉行檢查一次,對此檢查方法是先計算各欄及各行所記入各數字之和,是否與原有各數字之和相符合,而後再計算各欄數字和之總計是否與各行數字和之總計相符合,如不符合時,則各欄或各行中必有錯誤處。

(6)對於調查時期與區域,必須明示於表之外側。

(7)表之體裁依使用中國數字與阿剌伯數字而不同,使用中國數字縱書時,普通將分類組別置於右側及上部,使用阿剌伯數字時,則置於左側及上部。

(8)各欄及各行內所記入數字,爲便於認識起見,必須將總數一項數字比普通各項數字爲大。若有數種總數時其合計之總數比分計之總數印刷字碼須大一號,總計之總數比合計之總數再大一號,餘類推。

(9)表題之排列須依表側之所在處爲起點以定其次序。例如表側在右方時,表題之順序是由右至左,否則相反。

# 第三章　整理

## 第一節　整理之意義

分類與表式既已決定後，其次之工作則為整理調查票以求填入之數字。對此工作約可分為次之二段：

(1)遠心工作　依所定之時間空間與事實三項標準，以分撰各觀察單位。例如將全國人口依各年齡別逐次分撰以求各年齡別人口數，

(2)求心工作　於已分撰之各觀察單位中，擇其屬於一定標準以內者使集合為一起。例如將上例各年齡別人口，依五歲或十歲年齡組別集合在一起。

此二段工作總稱之為整理事務。如欲迅速完成此等事務時，則有設立工作進行表之必要。第28表與第29表為江蘇省民政廳所整理各縣縣政府及其屬下各機關工作人員調查票時。所用各種工作進行表中之一部份。

工作進行表之使用方法，是先由下列各種整理方法將統計材料內所有各種情形，變為數字或其他之符號以記入於工作進行表內，再由此工作進行表求得各組所屬數目，而後始將此等數目依次移入於統計表內，使成為完善之統計表。由此知工作進行表實為由統計資料以作成統計表間之媒介物，但在移動數目至統計表上以前，須在工作進行表內核

第 28 表 工作進行表(收發)

着手 月 日 時 分
完成 月 日 時 分
擔任者
縣

| 機關別 | 交到票數 | 要重審票數 | 實際票數 |
|---|---|---|---|
| 總數 | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |

第 29 表 工作進行表(分類) 1, 職務及薪俸別

着手 月 日 時 分
完成 月 日 時 分
擔任者
縣 機關

| 職別 | 人員 | | | 月俸 | 未詳票數 | |
|---|---|---|---|---|---|---|
| | 計 | 男 | 女 | (單位元) | 人員欄 | 月俸欄 |
| 總數 | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |

算縱橫二方面之總和是否符合，以檢查數字之有無錯誤。

至實地整理方法約可分爲次之三種：

(1)劃線法(Tick system, Strichmethode)

(2)數票法(Card system, Verfahren mittelst Zählblättchen)

(3)機器法(Mechanic system, Mechanishemethode)以後依次說明：

## 第二節　劃線法

本法在觀察單位不能各別分離及材料不甚衆多時，得常常利用之。

其方法是先在調查票內所有各觀察事項上，施以特殊符號，而後吾人觀察此符號之數目，用「正」「▱」「卌」「⁛」等特殊符號記入於工作進行表內，如第 30 表內情形然。本法雖甚簡單，但遇調查事項複雜時，常生劃線上之錯誤。又當調查票衆多時，爲記入此等多數同一符號起

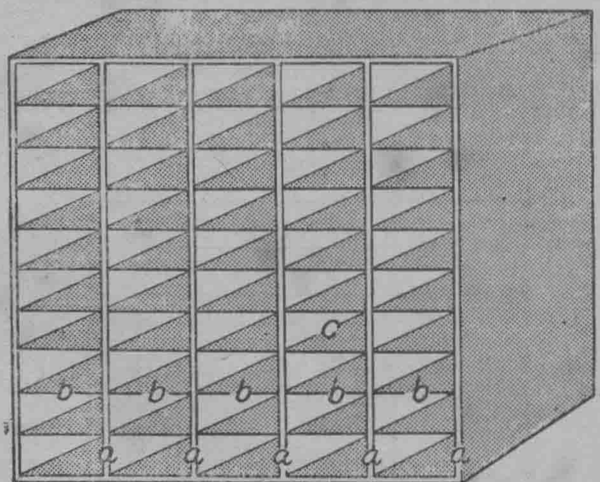
第 30 表

| 年齡 | 人數 | |
|---|---|---|
| | 符號 | 計 |
| 總數 | | 87 |
| 23 | 正正丅 | 12 |
| 24 | 正正正下 | 18 |
| 25 | 正正正正𤴓 | 24 |
| 26 | 正正正丅 | 17 |
| 27 | 正𤴓 | 9 |
| 28 | 正丅 | 7 |

見，工作進行表內須有廣大之空欄。且在工作半途中無檢查之可能，遇最後核算總數不符時，又非從頭再舉行劃線一次不可。

## 第三節 數票法

先將每一觀察單位及其品質，由列記式調查票以移至一頁小票上，（在單記票時即以其調查票當作爲小票）而後用徒手分撰小票內之品質相一致或在一定組別內者歸爲一起後，再用徒手計數其小票即能求得各品質所屬數目，並再將此項數目塡入於工作進行表內，則已完成整




第 一 圖

(a)爲明記其所屬處

(b)爲底板，可自由移動如一格之地位不足時得移去底板以擴張其放入小票之高度

(c)爲放入小票處

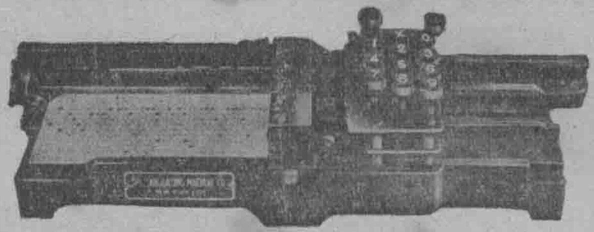
其箱之背面多以鐵絲網爲之使光線充足易於檢查

理之主要工作。本方法能將複雜之分類依次進行，但遇分類組別多數時，只在平面桌上實覺面積太小，且易於混亂。故為增高分類能力及防止混亂起見，使用一種分類箱，如第1圖。

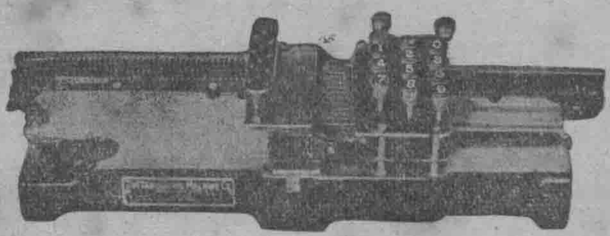
## 第四節 機器法

將列記式調查票上各單位之品質，先用數字符號定其號碼，而後用各種穿孔機(Key punch 第2圖)依其號碼穿孔於特種卡片〔Tabulanting card 第3圖(甲)為原有卡片圖形(乙)為已穿孔後之圖形〕上，每一單位用一張卡片，依卡片上孔穴之所在，即能明瞭其品質狀況。將多數已完結穿孔工作後之卡片，合為一起，放入電動分類機(Electric sorting machine 第4圖)之一方箱內，再加薄鐵片於其上，並同時將其欲分類之孔穴所屬項目，對準穿孔針，而後開始運轉機器，即能將各卡片依次分類。同時其分類之結果，亦能在左邊之數字記錄盤上，表現出來。吾人只要將此記錄盤上之數字，記入於工作進行表內，即已完成整理工作。其機器之分類速度每分鐘330張，工作能率實數倍於工人，故為現代各國大規模調查所樂用。

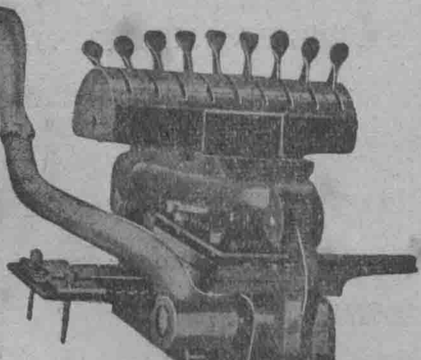
機器法內除上述各種機器外，尚有依分類之進行同時印刷其數字及其合計數之一種印刷製表機(Automatic printing-tabulator 第5圖)為貨銀統計上之必需品。



（甲）



（乙）

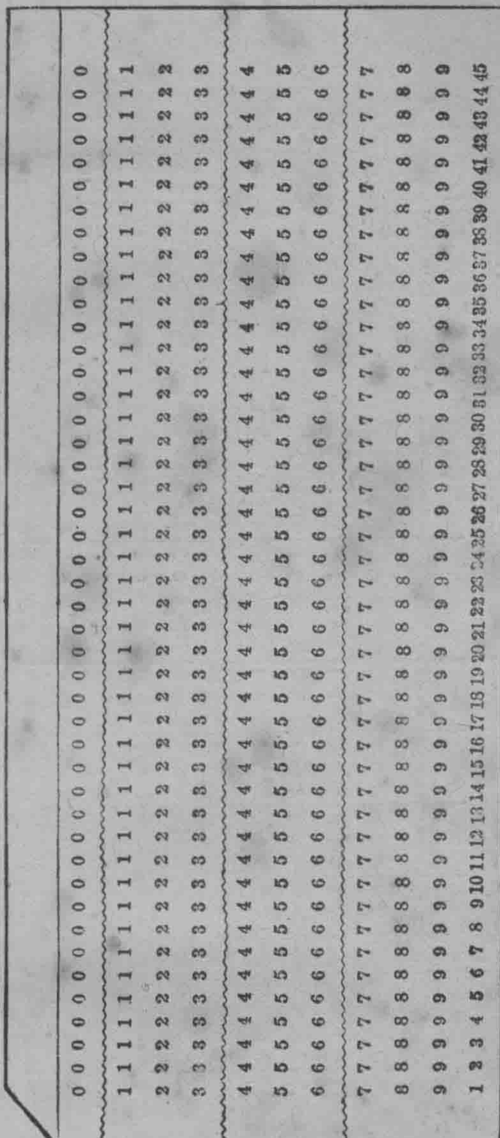


（丙）

第 2 圖

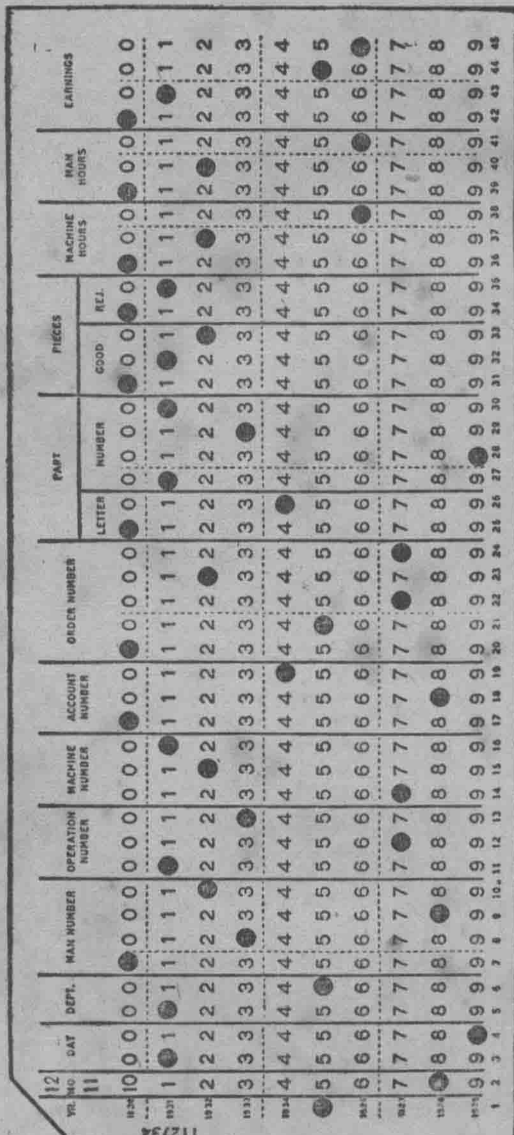
用以穿孔之卡片

(The tabulating card)

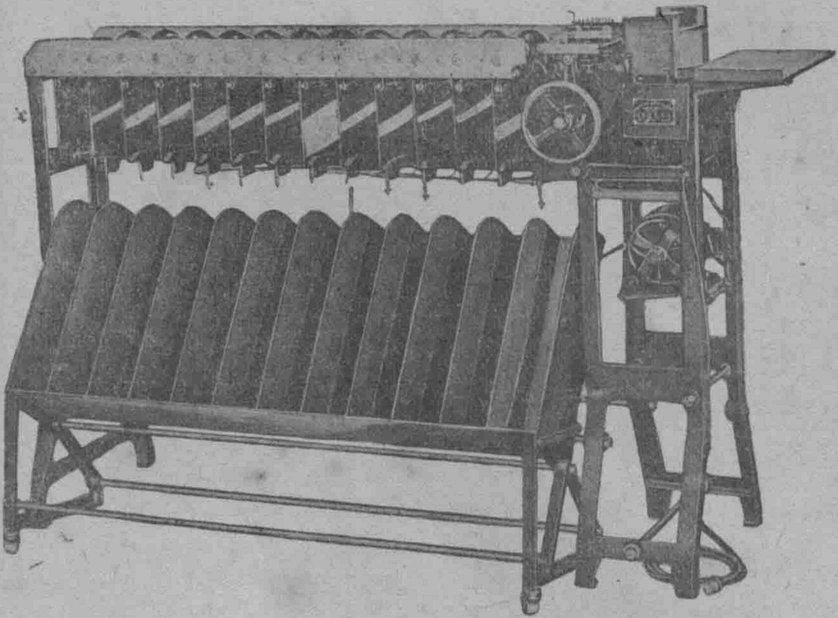


第 3 圖

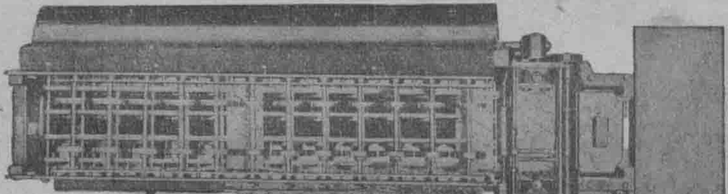
(甲)


YR.
MO.
DAY
DEPT.
MAN NUMBER
OPERATION NUMBER
MACHINE NUMBER
ACCOUNT NUMBER
ORDER NUMBER
PART
LETTER
NUMBER
PIECES
GOOD
REJ.
MACHINE HOURS
MAN HOURS
EARNINGS
112734

第 3 圖 (乙)

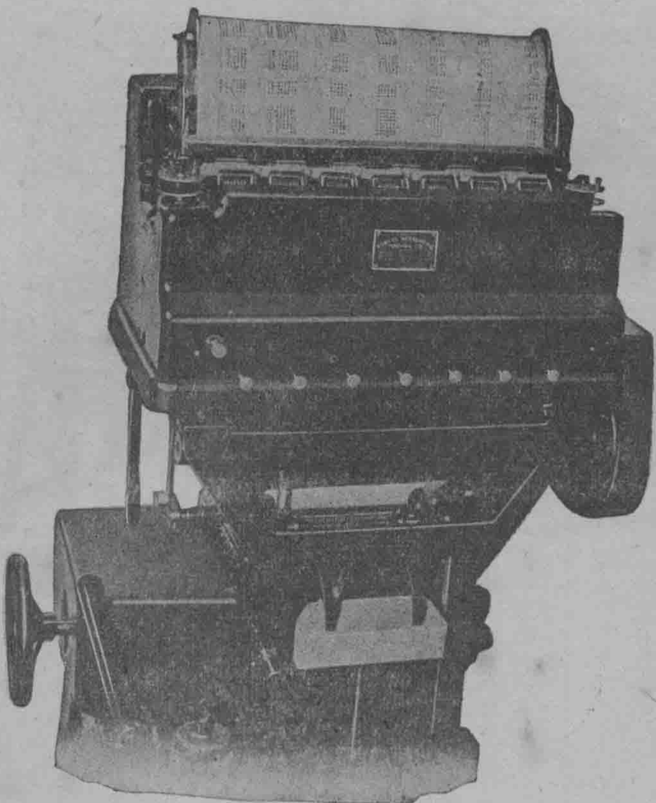


（甲）



（乙）

第 4 圖



第 5 圖

# 第四章　比例

## 第一節　比例之意義

由大量觀察而來之數字是爲絕對數，統計之目的雖在求此實數，但只此實數有時難以明瞭二者間之關係，例如雖知各國某年出生與死亡之總數，但只此總數實難知各國國民間之生死現象，及與他年相較後之生死變動狀況，故非求出其出生率與死亡率不足以互相比較。如第 31 表之甲乙二市其死亡實數雖已明瞭，但只此實數實不能比較二市之死亡情形。蓋因甲乙二市人口並非相等，其人口多者之死亡人數當然應比人口少者爲多，故必須與人口相對照後始能互相比較。在本表内如只比較其死亡實數，則甲市死亡人數比乙市爲大，但甲市人口亦比乙市爲多，故與人口相對照後求其死亡率每人口千人之死亡人數則乙市反比甲市爲大，二者情形完全相反。故知欲明瞭各種大量間之因果關係時，非求出其比例(ratio, Verhaltniszahl)不可。對此比例約可分爲二種，以後依次說明。

第　31　表

| | A 年死亡人數 | A 年人口 | 死亡率 |
|---|---|---|---|
| 甲市 | 20000 | 2000000 | 10 |
| 乙市 | 15000 | 500000 | 30 |

## 第二節　相對比例(Bezeihungszahlen)

本比例是用以比較兩種不同之大量現象或同一大量現象之在不同時期所測得之數值。例如比較人口與死亡數以求死亡率，比較前後二回人口以求人口增加率等是。其比例方法可分次之二種：

(1)將已觀察之社會大量(A)總體爲一，以求其所較他量(B)之數目。卽爲分子作一以求其分母之方法。其公式如次：

$$\frac{A}{B}=\frac{1}{x} \text{ 或 } x=\frac{B}{A} \qquad \text{(公式 2)}$$

上式內之 A,B 爲已知數，x 爲所求之數值。

(例)民國十七年年末江蘇省全省人口爲 34,129,683 人，土地面積爲 370,467 平方里，求每一平方里之人口密度？

在本例　　　A=370,467

　　　　　　B=34,129,683,

依公式得

$$\frac{\text{江蘇省之面積}}{\text{江蘇省之人口}}=\frac{370467}{34129683}=\frac{1}{92}$$

知每一平方里之人口密度爲 92 人。

(2)將欲比較之社會大量(B)爲百或千以求他量(A)之所當值。卽爲分母作百或千以求分子之方法。其公式爲

$$\frac{A}{B}=\frac{x}{100}=\frac{x'}{1000} \qquad \text{(公式 3)}$$

上式內之 A,B 爲已知數，x,x′ 爲未知數。

（例）1925 年調查得日本全國人口爲 59,736,822 人死亡數爲 1,210,706 人，求其死亡率，卽

$$A=1,210,706,$$

$$B=59,736,822.$$

依公式得

$$\frac{\text{全日本之死亡人數}}{\text{全日本之人口}}=\frac{1210706}{59736822}=\frac{20.3}{1000},$$

知其死亡率爲每千人死 20.3 人，將此與累年比較時，得知該國各年間死亡狀況之善惡程度，與世界各國相比較時得研究各國間之死亡情形。

## 第三節　配分比例(Gliederungszahlen)

於大量現象上求構成此總量之各部分對於總量之比例數。

設總量爲 A，構成此總量之各部分爲 $a_1,a_2,\cdots a_n$ 時，

$$a_1+a_2+\cdots+a_n=A$$

$$\frac{a_r}{A}=\text{配分比例}\ (r=1,2,\cdots n) \qquad \text{(公式 4)}$$

（例）第 32 表爲英國與日本各年齡組別人口死亡情形，表內千分比欄內各項數目，實由本法以求得之，所不同者只於各配分比例數上乘以 1000 而已。其算法如次：

日本千分比欄內各數值由

$$\frac{\text{(0 歲}-\text{5 歲) 年齡組別人口}}{\text{人口總數}}\times 1000$$

$$=\frac{517822}{1422096}\times 1000=364.1$$

$$\frac{(5\text{歲}-10\text{歲})\text{年齡組別人口}}{\text{人口總數}}\times 1000$$

$$=\frac{45914}{1422096}\times 1000=32.2,$$

..................................

等以求之。

第 32 表

英國及日本各年齡組別死亡情形(1920 年)

| | 日本 | | 英國 | |
|---|---|---|---|---|
| | 實數 | 千分比 | 實數 | 千分比 |
| 總數 | | | | |
| 歲 | 1,422,096 | 1000.0 | 466,130 | 1000.0 |
| 0－5 | 517,822 | 364.1 | 105,738 | 226.8 |
| 5－10 | 45,914 | 32.2 | 11,799 | 25.3 |
| 10－15 | 29,835 | 21.0 | 7,255 | 15.5 |
| 15－20 | 66,424 | 46.7 | 9,879 | 21.2 |
| 20－30 | 130,668 | 91.9 | 23,229 | 49.8 |
| 30－40 | 101,341 | 71.3 | 28,506 | 61.2 |
| 40－50 | 88,709 | 62.3 | 38,121 | 81.8 |
| 50－60 | 97,939 | 69.0 | 52,104 | 111.8 |
| 60－70 | 144,971 | 102.0 | 71,126 | 152.6 |
| 70－80 | 143,963 | 101.2 | 76,520 | 164.2 |
| 80－90 | 50,117 | 35.2 | | |
| 90以上 | 4,342 | 3.1 | 41,853 | 89.8 |
| 不詳 | 51 | 0.0 | | |

## 第四節 比例之使用與計算上各注意點

當算出比例以行推論時,須注意次之諸點:

(甲)須顧慮實數

比例雖爲重要品，但對於實數亦須顧慮。蓋因比例之價值依其算出所根據之實數而大有差異。雖同爲八成，其一萬中之八千與五十中之四十，其價值大不相同。又事象之實際狀況多有須於比例外，再知其實數後始能明瞭者，對此在比較上爲尤甚。

又在依時期而更動之變量上，年年取其配分比例時，其中某部分比例之增減與總體之增減常有不同一步調之事，有時其部分比例雖呈減少，但其實數反行增加者。例如比較各年間都市與農村之人口時，常呈此現象。其農村人口雖不減少，但因都市人口增加太速，以致配分比例上呈農村人口逐次減少之現象。參照第33表即能明白了解。

第 33 表

| | | 人口實數 | | 配分比例 | |
|---|---|---|---|---|---|
| | | A 年 | B 年 | A 年 | B 年 |
| 甲縣 | 總數 | 70000 | 100000 | 100 | 100 |
| | 城市 | 20000 | 40000 | 29 | 40 |
| | 鄉村 | 50000 | 60000 | 71 | 60 |

由上表知A年之配分比例城市爲29鄉村爲71，B年之配分比例城市由29增加至40鄉村由71減少至60，對此並非鄉村人口減少，依表知尚增加萬人，但因城市人口增加太速，以致成此現象，故知對於實數須深爲留意。

(乙)須計算同質(homogeneous)之數量

比例雖得作爲現象上一般的原因之測量器，但欲得一定原因，必須有同質之數量，故須在同質數量上所求出之比例，始能給與正確之結

論。反之在不同質數量上所求得之比例，實受有各種混合原因之影響，不能看做爲一定原因之表徵。故在理論上欲求得有價值之比例，必須於給與之大量中，完全除去與該現象無關之部分，並將此殘餘數量內擇其同質者分括爲一起，而後始能舉行計算。例如由有配偶者與總人口之比例以觀察該國家之婚姻狀況一法，實不甚優良。蓋因人口總數對於婚姻關係並非爲同質之數量，其中含有與婚姻完全無關之少年人口，以致所得比例大受少年人口之影響，故必須除去少年人口只就其宜婚姻年齡者之總數以計算其比例，較爲易於明瞭實際之狀況。且如此所求得之比例，對於比較目的上尤有巨大之價值，蓋因比較二地方之配偶狀況時，常有在宜婚姻年齡人口得較大之比例者，而在總人口上其結果適相反等事。如第 34 表。

第 34 表

| 縣別 | 總人口 | 宜婚姻年齡組別人口 | 有配偶人數 | 有配偶人數與總人口之比 | 有配偶人數與宜婚姻年齡人口之比 |
|---|---|---|---|---|---|
| 甲縣 | 1136182 | 734400 | 456602 | 40% | 62% |
| 乙縣 | 1776474 | 1086383 | 702957 | 39% | 65% |

觀上表知有配偶人數與宜婚姻年齡人口相比時則乙縣所得比例數大於甲縣，若與總人口相比較時，成爲甲縣之比例數大於乙縣，其結果適相反。故知在計算比例時，其與問題無關之部分實有除去之必要。但只除去其無關係部份，有時仍難得同質之數量，蓋因除去無關係部份後所得殘餘部份內，尚有因作用原因之不同，以致給與異質之數量，例如人口現象依男女年齡職業貧富城鄉等而呈巨大之差異，故必須將人口

大量依上列各區別以分開後,始能計算其比例。如不區分異質部份而即計算一般比例時,其所得結果實大受大量中各部份存在成分之影響。故在同一現象上取不同空間與時間之二組大量,依上述方法計算其區分與不區分之二種比例時,常因其部份存在成分之不同,大起矛盾於二者間。現將上表之已除去與問題無關之少年人口後所殘餘之宜婚姻年齡組別人口,再依其是否將男女分別以計算得各比例值如第35表。

第 35 表

| 縣別 | 實數 | | | | | | 有配偶人數與宜婚姻年齡組別人口之比 | | |
|---|---|---|---|---|---|---|---|---|---|
| | 總數 | | 男 | | 女 | | 總數 | 男 | 女 |
| | 婚姻年齡組別人口 | 有配偶人數 | 婚姻年齡組別人口 | 有配偶人數 | 婚姻年齡組別人口 | 有配偶人數 | | | |
| 甲縣 | 734400 | 456602 | 382373 | 232264 | 352027 | 224338 | 62% | 61% | 64% |
| 乙縣 | 1086383 | 702957 | 521173 | 346579 | 565210 | 356378 | 65% | 67% | 63% |

觀上表知依男女總數以計算比例時,則甲縣所得結果比乙縣爲低,若男女分別計算時其男方比例乙縣果然較高,但在女方比例則乙縣反比甲縣爲低。故在計算比例時除移去與問題無關之部分外,尙須將殘餘部分內異質部分分別計算而後始可。

(丙)注意比例之性質

依研究問題之不同,其所用比例方法亦隨之而異,例如配分比例是爲研究大量之內部構成上所不可缺物,當其計算時只知大量之構成狀況已足。反之相對比例除須知二大量之總數外,尙須施行同一分類於二大量上,而後始能求各同質部分之比例。

# 第五章　計算機及計算表

統計上所使用之數值，普通數目甚大，依此計算亦甚複雜，只用暗算及筆算實不能迅速給與正確之結果，於是學者發明計算機與計算表以補其缺。

## 第一節　計算機

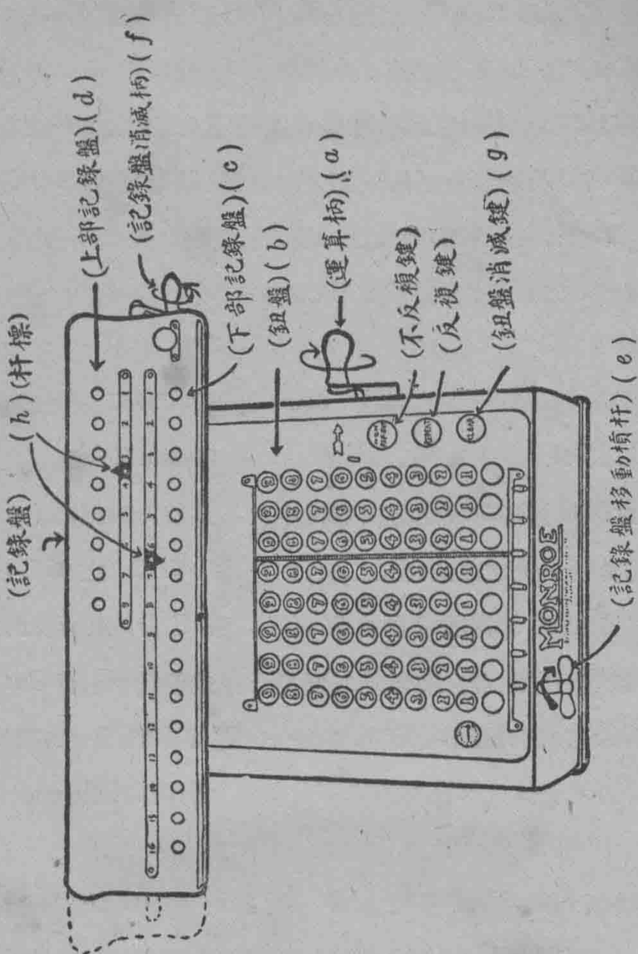
由學者預先發明種種形式之機器，計算者只運轉此等機器後卽得所求結果。現將普通最多使用之計算機列舉二種如次：

(甲)算盤

統計界是將算盤使用於加減上，至計算乘除及開方時則多使用次之蒙羅計算機。

(乙)蒙羅計算機(Monroe's calculating machine)本機得使用於加減乘除開方及開立方等計算上，其構造外形如次圖。圖內 a 爲運算柄(Operating crank)，依矢之方向一次迴旋時，其 b 鈕盤(Key board)上各鈕所捺成之數目卽一次相加，並表現其結果於下部記錄盤(Lower dial) c 內，其運算柄之迴轉次數表現於上部記錄盤(Upper dial) d 內。將運算柄依矢之反對方向迴轉一次，是將下部記錄盤 c 上所有數目減去鈕盤 b 上各扭所捺成數目一次。e 爲記錄盤之移動槓桿(Shifting lever)，將此係矢之方向迴轉半次，是將記錄盤向右移動一行，若舉行反

迴轉則將記錄盤向左移動。f 爲記錄盤消滅柄，將此依矢之方向迴轉一次是將 d 內數碼完全變爲「0」以行消滅，反對迴轉一次是將 c 內數碼消




第6圖　蒙羅計算機

滅。g 爲鈕盤消滅鍵，一捺此鍵則 b 上所有數碼皆行消去。現將其實際運用方法依實例說明如次：

(A)乘法

求 219×23 之計算

先在 b 盤(以後用 K.B. 表示之)上將最右起三列數碼內,第一列捺入 9 字鈕,第二列捺入 1 字鈕,第三列捺入 2 字鈕後,在 K.B. 上得 219 之數目,次將下部記錄盤 c (以後用 L.D. 表示之)上之第一號孔穴與第一列鈕相對齊而後將運算柄依矢之方向迴轉三次,則將 219 數目三次相加於 L.D. 上,現出 657 之數目,同時上部記錄盤 d (以後用 U.D. 表示之)上現出 3 之數字,其排列情形如次:

| | |
|---|---|
| U.D. | 00000003 |
| L.D. | 0000000000000657 |
| K.B. | 00000219 |

次依 e 之移動槓桿向矢之方向迴轉半次,是將記錄盤向右移轉一列後,再將運算柄依矢之方向迴轉二次,如此是於 L.D. 之 657 上再加以 2190 之二倍 4380 卽得所求之結果 5037,在 U.D. 上所現出之 23 數碼是證明運算柄之迴轉次數確爲二回與三回並無錯誤,若 U.D. 上現出 33 之數碼時,卽知在移動記錄盤後一次運算多迴轉一次成爲 219×33,當此時無須全部重算只將運算柄向矢之反對方向迴轉一次卽得。

(B)除法

計算除法時先將被除數捺在 K.B 上,並由 K.B. 上所捺成數碼之最左一列數鈕與 L.D. 之最左孔穴相對準後將運算柄依矢之方向迴轉一次將被除數運入於 L.D. 上,其 K.B. 上數目得由一捺鈕盤消滅鍵以消滅之,在 U.D 上所現出之運算柄迴轉次 1 字得依記錄消滅柄

依矢之方向迴轉一次以行消滅後,始能將除數捺入於K.B.上,但此時除數之最左端數字須捺入K.B.之最左端一列鈕上其他依次排列。至其運算方法,是將運算柄依矢之反對方向逐次迴轉,至鈴聲一鳴時卽須停止,並須反迴一次。聞到第二次鈴聲後,將移動槓桿向矢之反對方向迴轉半次,是將記錄盤向左移動一行後,再行開始運轉繼續進行。

例如計算 6824÷56.

先將L.D.之最左端與K.B.上所捺成數目6824之6字相對準後始能將運算柄向矢之方向迴轉一次,如此L.D.上現出6824之數目次捺鈕盤滅消鍵以消去K.B.上之數目,並依矢之方向迴轉記錄消滅柄一次以消滅U.D.上之1字後,始於K.B.上由左端起捺入56之數目。其面上情形如次:

| | |
|---|---|
| U.D. | 00000000 |
| L.D. | 6824000000000000 |
| K.B. | 56000000 |

成此情形後,開始將運算柄依矢之反對方向繼續迴轉,至背面鈴聲一鳴後當卽停止,並再依矢之方向迴轉一次,成爲次之情形:

| | |
|---|---|
| U.D. | 10000000 |
| L.D. | 1224000000000000 |
| K.B. | 56000000 |

由移動槓桿將記錄盤向左移動一行,使成爲

| | |
|---|---|
| U.D. | 10000000 |
| L.D. | 1224000000000000 |

K.B.　　56000000

將運算柄繼續依矢之反對方向迴轉至鈴聲一鳴爲止，並再向矢之方向迴轉一次後，再將槓桿向左移動一行，使其面上成爲

U.D.　　12000000

L.D.　　0104000000000000

K.B.　　56000000

再繼續同樣進行得

U.D.　　12100000

L.D.　　0048000000000000

K.B　　56000000

知其剩餘爲 48 已不能再減，其計算結果爲商 121 與剩餘 48。如倘欲計算其小數時，得將 h 之杆標置於 121 之右端，而後再繼續進行運算，其在杆標右方所得數值皆爲小數。

(c)開方

開方是應用

$$\sum_{K=1}^{n}(2K-1)=n^2,$$

及

$$(a+b)^2=a^2+(2a+b)b$$

二原理以計算一切。其計算方法仍依實例以行說明。

例如計算$\sqrt{730837156}$

先將欲舉行開方之數值移入於 L.D. 之左方內，並將 K.B. 之最

左端與數值之最左一行相對準,其面上情形如次:

| | |
|---|---|
| U.D. | 00000000 |
| L.D. | 7′30′83′71′560000000 |
| K.B. | 00000000 |

K.B. 上與 7 字相對之鈕列內捺入 1 字(若 L.D. 之最左一段分位有二位數字時,須由 K.B. 上之最左數起第二列鈕上開始進行),以舉行減法後,再捺入同列之 3 字以行相減,使成爲

| | |
|---|---|
| U.D. | 20000000 |
| L.D. | 330837156000000 |
| K.B. | 30000000 |

L.D. 上之 3 已不能再減去 3+2 之 5,故須將記錄盤向左移動一列並加 1 於 3 上使成爲 4,機面上成爲

| | |
|---|---|
| U.D. | 20000000 |
| L.D. | 330837156000000 |
| K.B. | 40000000 |

將 1,3,5,7,9,11,13 等數碼逐次捺入於 4 之右方鄰列內以舉行減法至不能再減爲止,其機面情形如次:

| | |
|---|---|
| U.D. | 27000000 |
| L.D. | 001837156000000 |
| K.B. | 53000000 |

將 K.B. 上之 53 加 1 使成爲 54 後再將記錄盤向左移動一列使成爲

| | |
|---|---|
| U.D. | 27000000 |
| L.D. | 001837156000000 |
| K.B. | 54000000 |

L.D. 之 18 少於 54，故尙須將記錄盤向左移動一列使成爲

| | |
|---|---|
| U.D. | 27000000 |
| L.D. | 001837156000000 |
| K.B. | 54000000 |

如此則 18371 大於 5400，得以施行減法，現逐次將 1,3,5 等數捺入 K.B. 之由最左數起第 4 列鈕上以舉行減法後，得

| | |
|---|---|
| U.D. | 27030000 |
| L.D. | 000216256000000 |
| K.B. | 54050000 |

加 1 於 5405 上使成爲 5406，並移動記錄盤向左一列得

| | |
|---|---|
| U.D. | 27030000 |
| L.D. | 000216256000000 |
| K.B. | 54060000 |

再逐次將 1,3,5,7 等數目捺入由左方數起第 5 列內以舉行減法得

| | |
|---|---|
| U.D. | 27034000 |
| L.D. | 000000000000000 |
| K.B. | 54067000 |

由 U.D. 知

$\sqrt{730837156}=27034$。

(D)開立方

本法之機面運轉既甚煩雜又爲統計界所少用,故不再述。

上述蒙羅計算機完全用人手以舉行運算，此外尙有用電氣以自動算計者,其原理完全相同。

此外尙有多種計算機,其原理及運算方法與蒙羅計算機大同小異,故不再述。

## 第二節 計算表

上述各種計算機,除算盤外,皆爲高價物品,故不適用於個人研究。因此學者特預先作成種種數值計算表以資應用。

數值計算表內爲吾人所常用者除對數表外,尙有次之數種:

(A) 巴羅計算表(P. Barlow: Tables of Squares, Cubes, Square-roots, Cube-roots and Reciprocals)。

揭載一萬以下各數值之平方立方平方根立方根及逆數之結果。

(B) 克累雷計算表(A. L .Crelle: Calculating Tables)。

揭載三行數字以下各數值間相乘結果。

(C) 披爾遜表(K. Pearson: Tables for Statisticians and Biometricians)。

現已出二册，第一册揭載各統計値之結果表凡 55 種，第二册揭載 51 種。

# 第六章　統計圖

## 第一節　統計圖之意義

將觀察之結果，除作成統計表外，尚須描畫統計圖 (Statistical figures charts) 使觀者一見卽能明瞭其現象之狀態，及相互間之因果關係。

統計圖是由點之多少，線之長短，面積與體積之大小，色彩之暗淡及曲線之傾斜角度等，以圖示由統計方法所得之數值。其功效有次之數種:

(1)世人對於數字多不慣熟，甚至看到數字卽感覺頭痛者亦有之。可是統計表完全是數字之排列，依此除少數有特殊研究者外，類多索然無味，功效殊尠。玆爲謀統計之民衆化與發揮統計之功效計，特將統計表內所具備之材料，圖示於紙面上，先依色彩之音階，繪線之縱横，引起閱者之興趣而後繁冗之數字，得能印入於腦海中。

(2)能將緊要之事實，圖示於紙面上，使世人一見卽能明白其大小高低之差，依此其大概情形亦能一見明瞭。

(3)得能捕捉事實之眞相，考察推移之狀況，在並記數種事實時，得能探求相互間之因果關係。

(4)依統表地圖得能察知事象之依地理的變遷狀況。

近來吾國統計事業日趨發達，因此統計圖之製作亦日益增加，惜所作成中，尚多漠然難解之圖形，不能給閱者以明晰之印象，其故實由於製圖者尚未十分了解統計圖之製作原理所致。夫天地間之現象，其有形事物較無形者易於了解，有形事物中之能整然排列者，當然較漫然散在者又易於印入人類之腦海。彼有名之博物學者阿軋細斯曾云：「顯微鏡與人以第二眼，圖畫與人以第三眼」。又西諺有云：「圖畫爲空字之文章，數字爲無聲之議論」等等。故知統計圖實得說爲與統計家以第二眼，與世人以第一眼。在歐洲大戰中，雙方均利用巧妙之圖畫互相宣傳，皆得良好之效果。又美國勞働者中，多不識字分子，故對於工場內危險處之警告，多用明瞭之繪畫統計圖，其結果亦甚優良。依此知今後統計圖之使用，日益增多，故對於製作上不能不加以詳細研究，使成爲完全明瞭之統計圖，而後世人得依統計表爲右眼，統計圖爲左眼，以分析一切問題。

統計圖之種類大別之爲一般統計圖與坐標統計圖二種，後者是用幾何學上之坐標以表現次數分配狀況，前者不使用坐標完全用圖形以表現各種統計之狀態。以下依次詳述。

## 第二節　一般統計圖

其作成方法是將統計表上所屬之事實，依排列順序用各種圖形以表示之。本法對於統計之說明與理解上，既有多大之裨益，而對於統計之民衆化上，尤有巨大之貢獻，其圖形對於表示之事實上，雖不能附加以新的意義，但能將原有之統計意義，明白表現於紙面上，使閱者一見

卽能理解一切。至其圖形依目的之不同，約可分爲次之二種：

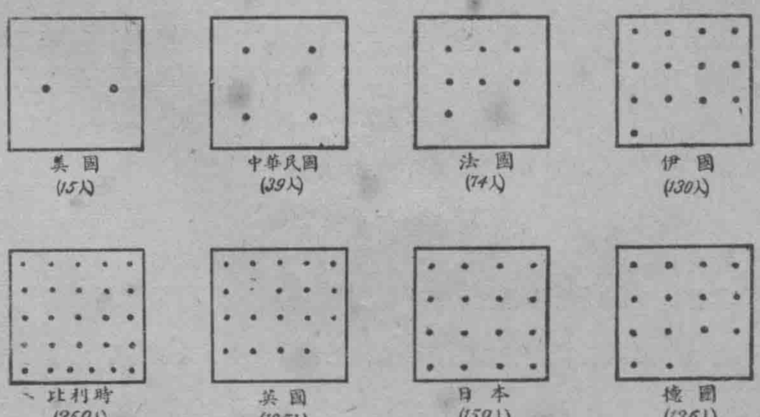
(1)只表示統計數值之統計圖，稱爲統計狀態圖(pictograms)。

(2)表示統計數值與地域分布關係之統計圖，稱爲統計地圖(Cartograms, Statistical maps)。

上述統計數值非只實數而已，各種比例與平均數亦含在其內。至統計狀態圖一般又可分爲點圖，長條圖，面積圖，體積圖及繪畫圖等五種，以下與統計地圖一併依次說明。

(A) 點圖

將欲比較各數值之幾單位作爲一點，係點之多少以比較各種所持數值。當比較各國人口密度時，以本法爲最妙。例如第 7 圖爲中英美意法德日比等八國在 1926 年之每一平方公里人口密度狀況，其作成方法是在同一面積之正方形內，取每十人爲一圓點，以記入該國人口密度所當點數。考察各正方形內點之疏密狀況，卽能比較各國人口密度，依圖




第 7 圖

知比利時之人口密度爲最大。

點圖與地圖相連結時，頗能顯其效果。例如工商業者欲觀察自己商品之販賣分布狀況，特將其所交易商店每數舖爲一圓點畫入地圖上後，卽能明瞭自己商權所及範圍，並得由此以定今後尙欲擴張營業之區域。將點圖用之於地圖一事，是屬於統計地圖，待以後在統計地圖內，再詳行說明。

(B) 長條圖

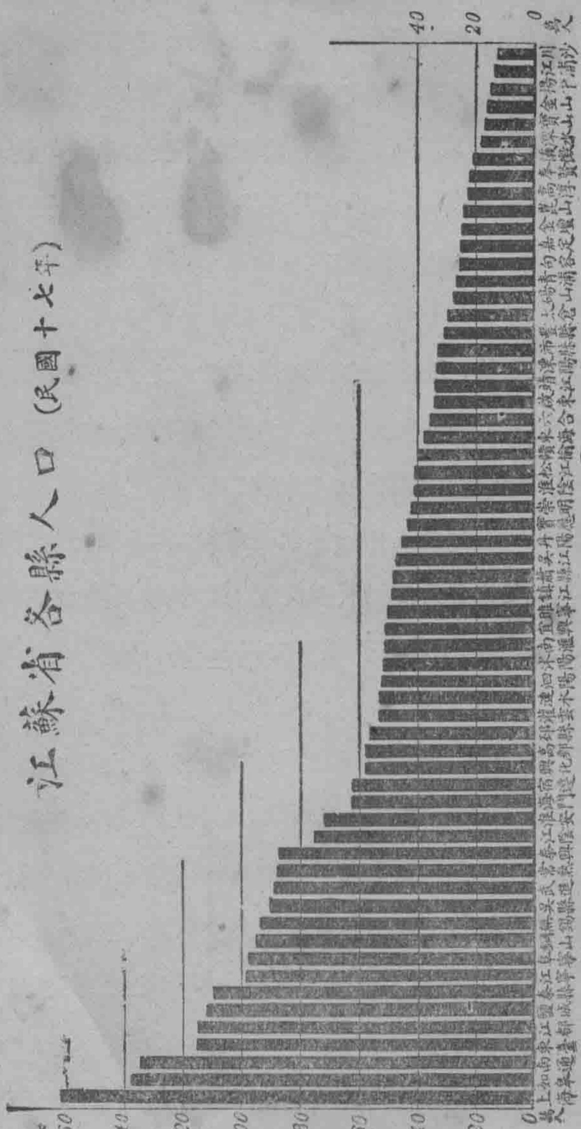
將欲比較之數值，用長條之長短度以表示之。此長條普通是直畫，故復稱之爲柱狀圖 (Histograms)。對此長條圖得分爲相異主體間之比較及同一主體之量的內部構成狀況比較二種。

第 8 圖爲相異主體間互相比較之長條圖，其各縣相互間完全無關，祇各自取其人數之相當長條於所占位置上，依其長條之長短以比較各縣人口之多少，依圖知江蘇省各縣人口以上海爲最多(連上海市在內)，川沙爲最少。

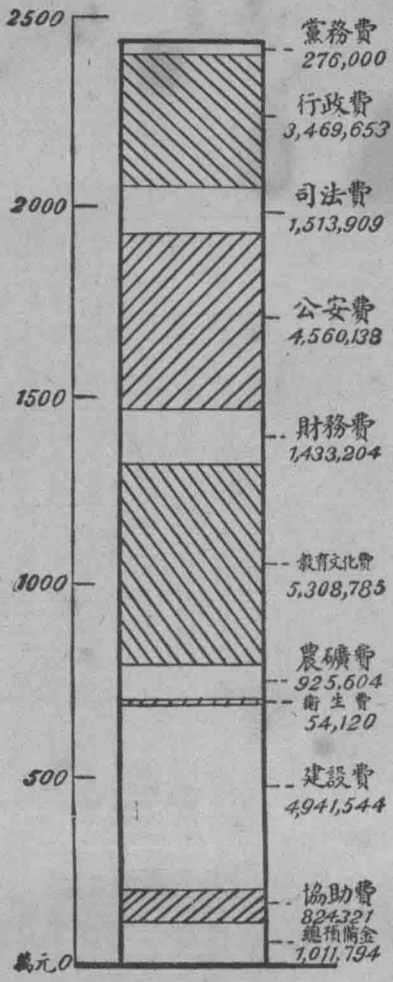
第 9 圖爲在同一主體間表示其內部構成狀況之長條圖，是在一根長條上，用顏色或其他種種之分別方法，以表示其構成部分之比例。本圖長條之全長爲江蘇省民國十九年度地方歲出經臨預算全數，其各部支出費用依其數目之多少取相當長度於此全數之長條上，依其所占地位之多少，卽知江蘇省各部支出費用所占全預算之比例情形，依圖知教育文化費爲最多，衞生費爲最少。

第 10 圖是將上述二種合併表示之長條圖，其國立私立……等爲專科以上各學校設立類別，並在各設立類別上再舉行內部分類。依圖知

江蘇省各縣人口 (民國十七年)




第 8 圖 相異主體間比較之長條圖

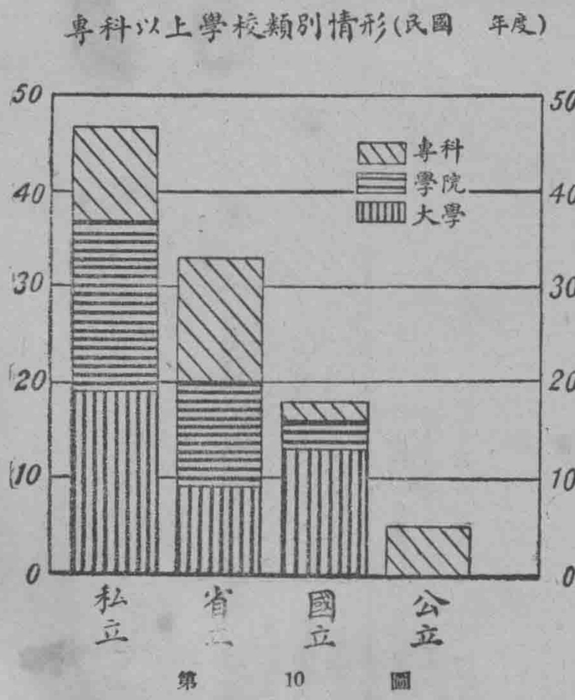



第 9 圖

表示內部構成比例之長條圖

(江蘇省地方十九年度歲出經臨預算細別)

民國二十年度全國專科以上各學校中以私立者爲最多，公立者最少。至其內容則國立學校幾全爲大學，而省立學校則以專科占其大多數，至公立學校則只有專科而已。




第 10 圖

相異主體及內部構成合併爲一體之長條圖

長條圖爲統計圖內之主要部分，現將其繪製時之注意點，列舉如次：

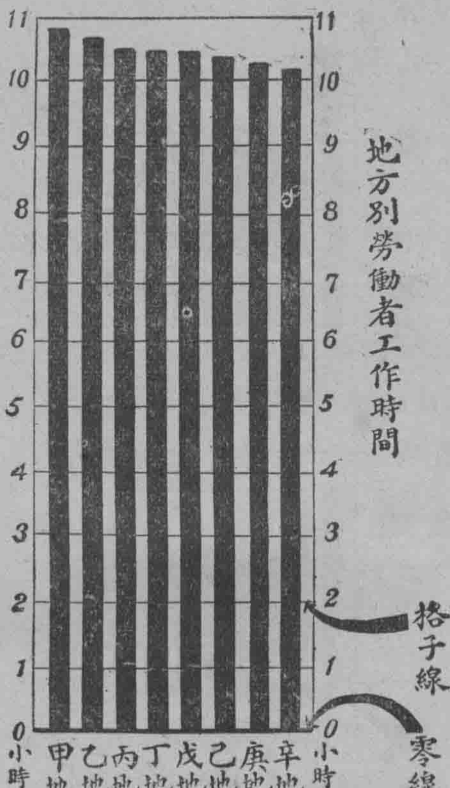
(甲)關於零線

(1)零線非畫不可，如第 11 圖矢之所指處。至第 12 圖因其不畫零線，以致圖上辛地之代表條長不及甲地之 $\frac{1}{3}$，一若辛地之工作時間不

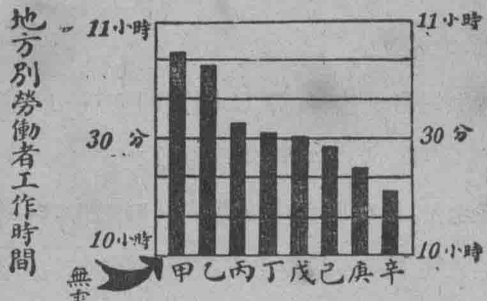
及甲地之$\frac{1}{3}$然，但其實際情形（參照第 11 圖）二者相差實甚微，故知不畫零線時實易起巨大之錯誤。

