統計概論

芮寶公著

中華書局印行

芮寶公著

統計概論

中華書局印行

引言

本書之目的,在供初習統計者參考或普通統計學程採作課本之用。故述理力求淺顯,關於公式之來源,避免引用高深數學之證明。在討論每一種方法之後,均附以實例,詳爲解述,俾事業界於工作之餘,欲求得統計知識者,閱之亦易領悟。

本書前四分之一敍述統計之基本理論,搜集與整理材料之方法,暨圖表之繪製。後四分之三則討論次數數列之分析方法,最後并殿以統計歸納與取樣問題一章,舉凡初級統計學之內容,似均已包括在內。

本書所用名詞悉以中國統計學社二十三年七月印行之統計譯名爲準;其未經該社通過者,則採用最通用之名詞。

本書所用符號悉照 F. C. Mills 所著 Statistical Methods 一書中所規定者。

作者前在上海復旦大學,無錫江蘇省立教育學院及國立中央大學法學院講授統計學時,曾採用本書爲講稿之一部,先後雖數經删改,謬誤之處,仍所難免,幸統計界同志,賜予指正。

本書承盛灼三朱君毅兩先生鼓勵付印,金國寶蔡正雅兩先生多所教益,均爲作者所深感.友人顧君夢五許君貴忻胞弟光懷幫同繪製書中圖表,兼爲校對,併此誌謝。

芮寶公　二十三年十月於南京

統計概論

目　次

表 次

圖 次

統計概論

第一章　概論

第一節　統計之定義

據一八六九年在荷蘭海牙舉行之第七次國際統計會議所報告,統計之定義,已有一百八十種之多,以後學者所倡用者,尚不在內。足見爲統計下一完善之界說殊非易事。茲採美儒金氏(Willford I. King)之定義,略加損益,迻譯如下:〔註一〕

統計者,乃根據所搜集之大量的事實,或估計之數目,加以分析,以研究一般自然現象或社會現象(Collective Natural or Social Phenomena)之方法也。

第二節　統計法與實驗法之比較

統計法與實驗法不同之處,約有三點:

1. 研究自然科學以實驗法爲主,研究社會科學則以統計法爲主。

一:金氏定義原文:—The science of statistics is the method of judging collective natural or social phenomena from the results obtained by the analysis of an enumeration or collection of estimates.

2. 實驗法對於觀察之對象,可以個別研究;而統計法則非根據大量觀察(Mass Observation)不可。

3. 實驗法之結果常絕對準確,如氫二氧一等於水;而統計法之結果,則不過逼近於準確而已。〔註一〕

第三節 大量觀察法

宇宙間之現象至為繁複,吾人殊不易明瞭其眞相,然若根據大量觀察法,取其大而遺其小,未嘗不可以得其梗概。猶之觀海,近望則波濤萬千,起伏無定;若從遠處望之,自有平線可尋。試就社會現象而言,雖若萬緒千頭,其間亦恆有一定之法則存在,所謂大數律(The Law of Large Numbers)是也。茲分兩項說明如下:

一、大數常性律(Law of Statistical Regularity) 統計學最重要之特色,在能執簡馭繁,探求眞理。曷為執簡馭繁?即研究某種現象時,不必一一調查其全體,但隨機(at random)抽取全體中之一部份加以研究,即能得相當準確之結果。如吾人欲調查全國工人之平均工資,固[illegible]調查,倘能抽取一部份適當的「樣本」(Samples),按合理的計算,以求其平均數,則此[illegible]

註一:參看 C. S. Li: Business statistics. PP. 2-3. [illegible]

全體工人工資平均之結果,當無大出入。設所取的「樣本」增多,其結果愈能代表全體,此大數常性之定律也。

二、大數惰性律(Law of Inertia of Large Numbers)

根據大數觀察時,吾人所取之「樣本」未必一一適與平均之結果相符,其中有過者,亦有不及者。如一地產麥之額,年有增減,但就全世界之產額言之,則比較的變動不多,故吾人可以根據已往若干年各國各地麥之產額,以預測全世界收穫之總量。設某地忽告災荒而歉收,遠遜預計之額,然他方面或有收成特別豐稔者以補償某地之不足,結果全世界預測之總量,影響甚微,此大數惰性之定律也。惟吾人須注意者,所謂大數惰性,非經常期間而不變也。不過謂根據大數觀察所得之結果,較之得諸少數「樣本」者,為有規則耳。

大數之有規律,不獨社會現象為然,即自然界之現象與機會亦莫不如此(詳見第十一、十二兩章)。本節所述之大數常性律,其理論係建立於數學中機率原理(The Theory of Probability)之上,而大數惰性律又係從大數常性律推演而出者也。

第四節　統計之功用

統計之功用,不勝枚舉,茲擇其重要者,略述如下:

一、統計與社會政策:近世各國無不倡言社會政策,社會立法;然欲醫治社會之疾病,必先明瞭其原因,然後對症下藥,方有改良之望。欲明社會疾病之原因則非取證於統計不可。有失業統計,於是知強迫保險之必要;有工資統計,於是知女工童工須特別加以保護,而有最低工資法之制定,卽其例也。

二、統計與公共衞生:疾病統計與死亡統計對於公共衞生有密切之關係。衞生當局之唯一參考,厥爲統計。蓋其以前設施之成績,固賴統計以表明;而以後施政之方針,尤賴統計以取決。當疾病發生之際,可由統計之報告,設法防止其蔓延。平時亦可藉統計以宣傳,灌輸衞生常識於民間,減少疾病之發生。

三、統計與商業:現代商業範圍擴大,故問題亦日趨複雜,內部如浪費之減少,工人效能之增進,總分公司之營業,售貨員之成績等,外部如供需之狀況以及市場之變遷,商情之盛衰,季節之影響等等,皆與商業之成敗有莫大之關係,

故近代歐美各大公司,皆特設統計部以專司各種調査之職。

四、統計與財政:財政以收支適合爲原則,支出雖較能預定,而收入則頗難預言。例如所得稅之多寡,須視人民收入之數額而定。關稅之收入,大部須視輸入貨品之數量與種類而定。然歐美各國之財政專家每能根據歷年之統計以預測未來之收入,其結果雖不必與實收數目完全相符,然適合者其常,而相差極大者僅例外耳。(註一)

統計之功用,不僅限於事業家,即理論家與科學家亦常利用統計以爲其理論或假設之證明。如經濟學家應用統計而立人口,工資,物價之定則;社會學家以此證明火酒之出售,與犯罪,貧窮及自殺事件之關係;教育心理學家以此證明腦力與年齡之關係;生物學家以此證明變化與遺傳之定律,凡此皆統計之功用也。

第五節　統計之程序

統計方法之程序,須視問題之範圍與性質而定,

註一:所舉統計之功用四點,係節錄金國寶先生編統計學講義。

初無一成不變之程式,惟就大體言之,可分爲下列五種步驟:

一、事前規畫

二、搜集材料

三、整理材料

四、分析材料

五、解釋結果

上述第一至第四四種步驟當於以下各章次第論之,惟第五步解釋結果一項,無一定之方法可述,而其重要之程度,殊不亞於以前四步,以其爲全盤工作之結論也。金氏嘗謂能正確的解釋結果,簡潔明瞭的下一結論,爲優越之統計家所必備之條件。信不誣也。〔註一〕

進行統計工作之程序,除上述五項外,更有不得不注意者,即校核之工作。校核之工作,不僅適用於最後兩步,幾乎在進行統計工作之歷程中,隨時隨事均有校核之必要,蓋必如此而後可將錯誤減至最小限度也。

註一: The power······ to interpret the results correctly and state the conclusions lucidly and succintly is one of the characteristics indispensable in a good statistician.

第二章 事前規畫

在未進行搜集材料以前,下列三事為吾人預先所應規畫:一、確定問題;二、規定適當之因子;三、選定單位.茲分別論之。

第一節 確定問題

當調查工作未進行之先,統計家第一步須確定問題,認清調查之目的,蓋問題稍有變更,調查之手續即將一部或全部不同。如吾人欲調查一地工人之工資,必先決定所欲調查者為貨幣工資(Money Wage or Nominal Wage)抑為實際工資(Real Wage);其次更須決定所欲調查者為計件工資(Piece rate)抑為計時工資(Time rate).此外,調查之目的,僅欲知工人個人每年之收入乎?抑工人全家之收入乎?凡此種種問題之範圍,皆不相同,因此,進行之方法亦各異,故於調查之先,須將目的認清。

第二節 規定適當之因子

統計之為用,在乎比較,而比較時相對數量尤較絕對數量為重要。所謂相對數量者,在統計上恆以百分數(Percentage)或比率(Ratio)表示之。例如吾人欲比較各城市人口之死亡率,通常以每千人中死亡若

干之比率比較之;比較各城市自殺之人數佔人口總數之成數時,則以每十萬人中自殺者若干人之比率比較之。統計之結果既常爲一種比率(但並非盡爲比率),則構成分子與分母之兩種因子,必以能適合吾人調査之目的爲條件,否則兩種分子分母不同之比率互相比較,結果必將謬誤也。如一八九八年美國與西班牙戰爭時,美國某報取戰時海軍死亡率與紐約居民之死亡率相比較,其結果海軍死亡率爲千分之九,而紐約居民之死亡率爲千分之十六,於是謂戰時之海軍尙較紐約之居民爲安全。不知因子不同,不能比較。蓋老人與兒童之死亡率高,而城市之中乃有不少老人與兒童,至海軍人員皆當壯年,且其身體曾經嚴格檢查,其死亡率自不能與一般城市居民相提並論。苟欲以海上生活與城市生活二者對於壽命之影響相比較,則必先在城市之中選出若干與海軍同等年齡之居民,並行使同樣嚴格之身體檢査,然後舉行登記或調査,求出每年兩者之死亡率互相比較,始有意義。又如比較兩地暗殺事件之多寡與人口之比率,則人口總數之中,老人與婦孺均須除外,蓋暗殺事件,大多數皆壯年男子爲之也。

第三節 選定單位

統計常不離乎數目,而數目必有其單位,蓋抽象之數字如一千或一萬在統計上直毫無意義也。定統計之單位,初視之似甚簡易,但實際上則頗感困難。如「田莊」之意義,盡人而知,然如用之於統計上,則「田莊」係指數畝之「田莊」乎?抑數十畝或數百畝之「田莊」乎?一「刑事犯」名詞,至爲淺顯,然用之於統計上,亦將有不少困難,蓋所謂「刑事犯」,指殺人者乎?抑放火者乎?倘一人犯殺人之罪,以行賄法官得免於法,其人卽非刑事犯乎?對於此種問題,吾人顯有不可踰越之障礙;故事先須擇定單位。如單位之字義不能十分明瞭,則對於此單位,必須下一嚴確之定義,庶搜集之材料不致錯誤。統計單位,按照瓦特金氏(G. P. Watkins)之分類,有二大類,每大類中,更分二小類如下:

一、單獨事物(Individual things)可以計數者:

甲、自然的事物(Natural kinds),如人口,動植物皆各以其原來之個體爲單位。此類單位意義最爲明瞭且甚正確。

乙、人爲的事物(Produced kinds),此種單位係按人力製造事物之個體計算,如桌、椅、房屋等是。

二、測量單位(Mensurational Units):

甲、度量衡的單位(The Physical Measures),如路之遠近以里計,器皿之容量以立方尺計,貨物之輕重以斤計。

乙、金錢的單位(The Picuniary Units),如我國之銀元,英國之金鎊,法國之法郎。

選擇統計單位,其定義不特須正確,並須有普遍性;穩定性,無論何種抽象之事物,宜以比較的具體之單位表示之,以免籠統含渾之弊。我國市用制度量衡雖經實業部頒布,似尚在推行之中,各地使用之度量衡依然紛亂,統計時必須輾轉折合,此我國從事統計工作者特有之困難也。

第三章 搜集材料

第一節 材料之類別

統計材料,可分爲二種:一爲原始材料 (Primary materials), 一爲次級材料 (Secondary materials)。原始材料者,乃由統計者直接搜集而得之材料。此項材料之優點在於詳實,惟整理需時,搜集不易,非私人勞力,財力所易舉辦。次級材料者,乃他人所刊布已經整理之原始材料。此項材料既經他人整理,必不如原始材料之詳實,況整理時難免無錯誤發生,雖經發覺亦無法改正。故就確度言,原始材料自較次級材料爲可靠。若就搜集時所費之勞力與財力而言,則次級材料未始無一日之長,斯在統計者能善爲利用之耳。

第二節 材料之來源

原始材料之來源,非由實地調查,即由通函訪問,至次級材料之來源,大約有下述四種。

一、政府方面:統計之意識,實與國家之組織而俱生;有雛形之政治組織,即有簡單之統計。近世國家組織隨文明進步而益臻完密,故政府機關莫不以編製統計爲主要職務之一。舉凡一國之人口,政治,經濟,教育,莫不資統計以決定

施政之方針。我國政府機關之統計,以海關貿易冊歷史為最久。〔註一〕國民政府成立後,各部院會及各省市政府幾無不有統計科或股之設置,民國二十年國民政府主計處成立,並特設統計專局掌理全國統計事務。數年以來,所搜集之材料,當不在少數,其中如財政部國定稅則委員會之物價統計,上海市政府社會局之勞工統計,均其最著者也。

二、公共團體:政府機關以外,近幾年來,各公共團體亦有藉統計以表現其業務之狀況或解答各種問題者,教育如中華職業教育社,工商如上海工廠聯合會,紗廠聯合會等均有相當之統計報告,可資採用。

三、學術機關:材料之來自此方面者,多屬專門性質,如天津南開大學與上海中國經濟統計研究所之經濟統計,中華教育基金董事會社會調查部及河北定縣中華平民教育促進會社

註一:我國海關刊行貿易統計年冊,始於咸豐九年(1859),貿易報告年冊始於同治三年(1864),然其內容至為簡略。同治六年(1867)以後始漸詳備,迄今已有六十餘年之歷史。

會調查部之社會統計,在我國學術界均有相當之地位。

四、個人方面:此方面之材料比較最少,然苟有之,其可靠性必高,因此項統計既爲個人所編製,則其人必於統計有精深之研究,濃厚之興趣,乃能有此毅力。我國個人編製統計者尙不多覯,其在美國,若伯卜生 (Roger W. Babson) 之商情預測表,費暄 (Irving Fisher) 之物價指數,均爲統計界極有價值之資料。

第三節　搜集材料之方法

統計材料之來源已如上述。惟原始材料固不易獲得,卽次級材料亦非咄嗟可致。是在統計者愼選方法,持以毅力,運以精心,以從事耳。茲將各種搜集材料之方法臚述於次:

一、用原有表册法　此法爲搜集次級材料之常法,如調查一城中小學兒童之數目,可利用各校之學生點名簿或已編好之學生出席統計表。調查工廠出品之成本,數量等,可根據其各種賬簿或已編之統計。搜集此種材料時,可親自訪問,可通信或派人徵求。此外各種書報,雜誌上之統計資料,平時亦應留心保存,庶用時

免搜集之勞。惟應用次級材料,須注意下列事項:

a. 原來搜集此項材料者之能力及爲人是否可靠。

b. 原始材料之來源,

c. 調查之目的。

d. 搜集材料之方法,

e. 數字之確度。

如於以上各點均已滿意,然後始可應用。

二、實地調查法　此法爲搜集原始材料最佳之方法,因用此法所搜集之材料,其確度常較用他法搜集者爲高。惟採用此法,在未進行調查之先,應確定調查之範圍爲全體調查(Complete Investigation)抑抽樣調查(Sampling Method),時期爲數日,數月或全年。實地調查法更可分爲四種:甲、親自調查法(Personal Investigation),乙、通信估計法(Estimates from Correspondents),丙、被查人塡報法(Schedules to be filled by Informants),丁、僱員調查法(Schedules in charge of Enumerators)。何去何從,應由統計者熟審問題之性質,與夫希望之確度及經費之多寡而擇定。茲分述於下:

甲、親自調查法　此法費用甚省，最合宜於精深之研究，但搜集之範圍，未免太狹。普來(Le Play)研究歐洲工人家庭收支狀況，即採用此法。首住於一工人家庭中，調查數月，然後再移至另一家庭，如是繼續爲之，費數年之光陰，始得一精確之結果，與後世大規模調查所得之結果，相去不遠。但普氏採用此法以研究工人家庭，是否可資取法，殊堪疑問。蓋工人之家庭甚多，個人之精力及時間有限，即使普氏費去畢生之光陰，所調查之家庭，亦屬有限也。全體調查有時固不可能，但亦須抽取一部分適宜的「樣本」，庶可爲全體之代表。況個人成見往往甚深，每於不知不覺中，影響其結果，斯亦此法美中不足之處也。

乙、通信估計法　如吾人研究某種問題，僅欲知其大概或近似的結果，則搜集材料時可用通信估計法，蓋此法既簡便易行，且甚經濟。例如搜集農產物產量之報告，即可採用此法。報告者對於當年收穫之確實數量，雖不能預知，若請彼等估計本年之收穫量約

較上年或常年 (Normal year) 增加幾成或減少幾成,彼等似不難回答。此種各個人之估計,雖未必準確,其中有估計過多者,有估計過少者,但此項錯誤係可以互相補償的(Compensating)而非累積的(Cumulative),故平均結果雖非十分準確,要與實數相差不遠。徵集此項報告,可事先備好黏有郵票之空白(或印就的)信封附於發出之信內,請報告者直接寄回,或由各地代理人徵集彙轉。

丙、被查人塡報法　此種調查方法與乙法皆用於範圍極大之調查,而其所得之結果則較之乙法爲可靠。惟採用此法最重要之條件須得被查人之合作:第一,被查人對於調查之問題須有興趣;第二,被查人須有相當知識以塡答各項問題。故問題必須十分簡單,事先尤須引起其興味,否則,非有政府之法令爲後盾,難得良善之結果。此法亦甚經濟;有時雖有一部分被查人置之不答,或答而不全,但可照統計者希望之數加倍發出調查表以杜此弊。迨收回時,擇其可用者加以整理,以爲全體之代表。現代工資統計,失

業統計,地方歲出統計,氣候統計等,大率皆用此法。

丁,僱員調查法　此法係由調查機關僱用專門職員分任調查事務,就確度言,爲大規模調查最好之方法。惟此法需費浩繁,政府機關行之則可,私人行之,財力恐有未逮。採用此法,調查表之內容,不妨較爲詳細;問題之範圍亦不妨略爲擴大。至於調查員之選擇,以富有理解力而誠懇耐勞者爲佳。如於調查之先,能與調查員以相當之訓練,於調查工作之進行,必能裨益不少。

上述四法中,丙丁兩法於調查前必須製好表格;甲乙兩法雖不必定用表格,然普通仍以用表格者居多。茲將編製調查表格及選擇問語時應注意之事項,附述於後:

編製調查表時應注意之事項:

a. 調查表之式樣以矩形爲準,大小須合度,紙張質料不宜過薄。

b. 標題須簡括,俾一覽而知調查之主旨。

c. 表中項目須依類排列,必要時應分爲數部。

d.表中線格、字體,可分粗細、大小,以示問語之爲主要或次要。

e.每一問語之後,須留有相當地位以備填答。

f.表中須註明調查之機關,最好能附一簡單之說明解釋調查之意義,俾被查者不致因誤會而僞報。

選擇問語時應注意之事項:

a.問語不宜過多。

b.問語須簡明易曉;最好使被查者能以「是」「否」或數字作答。

c.問語不可引起被查者厭惡之心。

d.問語不宜過於尋根究底,涉及隱私。

e.問語最好能有前後可以互相證實之處.

f.問語須直接,懇切,所問者應卽爲所欲搜集之材料。

三、估計法　此法所以濟一二兩法之窮,非至萬不得已時,不宜採用,以用此法所得之材料,常不甚精確也。例如欲知一國蘊藏於地下之鑛產,既無法計算其實數,亦無從得原有之紀載,乃不得不用估計法以救濟之。惟應用此法,須

注意下列二端:

甲、當有根據 估計者非憑空捏造之謂,必有相當之根據,然後估計始有價值。所可根據者有三:

a. 根據以前之統計或估計 如各國之人口每五年或十年調查一次,若某年非辦理人口統計之年,吾人欲知當時之人口概數,即可根據已往之統計以估計之。又如估計山西省煤之蘊藏量,多謂盡量開採,可供全世界二千年之用,此種估計雖未必精確,然係依據前人雷叔文(F. Van Rithofen) 之估計也。

b. 根據直接取得之材料 此法係由估計者直接搜集材料以為根據。如人壽保險公司之生命表(Mortality Table),即係根據已往數十年之死亡率及死亡與年齡,體魄等之關係而估計者也。

c. 根據間接相關之材料 如根據每人每年平均食鹽或糖之消費量與全國之總消費量,估計全國人口之約數,即為利用間接材料之一例。

乙.審查確度　估計之結果是否正確,此用估計法者不可不注意者也.審查估計之結果可依下述四項標準:

a. 如估計時所用之方法適當,並無臆測之意見參入,則結果近於正確.

b. 估計之事實苟非易變,則結果易致正確

c. 估計之結果如有多數人相同者亦近於正確.

d. 多數人估計之結果,雖微有差異,若此種差異可以互相補償者,則此估計亦近於正確.

總之,估計法爲搜集材料方法中之下乘,苟有他法,不宜輕易採用;既不得已而採用矣,應審愼周詳,力求根據之充分,免貽人以臆造之嫌.

第四章　整理材料

材料既經搜集,進一步之工作即爲整理材料。惟統計之法則,根據於大量觀察,搜集之材料必多,其中未必盡屬可用之材料,故於整理之先,須用抽查法(Sample Testing)或校讀法(Check Reading)審核一過,以別其正誤。若材料中發現有可疑之處或一部分材料與大多數有矛盾之處,應趕即設法覆查,或竟棄而不用。但無論如何,於此等處應加顯明之標記或審查者之意見,切不可塗擦删改,以損其眞。審核之工作既畢,乃可着手整理。整理統計材料之方法有三:曰謄錄法;曰記號法;曰活頁卡片法。吾人究應採取何法,當視材料之性質及繁簡以爲斷,蓋三法皆所以分別材料之異同,彙類歸納,爲製表之預備者也。玆分述三法於下:

第一節　謄錄法

所謂謄錄法者,係將所搜集之材料彙於一處,保存其實在情形,備作詳細硏究之用。玆以調查某地同級學生各科之成績爲例,設吾人所搜集者爲各個學生之成績單,用謄錄法以整理之,則將全體學生之各科成績彙於一紙,其式如下:

表

某地若干名同級學生各科成績一覽

姓名	國文	英文	數學	歷史

*表式見周調陽著教育統計學第13頁。

此法之長處在於詳細,然若學生之數目極多,未免手續過繁,殊不經濟。

第二節　記號法

記號法係將性質相同或相近之事實彙於一欄,用記號計其數目,備作研究分析之用。普通應用之記號有下列數種:

一、一 丅 下 正 正

二、丨 ∟ ⊔ 口 ⊠

三、╱ ⫽ ⫻ ※ ※

此三種記號均係由一至五,所以便於計數也。玆仍以調查學生各科成績爲例,以示記號法之應用如下:

表　二

某校學生各科成績分配

成績	國文	英文	數學
30—35	一	/	|
35—40	丅	//	L
40—45	𤴓	////	口
45—50	正 一	※ /	⧄ |
50—55	正 下	※ ///	⧄ 凵
55—60	正 正	※ ※	⧄ ⧄
60—65	正 正 正	※ ※ ※	⧄ ⧄ ⧄
65—70	正 正 正	※ ※ ※	⧄ ⧄ ⧄
70—75	正 正 𤴓	※ ※ ////	⧄ ⧄ 口
75—80	正 正 丅	※ ※ //	⧄ ⧄ L
80—85	正 𤴓	※ ////	⧄ 口
85—90	正	※	⧄
90—95	丅	//	L
95-100	一	/	|

應用此法整理材料之優點有三：一、因用記號，可免謄錄之勞。二、無論材料如何繁多，整理之手續均甚簡便。三、經過此法整理後之材料，其分配狀況可一望而知其大概。惟此法亦非無缺點：一、應用此法，則材料之個別情形無從顯示。二、整理時記號如有錯誤，不易發覺。此外應用此法時須注意者，尚有分組問題，如上例學生之成績由三十分至一百分共分十四組。分組既定，然後始可按學生各科之成績，一一記入相當之組內。

第三節　活頁卡片法

上述兩法皆甚呆板,材料既經整理之後,不便移動或有所增刪。應用活頁卡片法則無此弊。因每一材料,可記於一獨立之卡片上,卡片之數量可伸縮自如,吾人可隨時删去已廢材料之卡片,加入新材料。且應用此法,分類尤稱靈便。茲仍以調查學生成績之材料爲例。既將各個學生各科之成績錄入獨立之卡片,則吾人可隨意按性別,省籍或年齡之大小,比較學生成績之優劣。論其手續不過略一轉移卡片之勞,至易易也。卡片之質料宜稍厚,大都用白色,然如材料之種類太繁,亦可用數種顏色以分別之。至卡片之大小,以能容納登記之材料爲度,縱橫約爲三與五之比。

材料既經登記於片上,當按類依次排列之,置於卡片櫥中妥爲保管,俾應用時易於檢查。活頁卡片法爲整理材料最佳之方法,近世統計之進步,得力於此法不少。若統計材料極多,則分類時並可利用機器,分類尤稱簡便。整理材料之方法,略具於斯,以下請論分析材料,若表列,若圖示,若各種計算法,皆屬分析材料之範圍也。

第五章 表列

製表一事,驟視之似甚簡易,但若實地編製一較複雜之表其手續決不如想像之簡單,初學者一經嘗試,必覺困難叢生。製表爲分析材料之第一步,果能編製得法,分析工作可謂已成其大半,學者切勿忽略視之。

第一節 表之功用

表之功用約有六端,茲分述於後:

一、便於比較 凡有關係之事實,在表中可列於相近之欄,其異同之點自可一目瞭然,遠較以文字敍述爲便利。

二、易於記憶 同類之事實,彙於一處,可引起閱者之聯念作用,易於記憶。

三、便於總計 材料之總數,原不必在表中計算,但表既製成,排列整齊,甚便於總計也。

四、減省重複之說明 繁賾之事實製成統計表後,則綱舉目張,可以省去許多重複之說明,此近人文字中所以常附有統計表也。

五、可以發現規律 散漫之材料,製成統計表後,因排列有序,可發現一定之規律,近世社會科

學之定律或假設從統計表中發現或證實者,固不勝枚舉,即研究自然科學者,亦多將其試驗之結果彙於一表,以檢尋其中之規律。

六、可發現各組事實間之關係　列於一表之各組材料,其間每有一種關係,此種關係在散漫之材料中,或不易發現,但若列於一表,因排列得法,自易顯明。

第二節　表之種類

統計表之種類甚多,分類殊不一致,茲擇其中最普通之三種分類說明如下:

一、按所列事實爲原始的或非原始的,可分爲原始表 (Original Table) 及次級表(Secondary Table)二種:

a. 原始表　原始表係將所有搜得之材料,逐一記載,彙於一表,保存其實在情形,備作詳細研究之用。如第四章第一節之例,用謄錄法製成之表(參看表一),即原始表也。

b. 次級表　次級表者,縮合原始表之材料而爲賅簡記載之表也。此表之材料,僅爲原始表之一部分,如下例某地同級學生各科成績比較表,係根據學生各科成績之原始表

(表一)經過計算手續而製成,即次級表是也。

表 三*

某地同級學生各科成績優劣比較

科目	算術平均數	中位數	標準差	四分差
國文				
英文				
數學				
歷史				

*表式見周調陽著教育統計學第14頁。

二、按所列事實之繁簡,可分為單項表,二項表,三項表,四項表及多項表多種:

a. 單項表 單項表者,按一種標準編製之統計表也。如表四,係按成績統計學生之人數。

表 四

某級學生國文成績

分數	人數
20—30	2
30—40	4
40—50	7
50—60	13
60—70	26
70—80	18
80—90	7
90-100	1
總計	78

b.二項表　二項表者,按兩種標準編製之統計表也。如表四中按成績分組之外,更於人數欄中分學生之性別,即成二項表。三項表四項表以及多項表可以類推,茲不贅述。表中事實之繁簡,原無一定之限制,惟爲便於比較及記憶起見,非極有關係或不能分列之事實,不宜列入一表。金氏嘗謂:「每表須成一單位,」即此旨也。

三、按統計數列之種類,可分爲時間數列表,地理數列表,次數數列表或次數分配表三種:

a.時間數列表　時間數列表者,係按時間之先後,將同一事實在各時期之數目,順序排列之表也。如下例中國對外貿易狀況表,按時期之先後,排列進口,出口及入超之價值,即一時間數列表。

表　五

民國十年至二十一年我國對外貿易狀況

(單位:關銀一萬兩)

年份	進口價值	出口價值	入超價值
民國十年	90,612	60,126	30,486
民國十一年	94,505	65,489	29,016
民國十二年	92,340	75,292	17,048
民國十三年	101,821	77,178	24,643
民國十四年	94,786	77,635	17,151
民國十五年	112,422	86,429	25,993
民國十六年	101,293	91,862	9,431
民國十七年	119,597	99,135	20,461
民國十八年	126,578	101,569	25,009
民國十九年	130,976	89,484	41,492
民國廿年	143,349	90,948	52,401
民國廿一年	104,925	49,264	55,661

b.地理數列表　此表係以地域爲主體,將同一事實在各地方之數目,依次排列,以爲比較之用。下列一九二九年中、俄、美、德、日、英、法七國人口與國富比較表,卽其一例。

表　六*

一九二九年中俄美德日英法七國人口與國富比較

國別	人口概數(單位: 100,000)	國富總額(單位: 100,000日金圓)
中國	4,850	328,890
俄國	1,600	1,041,020
美國	1,160	7,623,560
德國	680	716,850
日本	600	1,023,430
英國	470	2,363,200
法國	410	1,035,300

*表中數字除中國人口概數係內政部發表者外,餘均爲日本內閣統計局所發表。

c.次數數列表　此表係將同一時期之事實,無地域關係者,依其分佈之情形或次數之多寡,順序排列而成。次數數列,又可細分爲二:一爲屬於質的統計(Statistics of Attributes)之次數數列;一爲屬於量的統計(Statistics of Variables)之次數數列。如表七按學校之種類,統計學生之人數,即爲屬於質的統計之次數數列表;表八按學生身體之重量而分類,則爲屬於量的統計之次數數列表。

表七

某地中等以上學校男女生人數

校別	學生人數		
	男	女	總計
大學校	6,299	124	6,423
專門學校	5,515	380	5,895
師範學校	340	220	560
中學校	5,859	1,027	6,886
職業學校	694	603	1,297
總計	18,707	2,354	21,061

表八

某校93名十八歲男生體重分配

體重磅數	人數
90—94.99	1
95—99.99	2
100—104.99	4
105—109.99	9
110—114.99	14
115—119.99	15
120—124.99	21
125—129.99	10
130—134.99	9
135—139.99	5
140—144.99	3
總計	93

上述三種數列中,時間數列與次數數列之性質最不相同,地理數列與次數數列則頗相近。就研究之方法

言,時間數列之分析較爲複雜,次數數列與地理數列之分析則較爲簡單,且大致相同,無須分論。就製表之工作言則以編製次數分配表之手續較繁,玆將其步驟附述於後:

一、求全距(Range) 全距者,卽材料中最大一項與最小一項間之距離,從最大一項之數值中減去最小一項之數值卽得。如下列二百一十名工人每月所得之工資,乃一次數數列。其中工資最高者爲$32.00,最低者爲$22.55,由$32.00減去$22.55,得$9.45,卽全距也。

二百一十名工人每月所得之工資

$26.25	$28.60	$23.10	$24.60	$27.55	$27.00	$27.75
26.70	27.80	27.55	25.75	29.50	28.80	24.75
28.20	25.35	25.25	26.75	27.80	26.30	27.60
27.70	29.30	24.00	28.60	25.70	28.25	25.75
24.30	25.55	26.75	29.10	26.10	26.30	27.00
27.60	27.90	24.60	24.50	28.55	23.40	27.80
26.15	31.00	26.30	26.80	27.15	27.75	27.25
27.95	28.55	30.75	28.85	27.55	26.60	28.35
27.30	30.00	28.15	30.55	28.15	28.35	28.00
26.75	26.25	23.00	24.60	30.60	29.60	27.55

30.25	27.00	27.90	24.25	25.25	24.50	25.55
29.55	27.25	25.80	25.85	28.30	26.15	28.90
25.75	27.55	28.60	29.30	27.30	28.55	25.25
26.60	28.60	27.10	26.80	23.50	26.55	26.30
26.30	25.30	27.50	27.30	27.00	27.95	25.15
26.55	27.80	24.55	32.00	27.45	27.80	27.30
30.70	26.30	28.60	29.20	30.00	28.25	28.10
26.90	25.15	25.75	28.30	24.55	25.70	27.90
26.55	28.65	26.75	27.40	28.30	24.75	26.00
28.10	23.00	28.10	26.40	26.55	28.55	27.15
24.10	27.15	29.75	27.40	27.00	27.55	29.80
28.15	25.80	27.20	24.10	28.30	26.30	29.30
29.00	27.30	26.35	28.50	23.50	25.00	27.55
26.10	26.00	27.80	24.25	29.90	25.15	23.60
24.55	24.15	26.30	27.45	26.55	27.90	25.80
26.85	25.75	29.00	26.50	25.25	30.15	26.30
28.60	27.80	27.10	25.75	27.30	26.80	26.25
28.70	25.30	26.60	26.30	24.10	27.55	30.10
24.35	27.60	24.10	27.90	26.30	28.55	22.55
27.30	27.55	25.35	28.10	26.80	25.60	24.10

二、決定組距(Class-interval)

全距求出之後當進而

分組。組距者，卽每組間之距離也。各組之組距須一律，不可過大，亦不可過小，須視全距之情形而定。如工人最高最低工資間之全距爲$9.45，設以$1.00爲組距，可將二百一十個工人分爲十一組(參看表九)；以$.50爲組距，則可分爲二十組(參看表十)。據勒格教授(H.O.Rugg)之意見，一次數分配表中之組數，最好介於十與二十之間，蓋組數少於十，足以減少次數分配之精確，多於二十，則計算時未免太煩，且次數分配集中之趨勢，亦不易顯明。

表九

210名工人每月所得工資分配

工資 (組距=$1.00)	次數
$22.00—$22.99	1
23.00— 23.99	7
24.00— 24.99	21
25.00— 25.99	27
26.00— 26.99	42
27.00— 27.99	54
28.00— 28.99	34
29.00— 29.99	13
30.00— 30.99	9
31.00— 31.99	1
32.00— 32.99	1
總數	210

表十

210名工人每月所得工資分配

工資 (組距=$0.50)	次數
$22.50—$22.99	1
23.00— 23.49	4
23.50— 23.99	3
24.00— 24.49	11
24.50— 24.99	10
25.00— 25.49	12
25.50— 25.99	15
26.00— 26.49	22
26.50— 26.99	20
27.00— 27.49	24
27.50— 27.99	30
28.00— 28.49	17
28.50— 28.99	17
29.00— 29.49	7
29.50— 29.99	6
30.00— 30.49	5
30.50— 30.99	4
31.00— 31.49	1
31.50— 31.99	0
32.00— 32.49	1
總數	210

三、劃定組限(Class limits) 組限者,即組距之界限。組限有高限(Upper limit)低限(Lower limit)之分,如表九第一組 \$22.00—\$22.99, 22.99 爲高限, 22.00 則爲低限,蓋 22.00 小於 22.99 也。組限須詳細書明如 \$22.99, 俾前後兩組之次數可以分清,而無混淆之弊。然此係指連續數列(Continuous series)而言,設爲間斷數列或非連續數列(Discrete series),各項變數雖不可分爲小數,亦可用整數劃清界限如11—20,21—30。組限之起訖,最好爲五或十之倍數,如表四及表八。至於各組之中點,能爲整數最佳,因吾人應用次數分配表時有一假定,即任何一組中之變數,其平均大小,可以該組之中點爲代表,中點如爲整數,則以後計算時,自較便利。

四、排列次數 組距及組限既經決定,於是將各變數按其大小列入相當之組中,即成一次數分配表。惟排列時可先用記號,然後將記號改爲數字(參看第四章第二節)。

第三節 製表之規律

製表之規律,可分四項說明之:

一、關於表之標題者:

a.表之標題當力求明顯易解,並須切合表之內容。

b.標題之位置,應在表之頂端。

c.標題中所舉各點,其次序應與表中所列之項目一致(參閱表六)。

d.表之內容有地域及時間區別時,標題中亦應註明(參閱表五)。

e.如表之內容甚長,佔數頁地位時,各頁均須註明標題,除第一頁外,餘頁並須註明「續前」二字。

二,關於表之項目者:

a.表中項目之次序須按下列各種標準排列:

1.重要之程度。

2.等級之高低(參閱表七)。

3.時間之先後(參閱表五)。

4.數量之大小。

5.筆畫之簡繁。

6.地方區域之位置。

b.表中項目最好一律由左至右橫寫。

c.大項目下可分細目,惟須低一格並用線格分開(參閱表七)。

d.對於某一項目特別注重時,可用較重之字體顯明之。

e.凡橫幅過長之表,左部所註項目不便閱覽時,可於右端重列一次。

f.爲便於檢查或引用起見,項目較多之表,可於項目上面或旁面用羅馬字標明次序。

三、關於表之線格者:

a.表中各縱行間須以直線劃分之,細目間用單直線,大項目間則用較粗之直線。橫行間則無須。

b.表之左右兩邊,無須用邊線隔絕。

c.表之頂端須劃雙橫線或粗橫線,下端則劃單橫線足矣。

d.表中上部項目之下及下部列總數或平均數之上,均須用橫線劃分之。

e.表中排列數字處不可用橫線分開,以免障礙視線。

四、關於數字之排列者:

a.表中數字,須一律用阿剌伯字,以其整齊且可節省篇幅。

b.表中各縱行數字之位數須上下相對,以便

加減比較。如有小數,則小數點尤須列在一垂直線上,以免計算錯誤。

c. 數字多至四位以上,須用分段點,普通每三位分作一段。

d. 無數字之空格須用短點線或短橫線補充之。

第六章 圖示

第一節 圖表功用之比較

吾人通常表現事實之方法有三:一、文字,二、表格,三、圖示。此三種方法各有短長,其中文字一項不在討論之列。至表格之功用已於前章第一節說明,本節所討論者,乃圖表兩者功用之比較,茲分述於下:

一、表之功用在於排列及分析材料;圖之功用則在表明此項排列及分析之結果。故製表之工作在前,繪圖之工作在後。

二、表之優點在於詳細,可供專家之參考,而其弱點則對於一般讀者每嫌乾燥無味。圖之優點在於簡單明顯,至於間有不甚詳盡之處,即其弱點。故能圖表並用,俾讀者各擇其宜,最善。

三、表中所列之事實多用數字表現,每陷於抽象;不若圖之具體化,易使閱者領悟。

第二節 圖之種類

統計圖之種類,就普通常用者而言,約可分為七種:(一)直條圖(Bar Diagram),(二)平面圖(Surface Diagram),(三)立體圖(Volume Diagram),(四)形像圖(Figurative Diagram),(五)組織圖(Organization Chart),(六)統計地圖

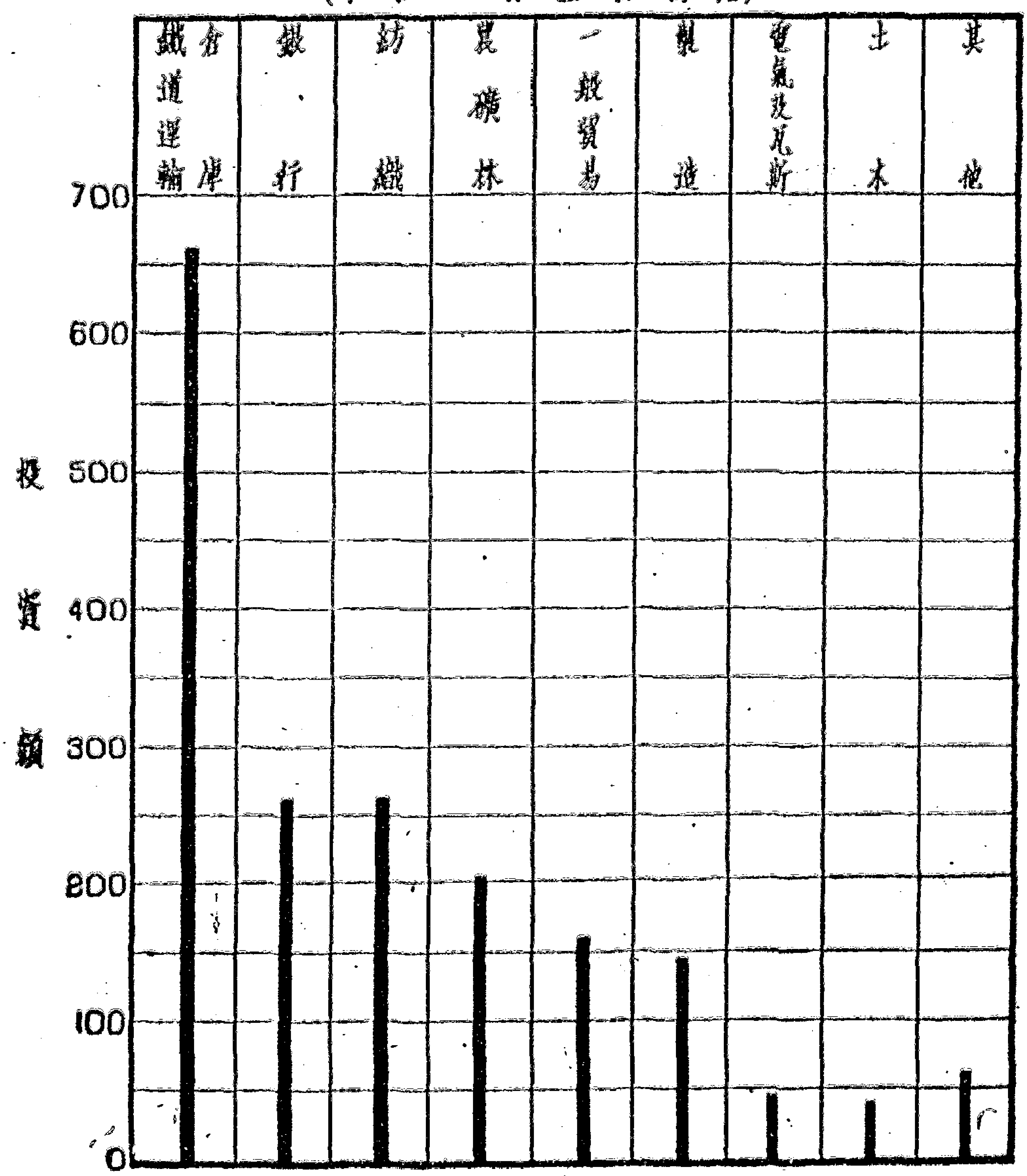
(Statistical Maps), (七)曲線圖 (Curves or Graphs)。茲依次分述如下:

一、直條圖　此圖係以直條之長短,表示事實數

圖一

日本對華投資

(單位: 日金百萬元)



材料來源:1919年太平洋會議時日本委員所發表

量之多寡,材料之無連續性者多用之。至直條之排列可由左至右,亦可由下至上。直條圖更可分爲三種:

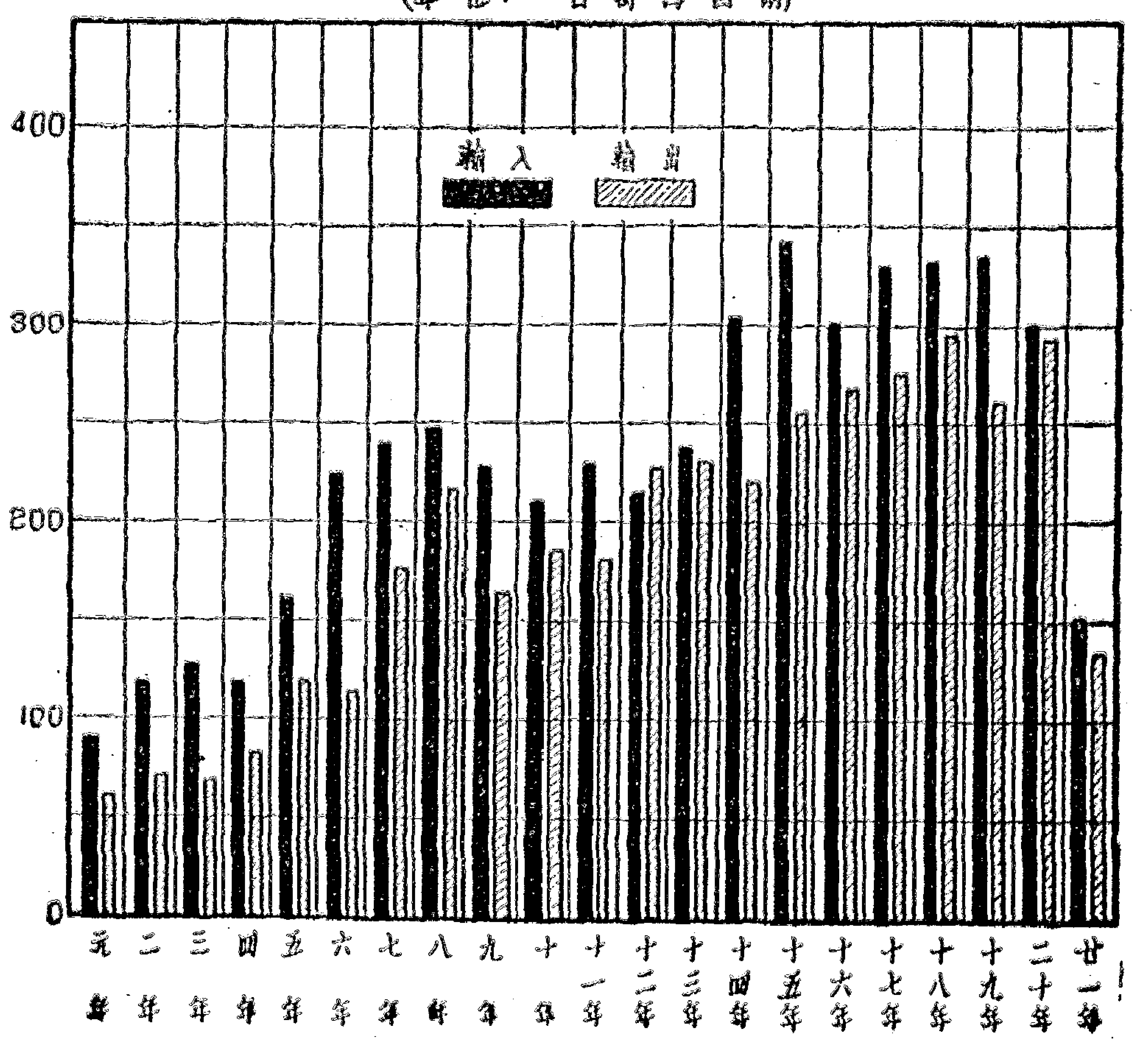
甲、單式直條圖　此圖係以一直條表示一單獨事實之數量,直條宜稍粗,各直條間之距離不可太狹,例如圖一。

乙、組合直條圖　組合直條圖者,以兩道或兩

圖二

民國元年至二十一年我國對日貿易

(單位：百萬海關兩)



材料來源：中國經濟學社中日貿易研究所出版之中日貿易統計(包括朝鮮台灣)

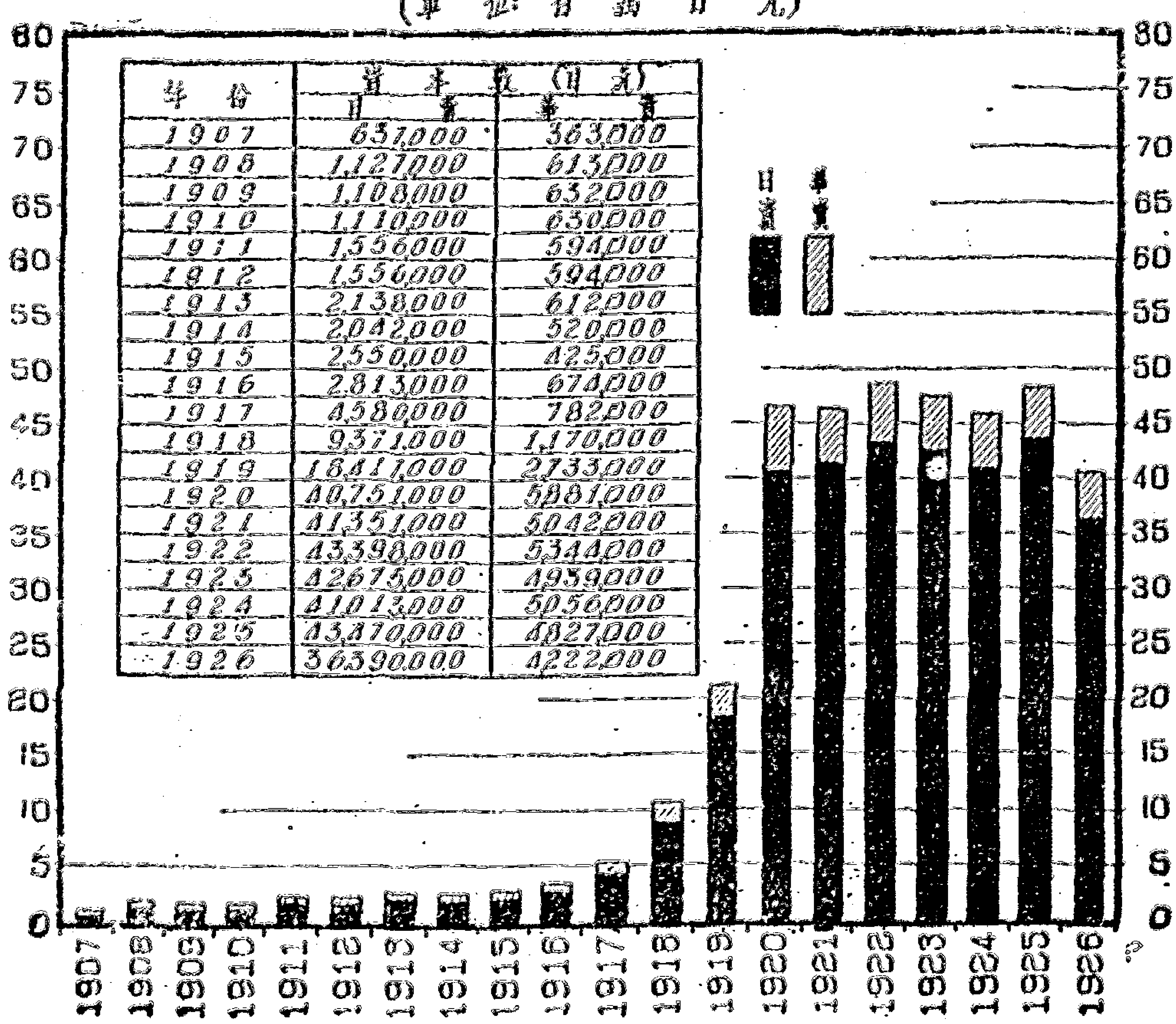
道以上之直條爲一組,代表一大項目,而以每組中各直條表示大項目中之小項目。例如圖二以兩直條代表我國對日貿易之總值,黑條表示日貨輸入,線條表示華貨輸日,即組合直條圖也。

丙,區分直條圖 區分直條圖者,以一直條代

第三圖

中日合資商行資本額

(單位:百萬日元)



年份	資本數(日元) 日資	華資
1907	631,000	363,000
1908	1,127,000	613,000
1909	1,108,000	632,000
1910	1,110,000	630,000
1911	1,556,000	594,000
1912	1,556,000	594,000
1913	2,138,000	612,000
1914	2,042,000	520,000
1915	2,550,000	425,000
1916	2,813,000	674,000
1917	4,580,000	782,000
1918	9,371,000	1,170,000
1919	18,411,000	2,733,000
1920	40,751,000	5,881,000
1921	41,351,000	5,042,000
1922	43,398,000	5,344,000
1923	42,675,000	4,959,000
1924	41,013,000	5,056,000
1925	43,870,000	4,827,000
1926	36,390,000	4,222,000

表一件事實之全體,再按其中各部數値之此例,分直條爲若干段,每段表示一部。例如圖三以各直條代表各年中日合資商行資本額,每條中更分二段表示日資及華資佔總額之成分。

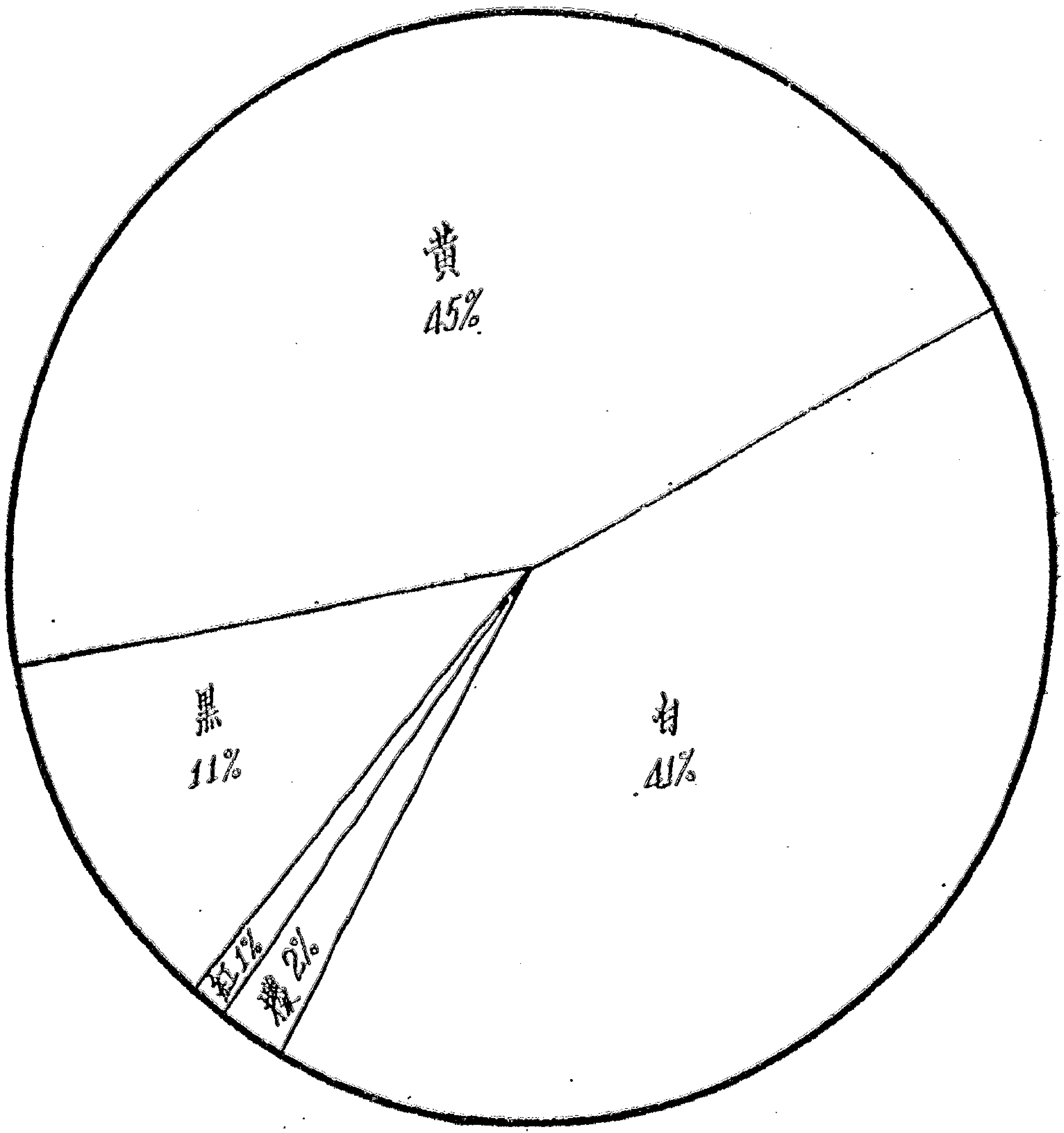
二,平面圖 此種圖形係以面積表示一事實數量之大小,如圓形圖,方形圖,均爲常用之平面圖。此外尙有三角形圖,以及多邊形圖等等,茲從略。

