萬有文庫
第二集七百種
王雲五主編

統計學原理

(一)

鮑萊著
李植泉譯

商務印書館發行

萬有文庫

第二集七百種

王雲五主編

統計學原理

(二)

鮑萊著

李植泉譯

商務印書館發行

統計學原理

(一)

鮑萊 著

李慎泉 譯

漢譯世界名著

統計學原理

（二）

鮑萊 著

李植泉 譯

漢譯世界名著

統計學原理

(二)

鮑萊 著

李愼泉 譯

漢譯世界名著

統計學原理

(四)

鮑萊著

李慎泉譯

漢譯世界名著

王雲五主編
萬有文庫
第二集七百種
統計學原理
四冊
Elements of Statistics

中華民國二十六年三月初版

原著者 Arthur L. Bowley
譯述者 李植泉
發行人 王雲五 上海河南路
印刷所 商務印書館 上海河南路
發行所 商務印書館 上海及各埠

韋

萬有文庫

第二集七百種

總編纂者
王雲五

商務印書館發行

第一編

目錄

插表目錄

插圖目錄

第二編

目錄

附錄　數學摘錄

統計學原理

第一編　基本統計方法

第一章　統計學之意義及其範圍

統計學之定義　統計學一詞,各方所下定義甚多,對於統計學之範圍,統計學著作家,復各有不同之見解。依本書之主旨,雖無區文別字之必要,但顧其名而思其義,理應加以解釋,統計學研究之範圍如何,亦應討論及之。茲爲便於研究計,請就常見之定義討論始。

計數學　舉例言之,統計學亦可名之爲『計數學』(Science of Counting)。吾人初聞此言,必以爲計數者,本爲輕而易舉之事,任人皆優爲之,卽用機械之力,亦能自動算出,何待乎學?詎知事實上,使遇極大之數,如一國之人口者,則計數初非易易,個人力量絕難應付裕如;時間也,空間也,無往不爲計數之障礙,且數目過大,逾某種限度時,欲求絕對之確實,勢乃有所不能矣。

算學與統計學之分野　極大之數,枚計之時,欲求每一單位

各個均確實無誤，乃不可能，實則只有大略估計一途。由此一點，吾人可知算術與統計學之分野：用算術得來者，鑿確無疑，統計則僅事估計而已，統計雖有時，亦頗確實，足供其用途之需，然確實程度，去數學遠甚。

統計乃合作之計數 統計所涉及之數目，大都甚巨，即用估計方法，亦斷非一人之力，所能勝任，必也憑藉有組織之團體，集多數人之合作，始克有功。數字之搜集，加入工作者，雖有多人，得來數字之計算，雖僅爲算術問題，然實際進行之時，必遭遇兩種困難。第一，調查事項之定義，不易解說明白，使調查員在訪問之時，何者應加詢問，何者應行放棄，有一定之原則，可資遵循也。蓋單簡之調查對象，如房間數，如年齡數，甚或以個人爲對象，其定義若何，均不免有疑問發生，調查表上雖有寥寥數語關於類別之說明，但一類之內，究包括何等項目(item)，殊難解釋明白也。第二，工作人員衆多，數字上之錯誤，必所難免也；蓋人數一多，則疏忽者有之，愚魯者有之，謄錄之筆誤，列表之誤謬，在所不免也。由此觀之，經多數人之合作，得出之總數，欲求其精確無疵，是誠難矣。雖然，如上所述之估計，固有時亦併入統計學一詞之內，然如謂估計卽爲統計學之定義，實亦無乃失之過濫，且此一定義，並未將統計方法之特殊性質說明也。

統計學乃一種方法 實際上，統計學之定義，不妨用歸納法

(aposteriori) 規定之。在遇有大量數字，描寫羣類之大數，以及關於異時異地之總計數或平均數之數列時，均不得不用特殊之方法——此種方法，一則須以大數之特別性質爲依據，一則須能論敍複雜之羣類，以便使人領會，三則須能分析結論之確度(accuracy)，四則須能測量相差之重要性 (significance of difference)，五則須能將估計之數，彼此互作比較。適用此種方法之估計，乃卽屬於統計學之範圍；本書之主旨，卽在研究此等方法也。

二 **統計方法應用之普遍性** 在此定義之下，須知統計學，既非經濟學之分支，更非任何學科之附屬物。統計學一科，其地位恰似外國語或代數學：在任何時間任何環境之下，均有其效用。

在自然科學上之應用 論及統計方法與各科學術之聯繫，乃至有興趣之問題。試先就自然科學方面論之：統計方法與天文學有接觸者兩點。第一，天文學家爲測量某星位，從若干微有差異之觀察中，急於求得其最優值，因而採用最小二乘法(the least square method)。蓋物體測量，每一數量必須觀察若干次數，然各個觀察不論如何謹慎，終不能使其絕對相合。物體測量如此，社會學上之觀察亦然，雖數字來源相同，而各統計家所求得之平均數，彼此一律甚難。對於此類之數量，必須就其最或然值歸納之；惟其如此，乃不得不應用差誤律 (Law of error)，而此則又

非藉重最小二乘法不可者也。

逐步改進之確度 第二，統計方法與天文學方法之相似點，及其與地質學及與其他實用科學所用方法相似之點，大致相同。科學測量之研究，須先就一數量，作草率之觀察，如地球與太陽之距離也，地質層之厚度也，元素之原子量也，物體之比重也，莫不皆然。既有多次關於數量之觀察，更進而求工具之精確，方法之改良，於是觀察之數量，乃逐漸達於確實。如此由粗而細，由細而精，逐步進展之過程，甚為重要，蓋以吾人今日之知識階段，許多統計測量，往往因資料缺乏，以致不能精確作成，故批評者因此而否認先事估計之價值；惟自科學觀點論之，此種批評不無相當錯誤，蓋測量既依據論理原則，與其一無所有也，何妨有錯誤之測量？況且測量之差誤，果有限度可尋，仍不難逐步改進也。

統計學與生物學 統計調查與一切之科學實驗，相似情形，上已略述一二，現請更進而考查統計在生物學上之應用。在皮爾生(Karl Pearson)教授，未發表其所作之調查(註一)以前，進化與遺傳之全部理論，謂其有統計基礎之實在性，尚未為人所公認。但自彼之調查發表以後，數理統計方面極重要之新貢獻，乃相繼出現於此途矣。然此問題之性質如何？茲請略一言之。設有許多觀察於此，譬如以人體之身長為喻，從此觀察數字得出一代表形態，即平均數，以此平均數為中心，將所有數量，依一定之法則，

組成羣類。然則吾人現在之問題，卽在考察平均數，或依平均數組成之羣類，二者何一發生變動耶？其變動爲何如耶？例如生物界之變化，歷經若干年代之考查，始知差異之所在，由此差異之點，乃構成進化論所根據之資料。此乃應用統計方法之效果，如動植物之化石，如動物之種族，以及其他多種之羣類，亦莫不恃此以研究之。設無此種工具，許多已成立之論辯，大部之力量，必將爲之失去，然則所謂學理者，將僅以個人對現象之印象爲基礎矣，尙有何科學測量可言哉！故在此方面應用之方法，可轉爲研究社會問題之一助。平均工資，依平均工資組成之羣類，以及此種數量之變動，亦成完全相似之問題矣；各種因素之效果，彼此所成之相關，亦可用一數學公式計算之矣；實際上，除此之外，欲測量複雜羣類之數字變動，恐更無其他確實之方法矣。關於人體測量學之試驗，幾多有價值之資料，均恃統計方法，爲之搜集，於是統計學家對事實之知識多有增加，若干重要之原理乃因是而出現矣。

統計學與氣象學　氣象學與統計學相似之處甚多。所有關於氣象學之測量，不外溫度、氣壓、空氣溼度、及風力數種。研究此問題，必須集多量之觀察，查其代表形態——平均數，並測量其變動之情況。夫用以記載各年平均溫度之表格，與英國人事登記總監(Registrar-General)發表之出生、死亡、結婚數字，本多

相同。設無統計方法可資應用，各種數列之平均數雖可求出，然亦不過仍爲數字而已，焉能引出合理之結論乎？今有統計學，各年數字之變動，其爲偶然，抑甚重要，其爲繼長增高，抑或週期變化，以及是否合乎何種法則，均不難應刃而解矣。然則由此推而論之，欲進而用以預測將來之人口數變動，及其他類似之問題，甚爲重要，誠昭然若揭矣。

統計學與人口學 卽以統計學與人口學(demography)而論，設人口學之內容，不僅限於人口數、出生率、結婚率、死亡率，以及依年齡、性別、地域分之分配——卽人口普查及人事登記總監之報告數字——且包括所有產業及社會調查，卽依業別分之人口分配，與依所得，工資、物價、生產、國外貿易，以及運輸等等之分配數字而言，則所謂人口學者，凡治社會學經濟學者有直接關係之大部統計調查，均將包括在內。人口學之正確範圍如何，姑置不論，於此可另得一統計學之定義曰：統計學者，乃測量整個社會有機體而從各方表明之之科學也。

統計學對於整個社會有機體之關係 嘗閱一論文，文中依勒布雷(Le Play)之方式，曾作一家庭之調查，記載家中每人之職業及收入，消費之方式，此一家庭之經濟地位，但此項研究，實不合統計原則。人口學之研究，乃集多類之家庭，作一種數量之觀察，從事某一產業之家庭，共有若干？其平均進款、消費、儲蓄、

幾何？此統計學之所事也。論文所用方法，以個體爲萬物，而統計方法，視個體若無物，性質所限也。當吾人測量羣類之時，個體之特質如何，概不過問，然若特質爲多人所同具，斯乃變爲重要矣。統計學直可謂之平均數學。測量繁雜之羣類，例如收入與工資，藝術家以一夕之工作，能得百鎊之金，而拙笨之勞工，一日之所得，僅爲六便士，然對一般平均數，影響甚微，彼等屬於一組類也；惟在一九一四年之前，有技能之工匠，大部每星期能得四十先令之收入，而大部無技能之勞動者，每星期所得尙不足十五先令，此二種收入之統計，必須作各別之觀察。至於確切之分類，乃爲一程度問題。二者之不同，乃所作特種調查之性質有異也。總之，遇有繁雜之羣類，欲以統計方法，而加以估計時，其目的乃在舉示其概略，期以一隅而概其全體意義而已。其表示之法，切忌細述情節，只須如畫家之繪一樹木，用烘托方法已足，無須分清枝葉也。用概略舉示方法，固每有含糊不清及稍欠準確之弊，然詳細情形，有成竹在胸，或一目可以瞭然，所得印象，必無錯誤之虞。蓋此種方法，其中含有重要原理：羣類之各個分子，變化者漸，全類變化者亦必緩。各個原子之運動，測驗固甚困難，一固體之運動法則，判明當較易易。多量數字，以及由數字得來之平均數，例如社會現象之量數，均有極大之惰性，存乎其中。人口總數，收入總數，出生及死亡率，平均工資等等之變化，莫不甚微，而關於