(2)零線須粗於別的格子線（如第 11 圖）若遇百分比較時，其100%線亦須畫粗線（如第 15 圖），因其亦可爲零線故也。

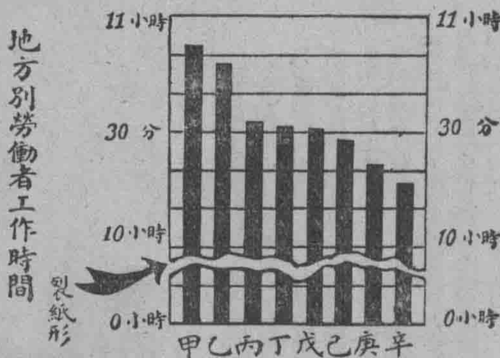
(3)爲紙面經濟起見，得在圖之中間用裂紙形以省略一部分，（如第 13 圖），或不劃零線只將下端零線處之直線變爲裂紙形，將長條之上部移至零線附近（如第 14 圖）。



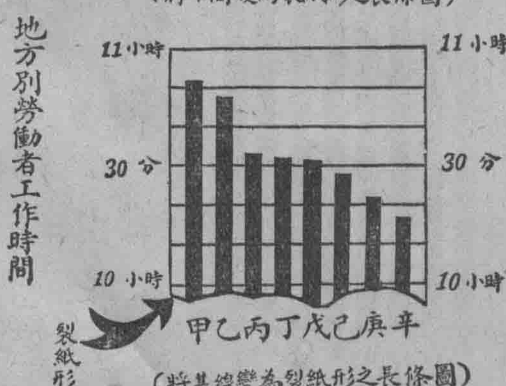

第 11 圖　（正確的長條圖）



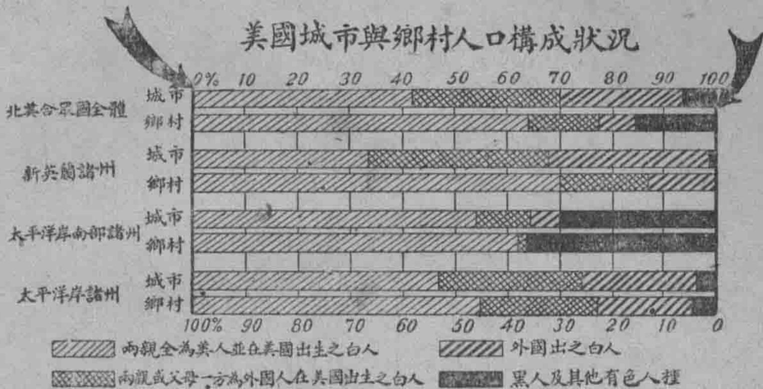

第12圖 (不正確之長條圖)




第13圖 (將中間變為裂紙形之長條圖)




第14圖 (將基線變為裂紙形之長條圖)




第 15 圖

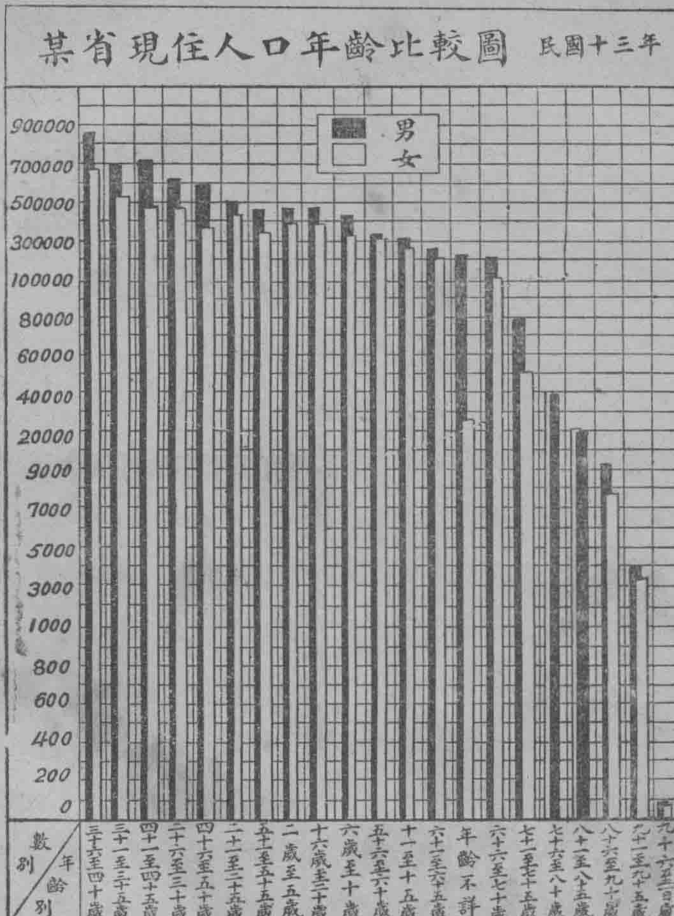
(乙)關於格子上

(1)各格間之距離均須相等，切不可有上下不相同之事。如對數尺寸，在長條圖上實不能使用。

(2)各等幅格子間距離之所代表數值亦須相等，切不可爲省略起見，用不同數值以代表各距離。如第16圖其千人以下每一格之代表人數爲百人，千人至萬人處每一格代表千人，一萬至十萬處每一格代表萬人，十萬至九十萬處每一格代表十萬人其同幅各格代表人數，依地方而不同，以至91歲至95歲之老人，只就長條之長短以觀察時，幾占36歲至40歲所有人數之$\frac{1}{3}$，但考察其實數時則91歲至95歲之老人只有36歲至40歲者之$\frac{1}{200}$，錯誤處實甚明瞭。

(3)格子線不可太細。若因細別階級以致格子線太多時，得將附有0之格子線稍爲放粗，但仍須比零線爲　。

(4)於格子線上附以數字時，須兩端皆附加之若遇百分比較時，其


某省現住人口年齡比較圖
民國十三年
男
女
900000
700000
500000
300000
100000
80000
60000
40000
20000
9000
7000
5000
3000
1000
800
600
400
200
0
數別
年齡別
三十六至四十歲
三十一至三十五歲
四十一至四十五歲
二十六至三十歲
四十六至五十歲
二十一至二十五歲
五十一至五十五歲
二歲至五歲
十六歲至二十歲
六歲至十歲
五十六至六十歲
十一至十五歲
六十一至六十五歲
年齡不詳
六十六至七十歲
七十一至七十五歲
七十六至八十歲
八十一至八十五歲
八十六至九十歲
九十一至九十五歲
九十六至一百歲

第16圖 錯誤之長條圖

兩端百分比以相反記入爲妙(如第 15 圖)。

(5)格子上數字不必全部附加,只將其有 0 之整數(round number)記入可也。例如只記入 10 萬 20 萬等數字,而不記入 11 萬,12 萬,……。

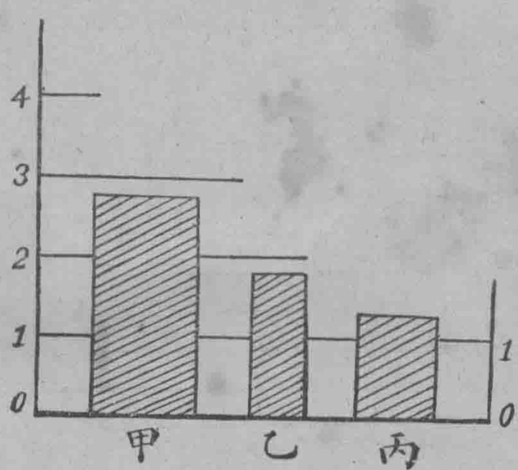
(6)於零線處必須附以零之文字。

(7)數字用中國數字或阿剌伯數字均可,至其單位之附加則多只於零線處附以所當單位,其他各數字處皆不附加之(如第 11 圖之只附加小時二字於零線處)。

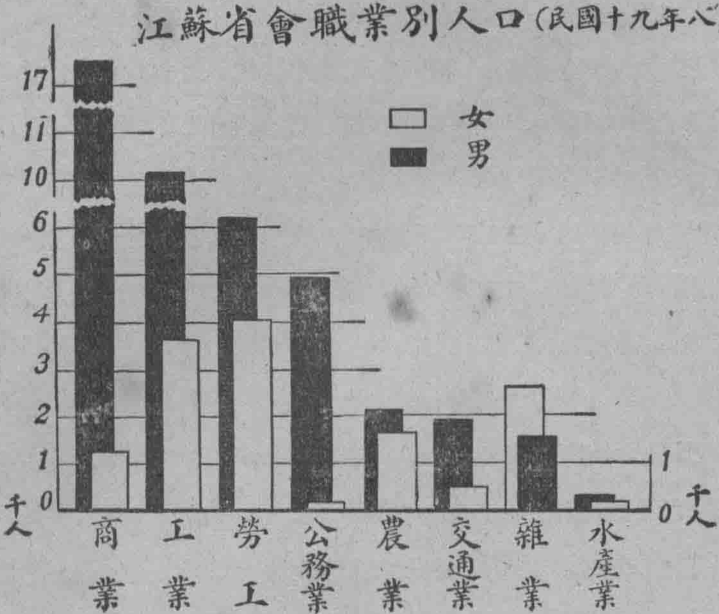
(丙)關於條狀上

(1)長條之幅度須彼此相等,切不可有如第 17 圖之甲乙丙三長條之幅度彼此不相等情形。

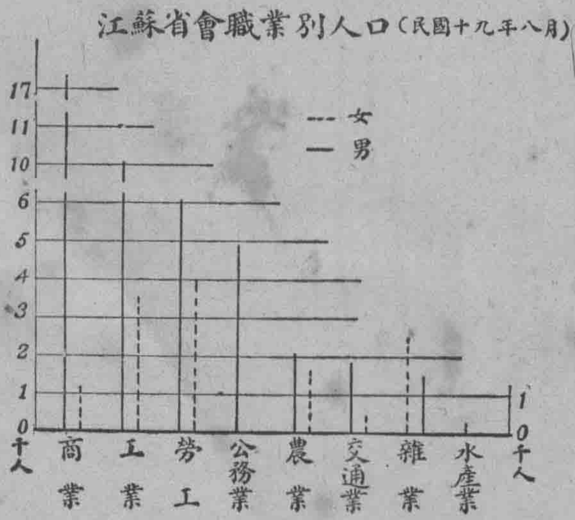
(2)長條之幅度太細時,圖上未免太空,(如第 19 圖故必須給與相當之粗度,(如第 18 圖)既較爲雅觀又便於比較。




第 17 圖




第 18 圖 適宜之長條圖

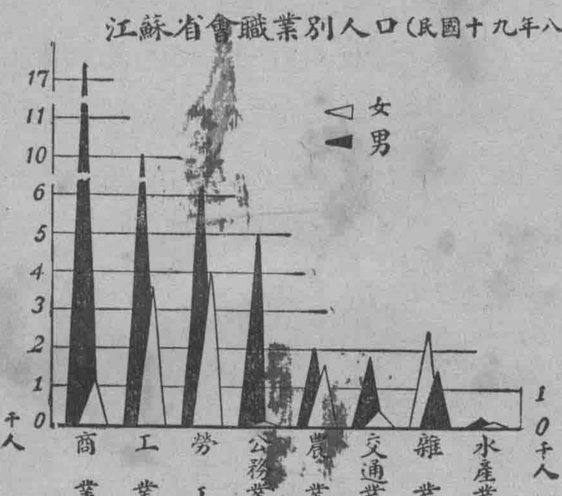



第 19 圖 太細之長條圖

(3)長條之上下須同一大小，如筍狀喇叭等形，務求排斥，因其當比較時，有迷目之虞也。如第 20 圖爲筍狀之長條圖，其上端旣難以畫成，且相互間又難以得正確之比較情形，如第 21 圖爲喇叭形之長條圖，因其帶有面積圖之色彩，使觀者起大者愈大小者愈小之不正確現象。

(4)長條務須直畫。如第 22 圖內甲乙丙三統計數值之比雖約爲 3:2:1，但在實際上甲條之長度幾等於丙條之六倍，乙條上亦有此情形，使觀者起大者愈大小者愈小之觀念。

(5)圖內有一部分長條太長，不能在圖上全部容納時，得增加其關於同主體之條數。(如第 23 圖爲中美日三國鐵路長度比較狀況，因美國路程較之中日二國多至數十倍，故特定二萬哩爲一條以增加美國所當長條數，但當此時其二條間連接處之橫條長不算入在內)。若相差不十分

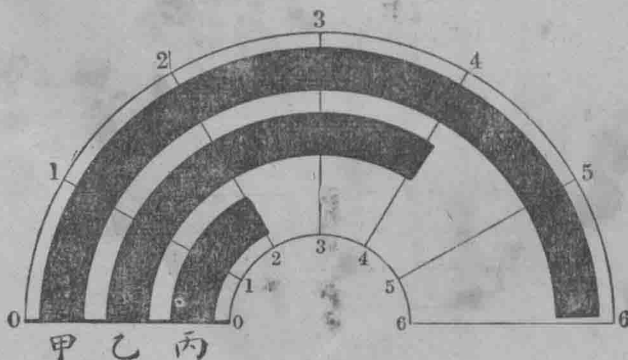



第 20 圖 形狀不好之長條圖

某廳職員體格比較圖之一

民國十八年一月二十三日檢查


第 21 圖 形狀不妥之長條圖




第 22 圖 不正確之長條圖

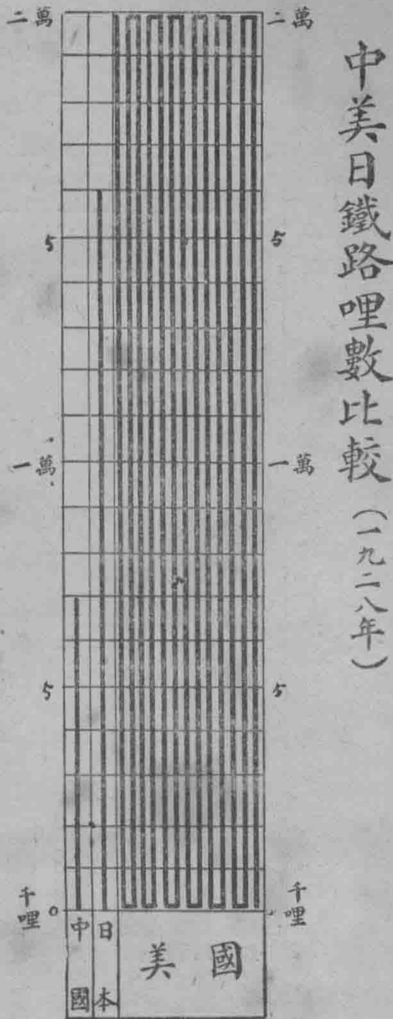
巨大時，得折曲其長條。(如第 24 圖)。或用裂紙形，切斷其長條。(如第 18 圖及第 25 圖)。

(6)表示數值之構成上有分別長條之色彩時，須將暗黑色置於下方，明淡色放在上方，使給與閱者以安定之概念。

(7)遇色別時，應將色之區別明示於圖表上適宜處(如第15圖)。

(8)長條之排列除已有特定排列外，以依長短之次序爲是(如第8圖)但遇年份比較時，以依年份次序爲是(如第 26 圖)。

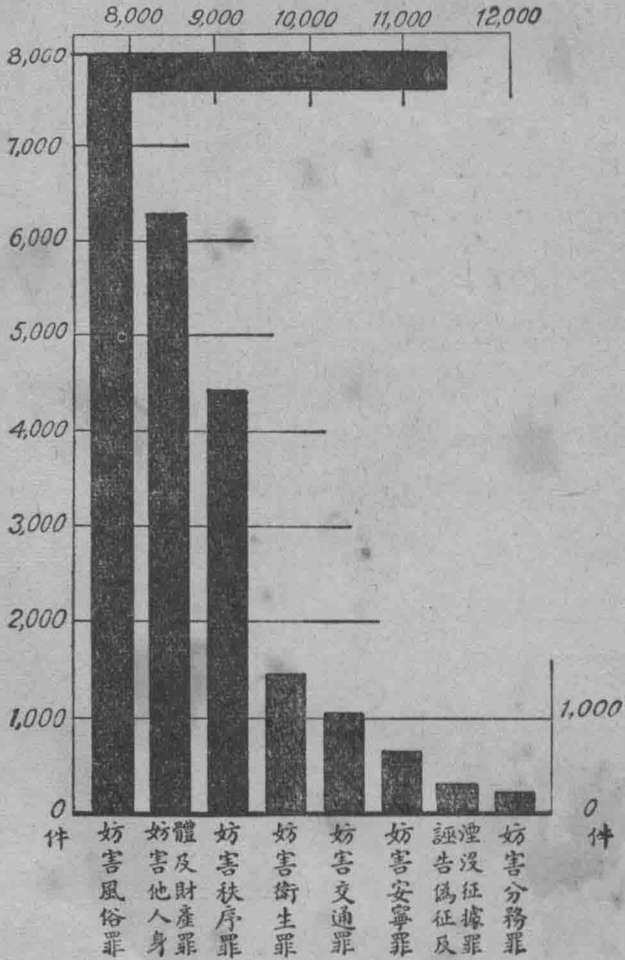
(9)將長條橫置時




第 23 圖

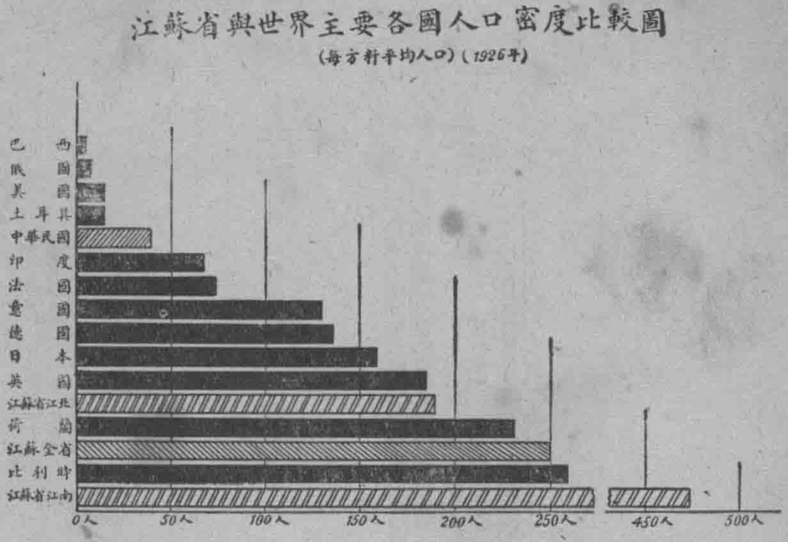
除已有特定排列者外，以長者放在下方較爲安定(如第 25 圖)。但歐美各國之統計書上，多以長者放在上方。

江蘇全省違警罪發生件數類別情形(民國十八年)




第 24 圖

(10)一欄內放二個以上之長條時，則將較短者之一部分嵌入於長的長條內(如第 18 圖)。




第 25 圖

(丁)關於文字之記入

(1)表示長條之所屬年份及主體之名稱等文字，爲統計圖上之一部份，故必須記入之。

(2)不可於長條上直接附以文字。如第 26 圖將各年商品輸入額之文字記入於長條上部，使閱者有引起將文字看作爲長條一部分之虞。

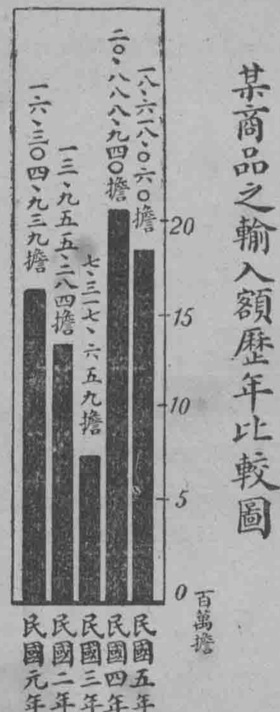
(戊)關於圖之標題

(1)須附以能明瞭內容之簡單標題。

(2)標題須記入於圖之外側，但遇內部有空白時，亦得在不妨礙長

條處割斷格子線以記入之。

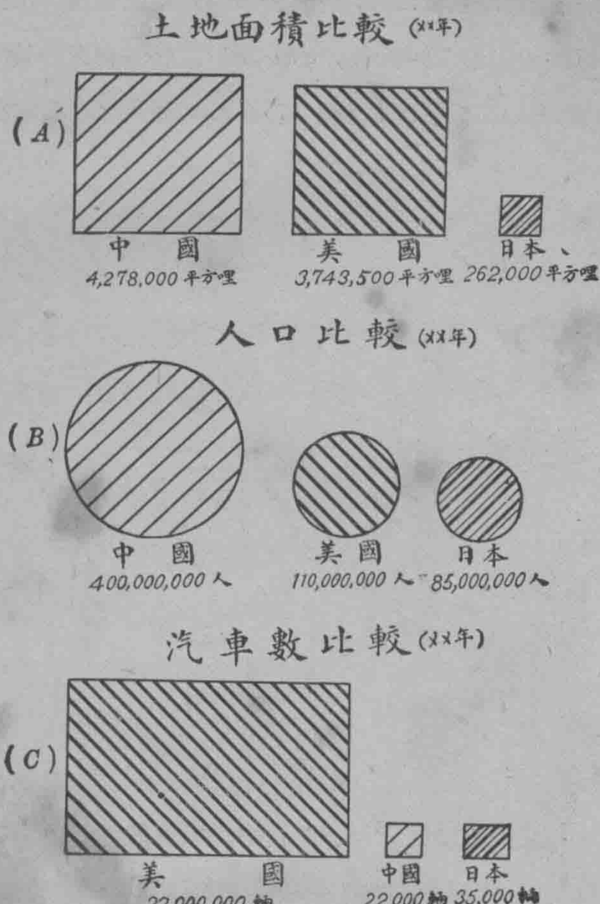
(3)標題之下方空處，得記入線紋之分配，單位之取法，材料之來源等例樣。




第　26　圖　(不好之文字記入例)

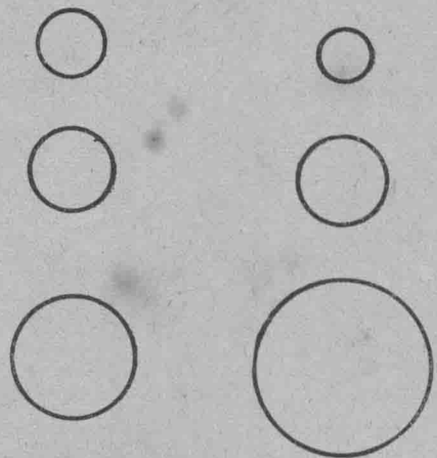
(C)面積圖

面積圖是依線紋所包圍圖形之面積以比較其數値之大小。其圖形爲一般所常用者是正方形，矩形，三角形，多角形及圓形等數種。如第27 圖內(A)是依正方形面積之大小，以比較中美日三國土地面積，(B)是依圓形面積以比較中美日三國人口數。(C)是依矩形面積以比較中美日三國汽車數。


土地面積比較 (××年)
(A)
中國
4,278,000平方哩
美國
3,743,500平方哩
日本
262,000平方哩
人口比較 (××年)
(B)
中國
400,000,000人
美國
110,000,000人
日本
85,000,000人
汽車數比較 (××年)
(C)
美國
22,000,000輛
中國
22,000輛
日本
35,000輛

第 27 圖

面積圖是爲二元的圖形，故在比較上比一元的長條圖爲難。例如甲乙丙三省，其土地面積爲 4:2:1，現將其數值圖示之如第 28 圖(A)，但世人觀此常以爲尙未到此比例數，反取 (B) 之各圓爲近於該比例數之圖形，但在實際上(B)圖各圓是就半徑爲 4:2:1 所畫成者，其面積已成



(A)　　　　　　(B)

第　28　圖

爲 16:4:1，故知使用面積圖以舉行比較時，常使觀者不能卽得其比較實情。但在此二元的圖形中，將其一元變爲等大時，卽將二元的面積圖變爲一元的圖形。次之扇形統計圖 (Pie diagrams) 卽爲此類統計圖之一種。

扇形統計圖之作成方法，是將某大量之構成部分依配分比例以分割圓周。其分割方法如次：

設A為某大量，

$$a_1, a_2 \cdots\cdots a_n \text{ 為構成部分時}$$

$$a_1 + a_2 + \cdots\cdots a_n = A$$

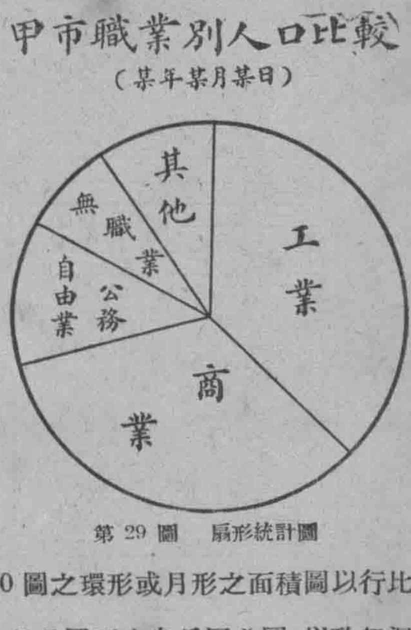
其用以分割圓周之比例為

$$\frac{a_r}{A} \times 360^\circ, \qquad (r=1,2,\cdots\cdots n)。$$

由此配分比例所求得之度數，依分角器以定圓周上各分割點後，再用直線連結此等分割點與圓心，使分全圓為若干部分後，用色彩或其他特殊符號，以表示各數值之所屬。依其所占面積之大小卽知所屬數值之多少。至各部分百分比之取擇上，仍用配分比例

$$\frac{a_r}{A} \times 100, \qquad (r=1,2,\cdots\cdots n)。$$

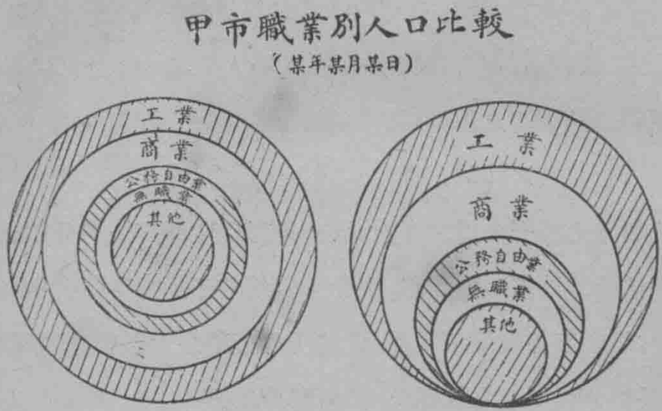
以求出各相當比例值後，將此等數值填入於所屬內卽得。第 29 圖為甲市職業別人口比較狀況，其圖內工業一項面積最大，知甲市人口之職業以屬於工業者為最多。此扇形統計圖因其半徑之長度為一定，故只觀察圓周之長短，卽能明白各數值之多少，將二元的圖形一變而為一元的圖形，在比較上甚為便利。




第 29 圖 扇形統計圖

在此等表示構成部分上，若將 29 圖之扇形統計圖一改而為如第 30 圖之環形或月形之面積圖以行比較各業人口時，反使觀者難以理解，且其圖形本身為同心圓，以致起迴

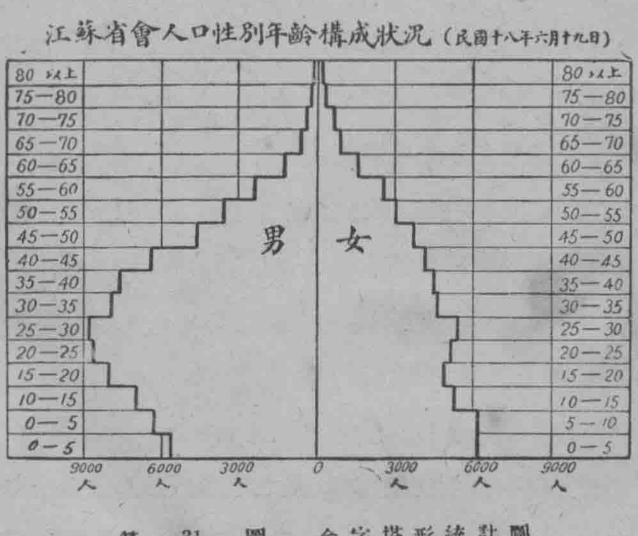
轉之勢，使觀者頭昏目眩，故此等面積圖以不畫爲宜。




第 30 圖 嵌圓形於圓內之面積圖

此外得屬於面積者，尚有

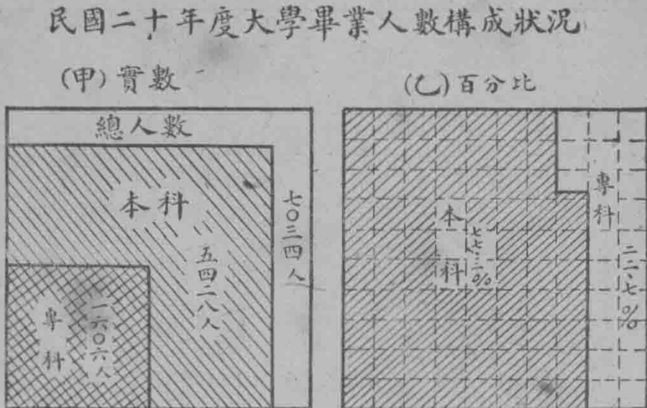
(1)金字塔(pyramid)形統計圖。如第31圖，是以民國十八年六月




第 31 圖 金字塔形統計圖

十九日所查得江蘇省會五歲年齡階級別人口爲基礎所作成之金字塔形統計圖。

(2)正方形面積統計圖。如第 32 圖是以民國二十年度大學畢業人數爲基礎所作成之實數與構成百分比圖。



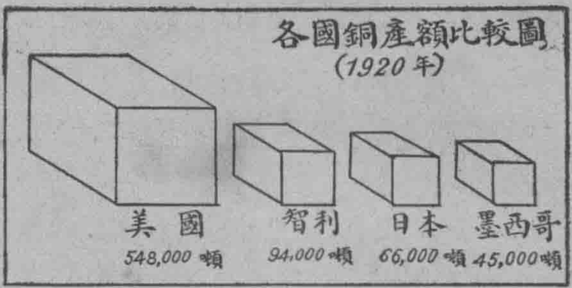

第 32 圖　正方形面積圖

(D)體積圖

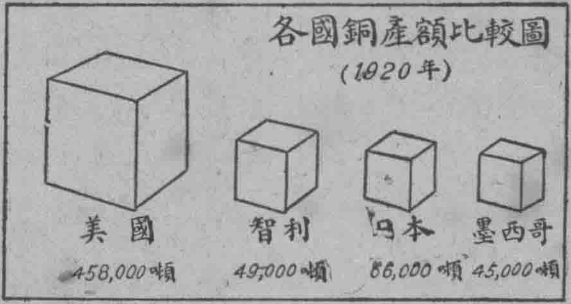
體積圖(Volume diagrams)是於紙面上畫角柱,圓柱,正立方體及其他各種立體形,使觀者想像其體積之大小以比較各數值。

當描畫立方體時用透視畫法(perspective 如第 33 圖)或斜方平行投影法(oblique parallel projection 如第 34 圖)均可。

比較面積圖時之困難情形已如上述,但在體積圖上,除長與幅之比較外,尚須比較其高度,成爲三元的圖形,在比較上非常困難,故多不使用之。




第 33 圖 體 積 圖

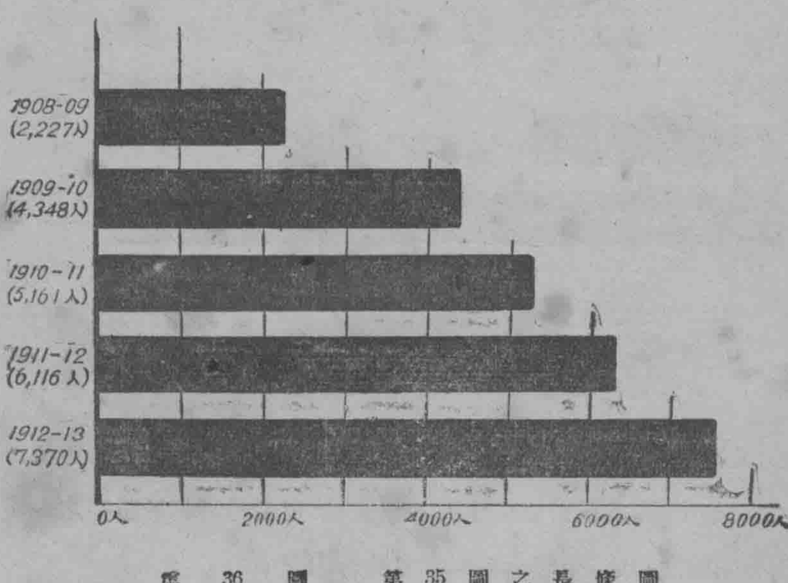



第 34 圖 體 積 圖

(E)繪畫圖

繪畫圖 (pictures) 是依普通繪畫方法繪出其相當各物之實況，而後依其實狀之大小以表示各數值之多少。對此實狀大小之取擇上，有時使用其長度，有時使用其面積。但使用長度時其大者愈大小者愈小，於比較上未免過大。如第 35 圖爲某處求職人數比較圖其作成方法是各取其求職人數作爲人之身長以作成人之形像，對此其 1912－1913 年之求職人數，一若有 1908－1909 年之七八倍，但在實際上考察第 36 圖時，其相差實不到四倍，故知使用長度時實使觀者起過大之錯誤。至使用面

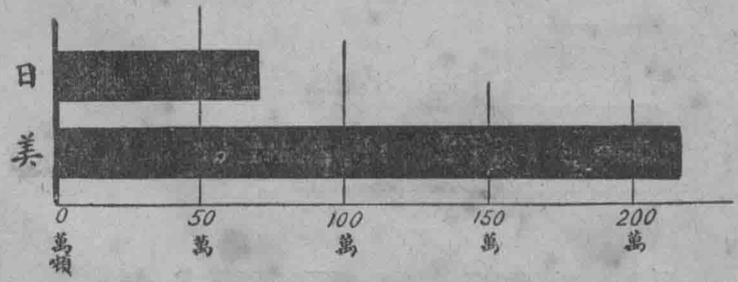

第 35 圖　某處之歷年求職人數




第 36 圖　第 35 圖之長條圖

積圖時，則其面積之計算既甚困難，且在繪成後又難於比較，如第 37 圖是依軍艦面積之大小以比較日美二國海軍之實力，對此軍艦面積之計算既甚困難，在畫成後又使閱者難以明瞭二者相差之程度，且該圖上又無年月之註明，愈使人難解。故此等繪畫統計圖只能使用於啓蒙及宣傳上，如在二者間相差愈大愈妙時，則使用其長度，愈小愈好時則使用其面積，以比較一切。


第 37 圖 難解之繪畫圖




第 38 圖 第 37 圖之長條圖

對於繪畫統計圖如只亟亟於引起世人之注目，而不留意於數值比較上之正確度時，其統計圖未免失其價值，故在製圖時對於數字之表現上，須有明確之規定，使閱者得知其實際情形。

將繪畫圖與他種統計圖相合併時，得成更有價值之統計圖。現略舉數例如次：

第 39 圖爲帶有長條圖色彩之繪畫圖。圖內是每一人形是代表千萬人口，並用黑色以表示農村人口，白色以表示城市人口。依圖知中俄日諸國農村人口占大部分，英法德諸國城市人口占多數，至美國則雙方相差無幾。

第 40 圖爲與長條合併之繪畫圖，一見即能明瞭各長條是用以表示各區牛之頭數。

第 41 圖爲美人年收入在九百美金之家庭費用支配狀況圖，其圖形之作成是於扇形統計圖之各部分上加以各項費用之代表形狀，依各代表物所占位置之如何，即能明白各項費用支出狀況。

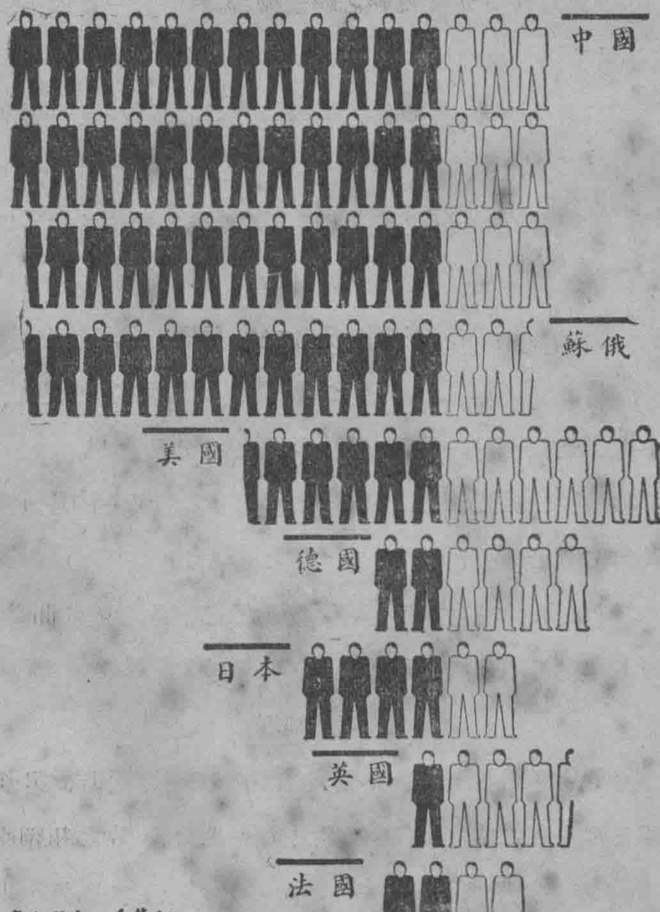
第 42 圖是於地圖上加入各統計值所當實物狀況，依圖知吾國各省火柴工廠分佈情形。

第 43 圖爲繪畫與坐標統計圖所合成之統計圖，依其曲線之高低，即能明瞭各年生活標準變動情形。

(F)統計地圖

統計地圖是用以表示統計數值與地域分布上之關係。其作成方法是於地圖上加入色之濃淡或異色，線之大小或疏密，點之粗細或多少等記號，以表示其各該所當數值。在地圖本身性質上，本已表示出地域鄰接之順序。故在地圖上更加以數值時，能將統計數值依地域的鄰接順序，以表示一切。比較其數值之大小與位置之所在，即能明瞭數值之集中或散漫等狀況，並能觀察其地方的變化與程度，使普通人一見即能明

# 世界各重要國之人口
## 及農村人口所佔之比數
(1926年)




第 39 圖

數字之單位爲千頭

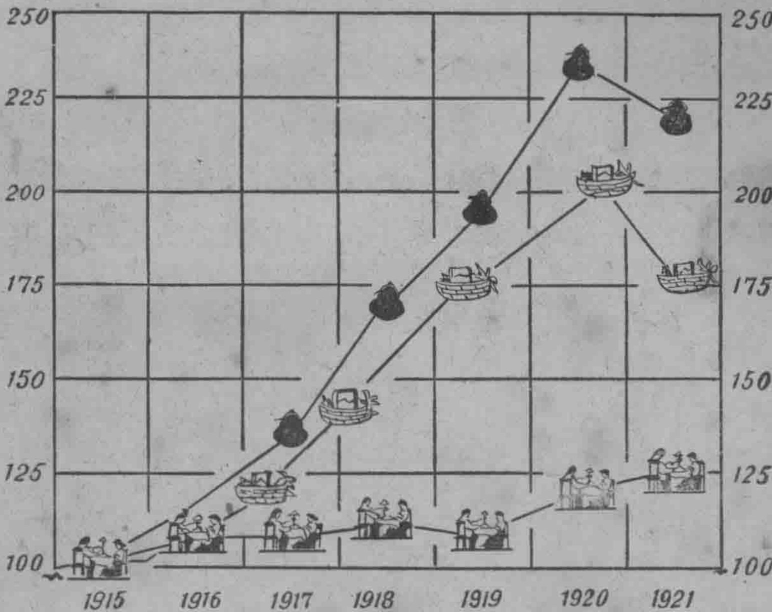
第40圖　與長條圖合併之繪畫圖

第41圖　與扇形合併之繪畫圖

# 中國各省火柴工廠

(1928年)


第 42 圖




第 43 圖　甲國生活標準之變動狀況

白各地方之位置風土地勢及其他之特質。依此地圖本身,無須十分正確,只求其能使觀者明白其爲某處已足。

統計地圖上用以表示數值之方法分爲次之三種:

(1)着色圖

用相異之色彩或同色之濃淡度,以區別數值之大小。在使用相異之色彩時,普通是用近於暗色之色彩以表示較大之數值。若將所用之色彩能巧妙分配時,實能得美麗與引人興趣之圖形,只因其使用色彩以致製作費用較大,故本法未必優於次之二種。第 44 圖爲江蘇省各縣等級及位置圖,是用三種色彩以表示縣之等級,依圖知江南多紅色,故知多一

等縣，江北多藍色故知多三等縣。（圖內因印刷關係，特將紅色改爲網狀形線，黃色改爲斜平行線，藍色改爲全白，以表示一切）。

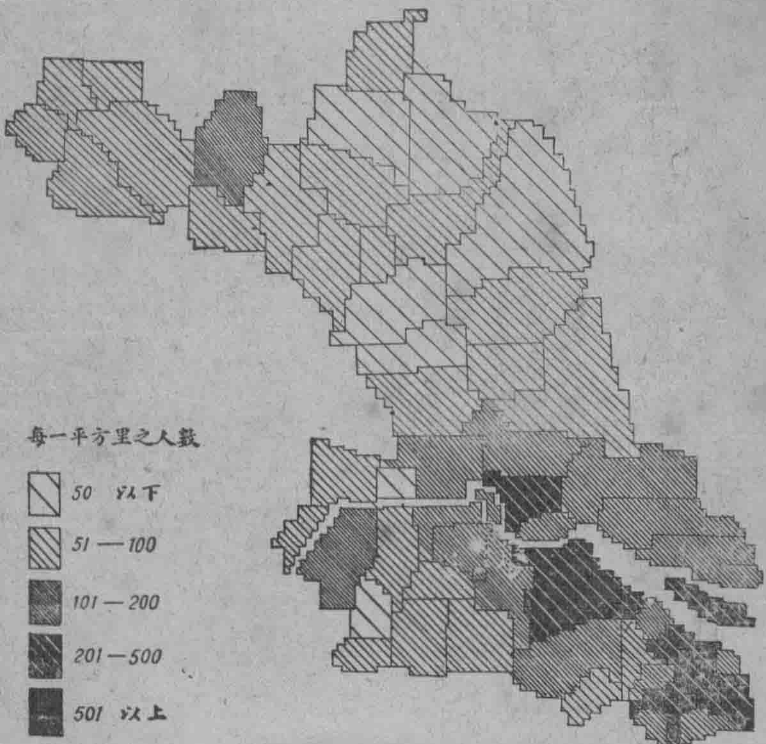
(2)線圖

將平行直線或網狀形直線引入於相當地域內，至其線之粗細與疏密則依數値之大小以決定之。第 45 圖爲江蘇省各縣人口密度比較圖是

江蘇省各縣等級及位置圖（民國二十年八月調查）


第 44 圖

江蘇省各縣人口密度比較（民國十七年年末）




第 45 圖

用平行之粗細直線以代表各數值所作成之統計地圖，依圖知長江兩岸所引線紋較爲粗密，故知兩岸人口密度較他處爲大。

着色圖及線圖在各地方之統計材料有相當差異及在同一地域內平等分布時，實甚適用。但實際上多使用於統計材料上只有微小之差異，及一地域內成爲不平等之分布上。因此易起其鄰接二地域一見成爲急激之變化或全無變化，及在同一地域內成爲平等分布等之錯誤，若欲除去此等錯誤，則須使用次之點圖。

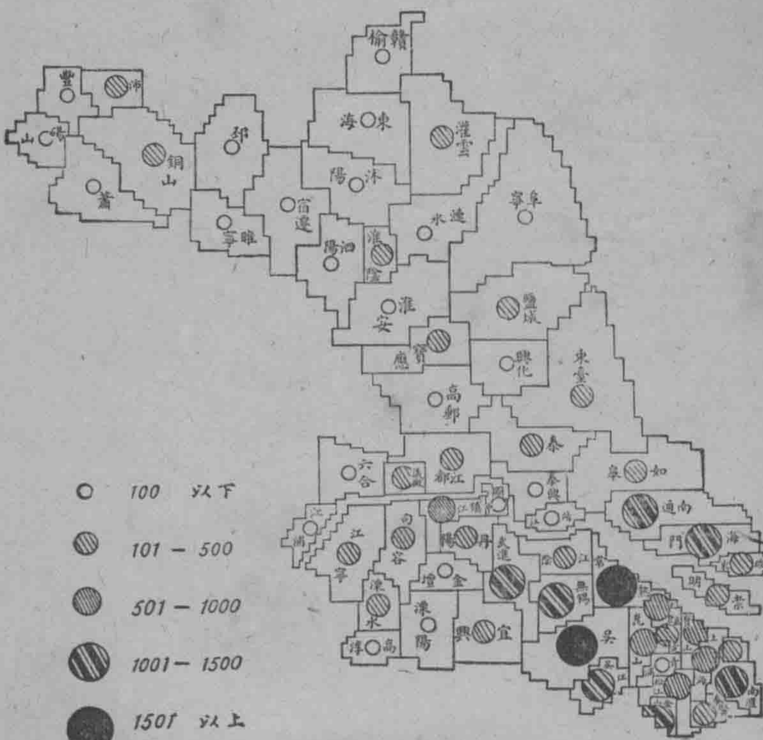
(3)點圖

依其使用圓點之種類，又得分次之三種，以行說明。

(a)用大小相異之圓點，以表示其數值之大小。至其點之所在位置，依統計材料之性質而不同。如第 46 圖爲比較江蘇省各縣違警罪發生件數之統計地圖，各代表圓點均取在縣城所在處。至本圖之便利處，是在能一見卽得比較各縣數值之大小。

(b)用陰影相異之同大圓點以表示各數值。至其圓點內之陰影，依數值之大小分爲五種，將全部暗色之圓點表示其最大階級，其下依次用 $\frac{3}{4}$，$\frac{1}{2}$，$\frac{1}{4}$，等部分暗色之圓點，以表示其次各階級數值，其最小階級數值則以全白圓點以表示之。其陰影之分配狀況，如第 47 圖。至其記入於地圖上之方法，是先將每一全部暗色圓點之所當數值，預行定妥後，用此數值以除所欲記入某區域所當數值，如此得全部暗色圓點之數目，其剩餘數值依上述陰影以分配後，得該區域所應塡入之圓點狀況。如第 48 圖爲江蘇省各縣公安能力狀況，圖內之每一全黑圓點代表 400 人，以下每百人分一階級。例如吳縣之公安人員爲 2455 人，現用 400 除

## 江蘇省違警罪發生件數縣別情形（民國十八年）

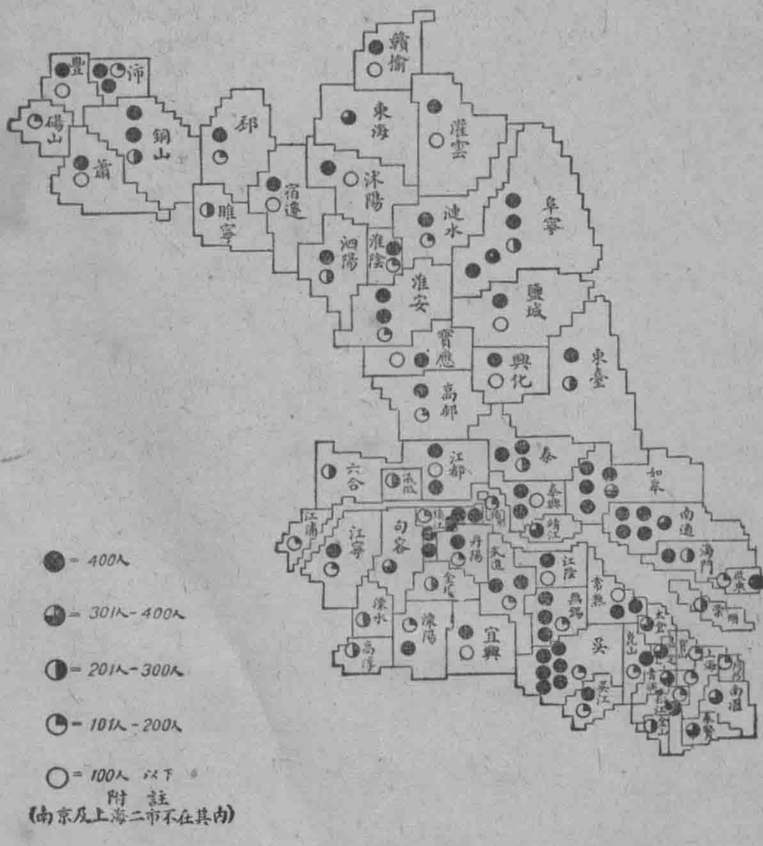



附註
南京及上海二市不在其內

第 46 圖

第 47 圖

## 江蘇省各縣公安能力狀況(民國二十年)



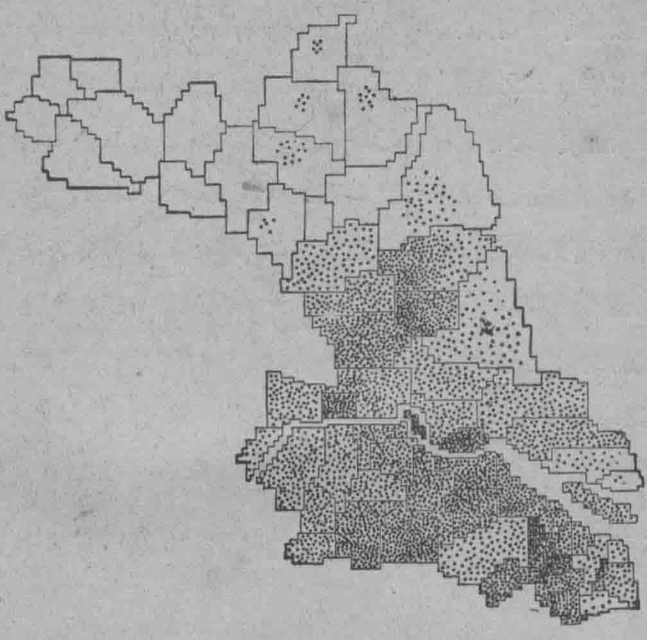

第 48 圖

之得商 6 與剩除 55，由此知該縣應塡圓點爲 6 個全黑圓點，與一個 $\frac{1}{4}$ 陰影圓點，將如此方法繼續進行卽得圖上情形。

(c)用極細圓點，依其數値之大小，集中散在之程度，以印入之。如第 49 圖爲江蘇省各縣秈粳稻產額之分布狀況，其江南各縣點數比江北爲密，故知江南秈粳稻產額比江北爲多，其印入圖內各點雖皆依同一數値（本圖每點代表 200 萬斤）預定在先，但有數縣因數値太大，以致印入所當區域內之圓點數非常衆多，難以分別淸楚，此點爲其不妥處。