甲,圓形圖 圓形圖更可分爲三種:

a. 單圓圖 此圖係以圓形之面積代表一事實之全體,再按一事實各部之大小分全面積爲若干扇形以代表各部。例如圖四,圓形之全面積係代表全世界之人口總數,其中分爲五個扇形,代表黃,白,黑,棕,紅五色人種所佔之百分數。繪製此圖之手續甚爲簡單,其扇形之劃分法,係按圓周360度計算,每百分之一等於3.6度。惟應用此圖,項目不宜過多;過多則扇形狹小,不便塡載文字或數字矣。圖四表示棕紅二色人種之扇形,卽嫌太狹。

圖四

黃,白,黑,棕,紅五色人種百分比較

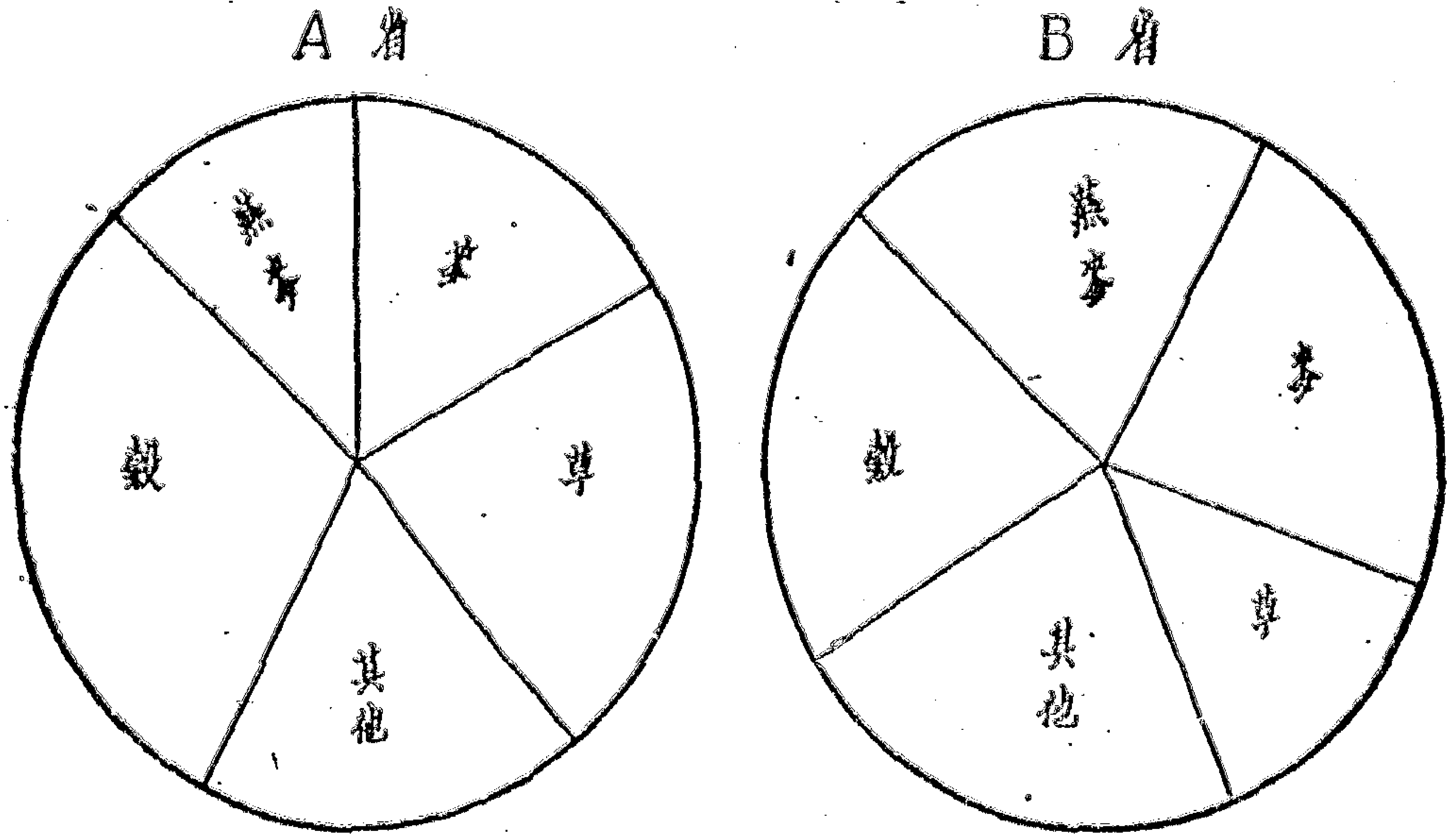


b. 多圓圖　多圓圖係用兩個以上之圓形代表兩件以上事實之數量,以便互相比較。例如圖五以兩個圓形代表 A, B 兩省耕地之面積,每一圓形中更可分爲數扇

形,比較耕地之全面積中種麥者若干,種穀者若干。

圖 五*

AB兩省各種農作物所佔耕地面積比較



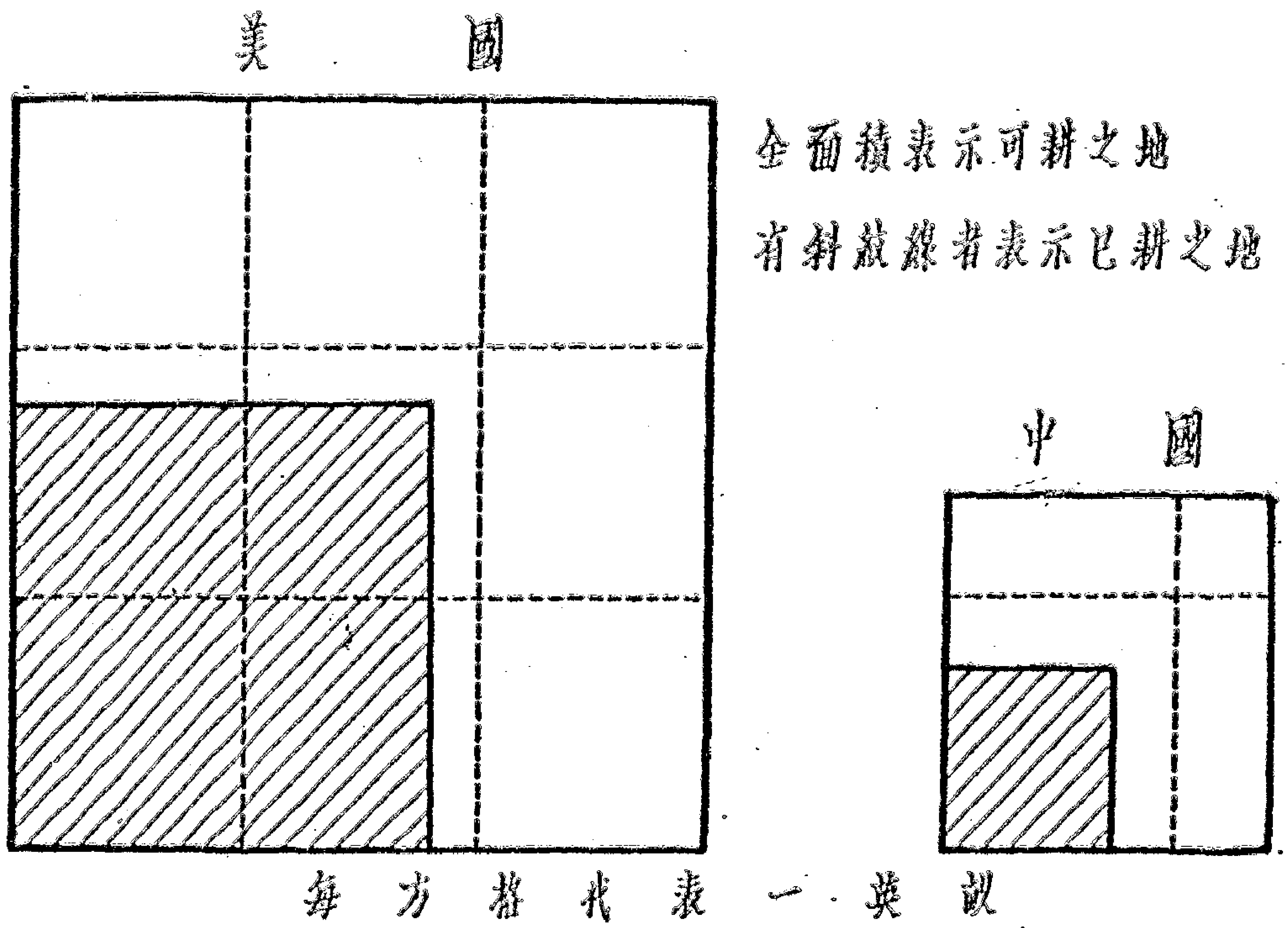
* 見 King, W. I.: Elements of Statistical Method, P. 96, Fig. 4.

c. 疊圓圖　此圖係從同一中心點用不同的半徑繪成兩個以上之圓周,代表兩種以上之事實。或以數個圓形相疊,依圓形面積之大小,比較各種事實之多寡,例從略。

乙、方形圖　方形圖係以正方形或長方形表示一事實之數量,依面積之大小,表明數量之多寡,例如圖六,卽正方圖之一例。

圖六★

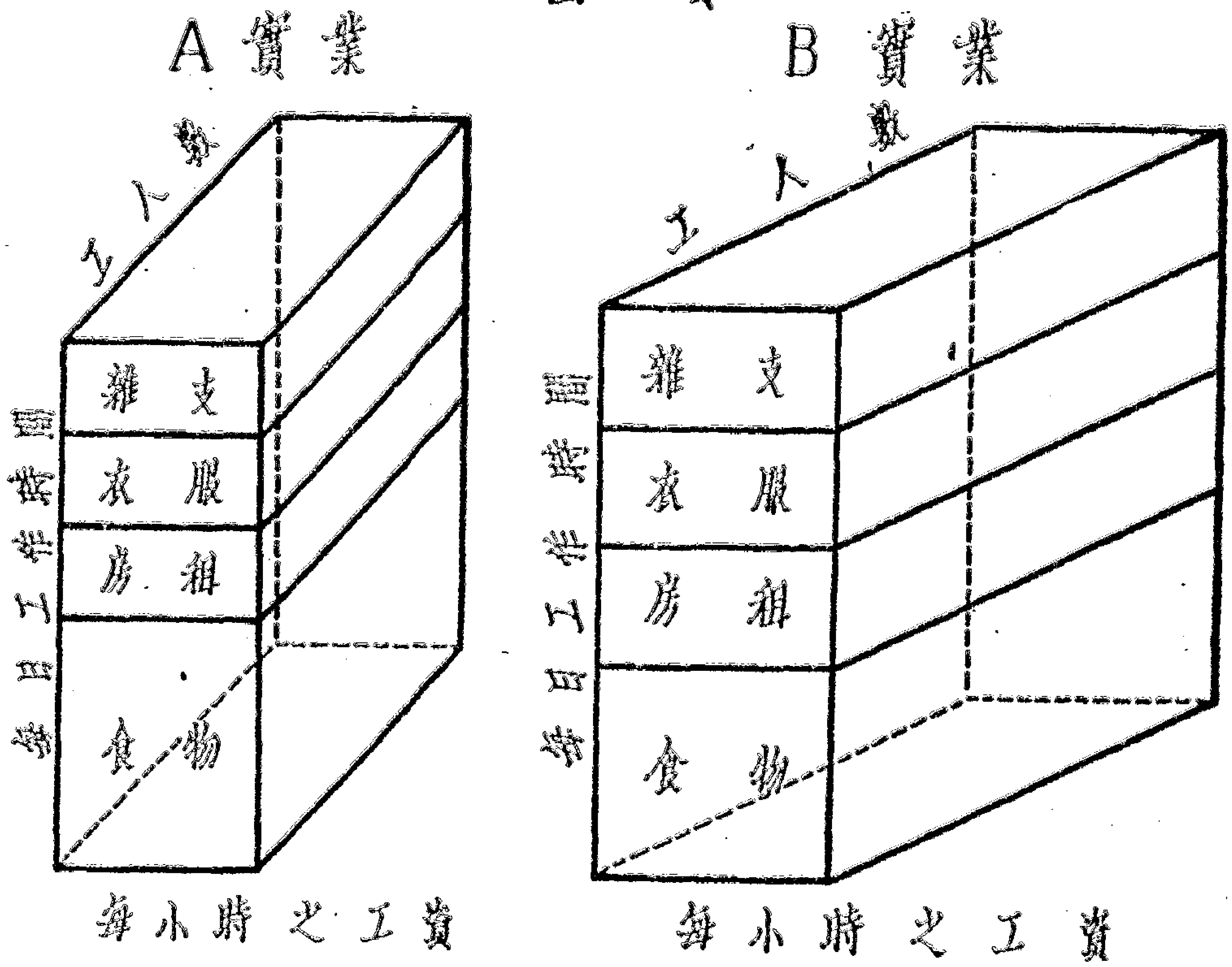
中美每人墾地額比較



★見中國經濟研究所出版經濟統計月誌第一卷第二期

三、立體圖　此圖係以一立體形代表一事實之全體，以長，寬，高或正面，側面及兩端之面積，表示各部之大小。如立體形，圓柱形，稜椎體形，球形等，均屬此類。圖七所示，即爲立體圖；可以作A、B兩種工業工人人數，每小時所得工資，每日工作時間，及工人支出分配等多方面之比較。惟應用此類圖形，比較之項目，不宜過多，過多則不便比較，效用將低減也。

圖七★

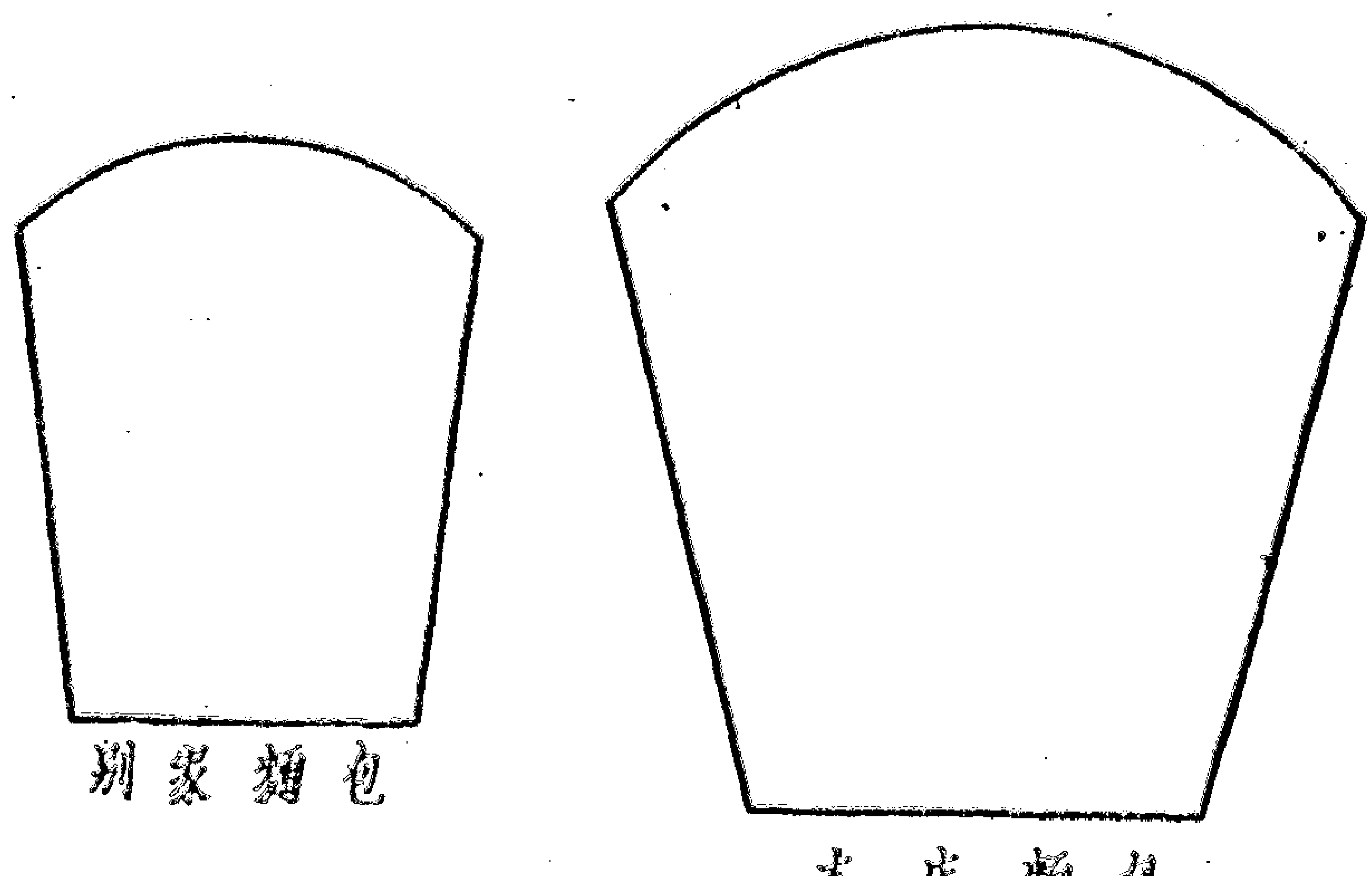


★ 見 *King, W. I.: Elements of Statistical Method, P. 95, Fig.*

四、形像圖　此圖係將所欲表示之事物，按照其原來狀態繪成圖形，以圖形之大小，高低，長短或多寡表示其數量。圖八為一麵包商說明其貨品較諸別家分量為重之簡單形像圖，用於廣告，收效甚宏。此外欲比較兩國商船噸數之多寡，可繪兩隻大小不同之商船表示之。如欲比較兩國飛機之多寡，亦可以形像圖表示之。總之，凡比較具體事物之數量，均可利用此類圖形，其優點在能引起閱者之興味，惟用此圖

作精確之比較，尚非藉助於數字不可；蓋僅憑目力，不易比較；且易得錯誤之觀念也。

圖 八★

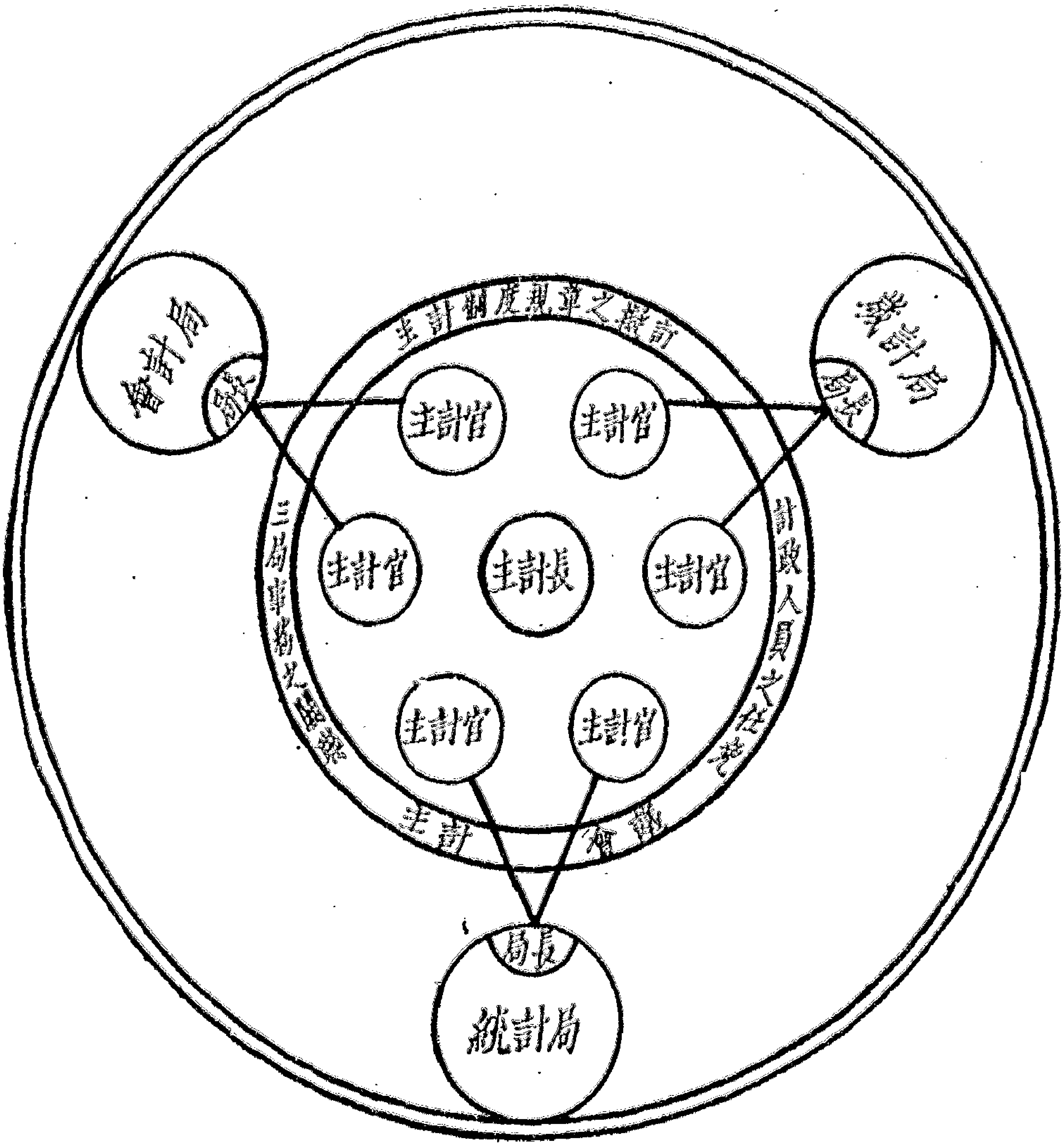


★見 King, W. I.: Elements of Statistical Method, P. 93, Fig. 1.

五、組織圖　此圖係表示各種機關之組織系統，以顯明其間隸屬或統轄之關係者；如圖九即其一例。

九圖

國民政府主計處組織圖



六、統計地圖　統計地圖者，表示統計事項地域上分配關係之圖也。統計地圖之繪製，有以各種顏色表示各地域某種事實之數量者，曰顏色圖(Colored maps)；有以横線之疏密或陰影

之深淺表示各地域某種事實之數量者,曰橫線圖或陰影圖(Cross-hatched or Shaded maps);有以圓點之大小,多寡或濃淡表示各地域某種事實之數量者,曰點圖(Dotted maps);有以細針插於圖上,按針之多寡表示各地域某種事實之數量及其移動情形者,曰針圖(Pin maps),例從略。

七、曲線圖　曲線圖係以曲線表示統計事項增減變遷之趨勢,閱此圖較之直條圖,或平面圖,立體圖易得連續之觀念。若吾人所表示者爲時間數列,則尤非用曲線圖不可。曲線圖之種類甚多,按照其縱橫線格所用之尺度(Scale)可分爲等差曲線圖(Arithmetic Curves or Graphs)及等比曲線圖或對數曲線圖(Geometrical Curves or Logarithmic Curves)二種,分述如下:

甲、等差曲線圖　等差曲線圖之縱橫線格以相等之距離,表示事實之相等數量。此種圖形,更可分爲二種:

a. 按曲線所表示之材料,可分爲:

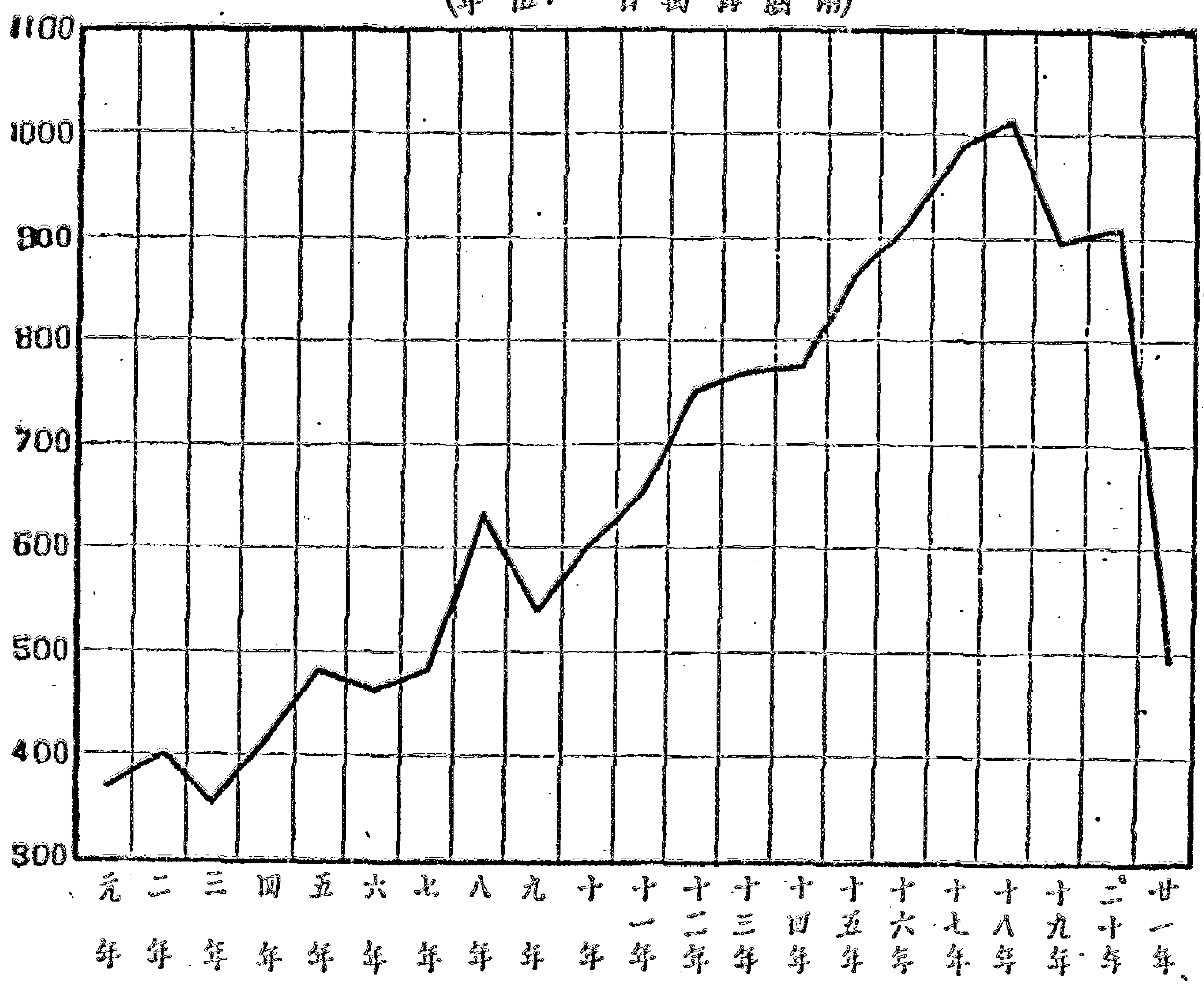
1. 時間曲線圖(Historical Curve)　時間曲線圖所表示者爲時間數列,此圖之繪

製,通常以橫格表示時間,縱格表示數量,如圖十。

圖十

民國元年至二十一年我國輸出價值

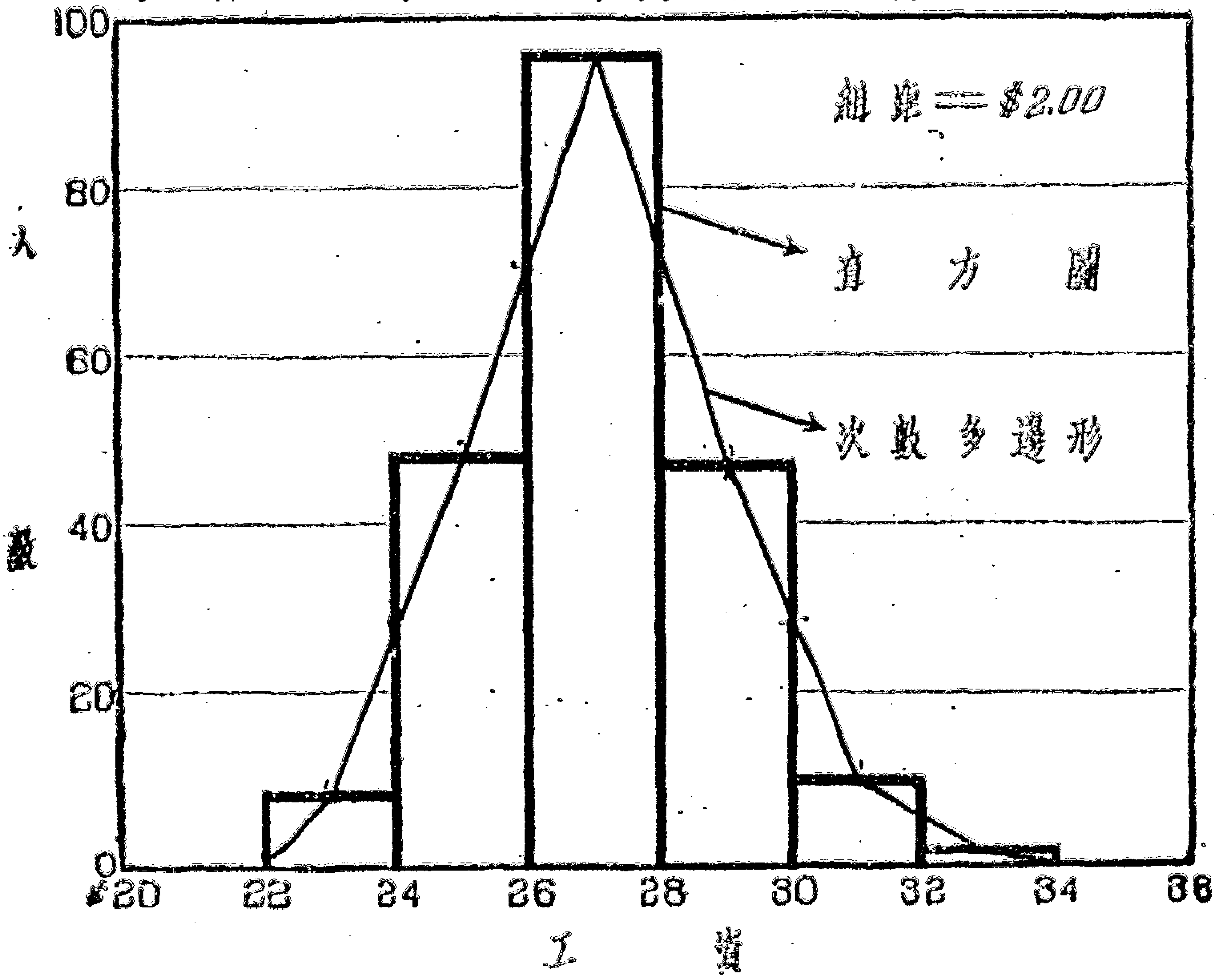
(單位: 百萬海關兩)



2. 次數曲線圖 (Frequency Curve) 次數曲線亦名次數多邊形 (Frequency polygon), 所表示者爲次數數列。此圖之繪製通常以橫格表示自變數(Independent variables), 縱格表示因變數 (Dependent variables), 如圖十一。

圖十一

二百一十名工人每月所得工資分配



b. 按曲線之形狀可分爲:

1. 次數多邊圖(參看圖十一).

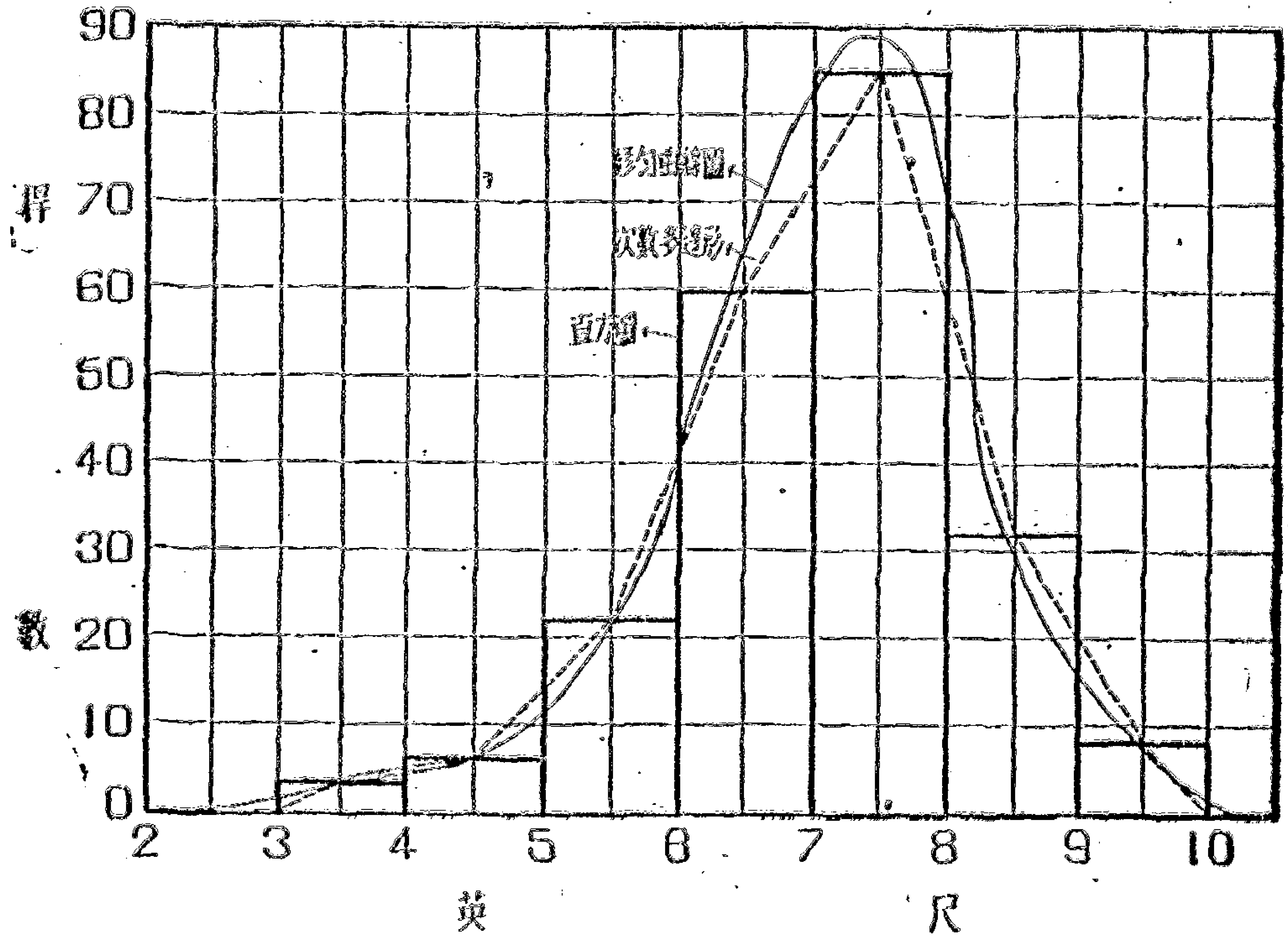
2. 直方圖(Rectangular Histogram) 次數多邊形或次數曲線,係以一點表示一組之次數,連接各點,遂成一曲線.直方圖則以一直方形表示一組之次數,其縱軸之高低與次數成正比例,其橫軸上左右兩線則爲組限.表示次數之分配,

曲線圖與直方圖之功用相等,然若計算次數面(Frequency Surface),則以直方圖較爲精確(參看圖十一)。

3. 修勻曲線圖(Smoothed Curve) 修勻曲線圖者,乃由曲線圖(時間曲線圖或次數曲線圖)或直方圖改造而成之平滑曲線也。修勻之方法,最簡單者有二:一爲隨手法(Free-hand method),一爲移動平均法(Moving average method)。隨手法者,依多邊形之形勢,配合一平滑之曲線,其掩有之面積,約等於原來曲線圖或直方圖下所包含之面積(例如圖十二)。移動平均法,多用於修勻時間曲線,其法係將數期(假定取三期)連貫事實之數量相加,以期數(三)除之,作爲當中一期(卽第二期)之數量。然後去第一期之數量,加入第四期之數量,復以三除之,作爲第三期之數量。如此繼續爲之,用求得之平均數繪成曲線,必較原來之曲線爲平滑。五期、七期移動平均法可以類推,例如圖十三。

圖十二

217玉蜀黍桿高度分配

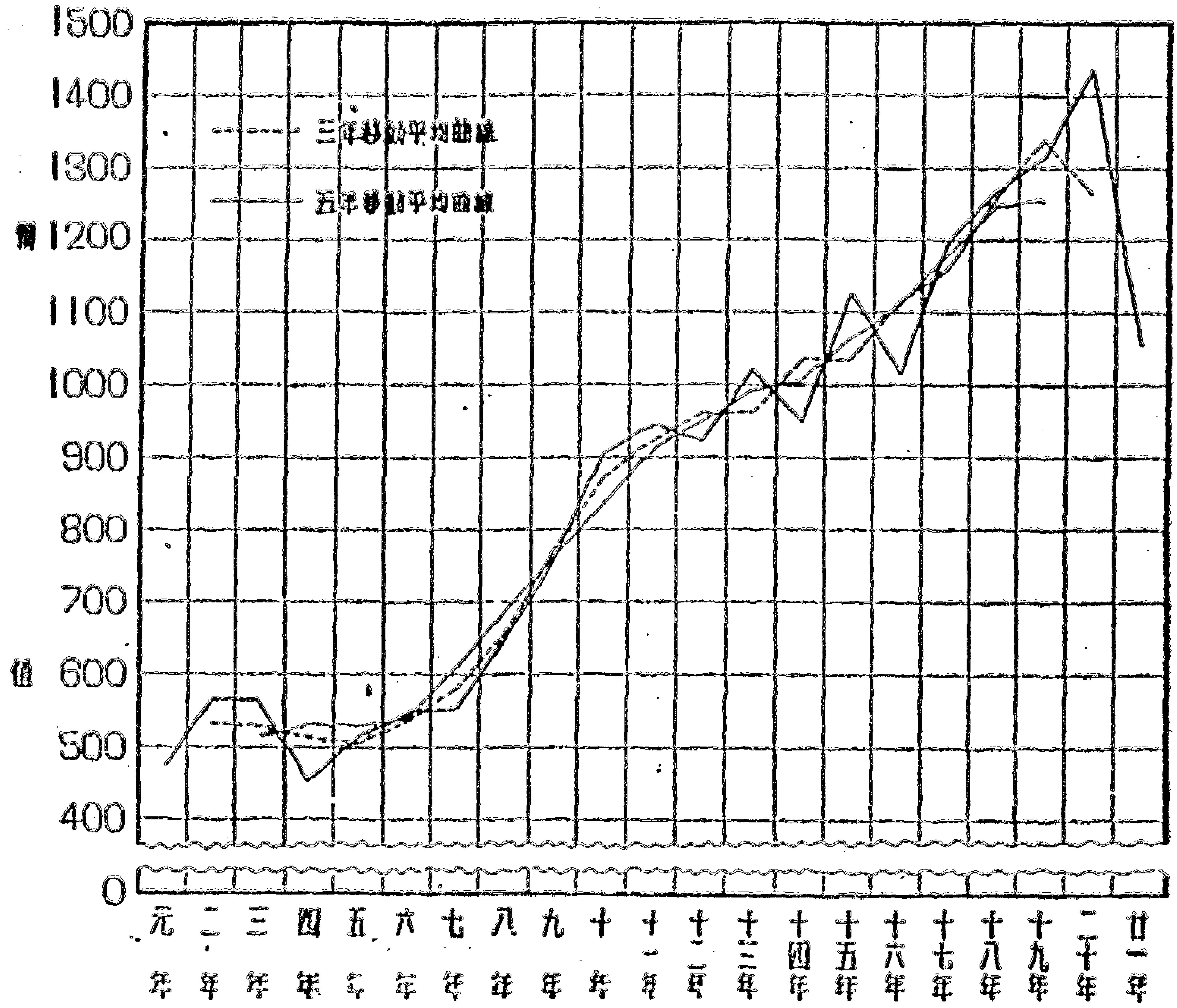


	輸入淨值 (單位百萬兩關銀)	移動平均數 三年	移動平均數 五年
民國元年	473		
二年	570	537	
三年	569	531	516
四年	454	513	532
五年	516	507	529
六年	550	540	544
七年	555	584	606
八年	647	655	684

圖十三

民國元年至廿一年我國輸入淨值

單位：百萬兩關銀



九年	762	772	763
十年	906	871	837
十一年	945	925	911
十二年	923	962	948
十三年	1,018	963	992
十四年	948	1,030	1,005
十五年	1,124	1,028	1,060

十六年	1,013	1,111	1,109
十七年	1,196	1,158	1,182
十八年	1,266	1,257	1,244
十九年	1,310	1,336	1,251
二十年	1,433	1,264	
二十一年	1,049		

細察上圖,可知五年移動平均曲線較之三年移動平均曲線爲平滑,若年數增多(如七年或九年等),則繪成之曲線愈益修勻。然年數愈多,則修勻之曲線愈短,三年移動平均曲線較原來曲線短二年,五年移動平均曲線較之三年移動平均曲線更短二年,此用移動平均法修勻曲線之缺點也。

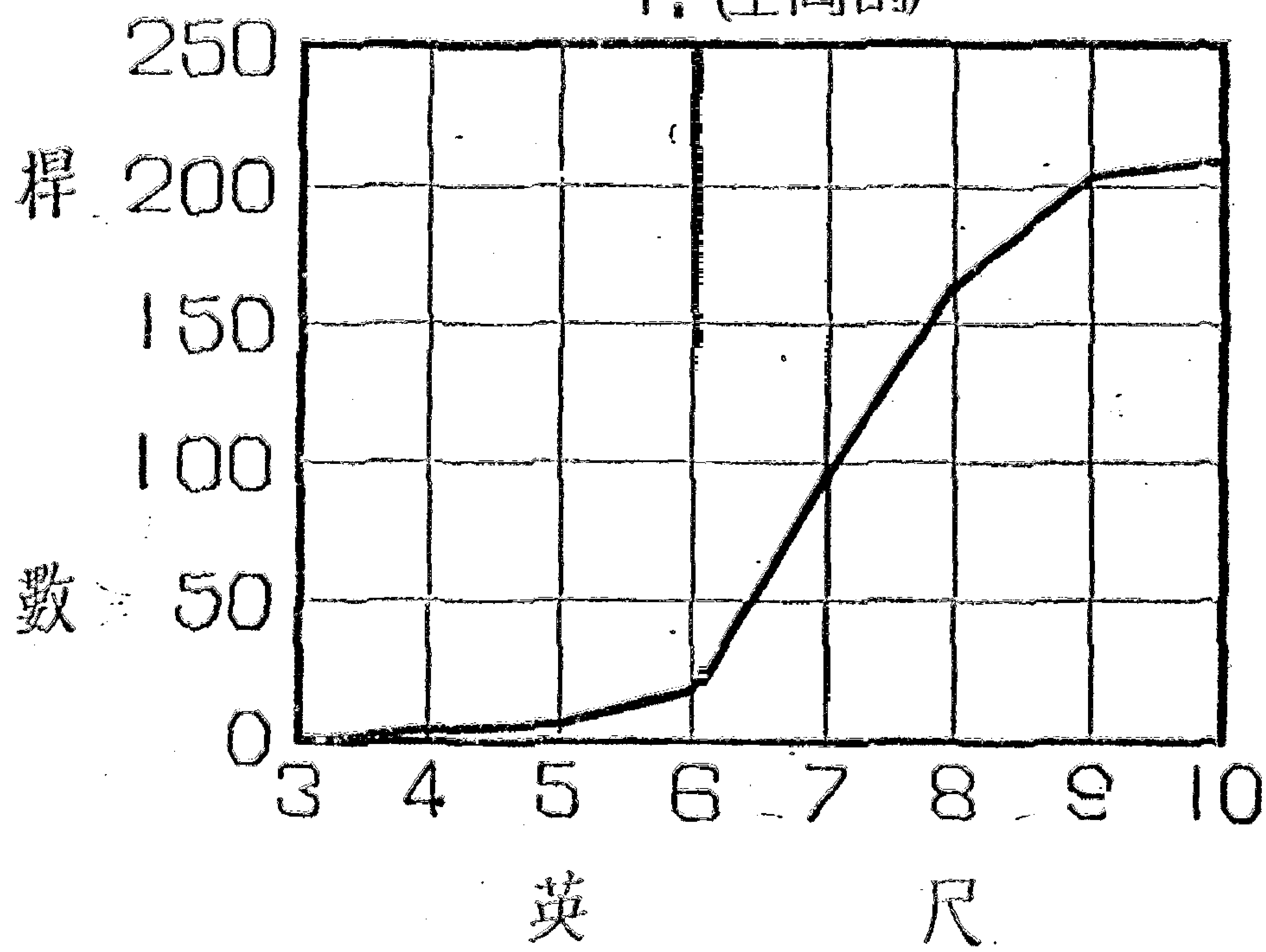
4. 累積次數曲線圖(Cumulative Frequency Graph) 繪製此圖與次數曲線圖不同之點,在後者係按次數分配表中各組原來之次數(亦稱簡單次數)繪成曲線,而累積次數曲線圖則係按累積次數繪成之曲線。所謂累積次數者,係將次數分配表中各組原來之次數依

次遞加。遞加之方法有二:一爲由數值較小之次數一端加起,如表十一第三欄,其累積次數按「以下」(less than)法讀之,如四呎高度以下者三桿,五呎以下者十桿;一爲由數值較大之次數一端加起,如表十一第四欄,其累積次數按「以上」(more than)法讀之,如三呎高度以上者二百十七桿,四呎以上者二百十四桿。累積次數之加法既有二種,故累積次數曲線之形式亦有二種:一爲上向的(less than form)累積次數曲線,圖形由左下角趨向右上角,按表十一第三欄之累積次數繪製即成此圖,例如圖十四甲;一爲下向的(more than form)累積次數曲線,圖形由左上角趨向右下角,按表十一第四欄之累積次數繪製,即成此圖,例如圖十四乙。惟當注意者,繪製上向的累積次數曲線時,各組之累積次數須繪於各組之高限上;繪製下向的累積次數曲線時,各組之累積次數須繪於各組之低限

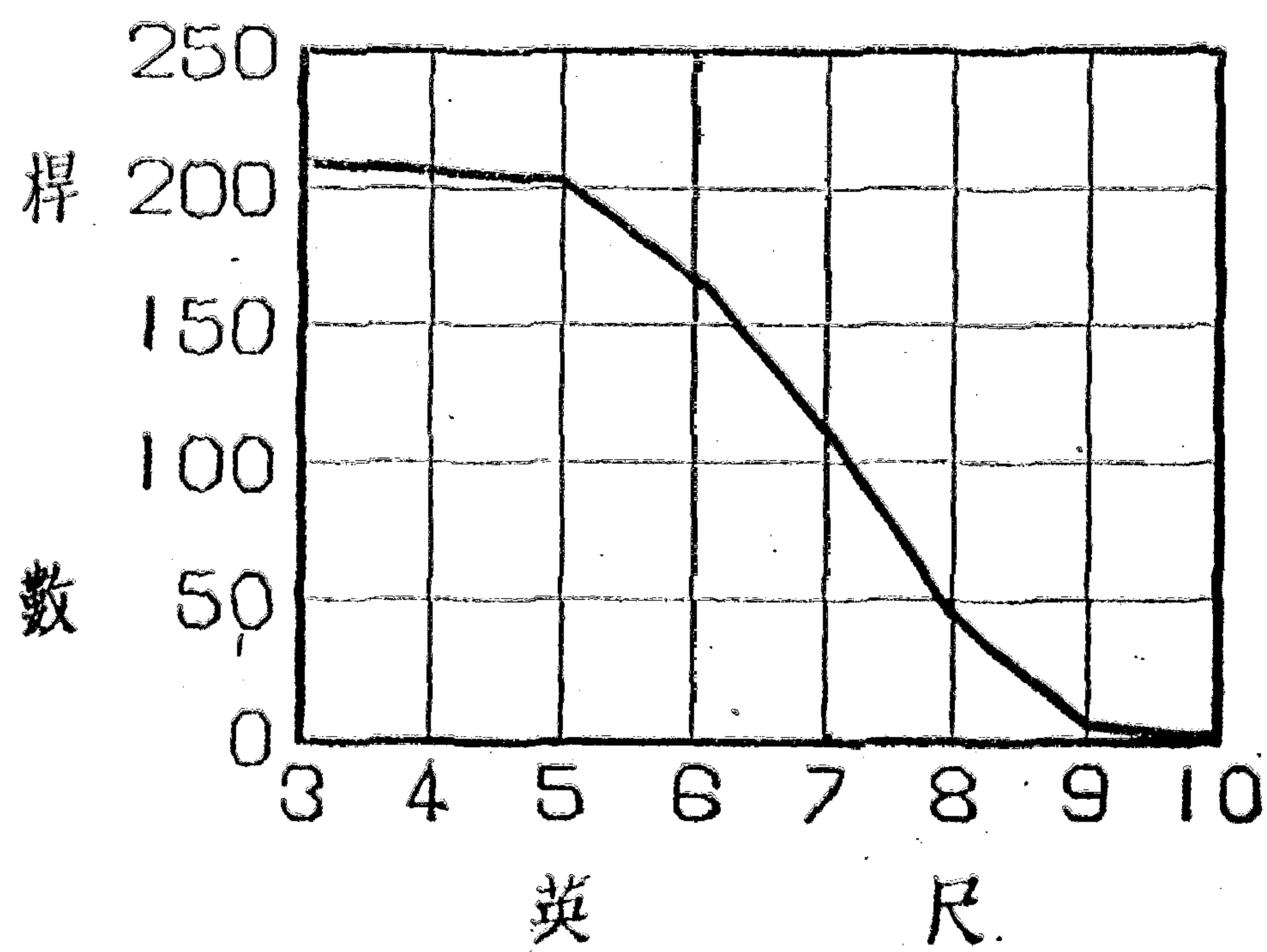
圖十四

217玉蜀黍桿高度累積分配

甲.(上向的)



乙.(下向的)



表十一

217玉蜀黍桿高度分配

高度(呎)	桿數 簡單次數	累積次數「以下」	「以上」
3—4	3	3	217
4—5	7	10	214
5—6	22	32	207
6—7	60	92	185
7—8	85	177	125
8—9	32	209	40
9-10	8	217	8
總數	217		

上。從累積次數曲線圖上可以尋出衆數,中位數,四分位數,十分位數及百分位數等之數值,容於下章再論之。

5. 累積時間曲線圖 (Cumulative Historical Curve) 累積時間曲線圖之繪法與累積次數曲線同,惟所遞加者,非各組之次數,而爲各期之數值或次數。此項曲線亦分上向與下向兩種,惟上向者非表示某組「以下」之次數,乃係表示某期「以前」及某期之數值或次數;下向者非表示某組「以上」之次數

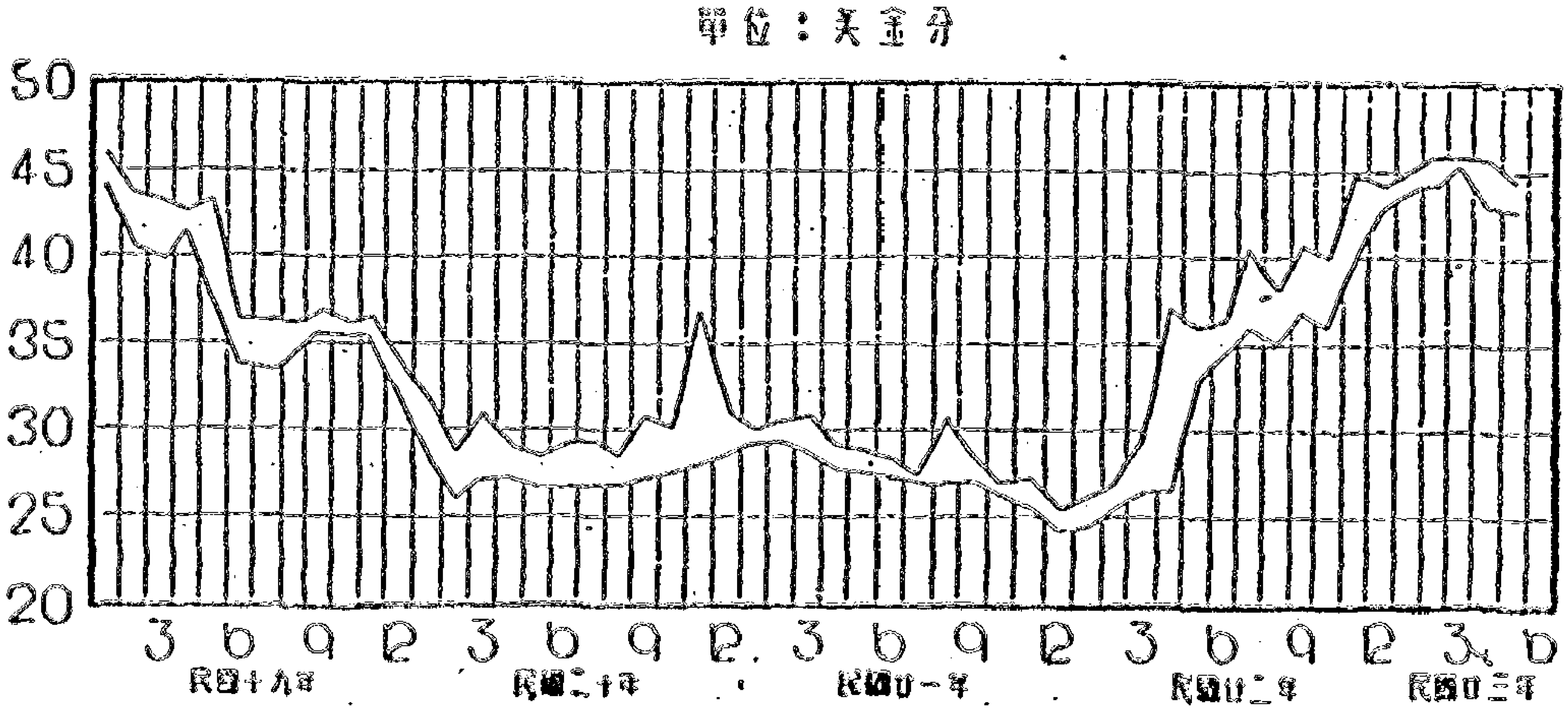
乃係表示某期及某期「以後」之數值或次數,此累積時間曲線圖與累積次數曲線圖相異之點也(例從略)。

6. 距限曲線圖 (Zone Curve) 以上所述各種曲線圖,每期或每組之數值祇有一種,設欲表示一期或一組中兩種不同之數值,則非用距限曲線圖不可。如表示一期中利率之高低,物價之變動,皆可用此種曲線表示。圖十五表示近五年來各月紐約銀市最高及最低價,即距限曲線圖之一例。