一家一戶之類似數量，則更動甚劇。統計測量之所以有效者，即恃此大量數字之恆性，而統計測量之應用，亦唯對於大量數字，統計方法始能有所發揮也。

統計學與經濟學 統計學與經濟學之關係，至爲單純，誠如馬夏爾(Marshall)教授所云，『統計資料爲經濟學家之根據』。統計學家供給事實，而經濟學家乃以之爲學說之測驗，或竟以之爲基礎。蓋經濟學家從事者，爲關於羣類之現象，個體事實乃羣類之一分子，欲得其研究之材料，非藉重平均數學——統計學——不可也。討論國家經濟之時，凡貿易量，貨幣購買力等問題，欲充實純粹之理論，又非大數科學——統計學——供給事實不可也。化學之實驗，類乎統計，化學原理之研究，猶之經濟學。故統計家負搜集排列整理分析之責，而不當歸納之任；即在調查關於因果現象之時，亦只舉出證據而已，結論無須求也。然若以統計學家而兼經濟學家，則舉行實驗，配合學理，二者集於一人之身，設非徹底明瞭實驗方法，及其難點所在，或本人缺乏機巧，不能卒底於成，則只研究理論，必難期以成功，故研究經濟學者，一則非精通統計方法；二則須洞悉難點爲何；三則可決定數字何處方可正確；四則能以批評統計證據；五則鑑別估計之信度不爲功。

統計學與個人經驗 統計之功能，即在擴大個人之經驗，蓋個人所見至爲有限，僅爲社會有機體某一部門之一小部；聽聞固

能增長知識，然所聞亦不外於友朋談話間，閱報章雜誌，或專家著作幾途。即使依判斷力，亦能就人羣或物類，得知數字之重要；然欲其不受環境之支配，而免其偏頗之見，乃絕不可能，且其所得資料，欲求分析正當，而期確實，亦大為不易也。故苟欲就之而加以詳盡之考查，乃即入於統計調查之範圍，即進行中間所不可免之難題，則亦非統計學不能解救。統計學之用途，否認者固多，於此大可打破其固執。統計之估計，或善或劣，或確或僞，固不敢必，然勿論如何，統計數字必較之觀察之偶然印象，確實多多，而事實之性質，亦惟用統計方法，始能發抉真象也。

統計學為比較的　統計學主要實際用途之一，即在示事實之相對重要性，而個人則只能妄加臆斷者也。統計學之特質，即在便於比較。某一數量之絕對數值，本無意義可言，必有數種相似之數量，然後乃可加以比較，於是乃有意義發生矣。例如英國之乞丐若干，並無陳說之價值，今若知英國全人口數，乃大不相同矣。又如倫敦東區每人飲水供給加侖數一言，設無其他市鎮之給水量，殊無些須之意義。即以工資調查之平均工資而言，亦須另有空間不同或時間不同之其他工資數，全部意義始能釋然；可見統計數字，無往不需其他之比較；非然者，雖有一種數量，其作用如何，殊難領略也。

公務統計　設欲測量之物件，羣類甚廣，非吾人獨力，或某

一機關，所可舉辦。固然，私人團體，舉辦之調查，昭著成效者，所在多有　例如布斯(Booth)氏之『人民之生活及勞動』(Life and Labour of the People)，里昂萊維(Leone Levi)氏之『工資與收入』，朗垂(Rowntree)氏之『貧乏』(Poverty)，即本著者及倍乃提赫斯特(Burnett-Hurst)，亦曾用抽樣方法（例如在『生計與貧乏』(Livelihood and Poverty)），研究可以控制之數量——然就一般而論，測量一部社會或工業團體，苟欲求結果完好，必須由中央政府或地方政府舉行之，而此亦卽統計事業之龐雜與不全所由起也。蓋數字資料之收集，乃政府本身職責所關，人口幾何，土地面積總數若干，細數若干，各級政府，爲其本身目的計，不可不知；多量之數字羣類，由官廳登記卽可得之。故政府之數字，均另有其主要之目的，至於用作統計之研究，乃其副作用耳；政府經手事項，均須分別記載之，以爲管理備查之需，產業之登記，亦有其特別管理之用，凡此種種數字，大部多須公布也。於是用此公布之數字，吾人乃得個人收入統計、教育統計、進口貿易統計、鐵路統計、礦產統計、工廠統計等等。雖有少數數字，純爲科學研究，而搜集得來，然多有爲行政目的而頒布之表册，同時亦可爲他種研究之用，研究社會學者尤不可少，卽人口普查之材料，亦多由此項得來。如就英國而論，試一翻閱各年「英國統計一覽」(Statistical Abstract for the United Kingdom)，

「勞工統計年報」(Annual Abstract of Labor Statistics),及人事登記總監 (Registrar-General) 之常年報告;各種刊物,非由公務統計撮要而成,即係以公務統計作參考也。

公務統計之不完善 爲行政目的而搜集之數字,不若研究社會學與經濟學者所用資料之精確,事乃至爲顯然。即使政府與學者之願望,大致相同,其分類與列表,依然未必適合科學之條件。近年以來,非純爲行政目的而調查之資料,確已進步多多,例如商業部之勞工局(現已併歸勞工部)之統計,即屬此類。然尙待改進之處,猶有極多,因如此雖不免多所耗費,而以與全國收入相比,實微末不足道也。第一應改進者,人口普查 (Census)——至少一部工作——應改爲五年舉行一次;第二,工作人員,已往均十年組織一次,統計整理就緒,發表報告以後,即行解散,今後應改爲常設機關,使承辦大規模之全國他種調查。如是,多種主要商品之批發價格與零售價格,不難列表加以分析而發表矣;國內陸路運輸,亦可與國外水路貿易,同樣列成表格;不致只知國外貿易而不知國內貿易矣;國內生產統計,亦不致僅拘拘於農業、礦業、鋼鐵業,而擴張及於一九〇七年生產統計所列之全部,於是全國重要產業之每年出產量,均不難得知矣。

中央統計機關 總而言之,欲求調整所有現行統計,或直接舉行調查,或指派相當機關執行,以期完成全國連續統計記載,

非設立中央統計機關不爲功。其實，經調整之調查，必須使其精益求精，並推廣其範圍，此種迫切之需要，不待深通經濟學與統計學而後知，中央調整機關之有無，利害昭彰，例證甚多，俯拾卽是。當研究國民所得之時，吾人可得者，只有工資統計，應納稅之所得統計；但屬於免稅之薪金，以及在國外投資所得之一部，則無從探悉，有之，只得藉助於估計方法而已。對於貨幣購買力之變動，固可由「經濟雜誌」(Economist) 及商業新聞，得知批發價格之情況，但因零售價格記錄不全，許多有趣之研究，只得付之闕如。關於工資狀況，尙可推算實際標準及平均工資，但因產業普查之缺乏，旣未能曉其某級工資共有若干人數，復難得知僱傭階級與被僱傭階級之相對人數。設將來有公衆需要此種資料之一日，必須有極開明之政府，犧牲時間，不避煩難，提供爲數並不爲多之金錢，以便以有計劃之嘗試，以補今日之不足；但所謂開明，並不全賴政府，任何人均有其應盡之職責，而欲促進此需要，又非借鏡他國之成規，並建立統計學識不可也。

統計之不能使人盡信　現今需要之所以未爲大衆所感覺者，夷考其故，未嘗非由於統計之已失人信仰之結果，此種情形，雖甚普遍，但未見其盡不合理，由俗諺『萬事萬物，統計皆能證之』一語可見一斑。惟此乃一般批評者之本身錯誤：蓋大衆所根據之資料，供給者爲新聞紙。

不可信之原因　但新聞紙上之統計，並非完善之統計；報告者既恃不正確之估計，大衆又無辨別之能力，既不知其是否用恰當之資料而估計得來，亦不知何者確可用統計估量方法。且也，報紙有時斷章取義，以工資調查平均數，及「勞工公報」(Labour Gazette) 失業人數等等資料之一部，作全體事實看待；殊不知估計之數字，本有其用途，必須適合其目的，乃可謂之正確而完全；若以此不可靠之前提，雖於選擇資料時，已盡精審機敏之能事，然亦不宜任意用爲發表主張之根據。總之，統計學家所得之結果雖被人用爲發表主張之基礎，但因此而認爲統計學家之過失，無乃失之過當。蓋以統計爲論據之錯誤主張，普通多由於：（一）忽略數字之上下文，不顧材料之是否完全，竟然斷章取義，或取某一羣類之數字，強用於現象完全不同之另一羣類；（二）以對於羣類一部之估計，視爲全部之數字；（三）僅採用合乎主張之片面數字，而忽略其餘；（四）自效果立卽論及原因，不問事實，此項差誤，統計學上最常發生。綜觀以上四種，均爲論理上之基本錯誤，統計學亦應加注意者也。

統計上之限制　統計工作上有若干限制，恐統計學家未必常能深切認識。一現象，最多只能就現象之數字部分，加以度量；至度量事實之時，更不應以將所欲度量者加以度量爲滿足，其相聯之數量，亦須度量之。例如，欲度量貧乏現象之緩急伸縮，因貧

乏二字，無從下以定義，更無從加以度量，而貧民之人數，又不能計數，吾人所能爲力者，只能就官方所認定之貧民人數，加以私人方面之估計，從事說明之；然貧民人數之密度，仍不得而知也。又如欲從事健康統計；所能着手之主要數量，僅有死亡率，平均壽長，及最流行之數種疾病，數項均爲互不相連之問題。統計學家對社會學之貢獻，即爲客觀之度量，而此又係資料中較不重要之部分，然此較不重要之部分，對於解決社會問題，其功用恰與建築房屋所用之度量，同爲確實之度量也。

（註一） 參閱科學綱要(The Grammar of Science)一九〇〇年版，自第十章起，參考書目，亦見該書。

第二章　統計調查方法

統計調查，驟觀之似無共同之方法，而實際上各種程序又相差甚遠，欲舉出一致之綱領，概括一切，自非易易，下列所舉(註一)，僅爲一般所用，但至少尙能將各種方法，通盤加以硏究也：

（一）材料之搜集，

（二）材料之列表，

（三）提要，

（四）結果之評判。

前三項，容俟以下分章詳細討論之。

應具之預備智識　現請先論學習統計方法者，應有之要件。搜集材料與製列表格之時，常識最爲緊要，而經驗尤爲主要之導師；簡單之算術，實際應用之時，熟練最爲緊要；但調查之各個部分，彼此因有連繫，在進行一部之時，應對整體了解爲宜。提要之時，各項算法，雖無須代數符號，但代數平均數，仍以善於利用爲便，且爲解說曲線計，對幾何亦應有充分之認識。估計數字之評判，與整理結果之解說，必須應用較高深數學之公式，於是公式之來源及應用，自以明瞭爲宜。此種公式，於(1)比較複雜羣類，(2)估量對平均數差度之重要性，或時間數列之離中差量時，需

用特多，而研究相關時，尤爲緊要。

(一)搜集材料：空白表格　搜集材料，普通多用散發空格調查票辦法，空白票或由調查員代填，或由被調查者自填，而空白票內容之適當籌劃，乃即一完善調查中主要難題之一。在未發出調查票之先，必須釐定全部之工作計劃，且調查結果將得何等數字，事先亦應熟加籌慮，以期規定編制細節，可以正確適用，並調整應用工具，庶免臨時有誤。前已論及，吾人所欲度量之物體，未必全可施以度量，必也求出其關係最切近，而可以度量之數量。例如欲問勞動階級之每年平均收入，其每週平均工資，乃不可不首先度量之。次在研究某一特別問題，其要件須具有相當之專門知識；不然，爲測驗調查方法以增經驗計，必須事先舉行小規模調查。所欲調查之人物，務須窮極搜索，巨細靡遺，並事先即加詢問。詢問之問題，以能使得來之答案，立即可供列表之用爲宜，故列表計劃，事先亦應規定妥貼。問話之決定，以明瞭易答不致誤解爲主，答案必使十分確定，庶答者可以『然』『否』作答，而免含混之弊。又問時不可觸人之所忌，以免引起反感，不可有窮究查問語氣，以免引人猜疑，問話務能斬釘截鐵，以免答非所問，必須包羅周詳，以免掛一漏萬。票上須附必要之說明，以爲問答之標準，而免發生錯誤，但又不可過繁；俾免不便遵循。確切之程度如何，尤須先事定奪，時日之問題，答案以至月數爲止境乎，