江蘇省各縣秈粳稻產量狀況（民國十九年）

每點＝2,000,000斤



第　49　圖

## 第三節　坐標統計圖

上述各種統計圖，是於相異的主體間以比較其數值。例如比較各縣人口，其主體之各縣完全不能依數量以表示之。但在身長統計上。其人數既爲一變數 (variable)，同時其人數所屬主體之身長得爲數量之表示，故亦得稱之爲變數，如此變數成爲二個，依其身長之變化人數亦隨之而變化，此二變數相應變化，吾人依數學上之理由，得稱此二變數間有函數 (function) 的關係。又當觀察人口身長等之歷年變化狀況時，其人口身長等當然爲一變數，年月日等雖不能算爲純粹之變數，但因年月日是依相等距離以表示一切，故亦得算爲一種變數。如此成爲二個變數，且其變數間因依年月日之變遷而人口身長之數值亦隨之變化，故亦得稱爲有函數的關係。將此等關係圖示時，完全成爲曲線圖形。現應用此原理於統計上，將依統計方法所得之事實，用曲線形狀以表示一切。此等曲線統計圖形，吾人稱之爲坐標統統圖 (Coordinate graph)，其種類分爲

(1)次數分配圖，

(2)累加次數分配圖，

(3)比率分配圖，

(4)歷史數列圖，

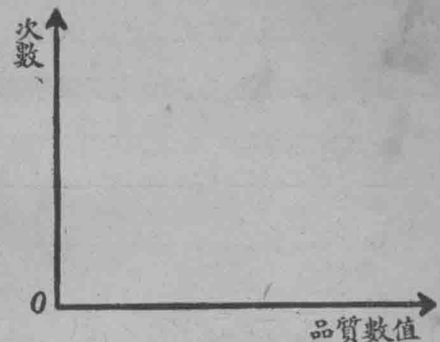
(5)三元坐標統計圖，

以下依次說明。

### (A)次數分配圖

於紙(普通用方格子紙)上定二直交軸，將品質數值取在橫軸方向

上，各品質值所屬次數取在縱軸方向上，如第50圖。並將橫軸上每一區劃之距離當品質值之幾單位，及縱軸上每一區劃當次數之幾單位等預先規定妥當。此二種比例之規定方法，並無一定之規則，只適應品質值及次數之範圍，使最大次數不成爲無理之巨大，及最小次數得能相當表現於紙面上已足。縱橫軸之區劃既已定妥後，次依各品質值及所屬次數，卽得各適應點於紙面上，用線連接此等適應點，卽成一折線之圖形，本圖形稱爲次數分配圖。



第　50　圖

在次數分配圖上，因其品質值有連續性與不連續性二種，故亦得分爲連續性與不連續性二種，以行說明。

在品質值爲不連續性時，其圖形之作成是先將表現各組值所當點取在橫軸上，次於此等所當點上各立以長度與其所當次數爲比例之垂直線，本法與一般圖示法之長條圖相類似，普通爲易於明瞭垂線長度之變化狀態起見，特用直線順次連接垂線之頂點。第51圖爲由第36表材料所作成之次數分配圖，其品質之各戶人數完全爲不連續性，故依上法求得如圖上之形狀。其順次連接各垂線頂點之直線，只用以易於明瞭次數之變化狀況而已，並無其他特別作用。

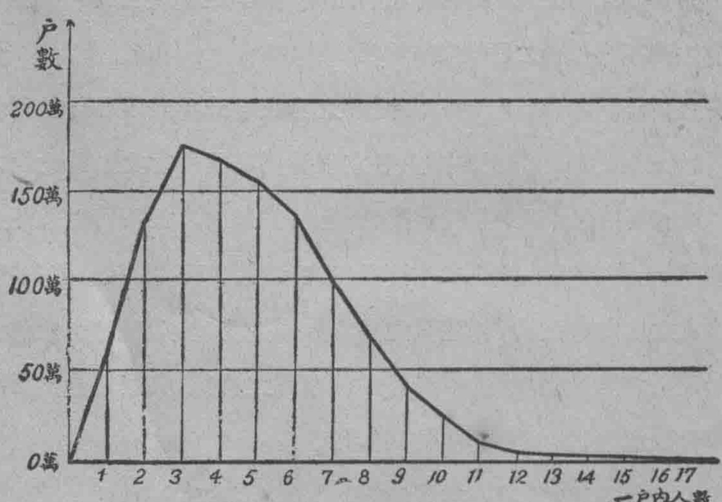
在品質值爲連續性時，其圖示方法又得分爲二種。

(1)於各組中心之相當點上，立以長度與所當次數爲比例之垂直線

並以直線順次連接各垂直線之頂點。起點在最小組之直前組中心所當橫軸上,終點在最大組之直後組中心所當橫軸上,如此得一折線,此折線稱爲次數分配折線,亦只用以易於明瞭次數之變化狀況而已,並無特別

第 36 表

| 各戶人數 | 戶數 |
|---|---|
| 一人 | 六三一,〇〇〇 |
| 二人 | 一,三四八,〇〇〇 |
| 三人 | 一,七六二,〇〇〇 |
| 四人 | 一,六九一,〇〇〇 |
| 五人 | 一,五七二,〇〇〇 |
| 六人 | 一,四〇九,〇〇〇 |
| 七人 | 一,四〇一,〇〇〇 |
| 八人 | 七〇八,〇〇〇 |
| 九人 | 四三六,〇〇〇 |
| 一〇人 | 二七一,〇〇〇 |
| 一一—一五人 | 二三一,〇〇〇 |
| 一六—二〇人 | 二三,〇〇〇 |
| 二一—二五人 | 八,〇〇〇 |
| 二六人以上 | 三,〇〇〇 |
| 合計 | 一一,一三四,〇〇〇 |

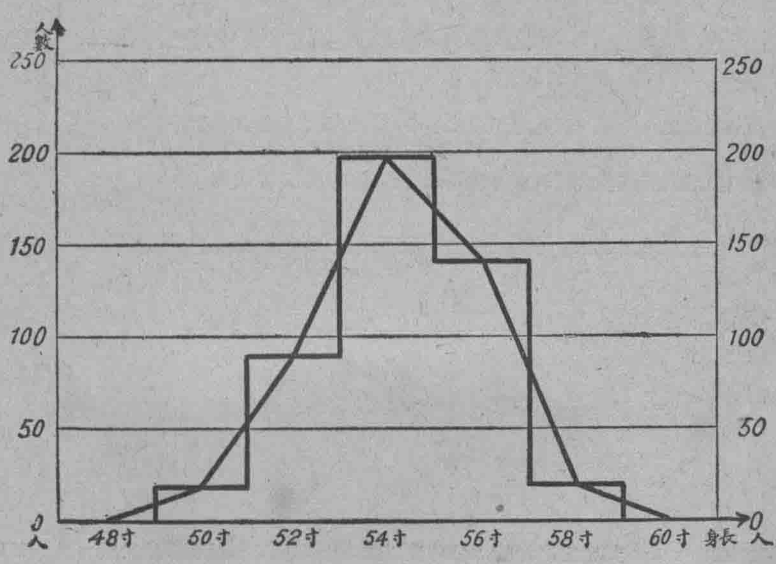



第 51 圖

作用。其折線與橫軸所包圍之面積與次數總和成為正比例，其包圍之圖形稱為次數多邊形(Frequency polygon, Häufigkeitspolygon)。第 52 圖內之略成為三角形之折線是由第 37 表之材料，依本法所作成之次數分配圖，其各組中心之垂直線為圖折線形之明晰起見，特不引入之。

第 37 表

| 身長組別 | 組之中心 | 人數 |
|---|---|---|
| 4 尺 9 寸——5 尺 1 寸 | 50 | 19 |
| 5 尺 1 寸——5 尺 3 寸 | 52 | 90 |
| 5 尺 3 寸——5 尺 5 寸 | 54 | 195 |
| 5 尺 5 寸——5 尺 7 寸 | 56 | 141 |
| 5 尺 7 寸——5 尺 9 寸 | 58 | 20 |
| 總數 | | 465 |




第 52 圖

至次數折線與橫軸所包圍之面積與總次數成爲正比例之證明如次:

設組之個數爲 K，各組之次數爲

$$f_1, f_2, f_3, \cdots\cdots\cdots\cdots f_K,$$

且

$$f_1 + f_2 + \cdots\cdots\cdots\cdots + f_K = n,$$

其分配圖內垂直線長度之每一單位適等於 r 個次數時，表示各組次數之垂直綫長度成爲

$$\frac{f_1}{r}, \quad \frac{f_2}{r}, \quad \frac{f_3}{r}, \quad \cdots\cdots\cdots \quad \frac{f_K}{r}。$$

次定品質值所當橫軸之每一單位長適等於 S 個品質時，表示組距(a)之長度爲$\frac{a}{s}$。

如此由最左點起至第一組之中心爲止所成三角形之面積爲 $\frac{1}{2}\cdot\frac{f_1}{r}\cdot\frac{a}{s}$,

其第一組中心與第二組中心所成梯形之面積爲$\frac{1}{2}\left(\frac{f_1}{r}+\frac{f_2}{r}\right)\cdot\frac{a}{s}$,

以下各相鄰二組中心所成梯形之面積爲

$$\frac{1}{2}\left(\frac{f_2}{r}+\frac{f_3}{r}\right)\cdot\frac{a}{s}, \qquad \frac{1}{2}\left(\frac{f_3}{r}+\frac{f_4}{r}\right)\cdot\frac{a}{s}, \quad \cdots\cdots$$

等，其由最後組之中心至其先端想像組之中心上所作成三角形之面積爲$\frac{1}{2}\cdot\frac{f_K}{r}\cdot\frac{a}{s}$。

將此等面積相加後卽得折線與橫軸所包圍之總面積，其數式爲

$$\frac{1}{2}\ \frac{f_1}{r}\ \frac{a}{s}+\frac{1}{2}\left(\frac{f_1}{r}+\frac{f_2}{r}\right)\frac{a}{s}+\frac{1}{2}\left(\frac{f_2}{r}+\frac{f_3}{r}\right)\frac{a}{s}+\cdots\cdots$$

$$+\frac{1}{2}\left(\frac{f_{K-1}}{r}+\frac{f_K}{r}\right)\frac{a}{s}+\frac{1}{2}\cdot\frac{f_K}{r}\cdot\frac{a}{s}$$

$$=\frac{1}{r}\ \frac{a}{s}(f_1+f_2+\cdots\cdots+f_K)$$

$$=\frac{1}{r}\cdot\frac{a}{s}\cdot n$$

在此式內 r, s, a 爲預定之常數，故知折線與橫軸所包圍之面積實與總次數 n 成爲正比例。

〔例〕在第 52 圖內其縱軸上每半英寸表示 50 人，橫軸上每半英寸表示其組距 2 寸，由此得

$$r=50/0.5=100,$$

$$S=2/0.5=4,$$

$$a=2,$$

次數折線與橫軸所包圍面積爲

$$\frac{1}{r}\ \frac{a}{s}\ n=\frac{1}{100}\times\frac{2}{4}\times465=2.325\ \text{平方英寸。}$$

(2) 於各組兩端之相當點上，立以長度與該組次數爲比例之垂直線，並以直線順次連接各垂直線頂點，如此於各組上得長度與次數爲比例之矩形，由此所成圖形之總體，稱爲直方圖(histogram, Block diagram, Stafelbild)。第 52 圖內之矩形折線是由本法所作成之次數分配圖，其各矩形面積之總和亦與總次數 n 成爲正比例。

(證)各組上矩形之高爲

$$\frac{f_1}{r},\ \frac{f_2}{r},\cdots\cdots\cdots\cdots\ \frac{f_K}{r}$$

各矩形之面積爲

$$\frac{f_1}{r}\cdot\frac{a}{s},\ \frac{f_2}{r}\cdot\frac{a}{s},\cdots\cdots\cdots\cdots\ \frac{f_K}{r}\cdot\frac{a}{s},$$

其面積總和爲

$$\frac{f_1}{r}\ \frac{a}{s}+\frac{f_2}{r}\ \frac{a}{s}+\cdots\cdots\cdots\cdots+\frac{f_K}{r}\ \frac{a}{s}$$

$$=\frac{1}{r}\ \frac{a}{s}(f_1+f_2+\cdots\cdots\cdots\cdots+f_K)$$

$$=\frac{1}{r}\cdot\frac{a}{s}\cdot n。$$

由上式得知其總面積與總次數 n 成爲正比例。

次數折線上順次相接之直線,得於連接點上改換其傾斜角度,且此傾斜角度依觀察次數之逐次增加及各組組距逐次縮小而變更其傾斜程度,使結局成爲一連續之曲線,此極限的曲線,稱爲次數曲線(Frequency curve)。次數曲線是能表現次數分配之理想的狀態。但由圖形上之折線以探求理想的曲線一事,爲事實所不能,故只能依次之三種方法,以求近似的曲線。

(1)用目光測量方法將折線變爲曲線,稱爲手描法(Free-hand method)。

(2)逐次舉行平均法以求近似的曲線,稱爲移動平均法(Moving-average method)。

(3)使用最小自乘法以求曲線,稱爲數學法(Mathematical method)。

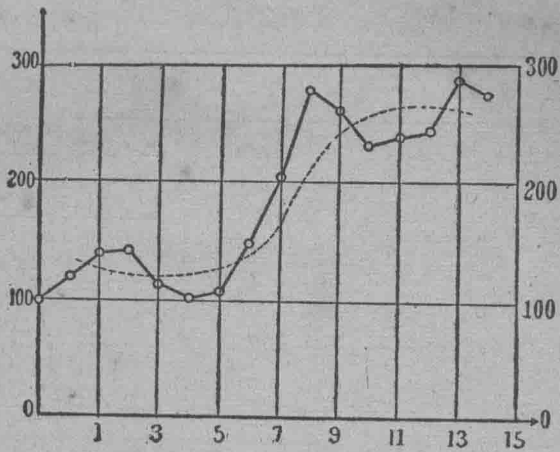
以下依次說明:

## (1)手描法

手描法是由目光測量其折線與橫軸所包圍面積仍等於曲線與橫軸所包圍面積然,順次改正折線爲圓滑之曲線。其曲線形狀雖依畫者而不同,但大體上是大同小異之近似的曲線,第53圖之細點線,是由本法將粗折線以改成之曲線。

## (2)移動平均法

當各階級所屬次數(例如年齡別人數)成爲不規則時,其折線形狀實甚複雜,但用本法逐次進行,得能求得圓滑的近似曲線,其作成方法如次:




第 53 圖

在第38表(甲)之第一第二兩欄爲給與之次數分配狀況,現由此狀況以求次數曲線。先由表內第三欄之平均計算,求得 $A', B', \cdots\cdots G'$ 等

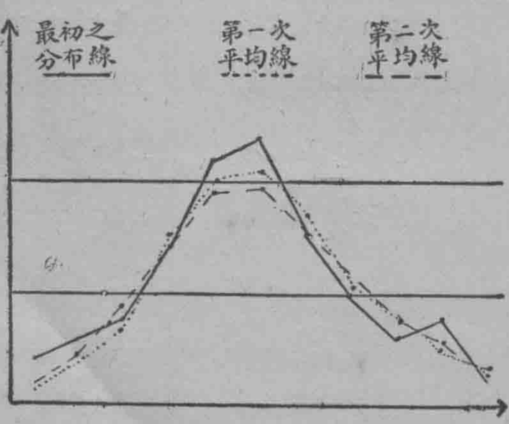
第 38 表 (甲)

| 階級 | 次數 | 第一次三數平均 | 第一次五數平均 |
|---|---|---|---|
| 1 | A | $\frac{另+A+B}{3}=A'$ | $\frac{另+另+A+B+C}{5}=A_1$ |
| 2 | B | $\frac{A+B+C}{3}=B'$ | $\frac{另+A+B+C+D}{5}=B_1$ |
| 3 | C | $\frac{B+C+D}{3}=C'$ | $\frac{A+B+C+D+E}{5}=C_1$ |
| 4 | D | $\frac{C+D+E}{3}=D'$ | $\frac{B+C+D+E+F}{5}=D_1$ |
| 5 | E | $\frac{D+E+F}{3}=E'$ | $\frac{C+D+E+F+G}{5}=E_1$ |
| 6 | F | $\frac{E+F+G}{3}=F'$ | $\frac{D+E+F+G+另}{5}=F_1$ |
| 7 | G | $\frac{F+G+另}{3}=G'$ | $\frac{E+F+G+另+另}{5}=G_1$ |

（乙）

| 階級 | 次數 | 第一次平均 | 第二次平均 |
|---|---|---|---|
| 1 | A | $\frac{\text{另}+A+B}{3}=A'$ | $\frac{\text{另}+A'+B'}{3}=A''$ |
| 2 | B | $\frac{A+B+C}{3}=B'$ | $\frac{A'+B'+C'}{3}=B''$ |
| 3 | C | $\frac{B+C+D}{3}=C'$ | $\frac{B'+C'+D'}{3}=C''$ |
| 4 | D | $\frac{C+D+E}{3}=D'$ | $\frac{C'+D'+E'}{3}=D''$ |
| 5 | E | $\frac{D+E+F}{3}=E'$ | $\frac{D'+E'+F'}{3}=E''$ |
| 6 | F | $\frac{E+F+G}{3}=F'$ | $\frac{E'+F'+G'}{3}=F'$ |
| 7 | G | $\frac{F+G+\text{另}}{3}=G'$ | $\frac{F'+G'+\text{另}}{3}=G''$ |

值，由此等數值所畫成折線，其傾斜程度實比由原次數 A，B，……G，等數值所畫成者為圓滑（參考第 54 圖）。若將其平均計算用第四欄之五值逐次平均法時，亦能達到所求之目的。當次數分布組數衆多時，得使用




第 54 圖

七值九值等奇數的移動平均法，以作成次數曲線。由現述方法所求得之平均數稱爲第一次平均數。

更如第 38 表（乙）內第四欄，是將三值平均之第一次平均數 A′,B……G′等看做爲最初之分配次數，再於此等 A′,B′……G′上施行求第一次平均數時所使用之方法，以求得第二次平均數 A″,B″……G″等數值，由此等數值所畫成之折線，其傾斜程度更比由第一次平均數所畫成者爲圓滑。例如第 54 圖內三折線，是由第 39 表內第二欄之最初分配次

第 39 表

| 身長組別（尺） | 人數 | 三數第一次（平均） | 三數第二次（平均） |
|---|---|---|---|
| 4.90——4.95 | 2 | $\frac{0+2+0}{3}=0.7$ | $\frac{0+0.7+2.0}{3}=0.9$ |
| 4.95——5.00 | 0 | $\frac{2+0+4}{3}=2.0$ | $\frac{0.7+2.0+3.7}{3}=2.1$ |
| 5.00——5.05 | 4 | $\frac{0+4+7}{3}=3.7$ | $\frac{2.0+3.7+7.3}{3}=4.3$ |
| 5.05——5.10 | 7 | $\frac{4+7+11}{3}=7.3$ | $\frac{3.7+7.3+10.0}{3}=7.0$ |
| 5.10——5.15 | 11 | $\frac{7+11+12}{3}=10.0$ | $\frac{7.3+10.0+10.3}{3}=9.3$ |
| 5.15——5.20 | 12 | $\frac{11+12+8}{3}=10.3$ | $\frac{10.0+10,3+8.3}{3}=9.5$ |
| 5.20——5.25 | 8 | $\frac{12+8+5}{3}=8.3$ | $\frac{10.3+8.3+5.3}{3}=8.0$ |
| 5.25——5.30 | 5 | $\frac{8+5+3}{3}=5.3$ | $\frac{8.3+5.3+4.0}{3}=5.9$ |
| 5.30——5.35 | 3 | $\frac{5+3+4}{3}=4.0$ | $\frac{5.3+4.0+2.7}{3}=4.0$ |
| 5.35——5.40 | 4 | $\frac{3+4+1}{3}=2.7$ | $\frac{4.0+2.7+1.7}{3}=2.8$ |
| 5.40——5.45 | 1 | $\frac{4+1+0}{3}=1.7$ | $\frac{2.7+1.7+0}{3}=1.5$ |

數，第三欄之第一次平均數及第四欄之第二次平均數所畫成之折線各線之傾斜情形，一見卽知由第一次平均數所畫成者，比由最初分配次數所畫成者爲圓滑，其由第二次平均數所畫成者，更比由第一次平均數所畫成者爲圓滑。由此知將此平均法逐次舉行後，實能得一圓滑之近似曲線。本法簡單容易故爲統計家所樂用。

(3)數學法

設各品質單位值爲

$$X_1, X_2, \cdots\cdots X_n,$$

次數爲

$$Y_1, Y_2 \cdots\cdots Y_n,$$

時，代表各單位值之橫坐標 X 與代表各次數之縱坐標 Y 相交得交點P，如此共有 n 點

$$P_1(X_1, Y_1), P_2(X_2, Y_2), \cdots\cdots P_n(X_n, Y_n)$$

分佈在一平面上，現求能平均的通過此等交點間之曲線。設所求曲線之方程式爲

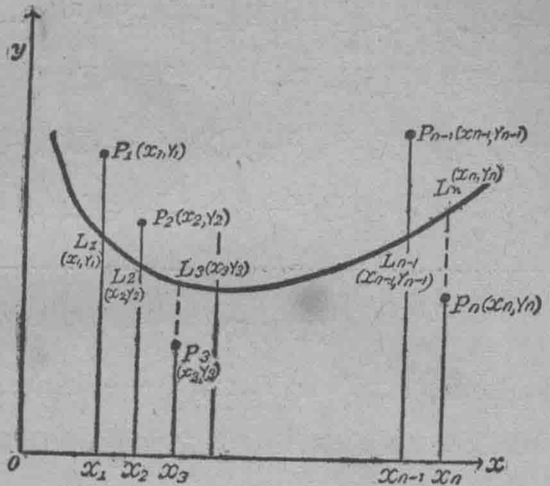
$$y = a_0 + a_1x + a_2x^2 + a_3x^3 + \cdots\cdots,$$

式內各係數 $a_0, a_1, a_2, a_3 \cdots\cdots$ 等，須由下法以求得之，只要求得此等係數之所當值後，卽能得所求曲線之方程式，最後由此方程式以畫其所當曲線，此曲線卽爲所求之曲線，

現定各交點 P 與曲線在縱坐標上所差距離，依第 55 圖知爲

$$P_rL_r = Y_r - y_r = Y_r - (a_0 + a_1x_r + a_2x_r^2 + a_3x_r^3 + \cdots\cdots)$$

$$(r = 1, 2, \cdots\cdots n),$$




第　55　圖

此偏差 $P_rL_r$，因交點之在曲線上或曲線下而異其符號，故欲曲線與各點所差距離成爲最小，必須使用最小自乘法 (Method of least squares)而後始能決定曲線之方程式。

依最小自乘法知由

$$S=\sum_{r=1}^{n}(Y_r-y_r)^2=\sum_{r=1}^{n}[Y_r-(a_0+a_1x_r+a_2x_r^2+a_3x_r^3+\cdots\cdots)]^2$$

$=$最小

以決定各係數 $a_i(i=0,1,2,\cdots\cdots)$,但依微分學定理得

$$\frac{\partial s}{\partial a_i}=0,\qquad (i=0,1,2,\cdots\cdots),$$

由此得一聯立方程式

$$\Sigma Y_r = na_0 + a_1\Sigma x_r + a_2\Sigma x_r^2 + a_3\Sigma x_r^3 + \cdots\cdots,$$
$$\Sigma x_r Y_r = a_0\Sigma x_r + a_1\Sigma x_r^2 + a_2\Sigma x_r^3 + a_3\Sigma x_r^4 + \cdots\cdots,$$
$$\Sigma x_r^2 Y_r = a_0\Sigma x_r^2 + a_1\Sigma x_r^3 + a_2\Sigma x_r^4 + a_3\Sigma x_r^5 + \cdots\cdots,$$
$$\Sigma x_r^3 Y_r = a_0\Sigma x_r^3 + a_1\Sigma x_r^4 + a_2\Sigma x_r^5 + a_3\Sigma x_r^6 + \cdots\cdots,$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

(公式 5)

在上之聯立方程式內,$\Sigma Y_r$, $\Sigma x_r Y_r$, $\Sigma x_r^2 Y_r$……及$\Sigma x_r$, $\Sigma x_r^2$, $\Sigma x_r^3$,……等爲已知數,故解此聯立方程式時則得 $a_i$ 之所當值,將此等所當值代入

$$y = a_0 + a_1x + a_2x^2 + a_3x^3 + \cdots\cdots$$

內,卽得所求之曲線方程式。

〔例〕求第 40 表之次數曲線。

第　40　表

| X | Y | x | $x^2$ | xY | $x^2Y$ | $x^3$ | $x^4$ |
|---|---|---|---|---|---|---|---|
| 1 | 149 | -3 | 9 | -447 | 1341 | -27 | 81 |
| 2 | 203 | -2 | 4 | -406 | 812 | -8 | 16 |
| 3 | 278 | -1 | 1 | -278 | 278 | -1 | 1 |
| 4 | 266 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5 | 236 | 1 | 1 | 236 | 236 | 1 | 1 |
| 6 | 241 | 2 | 4 | 482 | 964 | 8 | 16 |
| 7 | 240 | 3 | 9 | 720 | 2160 | 27 | 81 |
| 計 | 1613 | 0 | 28 | 107 | 5791 | 0 | 196 |

設所求曲線之方程式爲

$$y = a_0 + a_1x + a_2x^2,$$

依

$$S=\Sigma[Y_r-(a_0+a_1x_r+a_2x_r^2)]^2$$

$$=最小,$$

須求

$$\frac{\partial s}{\partial a_0}=2\Sigma[Y_r-(a_0+a_1x_r+a_2x_r^2)]=0,$$

$$\frac{\partial s}{\partial a_1}=2\Sigma[Y_r-(a_0+a_1x_r+a_2x_r^2)]x_r=0,$$

$$\frac{\partial s}{\partial a_2}2\Sigma[Y_r-(a_0+a_1x_r+a_2x_r^2)]x_r^2=0。$$

由此得聯立方程式

$$\Sigma Y_r=na_0+a_1\Sigma x_r+a_2\Sigma x_r^2,$$

$$\Sigma x_rY_r=a_0\Sigma x_r+a_1\Sigma x_r^2+a_2\Sigma x_r^3,$$

$$\Sigma x_r^2Y_r=a_0\Sigma x_r^2+a_1\Sigma x_r^3+a_2\Sigma x_r^4。$$

又在上表內之順次相差均為一，故得改用第三欄之 x，使 $\Sigma x$，$\Sigma x^3$ 均等於 0，將計算趨於簡單。現將如此改換後各所當值代入上之聯立方程式內，得

$$1613=7a_0 \qquad\qquad +28a_2$$

$$107=\qquad 28a_1$$

$$5791=28a_0 \qquad\qquad +196a_2$$

解此得

$$a_0=261.9,$$

$$a_1=3.8,$$

$$a_2=-7.9,$$

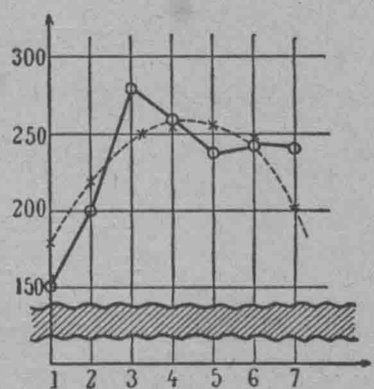
由此得所求曲線之方程式為

$$y=261.9+3.8x-7.9x^2,$$

現由此方程式求各 x 之所當 y 值如次表：

| X | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| x | -3 | -2 | -1 | 0 | 1 | 2 | 3 |
| Y | 179 | 223 | 250 | 262 | 258 | 248 | 202 |

由此等 y 值得第 56 圖內之虛曲線。




第　56　圖

(4)次數曲線之種類

曲線之種類雖甚衆多，但在統計上所最多使用者為次之數種：

(a)對稱形(Symmetrical frequency curve),

(b)非對稱形,

1. 左傾非對稱形(left-sided asymmetrical f. c.),

2. 右傾非對稱形(right-sided asymmetrical f. c.),

(c)J 字形(J-shaped frequency curve),

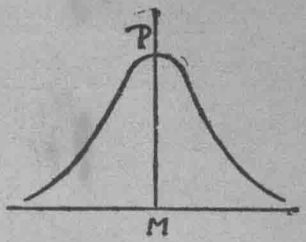
(d)U 字形(U-shaped frequency curve),

(c)複雜形(multimodal frequency curve),

以下依次說明。

(a)對稱形之曲線形狀如第 57 圖，其由曲線之最高點 P 向左右之

傾斜程度實相等。現以由P點至基線之垂線PM，爲折縫以打疊圖形時，其左右二曲線實相一致。本曲線形在社會現象上雖甚稀少，但在自然現象上則甚衆多。

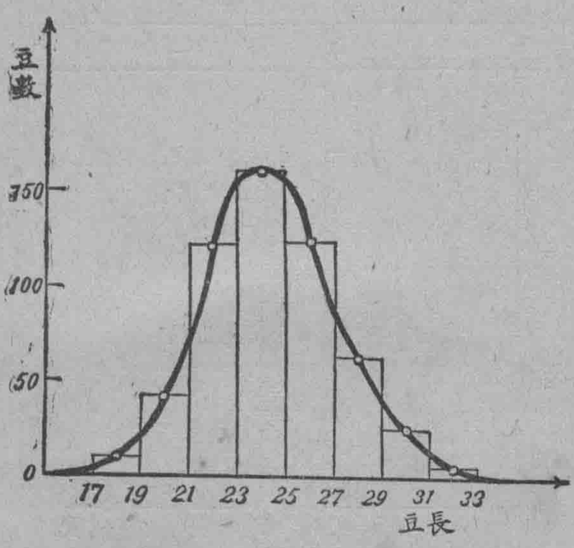



第 57 圖

〔例〕

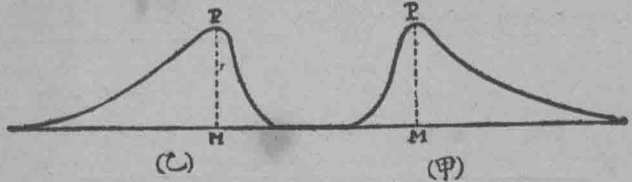
第 41 表

| 豆長 (mm) | 豆數 |
|---|---|
| 17——19 | 10 |
| 19——20 | 44 |
| 21——23 | 122 |
| 23——25 | 160 |
| 25——27 | 128 |
| 27——29 | 64 |
| 29——31 | 25 |
| 31——33 | 5 |
| 計 | 558 |




第 58 圖

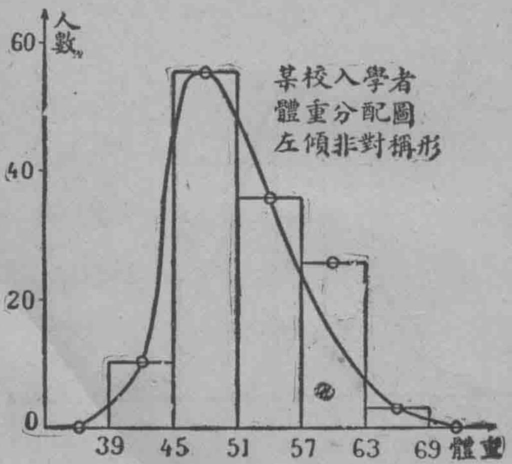
(b)非對稱形曲線之由最高點P所下垂線PM,必偏於圖之一邊。其左右二曲線之傾斜程度完全不相等,其偏於左方者(如第59圖甲)稱爲左傾非對稱形,偏於右方者(第59圖乙)稱爲右傾非對稱形。




第　59　圖

〔例I〕左傾之例　　第　42　表

| 體　重　(瓩) | 人　數 |
|---|---|
| 39——45 | 10 |
| 45——51 | 56 |
| 51——57 | 36 |
| 57——63 | 27 |
| 63——69 | 3 |
| 合　計 | 132 |

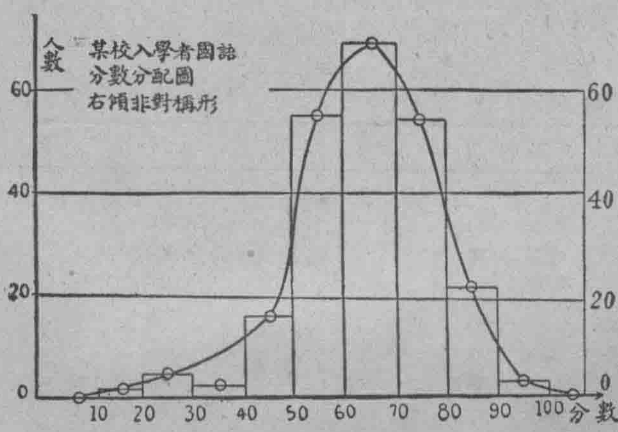



第　60　圖

〔例 II〕右傾之例

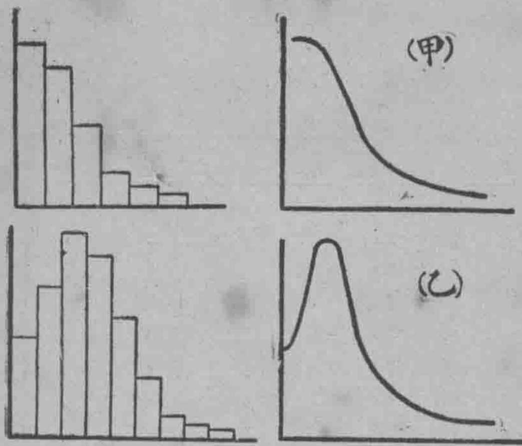
第 43 表

| 分數 | 人數 |
|---|---|
| 0——10 | 0 |
| 10——20 | 2 |
| 20——30 | 4 |
| 30——40 | 2 |
| 40——50 | 16 |
| 50——60 | 57 |
| 60——70 | 69 |
| 70——80 | 56 |
| 80——90 | 21 |
| 90——100 | 2 |
| 計 | 229 |




第 61 圖

(c) J 字形之曲線形狀如第 62 圖甲,因其形狀酷似 J 字,故有此名。非對稱形曲線中之極端者,亦得看做爲 J 字形,如第 62 圖乙。

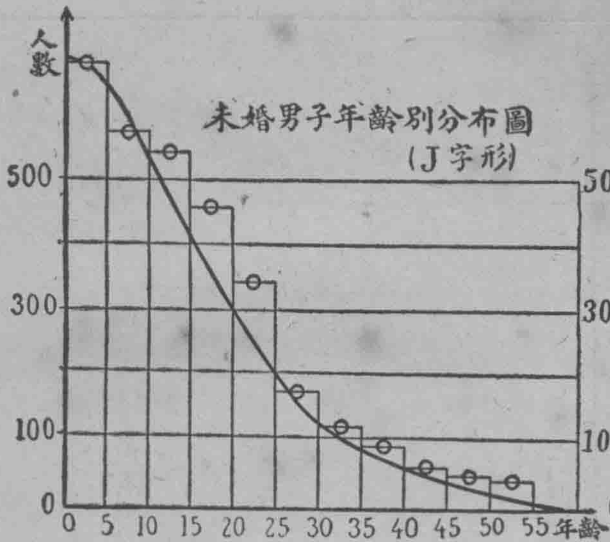



第 62 圖

〔例〕第 44 表爲甲國人口萬人中男子未婚者之年齡別分配表，將該表圖示時成爲第 63 圖之 J 字形曲線。

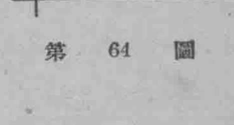
第 44 表

| 年 齡 | 男 子 數 |
|---|---|
| 0—5 | 674 |
| 5—10 | 562 |
| 10—15 | 534 |
| 15—20 | 465 |
| 20—25 | 339 |
| 25—30 | 185 |
| 30—35 | 108 |
| 35—40 | 81 |
| 40—45 | 59 |
| 45—50 | 57 |
| 50—55 | 47 |
| 以下略 | |




第 63 圖

(d) U 字形之曲線形狀如第 64 圖，在曲線之左右二端次數爲最大，其中間則逐次減少，因其形狀酷似 U 字，故有此名。第 70 圖之各曲線(除日本外)均成此形。

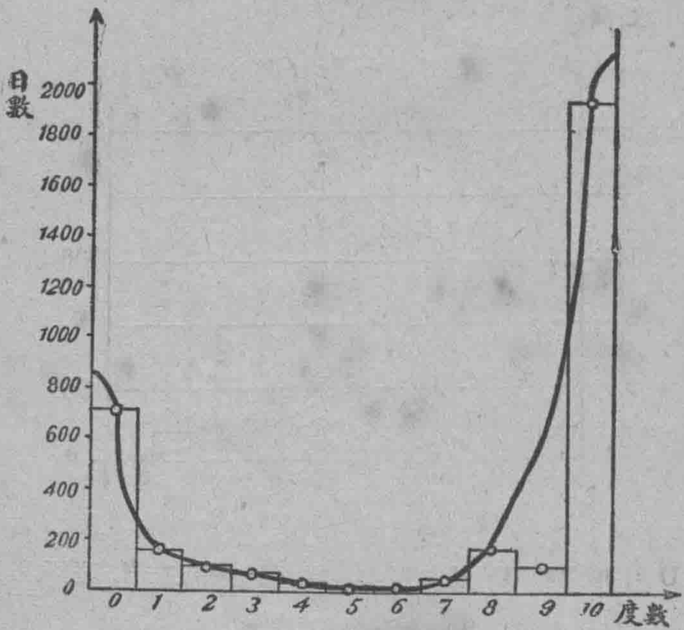


第 64 圖

〔例〕德國某氣象臺之十年間觀察曇之結果，如第 45 表，其分配曲線如第 65 圖成爲 U 字形。

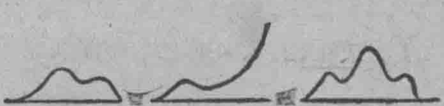
第 45 表

| 曇之度數 | 日數 | 曇之度數 | 日數 |
|---|---|---|---|
| 0.0 | 751 | 0.7 | 71 |
| 0.1 | 179 | 0.8 | 194 |
| 0.2 | 107 | 0.9 | 117 |
| 0.3 | 69 | 1.0 | 2089 |
| 0.4 | 46 | | |
| 0.5 | 9 | | |
| 0.6 | 21 | 計 | 3953 |




第 65 圖

(e)在實際問題上所作成之次數曲線，常不能若上述之簡單，多成爲複雜形之曲線。如第 66 圖爲其中之一部分。(但此處所述之複雜形曲線，並非是由觀察次數稀少以致所作成之曲線成爲不規則之複雜形，若增加其次數時仍能成爲單純形之曲線)。



第 66 圖

此等複雜形曲線多能看做爲二個以上之簡單形曲線之合成物。例如第 46 表內甲乙二分配總數均爲 100，其最大次數雙方均只有一個，現從甲之次數分配依公平選擇方法取出 40 個，(各組分配比例仍不變)

再從乙之次數分配公平選出 60 個，將此雙方所選出者合為一起成為一新的次數分配，其情形如第 47 表。

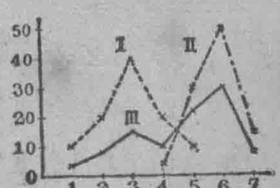
第 46 表

| 甲 | | 乙 | |
|---|---|---|---|
| 單位值 | 次數 | 單位值 | 次數 |
| 1 | 10 | 4 | 5 |
| 2 | 20 | 5 | 30 |
| 3 | 40 | 6 | 50 |
| 4 | 20 | 7 | 15 |
| 5 | 10 | | |
| | 100 | | 100 |

第 47 表

| 甲 | | 乙 | | 新 | |
|---|---|---|---|---|---|
| 單位值 | 次數 | 單位值 | 次數 | 單位值 | 次數 |
| 1 | 4 | | | 1 | 4 |
| 2 | 8 | | | 2 | 8 |
| 3 | 16 | | | 3 | 16 |
| 4 | 8 | 4 | 3 | 4 | 11 |
| 5 | 4 | 5 | 18 | 5 | 22 |
| | | 6 | 30 | 6 | 30 |
| | | 7 | 9 | 7 | 9 |
| 計 | 40 | 計 | 60 | 計 | 100 |

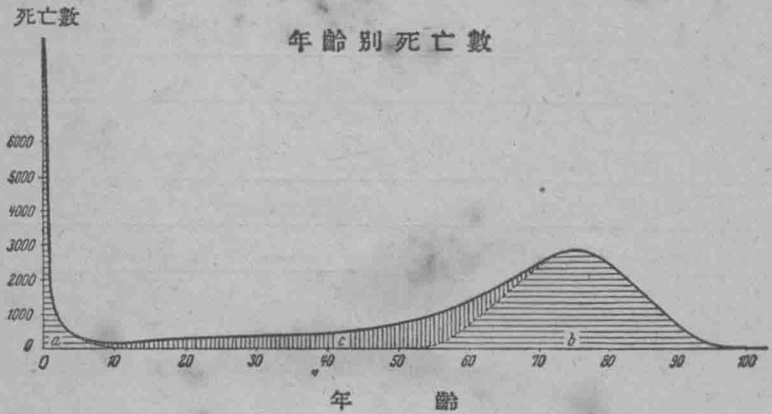
圖示上述三種分配情形得第 67 圖，圖內之 I 為由原有甲之次數分配(總數 100) II 為由原有乙之次數分配(總數 100) 所作成之次數多角形，III 為由上述方法以合成後之新的次數分配所作成之次數多角形，其峯尖有




第 67 圖

二個。依此知有二個峯尖之複雜形曲線，得由只有一個峯尖之簡單曲線以合成之。

關於複雜形次數曲線之分解上，爲近代統計學者所最感興趣者，要推死亡者年齡別次數分配曲線。勒克斯(Wilhelm Lexis)將此分解爲三部分，(如第 68 圖)第一爲 J 字形，是 0 歲至 10 歲的一部分(圖內之 a)第二爲對稱形，是當時所稱常態年齡階級以約 70 歲爲中心之一部分




第 68 圖

(圖內之 b)，第三是在第一與第二之間(圖內之 c)。且勒克斯又依此以分解生存者爲

(1)常態生存者即壽命達 60 歲以上者，

(2)不能生存者即在最初數年內夭卒者，

(3)劣等生存者即活至 10 歲至 60 歲之間者等三大集團。

又披爾遜(K. Pearson)將上述死亡年齡分配曲線分解爲

(1)幼年死亡者(J 字形)，

(2)少年死亡者(稍偏於左方之非對稱形)，

(3)青年死亡者(近似的對稱形)，

(4)中年死亡者(近似的對稱形)，

(5)老年死亡者(稍偏於右方之非對稱形)，

等五個簡單分配曲線。

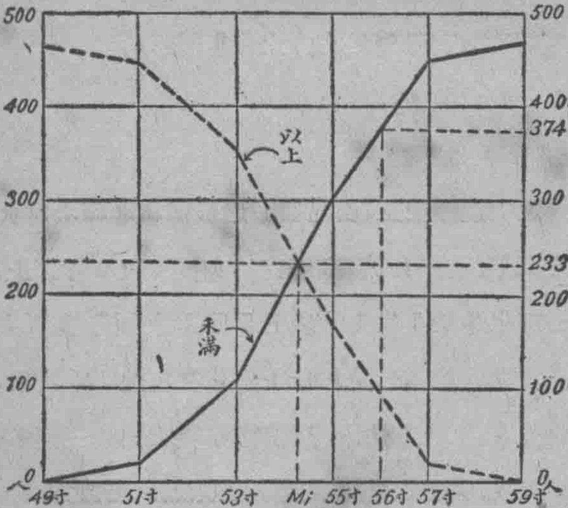
(B)累加次數分配圖

於普通之次數分配上，由品質值之最低組起順次將次數相加後，得累加次數分配狀況。例如第 48 表內於身長 4 尺 9 寸以上，5尺1寸未滿之 19 人上，加以身長 5 尺 1 寸以上 5 尺 3 寸未滿者 90 人後，得身長 5 尺 3 寸未滿者共 109 人。更於此上加以次組之 195 人後，將 5 尺 5 寸未滿者共 304 人。依此方法以行累加後，得表內第三欄各數值。此等數值稱爲累加次數。現將此累加次數分配狀況，依座標方法，於橫軸上記入組之境界值所當各點，於縱軸上取其累加次數。但當此時之累加次數之代表長，必須取在所屬××未滿之所當點上 例如在第 48 表內身長 5 尺 3 寸未滿者共 109 人，此 109 人之代表長，必須取在 5 尺 3 寸所當點上。

第　48　表

| 身長 | 人數 | 累加人數 | |
|---|---|---|---|
| | | 未滿 | 以上 |
| 4尺9寸——5尺1寸 | 19 | 19 | 465 |
| 5尺1寸——5尺3寸 | 90 | 109 | 446 |
| 5尺3寸——5尺5寸 | 195 | 304 | 357 |
| 5尺5寸——5尺7寸 | 141 | 445 | 161 |
| 5尺7寸——5尺9寸 | 20 | 465 | 20 |
| 計 | 465 | …… | …… |

依如此方法以圖示時，得累加次數多角形(ogive, summenpolygon)，第69圖內之實折線是由本法所作成者。此等圖形，稱爲累加次數分配圖。




第 69 圖

上述之累加分配是由品質值之最小組起順次累加其次數，依此知此等累加次數是爲給與某值以下各值之次數。若欲得某值以上各值之次數時，須由品質值之最大組數起順次累加其次數。如第48表內第四欄各數值，是由最高組起順次相加後所得結果，由此知身長4尺9寸以上者爲465人，身長5尺1寸以上者爲446人等分配狀況，將此分配狀況圖示時，得第69圖內之虛折線。故知累加次數分配上有未滿(less than)分配與以上(more than)分配二種，其圖形適相反。

表示累加次數之折線亦得依種種方法，將折線變爲曲線。在此累加

次數分配曲線上，有次之二種功用。

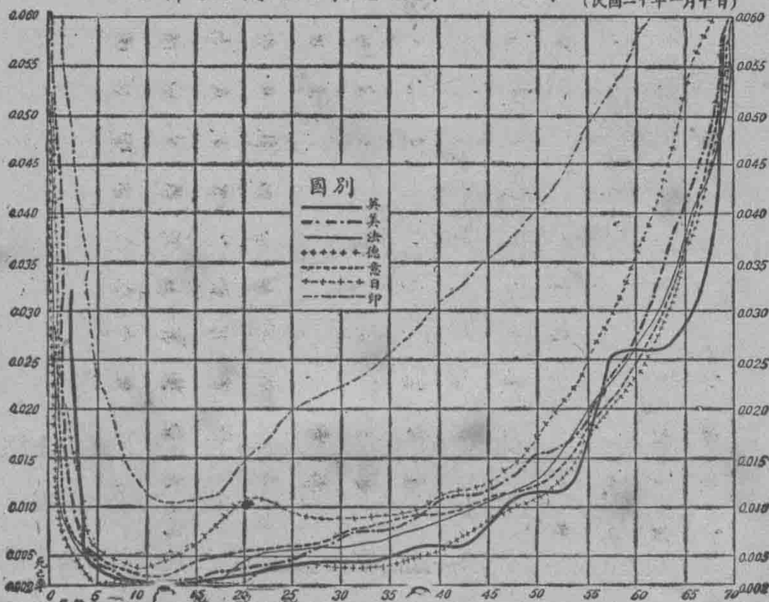
(1)由橫軸上任意一點立一垂直線與曲線相交，其由交點至橫軸之距離，是給與比此任意點所當品質值以下或以上各單位值之近似次數。

(2)能依簡單方法，求得該次數分配之中位數 (medial) 與最衆數 (mode)，此等求法請參照次章平均數。

## (C)比率分配圖

上述次數分布圖之縱軸上是表示實數（人數戶數等），但在統計學上常有不使用實數而改用百分比千分比等比率者。例如第 70 圖爲各國

世界主要各國死亡率比較圖 (民國二十年一月十日)

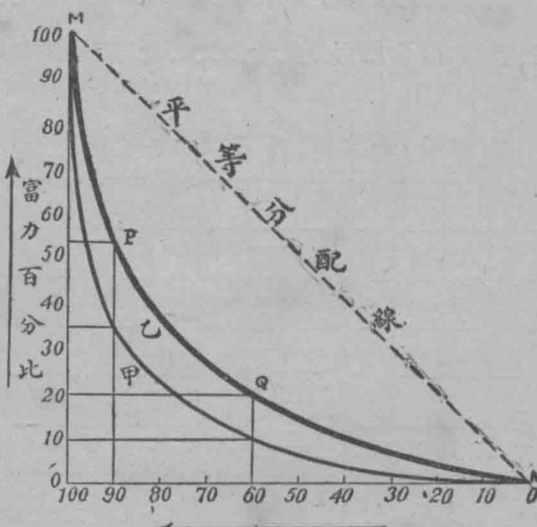



第 70 圖

年齡別死亡率比較狀況，其橫軸上雖仍爲實數之年齡，但在縱軸上是使用千分比，與上述次數分配圖所不同處。

在比率分配圖內縱橫二軸均用比率者，有羅蘭曲線(Lorenz curve)。

羅蘭曲線是用以表示一國國富，土地之所有，團體收入及工資分配等問題之圖形，爲美國羅蘭(M. O. Lorenz)所發明，故有此名。第 71 圖爲其假想例，其作成方法如次




第 71 圖

最初在縱橫二軸上切取等長之距離，並在此等長之直線上施以百分比之分點（縱軸由下至上橫軸由右至左），其橫軸上數字是表示一國之人口，縱軸上數字是表示一國之富力。其曲線之作成方法是先從國富全部當然屬於全國人民處着手進行，將國富（縱軸）100% 及人口（橫軸）100% 之所當點 M 爲起點。次將最富階級人口占全人口之 10% 者其所

有財產占全國富之百分之幾，現假定其占 47%，如此於縱軸上由 100% 處起向下引去 47% 至 53% 處，橫軸上由 100% 處起向右引去 10% 至 90% 處，將此二處所當距離爲座標定一點 P。更於第二位有產階級依其人口與富力之百分比同樣得一點於紙面上。將此方法逐次進行，至最後其國富爲 0 時當然其人口亦等於 0，將其所當點 N 作爲終點。如此得多數交點於圖面上，連接此等交點後得一曲線，此曲線稱爲羅蘭曲線，依此曲線之形狀，得判斷一國人民對於國富分配上之情形。若分配完全平等時，各人占有同等之國富，在最初 10% 之人民占有 10% 之國富時，其次各項 10% 之人口亦均須占有同等比例之國富，故其連接各相當點之曲線，成爲一直線。因此知當繪製某國國富分配情形，若其 MN 一線成爲直線時，即知該國國富完全平均分配於民衆。若有產階級之富力愈占多額時，則其曲線愈形彎曲，與直線相離愈大。在第 71 圖內甲線之彎曲比乙線爲大，故知甲國之國富分配比乙國爲不平均。其分配太不平等之國家實甚危險，故執政者當時時作成國富分配曲線以資參考，若見曲線逐次彎曲時，當使用種種補救方法，使曲線趨於平直，以資挽回。

詳情請參考 M. O. Lorenz, "Method of Measuring the Concentration of Wealth", Quarterly Publication of the Americal Statistical Association, June 1905. PP. 209-219。