圖十五

近五年來紐約銀價

單位：美金分



此外尚有各種形式不同之曲線如帶紋曲線圖 (Band Curve), 分歧曲線圖 (Divergence

Curve)等,然其繪製之方法與上述諸圖初無大異,故不贅。

乙、等比曲線圖　等比曲線圖之線格,以相等之距離,表示相等之比率,如吾人欲表現事實按倍數或比率增減者,應採用此種曲線圖。等比曲線圖之繪製,須利用對數表或對數尺,故亦名對數曲線圖。縱橫線格悉按等比尺度劃分者為全對數曲線圖,若僅縱格用等比尺度,橫格仍用等差尺度,則為半對數曲線圖。茲分述於次:

a. 半對數曲線圖 (Semi-logarithmic Curve)

此種曲線在經濟統計上為用甚廣。如下例甲乙兩種物價四十年來上漲之速率完全相等,然若用等差曲線圖表示之,其緩急殊不相同,如圖十六甲;試以半對數曲線圖表示之,則得斜度相等之兩平行直線,如圖十六乙。可見此等處用等差曲線顯屬謬誤也。

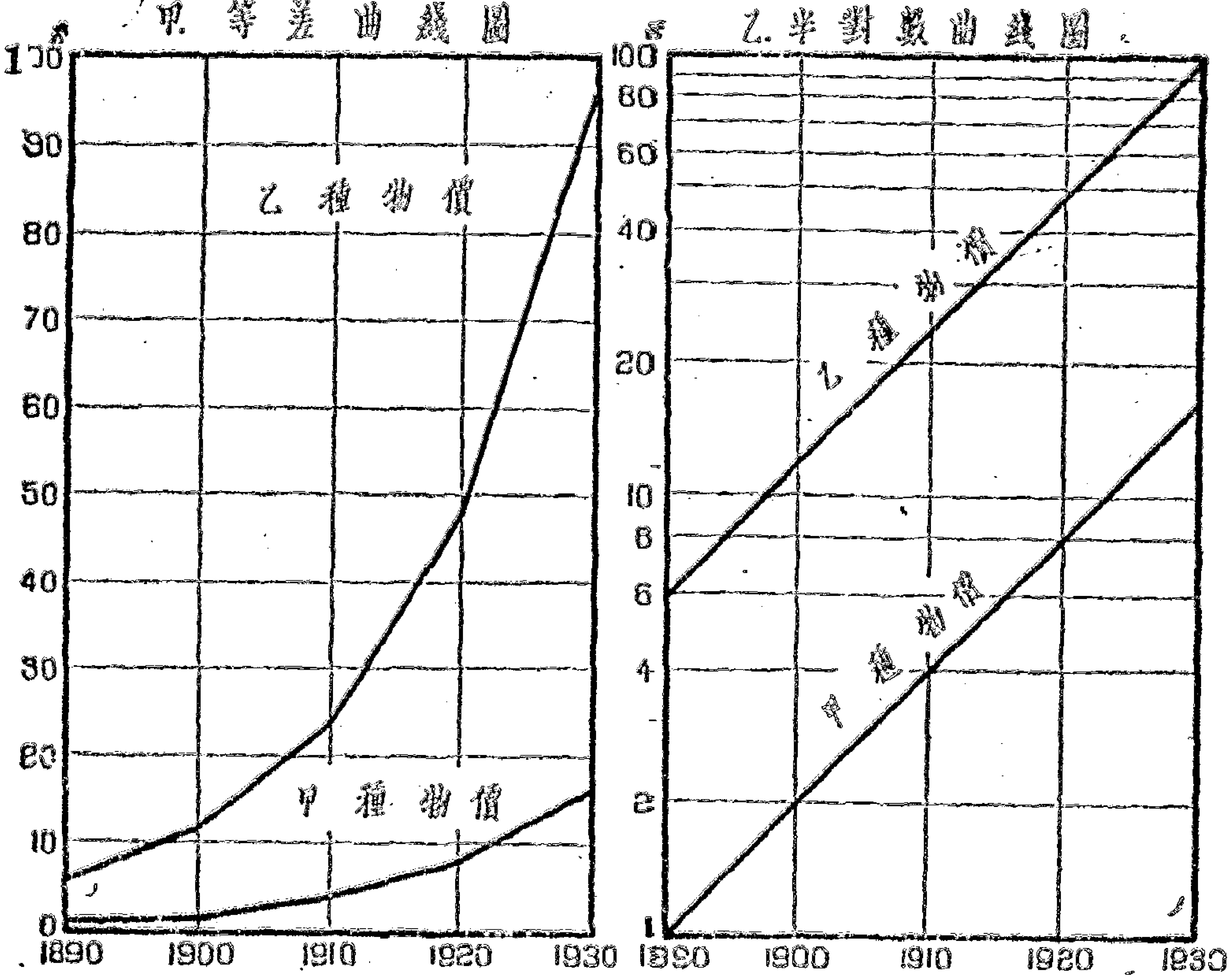
四十年來甲乙兩種物品之價格

1890年	甲 $1.00	乙 $ 6.00
1900年	2.00	12.00

1910 年	4.00	24.00
1920 年	8.00	48.00
1930 年	16.00	96.00

圖十六

甲乙兩種物價上漲速率比較



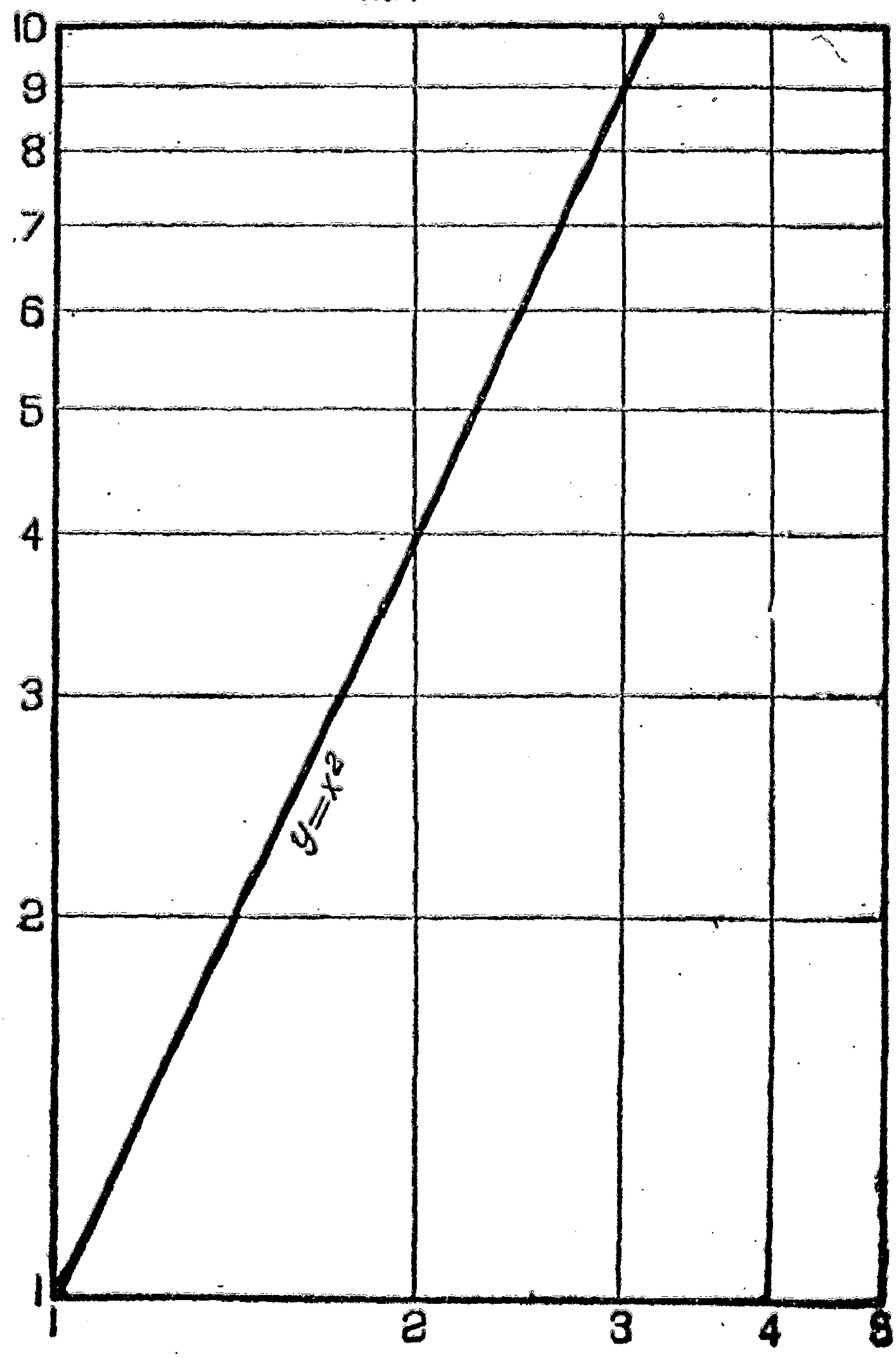
b. 全對數曲線圖(Logarithmic Curve)　全對數曲線圖之繪製,縱橫線格悉用等比尺度,如 $y=x^2$ 一方程式,用等差曲線圖表示之爲一拋物線,若用全對數曲線圖表示之,則成一直線,例如圖十七。惟當注意者,凡用等比尺度之線格,其起點決不能爲

$y=4$

$x=2$

零,如定欲將零度線表出,可用破裂紋劃分之(詳見下節).

圖十七

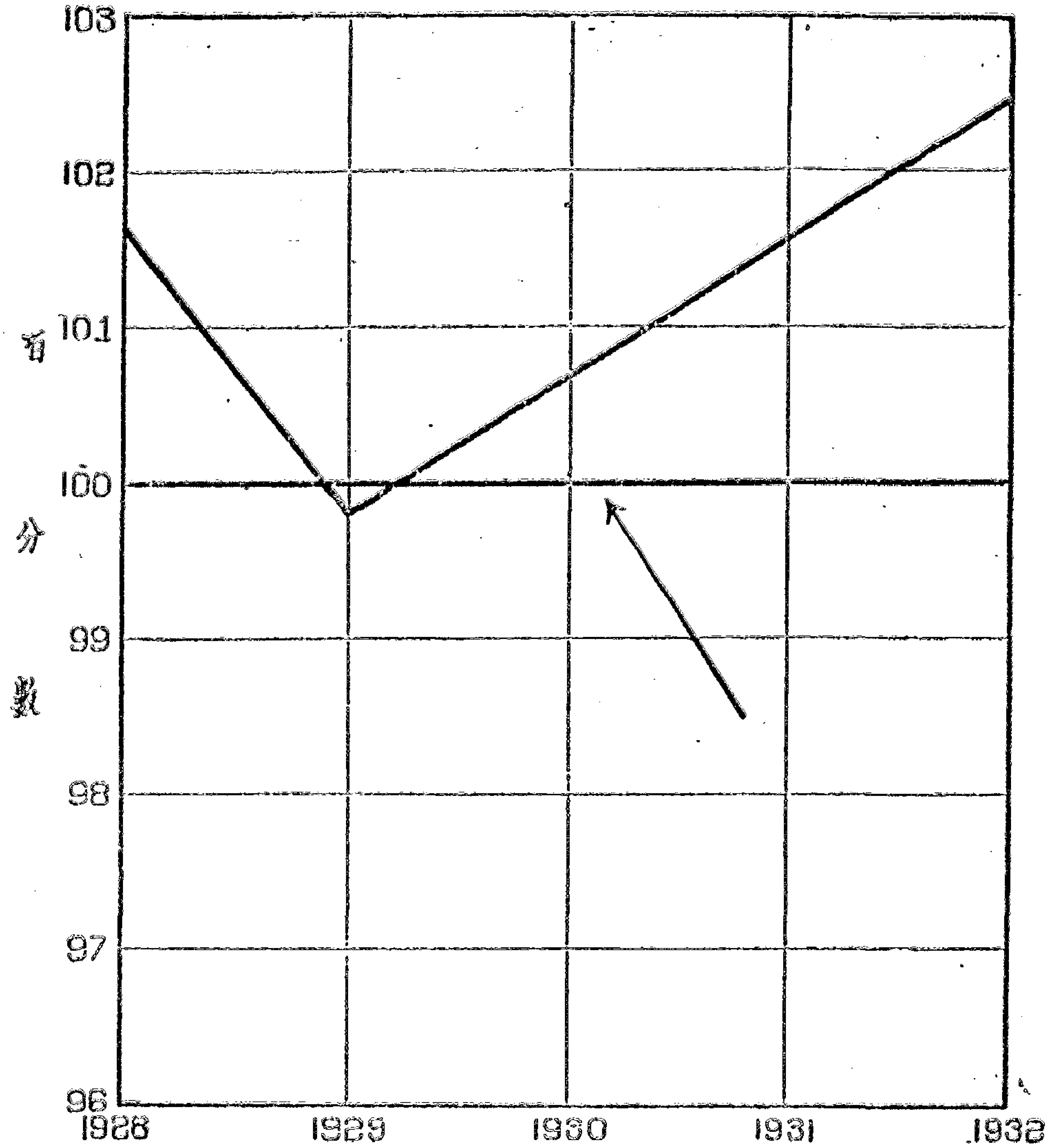


第三節 繪圖之規律

關於繪圖之規律,在一九一五年美國使用統計圖形各機關嘗推派代表組織一聯合委員會(Joint Committee on Standards for Graphic Presentation)討論之,主席爲柏林頓氏(Willard C. Brinton)。茲將其擬定之標準規律十七條迻譯於後:

1. 圖形之排列,例須由左至右。
2. 表示無連續性之數量時,最好用直條圖,因平面圖與立體圖常易誤解。
3. 如用曲線表示,非至不得已時,縱線之尺度,須以零度爲起點。
4. 倘曲線圖形太大,零度線不便表示時,可將零度線添在下面,中間用破裂紋表示其中一部份略去(參看圖十三)。
5. 用爲起點之零度線須比圖中線格稍粗,以示區別。
6. 如曲線係用百分法表示,則代表百分之線,或用爲比較之基線,須較其他線格略粗,例如圖十八。
7. 凡有時間關係之圖形,如所表示之年限,並非一個完全時期,則左右兩邊線,不必特別加粗。

圖十八



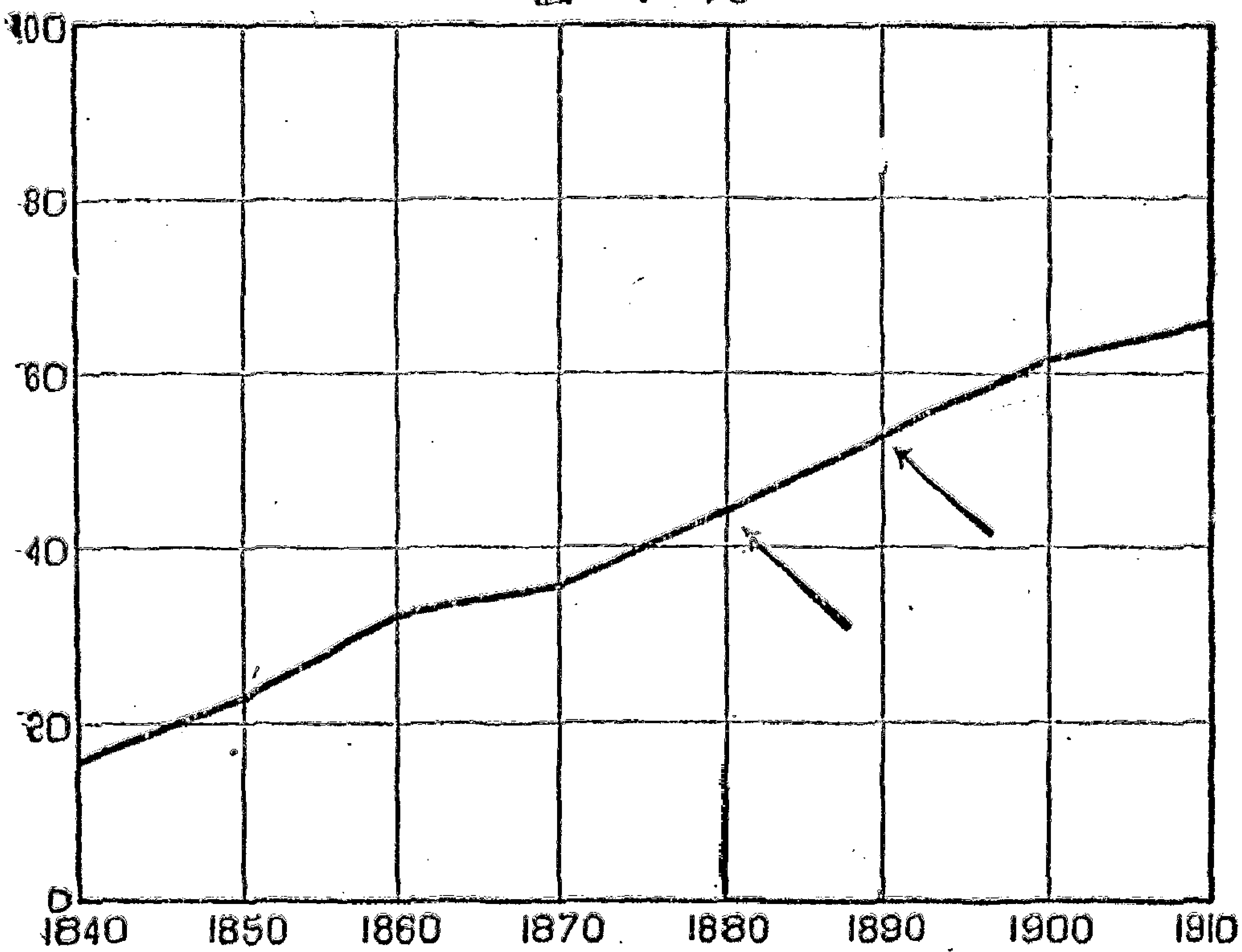
8. 用對數的格紙作曲線圖時,圖形之限制線,最好在對數尺度的十乘冪上。

9. 圖中縱橫線格不宜過多,以能幫助閱者便於觀察圖形上之數目及了解其意義爲度。

10. 圖中曲線應較線格爲粗。

11. 設圖中曲線係表示一系連續觀察之結果,最好將各次之結果均明白指出,例如圖十九。

圖十九



12. 圖中線格上之尺度,橫者須從左至右讀之,縱者須由下至上讀之。

13. 圖中數字須寫在左邊及下面,或沿圖之中軸線。

14. 圖形所表示之事實,或所根據之公式,均可記入圖中(參看圖十六,十七)。

15. 如詳細數字,圖形中不便記入,可以另列一表,附於圖旁(參看圖十三)。

16. 圖中數字或文字均宜使其能由下向上或由左向右讀之。

17. 圖之名稱(Titles)宜清楚完備,遇必要時,並應加以註釋。

第七章　平均數

第一節　平均數之意義及種類

統計數列分爲三類,前已言之。就中以時間數列與次數數列之性質最不相同,故其研究方法,必當分論。至地理數列與次數數列則極爲相近,其方法無須分述。學者苟能明瞭研究次數數列之方法,則分析地理數列,毫無困難矣。分析次數數列之方法,第一步在將散漫之資料製成次數分配表,其手續已於第五章第二節中論之。此項次數分配表製成後,雖可稍稍顯示次數分配之大概情形,尙不足供比較研究之用。例如有兩班學生於此,吾人如欲比較其成績之優劣,於分配表以外,應更就每班中求一代表的成績,以爲比較之根據,所謂代表的成績者,卽平均數也。故平均數非異常之事項,乃通常之事項,非極端之事項,乃中心之事項,平均數代表性之高低,視次數集中之程度而定。平均數有五種:一、算術平均數(Arithmetic average),二、中位數(Median),三、衆數(Mode),四、幾何平均數(Geometric average),五、倒數平均數(Harmonic average),請於以下各節分別論之。

第二節　算術平均數

算術平均數乃吾人頭腦中之狹義的平均數,亦即算術上之平均數,乃以次數總數除各項變數數值相加之和所得之商。計算算術平均數之公式有三,茲分述如下:

一、由未分組之材料求算術平均數之公式:

$$M=\frac{\Sigma m}{N}$$

式中: M爲算術平均數之符號

m爲各項變數之數值

Σ爲總加之符號,讀如Sigma

N爲次數之總數

所謂未分組之材料,乃未經製成次數分配表之意。其實材料即雖製成次數分配表,設各組之次數,均爲一時,亦可應用上述公式求算術平均數。例如有學生九名,其英文成績爲60、70、40,50、10、20,80、30、90,則算術平均數應爲:

$$\frac{60+70+40+50+10+20+80+30+90}{9}=\frac{450}{9}=50,$$

此爲從未分組之材料求得者。如將此九名學生之成績製成次數分配表,以十爲組距,則每組之次數均爲一,其平均結果亦爲五十分。惟當注意者,在材料未分組時,式中m所代表者爲各項之數值,在已分組之材料中,則m所代

表者,爲各組之中點,蓋假定各組次數之平均數值等於該組之中點也。

表十二

九名學生英文成績分配

分數	m	f
5-15	10	1
15-25	20	1
25-35	30	1
35-45	40	1
45-55	50	1
55-65	60	1
65-75	70	1
75-85	80	1
85-95	90	1
	$\Sigma m=450$	$N=9$

$$M=\frac{450}{9}=50$$

應用此式求算術平均數,每項變數之數值在結果中均佔同等之勢力,換言之,即輕重之程度相等,故亦名簡單算術平均數(Simple arithmetic average)。

二、由已分組之材料求算術平均數之公式:

$$M=\frac{\Sigma fm}{N}$$

式中:M 爲算術平均數之符號

m 爲各組之中點

f 爲各組之次數

Σ爲總加之符號

N爲次數之和

已分組之材料,乃將散漫之材料製成次數分配表之意,前例九名學生英文成績之分配,即其一例。惟材料之所以必須製成分配表,不外以其次數衆多,故有化繁爲簡之必要。普通次數分配表中各組之次數未必爲一。此處所謂已分組之材料,乃各組次數不盡爲一之分配表也。例如學生之成績5-15分者有二人,15-25分者有三人,25-35分者五人,35-45分者六人,45-55分者八人,55-65分者十二人,65-75分者九人,75-85分者三人,85-95分者一人,則其算術平均數爲

表十三

	m	f	fm
5-15	10	2	20
15-25	20	3	60
25-35	30	5	150
35-45	40	6	240
45-55	50	8	400
55-65	60	12	720
65-75	70	9	630
75-85	80	3	240
85-95	90	1	90
		N=49	Σfm=2,550

$$M=\frac{2,550}{49}=52.04$$

應用此式計算算術平均數須以各組之次數乘各組之中點,俾各項變數之數值在結果中保持相互間適當之比例,如得十分之學生二人,在四十九名學生中佔四十九分之二,故以二乘十。得二十分之學生三人,在四十九名學生中佔四十九分之三,故以三乘二十;餘類推。以其按各組之次數分別各組之輕重,隱寓加權(Weighting)之意味,故亦名加權算術平均數(Weighted arithmetic average)。惟權數(Weights)之原義,非專指次數而言,凡分別統計事項輕重程度所用之資料,或為數量,或為價值,或為抽象之數字,皆權數也。如吾人統計一家每日生活必需品柴、米、油、鹽等之消費值,必以各該品之消費量乘其價格,如柴幾斤,米幾升,油鹽各幾兩,此項數量斤、升、兩之類,即為各該品價格之權數。又如求學生各科之平均成績,常與各科以抽象之權數,以分別各科之輕重程度,如國文為三,歷史為二,數學為二·五之類。

三、計算算術平均數之簡便公式:

$$M=M'+C$$

式中:M 爲算術平均數之符號

M' 爲假定算術平均數之符號

C 爲校正數,等於$\frac{\Sigma d'}{N}$或$\frac{\Sigma fd'}{N}$

由已分組之材料求算術平均數已較由未分組之材料計算爲簡便,然如次數過多,其手續猶嫌煩瑣,尤以求各組次數乘各組中點時爲甚,故簡法尙矣。欲明簡法之理論,須先明瞭各項變數與其算術平均數之離中差(Deviations)之和等於零。蓋一切變數有大於其算術平均數者,亦有小於其算術平均數者,此項變數與其算術平均數間之差額,謂之離中差,在公式上恆以 d 代之。離中差有正有負,依數學原理,由算術平均數計算之各項離中差,其和應爲零,玆證明如下:

設 $m_1, m_2, m_3, \cdots\cdots m_n$ 爲各項數値

M 爲算術平均數

$d_1, d_2, d_3, \cdots\cdots d_n$ 爲各項變數對於算術平均數之離中差

則 $M = \frac{m_1+m_2+m_3+\cdots\cdots+m_n}{N}$ ……(1)

$NM = m_1+m_2+m_3+\cdots\cdots+m_n$ ……(2)

$0 = m_1+m_2+m_3+\cdots\cdots m_n - NM$ ……(3)

因N爲次數之總數

故 $0 = (m_1-M)+(m_2-M)+(m_3-M)+\cdots\cdots$
(m_n-M) ……………………………………(4)

但 $(m_1-M)=d_1$; $(m_2-M)=d_2\cdots\cdots$

故 $0 = d_1+d_2+d_3+\cdots\cdots+d_n$ ……………………(5)

$0 = \Sigma d$ ……………………………………………(6)

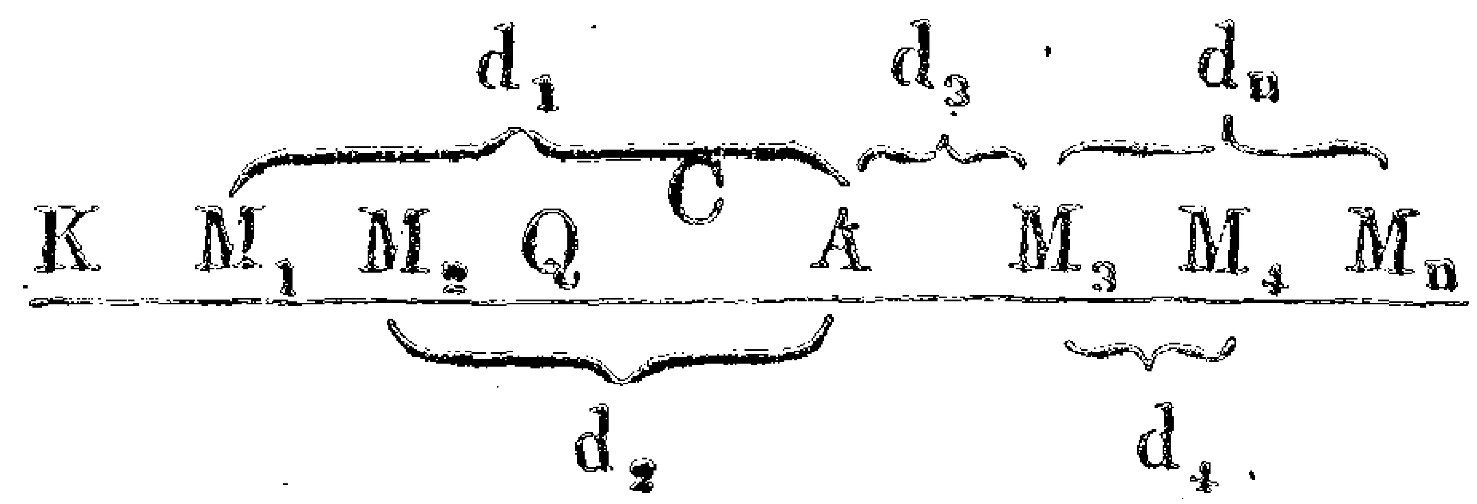
既知各項變數對於其算術平均數之離中差總和爲零,則對於算術平均數以外另一數值之離中差,其總和自不等於零。計算算術平均數之簡法,第一步乃先假定一數爲算術平均數,此假定平均數通常以分配中適中一組之中點充之。然後就此假定平均數求離中差,計其總和,而以次數總數除之,結果爲校正數。注意此校正數之正負符號,與假定算術平均數相加,卽爲眞確算術平均數 $(M=M'+C)$。玆證明此簡法之公式如下:

設 KM_1, KM_2, KM_3, KM_4,……KM_n爲各項變數

KA 爲眞確算術平均數

KQ 爲假定算術平均數

C 爲眞確算術平均數與假定算術平均數之差卽簡法公式中之校正數



則$d_1, d_2, d_3, d_4, \cdots\cdots\cdots\cdots d_n$爲各項變數對於眞確算術平均數之離中差。

而$d_1-c; d_2-c; d_3+c; d_4+c; \cdots\cdots d_n+c$爲各項變數對於假定算術平均數之離中差，通常以$d'$表示之。今將各項變數對於假定算術平均數之離中差相加而以次數總數(N)除之，則得：

$$\frac{\Sigma d'}{N}=\frac{-(d_1-c)-(d_2-c)+(d_3+c)+(d_4+c)+\cdots\cdots+(d_n+c)}{N}$$

$$=\frac{-d_1+c-d_2+c+d_3+c+d_4+c+\cdots\cdots+d_n+c}{N}$$

$$=\frac{\Sigma d+NC}{N}$$

已知$\Sigma d=0$

故　$\frac{\Sigma d'}{N}=C$

由上圖已知$KA=KQ+C$。今既證明$C=\frac{\Sigma d'}{N}$，則$KA=KQ+\frac{\Sigma d'}{N}$。KA者，眞確算術平均數也(M)；KQ者，假定算術平均數也(M')；故得計算算術平均數之簡便公式$M=M'+C$。惟以上係按未

分組之材料證明,若材料業經分組,C之求法亦同,惟須以各組之次數(f)乘各組之中點對於假定算術平均數之離中差(d')然後相加,而除以次數總數耳。此項簡法雖然在未分組或已分組之材料均可應用,但材料旣未分組,次數必少,用簡法計算算術平均數,是否可以節省時間與勞力,殊爲疑問。下列二例,不過用簡法釋明計算算術平均數之方法,實際上鮮有用簡法從未分組之材料求算術平均數者。

表十四

用簡法從未分組之材料

計算算術平均數

m	f	d'
10	1	−45
20	1	−35
30	1	−25
40	1	−15
50	1	−5
60	1	+5
70	1	+15
80	1	+25
90	1	+35
	N=9	Σd'=−45

設 $M'=55$

則 $C=\frac{-45}{9}=-5$

$$M = M' + C$$
$$= 55 - 5 = 50$$

表 十 五

用簡法從已分組之材料計算算術平均數

I	II	III	IV	V	VI	VII
	m	f	d'	fd'	d'	fd'
5—15	10	2	—40	—80	—4	—8
15—25	20	3	—30	—90	—3	—9
25—35	30	5	—20	—100	—2	—10
35—45	40	6	—10	—60	—1	—6
45—55	50	8	0	0	0	0
55—65	60	12	+10	+120	+1	+12
65—75	70	9	+20	+180	+2	+18
75—85	80	3	+30	+90	+3	+9
85—95	90	1	+40	+40	+4	+4
		N=49		Σfd'=430 —330 =100		Σfd'=43—33 =10

設$M' = 50$

則$C = \frac{100}{49}$(原來單位)

$M = M' + C$

$= 50 + 2.04$

$= 52.04$

$M' = 50$

$C = \frac{10}{49}$(以組距爲單位)

$= .204$

若以原來單位表示之，

則$C = .204 \times 10$(組距)

$= 2.04$

故$M = 50 + 2.04$

$= 52.04$

用簡法計算算術平均數固已簡捷,然計算各項或各組中點與假定算術平均數之離中差,倘可以組距爲單位,使手續愈加簡便。如表十五之組距爲十,第四欄所示,係用原來單位求出之d',設吾人以組距爲單位計算d',如第六欄所示,則較之用原來單位表示者小十倍,與次數(f)相乘時自較便捷。惟d'之計算,既以組距十爲單位,則求得之校正數(c)亦必較應得之校正數小十倍,故須以十乘之,俾其還原,然後與假定算術平均數相加乃得眞確算術平均數(參看表十五第六、七兩欄下面M之計算法)。茲將用簡便公式計算算術平均數之步驟列下:

1. 將資料製成次數分配表。
2. 擇適中一組之中點爲假定算術平均數。
3. 以組距爲單位表示各項變數對於假定平均數之離中差(d'),以假定平均數所在一組之離中差爲零,其下一組爲 -1, 其上一組爲 $+1$, 餘遞推。
4. 以各組之離中差(d')與各該組之次數相乘,注意其正負號,列於fd'欄下,然後相加,以求其總和(注意其正負號),是爲 $\Sigma fd'$

5. 以次數總數(N)除$\Sigma fd'$，所得之商，爲以組距爲單位之校正數(C)。

6. 以組距之數值乘第五步求得之校正數，即爲以原來單位表示之校正數。

7. 將此校正數與假定算術平均數相加(注意正負號)，即得眞確算術平均數。

第三節　中位數

中位數爲全體變數依大小次序排列後中間一項或一點之數值，在中位數之兩端各有全體變數之半。換言之，即小於中位數之變數與大於中位數之變數各爲全體變數百分之五十。求中位數之公式有二：一爲由未分組之材料求中位數，一爲由已分組之材料求中位數，茲分述於下：

一、由未分組之材料求中位數之公式(如雖分組而各組次數均爲一時，亦可用此公式)：

$$Md=\frac{N+1}{2}$$

式中：Md爲中位數之符號

N 爲次數總數

應用此式求中位數須先將各變數按大小順次排列，然後計其次數，如表十二，九名學生之英文成績，由小至大順序排列應爲10,20,30,40,50,60,70,80,90，按照

公式,則中位數爲五$\left(\frac{9+1}{2}=5\right)$,意即第五項之數值50爲中位數,五,其地位也。若次數爲偶數,則中位數之地位在居中兩項間之中點,其數值應爲居中兩項相加之半。

二、由已分組之材料求中位數之公式:

1.

$$Md=\frac{N}{2}$$

式中:Md爲中位數所在之地位

N爲次數總數

2.

$$Md=L+\frac{i}{f}\times c\cdots\cdots\cdots\cdots\cdots\cdots(\text{甲})$$

或

$$Md=U-\frac{i}{f}\times c\cdots\cdots\cdots\cdots\cdots\cdots(\text{乙})$$

式中:Md爲中位數之數值

L爲中位數所在組之低限

U爲中位數所在組之高限

f爲中位數所在組之次數

i在(甲)爲中位數所在組之低限與中位數所在點間之次數

在(乙)爲中位數所在組之高限與中位數所在點間之次數

c 爲組距

由未分組之材料求中位數至爲易易;若由已分組之材料求中位數,則其手續較爲繁複。第一步吾人須先按 $Md=\frac{N}{2}$ 公式決定中位數之地位,即中位數所在之一點,然後用 $Md=L+\frac{i}{f}\times c$ 或 $Md=U-\frac{i}{f}\times c$ 公式求出該點之數值。茲用二百十七玉蜀黍桿高度之分配爲例,計算其中位數如下:

表十六

二百十七玉蜀黍桿高度累積分配

呎	f		c. f. 以下	c. f. 以上
3-4	3		3	217
4-5	7		10	214
5-6	22		32	207
6-7	60		92	185
7-8	85	$\begin{cases}16.5\\68.5\end{cases}$	177	125
8-9	32		209	40
9-10	8		217	8
	N=217			

(1)先用 $Md=\frac{N}{2}$ 公式決定中位數所在之一點

$$Md=\frac{217}{2}=108.5。$$

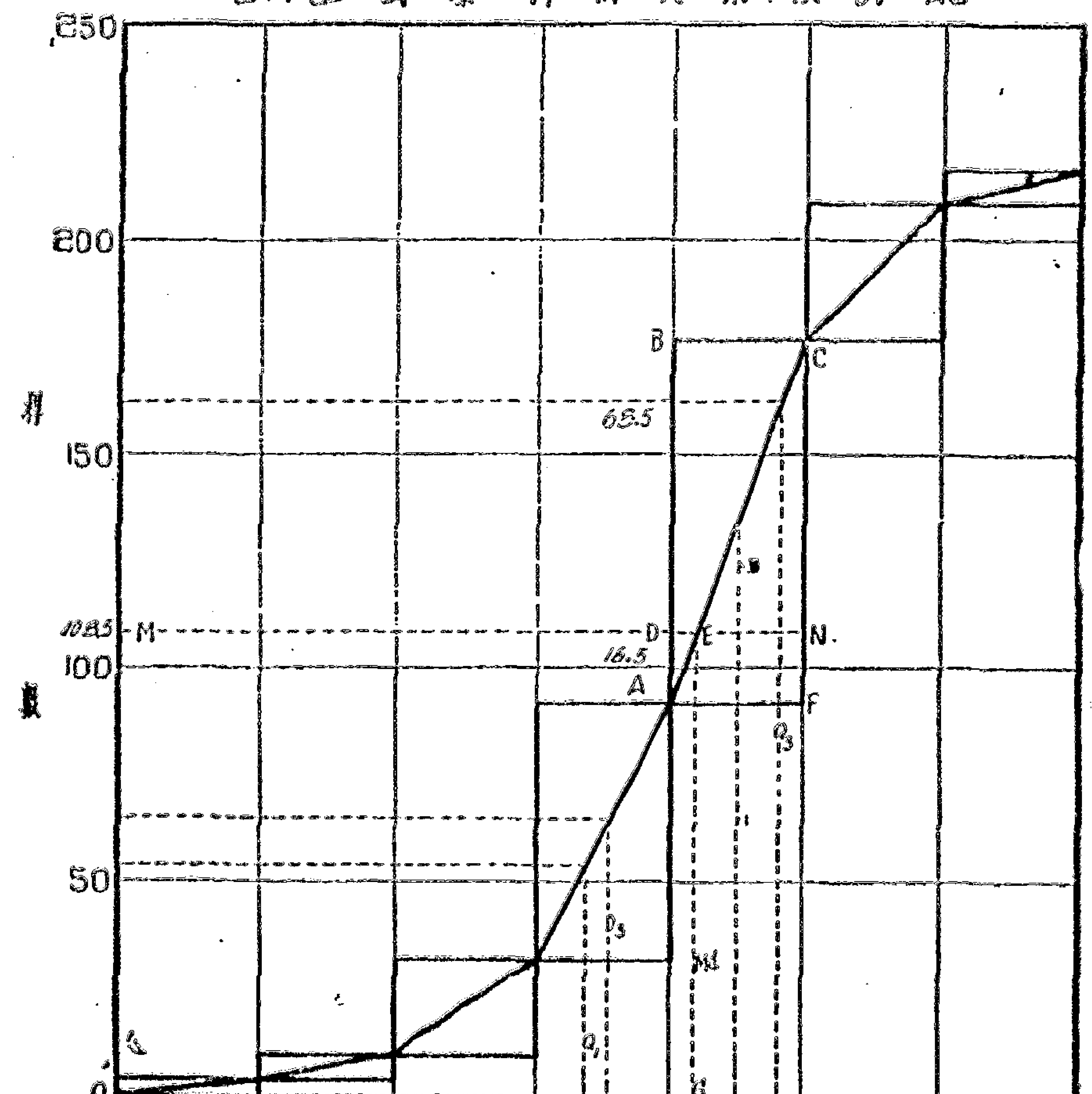
(2)既知中位數所在點爲108.5,故中位數之數值必介乎7呎與8呎之間,因由該組之累

積次數所示,知中位數確在該組也。然該組之低限為 7 呎,高限為 8 呎,中位數究較低限高幾何,高限低幾何,是不得不用補插法以決定之矣。查中位數所在組之累積次數無論為「以下」或「以上」均超過 108.5 之地位,由「以下」的累積次數而論,超過 68.5 (177-108.5),即從中位數所在組之次數 (85) 中取 16.5 加於較小一組之累積次數 (92) 即得中位數所在之地位 $(92+16.5=108.5)$, 16.5 即甲式中之 i。由「以上」的累積次數而論,超過 16.5 (125-108.5),即從中位數所在組之次數 (85) 中取 68.5 加於較大一組之累積次數 (40) 即得中位數所在之地位 $(40+68.5=108.5)$, 68.5 即乙式中之 i。吾人將材料製成次數分配表時本有一假定,即各組之次數其分配均屬平勻,可以該組之中點為該組各項之平均數值。本例之中位數在 7-8 之一組 85 項中,由小至大計數為第 16.5 項,故其數值應為 $7+\frac{16.5}{85}\times 1=7.19$ $(Md=L+\frac{i}{f}\times c)$。若由大至小計數為 68.5 項,故其數值應為 $8-\frac{68.5}{85}\times 1=7.19$ $(Md=U-\frac{i}{f}\times c)$。

用補插法決定中位數數值之公式可以下圖證明之:

圖二十一

217玉蜀黍桿高度累積分配



上圖為一累積次數曲線,吾人由縱軸上可以尋出中位數之地位(108.5),從中位數所在之點引出MN一線與底線平行,此線與累積次數曲線在E點

相交,E點在橫軸上之數值,即中位數之數值,E點所在之組即中位數所在組。求E點在橫軸上之數值,可由E點引一垂直線至底線,如圖二十之EG。OG即中位數之數值,用目力觀察,約為7.2弱。然OG=ME,故吾人亦可逕求ME之數值。圖中ME=MD+DE或ME=MN−EN,但MD=7,為中位數所在組之低限,MN=8,為中位數所在組之高限,以補插法計算者僅DE或EN之數值而已。依幾何學原理△ABC與△ADE為相似三角形,故

$$\frac{AD}{AB}=\frac{DE}{BC}$$

$$DE=\frac{AD\times BC}{AB}$$

查 AD=16.5 (中位數所在組之低限與中位數所在點間之次數)

AB=85 (中位數所在組之次數)

BC=1 (組距)

故 $DE=\frac{16.5\times 1}{85}=.19$(即甲式中之$\frac{i}{f}\times c$)

又知△ACE與△ECN為相似三角形,故

$$\frac{CN}{CF}=\frac{EN}{AF}$$

$$EN=\frac{CN\times AF}{CF}$$

查 CN=68.5 (中位數所在組之高限與中位數所在點間之次數)

AF＝1 (組距)

CF＝85 (中位數所在組之次數)

故 $EN=\frac{68.5\times1}{85}=.81$ (卽乙式中之$\frac{i}{f}\times c$)

吾人由中位數所在組之低限上加.19,或由中位數所在組之高限減去.81,均爲7.19,此卽中位數之數值也。茲將用甲式求中位數之步驟列下:

1. 將資料製成次數分配表。
2. 以二除次數總數。決定中位數所在之地位及中位數兩端之次數。
3. 將各組次數(由小至大)累積相加,以決定中位數所在之一組。
4. 計算此組低限以上至中位數所在點間之次數(i)。
5. 以該組原有次數(f)除(i),然後乘以組距,再加於該組之低限上(L),卽得中位數之數值。

中位數之外,尙有所謂四分位數(Quartiles),十分位數(Deciles),百分位數者(Percentiles),其計算方法與中位數同。所異者中位數係將全體次數等分爲二,四分位數係將全體次數等分爲四,十分位數係將全體次數等分爲十,百分位數係將全體次數等分爲百。中位數之兩端各有全體次數之半,第二四分位數,第五

十分位數及第五十百分位數之兩端,亦各有全體次數之半,故數值相等。至若第一四分位數所在之地位下端應有全體次數四分之一,上端應有全體次數四分之三。第三四分位數所在之地位下端應有全體次數四分之三,上端應有全體次數四分之一。十分位數與百分位數之地位可以類推。中位數吾人以Md代之,四分位數則以Q代之,十分位數則以D代之,百分位數則以P代之。求中位數時吾人須先決定其地位,然後用公式甲或乙計算其數值,求四分位數、十分位數及百分位數時亦然。茲用玉蜀黍稈高度之材料,略舉數例於下:

1. 第一四分位數:

a. 決定第一四分位數之地位: $\frac{N}{4}=\frac{217}{4}=54.25.$

b. 用甲式求第一四分位數之數值:

$$Q_1=6+\frac{22.25}{60}\times 1=6.37.$$

2. 第三四分位數:

a. 決定第三四分位數之地位:

$$\frac{N}{4}\times 3=\frac{217\times 3}{4}=162.75.$$

b. 用甲式求第三四分位數之數值:

$$Q_3=7+\frac{70.75}{85}\times 1=7.83.$$

3. 第三十分位數:

a.決定第三十分位數之地位:

$$\frac{N}{10}\times 3=\frac{217}{10}\times 3=65.1.$$

b.用甲式求第三十分位數之數值:

$$D_3=6+\frac{33.1}{60}\times 1=6.55.$$

4.第五十三百分位數:

a.決定第五十三百分位數之地位:

$$\frac{N}{100}\times 53=\frac{217}{100}\times 53=115.01.$$

b.用甲式求第五十三百分位數之數值:

$$P_{53}=7+\frac{23.01}{85}\times 1=7.27.$$

上例求數值時均用甲式,如用乙式,其方法相同,茲不贅述。中位數之數值,可從累積次數曲線圖中求出,四分位數,十分位數及百分位數之數值亦可從此項累積次數曲線圖中求出,由縱軸上第一四分位數所在點引出之線與底綫平行者與此曲線(上向的)相交之點,其在横軸上之數值,卽爲第一四分位數之數值。若爲下向的累積次數曲線,則該點之位置移向右方,其在横軸上之數值乃第三四分位數之數值而非第一四分位數之數值矣(餘類推)。下向的累積次數曲線如與上向者同繪於一圖,則此兩曲線之相交點在横軸上之數值卽爲中位數之數值,學者宜注意及之。

第四節 衆數

衆數之觀念最易明瞭,統計材料中次數最密集之一點,即爲衆數。如最普通之工資,最普通之成績,均屬衆數,以其爲最普通之事實,故可用爲全體事項之代表。衆數之求法,可分粗率的及精確的兩種方法,用粗率的方法求出之衆數,謂之視察衆數(Inspection Mode),用精確的方法求出之衆數,謂之理論衆數(Theoretical Mode)。惟眞正精確的衆數非儘量增多材料,配合一常態曲線(Normal Curve of Error)後不能求得,此處所謂理論衆數不過較之視察衆數稍近正確耳。

一、視察衆數　求視察衆數之方法有三:一。爲從次數分配表中求出,一爲從次數多邊圖或直方圖中求出,一爲從累積次數曲線圖中求出,茲分述於下:

a. 從次數分配表中求衆數,以次數最多一組之中點爲衆數之數值,如表十六中次數最多一組之中點爲7.5,即以7.5爲衆數之數值。

b. 從次數多邊圖或直方圖中求衆數,以次數多邊圖最高一點在橫軸上之數值或直方圖中最高一直方之中點在橫軸上之數值爲衆數。如圖十二次數多邊形最高一點及直方形中

最高一直方之中點在橫軸上之數值均爲7.5,即衆數之數值也。

c.從累積次數曲線中求衆數 累積次數曲線平峭之程度,視各組次數之多寡而定,如圖二十,其始甚平,繼則漸峭,迨至項部又成平勢。此圖最峭之點爲 7—8 一組之中點,就此點而求其在底線上之數值,卽爲衆數,蓋 7—8 一組中之次數最多,故曲線呈最峭之勢也(參閱圖二十)。

二、理論衆數 精確之理論衆數殊不易求得,而視察衆數又多粗率不可恃,於是由統計學家之經驗,倡用兩法,其結果較之視察衆數稍近於正確,卽此處所謂理論衆數,亦名近似衆數(Approximate Mode), 玆分述於次:

a.由視察衆數求比較近似的衆數 此法係根據鄰近衆數所在組(卽次數分配表中次數最多之一組)上下兩組之次數以糾正視察衆數之錯誤。蓋次數分配如屬對稱(Symmetrical),換言之,卽衆數所在組上下兩端之次數相等,則吾人不妨以衆數所在組之中點爲衆數之數值;如次數分配爲不對稱 (Asymmetrical), 則衆

數之地位須視上下兩端次數之多寡而定。如小於衆數一端之次數較大於衆數一端之次數爲多,則衆數之數值小於衆數所在組之中點;反之,則衆數之數值大於衆數所在組之中點。蓋衆數所在組上下兩端之次數不啻爲兩種勢力,足以左右衆數之數值,如勢均力敵,則衆數不動,否則衆數常向次數較多之一端移動。統計學家根據此種理論得一求近似衆數之公式如下:

$$MO = L + \frac{f_2}{f_2 + f_1} \times c$$

式中: MO 爲衆數之符號

L 爲衆數所在組之低限

f_2 爲衆數所在組較大一組之次數

f_1 爲衆數所在組較小一組之次數

c 爲組距

茲以表十六之材料爲例,求得衆數如下:

$$MO = 7 + \frac{32}{32 + 60} \times 1 = 7.35$$

本例衆數所在組較大一組之次數較較小一組之次數爲少,故衆數之數值,小於衆數所在組之中點(參閱表十六及圖十二)。

上式爲由衆數所在組之低限加一數量求得衆

數,吾人亦可由衆數所在組之高限減一數量求得衆數,其式如下:

$$MO=U-\frac{f_1}{f_1+f_2}\times c$$

式中: MO 爲衆數之符號

U 爲衆數所在組之高限

f_1 爲衆數所在組較小一組之次數

f_2 爲衆數所在組較大一組之次數

c 爲組距

茲仍以表十六之材料爲例,求得衆數如下:

$$MO=8-\frac{60}{60+32}\times 1=7.35$$

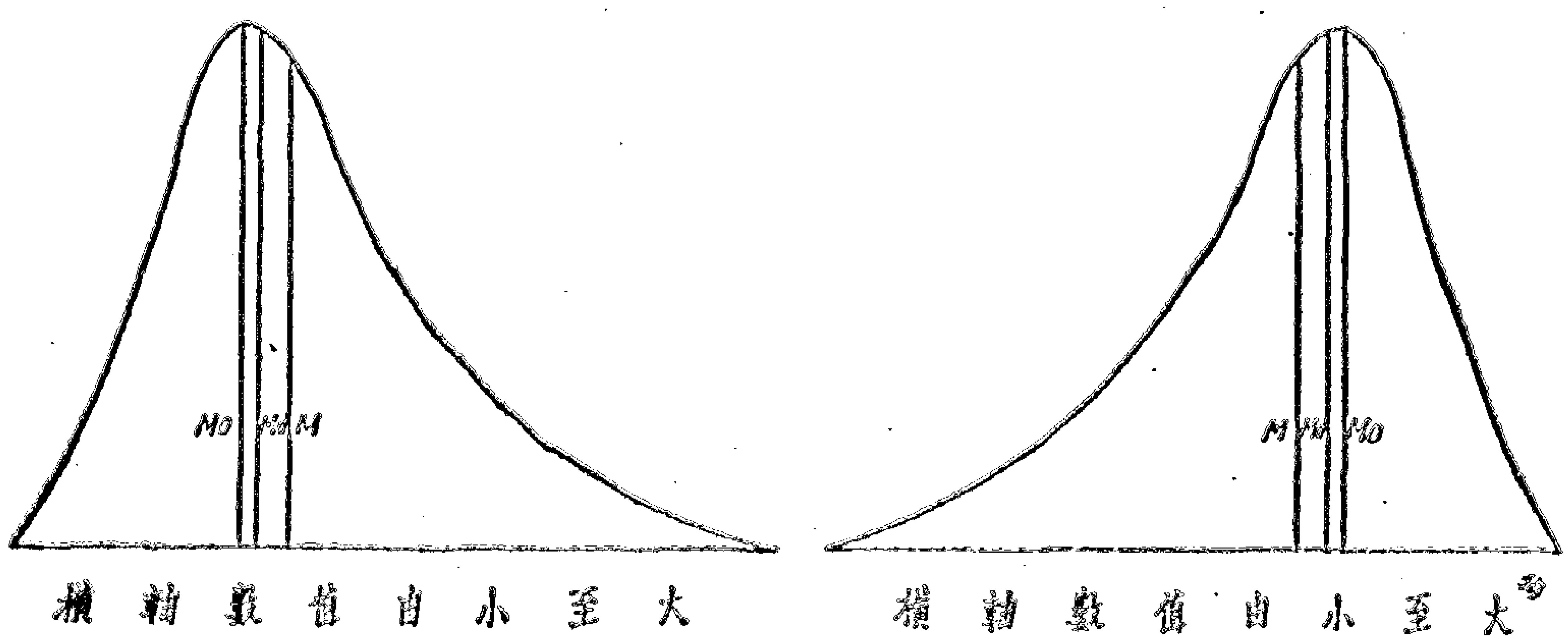
上例 f_2 與 f_1 所代表者爲衆數所在組較大一組與較小一組之次數,有時吾人取衆數上下兩組或三組之次數以決定衆數之數値,其方法相同,無庸贅述。

b. 皮爾生(Karl Pearson)計算衆數之經驗的法則

當次數分配爲對稱時,算術平均數,中位數,衆數三者合而爲一;其數値相等;如分配爲不對稱,而偏斜度不高時,三者之數値各不相同,但亦有相當之關係。即算術平均數與衆數相距最遠,中位數居於二者之間,中位數與算術平均數之距離約等於算術平均數與衆數間距離三分之一。如次數分配呈正的偏態,則算

術平均數大於衆數;如爲負的偏態,則算術平均數小於衆數。所謂正的偏態者,即次數分配偏於數值較大之一端,如圖二十一甲;負的偏態者,即次數分配偏於數值較小之一端,如圖二十一乙。

圖二十一



根據上述理論,皮爾生倡用一種經驗的法則,由算術平均數及中位數之數值求出衆數,其公式如下:

$$MO = M - 3(M - Md)$$

利用此式,如已知衆數及中位數之數值,或衆數及算術平均數之數值,亦可求算術平均數或中位數。但算術平均數與中位數有較精確之方法可以求出,故此式僅可用於求衆數,且非至不得

已時不宜輕用,因次數分配偏斜度稍高,則上述算術平均數,中位數及衆數三者之關係即不存在,由此式求出之衆數不免謬誤也。

上述諸法(除皮爾生之經驗的法則外)均假定次數分配中,有一次數最多之組,一望而知為衆數所在組。然若次數分配為不規則的,則衆數所在組頗不易決定,遇有此種情形,應用併組法 (Grouping method) 決定之,其法如下:

表十七

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

上表次數分配極不規則,次數最多之一組為第九組(即中點13之一組)然若一計該組隣近上下兩組之次數,則以13為衆數似又不及以9為衆數,因9一組內之次數僅較13一組之次數少一,而合計隣近上下兩組之次數則較13一組隣近上下兩組之次數為

多。吾人如欲知衆數究在何組，應將組距加大，或較原來之組距加大一倍，或加大兩倍。表十七所示第一步係從第一組起每兩組次數相加，即組距加大一倍；第二步係從第二組起每兩組次數相加，組距仍係加大一倍；第三步係從第一組起每三組次數相加，即組距加大兩倍；第四步係從第二組起每三組次數相加，組距亦係加大兩倍；第五步係從第三組起每三組次數相加，組距仍係加大兩倍。如此繼續爲之，由第一步併組之結果，知衆數在 13–14 一組；由第二步併組之結果，知衆數在 8–9 一組；由第三步併組之結果，知衆數在8–10一組；由第四步併組之結果，知衆數在9–11 一組；由第五步併組之結果，知衆數在 7–9 一組。各步併組之結果除第一步外，最多次數常在第五組及其隣近之上下一二組內，故知衆數之數値，實近於 9 也。惟吾人遇不規則的次數分配時，有當注意者，即此分配是否因包含性質不同之材料而有一個以上之衆數，如其有之，應分爲數分配，然後各求其衆數，否則不獨應用併組法後仍無明顯之衆數可尋，且次數分配亦將失其效用矣。

第五節　幾何平均數

算術平均數，中位數及衆數三者爲平均數量之

方法,有時吾人欲求比(Ratio)或率(Rates)之平均數,則非用幾何平均數與倒數平均數不可。幾何平均數者,乃 n 個比或率相乘後開 n 次方所得之方根也。計算幾何平均數之公式亦分簡單的與加權的兩種:

1.簡單幾何平均公式:

$$Mg=\sqrt[n]{r_1\times r_2\times r_3\times\cdots\cdots\cdots\times r_n}$$

式中: Mg 為幾何平均數之符號

r 為各個比或率

n 為比或率之次數

2.加權幾何平均公式:

$$Mg=\sqrt[n]{r_1^{W_1}\times r_2^{W_2}\times r_3^{W_3}\times\cdots\cdots\times r_n^{W_n}}$$