抑日數爲止境乎，工資收入等問題，答案以至元爲止境耶？抑以角分爲止境耶？此外，甚至空白票上之文字及空間大小，雖一小至每方寸之地位，或一字之微，亦須煞費考量，蓋不慮於始，鮮克有成，與其事後無法補救，無寧籌議於事先，故調查表格事先妥爲制定，乃一勞永逸之計也。

（二）列表法　列表方法，不一而足，須視問題之性質而定。並在各欄填列相當之總計數，以針對問題之要領。所調查之問題，人數若干，工資幾何，價格多寡，性質不一，列表方法，乃不得不隨之而異。人口普查，製表多藉機械之力；工資調查，重在各業對平均工資之散佈情況，故須用特別方法；貿易統計，採用細類極多，各類又均有煩雜之問題，不可一概而論。總而言之，調查之計劃，各類各有其特點，不可不知；而列表結果所示之總計數，必須表示各項之數目，蓋唯有精重表列法，乃可將備極混雜之變動事件，以簡馭繁，條分縷析，於是全類現象，乃可瞭如指掌也。

（三）平均及扼要　原始資料，整理至此，即可從事分析工作。分析之時，第一須從所得數中，提出有效數字；第二求出其總數及平均數，以便得其全部概念；第三以簡短語句，總扼資料之精華；第四一則須用全部資料歸納得出幾個平均數，二則須將所得少數平均數之含義，以簡要明達之文字一表而出之，此總括之結論，即爲一般所引用者也。以上分析工作，欲期其精巧，必須熟知

平均法則，並利用圖式。第五，欲補充數字之缺陷，插補中間年份之估計數字，必須應用內插補法，惟此法危險性大，用時不可不慎。第六，最後卽用文字敍明其過程——發端及結局，並估計其確度，然後調查工作，乃告完全矣。

(四)結果之評判 實習統計者，請勿以既得調查結果爲滿足，務須洞悉方法之種種步驟。在未評判一調查結果之前，一則須備悉數字之來源，是否包括全部數字；再則須知表列之資料，究自何處得來；三則須知調查之時，是否有偏見之存在；四則須知各個答案正確至何種程度；五則須知被調查者有關之事實，其本人是否確知無疑，且卽使所知確實，能否據實以告。故調查結果作成報告發表之時，應將蒐集資料全部計劃，逐步敍明，尤須附列原始空白調查票樣張，以便據以評判，而斷定列表結果，是否確由原始答案而來。蓋極爲有用之評判，常以內在證據(internal evidence)爲有效。各類實得之數，是否與其他位之重要性成比例，特別重要之數字，是否並無充分之證據，均不難探知；數字之連續性，可以檢查，驟然中斷之原由，亦可察出；已塡妥之調查表格，可以依樣本分爲數類，然後考查各類與整個結論適應程度，則調查表格是否代表甚爲普遍之數字，從是可知矣；精審研究詳細之表格，最末數字之正確至何程度，可以揭曉。故以評判態度爲此等檢查，則所得資料，是否足以求出精確之結果，吾人

不難得知，於是省略數字之影響，資料之不足，欲行估計其效果之重輕，吾人不可不加之意矣。

統計之最重要功能，在舉表明各類現象相互關係之證據；蓋所得之消息，可謂爲行動之指南針，欲知行動必如何而後可產生期望之後果，不能不用此指南針，而欲用此指南針，則唯有探討既往所生效果之緣由也。因此之故，一量數之變動，是否引起其他數量之變動，此乃必須決定之問題，此問題雖不必易於解答，但自應用數學即機率數學(mathematics of probability)方法，以研究統計學以來，此問題之解決，並非絕無辦法，實際言之，近年統計學之進步，主要實受此門數學之賜也。

此種問題，重要固重要矣，然深奧難解，普通學生無相當專門學識者，曷克勝解決之任。故本書編訂計劃，擬將需用專門或數學學識之各種問題，留待第二編討論，本編所研究者，乃無需特別高深訓練者也。

（註一）　例如白狄雍博士 (Bertillon) 著『統計學初級教本』(Cours élémentaire de Statistque)，即係如此，本書下文所論採用之處頗多。

第三章　單位之定義與資料之搜集

第一節　單位之定義

調查伊始，第一問題，『將計數者何』，迨列表旣成，最後又有一問題『已計數者何』。前一問題之答案，卽開端之定義，後一問題之答案，示定義應用時必有之變更。立一定義，其事甚難，基本困難，一則由於需要將尋常用語所含（分明或不分明）概念，解釋爲可以枚舉之實體；二則由於擇定可以計數之事物，何者與實體最爲接近，而可得吾人所欲知之消息也。

問題焦點所在與所得資料　例如，吾人欲調查失業現象中之麕聚雜居及工作損失問題。麕聚雜居現象，以數字表示之，惟有查房間數或空間對人數之關係，而此問題，又往往因家屬之性別及年齡而異，且因房間之通風及光線而有不同。實際上，可以計數者僅有（與其需要並無密切關係）人數，至房間（房間定義只能武斷規定之）數，或其立體空間，倘能筆之於記錄。工作損失，現之於數字，須查現未作工之尋常工作日數。實際上，所計數之失業人數，乃指在工團或勞工交易所完成某種手續——如定在每日某時簽名於簽名簿——者而言。『失業人數』之定義，視

登記之規章而定，且在失業羣中加入登記者，亦僅該團體所轄範圍內之一部失業者而已。人口普查報告所用之『廛聚雜居』，意指一住宅內人數超過房間數兩倍以上之人數而言，至所謂房間也者，乃已將浴室，洗濯室等等除外也。

講述統計上總數及平均數之字眼，如人口、進口貨、進口船噸數、平均價格、生活費用、有業人數、工資、所得、資本等，均爲專門術語，較之言談或文字中所指，意義確定多多，卽用法亦與普通應用者有基本之區別。此等名詞，均可給以確定之定義，而其定義只於已列有總計數字之原始報告中始能確定之，然卽在此種報告中意義仍有極爲曖昧之事也。茲爲示明如何可以確定所得總計數字之意義起見，特將調查原始資料之檢查方法，於下節分述之。

報告之明確要件　陳述總計數及平均數之時，務須儘量求其清晰明確，切忌冗長之語句，表格上之簡要標題，內容過於複雜者，則應列以定義。例如考察英國煤礦生產時，吾人不可以『每人生產量』一詞爲了事，而應書如『大不列顛全部煤礦，一九二〇年，一月二十五日起，一週之出煤噸數，而以該週掘煤工人平均人數除之』；如嫌此言過於贅累，則或於前後文中，或於小標題，抑於底註內，列與上述要點，並將僱工平均人數之計算，加以說明。

欲用百分數，必須表明計算之根據，不然則絕不可用。例如某種商品，前些日之價格，爲八十鎊，今日忽高至一百鎊，則其增價爲『前日價格』之百分之二十五，爲『現在價格』之百分之二十。今如降低『現在價格』之百分之二十五，則必成爲七十五鎊；但如降低前日價格之百分之二十五，則必復原爲八十鎊矣。執此以言工資，如工資對『一定標準』按四次各提升百分之十，則該標準百分之一〇〇，一一〇，一二〇，一三〇，一四〇，卽爲各種之工資；但如以各該期起始日工資之百分數計算，則各期中之增價，不過分別爲一〇，九・一，八・三，七・七而已。

複雜之定義，保持明晰之有效方法，在設計製表程序中，尤有特別之重要性，茲特例釋如下。

屬性及特質　茲取英國所得稅委員會表格之一爲例，在此表中，應課稅之所得總數，達一十九億七千萬鎊，其意義，見於該會之編制則例中。茲特將此總數之定義列下：

A． 所得

B． 據所得稅委員會之所知

C． 合乎課稅法令之規定

D． 減除折舊準備及其他

E． 爲英國私人或公司所有，不住於管轄區內而有納稅義務之人所有亦同

F. 在一九一八至一九一九會計年度應課稅者。

以上六則：各示一特質，總數中每一單位，無不畢具之；由此等性質之確切定義，所謂總計數之命義，乃得表明，且『所計數者何』一問題之答案，亦在其中矣。

記錄之事實（B），時間（F），及空間（E），數項特性，任何一者均萬不能少者也。

第二節　資料之搜集

第一目　人口普查

人口普查　前章所論之原則，最佳之例證，莫過於人口普查，蓋一則其計劃、宗旨及細目，人所共知；二則用以搜集原始資料之調查表格，可以顯示詳密調查上之困難也。

人口普查之目的　第一請論舉行人口普查之正確目的：人口普查之目的，乃供人口學之研究，惟有人口普查，乃可得知人口之總數及地域上之分佈情況，惟有人口普查，乃可得知兩性及各級年齡之人數，以及關於獨身、已婚、寡居等等之人事情況(civil condition)，亦惟有人口普查，乃可得知人民之國籍，此乃行政上所需消息之最少限度也。除此之外，尚有政治家及經濟學家，諸多欲知關於每一人民之事實，而欲蒐集普遍之資料，就英國而論，惟人口普查實優為之。

問題之選擇　然欲知之事項旣夥，所必調查之問題，亦因之而繁，各項問題，何取何捨？此宜以便利爲主，不可斤斤於理論。欲問之事項雖多，大要不外下列數則：（一）家庭之組織及人數，家庭在社會上之地位，家長經濟上之能力；（二）家中各人之就業性質，各人及全家之工資或收入，房屋之規模及租金，各人所受教育之程度，開始或就業退休之年齡，遷徙之情形，加入宗教或他種團體否，家中各人有何殘疾。如此種種之訪問，如欲免除調查表格項目之煩瑣，並期於適當期間內可將表列整理就緒，若干事項殊有刪略之必要，至於究應如何選擇之法，則非將欲行調查之問題，逐一考查其大概性質不可也。

回答問題之能力　選擇標準，第一須以被訪問者能以答復之問題爲限。蓋此項問題，如用以調查受過教育而謹謹守法之人士，自可訪得家庭遷移情形，及何年齡開始工作等等之全部消息；不知人口普查之特點，乃即在其範圍之普遍化，故所訂問語，務求知識程度最低之貧苦民衆，能以了解而能答復爲要；然惜此多爲人所遺忘而忽略之耳。

確定　第二，問語性質務須能以完全確定，庶答者不致有答話是否對題之懷疑。統計家認爲最有價値之答案，莫過於『是』『否』二字或一單簡之數目，確定之時間與空間，或用有確實意義之字句。諸如『許多』，『時常』，及『一部分』等等形容詞