〔例〕第 49 表爲美國麻省在 1889-1891 年間所有死亡者遺產分配情形，現由此材料以求羅蘭曲線。

(左表之第三欄內數值是由實際數值計算而來)。

現將下表數值順次累加後得累加次數分配表如第 50 表。

## 第 49 表

| 遺產(單位千金元) | 人數 | 各組之全遺產(單位千金元) |
|---|---|---|
| —— 0.5 | 23,151 | 9,260 |
| 0.5 —— 1 | 1,466 | 1,378 |
| 1 —— 5 | 5,715 | 15,432 |
| 5 —— 10 | 2,005 | 14 637 |
| 10 —— 25 | 1,612 | 25,791 |
| 25 —— 50 | 537 | 18,901 |
| 50 ——100 | 334 | 23,480 |
| 100 ——200 | 176 | 24,148 |
| 200 ——300 | 59 | 14,632 |
| 300 ——400 | 28 | 9,492 |
| 400 ——500 | 18 | 8,270 |
| 500 —— | 47 | 37,274 |

## 第 50 表

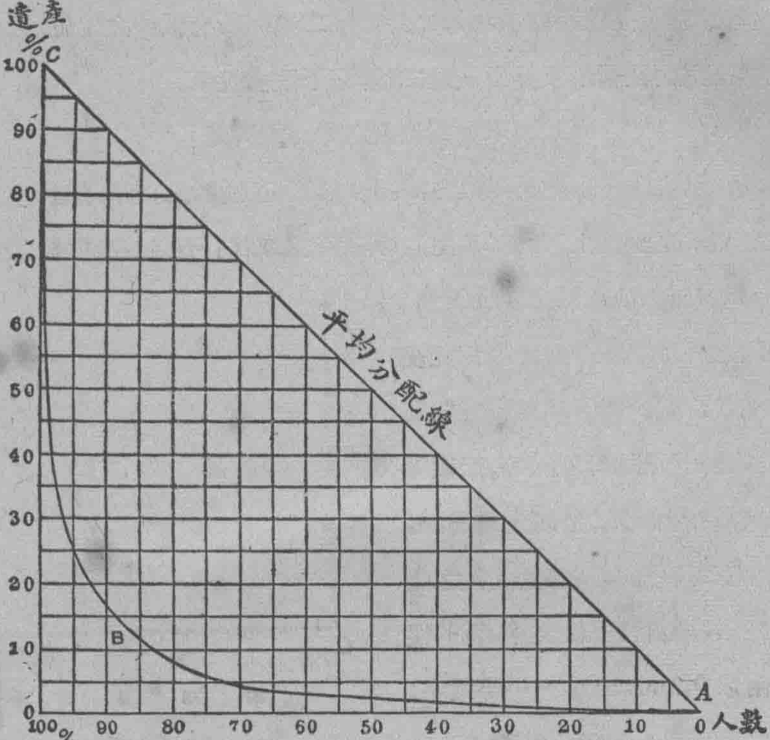
| 遺產組別 | 人數 | 百分比 | 遺產額 | 百分比 |
|---|---|---|---|---|
| 0.5 以下 | 23,151 | 65.86 | 9,260 | 4.57 |
| 1 以下 | 26,617 | 70.04 | 10,638 | 5.25 |
| 5 以下 | 30,332 | 86.30 | 26,070 | 12.86 |
| 10 以下 | 32,337 | 92.00 | 40,707 | 20.08 |
| 25 以下 | 33,949 | 96.59 | 66,498 | 32.81 |
| 50 以下 | 34,486 | 98.12 | 85,399 | 42.13 |
| 100 以下 | 34,820 | 99.07 | 108,879 | 53.72 |
| 200 以下 | 34,996 | 99.57 | 133,027 | 65.63 |
| 300 以下 | 35,055 | 99.73 | 147,659 | 72.85 |
| 400 以下 | 35,083 | 99.81 | 157,151 | 77.53 |
| 500 以下 | 35,101 | 99.87 | 165,421 | 81.61 |
| 全遺產 | 35,148 | 100.00 | 202,695 | 100.00 |

上表內第二第四兩欄數值是由第 49 表之第二第三兩欄數值累加而成，第三欄之百分比是以全人數 35148 看做爲 100 以計算各數值，例如人數 23151 之所當百分比爲

$$\frac{23151}{35148}\times 100=65.86,$$

即23151人當全人數之65.86%,同樣求得其他所當百分比。至第五欄之百分比求法亦完全相同。

現將第三欄人數百分比取在橫軸上(方向由右至左),第五欄遺產額百分比取在縱軸上以圖示各所當值時,得羅蘭曲線如第72圖之ABC曲線。




第 72 圖

## (D)時間數列圖

將各時期或各期間所有數值順次排列後之數列。稱為時間數列，依時間數列上各數值得研究事象之變動狀態，明瞭二量間之相互關係。但欲詳細研究此等變動狀態與二量間之關係時，則以使用圖示方法為最便利。由此所得圖形，稱為時間數列圖。

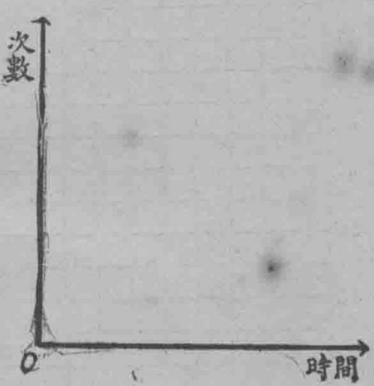
關於時間數列圖約可分為

(1)實數圖　是使用各時期或各期間所有數值之實數，以表明其實數依時間之經過所起之增減(實數之差)情形。

(2)比圖　是使用順次相鄰二時期或二期間所有數值之比值，以表明其相比之增減情形及變動狀態二種，以下依次說明。

### (1)表示實數相差之時間數列圖

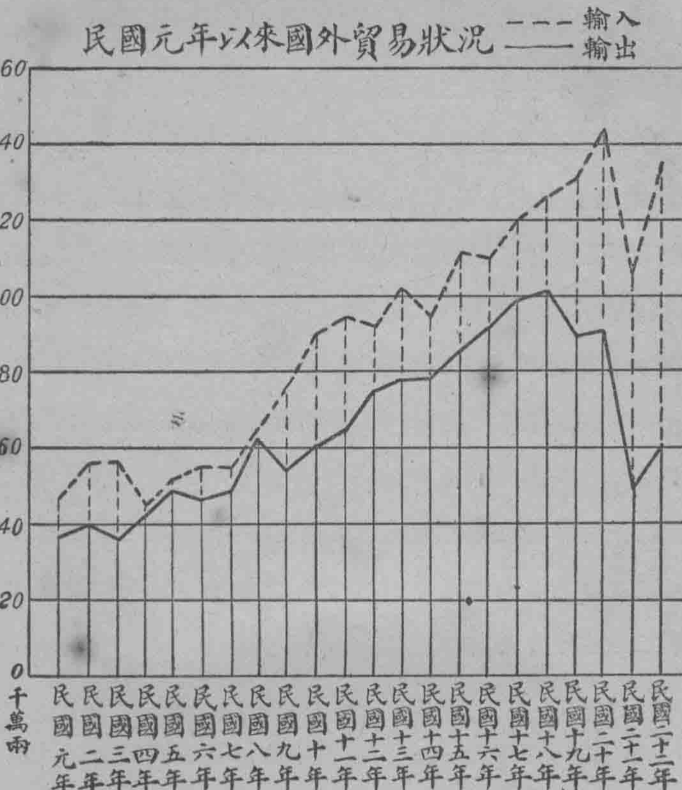
其圖示方法仍用直角座標方法，將時間取在橫軸上，各時期或各期間之數值取在縱軸上(如第73圖)，其時間之取法，不必將座標軸之交點定為零點，只依數列之表現範圍，適宜決定一時期或一期間以為此交點之相當時間。但各時期或各期間之相當點間距離，仍須相等。至縱軸上其與橫軸相交處，仍須定為零點。其縱軸上每一區劃適等於次數之幾單位一事，亦須預先決定後，再於縱軸上附以0,100,200,……等數字。



第 73 圖

在歷史系列圖上其零線之重要性與長條圖同，故其線紋亦須比其他各

橫格子線爲粗，如不引用零線時，仍須於底部引以裂紙形之曲線，以示區別。第 74 圖爲表示差之時間系列圖。其橫軸上記入各時期之年份，縱軸上記入各年輸入輸出之實數，依其各年相差程度，得知吾國歷年貿易增減情形。依圖知吾國貿易額雖逐年增加，但輸入貿易額無論何年均比輸出額爲多，歷年金錢輸出之巨大，因此得以明瞭，國人對此宜猛省之。




第 74 圖

至圖上橫軸與縱軸之尺寸，須有適當之配合，既須注意紙面之大小及使全圖一目瞭然外，尙須注意其數值之變動，不使有觀作過大或過小處，並其全期間與各時期之變動特性，亦須明示於圖形上。

關於圖示統計數值，得分動靜二態以行考究。

(甲)屬於動態統計者。

爲各年之出生數米之產額等之由各期間所累加之數值。故在圖示此等數值時，得分

1.於橫軸上表示各期間之區劃點處，立以長度與所給數值相比例之縱線，再於此等縱線頂點處順次以直線相連接後，即得所求之圖形。圖內以直線相連成之折線，只用以明瞭數值之變化狀態，並無其他特殊功用。

2. 於此等區劃點之中央處，立各期間所屬數值之相當縱線後，仍得所求折線。

此二種方法只將縱線之底點所在處，稍有不同。例如圖示各年之出生數時，第一方法是在各年之 12 月 31 日所當點處，立與該年出生數所當縱線，但在第二方法則將各縱線移至該年之 6 月 30 日所當處，其他完全相同。故在實際繪製時間數列之統計圖時，其給與之數值爲各期間之實數時，則二種方法均能適用。若將圖示之數值改爲平均數或比例數時，則以第二方法較爲妥當。

(2)屬於靜態統計者

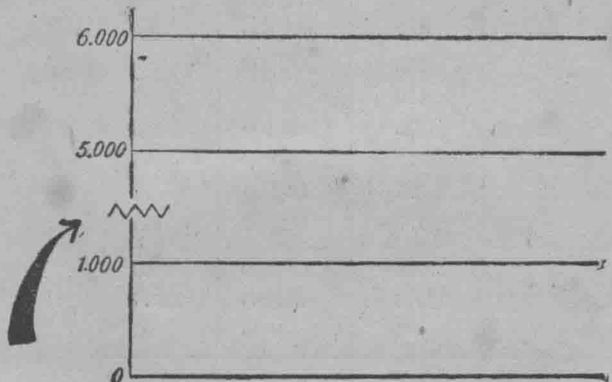
爲各年年末之人口數等數值，其圖示方法是於橫軸上與各年年末所當點處，立以長度與各數値相比例之縱線，並將各縱線之頂點順次以直線相連接後，即得所求之折線。此折線是實際表示所有數值之變動狀

態，若於橫軸上任意一點立一縱線使與折線相交，其交點之縱座標，爲給與橫軸上任意點所當時期之近似的數值。因其有此性質，故多將折線變爲曲線，以便於研究。但在自然現象之數列，例如每月氣溫氣壓等之記錄，多依各期間之平均數，以發表於世，故其縱線之取法，須與動態之第二方法相同，立在各期間之中央以圖示一切。

以上所述爲圖示一數列之方法，若欲並列二種數列以研究其間之關係時，其數值之變動能局限於比較的小範圍內，並能在一圖面上圖示時，則無甚問題。但在二數列間有巨大差異，只依上述方法尙不能在一圖面上圖示時，則須應用次之三法。

(甲)縮短二曲線間之距離使互相接近。

例如上之曲線大體爲在 5000 以上之數值，下之曲線大體爲在1000 以下之數值時，只要將下一曲線之 2000 所當橫格子線，改作爲 5000 卽得但當此時須切斷 1000 與 5000 所當二格子線間之縱座標軸，或於此切斷處只引以波浪曲線亦可。如第 75 圖內矢之所指處。



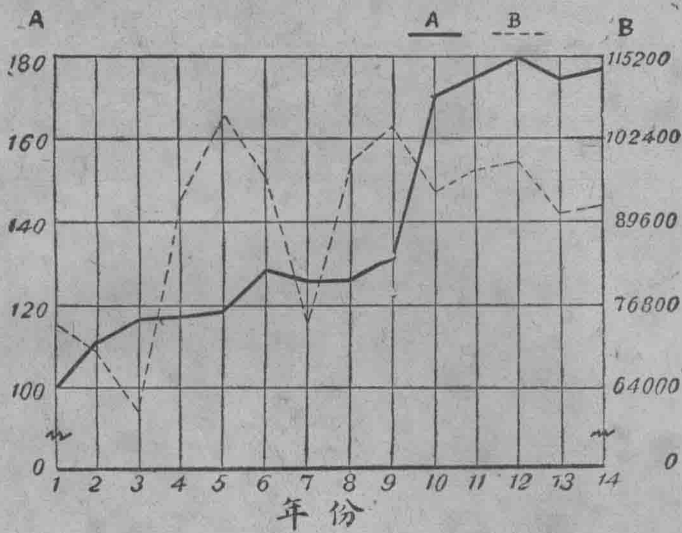

第　75　圖

(乙)用比例以圖示二相差巨大之數列。

例如第51表內AB二數列,其B之數值數百倍於A,甲之方法在此處已不能適用,故須別想方法。其方法是先求二數列內各數值之算術平均數,並將二平均數內之小者139.5去除大者89285後,得商數640。現用此商數640去除B之各數值後,得一新數列B′。此B′數列與A數列大小相似,故只將A,B′二數列圖示於紙面上,即得所求之圖形。當此時爲觀者便利起見,須將B數列各實數記入於最右之縱線上,其情形如第76圖。

第 51 表

| 年次 | A | B | B′ |
|---|---|---|---|
| 1 | 100 | 74,000 | 115 |
| 2 | 112 | 70,000 | 109 |
| 3 | 117 | 60,000 | 94 |
| 4 | 117 | 92,000 | 144 |
| 5 | 118 | 106,000 | 166 |
| 6 | 129 | 96,000 | 150 |
| 7 | 126 | 74,000 | 116 |
| 8 | 127 | 100,000 | 156 |
| 9 | 132 | 104,000 | 163 |
| 10 | 170 | 94,000 | 147 |
| 11 | 175 | 98,000 | 153 |
| 12 | 179 | 99,000 | 155 |
| 13 | 175 | 91,000 | 142 |
| 14 | 177 | 92,000 | 144 |
| 計 | 1,954 | 1,250,000 | 1,954 |
| 平均 | 139.5 | 89.285 | 139.5 |
| 比 | 1 | 640 | 1 |




第 76 圖

本法對於考察大數值之數列上，雖有不便處，但在觀察其由平均數之偏差以比較二數列時，其方法未始不可。

(丙)用指數以圖示二相差巨大之數列。

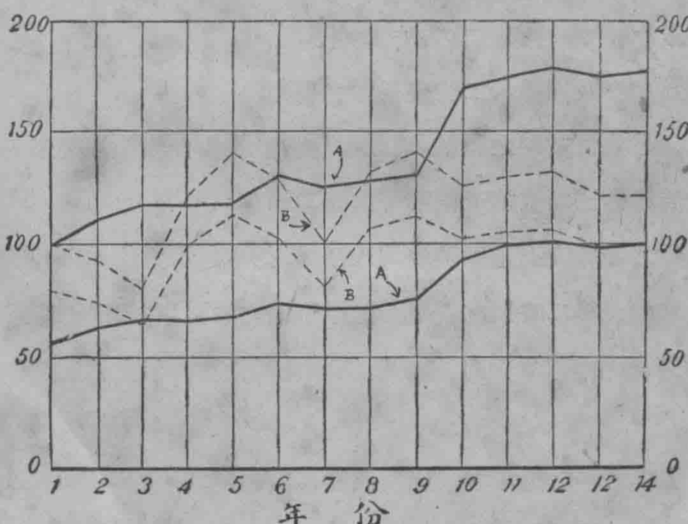
本法是將二數列之數值，均改換爲指數後由此指數以圖示一切。至指數之計算方法參照第五編第一章第二節。

指數之計算依上述知得分爲三種，依此本圖示方法亦得分爲三種。現用最初與最後二時期之數值爲標準時期之數值以計算第 51 表內 A，B 二數列後，得第 52 表內各數值。現將此等數值圖示之得第 77 圖內之四折線。

本法在考察數值之增減比較上實甚便利，其依同一比例以增減者能在縱線上依同一變化以表示在之。其 A，B 二線在最初或最後之時期

## 第 52 表

| 年份 | 以最初之數值爲100時之指數 A | 以最初之數值爲100時之指數 B | 以最後之數值爲100時之指數 A | 以最後之數值爲100時之指數 B |
|---|---|---|---|---|
| 1 | 100 | 100 | 56 | 80 |
| 2 | 112 | 95 | 63 | 76 |
| 3 | 117 | 81 | 66 | 65 |
| 4 | 117 | 124 | 66 | 100 |
| 5 | 118 | 143 | 67 | 115 |
| 6 | 129 | 130 | 73 | 104 |
| 7 | 126 | 100 | 71 | 80 |
| 8 | 127 | 135 | 72 | 109 |
| 9 | 132 | 142 | 57 | 113 |
| 10 | 170 | 127 | 96 | 102 |
| 11 | 175 | 132 | 99 | 106 |
| 12 | 179 | 134 | 101 | 108 |
| 13 | 175 | 123 | 99 | 99 |
| 14 | 177 | 124 | 100 | 100 |



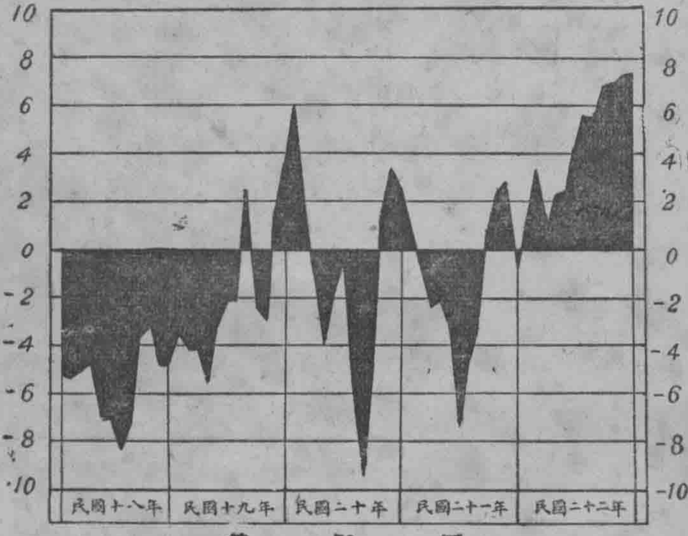

第 77 圖

成爲互相一致處，雖爲本法之缺點。但依此能將二數列互相接近，並能由同一點出發以觀察二數列之變動狀況，實能與其缺點互相對消。又其圖上各數值是爲相對值而非絕對值未免有所不妥，但能記入原有數值於圖傍時，卽能補救其缺點。

此外關於相差時間數列圖示上，尙有

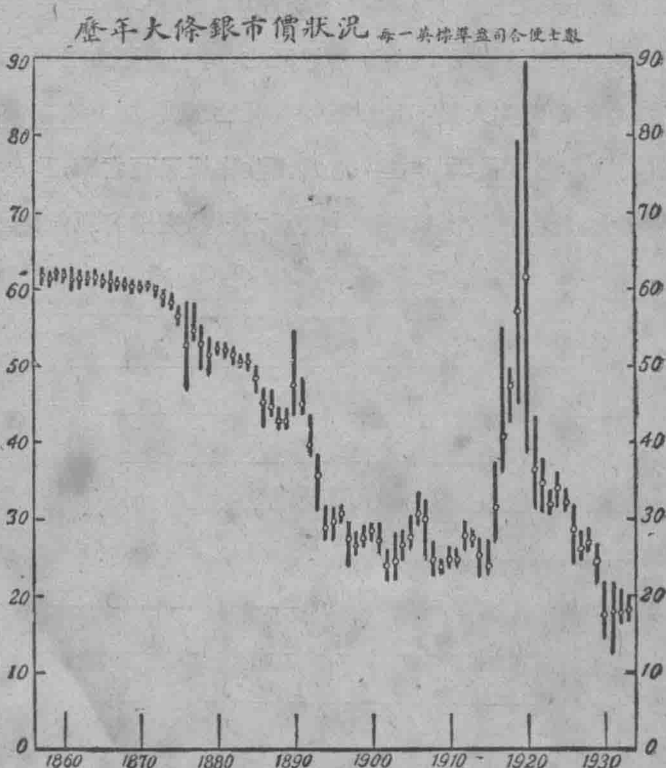
a. 陰陽時間數列圖 (silhouette excess and deficit chart) 用以表示超過與不足之情形。第 78 圖爲南京市近數年來人口自然增加率情形，其在零線上方者表示出生率超過死亡率，零線下方者表示死亡率超過出生率，卽爲出生人數不足以塡補其死亡者。依圖知南京市在前數年是正負互見，從民國 22 年起一直爲出生率超過死亡率，且愈後其自然增加率愈大，此亦由於南京市衞生設備之日趨完備有以致之也。

南京市人口自然增加率最近狀況




第　78　圖

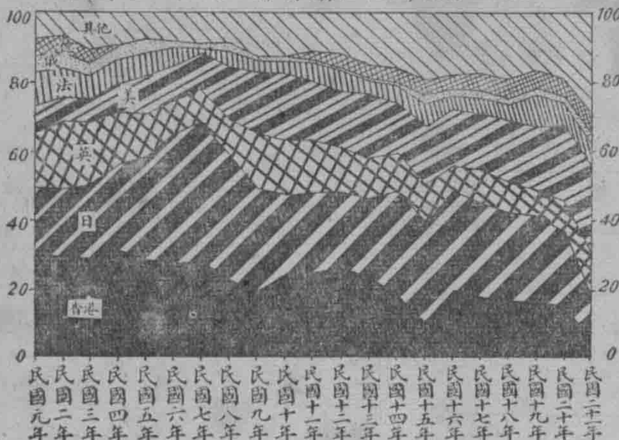
b. 距限時間數列圖 (zone chart) 將各時期之最大數最小數及其平均數，同時表現在一幅圖面上。第 79 圖爲歷年銀價變動情形，圖內各小圓圈表示各年平均銀價，最高點之所當值爲該年之最大銀價，最低點所當值爲該年最低銀價，其最高點與最低間之距離，是表示該年銀價變動範圍。依圖知 1920 年左右，其銀價爲最高，其變動情形亦最急烈，以後銀價逐次減低，而以 1931 年爲銀價最低的一年。




第 79 圖 距限統計圖

c. 帶形時間數列圖 (band chart) 用以表示歷年總額之內部構成狀況。第 80 圖爲民國元年以來進口洋貨價值按國百分比較情形，依圖知歷年輸入額中以日本占最大部分，但從 1930 年以後，則爲美國所追出。

民元以來進口洋貨價值按國百分比較

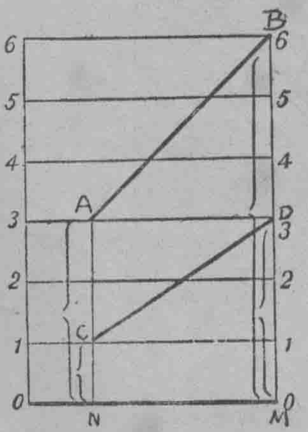



第 80 圖 帶形統計圖

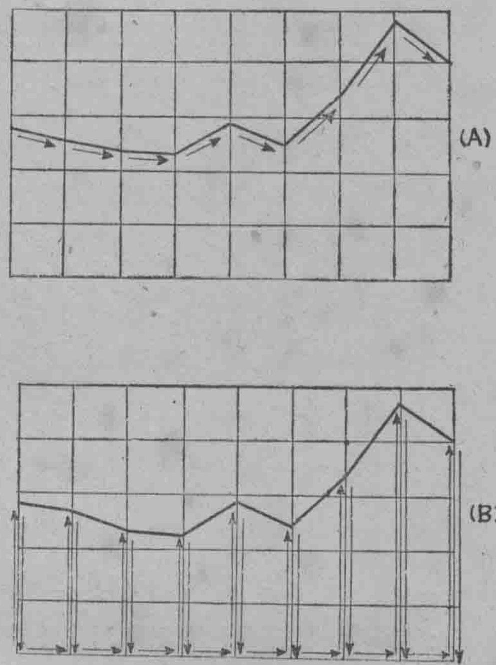
(2)表示實數相比之時間數列圖

上述圖示方法是用以表示數列內各數值相差之狀況，但只此絕對差數，有時難以判明數值變動之實情，並易於引起誤解。蓋因在普通方格子紙上所繪成之歷史數列圖內，爲吾人所最注意處，是線之傾斜方向與角度，其傾斜急激時知其增減衆多，傾斜緩慢時知其增減稀少。但只注意線之傾斜程度，而不注意各點與零線之距離時，實難以比較二線之增減速度。例如第 81 圖其 NM 爲零線，AB 與 CD 爲欲比較二曲線之一部分，其 CD 之傾斜程度雖比 AB 爲緩，但與零線相對照時，知 BM

與 AN 之比只爲 2:1,而 DM 與 C N 之比反爲 3:1,故知 CD 之增加率實比 BM 爲大。依此知觀察普通方格子紙上所繪成之圖形時,其目光之轉移途徑不能依第 82 圖 A 內所示矢之方向,須依 B 內矢之方向——與零線對照後,順次移動其目光。




第 81 圖




第 82 圖 對於普通圖表上曲線之觀察方法

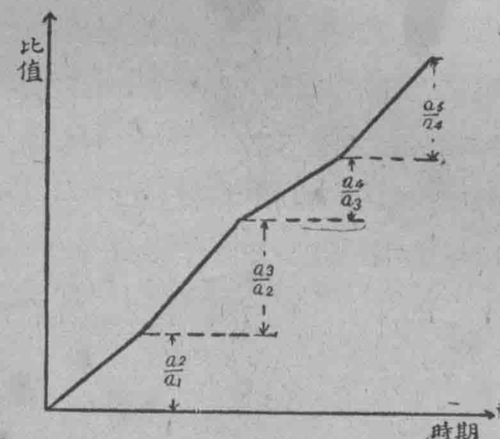
依上述知由普通方格子紙所作成之歷史數列圖以觀察各期數值之增減率,實難得正確之結果,故必須別想方法,使能正確地依增減率之狀況,以判斷數值間之關係。爲副此目的起見所發見各方法中之最簡單者,爲先計算順次二期數值之比,而後將此比值圖示於方格子紙上一法。例如依等期間相隔各時期之數值爲

$$y_1,\quad y_2,\quad y_3,\quad y_4,\cdots\cdots\cdots\cdots,$$

由此等數值計算得順次二期數值之比為

$$\frac{y_1}{y_1},\quad \frac{y_2}{y_1},\quad \frac{y_3}{y_2},\quad \frac{y_4}{y_3},\cdots\cdots\cdots\cdots,$$

現在取橫軸在縱軸之 1 處並在橫軸上決定各時期之相當點後，再依各期之比值順次決定各相當點於圖面上，並用直線連接之，即得所求之圖形。但此等比值並非表示與基線相隔之距離，乃為與通過即前比值所當點並與橫軸相平行一直線間相隔之距離，其情形如第 83 圖，由




第　83　圖

此知欲直接依基線以定各點時，須將各比值順次相加使成為

$$\frac{y_2}{y_1},\quad \frac{y_2}{y_1}+\frac{y_3}{y_2},\quad \frac{y_2}{y_1}+\frac{y_3}{y_2}+\frac{y_4}{y_3},\cdots\cdots\cdots\cdots$$

後，始能由此等數值直接依基線作成與第二，第三，第四，……等時期相對應各點，因其須如此作法，未免繁雜，故特加以修改如次：

將時間數列之全體看做為一等比級數以行考慮。但在此時若仍依前法以圖示各數值，則仍遭同等困難。故須於此等等比級數各數值上，施以對數後，將等比級數一變而為等差級數，即將給與之等比級數

$$y_1, \quad y_2 = y_1 r, \quad y_3 = y_1 r^2, \cdots\cdots\cdots\cdots$$

變爲

$$\log y_1, \ \log y_2 = \log y_1 + \log r, \ \log y_3 = \log y_1 + 2\log r,$$

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

等依 $\log r$ 爲公差之等差級數,將此等對數値圖示於方格子紙上時,易得一直線圖形。現將此原理應用於歷史數列圖上,先依各數値之對數求得各適應點於圖面上後,依連接此等點所作成之曲線形狀得觀察所給與數列依時間之經過所起增減狀況。將本圖形與比之絕對値所作成之圖形相比較時,知本圖形內橫軸上 1,2,3……等時期所當之 $\log y_1$,$\log y_2$,$\log y_3$……等縱線長之順次相差,依對數之性質知爲各數値比之對數,即

$$\log y_2 - \log y_1 = \log \frac{y_2}{y_1}$$

$$\log y_3 - \log y_2 = \log \frac{y_3}{y_2}$$

$$\log y_4 - \log y_3 = \log \frac{y_4}{y_3}$$

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

依此知本圖形是表示比之對數値,雖依對數以表現比値,但其目的仍爲給與相比値起見,故仍得稱爲比圖(ratio chart)。或因其爲圖示所給與數列上各數値之對數,故又稱之爲對數圖(logarithmic chart)。上述對數圖只於各期所有數値上使用對數,其橫軸上各時期仍用普通數値,故與橫軸上亦用對數値之圖形易於區別起見,特稱本圖形爲半對數圖(semi-logarithmic chart)。

至半對數圖形之作成方法，若仍用普通方格子紙時，須將依對數表所查得各時期所當數值之對數值，一一圖示於紙面上。其一一由對數表以搜求各數值所當對數值一事，旣甚煩瑣，且將其所求得之對數值一一圖示於紙面上一事，又甚麻繁。故爲簡單起見，特預先於縱軸上施以對數值之格子，而後只依數列之實數，以求得各點可也。依此原理所作成之方格子紙，稱爲半對數方格子紙(semi-log section paper)。

半對數圖形均在此半對數方格子紙上作成一切，故對於半對數方格子紙之作成方法說明如次：

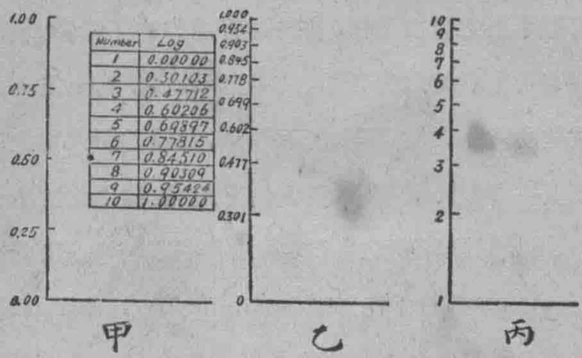
先由對數表求得 1 至 10 各自然數之對數值爲

$$\log 1=0, \qquad \log 2=0.30103,$$
$$\log 3=0.47712, \qquad \log 4=0.60206,$$
$$\log 5=0.69897, \qquad \log 6=0.77816,$$
$$\log 7=0.84510, \qquad \log 8=0.90309,$$
$$\log 9=0.95429, \qquad \log 10=1。$$

現於縱軸上取一定之長度，其下端定爲 0，上端定爲 1，並適宜等分其中間如第 84 圖(甲)之狀況。次將 1 至 10 各自然數之對數值記入於上記縱軸上，其情形如圖內(乙)。最後將記入對數值處記入 1 至 10 之自然數，如圖內(丙)之狀況後，卽得所求半對數圖形之縱軸。

由 1 至 10 各整數之記入方法已如上述，其1·1，1·2，……等有小數之數值記入方法，仍與上述方法相似，先由對數表求得各值所當對數值後，依上述方法逐次進行卽得。

將 1 至 10 之對數記入妥定後，其 10 至 100 之對數記入方法，是




第 84 圖

將1至10各數值所作成之縱軸狀況，完全不改變地移接於1至10各數值所作成縱軸之10的所當點上，並將1,2,3……等數值改爲10,20,30……等數字後卽得。至由100到1000及由1000到10000等之記入情形，完全相同。參照附添之半對數方格子紙卽能明白了解。

半對數方格子紙旣已作成後，其半對數圖形之作成方法，實與作成次數分配圖相同。所不同者只在作成次數分配圖時，是使用普通之方格子紙，但此處是使用半對數方格子紙而已。又半對數方格子紙之縱軸上尺寸，旣已變爲對數，故當所欲決定某時期數值之相當點，在格子線間時，須將格子線間之數值仍須依對數尺寸之距離，以定其點之所在。

半對數圖形之作成方法已如上述，其圖形之特徵有次之數種：

a、半對數圖形上，二數值之在縱軸上所記數字間之距離差，不爲二值之實數差，而爲二值變化之比例。依此將由100增加至150，與150增加至300之情形，在普通圖形上，其縱軸上所記入二數值間距離差，實有巨大之差異，但在半對數圖形上則成爲完全相等。參照第85圖(A)內矢之所指處，卽能明白了解。其原因實由於對數上有

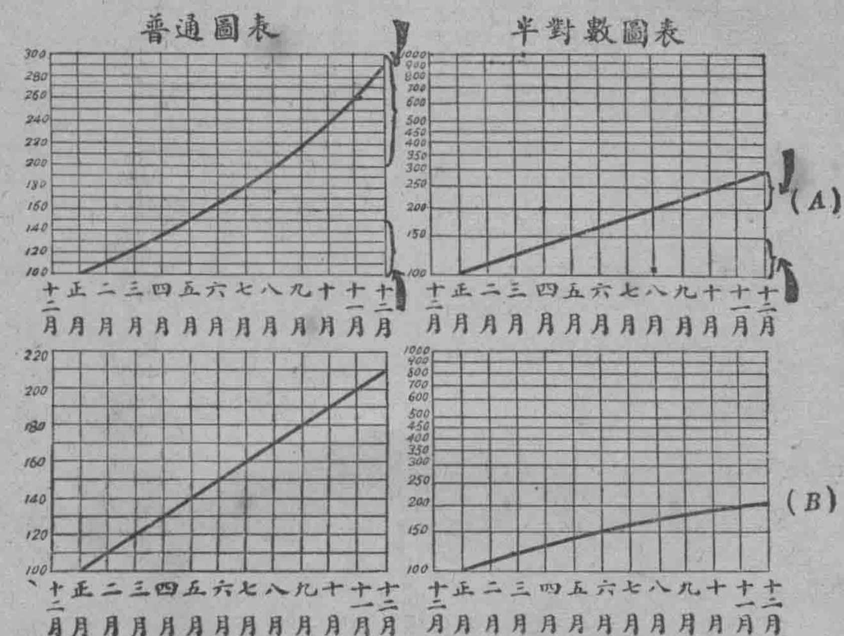
$$\log A-\log B=\log\frac{A}{B}$$

之性質，故成爲

$$\log 300-\log 200=\log\frac{300}{200}=\log 1.5,$$

$$\log 150-\log 100=\log\frac{150}{100}=\log 1.5,$$

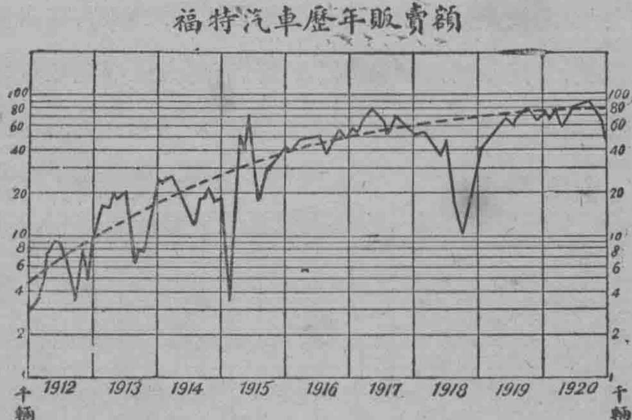
其距離完全相等。依此知實數依等比級數以增加之數列，在對數表列上成爲等差級數，其圖示情形如第 85(A)圖。若實數依等差級數以增加之數列，在對數數列之增加率反逐次減少，其圖示情形如第 85 圖(B)。




第　85　圖

2. 在半對數圖形上，完全無零線，其基線常用 1 以表示之。故在數列內各數值均爲巨大時，得將各數值依同比例以行縮小，或將基線定在相當巨大數值相當處，使數列全體移至基線之附近，以便於製作。

3. 能將前後相差巨大之數列，在同一紙面上以行製作。例如第 86 圖爲美國福特公司每年汽車販賣額，其間雖有三千輛與十萬輛之巨大差異，但因其使用半對數圖示方法，故仍能在一幅紙上畫出其全部情形。




第 86 圖

4. 能在圖形上施行乘除之計算。

半對數圖之特徵已如上述，其作成之曲線性質一若難以理解然，但能將次之諸規則完全理解後，亦不覺困難。

(1)線與横軸平行時，(如第 87 圖內之 A 線) 是表示各數值全無增減。

(2)依直線形狀上昇時(圖內 B 線)，是表示各數值依一定比率以增加一切。

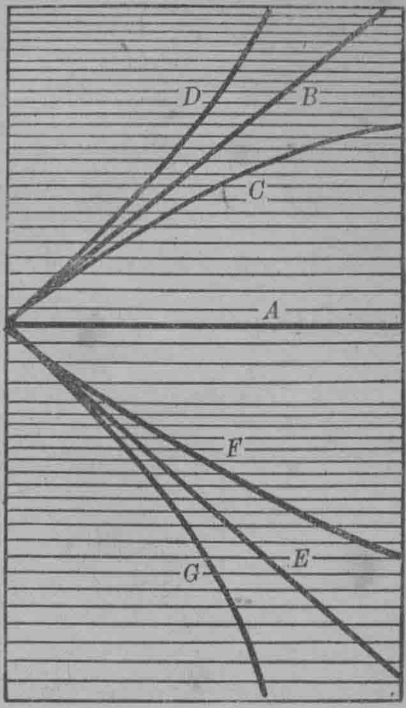
(3)雖爲上向但帶有下曲之圓形時(圖內C線)，表示各數值依量之增加而逐次減少其增加率。

(4) 既爲上向並又帶有上曲圓形時，(圖內D線)，是表示各數值依量之增加而逐次擴大其增加率。

(5) 依直線形狀以下向時，(圖內E線)，是表示各數值依一定比率以減少一切。

(6) 雖爲下向但帶有上曲圓形時(圖內F線)，是表示各數值之減少率依量之減少而逐次縮小。

(7) 既爲下向並又帶有下曲圓形時(圖內G線)，是表示各數值之減少率依量之減少而逐次擴大。




第 87 圖

半對數圖表上各線紋之狀況

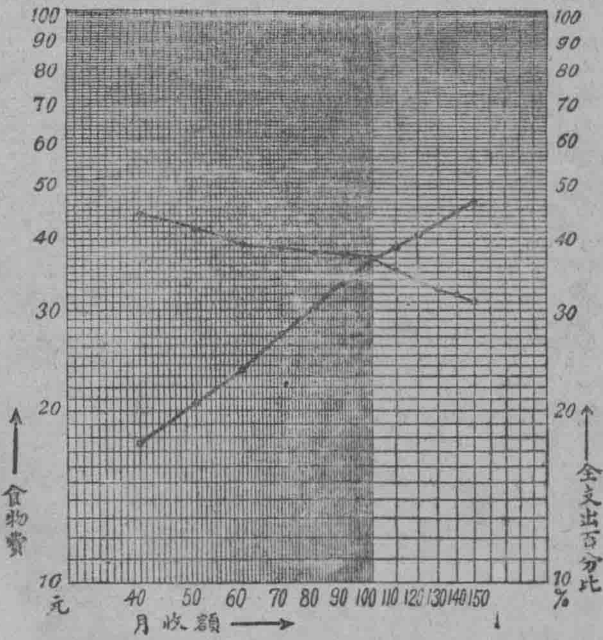
(8)當二線成爲互相平行時，其增加(減少)率實相等。

(9)二線中之傾斜度高大者之增加(減少)率比他者爲大。

能將上述9項規則熟記後，半對數圖表實甚易於明白了解。

至縱橫二軸均用對數之對數圖，是用以表示某變數依一定比例以行變化時，其對應變數所起變化比率之狀況。例如第88圖爲某市勞動者之每月收入與食費支出之分布情形。其縱橫二軸均使用對數值，圖內

之粗線是表示食費實數增減比例與收入增減比例間之關係情形，細線是表示食費對於全支出百分比之增減比例與收入增減比例間之關係情形，若收入二倍時其食費之支出亦大二倍，收入三倍時支出亦爲三倍等依正確比率以行變化時，則表示食費實數之粗線須與橫軸成爲 45°，現該粗線與橫軸所成角度比 45°爲小，故知食費之增加比率比收入之增加比率爲少。又其食費對於全支出之百分比，若爲一定不變時，則表示此比例數之細線須與橫軸相平行，現該線是爲下向趨勢，故知食費對全支出之百分比，依收入之增加而逐次減少。




第 88 圖

兩軸均用對數之圖形在經濟統計上時常使用，例如表示收入分配之派來脫線(Pareto's line)爲其中之最著者。

派來脫線爲伊之經濟學家派來脫(Vilfreds Pareto)所發見，用以研究收入之分配。(Pareto Cours d'economie Politique. Tome II live III Chap. I 1897)。

其化學公式爲

$$y=\frac{A}{x^{\alpha}},$$

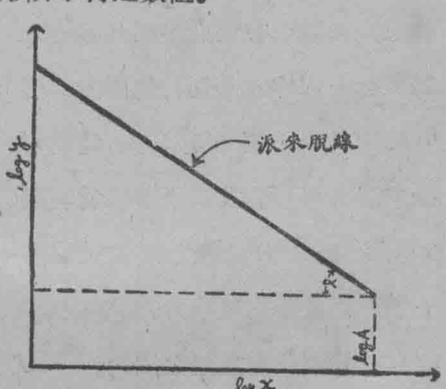
或 $$\log y=\log A-\alpha\log x,$$

式內之 x =一定收入額，

y=占有 x 額以上收入之人員數，

A, $\alpha$= 由各社會實際情形所求得之數值。

其圖示情形如第 89 圖，但在實際計算時其圖示之派來脫線多不爲純然之直線，而爲近似的直線。(詳細情形請參照上述派來脫著書及下列諸書)：




第 89 圖

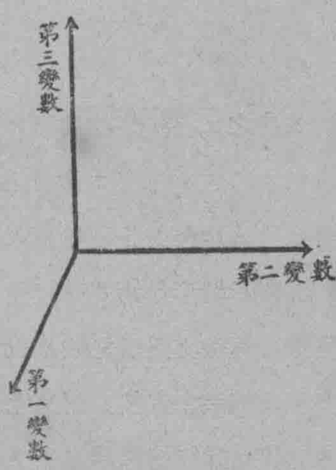
Mitchell; Income in United States. Vol. II., Chap. 28.

Pigou; Economics of Welfare, Pt. V, Chap II, "Perato's Law"

Bowley; The Measurment of Social Phnomena P. 106.

### (E)三元座標統計圖

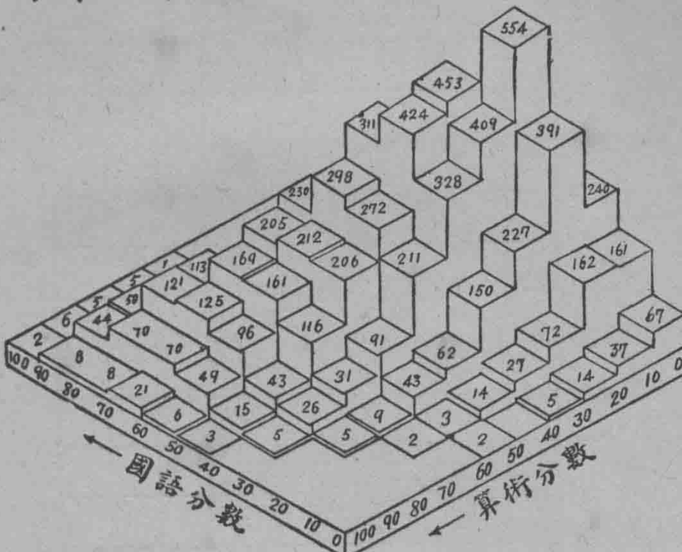
上述各種座標統計圖，其變數均只有二個，若遇有三個變數，例如在夫年30歲妻年20歲之夫婦有100對，夫年25歲妻年21歲之夫婦有80對……等之材料上，其夫妻之年齡均已各爲變數，而夫婦對數又成爲一變數，如此共有三個變數，已不能應用上述各法以舉行圖示。當此時須在含有第一變數與第二變數所當軸之平面上，再立一垂直軸在此軸上，取其長度以表示第三變數所當量而後始可。其軸之取配情形如第90圖。由此方法所繪得之統計圖形，已非曲線圖形而成爲紙上排列長短各針之圖形，如第91圖爲英國936名小學女生之國語算術試驗成績，第二門分數取在平面之縱橫二軸上，人數取在與此平面相垂直之第三軸上以作成一切。依圖知二門功課均爲100分者有2人，均爲零分者則有67人，其間最多者爲國語30分，算術10分左右之學生。又圖內之積木偏重於右邊後方，知學生成績內算術成績較國語爲劣，知雖在英國亦有算術難於國語之傾向。



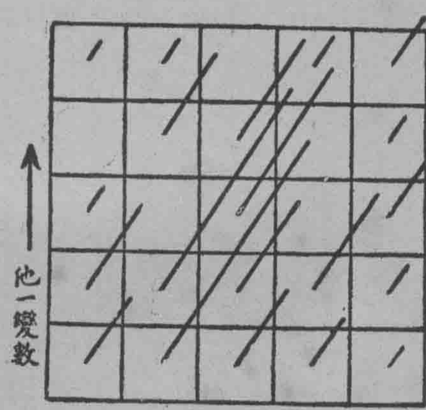

第 90 圖

在此種圖形上，其陰面各數值完全不能表現出來，故爲改正此等缺點起見，主張採用次之方法。將平面上縱橫二軸取爲二個變數所當軸外，其發現次數不取在與平面相垂直之第三軸上而仍取在同一平面上，但

某年英國小學女生國語算術試驗成績




第 91 圖　立體圖形

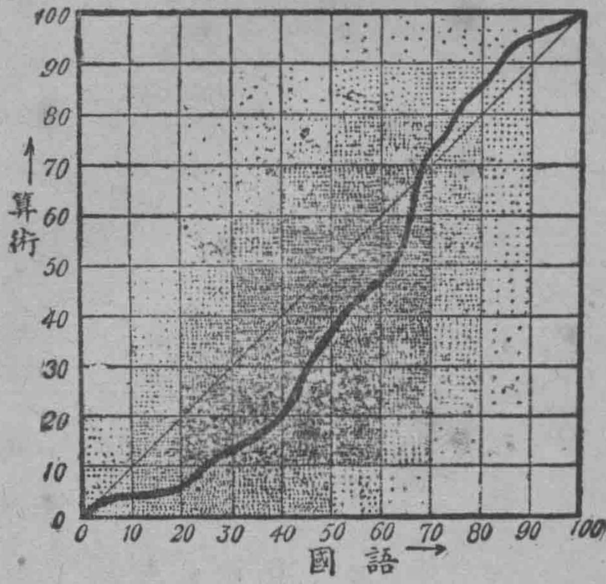



第 92 圖

與縱橫軸稍偏之直線上，其圖示情形如第 92 圖。依圖知各種數值均能在平面圖上以表現一切。

此外尚有將發現次數用點以記入於縱橫二軸所隔成之方格子內。如第 93 圖是將第 91 圖內英國女生考試成績，依現述方法所作成者。此等圖形在

議論相關上非常適用，待以後在相關論 (theory of correlation) 內再行詳述。




第 93 圖　相關圖形

# 第七章　平均數

## 第一節　平均數之意義

統計表之製作方法已如上述，知依一定事項以區分一切，對此各區分組別內所屬數目，實各有多少交互之不同，故多數特依移動方法改變爲由少漸次增加以至於極點，而後再由極點漸次減少之次數分配表，使便於比較。例如第 53 表爲某校入學考試成績統計表，依表知其分布狀

第　53　表

| 分數 | 人數 | 分數 | 人數 | 分數 | 人數 | 分數 | 人數 |
|---|---|---|---|---|---|---|---|
| 30 | 1 | 43 | — | 56 | 7 | 69 | 23 |
| 31 | — | 44 | — | 57 | 11 | 70 | 95 |
| 32 | — | 45 | 7 | 58 | 16 | 71 | 6 |
| 33 | — | 46 | — | 59 | 16 | 72 | 52 |
| 34 | — | 47 | 2 | 60 | 55 | 73 | 50 |
| 35 | 1 | 48 | 2 | 61 | 2 | 74 | 17 |
| 36 | — | 49 | 1 | 62 | 13 | 75 | 63 |
| 37 | — | 50 | 13 | 63 | 37 | 76 | 23 |
| 38 | — | 51 | 1 | 64 | 14 | 77 | 11 |
| 39 | — | 52 | 1 | 65 | 68 | 78 | 5 |
| 40 | 6 | 53 | 1 | 66 | 31 | 79 | 1 |
| 41 | — | 54 | 1 | 67 | 43 | 計 | 773 |
| 42 | — | 55 | 27 | 68 | 50 | | |

態，既不規則又甚複雜，在觀察成績分配狀況時，上表非常不便。故特在上表內取幾分爲一組，將全部分數分爲若干組，如第 54 表(a)(b)二表

是依四分組別，第55表(a)(b)(c)三表是依五分組別，以區分一切。在此五表內，以何種爲最適宜，則須擇其服從次之二條件者：

第 54 表 (a)

| 分數 | 組之中心 | 人數 | 分數 | 組之中心 | 人數 |
|---|---|---|---|---|---|
| 30——34 | 32 | 1 | 58——62 | 60 | 89 |
| 34——38 | 36 | 1 | 62——66 | 64 | 132 |
| 38——42 | 40 | 6 | 66——70 | 68 | 147 |
| 42——46 | 44 | 7 | 70——74 | 72 | 203 |
| 46——50 | 48 | 5 | 74——78 | 76 | 114 |
| 50——54 | 52 | 16 | 78——82 | 80 | 9 |
| 54——58 | 56 | 46 | 計 | | 773 |

第 54 表 (b)

| 分數 | 組之中心 | 人數 | 分數 | 組之中心 | 人數 |
|---|---|---|---|---|---|
| 29——33 | 31 | 1 | 57——61 | 59 | 98 |
| 33——37 | 35 | 1 | 61——65 | 63 | 66 |
| 37——41 | 39 | 6 | 65——69 | 67 | 192 |
| 41——45 | 43 | – | 69——73 | 71 | 176 |
| 45——49 | 47 | 11 | 73——77 | 75 | 153 |
| 49——53 | 51 | 16 | 77——81 | 79 | 17 |
| 53——57 | 55 | 36 | 計 | | 773 |

第 55 表 (a)

| 分數 | 組之中心 | 人數 | 分數 | 組之中心 | 人數 |
|---|---|---|---|---|---|
| 26——31 | 28.5 | 1 | 56——61 | 58.5 | 105 |
| 31——36 | 33.5 | 1 | 61——66 | 63.5 | 134 |
| 36——41 | 38.5 | 6 | 66——71 | 68.5 | 242 |
| 41——46 | 43.5 | 7 | 71——76 | 73.5 | 188 |
| 46——51 | 48.5 | 13 | 76——81 | 78.5 | 40 |
| 51——56 | 53.5 | 31 | 計 | | 773 |