式中:Mg 為幾何平均數之符號

r 為各個比或率

w 為 r 之次數或權數

n 為比或率之次數總數或權數之和

但實際上幾何平均數之計算須利用對數(Logarithm),上列二式可化為:

1.簡單幾何平均公式:

$$\log Mg=\frac{\log r_1+\log r_2+\log r_3+\cdots\cdots\cdots+\log r_n}{n}$$

2.加權幾何平均公式:

$$\log Mg=\frac{W_1\log r_1+W_2\log r_2+W_3\log r_3+\cdots\cdots+W_n\log r_n}{n}$$

蓋按對數之法則,各項比或率之幾何平均數之對數,等於各項比或率對數之算術平均數。在加權公式中各項比或率之權數爲各該比或率之指數(Exponents),利用對數法,則此項權數變爲各該比或率之係數(Coefficients) 或乘數(Multipliers)。惟利用對數計算之結果爲幾何平均數之對數,吾人須由對數表中求得其眞數(Natural number),方爲幾何平均數。故利用對數計算幾何平均數,含有兩步手續,一爲由眞數求對數,一爲由對數求眞數,學者於此兩步手續如能了然,則計算幾何平均數毫無困難矣。茲舉例如次:

1. 簡單幾何平均數

表十八

五項物品之價比

(假定以某年之物價作爲100)

	r	logr
甲	120	2.07918
乙	125	2.09691
丙	140	2.14613
丁	186	2.26951
戊	201	2.30320

$5\,\underline{)\,10.89493} = \Sigma \log r$

$$2.17899$$

$$\log Mg = 2.17899$$

$$Mg = 151.0$$

2.加權幾何平均數

表十九

五項物品之價比

（假定以某年之物價作爲100）

	r	權數 (w)	logr	wlogr
甲	120	1	2.07918	2.07918
乙	125	2	2.09691	4.19382
丙	140	3	2.14613	6.43839
丁	186	3	2.26951	6.80853
戊	201	1	2.30320	2.30320
		$\Sigma w = 10$	10	21.82312 $= \Sigma(w\log r)$
				2.18231

$$\log Mg = 2.18231$$

$$Mg = 152.2$$

幾何平均數在經濟統計中常用以計算物價指數。例如有甲乙二物，甲物之價由100%跌至10%，乙物之價由100%漲至1000%，若用算術平均數計算，則平

均爲505%($\frac{1,000+10}{2}=505$);若用幾何平均數計算,則平均仍爲 100% ($\sqrt{1000\times10}=100$)。就此例而論,自以幾何平均數之結果爲正確,因一跌十倍,一漲十倍,漲落相銷,結果應仍爲100%也。

此外任何事物依複利率增減者,如欲求其平均增減之速率,亦當用幾何平均數。複利率之公式爲$P_n=P_0(1+r)^n$,茲舉數例說明其應用於下:

例一:求複利率 例如有本金千元,按複利計息,十二年後,本利共得一千六百元。按算術平均數求利率爲 5 %,然此決非眞相,其眞正利率應爲 4 %,算法如下:

$$P_n=P_0(1+r)^n$$

式中: P_0爲最初之本金

P_n爲 n 年或期後之本金

r 爲利率

n 爲時期

$$\frac{P_n}{P_0}=(1+r)^n$$

$$\log P_n-\log P_0=n\log(1+r)$$

$$\log1600-\log1000=12\log(1+r)$$

$$3.20412-3.00000=12\log(1+r)$$

$$\frac{.20412}{12}=\log(1+r)$$

$$.01701 = \log(1+r)$$

$$1.04 = 1+r$$

$$r = .04 \text{ 卽 } 4\%$$

例二:求人口增加率　人口之增加,大概亦依複利率,故欲知兩調查年度間人口增加之速率,亦須用幾何平均數。例如某城1920年之人口爲70,000,1930年之人口爲100,000,則平均每年增加之速率爲百分之三·六。算法如下:

$$\log 100{,}000 - \log 70{,}000 = 10\log(1+r)$$

$$5.00000 - 4.84510 = 10\log(1+r)$$

$$\frac{0.15490}{10} = \log(1+r)$$

$$0.01549 = \log(1+r)$$

$$1.036 = 1+r$$

$$r = .036 \text{ 卽 } 3.6\%$$

例三:求能力增進率　能力之增進,亦依複利率,故教育測驗上計算此項增進率時,亦用幾何平均數。例如一組學生某種技能學習三月之後,已進步百分之九十,則每月平均之增進率應爲百分之二三·九。算法如下:

$$\log 190 - \log 100 = 3\log(1+r)$$

$$2.27875-2.00000=3\log(1+r)$$

$$\frac{.27875}{3}=\log(1+r)$$

$$.09292=\log(1+r)$$

$$1.239=1+r$$

$r=.239$即23.9%

第六節 倒數平均數

倒數平均數乃各個比或率倒數之算術平均數之倒數。凡求每小時速率之平均,非用此種平均數不可。計算倒數平均數之公式,亦分簡單的與加權的兩種:

1. 簡單倒數平均公式:

$$H=\frac{1}{\dfrac{\dfrac{1}{r_1}+\dfrac{1}{r_2}+\dfrac{1}{r_3}+\cdots\cdots+\dfrac{1}{r_n}}{n}}$$

式中:H為倒數平均數之符號

r為所欲平均之各項比或率

n為次數總數

2. 加權倒數平均公式

$$H=\frac{1}{\dfrac{\dfrac{W_1}{r_1}+\dfrac{W_2}{r_2}+\dfrac{W_3}{r_3}+\cdots\cdots+\dfrac{W_n}{r_n}}{n}}$$

式中:H為倒數平均數之符號

r 爲所欲平均之各項比或率

W 爲各項次數或權數

n 爲次數總數或權數之和(ΣW)

惟倒數平均數之意義極其晦澀,故其應用之範圍不廣。茲舉二例說明簡單倒數平均數之應用,其加權公式之例從略。

例一:求每小時之平均速率 例如有學生二人,同受數學測驗,甲一小時做對四題,乙一小時做對二題,若用算術平均計算,其平均速率爲每小時三題$\left(\frac{4+2}{2}=3\right)$。但此甲乙二生每小時之平均速率並非每小時三題,而爲二題又三分之二,蓋若爲三題,則每做一題,平均需二十分鐘,今甲生每做一題需十五分鐘,乙生每做一題需三十分鐘,平均爲二十二分鐘又二分之一$\left(\frac{15+30}{2}=22.5\right)$,以之除每小時之分數應爲$2\frac{2}{3}$也。凡遇此種問題,應求其倒數平均數。

$$H=\frac{1}{\dfrac{\frac{1}{4}+\frac{1}{2}}{2}}$$

$$=\frac{1}{\frac{\frac{2+1}{4}}{2}}$$

$$=\frac{1}{\frac{3}{8}}$$

$$=\frac{8}{3}$$

$=2\frac{2}{3}$ 每小時平均計算之題數

例二:求平均物價　在經濟統計中求平均物價,而物品之價格係以每元若干個表示者,亦須用倒數平均數。例如有甲、乙、丙三種物品,甲價每元四個,乙價每元五個,丙價每元二十個,若用算術平均數計算,則每元平均可買 $9\frac{2}{3}$ 個 $\left(\frac{4+5+20}{3}=9\frac{2}{3}\right)$,每個平均價為 10.34 分。但依原價而論,甲每個價 25 分,乙每個價 20 分,丙每個價五分,其算術平均數為 $16\frac{2}{3}$ 分,平均每元僅能買六個。遇有此種情形,吾人亦須求其倒數平均數。算法如下:

$$H=\frac{1}{\frac{\frac{1}{4}+\frac{1}{5}+\frac{1}{20}}{3}}$$

$$=\frac{1}{\frac{\frac{5+4+1}{20}}{3}}$$

$$=\frac{1}{\frac{10}{60}}$$

$$=\frac{60}{10}$$

$=\quad 6$ 每元平均可買之個數

倒數平均數之計算,可利用倒數表(Table of Reciprocals),然如祗二三項,可逕用$H=\frac{2ab}{a+b}$或$H=\frac{3abc}{ab+ac+bc}$之公式計算,式中a, b, c,即爲所欲平均之各項比或率,二式均由簡單倒數平均公式蛻化而來,故其結果與逐步計算而得者,初無二致。

第七節　各種平均數之優劣及關係

平均數計有五種,其計算方法已如上述,茲請論其優劣及關係如下:

甲、各種平均數之優劣:

一、算術平均數之計算,一切變數鉅細靡遺,但其結果常受極端數量(尤其是大的數量)之影響,中位數與衆數則不受兩極端(極大或極小)數量之影響。然中位數猶與衆數有別;蓋中位數之地位爲全體變數依次排列後居中之一項或一點,若兩端增減幾項,中位數必隨之移動,而於衆數則毫無影響。幾何平均數受極端數量之影響亦少,故亦較算術平均數爲穩定。倒

數平均數與算術平均數相反,常受小的極端數量(大數量之倒數)之影響,故算術平均數之結果常偏高,倒數平均數之結果常偏低。

二、如所研究之材料次數無多,且極散漫,並無集中之傾向,則衆數爲不適用。例如一城人民之財富均不相同,獨有三人有同等之財產爲一萬元,若遽視爲此城之平均財富,殊有未妥。故就此點言,衆數不及中位數,算術平均數將一切數量悉算在內,尤無此弊。

三、算術平均數,幾何平均數,倒數平均數三者均由計算而得,故可用代數方法處理之,而中位數與衆數則不能。

四、五種平均數中,就決定之難易言,以中位數爲最簡便,衆數之近似的結果雖亦易求,但眞確衆數則不易決定。算術平均數非計算不可,手續似較繁。但從他方面論,中位數與衆數之決定,須先將一切變數依次排列或繪成次數曲線,而算術平均數則無須此種手續。倒數平均數之計算亦繁。至於幾何平均數則非利用對數不可。

五、吾人所研究之材料若兩極端變數之數量不

甚清楚,則以用中位數或衆數爲善。蓋衆數對於兩極端數量之次數完全不管,中位數則知其次數已足,變數之大小可以不問,非若計算算術平均數,幾何平均數,倒數平均數須詳知各項數量之大小也。

六、統計材料之不能量度者,如心理狀態之類,以用中位數求其平均爲適宜,算術平均數則不適用矣。

七、就普通人了解之難易言,則以算術平均數最爲簡單淺顯,倒數平均數之意義最爲晦澀,故統計分析上除非不得已時甚少用之也。幾何平均數對於等比之變化各與以同等之地位,故計算事物平均變動之比率時非用之不可。

乙、各種平均數之關係:

一、在完全對稱之次數分配中,算術平均數,中位數及衆數之數値相等。

二、在偏斜不甚之次數分配中,中位數之地位介於算術平均數與衆數之間。中位數與算術平均數間之距離約等於算術平均數與衆數間距離三分之一,此種關係可以下列公式表示之:

$$M-Mo=3(M-Md)$$

或

$$Mo=M-3(M-Md)$$

三、任何數量之算術平均數必大於其幾何平均數,幾何平均數又必大於倒數平均數,但有一例外,即各數量一一相等時,算術平均數,幾何平均數及倒數平均數亦皆相等。

四、任何二數(祇限二數)之幾何平均數等於其算術平均數與倒數平均數之幾何平均數。例如2與8兩數之倒數平均數爲$3\frac{1}{5}$,算術平均數爲5,幾何平均數爲4,4即$3\frac{1}{5}$與5之幾何平均數也。茲以代數式證明此關係如下:

設:

$$M=\frac{a+b}{2}$$

$$H=\frac{2ab}{a+b}$$

$$Mg=\sqrt{ab}$$

則: $\sqrt{\frac{a+b}{2}\times\frac{2ab}{a+b}}=\sqrt{ab}$

五、次數分配之離中趨勢如依算術定律支配者,

衆數與中位數往往與算術平均數相近。如依幾何定律支配,則衆數與中位數常與幾何平均數相近。

第八章　離中趨勢及偏斜度

第一節　離中趨勢之意義

統計材料製成次數分配表後,雖可稍稍窺測其集中之趨勢,然不若平均數所表示之中心傾向爲顯明。但僅有次數分配表及平均數,於次數分配之性質尙不能謂已窺其全豹。平均數之外尙須有離中趨勢與偏斜度 (Dispersion and skewness) 之測定,方足以明次數分配之眞相。平均數係表示一切數量集中之程度,而離中趨勢則係表示其離中之程度,平均數在次數曲線底線上爲一點,離中趨勢則爲一定之距離;平均數之意義隨離中趨勢之大小而定,離中趨勢大,則平均數之代表性低,離中趨勢小,則平均數之代表性高。至於偏斜度乃表示中心數值左右兩端次數之分配對稱或偏斜之情形。平均數已於上章詳述之矣,本章所討論者乃離中趨勢與偏斜度之問題也。

表示離中趨勢之單位有用原有事項之單位者,有用抽象之數字者,前者曰絕對離中趨勢 (Absolute Measure of Dispersion),後者曰相對離中趨勢或離中係數(Coefficients of Dispersion)。吾人研究一種事實分配之

離中程度時,求出其絕對離中趨勢已可知其大概,如各種物價漲落之緩急,一地財富分配之不齊,均可用絕對離中趨勢表示之。設吾人之目的為比較各時期物價變動程度之高低,各地財富分配差異之大小,則非將絕對離中趨勢化為相對離中趨勢不能比較,蓋各年之平均物價未必相同,各地之平均財富尤難一致也。計算絕對離中趨勢之方法較普通者有五:一、全距(Range)二、四分差(Quartile Deviation)三、平均差(Average Deviation)四、標準差(Standard Deviation)五、機差(Probable Error);計算相對離中趨勢則以皮爾生(Pearson)氏之離中係數應用最廣,茲分述於後。

第二節 全距

全距為測定離中趨勢最簡便之一法,從最大一項中減去最小一項,其差即全距也。如材料業經分組製成次數分配表,則全距為最大一組高限與最小一組低限間之差。以全距之大小,為測定分配疏密之標準。如表十一玉蜀黍桿之一例,最小一組之低限為3呎,最大一組之高限為10呎,$10-3=7$ 即全距也。全距之求法雖易,然其結果殊不可靠,蓋全距之大小,僅依極端兩項之數值而定,一二極大或極小事項之增減,影響於全距極大。況兩次數分配之全距相等而離中

趨勢不等者有之,離中趨勢相等而全距不等者亦有之,有此數端,全距不能用爲測定離中趨勢精密之尺度,自無疑義。

第三節　四分差

中位數爲全體變數依大小次序排列後中間一項或一點之數值,故求中位數時,將全體變數等分爲二。四分位數係將全體變數等分爲四,換言之,即在中位數兩端之變數中再求出其中點。在中位數較大一端之中點爲第三四分位數 (Upper quartile),在中位數較小一端之中點爲第一四分位數 (Lower quartile)。第一四分位數與第三四分位數之間含有全體變數二分之一,四分差者,乃利用第一四分位數與第三四分位數間距離之半,以間接表示離中之趨勢者也。蓋第三四分位數與第一四分位數間雖必含有全體變數之半,但其距離之長短,則視離中之狀態而定。離中趨勢愈大,則距離愈大;離中趨勢愈小,則距離愈短。就此距離而二分之,即得四分差,其公式如下:

$$Q.D.=\frac{Q_3-Q_1}{2}$$

式中:Q.D.爲四分差 (Quartile Deviation)之符號

Q_3爲第三四分位數

Q_1 爲第一四分位數

茲仍以玉蜀黍桿之材料爲例,計算四分差如下:

$$Q_3 = 7 + \frac{70.75}{85} \times 1 \doteq 7.83$$

$$Q_1 = 6 + \frac{22.25}{60} \times 1 = 6.37$$

$$Q.D. = \frac{7.83 - 6.37}{2} = .73$$

(關於 Q_3 及 Q_1 之詳細算法,請參閱第七章第三節)

如以底線上 Q_1 與 Q_3 中間一點之數值爲 K,則全體變數之半必在 K ± Q.D. 之距離中。K 之求法,或由 Q_3 減去 Q.D.,或於 Q_1 加上 Q.D.:

$$K = 7.83 - .73 = 7.10$$

或

$$K = 6.37 + .73 = 7.10$$

由此可知二百十七玉蜀黍桿之半當在 7.10 ± .73之距離中。再次數分配如屬完全對稱,則 K 之數值應與中位數相等,今 K=7.10,而 Md = 7.19(參閱第七章第三節),可見 217 玉蜀黍桿高度之分配,尚稍有偏斜也。四分差以全體變數當中一半之差異狀態代表全體變數之離中趨勢,不受兩端一二項極大或極小變數之影響,較之全距似勝一籌,然其代表性究不充分,

況全距及四分差均非眞正離中趨勢,不過用間接方法以觀察離中之程度而已。根據各項變數之離中差而測定離中趨勢之方法有二,所謂平均差與標準差是也。

第四節　平均差

平均差者,各項變數對於平均數離中差之平均也。各項變數有大於平均數者,亦有小於平均數者,故離中差有正有負,但計算平均差時,此項離中差無論爲正爲負,須一律相加,以求其平均,否則平均差縱不等於零,亦必甚小。平均差之計算,可從中位數,亦可從算術平均數或衆數,就理論言,則以從中位數計算較當,因平均差之數値以從中位數計算者爲最小也。計算平均差之公式有三:

1. 由未分組之材料求平均差之公式(如雖分組而各組次數均爲一時,亦可用此公式):

$$A.D.=\frac{\Sigma d}{N}$$

式中:A. D. 爲平均差(Average Deviation)之符號

d 爲各種變數與平均數之離中差

Σ 爲總加之符號

N 爲次數總數

茲舉例如下:

表二十

m	f	d
3	1	6
6	1	3
9	1	0
12	1	3
15	1	6

$Md=9 \qquad \Sigma d=18$

$$A.D.=\frac{18}{5}=3.6$$

上例算術平均數與中位數二者相等,第一項與平均數之離中差應爲－6,第五項與平均數之離中差應爲＋6,因計算平均差時此項正負號一概不問,故$\Sigma d=18$。

2.由已分組之材料求平均差之公式:

$$A.D.=\frac{\Sigma fd}{N}$$

式中: A.D.爲平均差之符號

d 爲各項變數與平均數之離中差

f 爲各組之次數

Σ 爲總加之符號

N 爲次數總和

茲舉例如下:

表二十一

組限	m	f	d	fd
3—4	3.5	3	3.69	11.07
4—5	4.5	7	2.69	18.83
5—6	5.5	22	1.69	37.18
6—7	6.5	60	.69	41.40
7—8	7.5	85	.31	26.35
8—9	8.5	32	1.31	41.92
9—10	9.5	8	2.31	18.48
		N=217		Σfd=195.23

Md=7.19(參閱第七章第三節)

$$A.D.=\frac{195.23}{217}=.9$$

3. 簡法:由眞確平均數計算平均差之手續殊嫌煩瑣,因平均數未必確爲整數,各項變數與平均數之離中差往往爲小數(如上例),與各組次數相乘時甚費時間。因此,遇組數及次數均多之分配表,可用較爲簡便之方法計算其平均差。簡法之公式如下:

$$A.D.=\frac{\Sigma fd'+(Ns-Nl)c}{N}$$

式中:A.D.爲平均差之符號

d' 爲各項變數與假定平均數之離中差

f 爲各組之次數

Σ 爲總加之符號

Ns 爲從假定平均數計算離中差較從眞確平均數少算之次數

Nl 爲從假定平均數計算離中差較從眞確平均數多算之次數

c 爲眞確平均數與假定平均數間之差

N 爲次數總和

用簡法計算平均差須求出眞確平均數,以眞確平均數所在組之中點爲假定平均數,然後就假定平均數以計算各項變數之離中差(d'),再與各組次數相乘得(fd'),更將各組之 fd'相加而得 Σfd'。如表二十二之例,眞確中位數爲 7.19,假定中位數爲 7.5,d' 欄所示爲各組中點與假定平均數之離中差(無小數), fd' 一欄則爲各組次數乘各組中點與假定平均數之離中差之結果。惟假定平均數旣與眞確平均數之數値不同,則就假定平均數計算之離中差與從眞確平均數計算而得者必有多少之差誤,此項差誤必須校正,方可得

正確之平均差,式中$(N_s-N_l)c$即此校正數也。蓋假定平均數如較眞確平均數爲大,中位數以下(即小於中位數之各組)各組變數之離中差必多算,中位數以上(即大於中位數之各組)各組之離中差必少算;反之,假定平均數如較眞確平均數爲小,則結果適相反。但平均數所在組各變數之離中差悉屬少算,因應用簡法計算,該組各變數之離中差均爲零也。各組變數之離中差多算或少算之數值,等於假定平均數與眞確平均數間之差(c),本例$c=.31$,有一項多算,則$\Sigma fd'$中應減去一倍.31,有一項少算,則$\Sigma fd'$中應加一倍.31。今多算92項,少算125項,故$\Sigma fd'$中應加33倍.31。學者如覺不易判別各組d'爲多算抑爲少算,可將從眞確平均數計算之離中差同時列出如第六欄,以資比較。

表二十二

I	II	III		IV	V	VI
組限	m	f		d'	fd'	d
3—4	3.5	3		4	12	3.69
4—5	4.5	7	多	3	21	2.69
5—6	5.5	22	算	2	44	1.69
6—7	6.5	60		1	60	.69

7—8	7.5	85	少算	0	0	.31
8—9	8.5	32		1	32	1.31
9—10	9.5	8		2	16	2.31
		N=217			Σfd'=185	

$\mathrm{Md}=7.19$

假定平均數爲7.5

$c=7.5-7.19=.31$

$\mathrm{Ns}=125$

$\mathrm{Nl}=92$

$$\mathrm{A.D.}=\frac{185+(125-92)\times .31}{217}$$
$$=\frac{185+10.23}{217}$$
$$=\frac{195.23}{217}$$
$$=\ .9$$

惟當注意者,應用簡法計算平均差,假定平均數必須與眞確平均數同在一組,否則結果必致錯誤。本例組距爲一,故計算離中差時卽以一爲單位,若組距不等於一,亦可以組距爲單位計算離中差以省手續。玆將用簡法求平均差之步驟列下:

1. 計算眞確中位數。

2. 以中位數所在組之中點爲假定平均數。

3. 以組距爲單位,計算各組中點與假定平均數

之離中差(d')。

4. 以各組次數乘第三步求得之離中差,得fd'欄各數,然後一一相加,不問離中差之爲正爲負,得 Σfd'。

5. 求校正數(c),卽中位數與假定平均數之差。如 d'以組距爲單位計算,此項校正數亦須以組距爲單位,其法卽以組距除之。

6. 全體變數對於假定平均數之離中差(d')有大於對眞確中位數之離中差者,亦有小於對眞確中位數之離中差者。前者次數之和爲公式中之Nl,後者次數之和爲公式中之Ns,求出Ns－Nl之差數,注意其正負號,然後與校正數相乘,是爲總校正數。

7. 將總校正數(注意其正負號)與 Σfd' 相加,而除以次數總和,卽得以組距爲單位之平均差。

8. 以組距乘第七步求得之結果,卽爲以原有單位表示之平均差。

第五節　標準差

平均差以各變數與平均數之離中差之平均測定次數分配之離中趨勢,固較全距及四分差爲精密,但計算離中差時,對於正負符號一概不問,就數學理

論言,未免牽強。皮爾生(Karl Pearson)氏因之發明一法,以一切離中差(各變數與算術平均數,中位數或衆數之差)自乘,以消去其負號,然後求各乘方之算術平均數,再開方以資還原,是爲標準差(Standard Deviation),亦名均方差(Root-mean-square Deviation).在統計學上恆以σ(Sigma)表示之。平均差之計算,常以中位數爲中心,而標準差之計算,則常以算術平均數爲中心,蓋平均差以從中位數計算者爲最小,而標準差則以從算術平均數計算者爲最小也。計算標準差之公式甚多,茲擇其中較普通及較簡捷者四種分述於後:

一,由未分組之材料求標準差之公式(如雖分組而各組次數均爲一時,亦可用此公式):

$$\sigma=\sqrt{\frac{\Sigma d^2}{N}}$$

式中:σ爲標準差之符號

d爲各變數與平均數之離中差

Σ爲總加之符號

N爲次數總數

茲舉例如下:

表二十三

m	f	d	d^2
3	1	-6	36

6	1	-3	9
9	1	0	0
12	1	+3	9
15	1	+6	36
	5		$\Sigma d^2=90$

$M=9$

$\sigma=\sqrt{\frac{90}{5}}$

$=\sqrt{18}$

$=4.24$

二、由已分組之材料求標準差之公式:

$$\sigma=\sqrt{\frac{\Sigma fd^2}{N}}$$

式中: σ 爲標準差之符號

d 爲各變數與平均數之離中差

f 爲各組次數

Σ 爲總加之符號

N 爲次數總和

茲舉例如下:

表二十四*

m	f	fm	d	d^2	fd^2
3	1	3	-6.5	42.25	42.25
6	2	12	-3.5	12.25	24.50

9	3	27	−.5	.25	.75
12	4	48	+2.5	6.25	25.00
15	1	15	+5.5	30.25	30.25
	N=11	Σfm=105			Σfd²=122.75

$$M=\frac{105}{11}$$
$$=9.5$$
$$\sigma=\sqrt{\frac{122.75}{11}}$$
$$=\sqrt{11.16}$$
$$=3.3$$

*計算離中差時可以組距爲單位俟 σ 求得後再行還原。

標準差之計算第一步須先求出算術平均數但算術平均數無論在未分組或已分組之材料中未必恰爲整數平均數如非整數離中差必含小數計算離中差之自乘時手續未免太繁故遇此等情形應用簡法計算標準差最簡捷之方法有二茲分述之。

三簡法一:此法之理論係先定一假定平均數就此假定平均數計算其標準差假定平均數可於次數分配中擇適中一組之中點充之惟此項假定平均數未必適與眞確平均數相等就假定平均數計算之標準差與由眞確平均數

計算者自有多少之差誤,故就假定平均數求得標準差後,須減去此項差誤,方爲眞確標準差,其公式如下:

甲,用簡法一由未分組之材料求標準差之公式:

$$\sigma = \sqrt{\frac{\Sigma d'^2}{N} - c^2}$$

式中: σ 爲標準差之符號

d' 爲各變數與假定平均數之離中差

Σ 爲總加之符號

N 爲次數總數

c 爲改正數等於 $\frac{\Sigma d'}{N}$

乙,用簡法一由已分組之材料求標準差之公式:

$$\sigma = \sqrt{\frac{\Sigma f d'^2}{N} - c^2}$$

式中: σ 爲標準差之符號

d' 爲各組中點與假定平均數之離中差

f 爲各組之次數

Σ 爲總加之符號

N 爲次數總和

c 為改正數等於$\frac{\Sigma fd'}{N}$

上述甲式實即等於第一式,乙式等於第二式,茲將甲式之所以等於第一式用代數法證明於下,至於乙式等於第二式之證明,其法相同,故不贅。

證明:

$$\sqrt{\frac{\Sigma d'^2}{N}-\left(\frac{\Sigma d'}{N}\right)^2} = \sqrt{\frac{\Sigma d^2}{N}}$$

因 $M=M'+c$

$d=m-M$

$\therefore\ d=m-M'-c$

$d=d'-c$

$d'=d+c \cdots\cdots\cdots\cdots\cdots\cdots(1)$

$d'^2=d^2+2cd+c^2 \cdots\cdots\cdots\cdots\cdots\cdots(2)$

$\Sigma d'^2=\Sigma d^2+2c\Sigma d+Nc^2 \cdots\cdots\cdots\cdots\cdots\cdots(3)$

但 $\Sigma d=0$

故 $\Sigma d'^2=\Sigma d^2+Nc^2 \cdots\cdots\cdots\cdots\cdots\cdots(4)$

$\frac{\Sigma d^2}{N}=\frac{\Sigma d'^2}{N}-c^2 \cdots\cdots\cdots\cdots\cdots\cdots(5)$

$\sqrt{\frac{\Sigma d^2}{N}}=\sqrt{\frac{\Sigma d'^2}{N}-\left(\frac{\Sigma d'}{N}\right)^2} \cdots\cdots(6)$

茲就表二十五及表二十六兩次數分配分別用甲乙二式計算其標準差如下:

表 二 十 五

(根據甲式計算)

m	f	d'	d'^2
2	1	−4	16
4	1	−2	4
6	1	0	0
8	1	+2	4
10	1	+4	16
12	1	+6	36
	N=6	$\Sigma d'=6$	$\Sigma d'^2=76$

設　$M'=6$

$$\sigma=\sqrt{\frac{76}{6}-\left(\frac{6}{6}\right)^2}$$

$$=\sqrt{12.7-1}$$

$$=\sqrt{11.7}$$

$$=\ 3.4+$$

表二十六

(根據乙式計算)

(1)	(2)	(3)	(4)	(5)	(6)	(7)
m	f	d'	fd'	fd'^2	$(d'+1)^2$	$f(d'+1)^2$
3	1	−2	−2	4	1	1
6	2	−1	−2	2	0	0
9	3	0	0	0	1	3
12	4	+1	+4	4	4	16

15	1	+2	+2	4	9	9
	N=11		Σfd'=+2	Σfd'²=14		29

設　$M'=9$

$$c=\frac{2}{11}=.18\text{(以組距爲單位)}$$

$$\Sigma fd'^2=14\text{(以組距爲單位)}$$

$$\sigma=\sqrt{\frac{14}{11}-.18^2}$$

$$=\sqrt{1.27-.03}$$

$$=\sqrt{1.24}$$

$$=1.1\text{(以組距表示之標準差)}$$

$$\sigma=3.3\text{(以原來單位表示之標準差)}$$

表二十六中離中差之計算係以組距(3)爲單位,如將校正數(c)改爲原來單位加於假定算術平均數之上,即得眞確算術平均數,惟本例係計算標準差,算術平均數之數値無須求出,故(c)可不必還原,迨 σ 求得後始乘以組距(3),是爲以原來單位表示之標準差。表中第六第七兩直行則爲校核標準差計算之正誤而作,此法係薛立愛(Charlier)氏所倡,故名薛立愛氏校核法(Charlier Check),其法係從假定算術平均數所在組下一組(即較小一組)之中點計算各組變數之離中差,結果等於$(d'+1)$,將各組之$(d'+1)$自乘列於第六直行下,再乘以各組之次數得 $f(d'+1)^2$,列於第

七直行下。$f(d'+1)^2$ 相加之和 $[\Sigma f(d'+1)^2]$ 與計算標準差公式(簡法一之公式)中之各項有一定之關係,蓋:

$$\Sigma f(d'+1)^2=\Sigma f[(d')^2+2d'+1]$$
$$=\Sigma f(d')^2+2\Sigma fd'+\Sigma f$$
$$=\Sigma f(d')^2+2\Sigma fd'+N$$

如 $\Sigma f(d'+1)^2=\Sigma f(d')^2+2\Sigma fd'+N$,則計算正確,否則必有錯誤。查表二十六中,$\Sigma f(d')^2$,$\Sigma fd'$,$N$,各項之數值均已算出,代入上式,則得:

$$29=14+2\times2+11$$
$$=29$$

可見計算並無錯誤也。用甲式計算之標準差亦可用此法校核其結果,茲不贅述。

用簡法一計算標準差之方法,可分為下列九步:

1. 將資料製成次數分配表。
2. 選擇適中一組之中點為假定算術平均數(M')。
3. 以組距為單位,計算各組中點對於假定算術平均數之離中差(d')。以假定算術平均數所在組之 d' 為零,其下一組為 -1,其上一組為 $+1$,餘遞推。
4. 以各組之次數乘各組之 d',注意其正負號,列

在 fd' 欄下,然後相加即得 $\Sigma fd'$。

5. 以次數總和(N)除 $\Sigma fd'$,結果為校正數(c)。

6. 以每一組之 d' 乘各該組之 fd',列在 fd'^2 欄下相加而求 fd'^2 之和即得 $\Sigma fd'^2$。

7. 以次數總數(N)除 $\Sigma fd'^2$,得由假定算術平均數求出標準差之平方(以組距為單位)。

8. 從 $\frac{\Sigma fd'^2}{N}$ 中減去校正數之平方 (c^2),然後開方,其結果即等於由眞確算術平均數計算之標準差(以組距為單位)。

9. 以組距乘第八步求得之結果,即為以原來單位表示之標準差。

四、簡法二:簡法一以假定算術平均數為中心計算標準差,免除離中差小數自乘之煩,固已簡捷,然求由假定算術平均數之離中差及其平方之手續,究尙未能免去。應用簡法二,則此項手續概可省去,因標準差可由各變數直接計算也。簡法二之公式如下:

甲、用簡法二由未分組之材料求標準差之公式:

$$\sigma = \sqrt{\frac{\Sigma m^2}{N} - \left(\frac{\Sigma m}{N}\right)^2}$$

或

$$\sigma=\sqrt{\frac{N\Sigma m^2-(\Sigma m)^2}{N^2}}$$

式中：σ 爲標準差之符號

m爲各變數之數值

Σ 爲總加之符號

N爲次數總數

乙、用簡法二由已分組之材料求標準差之公式：

$$\sigma=\sqrt{\frac{\Sigma fm^2}{N}-\left(\frac{\Sigma fm}{N}\right)^2}$$

或

$$\sigma=\sqrt{\frac{N\Sigma fm^2-(\Sigma fm)^2}{N^2}}$$

式中：σ 爲標準差之符號

m爲各組之中點

f 爲各組之次數

Σ爲總加之符號

N爲次數總和

簡法二之甲式係由第一式蛻化而來，乙式則由第二式蛻化而來，其演法相同，茲將後者之演式列下：

證明：

$$\sqrt{\frac{\Sigma fd^2}{N}}=\sqrt{\frac{N\Sigma fm^2-(\Sigma fm)^2}{N^2}}$$

因　$d=m-M$

$$\frac{\Sigma fd^2}{N}=\frac{\Sigma f(m-M)^2}{N}$$

$$=\frac{\Sigma fm^2}{N}-2\frac{\Sigma fmM}{N}+\frac{N(M)^2}{N}$$

但 $$M=\frac{\Sigma fm}{N}$$

$$\frac{\Sigma fd^2}{N}=\frac{\Sigma fm^2}{N}-2\left(\frac{\Sigma fm}{N}\right)^2+\left(\frac{\Sigma fm}{N}\right)^2$$

$$=\frac{\Sigma fm^2}{N}-\left(\frac{\Sigma fm}{N}\right)^2$$

$$\sqrt{\frac{\Sigma fd^2}{N}}=\sqrt{\frac{\Sigma fm^2}{N}-\left(\frac{\Sigma fm}{N}\right)^2}$$

或

$$\sigma=\sqrt{\frac{N\Sigma fm^2-(\Sigma fm)^2}{N^2}}$$

用簡法二計算標準差手續極爲簡便,因離中差可無須求出。如有計算機 (Calculating machine), 則標準差可在計算機上直接計算,尤稱簡便。玆舉一例以示用簡法二乙式計算標準差之方法如下:

表二十七

m	f	fm	m^2	fm^2
3	1	3	9	9
6	2	12	36	72
9	3	27	81	243
12	4	48	144	576

15	1	15	225	225
	N=11	105	495	1,125

$$\sigma=\sqrt{\frac{11\times1125-(105)^2}{11^2}}$$

$$=\sqrt{\frac{12,375-11,025}{121}}$$

$$=\sqrt{\frac{1,350}{121}}$$

$$=\sqrt{11.16}$$

$$=3.3$$

第六節　機差

機差亦名機誤,其最大之功用在於測量以一部材料之結果代表全體結果之錯誤,換言之,即由取樣(Sampling)所生之錯誤。然若次數分配近於常態或偏斜不甚時,亦可用以測量次數分配之離中趨勢。因次數分配如屬常態的,在算術平均數左右加減一機差之距離間,當有全體變數之半。視機差數值之大小,可以測定分配之疏密。機差之符號,吾人常以P. E.(Probable Error)代之,在測量次數分配之離中趨勢時,其意義與四分差相同,但機差之觀念僅限於變數較多且近於常態之分配,如遇偏斜或變數甚少之次數分配,則四分差之意義較爲確切,而機差不適用矣。機差之

數值,可由標準差計算而得,蓋在常態之分配中,機差與標準差常有一定之關係,其公式如下:

$$P.E. = .67449\,\sigma$$

第七節　各種離中趨勢計算法之特點及關係

甲,各種離中趨勢計算法之特點:

一,就計算之手續言,以全距及四分差爲最易;但就精確之程度言,則全距不如四分差,四分差又不如平均差及標準差。機差之計算當 σ 已求得後,亦甚易易。

二,就受極端變數之影響而論,全距最不足恃,因全距之數值,僅依最大與最小兩項而定,一二項之去留,可以大變全距之數值。標準差與平均差之大小與全體變數均有關係,故較穩定,而兩者相較,尤以平均差所受極端變數之影響爲最小。

三,就數學理論言,則平均差不如標準差,因其不問離中差之正負符號,一律相加,殊屬牽强也。

四,就代數方法之處理而論,則以標準差爲最優,蓋標準差之數理的意義明白確切,而四分差則不能用代數方法處理之。

五、就取樣法所生之變動而言,通常以標準差受到之影響爲最小。

六、在統計應用上,以標準差之效用爲最廣。

七、次數分配,如屬完全對稱,則機差亦甚有用,如爲偏斜之分配,則用機差不如用四分差。

八、機差可爲測定數量可靠性之指標,此即其最大之功用也。

乙、各種離中趨勢之關係:

一、全距乃次數曲線圖中底線上之一定距離,全體變數盡在此距離之中。

二、四分差乃第三四分位數與第一四分位數間距離之半,如在第三及第一四分位數間之中點左右各取一定距離等於四分差之數値,則在此距離中當有全體變數之半。

三、在完全對稱或不甚偏斜之次數分配,由算術平均數計算之平均差約等於標準差五分之四。在算術平均數左右各取一定之距離,等於 3.75 倍之平均差,則此距離約可包含全體變數百分之九十。

四、在完全對稱或不甚偏斜之次數分配,從算術平均數左右各取一標準差之距離,則此距離

中約包含全體變數三分之二(在常態曲線中實爲 68.26%),若各取二標準差之距離,則約包含全體變數百分之九十五(在常態曲線中實爲 95.46%);若各取三標準差之距離,則約包含全體變數百分之九十九(在常態曲線中實爲 99.73%);故六倍標準差之距離約等於全距之長。

五、在完全對稱之次數分配,機差之值與四分差相等,約等於 0.6745σ,若從算術平均數左右各取一機差之距離,則此距離中包含全體變數百分之五十;若取四倍機差之距離,則此距離中約包含全體變數百分之九十九。

綜上所述,可得各種離中趨勢之關係如下:

$$\text{A. D.}=.7979\sigma\text{或}\frac{4}{5}\sigma$$

$$\text{A. D.}=1.1843\text{P.E.}$$

$$\text{A. D.}=1.1843\text{Q.D.}$$

$$\text{P. E.}=.6745\sigma\text{或}\frac{2}{3}\sigma$$

$$\text{P. E.}=.8453\text{A.D.}$$

$$\text{P. E.}=\text{Q.D.}$$

$$\sigma=1.2533\text{A.D.}$$

$$\sigma=1.4826\text{P.E.}$$

$\sigma = 1.4826$Q.D.

Q. D. $= .6745\sigma$

Q. D. $= .8453$A.D.

Q. D. $=$ P.E.

第八節　離中係數

上述各法所求得之離中趨勢，悉以原來單位表示，謂之絕對離中趨勢。如吾人所研究者爲一單獨事項離中之狀態，則以絕對離中趨勢表示之自無不可，然若欲比較兩種以上不同事項之離中程度則絕對離中趨勢爲不適用。蓋事物之單位不同者如呎之於斤，磅之於斛，固無比較之可能，卽事物之單位相同，而平均數不同者，亦不能藉絕對離中趨勢以比較其離中之程度。如犬與馬之體重可悉以磅計，但馬之體重遠過於犬，其絕對離中趨勢自亦較犬爲大，吾人殊不能遽謂馬之體重分配，其離中程度遠較犬之體重分配爲高。必也，將絕對離中趨勢化爲相對離中趨勢亦名離中係數，然後比較始有意義。化絕對離中趨勢爲相對離中趨勢之方法，不外以平均數或原來變數之數值除絕對離中趨勢而以百分數表示之。如全距之數值係由最大一項減最小一項而得，則全距之相對離中趨勢可以最大一項與最小一項兩者數值之和

除全距;平均差可由 M, Md, 或 Mo 計算,其相對離中趨勢之求法可以計算時取爲中心之平均數除之;四分差等於第三四分位數及第一四分位數間距離之半,其相對離中趨勢可以第三及第一四分位數兩者數值相加之半除之;至於標準差之計算普通均以算術平均數爲中心,故其相對離中趨勢之求法係以算術平均數除之。玆用公式表示各相對離中趨勢之求法如下:

1.全距之相對離中趨勢:

$$\frac{L-S}{L+S}$$

式中:

L 代表最大一項之數值

S 代表最小一項之數值

2.平均差之相對離中趨勢:

$$\frac{A.D.}{M} \text{或} \frac{A.D.}{Md} \text{或} \frac{A.D.}{Mo}$$

3.四分差之相對離中趨勢:

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

4.標準差之相對離中趨勢:

$$\frac{\sigma}{M}$$

相對離中趨勢在應用上最普通者,莫過於第四式,此

法爲皮爾生氏所倡用,謂之離中係數,恆以V表示之,茲就表二十七之材料舉例如下:

$$M=\frac{105}{11}=9.5$$

$$\sigma=3.3$$

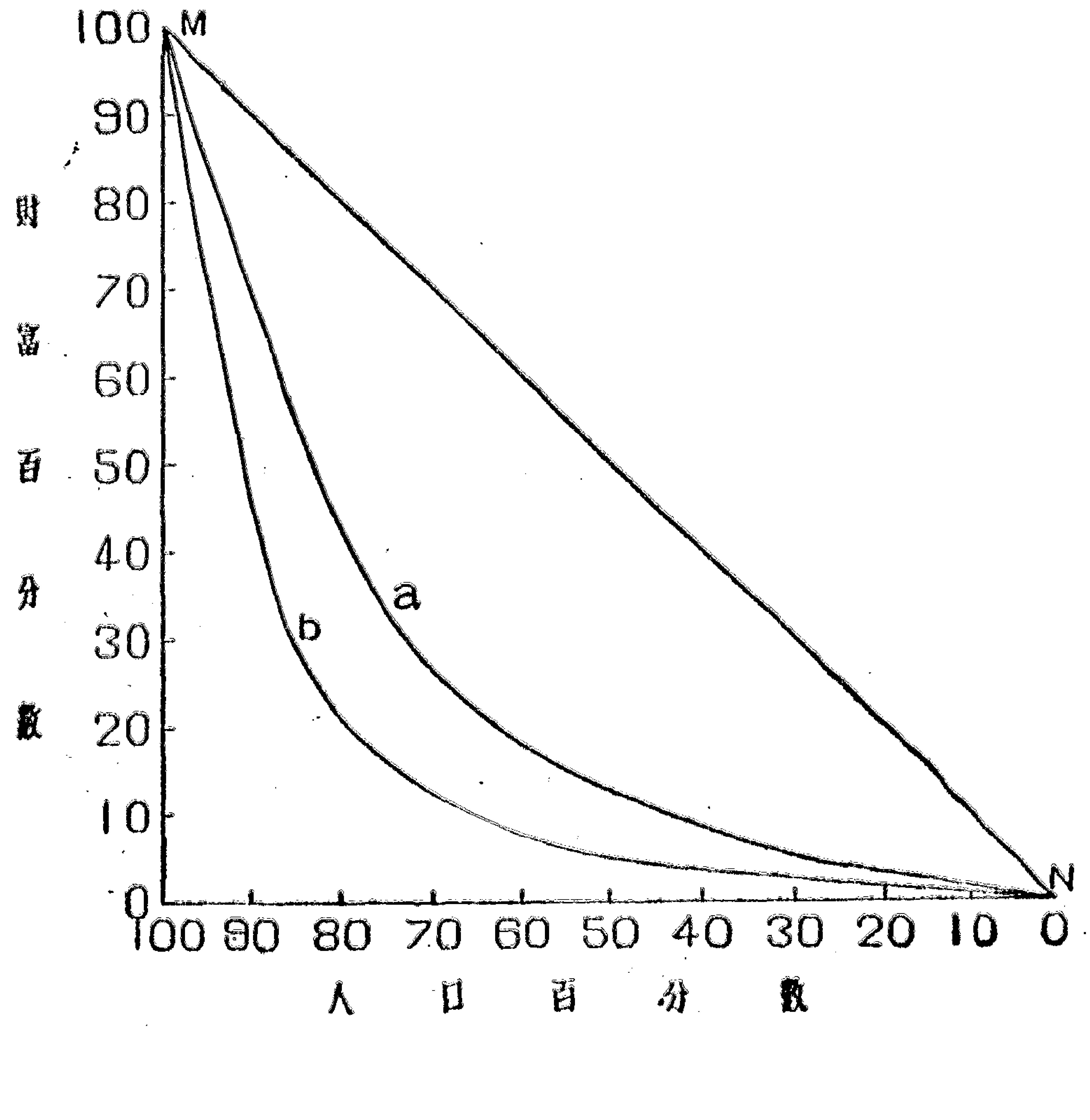
$$\therefore V=\frac{3.3}{9.5}\times 100\%=35\%$$

(乘以100%,係將係數化爲百分數之意)

第九節　羅倫曲線

圖二十二

羅倫曲線



絕對離中趨勢與相對離中趨勢之計算方法已詳言之矣。設吾人所研究者僅爲一種事實,於其分配之離中情形,但求知其大概,無須得一數字的測量,則可用一種曲線表示之。此曲線爲羅倫博士(Dr. Lorenz)所發明,故名羅倫曲線(Lorenz Curve),其最大之功用在於研究財富分配之情形,他若土地,工資,收入等之分配情形亦可用以表現。曲線之繪法,請參閱上圖。

上圖所表示者爲財富之分配情形,以橫軸表示人口百分數,縱軸表示財富百分數。假定一地人民之財富分配極其平均,則百分之十人口應掌有該地財富總額百分之十;百分之五十人口應掌有該地財富總額百分之五十。此種平均分配之情形,可以MN一直線表示之,謂之平均分配線(Line of Equal Distribution)。但實際上財富之分配,絕難平均,此種情形,可以a,b等線表示之。凡表示實際分配之線,離平均分配線愈近則財富之分配愈均勻;反之,離平均分配線愈遠,則財富之分配愈不均勻。此僅就財富之分配而言,如用羅倫曲線研究土地,工資,收入等之分配情形,其理亦同。此項曲線表示一種事實之離中情形,雖不若以數字表示之清晰,但能觀察各部之分配狀況。與絕對離中趨勢或相對離中趨勢之以一簡單數字表示全

部分配狀態者各有所長,可以相輔爲用。

第十節　偏斜度

平均數表示次數分配集中之傾向,離中趨勢則係表示其離中之程度,欲明瞭次數分配之眞相,此二者以外尙須有偏斜度之測定。偏斜度者,中心數値兩旁次數分配對稱或不對稱之狀態也。表示偏斜度或用原來單位,或用抽象之數字,與離中趨勢同。設吾人所研究者僅一種次數分配,則不妨以原來單位表示其偏斜度;設欲比較兩種以上次數分配偏斜之程度,必將用原來單位表示之偏斜度化爲抽象之數字始有比較之可能;所謂偏斜係數(Coefficients of Skewness)是也。求偏斜係數之方法不一,茲擇其中最普通之兩法分述於下:

一、皮爾生氏計算偏斜係數之公式　皮氏之方法係利用算術平均數,中位數,及衆數三者間之距離爲測定偏斜度之標準。蓋次數分配如屬對稱,則三者之數値相等,偏斜度爲零;如次數分配爲不對稱而偏斜度不高時,則三者之數値不等,但有相當之關係。大概算術平均數與衆數之距離約三倍於算術平均數與中位數之距離。算術平均數之數値或大於衆數,或

小於衆數,視偏斜度之爲正爲負而定(參閱第七章第四節),而中位數之數值,則常介於算術平均數與衆數二者之間。皮氏嘗利用此三種平均數之關係倡用一求衆數之經驗的法則[Mo=M−3(M−Md)],玆所述之求偏斜係數之公式,卽根據此法則而來,其式如下:

$$sk=\frac{M-Mo}{\sigma}$$

式中:sk 爲偏斜係數之符號

M 爲算術平均數

Mo爲衆數

σ 爲標準差

玆就二百十七玉蜀黍桿高度之分配舉例如下:

表二十八

(組距呎)	m	f	fm	c.f.	d	fd²
3—4	3.5	3	10.5	3	−3.59	38.67
4—5	4.5	7	31.5	10	−2.59	46.97
5—6	5.5	22	121.0	32	−1.59	55.66
6—7	6.5	60	390.0	92	−.59	21.60
7—8	7.5	85 {16.5 / 68.5	637.5	177	+.41	15.50
8—9	8.5	32	272.0	209	+1.41	63.68

9—10	9.5	8	76.0	217	+2.41	46.48
		217	1538.5			288.56

$$M=\frac{\Sigma fm}{N}$$
$$=\frac{1538.5}{217}$$
$$=7.09$$

$$Mo=L+\frac{f_2}{f_2+f_1}\times 1=7+\frac{32}{32+60}\times 1=7.35$$

$$\sigma=\sqrt{\frac{\Sigma fd^2}{N}}$$
$$=\sqrt{\frac{288.56}{217}}$$
$$=\sqrt{1.33}$$
$$=1.15$$

$$sk=\frac{M-Mo}{\sigma}$$
$$=\frac{7.09-7.35}{1.15}=\frac{-.26}{1.15}=-.23$$

$$Md=L+\frac{i}{f}\times c$$
$$=7+\frac{16.5}{85}\times 1=7.19$$

$$sk=\frac{3(M-Md)}{\sigma}=\frac{3(7.09-7.19)}{1.15}$$
$$=\frac{-.30}{1.15}=-.26$$

如玉蜀黍桿高度之分配爲對稱,則算術平均數之數值等於衆數,M—Mo=0,卽分配無偏斜之意。今算術平均數小於衆數,M—Mo=—.26,是此二百十七玉蜀黍

桿高度之分配呈負的偏斜度(參閱第七章第四節),以 σ 除之者,所以化此項用原來單位表示之偏斜度爲偏斜係數也。理論衆數之數值每不易決定,故皮氏之公式可以改爲下式。

$$sk = \frac{3(M-Md)}{\sigma}$$

式中: sk 爲偏斜係數之符號

M 爲算術平均數

Md 爲中位數

σ 爲標準差

蓋 $M-Mo=3(M-Md)$ 也。按此式計算之偏斜係數爲 $-.26$,詳見表二十八。

二、鮑來 (A. L. Bowley) 氏計算偏斜係數之公式

鮑氏之方法,係利用第一及第三四分位數與中位數之關係爲測定偏斜度之標準。蓋次數分配如屬對稱,則中位數與第一四分位數間之距離,應等於第三四分位數與中位數間之距離,否則此二距離必不相等。設以 q_2 表示第三四分位數與中位數間之距離,q_1 表示中位數與第一四分位數間之距離,則當次數分配爲對稱時,q_2 恰當於 q_1, $q_2-q_1=0$; 如次數分配呈正的偏斜,則 $q_2>q_1$; 如呈負的偏斜,則

$q_2<q_1$，鮑氏之公式即利用q_2與q_1之數值爲測定偏斜度之大小與正負之標準,其公式如下:

$$sk=\frac{q_2-q_1}{q_2+q_1}$$

式中:sk爲偏斜係數之符號

q_2爲第三四分位數與中位數間之距離

q_1爲中位數與第一四分位數間之距離

玆仍就二百十七玉蜀黍稈高度之分配,用鮑氏之公式測定其偏斜度如下:

表二十九

(組距呎)	f	c.f.
3—4	3	3
4—5	7	10
5—6	22	32
6—7	60	92
7—8	85	177
8—9	32	209
9—10	8	217
	217	

$$Q_1 = 6 + \frac{22.5}{60} \times 1 = 6.37$$

$$Md = 7 + \frac{16.5}{85} \times 1 = 7.19$$

$$Q_3 = 7 + \frac{70.75}{85} \times 1 = 7.83$$

$$q_1 = Md - Q_1 = 7.19 - 6.37$$

$$= .82$$

$$q_2 = Q_3 - Md = 7.83 - 7.19$$

$$= .64$$

$$sk = \frac{.64 - .82}{.64 + .82} = -\frac{.18}{1.46}$$

$$= -.12$$

本例玉蜀黍桿之分配呈負的偏斜,故 q_1 大於 q_2,所以更除以 q_2+q_1 者,化原來單位表示之偏斜度爲偏斜係數也。用此式計算偏斜係數,其結果不出 0 與 ±1 之範圍。據鮑氏之意見,偏斜係數如爲 .1,則次數分配已呈相當之偏斜,如達 .3,則偏斜顯然。惟當注意者,吾人如欲比較兩種次數分配之偏斜度,必用同一公式求其偏斜係數。鮑氏之公式與皮氏之公式所根據之理論不同,故其結果無比較之可能也。

第九章　相關

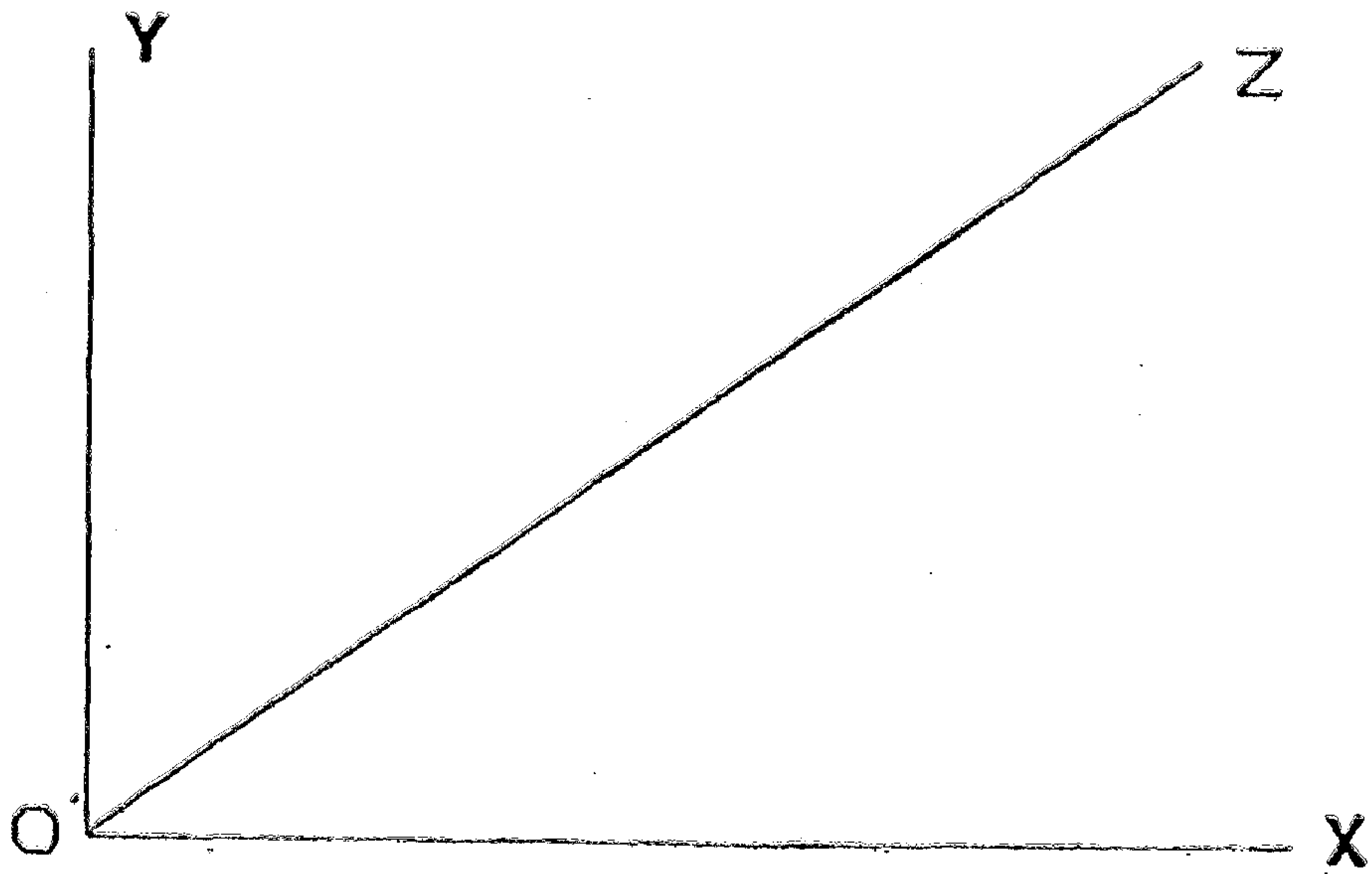
第一節　相關之意義及種類

七八兩章所述為平均數,離中趨勢及偏斜度之問題,吾人可用以測度一次數分配集中之程度,離中之趨勢與夫偏斜之狀態。設吾人所研究者為兩種事實間相互之關係,則非求其相關 (Correlation) 不可,如麥之產額與麥價,米之收成與雨量,其間均有一定之關係。惟所謂相關,非謂任何兩種事實均有關係也,必也,當此一事實有變動時,另一事實亦有所感應而生變動,則此兩種事實始得謂之相關。相關之種類,就方向言有正相關(Positive Correlation)與負相關 (Negative Correlation)二種。當某種事實增加或減少時,他種事實因受感應亦隨之增加或減少,如方向相同,則為正相關;如方向相反,則為負相關。前者之例如批發物價與零售物價,批發物價漲,則零售物價亦隨之而漲;批發物價跌,則零售物價亦隨之而跌。後者之例如物品之產量與價格,產量增多,則價格低落;產量減少,則價格增高。完全的正相關與負相關可以圖二十三,二十四表示之,至若兩種事實中某種事實發生變動而他種事實毫不發生影響者,則此兩種事實謂之不相關或