及副詞，所含之數目意義，各人感覺不同；且用此等字義，描寫一人物情狀，雖亦頗能表示清楚，但製表時乃全歸無用，而人口普查中徵求答案之用途，乃惟在用以製表也。例如，問及教育程度，則所問者非『曾受良好教育乎，抑受中常教育，或毫無教育乎』？吾人所希望者，惟求其『報明停止求學時之年齡，讀書幾年』而已。但此項含糊之問題，有時由於首次嘗試之缺乏經驗，或由於被調查者之疏忽，於是所得之文句，全失其實用之價值，反或引起重大之錯誤。調查工資及收入之時，問題若不具有一定之意義，結果必難期十分確實，例如工資調查，即需要精審之定義，並須有若干附帶問題，方可查得適當之估計數字；蓋問題之簡單如『君之每週或每年工資若干』者，答法亦有若干種，然答數之毫無價值，則無可疑義矣（註一）。

忠實 第三，規定問題尚有一要件，一則必須能得忠實之答復；二則必須能以避免偏頗之答復。此一準則，如欲嚴格恪遵，則對人口調查票之款式中，『何者勿得遺漏』之懷疑，幾已無須乎此。關於此點，問案之最難者莫過於問『君為僱主抑為被僱傭者』？固然，身兼僱主而同時又有被僱傭者之身分者，所在多有，故在第一次發現之後，以後如再舉行調查，應將此項問案捐除，蓋即使有此問案，然一般心理，多欲自擡身價，因而調查所得答復，未必足以信賴；反之，無中生有，錯誤必多，是以在求『忠實』之原

則下，此關於社會階級之問案，應行剔出也。

樂於答復　第四原則，設問須使人樂於答復；反之，有究查語氣之問話，或使人易慮及有變更法律及徵課新稅之可能者，均宜力圖諱避。關於工會，幫團等會員身分，或關於保險之問題，均足使人認爲有查究性質。多數之人，以爲進款若干，純爲個人私事，與外人無干，因而不肯說出。問及房租時，被問者則以爲此或乃加稅之先聲，因而引起彼等之疑懼。關於宗教之問題，彼等答案殊難引爲滿意，此觀一八九〇年戶口普查委員會所得結果(註二)可知，故在英國而舉行調查，依以上四條定則，應以不問宗教問題爲是，蓋若干人所奉宗教，作何名稱，一遇詰問，瞠然莫對之事，屢見不鮮，有者以爲問題太大，茫然莫知所措，此事既所在多有，而有人故弄玄虛，與事實恰相反對，或全然不肯回答者，尤比比皆是也。

次頁所附人口普查表，其問題一，二，三，四，十四及十五，依上述理由，皆不屏除，因問題均頗確定，一般戶主，大都能以答復，且均樂於答復，而答復乃可無誤。惟第五問題，遇有離婚，分居，或不合法之結合，則不免有欠確實之答復。至問題第六，七，八，九，在一九一一年，始行加入，雖亦多欠確實，但許多重要之新穎消息於此而得。至於第十及十一兩問題，對於職務分類（例如書記或木匠）及職業分類（例如紡織業所用之書記及木匠），區別

一九一一年英格蘭及威爾士人口普查

	姓名	對戶主之關係	年齡（截至最近生日爲止）及性別		結婚情況				
	不論自家人，來訪親友，寄膳者，或僕役，凡（一）於一九一一年四月二日（星期日）夜宿於本宅，而在半夜生存者，或（二）於四月三日（星期一）早晨到達此宅而未經別處查過者均須填報，除此以外之人則不得計算在內	註明是否爲「戶主」「妻」「兒」「女」或係其他親屬，抑爲「戚友」「寄膳者」或「僕役」	一歲以下之嬰兒年齡，以月計，如「一月以下」，「一月」……等		凡在十五歲或十五歲以上之人口，在名下宜註明「未婚」，「已婚」，「鰥夫」，「寡居」	在本表所填現有已婚婦女名下，請填明下列所舉之人數			
						現有已婚者，已結婚若干週年如結婚不滿一年者請註明「一年以下」	現有已婚者所生兒童人數（如並未生育或生產時並無生存之兒童請在第七欄內答一無字）		
			男性年齡	女性年齡			出生時生存之兒童總人數	現尚生存之兒童人數	業已死亡之兒童人數
	1	2	3	4	5	6	7	8	9
1									
2									
3									

（第一表）

一九一一年英格蘭及威爾士人口普查（續）

十歲以上者之職業				出生地	生於外國者之國籍	殘疾
個人職業	所在機關	役人者或役於人者，抑係自役者	是否在家工作	(1) 如在英國聯合王國落生者請註明出生郡名鎮名或教區名 (2) 如生於英帝國其他各部者，請註明屬地，殖民地，……以及邦名或省名 (3) 如生於外國，請書其國名 (4) 如在航程中出生者，請書明『生在海中』 附註：——凡非生於英格蘭及威爾士者，請報名『常住』或『暫居』	請照下列各項擇一填報： (1)『世爲英國臣民』， (2)『歸化之英國臣民』並註歸化年月 (3) 如屬外國國籍，應書屬『法』『德』『俄』或其他	本表所填之人如有下列殘疾者，須在各該名下填註明白，並記發生廢疾之年歲： (1)『全聾』或『聾啞』， (2)『全瞎』 (3)『瘋狂』 (4)『白癡』或『意志薄弱』
填答務須依所服役之商業，製造業，自由職業，……細別門類詳確填註。如在工商業服役者，工作種類及職務名稱亦須詳細填明	此條須填答工作之機關，但如第十條已填註明白者，此條可勿須再填。私人所用僕役，此條可無須填。惟在公共團體（如政府，市政廳……等）工作之僕役仍須說明所在團體機關名稱	凡在工商業工作者，在其名下，請註明 (1) 僱主（即除僕役外，尚僱用他人工作者）(2) 工作者（即爲僱主工作者）(3) 自役（即旣非爲他人工作亦不役使他人者）	從事工商業而在家中工作者，於名下，請填「在家」二字			
10	11	12	13	14	15	16

須由，或代，戶主，或住所之主持人填寫之項：

本住所（包括家宅，共同住屋，或分租房舍）之房間數請填註於下。廚房作爲一間，但洗濯間，梯頂臺，接待室，藏衣室，浴室，以及貨棧，寫字間，店房，等均不得作爲間數計算	填表人保證，填答時確盡所知，正確報告，並無虛飾 簽字蓋章：________ 通信處：________

（續第一表）

不無困難，惟自一九一一年以來，此兩問題，業已加以改訂，以便製定二項表格，然第十二問，對於有工作者之情形，並不能完全訪明，而十三問復多欠確實。兼任兩個同等重要職務之人，迄無釐定條文。此外，尙有第十六問題，尙不完全確定，而答案亦無大用。至若其他各項之優點，人口普查委員會之報告書，曾有詳盡議論（註三）可以參閱，此時所論不過在表明問題取捨之理由而已。

填表人員 關於塡寫表格之人員，乃爲一重要問題，而爲前此所未暇論及者。表格由戶主塡報者，英國人口普查卽行之，此須設問性質簡單，文字淺鮮，與用指派調查員查詢代塡者不同。但後者一法，表格可較複雜，設問不妨詳盡，且問題爲一部無知人民有答非所問之可能者，亦無妨礙，蓋調查員可以儘量研詰，必能得一答復，答復之顯有錯誤者，乃可隨時排除，且爲製表便利計，更可反復詢問，至答復對題而後已也。規模宏大，問題複雜之調查，如人口普查之類，必須集合衆多之調查員，加以短期之訓練，指導事項及通盤計劃，必須儘量單純；至調查範圍縮小時，訓練則不厭求詳，問題則不妨稍繁，此在第十，十一，十二及十三等項，尤關重要者也。

空白表格之款式 調查表格之一般形式，頗有加以注意之必要。目的如在欲知家庭之結構，答案以能在一張上塡答爲最佳，

故表格之地位，應充分寬裕，庶足以爲最大家庭之用，以免黏合多張表格之煩，而製表尤爲便利。又表格必須預備充分之空間，以便不善寫字者填答問題之用，但紙張仍不可過大，以能以展鋪寫字臺上爲宜。至於說明，務應清楚，不可模糊，並須貼近填答案之處；爲達此目的，不妨善用大寫字，斜體字，各種不同字體。第一表所舉人口普查表，所用者即其概略之縮影也。

調查目的之宣示　調查表格上，須將舉行調查之目的，及搜得數字之應用，加以說明，以期邀得被調查者之贊助，而消除一般之誤解。至於說明之範圍，程度不一，大部繫於調查之是否強迫，如人口普查，抑爲情願，如工資調查。上文所舉之人口普查，因人民已知其一般形態，而樂於完成其任務，故無須再加序文，但調查之已爲國會批准，人民有服從之義務一節，仍有宣布之必要。宣示之法，須印於表格摺疊之後面，庶表格不必展開，即可入覽，此外一方面固須將不肯填繳表格或虛僞謊報者之罪名，以及前次違犯規則者之處罰，印揭背面，以資恐嚇，一方復須聲明，準代填答者保守祕密，以免他人之利用或加害。但訪問如非屬強迫性質，必須附印信函，隨表格遞送，請求填答者熱心贊助，以底於成。

附帶訪問　表格之主要部分歸戶主填報，其餘部分，可由調查員填寫，如此則費事不多，而若干之附帶消息，可以搜集得來。

在表格之外角,註明人事登記區、分支區、調查區、及通信處等,以便依號次製列成表,各區域之材料,隨手可以翻出也。調查員攜帶空白表格,親往各家訪問,此乃爲英國人口普查所不能,然用此辦法,關於房屋式樣及街道情形,乃一覽而知,若在嚴重調查之時,更可僱用專門助手,每至一家,瀏覽一過,該戶之詳細種切,即能書成若干有趣味之記載矣。

縱橫行列 現依上述準則,進而批評表格之細節。第一點,表格之行列,安排上有時竟不能完善。蓋勞動者,既根本不能執筆,甚至雖一點一劃之微,亦須經人指教,調查表格行列縱橫,自難期其領悟,故若非有人從旁協助,即將有失措之虞也。下列問題:——

請告知君之姓名＿＿＿＿＿＿

請告知君之年齡＿＿＿＿＿＿

請告知君之性別＿＿＿＿＿＿

未婚,已婚或喪偶＿＿＿＿＿

等等。

答語緊隨問題之後,塡註自較容易,然如此則調查表格,不論成人或兒童,必須人一各張,如法國所行者即是。

問題之評判 表中第一問題,本爲人口普查之一般目的,應如何確切肯定爲宜,然若略一考察,乃不容不有所懷疑,蓋所謂

夜宿云者。設有一在晨四時，始行回家之打更人，或晨二時始歸之印刷工人，將何以置之耶？戶主又何能盡知住所內有何人改宿於別處耶？關於此點，若一一加以說明，將不勝其煩，而填答者尤感頭緒紛亂，莫明所以，故究竟如何答法，必須對調查員講授明白。