第 55 表 (b)

| 分數 | 組之中心 | 人數 | 分數 | 組之中心 | 人數 |
|---|---|---|---|---|---|
| 27.5—32.5 | 30 | 1 | 57.5—62.5 | 60 | 102 |
| 32.5—37.5 | 35 | 1 | 62.5—67.5 | 65 | 193 |
| 37.5—42.5 | 40 | 1 | 67.5—72.5 | 70 | 226 |
| 42.5—47.5 | 45 | 9 | 72.5—77.5 | 75 | 164 |
| 47.5—52.5 | 50 | 17 | 77.5—82.5 | 80 | 6 |
| 52.5—57.5 | 55 | 47 | 計 | | 773 |

第 55 表 (c)

| 分數 | 組之中心 | 人數 | 分數 | 組之中心 | 人數 |
|---|---|---|---|---|---|
| | | | 54.5—59.5 | 57 | 77 |
| 29.5—34.5 | 32 | 1 | 59.5—64.5 | 62 | 121 |
| 34.5—39.5 | 37 | 1 | 64.5—69.5 | 67 | 215 |
| 39.5—44.5 | 42 | 6 | 69.5—74.5 | 72 | 220 |
| 44.5—49.5 | 47 | 12 | 74.5—79.5 | 77 | 103 |
| 49.5—54.5 | 52 | 17 | 計 | | 773 |

(1)取其次數分配狀況之最依規則順序,依次排列者。

(2)取各組之中心數值是以 0 或 5 之數字爲終者,因能如此取法,則以後計算上甚爲便利。

在上述五表內以第 55 表(b)爲最適合,故知該表爲最適宜之次數分配表。至各表內各組相連處有同樣數字各二處,其爲組之下限者是表示未滿之意義。例如第 54 表(a)內各組,是爲 30 分到 34 分未滿,34 分到 38 分未滿,……。

至次數分配表內各組範圍之決定方法依組之品質而不同。在連續性之品質上,則將適爲組限數值之次數平均分配於所當二組上。其所以如此者,蓋因長度重量等連續性之品質,當測定時,均不免含有誤差。例

如森林學者 Vater 在森林實驗上測量九歲松樹凡 125 株之結果，如第 56 表，將本表加以整理得次數分配表如第 57 表，在本表內高度 185 糎之松樹實際未必一定爲 185 糎，或多或少亦說不定。故不能將該樹全部屬於 185－205 之組內，須將其平均分配於雙方組內。

第 56 表

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

第　57　表

| 高度 | 組之中心 | 樹數 | 高度 | 組之中心 | 樹數 |
|---|---|---|---|---|---|
| 45——65 | 55 | 1 | 185——205 | 195 | 26.5 |
| 65——85 | 75 | 1 | 205——225 | 215 | 13 |
| 85——105 | 95 | 1 | 225——245 | 235 | 5 |
| 105——125 | 115 | 6 | 245——265 | 255 | 4 |
| 125——145 | 135 | 11 | 265——285 | 275 | 1 |
| 145——165 | 155 | 21 | | | |
| 165——185 | 175 | 34.5 | 計 | | 125 |

反此在不連續性之品質上，如人數豆數等之數值，是非常確定，凡測得爲 5 者，決不致變爲 4 或 6。因此其次數分配情形，不必如連續性然，只依普通方法如第 54 表與第 55 表之分配情形卽可。

在次數分配表內其「某某——某某」之各組數值，有時得不必明白表現出來，只書其中心數値亦可。

次數分配表之作成方法已如上述，其代表形式如第 58 表。但祗此

第　58　表

| 組別 | 組之中心 X | 次數 f |
|---|---|---|
| $L_1$——$u_1$ | $X_1$ | $f_1$ |
| $L_2$——$u_2$ | $X_2$ | $f_2$ |
| ⋮ | ⋮ | ⋮ |
| $L_m$——$u_m$ | $X_m$ | $f_m$ |
| 計 | | $N=\Sigma f$ |

次數分配表，有時又難以比較其相互間之狀況。例如欲解決壯丁身長是否每年增加，國民平均壽命是否年年延長等問題，若只依各身長別壯丁次數分配表，及年齡別死亡表以比較時，實仍不能解決。必須先就各年

之次數分配表，求出其代表或中心物，而後始能互相比較。本章所述平均數，實用以表示此代表物或中心物也。

平均數一般可分為次之五項：

(1)算術平均數(Arithmetic mean, Das arithemetiche Mittel),

(2)中位數(Median, Zentralwert, Meadianwert),

(3)最衆數(Mode, Der häufigste(dichteste) wert),

(4)幾何平均數(Geometric mean, Das geometriche Mittel),

(5)調和平均數(Harmonic mean, Das harmonische Mittel)。

此等平均數內所應具備條件，依優爾(Yule: An introduction to the theory of statistics 1924. P. 108)之主張，須有次之六種。

(1)須為確定數值，使觀察者難輸入主觀數字。

(2)須關係於全部觀察上，不然不足看做為總體之代表物。

(3)須有簡易之性質，不可有太抽象之數學性質。

(4)須計算簡單。

(5)對於樣本部分選定之影響，須非常稀少。

(6)能代數的運用。

與此等條件最適合者為算術平均數，以後將各平均數之公式與來源及其應用情形依次詳述。

## 第二節　算術平均數

### (A) 實用方法

將各觀察單位值相加後，以單位個數除之即得各單位值之算術平

均數。其數式如次：

$$X_1, X_2, \cdots\cdots\cdots\cdots X_n$$

爲 n 個單位值，M 爲算術平均數時，

$$M=\frac{X_1+X_2+\cdots\cdots+X_n}{n}$$

簡寫爲　(公式 6)

$$M=\frac{1}{n}\Sigma X$$

〔例〕有身長五尺二寸五尺三寸五尺四寸五尺五寸五尺六寸等五人，求其平均身長爲幾何？

依題得

$X_1=52$，　$\Sigma X=52+53+54+55+56$

$X_2=53$，　$=270$ 寸

$X_3=54$，

$X_4=55$，　$\because M=\frac{1}{n}\Sigma X=\frac{270}{5}=54$ 寸

$X_5=56$，

在普通統計表上，每遇同值之單位甚衆。現假定

| | |
|---|---|
| 單　位　值 | $X_1, X_2, X_3 \cdots\cdots\cdots X_m$ |
| 次　　數 | $f_1, f_2, f_3, \cdots\cdots\cdots f_m$ |

時，其求算術平均數之公式變爲

$$M=\frac{f_1X_1+f_2X_2+\cdots\cdots+f_mX_m}{f_1+f_2+\cdots\cdots+f_m}$$

簡寫爲　(公式 7)

$$M=\frac{\Sigma fX}{\Sigma f}$$

〔例〕依第59表之身長次數分配表以求平均身長。依表知吾人所需

第 59 表

| 身 長 | 人 數 |
|---|---|
| 49——51 | 19 |
| 51——53 | 90 |
| 53——55 | 195 |
| 55——57 | 141 |
| 57——59 | 20 |
| 計 | 465 |

要之單位值爲各組之中位數，如4尺9寸——5尺1寸之中位數爲5尺，故知 $X_1=50, f_1=19$，餘類推。由此得

$$X_1=50,\quad f_1=19,\quad f_1X_1=19\times50=950,$$

$$X_2=52,\quad f_2=90,\quad f_2X_2=90\times52=4680,$$

$$X_3=54,\quad f_3=195,\quad f_3X_3=195\times54=10530,$$

$$X_4=56,\quad f_4=141,\quad f_4X_4=141\times56=7896,$$

$$X_5=58,\quad f_5=20,\quad f_5X_5=20\times58=1160,$$

依公式得平均身長爲

$$M=\frac{\Sigma fX}{\Sigma f}=\frac{950+4680+10530+7896+1160}{19+90+195+141+20}$$

$$=\frac{25216}{465}=54.23\text{ 寸弱。}$$

上之計算方法未免太繁，故統計界多使用第60表所定格式之計算方法。

第 60 表

| 身長之組 | 組之中心 X | 人數 f | 積 f.X |
|---|---|---|---|
| 尺寸 尺寸 | | | |
| 49——51 | 50 | 19 | 950 |
| 51——53 | 52 | 90 | 4680 |
| 53——55 | 54 | 195 | 10530 |
| 55——57 | 56 | 141 | 7896 |
| 57——59 | 58 | 20 | 1160 |
| 計 | | 465 | 25216 |
| $M=\frac{\Sigma fx}{\Sigma f}=\frac{25216}{465}=54.23$ 寸 | | | |

若各組所當數值及所屬次數衆多時,上之計算實甚麻煩,故多施用特殊方法,使計算趨於簡便,現將其簡便計算之原理及應用方法說明如次。

取A爲各單位值中一任意值(Arbitrary value),與單位值 X 相差數爲 $x'$ 此 $x'$ 稱爲 X 之由 A 所測得之偏差(Deviation from A)其數式如次。

$$\begin{array}{lcl} X_1-A=x'_1 & & X_1=A+X'_1, \\ X_2-A=x'_2, & 即 & X_2=A+x'_2, \\ \cdots\cdots & & \cdots\cdots \\ X_n-A=x'_n, & & X_n=A+x'_n, \end{array}$$

$$\begin{aligned} \Sigma X &= (A+x'_1)+(A+x'_2)+\cdots\cdots+(A+x'_n) \\ &= nA+(x'_1+x'_2+\cdots\cdots+x'_n) \\ &= nA+\Sigma x', \end{aligned}$$

兩邊除以 n 得

$$\frac{\Sigma X}{n}=A+\frac{\Sigma x'}{n},$$

$$\therefore\ M=A+\frac{\Sigma x'}{n} \qquad \text{(公式 8)}$$

〔例〕求身長 52 寸, 53 寸, 54 寸, 55 寸, 56 寸等 5 人之平均身長爲幾何?

任意取 $A=52$ 寸,

$$x'_1=X_1-A=52-52=0,$$

$$x'_2=X_2-A=53-52=1,$$

$$x'_3=X_3-A=54-52=2,$$

$$x'_4=55-52=3,$$

$$x'_5=56-52=4,$$

$$\Sigma x'=0+1+2+3+4=10,$$

$$\frac{\Sigma x'}{n}=\frac{10}{5}=2,$$

$$\therefore\ M=A+\frac{\Sigma x'}{n}=52+2=54\text{ 寸},$$

其結果與前法相同。

至公式 7 之簡便算法如次:

$$X_r=A+x'_r,$$

$$f_rX_r=Af_r+f_rx'_r,$$

$$\Sigma_r f_rX_r=A\sum_r f_r+\sum_r f_rx'_r$$

$$\therefore\ M=A+\frac{\Sigma fx'}{\Sigma f} \qquad \text{(公式 9)}$$

將公式 9 與公式 7 比較時，其公式上雖祇將 $\Sigma fX$ 換以 $\Sigma fx'$ 而已，但在計算上則公式 9 大爲簡便。現將本法應用於公式 7 之例上則如次：

$$\text{任意取}\quad A=50,$$

$$x'_1=X_1-A=50-50=0,\qquad f_1x'_1=19\times 0=0,$$

$$x'_2=X_2-A=52-50=2,\qquad f_2x'_2=90\times 2=180,$$

$$x'_3=X_3-A=54-50=4,\qquad f_3x'_3=195\times 4=780,$$

$$x'_4=X_4-A=56-50=6,\qquad f_4x'_4=141\times 6=846,$$

$$x'_5=X_5-A=58-50=8,\qquad f_5x'_5=20\times 8=160,$$

$$\Sigma fx'=0+180+780+846+160=1966,$$

$$\frac{\Sigma fx'}{\Sigma f}=\frac{1966}{465}=4.23,$$

$$\therefore M=A+\frac{\Sigma fx'}{\Sigma f}=50+4.23=54.23\text{ 寸。}$$

其結果與前相一致。

依上法知組距巨大時，其計算尙覺麻繁。故在各組距相等時，依次法進行，較爲簡便。

(1)取任意一組之中位數爲 A（以取次數最大一組之中心數值較爲便利）。

(2)以組距爲單位所測得之偏差爲 $\bar{x}'$（$x'=a\bar{x}'$ a＝組距，故知 $\bar{x}'$ 爲零數左右之整數）。

(3)將 $f\bar{x}'$ 之正部分與負部分分別計算（先求正部分之和及負部分之和，而後再求二者之代數和，$\Sigma f\bar{x}'$）。

(4)用 $\Sigma f$ 除此 $\Sigma f\bar{x}'$ 後再乘以組距 a $\left(\text{即 } a\times\frac{\Sigma f\bar{x}'}{\Sigma f}\right)$。

(5)於(4)之結果上加以A，即得吾人所求之結果。

即 $$M=A+a\frac{\Sigma f\bar{x}'}{\Sigma f}$$ (公式 10)

其代表形式如第 61 表。

第 61 表

| 組別 | 代表值 X | 次數 f | 偏差 $x'$ | 以組距爲單位之偏差 $\bar{x}'=\frac{x'}{a}$ | 積 $f\bar{x}'$ |
|---|---|---|---|---|---|
| $L_1$——$u_1$<br>$L_2$——$u_2$<br>⋮<br>$L_m$——$u_m$ | $X_1$<br>$X_2$<br>⋮<br>$X_m$ | $f_1$<br>$f_2$<br>⋮<br>$f_m$ | $x'_1$<br>$x'_2$<br>⋮<br>$x'_m$ | $\bar{x}'_1$<br>$\bar{x}'_2$<br>⋮<br>$\bar{x}'_m$ | $f_1\bar{x}'_1$<br>$f_2\bar{x}'_2$<br>⋮<br>$f_m\bar{x}'_m$ |
| 計 | | $\Sigma f$ | | | $\Sigma f\bar{x}'$ |

$$a=u_i-L_i,\qquad \bar{d}=\frac{\Sigma f\bar{x}'}{\Sigma f},\qquad d=a\cdot\bar{d}$$

$$M=A+d=A+\frac{\Sigma f\bar{x}'}{\Sigma f}\times a$$

將本法應用於公式 7 之例上則如第 62 表〔例〕

第 62 表

| 身長組別 (尺寸 尺寸) | 組之中心 X | 人數 f | $\bar{x}'$ | $f\bar{x}'$ |
|---|---|---|---|---|
| 49——51 | 50 | 19 | −2 | −38 |
| 51——53 | 52 | 90 | −1 | −90 |
| | | | | −128 |
| 53——55 | A=54 | 195 | 0 | |
| 55——57 | 56 | 141 | 1 | 141 |
| 57——59 | 58 | 20 | 2 | 40 |
| 計 | | 465 | | 181<br>−128<br>53 |

$$\frac{\Sigma f\bar{x}'}{\Sigma f}=\frac{53}{465}=0.114$$

$$M=A+\frac{\Sigma f\bar{x}'}{\Sigma f}\times a=54+0.228=54.228\text{寸}$$

(B) 由次數曲線以求算術平均數

假定次數曲線之方程式爲 $y=f(x)$ 時，

$$M=\frac{\int_{-\infty}^{\infty}xf(x)dx}{\int_{-\infty}^{\infty}f(x)dx} \qquad \text{(公式 11)}$$

(C) 算術平均數之性質

(甲)優點

(1)性質簡單明瞭易於了解，且計算容易，無須高深數學。

(2)能代數的運用。

(a) 算術平均數與求出此平均數之各單位值相差數之總和等於零。

〔證〕M 爲單位值 $X_1, X_2, \cdots\cdots X_n$ 之算術平均數，

$$X_1-M=x_1,$$

$$X_2-M=x_2,$$

$$\cdots\cdots\cdots\cdots,$$

$$X_n-M=x_n,$$

將此等方程式邊邊相加得

$$\Sigma X-nM=\Sigma x,$$

但算術平均數 M 是由

$$M=\frac{1}{n}\Sigma X, \quad \text{即} \quad nM=\Sigma X.$$

以求得之，故將此式與上式相比較時，得

$$\Sigma x=0 \qquad \text{(公式 12)}$$

知各單位值與 M 之偏差之和爲零。

(b)如單位值 X，由 n 個部分數列所合成時，其 X 之平均數得由各部分之平均數以表示之。

〔證〕設數列 $X_1X_2\cdots\cdots X_{n_1}$ 之算術平均數爲 M，

數列 $Y_1Y_2\cdots\cdots Y_{n_2}\cdots\cdots$ 爲 $M_2$

時此等全體之算術平均數 M 得由

$$M=\frac{(X_1+X_2+\cdots\cdots+X_{n_1})+(Y_1+Y_2+\cdots\cdots+Y_{n_2})}{n_1+n_2}$$

$$=\frac{\Sigma X+\Sigma Y}{n_1+n_2}=\frac{n_1M_1+n_2M_2}{n_1+n_2}$$

一般

$$NM=N_1M_1+N_2M_2+\cdots\cdots\cdots+N_nM_n \quad (公式13)$$

但 $$N=\Sigma N_i$$

〔例〕甲講堂內學生共 56 人，其平均身長爲 54.3寸，乙講堂內學生共 48 人，其平均身長爲 53.8 寸，問二講堂合併後全體學生之平均身長爲幾何？

依題意知

$N_1=56,\qquad M_1=54.3,$

$N_2=48,\qquad M_2=53.8,$

$N=N_1+N_2=56+48=104,$

依公式得

$$104\,M=56\times54.3+48\times53.8$$

$$\therefore\ M=54.1\ 寸弱。$$

(3)得省畧將各單位值依順序整列（如求中位數）以求代表物之手續。

(4)只要明白單位值之個數及總和後，雖不知各個詳情，亦能求得

其平均數。

單位値之個數爲 N，各單位値之總和爲 S 時，M 得由次式以求得之。

$$M=\frac{S}{N} \qquad \text{(公式 14)}$$

〔例〕只知職工人數及所發工資總數而不知各人實得國幣數時，亦能依上式求得其平均工資爲多少。

算術平均數有此性質，故常用爲各種估計之基礎。

(5)其平均數與各個單位値皆有密切之關係。

(6)誤差法則上之最確値卽爲此算術平均數，故在誤差法則上含有重大使命。

(乙)劣點

(1)其平均數常爲實際上所無之數値。例如依算術平均法求得甲省之每戶平均人數爲 4.3 人，但此數目爲實際上所無，只一抽象數而已。

(2)遇有特大或特小之單位値時，其影響於平均數上甚大。

(3)缺乏數列之極端數値時，則不能正確決定之。

(4)其平均數不能如中位數最密數然由圖形上以求得之。

(5)不能計量之物，不能求得算術平均數，此處不及中位數。

(D) 算術平均數之應用範圍

(1)算術平均數之適用範圍最廣，幾可適用於統計之全部份。但一般是最適合於測定含有靜的性質之現象上。

(2)倘數列之各値，應各得其相當之權。

(3)倘吾人欲得到最高之可靠性。

(4)倘以後須計算標準差及相關係數等。

## 第三節 中位數

### (A) 實用方法

將各觀察單位值，依大小順序以行排列，在此排列內最中之單位值，即爲吾人所求之中位數，普通用 $M_i$ 以表示之。(本法由遺傳學者 F. Galton 所最初使用) 對此單位值之個數 N 爲奇數 ($N=2K+1$) 時，則由小或由大數起第 $K+1$ 號單位值 $X_{K+1}$，即爲所求之 $M_i$。

〔例〕將 65，75，65，70，80，50，40，依大小排列之如次：

40， 50， 65， 65， 70， 75， 80，

其最中(第四號)一數值爲 65，故知 $M_i=65$。

若單位值之個數 N 爲偶數 ($N=2K$) 時，其中間單位值有二個 ($X_K$, $X_{K+1}$)，當此時則取二單位值之算術平均數

$$\frac{1}{2}(X_K+X_{K+1}),$$

爲該數列之中位數。

〔例〕

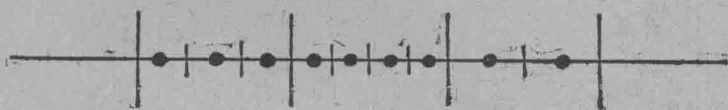
在 40，50，65，75，80，85 內，依其 $N=6$ 知中間單位值有二個爲 65 與 75，此二單值之算術平均數則爲所求之中位數。

$$\therefore\ M_i=\frac{1}{2}(65+75)=70。$$

本法在各單位值能依次排列時，則甚簡單。但遇觀察結果依組別以給與時，則不能卽刻求出其中位數，須依賴數學助力，而後始可。但當此時，有次之三假定。

假定 1　將各數值看做為連續性。雖如人數房間數等品質，亦得附以小數。

假定 2　一組內所屬數值，均看做為平均分配。卽將組距依所屬次數以行等分，並假定各單位值是在各等分區域之中間如第 94 圖，圖內之·是表示各單位值之位置。



第 94 圖

假定 3　將第$\frac{N}{2}$號數值看做為中位數。

此處之所以取第$\frac{N}{2}$號而不取第$\frac{N+1}{2}$號者，因在實際上 N 之數值相當巨大，且在上述假定之下其取$\frac{N}{2}$與$\frac{N+1}{2}$所得結果，本無甚差異。更在以後求中位數之公式內，如取$\frac{N+1}{2}$而不取$\frac{N}{2}$時，則從最小算起所求得之 $M_i$ 與從最大算起所求得之 $M_i$ 不相一致，必須改為$\frac{N}{2}$而後始能互相一致。

在此三假定之下，求統計材料依組別分配以給與時之中位數方法如次：

各組次數為

$f_1, f_2 \cdots\cdots, f_m$，且 $\Sigma f = N$時，先求能滿足關係式

$$f_1 + \cdots\cdots + f_{r-1} < \frac{N}{2} < f_1 + \cdots\cdots + f_{r-1} + f_r$$

之 $f_r$，此 $f_r$ 為單位值在 $L_r - U_r$ 內之次數，由此知所求中位數 $M_i$ 必為

$L_r-U_r$ 內之一值，現假定該值與 $L_r$ 之距離爲 x，$L_r$ 與 $L_r+x$ 間所有次數假定其爲 δ，依第 63 表知

第 63 表

| 組別 | 次數 |
|---|---|
| $L_1$ — $U_1$ | $f_1$ |
| ⋮ | ⋮ |
| $L_{r-1}$ — $U_{r-1}$ | $f_{r-1}$ |
| $L_r$ — $U_r$ | $f_r$ |
| $L_{r+1}$ — $U_{r+1}$ | $f_{r+1}$ |
| ⋮ | ⋮ |
| $L_m$ — $U_m$ | $f_m$ |
| 計 | $\Sigma f=N$ |

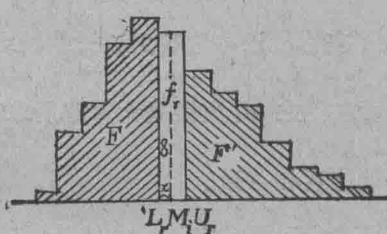
$$f_1+f_2+\cdots\cdots+f_{r-1}=F,$$

$$f_r=F_r,$$

$$f_{r+1}+f_{r+2}+\cdots\cdots+f_m=F',$$

$L_r-U_r$ 之組距 $=a$ 時，得

$$\delta=\frac{N}{2}-F \qquad (1)$$




第 95 圖

並應用一次補間(linear interpolation)原理

得
$$\frac{\delta}{f_r}=\frac{x}{a}, \qquad (2)$$

解此二式得

$$x=\frac{\frac{N}{2}-F}{f_r}a,$$

$$\therefore M_i = L_r + x = L_r + \frac{\frac{N}{2} - F}{f_r} a$$

或　(公式 15)

$$M_i = U_r - x' = U_r - \frac{\frac{N}{2} - F'}{f_r} a$$

〔例 I〕

第 64 表

| 組別 | 次數 |
|---|---|
| 35——45 | 1 } $F=3$ |
| 45——55 | 2 } |
| 55——65 | 4　$f_r=4$ |
| 65——75 | 3 } $F'=5$ |
| 75——85 | 2 } |
| 計 | 12 |

$$\frac{1}{2}N=6, L_r=55, u_r=65, a=10$$

$$M_i = L_r + \frac{\frac{N}{2} - F}{f_r} a$$

$$= 55 + \frac{6-3}{4} \times 10$$

$$= 55 + 7.5$$

$$= 62.5_{\circ}$$

或
$$M_i = U_r - \frac{\frac{N}{2} - F'}{f_r} a = 65 - \frac{6-5}{4} \times 10$$
$$= 65 - 2.5 = 62.5$$

二者相一致。

〔例 II〕若次數分配表依第 65 表(甲)之形式以給與時,須改爲(乙)而後始能計算其中位數。

第 65 表

(甲)

| 代表值 | f |
|---|---|
| 45 | 3 |
| 55 | 19 |
| 65 | 30 |
| 75 | 15 |
| 85 | 2 |
| 95 | 1 |
| 計 | 70 |

(乙)

| 組別 | 代表值 | f | |
|---|---|---|---|
| 40——50 | 45 | 3 | F =22 |
| 50——60 | 55 | 19 | |
| 60——70 | 65 | 30 | $f_r$ =30 |
| 70——80 | 75 | 15 | F′=18 |
| 80——90 | 85 | 2 | |
| 90——100 | 95 | 1 | |
| 計 | | 70 | |

$$\frac{N}{2} = 35, \quad L_r = 60, \quad U_r = 70, \quad a = 10$$

$$M_i = L_r + \frac{\frac{N}{2} - F}{f_r} a = 60 + \frac{35-22}{30} \times 10$$
$$= 60 + 4.33 = 64.33$$

或
$$M_i = U_r - \frac{\frac{N}{2} - F'}{f_r} a = 70 - \frac{35-18}{30} \times 10$$
$$= 70 - 5.67 = 64.33$$

## (B) 由次數曲線以求中位數

設次數曲線之方程式爲

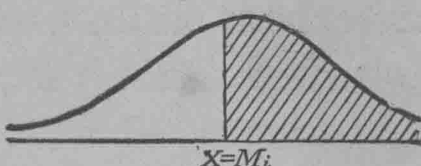
$$y=f(x),$$

此曲線與 x 軸所包圍之面積是代表次數之全體，故在此面積內用 $x=X$ 之縱線分爲二部，若 $x=X$ 縱線之左部面積等於右部面積時，此 X 卽爲所求之中位數。其數式如次：

$$\int_{-\infty}^{X} f(x)\,dx=\int_{X}^{+\infty} f(x)dx,$$

卽

$$\int_{-\infty}^{X} f(x)dx-\int_{X}^{\infty} f(x)dx=0 \qquad \text{(公式 16)}$$




第　96　圖

能滿足此積分方程式之 X，卽爲所求之中位數。

## (C) 由累加次數曲綫以求中位數。

設次數曲線之方程式爲

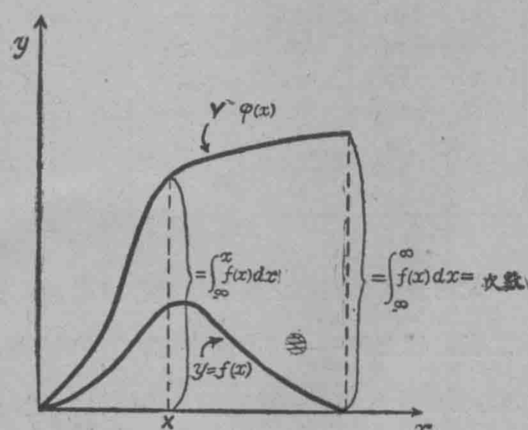
$$y=f(x),$$

累加次數曲線之方程式爲

$$Y=\varphi(x)$$

時，依累加次數曲線之性質，知

$$Y=\varphi(x)\int^{X} f(x)dx,$$




第　97　圖

在此時中位數所當之 Y，適等於總次數之半。其數學公式爲

$$\frac{\int_{-\infty}^{\infty} f(x)dx}{2}=\varphi(x)$$

即
$$\int_{-\infty}^{\infty} f(x)dx=2\varphi(x) \qquad \text{(公式 17)}$$

滿足此方程式之 X 卽爲所求之中位數。

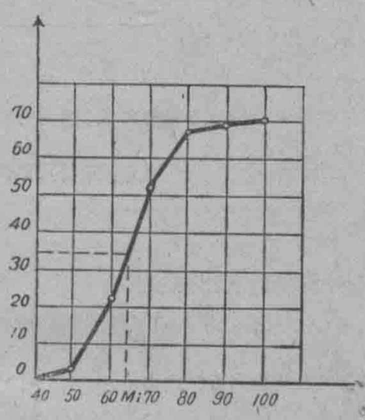
在實際上由累加次數曲線以求中位數方法如次：

先由次數分配表求得累加次數以作成累加次數曲線後，在此最大累加數（總次數）之半所當點處引一與橫軸相平行直線使與曲線相交，由此交點引垂直線與橫軸相相交，此交點所當數值卽爲所求中位數。

〔例〕

第 66 表

| 組別 | 次數 | 累加次數 |
|---|---|---|
| 40—50 | 3 | 3 |
| 50—60 | 19 | 22 |
| 60—70 | 30 | 52 |
| 70—80 | 15 | 67 |
| 80—90 | 2 | 69 |
| 90—100 | 1 | 70 |
| 計 | 70 | |




第 98 圖

先由表內最右欄之累加次數作成累加次數折線如第 98 圖，在此圖內之在縱軸上與 $\frac{N}{2}=35$ 所當處，引一與橫軸相平行之直線，使與折線相交，再由此交點引一垂直線使與橫軸相交，依此交點所當值知 $M_i=63$，但依公式 15 求得

$$M_i = L_r + \frac{\frac{N}{2} - F}{f_r} a$$

$$= 60 + \frac{30-32}{30} \times 10 = 62.7 \fallingdotseq 63$$

知相差無幾。

(D) 中位數之性質

(甲)優點

(1)性質簡單明瞭，易於了解。

(2)分配能依大小順序整列時，其發見手續，甚為簡單。雖依組以分配時，仍能依一次補間法以求得之。

(3)對於特大與特小等單位值之影響甚少，故雖有無限大及無限小之極端單位值時，仍能毫無關係以求得之。

(4)中位數之位置較為確定，不受少數單位值變動之影響。

(5)在特定之觀察數列，不依計量方法之不同而異其結果。

〔例〕在第 67 表內(a)(b)二欄，為計量單位不同之同一數列。

第　67　表

| 號碼 | (a) 一斗米之價錢(元) | (b) 一元能購得數量 |
|---|---|---|
| I | 0.82 | 1.22 |
| II | 0.80 | 1.25 |
| III | 0.72 | 1.39 |
| IV | 0.70 | 1.43 |
| V | 0.65 | 1.54 |
| VI | 0.53 | 1.89 |
| VII | 0.50 | 2.00 |
| 總和 | 4.72 | 10.72 |
| 算術平均數 | 0.67 | 1.53 |
| 中位數 | 0.70 | 1.43 |

在本表內(a)(b)二欄之中位數均爲第四號之單位值，而在算術平均數則不然,在(a)欄內是在第四第五二號之中間，(b)欄是爲第五號，由此知中位數不依計量方法之不同,而異其結果。

(6)得在圖形上求得之。

(7)在不能依數量以表示各單位值只在知其順序時，仍能依中位數,以求得其平均數。

(乙)劣點

(1)不能代數的運用。

(2)雖知單位值之個數與總和,仍不能決定之。

(3)在單位值之個數適爲偶數時,其中位數變爲實際上所不存在之數值(組別分配時亦有此缺點)。

(4)因其不使用觀察數列全體,故與分配次數內之極大值及極小值完全無關,在重視極端值之研究上,不甚適用。

(E)中位數之應用範圍

(1)中位數是在統計資料有密集於中央之傾向時最爲適宜之平均數值,從來將中位數常適用於人口統計方面之表示年齡構成上,例如蓋然壽命是爲同時出生之同一年齡者活至已成爲半數死亡時之壽命，是由死亡生剩表以求其中位數。又如中央年齡是由生存者之年齡構成所求得之中位數。

(2)倘極端值對於平均數所發生之影響過大，而吾人欲免除之。例如計算工廠工人之平均工資,或計數某地方人民之平均資產,則欲免除經理或巨富者之影響,可用中位數以代表平均數。

(3)倘事實之一部或全部缺乏精確數量。

(4)倘以後有計算四分位差及平均偏差之必要者。

(5)此外在物價統計用作爲平均指數，尤在計算季節指數上最多使用之。

## 第四節　最衆數

統計表中之占最大次數之單位值，稱爲最衆數，用 $M_0$ 以表示之。此最衆數由 Karl Pearson 依 Silk hat 型之暗示以求得之數值。其原文爲 Mode ，是脫胎於法文 La mode 一字，La mode 之意義是爲流行，故知最衆數之意義與流行某種服飾之意義相同，其數值爲製造業者所最需要。例如鞋襪製造者爲使其製造品能適合多數人士起見，於事前有知人足上最衆數之必要。

最衆數在各單位值能排列清楚時，其求法甚爲簡單，例如第 68 表

第　68　表

| 年齡 | 人數 | 年齡 | 人數 | 年齡 | 人數 | 年齡 | 人數 |
|---|---|---|---|---|---|---|---|
| 7 | 2 | 18 | 16 | 29 | 50 | 40 | 16 |
| 8 | 4 | 19 | 17 | 30 | 45 | 41 | 18 |
| 9 | 5 | 20 | 18 | 31 | 40 | 42 | 13 |
| 10 | 3 | 21 | 19 | 32 | 38 | 43 | 9 |
| 11 | 8 | 22 | 20 | 33 | 36 | 44 | 11 |
| 12 | 6 | 23 | 23 | 34 | 37 | 45 | 6 |
| 13 | 10 | 24 | 26 | 35 | 32 | 46 | 4 |
| 14 | 7 | 25 | 30 | 36 | 26 | 47 | 5 |
| 15 | 9 | 26 | 28 | 37 | 20 | 48 | 3 |
| 16 | 14 | 27 | 32 | 38 | 25 | 49 | 4 |
| 17 | 13 | 28 | 43 | 39 | 22 | 50 | 1 |

內，其人數欄內最多一行爲 50 人，此 50 人所屬年齡組爲 29 歲，由此知本表內年齡之最衆數爲 29 歲。然將前表改爲四歲組別之次數分配表如第69表時，其最衆數雖能由表卽知其在(27－30)之年齡組內，但欲確定其數值，則甚困難。依此學者發明種種方法以補救之。現將各補救方法中，擇其主要者，列舉五種如次：

第 69 表

| 年齡 | 人數 |
|---|---|
| 7—10 | 14 |
| 11—14 | 31 |
| 15—18 | 52 |
| 19—23 | 74 |
| 23—26 | 107 |
| 27—30 | o→170 |
| 31—34 | 151 |
| 35—38 | 103 |
| 39—42 | 69 |
| 43—46 | 30 |
| 47—50 | 13 |

(1)編羣法(Aprocess grouping)卽爲交換分類組別之方法。

(2)金氏方法

(3)披爾遜實驗公式

(4)由次數曲線以求最衆數其方法

(5)由累加次數曲線以求最衆數方法

以下依次說明。

(A) 編羣法

在次數分配表內，若祇取其占有最大次數所屬組之中位數，作爲所求之最衆數時，其方法實無甚價值。蓋因有依所選定組距之不同而異其最衆數所屬組之事，參照次例卽能明白了解。

〔例〕第 70 表爲學生千名之體重分配狀況，其第二欄是爲依五磅組距所定各組人數，第三欄及第四欄爲依十磅組距所定各組人數，第五至第七欄爲十五磅組距所定人數，第八至第十一欄爲二十磅組距所定人

## 第 70 表　依編羣法以計算最密數

| 體重 | 人數 | 十磅別各組之次數 | | 十五磅別各組之次數 | | | 二十磅別各組之次數 | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| (五磅別) | | 最初組之下限值爲90磅 | 最初組之下限值爲95磅 | 最初組之下限值爲90磅 | 最初組之下限值爲95磅 | 最初組之下限值爲100磅 | 最初組之下限值爲90磅 | 最初組之下限值爲95磅 | 最初組之下限值爲100磅 | 最初組之下限值爲105磅 |
| | f | f | f | f | f | f | f | f | f | f |
| (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| 90——95 | 6 | 13 | | 23 | | | 41 | | | |
| 95——100 | 7 | | 17 | | 35 | | | 100 | | |
| 100——105 | 10 | 28 | | | | 93 | | | 147 | |
| 105——110 | 18 | | 83 | 164 | | | | | | 275 |
| 110——115 | 65 | 146 | | | 257 | | 391 | | | |
| 115——120 | 81 | | 192 | | | 326 | | 451 | | |
| 120——125 | 111 | 245 | | 370 | | | | | 487 | |
| 125——130 | 134 | | 259 | | 376 | | | | | 461 |
| 130——135 | 125 | 242 | | | | 327 | 402 | | | |
| 135——140 | 117 | | 202 | 277 | | | | 331 | | |
| 140——145 | 85 | 160 | | | 214 | | | | 249 | |
| 145——150 | 75 | | 129 | | | 164 | | | | 189 |
| 150——155 | 54 | 89 | | 114 | | | 135 | | | |
| 155——160 | 35 | | 60 | | 81 | | | 14 | | |
| 160——165 | 25 | 46 | | | | 59 | | | 64 | |
| 165——170 | 21 | | 34 | 39 | | | | | | 44 |
| 170——175 | 13 | 18 | | | 23 | | 27 | | | |
| 175——180 | 5 | | 10 | | | 14 | | 16 | | |
| 180——185 | 5 | 9 | | 11 | | | | | 12 | |
| 185——190 | 4 | | 6 | | 7 | | | | | 7 |
| 190——195 | 2 | 3 | | | | 3 | 4 | | | |
| 195——200 | 1 | | 1 | 2 | | | | 2 | | |
| 200——205 | 0 | 1 | | | | | | | | |
| 205——210 | 1 | | | | | | | | | |

數，各欄組別界限值順次相差，皆爲五磅。本表內依壯體數字現出之次數，爲依各分類法所得次數分配表內之最大次數。觀各最大次數之位置，知依組距與界限決定方法之不同而生差異，茲將此等最衆組(modal class)列表如第71表。觀表知最衆組在125磅—130磅者與在130磅—135磅者各八次，爲最多數。由此得推想所求之最衆數必在此二分類組間。又因其所屬回數均爲八回，故得決定其二分類組相鄰之中位數(即 130 磅)定作爲所求之最衆數。

第　71　表

| 欄 | 最　衆　組　(磅) |
|---|---|
| (2) | 125——130 |
| (3) | 120——130 |
| (4) | 125——135 |
| (5) | 120——135 |
| (6) | 125——140 |
| (7) | 130——145 |
| (8) | 130——150 |
| (9) | 115——135 |
| (10) | 120——140 |
| (11) | 125——145 |

125 磅——130 磅者 8 次
130 磅——135 磅者 8 次

總之編羣法是爲先由變更種種組距與界限值，以製成多數次數分配表後，再由此等分配表內求出其最衆組，並在此等最衆組中，擇其最多發生之共通組之中位數，作爲所求之最衆數。

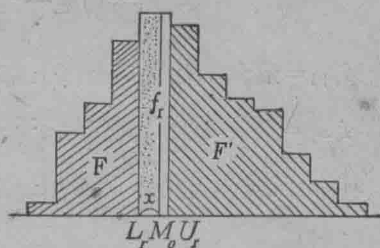
(B) 金氏方法

次數分配表完全爲對稱時，最衆數當然在含有次數最多一組之中

央。若次數配列並非完全對稱成爲近似的對稱時，最衆數是偏在次數最多組之一方。金氏依此原理，取次數分配表內含有最大次數組之次數爲 $f_r$，其兩側各次數和爲 F 與 F′ 時，現以 F，F′ 爲反比例以分 $f_r$ 爲二部分，其分線之橫座標即爲所求之最衆數。

第 72 表

| 組別 | 次數 |
|---|---|
| $L_1$ — $U_1$ | $f_1$ } F |
| ⋮ | ⋮ |
| $L_{r-1}$ — $U_{r-1}$ | $f_{r-1}$ |
| $L_r$ — $U_r$ | $f_r$ |
| $L_{r+1}$ — $U_{r+1}$ | $f_{r+1}$ } F′ |
| ⋮ | ⋮ |
| $L_m$ — $U_m$ | $f_m$ |




第 99 圖

設次數分配表內含有最大次數之組爲 $L_r-U_r$，組距爲 a，假定最衆數 $M_0$ 與下界 $L_r$ 之差爲 x，依上之理論得

$$x:(a-x)=F':F$$

由此得

$$x=\frac{F'}{F+F'}a。$$

$$\therefore\ M_0=L_r+x=L_r+\frac{F'}{F+F'}a$$

或 (公式 18)

$$M_0=U_r-x'=U_r-\frac{F}{F+F'}a$$

關此有許多書上只使其鄰接二組所屬次數以計算一切，其公式如次：

$$M_0=L_r+\frac{f_{r+1}}{f_{r+1}+f_{r-1}}a$$

或 (公式 19)

$$M_0=U_r-\frac{f_{r-1}}{f_{r+1}+f_{r-1}}a$$

$f_{r+1}$＝最大次數所屬組以上鄰接一組之數次。

$f_{r-1}$＝最大次數所屬組以下鄰接一組之次數。

但由此式以算得者，不若上式之為正確合理。

〔例〕

第 73 表

| 組別 | 次數 | |
|---|---|---|
| 0——5 | 2 | F=67 |
| 5——10 | 14 | |
| 10——15 | 20 | |
| 15——20 | 31 | |
| 20——25 | 36 | $f_r=36$ |
| 25——30 | 34 | F′=122 |
| 30——35 | 25 | |
| 35——40 | 31 | |
| 40——45 | 19 | |
| 45——50 | 11 | |
| 50——55 | 6 | |
| 55——60 | 5 | |
| 60——65 | 1 | |
| 計 | 225 | |

依第 73 表知最大次數組為 20－25，由此得

$L_r=20,\qquad U_r=25,$

$F=67,\qquad F'=122,$

$a=5$。

$$\therefore\ M_0=L_r+\frac{F'}{F+F'}\times a=20+\frac{122}{67+122}\times 5=23.23,$$

$$\text{或}\ M_0=U_r-\frac{F}{F+F'}\times a=25-\frac{67}{67+122}\times a=23.23。$$

(C) 披爾遜實驗公式

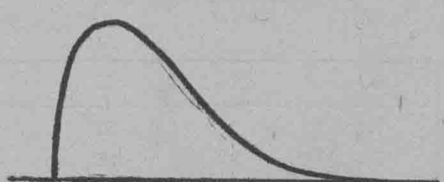
披爾遜於 1895 年依伊所定次數曲線 Type III 上，用實驗方法求得

$$M_0=M-\frac{M-M_i}{C}$$

但　(公式 20)

$$C=0.3309-\frac{0.0846(M-M_i)^2}{\sigma^2-9\ (M-M_i)^2},$$

故在已知次數分配之 M，$M_i$，及 $\sigma$，即能求得 $M_0$。至 Tyhe III 之曲線形，是一方限定一方不限定之不對稱形曲線。但在實際上除最衆數是特別接近於一方，及曲線上其他一邊是特別接近橫軸者(例如收入之分狀況)如第 100 圖外，(參照第 106 圖) 使用 C=0.33 已足應用，即取 $M_0$ 之公式爲



第 100 圖

$$M_0=M-\frac{M-M_i}{0.33}=M-3.03(M-M_i)\qquad\text{(公式 21)}$$

在實際上爲簡便起見多使用次之公式以計算最衆數,

$$M_0 = M - 3(M - M_i) \qquad \text{(公式 22)}$$

本公式之眞實性雖至今倘未證明,但依經驗上之應用知其爲最近於眞正最衆數之一數值,故仍多使用之。

〔例〕在第 70 表內求得

$$M = 134.4 \text{ 磅},$$

$$M_i = 132.7 \text{ 磅},$$

依公式 22 知該分配狀況之最衆數爲

$$M_0 = M - 3(M - M_i)$$

$$= 134.4 - 3(134.4 - 132.7)$$

$$= 129.3 \text{ 磅}。$$

(D) 由次數曲線以求最衆數

在次數分配不甚整齊而多凸凹時,欲定其最大次數所在處,實甚困難。當此時是多先取相當組距之組以作成圓滑的分配圖,在此圖內最大縱線所對應之橫座標則爲所求之最衆數。

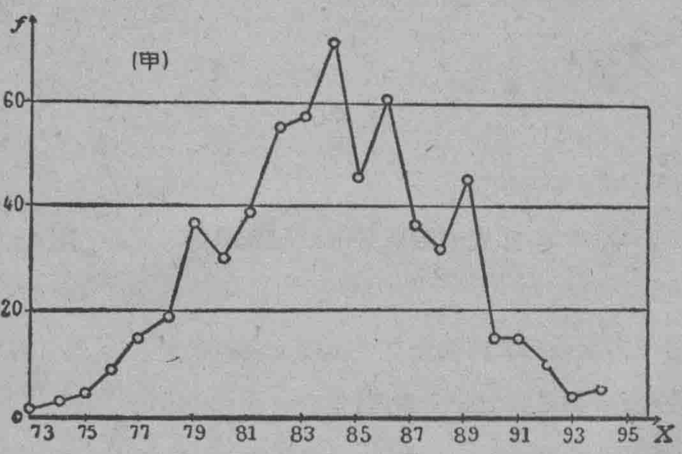
〔例〕 (甲)

| X | f | X | f | X | f |
|---|---|---|---|---|---|
| 73 | 1 | 81 | 39 | 89 | 45 |
| 74 | 3 | 82 | 55 | 90 | 15 |
| 75 | 4 | 83 | 57 | 91 | 16 |
| 76 | 9 | 84 | 72 | 92 | 11 |
| 77 | 15 | 85 | 45 | 93 | 5 |
| 78 | 19 | 86 | 61 | 94 | 6 |
| 79 | 37 | 87 | 37 | | |
| 80 | 30 | 88 | 32 | 計 | 608 |

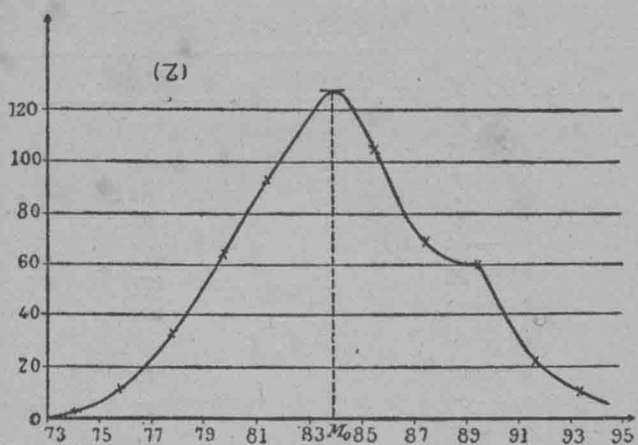
（乙）

| 組　　　別 | f |
| --- | --- |
| 73——75 | 4 |
| 75——77 | 13 |
| 77——79 | 34 |
| 79——81 | 67 |
| 81——83 | 94 |
| 83——85 | 129 |
| 85——87 | 106 |
| 87——89 | 69 |
| 89——91 | 60 |
| 91——93 | 21 |
| 93——95 | 11 |
| 計 | 608 |

甲表爲不整齊之次數分配，將此圖示時成爲第 101 圖甲之凸凹衆多之次數分配圖，在此圖內實難以確定最大次數所在處，故特依以 2 爲組距以作成乙表後，再由此表以圖示之則得乙圖，在乙圖內卽得最大次數之所在，由此卽知 $M_0=84.3$。




第　101　圖

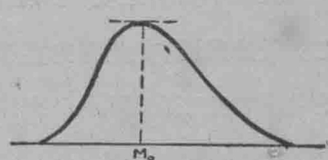



第　101　圖

至數理之說明如次：

設次數曲線之方程式爲

$$y=f(x),$$




第　102　圖

最衆數是爲在此次數曲線內占最大 y 時之 x 所當值，故在 $y=f(x)$ 之曲線上，能滿足微分方程式

$$\begin{cases} \dfrac{dy}{dx}=0 \\ \dfrac{d^2y}{dx^2}<0 \end{cases}$$

之 x 所當值，卽爲所求之最衆數。

理論上雖如此，但在實際依次數分配表以行計算時，多依次之二法以求 $M_o$。

(1)在次數分配表內只含有一個最大次數時之求法。

設在所給與之次數分配表內，定

$f_r$ 爲最大次數，

$f_{r-1}$ 是比 $f_r$ 爲低各組中與 $f_r$ 最爲鄰近一組之次數，

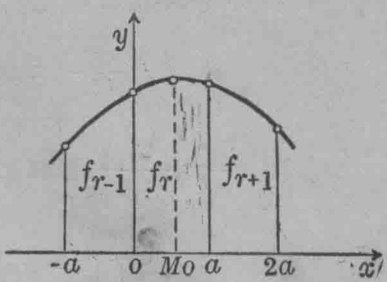
$f_{r+1}$ 是比 $f_r$ 爲高各組中與 $f_r$ 最爲鄰近一組之次數，

a＝組距，

依數學智識，知無論任何曲線之一小部分，得看做爲拋物線形。現定次數曲線內三組所作成一部分曲線之方程式爲

$$y=b_1x^2+b_2x+b_3,$$

並以最大次數所屬組之下界値爲原點，各組之界限點成爲 －a，o，a，2a (取組距單位亦可)，但依積分學並參照第 103 圖得




第 103 圖

$$f_{r-1}=\int_{-a}^{0} y\,dx=\int_{-a}^{0}(b_1x^2+b_2x+b_3)dx$$

$$=\left|\frac{b_1}{3}x^3+\frac{b_2}{2}x^2+b_3x\right|_{-a}^{0}$$

$$=\frac{b_1}{3}a^3-\frac{b_2}{2}a^2+b_3a,$$

同樣得

$$f_r=\int_0^a y\,dx=\frac{b_1}{3}a^3+\frac{b_2}{2}a^2+b_3a,$$

$$f_{r+1}=\int_a^{2a} y\,dx=\frac{7}{3}b_1a^3+\frac{3}{2}b_2a^2+b_3a,$$

用遞差法 (Method of difference)得

$$\Delta f_{r-1}=f_r-f_{r-1}=b_2a^2,$$

$$\Delta f_r=f_{r+1}-f_r=2b_1a^3+b_2a^2,$$

$$\Delta^2 f_{r-1}=\Delta f_r-\Delta f_{r-1}=2b_1a^3,$$

$$\therefore\quad \frac{b_2}{2b_1}=\frac{\Delta f_{r-1}}{\Delta^2 f_{r-1}}a。$$

但使 $y=b_1x^2+b_2x+b_3$ 為極大時之條件，為

$$\frac{dy}{dx}=2b_1x+b_2=0,$$

由此得 $x=-\frac{b_2}{2b_1}$，現將上之差數值代入後，得

$$x=-\frac{\Delta f_{r-1}}{\Delta^2 f_{r-1}}a,$$

知在此 x 所當處之 y 成為最大，此 x 是以最大次數組之下界值為原點，故知所求之最衆數為

$$M_0=L_r+x=L_r-\frac{\Delta f_{r-1}}{\Delta^2 f_{r-1}}a \qquad \text{(公式 23)}$$

〔例〕 第 74 表

| 組別 | 次數 |
|---|---|
| ⋮ | ⋮ |
| 79——81 | 67 |
| 81——83 | 94……$f_{r-1}$ |
| 83——85 | 129……$f_r$ |
| 85——87 | 106……$f_{r+1}$ |
| 87——89 | 69 |
| ⋮ | ⋮ |