零相關(Zero Correlation),可以圖二十五表示之。

圖二十三

完全正相關



X軸表示甲種事實

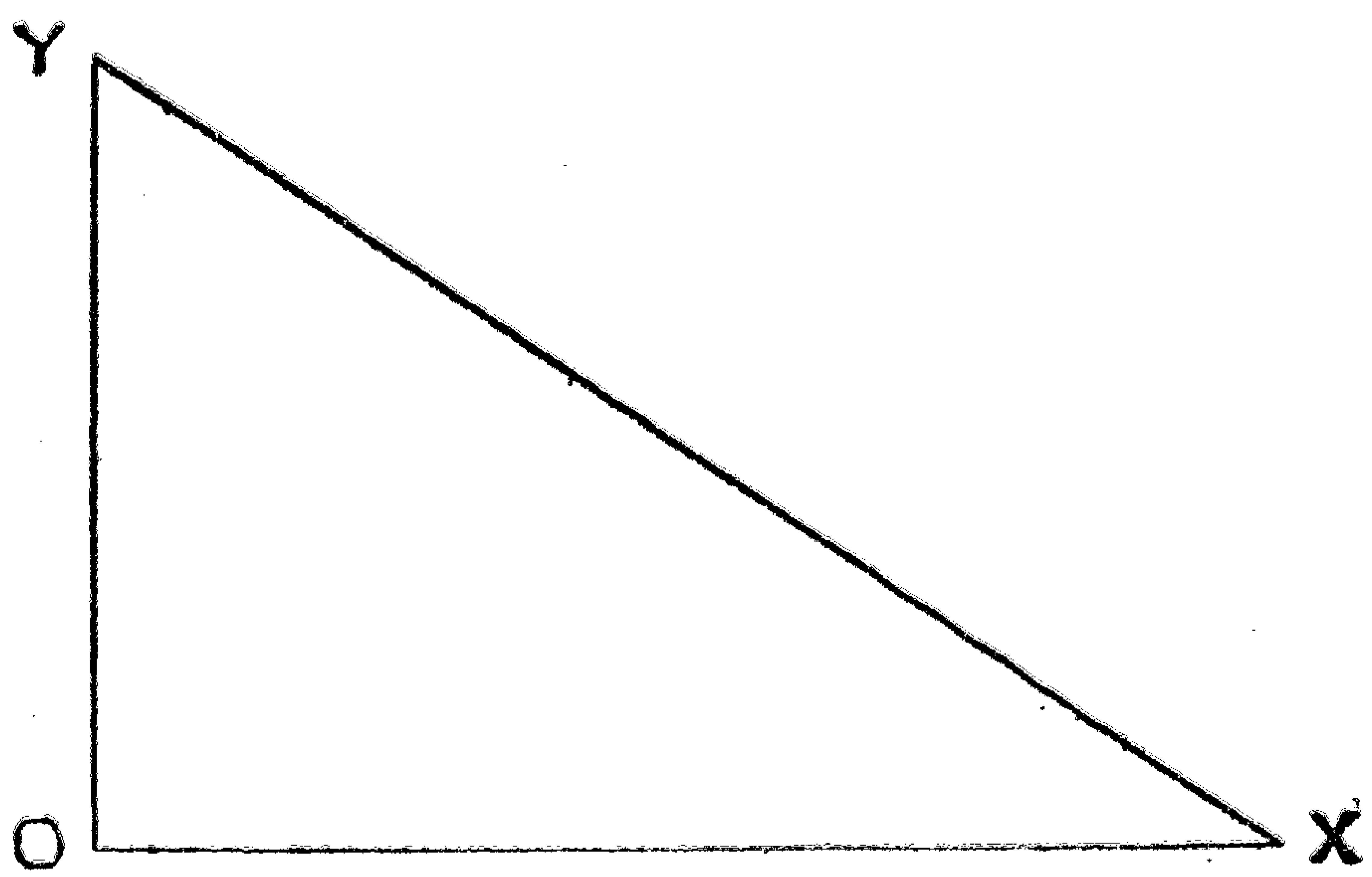
Y軸表示乙種事實

OZ爲甲乙兩種事實之完全正相關線

相關之程度如屬完全正的(Perfect Positive Correlation),則結果應爲 +1;如屬完全負的(Perfect Negative Correlation),則結果應爲 −1;如毫不相關,則結果爲零。事實上相關之程度,無論爲正爲負,鮮能達完全之地位,故相關之結果恆介於 +1 與 −1 之間。至於毫不相關

圖二十四

完全負相關



X軸表示甲種事實

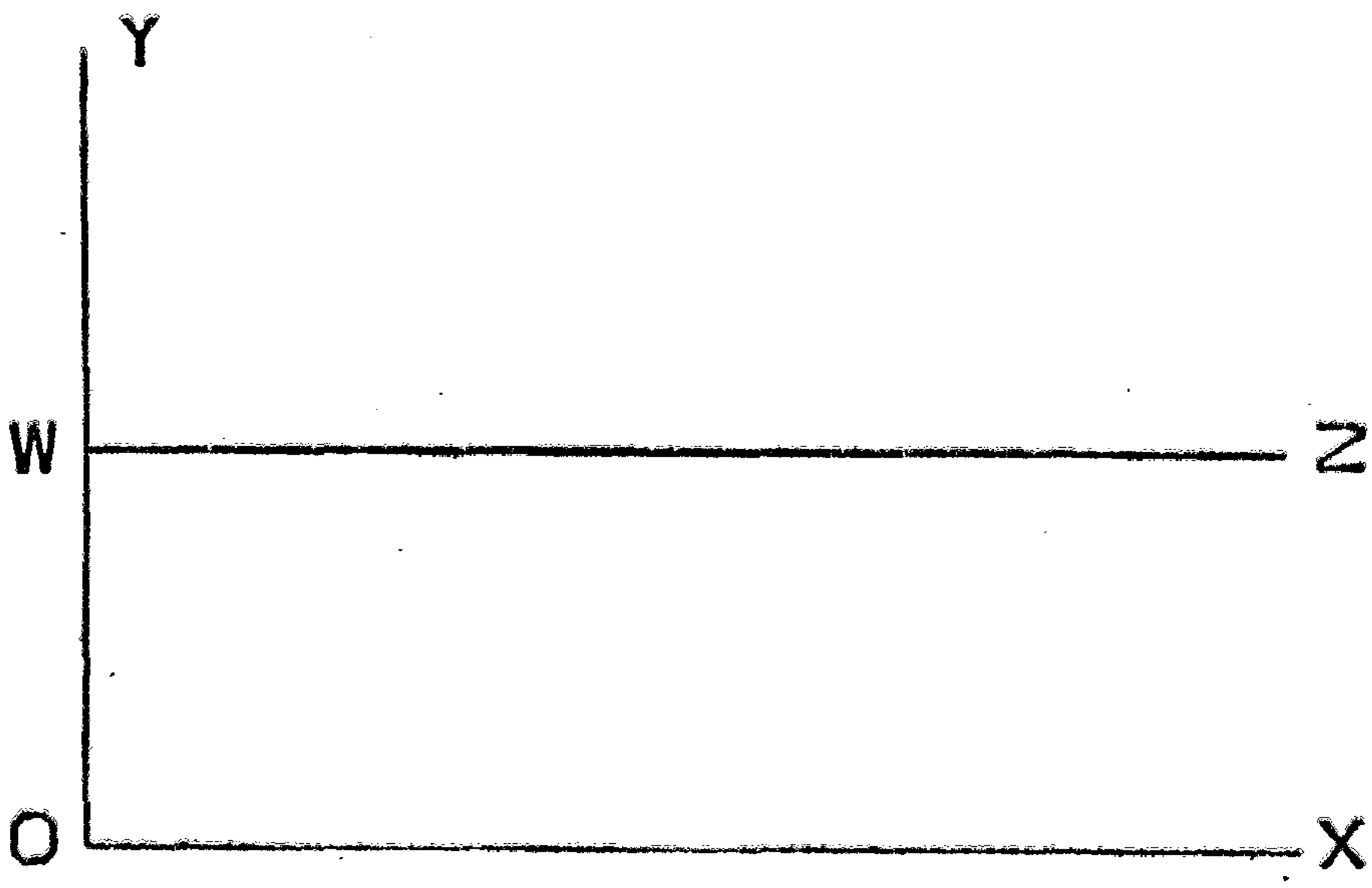
Y軸表示乙種事實

YX爲甲乙兩種事實之完全負相關線

之事實亦不多覯,是在統計者判別其關係爲偶然的抑互爲因果的以確定其相關之有無耳。計算相關之方法頗多,茲擇其中應用較廣者分(一)由量的統計求相關及(二)由質的統計求相關二類臚述於後。所謂量的統計者,乃屬於數量的,可以精密測量之次數數列,計算此類材料相關之方法本章所介紹者爲乘積率

圖二十五

零相關



X軸表示甲種事實

Y軸表示乙種事實

WZ為甲乙兩種事實之零相關線

法(Product Moment Method)及相關率 (Correlation Ratio)二種;所謂質的統計者,乃屬於品質的,不可爲精密測量之次數數列,計算此類材料相關之方法本章所介紹者爲均方相關法(Method of Mean Square Contingency),請於以下各節論之。

第二節　乘積率法

乘積率法爲皮爾生教授所倡用,乃求變數的統計中直線相關最佳之方法。所謂直線相關(Linear Correlation)者,非必如圖三十三及三十四所示之謹嚴,凡兩種變數間相互之關係用直線表示較用曲線表示爲貼切者,悉爲直線相關。用乘積率法計算所得之結果,謂之相關係數,卽所以表示相關程度之高低者。求相關係數之手續,可分列相關表(Correlation Table)與不列相關表二種。當變數不多時,以不列相關表較爲簡捷;若變數甚多,則以列相關表爲宜,蓋不特計算之手續較省,且甚便於繪迴歸線(Lines of Regression)也。

一、不列相關表計算相關係數之方法　乘積率法之公式爲:

$$r = \frac{\Sigma xy}{N\sigma_x\sigma_y}$$

式中: r爲相關係數之符號

x爲X變數中各項與其算術平均數之離中差

y爲Y變數中各項與其算術平均數之離中差

N爲X Y兩種變數之次數

σ_x爲X變數之標準差

σ_y 爲 Y 變數之標準差

按此式計算相關係數,第一步先將兩種變數中之自變數(X)順次排列,另一變數(Y)中各項之位置則隨其 X 變數中相當之一項(即同在一對之變數)而定。第二步求出 X 變數與 Y 變數之算術平均數,並計算 X 變數與 Y 變數中各項對於其算術平均數之離中差,此項離中差即公式中之 x 及 y。第三步將兩種變數中各對變數之離中差相乘,注意其正負符號,得各項乘積(xy),各項乘積之總和(Σxy)爲正爲負,即爲判別相關爲正爲負之標準,蓋以 $N\sigma_x\sigma_y$ 除 Σxy,即得相關係數也。計算標準差之方法已詳第八章第四節,無須贅述,茲就二十對夫妻之年齡用此法計算其相關係數如下:

$$r=\frac{\Sigma xy}{N\sigma_x\sigma_y}=\frac{287}{20\times4.02\times4.17}$$

$$=\frac{287}{335.27}=+.856$$

表三十

用乘積率法不列相關表計算二十對夫妻年齡之相關係數

(見 King, W. I. : Elements of Statistical Method, P. 209)

X			Y			
夫之年齡	x 離中差	x^2	妻之年齡	y 離中差	y^2	xy
22	—8	64	18	—8	64	+64
24	—6	36	20	—6	36	+36
26	—4	16	20	—6	36	+24
26	—4	16	24	—2	4	+ 8
27	—3	9	22	—4	16	+12
27	—3	9	24	—2	4	+6
28	—2	4	27	+1	1	—2
28	—2	4	24	—2	4	+4
29	—1	1	21	—5	25	+5
30	0	0	25	—1	1	0
30	0	0	29	+3	9	0
30	0	0	32	+6	36	0
31	+1	1	27	+1	1	+1
32	+2	4	27	+1	1	+2
33	+3	9	30	+4	16	+12
34	+4	16	27	+1	1	+ 4
35	+5	25	30	+4	16	+20
35	+5	25	31	+5	25	+25
36	+6	36	30	+4	16	+24
37	+7	49	32	+6	36	+42
M=30	$\Sigma x^2=324$		M=26	$\Sigma y^2=348$		$\Sigma xy=+287$

上表二十對變數離中差之乘積,祇有一負數,餘均爲正,故本例夫妻年齡之相關爲正,且其相關之程度甚高也。兩種變數之標準差係按 $\sigma=\sqrt{\frac{\Sigma x^2}{N}}$ 公式計算, $\sigma_x=4.02$, $\sigma_y=4.17$。在不列相關表計算相關係數之公式可化爲

$$r=\frac{\Sigma xy}{\sqrt{\Sigma x^2\,\Sigma y^2}}$$

使計算手續較為簡便,蓋

$$N\sigma_x\sigma_y=N\sqrt{\frac{\Sigma x^2}{N}}\sqrt{\frac{\Sigma y^2}{N}}$$

$$=N\frac{1}{N}\sqrt{\Sigma x^2\,\Sigma y^2}=\sqrt{\Sigma x^2\,\Sigma y^2}$$

計算$\sqrt{\Sigma x^2\,\Sigma y^2}$遠較$N\sigma_x\,\sigma_y$為簡捷也.

茲演算如下:

$$r=\frac{\Sigma xy}{\sqrt{\Sigma x^2\,\Sigma y^2}}$$

$$=\frac{287}{\sqrt{324\times348}}$$

$$=\frac{287}{\sqrt{112,752}}$$

$$=\frac{287}{335.8}$$

$$=+.855$$

(小數第三位略有出入,乃計算時小數舍入之關係)

二、列相關表計算相關係數之方法　用此法計算相關之手續可分兩步:一、製相關表,二、計算相關係數,茲分述於次:

甲、製相關表　所謂相關表,係將兩種事實,一沿表之左邊由下至上,一沿表之上方或下

表三十一

5317對夫妻年齡相關表

妻之年齡

夫之年齡

m	f	d'	fd'	fd'²	17.5	22.5	27.5	32.5	37.5	42.5	47.5	52.5	57.5	62.5	67.5	72.5	77.5	82.5	87.5		
	f				23	414	808	854	781	669	550	437	317	226	134	68	27	8	1	5317	
		d'			−6	−5	−4	−3	−2	−1	0	+1	+2	+3	+4	+5	+6	+7	+8		
			fd'		−138	−2070	−3232	−2562	−1562	−669	−10233 +2851	+437	+634	+678	+536	+340	+162	+56	+8		
				fd'²	828	10350	12928	7686	3124	669		437	1268	2034	2144	1700	972	392	64	44596	
17.5	4	−6	−24	144	+36 (2) +72	+30 (2) +60															20.00
22.5	240	−5	−1200	6000	+30 (16) +480	+25 (173) +4325	+20 (46) +920	+15 (4) +60	+10 (1) +10												23.35
27.5	688	−4	−2752	11008	+24 (4) +96	+20 (185) +3700	+16 (402) +6432	+12 (84) +1008	+8 (10) +80	+4 (2) +8	(1) 0										26.93
32.5	817	−3	−2451	7353	+18 (1) +18	+15 (41) +615	+12 (265) +3180	+9 (411) +3699	+6 (84) +504	+3 (12) +36	(2) 0	−3 (1) −3									31.08
37.5	793	−2	−1586	3172		+10 (9) +90	+8 (69) +552	+6 (251) +1506	+4 (369) +1476	+2 (80) +160	(12) 0	−2 (2) −4	−4 (1) −4								35.60
42.5	700	−1	−700	700		+5 (3) +15	+4 (17) +68	+3 (71) +213	+2 (219) +438	+1 (309) +309	(66) 0	−1 (12) −12	−2 (2) −4	−3 (1) −3							40.04
47.5	595	0	−8713 +3730			(1) 0	(6) 0	(20) 0	(66) 0	(178) 0	(252) 0	(59) 0	(10) 0	(2) 0	(1) 0						44.90
52.5	483	+1	+483	483			−4 (2) −8	−3 (8) −24	−2 (19) −38	−1 (57) −57	(146) 0	+1 (195) +195	+2 (44) +88	+3 (10) +30	+4 (2) +8						49.51
57.5	369	+2	+738	1476			−8 (1) −8	−6 (3) −18	−4 (8) −32	−2 (18) −36	(46) 0	+2 (110) +220	+4 (141) +564	+6 (35) +210	+8 (6) +48	+10 (1) +10					53.99
62.5	277	+3	+831	2493				−9 (1) −9	−6 (3) −18	−3 (8) −24	(16) 0	+3 (39) +117	+6 (81) +486	+9 (101) +909	+12 (23) +276	+15 (4) +60	+18 (1) +18				58.42
67.5	175	+4	+700	2800				−12 (1) −12	−8 (1) −8	−4 (3) −12	(6) 0	+4 (11) +44	+8 (26) +208	+12 (53) +636	+16 (58) +928	+20 (13) +260	+24 (2) +48	+28 (1) +28			62.64
72.5	104	+5	+520	2600					−10 (1) −10	−5 (1) −5	(2) 0	+5 (5) +25	+10 (8) +80	+15 (18) +270	+20 (31) +620	+25 (31) +775	+30 (6) +180	+35 (1) +35			66.44
77.5	50	+6	+300	1800						−6 (1) −6	(1) 0	+6 (2) +12	+12 (3) +36	+18 (5) +90	+24 (10) +240	+30 (14) +420	+36 (12) +432	+42 (2) +84			69.30
82.5	18	+7	+126	882								+7 (1) +7	+14 (1) +14	+21 (1) +21	+28 (2) +56	+35 (4) +140	+42 (5) +210	+49 (3) +147	+56 (1) +56		73.33
87.5	4	+8	+32	256											+32 (1) +32	+40 (1) +40	+48 (1) +48	+56 (1) +56			75.00
	5317			41167																	
					23.37	26.27	30.27	34.96	39.81	44.71	49.55	54.38	59.08	63.45	68.10	72.57	76.39	78.75	82.50		

統計概論一五〇頁後揷頁

方由左至右,按組排列,以顯示其間相互關係之表也。編製之手續與次數分配表略同,惟次數分配表中僅列一種數列,此則並列兩種數列耳。玆將編製之步驟列下:

a. 求兩種次數數列之全距,並決定其組距及組限,如表三十一之例,夫妻年齡之全距均爲七十五,以 5 爲組距,可各分爲十五組。

b. 以一種數列中之變數爲X,另一種數列中之變數爲Y,將Y變數各組之組限自小而大沿表之左端由下至上註明;X變數各組之組限自小而大沿表之下端或上端由左至右註明。

c. 將兩種變數中各對變數用記號法記入相當之方格中,即成一散佈圖(Scatter Diagram),由此圖可以看出相關程度之高低及相關之爲正爲負。

d. 將散佈圖中各方格之記號改爲數字,即爲相關表(參閱表三十一)。

乙、計算相關係數 相關表製成後即可從事於相關係數之計算,其公式與不列相關表

時同,但既列相關表,實際計算時,可以

$$r=\frac{\Sigma x'y'-NCxCy}{N\sigma_x\sigma_y}$$ 公式代

$$r=\frac{\Sigma xy}{N\sigma_x\sigma_y}$$ 公式,以省手續。

茲證明此簡式等於原來之公式如下:

設 x 爲 X 變數中各項與其平均數之離中差

y 爲 Y 變數中各項與其平均數之離中差

x' 爲 X 變數中各項與其假定平均數之離中差

y' 爲 Y 變數中各項與其假定平均數之離中差

Cx爲 x' 之校正數(即 $x'-x$)

Cy爲 y' 之校正數(即 $y'-y$)

則 $x'=x+Cx$

$y'=y+Cy$

$\Sigma x'y'=\Sigma(x+Cx)(y+Cy)$

$=\Sigma xy+Cx\Sigma y+Cy\Sigma x+NCxCy$

但 $\Sigma y=0$

$\Sigma x=0$

故 $\Sigma x'y'=\Sigma xy+NCxCy$

$\Sigma xy=\Sigma x'y'-NCxCy$

$$\frac{\Sigma xy}{N\sigma_x\sigma_y}=\frac{\Sigma x'y'-NCxCy}{N\sigma_x\sigma_y} \text{或} \frac{\frac{\Sigma x'y'}{N}-CxCy}{\sigma_x\sigma_y}$$

按簡式計算之各步,詳見表三十一,該表之材料爲根據一九〇一年英格蘭(England)與威爾斯(Wales)人口調查所得5,317,520對夫妻之年齡,因次數太多,改以一千爲單位,計算之結果爲+.915,可見夫妻年齡相關之程度甚高也。

$$r=\frac{\frac{\Sigma x'y'}{N}-CxCy}{\sigma_x\sigma_y}$$

今 $\Sigma x'y'=39,647-355=39,292$

$N=5,317$

$$Cy=\frac{2,851-10,233}{5,317}=-1.388$$

$$Cx=\frac{3,730-8,713}{5,317}=-\ .937$$

$$CxCy=(-1.388)(-.937)=1.3$$

$$\sigma_y=\sqrt{\frac{44,596}{5,317}-(-1.388)^2}=2.54$$

$$\sigma_x=\sqrt{\frac{41,167}{5,317}-(-\ .937)^2}=2.62$$

$$\therefore r=\frac{\frac{39,292}{5,317}-1.3}{2.54\times2.62}=+.915$$

(注意: $x',y',Cx,Cy,\sigma_x,\sigma_y$ 之計算,可悉以組距爲單位,無庸化爲原來單位,以省手續。)

玆將用簡式計算相關係數之步驟列下:

a. 將各橫行中各方格之次數相加列在 fy 一行,各縱行中各方格之次數相加列在 fx 一行;再將 fy 與 fx 行中次數分別相加,得次數總數。

b. 擇兩種變數中適中一組之中點爲各該變數之假定算術平均數,如本例夫妻年齡之假定平均數均爲 47.5。

c. 以組距爲單位,計算各組中點與各該變數假定算術平均數之離中差,分別列於 d' 行中。

d. 以兩種變數中各組之次數乘各該組之離中差,得 fx' 及 fy'。

e. 將所有 fx' 與 fy' 之積用代數法相加(意即注意其正負符號),以求其總和 $\Sigma fx'$ 及 $\Sigma fy'$。

f. 以次數總數(N)分別除 $\Sigma fx'$ 及 $\Sigma fy'$,求校正數 Cx 及 Cy。

g. 以校正數自乘得 C^2x 及 C^2y。

h. 以各組之 x' 及 y' 乘各該組之 fx' 及 fy' 求 fx'^2 及 fy'^2,然後分別相加得 $\Sigma fx'^2$ 及 $\Sigma fy'^2$。

i. 用計算標準差簡法一之公式

$$\sigma = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2}$$

求 X 變數及 Y 變數之標準差,是爲以組距爲

單位之σ_x及σ_y。

j. 以組距爲單位,計算$\Sigma x'y'$,其法有二:

1. 先求直行中各方格內次數乘各該組y'之積,然後將每一直行各方格之fy'相加而乘以該直行之x',是爲該直行之$\Sigma x'y'$。將各直行之$\Sigma x'y'$一一求出,其和卽爲全體變數之$\Sigma x'y'$。如表三十一中第一直行各方格內之次數爲2、2,其y'爲-6、-5,$(-6\times2)+(-5\times2)=-22$,更乘以$x'(-6)$,得132 $(-22\times-6=132)$,卽爲該直行之$\Sigma x'y'$。其餘各直行之$\Sigma x'y'$求法相同,茲計算如下:

第一直行:$[(-6\times2)+(-5\times2)]\times-6 \quad =132$

第二直行:$[(-6\times16)+(-5\times173)+(-4\times46)+(-3\times4)+(-2\times1)]\times-5$

$=5,795$

第三直行:$[(-6\times4)+(-5\times185)+(-4\times402)+(-3\times84)+(-2\times10)+(-1\times2)]\times-4 \quad =11,324$

第四直行:$[(-6\times1)+(-5\times41)+(-4\times265)+(-3\times411)+(-2\times84)+(-1\times12)+(1\times1)]\times-3 \quad =8,049$

第五直行：$[(-5\times9)+(-4\times69)+(-3\times251)+(-2\times369)+(-1\times80)+(1\times2)+(2\times1)]\times-2=3,776$

第六直行：$[(-5\times3)+(-4\times17)+(-3\times71)+(-2\times219)+(-1\times309)+(1\times12)+(2\times2)+(3\times1)]\times-1=1,024$

第七直行：

第八直行：$[(-4\times2)+(-3\times8)+(-2\times19)+(-1\times57)+(1\times195)+(2\times44)+(3\times10)+(4\times2)]\times1=194$

第九直行：$[(-4\times1)+(-3\times3)+(-2\times8)+(-1\times18)+(1\times110)+(2\times141)+(3\times35)+(4\times6)+(5\times1)]\times2=958$

第十直行：$[(-3\times1)+(-2\times3)+(-1\times8)+(1\times39)+(2\times81)+(3\times101)+(4\times23)+(5\times4)+(6\times1)]\times3=1,815$

第十一直行：$[(-3\times1)+(-2\times1)+(-1\times3)+(1\times11)+(2\times26)+(3\times53)+(4\times58)+(5\times13)+(6\times2)+(7\times1)]\times4=2,120$

第十二直行：$[(-2\times1)+(-1\times1)+(1\times5)+(2\times8)+(3\times18)+(4\times31)+(5\times31)+(6\times6)+(7\times1)]\times5$ $=1,970$

第十三直行：$[(-1\times1)+(1\times2)+(2\times3)+(3\times5)+(4\times10)+(5\times14)+(6\times12)+(7\times2)]\times6$ $=1,308$

第十四直行：$[(1\times1)+(2\times1)+(3\times1)+(4\times2)+(5\times4)+(6\times5)+(7\times3)+(8\times1)]\times7$ $=651$

第十五直行：$[(4\times1)+(5\times1)+(6\times1)+(7\times1)]\times8$ $=176$

$\Sigma x'y'=39,292$

2. 先求每方格中次數乘其所在直行及橫行 x' 及 y'；得各方格之 $fx'y'$，書於各方格中次數之下。如表三十一中妻之年齡 17.5 及夫之年齡 17.5 兩組相交一方格中次數爲 2，x' 及 y' 均爲 -6，$(-6)\times(-6)=36$，書於該方格中次數之上，更以次數乘之得 72 $(36\times2=72)$，書於次數之下。將各方格中之 $fx'y'$ 相加，注意其正負符號，即得全體變數之 $\Sigma x'y'$。

本例$\Sigma x'y'=39,292$。

k. 以$N\sigma_x\sigma_y$之積除$\Sigma x'y'-NCxCy$或以$\sigma_x\sigma_y$之積除$\left(\frac{\Sigma x'y'}{N}-CxCy\right)$,所得之商卽爲相關係數(r)。

相關係數爲相關程度高低之數字的測量,如相關係數小於.3,則有無關係尙不能必;如大於.5,則必爲相關無疑;若相關係數達.8以上,則關係異常密切。以上所述係就相關係數判別相關程度之高低,此外更可藉相關係數之機差爲判別之標準,蓋相關係數如小於其機差,則無相關之可言,如大於其機差六倍,則關係確立無疑。計算機差之公式如下:

$$P.E.=\frac{.6745(1-r^2)}{\sqrt{N}}$$

式中:N爲兩種變數之次數,N愈大,則機差愈小,N愈小,則機差愈大。

例如5,317對夫妻年齡之相關係數爲+.915,其機差爲.0015,$\frac{.6745(1-.915^2)}{\sqrt{5,317}}$ 僅及相關係數千分之1.6,可見夫妻年齡相關之程度甚高也。相關係數普通應與其機差同時表出,書作r±P.E.,意卽眞正之相關係數在此範圍內之機會爲百分之五十,在r±4P.E.範圍內之機會爲百分之九十九。

相關係數所以表明兩種變數關係之大小及正負,但不能表示兩種變數中X數某項有一單位變動

時,影響於Y數中相當之一項幾何,或Y數某項有一單位變動時,影響於X數中相當之一項幾何。蓋X與Y兩數列之標準差苟非完全相等,則X數影響Y數或Y數影響X數者其大小常不一致。欲明瞭二者相互影響之大小,必求其迴歸係數(Coefficients of Regression)。迴歸係數有二:一爲Y變數迴歸X變數之係數;一爲X變數迴歸Y變數之係數。前者以b_1,或b_{yx}代之,後者以b_2,或b_{xy}代之,茲將求此二迴歸係數之公式列下:

1. Y迴歸X之係數:

$$b_{yx}=r\frac{\sigma_y}{\sigma_x}$$

2. X迴歸Y之係數:

$$b_{xy}=r\frac{\sigma_x}{\sigma_y}$$

就上例計算5.317對夫妻年齡相關之結果,

$$b_{yx}=.915\times\frac{2.54}{2.62}$$

$$=.915\times.97=.89$$

$$b_{xy}=.915\times\frac{2.62}{2.54}$$

$$=.915\times1.03=.94$$

(如σ以組距爲單位而X數與Y數之組距不同,應先將兩標準差各乘以組距)

意卽夫之年齡有一單位變動時妻之年齡應有.89單位之變動;妻之年齡有一單位變動時,夫之年齡應有.94單位之變動。惟吾人有當注意者,兩迴歸係數乘積之平方根等於相關係數($b_{yx} \cdot b_{xy} = r\frac{\sigma_y}{\sigma_x} \cdot r\frac{\sigma_x}{\sigma_y} = r^2$),故知$b_{yx}$爲.89,$b_{xy}$爲.94,亦可求得相關係數爲.915($r = \sqrt{.89 \times .94} = .915$)。迴歸係數旣經求出,則吾人可據迴歸方程式以繪迴歸線。所謂迴歸方程式者乃表示Y變數迴歸X變數或X變數迴歸Y變數之方程式也。

1. Y迴歸X之方程式:

$$y = r\frac{\sigma_y}{\sigma_x}x$$
$$= .89x$$

2. X迴歸Y之方程式:

$$x = r\frac{\sigma_x}{\sigma_y}y$$
$$= .94y$$

式中: x 爲X變數中各項與其算術平均數之離中差

y 爲Y變數中各項與其算術平均數之離中差

惟就此二方程式,吾人所知者僅爲某變數有一單位變動時影響於他變數者若干,尙不能據某變數中之一項推知他變數中相當一項之數値,如欲作此

項推測,須將上列二方程式改爲下式:

1. 推測Y迴歸X之方程式:

$$Y-\bar{Y}=.89(X-\bar{X})$$

2. 推測X迴歸Y之方程式:

$$X-\bar{X}=.94(Y-\bar{Y})$$

式中:X爲X變數中各項之數值

Y爲Y變數中各項之數值

$\bar{X}$爲X數之算術平均數

$\bar{Y}$爲Y數之算術平均數

依上式推測,須先求出X變數與Y變數之算術平均數。按表三十一計算5,317對夫妻年齡之相關係數時,夫妻年齡之假定算術平均數均爲47.5,但$Cx=-.937\times5=4.685$,$Cy=-1.388\times5=-6.940$,故夫之年齡眞確算術平均數應爲$47.5+(-4.7)=42.8$,妻之年齡眞確算術平均數應爲$47.5+(-6.9)=40.6$。夫妻年齡之算術平均數既經求得,則知一夫之年齡,即可推知其妻之年齡;知一妻之年齡,即可推知其夫之年齡。茲舉數例於次:

1. Y迴歸X:

設夫之年齡爲40,

則妻之年齡應爲38.11。

$$Y-40.6=.89(40-42.8)$$

$$Y=40.6-2.49$$

$$=38.11$$

設夫之年齡爲39,

則妻之年齡應爲37.22。

$$Y-40.6=.89(39-42.8)$$

$$Y=40.6-3.38$$

$$=37.22$$

2. X 迴歸 Y:

設妻之年齡爲40

則夫之年齡應爲42.24

$$X-42.8=.94(40-40.6)$$

$$X=42.8-.56$$

$$=42.24$$

設妻之年齡爲30

則夫之年齡應爲32.84

$$X-42.8=.94(30-40.6)$$

$$X=42.8-9.96$$

$$=32.84$$

按上式推算,可知夫之年齡有一年變動時,妻之年齡有.89年之變動;妻之年齡有十年變動時,夫之年

齡有9.4年之變動。茲列表如下:

表三十二

甲,Y 迴歸 X

X	Y
80	73.11
70	64.81
60	55.91
50	47.01
40	38.11
30	29.21

乙,X 迴歸 Y

Y	X
70	70.44
60	61.04
50	51.64
40	42.24
30	32.84
20	23.44

註:甲表係據夫之年齡推測妻之年齡。

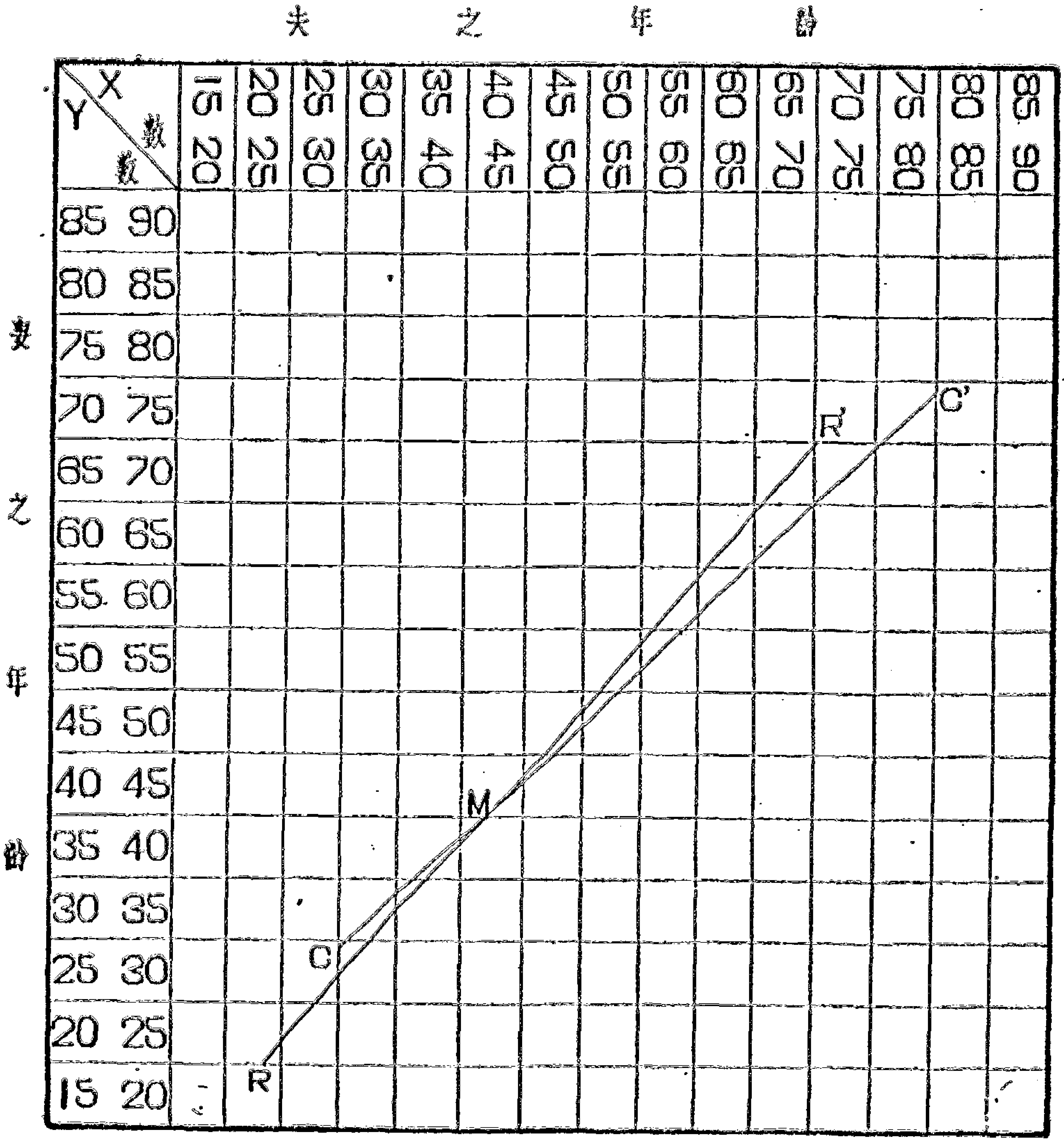
乙表係據妻之年齡推測夫之年齡。

就甲表中所列各對之數值,任取二對在相關表上以兩點表示之,更將此兩點連成一直線,則其餘各對數值均應落在此直線上,此直線爲CC'線,乃表示Y迴歸X之直線。就乙表中所列各對之數值,任取二對在相關表上以兩點表示之,更將此兩點連成一直線,則其餘各對數值均應落在此直線上,此直線爲RR'線,乃表示X迴歸Y之直線。如圖二十六中CC'線C點之橫坐標爲30,縱坐標爲29.21,C'點之橫坐標爲80,縱坐標爲73.11。RR'線R點之橫坐標爲23.44,縱坐標爲20;R'點之橫坐標爲70.44,縱坐標爲70。CC'線與RR'

線相交於M,M點之橫坐標爲42.8,即X變數(夫之年齡)之算術平均數;縱坐標爲40.6,即Y變數(妻之年齡)之算術平均數。

圖二十六

迴歸線



根據迴歸方程式而推測之結果,非絕對可靠,不過據X變數中一項可以推知Y變數中相當一項之

極可能的結果,據Y變數中一項可以推知X變數中相當一項之極可能的結果而已。此項推測之結果,其可靠程度究屬如何,可藉推測的標準誤(Standard Error of Estimates)測定之,其公式如下:

$$S=\sigma\sqrt{1-r^2}$$

式中: S為推測的標準誤之符號

σ為標準差

r為相關係數

計算據X推測Y之標準誤,應以Y變數之標準差乘$\sqrt{1-r^2}$;如據Y而推測X,則標準誤之求法應以X變數之標準差乘$\sqrt{1-r^2}$。本例X變數之標準差為13.1(2.62×5=13.10), Y變數之標準差為12.70(2.54×5=12.7),相關係數為+.915,故:

$$S_y=12.70\sqrt{1-(.915)^2}$$
$$=5.08$$
$$S_x=13.10\sqrt{1-(.915)^2}$$
$$=5.24$$

標準誤之解釋乃謂推測之結果係一極可能或極近似之數目,事實上推測之結果,在加減一標準誤範圍內之機會當有68.26%,在加減三標準誤範圍內之機會當有99.73%。如知夫之年齡為80,推知其妻之

年齡爲73.11(參看表三十二),此項推測之結果係一極近似之數目,事實上妻之年齡不出73.11±5.08範圍之機會爲百分之六八・二六,不出73.11±(3×5.08)範圍之機會爲百分之九九・七三。如知妻之年齡爲50,推知其夫之年齡爲51.64 (參看表三十二),則事實上夫之年齡不出51.64±5.24範圍之機會當爲百分之六八・二六,不出51.64±(3×5.24)範圍之機會爲百分之九九・七三。標準誤以外,吾人亦可以機差測定此項推測可靠之程度,其解釋則爲事實上之X或Y在推測的結果加減一機差範圍內之機會爲百分之五十,加減四機差範圍內之機會爲百分之九十九。計算機差之公式如下:

1. 如Y爲推測的結果:

$$P.E._y = 0.6745\sigma_y\sqrt{1-r^2}$$

2. 如X爲推測的結果:

$$P.E._x = 0.6745\sigma_x\sqrt{1-r^2}$$

惟當注意者,相關之程度,如達完全之地位,則相關係數等於一,標準誤及機差均爲零,意即據某變數中一項可以推知另一變數中相當之一項,而此項推測之結果可謂絕對可靠;反之,相關係數愈小,則標準誤與機差之數值愈大,此項推測可靠之程度愈低。

第三節 相關率

乘積率法爲計算直線相關(Linear Correlation)最佳之方法,然若遇非直線相關(Non-linear Correlation),則乘積率法爲不適用。皮爾生教授因之倡用一法爲計算非直線相關之用,謂之相關率。吾人欲判別兩種

圖二十七

5317對夫妻年齡散佈圖★

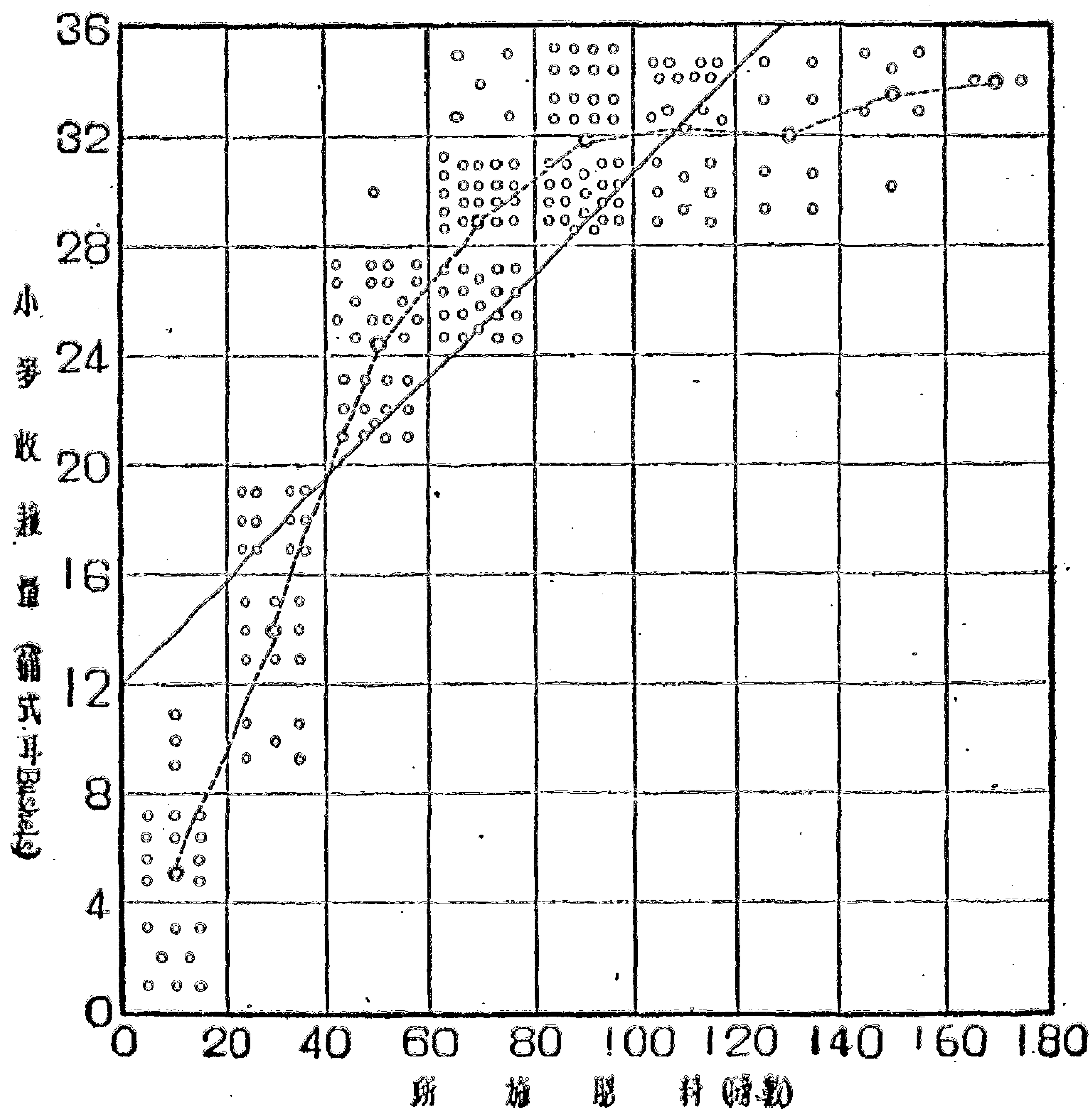
妻之年齡 \ 夫之年齡	15–20	20–25	25–30	30–35	35–40	40–45	45–50	50–55	55–60	60–65	65–70	70–75	75–80	80–85	85–90
85–90														1	
80–85											1	1	2	2	1
75–80										1	2	6	12	5	1
70–75									1	4	13	31	14	4	1
65–70							1	2	6	23	58	31	10	2	1
60–65						1	2	10	35	101	53	18	5	1	
55–60					1	2	10	44	141	81	26	8	3	1	
50–55				1	2	12	59	195	110	39	11	5	2	1	
45–50			1	2	12	66	252	145	46	16	6	2	1		
40–45			2	12	80	378	178	57	18	8	3	1	1		
35–40		1	10	84	359	219	66	19	8	3	1	1			
30–35		4	84	44	251	71	20	8	3	1	1				
25–30		46	482	265	69	17	6	2	1						
20–25	2	173	185	41	9	3	1								
15–20	2	16	4	1											

★因次數太多不便用記號表示,故仍用數字。

變數之關係為直線的或非直線的,可先將此兩種變數列入相關表或散佈圖,如連接各行變數均點(Means)所成之線近於直線,則為直線相關(如圖二十七),可用乘積率法求其相關係數;如非直線(如圖二十八),則當求其相關率。

圖二十八

193畝麥田每畝小麥收穫量
與其所施肥料數量散佈圖



用乘積率法求相關係數,不論以某種變數爲自變數或因變數,其結果常一致;求相關率則當先分別孰爲自變數,孰爲因變數,然後着手計算;蓋以X爲自變數所求得之相關率與以Y爲自變數所求得之相關率不相等也。相關係數求得後,可以算出兩種變數之迴歸係數,據某數中之一項可以推知他數相當一項之數值,在非直線相關則不能。相關係數爲正爲負,視相關之正負而定;相關率有時竟不冠以符號,因既爲非直線相關,其關係或一部份爲正,一部份爲負,不能肯定其全部之關係爲正或爲負也。相關係數吾人恆以r代之,相關率則以η代之(讀如eta),其公式有二:

1. 以X爲自變數:

$$\eta_{yx}=\frac{\sigma_{my}}{\sigma_y}$$

2. 以Y爲自變數:

$$\eta_{xy}=\frac{\sigma_{mx}}{\sigma_x}$$

上式η右下方之附字yx及xy,第一字係指因變數,第二字指自變數。式中σ_y爲Y變數之標準差,σ_x爲X變數之標準差,已爲吾人所熟知,惟σ_{my}及σ_{mx}究爲何物,尙有一加解釋之必要。σ_{my}者,乃各直行均點對於全表Y數平均數之標準差也(簡稱直行均點標

準差),其算法如下:

$$\sigma_{my}=\sqrt{\frac{S[n_x(\overline{Y}_x-\overline{Y})^2]}{N}}$$

式中: S　為總加之符號

n_x 為每一直行次數之和

$\overline{Y}_x$ 為每一直行變數之平均數

$\overline{Y}$ 為 Y 數之算術平均數

N　為次數總數

σ_{mx} 者,乃各橫行均點對於全表 X 數平均數之標準差也(簡稱橫行均點標準差),其算法如下:

$$\sigma_{mx}=\sqrt{\frac{S[n_y(\overline{X}_y-\overline{X})^2]}{N}}$$

式中: S　為總加之符號

n_y 為每一橫行次數之和

$\overline{X}_y$ 為每一橫行變數之平均數

$\overline{X}$ 為 X 數之算術平均數

N　為次數總數

茲就一百九十三畝麥田小麥收穫量及其所施之肥料一例,計算其相關率如表三十三。本例係以小麥之收穫量為 Y 變數,所施肥料之數量為 X 變數,故所用公式為

$$\eta_{yx}=\frac{\sigma_{my}}{\sigma_y}$$

至於

表三十三

193畝麥田每畝小麥收穫量與其所施肥料數量相關表

X——所施肥料(磅數)；Y——小麥收穫量(蒲式耳)

Y \ X	0-19.9	20-39.9	40-59.9	60-79.9	80-99.9	100-119.9	120-139.9	140-159.9	160-179.9	n_y	$\bar{X}_y$
32-35.9				5	16	12	4	5	2	44	107.27
28-31.9			1	20	21	8	4	1		55	88.91
24-27.9			16	19						35	60.86
20-23.9			13							13	50.00
16-19.9		12								12	30.00
12-15.9		8								8	30.00
8-11.9	3	5								8	22.50
4- 7.9	10									10	10.00
0- 3.9	8									8	10.00
n_x	21	25	30	44	37	20	8	6	2	193	
$\bar{Y}_x$	5.05	15.12	24.4	28.73	31.73	32.4	32.0	33.33	34.0		
$(\bar{Y}_x-\bar{Y})$	−19.9	−10.0	−.6	+3.7	+6.7	+7.4	+7.0	+8.3	+9.0		
$(\bar{Y}_x-\bar{Y})^2$	396.1	100.0	.36	13.69	44.9	54.8	49.0	68.9	81.0		
$n_x(\bar{Y}_x-\bar{Y})^2$	8318.0	2500.0	10.8	602.8	1661.3	1096.0	392.0	413.4	162.0	1,156.3	

m	f	d'	fd'	fd'²
34	44	+3	132	396
30	55	+2	110	220
26	35	+1	35	35
22	13	0	0	0
18	12	−1	−12	12
14	8	−2	−16	32
10	8	−3	−24	72
6	10	−4	−40	160
2	8	−5	−40	200
	N=193		277−132	1,127
			=145	

$$\bar{Y}=22+\frac{145}{193}\times 4=25$$

$$\sigma_y=\sqrt{\frac{1127}{193}-\left(\frac{145}{193}\right)^2}\times 4$$

$$=\sqrt{5.6-.6}\times 4$$

$$=2.3\times 4$$

$$=9.2$$

$$\sigma_{my}=\sqrt{\frac{15,156.3}{193}}=\sqrt{78.6}=8.9$$

$$\eta_{yx}=\frac{\sigma_{my}}{\sigma_y}=\frac{8.9}{9.2}=.967$$

$$\eta_{xy}=\frac{\sigma_{mx}}{\sigma_x}$$

一式之計算,方法相同,例從略。

相關率之計算已詳上表,茲將本例之步驟列下:

a. 以每畝麥田所施之肥料數量爲 X 變數,每畝麥田小麥收穫量爲 Y 變數,製相關表。

b. 將表中各直行次數及各橫行次數相加求 n_x 及 n_y(此與計算相關係數時求 f_x 及 f_y 之法相同)。

c. 求 Y 數之算術平均數($\bar{Y}$)。

d. 求每一直行 Y 數之平均數($\bar{Y}_x$)。

e. 求 Y 之標準差(σ_y)。

f. 由每一直行 Y 數之平均數中減去 Y 之平均數($\bar{Y}_x-\bar{Y}$)。

g. 將各項($\bar{Y}_x-\bar{Y}$)自乘求($\bar{Y}_x-\bar{Y})^2$。

h. 以每一直行之次數(n_x)乘各該行之 $(\bar{Y}_x-\bar{Y})^2$, 得 $n_x(\bar{Y}_x-\bar{Y})^2$。

i. 將所有之 $n_x(\bar{Y}_x-\bar{Y})^2$ 相加,是爲 $S[n_x(\bar{Y}_x-\bar{Y})^2]$。

j. 以次數總數除 $S[n_x(\bar{Y}_x-\bar{Y})^2]$,結果卽爲直行均點標準差($\sigma_{my}$)。

k. 以 σ_y 除 σ_{my}, 卽得相關率(η_{yx})。

相關率之解釋與相關係數大略相同,其確度亦恆以機差表示,其式如下:

$$\text{P. E. } \eta = \frac{0.6745(1-\eta^2)}{\sqrt{n}}$$

本例相關率爲·967,其機差應爲·0003:

$$\left(\frac{\cdot 6745(1-\cdot 967^2)}{\sqrt{193}} = \frac{\cdot 6745 \times \cdot 065}{13.9} = \frac{\cdot 044}{13.9} = \cdot 0003\right)$$

僅及相關率萬分之三,可見相關程度甚高也。

惟 η 之數値可恃與否,須視材料之多寡與分組之精粗而定。如 N 甚小,則 η 之數値殊不可靠。如次數少而分組多,則 η 求出後須用校正法加以校正。此校正法爲皮爾生教授所提示,謂之相關率校正法(Correction of Correlation Ratio),其式如下:

$$\text{校正的}\eta^2 = \frac{\eta^2 - \frac{(k-3)}{N}}{1 - \frac{(k-3)}{N}}$$

式中:η^2 爲相關率之平方

k 爲行數(本例爲直行行數,設所求者爲 η_{xy},則 k 代表橫行行數)

N 爲次數總數

按此法校正,如 N 甚大而分組適當,則校正後之 η 與未校正時相差不大;反之,如 N 小而分組太多,則校正後之 η 將遠較原來之 η 爲小,蓋當每組次數少

至一時,則$\sigma_{my}=\sigma_y$,本例校正後的η,結果爲·966,與原來之η出入甚微,但若組數增多,次數減少,則結果將不同矣。

$$\text{校正的}\eta^2=\frac{\cdot967^2-\frac{9-3}{193}}{1-\frac{9-3}{193}}$$

$$=\frac{\cdot935-\cdot031}{1-\cdot031}$$

$$=\frac{\cdot904}{\cdot969}$$

$$=\cdot933$$

$$\eta=\cdot966$$

判別直線相關與非直線相關固可就散佈圖上兩種變數分佈之趨勢以爲決定,此外亦可視相關係數與相關率數值之大小以爲斷。設有兩種相關之變數於此,吾人既求其相關係數,復求其相關率,如爲直線相關,則二者之數值應相等;如爲非直線相關,則相關係數之數值恆小於相關率,r與η相差愈大,愈能確定爲非直線相關。統計學家常利用η^2與r^2二者之差爲判別直線相關或非直線相關之標準,其式如下:

$$\zeta=\eta^2-r^2$$

(ζ讀如zeta)