人口之意義 再就『某區之人口』一詞而論，其意義何如，亦不無可疑。蓋在法國，有所謂『事實人口』(la population de fait)，即某一瞬間某一地域所發現之人口，與『法律人口』(la population de droit)，即某一地域之常住人口，加上臨時他往減去臨時來住之人口；事實人口與法律人口，二者迥不相同。例如市區人口，乃即法律人口減去監獄囚犯，醫院病人，住校學生，寺院僧尼，以及軍人等（註四），而英國人口普查，所查則僅為『事實人口』也。又在美國，一八九〇年，有所謂『憲法人口』(constitutional population)，乃將印地安人保留地(Indian Reservation)特區(territory)，及哥倫比亞區之居民除外；有所謂『普通人口』，即將特區包括在內（印地安人保留地，印地安特別區，及阿拉斯加仍除外）；更有所謂『總人口』(Total population)，即除印地安人保留地外，其餘各地一概包括在內（註五）。在一九一〇年時，普通所指之美國人口，即僅就美國大陸而言，即四十八州〔向為特區之亞里綜那(Arizona)及新墨西哥(New Mexico)包括在

內〕及哥倫比亞區也。若再加上阿拉斯加、夏威夷、及波特里哥(Porto Rico),海外之陸海軍人,可謂之總人口,但此外尙有一總人口數,卽尙須包括菲律濱、幾亞穆(Guam)、沙謀亞(Samoa)、及巴拿馬運河區域等地之人口估計數字在內。因哥倫比亞區及印地安保留地之人口,不在攤派稅捐範圍之內,故尙須從大陸人口中減去。須知鏈奈爾島(Channel Island)曼島羣島(Isle of Man)之人口,本在英國人口普查挨戶淸查之列,但普通所提之英國總人口,此兩處之人口不與也。又海上或海外之陸海軍人人數,亦係單列一表,亦不在總人口之內也。

年齡 表中第三第四兩行塡錯之原因有二:一則因年歲較大之人,有時並不知其確實之年歲爲幾何,乃隨手塡爲相近之整略數,於是塡完送回之表格,年齡多爲四十,五十,或六十。欲免此中差誤,製表時應按 35-45,45-55……等歲分組,如欲製更詳之表,最好依 3-7,8-12,13-17……等歲分組;二則多數婦女慣於低報年齡,在此情形之下,補救之法,可對照人事登記簿册,以求較近之總計數也。

尙有須加注意者,報告年齡,旣以『上次生辰』爲準,則年齡平均必較實在年齡,低六個月,而在事實上,如有報十七歲者,則其實在年齡,確亦較十七歲多數月不等其分配均勻一致,最多至十八歲爲止也。

職業　表中第十，十一，十二，及十三等四欄，有極重要而應加批評之點，但人口普查委員會關於產業調查上所有一切問題，一時不便討論；惟可得而言者，一職業調查，欲求其完善，最好單獨舉行並與人口普查稍異其原則。此蓋無可疑，但在改弦更張之意見未一致前，則與其因不能得到確實答案而放棄之，反不若於現有人口普查表格中添加適當關於職業問題之爲愈。總而言之，職業調查理應與人口普查，合併舉行，非然者，解釋之巨大困難，將層出不窮矣。困難之點何在，此則觀於『生產調查報告』中之服役人數統計，與人口普查有業人數統計，對照鈎稽時可知之。

吾人批評特殊問題時，必須將目的物，常存於心目中，玆所

紡織業

	綿織業	毛織業	蔴織業	總計
僱主				
經理				
書記				
監工員				
紡工				
織工				
勞動者				
童				
總計				

（第二表）

討論者，目的物有二：一查各工業各部門服務人數，即縱的區分也；二查各級職務（如勞動者、工頭、僱主……等，或鐵匠、木匠、織工……等）上分類之人數，即橫的區分也，於是列成表格，結果如上。

訪問之最低限度，必須能得下列回答：

法律業——律師業——司事員

礦　業——煤炭業——挖煤工人

金屬業——鐵　業——打鐵匠

執此以論，僅一『請告知君之職業』問題，必不能得到此類消息。挖煤工人，僅云礦工而已；打鐵匠，亦不過提一鐵匠而已，而爲書記者，則將云爲司事員而已，挖煤工人，僅言礦工而已矣。故欲將所需要者，加以解釋，錯誤之事，欲求避免，必須在空白表格背面印成填表須知，以便填答者有所遵循。而填表須知必須標列特別字體，以期清楚明瞭，尤須簡短切題。如回答者有正確填註之心，並有相當教育程度，足以瞭解簡單之說明，填寫錯誤之事，每不易有，惟恐填答者多未之見，或閱而不行耳，此乃與研究空白調查表格，大有關係者也。表格發交被調查人，令其自填，人力時間，節省甚多。故填報之法，可有兩種途徑，一將表格盡力化至最簡，而結果可期最良，否則，不必令被調查人填寫，僅由其報告所知而由調查員代填，蓋調查員深諳填表須知，調查之法律強迫

效力，亦所熟知也。然後者一途，時間費用兩不經濟耳。

現在之調查制度，加以製表之方法錯誤（或亦爲事實所不可免），均足爲英國迄未有可靠職業調查之原因。現存之數字，一則由於資料之不確實，二則由於製表之不良，已大失其價值，而吾人若干重要計算，尙以之爲根據，夫人皆知其不可也。

新問題之結果　在一八九一年，曾有一番努力，企圖分淸有無技能工人，僱主與被僱者之相對數字，卽現在第十二欄之問題也。第十二欄之標題，本不足爲淸晰之規範，文字既無『請告知』，『請填寫』等平常字眼，而填答方法，究係用『是』『非』，抑在相當欄內填註記號，在表格前面，亦無說明，至三個項目，界限亦復有欠明確；然則在一八九一年之企圖，由下列報告書中（註六）看來，始終未能作到也：

『填回之表格，內有極多，根本並無記號；更有甚多，二欄乃至三欄，同時均有十字號，卽其中只有一欄記十字號者，亦有充分之理由，足以判定十字號已錯置欄數也。誤置他欄之事，除填答者有意點錯外，似幾難有他種原因，蓋性質愚魯，頭腦庸俗之人，往往以擡高職務地位爲榮，且據以教唆他人。因此之故，填表之時，一人之見如此，有聯帶關係之多數人，亦由僱員地位，而填入僱主階級矣。蓋收回之表格，內中若干項目下，填僱主者較填寫僱員者爲多，填店主者，比填店員者爲多……也。此種令人不

解之情形，若非恃此僅有之假設，必難得有解說。故此項答案，吾人認爲大不可靠，無論何張表格，均不便有所利用也』。

雖然如此，此種問題仍行列入如故，製表亦如故，而統計家居然予以相當信仰，竟行採用之矣。

此種企圖，及其結果，對於設計調查表格者，關係何等重要，不言可喻，此吾人所以一再提及，不憚詞費者也。

在未結束本題之先，尙有一事，不容忽略者，即在英國欲由英國人口普查中，將依賴某一產業生活之人數，換言之，某業之僱主人數，及其家中人數（如尙未調查得來）與所用僕役人數，以及僱員及其依賴者之總人數，直接歸納起來，斯時尙不可能。但此乃在考查一國中各種產業之比較重要性時所必需，在其他各國雖已不無適當辦法，而在英國迄無可用之表格，故只得靠統計家之估計，始能求得總計數字也（註七）。

欲知人口普查表格答覆所給之消息，如何可以作成詳細分類數字，只須查閱布斯氏編『人民之生活及勞動』(Life and labor of the people) 一書中，第五編第四十六頁所列者即是。

就一般論之，統計家之工作，必須利用既得之資料，故先決問題，必須了解蒐集資料時所用之定義，以及已發表表格所受之限制。蓋極荒謬之編製，一經妙手整理，亦常能得出有價值而正確之結果也。

第二目　工資調查

工資調查（如於一八八六及一九〇六兩年所舉行者是），與一般人口普查，不同之點甚多，舉其重要者，約有二端：（一）工資調查表格之塡答，屬於自由性質，而人口普查，則有強迫效力；（二）工資調查，塡答須具有較高之知識程度。

工資調查之主旨　與前相同，欲舉行工資調查，必須先研究其主旨何在。其主旨云何？一則表明英國人民之收入；二則比較各業之工資率；三則調查各級工資率上之比較人數。

時間單位　爲達此目的，計算之數量，以何者爲最佳？就先決問題而論，時間單位將用一日乎？一週抑一年乎？如所得之資料，爲每日之工資，則一週之間，各種職業，日數自四天至七天不等，必難據以核算每週之工資。至每週工資可謂爲較確定之日數矣，然因各業多受季節影響，一季工作異常忙碌，而他季則不然，更有若干之產業，各季節工作相差甚大者，故在某一星期，恰爲一業營業最盛之季，而爲另一業生意最淡之時，二者用相比較，其間相去，豈可以道里計耶？爲今之計，一則爲避免此種差誤，二則因一年工作若干全星期(full week)，除在少數向不停工之產業外，多甚難以計算，最佳莫過於用一年爲時間單位，惟惜直接推算個人之每年收入，實際上又不可能耳。蓋在大工業中，人員更換，時日靡常，一人所領工資，一年恆三易其僱主，各人工資總

數若干，雖僱主亦不得而知；即就常年不更動人員之廠家而論，每週發放工資總額，一般並無表格之製列，似難期收綜計便利之效；若重新加算，則卷帙浩繁，計算人工又將大至驚人矣。至若工人方面，一年中收入數字，大多數並無詳確之記載，所能為力者，只有藉重個人之精審研究，加以估計，然謂其欠有確實性，則無可疑；在最多情形下，欲使工人自述其過去十二個月間有關收入數字，如其不願，固無待言，即使情願，恐亦有所不能也。

由此觀之，舉行工資調查，必須採用較短時間之單位也，昭然明甚，惟大多數之工資既按星期發放，則最自然之單位，莫過於用一星期矣。用星期計算，其每週工資，用以估計常年收入所引起之附帶問題，可暫置勿論，惟對於第一問題，即調查之最優數量，乃只得一間接之答復；唯一之直接的個個數量，乃為每週工資，而此只可用大量估計方法，加以補充也。

調查僱主與僱員 先決問題，討論既如上述，現又有一第二問題：需向之訪問者為何人？供給資料者，乃為僱主與僱員，而在一理想之調查中，二方面之消息同須求得答復，固矣；然為求手續之單簡，費用之節省，資料之確實，唯有訪問僱主時，方能收效，訪問僱員時，則有所不能也。

訪問僱員時，程序有如下列：（一）籌擬一類似人口普查表格之調查票，一面將調查之目的，作一簡短之敍述，一面擬列若

干簡要清晰，單簡適當之問題，每一問題之後，各留適宜之空白，以足以填答最少限度之報告爲合格；(二)表格發交調查員取用，填畢交回集齊之後，預備相當時間，使調查員有審查改正機會。此法所需之組織廣大，費用浩繁，較之人口普查，有過之無不及，事乃至屬明顯，然其舉行結果，所可求得之確實消息之最高量，能否足供應用之最小量，尙在可疑之列。雖然，調查之一部分工作，藉用工會之力舉辦之，亦無不可也。

訪問僱主之方法列下：就所欲調查之範圍，抽選若干產業，分別查得各僱主之住址，將適當之空白表格，備具解釋函件，逕行寄交所有僱主，請其填畢寄回。用此方法，較之前論調查工人之計劃，簡易多多，省費多多，且所用表格，大爲減少，整理人員，數亦至少，蓋在商人方面，表格填畢寄還，較留存待人走取，反覺簡便。惟調查之時，旣無當面之接洽，表中設問，尤須特別清楚，因事後改正錯誤，往返通信，雙方均感煩擾也。一八八六年，所用之表格，玆節錄一部分分列於下。