由第 74 表知

$$\Delta f_{r-1}=f_r-f_{r-1}$$
$$=129-94=35,$$
$$\Delta f_r\ =f_{r+1}-f_r$$
$$=106-129=-23,$$
$$\Delta^2 f_{r-1}=\Delta f_r-\Delta f_{r-1}$$
$$=-23-35=-58,$$
$$x=-\frac{\Delta f_{r-1}}{\Delta^2 f_{r-1}}a=-\frac{35}{-58}\times 2=1.2,$$

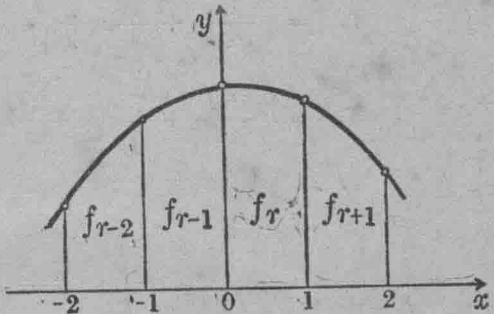
但 $L_r=83$，

$$\therefore\quad M_0=L_r+x=83+1.2=84.2。$$

至計算 x 之遞差法，以使用第 75 表之差表 (Difference table) 較爲便利。例如依上例如計算 $\Delta^2 f$ 如第 75 表，取 $\Delta f$ 及 $\Delta^2 f$ 二欄之最上一數值以相比時卽得所求之 x。

第 75 表

| f | $\Delta f$ | $\Delta^2 f$ |
|---|---|---|
| 94 | | |
| | 35 | |
| 129 | | −58 |
| | −23 | |
| 106 | | |



第 104 圖

$$\therefore\quad x=-\frac{35}{-58}\times 2=1.2。$$

(2)在次數分配表內有二個相同最大次數組時之求法。

設 $f_{r-2}$，$f_{r-1}$，$f_r$，$f_{r+1}$爲鄰接四組之次數，且 $f_{r-1}$，$f_r$ 比其他各次數爲大，並有 $f_{r-1}=f_r$ 之關係時求該次數分配表之最衆數。

現以組距 a 爲單位，$f_{r-1}$ 與 $f_r$ 二組之境界點爲原點並定覆於四組上一部分曲線之方程式爲

$$y=b_1x^3+b_2x^2+b_3x+b_4$$

時，

$$f_{r-2}=\int_{-2}^{-1} ydx=\int_{-2}^{-1}(b_1x^3+b_2x^2+b_3x+b_4)dx$$

$$=\left|\frac{1}{4}b_1x^4+\frac{1}{3}b_2x^3+\frac{1}{2}b_3x^2+b_4x\right|_{-2}^{-1}$$

$$=-\frac{15}{4}b_1+\frac{7}{3}b_2-\frac{3}{2}b_3+b_4,$$

同樣得

$$f_{r-1}=-\frac{1}{4}b_1+\frac{1}{3}b_2-\frac{1}{2}b_3+b_4,$$

$$f_r=\frac{1}{4}b_1+\frac{1}{3}b_2+\frac{1}{2}b_3+b_4,$$

$$f_{r+1}=\frac{15}{4}b_1+\frac{7}{3}b_2+\frac{3}{2}b_3+b_4,$$

$$\left.\begin{aligned}
f_{r+1}-f_{r-2}&=\frac{15}{2}b_1+3b_3,\\
f_r-f_{r-1}&=\frac{1}{2}b_1+b_3,\\
f_{r+1}+f_{r-2}&=\frac{14}{3}b_2+2b_4,\\
f_r+f_{r-1}&=\frac{2}{3}b_2+2b_4,
\end{aligned}\right\}$$

解上之各式得

$$b_1=\frac{1}{6}(f_{r+1}-f_{r-2})$$

$$b_2=\frac{1}{4}\left[(f_{r+1}+f_{r-2})-(f_r+f_{r-1})\right] \quad \text{(公式 24)}$$

$$b_3=-\frac{1}{12}(f_{r+1}-f_{r-2})$$

但在 $y=b_1x^3+b_2x^2+b_3x+b_4$ 之三次拋物線內欲求 y 之極大值，須

$$\frac{dy}{dx}=3b_1x^2+2b_2x+b_3=0,$$

解此得

$$x=\frac{1}{3b_1}\left[-b_2\pm\sqrt{b_2^2-3b_1b_3}\right],$$

將此代入

$$\frac{d^2y}{dx^2}=6b_1x+2b_2$$

內，得

$$\left[\frac{d^2y}{dx^2}\right]_x=\pm2\sqrt{b_2^2-3b_1b_3},$$

故在

$$x=\frac{1}{3b_1}\left[-b_2-\sqrt{b_2^2-3b_1b_3}\right]$$

時，

$$\frac{d^2y}{dx^2}<0,$$

知該時之 y 爲最大。由此得所求之最衆數爲

$$M_0 = L_r + x \cdot a$$

(公式 25)

$$= L_r - \frac{a}{3b_1}\left(b_2 + \sqrt{b_2{}^2 - 3b_1 b_3}\right)$$

〔例 I〕

第 76 表

| 組別 | 次數 | |
|---|---|---|
| 30——40 | 2 | |
| 40——50 | 3 | |
| 50——60 | 5 | $f_{r-1}$ |
| 60——70 | 5 | $f_r$ |
| 70——80 | 2 | $f_{r+1}$ |
| 80——90 | 2 | |
| 90——100 | 1 | |
| 計 | 20 | |

由第 76 表知 $L_r = 60, f_{r-2} = 3$,

$$f_{r-1} = 5, f_r = 5, f_{r+1} = 2,$$

依公式 24 得

$$b_1 = \frac{1}{6}(2-3) = -\frac{1}{6},$$

$$b_2 = \frac{1}{4}\Big[(3+2) - (5+5)\Big] = -\frac{5}{4},$$

$$b_3 = -\frac{1}{12}(2-3) = \frac{1}{12},$$

將此等各值代入於 x 內，得

$$x=\frac{1}{3\times\left(-\frac{1}{6}\right)}\left[-\left(\frac{-5}{4}-\sqrt{\left(-\frac{5}{4}\right)^2-3\times\left(-\frac{1}{6}\times\frac{1}{12}\right)}\right)\right]$$

$$=-0.02,$$

$$\therefore\quad M_o=L_r+x.a=60+(-0.02)\times 10=59.8$$

〔例 II〕

第 77 表

| 組別 | 次數 | |
|---|---|---|
| 65——70 | 1 | |
| 70——75 | 2 | |
| 75——80 | 5 | |
| 80——85 | 17 | |
| 85——90 | 24 | |
| 90——95 | 36 | |
| 95——100 | 41 | |
| 100——105 | 59 | $f_{r-2}$ |
| 105——110 | 65 | $f_{r-1}$ |
| 110——115 | 65 | $f_r$ |
| 115——120 | 51 | $f_{r+1}$ |
| 120——125 | 40 | |
| 125——130 | 17 | |
| 130——135 | 19 | |
| 135——140 | 4 | |
| 140——145 | 2 | |

由第 77 表得

$$b_1=\frac{1}{6}(f_{r+1}-f_{r-2})=\frac{1}{6}(51-59)=\frac{4}{3},$$

$$b_2=\frac{1}{4}\left[(f_{r+1}+f_{r-2})-(f_r+f_{r-1})\right]$$

$$=\frac{1}{4}\Big[(51+59)-(65+65)\Big]=-5,$$

$$b_3=-\frac{1}{12}(f_{r+1}-f_{r-2})=-\frac{1}{12}(51-59)=\frac{2}{3},$$

$$x=\frac{1}{3b_1}\Big[-b_2-\sqrt{b_2{}^2-3b_1b_3}\Big]$$

$$=\frac{1}{4}\Big[5-\sqrt{25+\frac{8}{3}}\Big]=0.065,$$

$$M_0=L_r+x.a$$

$$=110+0.065\times5=110.33。$$

在組別分配上所求得之最衆數，並非爲確定物，故在二等大次數相並之次數分配表以求其 $M_0$ 時，得先變更分配表上組之組合，使變成爲只有一個最大次數之分配表，而後再依(1)之方法以求其最衆數。

〔例〕將上述例 II 之組距變取爲 10 時，卽將第 77 表分配情形變成第 78 表

第 78 表

| 組別 | 次數 | |
|---|---|---|
| 65—75 | 3 | |
| 75—85 | 22 | |
| 85—95 | 60 | |
| 95—105 | 100 | $f_{r-1}$ |
| 105—115 | 130 | $f_r$ |
| 115—125 | 91 | $f_{r+1}$ |
| 125—135 | 36 | |
| 135—145 | 6 | |

依第 78 表得

| f | Δf | $\Delta^2$f |
|---|---|---|
| 100 | | |
| | 30 | |
| 130 | | −69 |
| | −39 | |
| 91 | | |

由此得

$$x=-\frac{\Delta f_{r-1}}{\Delta^2 f_{r-1}}\times a=\frac{30}{69}\times 10$$

$$=4.35,$$

$$\therefore\quad M_0=L_r+x=105+4.35$$

$$=109.35。$$

其結果雖與上例不相一致，但亦相差無幾。

(E) 由累加次數曲線以求最衆數

設累加次數曲線之方程式爲

$$Y=\varphi(x)=\int^{x} f(x)dx$$

由此得

$$\frac{dY}{dx}=\frac{d\varphi(x)}{dx}=f(x),$$

$$\frac{d^2Y}{dx^2}=\frac{df(x)}{dx}=f'(x),$$

但在最衆數之 x 必能滿足

$$f'(x)=0 \quad 即 \frac{d^2Y}{dx^2}=0$$

之方程式，故知累加次數曲線上彎曲點(point of inflection)之橫座標所當值，則爲所求之最衆數。

但在實際上依累加次數曲線以求最衆數時，多使用次之手續。

依最衆數之性質，知該數前後各數值之次數實甚衆多，故在累加次數分配上，其累加次數在此數值附近之增加特別急激，依此其分布曲線在此數值之附近處，對於橫軸之傾斜角度，特別增加現應用此性質由最下部起順次於曲線上各點引以切線，其切線最初是在曲線與橫軸之間

(在末滿分布上)，過某點後其切線一躍在曲線上，改換此狀態之某點，則爲所求最乘數之所當點。依此知欲在累加次數曲線上求最乘數時，只要將定規沿曲線滑行至與曲線相交時，其交點則爲所求一點，此點之橫座標則爲所求之最乘數。

## (F) 最乘數之性質

### (甲)優點

(1)性質簡單明瞭，易於了解。

(2)不受極端值之影響。

(3)能由少數之選定樣本值，以求得正確之最乘數。

(4)常能代表實際所有之數值，使與吾人之常識相一致。

(5)得在圖形上求得之。

### (乙)劣點

(1)不能代數的運用。

(2)不能依簡單之計算以求得之。

(3)不適用於重視極端值之研究上。

(4)依組距之大小及地位而變更其數值，故難於決定確立不變之數值。

## (G) 最乘數之應用範圍

(1)在不對稱之次數分布上，常使用最乘數爲代表值。

(2)若所需要求者，爲用迅速方法決定密集度之近似數值。

(3)若所需要者，爲求得發現次數最多之數值。例如工資，結婚年齡，與死亡年齡等。

## 第五節 幾何平均數

### (A) 普通幾何平均數

將各觀察單位值之乘積上,開以單位個數次方後所得之結果值,稱爲幾何平均數,用 G 以表示之。其數式如次:

單位值: $X_1, X_2, \cdots\cdots\cdots\cdots X_n$,

$$G=\sqrt[n]{X_1 X_2 \cdots\cdots\cdots X_n}=(X_1 X_2 \cdots\cdots\cdots X_n)^{\frac{1}{n}} \quad \text{(公式26)}$$

依公式知欲求幾何平均數時, 必須經過開數次方之運算, 實甚麻繁。因此當計算時, 多使用對數以期計算之簡便。現將上式兩邊施以對數得

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

簡寫之爲 (公式 27)

$$\log G=\frac{1}{n}\Sigma \log X$$

〔例〕求 102, 121, 93 之幾何平均數。

依題算得如第 79 表

第 79 表

| X | logX |
|---|---|
| 102 | 2.00860 |
| 121 | 2.08279 |
| 93 | 1.96848 |
| 計 | 6.05987 |

依表得

$$\log G = \frac{1}{n}\Sigma \log X$$

$$= \frac{1}{3} \times 6.05987$$

$$= 2.01996,$$

依對數表得

$$G = 104.70。$$

(B) 加權幾何平均數

於各單位值上加以權度後，所求得之幾何平均數，稱爲加權幾何平均數(weighted geometric mean)現取

$$f_1,\ f_2,\ \cdots\cdots\cdots\cdots f_m,$$

爲各單位值

$$X_1,\ X_2,\ \cdots\cdots\cdots\cdots X_m,$$

之權度，$G_W$ 爲加權幾何平均數時，

$$G_w = \sqrt[(f_1+f_2+\cdots\cdots+f_n)]{X_1^{f_1} . X_2^{f_2} \cdots\cdots\cdots X_m^{f_m}}$$

(公式 28)

$$= (X_1^{f_1} X_2^{f_2} \cdots\cdots\cdots X_m^{f_m})^{\frac{1}{f_1+f_2+\cdots\cdots+f_m}}$$

施對數於本式之兩端，得

$$\log G_w = \frac{1}{f_1+f_2+\cdots\cdots+f_m}(f_1 \log X_1 + f_2 \log X_2 + \cdots\cdots + f_m \log X_m),$$

即 $$\log G_w = \frac{1}{n}\Sigma f \log X$$ (公式 29)

〔例〕

在計算物價指數時，常於各物品上依其需要情形加以權度，如第 80 表爲穀類物價指數之計算情形。

第 80 表

| 物品 | 權度 f | 各物品價格之比例 X | log X | f log X |
|---|---|---|---|---|
| 本國米 | 65 | 1.93 | 0.28556 | 18.56140 |
| 外國米 | 4 | 1.58 | 0.19866 | 0.79464 |
| 大麥 | 5 | 1.68 | 0.22531 | 1.12655 |
| 小麥 | 9 | 1.50 | 0.17609 | 1.58481 |
| 外國小麥 | 2 | 1.70 | 0.23015 | 0.46090 |
| 計 | 85 | | | 22.52830 |

依表知 $N=85$，$\Sigma f \log X = 22.52830$，

$$\therefore \log G_w = \frac{1}{N}\Sigma f \log X$$

$$= \frac{1}{85}\times 22.52830$$

$$= 0.26504,$$

$$\therefore G_w = 1.841。$$

知穀類之物價指數爲 $1.84\times100=184$。

(C) 幾何平均數之性質

(甲)優點

(1)關係於觀察數列之全體。

(2)關於極大值及極小值之影響，較算術平均數爲少。

(3)能代數的運用

(a)全體之幾何平均數得由部分平均數以表示之。

〔證〕

$X_1, X_2 \cdots\cdots\cdots\cdots X_k$ 之幾何平均數爲 $G_x$,

$Y_1, Y_2 \cdots\cdots\cdots\cdots Y_l$ 之幾何平均數爲 $G_y$,

$X_1, X_2 \cdots\cdots\cdots\cdots X_k, Y_1, Y_2 \cdots\cdots\cdots\cdots Y_l$ 之幾何平均數爲 G, 並取 $k+l=n$ 時,

$$G^n = X_1 \cdot X_2 \cdots\cdots X_k \cdot Y_1 Y_2 \cdots\cdots Y_l = G_X^k G_Y^l,$$

施以對數得

$$n \log G = k \log G_X + l \log G_Y,$$

$$\therefore \log G = \frac{1}{N}(k \log G_x + l \log G_Y)。$$

一般

$$\log G = \frac{1}{N} \sum_r N_r \log G_r \qquad (公式 30)$$

(b)二數列相互對應時,其對應數之比的幾何平均數等於各部分幾何平均數之比。

〔證〕

$X_1, X_2, X_3$ 之幾何平均數爲

$$G_X = \sqrt[3]{X_1 X_2 X_3},$$

$Y_1, Y_2, Y_3$ 之幾何平均數爲

$$G_Y = \sqrt[3]{Y_1 Y_2 Y_3}$$

$\frac{X_1}{Y_1}, \frac{X_2}{Y_2}, \frac{X_3}{Y_3}$ 之幾何平均數爲

$$G = \sqrt[3]{\frac{X_1}{Y_1} \cdot \frac{X_2}{Y_2} \cdot \frac{X_3}{Y_3}}$$

$$=\frac{\sqrt[3]{X_1\cdot X_2\cdot X_3}}{\sqrt[3]{Y_1Y_2Y_3}}=\frac{G_X}{G_Y}$$

$$\therefore\quad G=\frac{G_X}{G_Y}\qquad\text{(公式31)}$$

〔例〕三種商品A,B,C之今昨二年之比率如第81表之第三欄內各數值，在欲計算此等比率之幾何平均數時，毋庸計算此等比值，只要已知各年之幾何平均數 $G_X$, $G_Y$ 時卽能依公式31求得比值之幾何平均數爲

$$G=\frac{G_x}{G_y}=\frac{103.92}{92.15}=112.77。$$

第 81 表

| | 今年 (X) | 去年 (Y) | 今年與去年之比 |
|---|---|---|---|
| 商品A | 110 | 103 | 110/103 |
| 商品B | 92 | 80 | 92/80 |
| 商品C | 88 | 95 | 88/95 |
| 幾何平均數 | $G_X=103.92$ | $G_Y=92.15$ | $G=112.77$ |

(c)數種同次數(n)之數列各單位值相乘，結果一數列之幾何平均數等於各列本身之幾何平均數之積。

〔證〕

$X_1, X_2, \cdots\cdots X_n$ 之幾何平均數爲

$$G_X=\sqrt[n]{X_1\cdot X_2\cdots\cdots X_n},$$

$Y_1Y_2\cdots\cdots Y_n$ 之幾何平均數爲

$$G_Y=\sqrt[n]{Y_1\cdot Y_2\cdots\cdots Y_n},$$

$Z_1, Z_2, \cdots\cdots Z_n$ 之幾何平均數爲

$$G_Z=\sqrt[n]{Z_1\cdot Z_2\cdots\cdots Z_n}$$

$$G_XG_YG_Z=\sqrt[n]{(X_1X_2\cdots\cdots X_n)(Y_1Y_2\cdots\cdots Y_n)(Z_1Z_2\cdots\cdots Z_n)}=G$$

$$\therefore\quad G=G_XG_YG_Z$$

一般

$$G=G_1 \cdot G_2 \cdot G \cdots\cdots G_m。 \quad \text{(公式 32)}$$

(乙)劣點

(1)數列內有零或負之單位值時,不能求其幾何平均數。

(2)所求得之結果常爲實際所無之數值。

(3)計算複雜爲普通人所難以應用。

(D) 幾何平均數之應用範圍

本平均數對於變動比率之平均上有特殊之効用,例如在

(1)重視變化率之變動。

(2)計算依複利的增加或減少諸現象(利息,人口增加等)之平均增加率。

例如人口平均增加率之

$$1+r=\sqrt[n]{\frac{P_n}{P_0}}=\sqrt[n]{\frac{P_n}{P_{n-1}}\cdot\frac{P_{n-1}}{P_{n-2}}\cdots\cdots\frac{P_2}{P_1}\cdot\frac{P_1}{P_0}},$$

由此知 $1+r$ 一數值實爲 $\frac{P_n}{P_{n-1}}, \frac{P_{n-1}}{P_{n-2}}, \cdots\cdots \frac{P_2}{P_1}, \frac{P_1}{P_0}$ 等之幾何平均數。

(3)物價指數,等之計算上爲不可缺物。

總之在測定含有動的性質之現象的變化上,實爲最適宜之平均數

## 第六節　調和平均數

(A) 普通調和平均數

依各觀察單位值之逆數以求出其算術平均數,此算術平均數之逆

數稱爲調和平均數，用 H 以表示之。其數式如次：

單位值： $X_1, X_2, \cdots\cdots X_n$

$$\frac{1}{H}=\frac{1}{n}\left(\frac{1}{X_1}+\frac{1}{X_2}+\cdots\cdots+\frac{1}{X_n}\right),$$

$$\therefore\ H=\frac{n}{\frac{1}{X_1}+\frac{1}{X_2}+\cdots\cdots\cdots+\frac{1}{X_n}}$$

簡寫爲 (公式 33)

$$H=\frac{n}{\Sigma\frac{1}{X}}$$

〔例〕求 2,3,4,5,6 之調和平均數。

依公式 33，得

$$H=\frac{5}{\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\frac{1}{5}+\frac{1}{6}}$$

$$=\frac{300}{87}$$

$$=3.45\cdots\cdots\cdots\cdots$$

(B) 加權調和平均數

調和平均數亦能如幾何平均數然，有加權調和平均數 (weighted harmonic mean)，用 $H_w$ 以表示之。其數式如次：

單位值： $X_1, X_2, \cdots\cdots\cdots\cdots X_m$,

權 度： $f_1, f_2, \cdots\cdots\cdots\cdots f_m$,

$$\frac{1}{H_w}=\frac{1}{\Sigma f}\left(\frac{f_1}{X_1}+\frac{f_2}{X_2}+\cdots\cdots+\frac{f_m}{X_m}\right),$$

$$\therefore \quad H_w = \frac{\Sigma f}{\Sigma \frac{f}{X}} \qquad \text{(公式 34)}$$

在計算調和平均數時，以使用逆數表較為便利。

〔例〕

第 82 表

| (1) X | (2) f | (3) $\frac{1}{X}$ | (4) $f\frac{1}{X}$ |
|---|---|---|---|
| 1.60 | 3 | 0.625 | 1.875 |
| 1.80 | 8 | 0.556 | 4.448 |
| 2.00 | 13 | 0.500 | 6.500 |
| 2.20 | 20 | 0.455 | 9.100 |
| 2.40 | 7 | 0.417 | 2.919 |
| 2.60 | 5 | 0.385 | 1.925 |
| 2.80 | 1 | 0.357 | 0.357 |
| 3.00 | 4 | 0.333 | 1.332 |
| 3.20 | 2 | 0.313 | 0.626 |
| 3.40 | 1 | 0.294 | 0.294 |
| 3.60 | 1 | 0.278 | 0.278 |
| 計 | 65 | | 29.564 |

依公式 34，得

$$H_w = \frac{\Sigma f}{\Sigma \frac{f}{X}}$$

$$= \frac{65}{29.564}$$

$$= 2.16。$$

(C) 調和平均數之性質

(甲)優點

(1)能代數的運用。

(2)得用於時間之分配及相似事實之平均上，故為經濟統計界計算平均價格上之必需品，

(乙)劣點

(1)計算麻繁。

(2)為普通所難以理解。

(D) 調和平均數之應用範圍

本平均數適用於時間變化率，及對於某種一定單位之變量的平均上。

〔例 I〕某汽車依每時$x_1, x_2, \cdots\cdots x_n$之速度各行S哩，求其平均速度。

在此等問題上，若取$x_1, x_2, \cdots\cdots x_n$之算術平均數

$\frac{x_1+x_2\cdots\cdots x_n}{n}$，尚不能說是平均速度，蓋因平均速度必須用全時間去除全路程而後始可。

$$平均速度=\frac{nS}{\frac{S}{x_1}+\frac{S}{x_2}+\cdots\cdots+\frac{S}{x_n}}$$

$$=\frac{n}{\frac{1}{x_1}+\frac{1}{x_2}+\cdots\cdots+\frac{1}{x_n}}$$

由此知$x_1, x_2, \cdots\cdots x_n$之調和平均數，即為所求之平均速度。

〔例 II〕有n種商品，甲種每元能購$P_1$個，乙種每元能購$P_2$個，……求此等商品之平均價格。

若取$P_1, P_2, \cdots\cdots P_n$之算術平均數當作為平均價格時，則商店須受

損失。故必須由

$$\frac{n}{\frac{1}{P_1}+\frac{1}{P_2}+\cdots\cdots+\frac{1}{P_n}}$$

所求得者，卽爲$P_1,P_2,\cdots\cdots P_n$之調和平均數，始能作爲平均每元所能購買數量。

## 第七節　其他之平均數

上述五種平均數爲一般所最多使用者，此外尙有次之數種。

### (A)總和平均數

在某種現象之二數列上，其各數列單位值各自總和後之相比値，稱爲總和平均數(summary mean)，用 $M_S$ 以表示之。其數式如次：

甲數列：$X_1,X_2,\cdots\cdots\cdots\cdots X_n$,

乙數列：$Y_1,Y_2,\cdots\cdots\cdots\cdots Y_n$,

$$M_S=\frac{X_1+X_2+\cdots\cdots\cdots+X_n}{Y_1+Y_2+\cdots\cdots\cdots+Y_n},$$

簡寫爲　(公式 35)

$$M_S=\frac{\Sigma X}{\Sigma Y}。$$

### (B)自乘平均數

開平方於各單位自乘之算術平均數上所得結果値，稱爲自乘平均數(quadratic mean)，用 $M_Q$ 以表示之。其數式如次：

單位值：$X_1,X_2,\cdots\cdots\cdots\cdots X_n$,

$$M_Q=\sqrt{\frac{1}{n}(X_1{}^2+X_2{}^2+\cdots\cdots\cdots+X_n^2)},$$

簡寫為 (公式 36)

$$M_Q=\sqrt{\frac{1}{n}uX^2}。$$

(C)反調和平均數

用數列各單位值之和去除原數列各單位值自乘之和後所得結果值，稱為反調和平均數 (contra-harmonic mean)，以 Hc 以表示之。其數式如次：

單位值： $X_1, X_2, \cdots\cdots\cdots\cdots X_n$,

$$H_c=\frac{X_1^2+X^2+\cdots\cdots\cdots+X_n^2}{X_1+X_2+\cdots\cdots\cdots+X_n},$$

簡寫為 (公式 37)

$$M_c=\frac{\Sigma X^2}{\Sigma X}。$$

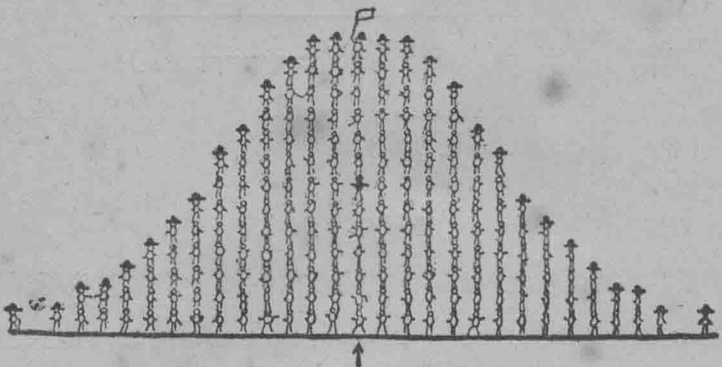
## 第八節 各平均數間之關係

由同一數列所作成各種平均數間，實有次之諸關係。

(1)次數分配為完全對稱時，其算術平均數 M 中位數 Mi 及最衆數 $M_0$ 三者互相一致。

$$M=M_i=M_0,$$

第 105 圖為完全對稱分配情形，連接各人帽子所在處得一對稱次數曲線，其最高處所插旗是表示最衆數所在位置，完全塗黑一人表示中位數所在處，又下方矢之所指處，表示算術平均數之所當位置，依圖知三者全在一垂直線上，故知三者所當品質值均相等。



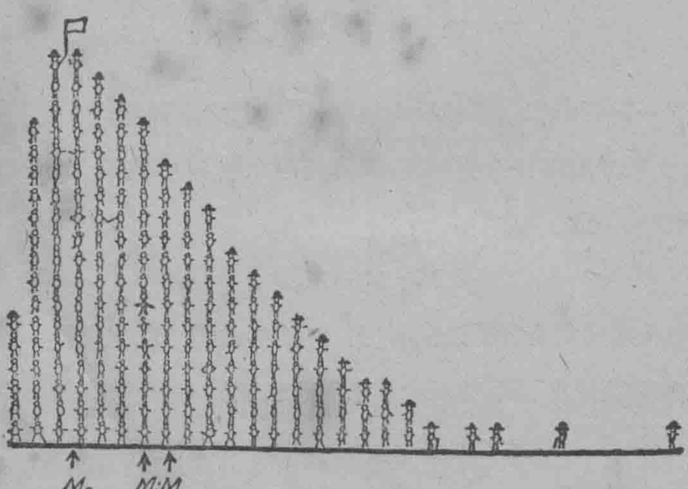
第 105 圖

(2)次數分配成爲不對稱時，其中位數。是在算術平均數與最衆數之間。其排列情形如次：

(a)在左傾不對稱分配上，

$$M>Mi>M_0,$$

其詳情參照第 106 圖。




第 106 圖

(b)在右傾不對稱分布上其情形適相反。

$$M<Mi<M_0$$

(3)由同一數列所算出之算術平均數 M 幾何平均數 G 調和平均數 H 間有次之關係，

$$M>G>H。\qquad (公式 38)$$

〔證〕現只就二個單位值 $X_1, X_2$ 以行證明，其一般之證明由閱者自行推廣之。

(a)求 $M>G$ 之證明，

由 $M=\frac{1}{2}(X_1+X_2), G=\sqrt{X_1 \cdot X_2}$，

得

$$M-G=\frac{1}{2}(X_1+X_2)-\sqrt{X_1X_2}$$

$$=\frac{1}{2}(X_1+X_2-2\sqrt{X_1X_2})$$

$$=\frac{1}{2}(\sqrt{X_1}-\sqrt{X_2})^2。$$

只要 $X_1, X_2$ 爲二個不同正數值時，上式右邊必爲正，

$$\therefore\ M-G>0，即\ M>G。$$

(b)求 $G>H$ 之證明，

$$由\ G=\sqrt{X_1X_2},\ H=\frac{2X_1X_2}{X_1+X_2},$$

得

$$G-H=\sqrt{X_1X_2}-\frac{2X_1X_2}{X_1+X_2}$$

$$=\sqrt{X_1X_2}\frac{X_1+X_2-2\sqrt{X_1X_2}}{X_1+X_2}$$

$$=\frac{\sqrt{X_1X_2}}{X_1+X_2}(\sqrt{X_1}-\sqrt{X_2})^2,$$

只要 $X_1, X_2$ 爲二個不同正數值時，上式右邊必爲正。

$$\therefore \quad G-H>0, \text{即} \quad G>H$$

由(a)，(b)得

$$M>G>H$$

(4) 二單位值之幾何平均數等於其算術平均數與調和平均數之幾何平均數，

$$G=\sqrt{M \cdot H} \qquad \text{(公式 39)}$$

本關係只限於二個單位值，以上卽不成立。

# 第八章　差量

## 第一節　差量之意義

在研究次數分配狀態上，雖有平均數得以決定集團之中心是在何處，但只此平均數實尚覺不足。參考次之二例，卽能明白了解。

〔例 I〕 A,B 二數列之實際排列情形如次：

A：6, 6, 7, 7, 7, 8, 8, 8, 8, 8, 9, 9, 9, 10, 10,

B：1, 1, 2, 3, 3, 5, 5, 8, 8, 8, 9, 9, 11, 22, 25,

由 A,B 二數列計算得

$$\text{A 之}\begin{cases} M=8 \\ M_i=8 \\ M_0=8 \\ n=15 \end{cases} \qquad \text{B 之}\begin{cases} M=8 \\ M_i=8 \\ M_0=8 \\ n=15 \end{cases}$$

其平均數與個數雖均相等，但其最大單位値與最小單位値之差，則 A 為 5 而 B 為 25，相差幾至五倍。由此知 A 之各個單位値與平均數相差無幾，是密集於平均數之左右。反之，B 數列上各單位値與平均數相差實相當巨大，是分散於平均數之附近。

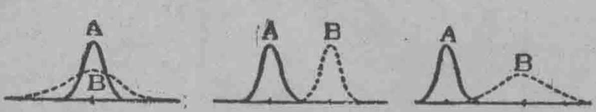
〔例 II〕甲乙二人一週內每日收入情形如第 83 表：

第 83 表

| | 星期一 | 星期二 | 星期三 | 星期四 | 星期五 | 星期六 | 總 計 | 平 均 |
|---|---|---|---|---|---|---|---|---|
| | 元 | 元 | 元 | 元 | 元 | 元 | 元 | 元 |
| 甲 | 2.1 | 2.1 | 2.0 | 1.9 | 1.9 | 2.0 | 12.0 | 2.0 |
| 乙 | 2.5 | 2.9 | 0.7 | 2.0 | 2.0 | 1.9 | 12.0 | 2.0 |

依表雖能知甲乙二人平均每日收入皆爲二元，但只此平均數尙未能卽刻斷定二人之收入爲完全相等。蓋因二人之每日收入有所不同，依表知甲之收入前後無甚差異，每日均在平均數左右。反之，乙之收入是有時甚多，有時甚少。

觀上述二例，知欲研究分配狀態時，除平均數外，尙須明白該集團中各單位値間相差程度之必要。此相差程度吾人稱之爲差量 (dispersion, streunngsmasse, verteilungsmasse)，與平均數完全不同，在第 107 圖內左圖之 AB 二分配曲線其平均數雖相等，但單位値間之差量




第 107 圖

則不同，是 A 比 B 爲小。又中圖之 AB 二分配曲線其差量雖相等，但平均數則不同。右圖則 AB 二分配曲線之差量與平均數均完全不相等。

至差量本身分爲絕對的差量 (absolutedispersion) 與相對的差量 (relative dispersion) 二種。其測定方法在絕對的差量上有

(1) 全距 (range)

(2) 四分位差 (quartile deviation, quartile)

(3)平均差(mean deviation, durchschnittliche abweichung)

(4)標準差(standard deviation, standard abweichung)

(5)其他之差量，

在比較的差量上有

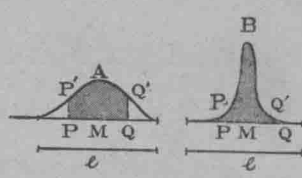
(1)變化係差(coefficient of variation, der variabilitatskoeffizient) 等，以下依次詳述。

## 第二節　全距

測定差量中爲吾人所最易引起注意者，是單位值分配之全距 ($R_g$)，卽爲單位值內之最大單位值($X_{max}$)與最小單位值($X_{min}$)之差。其數學公式如次：

$$R_g = X_{max.} - X_{min.} \quad \text{(公式 40)}$$

依此差數之大小以考察差量一事，雖得爲方法之一。但在實際上有多數數列，其大部分之單位值，是會集於平均數之附近，對於此點亦不得不注意之。例如第 108 圖內 A,B 二次數分配曲線之全距均等於 1，但在平均數附近狀況，則 B 之諸單位值較 A 之諸單位值爲密集於平均數之附近，對此只要在 A,B 二曲線上各以其平均數 M 爲中點取定長 PQ，在此 PQ 上之面積 PP'QQ' 內之次數，依圖知 A 之分配上在此面積外之次數相當衆多，而在 B 之分配上幾將全部次數含在此面積內，依此知 A 之分配上尙有多數單位值散布於 PQ 範圍外而在 B 方則大部分密集於平均數附近，故知 A 之



第 108 圖

差量實較 B 爲小，與由全距所定情形完全不相同。且此全距大受極大值(極小值)之影響，成爲不安定之數量。故知只依全距以決定差量，有時實甚危險，對此宜注意之。

## 第三節　四分位差

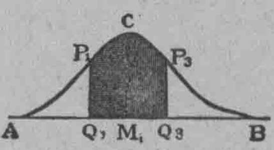
### (A)意義及其計數公式

將觀察單位值依大小順序排列後依中位數 $M_i$ 分爲二部分。其第一部分之中位數稱爲下四分位數(lower quartile)，用 $Q_1$ 以表示之。第二部分之中位數稱爲上四分位數 (upper quartile)，用 $Q_3$ 以表示之。如此數列全體被 $Q_1$, $M_i$, $Q_3$ 三值等分爲四部分。其情形如第 109 圖，

卽面積 $AQ_1P_1 = P_1Q_1M_iC$

$= CM_iQ_3P_3$

$= P_3Q_3B$。




第 109 圖

知單位值全體之半數，在 $Q_1$ $Q_3$ 之間。由此知依距離 $\overline{Q_1Q_3}$ 之大小卽能測定差量之大小。

在次數分配完全爲對稱時，中位數與下四分位數之差，適等於上四分位數與中位數之差。其公式爲

$$M_i - Q_1 = Q_3 - M_i。$$

但在其他之分布上則此二差數必不相等，依此特定 $M_i - Q_1$ 與 $Q_3 - M_i$ 之和之半爲一差量，此差量稱之爲四分差(quartile deviation)，用 Q 以表示之。

$$Q=\frac{1}{2}\{(M_i-Q_1)+(Q_3-M_i)\}=\frac{1}{2}(Q_3-Q_1)\quad\text{(公式 41)}$$

依上式知 $Q$ 爲 $Q_3$ 與 $Q_1$ 間距離之半,故或稱 $Q$ 爲 semi-interquartile range。

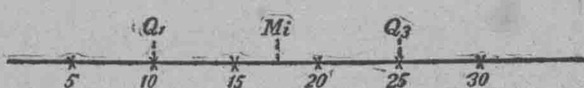
〔例 I〕 求數列 5, 10, 15, 20, 25, 30 之四分位差。

先由中位數 $M_i$ 分爲二部分,

第一部分 5, 10, 15, 之中位數,即下四分位數,爲 $Q_1=10$,

第二部分 20, 25, 30, 之中位數,即上四分位數爲 $Q_3=25$,

$$\text{四分位差}\ Q=\frac{Q_3-Q_1}{2}=\frac{25-10}{2}=7.5。$$



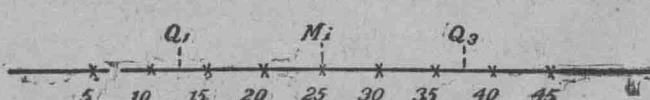

〔例 II〕 求數列 5, 10, 15, 20, 25, 30, 35, 40, 45 之四分位差。

先由中位數 25 分爲二部分,

第一部分 5, 10, 15, 20 之中位數即 $Q_1=\frac{10+15}{2}=12.5$,

第二部分 30, 35, 40, 45 之中位數即 $Q_3=\frac{35+40}{2}=37.5$,

$$\text{四分位差}\ Q=\frac{Q_3-Q_1}{2}=\frac{37.5-12.5}{2}=12.5。$$




至其實際計算方法,得分爲次之二種以行說明。

(1)單位值各別分布時之求法

設單位值 $X_1, X_2, \cdots X_n$，是已由小至大順次排列妥當，其個數 n 不管其為偶數$(n=2m)$或奇數$(n=2m+1)$，凡比中位數為大或為小之單位值個數，均為 m 個，此

m 為奇數$(m=2k+1)$時，

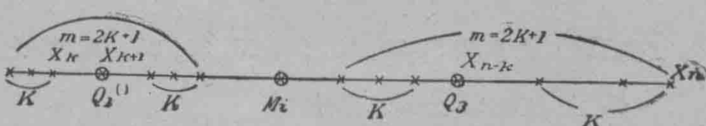
$$Q_1 = X_{k+1}, \qquad Q_3 = X_{n-k} \text{。}$$

m 為偶數$(m=2k)$時，

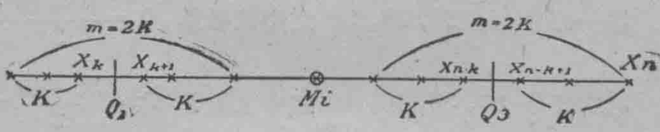
$$Q_1 = \frac{X_k + X_{k+1}}{2}, \qquad Q_3 = \frac{X_{n-k} + X_{n-k-1}}{2}$$

現將其取法圖示如第 110 圖

$m=2k+1$ 時




$m=2k$ 時




第 110 圖

〔例 I〕 求 2，5，8，10，13，18，20，30，之四分位差。

$$\begin{array}{cc|cc|cc|cc} & & Q_1 & & M_i & & Q_3 & \\ 2 & 5 & 8 & 10 & 13 & 18 & 20 & 30 \end{array}$$

依此知

$$Q_1=\frac{5+8}{2}=6.5,\qquad Q_3=\frac{18+20}{2}=19,$$

$$Q=\frac{1}{2}(Q_3-Q_1)=\frac{19=6.5}{2}=6.25。$$

〔例 II〕

$$\begin{array}{ccccccccccc} 18, & 25, & 30, & 41, & 51, & 56, & 80, & 82, & 92, & 93, & 103, \\ & & \| & & & \| & & & \| & & \\ & & Q_1 & & & M_i & & & Q_3 & & \end{array}$$

$$Q=\frac{Q_3-Q_1}{2}=\frac{92-30}{2}=31。$$

(2)單位值依次數分配以給與時之求法。

第　84　表

| 組　　別 | 次　　數 |
|---|---|
| $L_1$——$u_1$<br>$L_{i-1}$——$u_{i-1}$ | $\left.\begin{matrix} f_1 \\ f_{i-1} \end{matrix}\right\} F_1$ |
| $Q_1 \rightarrow L_i$——$u_i$ | $f_i$ |
| $Q_2 \rightarrow L_j$——$u_j$ | $f_j$ |
| $L_{j+1}$——$u_{j+1}$<br>$L_m$——$u_m$ | $\left.\begin{matrix} f_{j+1} \\ f_m \end{matrix}\right\} F_2$ |
| 計 | N |

依第 84 表得

$$f_1+f_2+\cdots+f_{i-1}=F_1,$$

$$並\ F_1<\frac{N}{4}<F_1+f_i$$

時知 $Q_1$ 所當值在第 i 組內。

現定第 i 組之下界限爲 $L_i$，組距爲 a 時，依一次補間方法得

$$Q_1=L_i+a\frac{\frac{N}{4}-F_1}{f_i} \qquad \text{(公式 42)}$$

又 $$f_{j+1}+f_{j+2}+\cdots+f_m=F_2,$$

並 $$F_2+f_j>\frac{N}{4}>F_2$$

時，知 $Q_3$ 在第 j 組內。故當第 j 組之上界限爲 $u_j$ 時仍依一次補間法得

$$Q_3=u_j-a\frac{\frac{N}{4}-F_2}{f_j} \qquad \text{(公式 43)}$$

由此求得 $Q_1, Q_3$ 後，即能依公式 41

$$Q=\frac{1}{2}(Q_3-Q_1)$$

以求其四分位差。

〔例 I〕 第 85 表

| 組別 | 次數 |
|---|---|
| 40——50 | 3 $=F_1$ |
| 50——60 | 19 $=f_i$ |
| 60——70 | 30 |
| 70——80 | 15 $=f_j$ |
| 80——90 | 2 } $=F_2$ |
| 90——100 | 1 } |
| 計 | 70$=$N |

由第 85 表得 $\frac{N}{4}=17.5$，$F_1=3$，$F_2=3$，

$$a=10, \quad L_i=50, u_j=80,$$

$$Q_1 = L_i + a\frac{\frac{N}{4} - F_1}{f_i}$$

$$= 50 + 10 \times \frac{17.5 - 3}{19} = 57.63,$$

$$Q_3 = u_j - a\frac{\frac{N}{4} - F_2}{f_j} = 80 - 10 \times \frac{17.5 - 3}{15} = 70.33,$$

$$\therefore\ Q = \frac{1}{2}(Q_3 - Q_1) = \frac{70.33 - 57.63}{2} = 6.35。$$

〔例 II〕 某校入學考試成績如第 86 表，求其四分位差。

第　86　表

| 分　數 | 人　數 | |
|---|---|---|
| 0—5 | 2 | |
| 5—10 | 3 | |
| 10—15 | 3 | |
| 15—20 | 9 | $F_1=41$ |
| 20—25 | 2 | |
| 25—30 | 7 | |
| 30—35 | 15 | |
| 35—40 | $44=f_i$ | |
| 40—45 | 21 | |
| 45—50 | 45 | |
| 50—55 | 13 | |
| 55—60 | 27 | |
| 60—65 | 12 | |
| 65—70 | $26=f_j$ | |
| 70—75 | 13 | |
| 75—80 | 24 | |
| 80—85 | 4 | |
| 85—90 | 9 | $F_2=61$ |
| 90—95 | 3 | |
| 95—100 | 8 | |
| 計 | $300=N$ | |

依表知 $\frac{N}{4}=75,\ F_1=41,\ f_j=44,$

$$a=5,\quad F_2=61,\ f_j=26,$$

$$L_i=35,\ u_j=70,$$

依公式得 $$Q_1=L_i+a\,\frac{\frac{N}{4}-F_1}{f_i}$$

$$=35+5\times\frac{75-41}{44}=37.727,$$

$$Q_3=u_j-a\,\frac{\frac{N}{4}-F_2}{f_j}$$

$$=70-5\times\frac{75-61}{26}=67.308$$

$$\therefore\quad Q=\frac{1}{2}(Q_3-Q_1)=\frac{1}{2}(67.308-37.727)$$

$$=14.79。$$

(B)四分位差之性質

(甲)優點

(1)簡單明瞭,易於了解。

(2)依單位值之大小順序排列後,其兩邊極端項所屬數值雖不加入在內,亦能簡單算得其四分位差。

(3)與標準差有密切之關係,在對稱的分配上四分差適等於標準差之$\frac{2}{3}$。

即
$$\frac{Q}{\sigma}=\frac{2}{3} \quad \text{或} \quad 9Q=6\sigma \qquad (\text{公式 } 44)$$

由此知 9Q 之範圍內，含有總次數之 99%。

(乙)劣點

(1)不能代數的運用。

(2)不關係於全體觀察單位值上。

(3)本差量之計算方法過於粗草，故其信賴度較爲薄弱，但在經濟統計等研究上，仍多使用之。

(C)十分位數及百分位數

現乘說明四分位差之便，附加說明十分位數(deciles)之意義。

將單位值之全體依九個數值

$$D_1,\ D_2,\ D_3,\ D_4,\ D_5,\ D_6,\ D_7,\ D_8,\ D_9,$$

以十等分之，即

比 $D_1$ 爲小各單位值之個數

=比 $D_1$ 爲大比 $D_2$ 爲小各單位值之個數

=比 $D_2$ 爲大比 $D_3$ 爲小各單位值之個數

= ……………………………………………

=比 $D_9$ 爲大各單位值之個數

=總次數之 $\frac{1}{10}$。

此九個數值稱爲十分位數，其中之 $D_5=M_i$。

依同樣方法得求出百分位數(percentiles)。至由次數分配表以求

百分位數之公式如次:

第 87 表

| 組別 | 次數 |
|---|---|
| $L_1$——$u_1$ | $\left.\begin{matrix} f_1 \\ f_{i-1} \end{matrix}\right\} F$ |
| $L_{i-1}$——$u_{i-1}$ | |
| $P_p \rightarrow L_i$——$u_i$ | $f_i$ |
| $L_{i+1}$——$u_{i+1}$ | $\left.\begin{matrix} f_{i+1} \\ f_m \end{matrix}\right\} F'$ |
| $L_m$——$u_m$ | |
| 計 | N |

設 $N$＝總次數,

$P_p$＝百分位數,

$$P=\frac{P_p}{100},$$

$f_i$＝含有 $P_p$ 一組之次數,

$Q$＝組距,

$F$＝ $f_i$ 以下各次數之和,

$F'$＝ $f_i$ 以上各次數之和,

$L_i$＝含有 $P_p$ 一組之下限值,

$u_i$＝含有 $P_p$ 一組之上限值,

先由

$$F<PN<F+f_i$$

以定 $f_i$ 之所在,次依一次補間公式求得

$$P_p=L_i-\frac{PN-F}{f_i}\cdot a \qquad \text{(公式 45)}$$

若由他端(最大端)以求百分位數之公式爲

$$P_P = u_i - \frac{(1-P)N - F'}{f_i} a \qquad (公式\ 46)$$

上述之中位數四分位數及十分位數等均得包含在此百分位數之內。

其

$$M_i = P_{.50},$$

$$Q_1 = P_{.25}, \quad Q_3 = P_{.75},$$

$$D_1 = P_{.10}, \quad D_2 = P_{.20}, \cdots\cdots D_9 = P_{.90}。$$

〔例〕 依第 86 表之次數分配,以求 $P_{20}$。

依題得

$$PN = 300 \times .20 = 60,$$

$$41 < 60 < 41 + 44,$$

故知

$$f_i = 44, \quad L_i = 35, \quad u_i = 40,$$

$$a = 5, \quad F = 41, \quad F' = 215,$$

依公式得

$$P_{.20} = L_i + \frac{PN - F}{f_i} \cdot a$$

$$= 35 + \frac{60 - 41}{44} \times 5$$

$$= 37.16;$$

若由最大端求起依公式得

$$P_{\cdot 20}=u_i-\frac{(1-P)N-F'}{f_i}\cdot a$$

$$=40-\frac{240-215}{44}\times 5$$

$$=37.16,$$

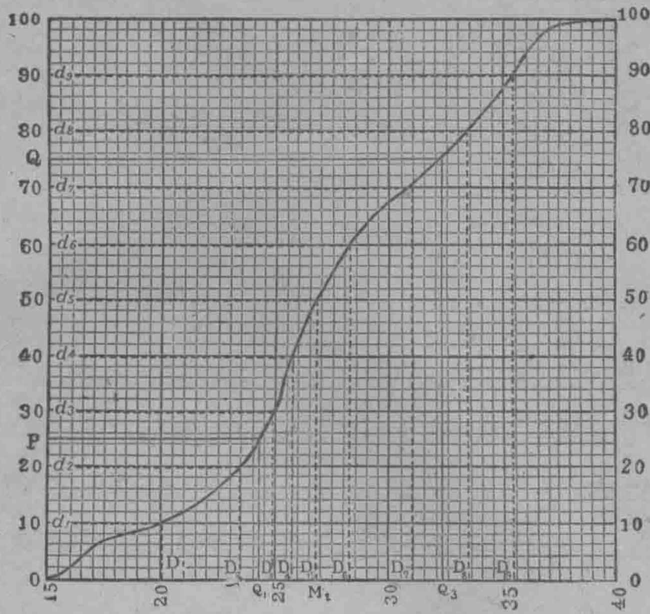
二者完全一致。

此等四分位數十分位數及百分位數均得由累加次數曲線以求得之。

〔例〕 第 88 表為 1885 年英國勞工工資(每週先令數)分配情形，由此作成累加次數曲線如第 111 圖。

第 88 表

| 工 資 | 人數% | 累積人員 | 工 資 | 人數% | 累積人員 | 工 資 | 人數% | 累積人員 |
|---|---|---|---|---|---|---|---|---|
| 15—16 | 2 | 2 | 23—24 | 5 | 23 | 31—32 | 4 | 74 |
| 16—17 | 4 | 6 | 24—25 | 9 | 32 | 32—33 | 4 | 78 |
| 17—18 | 1 | 7 | 25—26 | 12 | 44 | 33—34 | 5 | 83 |
| 18—19 | 1 | 8 | 26—27 | 8 | 52 | 34—35 | 5 | 88 |
| 19—20 | 2 | 10 | 27—28 | 7 | 59 | 35—36 | 6 | 94 |
| 20—21 | 2 | 12 | 28—29 | 5 | 64 | 36—37 | 4 | 98 |
| 21—22 | 3 | 15 | 29—30 | 3 | 67 | 37—38 | 1 | 99 |
| 22—23 | 3 | 18 | 30—31 | 3 | 70 | 38—39 | 0.8 | 99.8 |
| | | | | | | 39—40 | 0.2 | 100.0 |