惟ζ之數值如不甚大,吾人尚不能遽斷爲非直線相關,蓋η與r均有受取樣影響之可能,縱爲直線

相關,彼此或倘不能完全吻合。本例校正的相關率爲·966,如用乘積率法求其相關係數,則爲 +·793。按上式計算,$\zeta=(\cdot966)^2-(\cdot793)^2=\cdot933-\cdot629=\cdot304$,相差甚大,故可斷言其爲非直線相關。$\zeta$之數值究應大至如何程度始能確定關係之爲非直線的,當視其標準誤而定,請於另章論之。

第四節 均方相關法

均方相關法爲由質的統計計算相關之方法。所謂質的統計者,乃不可以數量精密測量之材料,如人類之性情有溫和的,紆緩的,暴躁的;眼睛之顏色有黑,灰,藍,櫻等色。遇有此等材料,乘積率法與相關率均不適用,蓋吾人祇能將其分類而不能分組也。求品質的統計相關之方法,以均方相關法爲最佳,此法爲皮爾生教授所倡用,其式如下:

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

式中:C 爲均方相關之係數

N 爲次數總數

$$S=\Sigma\left\{\frac{(n_{rc})^2}{\frac{n_r n_c}{N}}\right\}$$

n_{rc} 代表相關表中各方格中之次數

n_r 代表每一橫行次數之和

n_c 代表每一直行次數之和

$n_r n_c$ 代表每方格所在橫行次數之和與所在直行次數之和相乘之積

用均方相關法計算相關,亦須先列相關表(惟係分類的,與前舉二例稍有不同),其求得之結果謂之均方相關係數,茲舉例如下:

表三十四

1000對父子眼睛顏色相關表

父之眼睛顏色

子之眼睛顏色		藍	灰	深灰	褸	總計
	藍	194	70	41	30	335
	灰	83	124	41	36	284
	深灰	25	34	55	23	137
	褸	56	36	43	109	244
	總計	358	264	1 0	198	1000

I.求 $\frac{n_r n_c}{N}$	II.求 $(n_{rc})^2$	III.求 $\frac{(n_{rc})^2}{\frac{n_r n_c}{N}}$ = S_1
$\frac{335\times358}{1000}=119.9$	$(194)^2=37,636$	$\frac{37,636}{119.9}=313.9$
$\frac{284\times358}{1000}=101.7$	$(83)^2=6,889$	$\frac{6,889}{101.7}=67.7$
$\frac{137\times358}{1000}=49.0$	$(25)^2=625$	$\frac{625}{49}=12.8$
$\frac{244\times358}{1000}=87.4$	$(56)^2=3,136$	$\frac{3,136}{87.4}=37.0$

$\frac{335\times264}{1000}=88.4$	$(70)^2=4,900$	$\frac{4,900}{88.4}=55.4$
$\frac{284\times264}{1000}=75.0$	$(124)^2=15,376$	$\frac{15,376}{75.0}=205.0$
$\frac{137\times264}{1000}=36.2$	$(34)^2=1,156$	$\frac{1,156}{36.2}=31.9$
$\frac{244\times264}{1000}=64.4$	$(36)^2=1,296$	$\frac{1,296}{64.4}=20.1$
$\frac{335\times180}{1000}=38.3$	$(41)^2=1,681$	$\frac{1,681}{38.3}=43.9$
$\frac{284\times180}{1000}=51.1$	$(41)^2=1,681$	$\frac{1,681}{51.1}=32.9$
$\frac{137\times180}{1000}=24.7$	$(55)^2=3,025$	$\frac{3,025}{24.7}=122.5$
$\frac{244\times180}{1000}=43.9$	$(43)^2=1,849$	$\frac{1,849}{43.9}=42.1$
$\frac{335\times198}{1000}=66.3$	$(30)^2=900$	$\frac{900}{66.3}=13.6$
$\frac{284\times198}{1000}=56.2$	$(36)^2=1,296$	$\frac{1,296}{56.2}=23.1$
$\frac{137\times198}{1000}=27.1$	$(23)^2=529$	$\frac{529}{27.1}=19.5$
$\frac{244\times198}{1000}=48.3$	$(109)^2=11,881$	$\frac{11,881}{48.3}=246.0$
		$S=1,287.4$

$$C=\sqrt{\frac{1287.4-1000}{1287.4}}=\sqrt{\frac{287.4}{1287.4}}$$

$$=\sqrt{.232}\quad=.473$$

均方相關法之意義在將各方格中之實得次數與全憑機遇之次數一一比較。如本例一千對父子中藍眼之父共爲358人,藍眼之子共爲335人,父子眼睛均爲藍色者共194對(第一直行與第一橫行相交一方

格中之次數),設父子眼睛顏色毫無關係,全憑機遇湊合,則此方格中之次數應爲 120 $\left(\frac{335}{1000}\times\frac{358}{1000}=\frac{119.9}{1000}\right)$,今實得次數爲194,足見其有關係而非全憑機遇也。將各方格之機遇次數一一求出(如表三十四中I所示)除各該方格中實得次數之自乘方(如表三十四中II,III所示),然後將所得之商一一相加,其和即爲式中之S,S如等於N,則C=0,表示兩種品質毫無關係;反之,如S較N愈大,則相關之程度愈高。通常求得之C不附以正負符號,但爲解釋方便起見,如相關表內某種品質之高度與另一品質之高度相聯,則C爲正。如某種品質之高度與另一品質之低度相聯,則C爲負。本例父之眼色藍者,子亦多爲藍色;父之眼灰色者,子亦多爲灰色,故爲正相關,C應附以正號。均方相關法有一缺點,學者宜注意之。蓋用此法計算相關,對於同種材料,如分組不同,C即因之而異。本例橫直各分四行,設改爲5×5行,則C將不同。據友爾(G. U. Yule)教授之意見:

當行數爲 2×2時	C不能大於.707
當行數爲 3×3時	C不能大於.816
當行數爲 4×4時	C不能大於.866
當行數爲 5×5時	C不能大於.894

當行數爲 6× 6 時　　　C不能大於.913

當行數爲 7× 7 時　　　C不能大於.926

當行數爲 8× 8 時　　　C不能大於.935

當行數爲 9× 9 時　　　C不能大於.943

當行數爲 10×10 時　　　C不能大於.949

故採用此法,相關表中至少應有5×5格,C始能近於一。本例僅有4×4格,所得之C似尙不足以充分表示父子眼色相關之程度。惟分類亦不宜太細,因分類太細,不特計算之手續加繁,且C易受少數偶然現象之影響,而爲之牽動。

此外,C與相關係數(r)之關係亦有一述之必要,大概分類適當(至少5×5格),取樣較多,而兩種品質俱呈常態的分配時,C可視爲r,如與上述條件相去愈遠,則C與r相差愈大。

均方相關法之理論略具於斯,但計算均方相關係數時尙可將上式改爲下式,以省手續。

$$C=\sqrt{\frac{P-1}{P}}$$

式中:$P=\frac{S}{N}$

用簡法計算均方相關係數係將求各方格中機遇次數$\left(\frac{n_r n_c}{N}\right)$,及機遇次數除$(n_{rc})^2$數步手續併而

爲一。如表三十四第一直行內計有四方格,由上至下第一方格內次數爲194,第二方格內次數爲83,第三方格內次數爲25,第四方格內次數爲56。以第一方格內之機遇次數除194之平方等於

$$\frac{(194)^2}{\frac{335\times358}{1000}}=\frac{(194)^2\times1000}{335\times358}$$

第二方格內之機遇次數除83之平方等於

$$\frac{(83)^2}{\frac{284\times358}{1000}}=\frac{(83)^2\times1000}{284\times358}$$

第三方格內之機遇次數除25之平方等於

$$\frac{(25)^2}{\frac{137\times358}{1000}}=\frac{(25)^2\times1000}{137\times358}$$

第四方格內之機遇次數除56之平方等於

$$\frac{(56)^2}{\frac{244\times358}{1000}}=\frac{(56)^2\times1000}{244\times358}$$

每一$\frac{N(n_{rc})^2}{n_r n_c}$之分子分母中均可提出一公因數。本例第一直行四$\frac{N(n_{rc})^2}{n_r n_c}$之和等於

$$\frac{1000}{358}\left[\frac{(194)^2}{335}+\frac{(83)^2}{284}+\frac{(25)^2}{137}+\frac{(56)^2}{244}\right]$$

蓋N爲分子之公因數,而358爲分母之公因數也。然N(1000)又爲其他直行$\Sigma\left[\frac{N(n_{rc})^2}{n_r n_c}\right]$之公因數,故可置之不顧,如此則求得之S應小N倍。設以P.代

表用簡法求得之 S,則

$$P=\frac{S}{N}$$

原來公式因之改為

$$C=\sqrt{\frac{P-1}{P}}$$

茲列式證明如下:

$$因\ P=\frac{S}{N}$$

$$故\ \sqrt{\frac{S-N}{S}}=\sqrt{\frac{NP-N}{NP}}$$

$$=\sqrt{\frac{P-1}{P}}$$

茲仍就上例用簡法計算均方相關係數於下:

$$第一直行之\Sigma\left[\frac{(n_{rc})^2}{n_r n_c}\right]:\frac{1}{358}\times\left(\frac{194^2}{335}+\frac{83^2}{284}+\frac{25^2}{137}+\frac{56^2}{244}\right)$$

$$=.43$$

$$第二直行之\Sigma\left[\frac{(n_{rc})^2}{n_r n_c}\right]:\frac{1}{264}\times\left(\frac{70^2}{335}+\frac{124^2}{284}+\frac{34^2}{137}+\frac{36^2}{244}\right)$$

$$=.31$$

$$第三直行之\Sigma\left[\frac{(n_{rc})^2}{n_r n_c}\right]:\frac{1}{180}\times\left(\frac{41^2}{335}+\frac{41^2}{284}+\frac{55^2}{137}+\frac{43^2}{244}\right)$$

$$=.22$$

$$第四直行之\Sigma\left[\frac{(n_{rc})^2}{n_r n_c}\right]:\frac{1}{198}\times\left(\frac{30^2}{335}+\frac{36^2}{284}+\frac{23^2}{137}+\frac{109^2}{244}\right)$$

$$=.30$$

$$P=1.26$$

按簡法計算,C 為 .454, 較表三十四之結果略小,殆由小數舍入之關係而微有出入也。

第十章　淨相關與複相關

第一節　淨相關與複相關之意義

前章所述為測度兩種變數或品質相關之方法,惟任何兩種變數或品質甲與乙相關之外,與丙,丁或亦有關係。如表三十三193畝麥田每畝小麥收穫量與其所施肥料數量之例,小麥收穫量之多寡固與所施肥料之數量有密切之關係,但與雨量,溫度等其他因子亦必有其相當之關係。相關率之結果僅告吾人小麥收穫量與其所施肥料數量之關係。既未嘗將雨量,溫度等因子計入,亦未摒除,如此求得之結果,統計學家恆稱之為簡相關(Simple Córrelation),第九章所述者均屬此類。所謂淨相關(Partial Correlation)者,乃兩種變數或品質間之純淨關係,而其他因子均經摒除淨盡(意卽假定其他因子之影響相同而不變),如小麥收穫量為因變數,吾人僅求其與所施肥料數量之關係,而假定雨量,溫度等均相同,則所得之結果謂之淨相關。所謂複相關(Multiple Correlation)者,乃一因變數或品質與多種自變數或品質之關係,如小麥收穫量為因變數,吾人不僅求其與所施肥料數量之關係,吾人所求者為小麥收穫量與所施肥料數量及雨量,溫度

等總合之關係,則所得之結果,謂之複相關。猶之化學家之作試驗,淨相關譬如兩種純淨原素之化合,複相關譬如多種純淨原素之化合,簡相關則如兩種含有雜質不純淨的原素之化合。社會現象,錯綜紛紜,吾人自難將一切相關之事實一一舉而出之,然在相當範圍內,如能摒除不切合之因子,就所欲研究者求出其間之關係,自較僅顧及一較重要之自變數與其因變數之簡相關爲愈也。下表爲美國甘塞斯(Kansas)州1890年至1922年玉蜀黍收穫量及各年六月,七月,八月各月平均溫度表,吾人可單求各年玉蜀黍收穫量與各年六月,七月,或八月平均溫度之關係,亦可求各年玉蜀黍收穫量與各年六,七,八三個月溫度總合之關係。惟六,七,八三月之溫度與玉蜀黍之收穫量均有關係,若吾人單求各年玉蜀黍收穫量與六,七,八三月中任何一月平均溫度之關係,而將其他二月溫度之影響摒去,則結果爲淨相關;若吾人將六,七,八三月溫度與玉蜀黍收穫量之關係一併計入,則結果爲複相關。

表三十五

美國甘塞斯州1890—1922年各年玉蜀黍

收穫量與6,7,8各月平均溫度表

1 年份	2 玉蜀黍每畝之收穫量(蒲式耳)	3 六月平均溫度(華氏表)	4 七月平均溫度(華氏表)	5 八月平均溫度(華氏表)
1890	15.6	77.6	83.1	76.1
1891	26.7	70.7	74.0	75.1
1892	24.5	73.4	77.5	76.5
1893	21.3	74.7	79.5	73.8
1894	11.2	74.2	77.8	78.0
1895	24.3	71.7	74.9	76.0
1896	28.0	74.1	78.1	78.7
1897	18.0	76.6	80.2	76.0
1898	16.0	75.0	77.7	78.2
1899	27.0	73.9	76.2	80.6
1900	19.0	74.9	77.9	81.0
1901	7.8	77.3	85.0	79.1
1902	29.9	70.9	76.8	78.2
1903	25.6	67.2	78.3	75.3
1904	20.9	70.4	75.6	74.6
1905	27.7	75.5	74.5	78.7
1906	28.9	71.8	73.8	76.3
1907	22.1	72.0	78.4	78.1
1908	22.0	72.1	75.8	76.2
1909	19.9	73.1	78.1	80.1
1910	19.0	72.2	79.5	75.7
1911	14.5	80.5	78.6	76.4
1912	23.0	69.3	79.9	77.4
1913	3.2	74.2	82.1	84.2
1914	18.5	78.2	79.9	78.2
1915	31.0	69.2	74.0	70.1
1916	10.0	70.3	81.2	79.6
1917	13.0	72.8	80.8	73.4
1918	7.1	78.4	78.3	82.3
1919	15.2	72.3	80.2	78.3
1920	26.5	72.8	77.6	72.9
1921	22.2	74.4	79.2	78.6
1922	19.3	75.2	77.0	80.1

第二節　淨相關之計算法

上例吾人已知玉蜀黍之收穫量與六,七,八三月之溫度均有關係,設吾人單欲求玉蜀黍之收穫量,與

六月平均溫度之關係,則必須將七,八兩月溫度之影響除去。摒除之法可選取七,八兩月平均溫度相同之若干年,然後就此若干年之材料,求玉蜀黍收穫量與六月平均溫度之關係。但七八兩月平均溫度相同之年殊不易得,故此法不甚適用,於是不得不利用數學方法以處理之矣。

計算淨相關之公式視變數種類之多寡而異。如自變數爲二,則所得之係數謂之一次係數(Coefficient of first order);自變數爲三,則所得之係數謂之二次係數(Coefficient of second order);自變數爲四,則所得之係數謂之三次係數(Coefficient of third order);餘遞推。至於簡相關之係數,因祇一自變數,故謂之零次係數(Coefficient of zero order)。各次淨相關係數均由其較低一次之係數求得,故計算某次淨相關係數,必先求簡相關係數,然後依次遞求。

一次淨相關係數之公式爲:

$$r_{12\cdot3}=\frac{r_{12}-r_{13}r_{23}}{(1-r_{13}^2)^{\frac{1}{2}}(1-r_{23}^2)^{\frac{1}{2}}}*$$

式中: $r_{12\cdot3}$ 爲一次淨相關係數之符號(按即因變數與第一自變數之淨相關而摒去第二自變數之影響者)

r_{12}為因變數與第一自變數簡相關之符號

r_{13}為因變數與第二自變數簡相關之符號

r_{23}為第一自變數與第二自變數簡相關之符號

二次淨相關係數之公式為:

$$r_{12\cdot34}=\frac{r_{12\cdot3}-r_{14\cdot3}\cdot r_{24\cdot3}}{(1-r_{14\cdot3}^{2})^{\frac{1}{2}}(1-r_{24\cdot3})^{\frac{1}{2}}}\ *$$

式中: $r_{12\cdot34}$為二次淨相關係數之符號

$r_{12\cdot3}$為一次淨相關係數之符號(因變數與第一自變數之淨相關)

$r_{14\cdot3}$為一次淨相關係數之符號(因變數與第三自變數之淨相關)

$r_{24\cdot3}$為一次淨相關係數之符號(第一自變數與第三自變數之淨相關)

*$r_{13\cdot2}$, $r_{13\cdot24}$, $r_{14\cdot23}$等之公式可由上式類推

依次遞推,則 n 次淨相關係數之公式爲:

$$r_{12\cdot345\cdots n}=\frac{r_{12\cdot345\cdots(n-1)}-r_{1n\cdot345\cdots(n-1)}\,r_{2n\cdot345\cdots(n-1)}}{\left[1-r^2_{1n\cdot345\cdots(n-1)}\right]^{\frac{1}{2}}\left[1-r^2_{2n\cdot345\cdots(n-1)}\right]^{\frac{1}{2}}}$$

r 下之附字所以表示淨相關之次數,如有三個附字,則爲一次淨相關;四個附字,則爲二次淨相關,餘遞推。在點之前者謂之前附字(Primary subscripts),在點之後者謂之後附字 (Secondary subscripts)。淨相關符號(r)下之前附字均爲兩個,表示兩種變數其淨相關爲吾人所欲求者,後附字表示與因變數相關之自變數,其影響須摒去者。前附字中之第一字係表示因變數,第二字表示自變數,次序恆有一定,後附字之次序則無關重要,$r_{12\cdot3456}$ 與 $r_{12\cdot4365}$ 其意義初無少異也。

淨相關之公式已如上述,茲就表三十五之例計算其淨相關係數。惟吾人有當注意者,表三十五中自變數有三,應按二次淨相關係數之公式計算,但欲求二次淨相關係數必先求一次淨相關係數;欲求一次淨相關係數,必先求零次相關係數(即簡相關)。茲以 X_1 代玉蜀黍收穫量,X_2 代六月間平均溫度,X_3 代七月間平均溫度,X_4 代八月間平均溫度,按乘積率法($r=\frac{\Sigma xy}{\sqrt{\Sigma x^2\,\Sigma y^2}}$)計算 r_{12}, r_{13}, r_{14}, r_{23}, r_{24}, r_{34} 如下:

六七月 六八月 七八月

表　三　十　六

1890—1922年各年玉蜀黍收穫量與各年六月間平均溫度零次相關係數計算表

年份	X_1			X_2			x_1x_2	
	玉蜀黍收穫量(蒲式耳)	x_1	x_1^2	六月間平均溫度(華氏表)	x_2	x_2^2		
1890	15.6	−4.4	19.36	77.6	+4.1	16.81	−18.04	
1891	26.7	+6.7	44.89	70.7	−2.8	7.84	−18.76	
1892	24.5	+4.5	20.25	73.4	−0.1	0.01	−0.45	
1893	21.3	+1.3	1.69	74.7	+1.2	1.44		+1.56
1894	11.2	−8.8	77.44	74.2	+0.7	0.49	−6.16	
1895	24.3	+4.3	18.49	71.7	−1.8	3.24	−7.74	
1896	28.0	+8.0	64.00	74.1	+0.6	0.36		+4.80
1897	18.0	−2.0	4.00	76.6	+3.1	9.61	−6.20	
1898	16.0	−4.0	16.00	75.0	+1.5	2.25	−6.00	
1899	27.0	+7.0	49.00	73.9	+0.4	0.16		+2.80
1900	19.0	−1.0	1.00	74.9	+1.4	1.96	−1.40	
1901	7.8	−12.2	148.84	77.3	+3.8	14.44	−46.36	
1902	29.9	+9.9	98.01	70.9	−2.6	6.76	−25.74	
1903	25.6	+5.6	31.36	67.2	−6.3	39.69	−35.28	
1904	20.9	+0.9	0.81	70.4	−3.1	9.61	−2.79	
1905	27.7	+7.7	59.29	75.5	+2.0	4.00		+15.40
1906	28.9	+8.9	79.21	71.8	−1.7	2.89	−15.13	
1907	22.1	+2.1	4.41	72.0	−1.5	2.25	−3.15	
1908	22.0	+2.0	4.00	72.1	−1.4	1.96	−2.80	
1909	19.9	−0.1	0.01	73.1	−0.4	0.16		+0.04
1910	19.0	−1.0	1.00	72.2	−1.3	1.69		+1.30
1911	14.5	−5.5	30.25	80.5	+7.0	49.00	−38.50	
1912	23.0	+3.0	9.00	69.3	−4.2	17.64	−12.60	
1913	3.2	−16.8	282.24	74.2	+0.7	0.49	−11.76	
1914	18.5	−1.5	2.25	78.2	+4.7	22.09	−7.05	
1915	31.0	+11.0	121.00	69.2	−4.3	18.49	−47.30	
1916	10.0	−10.0	100.00	70.3	−3.2	10.24		+32.00
1917	13.0	−7.0	49.00	72.8	−0.7	0.49		+4.90
1918	7.1	−12.9	166.41	78.4	+4.9	24.01	−63.21	
1919	15.2	−4.8	23.04	72.3	−1.2	1.44		+5.76
1920	26.5	+6.5	42.25	72.8	−0.7	0.49	−4.55	
1921	22.2	+2.2	4.84	74.4	+0.9	0.81		+1.98
1922	19.3	−0.7	0.49	75.2	+1.7	2.89	−1.19	
	33 | 658.9 M=20		Σx_1^2 =1573.8	33 | 2426.9 M=73.5		Σx_2^2 =275.70	Σx_1x_2 = −382.16+70.54 = −311.62	

$$r_{12}=\frac{-311.62}{\sqrt{1573.83\times835.70}}=\frac{-311.62}{39.67\times16.60}=\frac{-311.62}{658.52}=-.473$$

表 三 十 七

1890—1922年各年玉蜀黍收穫量與各年七月間平均溫度零次相關係數計算表

年份	X_1			X_3			x_1x_3	
	玉蜀黍收穫量(蒲式耳)	x_1	x_1^2	七月間平均溫度(華氏表)	x_3	x_3^2		
1890	15.6	−4.4	19.36	83.1	4.88	23.81	−21.47	
1891	26.7	+6.7	44.89	74.0	−4.22	17.81	−28.27	
1892	24.5	+4.5	20.25	77.5	−0.72	0.52	− 3.24	
1893	21.3	+1.3	1.69	79.5	1.28	1.64		1.66
1894	11.2	−8.8	77.44	77.8	−0.42	0.18		3.70
1895	24.3	+4.3	18.49	74.9	−3.32	11.02	−14.27	
1896	28.0	+8.0	64.00	78.1	−0.12	0.01	− 0.96	
1897	18.0	−2.0	4.00	80.2	1.98	3.92	− 3.96	
1898	16.0	−4.0	16.00	77.7	−0.52	0.27		2.08
1899	27.0	+7.0	49.00	76.2	−2.02	4.08	−14.14	
1900	19.0	−1.0	1.00	77.9	−0.32	0.10		0.32
1901	7.8	−12.2	148.84	85.0	6.78	45.97	−82.72	
1902	29.9	+9.9	98.01	76.8	−1.42	2.02	−14.06	
1903	25.6	+5.6	31.36	78.3	0.08	0.01		0.45
1904	20.9	+0.9	0.81	75.6	−2.62	6.86	− 2.36	
1905	27.7	+7.7	59.29	74.5	−3.72	13.84	−28.64	
1906	28.9	+8.9	79.21	73.8	−4.42	19.54	−39.34	
1907	22.1	+2.1	4.41	78.4	0.18	0.03		0.38
1908	22.0	+2.0	4.00	75.8	−2.42	5.86	− 4.84	
1909	19.9	−0.1	0.01	78.1	−0.12	0.01		0.01
1910	19.0	−1.0	1.00	79.5	1.28	1.64	− 1.28	
1911	14.5	−5.5	30.25	78.6	0.38	0.14	− 2.09	
1912	23.0	+3.0	9.00	79.9	1.68	2.82		5.04
1913	3.2	−16.8	282.24	82.1	3.88	15.05	−65.18	
1914	18.5	−1.5	2.25	79.9	1.68	2.82	− 2.52	
1915	31.0	+11.0	121.00	74.0	−4.22	17.81	−46.42	
1916	10.0	−10.0	100.00	81.2	2.98	8.88	−29.80	
1917	13.0	−7.0	49.00	80.8	2.58	6.66	−18.06	
1918	7.1	−12.9	166.41	78.3	0.08	0.01	− 1.03	
1919	15.2	−4.8	23.04	80.2	1.98	3.92	− 9.50	
1920	26.5	+6.5	42.25	77.6	−0.62	0.38	− 4.03	
1921	22.2	+2.2	4.84	79.2	0.98	0.96		2.15
1922	19.3	−0.7	0.49	77.0	−1.22	1.49		0.85
	33 | 658.9		Σx_1^2 =1573.83	33 | 2581.5		Σx_3^2 =220.08	Σx_1x_3 =16.64−438.18	
	M=20			M=78.22			=421.54	

$$r_{13}=\frac{-421.54}{\sqrt{1573.83\times220.08}}=\frac{-421.54}{39.67\times14.84}=\frac{-421.54}{588.70}=-.716$$

表三十八

1890—1922年各年六月間平均溫度與七月間平均溫度零次相關係數計算表

Y23

年份	X_2 六月間平均溫度(華氏表)	x_2	x_2^2	X_3 七月間平均溫度(華氏表)	x_3	x_3^2	x_2x_3	
1890	77.6	+4.1	16.81	83.1	4.88	23.81	20.01	
1891	70.7	−2.8	7.84	74.0	−4.22	17.81	11.82	
1892	73.4	−0.1	0.01	77.5	−0.72	0.52	0.07	
1893	74.7	+1.2	1.44	79.5	1.28	1.64	1.53	
1894	74.2	+0.7	0.49	77.8	−0.42	0.18		−0.29
1895	71.7	−1.8	3.24	74.9	−3.32	11.02	5.98	
1896	74.1	+0.6	0.36	78.1	−0.12	0.01		−0.07
1897	76.6	+3.1	9.61	80.2	1.98	3.92	6.14	
1898	75.0	+1.5	2.25	77.7	−0.52	0.27		−0.78
1899	73.9	+0.4	0.16	76.2	−2.02	4.03		−0.81
1900	74.9	+1.4	1.96	77.9	−0.32	0.10		−0.45
1901	77.3	+3.8	14.44	85.0	6.78	45.97	25.76	
1902	70.9	−2.6	6.76	76.8	−1.42	2.02	3.69	
1903	67.2	−6.3	39.69	78.3	0.08	0.01		−0.50
1904	70.4	−3.1	9.61	75.6	−2.62	6.86	8.12	
1905	75.5	+2.0	4.00	74.5	−3.72	13.84		−7.44
1906	71.8	−1.7	2.89	73.8	−4.42	19.54	7.51	
1907	72.0	−1.5	2.25	78.4	0.18	0.03		−0.27
1908	72.1	−1.4	1.96	75.8	−2.42	5.86	3.39	
1909	73.1	−0.4	0.16	78.1	−0.12	0.01	0.05	
1910	72.2	−1.3	1.69	79.5	1.28	1.64		−1.66
1911	80.5	+7.0	49.00	78.6	0.38	0.14	2.66	
1912	69.3	−4.2	17.64	79.9	1.68	2.82		−7.05
1913	74.2	+0.7	0.49	82.1	3.88	15.05	2.72	
1914	78.2	+4.7	22.09	79.9	1.68	2.82	7.90	
1915	69.2	−4.3	18.49	74.0	−4.22	17.81	18.15	
1916	70.3	−3.2	10.24	81.2	2.98	8.88		−9.54
1917	72.8	−0.7	0.49	80.8	2.58	6.66		−1.81
1918	78.4	+4.9	24.01	78.3	0.08	0.01	0.39	
1919	72.3	−1.2	1.44	80.2	1.98	3.92		−2.38
1920	72.8	−0.7	0.49	77.6	−0.62	0.38	0.43	
1921	74.4	+0.9	0.81	79.2	0.98	0.96	0.88	
1922	75.2	+1.7	2.89	77.0	−1.22	1.49		−2.07
	33 | 2426.9 M=73.5		Σx_2^2 =275.70	33 | 2581.5 M=78.22		Σx_3^2 =220.08	Σx_2x_3=127.20−35.12 =92.08	

$$r_{23}=\frac{92.08}{\sqrt{275.70\times220.08}}=\frac{92.08}{16.60\times14.84}=\frac{92.08}{246.34}=.374$$

表 三 十 九

1890—1922年各年玉蜀黍收穫量與各年八月間平均溫度零次相關係數計算表

年份	X_1			X_4			x_1x_4	
	玉蜀黍收穫量(蒲式耳)	x_1	x_1^2	八月間平均溫度(華氏表)	x_4	x_4^2		
1890	15.6	−4.4	19.36	76.1	−1.29	1.66	5.68	
1891	26.7	6.7	44.89	75.1	−2.29	5.24		−15.34
1892	24.5	4.5	20.25	76.5	−0.89	0.79		−4.01
1893	21.3	1.3	1.69	73.8	−3.59	12.89		−4.67
1894	11.2	−8.8	77.44	78.0	0.61	0.37		−5.37
1895	24.3	4.3	18.49	76.0	−1.39	1.93		−5.98
1896	28.0	8.0	64.00	78.7	1.31	1.72	10.48	
1897	18.0	−2.0	4.00	76.0	−1.39	1.93	2.78	
1898	16.0	−4.0	16.00	78.2	0.81	0.66		−3.24
1899	27.0	7.0	49.00	80.6	3.21	10.30	22.47	
1900	19.0	−1.0	1.00	81.0	3.61	13.03		−3.61
1901	7.8	−12.2	148.84	79.1	1.71	2.92		−20.86
1902	29.9	9.9	98.01	78.2	0.81	0.66	8.02	
1903	25.6	5.6	31.36	75.3	−2.09	4.37		−11.70
1904	20.9	0.9	0.81	74.6	−2.79	7.78		−2.51
1905	27.7	7.7	59.29	78.7	1.31	1.72	10.09	
1906	28.9	8.9	79.21	76.3	−1.09	1.19		−9.70
1907	22.1	2.1	4.41	78.1	0.71	0.50	1.49	
1908	22.0	2.0	4.00	76.2	−1.19	1.42		−2.38
1909	19.9	−0.1	0.01	80.1	2.71	7.34		−0.27
1910	19.0	−1.0	1.00	75.7	−1.69	2.86	1.69	
1911	14.5	−5.5	30.25	76.4	−0.99	0.98	5.44	
1912	23.0	3.0	9.00	77.4	0.01	0.00	0.03	
1913	3.2	−16.8	282.24	84.2	6.81	46.38		−114.41
1914	18.5	−1.5	2.25	78.2	0.81	0.66		−1.22
1915	31.0	11.0	121.00	70.1	−7.29	53.14		−80.19
1916	10.0	−10.0	100.00	79.6	2.21	4.88		−22.10
1917	13.0	−7.0	49.00	73.4	−3.99	15.92	27.93	
1918	7.1	−12.9	166.41	82.3	4.91	24.11		−63.34
1919	15.2	−4.8	23.04	78.3	0.91	0.83		−4.37
1920	26.5	6.5	42.25	72.9	−4.49	20.16		−29.19
1921	22.2	2.2	4.84	78.6	1.21	1.46	2.66	
1922	19.3	−0.7	0.49	80.1	2.71	7.34		−1.90
	33 | 658.9 M=20		Σx_1^2 =1573.83	33 | 2553.8 M=77.39		Σx_4^2 =257.14	Σx_1x_4 = 98.76 − 406 36 = −307.60	

$$r_{14} = \frac{-307.60}{\sqrt{1573.83 \times 257.14}} = \frac{-307.60}{39.67 \times 16.07} = \frac{-307.60}{637.50} = -.483$$

表四十

1890—1922年各年七月間平均溫度與八月間平均溫度零次相關係數計算表

∴ r_{24}

年份	X_3 七月間平均溫度(華氏表)	x_3	x_3^2	X_4 八月間平均溫度(華氏表)	x_4	x_4^2	x_3x_4	
1890	83.1	4.88	23.81	76.1	−1.29	1.66	−6.30	
1891	74.0	−4.22	17.81	75.1	−2.29	5.24		+9.66
1892	77.5	−0.72	0.52	76.5	−0.89	0.79		+0.64
1893	79.5	1.28	1.64	73.8	−3.59	12.89	−4.60	
1894	77.8	−0.42	0.18	78.0	0.61	0.37	−0.26	
1895	74.9	−3.32	11.02	76.0	−1.39	1.93		+4.61
1896	78.1	−0.12	0.01	78.7	1.31	1.72	−0.16	
1897	80.2	1.98	3.92	76.0	−1.39	1.93	−2.75	
1898	77.7	−0.52	0.27	78.2	0.81	0.66	−0.42	
1899	76.2	−2.02	4.08	80.6	3.21	10.30	−6.48	
1900	77.9	−0.32	0.10	81.0	3.61	13.03	−1.16	
1901	85.0	6.78	45.97	79.1	1.71	2.92		+11.59
1902	76.8	−1.42	2.02	78.2	0.81	0.66	−1.15	
1903	78.3	0.08	0.01	75.3	−2.09	4.37	−0.17	
1904	75.6	−2.62	6.86	74.6	−2.79	7.78		+7.31
1905	74.5	−3.72	13.84	78.7	1.31	1.72	−4.87	
1906	73.8	−4.42	19.54	76.3	−1.09	1.19		+4.82
1907	78.4	0.18	0.03	78.1	0.71	0.50		+0.13
1908	75.8	−2.42	5.86	76.2	−1.19	1.42		+2.87
1909	78.1	−0.12	0.01	80.1	2.71	7.34	−0.33	
1910	79.5	1.28	1.64	75.7	−1.69	2.86	−2.16	
1911	78.6	0.38	0.14	76.4	−0.99	0.98	−0.38	
1912	79.9	1.68	2.82	77.4	0.01	0.00		+0.02
1913	82.1	3.88	15.05	84.2	6.81	46.38		+26.42
1914	79.9	1.68	2.82	78.2	0.81	0.66		+1.36
1915	74.0	−4.22	17.81	70.1	−7.29	53.14		+30.76
1916	81.2	2.98	8.88	79.6	2.21	4.88		+6.59
1917	80.8	2.58	6.66	73.4	−3.99	15.92	−10.29	
1918	78.3	0.08	0.01	82.3	4.91	24.11		+0.39
1919	80.2	1.98	3.92	78.3	0.91	0.83		+1.80
1920	77.6	−0.62	0.38	72.9	−4.49	20.16		+2.78
1921	79.2	0.98	0.96	78.6	1.21	1.46		+1.19
1922	77.0	−1.22	1.49	80.1	2.71	7.34	−3.31	
	33 | 2581.5 M=78.22		$\Sigma x_3^2=220.08$	33 | 2553.8 M=77.39		$\Sigma x_4^2=257.14$	$\Sigma x_3x_4=-44.79+112.94=68.15$	

$$r_{34}=\frac{68.15}{\sqrt{220.08\times257.14}}=\frac{68\cdot15}{14.84\times16.07}=\frac{68.15}{238.48}=.286$$

表四十一

1890—1922年各年六月間平均溫度與八月間平均溫度零次相關係數計算表

Y24

年份	X_2			X_4			x_2x_4	
	六月間平均溫度(華氏表)	x_2	x_2^2	八月間平均溫度(華氏表)	x_4	x_4^2		
1890	77.6	+4.1	16.81	76.1	−1.29	1.66	−5.29	
1891	70.7	−2.8	7.84	75.1	−2.29	5.24		+ 6.41
1892	73.4	−0.1	0.01	76.5	−0.89	0.79		+ 0.09
1893	74.7	+1.2	1.44	73.8	−3.59	12.89	−4.31	
1894	74.2	+0.7	0.49	78.0	0.61	0.37		+ 0.43
1895	71.7	−1.8	3.24	76.0	−1.39	1.93		+ 2.50
1896	74.1	+0.6	0.36	78.7	1.31	1.72		+ 0.79
1897	76.6	+3.1	9.61	76.0	−1.39	1.93	−4.31	
1898	75.0	+1.5	2.25	78.2	0.81	0.66		+ 1.22
1899	73.9	+0.4	0.16	80.6	3.21	10.30		+ 1.28
1900	74.9	+1.4	1.96	81.0	3.61	13.03		+ 5.05
1901	77.3	+3.8	14.44	79.1	1.71	2.92		+ 6.50
1902	70.9	−2.6	6.76	78.2	0.81	0.66	−2.11	
1903	67.2	−6.3	39.69	75.3	−2.09	4.37		+13.17
1904	70.4	−3.1	9.61	74.6	−2.79	7.78		+ 8.65
1905	75.5	+2.0	4.00	78.7	1.31	1.72		+ 2.62
1906	71.8	−1.7	2.89	76.3	−1.09	1.19		+ 1.85
1907	72.0	−1.5	2.25	78.1	0.71	0.50	−1.07	
1908	72.1	−1.4	1.96	76.2	−1.19	1.42		+ 1.67
1909	73.1	−0.4	0.16	80.1	2.71	7.34	−1.08	
1910	72.2	−1.3	1.69	75.7	−1.69	2.86		+ 2.20
1911	80.5	+7.0	49.00	76.4	−0.99	0.98	−6.93	
1912	69.3	−4.2	17.64	77.4	0.01	0.00	−0.04	
1913	74.2	+0.7	0.49	84.2	6.81	46.38		+ 4.77
1914	78.2	+4.7	22.09	78.2	0.81	0.66		+ 3.81
1915	69.2	−4.3	18.49	70.1	−7.29	53.14		+31.35
1916	70.3	−3.2	10.24	79.6	2.21	4.88	−7.07	
1917	72.8	−0.7	0.49	73.4	−3.99	15.92		+ 2.79
1918	78.4	+4.9	24.01	82.3	4.91	24.11		+24.06
1919	72.3	−1.2	1.44	78.3	0.91	0.83	−1.09	
1920	72.8	−0.7	0.49	72.9	−4.49	20.16		+ 3.14
1921	74.4	+0.9	0.81	78.6	1.21	1.46		+ 1.09
1922	75.2	+1.7	2.89	80.1	2.71	7.34		+ 4.61
	33 | 2426.9 M=73.5		Σx_2^2 =275.70	33 | 2553.8 M=77.39		Σx_4^2 =257.14	Σx_2x_4=130.05 −33.30=96.75	

$$r_{24}=\frac{96.75}{\sqrt{275.70\times257.14}}=\frac{96.75}{16.60\times16.07}=\frac{96.75}{266.76}=.363$$

零次相關係數既經求得,即可據以計算一次淨相關係數。茲以玉蜀黍收穫量(X_1)爲因變數,六月平均溫度(X_2)爲自變數,而假定七月溫度(X_3)相同,計算 $r_{12\cdot3}$ 如下:

$$r_{12\cdot3}=\frac{r_{12}-r_{13}\cdot r_{23}}{(1-r_{13}^2)^{\frac{1}{2}}(1-r_{23}^2)^{\frac{1}{2}}}$$

$$=\frac{-.473-(-.716\times.374)}{\sqrt{1-(-.716)^2}\cdot\sqrt{1-.374^2}}$$

$$=\frac{-.473-(-.2678)}{.6940\times.9290}$$

$$=\frac{-.2052}{.6447}$$

$$=-.3183$$

計算 $r_{14\cdot3}$, $r_{24\cdot3}$, $r_{13\cdot2}$, $r_{14\cdot2}$, $r_{34\cdot2}$, $r_{12\cdot4}$, $r_{13\cdot4}$, $r_{23\cdot4}$ 等之方法相同,茲將其結果列表如下,算式從略。

一次淨相關之係數如 $r_{12\cdot3}$,其結果等於 $r_{21\cdot3}$,$r_{23\cdot4}$ 等於 $r_{32\cdot4}$,$r_{34\cdot2}$ 等於 $r_{43\cdot2}$。上表已將 $r_{12\cdot3}$,$r_{23\cdot4}$ $r_{34\cdot2}$ 等求出,故 $r_{21\cdot3}$,$r_{32\cdot4}$ 無須重複計算。

計算各次淨相關係數,以求 $(1-r^2)^{\frac{1}{2}}$ 之一步手續較繁,所好 r 之數值不出 0 與 ±1 之範圍,故統計學家製就一表,$(1-r^2)^{\frac{1}{2}}$ 之數值由表中一查即得,無

表四十二*

玉蜀黍收穫量與六,七,八各月平均溫度一次淨相關係數計算表

零次係數		$(1-r^2)^{\frac{1}{2}}$	分子中兩零次相關係數之乘積	分子之全部	分母	一次淨相關係數	
附字	r					附字	r
12	−.473		−.2678	−.2025	.6447	12.3	−.3183
13	−.716	.6940					
23	.374	.9290					
14	−.483		−.2048	−.2782	.6642	14.3	−.4188
13	−.716	.6940					
43	.286	.9570					
24	.363		+.1070	+.2560	.8891	24.3	+.2879
23	.374	.9290					
43	.286	.9570					
13	−.716		−.1769	−.5391	.8200	13.2	−.6574
12	−.473	.8827					
32	.374	.9290					
14	−.483		−.1717	−.3113	.8236	14.2	−.3780
12	−.473	.8827					
42	.363	.9330					
34	.286		+.1358	+.1502	.8668	34.2	+.1733
32	.374	.9290					
42	.363	.9330					
12	−.473		−.1753	−.2977	.8185	12.4	−.3637
14	−.483	.8773					
24	.363	.9330					
13	−.716		−.1381	−.5779	.8396	13.4	−.6883
14	−.483	.8773					
34	.286	.9570					
23	.374		+.1038	+.2702	.8929	23.4	+.3026
24	.363	.9330					
34	.286	.9570					

* $(1-r^2)^{\frac{1}{2}}$ 係由表四十三查得,該表 r 僅兩位小數,與按三位小數計算之結果,稍有出入。

須逐一計算,茲將此表列下:

表四十三

按 r 之數值查 $\sqrt{1-r^2}$

r	$\sqrt{1-r^2}$	r	$\sqrt{1-r^2}$	r	$\sqrt{1-r^2}$	r	$\sqrt{1-r^2}$
.00	1.0000	.26	.9656	.51	.8602	.76	.6499
.01	.9999	.27	.9629	.52	.8542	.77	.6380
.02	.9998	.28	.9600	.53	.8480	.78	.6258
.03	.9995	.29	.9570	.54	.8417	.79	.6131
.04	.9993	.30	.9539	.55	.8352	.80	.6000
.05	.9987	.31	.9507	.56	.8285	.81	.5864
.06	.9982	.32	.9474	.57	.8216	.82	.5724
.07	.9975	.33	.9440	.58	.8146	.83	.5578
.08	.9968	.34	.9404	.59	.8074	.84	.5472
.09	.9959	.35	.9367	.60	.8000	.85	.5268
.10	.9950	.36	.9330	.61	.7924	.86	.5103
.11	.9939	.37	.9290	.62	.7846	.87	.4931
.12	.9928	.38	.9250	.63	.7766	.88	.4750
.13	.9915	.39	.9208	.64	.7684	.89	.4560
.14	.9902	.40	.9165	.65	.7599	.90	.4359
.15	.9887	.41	.9121	.66	.7513	.91	.4146
.16	.9871	.42	.9075	.67	.7424	.92	.3919
.17	.9854	.43	.9028	.68	.7332	.93	.3676
.18	.9837	.44	.8980	.69	.7238	.94	.3412
.19	.9818	.45	.8930	.70	.7141	.95	.3122
.20	.9798	.46	.8879	.71	.7042	.96	.2800
.21	.9777	.47	.8827	.72	.6940	.97	.2431
.22	.9755	.48	.8773	.73	.6834	.98	.1990
.23	.9732	.49	.8717	.74	.6726	.99	.1411
.24	.9708	.50	.8660	.75	.6614	1.00	.0000
.25	.9682						

一次淨相關係數既經求得,即可從事計算二次淨相

關係數。本例二次淨相關係數有三：$r_{12\cdot34}$，$r_{13\cdot24}$，$r_{14\cdot23}$，但爲校對錯誤起見，每一二次淨相關係數可從不同之一次淨相關係數求出；蓋$r_{12\cdot34}$之公式原爲：

$$r_{12\cdot34}=\frac{r_{12\cdot3}-r_{14\cdot3}\cdot r_{24\cdot3}}{(1-r_{14\cdot3}^2)^{\frac{1}{2}}(1-r_{24\cdot3}^2)^{\frac{1}{2}}}$$

然亦可書作：

$$r_{12\cdot43}=\frac{r_{12\cdot4}-r_{13\cdot4}\cdot r_{23\cdot4}}{(1-r_{13\cdot4}^2)^{\frac{1}{2}}(1-r_{23\cdot4}^2)^{\frac{1}{2}}}$$

因$r_{12\cdot34}$與$r_{12\cdot43}$之意義相同，其結果亦無二致也。茲將計算$r_{12\cdot34}$之算式列下，其$r_{13\cdot24}$與$r_{14\cdot23}$各步計算之結果，見表四十四，算式從略。

$$r_{12\cdot34}=\frac{r_{12\cdot3}-r_{14\cdot3}\cdot r_{24\cdot3}}{(1-r_{14\cdot3}^2)^{\frac{1}{2}}(1-r_{24\cdot3}^2)^{\frac{1}{2}}}$$

$$=\frac{-.3183-(-.4188\times.2879)}{.9075\times.9570}$$

$$=\frac{-.3183+.1206}{.8685}$$

$$=\frac{-.1977}{.8685}$$

$$=-.2276$$

表四十四*

玉蜀黍收穫量與六，七，八各月平均溫度二次淨相關係數計算表

一次淨相關係數		$(1-r^2)^{\frac{1}{2}}$	分子中兩一次淨相關係數之乘積	分子之全部	分母	二次淨相關係數	
附字	r					附字	r
12.3	—.3183		—.1206	—.1977	.8685	12.34	—.2276
14.3	—.4188	.9075					
24.3	+.2879	.9570					
13.2	—.6574		—.0655	—.5919	.9115	13.24	—.6494
14.2	—.3780	.9250					
34.2	+.1733	.9854					
14.2	—.3780		—.1139	—.2641	.7403	14.23	—.3567
13.2	—.6574	.7513					
43.2	+.1733	.9854					
12.4	—.3637		—.2083	—.1554	.6904	12.34	—.2251
13.4	—.6883	.7238					
23.4	+.3026	.9539					
13.4	—.6883		—.1101	—.5782	.8900	13.24	—.6497
12.4	—.3637	.9330					
32.4	+.3626	.9539					
14.3	—.4188		—.0916	—.3272	.9067	14.23	—.3609
12.3	—.3183	.9474					
42.3	+.2879	.9570					

* $(1-r^2)^{\frac{1}{2}}$ 係由表四十三查得,該表 r 僅兩位小數,與按三位小數計算之結果,稍有出入。

二次淨相關係數之意義前已言之,茲將簡相關之係數與二次淨相關之係數并列於下,吾人試一加

比較,則於玉蜀黍收穫量與六,七,八各月平均溫度之關係,當更了然矣。

$$r_{12} = -.473 \qquad r_{12.34} = -.2276$$
$$r_{13} = -.716 \qquad r_{13.24} = -.6494$$
$$r_{14} = -.483 \qquad r_{14.23} = -.3567$$

玉蜀黍收穫量與六月平均溫度之簡相關係數顯較二次淨相關係數爲高,其故蓋由六月之平均溫度與七,八兩月之平均溫度均有相當關係,且爲正相關故六月溫度在簡相關中其地位較固有者爲重要,然而非實情也。玉蜀黍收穫量與七,八兩月之二次淨相關係數亦較其簡相關係數爲小,理由相同,茲不贅述;惟無論簡相關或淨相關,七月平均溫度與玉蜀黍收穫量之係數均較其他二月平均溫度與玉蜀黍收穫量之係數爲大,於此可見在本例中七月溫度於玉蜀黍之生長,實較六八兩月爲重要。

淨相關係數之求法已如上述,茲進而一論淨相關之迴歸方程式;蓋淨相關係數之所示,僅爲相關之程度,如欲根據已知各種變數之數值以推測另一變數極可能之數値,則非列迴歸方程式不可。淨相關之迴歸方程式,亦視變數之多寡而定,如變數爲三(一次淨相關),其迴歸方程式應爲:

$$x_1 = b_{12\cdot3}x_2 + b_{13\cdot2}x_3$$

式中:x_1為因變數中各項與其算術平均數之離中差

x_2為第一自變數中各項與其算術平均數之離中差

x_3為第二自變數中各項與其算術平均數之離中差

$b_{12\cdot3}$為因變數與第一自變數之淨迴歸係數

$b_{13\cdot2}$為因變數與第二自變數之淨迴歸係數

如變數為四(二次淨相關),則其迴歸方程式應為:

$$x_1 = b_{12\cdot34}x_2 + b_{13\cdot24}x_3 + b_{14\cdot23}x_4$$

式中:x_1為因變數中各項與其算術平均數之離中差

x_2為第一自變數中各項與其算術平均數之離中差

x_3為第二自變數中各項與其算術平均數之離中差

x_4為第三自變數中各項與其算術平均數之離中差

$b_{12\cdot34}$為因變數與第一自變數之淨迴歸係數

$b_{13\cdot24}$為因變數與第二自變數之淨迴歸係數

$b_{14\cdot23}$為因變數與第三自變數之淨迴歸係數

惟從上式吾人僅可知當各種變數變動若干單位時,影響於某種變數若干,尚不能據各種變數之數値直接推知某種變數之數値。如欲直接推測,須將上式略略改變;蓋

$$x_1 = (X_1 - \overline{X}_1)$$
$$x_2 = (X_2 - \overline{X}_2)$$
$$x_3 = (X_3 - \overline{X}_3)$$
$$x_4 = (X_4 - \overline{X}_4)$$

故

$$x_1 = b_{12\cdot3}x_2 + b_{13\cdot2}x_3$$

可化為:

$$(X_1 - \overline{X}_1) = b_{12\cdot3}(X_2 - \overline{X}_2) + b_{13\cdot2}(X_3 - \overline{X}_3)$$

又

$$x_1 = b_{12\cdot34}x_2 + b_{13\cdot24}x_3 + b_{14\cdot23}x_4$$

可化為:

$$(X_1-\overline{X}_1)=b_{12\cdot34}(X_2-\overline{X}_2)+b_{13\cdot24}(X_3-\overline{X}_3)$$
$$+b_{14\cdot23}(X_4-\overline{X}_4)$$

式中：X 爲各變數中各項之數值，$\overline{X}$ 爲各種變數之算術平均數，移項則三種變數之方程式爲：

$$X_1=b_{12\cdot3}X_2+b_{13\cdot2}X_3+a$$

式中：a 爲本方程式之常數等於

$$\overline{X}_1-b_{12\cdot3}\overline{X}_2-b_{13\cdot2}\overline{X}_3$$

四種變數之方程式爲：

$$X_1=b_{12\cdot34}X_2+b_{13\cdot24}X_3+b_{14\cdot23}X_4+a$$

式中：a 爲本方程式之常數等於

$$\overline{X}_1-b_{12\cdot34}\overline{X}_2-b_{13\cdot24}\overline{X}_3-b_{14\cdot23}\overline{X}_4$$

五種變數，六種變數之方程式可以類推，茲不贅述。惟欲作推測，必須先將迴歸係數求出，如 $b_{12\cdot3}$, $b_{13\cdot2}$, $b_{12\cdot34}$, $b_{13\cdot24}$ 及 $b_{14\cdot23}$ 之類。求淨迴歸係數之公式亦視變數種類之多寡而異，如爲三種變數，其因變數與第一自變數淨迴歸係數之公式應爲：

$$b_{12\cdot3}=r_{12\cdot3}\times\frac{\sigma_{1\cdot23}}{\sigma_{2\cdot13}}$$

式中：$b_{12\cdot3}$ 爲含有三種變數淨迴歸係數之符號（因變數與第一自變數之淨迴歸）

$r_{12\cdot3}$ 爲一次淨相關係數（因變數與第一自變數之淨相關）

$\sigma_{1.23}$ 為根據 X_2 及 X_3 推測 X_1 之標準誤亦名淨標準誤等於 $\sigma_1\sqrt{1-r_{12}^2}\sqrt{1-r_{13.2}^2}$

$\sigma_{2.13}$ 為根據 X_1 及 X_3 推測 X_2 之標準誤等於 $\sigma_2\sqrt{1-r_{12}^2}\sqrt{1-r_{23.1}^2}$

$b_{13.2}$ 之公式可以類推。

如為四種變數,其因變數與第一自變數淨迴歸係數之公式應為:

$$b_{12.34}=r_{12.34}\times\frac{\sigma_{1.234}}{\sigma_{2.134}}$$

式中:$b_{12.34}$ 為含有四種變數淨迴歸係數之符號(因變數與第一自變數之淨迴歸)

$r_{12.34}$ 為二次淨相關係數(因變數與第一自變數之淨相關)

$\sigma_{1.234}$ 為根據 X_2, X_3 及 X_4 推測 X_1 之標準誤等於 $\sigma_1\sqrt{1-r_{12}^2}\sqrt{1-r_{13.2}^2}\sqrt{1-r_{14.23}^2}$

$\sigma_{2.134}$ 為根據 X_1, X_3 及 X_4 推測 X_2 之標準誤等於 $\sigma_2\sqrt{1-r_{12}^2}\sqrt{1-r_{23.1}^2}\sqrt{1-r_{24.13}^2}$

由此類推,可知含有 n 個變數之淨迴歸係數,其普通公式應為:

$$b_{12.345\cdots\cdots n}=r_{12.345\cdots\cdots n}\times\frac{\sigma_{1.2345\cdots n}}{\sigma_{2.1345\cdots n}}$$