工　資　調　查

絲廠工資率報告
廠名或店名________
地　　址________

注意——經理與職員之薪金，可請勿庸填列，本表所調查者，僅爲工人工資。
一八八六年——所僱之人數……………人數________
一八八五年全年所付之工資……………金鎊________
一八八五年所付最高之每週工資　金鎊________日期________
該一星期領工資之人數……………人數________
一八八五年全年所付最低之每週工資　金鎊________日期________
該一星期領工資之人數……………人數________
請示現時加工時間之平均工資率：換言之，加工時間，按一時一刻計算，抑按一小時半計算……等，或用其他方法，請書明其方法________
請示知，現時有否加工之工作，如有，有若干；或現時有無少於規定時間之工作，如有，少多少________

一八八六年一月一日各家絲廠所僱工人，

現行每週工資率及工作時間

職務分類 注意——下列職務分類表，如不適用，必要時，可請斟酌改動。	工作全星期時，現行每週工資率及工作時間數，但加工時間不在此限 注意——請告知每星期工作時間數，工資按時計算，抑按件計算，如以件計，請說明每星期所得工資之總額，加工不計。											
	男性						女性					
	男子			幼童			婦女 十八歲以上者			幼童 十八歲以下者		
	工人人數	工資率	工作時間	工人人數	工資率	工作時間	工人人數	工資率	工作時間	工人人數	工資率	工作時間
絲經部												
分絲工 時工												
件工												
捲絲工 時工												
件工												
洗絲工 時工												
件工												
紡絲工 時工												
件工												
雙線工 時工												
件工												
其他												
紡絲部												
選絲工 時工												
件工												
鍋爐工 時工												
件工												
修整工 時工												
件工												
梳刷工 時工												
件工												
其他												
織綢部												
捲絲工 時工												
件工												
排經工 時工												
件工												
揀經工 時工												
件工												
摺疊工 時工												
件工												
裝填工 時工												
件工												
其他												

（第三表）

全年工資　計量工人階級之全年工資，爲調查最終目的之一。全年工資，包括之項目頗多，重要事項，不外如下所列：尋常每週工資，加工工資，特別工作(例如派赴遠地工作之建築工人)，或特別季節（例如農場收穫期）所給之特別償付；與非現金之付款，例如房屋之免租或減租，免費或廉價之煤炭，以及以批發價格或廉價優待售與之特種貨物（如紡織工廠之布疋，農場工人所買之番薯）等是也。

如用實物償付工資之事實，果爲普遍之現象，或佔重要之部份，則進行調查，必須以勞工委員會附屬之農業委員爲前例，另用一徹底不同之方法。但如所關僅爲一單純之科目，例如房租免付，則無須改弦更張，只於前列之調查表格上填一項問題已足。此種情形，繅絲工業，尙屬少見。總之，以上數點，指示吾人，在未擬定妥善調查表格時，主持調查人員，對於實際情形，必須有充分瞭解，方能擬定恰當之表式也。

上文曾提及每週工資，加工工資及特別償付等問題，茲以加工工資與特別償付二者與每週工資合併討論之。平常每週工資，在大多數產業中，大多數部門內，乃爲普遍而意義確定之數量。任何工頭一人，手下管轄工人若干，每人在平常全週工資幾何，大概均能報告。在許多方面，工團（如建築業是）均曾對於每小時或每星期之工資，有所規定。在另外若干方面，如在紡紗業，件

工工資數額，亦曾加以規定，庶因工作之難易，使每週工作之工資，可以得到確定數額；就一般情形而論，苟取工資薄，略加涉獵，卽可知每部門之工人，所得工資，均有傾向平均數之趨勢。由此進行，平常整個星期之平均每週工資，必可從而求得，並可具有頗大之確度。然行程並非止此而已，此乃爲計算全年收入之一部工作，此外尙非求得一年共工作若干全週不可。關於此層，前列表式中所用之方法；尙有硏究之餘地。所欲探問之問題，已見於前列工資調查表格上，由此諸項問題，可以查考問題焦點所在與所得資料之差別。

問題焦點所在及資料　問題焦點之所在：全年工資減去停頓之星期工資並加上加工之期間工資，等於若干全週工資？於此，第一待決之問題爲：吾人將一律減去因病假而損失之工作時間兩個星期耶？抑按實在不能作工之時間據實計算耶？患病原爲個人之不幸，並非一般之損失，在可能範圍之內，自以屏除不計爲宜。至某一季節之加工，雖在一年間未必有多過於常態星期之工資，加入平均計算，但與其他季節之停頓期間，終有立趨平衡之趨勢，而以在採取一小時作一點十五分，或一小時半計算之基礎時爲尤甚。故估計一年之收入，以如是之多之常態星期工資爲準，不唯得手續之便利，抑且有論理之根據。舉例言之，如一方病假日期爲二星期，工廠停閉之日期爲三星期，另一方面，某一忙迫

之月份中，加工之工資，恰等於常態星期兩星期之所得，結果一年即有四十九個全週之工資。由此得來之數字，即爲一年工廠以工資形式而付出之總額，再用恃工廠爲生者（假設全行加入工作）之常態星期工資總額除之。譬如，在常態星期，一千二百工人(男工、女工及童工)，如全行加入工作，可賺工資一千鎊，此數即爲依賴某一紗廠生存之平均人數，又如全年以工資形式付出之數額爲四萬八千鎊，則全年工資即等於四十八個常態星期之工資，而工資平均，必爲四十鎊矣。然工廠所付工資之總數，多有帳簿可稽，依賴工廠作工爲生之人數，常不能確實查出；蓋一大規模之產業中，人員不斷變動，設有工人離此他去，彼究已入他廠工作，抑從此即告失業，雖身爲經理者，亦難得知。故如以一年中曾在廠工作之總人數爲基準，既失之誇大，但如以常態星期之工作人數，又未免失之過小。唯一可用者，似只有一年中工作最忙之星期，所有作工之人數而已；蓋工廠工作忙碌之時，除因病缺席者外，其他未能到廠工作之人，絕不能謂之賴工廠而生存，反之，如其同時並未在他廠工作，則彼加入永久失業羣中，乃可斷言矣；夫工廠生意繁忙之時，請假遊玩者，實乃絕無此人，且同業工廠，生意閒忙，大都一致，此廠忙時，彼廠未必閒，而彼廠閒時，此廠亦未必忙也。即就之問題——全年共計工資若干；工作最忙之星期，共用工人若干——既得答覆，假設答覆適當，則吾人所

欲得知之消息，即可全然明瞭；因全年付出工資總數，用工作最忙之星期所僱工人人數之最高額除之，即可得出平均全年工資也。如欲求常態星期之等量週數(equivalent mumber of normal weeks)，即用賺平均工資之工人人數最高量(可由調查表格第二面得之)乘之，則所得乘積乃表示每週工資之總量(假設全體工人全行被僱加入工作)，然後用此乘積，除全年所付工資總數即可矣。

為例釋上項手續起見，不妨用最近所舉行工資調查之資料，與一九〇六年所得者比較，蓋最近所得消息，較一九〇六年之調查，詳備甚多也。

全年工資與每週工資　一九〇六年工資調查，所得之種切，如下所列。惟須聲明者，此處所舉，乃僅就英國之紗廠為例，且下列數字，又僅指有答復寄回之廠家而言。

T＝一九〇六年之工資總計＝£10,195,229。

W＝工資總額之每週報告十二個星期（註八）之平均數

＝£204,173。

N＝工人總人數之每週報告，十二個星期之平均數＝212,503。

M＝十二個星期中工人總人數最高額＝2[illegible]3,472。

A_e＝某一星期所僱工人之平均工資＝19.43先令。

A_f＝某一星期工人未曾加工亦未停工者之平均工資

$=19$先令7便士。

由此可得結果如下：——

$A=\frac{W}{N}=$十二個標準星期每星期平均工資$=19.21$先令

$E_a=\frac{T}{N}=$平均人數之平均全年工資$=£47.98$

$n_1=\frac{E_a}{A}=$一年中實得平均工資時之星期數$=49.95$星期。

一年共有五十二個星期，茲實得四九・九五個星期，相差有二・○五個星期；以意度之，此數乃卽放假之日數，約合工作日數八日至十五日，惟此中未必全爲假期，此外工廠因故停工之日數，亦當包括在內也。

$E_m=\frac{T}{M}=£47.79=$總人數最高額之平均全年工資，所謂最高額卽恃工廠爲生之人數，至相差數額，乃由於失業關係也。

$n_2=\frac{E_m}{A}=49.73=$總人數最高額賺平均工資之星期數。

$n_1-n_2=0.21=$一九○六年因失業而損失之星期估計數。此外，尙須加上在工作最忙之星期中，仍係未找得工作之失業者，所損失之星期估計數。

$\frac{A_t}{A_e}=1.008$，在此情形之下，就全般情形而論，加工之時間乃較停工之時間爲多，故工資較之純爲全時工作 (full-time work) 所得者尙多得千分之八。

以此千分之八，加於 n_2 上，則所僱工人人數最高量，在一年

中，能賺全時工資之星期數，爲五〇·一。

在一八八六年，查得記錄上一八八五年工資總數爲 T＝£3，146，566。

一八八六年常態星期之平常工資，爲A_f＝一五·二先令。一八八五年記錄上最高之總人數，爲M＝87.887。故最高總人數之平均全年工資，爲 $E_m=\frac{T}{M}=£35.8$；又如A_f等於A（一年中平均每週工資），則 $n_2=E_m\div A_f-47.1$，此卽最高總人數賺平均工資之星期數。至A_f與A_e二者，以缺乏資料之故，此時殊不能加以比較也。

總而言之，無論恃如何觀點，此種方法均不無堪以批評之餘地，事實昭然若揭，無待繁舉，上文所論，與其謂爲常態工資對全年工資相互關係之通盤記述，毋寧視同一種例釋，目的僅在舉示中心問題及資料之性質也。

關於工作時間之損失，除因放假日及在工作最忙時仍係失業者外，尙有因病請假對於工作時間之犧牲一項，據人估計結果（註九）此項損失，平均每年有一·七個星期云。

法國方法　法國之工資調查，據一八九八年（註一〇）所發表之調查結果報告觀之，其所用工作日數估計方法，與英國逈不相同。查法國蒐集之資料，有下列數項：

（一）各門工業每月人員之更動，查得平均在一年中佔有

百分之四。換言之，每一廠家，在所僱一百人中，繼續工作經過十二個月之久者，只有九十六人。

（二）在一年行程中，每一廠家，各月所僱人數之最高額最低額之相差，查得佔平均人數中之百分之一九。由此可見，按平均計算，陷於失業者，至少當得此數之半。

（三）在一年中，每一廠家隨時僱用之工人人數，據查在長期僱用人數一百人中，爲一百四十，由此得一當然之結論：失業者之平均人數，至多不過一百四十分之四十，換言之，即百分之二十八。然則所謂百分之九，所謂百分之二十八，乃即平均工作之缺乏(Lack of work)之上下限 (limit)。是以如用此方法，所得之資料，較之用英國式方法，多爲詳盡，或且較爲可靠，亦未可知。又由法國工會團體得來之資料，得知平均工作之缺乏，數爲百分之二十，至英國應用前述方法所得之數字，據一八八六年之全部工資調查所示，則爲百分之十二也。