第 111 圖

現將累加人數 100 人,依十等分得

$d_1=10$, $d_2=20$, …… $d_9=90$,

由此等縱軸上之相當點處引與橫軸平行之直線使與曲線相交後,再由此等交點引垂直線與橫軸相交各點之所當值,則為所求之十分位數。依圖得

$D_1=20$, $D_2=23.5$, $D_3=24.9$,

$D_4=25.7$, $D_5=M_i=26\cdot8$ $D_6=28.2$,

$D_7=31$, $D_8=33.4$, $D_9=35.4$。

至四分位差之求法,是先在縱軸上取 25 及 75 之所當點 P,Q 後,

由此二點依上述方法得

$$Q_1=24.2, \qquad Q_3=32,$$

$$\therefore \quad Q=\frac{1}{2}(Q_3-Q_1)=\frac{32-24.2}{2}=3.9。$$

## 第四節 平均差

### (A)意義及其計算公式

前述之四分位差只由 $\overline{Q_1\,Q_3}$ 之長以定其差量而已，對此 $Q_1$ $Q_3$ 間所有各單位值與平均數相差之距離，尚未詳細加入，故在只知大體之差量時，上法尚能使用。如欲知精密差量，則必須將各單位值與平均數相差之距離加入在內。本節之平均差及次節之標準差，皆為測定精密差量之方法。

設各單位值為 $X_1$, $X_2$, ……$X_m$,

次數為 $f_1$, $f_2$, …… $f_m$,

其中位數為 $M_i$ 時，各單位值與 $M_i$ 相差之距離為

$$X_1-M_i, \quad X_2-M_i, \quad \cdots\cdots \quad X_m-M_i,$$

此等差數稱為由中位數所起各單位值之偏差(deviation from $M_i$)。依此知單位值比中位數為大者，其偏差為正。比中位數為小者，其偏差為負。但在專為測量各單位值與中位數相差程度起見，可勿必注意其正負號，只取各偏差之絕對值已足。現由此等偏差之絕對值，求其算術平均數，依此平均數之大小，以定其差量。如此所得之平均數，稱為平均差，用 Δ (delta)以表示之。其數式如次:

取各單位值與中位數相差距離之絕對值爲

$$|X_1-M_i|,\quad |X_2-M_i|,\cdots\cdots|X_m-M_i|,$$

各組所屬次數爲

$$f_1,\quad f_2\cdots\cdots\cdots\cdots\cdots f_m,$$

時,依上之定義得平均差之數式爲

$$\Delta=\frac{f_1|X_1-M_i|+f_2|X_1-M_i|+\cdots\cdots+f_m|X_m-M_i|}{f_1+f_2+\cdots\cdots+f_m}$$

簡寫爲 (公式 47)

$$\Delta=\frac{1}{N}\Sigma f|X-M_i|$$

其計算形式如第 89 表

第 89 表

| 組別 | 代表值(X) | 次數(f) | $\|X-M_i\|$ | $f\|X-M_i\|$ |
|---|---|---|---|---|
| $L_1-u_1$ | $X_1$ | $f_1$ | $\|X_1-M_i\|$ | $f_1\|X_1-M_i\|$ |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| $L_m-u_m$ | $X_m$ | $f_m$ | $\|X_m-M_i\|$ | $f_m\|X_m-M_i\|$ |
| 計 | | N | | $\Sigma f\|X-M_i\|$ |

計算步驟

(1)求各組之代表值 X,

(2)求次數分配之中位數 $M_i$,

(3)計算偏差之絕對值 $|X_k-M_i|$,

(4)求 $|X_k-M_i|$ 之算術平均數

$$\frac{1}{N}\Sigma f\,|X-M_i|\,,$$

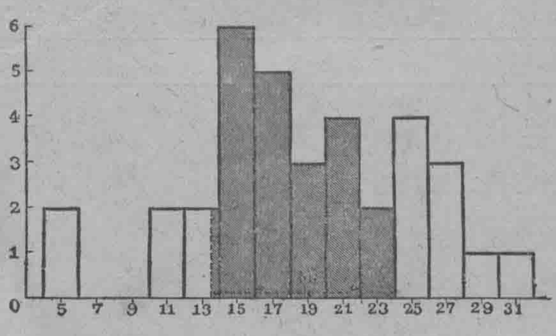
此平均數卽爲所求之平均差。

〔例〕求某考試成績之平均差。

先依第 90 表之次數分配狀況求得各單位値之中位數爲 $M_i=$ 18.7,

第 90 表

| 分數 | X | 人數 f | $X-M_i$ | $f\|X-M_i\|$ |
|---|---|---|---|---|
| 4——6 | 5 | 2 | −13.7 | 27.4 |
| 6——8 | 7 | 0 | −11.7 | 0 |
| 8——10 | 9 | 0 | − 9.7 | 0 |
| 10——12 | 11 | 2 | − 7.7 | 15.4 |
| 12——14 | 13 | 2 | − 5.7 | 11.4 |
| 14——16 | 15 | 6 | − 3.7 | 22.2 |
| 16——18 | 17 | 5 | − 1.7 | 8.5 |
| 18——20 | 19 | 3 | + 0.3 | 0.9 |
| 20——22 | 21 | 4 | + 2.3 | 9.2 |
| 22——24 | 23 | 2 | + 4.3 | 8.6 |
| 24——26 | 25 | 4 | + 6.3 | 25.2 |
| 26——28 | 27 | 3 | + 8.3 | 24.9 |
| 28——30 | 29 | 1 | +10.3 | 10.3 |
| 30——32 | 31 | 1 | +12.3 | 12.3 |
| 計 | | 35 | | 176.3 |




第 112 圖

其 $X_1=5$ 與中位數之偏差爲

$$X_1-M_i=5-18.7=-13.7,$$

其餘各偏差之計算皆與此相同。現進行表內之計算得

$$N=35, \qquad \Sigma f\,|X-M_i|=176.3,$$

$$\therefore \quad \Delta=\frac{1}{N}\Sigma f\,|X-M_i|=\frac{176.3}{35}=5.04,$$

其分布情形如第 112 圖。

(B)簡便算法

在實際計算平均差多用次之簡便算法。

設中位數 $M_i$ 在單位值 $X_r$ 與 $X_{r+1}$ 之間,其以 $M_i$ 爲標準所得各偏差之絕對值 $|X-M_i|$ 之和爲 $\Sigma$,若以 $X_{r+1}$ 看做爲假中位數 (trial median) $M'_i$ 以測得各假偏差(trial deviation)之絕對值 $|X-M_i'|$ 之和爲 $\Sigma'$ 時,$\Sigma$ 與 $\Sigma'$ 間有次之關係

$$\Sigma'=\Sigma-d'(N-2F) \qquad \text{(公式 48)}$$

此處之

$$d'=M_i'-M_i=X_{r+1}-M_i$$

$$F=f_1+f_2+\cdots\cdots+f_r$$

$$N=f_1+f_2+\cdots\cdots+f_m$$

其分配情形如第 91 表:

第 91 表

| $X$ | $f$ | $x=\lvert X-M_i\rvert$ | $x'=\lvert X-M_i'\rvert$ | $fx$ | $fx'$ |
|---|---|---|---|---|---|
| $X_1$ | $f_1$ | $x_1=M_i-X_1$ | $x'_1=x_1+d'$ | $f_1x_1$ | $f_1x_1'$ |
| $X_2$ | $f_2$ | $x_2=M_i-X_2$ | $x_2'=x_2+d'$ | $f_2x_2$ | $f_2x_2'$ |
| $\vdots$ | $\vdots$ F | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| $M_i''=X_r$ | $f_r$ | $x_r=M_i-X_r$ | $x_r'=x_r+d'$ | $f_rx_r$ | $f_rx_r'$ |
| $d''$ $M_i\rightarrow$ $d'$ | | | | | |
| $M_i'=X_{r+1}$ | $f_{r+1}$ | $x_{r+1}=X_{r+1}-M_i$ | $x'_{r+1}=x_{r+1}-d'$ | $f_{r+1}x_{r+1}$ | $f_{r+1}x'_{r+1}$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| $X_m$ | $f_m$ | $x_m=X_m-M_i$ | $x'_m=x_m-d'$ | $f_mx_m$ | $f_mx'_m$ |
| 計 | N | | | $\Sigma$ | $\Sigma'$ |

依表得

$$\begin{aligned}\Sigma' &= f_1x_1'+f_2x_2'+\cdots\cdots+f_mx'_m\\ &= f_1(x+d')+f_2(x_2+d')+\cdots\cdots+f_r(x_r+d')\\ &\qquad +f_{r+1}(x_{r+1}-d')+f_{r+2}(x_{r+2}-d')+\cdots\cdots\\ &\qquad +f_m(x_m-d')\\ &= f_1x_1+f_2x_2+\cdots\cdots+f_mx_m+d'(f_1+f_2+\cdots\cdots+f_r)-d'\\ &\qquad (f_{r+1}+\cdots\cdots f_m)\\ &= \Sigma fx+d'[F-(N-F)]\\ &= \Sigma-d'(N-2F),\end{aligned}$$

卽　$\Sigma=\Sigma'+d'(N-2F)$,

用 N 除此兩邊,得

$$\Delta=\frac{\Sigma f|X-M_r'|+d'(N-2F)}{N} \qquad \text{(公式 49)}$$

若各組之組距均等於 a 時,得用組距爲單位以求平均差。其公式爲

$$\bar{\Delta}=\frac{\Sigma f|\overline{X'}|+\frac{d'}{a}(N-2F)}{N} \qquad \text{(公式 50)}$$

此處之

$$|\overline{x'}|=\frac{|M_i'-X|}{a}$$

$$\therefore \quad \Delta=a\bar{\Delta} \qquad \text{(公式 51)}$$

〔例〕 第 92 表

| 組別 | X | f | $\lvert\overline{x'}\rvert=\left\lvert\frac{65-x}{a}\right\rvert$ | $f\lvert\overline{x'}\rvert$ |
|---|---|---|---|---|
| 40—50 | 45 | 3 | 2 | 6 |
| 50—60 | 55 | 19 | 1 | 19 |
| 60—70 | 65 | 30 | 0 | 0 |
| 70—80 | 75 | 15 | 1 | 15 |
| 80—90 | 85 | 2 | 2 | 4 |
| 90—100 | 95 | 1 | 3 | 3 |
| 計 | | 70 | | 47 |

依第 92 表內分配情形計算得單位值之中位數爲

$$M_i=60+10\times\frac{35-22}{30}=64.33=6.433a,$$

$$M_i=65=6.5a,$$

$$d' = M_i' - M_i = (6.5 - 6.433)a = 0.067a,$$

$$F = 3 + 19 = 22,$$

$$\bar{\Delta} = \frac{47 + 0.067 \times (70 - 44)}{70} = \frac{48.742}{70} = 0.6963,$$

$$\therefore \quad \Delta = a\,\bar{\Delta} = 10 \times 0.6963 = 6.96$$

上之 $M_i'$ 是取在 $X_{r+1}$ 上，若取 $X_r$ 為假中位數時，其以 $X_r$ 為基本所測得諸偏差之和 $\Sigma''$ 與 $\Sigma$ 間有次之關係

$$\Sigma'' = \Sigma + d''(N - 2F) = \Sigma - d''(2F - N),$$

(但此處之 $d'' = M_i - X_r$)

由此仍能得求平均差之簡便公式

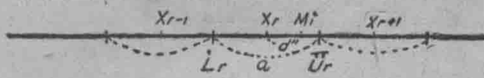
$$\Delta = \Delta' + \frac{1}{N}(2F - N)d'' \qquad \text{(公式 52)}$$

(C)平均差之校正

上述求平均差公式，是將各組所屬次數均集中於所當代表值上以計算一切。對此在不含有中位數之組上不覺有所矛盾處，但在含有中位數之組上則發生相當誤差，故須加以校正而後始可。現將其校正方法，說明如次：

| 組別 | 代表值 | 次數 |
|---|---|---|
| $L_1$——$u_1$ | $X_1$ | $f_1$ |
| ⋮ | ⋮ | ⋮ |
| $L_{r-1}$——$u_{r-1}$ | $X_{r-1}$ | $f_{r-1}$ |
| $L_r$——$u_r$ | $X_r$ | $f_r$ |
| $L_{r+1}$——$u_{r+1}$ | $X_{r+1}$ | $f_{r+1}$ |
| ⋮ | ⋮ | ⋮ |
| $L_m$——$u_m$ | $X_m$ | $f_m$ |
| 計 | | N |

設中位數 $M_i$ 是在組 $L_r - U_r$ 內,並取各組之中位數爲其代表值,如此得將 $M_i$ 附近情形圖示如第 113 圖。




第 113 圖

先取以 $X_r$ 爲基準所測得之平均差爲 $\Delta'$ 時,得

$$\Delta' = \frac{1}{N}\Big\{f_1(X_r - X_1) + f_2(X_r - X_2) + \cdots\cdots + f_{r-1}(X_r - X_{r-1}) + f_{r+1}(X_{r+1} - X_r) + \cdots\cdots + f_m(X_m - X_r)\Big\}, \cdots\cdots (1)$$

次以 $M_i$ 爲基準所測得一般的平均差爲

$$\Delta = \frac{1}{N}\Big\{f_1(M_i - X_1) + f_2(M_i - X_2) + \cdots + f_{r-1}(M_i - X_{r-1}) + P + f_{r+1}(X_{r+1} - M_i) + \cdots\cdots + f_m(X_m - M_i)\Big\}, \cdots\cdots (2)$$

(2)－(1)並取 $d'' = M_i - X_r$ 時,得

$$\begin{aligned}\Delta - \Delta' &= \frac{1}{N}\Big\{(f_1 + f_2 + \cdots\cdots + f_{r-1})(M_i - X_r) - (f_{r+1} + \cdots\cdots + f_m)(M_i - X_r) + P\Big\}\\ &= \frac{1}{N}\Big\{(f_1 + f_2 + \cdots\cdots + f_{r-1})d'' - (f_{r+1} + \cdots\cdots + f_m)d'' + P\Big\}, \cdots\cdots (3)\end{aligned}$$

依(3)式知在求得 P 之所當值後,卽得所求之正確平均差。現爲計算簡便起見,特取組距爲單位以求 P,在以組距爲單位時,得變換各所

當值爲

$$L_rU_r=1,$$

$$X_rM_i=\frac{M_i-X_r}{\text{組距}}=\frac{d''}{a}=C,$$

$$L_rX_r=X_rU_r=\frac{1}{2},$$

$$L_rM_i=L_rX_r+X_rM_i=\frac{1}{2}+C,$$

$$M_iU_r=X_rU_r-X_rM_i=\frac{1}{2}-C,$$

且屬於 $L_r-U_r$ 組之次數爲 $f_r$ 個，現先假定此等 $f_r$ 個次數是平均分配於組距 $L_rU_r$ 上後，再將此 $f_r$ 個次數依在 $L_rU_r$ 上由 $M_i$ 所分距離 $\frac{1}{2}+C$ 及 $\frac{1}{2}-C$ 以舉行分配時，知此 $f_r$ 個次數中在 $L_rM_i$ 上者爲 $\left(\frac{1}{2}+C\right)f_r$ 個，在 $M_iU_r$ 上者爲 $\left(\frac{1}{2}-C\right)f_r$ 個。又因 $L_rM_i$ 及 $M_iU_r$ 之中點與 $M_i$ 相差距離各爲

$$\frac{1}{2}\left(\frac{1}{2}+C\right), \qquad \frac{1}{2}\left(\frac{1}{2}-C\right),$$

故在假定 $L_rM_i$ 及 $M_iU_r$ 上所有分配次數是均集中於其中點時，在此組內所有各單位值與 $M_i$ 之偏差的絕對值之和，得定爲

$$P=f_r\left(\frac{1}{2}+C\right)\times\frac{1}{2}\left(\frac{1}{2}+C\right)+f_r\left(\frac{1}{2}-C\right)\times\frac{1}{2}\left(\frac{1}{2}-C\right)$$

$$=\left(\frac{1}{4}+C^2\right)f_r。$$

繼此 P 之所當值代入(3)式內並移項後得

$$\overline{\Delta}=\overline{\Delta}'+\frac{1}{N}\left\{\left(f_1+f_2+\cdots\cdots+f_{r-1}\right)\frac{d''}{a}-\left(f_{r+1}+\cdots\cdots+f_m\right)\frac{d''}{a}\right.$$

$$\left.+f_r\left(\frac{1}{4}+C^2\right)\right\}$$

$$=\overline{\Delta}'+\frac{1}{N}\left\{\left[\left(f_1+f_2+\cdots\cdots+f_{r-1}+f_r\right)-\left(f_{r+1}+\cdots+f_m\right)\right]C\right.$$

$$\left.+\left(\frac{1}{4}-C+C^2\right)f_r\right\}$$

$$=\overline{\Delta}'+\frac{1}{N}\left\{\left(F-\left(N-F\right)\right]C+\left(\frac{1}{2}-C\right)^2f_r\right\}$$

$$=\overline{\Delta}'+\frac{1}{N}\left\{\left(2F-N\right)C+\left(\frac{1}{2}-C\right)^2f_r\right\}。$$

所求校正後之平均差爲

$$\Delta=a\overline{\Delta}$$

$$=\Delta'+\frac{1}{N}\left\{(2F-N)d''+\left(\frac{1}{2}-C\right)^2af_r\right\}$$

(公式 53)

知所求之平均差,是於以中位數所存在一組之代表值 $(X_r)$ 爲基準所求得之平均差 $\Delta'$ 上,再加以

$$+\frac{1}{N}\left\{\left(2F-N\right)d''+\left(\frac{1}{2}-C\right)^2af_r\right\}$$

之數值後即得。

（例）

第 93 表

| 組別 | X | f | $\left\lvert\frac{x-x_r}{a}\right\rvert$ | $f\left\lvert\frac{x-x_r}{a}\right\rvert$ |
|---|---|---|---|---|
| 137.5—142.5 | 140 | 2 | 4 | 8 |
| 142.5—147.5 | 145 | 9 | 3 | 27 |
| 147.5—152.5 | 150 | 45 | 2 | 90 |
| 152.5—157.5 | 155 | 143 | 1 | 143 |
| 157.5—162.5 | 160 | 200 | 0 | 0 |
| 162.5—167.5 | 165 | 132 | 1 | 132 |
| 167.5—172.5 | 170 | 42 | 2 | 84 |
| 172.5—177.5 | 175 | 7 | 3 | 21 |
| 177.5—182.5 | 180 | 1 | 4 | 4 |
| 計 | | 584 | | 515 |

依第 93 表內分配情形求得

$$M_i=159.75, \qquad X_r=160,$$

$$f_r=200, \qquad F=402,$$

$$d''=M_i-X_r=159.72-160=-0.25,$$

$$C=\frac{d''}{a}=\frac{-0.25}{5}=-0.05,$$

$$\Sigma f\left|\frac{X-X_r}{a}\right|=515,$$

$$(2F-N)C=(804-584)(-0.05)=-11,$$

$$\left(\frac{1}{2}-C\right)^2 f_r=\left(\frac{1}{2}+0.05\right)^2\times 200=60.5$$

$$\therefore\quad \Sigma f\left|\frac{X-X_r}{a}\right|+\left(2F-N\right)C+\left(\frac{1}{2}-C\right)^2 f_r=564.5,$$

$$\overline{\Delta}=\frac{565.4}{584}=0.966,$$

$$\Delta=a\overline{\Delta}=5\times0.966=4.83。$$

(D)由次數曲線以求平均差

設次數曲線之方程式爲

$$y=f(x)$$

時，求平均差之數式爲

$$\Delta=\frac{\int_{-\infty}^{\infty}|X-M_i|f(X)dX}{\int_{-\infty}^{\infty}f(X)dX} \qquad \text{(公式 54)}$$

(E)平均差上使用中位數之理由

在平均差之計算公式內，所以專使用中位數爲基準而不使用其他各平均數者，實欲使所求得之平均差成爲最小，於應用上較爲便利故也。現將其成爲最小之證明方法，列舉三種如次：

(1)幾何的證明

用線長以代替偏差之絕對值以行證明。(但不注意線之正負符號)。

在二點 $X_1$，$X_n$ 間一點 P 與此二點相差距離之和，不管 P 在 $X_1X_n$ 間之何處，其和常爲一定等於 $X_1X_n$，

$$X_1P+PX_n=X_1X_n。$$

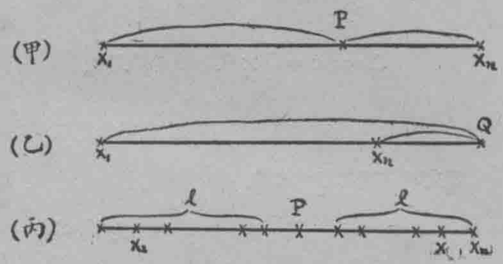
〔參照第 114 圖(甲)〕。

若取在 $X_1X_n$ 之外側一點 Q 爲基準以測得與 $X_1$，$X_n$ 相差距離之和，常大於 $X_1X_n$，即

$$X_1Q+X_nQ=X_1X_n+2X_nQ>X_1X_n。$$

〔參照第114圖(乙)〕

由此知以 $X_1X_n$ 間之P為基準所測得之偏差和,實比以 $X_1X_n$ 外之Q為基準所測得者為小。




第 114 圖

現在 P 點之兩側各取 l 個數值,〔參照第114圖(丙)〕當此時以 P 為基準所測得與各數值相差距離之和,卽為以 P 為基準所得各偏差之和,此偏差和實比在其他各數值間之一點為基準所測得者為小。故在直線上有偶數個 (2m) 數值時,上之 P 點得定在第 m 個及第 m+1 個之間,若其個數為奇數(2m+1)個時,上之 P 點是得定與第 m+1 個相一致。依如此方法所定得之 P 點,實不外為該數列之中位數。由此得知以中位數為基準所測得之平均差,實為其中之最小者也。

(2)代數的證明

設以某值 B 為基準所求得之平均差為 $\Delta_B$,在此 B 之左方假定其有總次數 n 中之 l 個。換言之,此 B 之數值不大於依大小順序排列淸楚中之由最小數起第 l+1 個所當數值。

次將此 B 之所當點向左移動距離 b 後得第二點 D,現以 D 為基準所求得之平均差為 $\Delta_D$,若此 D 點所當數值不小於依大小順序排列淸楚後之第 l 個所當數值時,在 B 點之左方共有 l 個次數,在 D 之右方共有(n−l)個次數。

將基點舉行上述移動後，在其右之 $(n-l)$ 個次數，均各增加 $b$ 之偏差，共增加 $(n-l)\,b$ 之偏差。反之在其左方之 $l$ 個次數各減少 $b$ 之偏差，共減少 $lb$ 之偏差。依此得以 D 爲基準之平均差爲

$$\Delta_D=\Delta_B+\frac{(n-l)b-lb}{n}=\Delta_B+\frac{b}{n}(n-2l)。$$

在此式內當 $n-2l>0$，卽 $l<\frac{n}{2}$ 時，因 $b_1n$ 均爲正數，故得

$$\Delta_D>\Delta_B。$$

知在 $l<\frac{n}{2}$ 範圍內，依其基點之逐次向右移動，而漸次減小其平均差。

若取 $n-2l<0$，卽 $l>\frac{n}{2}$ 時，成爲

$$\Delta_D<\Delta_B。$$

知在 $l>\frac{n}{2}$ 範圍內，依其基點之逐次向左移動而漸次減小其平均差。

至 $n-2l=0$，卽 $l=\frac{n}{2}$ 時，始成爲

$$\Delta_D=\Delta_B。$$

知在 $l=\frac{n}{2}$ 範圍內，卽當 $n=$ 偶數時，在第 $\frac{n}{2}$ 及第 $\left(\frac{n}{2}+1\right)$ 次數所當數值間各值所作成之平均差成爲一定。且此平均差因其左右各平均差均漸次向此減小，故知其爲最小值。若在 $n=$ 奇數時，成爲以第 $\frac{n+1}{2}$ 次數之數值爲基準所作成之平均差爲最小。但當 $n=$ 偶數時，在第 $\frac{n}{2}$ 及第 $\left(\frac{n}{2}+1\right)$ 間之數值，及當 $n=$ 奇數時其第 $\frac{n+1}{2}$ 次數所當數值，

是爲該數列之中位數。故知以中位數爲基準所作成之平均差，實爲其中之最小者也。

(3)微積分的證明。

設取任意一平均數 X 爲基準所作成之平均差爲

$$\Delta_X=\frac{\int_{-\infty}^{\infty}|x-x|f(x)dx}{\int_{-\infty}^{\infty}f(x)dx},$$

在此式內其分子之 X 愈大時，則所求得之 $\Delta_X$ 亦愈大，故不能求其極大值，只能研究其極小情形。但欲 $\Delta_X$ 成爲極小，只須分子變爲極小巳足，現就分子以研究之。

取

$$\begin{aligned}\varphi(X)&=\int_{-\infty}^{\infty}|x-X|f(x)dx\\&=\int_{-\infty}^{X}|x-X|f(x)dx+\int_{X}^{\infty}|x-X|f(x)dx\\&=\int_{-\infty}^{X}(X-x)f(x)dx+\int_{X}^{\infty}(x-X)f(x)dx\\&=X\int_{-\infty}^{X}f(x)dx-\int_{-\infty}^{X}xf(x)dx+\int_{X}^{\infty}xf(x)dx-X\int_{X}^{\infty}f(x)dx,\end{aligned}$$

關於 X 以舉行微分得

$$\varphi'(X)=\int_{-\infty}^{X}f(x)dx+Xf(X)-Xf(X)-Xf(X)-\int_{X}^{\infty}f(x)+Xf(X),$$

欲 $\varphi(X)$ 成爲極小時必須

$$\varphi'(X)=0,$$

由此求得

$$\int_{-\infty}^{X}f(x)dx-\int_{X}^{\infty}f(x)dx=0,$$

卽 $$\int_{-\infty}^{X}f(x)dx=\int_{X}^{\infty}f(x)dx。$$

依公式 16 知能滿足此積分方程式之 X 卽爲中位數，故知以中位數爲基準所作成之平均差,爲其中之最小者也。

## (F)平均差之性質

### (甲)優點

(1)簡單明瞭易於了解。

(2)受各單位值之影響。

(3)計算容易。

(4)平均差實爲平均數之一變型(偏差之平均),故與算術平均數之性質大體類似。

(5)於中位數之左右取平均差長之距離時，其所定範圍 ( $M_i-\Delta$, $M_i+\Delta$) 內之次數已占總次數之 57.62%。

### (乙)劣點

(1)受極端值之影響。

(2)不能代數的運用。以其有此性質,故在理論的硏究上不甚適合，但因其性質之簡單明瞭與計算之容易，故仍爲統計學家生物學家等所樂用。

## 第五節 標準差

### (A)意義及其計算公式

設次數分配情形爲

| 單位值 | $X_1$, $X_2$…………$X_r$…………$X_n$ |
|---|---|
| 次數 | $f_1$, $f_3$…………$f_r$…………$f_n$ |

由此求得其算術平均數爲 M，各單位值與 M 相差距離爲

$$x_1=X_1-M,\ x_2=X_2-M,\cdots\cdots\cdots\cdots x_m=X_m-M。$$

此等距離稱爲由 M 所測得各單位值之偏差，在此等偏差上實各有正負符號之不同，現爲計算上便利起見，特各取其平方，使勿必再顧慮正負符號。現取各偏差值各自平方後之算術平均數爲

$$\frac{f_1x_1{}^2+f_2x^2+\cdots\cdots+f_mx^2{}_m}{f_1+f_2+\cdots\cdots+f_m}。$$

依平均差處所述理由，知如此所得平均數雖亦得定爲一種差量，但卽以此平均數當做爲差量則尙有未妥處。蓋因前述各差量(四分位差，平均差等)其差量與單位值均爲同種類物(其元次相等)，但在現述之平均數則其種類不相同，故須於該數值上再開方(取其正值)後始能定爲所求之差量，此差量稱爲標準差，用 $\sigma$ 以表示之。對此數值之稱呼有次之數種：

Pearson 稱爲標準差(standard deviation),

Gauss 稱爲平均誤差(mean error),

Airy 稱爲平均平方誤差(mean square error),

但在統計界上則均使用標準差之名稱。此外

Edgeworth 稱 $\sigma^2, 2\sigma^2$ 爲移動(fluctuation),

Airy 稱 $\sqrt{2}\sigma$ 爲模差(modulus),

Lexis 稱 $\dfrac{1}{\sqrt{2}\,\sigma}$ 爲可靠性(precision)。

至標準差之數式如次：

$$\sigma=\sqrt{\frac{f_1(X_1-M)^2+f_2(X_2-M)^2+\cdots\cdots+f_m(X_m-M)^2}{f_1+f_2+\cdots\cdots+f_m}}$$

簡寫爲 (公式 55)

$$\sigma=\sqrt{\frac{\Sigma fX^2}{f}}。$$

其計算形式如第 94 表。

第 94 表

| 組別 | 代表值 X | 次數 f | x=X-M | fx | $fx^2$ | $f(x+1)^2$ |
|---|---|---|---|---|---|---|
| $L_1-U_1$ | $X_1$ | $f_1$ | $x_1$ | $f_1x_1$ | $f_1x_1^2$ | $f_1(x_1+1)^2$ |
| $L_2-U_2$ | $X_2$ | $f_2$ | $x_2$ | $f_2x_2$ | $f_2x_2^2$ | $f_2(x_2+1)^2$ |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| $L_m-U_m$ | $X_m$ | $f_m$ | $x_m$ | $f_mx_m$ | $f_mx^2_m$ | $f_m(x_m+1)^2$ |
| 計 | M=… | $\Sigma f$ | | $\Sigma fx$ | $\Sigma fx^2$ | $\Sigma f(x+1)^2$ |

在上表內之最右一欄，是用以檢查計算之有無錯誤。其檢查方法如次：

依數學上定理得

$$(x_k+1)^2=x_k^2+2x_k+1,$$

雙方乘以 $f_k$ 並將 K 由 1 至 M 逐次相加得

$$\Sigma f_k(x_k+1)^2=\Sigma f_kx_k^2+2\Sigma f_kx_k+\Sigma f_k,$$

因其有此關係式，故只要檢查

$$\Sigma fx^2+2\Sigma f+\Sigma f$$

之所當值是否與最右欄之總和

$$\Sigma f(x+1)^2$$

相一致,若不一致時則計算中必有錯誤。

〔例 I〕

第 95 表

| X | x | $x^2$ | $(x+1)^2$ |
|---|---|---|---|
| 23.7 | −0.2 | 0.04 | 0.64 |
| 24.1 | +0.2 | 0.04 | 1.44 |
| 23.9 | 0 | 0 | 1.00 |
| 23.8 | −0.1 | 0.01 | 0.81 |
| 24.0 | +0.1 | 0.01 | 1.21 |
| M=23.9 | 0 | 0.10 | 5.10 |

依第 95 表得

$$n=5,$$

$$\Sigma x^2+2\Sigma x+n=5.10,$$

與 $\Sigma(x+1)^2$ 之結果相一致,知計算無誤,故得

$$\sigma=\sqrt{\frac{\Sigma x^2}{n}}=\sqrt{\frac{0.10}{5}}=0.141\cdots\cdots。$$

〔例 II〕

第 96 表

| 組 別 | X | f | x | fx | $fx^2$ | $f(x+1)^2$ |
|---|---|---|---|---|---|---|
| 40— 50 | 45 | 3 | −19.6 | − 58.8 | 1152.48 | 1039.88 |
| 50— 60 | 55 | 19 | − 9.6 | −182.4 | 1751.04 | 1405.24 |
| 60— 70 | 65 | 30 | + 0.4 | + 12.0 | 4.80 | 58.80 |
| 70— 80 | 75 | 15 | +10.4 | +156.0 | 1622.40 | 1949.40 |
| 80— 90 | 85 | 2 | +20.4 | + 40.8 | 832.32 | 915.92 |
| 90—100 | 95 | 1 | +30.4 | + 30.4 | 924.16 | 985.96 |
| 計 | M=64.6 | 70 | | − 2.0 | 6287.20 | 6353.20 |

依第 96 表得

$$\Sigma fx^2+2\Sigma fx+\Sigma f=6287.20-4.0+70$$

$$=6353.20$$

與 $\Sigma f(x+1)^2$ 之結果相一致，知計算無誤。

$$\therefore\quad \sigma=\sqrt{\frac{\Sigma fx^2}{\Sigma f}}=\sqrt{\frac{6287.20}{70}}=9.48。$$

(B)簡便算法

在各單值內任意取一單位值，現假定其為 A。此 A 與各單位值之差為

$$x_1'=X_1-A,\ x_2'=X_2-A,\cdots\cdots\cdots x_m'=X_m-A,$$

與 M 之差為

$$d=M-A$$

時，得依

$$\sigma^2=\frac{\Sigma fX'^2}{\Sigma f}-d^2 \qquad \text{(公式 56)}$$

以求得標準差。

〔證〕

$$\Sigma f_k x'^2_k=\Sigma f_k(X_k-A)^2$$

$$=\Sigma f_k[(X_k-M)+(M-A)]^2$$

$$=\Sigma f_k[(X_k-M)+d]^2$$

$$=\Sigma f_k[(X_k-M)^2+2d(X_k-M)+d^2]$$

$$=\Sigma f^K(X_K-M)^2+2d\Sigma f_K(X_K-M)+d^2\Sigma f_k$$

但

$$\Sigma f_k(X_K-M)=\Sigma f_K X_K-M\Sigma f_K=0,$$

$$\therefore\quad \Sigma f_K x'^2_K=\Sigma f_K(X_K-M)^2+d^2\Sigma f_K,$$

$$\frac{\Sigma f_K x'^2_K}{\Sigma f_K}=\frac{\Sigma f_K(X_K-M)^2}{\Sigma f_K}+d^2=\sigma^2+d^2,$$

$$\therefore\quad \sigma^2=\frac{\Sigma f_K x_K'^2}{\Sigma f_K}-d^2。$$

若用 S 以表示由一任意單位值 A 所求得之標準差時，上式成爲

$$\sigma^2=S^2-d^2。\qquad\text{(公式57)}$$

若取 A＝0 時，公式 56 變爲

$$\sigma^2=\frac{\Sigma fX^2}{\Sigma f}-M^2。\qquad\text{(公式58)}$$

現應用 $\sigma^2=S^2-d^2$ 之理論，以求標準差之簡便算法如次：

取次數分配內最大次數所當單位值 $X_r$ 爲 A，以計算標準差 $\sigma$ 之手續如第 97 表。

第　97　表

| 組　別 | 代表值 X | 次數 f | $x'=X-X_r$ | $fx'$ | $fx'^2$ | $f(x'+1)^2$ |
|---|---|---|---|---|---|---|
| $L_1—U_1$ | $X_1$ | $f_1$ | $x_1'$ | $f_1x_1'$ | $f_1x_1'^2$ | $f_1(x_1'+1)^2$ |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| $L_{r-1}—U_{r-1}$ | $X_{r-1}$ | $f_{r-1}$ | $x'_{r-1}$ | $f_{r-1}x'_{r-1}$ | $f_{r-1}x'^2_{r-1}$ | $f_{r-1}(x'_{r-1}+1)^2$ |
| $L_r—U_r$ | $A=X_r$ | $f_r$ | $x'_r$ | $f_rx'_r$ | $f_rx'^2_r$ | $f_r(x_r'+1)^2$ |
| $L_{r-1}—U_{r-1}$ | $X_{r+1}$ | $f_{r+1}$ | $x'_{r+1}$ | $f_{r+1}x'_{r+1}$ | $f_{r+1}x'^2_{r+1}$ | $f_{r+1}(x'_{r+1}+1)^2$ |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| $L_m—U_m$ | $X_m$ | $f_m$ | $x_m$ | $f_mx'_m$ | $f_mx'^2_m$ | $f_m(x'_m+1)^2$ |
| 計 | | $\Sigma f$ | | $\Sigma fx'$ | $\Sigma fx'^2$ | $\Sigma f(x'+1)^2$ |

在 $\sigma^2=S^2-d^2$ 內之

$$S^2=\frac{\Sigma fx'^2}{\Sigma f},$$

$$d=M-A=\frac{\Sigma fx}{\Sigma f}-X_r$$

$$=\frac{1}{\Sigma f}[f_1X_1+f_2X_2+\cdots\cdots+f_mX_m-X_r(f_1+f_2+\cdots\cdots+f_m)]$$

$$=\frac{1}{\Sigma f}[f_1(X_1-X_r)+f_2({}_2X-X_r)+\cdots\cdots+f_m(X_m-X_r)]$$

$$=\frac{\Sigma fx'}{\Sigma f},$$

$$\therefore \qquad \sigma^2=\frac{\Sigma fX'^2}{\Sigma f}-\left(\frac{\Sigma f}{\Sigma f}\right)^2。 \qquad \text{(公式 59)}$$

上式之右方各數値均得由第 97 表內最下一行以求得之，故知依表卽能求得 $\sigma$。

〔例〕

第 98 表

| 單位値 X | 次數 f | $x'=X-160$ | $fx'$ | $fx'^2$ | $f(x'+1)^2$ |
|---|---|---|---|---|---|
| 140 | 2 | −20 | − 40 | 800 | 722 |
| 145 | 9 | −15 | − 135 | 2025 | 1764 |
| 150 | 48 | −10 | − 480 | 4800 | 3888 |
| 155 | 143 | − 5 | − 715 | 3575 | 2288 |
| 160 | 200 | 0 | (−1370) | 0 | 200 |
| 165 | 131 | 5 | 655 | 3275 | 4716 |
| 170 | 42 | 10 | 420 | 4200 | 5082 |
| 175 | 7 | 15 | 105 | 1575 | 1792 |
| 180 | 1 | 20 | 20 | 400 | 441 |
| 計 | 583 | | +1200<br>−1370<br>− 170 | 20650 | 20893 |

依第 98 表之最下行得

$$\Sigma f=583,\qquad \Sigma fx'=-170,$$

$$\Sigma fx'^2=20650,\qquad \Sigma f(x'+1)^2=20893,$$

$$\Sigma fx'^2+2\Sigma fx'+\Sigma f=20893,$$

與 $\Sigma f(x'+1)^2$ 之結果相一致，知計算無誤。故得

$$\sigma^2=\frac{\Sigma fx'^2}{\Sigma f}-\left(\frac{\Sigma fx'}{\Sigma f}\right)^2$$

$$=\frac{20650}{583}-\left(-\frac{170}{583}\right)^2$$

$$=35.335,$$

$$\therefore\quad \sigma=5.94。$$

但在實際上若遇組距完全相等之次數分配時，多用第 99 表之簡便算法以求其標準差。

第 99 表

| 組 別 | 代表值 X | 次數 f | $x'=X-X_r$ | $\bar{x}'=\frac{x'}{a}$ | $f\bar{x}'$ | $f\bar{x}'^2$ | $f(\bar{x}'+1)^2$ |
|---|---|---|---|---|---|---|---|
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| —— | $X_{r-2}=A-2a$ | $f_{r-2}$ | $-2a$ | $-2$ | $-2f_{r-2}$ | $4f_{r-2}$ | $f_{r-1}$ |
| —— | $X_{r-1}=A-a$ | $f_{r-1}$ | $-a$ | $-1$ | $-f_{r-1}$ | $f_{r-1}$ | 0 |
| —— | $X_r=A$ | $f_r$ | 0 | 0 | $(-\cdots\cdots)$ | 0 | $f_r$ |
| —— | $X_{r+1}=A+a$ | $f_{r+1}$ | $a$ | 1 | $f_{r+1}$ | $f_{r+1}$ | $4f_{r+1}$ |
| —— | $X_{r+2}=A+2a$ | $f_{r+2}$ | $2a$ | 2 | $f_{r+2}$ | $4f_{r+2}$ | $9f_{r+2}$ |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 計 | | $\Sigma f$ | | | $(+\cdots\cdots)$ $\Sigma f\bar{x}'$ | $\Sigma f\bar{x}'^2$ | $\Sigma f(\bar{x}'+1)^2$ |

由上表得

$$\bar{S}^2=\frac{\Sigma f\bar{x}'^2}{\Sigma f},\qquad \bar{d}=\frac{\Sigma f\bar{x}'}{\Sigma f},$$

$$\bar{\sigma}^2=\bar{S}^2-\bar{d}^2,$$

但

$$S^2=\frac{\Sigma fx'^2}{\Sigma f}=\frac{\Sigma f(a\bar{x}')^2}{\Sigma f}=a^2\frac{\Sigma f\bar{x}'^2}{\Sigma f}=a^2\bar{S}^2,$$

$$d=\frac{\Sigma fx'}{\Sigma f}=\frac{\Sigma fa\bar{x}'}{\Sigma f}=a\frac{\Sigma f\bar{x}'}{\Sigma f}=a\bar{d}$$

$$\sigma^2=S^2-d^2=a^2\bar{S}^2-a^2\bar{d}^2$$

$$=a^2(\bar{S}^2-\bar{d}^2)=a^2\bar{\sigma}^2,$$

$$\therefore\quad \sigma=a\bar{\sigma}。\qquad\text{(公式 60)}$$

〔例〕

第　100　表

| 組別 | X | f | $\bar{x}'$ | $f\bar{x}'$ | $f\bar{x}'^2$ | $f(\bar{x}'+1)^2$ |
|---|---|---|---|---|---|---|
| 40——50 | 45 | 3 | −2 | −6 | 12 | 3 |
| 50——60 | 55 | 19 | −1 | −19 | 19 | 0 |
| 60——70 | A=65 | 30 | 0 | (−25) | 0 | 30 |
| 70——80 | 75 | 15 | 1 | 15 | 15 | 60 |
| 80——90 | 85 | 2 | 2 | 4 | 8 | 18 |
| 90——100 | 95 | 1 | 3 | 3 | 9 | 16 |
| 計 | | 70 | | +22<br>−25<br>−3 | 63 | 127 |

依表知

$$A=65,\qquad a=10,\qquad \Sigma f=70,$$

$$\Sigma f\bar{x}'=-3\qquad \Sigma f\bar{x}'=\sigma 3,\qquad \Sigma f(\bar{x}'+1)^2=127,$$

$$\Sigma f\bar{x}'^2+2\Sigma f\bar{x}'+\Sigma f=63-6+70=127,$$

與 $\Sigma(\bar{x}'^2+1)^2$ 之結果相一致，知計算無誤。故能依表求得

$$\bar{d}=\frac{\Sigma f\bar{x}'}{\Sigma f}=\frac{-3}{70}=-0.0429,$$

$$\bar{S}^2=\frac{\Sigma f\bar{x}'^2}{\Sigma f}=\frac{63}{70}=0.9,$$

$$\bar{\sigma}^2=\bar{S}^2-\bar{d}^2=0.9-0.0018=0.8981,$$

$$\bar{\sigma}=0.948,$$

$$\therefore\quad \sigma=a\cdot\bar{\sigma}=10\times 0.948=9.48。$$

(c)由次數曲線以求標準差

設次數曲線之方程式爲

$$y=f(x)$$

時，求標準差之公式爲

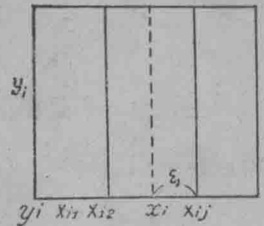
$$\sigma^2=\frac{\int_{-\infty}^{\infty}(X-M)^2f(X)dX}{\int_{-\infty}^{\infty}f(X)dX}\qquad\text{(公式 61)}$$

(D)標準差之校正

上述求標準差公式，是將各組所屬次數均集中於所當代表值(各組之中位數)上以計算一切，對此實不免含有誤差。薛伯 (Sheppard) 特

想出一種方法以補正其缺點。但其所加補正實甚微小，故在一般上不加入在內，亦無關重要。現將其補正方法說明如次（關於薛伯之一般補正方法在本編第十章第二節內再爲詳述）：

先假定一組內所屬次數是平均分配於該組距內，即第 i 組內所屬數値 $X_{i1}$，$X_{i2}$，……$X_{ik}$ 是依等距離以行相隔，且此等數値所屬次數亦均假定其等於 $y_i$。現定該組之代表値爲 $x_i$，而各 $X_{ij}(j=1,2\cdots\cdots,k)$ 與此 $x_i$ 相差距離用 $E_j$ 以表示時



第 115 圖

$$X_{ij}-X_i=D_j,$$

$$\therefore\ x_i=X_{ij}-\varepsilon_j。$$

又因

$$M-x_i=M-(X_{ij}-\varepsilon_j)=M-X_{ij}+\varepsilon_j,$$

$$\therefore\ (M-x_i)^2=(M-x_{ij})^2+2(M-X_{ij})E_j+\varepsilon_j^2,$$

現用 1, 2,……等依次代入 j 內並行相加得($f_i$＝第 i 組次數)

$$f_i(M-x_i)^2=\Sigma(M-X_{ij})^2+2M\Sigma\varepsilon_j-2\sum_j X_{ij}\varepsilon_j+\Sigma\varepsilon^2{}_j$$

但　　$\Sigma\varepsilon=0,\qquad \Sigma X\varepsilon=0,\qquad \Sigma\varepsilon^2=f_i\frac{\Sigma\varepsilon^2}{f_i},$

此$\frac{\Sigma\varepsilon^2}{f_i}$是用以計算各組本身標準差之數量，且此數量在各組上均得看做爲相等，現用 $\sigma\varepsilon^2$ 以表示時，上式成爲

$$f_i(M-x_i)^2=\sum_j(M-X_{ij})^2+f_i\sigma\varepsilon^2。$$

現用 1, 2,……等代入 i 內就組之全體以行考慮後，再舉行相加得

$$\sum_i f(M-x_i)^2=\sum_{i,j}(M-X_{ij})^2+\sigma\varepsilon^2\sum_i f_i,$$

用 $\Sigma f$ 以除其兩邊，得

$$\frac{\Sigma f(M-x)^2}{\Sigma f}=\frac{\Sigma(M-X)^2}{\Sigma f}+\sigma\varepsilon^2。$$

上式之左邊爲以前所述之標準差，用 $\sigma^2$ 以表示之，右邊之第一項是在新假定之下所起標準差用 $\Sigma_C{}^2$ 以表示時，得

$$\sigma^2=\sigma_C{}^2+\sigma\varepsilon^2,$$

依此得

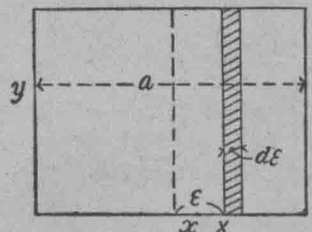
$$\sigma_C{}^2=\sigma^2-\sigma\varepsilon^2。$$

上式之 $\sigma\varepsilon^2$ 是由假定各組內含有無數數值並依次之積分方法以求得者(a＝組距)，

$$f_i=\Sigma f_{ij}=\int_{-\frac{a}{2}}^{\frac{a}{2}} y d\varepsilon=y\cdot a$$

$$\Sigma\varepsilon_j{}^2=\int_{-\frac{a}{2}}^{\frac{a}{2}} yE^2 d\varepsilon=y\frac{a^3}{12},$$

$$\therefore\quad \sigma^2\varepsilon=\frac{\Sigma\varepsilon_j{}^2}{f_i}=\frac{y\frac{a^3}{12}}{y\cdot a}=\frac{a^2}{12}。$$



第 116 圖

此 $\sigma\varepsilon$ 爲 Sheppard 之補正值，將此代入前式，得

$$\sigma_c{}^2=\sigma^2-\frac{a^2}{12}。 \qquad \text{(公式 62)}$$

〔例〕

第 101 表爲某種花類雄蕊數花朵別分配表

第 101 表

| 雄蕊數 X | 花朵數 f | x′ | fx′ | fx′² | f(x′+1)² |
|---|---|---|---|---|---|
| 27 | 1 | −10 | −10 | 100 | 81 |
| 28 | 1 | −9 | −9 | 81 | 64 |
| 29 | 1 | −8 | −8 | 64 | 49 |
| 30 | 3 | −7 | −21 | 142 | 108 |
| 31 | 8 | −6 | −48 | 288 | 200 |
| 32 | 10 | −5 | −50 | 250 | 160 |
| 33 | 19 | −4 | −76 | 304 | 171 |
| 34 | 34 | −3 | −102 | 306 | 136 |
| 35 | 52 | −2 | −104 | 208 | 52 |
| 36 | 62 | −1 | −62 | 62 | 0 |
| A=37 | 56 | −0 | (−490) | 0 | 56 |
| 38 | 40 | 1 | 40 | 04 | 160 |
| 39 | 37 | 2 | 74 | 148 | 333 |
| 40 | 24 | 3 | 72 | 216 | 384 |
| 41 | 10 | 4 | 40 | 160 | 250 |
| 42 | 3 | 5 | 15 | 75 | 108 |
| 43 | 1 | 6 | 6 | 36 | 49 |
| 計 | 362 | | +247<br>−490<br>−243 | 2485 | 2361 |

依表先檢查得

$$\Sigma fx'^2+2\Sigma fx'+\Sigma f=2485-486+302=2361$$

與 $\Sigma f(x'+1)^2$ 之結果相一致，知計算無誤。再依表求得

$$A=37,d=\frac{\Sigma fx'}{\Sigma f}=\frac{-243}{362}=-0.671,$$

$$S^2=\frac{\Sigma fx^2}{\Sigma f}=\frac{2485}{362}=6.864\cdots\cdots,$$

$$\sigma^2=S^2-d^2=6.864\cdots\cdots-0.4489=6.41,$$

$$\therefore\ \sigma=2.5\cdots\cdots。$$

但將第 101 表內分配情形依組距＝3 以變其分配時，得第 102 表。

第　102　表

| 組　別 | X | f | $\bar{x}'$ | $f\bar{x}'$ | $f\bar{x}'^2$ | $f(\bar{x}'+1)^2$ |
|---|---|---|---|---|---|---|
| 27—30 | 28.5 | 3 | −3 | −9 | 27 | 12 |
| 30—33 | 31.5 | 21 | −2 | −42 | 84 | 21 |
| 33—36 | 34.5 | 105 | −1 | −105 | 105 | 0 |
| A→36—39 | 37.5 | 158 | 0 | (−156) | 0 | 138 |
| 39—42 | 40.5 | 71 | 1 | 71 | 71 | 284 |
| 42—45 | 43.5 | 4 | 2 | 8 | 16 | 36 |
| 計 | | 362 | | +79<br>−156<br>−77 | 303 | 511 |

先檢查得

$$\Sigma f\bar{x}'^2+2\Sigma f\bar{x}'+\Sigma f=303-154+362=511,$$

與 $\Sigma f(\bar{x}'+1)^2$ 之結果相一致，知計算無誤。次依表得

$$a=3,\ \bar{d}=\frac{-77}{362}=-0.213,$$

$$\bar{S}^2=\frac{303}{362}=0.8370,$$

$$\bar{\sigma}^2=\bar{S}^2-\bar{d}^2=0.8970-0.0454=0.7916,$$

$$\bar{\sigma}=0.89,$$

$$\therefore\ \sigma=a.\bar{\sigma}=3\times0.89=2.67。$$

與由第 101 表所求得者不相一致。現在此依組別分配所求得之標準差

上施以薛伯補正。

依上表得

$$\sigma^2=a^2\bar{\sigma}^2=9\times0.7916=7.12,$$

$$\frac{a^2}{12}=\frac{1}{12}\times9=0.75,$$

$$\therefore\quad \sigma_C{}^2=7.12-0.75=6.37,$$

$$\sigma_C=2.5\cdots\cdots。$$

與由第 101 表所求得者幾相一致。薛伯補正之有效，由此得以明矣。

## (E)標準差上使用算術平均數之理由

在標準差計算中不使用其他平均數而專使用算術平均數者，實欲使所求得之標準差成爲極小故也。

〔證明一〕取任意一平均數爲 A，由此所作成之標準差爲

$$S=\sqrt{\frac{\Sigma f x'^2}{\Sigma f}}\qquad(\text{但}x'=X-A),$$

此標準差數值實依 A 之選定情形而不相同。現特將上式依公式 57 變形爲

$$S^2=\sigma^2+(M-A)^2,$$

在此式內 $(M-A)^2$ 一數值不能變爲負數，故知 $S^2$ 決不能比 $\sigma^2$ 爲小，只在 $M-A=0$ 卽 $A=M$ 時，始能使

$$S^2=\sigma^2。$$

知在各種 S 值中，其最小者，實爲以算術平均數爲基準所求得之標準差，因其有此性質，故特採算術平均數爲基準以計算標準差，使其所得結果成爲最小也。

〔證明二〕設取任意一平均數 X 爲基準所作成之標準差爲

$$\sigma^2=\frac{\int_{-\infty}^{\infty}(x-X)^2f(x)dx}{\int_{-\infty}^{\infty}f(x)dx}.$$

欲此式成爲最小只須取分子成爲最小已足，現取

$$\varphi(x)=\int_{-\infty}^{\infty}(x-X)^2f(x)dx,$$

微分之，得

$$\varphi'(x)=-2\int_{-\infty}^{\infty}(x-X)f(x)dx=0,$$

由此得

$$\int_{-\infty}xf(x)dx=X\int_{-\infty}^{\infty}f(x)dx,$$

卽

$$X=\frac{\int_{-\infty}^{\infty}xf(x)dx}{\int_{-\infty}^{\infty}f(x)dx}。$$

依公式 11 知上式右邊所得結果，是爲該分配之算術平均數，現旣與 X 相等，故依微分學知識，知當 X 等於算術平均數時之標準差成爲最小也。

## (E)標準差之性質

### (甲)優點

(1)關係於各觀察單位值。

(2)在相加前之平方方法，能避去由不注意正負符號所起代數上之錯誤。

(3)有正確之數學意義能代數的運用。

(a)統計數列是由若干部分數列所合成時，總數列之標準差得由

各部分數列之標準以表示之。

〔證〕設數列之總體是由二部分數列所合成，且各數列上各種符號如第 103 表。

第 103 表

| | 平均數 | 總次數 | 標準差 | d |
|---|---|---|---|---|
| 總體 | $M$ | $N=N_1+N_2$ | $\sigma$ | |
| 第一部分 | $M_1$ | $N_1$ | $\sigma_1$ | $d_1=M_1-M$ |
| 第二部分 | $M_2$ | $N_2$ | $\sigma_2$ | $d_2=M_2-M$ |

依公式 57 得

$$S_1^2=\sigma_1^2+d_1^2,\qquad S_2^2=\sigma_2^2+d_2^2,$$

但

$$\sigma^2=\frac{\sum_{i=1}^{N}(X_i-M)^2}{N}=\frac{\sum_{i=1}^{N_1}(X_i-M)^2+\sum_{i=N_1+1}^{N}(X_i-M)^2}{N},$$

$$\therefore\ N\sigma^2=\sum_{i=1}^{N_1}(X_i-M)^2+\sum_{i=N_1+1}^{N}(X_i-M)^2$$

$$=N_1\frac{\sum_{i=1}^{N_1}(X_i-M)^2}{N_1}+N_2\cdot\frac{\sum_{i=N_1+1}^{N}(X_i-M)^2}{N_2}$$

（但在此處將 M 看做爲各數列之假平均）

$$=N_1S_1^2+N_2S_2^2$$

$$=N_1(\sigma_1^2+d_1^2)+N_2(\sigma_2^2+d_2^2),$$

即 $N\sigma^2=N_1\sigma_1^2+N_2\sigma_2^2+N_1d_1^2+N_2d_2^2$.