推測的標準誤或淨標準誤之普通公式應為:

$$\sigma_{1.2345\cdots n}=\sigma_1\sqrt{1-r_{12}^2}\sqrt{1-r_{13.2}^2}\sqrt{1-r_{14.23}^2}$$

$$\times\sqrt{1-r^2_{15.234}}\cdots\cdots,\sqrt{1-r^2_{1n.2345\cdots(n-1)}}$$

表三十三之例含有四種變數,欲作推測,須先求出三個迴歸係數,$b_{12.34}$,$b_{13.24}$及$b_{14.23}$。按公式計算,

$$b_{12.34}=-.3887$$

$$b_{13.24}=-1.545$$

$$b_{14.23}=-.6388$$

茲將其算式列下:

$$b_{12.34}=r_{12.34}\times\frac{\sigma_{1.234}}{\sigma_{2.134}}$$

$$=-.2276\times\frac{4.27}{2.50}=-.2276\times1.708$$

$$=-.3887$$

$$r_{12.34}=-.2276$$

$$\sigma_{1.234}=\sigma_1\sqrt{1-r^2_{12}}\sqrt{1-r^2_{13.2}}\sqrt{1-r^2_{14.23}}$$

$$=6.9\times\sqrt{1-(-.473)^2}\sqrt{1-(-.6574)^2}$$

$$\times\sqrt{1-(-.3567)^2}$$

$$=6.9\times.8827\times.7513\times.9330$$

$$=4.27$$

$$\sigma_{2.134}=\sigma_2\sqrt{1-r^2_{23}}\sqrt{1-r^2_{24.3}}\sqrt{1-r^2_{12.34}}$$

$$=2.89\sqrt{1-(-.374)^2}\sqrt{1-(-.2879)^2}$$

$$\times\sqrt{1-(-.2276)^2}$$

$$=2.89\times.9290\times.9570\times.9732=2.50$$

$$b_{13\cdot24}=r_{13\cdot24}\times\frac{\sigma_{1\cdot324}}{\sigma_{3\cdot124}}$$

$$=-.6494\times\frac{4.27}{1.795}=-.6494\times2.379$$

$$=-1.545$$

$$r_{13\cdot24}=-.6494$$

$$\sigma_{1\cdot324}=4.27$$

$$\sigma_{3\cdot124}=\sigma_3\sqrt{1-r_{23}^2}\sqrt{1-r_{34\cdot2}^2}\sqrt{1-r_{13\cdot24}^2}$$

$$=2.58\sqrt{1-(.374)^2}\sqrt{1-(.1733)^2}$$

$$\times\sqrt{1-(-.6494)^2}$$

$$=2.58\times.9290\times.9854\times.7599$$

$$=1.795$$

$$b_{14\cdot23}=r_{14\cdot23}\times\frac{\sigma_{1\cdot423}}{\sigma_{4\cdot123}}$$

$$=-.3567\times\frac{4.27}{2.384}$$

$$=-.3567\times1.791=-.6388$$

$$r_{14\cdot23}=-.3567$$

$$\sigma_{1\cdot423}=4.27$$

$$\sigma_{4\cdot123}=\sigma_4\sqrt{1-r_{34}^2}\sqrt{1-r_{24\cdot3}^2}\sqrt{1-r_{14\cdot23}^2}$$

$$=2.79\sqrt{1-(.286)^2}\sqrt{1-(.2879)^2}$$

$$\times\sqrt{1-(-.3567)^2}$$

$$=2.79\times.9570\times.9570\times.9330$$

$$=2.384$$

將迴歸係數代入方程式,則得:

$$(X_1-\overline{X}_1)=-.3887(X_2-\overline{X}_2)-1.545(X_3-\overline{X}_3)$$
$$-.6388(X_4-\overline{X}_4)$$

已知 $\overline{X}_1=20°$,$\overline{X}_2=73.5°$,$\overline{X}_3=78.22°$,$\overline{X}_4=77.39°$(參看表三十六,三十七,三十八,三十九),故本例之迴歸方程式可化為:

$$(X_1-20)=-.3887(X_2-73.5)-1.545(X_3-78.22)$$
$$-.6388(X_4-77.39)$$

$$X_1=20+28.57+120.85+49.44-.3887X_2-1.545X_3$$
$$-.6388X_4$$

$$X_1=218.86-.3887X_2-1.545X_3-.6388X_4$$

迴歸係數既已求得,則根據上列方程式吾人於四種變數中如已知任何三種變數之數值,不難推知其他一種變數之數值。茲舉例以明之,如本例1922年六月平均溫度為75.2°,七月平均溫度為77°,八月平均溫度為80.1°,則該年玉蜀黍每畝最可能之收穫量應為19.38蒲式耳,算式如下:

$$X_1=218.86-.3887\times75.2-1.545\times77-.6388\times80.1$$
$$=218.86-29.23-118.97-51.17$$

$=218.86-199.37$

$=19.49$

查1922年實際玉蜀黍每畝之收穫量爲19.3蒲式耳,而推測之結果爲19.49蒲式耳,相差極微。此項推測之結果,其可靠之程度若何,可以推測的標準誤(Standard Error of Estimate)測定之。標準誤之求法與淨標準誤同,在含有四種變數之淨相關中,如所推測者爲第一種變數,則推測的標準誤爲$\sigma_{1.234}$;如所推測者爲第二種變數,則推測之標準誤爲$\sigma_{2.134}$,餘可遞推。本例推測的標準誤爲$\sigma_{1.234}$,其數值應爲:

$$\sigma_{1.234}=\sigma_1\sqrt{1-r_{12}^2}\sqrt{1-r_{13.2}^2}\sqrt{1-r_{14.23}^2}$$

$$=6.9\times.8827\times.7513\times.9330=4.27$$

意卽某年六月平均溫度如爲75.2°,七月平均溫度爲77°,八月平均溫度爲80.1°,則該年玉蜀黍之實際收穫量不出(19.49±4.27)蒲式耳範圍之機會爲百分之六八·二六;不出(19.49±2×4.27)蒲式耳範圍之機會爲百分之九五·四五,標準誤以外,吾人亦可以機差測定此項推測可靠之程度,其式如下:

$$P.E._{1.234}=.6745\sigma_{1.234}$$

$$=.6745\times4.27$$

$$=2.880$$

意卽某年六月平均溫度如爲75.2°,七月平均溫度爲 77°,八月平均溫度爲80.1°,則該年玉蜀黍之實際收穫量不出(19.38±2.880)蒲式耳範圍之機會爲百分之五十。

淨相關之迴歸方程式及迴歸係數之求法略具於斯,以下請論複相關。

第三節　複相關之計算法

複相關乃一種因變數與多種自變數總合之關係,前已言之。複相關係數之大小,常視因變數之標準差與其推測的標準誤數値之關係而定,其公式如下:

$$R_{1\cdot2345\cdots\cdots n}=\sqrt{\frac{1-\sigma^2_{1\cdot2345\cdots\cdots n}}{\sigma^2_1}}$$

式中:$R_{1\cdot2345\cdots\cdots n}$ 爲 n 種變數複相關係數之符號

$\sigma_{1\cdot2345\cdots\cdots n}$ 爲因變數之推測的標準誤

σ_1爲因變數之標準差

本例計有四種變數,複相關之係數可按下式計算,結果爲.78。

$$R_{1\cdot234}=\sqrt{1-\frac{\sigma^2_{1\cdot234}}{\sigma^2_1}}$$

$$=\sqrt{1-\frac{(4.27)^2}{(6.90)^2}}$$

$$= \sqrt{1-\frac{18.23}{47.61}}$$

$$= \sqrt{1-.383}$$

$$= \sqrt{.617}$$

$$= .785$$

惟

$$\sigma_{1.234}=\sigma_1\sqrt{1-r^2_{12}}\sqrt{1-r^2_{13.2}}\sqrt{1-r^2_{14.23}}$$

代入上式,則計算複相關係數之公式可化爲:

$$R_{1.234}=\sqrt{1-\left[(1-r^2_{12})(1-r^2_{13.2})(1-r^2_{14.23})\right]}$$

將 r_{12},$r_{13.2}$,$r_{14.23}$ 之數值代入,結果亦爲.78,算式如下:

$$R_{1.234}=\sqrt{1-\left\{\left[1-(-.473)^2\right]\left[1-(-.6574)^2\right]\left[1-(-.3567)^2\right]\right\}}$$

$$=\sqrt{1-\left\{\left[1-.2237\right]\left[1-.4322\right]\left[1-.1272\right]\right\}}$$

$$=\sqrt{1-\left\{.7763\times.5678\times.8728\right\}}$$

$$=\sqrt{1-.384}=\sqrt{.616}$$

$$= .785$$

按此式計算複相關係數,吾人可隨意變換公式之形式,極便校對。蓋 $R_{1.234}=R_{1.324}=R_{1.342}=R_{1.243}=R_{1.432}=R_{1.423}$, 複相關係數可從下列六式中任何一式求得也。

$$R_{1.234}=\sqrt{1-\left\{\left[1-r_{12}^2\right]\left[1-r_{13.2}^2\right]\left[1-r_{14.23}^2\right]\right\}}$$

$$R_{1.324}=\sqrt{1-\left\{\left[1-r_{13}^2\right]\left[1-r_{12.3}^2\right]\left[1-r_{14.23}^2\right]\right\}}$$

$$R_{1.342}=\sqrt{1-\left\{\left[1-r_{13}^2\right]\left[1-r_{14.3}^2\right]\left[1-r_{12.34}^2\right]\right\}}$$

$$R_{1.243}=\sqrt{1-\left\{\left[1-r_{12}^2\right]\left[1-r_{14.2}^2\right]\left[1-r_{13.24}^2\right]\right\}}$$

$$R_{1.432}=\sqrt{1-\left\{\left[1-r_{14}^2\right]\left[1-r_{13.4}^2\right]\left[1-r_{12.34}^2\right]\right\}}$$

$$R_{1.423}=\sqrt{1-\left\{\left[1-r_{14}^2\right]\left[1-r_{12.4}^2\right]\left[1-r_{13.24}^2\right]\right\}}$$

複相關係數前常不冠以正負號,因因變數與各自變數之關係有正有負,其總合之關係自難與以一定符號。惟當各自變數與因變數之關係均爲正或均爲負時,則 R 亦不妨冠以正號或負號。本例玉蜀黍收穫量與六,七,八三月溫度之關係均爲負,故複相關係數當爲 —.785。此外吾人當注意者,淨相關與複相關之方法,在應用上有兩種限制:第一,淨相關與複相關之方法,以應用於直線相關之材料爲宜。若干種變數中任何兩種變數之關係如爲非直線的,則所得之係數減小,而根據迴歸方程式推測之結果將不甚可靠。第二,材料須適當的多,如有多種變數,則材料尤須增多。設所研究之問題包含多種變數而每種變數之材料有限,則結果往往易得遠較實際關係爲大之係數,

吾人如能於應用時恪遵上述限制,則淨相關與複相關之方法,在統計分析上誠極有用之工具也。

第十一章 常態曲線

宇宙間任何一羣事物或現象之分配,頗有集中之趨勢,前已言之,意即適中者居多,極端者居少。如人類之智慧,上智與下愚均少,而平庸者多;又如體長,極長者與侏儒均少,而適中者居多。此種集中之趨勢可以對稱的鐘形曲線(Symmetrical bell shaped curve)表示之,蓋即所謂常態曲線(Normal Curve of Error)。常態曲線亦名機率曲線,又名高氏曲線(Gaussian curve),以高氏爲發現此曲線程式之第一人也。昔時統計學家恆認常態曲線可以表示一切分配之趨勢,此說自不盡然,因常態曲線之外,尚有他種曲線可用以表示某種分配之特殊情形,但常態曲線爲各種曲線中最重要之一種,要爲不可掩之事實。

欲明瞭常態曲線之理論,請先略述機率之原理(Theorems in Probability),因機率大小之分配,由此曲線而表明也。

第一節 簡單機率之法則

機率者,表示一事成敗機會之比率也。設一事有 n 種方式發現,其中 a 種爲成功的, b 種爲失敗的,則成功之機率爲$\frac{a}{n}$,失敗之機率爲 $\frac{b}{n}$. 在數學上通常

以 p 表示成功之機率，q 表示失敗之機率，列式如下：

$$p=\frac{a}{n}$$

$$q=\frac{b}{n}$$

$$\text{但}\ a+b=n$$

$$\therefore \frac{a}{n}+\frac{b}{n}=1$$

$$p+q=1$$

故機率為一比率 (ratio)，其分子為成功的或失敗的發現數目，分母為一切發現的總數，機率之大小恆在零與一之間，零表示絕對不可能，一則表示必然實現 (Certainty)，茲舉一簡單之例以明之。設擲幣一枚，其可能的發現有二，一為面，一為背，如面為吾人所希望之發現，則成功之機率為$\frac{1}{2}$，失敗之機率亦為$\frac{1}{2}$。設擲一骰，其可能之發現有六，如六點一面為吾人所希望之發現，則成功之機率為$\frac{1}{6}$，而失敗之機率為$\frac{5}{6}$。

惟一事之希望的發現如有種種各不相關之不同方式，則其發現之機率應為各項機率相加之和。例如一囊內有紅球二十，白球十六，黑球十四，每次只取一球則取得紅球之機率為 $\frac{20}{50}$，取得白球之機率為

(20+16+14)

$\frac{16}{50}$，取得黑球之機率為 $\frac{14}{50}$。設吾人之目的在一取而得紅球或白球，則成功之機率為$\frac{20}{50}+\frac{16}{50}=\frac{36}{50}$。

上例係就一單純事件而言,若吾人所求者為複合事件之機率,則應將各項機率相乘而求其積。如擲一骰,第一擲之結果不影響第二擲之結果,彼此獨立者謂之單純事件。若吾人取骰一粒,連擲二次,問第一擲得一點,第二擲得兩點之機率如何,則應以第一擲得一點之機率與第二擲得兩點之機率相乘:

$$p = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$$

蓋第一擲縱得一點,設第二擲非兩點;或第二擲得兩點,而第一擲非一點,均非吾人所希望之結果。第一擲與第二擲之結果互相關聯,故謂之複合事件,其機率應為各單獨事件機率之積(The probability of a compound event is the product of the probabilities of the separate events)。

通常吾人計算一事之機率,加法與乘法每須同時並用。例如擲骰二粒(為便於解說計,假定一為甲,一為乙),得到五點之機率如何?吾人已知每骰有六面,下列各對中任何二面組合均為五。

甲骰	乙骰
1	4
2	3
3	2

4　　　　1

甲骰一擲而得一點之機率爲$\frac{1}{6}$,乙骰一擲而得四點之機率亦爲$\frac{1}{6}$,但兩骰同擲甲得一而乙得四之機率則爲$\frac{1}{36}\left(\frac{1}{6}\times\frac{1}{6}\right)$。因其他三種組合之結果均爲五點,故成功之機率爲$\frac{4}{36}$或$\frac{1}{9}$:

$$p=\frac{1}{36}+\frac{1}{36}+\frac{1}{36}+\frac{1}{36}=\frac{4}{36}=\frac{1}{9}$$

又如吾人問擲骰二粒,至少得五點之機率如何,則結果大不相同,蓋旣云至少五點,則六點,七點乃至十二點均在其內,茲計算其機率如下:

兩骰擲得　五　點之機率$=\frac{4}{36}$

兩骰擲得　六　點之機率$=\frac{5}{36}$

兩骰擲得　七　點之機率$=\frac{6}{36}$

兩骰擲得　八　點之機率$=\frac{5}{36}$

兩骰擲得　九　點之機率$=\frac{4}{36}$

兩骰擲得　十　點之機率$=\frac{3}{36}$

兩骰擲得十一點之機率$=\frac{2}{36}$

兩骰擲得十二點之機率$=\frac{1}{36}$

機率之總和$=\frac{30}{36}$

故擲得五點或五點以上之機率爲$\frac{30}{36}$或$\frac{5}{6}$。細察上列各項機率,可知擲骰兩粒,以擲得七點之機率爲最大,兩端如擲得八,九,十,十一,十二及六點,五點之機率均逐漸減小,此項機率之分配,即可以常態曲線表示之。但各項機率之大小(指複雜事件之機率而言),當各單獨事件之p等於q時,可利用二項開展式(Binomial expansion)求之,無須一一計算。例如取幣二枚同時擲之(爲便於解說,假定二幣一爲甲,一爲乙),其可能之發現有四:

甲乙	甲乙	甲乙	甲乙
面面	面背	背面	背背

惟合而觀之,祇有三種結果,以一背一面之機率最大,最有實現之機會,蓋兩面之機率爲$\frac{1}{4}$,兩背之機率爲$\frac{1}{4}$,而一面一背之機率爲$\frac{1}{2}$。又如擲幣三枚(甲,乙,丙),其可能之發現有八:

甲乙丙	甲乙丙	甲乙丙	甲乙丙
面面面	面面背	面背面	面背背
甲乙丙	甲乙丙	甲乙丙	甲乙丙
背面背	背面面	背背面	背背背

惟合而觀之,祇有四種結果,其機率如下:

三幣俱面之機率$=\frac{1}{8}$

二面一背之機率$=\frac{3}{8}$

一面二背之機率$=\frac{3}{8}$

三幣俱背之機率$=\frac{1}{8}$

凡此種種之機率,吾人無須逐一計算,但以p與q表示成功及失敗之機率代入下式而開展之,即得各項之機率:

$$(p+q)^n$$

式中:p為成功之機率

q為失敗之機率

n為事件之數目,在擲幣試驗中,即為幣數

在擲幣試驗,得面與背之機率均為$\frac{1}{2}$,擲幣二枚,則依二項開展式,各項結果之機率應為:

$$\left(\frac{1}{2}+\frac{1}{2}\right)^2=\frac{1}{4}+\frac{1}{2}+\frac{1}{4}$$

意即兩幣俱面或兩幣俱背之機率各為$\frac{1}{4}$,一背一面之機率為$\frac{1}{2}$,如擲幣三枚,則各項結果之機率應為:

$$\left(\frac{1}{2}+\frac{1}{2}\right)^3=\frac{1}{8}+\frac{3}{8}+\frac{3}{8}+\frac{1}{8}$$

意卽三幣俱面或三幣俱背之機率各爲$\frac{1}{8}$，一面兩背或一背兩面之機率各爲$\frac{3}{8}$。

設吾人不僅欲知擲幣一次時各種結果之機率，且欲知連擲若干次時各種結果之可能的次數(按卽機遇次數亦名理論次數)，可依下式求之：

$$N(p+q)^n$$

式中：N 代表所擲次數，餘見前

故若擲幣三枚，連擲二百四十次，則得三面，二面一背，一面二背及三背之機遇次數爲：

$$240(p+q)^3=240\left(\frac{1}{2}+\frac{1}{2}\right)^3$$

$$=240\left(\frac{1}{8}+\frac{3}{8}+\frac{3}{8}+\frac{1}{8}\right)$$

$$=30+90+90+30$$

由上舉二例可知 n 爲偶數，則中間一項之機率或機遇次數爲最大；如 n 爲奇數，則中間相等二項之機率或機遇次數爲最大。既知一事每次各種發現之機率，吾人不難推求其若干次之機遇次數。取某種現象之機遇次數與實際次數相比較，則吾人之知識將擴大爲某種現象全部的，而不僅限於所取之樣本矣。

維爾頓(W. F. R. Weldon)氏嘗取骰十二粒作一試驗，以得一，二，三點各面爲失敗，四，五，六點各面爲成

功,共擲4096次。其實際次數見表四十五第二欄,其機遇次數,亦名理論次數(Theoretical frequencies)可以下列之二項開展式求之:

$$(p+q)^n = p^n + np^{n-1}q + \frac{n(n-1)}{1\cdot2}p^{(n-2)}q^2 + \frac{n(n-1)(n-2)}{1\cdot2\cdot3}p^{(n-3)}q^3 + \cdots\cdots q^n$$

蓋擲骰一枚得一,二,三各面之機率與得四,五,六各面之機率相等均為$\frac{1}{2}$,故將下式開展,即得各種結果之機遇次數,見表四十五第三欄:

$$4096\left(\frac{1}{2}+\frac{1}{2}\right)^{12} = 4096\left(\frac{1}{4096}+\frac{12}{4096}+\frac{66}{4096}+\frac{220}{4096}+\frac{495}{4096}+\frac{792}{4096}+\frac{924}{4096}+\frac{792}{4096}+\frac{495}{4096}+\frac{220}{4096}+\frac{66}{4096}+\frac{12}{4096}+\frac{1}{4096}\right)$$

表四十五

擲骰試驗之實際次數與理論次數之比較

1	2	3
得四,五,六點各面骰子數	實際次數	理論次數
0	0	1
1	7	12
2	66	66
3	198	220
4	430	495
5	731	792

6	948	924
7	847	792
8	536	495
9	257	220
10	71	66
11	11	12
12	0	1
	4096	4096

實際分配之算術平均數爲6.139,標準差爲1.712,至於理論分配之算術平均數及標準差,可以下列二式求之:

$$M=np$$

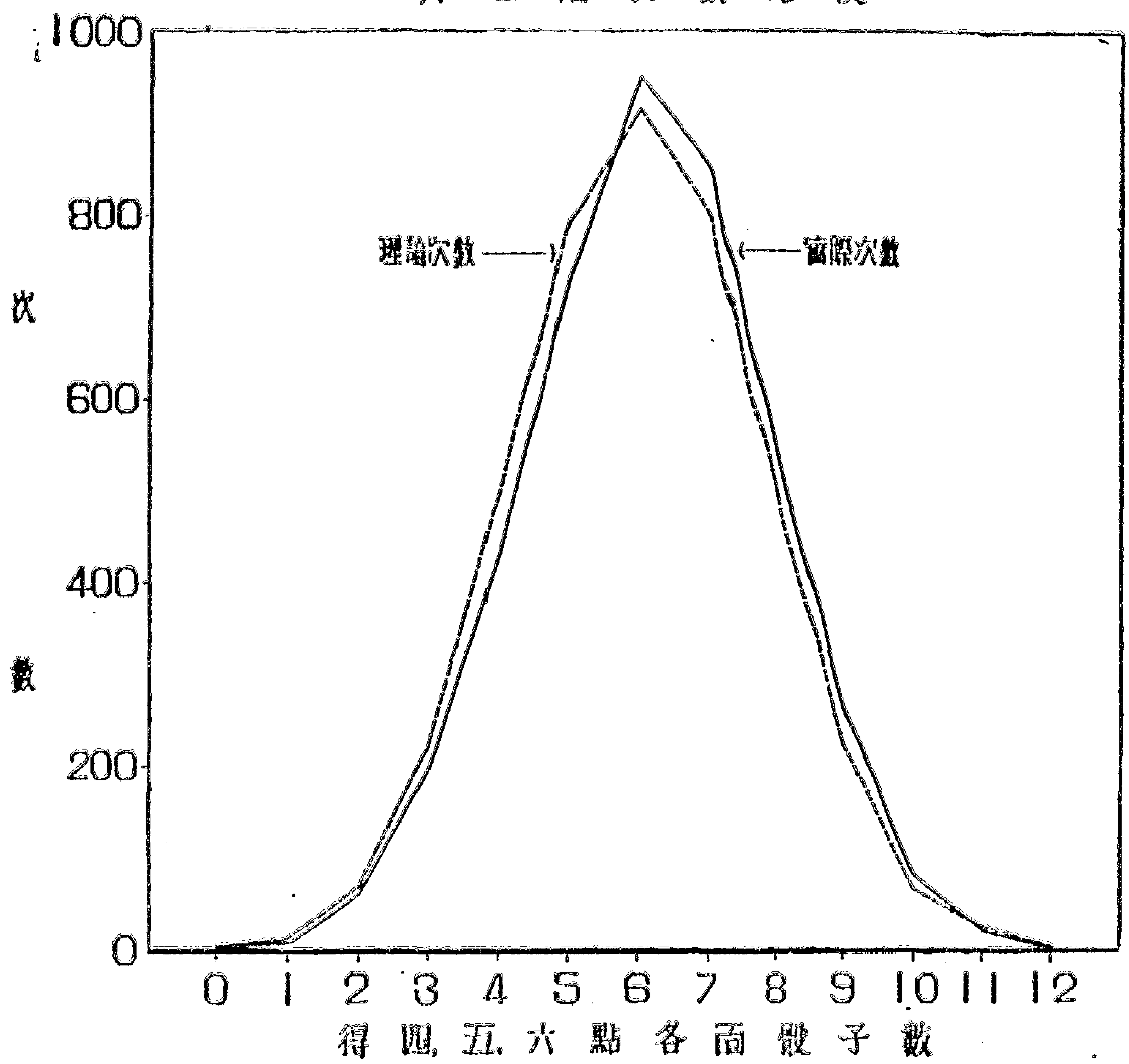
$$\sigma=\sqrt{npq}$$

本例$n=12, p=q=\frac{1}{2}$,故算術平均數爲 $6(12\times\frac{1}{2}=6)$,標準差爲$1.732\left(\sqrt{12\times\frac{1}{2}\times\frac{1}{2}}=\sqrt{3}=1.732\right)$。

茲以圖表示實際次數與理論次數,以覘二者之異同,設擲骰之次數儘量增加,則實際次數與理論次數相差愈微。

圖二十九

擲骰試驗之實際次數
與理論次數比較



上圖爲表示十二骰連擲4096次之次數多邊形,除底線外,計有十二邊,如骰數爲六,則多邊形應有六邊,骰數爲十四,則多邊形應有十四邊,圖形之邊與骰數相等,骰數愈多,則表示$(p+q)^n$開展之圖形邊數愈多,愈近於平滑,如n無限大,則所成之圖爲一常態曲線。惟常態曲線之繪製,另有較簡捷之方法,無須由二

項開展式計算,容於下節論之。

第二節　常態曲線之繪法

常態曲線之方程式有數種形式,但以下列一種最為普通:

$$y=y_o e^{-\frac{x^2}{2\sigma^2}}$$

式中:y_o為常態曲線下最高之縱線

e為納氏對數之底數(The base of Napierian logarithms)等於2.7182818

σ為標準差

x為曲線圖底線上由全體變數平均數所在之點至某一點之距離以標準差表示者(按即以標準差表示之離中差)

y為由底線上各點(平均數所在之點除外)向上引出縱線之高度

因y_o(常態曲線下最高之縱線)可以下式求之:

$$y_o=\frac{N}{\sigma\sqrt{2\pi}}$$

式中:N等於次數總數

σ為標準差

π等於3.1416

故常態曲線之方程式亦書作:

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

就上式可以略知常態曲線之性質,蓋當 x 等於零時,$e^{-\frac{x^2}{2\sigma^2}}=1, y=y_0$(常態曲線下縱線之最大値),故算術平均數與衆數合而爲一;在 y_0 左右相等距離一點之 y,其値相同,故 y_0 左右兩邊之次數相等且屬對稱,換言之,即算術平均數與中位數合而爲一。

按上式求常態曲線下各縱線之高度,手續異常繁瑣,因非先求 y_0 然後將各項 x(即以標準差表示之離中差)之數値代入公式逐項計算不可。爲簡省計算手續起見,統計學家嘗就公式中 y 與 y_0 之關係作成一表(見附錄一),利用此表則由底線上平均數所在點左右任何一點引出之縱線相當於 y_0 之分數可以一查即得。蓋常態曲線下之縱線常爲 y_0 之分數,且視距離 y_0 之遠近而有一定之比例。例如 x 爲 2σ,則由此點引出之縱線等於 $.13534y_0$,其算法如下:

$$y=y_0e^{-\frac{(2\sigma)^2}{2\sigma^2}}=y_0e^{-\frac{4\sigma^2}{2\sigma^2}}$$

$$=y_0e^{-2}=y_0\frac{1}{2.71828^2}$$

$$=y_0\times\frac{1}{7.389}=.13534y_0$$

如吾人已將 y_0 求得,則在 y_0 左右 2σ 距離處 y 之高度不難推知。蓋從附錄一常態曲線之縱線表 x/σ 下

2.0行一查卽知 $y=.13534y_o$，無須另行計算。

常態曲線公式中以求 $e^{\frac{-x^2}{2\sigma^2}}$ 之手續最繁，今 y 既可由表中直接查出，則吾人就一定之次數分配配合一常態曲線以與實際分配相比較，自無困難。配合常態曲線之手續計分八步如下：—

一、將實際分配繪爲次數多邊圖。

二、以實際分配平均數所在之一點爲常態曲線之中心 $(x=0)$。

三、沿圖形底線上中心點左右置若干以標準差爲單位之相等距離，如 $.1\sigma$, $.2\sigma$, $.3\sigma$, 或 $.01\sigma$, $.02\sigma$, $.03\sigma$ 等。距離分得愈小，則配合之曲線愈爲平滑。

四、按 $y_o=\frac{N}{\sigma\sqrt{2\pi}}$ 之公式求 y_o。

五、由平均數所在點豎立一縱線等於 y_o 之數值。

六、計算各項離中差而除以 σ，然後就附錄一表中一一查出 y/y_o，再與 y_o 之數值相乘。

七、就底線上所定各點依次豎立縱線等於第六步所得各項 y 之高度。

八、將所引各縱線之頂端連接，卽成一常態次數多邊形。如底線上所定距離間隔愈小，則圖形愈近於平滑。

配合常態曲線之方法已略述如上，茲就表四十六995

家電話用戶每年通話次數分配舉例以明之。

表四十六

995家電話用戶每年通話次數分配

1 通話次數	2 中點	3 家數	4 由假定平均數計算之離中差	5	6
	m	f	x'	fx'	fx'^{2}
0—— 50	25	0	—10	0	0
50 —— 100	75	1	— 9	— 9	81
100—— 150	125	9	— 8	— 72	576
150—— 200	175	19	— 7	—133	931
200—— 250	225	38	— 6	—228	1368
250—— 300	275	50	— 5	—250	1250
300—— 350	325	95	— 4	—380	1520
350 —— 400	375	85	— 3	—255	765
400—— 450	425	115	— 2	—230	460
450—— 500	475	132	— 1	—132	132
500—— 550	525	144	0	0	0
550—— 600	575	116	+1	116	116
600—— 650	625	79	+2	158	316
650—— 700	675	54	+3	162	486
700—— 750	725	31	+4	124	496
750—— 800	775	11	+5	55	275
800—— 850	825	5	+6	30	180
850—— 900	875	6	+7	42	294
900—— 950	925	2	+8	16	128
950——1000	975	1	+9	9	81
1000——1050	1025	1	+10	10	100
1050——1100	1075	1	+11	11	121
		995		—956	9,676

$$M = 525 + \frac{-956}{995} \times 50 = 525 - 48.04 = 476.96$$

$$\sigma = \sqrt{\frac{9,676}{995} - \left(\frac{-956}{995}\right)^2} = 2.953$$ (以組距爲單位之標準差)

上表計算者爲算術平均數及標準差,兩者旣經求得,卽可計算各組中點對於眞確算術平均數之離中差,然後一一以 σ 除之,是爲 x/σ。由附錄一表中再查出各縱線(y)與 y_0 之比率,於是計算 y_0 之値而與各項 y/y_0 一一相乘,卽得各組之理論次數;本例之理論次數如下:

表四十七

就995家電話用戶每年通話次數分配配合之常態曲線下各縱線高度計算表

通話次數 m	由眞確算術平均數計算之離中差(以組距爲單位) x	x/σ (σ 亦以組距爲單位)	y/y_0	理論次數(卽各縱線之高度) y_0
25	—9.0392	—3.06	.00819	1.10
75	—8.0392	—2.72	.02474	3.33
125	—7.0392	—2.38	.05888	7.91
175	—6.0392	—2.05	.12230	16.44
225	—5.0392	—1.71	.23176	31.15
275	—4.0392	—1.37	.39123	52.59
325	—3.0392	—1.03	.58834	79.08
375	—2.0392	—0.69	.78817	105.95
425	—1.0392	—0.35	.94055	126.43
475	—0.0392	—0.01	.99995	134.41
525	0.9608	0.32	.94856	127.51
575	1.9608	0.66	.80429	108.11
625	2.9608	1.00	.60653	81.53
675	3.9608	1.34	.40747	54.77
725	4.9608	1.68	.24385	32.78
775	5.9608	2.02	.13000	17.47
825	6.9608	2.36	.06174	8.30
875	7.9608	2.70	.02612	3.51
925	8.9608	3.03	.01015	1.36
975	9.9608	3.37	.00342	0.46
1025	10.9608	3.71	.00103	0.14
1075	11.9608	4.05	.00027	0.04
				994.37

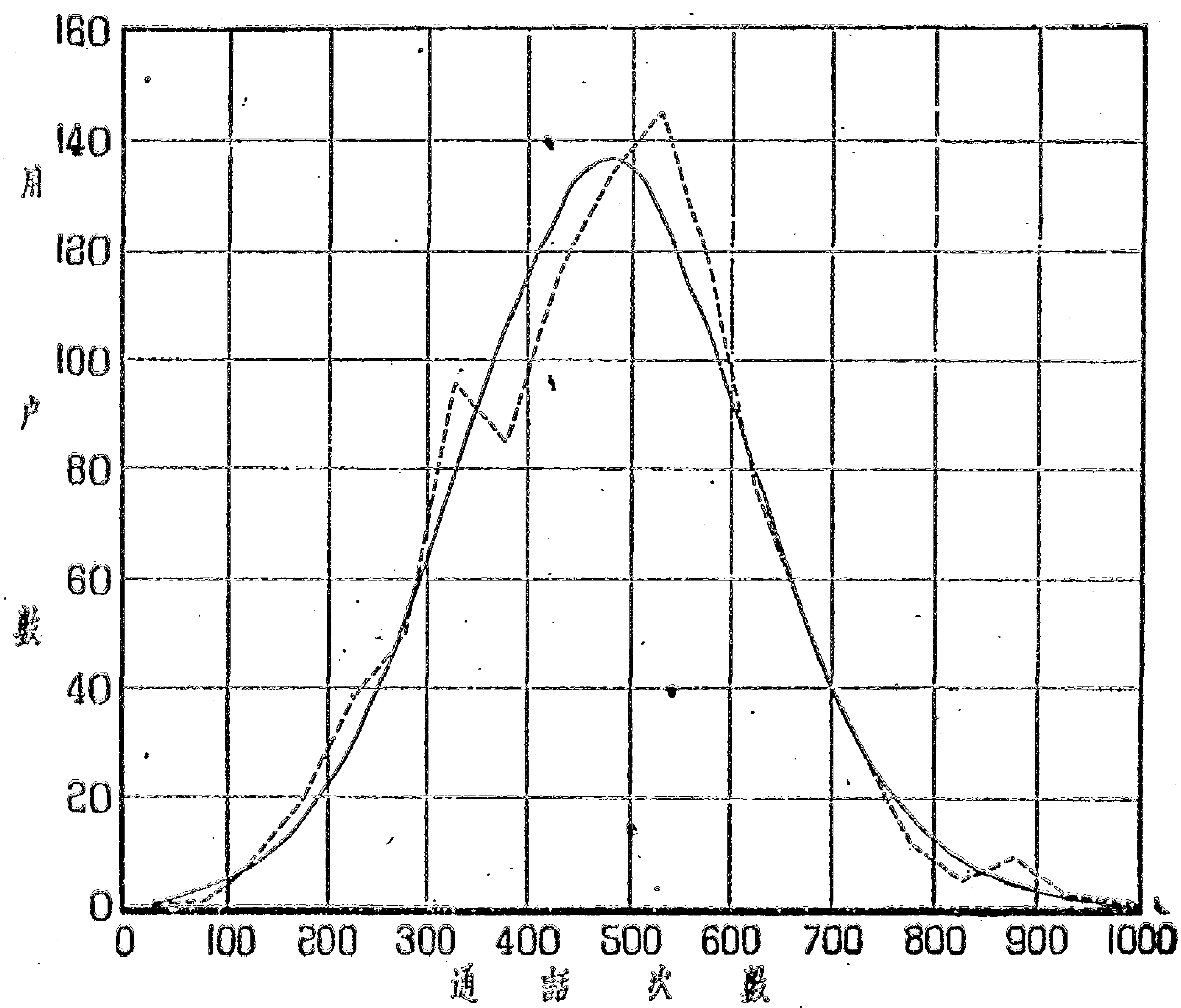
$$y_0 = \frac{995}{2.506628^* \times 2.953} = 134.42$$

計算y_0時,分母中之σ必須以組距爲單位,因必如此算得之理論次數與實際次數始有比較之可能。理論次數既經求得,即可據以配合常態曲線如下圖。

圖三十

995家電話用戶每年通話次數

常態分配圖



惟配合之曲線與依實際分配所繪之多邊形是

*$\sqrt{2\pi} = 2.506628$

否適合,尚有考核之必要(Testing the goodness of fit)。吾人就上圖觀察,知所配合之常態曲線頗與實際分配相合,但有數點,差異顯著。此項差異受取樣之影響乎?抑電話用戶通話次數本非常態之分配?二者之中,必居其一,吾人决不能不加以解答。如受取樣之影響,則家數增多,差異自可減小;如根本非常態分配,則不當配合常態曲線。欲判斷差異之原因屬於前者抑屬於後者,吾人可擇實際次數與理論次數差異最大數點查考其原因。然欲從事實際次數與理論次數之比較,必先有較精確之理論次數,而精確之理論次數必須按常態曲線下之面積計算。蓋常態曲線為近於平滑之曲線,以縱線高度表示各組之理論次數自不若以面積表示為精確也。常態曲線下面積之計算,容於下節論之。

第三節　常態曲線下之面積

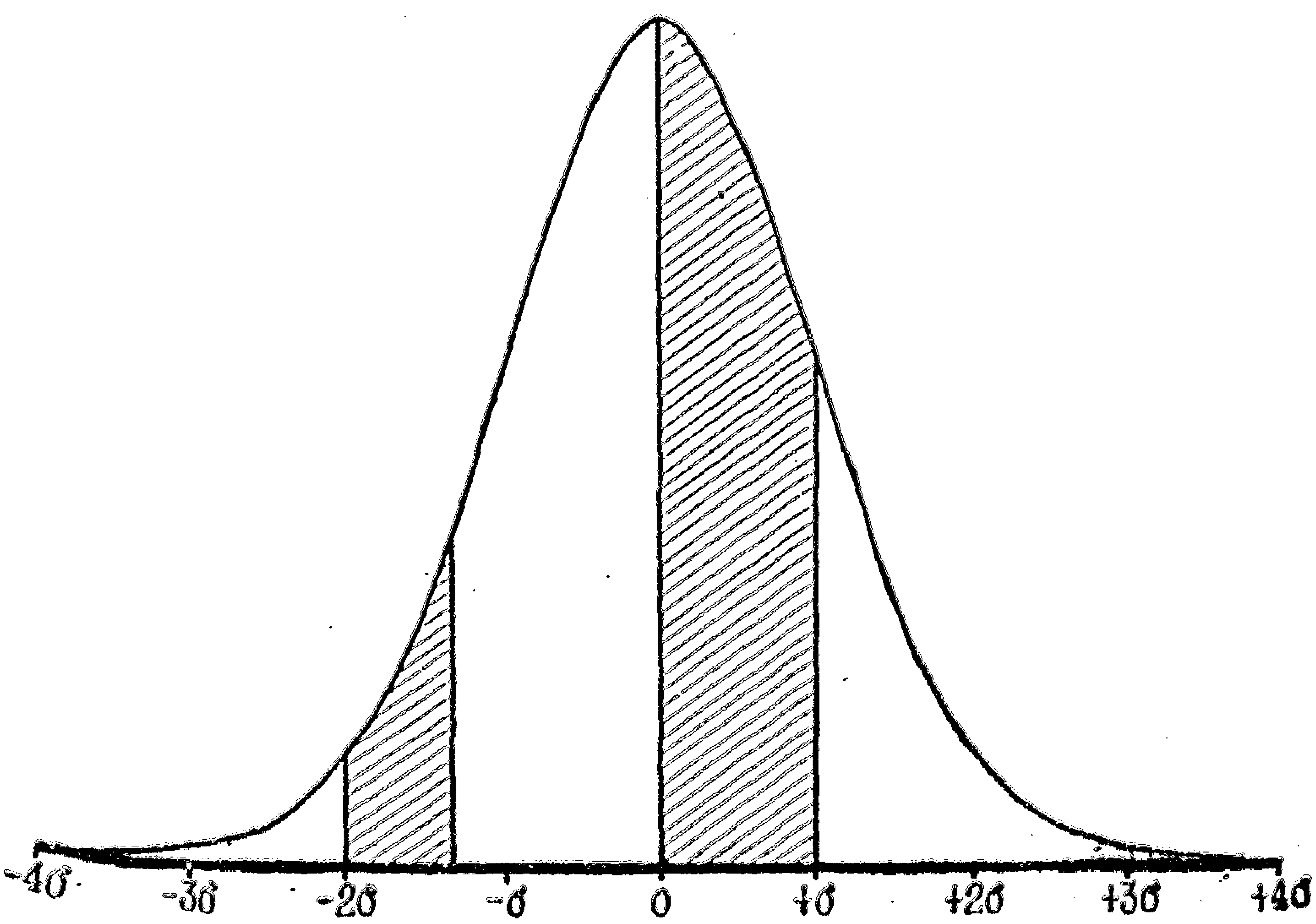
常態曲線下面積之全部代表次數總數,吾人如能知某部佔全面積之百分數,則該部面積代表之次數不難求得。各部面積佔全面積之百分數可由附錄二表中查出,此表假定常態曲線下全面積等於一,以σ為單位,將y_0左右之面積分作若干部份,視其縱線與y_0之距離(按即各縱線在底線上之位置與平均數

所在點間之距離),一一載明各部面積佔全面積之百分數。

因常態曲線兩邊對稱,故表中所載分數,正負均可適用。例如從y_0至$+1\sigma$距離處所豎立縱線間之面積應爲全體面積之.3413,由表中x/σ下1.0行中一查即得(從y_0至$-\sigma$距離處所豎立縱線間之面積亦爲全面積之.3413)。又如從-1.4σ與-2.0σ兩點各豎一縱線與常態曲線相交,則此二縱線間之面積究有幾何,依上表從y_0至-1.4σ之縱線間之面積等於全面積百

圖三十一

常態曲線下之面積



分之41.92,從y_0至-2.0σ縱線間之面積等於全面積百分之47.73,兩者之差爲百分之5.81(47.73%−41.92%=5.81%)即-1.4σ與-2.0σ二縱線間之面積(參看圖三十一)。以此項各部面積佔全面積之百分數乘次數總數(N)即得理論次數。

茲就常態曲線下之面積計算995家電話用戶每年通話之理論次數如下:

表四十八

995家電話用戶每年通話之理論次數計算表

1 組限	2 x/σ	3 y_0與x/σ兩縱線間之面積佔全面積之分數	4 y_0與x/σ兩縱線間之次數	5 理論次數	
0	−3.23	.4993810	496.88		
50	−2.89	.4980738	495.58	0——50	1.92*
100	−2.55	.4946139	492.14	50——100	3.44
150	−2.22	.4867906	484.36	100——150	7.78
200	−1.88	.4699460	467.60	150——200	16.76
250	−1.54	.4382198	436.03	200——250	31.57
300	−1.20	.3849303	383.01	250——300	53.02
250	−0.86	.3051055	303.58	300——350	78.63
400	−0.52	.1984682	197.48	350——400	106.90
450	−0.18	.0714237	71.07	400——450	126.41
500	+0.16	.0635595	63.24	450——500	134.31
550	+0.495	.1896931	188.74	500——550	125.50
600	+0.83	.2967306	295.25	550——600	106.51
650	+1.17	.3789995	377.10	600——650	81.85
700	+1.51	.4344783	432.31	650——700	55.21
750	+1.85	.4678432	465.50	700——750	33.19
800	+2.19	.4857379	483.31	750——800	17.81
850	+2.53	.4942969	491.83	800——850	8.52
900	+2.87	.4979476	495.46	850——900	3.63
950	+3.20	.4993129	496.82	900——950	1.36
1000	+3.54	.4997999	497.30	950——1000	.48
1050	+3.88	.4999478	497.45	1000——1050	.15
1100	+4.22	.4999878	497.49	1050以上	.05
					995.00

*按常態曲線下之面積計算,在-3.23σ下尚應有.62次,但就本例而論,於理不合,故加入0—50一組。

較精確之理論次數既經求得,然則實際分配與理論次數差異之原因,究由於實際分配不合於常態曲線,抑係受取樣之影響?欲明乎此,吾人可就差異最大數點計算其取樣之標準誤 (The standard error of Sampling)。當一事成敗之機率已求得時,其標準誤可依下式計算,已於第一節中言之:

$$\sigma=\sqrt{npq}$$

式中 p 表示成功之機率,q 表示失敗之機率。在次數分配中,吾人可以f/N代 p,$\frac{N-f}{N}$代 q,(f 代表常態曲線底線上由某一點所豎立縱線之高度亦即理論次數,N 代表次數總數),故取樣標準誤之公式可書作:

$$\sigma_s=\sqrt{N\times\frac{f}{N}\times\frac{N-f}{N}}$$

$$=\sqrt{\frac{f(N-f)}{N}}$$

本例實際分配與理論次數差異較大者計有四點,茲以 f 代理論次數,f_o 代實際次數(參看表四十七,四十八及圖三十),計算其差異如下:

m	f_o	f	f_o-f
325	95	78.63	+16.37
375	85	106.90	−21.90

525	144	125.50	+18.50
775	11	17.81	−6.81

第一點實際次數與理論次數相差16.37,第二點相差−21.90,第三點相差+18.50,第四點相差−6.81,然按$\sigma_s=\sqrt{\frac{f(N-f)}{N}}$公式計算其取樣之標準誤,則325一組之$\sigma_s$爲8.51,375一組之$\sigma_s$爲9.73,525一組之$\sigma_s$爲10.47,775一組之$\sigma_s$爲4.18。

325—組之σ_s:

$$\sigma_s=\sqrt{\frac{78.63\times(995-78.63)}{995}}=8.51$$

375—組之σ_s:

$$\sigma_s=\sqrt{\frac{106.90(995-106.90)}{995}}=9.73$$

525—組之σ_s:

$$\sigma_s=\sqrt{\frac{125.50(995-125.50)}{995}}=10.47$$

775—組之σ_s:

$$\sigma_s=\sqrt{\frac{17.81(995-17.81)}{995}}=4.18$$

取樣之標準誤既經求得,則吾人不難解答實際分配與理論次數差異之原因。如325一組實際次數與理論次數相差16.73,而σ_s爲8.51,相差之數等於σ_s之1.9倍,按附錄二表中$x/\sigma=1.9$時,則面積等於.47128,意卽由y_o至1.9σ距離之縱線間含有之次數等於次數總數百分之四七一二八,在$\pm1.9\sigma$兩縱線間之次數

等於次數總數百分之九四·二五六。換言之,實際次數(95)與理論次數相差至16.73之機會僅百分之六,故吾人可斷言此項差異係受取樣之影響。餘三組實際分配與理論次數之差異可用同法查考,其中差異最大者厥爲375之一組,相差-21.90,而σ_s爲9.73,相差之數等於σ_s之2.2倍,按附錄二表中$x/\sigma=2.2$時,則面積等於.48610,在$\pm 2.2\sigma$兩縱線間之次數等於次數總數百分之九七·二二,意即實際次數(85)與理論次數相差至-21.90之機會僅百分之三弱。吾人因之可斷言電話用戶通話次數之分配實有常態之形勢,其實際分配與理論次數有數點差異較著之原因,則係受取樣之影響。

就原有之次數分配而研究電話用戶通話次數,吾人所知者僅爲各組之實際次數,配合常態曲線後則吾人可知各組之理論次數,且可進一步推知底線上某點至某點之間應有之次數,是吾人之知識不僅限於研究一部分之少數材料矣。

常態曲線之用途甚廣,除試驗實際分配是否合於常態外,他如推算事業成績之人數分配,測驗問題難易之程度,以及定學校試題之分數等均可利用常態曲線之理論與方法,茲不詳述。

第四節 理論衆數之求法

第七章中所述計算衆數之方法,求得之結果僅爲近似衆數,眞正理論衆數,應按下式求之:

$$M_o = M - d$$

式中:M_o爲衆數

M 爲算術平均數

d 爲算術平均數與衆數間之距離,等於$\varkappa \times \sigma$

欲明瞭 d 之算法,請先一述 $\varkappa$ 之意義。$\varkappa$ 者,乃測度偏態之比較精確的數量,其值可依下式求之:

$$\varkappa = \frac{\sqrt{\beta_1}(\beta_2+3)}{2(5\beta_2-6\beta_1-9)}$$

式中:$$\beta_1 = \frac{\left[\frac{\Sigma fx'^3}{N} - 3\times\frac{\Sigma fx'}{N}\times\frac{\Sigma fx'^2}{N} + 2\times\left(\frac{\Sigma fx'}{N}\right)^3\right]^2}{\left\{\left[\frac{\Sigma fx'^2}{N} - \left(\frac{\Sigma fx'}{N}\right)^2\right] - \frac{1}{12}\right\}^3}$$

$$\beta_2 = \frac{1}{\left\{\left[\frac{\Sigma fx'^2}{N}-\left(\frac{\Sigma fx'}{N}\right)^2\right]-\frac{1}{12}\right\}^2}\left\{\left[\frac{\Sigma fx'^4}{N} - 4\times\frac{\Sigma fx'}{N}\times\frac{\Sigma fx'^3}{N} + 6\times\left(\frac{\Sigma fx'}{N}\right)^2\times\frac{\Sigma fx'^2}{N}\right.\right.$$
$$\left.\left.-3\times\left(\frac{\Sigma fx'}{N}\right)^4\right] - \frac{1}{2}\left[\frac{\Sigma fx'^2}{N} - \left(\frac{\Sigma fx'}{N}\right)^2\right] + \frac{7}{240}\right\}$$

註:$\frac{1}{12}$ 及 $\frac{7}{240}$ 爲解柏德(W. F. Sheppard)氏之改正數,均以組距爲單位。

在完全常態之分配,$\beta_1=0$, 故 $\varkappa$ 亦等於零。至於

β_2在完全常態分配中應等於 3,如小於 3,則繪成之曲線其頂部必較常態曲線爲平;如大於 3,則繪成之曲線其頂部必較常態曲線爲峭。茲就 9 9 5 家電話用戶通話次數分配之例,計算其β_1及β_2如下:

$$\frac{\Sigma fx'}{N}=\frac{-956}{995}=-.960804$$

$$\frac{\Sigma fx'^2}{N}=\frac{9,676}{995}=9.724623$$

$$\frac{\Sigma fx'^3}{N}=\frac{-22,952}{995}=-23.067337$$

$$\frac{\Sigma fx'^4}{N}=\frac{283,564}{995}=284.988945$$

$$\frac{1}{12}=.083333$$

$$\frac{7}{240}=.029167$$

$$\beta_1=\frac{\left[-23.067337-3\times(-.960804)\times 9.724623+2\times(-.960804)^3\right]^2}{\left\{\left[9.724623-(-.960804)^2\right]-.083333\right\}^3}$$

$$=\frac{10.170429}{662.632015}=.01534853$$

$$\beta_2=\frac{1}{\{[9.72-(-9.61)^2]-.083\}^2}\Big\{\Big[284.99-4\times(-.96)\times(-23.067)$$

$$+6\times(-.96)^2\times 9.72-3\times(-.96)^4\Big]-\frac{1}{2}\Big[9.72-(-.96)^2\Big]$$

$$+.029\Big\}=\frac{243.271411}{76.006070}=3.200683$$

將β_1及β_2之值代入$\varkappa$之公式,則$\varkappa=-.05558$;計算如下:

$$\varkappa=\frac{\sqrt{.01535}(3.2+3)}{2(5\times3.2-6\times.01535-9)}$$

$$=.05558$$

如算術平均數大於中位數,則$\varkappa$前應冠以正號,反之,應冠以負號;$\varkappa$數值之大小,即所以表示偏斜之程度。本例算術平均數為476.96,中位數為482.39,為負的偏態,$\varkappa$應為$-.05558$。$\varkappa$既經求得,即可依$d=\varkappa\times\sigma$之公式計算算術平均數與衆數間之距離。本例以組距為單位之標準差為2.953,以原來單位表示之,應為147.65($2.953\times50=147.65$),代入式中:

$$d=-.05558\times147.65$$

$$=-8.21$$

已知求理論衆數之公式為$M_o=M-d$,將算術平均數及d之數值代入式中即得理論衆數之數值。

$$M_o=476.96-(-8.21)$$

$$=476.96+8.21=485.17$$

由此式求得之衆數遠較由第七章中所述諸法求得者為精確,惟以計算手續異常繁複,故非必要時,用之者鮮。

第十二章 統計歸納與取樣問題

第一節 統計歸納之意義

前十一章所述爲統計分析之方法,多偏於統計技術方面。茲請進而一論應用此項統計方法時各方面之問題,如統計方法在研究經濟,社會各項問題中之地位何如?應用統計方法及其結果時有何假定?有何限制?凡此種種,雖無關於統計技術之本身,然欲確切認識統計方法之性質,吾人殊不能不一加探討。

應用統計方法以整理,分析所搜集之材料,製成次數分配表,研究其集中程度,離中趨勢,偏斜度等謂之統計敍述(Statistical Description);如吾人將此項搜集所得一部材料之結果引伸之,應用於研究材料範圍以外,代表一般或全部之結果,則謂之統計歸納(Statistical Induction or Statistical Inferrence)。統計之結果,其應用若僅限於所搜集之材料範圍以內,自屬可信無疑;設引伸之,擴大其應用之範圍,以一部之結果代表一般或全部之結果,是否可靠,有無根據,有無限制,此本章討論之主旨也。

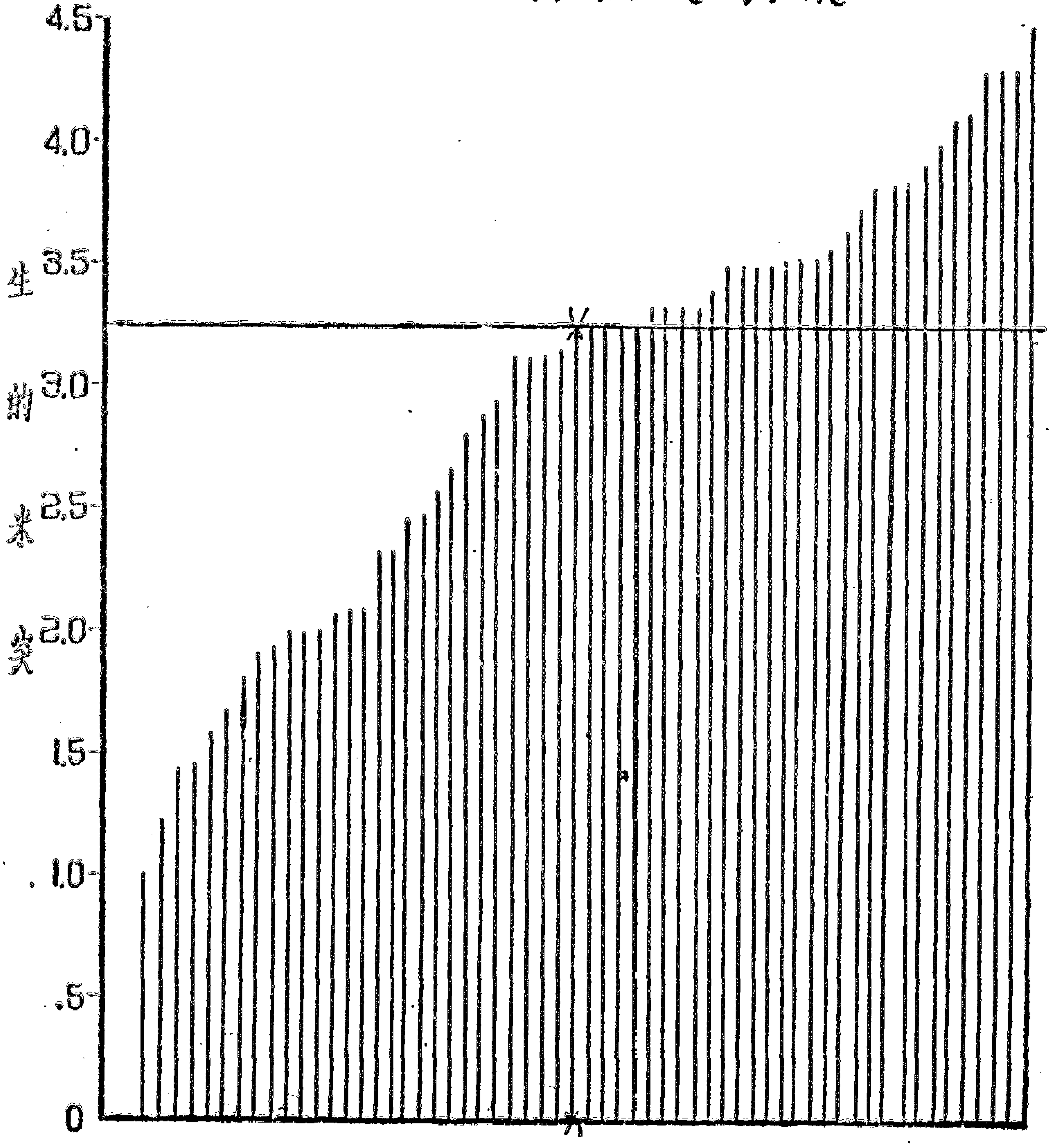
第二節 統計歸納之基本原則

宇宙間之事物或現象雖若千頭萬緒,錯綜紛紜,

然如根據大量觀察,常能發見其劃一性(Uniformity)及穩定性(Stability,)故就大羣(Population)中取其一部而研究之,其結果每能爲大羣之代表,已於第一章第三