僅一工資調查表格之第一頁，設問問題雖不甚多，而必須之考慮，上文已作冗長之討論，足見在空白調查表格，尙未籌擬妥貼之前，先決事項之智識，誠乃必不可少，於此已有良好之證明矣。關於表格上其他各點，現爲篇幅所限，不克多所評論，惟不容已於言者，即一八八六年關於工人個人工資之問題，頗有欠於詳盡之感也。舉在各工廠『紡絲工，件工』（見前第三表工資調查

表格）之例言之，所塡答之工資，乃爲僱用工人之平均工資，因而各個工人之工資，並無記錄可尋。而工資之一般分配情形，亦僅一概略之報告而已也。至一九〇六年之時，塡表須知，曾有『賺同額工資者，可以歸併塡答；如非屬同額工資者，每一行只得塡列一人』之規定，於是每一種工業每一種職業（occupation）之實際變動情形，乃得完全表現矣。

兩年份所用之表格，頗可作比較之硏究，蓋非如此，不足以體驗擬定設問問題時之困難也。

第三目　私人舉辦之調查

非公家主辦之調查，與官方舉行者，苟後者並無強迫塡答性質（如工資調查是），二者在本質上並無可以區別之點；惟私人所作調查，以缺乏組織或基金之故，調查之範圍，大多受若干之限制，然惟其如此，於是乃有施行抽樣方法（method of samples）（見本書第二編第二章）之自由，故爾在適當地域之內，儘可充分利用之也。

勞動階級概況 爲舉例起見，玆就某某數城鎭之工人階級經濟概況調查討論之，此項調查結果，在『生計與貧乏』（Livelihood and poverty）（註一一）中，所發表者卽是。惟現所討論者，並非前文所予確切定義之問題；此時之目的，簡括言之，在討論關於工人人數，工人家庭中依賴生存者之人數，及其工作收入與生活所

賃銀勞動者				工作時數		工資	
對戶主之關係	年齡	職業	僱主	上一星期	全時工作	上一星期	全時工作

非賃銀勞動者					
	性別				
	年齡				
	對戶主之關係				

住房情形						
	房租					
	房屋種類					
	人數					
	房間數					
	花園					

歸檔號碼

調查號碼

請閱後面

（第四表）

需，種種資料，如何可以求得而在實際上並不致發生困難之問題；並在將各部資料付諸審查之後，如果恃爲可信，卽用以製表之問題也。

在舉行此項調查之時，因被調查者，若無任何強烈之引誘力（例如被調查者，可得糖果），必不願塡寫表格，答覆問題，故一般只得用派員訪詢之法。是以所用表格，只須有簡略之說明，以調查員已經過選擇與訓練故也。至於應用薄紙表格，不如用卡片之便利，卡片之式樣，如上面附列者卽是。

空白卡片　次須考慮者，戶主或戶主之妻，實在所知之事實爲何，且當調查員反覆詰問時，答者將用何詞以對之。苟能得與主婦接談，則關於家中人等之消息，未滿二十歲者之眞實年齡，在外工作者之職業等等項，均不難誘出答覆；房屋之租金及形式，自亦易載之記錄（必要時且須加以核對）。由此項消息，直接卽可製成有價值之表格，表示各家庭之結構及收入能力，並可由此得出絕大之變動，不受標準型式之拘束，而爲人口普查及其他官方調查所不及也。

無如在徵課家庭所得稅時，竟發現諸多之困難。各家主婦，對於丈夫或年長兒童之工資，常多茫然不知，故在極多情形之下，不能卽時訪得。如能得有答覆，必要時可向僱主方面復查，以證明之，庶免有多報少報之弊。如所報職業屬實（一般亦確係如此），

卽可依各該城鎮之已知標準，估計常態星期之工資，則結果必能十分準確。卡片上所謂『上一星期工資』及『全時工作時間之工資』，二者必須分清，因前者乃可詢得確定之答覆，而後者時常僅爲一種估計而已。調查人員，既有兩項之報告，務設法尋求二項相差之理由，然後據以核正後者一項（因製表所用唯此一項），使得確定之數字。用以製表者，僅爲有全週工作下之情形，關於病假及失業之情形，概非所問。在『其他收入』一問題下，答案鮮有完全者；惟若干家庭工資收入顯有不足者，生活既已維持至今，必能指明其維持生計之進款，且因填答者所以致誤之處，僅屬漏填一種情形，故答案已示正面消息。勞動階級之家庭，大多數均有微末收入之生利財產，此項收入之主要來源，據一般之報告，卽住房餘屋或他處房屋出租也。

與最低生活程度之關係　工作收入之估計數字，固不能用以準確列成表示因全年收入不同而分級之各級人數表，然就其主要用途而論，亦尙足以完成其原來之使命。主要之用途爲何？曰在查明工人家庭中收入（慈善機關之賑給除外），足以維持某一生活水準之比例，究有若干，所謂某一水準，如郞特利(Rowntree)氏在『生計與貧乏』中所計算之最低生活程度者卽是。關於各家依賴生活者之性別與年齡及人數，與在外作工者之工作性質，全家人口所處之水平地位如何，在大多數情形之下，並無

可以置疑之點。遇有可疑之情形（列表時不用之），可藉重調查員之附註（在調查卡片之背面，留有空餘地位，即備調查員塡註解說之用），而爲合理之判斷。

調查片並不交與被調查之人，須由調查員於詢問之後隨時塡註。如恐各家調查片有混淆雜亂之弊，應在各片上排訂調查號碼。在分類之初步手續完成後，並應訂定歸檔號碼。然後將各卡片，付諸審查，其列表應用之數字，復須作精密之核算。登記所用之表格究應預備幾許行數，此須將卡片分爲數札，然後分別枚計之；此步手續，甚爲簡捷，然亦需繼續不斷而謹愼精密之核對。

調查範圍之指定 每作一種調查（如某一城鎭之勞動階級），其範圍如何，務須有精審之限定。由經濟觀點上之分區，是否與行政目的上之區域相合？如其不然，遠在郊外之人家，究竟包括在內與否？均應加以確切之指定。範圍既定，次須覓得全區內之確實戶數名單，並按此名單用抽樣方法，選舉若干戶數；舉行調查之時，即以名單所載者爲準，而名單所列亦即爲『範圍』之定義。至於所謂『勞動階級』，尙無公認之定義，而實際應用者，必須有待於辦理卡片時之決定。故在選擇戶數之先，凡各家租金在某一定數額之上，或住宅載在主要居留人名簿上者，均須屛除於調查範圍之外。即在被選而前往調查之時，如發覺戶主爲職員，教員或經理人者，即應放棄之。至於其他如商店助理員、佣金經

紀人、旅店店主、小商店主，在發現之時，即須有所決定，並須特別註明。由此論之，所謂勞動階級家庭之最終定義，乃由劃定界限方法得知，苟如予以整個文字上之定義，則有如下列所示：凡各戶房租每週不超過十二先令，而戶主非爲職員、教員……等等者，均爲勞動階級之家庭。如此規定限界之方法，在製成表列時，必須極爲一致；在寫作報告時，事先已有規定之事項，必須明白敍及，在處理邊際情形下之取捨問題，尤須特別重視也。

第四目　英國國外貿易統計

由原始表格，可做成其他極多之統計，研究此等原始表格，極有興趣，惟以限於篇幅，不能逐一討論，現在只得就較爲特異之調查研究之，此即用作國外貿易統計之調查也。

人口普查之時，答覆表格具有強迫性質，並由戶主自行填答；在舉行工資調查之時，答覆一隨各人情願，僱主即允填答，亦只填答一次即足；勞工局 (Department of Labour) 主持之調查，被調查者雖有答覆與否之自由，惟調查次數多屬按期性質，因有變爲『準官式』(quasi-official) 之傾向。至出入口貿易統計資料之蒐集，則將此三種調查之性質冶合於一爐。

供給消息者　國外貿易統計，所欲調查之事項，可以供給消息者，可以分爲三種：一爲貨物出售商；二爲經手放行貨物之海關官吏；三爲貨物收受者或其代理商。以英國之出口貿易而論，

因環境之許可，出口商或出口代理人，在發出貨物之時，例須將貨物之數量、價值、與售往之地點，報告於海關之統計科；在進口貿易方面，領貨代理商，須將行將登陸之貨物擬具報告文件，呈交海關官吏，然後海關官吏，視其貨載之性質，而分別決定辦理，如屬免稅類者，大略檢查通過，必行課稅者，即須詳細檢查；至遇有轉口情形，則在登陸之口岸，一如進口之貨物手續辦理，並在必要時，上船之口岸，亦須加以檢查也。

此項空白表格，既由海關官吏視為主管責任之一部而查核之，同時又為代理商人專為此事而塡者，故無須附具信函，復隨事實之需要，可使內容儘量複雜；表上並無任何問話，僅為空白表格而已。取得此項表格之後，即加以審查，以便得出商業部之出入口貨物總計數，並擇其可用之資料，列成表格也(註一二)。

問題主旨所在與資料 此種調查中，吾人所欲測量之數量，為凡有交換價值，自本國輸出，或由國外輸入，依商品品目及輸往地與來源地而分類之貨物容量或重量；在裝船或卸貨時之當時價值。惟當實際調查之時，吾人所能測量者；與此迥不相同，而為商業部所收得之價值與容量記錄數字也。故吾人須將表格，加以審查，以便決定（一）出入口貨何部曾有登記；（二）所塡價值是否正確；（三）所載數量是否確實；（四）商品名目是否確

是帆船
抑汽輪　汽　輪　官定號數
第一號　登記號數
某港口　第九八〇號報告（註一三）　登記日期

船名	噸數	英國船或外國船。如爲英國船，註明註册港口：如爲外國船，註明屬於何國。	海員人數		船主姓名是英國人抑係外國人	自何處來
			英籍海員	外籍海員		
馬利安尼	七〇〇	英國	12	——	亨德 (H. Hind)	法國阿維爾 (Havre)

載貨情形

一 裝貨之地點名稱以時間先後列下	二 符號	三 數目	四 貨物之包裝情形，散包貨物之情形，與意欲在本港口輸入煙葉，雪茄烟，或鼻煙每包內容大概單位。	五 輸往英國其他港口之貨物及包裝情形	六 將轉往他口，或存艙預備出口之貨物（如無此項則免填）	七 受託寄售貨物者之姓名
法國阿維爾	自	巴黎	至倫敦——包裹六	〇〇件，鮮貨	水果	斯密士
				包裹六八	件，貨物	
	COK	1392				
	AE	495/6				
	KG	340/9				
	FOT	1/50				
	AJ	3/6				
	CK	1	七〇箱酒			,,
	KC	40				
	ACD	20				
	WD	166	五箱毛織品轉運利物浦港			,,
	O&D	1	一箱白蘭地酒			,,

庫藏貨物

艙上餘存之物品，有｛三磅雪茄烟　四磅烟葉｝
外籍旅客　無
航師姓名　________
泊船碼頭　南奎 (South Quay)
代理人姓名及住址………C. J. C.