一般

$$N\sigma^2=\Sigma(N_i\sigma_i^2)+\Sigma(N_id_i^2)\qquad\text{(公式 63)}$$

〔系〕當 $M_1=M_2$ 時上式成爲

$$N\sigma^2=N_1\sigma_1^2+N_2\sigma_2^2$$

取 $M_1=M_2=\overline{M}$ 時依公式 13 得 $\overline{M}=M$

由此得

$$d_1=d_2=\overline{M}-M=0$$

將上式變爲

$$N\sigma^2=N_1\sigma_1^2+N_2\sigma_2^2$$

一般

當 $M_1=M_2=\cdots\cdots=M_n$ 時

$$N\sigma^2=N_1\sigma^2_1+N_2\sigma_2^2+\cdots\cdots+N_n\sigma_n^2 \qquad \text{(公式 64)}$$

(4)標準差與算術平均 M 幾何平均 G 及調和平均 H 間實有密切之數理關係。

(a)標準差與算術平均數及幾何平均數間有次之關係式

$$G=M\left(1-\frac{1}{2}\frac{\sigma^2}{M^2}\right) \qquad \text{(公式 65)}$$

詳情請參考

W. Scheibner: "Ueber Mittelwerth", Berichte der kgl Sachsischen Gesellschaft d. Wissenschaft. 1873, S. 564.

G. J. Fechner: Kollektivmasslehre, 1897

對於上式 Guncker 更設『若所測偏差(X)比其平均數(M)爲微小而 $\left(\frac{x}{M}\right)^2$ 得以無視時』之假定以求得其間之近似的關係式如次:

$$M^2-G^2=\sigma_2 \qquad \text{(公式 66)}$$

詳情請參考

G. Duncker: Die Methode der Variationsstatistik, Leipzig, 1899.

Crump: Journal of Royal Statistical Society, March, 1924.

(b)標準差與算術平均數及調和平均數間有次之關係。

在偏差比其平均數爲小時,近似的得

$$H=M\left(1-\frac{\sigma^2}{M^2}\right)。\qquad (公式 67)$$

詳情請參考 Scherbner, Loc, Cit. Qu. 8

(5)在對稱或輕微之不對稱次數分配上所求得之標準差及其他差量間有次之近似的數理關係。

(i) $$\Delta=\frac{4}{5}\sigma \qquad (公式 68)$$

(但此處之平均差 $\Delta$ 須以 M 爲基準)

(ii) $$Q=\frac{2}{3}\sigma \qquad (前公式 44)$$

(6)在已知其順列號碼時,雖未知各單位之確實數值亦能求得其標準差。

現用自然數

$$1,2,3,\cdots\cdots\cdots\cdots N,$$

以代替

$$X_1,X_2,X_3,\cdots\cdots\cdots\cdots X_N$$

時,其算術平均數與中位數相等,

$$\therefore \quad M=\frac{N+1}{2}。$$

以此 M 爲基準所得偏差 x 爲

$$-\frac{N-1}{2},\cdots\cdots-2,-1,0,1,2\cdots\cdots\frac{N-1}{2},$$

平方上述各數值並舉行相加後得

$$\Sigma x^2=2\left[1+2^2+3^2+\cdots\cdots\left(\frac{N-1}{2}\right)^2\right]$$

$$=2\left[\frac{1}{21}N(N^2-1)\right]$$

$$=\frac{1}{12}N(N^2-1),$$

$$\sigma^2=\frac{\Sigma x^2}{N}=\frac{1}{12}(N^2-1),$$

$$\therefore \quad \sigma=\sqrt{\frac{1}{2}(N^2-1)} \qquad \text{(公式 69)}$$

(7)與數學分科之誤差論(theory of error)有密切之關係，故在統計之數理研究上常帶有基本之任務。

(8)在對稱及輕微之不對稱分配上，在

($M-\sigma$, $M+\sigma$)範圍內含有總次數之 68.26%，

($M-2\sigma$, $M+2\sigma$)範圍內含有總次數之 95.46%，

($M-3\sigma$, $M+3\sigma$)範圍內含有總次數之 99.73%，

（詳細證明請參照第四編之大數法則論）

依此知標準差愈小時，各單位值愈密集於平均數之附近。且在算術平均數 M 之附近取六倍標準差之距離時，幾將次數之全部含在其內。

(9)樣本(sample)選擇之影響較其他差量爲少。

(2)劣點

(1)受有極大及極小值之影響。但將此劣點與其優點相較，卽知標準差實優多劣少，故在求正確可靠之差量上，實爲唯一之適當數值也。

## 第六節 其他之差量

### (A)均互差

本法爲意之 Gini (Corrado) 教授所最初創造，( Variabilità e Muitabilitià, Frscicolo I°, Bologna, 1912, p. 19 seq.)用於研究人口生物經濟等之相互變動上，此法至今雖尙不爲世所通用，但其方法之概念則甚簡單，且甚合於原理。現將其方法說明如次：

設 $X_1, X_2, \cdots\cdots X_n$ 爲 n 個單位值，其相互間之差量爲

$$X_i - X_k \qquad (i,k=1,2,\cdots\cdots n),$$

如此差量共有

$$_nC_2 = \frac{n(n-1)}{2}$$

個，但均互差(mean difference)是爲此等差量之算術平均數，用 g 以表示之。其數學公式如次：

$$g=\frac{1}{\frac{n(n-1)}{2}}\Big[(X_n-X_1)(X_n-X_2)\cdots\cdots(X_n-X_{n-2})(X_n-X_{n-1})$$

$$+(X_{n-1}-X_1)(X_{n-1}-X_2)\cdots\cdots(X_{n-1}-X_{n-2})$$

$$+\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$+(X_3-X_1)+(X_3-X_2)$$

$$+(X_2-X_1)\Big]$$

$$=\frac{2}{n(n-1)}\Big[(n-1)X_n+(n-3)X_{n-1}+(n-5)X_{n-2}+\cdots\cdots$$

$$+(1-n)X_1+(3-n)X_2+(5-n)X_3+\cdots\cdots\Big],$$

即 $$g=\frac{2}{n(n-1)}\Big[(n-1)(X_n-X_1)+(n-3)$$

(公式70)

$$(X_{n-1}-X_2)+(n-5)(X_{n-2}-X_3)+\cdots\cdots\Big]$$

(例)英格蘭與威爾斯各大都市全體在 1902 年之 52 週間每人口萬人每週死亡人數表依大小依順序排列,如第 104 表 a, b 二欄,此 a, b 二欄內所有數值之個數共為 52 個,

a 之個數十 b 之個數 =52,

依表內之計算求得均互差為

$$g=32659\times\frac{2}{52\times51}=24.63。$$

第 104 表

| 人口10,000人之每週死亡人數 | | 均互差之計算 | | |
|---|---|---|---|---|
| a | b | a−b (1) | 52−(2K−1) (2) | (1)×(2) |
| 244 | 136 | 108 | 51 | 5,508 |
| 233 | 139 | 94 | 49 | 4,606 |
| 226 | 141 | 85 | 47 | 3,995 |
| 209 | 143 | 66 | 45 | 2,970 |
| 206 | 144 | 62 | 43 | 2,666 |
| 201 | 145 | 56 | 41 | 2,296 |
| 196 | 149 | 47 | 39 | 1,833 |
| 196 | 150 | 46 | 37 | 1,702 |
| 196 | 151 | 45 | 35 | 1,575 |
| 191 | 152 | 39 | 33 | 1,287 |
| 183 | 154 | 27 | 31 | 899 |
| 182 | 155 | 27 | 29 | 788 |
| 182 | 159 | 23 | 27 | 621 |
| 181 | 160 | 21 | 25 | 525 |
| 179 | 164 | 15 | 23 | 345 |
| 177 | 165 | 12 | 21 | 252 |
| 177 | 166 | 11 | 19 | 209 |
| 177 | 166 | 11 | 17 | 187 |
| 176 | 167 | 9 | 15 | 135 |
| 176 | 169 | 7 | 13 | 91 |
| 176 | 169 | 7 | 11 | 77 |
| 174 | 169 | 5 | 9 | 45 |
| 174 | 170 | 4 | 7 | 23 |
| 174 | 170 | 4 | 5 | 20 |
| 173 | 172 | 1 | 3 | 3 |
| 173 | 172 | 1 | 1 | 1 |
| 計 9,028 | | | | 32,659 |

(B)對數標準差

對數標準差是用以測量幾何平均數附近之密集程度，雖其計算較

爲繁雜，但在經濟統計之研究上亦常用以爲測定差量之數值。其數學公式如次：

設 X＝單位値， f＝次數， G＝幾何平均數，

$\sigma_{\log}$ (或用 $\sigma_g$)＝對數標準差

時，其計算公式爲

$$\sigma_{\log}(\text{或}\sigma_g)=\sqrt{\frac{\Sigma f(\log X-\log G)^2}{\Sigma f}} \qquad \text{(公式 71)}$$

此 $\sigma_g$ 爲一對數值，故欲變爲一自然數的標準差時，非再用逆對數(anti-log)不可。

〔例〕

第 105 表

| 單位値 X | 次數 f | log X | d=log X—log G, =log X－0.34779 | $d^2$ | $fd^2$ |
|---|---|---|---|---|---|
| 1.600 | 3 | 0.20412 | －0.14367 | 0.02064 | 0.06192 |
| 1.800 | 8 | 0.25537 | －0.09252 | 0.00856 | 0.06848 |
| 2.000 | 13 | 0.30103 | －0.04676 | 0.00219 | 0.02847 |
| 2.200 | 20 | 0.34243 | －0.00537 | 0.00003 | 0.00060 |
| 2.400 | 7 | 0.38021 | 0.03242 | 0.00105 | 0.00735 |
| 2.600 | 5 | 0.41497 | 0.06718 | 0.00451 | 0.02255 |
| 2.800 | 1 | 0.44716 | 0.09937 | 0.00987 | 0.00987 |
| 3.000 | 4 | 0.47712 | 0.12933 | 0.01673 | 0.06692 |
| 3.200 | 2 | 0.50515 | 0.15736 | 0.02476 | 0.04952 |
| 3.400 | 1 | 0.53148 | 0.18369 | 0.03374 | 0.03374 |
| 3.600 | 1 | 0.55630 | 0.20851 | 0.04348 | 0.04348 |
| 計 | 65 | | | | 0.39290 |

由表求得 log G＝0.34719

$$\therefore\quad \sigma_{\log}=\sqrt{\frac{\Sigma f(\log X-\log G)^2}{N}}=\sqrt{\frac{0.39290}{65}}$$

$$=\sqrt{0.00604}=0.07772$$

## 第七節 差量之性質

欲明瞭差量之性質，必須先研究上述各種差量之共通性質。現擇其主要者略述數種如次，但其證明只說就最困難之標準差上施行之。

(A)各單位值不相一致時，其差量必爲大於零之正數。

〔證〕本性質依差量之意義卽能明白了解，其數式之證明如次：

$$\sigma=\sqrt{\frac{1}{N}\Sigma f x^2}$$

內之 f 皆正數，$x^2$ 是無論如何爲負數，故欲

$$\sigma=0$$

時非所偏差 x 均等於零不可，卽

$$x_1=X_1-M=0,\ x_2=X_2-M=0,\cdots\cdots x_m=X_m-M=0,$$

如此得

$$X_1=X_2=\cdots\cdots=X_m=M,$$

知各單位值相等時，其差量始能等於零，否則必爲大於零之正數。

(B)若不改變單位值，則雖 p 倍其次數時，其所得差量仍不變。

〔證〕不變單位值 X 而只變次數 f 爲 P.f 時，依算術平均數之定義得

$$\frac{\Sigma Pfx}{\Sigma Pf}=\frac{P\Sigma fx}{P\Sigma f}=\frac{\Sigma fx}{\Sigma f},$$

知所得平均數仍不變動，依此知偏差 $x=X-M$ 數值亦不變動，現定新次數分配表之標準差爲 $\sigma_1$ 時，

$$\sigma_1{}^2=\frac{\Sigma Pfx^2}{\Sigma Pf}=\frac{p\Sigma fx^2}{p\Sigma f}=\frac{\Sigma fx^2}{\Sigma f}=\sigma^2,$$

$$\therefore \quad \sigma_1=\sigma,$$

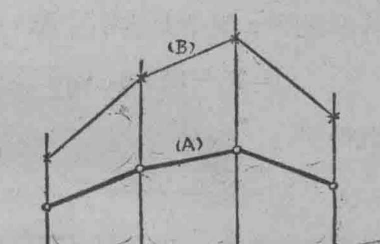
參觀第 117 圖卽知其差量不起變化。

第 106 表

| $X$ | $f$ | $(f)$ | $x$ | $(f)x^2$ |
|---|---|---|---|---|
| $X_1$ | $f_1$ | $pf_1$ | $x_1$ | $pf_1x_1{}^2$ |
| $X_2$ | $f_2$ | $pf_2$ | $x_2$ | $pf_1x_2{}^2$ |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| $X_m$ | $f_m$ | $pf_m$ | $x_m$ | $pf_mx^2{}_m$ |
| 計 | $\Sigma f$ | $p\Sigma f$ | | $p\Sigma fx^2$ |

(C)雖加常數 b 於單位值上而不變動次數時，其差量仍不變。

〔證〕將單位值 X 變爲 X+b 時，其算術平均數 M 變爲 M+b，偏差爲



第 117 圖

$$x=(X+b)-(M+b)=X-M$$

知仍不變，

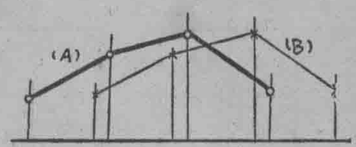
$$\therefore \quad \sigma_1{}^2=\frac{\Sigma fx^2}{\Sigma f}=\sigma_1{}^2,$$

即 $\sigma_1=\sigma$。(參照第 118 圖)

第 107 表

| (X) | f | x | $fx^2$ |
|---|---|---|---|
| $X_1+b$ | $f_1$ | $x_1$ | $f_1x_1^2$ |
| $X_2+b$ | $f_2$ | $x_2$ | $f_2x_2^2$ |
| ⋮ | ⋮ | ⋮ | ⋮ |
| $X_m+b$ | $f_m$ | $x_m$ | $f_mx^2_m$ |
| 計 | $\Sigma f$ | | $\Sigma fx^2$ |

(D) P 倍各單位值而不變更次數時之差量是 P 倍其原値。




第 118 圖

〔證〕單位值變爲 PX 次數仍爲 f 時之算術平均數變爲 PM，其偏差爲

$$PX-PM=P(X-M)=Px,$$

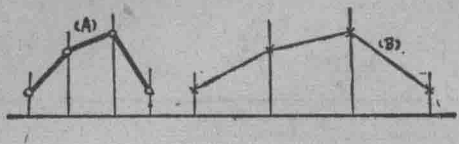
知 P 倍原有偏差，故得

$$\sigma_1^2=\frac{\Sigma f(Px)^2}{\Sigma f}=P^2\frac{\Sigma fx^2}{\Sigma f}=P^2\sigma^2$$

$\therefore\ \sigma_1=P\sigma$ (參照第 119 圖)

第 108 表

| (X) | f | (x) | $f(x)^2$ |
|---|---|---|---|
| $px_1$ | $f_1$ | $px_1$ | $pf_1x_1^2$ |
| $px_2$ | $f_2$ | $px_2$ | $pf_2x_2^2$ |
| ⋮ | ⋮ | ⋮ | ⋮ |
| $px_m$ | $f_m$ | $px_m$ | $pf_mx^2_m$ |
| 計 | $\Sigma f$ | | $p\Sigma fx^2$ |




第 119 圖

上述各性質能明白了解後，對於計算上實多補助處。

〔例〕

第 109 表

(1)

| X | f |
|---|---|
| 50 | 4 |
| 60 | 7 |
| 70 | 9 |
| 80 | 6 |
| 90 | 3 |

(2)

| X | f |
|---|---|
| 65 | 4 |
| 75 | 7 |
| 85 | 9 |
| 95 | 6 |
| 105 | 3 |

(3)

| X | f |
|---|---|
| 50 | 12 |
| 60 | 21 |
| 70 | 27 |
| 80 | 18 |
| 90 | 9 |

(4)

| X | f |
|---|---|
| 25 | 4 |
| 30 | 7 |
| 35 | 9 |
| 40 | 6 |
| 45 | 3 |

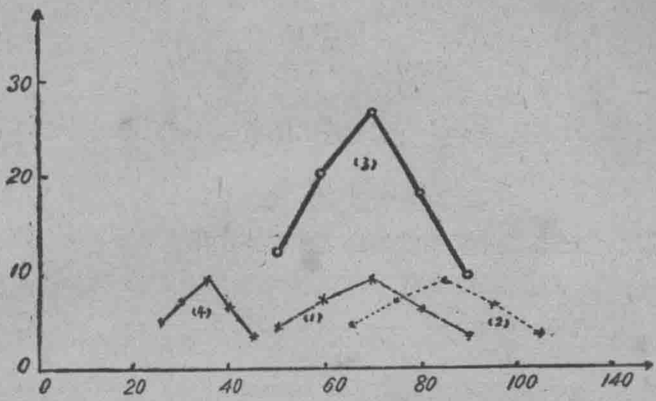
在此四個次數分配表內，若已知 (1) 之差量爲 $\sigma$ 時，其他各次數分配表之差量即能依上述各性質求得爲

(2)之差量 $= \sigma$（依（C）之性質），

(3)之差量 $= \sigma$（依（B）之性質），

(4)之差量 $= \frac{\sigma}{2}$（依（D）之性質）。

將此等次數分配圖示如第 120 圖時，即能依其曲線形狀以明瞭一切。




第 120 圖

## 第八節　比較的差量

上述各種差量是只關係於一個分布狀況上，故稱為絕對的差量。但欲將此絕對的差量以比較同種類之二個以上分配狀況時，實勢所不能。

例如比較甲乙二校之成績，甲校之滿分定為十分，現由其分數表求得

$$M=6.5, \qquad \sigma=1.8。$$

但乙校之滿分定為 100 分，由該校之分數表求得

$$M=62, \qquad \sigma=16.2。$$

若用如此所求得之差量，以比較二校學生分數分配上之差異程度，則完全錯誤。蓋因二校評分標準，完全不同故也。在同種類內互相比較時，已有如此之錯誤，何況比較異種類之分配狀況，故必須別想方法，以測定其相互間比較的差量。披爾遜(K. Pearson)用

$$V=\frac{\sigma}{M}\times 100 \qquad \text{(公式 72)}$$

之數值為比較的差量，此 V 稱為變化係數 (coefficient of Variation)。

〔例〕比較上述二校學生成績時，依本公式得

$$\text{甲校 } V_{\text{甲}}=\frac{1.8}{6.5}\times 100=27.5,$$

$$\text{乙校 } V_{\text{乙}}=\frac{16.2}{62}\times 100=26.1,$$

知甲校之差量比乙校為大，其結果與上例完全相反。

對於比較的差量之測定上此外雖尚有

$$\text{(1)相對的平均偏差}=100\times\frac{\triangle}{M_i} \qquad \text{(公式 73)}$$

$$\text{(2)相對的四分差}=100\times\frac{Q_3-Q_1}{Q_3+Q_1} \qquad \text{(公式 74)}$$

等數種，但為一般所不常用，故略之。

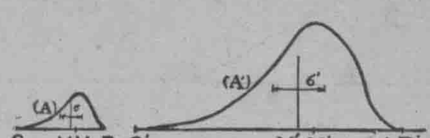
# 第九章　偏斜度

## 第一節　偏斜度之意義

依上述知平均法是用以求數列集中趨勢之代表物，差量是用以認識數列之恆同性與偏差性，但對於集中趨勢之左右二側，依如何形式以行分配則尙未談及。可是在實際上非對稱形之分布曲線，非常衆多，此等分布曲線與理想的對稱分布曲線相偏程度亦有研究之必要。此相差程度，吾人稱之爲偏斜度(measure of skewness)，以下將其測定方法，依次說明。

## 第二節　披爾遜公式

在理想的對稱分布曲線上，其算術平均數 M 與最衆數 $M_0$ 是相一致。但在非對稱的分布曲線上，則其 $M_0$ 必偏於 M 之左方或右方。現取此二值之差 $M-M_0$，以測量相偏之程度，在理想上若甚合法，但在實際上 $M-M_0$ 一數值，是爲測定偏斜之絕對數而非相對數，故在第 121 圖內 A,A′ 二曲線是爲相似的分布曲線，其分布距離 $\overline{C'D'}$ 之長如爲 $\overline{CD}$ 之 P 倍 (圖內 P=3)時，A′ 之 $M'-M_0'$ 亦非比 A 之 $M-M_0$ 大 P 倍不可，成爲 A′ 之偏斜度 P 倍於 A，但在實際




第　121　圖

上其二分配曲線完全相似，由此知只用 $M-M_0$ 數值有時仍難定其偏斜程度，故非改用相對數值不可。現先從簡單入手，取 $M-M_0$ 與分配距離之比當作爲偏斜度，在此時因有

$$\frac{M'-M_0'}{\overline{C'D'}}=\frac{P(M-M_0)}{P\overline{CD}}=\frac{M-M_0}{\overline{CD}}$$

之關係，覺本法倘能適用。但其所使用之分配距離倘非安定數值，若遇單位值上有極端數值時，其距離上大起變化，故須改用安定數值。披爾遜特用標準差 $\sigma$ 以代替距離，當此時因有

$$\frac{M'-M_0'}{\sigma'}=\frac{M-M_0}{\sigma}\qquad(\because\quad \sigma'=P\sigma)$$

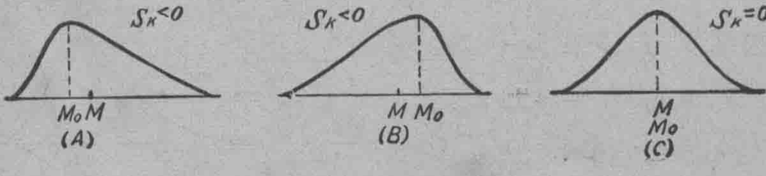
之關係，故得定

$$\frac{M-M_0}{\sigma}$$

爲所求之偏斜度，用 $S_k$ 以表示之，即

$$S_k=\frac{M-M_0}{\sigma},\qquad\text{(公式 75)}$$

依此 $S_k$ 之情形，得測次數曲線之偏斜程度。




第　122　圖

在 $S_k>0$ 時 $M>M_0$ 知曲線偏於左方〔第 122 圖(A)〕,

$S_k<0$ 時 $M<M_0$ 知曲線偏於右方〔第 122 圖(B)〕,

$S_k=0$ 時 $M=M_0$ 知曲線爲對稱形〔第 122 圖(C)〕。

以上是依 $M_0$, M 以求偏斜度，若使用中位數 $M_i$ 與算術平均數 M 時,則依

$$M-M0 \fallingdotseq 3(M-M_i)$$

之公式將上式變爲

$$S_k \doteqdot \frac{3(M-M_i)}{} \qquad \text{(公式 76)}$$

〔例〕

第 110 表

| 組　別 | f | X | $x'=X-A$ | $\bar{x}'=\frac{x'}{a}$ | $f\bar{x}'$ | $f\bar{x}'^2$ |
|---|---|---|---|---|---|---|
| 0—5 | 3 | 2.5 | −15 | −3 | −9 | 27 |
| 5—10 | 11 | 7.5 | −10 | −2 | −22 | 44 |
| 10—15 | 20 | 12.5 | −5 | −1 | −20 | 20 |
| 15—20 | 19 | A—17.5 | 0 | 0 | (−51) | 0 |
| 20—25 | 13 | 22.5 | 5 | 1 | 13 | 13 |
| 25—30 | 9 | 27.5 | 10 | 2 | 18 | 36 |
| 30—35 | 5 | 32.5 | 15 | 3 | 15 | 45 |
| 35—40 | 2 | 37.5 | 20 | 4 | 8 | 32 |
| 40—45 | 1 | 42.5 | 25 | 5 | 5 | 25 |
| 計 | 83 | | | | +59<br>−51<br>+8 | 242 |

依第 110 表得

A=17.5　　a=5　　N=83,

$\Sigma f\overline{x'}=8,\qquad \Sigma f\overline{x'^2}=242,$

$$M=A+\frac{\Sigma fx'}{N}a=17.5+0.1\times5=18.0。$$

$$M_i=L_r+\frac{\frac{N}{2}-F}{f_r}\times a=15+\frac{41.5-34}{19}\times5=17.0$$

$$M_0=L_{r-1}-\frac{\Delta f_{r-2}}{\Delta^2 f_{r-2}}a=10-\frac{9}{-10}\times5=14.5,$$

$$\because\left\{\begin{array}{lllll} f_{r-2} & 11 & & & \\ & & 9 & & \\ f_{r-1} & 20 & & -10 & \\ & & -1 & & \\ f_r & 19 & & & \end{array}\right\}$$

$$\bar{\sigma}^2=\bar{S}^2-\bar{d}^2=\frac{24^2}{83}-\left(\frac{8}{83}\right)^2=2.92-0.01=2.91,$$

$$\bar{\sigma}=1.70$$

$$\sigma=\bar{\sigma}\cdot a=1.70\times5=8.5,$$

$$S_k=\frac{M-M_0}{\sigma}=\frac{18.0-14.5}{8.5}=+0.41,$$

或

$$S_k\doteqdot\frac{3(M-M_i)}{\sigma}=\frac{3(18.0-17.0)}{8.5}=+0.35。$$

## 第三節 用中位數與四分位差以求偏斜度

在對稱曲線上，中位數 $M_i$ 是在上四分位數 $Q_3$ 與下四分位數 $Q_1$

之中央。但在非對稱曲線上則其上四分位數與中位數 $M_i$ 之距離，不能與由中位數到下四分位數相差距離相等。現利用此相差程度。

$$(Q_3-M_i)-(M_i-Q_1)=Q_3+Q_1-2M_i,$$

以定爲偏斜度，亦未始不可。但該値是爲絕對數値，實仍不免有上節所述錯誤處，故必須使變爲相對數値，用二個四分位數之差 $(Q_3-Q_1)$ 以除上述數値後，始能用作爲測定偏斜度之公式。其數式如次：

$$S_k=\frac{Q_3+Q_1-2M_i}{Q_3-Q_1} \qquad \text{(公式 77)}$$

〔例〕仍用上節例內數字用本節公式求其偏斜度。

第 111 表

| 組別 | f |
|---|---|
| 0——5 | 3 $F_1$ |
| 5——10 | 11 $F_1$ |
| 10——15 | 20……$f_i$ |
| 15——20 | 19 |
| 20——25 | 13……$f_j$ |
| 25——30 | 9 $F_2$ |
| 30——35 | 5 $F_2$ |
| 35——40 | 2 $F_2$ |
| 40——45 | 1 $F_2$ |
| 計 | N=83 |

依第 111 表求得

$M_i=17.0$,　$a=5$,　$F_1=14$,　$f_i=20$,

$L_i=10$,　$F_2=17$,　$f_j=13$　$U_j=25$,

$$\frac{N}{4}=20.75,$$

$$Q_1=L_i+a\times\frac{\frac{N}{4}-F_1}{f_1}=10+5\times\frac{20.75-14}{20}=11.69,$$

$$Q_3=U_j—a\times\frac{\frac{N}{4}-F_2}{f_j}=25-5\times\frac{20.75-17}{13}=23.56,$$

$$S_k=\frac{Q_3+Q_1-2M_i}{Q_3-Q_1}=\frac{23.56+11.69-2\times17.0}{23.56-11.69}=0.1。$$

## 第四節 用偏差以求偏斜度

利用偏差之一次方以作成用以檢討分配狀態之平均差，其二次方是作成標準差，現使用其三次方，則偏差之收縮程度愈形強大，故得用此以求偏斜度。其數式如次：

$$S_k=\frac{1}{\sigma}\sqrt[3]{\frac{\Sigma fx^3}{N}} \qquad \text{(公式 78)}$$

本法對於數理的根據甚爲充足，故在數理統計上，時常使用。但在不需要十分正確之社會科學上，多使用上述二法，以求其偏斜度。

〔例〕

第　112　表

| 組　別 | f | X | $x'=X-A$ | $\bar{x}'=\frac{x'}{a}$ | $f\bar{x}'$ | $f\bar{x}'^2$ | $fx'^3$ |
|---|---|---|---|---|---|---|---|
| 0——5 | 3 | 2.5 | −15 | −3 | −9 | 27 | −81 |
| 5——10 | 11 | 7.5 | −10 | −2 | −22 | 44 | −88 |
| 10——15 | 20 | 12.5 | −5 | −1 | −20 | 20 | −20 |
| 15——20 | 19 | 17.5 | 0 | 0 | (−51) | | (−189) |
| 20——25 | 13 | 22.5 | 5 | 1 | 13 | 13 | 13 |
| 25——30 | 9 | 27.5 | 10 | 2 | 18 | 36 | 72 |
| 30——35 | 5 | 32.5 | 15 | 3 | 15 | 45 | 135 |
| 35——40 | 2 | 37.5 | 20 | 4 | 8 | 32 | 128 |
| 40——45 | 1 | 42.5 | 25 | 5 | 5 | 25 | 125 |
| 計 | 83 | | | | +59<br>−51<br>+8 | 242 | +473<br>−189<br>+284 |

依第 112 表得

$$M=18.0,\qquad A=17.5,$$

$$\sigma=8.5,\qquad a=5,$$

但

$$\Sigma f_k x_k'^3=\Sigma f_k(X_k-A)^3$$

$$=\Sigma f_k\{(X_k-M)+(M-A)\}^3$$

$$=\Sigma f_k\{(X_k-M)^3+3(X_k-M)^2(M-A)$$

$$+3(X_k-M)(M-A)^2+(M-A)^3\}$$

$$=\Sigma f_k(X_k-M)^3+3(M-A)\Sigma f_k(X_k-M)^2$$

$$+(M-A)^3\Sigma f_k,$$

$$\therefore\ \Sigma f_k(X_k-M)^3=\Sigma f_k(X_k-A)^3-3(M-A)\Sigma f_k$$

$$(X_k-M)^2-(M-A)^3\Sigma f_k。$$

用 $\Sigma f_k$ 除上式之兩邊得

$$\frac{\Sigma f_k(X_k-M)^3}{\Sigma f_k}=\frac{\Sigma f_k(X_k-A)^3}{\Sigma f_k}-3(M-A)$$

$$\frac{\Sigma f_k(X_k-M)^2}{\Sigma f_k}-(M-A)^3$$

$$=a^3\frac{\Sigma f\bar{x}'^3}{\Sigma f}-3(M-A)\sigma^2-(M-A)^3,$$

至在本表內

$$\Sigma f\bar{x}'^3=284, \qquad \Sigma f=83,$$

$$\therefore \quad \frac{\Sigma f(X-M)^3}{\Sigma f}=5^3\times\frac{284}{83}-3(18.0-17.5)(8.5)^2$$

$$-(18.0-17.5)^3=536.961,$$

$$\sqrt[3]{\frac{\Sigma f(X-M)^3}{\Sigma f}}=\sqrt[3]{536.961}=8.12,$$

$$\therefore \quad S_k=\frac{1}{\sigma}\sqrt[3]{\frac{\Sigma f(X-M)^3}{\Sigma f}}=\frac{8.12}{8.50}=+0.96。$$

## 第五節 用十分位數以求偏斜度

在十分位數內,若次數分配爲完全對稱時,其 $D_5$ 與 $D_1$ 之差實等於 $D_9$ 與 $D_5$ 之差,即

$$D_5-D_1=D_9-D_5。$$

若次數分配成爲不對稱形時,上述二差數決不能相等,即

$$D_5-D_1\neq D_9-D_5。$$

現利用此原理定上述二差數相加之半

$$D_5-\frac{1}{2}(D_9+D_1)$$

所當値之如何以定偏斜度之情形。其計算公式如次：

$$S_k=D_5-\frac{1}{2}(D_9+D_1) \qquad \text{(公式 79)}$$

〔例〕依第 88 表得

$$D_1=20, \qquad D_5=26.8, \qquad D_9=35.4$$

應用公式 79，得

$$S_k=26.8-\frac{1}{2}(35.4+20)$$

$$=-0.9$$

$$<0$$

知由該表資料所作成之次數曲線是爲右傾曲線。

# 第十章　動差

## 第一節　動差之意義

依上述知在記述次數分配上各特徵之方法中，其最代表的是爲算術平均數(M)標準差($\sigma$)及偏斜度(S)。且在此等統計值之計算進行中，其最小分子是爲各單位值 $X_i$ 與算術平均數相差之偏差(x)。在算術平均數上此等偏差之和等於零，在標準差上是基於此等偏差自乘之總和，又在偏斜度上是基於此等偏差三次方之總和，至偏差四次方之總和又得用以研究分配之尖峯性(kurtasis)。例如依

$$K_2=\frac{1}{\sigma^4}\frac{\Sigma+x^4}{N},$$

知其在

$$K_2=3$$

時，次數分配是爲常態。在

$$K_2<3$$

時，分配成爲過平分配。在

$$K_2>3$$

時，分配成爲過尖分配。

在更欲詳細研究分配之特徵時，倘得使用偏差五次方之總和，六次方之總和等等數值以進行一切。依此知偏差各次方之總和數值，對於研

究次數分配上，實含有重大之功用。現爲總述上項各性質，特提用一簡單數式以表示之。此簡單數式是爲由此等偏差之 n 次方所作成之算術平均數，即

$$\frac{\Sigma fx^n}{N}=m_n \qquad \text{(公式 80)}$$

由本式所算得之數值，K. Pearson 稱之爲第 n 次動差（$n^{th}$ moment），用 $m_n$ 以表示之。（其所以稱之爲動差者，因動差之原文 moment 一字，在力學上所習見名詞，用以測量力之迴轉趨勢，此趨勢之大小在同一力上，則依着力點與原點距離之遠近而異。但統計學上之動差則與此意義相類似，若將各組之次數看做爲力學上之力，單位值與算術平均數相差之偏差看做爲着力點與原點之距離時，其性質完全相同，故 Pearson 特用 moment 一字以代表之。）即

$$第一動差=m_1=\frac{\Sigma f(X-M)}{N}=\frac{\Sigma fx}{N},$$

$$第二動差=m_2=\frac{\Sigma f(X-M)^2}{N}=\frac{\Sigma fx^2}{N},$$

$$第三動差=m_3=\frac{\Sigma f(X-M)^3}{N}=\frac{\Sigma fx^3}{N},$$

......................................................

$$第\ n\ 動差=m_n=\frac{\Sigma f(X-M)^n}{N}=\frac{\Sigma fx^n}{N}。$$

用此動差符號以表示標準差偏斜度及尖峯度時，得

$$\sigma + \sqrt{m_2},$$

$$S = \frac{1}{\sqrt{m_2}} \sqrt[3]{m_3}, \qquad \text{(公式 81)}$$

$$K_2 = \frac{m_4}{m_2^2},$$

又因在算術平均上其偏差之和爲 0，故得

$$m_1 = 0。$$

但在統計學上之動差，不只依以算術平均數爲基準，實得依任意之基準數值以計算一切。現用 $m'_n$ 以表示由任意基準數值所算得之第 n 動差時，其關於算術平均數之動差不過爲其中之特定者。

設各單位值(X)之關於任意值 A 的偏差爲

$$x_i' = X_i - A \qquad (i = 1, 2, 3, \cdots\cdots)$$

時，由此偏差所算得之各種動差，爲

$$\text{第一動差} = m'_1 = \frac{\Sigma f x'}{N},$$

$$\text{第二動差} = m'_2 = \frac{\Sigma f x'^2}{N},$$

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

$$\text{第 n 動差} = m'_n = \frac{\Sigma f x'^n}{N},$$

此等關於任意數值爲基準所求得之動差與以算術數爲基準所算得之動差間實有一定之關係，現將其內容說明如次：

在計算以算術平均數為基準之動差時，須使用各單位值與算術平均數之偏差，但此偏差因算術平均數常含有小數，以致亦含有小數，使計算趨於煩雜。故為使計算簡單起見，特依前述算術平均數及標準差等之簡便求法，先求出關於任意數值為基準之動差，而後再由此等動差以求出其關於算術平均數之動差。至二者間之關係式則如次：

設 $$m_n=\frac{\Sigma fx^n}{N},$$

$$m'_n=\frac{\Sigma fx'^n}{N},$$

但 $$x=X-M=(X-A)+(A-M)=x'-d,$$

$$\therefore\quad m_n=\frac{1}{N}\Sigma fx^n=\frac{1}{N}\Sigma f(x'-d)^n$$

$$=\frac{1}{N}\Sigma f[x'^n-ndx'^{n-1}+\frac{n(n-1)}{2}d^2x'^{n-2}-\cdots\cdots\cdots$$

$$+(-1)^nd^n]$$

$$=\frac{1}{N}\Sigma fx'^n-nd\frac{1}{N}\Sigma fx'^{n-1}+\frac{n(n-1)}{2}d^2\frac{1}{N}\Sigma fx'^{n-2}-\cdots$$

$$+(-1)^nd^n\frac{1}{N}\Sigma f.$$

$$=m'_n-ndm'_{n-1}+\frac{n(n-1)}{2}d^2m'_{n-2}-\cdots\cdots+(-1)^nd^n,$$

取 n=1,2,3 及 4 時，得

$$m_1 = 0,$$
$$m_2 = m'_2 - m'^2_1,$$
$$m_3 = m'_3 - 3m'_1 m'_2 + 2m_1^3, \qquad \text{(公式82)}$$
$$m_4 = m'_4 - 4m'_1 m'_3 + 6m'^2_1 m'_2 - 3m'^4_1;$$

若取原點爲其任意數值時,得

$$d = M,$$

$$\text{第 n 動差} = \overline{m}'_n = \frac{\Sigma f X^n}{N},$$

將上述關係成爲

$$m_n = \overline{m}'_n - nM\overline{m}'_{n-1} + \frac{n(n-1)}{2}M^2\overline{m}'_{n-2} - + \cdots\cdots,$$

$$+ (-1)^n M^n。$$

本節內所述動差,爲數理統計分析上重要工具之一。但在數理分析上,多將偏差看做爲連續變數,依此得用積分記號以代替總和,其關於任意數值 a 之第 n 動差之數式爲

$$\mu'_n = \frac{\int_{-\infty}^{\infty}(x-a)^n F(x)dx}{\int_{-\infty}^{\infty}F(x)dx} = \frac{1}{N}\int_{-\infty}^{\infty}(x-a)^n F(x)dx,$$

關於原點之第 n 動差爲

$$\bar{\mu}'_n = \frac{\int_{-\infty}^{\infty}x^n F(x)dx}{\int_{-\infty}^{\infty}F(x)dx} = \frac{1}{N}\int_{-\infty}^{\infty}x^n F(x)dx,$$

關於算術平均數之第 n 動差爲

$$\mu_n=\frac{\int_{-\infty}^{\infty}x(x-M)^nF(x)dx}{\int_{-\infty}^{\infty}F(x)dx}=\frac{1}{N}\int_{-\infty}^{\infty}(x-M)^nF(x)dx,$$

上述各式內之 F(x) 是表示次數分配之方程式，與普通之次數 f 相當。

## 第二節 薛伯氏校正法

一般計算統計值時均不計算各組內各個單位值之詳細情形，只以該組之中點所當值爲其代表物。因此其所求得之結果值上，均不免含有若干誤差。薛伯氏(X. F Sheppard)特想出種種方法以行校正。

設原點處之第 n 動差爲

$$\bar{\mu}_n=\int_{-\infty}^{\infty}x^n\varphi(x)dx,\quad \left[\text{但 }\varphi(x)=\frac{1}{N}F(x)\right]$$

用 Σ 記號以表示時得

$$\bar{\mu}_n=\sum_{i=-\infty}^{\infty}x_i^n\varphi(x_i)(x_i-x_{i-1}),$$

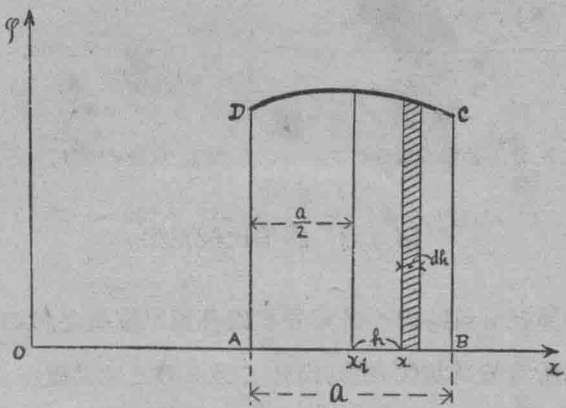
用次數分配表以記算時，須改爲

$$\bar{m}_n=\sum_i x_i\left[\int_{-\frac{a}{2}}^{\frac{a}{2}}\varphi(x_i+h)dh\right],$$

式內之 $x_i$ 爲第 i 組之中點所當值，a 爲組距，

$$\int_{-\frac{a}{2}}^{\frac{a}{2}}\varphi(x_i+h)dh=\square ABCD\text{ 之面積}$$

$$=\text{第 i 之組之次數},$$




第 123 圖

設 φ(x) 爲連續函數時,依 Taylr 定理以行展開如次:

$$\varphi(x_i+h)=\sum_{r=0}^{\infty}\frac{k^r}{r!}\varphi^{(r)}(x_i),$$

$$\int_{-\frac{a}{2}}^{\frac{a}{2}}\varphi(x_i+h)dh=\sum_{r=0}^{\infty}\frac{\varphi^{(r)}(x_i)}{r!}\int_{-\frac{a}{2}}^{\frac{a}{2}}h^r dh$$

$$=\sum_{r=0}^{\infty}\frac{\varphi^{(r)}(x_i)}{r!}\cdot\frac{1}{r+1}\left[\left(\frac{a}{2}\right)^{r+1}-\left(-\frac{a}{2}\right)^{r+1}\right]$$

$$\left\{\text{此處}\left[\left(\frac{a}{2}\right)^{r-1}-\left(-\frac{a}{2}\right)^{r+1}\right]\right.$$

$$=\left\{\begin{matrix}0 & \text{當 } r=\text{奇數}\\ \frac{1}{2^r}a^{r+1} & \text{當 } r=\text{偶數}\end{matrix}\right\}$$

$$=\sum_{k=0}^{\infty}\frac{\varphi^{(2k)}(x_i)}{(2k+1)!}\frac{1}{2^{2k}}a^{2k+1}$$

$$\therefore\ \overline{m}_n=\sum_{i=-\infty}^{\infty} x_i^{\,n}\sum_{k=0}^{\infty}\frac{\varphi^{(2k)}(x_i)}{(2k+1)!}\frac{1}{2^{2k}}a^{2k+1}$$

$$=\sum_{k=0}^{\infty}\frac{a^{2k+1}}{2^{2k}(2k+1)!}\sum_{i=-\infty}^{\infty}x_i^{\,n}\varphi^{(2k)}(x_i)$$

$$=\sum_{k=0}^{\infty}\frac{a^{2k}}{2^{2k}(2k+1)!}\sum_{i=-\infty}^{\infty}x_i^{\,n}\varphi^{(2k)}(x_i)\times a$$

$$=\sum_{k=0}^{\infty}\frac{a^{2k}}{2^{2k}(2k+1)!}\int_{-\infty}^{\infty}x^n\varphi^{(2k)}(x)dx$$

但函數 $\varphi(x)$ 及其微係數在 $x=-\infty$，及 $+\infty$ 處均比 $x^n$ 爲高次無限小，故上式內之積分得用部分積分法以求得之，即

$$\int_{-\infty}^{\infty}x^n\varphi^{(2k)}(x)dx=n(n-1)\cdots\cdots(n-\overline{2k-1})\int_{-\infty}^{\infty}x^{n-2k}\varphi(x)dx$$

$$=(2k)!\,{}_nC_{2k}\overline{\mu}_{n-2k},$$

$$\therefore\ \overline{m}_n=\sum_{k=0}^{\infty}\frac{a^{2k}}{2^{2k}(2k+1)!}(2k)!\,{}_nC_{2k}\overline{\mu}_{n-2k}$$

$$=\sum_{k=0}^{\infty}\frac{{}_nC_{2k}}{2^{2k}(2k+1)!}a^{2k}\overline{\mu}_{n-2k}$$

$$=\overline{\mu}_n+\frac{{}_nC_2}{2^2\cdot 3}a^2\overline{\mu}_{n-2}+\frac{{}_nC_4}{2^4\cdot 5}a^4\overline{\mu}_{n-4}+\cdots\cdots,$$

取 $n=0,\ 1,\ 2,\ 3,\ 4,\ 5$ 時，得

$$\overline{m}_0=\overline{\mu}_0,$$

$$\overline{m}_1=\overline{\mu}_1,$$

$$\overline{m}_2=\overline{\mu}_2+\frac{1}{12}a^2\overline{\mu}_0,$$

$$\overline{m}_3=\overline{\mu}_3+\frac{1}{4}a^2\overline{\mu}_1,$$

$$\overline{m}_4=\overline{\mu}_4+\frac{1}{2}a^2\overline{\mu}_2+\frac{1}{80}a^4\overline{\mu}_0,$$

$$\overline{m}_5=\overline{\mu}_5+\frac{5}{6}a^2\overline{\mu}_3+\frac{1}{16}a^4\overline{\mu}_1。$$

上式內之 $\overline{m}_n$ 是由次數分配表所算得者，故稱粗假動差（rough moment），$\overline{\mu}_n$ 是由次數曲線以精密算得者，故稱爲眞正動差（true moment），依此知眞正動差得由粗假動差表示之如次：

$$\overline{\mu}_0=\overline{m}_0$$

$$\overline{\mu}_1=\overline{m}_1$$

$$\overline{\mu}_2=\overline{m}_2-\frac{1}{12}a^2\overline{m}_0$$

$$\overline{\mu}_3=\overline{m}_3-\frac{1}{4}a^2\overline{m}_1 \qquad \text{(公式 83)}$$

$$\overline{\mu}_4=\overline{m}_4-\frac{1}{2}a^2\overline{m}_2+\frac{7}{140}a^4\overline{m}_0$$

$$\overline{\mu}_5=\overline{m}_5-\frac{5}{6}a^2\overline{m}_3+\frac{7}{48}a^4\overline{m}_1。$$

若取原點在平均數上時（M=0），依前節公式知只將 $\overline{\mu}_n$，$\overline{m}_n$ 上之「一」取去，並應用

$$m_0=1,$$

$$m_1=0,$$

卽得

$$\mu_2=m_2-\frac{1}{12}a^2,$$

$$\mu_3=m_3,$$

$$\mu_4=m_4-\frac{1}{2}a^2m_2+\frac{7}{240}a^4, \qquad (公式\ 84)$$

$$\mu_5=m_5-\frac{5}{6}a^2m_3,$$

詳情請參照

W. F. Sheppard: Proceedings of the London Mathematical Society Vol. XXIX. PP.353 以下。