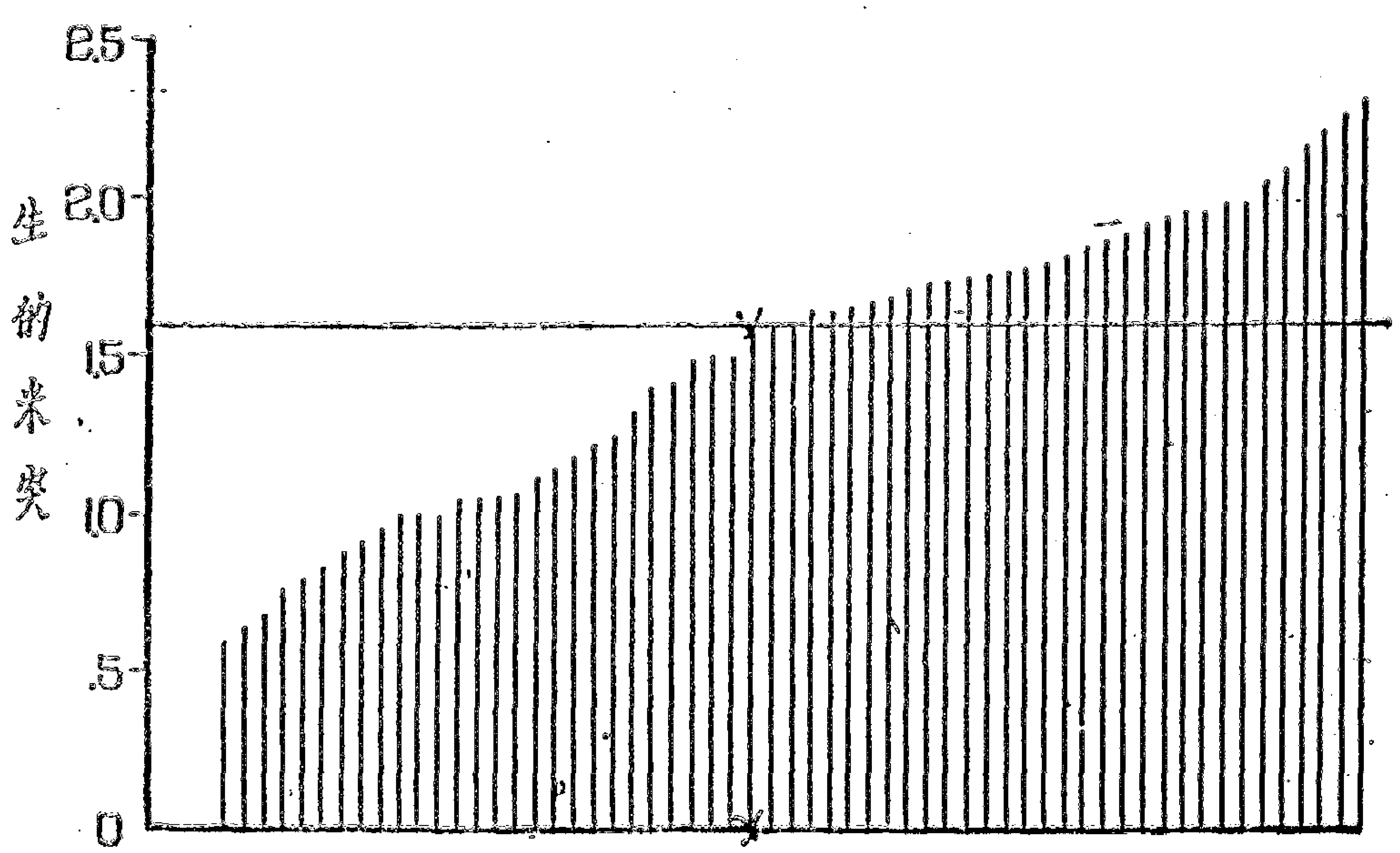
圖三十二

五十九個貝殼之長度



節中約略言之。英人愛爾段敦(Elderton)氏兄妹嘗就貝殼之長度,寬度而試驗之,第一次從一大羣之貝殼中取出五十九個,一一量其長度,寬度依次排列之,如圖三十二與三十三:

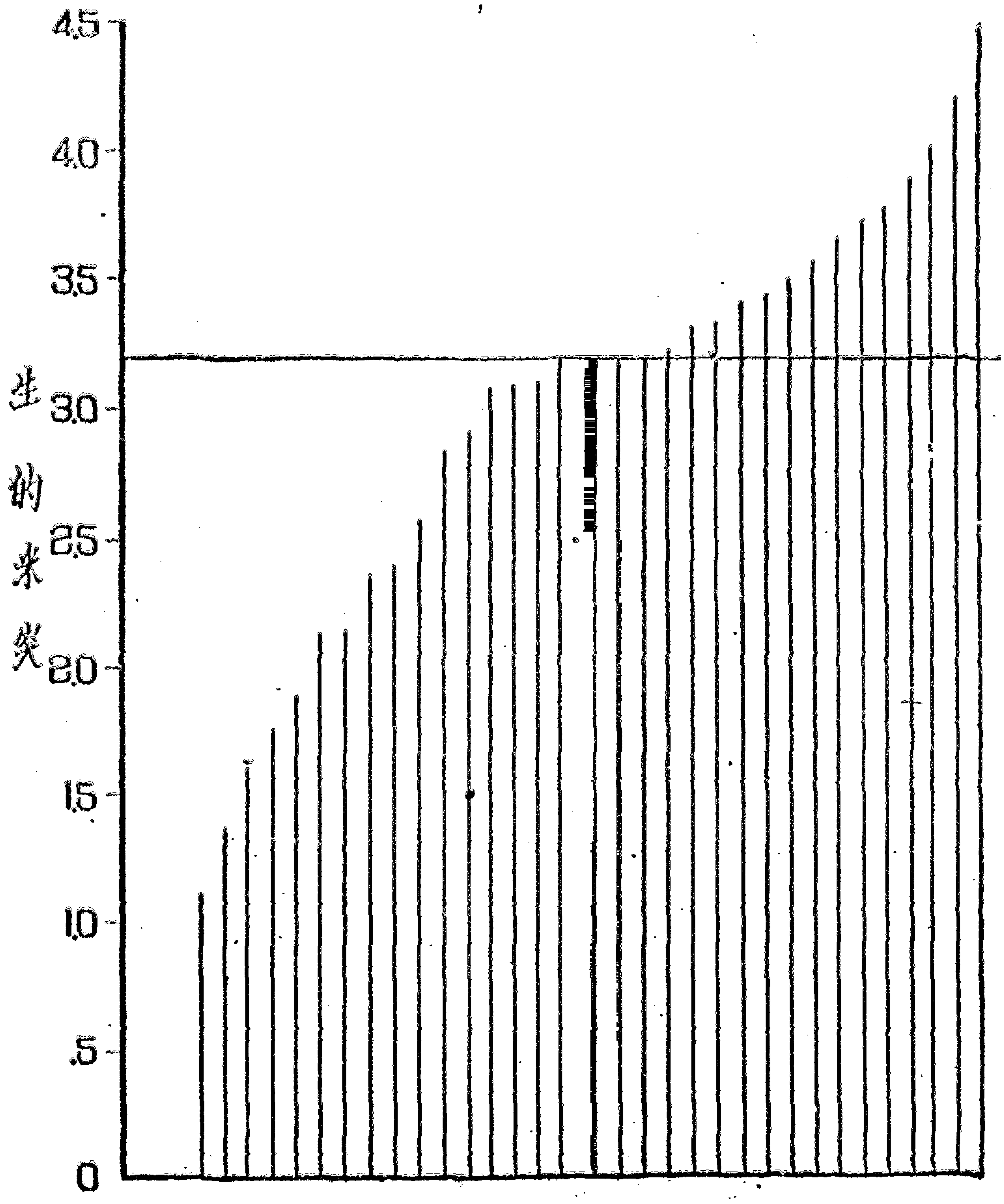
圖三十三
五十九個貝殼之寬度



第二次另取貝殼三十三個,一一量其長度及寬度如前排列之,如圖三十四,三十五:

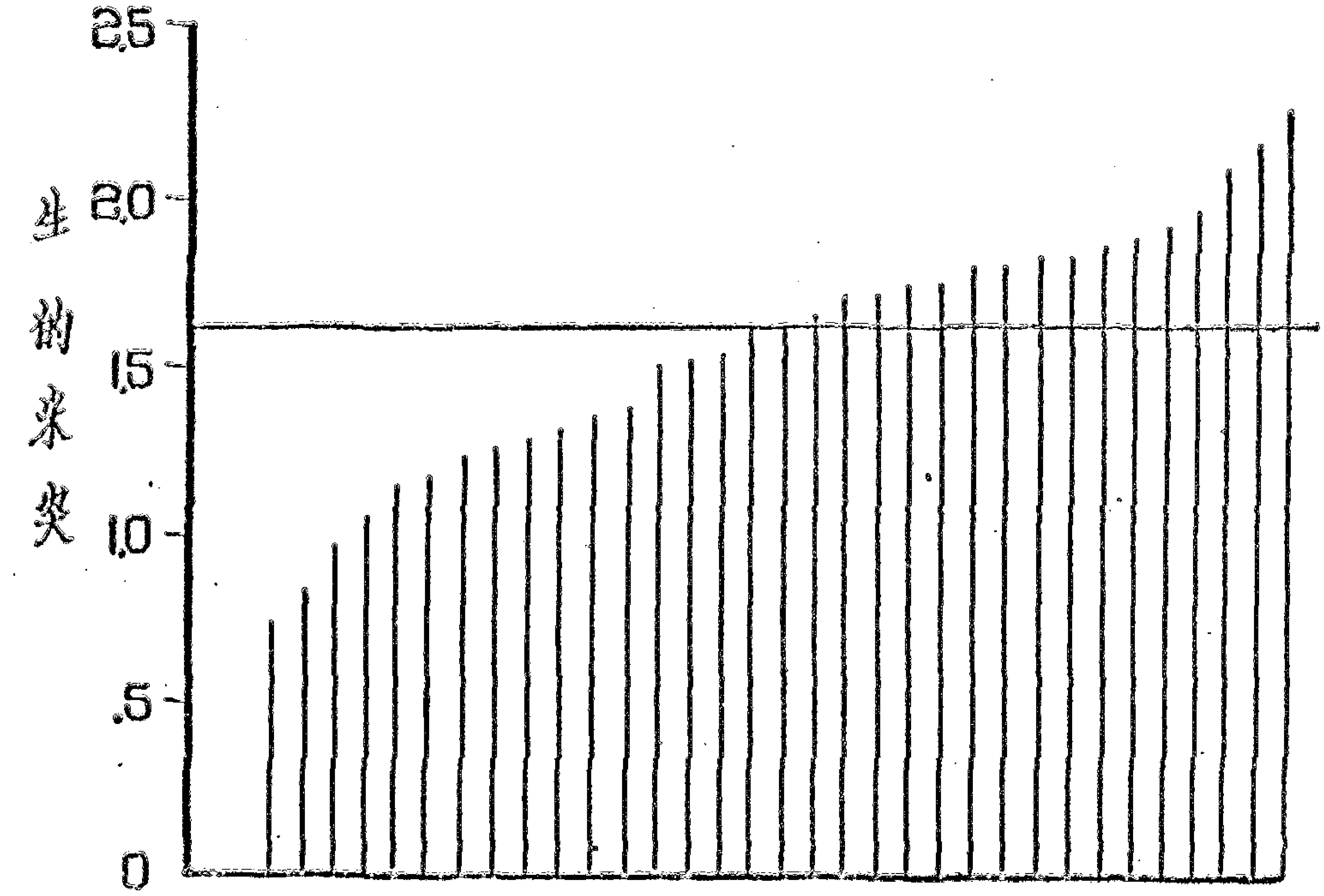
圖三十四

三十三個貝殼之長度



圖三十五

三十三個貝殼之寬度



讀者細察上圖,可知:(一)各圖之形狀兩端參差,而近中間之部分則平坦,此由多數貝殼與居中者之長度或寬度相仿,極長或極短,極寬或極狹之貝殼均屬少數。(二)圖三十二中貝殼長度之中位數與圖三十四中之中位數相仿;圖三十三中貝殼寬度之中位數與圖三十五中之中位數相仿。換言之,貝殼有適中之長度及寬度,而此適中之長度及寬度在各次取樣中常有一致之趨勢。貝殼如此,推而至於他項事物或現象亦然。故統計之結果,雖僅由一部分之材料求得,吾人亦得引伸之,用以代表一般或全部之結果,此統

計歸納之基本原則也。

第三節　簡單取樣或隨機取樣之條件

由一部份材料求得之結果,欲期其能充分代表全體,必先求取樣之適當。欲得適當之樣本,吾人必於一大羣中隨機的 (Taking at random) 抽取適當多的 (Sufficiently large) 數目,決不能有絲毫成見存乎其間。設稍有取舍之成見,則所得之樣本常有偏誤(bias)。所取之數目亦不宜過少,過少則不足以代表全體。友爾 (G. U. Yule) 教授嘗舉簡單取樣之條件三項,謂吾人對於此三項條件必一一遵守,然後所取之樣本始有代表全體之可能,而統計結果可靠之程度乃能確定。其條件之大意如下:

一、所取之樣本應各自完全獨立,不相關聯。例如編製物價指數,甲物之價格如與乙物之價格相關聯,則二者不宜列入同一指數。

二、所取之樣本應在同一地域並同一時間。例如研究某地人眼睛各種顏色所佔之百分數,則所取之樣本自以當地並同時之人口爲限。

三、取樣時不但每個樣本在一大羣中發現之機會應相等,並須每個樣本各部分發現之機會亦相等。如作擲幣試驗若干次,不獨各次所用

之幣應相同,且須各幣之形式,輕重相等,俾各幣在每一擲中發現面或背之機會亦相等。

上述三項條件果能一一遵行,則吾人由一次取樣之結果,可以按公式測定其可靠之程度,推知全體結果之範圍。測定統計結果可靠程度之公式甚多,請於下節論之。

第四節 測定統計結果可靠程度之各種公式

統計結果可靠之程度,非必按公式而後可以測定,如取骰十二粒連擲4,096次,以一,二,三各面之發現爲失敗,四,五,六各面之發現爲成功,將所得各次之結果製成次數分配表,計算其標準差爲1.712,與理論次數之標準差(即標準誤)1.732,相差極微(參閱第十一章第一節)。推而廣之,如吾人研究工人每月之平均工資,第一次根據250名工人之工資計算,欲知此項平均工資代表全體工人平均工資時可靠之程度,可繼續根據若干次之取樣,每次工人人數仍爲250,求得其平均數,將若干次之平均數製成次數分配表,而求其標準差。取樣次數愈多,則此項標準差愈近於取樣之標準誤。惟如此測定統計結果可靠之程度,取樣次

數必多,事實上頗不易行,故仍以按公式計算爲便。茲將測定統計結果可靠程度之各種重要公式列下,其推衍之敍述從略:

一、測定算術平均數可靠程度之公式:

甲、標準誤:

$$\sigma_M=\frac{\sigma}{\sqrt{N}}$$

式中:σ_M 爲算術平均數之標準誤的符號

σ 爲原分配(即所取樣本)之標準差

N 爲次數

乙、機差(亦名機誤):

$$P.E._M=.6745\frac{\sigma}{\sqrt{N}}$$

茲舉例說明上式之應用。假定某城共有工人十萬人,吾人欲研究其平均工資,勢不能一一調查,於是由十萬工人中隨機抽取 900 人,求得其平均工資爲 \$27.50,標準差 \$2.00,則取樣之標準誤爲 \$0.067:

$$\sigma_M=\frac{\$\ 2.00}{\sqrt{900}}$$

$$=\frac{\$\ 2.00}{30}=\$0.067$$

機差爲 \$0.045:

$$P.E._M=.6745\times\frac{\$\ 2.00}{\sqrt{900}}$$

$$=\ \$0.045$$

意即吾人若由十萬工人中另行抽取900人而求得其平均工資,則所得之結果當有百分之六八·二六之機會在\$27.50±\$.067範圍以內,二分之一之機會在\$27.50±\$.045範圍以內。故吾人報告結果時,如以標準誤表示可靠之程度,則應書作\$27.50±\$.067;如以機差表示可靠之程度,則應書作\$27.50±\$.045。上項解釋可適用於任何統計結果可靠程度測定之公式,惟究以標準誤抑機差為準,報告時應一併說明。表示統計結果可靠之程度,用標準誤或機差均無不可,但理論上似以用標準誤為宜(理由另詳),故以下僅列標準誤之公式,其機差之公式從略。

二、測定中位數可靠程度之公式:

$$\sigma_{Md}=1.25331\times\frac{\sigma}{\sqrt{N}}$$

三、測定四分位數(Q_1及Q_3)可靠程度之公式:

$$\sigma_{Q}=1.36263\times\frac{\sigma}{\sqrt{N}}$$

四、測定四分差可靠程度之公式:

$$\sigma_{Q.D.}=.78672\times\frac{\sigma}{\sqrt{N}}$$

五、測定標準差可靠程度之公式:

甲、如所取之樣本為常態的分配,可用下式:

$$\sigma_{\sigma}=\frac{\sigma}{\sqrt{2N}}$$

乙、不論所取樣本分配為何種形式,均可用下

式：

$$\sigma_\sigma=\sqrt{\frac{\mu_4-\mu_2^2}{4\mu_2\cdot N}}$$

式中：$\mu_2=\left[\frac{\Sigma f d'^2}{N}-\left(\frac{\Sigma f d'}{N}\right)^2\right]-\frac{1}{12}$

$$\mu_4=\left[\frac{\Sigma f d'^4}{N}-4\times\frac{\Sigma f d'}{N}\times\frac{\Sigma f d'^3}{N}+6\times\left(\frac{\Sigma f d'}{N}\right)^2\times\frac{\Sigma f d'^2}{N}-3\times\left(\frac{\Sigma f d'}{N}\right)^4\right]-\frac{1}{2}\left[\frac{\Sigma f d'^2}{N}-\left(\frac{\Sigma f d'}{N}\right)^2\right]+\frac{7}{240}$$

六、測定相關係數(假定為常態的分配)可靠程度之公式：

$$\sigma_r=\frac{1-r^2}{\sqrt{N}}$$

此式並適用於測定淨相關與複相關係數之可靠程度；用於相關率，相差亦不過遠。

七、測定迴歸係數(假定為常態的分配)可靠程度之公式：

$$\sigma_{b_{12}}=\frac{\sigma_1\sqrt{1-r^2}}{\sigma_2\sqrt{N}}$$

此式可用以測定任何一級迴歸係數之可靠程度。

八、測定 ζ 可靠程度之公式：

$$\sigma_\zeta=2\sqrt{\frac{\zeta}{N}}\sqrt{(1-\eta^2)^2-(1-r^2)^2+1}$$

判別兩種變數之關係為直線的或非直線的，

可視η^2與r^2兩者之差以爲斷,其公式爲$\zeta=\eta^2-r^2$,嘗於第九章第三節中言之,然ζ之值大至如何程度,始可決定爲非直線相關,則須按上式計算。茲仍以193畝麥田小麥收穫量與其所施肥料數量之相關率及相關係數爲例計算ζ及σ_ζ如下:

$$\zeta=\eta^2-r^2$$
$$=(.966)^2-(.793)^2$$
$$=.933-.629$$
$$=.304$$

將ζ之數值代入公式,則$\sigma_\zeta=.074$

$$\sigma_\zeta=2\sqrt{\frac{\zeta}{N}}\sqrt{(1-\eta^2)^2-(1-r^2)^2+1}$$
$$=2\times\sqrt{\frac{.304}{193}}\sqrt{(1-.933)^2-(-.629)^2+1}$$
$$=2\times\sqrt{.00158}\sqrt{.004489-.137641+1}$$
$$=2\times\sqrt{.00158}\sqrt{.866848}$$
$$=2\times\sqrt{.00137}$$
$$=2\times.037$$
$$=.074$$

ζ之數值爲.304較其標準誤大4.1倍强,可見

小麥收穫量與其所施肥料數量之關係爲非直線相關也。

九、測定 χ 可靠程度之公式:

$$\sigma_\chi=\sqrt{\frac{3}{2N}}$$

χ 爲測定偏斜度之較精確的數量，於第十一章第四節中嘗言及之，茲仍就995家電話用戶每年通話次數分配之例計算 σ_χ 如下:

$$\sigma_\chi=\sqrt{\frac{3}{1990}}$$
$$=.03883$$

χ 之值爲 $-.05558$（參閱第十一章第四節），當其標準誤1.43倍，故該分配之所以略呈偏態，似受取樣之影響，吾人不能遽斷爲不合常態也。

十、測定 d 可靠程度之公式:

$$\sigma_d=\sqrt{\frac{3}{2N}}\sigma$$

d 爲算術平均數與衆數間之距離，在常態分配中，其標準誤可按上式求出。茲仍以995家電話用戶每年通話次數之分配爲例，計算 σ_d 如下:

$$\sigma_d=\sqrt{\frac{3}{1990}}\times 147.65$$
$$=5.733$$

d 之數值爲8.21（參閱第十一章第四節），約當

其標準誤1.5倍弱,是算術平均數與衆數不能恰恰相等之原因顯然係受取樣之影響,蓋 d 之數值若不大於其標準誤三倍,吾人不能遽斷該分配爲非常態也。

十一、測定兩算術平均數差額標準誤之公式:

$$\sigma_D = \sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}$$

式中:σ_D 爲兩算術平均數差額標準誤之符號,D 表示差額,爲 Difference 之縮寫

σ_1 爲甲分配之標準差

σ_2 爲乙分配之標準差

N_1 爲甲分配之次數總數

N_2 爲乙分配之次數總數

例如有兩算術平均數於此,係由兩次取樣之結果計算而得,其差額係受取樣之影響乎,抑根本兩次樣本所從取之兩大羣即不相同,可用上式測定之。如 D 大於 σ_D 三倍,則吾人可斷言此項差額根本由於兩大羣之材料不同,否則此項差額必係受取樣之影響。

十二、測定取樣標準誤之公式:

$$\sigma_s = \sqrt{\frac{f(N-f)}{N}}$$

此式常用以測定實際次數與理論次數之差異,以斷實際分配之合乎常態與否,例見第十一章第三節,讀者參閱該節可也。

以上各種公式,除第一式外,悉以標準誤為準,吾人若欲化為機差,可各乘以 0.6745。惟常態曲線面積分表其橫距之計算係以標準差為單位,用標準誤,檢查機率時自較用機差為便也。

統計結果可靠之程度固可按公式測定,然亦非無限制,蓋公式所測定者為受取樣之錯誤,必先取得適當之樣本,而後可以應用上述公式,此吾人搜集材料時所以當慎之又慎,勿任取樣以外之錯誤侵入也。

附錄一

常態曲線下之縱線表

x/σ	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	1.00000	.99995	.99980	.99955	.99920	.99875	.99820	.99755	.99685	.99596
0.1	.99501	.99396	.99283	.99159	.99025	.98881	.98728	.98565	.98393	.98211
0.2	.98020	.97819	.97609	.97390	.97161	.96923	.96676	.96420	.96156	.95882
0.3	.95600	.95309	.95010	.94702	.94387	.94055	.93723	.93382	.93024	.92677
0.4	.92312	.91939	.91558	.91169	.90774	.90371	.89961	.89543	.89119	.88688
0.5	.88250	.87805	.87353	.86896	.86432	.85962	.85488	.85006	.84518	.84050
0.6	.83527	.83023	.82514	.82010	.81481	.80957	.80429	.79896	.79358	.78817
0.7	.78270	.77721	.77167	.76610	.76048	.75484	.74916	.74342	.73769	.73193
0.8	.72615	.72033	.71448	.70861	.70272	.69681	.69087	.68492	.67896	.67298
0.9	.66689	.66097	.65494	.64891	.64287	.63683	.63077	.62472	.61865	.61259
1.0	.60653	.60047	.59440	.58834	.58228	.57623	.57017	.56414	.55810	.55209
1.1	.54607	.54007	.53409	.52811	.52214	.51620	.51027	.50437	.49848	.49260
1.2	.48675	.48092	.47511	.46933	.46357	.45783	.45212	.44644	.44078	.43516
1.3	.42956	.42399	.41845	.41294	.40747	.40202	.39661	.39123	.38569	.38058
1.4	.37531	.37007	.36487	.35971	.35459	.34950	.34445	.33944	.33447	.32954
1.5	.32465	.31980	.31500	.31023	.30550	.30082	.29618	.29158	.28702	.28251
1.6	.27804	.27361	.26923	.26489	.26059	.25634	.25213	.24797	.24385	.23978
1.7	.23575	.23176	.22782	.22392	.22008	.21627	.21250	.20879	.20511	.20148
1.8	.19790	.19436	.19086	.18741	.18400	.18064	.17732	.17404	.17081	.16762
1.9	.16448	.16137	.15831	.15530	.15232	.14939	.14649	.14364	.14083	.13806
2.0	.13534	.13265	.13000	.12740	.12483	.12230	.11981	.11737	.11496	.11259
2.1	.11025	.10795	.10570	.10347	.10129	.09914	.09702	.09495	.09290	.09090
2.2	.08892	.08698	.08507	.08320	.08136	.07956	.07779	.07604	.07433	.07265
2.3	.07100	.06939	.06780	.06624	.06471	.06321	.06174	.06029	.05888	.05750
2.4	.05614	.05481	.05350	.05222	.05096	.04973	.04852	.04734	.04618	.04505
2.5	.04394	.04285	.04179	.04074	.03972	.03873	.03775	.03680	.03586	.03494
2.6	.03405	.03317	.03232	.03148	.03066	.02986	.02908	.02831	.02757	.02684
2.7	.02612	.02542	.02474	.02408	.02343	.02280	.02218	.02157	.02098	.02040
2.8	.01984	.01929	.01876	.01823	.01772	.01723	.01674	.01627	.01581	.01536
2.9	.01492	.01449	.01408	.01367	.01328	.01288	.01252	.01215	.01179	.01145
3.0	.01111	.00819	.00598	.00432	.00309	.00219	.00153	.00106	.00073	.00050
4.0	.00034	.00022	.00015	.00010	.00006	.00004	.00003	.00002	.00001	.00001
5.0	.00000									

附錄二

常態曲線下之面積表

x/σ	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.00000	.00399	.00798	.01197	.01595	.01994	.02392	.02790	.03188	.03586
0.1	.03983	.04380	.04776	.05172	.05567	.05962	.06356	.06749	.07142	.07535
0.2	.07926	.08317	.08706	.09095	.09483	.09371	.10257	.10642	.11026	.11409
0.3	.11791	.12172	.12552	.12930	.13307	.13683	.14058	.14431	.14803	.15173
0.4	.15542	.15910	.16276	.16640	.17003	.17364	.17724	.18082	.18439	.18793
0.5	.19146	.19497	.19347	.20194	.20540	.20884	.21226	.21566	.21904	.22240
0.6	.22575	.22907	.23237	.23565	.23891	.24215	.24537	.24857	.25175	.25490
0.7	.25804	.26115	.26424	.26730	.27035	.27337	.27637	.27935	.28230	.28524
0.8	.28814	.29103	.29389	.29673	.29955	.30234	.30511	.30785	.31057	.31327
0.9	.31594	.31859	.32121	.32381	.32639	.32894	.33147	.33398	.33646	.33891
1.0	.34134	.34375	.34614	.34850	.35083	.35314	.35543	.35769	.35993	.36214
1.1	.36433	.36650	.36864	.37176	.37286	.37493	.37698	.37900	.38100	.38298
1.2	.38493	.38686	.38877	.39065	.39251	.39435	.39617	.39796	.39973	.40147
1.3	.40320	.40490	.40658	.40824	.40988	.41149	.41309	.41466	.41621	.41774
1.4	.41924	.42073	.42220	.42364	.42507	.42647	.42786	.42922	.43056	.43198
1.5	.43319	.43448	.43574	.43699	.43822	.43943	.44062	44179	.44295	.44408
1.6	.44520	.44630	.44738	.44845	.44950	.45053	.45154	.45254	.45352	.45449
1.7	.45543	.45637	.45728	.45818	.45907	.45994	.46080	.46164	.46246	.46327
1.8	.46407	.46485	.46562	.46638	.46712	.46784	.46856	.46929	.46995	.47062
1.9	.47128	.47193	.47257	.47320	.47381	.47441	.47500	.47558	.47615	.47670
2.0	.47725	.47778	.47831	.47882	.47932	.47982	.48030	.48077	.48124	.48169
2.1	.48214	.48257	.48308	.48341	.48382	.48422	.48461	.48500	.48537	.48574
2.2	.48610	.48645	.48679	.48713	.48745	.48778	.48809	.48840	.48870	.48899
2.3	.48928	.48956	.48983	.49010	.49036	.49061	.49086	.49111	.49134	.49158
2.4	.49180	.49202	.49224	.49245	.49266	.49286	.49305	.49324	.49343	.49361
2.5	.49379	.49396	.49413	.49430	.49446	.49461	.49477	.49492	.49506	.49520
2.6	.49534	.49547	.49560	.49573	.49585	.49598	.49609	.49621	.49632	.49643
2.7	.49653	.49664	.49674	.49683	.49693	.49702	.49711	.49720	.49728	.49736
2.8	.49744	.49752	.49760	.49767	.49774	.49781	.49788	.49795	.49801	.49807
2.9	.49813	.49819	.49825	.49830	.49836	.49841	.49846	.49851	.49856	.49860
3.0	.49865									
3.1	.49903									
3.2	.49931									
3.3	.49952									
3.4	.49966									
3.5	.49977									
3.6	.49984									
3.7	.49990									
3.8	.49993									
3.9	.49995									
4.0	.49997									
4.5	.49999									
5.0	.50000									

附錄三

對數表 (10—99)

No.	0	1	2	3	4	5	6	7	8	9
10	0000	0043	0086	0128	0170	0212	0253	0294	0334	0374
11	0414	0453	0492	0531	0569	0607	0645	0682	0719	0755
12	0792	0828	0864	0899	0934	0969	1004	1038	1072	1106
13	1139	1173	1206	1239	1271	1303	1335	1367	1399	1430
14	1461	1492	1523	1553	1584	1614	1644	1673	1703	1732
15	1761	1790	1818	1847	1875	1903	1931	1959	1987	2014
16	2041	2068	2095	2122	2148	2175	2201	2227	2253	2279
17	2304	2330	2355	2380	2405	2430	2455	2480	2504	2529
18	2553	2577	2601	2625	2648	2672	2695	2718	2742	2765
19	2788	2810	2833	2856	2878	2900	2923	2945	2967	2989
20	3010	3032	3054	3075	3096	3118	3139	3160	3181	3201
21	3222	3243	3263	3284	3304	3324	3345	3365	3385	3404
22	3424	3444	3464	3483	3502	3522	3541	3560	3579	3598
23	3617	3636	3655	3674	3692	3711	3729	3747	3766	3784
24	3802	3820	3838	3856	3874	3892	3909	3927	3945	3962
25	3979	3997	4014	4031	4048	4065	4082	4099	4116	4133
26	4150	4166	4183	4200	4216	4232	4249	4265	4281	4298
27	4314	4330	4346	4362	4378	4393	4409	4425	4440	4456
28	4472	4487	4502	4518	4533	4548	4564	4579	4594	4609
29	4624	4639	4654	4669	4683	4698	4713	4728	4742	4757
30	4771	4786	4800	4814	4829	4843	4857	4871	4886	4900
31	4914	4928	4942	4955	4969	4983	4997	5011	5024	5038
32	5051	5065	5079	5092	5105	5119	5132	5145	5159	5172
33	5185	5198	5211	5224	5237	5250	5263	5276	5289	5302
34	5315	5328	5340	5353	5366	5378	5391	5403	5416	5428
35	5441	5453	5465	5478	5490	5502	5514	5527	5539	5551
36	5563	5575	5587	5599	5611	5623	5635	5647	5658	5670
37	5682	5694	5705	5717	5729	5740	5752	5763	5775	5786
38	5798	5809	5821	5832	5843	5855	5866	5877	5888	5899
39	5911	5922	5933	5944	5955	5966	5977	5988	5999	6010
40	6021	6031	6042	6053	6064	6075	6085	6096	6107	6117
41	6128	6138	6149	6160	6170	6180	6191	6201	6212	6222
42	6232	6243	6253	6263	6274	6284	6294	6304	6314	6325
43	6335	6345	6355	6365	6375	6385	6395	6405	6415	6425
44	6435	6444	6454	6464	6474	6484	6493	6503	6513	6522
45	6532	6542	6551	6561	6571	6580	6590	6599	6609	6618
46	6628	6637	6646	6656	6665	6675	6684	6693	6702	6712
47	6721	6730	6739	6749	6758	6767	6776	6785	6794	6803
48	6812	6821	6830	6839	6848	6857	6866	6875	6884	6893
49	6902	6911	6920	6928	6937	6946	6955	6964	6972	6981
50	6990	6998	7007	7016	7024	7033	7042	7050	7059	7067
51	7076	7084	7093	7101	7110	7118	7126	7135	7143	7152
52	7160	7168	7177	7185	7193	7202	7210	7218	7226	7235
53	7243	7251	7259	7267	7275	7284	7292	7300	7308	7316
54	7324	7332	7340	7348	7356	7364	7372	7380	7388	7396

附錄三

對數表（續）

No.	0	1	2	3	4	5	6	7	8	9
55	7404	7412	7419	7427	7435	7443	7451	7459	7466	7474
56	7482	7490	7497	7505	7513	7520	7528	7536	7543	7551
57	7559	7566	7574	7582	7589	7597	7604	7612	7619	7627
58	7634	7642	7649	7657	7664	7672	7679	7686	7694	7701
59	7709	7716	7723	7731	7738	7745	7752	7760	7767	7774
60	7782	7789	7796	7803	7810	7818	7825	7832	7839	7846
61	7853	7860	7868	7875	7882	7889	7896	7903	7910	7917
62	7924	7931	7938	7945	7952	7959	7966	7973	7980	7987
63	7993	8000	8007	8014	8021	8028	8035	8041	8048	8055
64	8062	8069	8075	9082	8089	8096	8102	8109	8116	8122
65	8129	8136	8142	8149	8156	8162	8169	8176	8182	8189
66	8195	8202	8209	8215	8222	8228	8235	8241	8248	8254
67	8261	8267	8274	8280	8287	8293	8299	8306	8312	8319
68	8325	8331	8338	8344	8351	8357	8363	8370	8376	8382
69	8388	8395	8401	8407	8414	8420	8426	8432	8439	8445
70	8451	8457	8463	8470	8476	8482	8488	8494	8500	8506
71	8513	8519	8525	8531	8537	8543	8549	8555	8561	8567
72	8573	8579	8585	8591	8597	8603	8609	8615	8621	8627
73	8633	8639	8645	8651	8657	8663	8669	8675	8681	8686
74	8692	8698	8704	8710	8716	8722	8727	8733	8739	8745
75	8751	8756	8762	8768	8774	8779	8785	8791	8797	8802
76	8808	8814	8820	8825	8831	8837	8842	8848	8854	8859
77	8865	8871	8876	8832	8887	8893	8899	8904	8910	8915
78	8921	8927	8932	8938	8943	8949	8954	8960	8965	8971
79	8976	8982	8987	8993	8908	9004	9009	9015	9020	9025
80	9031	9036	9042	9047	9053	9058	9063	9069	9074	9079
81	9085	9090	9096	9101	9106	9112	9117	9122	9128	9133
82	9138	9143	9149	9154	9159	9165	9170	9175	9180	9186
83	9191	9196	9201	9206	9212	9217	9222	9227	9232	9238
84	9243	9248	9253	9258	9263	9269	9274	9279	9284	9289
85	9294	9299	9304	9309	9315	9320	9325	9330	9335	9340
86	9345	9350	9355	9360	9365	9370	9375	9380	9385	9390
87	9395	9400	9405	9410	9415	9420	9425	9430	9435	9440
88	9445	9450	9455	9460	9465	9469	9474	9479	9484	9489
89	9494	9499	9504	9509	9513	9518	9523	9528	9533	9538
90	9542	9547	9552	9557	9562	9566	9571	9576	9581	9586
91	9590	9595	9600	9605	9609	9614	9619	9624	9628	9633
92	9638	9643	9647	9652	9657	9661	9666	9671	9675	9680
93	9685	9689	9694	9699	9703	9708	9713	9717	9722	9727
94	9731	9736	9741	9745	9750	9754	9759	9763	9768	9773
95	9777	9782	9786	9791	9795	9800	9805	9809	9814	9818
96	9823	9827	9832	9836	9841	9845	9850	9854	9859	9863
97	9868	9872	9877	9881	9886	9890	9894	9899	9903	9908
98	9912	9917	9921	9926	9930	9934	9939	9943	9948	9952
99	9956	9961	9965	9969	9974	9978	9983	9987	9991	9996

附錄四

平方表

No.	0	1	2	3	4	5	6	7	8	9
10	10000	10201	10404	10609	10816	11025	11236	11449	11664	11881
11	12100	12321	12544	12769	12996	13225	13456	13689	13924	14161
12	14400	14641	14884	15129	15376	15625	15876	16129	16384	16641
13	16900	17161	17424	17689	17956	18225	18496	18769	19044	19321
14	19600	19881	20164	20449	20736	21025	21316	21609	21904	22201
15	22500	22801	23104	23409	23716	24025	24336	24649	24964	25281
16	25600	25921	26244	26569	26896	27225	27556	27889	28224	28561
17	28900	29241	29584	29929	30276	30625	30976	31329	31684	32041
18	32400	32761	33124	33489	33856	34225	34596	34969	35344	35721
19	36100	36481	36864	37249	37636	38025	38416	38809	39204	39601
20	40000	40401	40804	41209	41616	42025	42436	42849	43264	43681
21	44100	44521	44944	45369	45796	46225	46656	47089	47524	47961
22	48400	48841	49284	49729	50176	50625	51076	51529	51984	52441
23	52900	53361	53824	54289	54756	55225	55696	56169	56644	57121
24	57600	58081	58564	59049	59536	60025	60516	61009	61504	62001
25	62500	63001	63504	64009	64516	65025	65536	66049	66564	67081
26	67600	68121	68644	69169	69696	70225	70756	71289	71824	72361
27	72900	73441	73984	74529	75076	75625	76176	76729	77284	77841
28	78400	78961	79524	80089	80656	81225	81796	82369	82944	83521
29	84100	84681	85264	85849	86436	87025	87616	88209	88804	89401
30	90000	90601	91204	91809	92416	93025	93636	94249	94864	95481
31	96100	96721	97344	97969	98596	99225	99856	100489	101124	101761
32	102400	103041	103684	104329	104976	105625	106276	106929	107584	108241
33	108900	109561	110224	110889	111556	112225	112896	113569	114244	114921
34	115600	116281	116964	117649	118336	119025	119716	120409	121104	121801
35	122500	123201	123904	124609	125316	126025	126736	127449	128164	128881
36	129600	130321	131044	131769	132496	133225	133956	134689	135424	136161
37	136900	137641	138384	139129	139876	140625	141376	142129	142884	143641
38	144400	145161	145924	146689	147456	148225	148996	149769	150544	151321
39	152100	152881	153664	154449	155236	156025	156816	157609	158404	159201
40	160000	160801	161604	162409	163216	164025	164836	165649	166464	167281
41	168100	168921	169744	170569	171396	172225	173056	173889	174724	175561
42	176400	177241	178084	178929	179776	180625	181476	182329	183184	184041
43	184900	185761	186624	187489	188356	189225	190096	190969	191844	192721
44	193600	194481	195364	196249	197136	198025	198916	199809	200704	201601
45	202500	203401	204304	205209	206116	207025	207936	208849	209764	210681
46	211600	212521	213444	214369	215296	216225	217156	218089	219024	219961
47	220900	221841	222784	223729	224676	225625	226576	227529	228484	229441
48	230400	231361	232324	233289	234256	235225	236196	237169	238144	239121
49	240100	241081	242064	243049	244036	245025	246016	247009	248004	249001
50	250000	251001	252004	253009	254016	255025	256036	257049	258064	259081
51	260100	261121	262144	263169	264196	265225	266256	267289	268324	269361
52	270400	271441	272484	273529	274576	275625	276676	277729	278784	279841
53	280900	281961	283024	284089	285156	286225	287296	288369	289444	290521
54	291600	292681	293764	294849	295936	297025	298116	299209	300304	301401

附錄四

平方表（續）

No.	0	1	2	3	4	5	6	7	8	9
55	302500	303601	304704	305809	306916	308025	309136	310249	311364	312481
56	313600	314721	315844	316969	318096	319225	320356	321489	322624	323761
57	324900	326041	327184	328329	329476	330625	331776	332929	334084	335241
58	336400	337561	338724	339889	341056	342225	343396	344569	345744	346921
59	348100	349281	350464	351649	352836	354025	355216	356409	357604	358801
60	360000	361201	362404	363609	364816	366025	367236	368449	369664	370881
61	372100	373321	374544	375769	376996	378225	379456	380689	381924	383161
62	384400	385641	386884	388129	389376	390625	391876	393129	394384	395641
63	396900	398161	399424	400689	401956	403225	404496	405769	407044	408321
64	409600	410881	412164	413449	414736	416025	417316	418609	419904	421201
65	422500	423801	425104	426409	427716	429025	430336	431649	432964	434281
66	435600	436921	438244	439569	440896	442225	443556	444889	446224	447561
67	448900	450241	451584	452929	454276	455625	456976	458329	459684	461041
68	462400	463761	465124	466489	467856	469225	470596	471969	473344	474721
69	476100	477481	478864	480249	481636	483025	484416	485809	487204	488601
70	490000	491401	492804	494209	495616	497025	498436	499849	501264	502681
71	504100	505521	506944	508369	509796	511225	512656	514089	515524	516961
72	518400	519841	521284	522729	524176	525625	527076	528529	529984	531441
73	532900	534361	535824	537289	538756	540225	541696	543169	544644	546121
74	547600	549081	550564	552049	553536	555025	556516	558009	559504	561001
75	562500	564001	565504	567009	568516	570025	571536	573049	574564	576081
76	577600	579121	580644	582169	583696	585225	586756	588289	589824	591361
77	592900	594441	595984	597529	599076	600625	602176	603729	605284	606841
78	608400	609961	611524	613089	614656	616225	617796	619369	620944	622521
79	624100	625681	627264	628849	630436	632025	633616	635209	636804	638401
80	640000	641601	643204	644809	646416	648025	649636	651249	652864	654481
81	656100	657721	659344	660969	662596	664225	665856	667489	669124	670761
82	672400	674041	675684	677329	678976	680625	682276	683929	685584	687241
83	688900	690561	692224	693889	695556	697225	698896	700569	702244	703921
84	705600	707281	708964	710649	712336	714025	715716	717409	719104	720801
85	722500	724201	725904	727609	729316	731025	732736	734449	736164	737881
86	739600	741321	743044	744769	746496	748225	749956	751689	753424	755161
87	756900	758641	760384	762129	763876	765625	767376	769129	770884	772641
88	774400	776161	777924	779689	781456	783225	784996	786769	788544	790321
89	792100	793881	795664	797449	799236	801025	802816	804609	806404	808201
90	810000	811801	813604	815409	817216	819025	820836	822649	824464	826281
91	828100	829921	831744	833569	835396	837225	839056	840889	842724	844561
92	846400	848241	850084	851929	853776	855625	857476	859329	861184	863041
93	864900	866761	868624	870489	872356	874225	876096	877969	879844	881721
94	883600	885481	887364	889249	891136	893025	894916	896809	898704	900601
95	902500	904401	906304	908209	910116	912025	913936	915849	917764	919681
96	921600	923521	925444	927369	929296	931225	933156	935089	937024	938961
97	940900	942841	944784	946729	948676	950625	952576	954529	956484	958441
98	960400	962361	964324	966289	968256	970225	972196	974169	976144	978121
99	980100	982081	984064	986049	988036	990025	992016	994009	996004	998001

附録五

倒數表

No.	0	1	2	3	4	5	6	7	8	9
10	100000	990099	980392	970874	961539	952381	943396	934579	925926	917431
11	909091	900901	892857	884956	877193	869565	862069	854701	847458	840336
12	833333	826446	819672	813008	806452	800000	793651	787402	781250	775194
13	769231	763359	757576	751880	746269	740741	735294	729927	724638	719425
14	714286	709220	704225	699301	694444	689655	684932	680272	675676	671141
15	666667	662252	657895	653595	649351	645161	641026	636943	632911	628931
16	625000	621118	617284	613497	609756	606061	602410	598802	595238	591716
17	588235	584795	581395	578035	574713	571429	568182	564972	561798	558659
18	555556	552486	549451	546448	543478	540541	537634	534759	531915	529101
19	526316	523560	520833	518135	515464	512821	510204	507614	505051	502513
20	500000	497512	495050	492611	490196	487805	485437	483092	480769	478469
21	476190	473934	471698	469484	467290	465116	462963	460830	458716	456621
22	454545	452489	450450	448431	446429	444444	442478	440529	438597	436681
23	434783	432900	431035	429185	427350	425532	423729	421941	420168	418410
24	416667	414938	413223	411523	409836	408163	406504	404858	403226	401606
25	100000	398406	396825	395257	393701	392157	390625	389105	387597	386100
26	384615	383142	381679	380228	378788	377359	375940	374532	373134	371747
27	370370	369004	367647	366300	364964	363636	362319	361011	359712	358423
28	357143	355872	354610	353357	352113	350877	349650	348432	347222	346021
29	344828	343643	342466	341297	340136	238983	337838	336700	335571	334448
30	333333	332226	331126	330033	328947	327869	326797	325733	324675	323625
31	322581	321543	320513	319489	318471	317460	316456	315457	314465	313480
32	312500	311526	310559	309598	308642	307692	306749	305810	304878	303951
33	303030	302115	301205	300300	299401	298508	297619	296736	295858	294985
34	294118	293255	292398	291545	290698	289855	289017	288184	287356	286533
35	285714	284900	284091	283286	282486	281690	280899	280112	279330	278552
36	277778	277008	276243	275482	274725	273973	273224	272480	271739	271003
37	270270	269542	268817	268097	267380	266667	265957	265252	264550	263852
38	263158	262467	261780	261097	260417	259740	259067	258398	257732	257069
39	256410	255754	255102	251463	253807	253165	252525	251889	251256	250627
40	250000	249377	248756	248139	247525	246914	246305	245700	245098	244499
41	243902	243309	242718	242131	241546	240964	240385	239808	239234	238663
42	238095	237530	236967	236407	235849	235294	234742	234192	233645	233100
43	232558	232019	231482	230947	230415	229885	229358	228833	228311	227790
44	227273	226757	226244	225734	225225	224719	224215	223714	223214	222717
45	222222	221730	221239	220751	220264	219780	219298	218818	218341	217865
46	217391	216920	216450	215983	215517	215054	214592	214133	213675	213220
47	212766	212314	211864	211416	210971	210526	210084	209644	209205	208768
48	208333	207900	207469	207039	206612	206186	205761	205339	204918	204499
49	204082	203666	203252	202840	202429	202020	201613	201207	200803	200400
50	200000	199601	199203	198807	198413	198020	197628	197239	196850	196464
51	196078	195695	195312	194932	194553	194175	193798	193424	193050	192678
52	192308	191939	191571	191205	190840	190476	190114	189753	189394	189036
53	188679	188324	187970	187617	187266	186916	186567	186220	185874	185529
54	185185	184843	184502	184162	183824	183486	183150	182815	182482	182194

附錄五

倒數表（續）

No.	0	1	2	3	4	5	6	7	8	9
55	181818	181488	181159	180832	180505	180180	179856	179533	179212	178891
56	178571	178253	177936	177620	177305	176991	176678	176367	176056	175747
57	175439	175131	174825	174520	174216	173913	173611	173310	173010	172712
58	172414	172117	171821	171527	171233	170940	170649	170358	170068	169779
59	169492	169205	168919	168634	168350	168067	167785	167504	167224	166945
60	166667	166389	166113	165838	165563	165289	165017	164745	164474	164204
61	163934	163666	163399	163132	162866	162602	162338	162075	161812	161551
62	161290	161031	160772	160514	160256	160000	159744	159490	159236	158983
63	158730	158479	158228	157978	157729	157480	157233	156986	156740	156495
64	156250	156006	155763	155521	155280	155039	154799	154560	154321	154083
65	153846	153610	153374	153139	152905	152672	152439	152207	151976	151745
66	151515	151286	151057	150830	150602	150376	150150	149925	149701	149477
67	149254	149031	148810	148588	148368	148148	147929	147710	147493	147275
68	147059	146843	146628	146413	146199	145985	145773	145560	145349	145138
69	144928	144718	144509	144300	144092	143885	143678	143472	143267	143062
70	142857	142653	142450	142248	142045	141844	141643	141443	141243	141044
71	140845	140647	140449	140252	140056	139860	139665	139470	139276	139182
72	138889	138696	138504	138313	138122	137931	137741	137552	137363	137174
73	136986	136799	136612	136426	136240	136054	135870	135685	135501	135318
74	135135	134953	134771	134590	134409	134228	134048	133869	133690	133511
75	133333	133156	132979	132802	132626	132450	132275	132100	131926	131752
76	131579	131406	131234	131062	130890	130719	130548	130378	130208	130039
77	129870	129702	129534	129366	129199	129032	128866	128700	128535	128370
78	128205	128041	127877	127714	127551	127389	127227	127065	126904	126743
79	126582	126422	126263	126103	125945	125786	125628	125471	125313	125156
80	125000	124844	124688	124533	124378	124224	124070	123916	123762	123609
81	123457	123305	123153	123001	122850	122699	122549	122399	122249	122100
82	121951	121803	121655	121507	121359	121212	121066	120919	120773	120627
83	120482	120337	120192	120048	119904	119761	119617	119474	119332	119190
84	119048	118906	118765	118624	118483	118343	118203	118064	117925	117786
85	117647	117509	117371	117233	117096	116959	116822	116686	116550	116414
86	116279	116144	116009	115875	115741	115607	115473	115340	115207	115075
87	114943	114811	114679	114548	114416	114286	114155	114025	113895	113766
88	113636	113507	113379	113250	113122	112994	112867	112740	112613	112486
89	112360	112233	112108	111982	111857	111732	111607	111483	111359	111235
90	111111	110988	110865	110742	110620	110497	110375	110254	110132	110011
91	109890	109769	109649	109529	109409	109290	109170	109051	108933	108814
92	108696	108578	108460	108342	108225	108108	107991	107875	107759	107643
93	107527	107411	107296	107181	107066	106952	106838	106724	106610	106496
94	106383	106270	106157	106045	105932	105820	105708	105597	105485	105374
95	105263	105152	105042	104932	104822	104712	104603	104493	104384	104275
96	104167	104058	103950	103842	103734	103627	103520	103413	103306	103199
97	103093	102987	102881	102775	102669	102564	102459	102354	102249	102145
98	102041	101937	101833	101729	101626	101523	101420	101317	101215	101112
99	101010	100908	109806	100705	100604	100503	100402	100301	100200	100100

參考書

中文

1.統計原理與方法

王仲武著　統計學原理及應用　商務印書館出版

王仲武著　統計公式及例解　商務印書館出版

金國寶著　統計學大綱　商務印書館出版

金國寶著　統計新論　中華書局出版

孟　森譯　統計通論　商務印書館出版

甯恩承譯　統計方法　商務印書館出版

趙文銳譯　統計學原理　商務印書館出版

陳其鹿著　統計學　商務印書館出版

壽景偉著　應用統計　商務印書館出版

閒一夔著　統計學概論　民智書局出版

唐啓賢著　統計學　黎明書局出版

許炳漢譯　統計方法概論　北新書局出版

陳炳權著　統計方法　大東書局出版

蔡毓聰著　統計學ABC　世界書局出版

2.數理統計

艾　偉著　高級統計學　商務印書館出版

3.社會統計

毛起鵕著　社會統計大綱　黎明書局出版

毛起鵕著　社會統計　世界書局出版

言心哲著　社會調查大綱　中華書局出版

李景漢著　實地社會調查方法　星雲堂書店出版

樊　弘著　社會調查方法　商務印書館出版

4.教育統計

朱君毅譯　心理與教育之統計法　商務印書館出版

朱君毅譯　教育統計學綱要　商務印書館出版

朱君毅著　教育統計學　商務印書館出版

朱君毅著　教育測驗與統計　商務印書館出版

郜爽秋著　教育圖示法　教育印書合作社出版

周調陽著　教育統計學　中華書局出版

劉廼敬著　實用統計學　南京書局出版

5.經濟統計

林光澂著　商業統計　商務印書館出版

金國寶著　物價指數淺說　商務印書館出版

楊西孟著　生活費指數編製法　商務印書館出版

楊西孟著　指數公式總論　商務印書館出版

趙人儁著　物價指數論提要　國定稅則委員會出版

英文

Bowley, Arthur L., "An Elementary Manual of Statistics," revised edition, 1920, P.S. King and Son, London.

"Elements of Statistics," revised edition, 1926, Scribner's.

Brinton, Willard C., "Graphic Methods for Presenting Facts," 1914, The Engineering Magazine Co., New York.

Chaddock, Robert Emmet "Principles and Methods of Statistics," 1925, Houghton Mifflin.

Crum, William Leonard, and Patton, Alson Currie, "An Introduction to the Methods of Economic Statistics," 1925, A. W. Shaw.

Davenport, C. B. "Statistical Methods, with Special Reference to Biological Variation," revised edition, 1914, John Wiley.

Davies, George R., "Introduction to Economic Statistics," 1922, Century.

Day, Edmund E., "Statistical Analysis," 1925, Macmillan.

Dittmer, Clarence G., "Introduction to Social Statistics," 1926, A. W. Shaw.

Elderton, W. Palin. "Frequency-Curves and Correlation," revised edition, 1927, Charles and Edward Layton, London.

Elderton, W. Palin and Ethel M., "Primer of Statistics," revised edition, 1923, A. and C. Black, London.

Fisher, Irving, "The Making of Index Numbers," revised edition, 1927, Houghton Mifflin.

Jerome, Harry, "Statistical Method," 1924, Harper.

Karsten Karl, G., "Charts and Graphs," 1923, Prentice-Hall.

Kelley, Truman L., "Statistical Method," 1923, Macmillan.

King, Willford I., "The Elements of Statistical Method," 1912, Macmillan.

Mills, Frederick C., "Statistical Methods Applied to Economics and Business," 1924, Henry Holt.

Riggleman, John R., "Graphic Methods for Pressenting Business Statistics," 1926, Mc Graw Hill.

Rugg, Harold O., "Statistical Methods Applied to Education," 1917, Houghton Mifflin.

Secrist, Horace, "An Introduction to Statistical Methods," revised edition 1925, Macmillan.

Secrist, Horace, "Readings and Problems in Statistical Methods," 1920, Macmillan.

Yule, G. Udny, "An Introduction to the Theory of revised edition 1924, C. Griffin and Co., Lon

新到書告

社會調查大綱

一冊　一元二角

言心哲著

全書共分兩編：第一編爲總論，敍述社會調查的原理及方法，於社會調查之性質，社會調查之歷史，社會調查之步驟，社會調查之組織，調查談話，圖表編製，以及各種實地調查方法，均詳爲論列。第二編爲分論，關於人口調查，教育調查，犯罪調查，失業調查，農村調查等，皆有專章論及。書中收集關於我國社會調查之材料甚多，書末並附有參考書目，搜羅宏富，內容完備，甚合於學校教本及專家參考之用。

中華書局出版

民國二十四年九月印刷
民國二十四年九月發行

統計概論（全一冊）

◎定價銀七角五分
（外埠另加郵匯費）

著者　芮寶公

發行者　中華書局有限公司
代表人陸費逵

印刷者　上海靜安寺路
中華書局印刷所

總發行所　上海棋盤街中華書局

分發行所　各埠中華書局

（九二五八）