余保證以上所塡，確爲本輪及載貨之實在情形，余已盡余之所知，報告詳情如上，並保證並無打破包裹情事，且自離開最後外國裝載口岸阿維爾後，並無中途卸貨情事。

船主亨德具結一八九六年十月十三日
海關外收發________會簽

（第五表）

實；（五）表格上所填輸往地點與來源地點是否確實分清。

填表舉例 貨船靠岸之時，船主必須送呈一紙報告，茲爲舉例計，特將縮小之樣品，摘錄於上。

課稅貨物 急於轉運之貨物，一經按照類似上列之特別款式，填報交與海關，可以立即開行。其餘貨物，一概分按徵稅及免稅辦理。在上列第五表中，設有十箱之酒，係備家中自用，只須繕具報告送去即可；下餘六十箱，乃送儲倉庫者，必須另作報告，說明品質，數量及價值；但七十箱全部則均視作進口貨。又設儲存於倉庫之酒，有二十箱，將轉運往他口，然後復行出口；是須繕呈報告，乃歸入復出口之外國貨一類矣。在駛進之海口靠岸之輪船上，艙中尚自存有酒二十箱，其中十箱，尚須留存至轉往他口時應用，則亦須作一報告呈上；餘下之十箱，提出移存別家倉庫，仍屬保稅倉庫性質，並無須作報告，但若提貨時則應依其情形如何，照上述四種手續辦理。此雖指酒類而言，然其他課稅之物品，手續亦不外此，換言之，仍當遵循同一程序也。

貨物之檢查 貨物報告情形不明，或與報告文字不符者，須開包檢視，內中何物逐一開列檢驗單 (bill of sight) 中，並作一報告送進。私人所有物，亦須分別加以檢查，記錄於特許證 (sufferance form) 上；此種私人用品，如果真實屬於個人所有品，可無須作何登記，但若有應行徵課關稅者，則一視爲普通進口貨

辦理矣。至如應課稅之物品，不論藏在所有品或商品中，一經查出，即行沒收，並不視爲進口貨也。

免稅貨物報告(註一四)

港口＿＿＿＿＿＿＿＿＿＿

船塢或碼頭＿＿＿＿＿＿＿＿＿＿

進口商店名＿＿＿＿＿＿＿＿＿＿（第　　號）

檢查情形	船名	船主姓名	進口船號數	報告日期	自何處來
	馬利安尼	享德	九八〇	一八九六年十月十三日	法國阿維爾

符號及號碼	包件數貨物情形，按照官定進口貨單名目分類	數量	價值金鎊
COK 1392	一件機製貨物N.O.E.彈子棒尖頭		28
AF 495/6	二件皮鞋……………………	十打	58
KG 340/9	十件機製棉製品，裝飾物		140
	花邊		280
	布疋，並非洋紗……………………	三百碼	8
FOT 1/10	十件皮手套……………………	一萬一千二百四十打	12,316
,, 11/5	五件寬面絲織品		10,400
,, 16/20	五件美術作品		
	石膏模型		380
	雕像		1,280
	繪圖……………………	三張	10,200
,, 21/5	五件成包書籍……………………	四百磅	300
,, 26/30	五件青銅裝飾物……………………	三百磅	38
,, 31/5	五件機製五金作裝飾用黃銅頭洋釘	四百磅	
,, 36/40	五件機製絲織衣服，斗篷，裝飾品		24
,, 41/50	十件機製貨物N.O.E.		1,816
	零星裝飾品		110
	不帶馬之馬車		160
	毛刷		78
	膠		110
	彈子棒粉		12
	鐵器		116
AJ 3/6	四件文房用墨水		48
CK 1	一件機鋼鐵機器退回英貨……………	三百磅	24

謹填報免稅貨物如上，並宣言所陳各節，並無虛僞。

具結人（進口商或代理商）瓊士(J. Jones)

（第六表）

生金銀須另用一特別表單填報，與普通報告情形不同，須單獨辦理之。

免稅貨物，如欲轉運其他口岸，即用原包封印輸送，普通罕加檢查，但統計方面，在海關中央機關，對於此種貨物，則視同應課稅之貨物，如法辦理也。

免稅貨物 課稅貨物之統計資料，取得之程序，既如上述；至免稅貨物就一般情形而論，乃佔貨物之大部，其報告之款式，既為英國國外貿易數值之原始資料，頗有加以注意之必要，茲覓得資料樣本如上。

資料之核證 如此得來之資料，平時即由主持統計事務之中央機關收下，並不再行詢問。惟代理商塡寫表格，往往多有不當，價值常有遺漏之事。遇有此種情形時，乃須責成輪船駛進之港口海關職員，飭令代理商人補塡完全，如所塡仍欠妥當，並須以其所備現行行市單核正之。至在表格中發現有顯著之錯誤，或內容有脫落，或價格迥異常度時，須由中央機關發出詢問書，交與當地海關：例如就上列之表格，可以引起下列之通訊：

(1)繪畫，價值十萬零二百鎊，價值過昂。請解釋。答：查對無誤；曾見其發票，圖畫為米萊(Millet)所繪。

(2)成包書籍；重量與價值有誤否？答：兩項均無錯誤；曾見其通知書(advice)；均為古本珍貴書籍。

（3）列爲『機製貨物，木片辮』之貨物：性質如何？請解釋，填寫有誤否，請說明。答：無誤；成辮之木片，有時摻雜馬尾及其他等物。

（4）番薯，四千磅，價六十二鎊。重量與價值有誤否？答：價值無誤。重量應爲四萬磅。

於是任何異常之項目，均可核對而加以核證矣。

差誤之可能性　遇有不易估價之貨物，或不易列表之雜項貨物，種種錯誤，必從中層出不窮矣；除此之外，差誤發生之本源，尙有一種情形，係當填寫表格之代理商人或海關職員，爲避免中央機關詢問之煩起見，以特別價值之貨物，填爲平常之價碼；然涉及主要商品及大量貨物時，凡接頭之人員，均熟知其價值之何若，填寫必不致有極大之出入，因此在重要之事項，卽有錯誤，亦可減至最小限度。由此論之，進口總價值，爲確實程度不同之極多數量之總和，凡稍一檢閱統計年報之項目者，卽可查出何項特別有陷於錯誤之可能。計此種貨物，有古書、有美術作品、有因時尙不同而變動之貨品、有賽馬用之跑馬等等，因其價值隨日俱變，輸出入差額中，甚難確定其確實價值也。

以寄售之貨物而論，寄售之羊毛佔進口羊毛之大部，代理商人並不能報出其價值。此種貨物，乃隨流動之市價而定，如就羊毛而言，其價格若干，卽全視下次售毛開價幾何。因寄售之貨品，

每批未必確能定妥應時價格，於是常有發生錯誤之可能；且因售賣之羊毛，乃以登陸及存入倉庫後之價值爲準，而登記爲進口貨值之價值，乃須不計卸貨及搬運費，是以恆有高估羊毛價值之事。

出口貿易資料　出口貨物之數量與價值，或由代理商人填妥，或由承運輪船公司代填，表格填畢後，在開船離岸後六日內，送呈海關職員代轉或直接送呈中央統計機關。茲覓得現行格式樣本，節錄如下：

英國貨物列舉表(註一五)

某港　　船名——『馬利安尼』船主亨德，開往阿維爾

（注意）　依海關法，出口貨列舉表，須於開船之日起，六日內遞交海關主管人員。

符　號	號　碼	包裝情形及件數	英國貨物之數量及說明，遵照公布出口貨物單之規定	價　值
KCL	641/2	二捆	粍原呢一，四〇〇碼	£340
CKD	140/1	二箱	機製鋼刀片三〇〇磅	24
RMO	10/12	三板條箱	機製鋼鐵打稻機一架	380
CL	140	一箱	棉織印花布疋一，八〇〇碼	50
				£794

謹填列如上並無虛言。

填報人（出口商或代理商）＿＿＿＿＿＿

地　　址＿＿＿＿＿＿

會簽者（海關職員）＿＿＿＿＿＿

一八九六年，十月十三日。

（第七表）

爲英國貨物所用表格，與爲外國——包括課税及免税——貨物所用者不同；轉運他口 (transhipment) 之出口貨物，因曾登記爲進口貨，斯時所用表格，亦有區別。在此情形之下，貨物之列舉及數量，大致可以無誤，惟價値之虛報，依然有其存在之動因。蓋貨物之須納從價税者，價値每多低估；貨物之摻雜他物者，每多冒爲眞貨，因而價値高估。此種差誤之計算，大爲不易也。

進口貨出口貨在統計上之定義　討論至此，更請就商業部報告 (Board of Trade Returns) 上出入口貨物之意義，給以確切之定義。例如在一九一三年之統計表上，進口貨價値達七萬六千九百萬金鎊，出口貨價値達六萬三千五百萬金鎊，就中有一萬一千萬鎊爲外國或殖民地之復出口貨。玆將定義條舉於下：

在進口貨之下，包括所有經過海關登陸之貨物，並包括甫經運到而留存艙上備用之物品，或買主未曾用過之退貨，但以下乃爲例外：

(a)本國人所捕之魚，用本國船隻，自漁場直接運到者；外國派駐本國之大使或公使直接攜來之貨物；自外人手中購來之舊船。又

(b)箱，袋及其他等等用爲包裹，行李包，船中存儲品，壓船重物，國有船隻上海陸軍用儲備品之用者；在保税倉庫範圍之內，轉移之貨物；輸送之貨物，經過一國而有直達貨物提單(through bill of lading)者（但仍須分作報告）；及貨物不下船而事先有聲明者。

在出口貨之下，包括所有在船貨提單上載明之貨物，但上段（b）項所列之各類，則不在此項範圍之內；自一八九九年起，自本國開出售與外國之新船，亦謂之出口貨。

貨物到達港口之後，即在本港口或他港口立即重新載回，或存入保稅倉庫而又運出者，仍謂之進口貨，但在出口貨方面，則別立名目而謂之『外國與殖民地產品之出口貨』(Exports of Foreign and colonial produce)。

生金銀與錢幣之輸出與輸入，不列爲出口貨或進口貨總値之內，須另作一表登記之。至私人攜帶之錢幣，以及進口或出口之大部金剛石（乃一極重要之項目），均未曾有何記錄也。

關於煤炭，處理方法，可爲上文所論各節說明之一助。預備作爲航程上應用之煤，只予以登記，並不記爲出口貨之列；但如作爲載貨之一種，則應包括於出口貨之內矣。

價値　進口貨之價値，乃以未及登岸之時之書面交換價値爲準，故凡付與外國商人之底價，付與輪船公司之水脚，付與保險公司之保險費等等費用，均包括在內，但貨物搬卸脚夫費並不計及也。出口貨物之價値，即爲『在船上交貨』(free on board)之價値。此項價値之正確定義，乃爲任何國家研究貿易差額上，

極端重要者。

經驗告知吾人，依貨物之輸往國家(country of destination)作出口貨之分類，與以輸自國(country of origin)作進口貨物之分類，均有莫大之困難。自一九〇四年以來，因不斷之努力，關於此等問題，確實程度，已有增進。在貿易報告中，自一九〇四年後，已用兩組表格，其中較新之一種，卽關於委託寄售貨物之國別表，現時猶佔較大之重要性。

應用戰爭時之國外貿易記錄，務須持以極大之謹愼。蓋上文所謂『國有船隻上海陸軍用儲備品』之一類，在戰時雖佔浩大之數量，而記錄上則付闕如也。

以上所論已多，而公家舉行之調查，尙未加以討論，一九〇七年之『生產調查』，可以作爲良好之例證，此項調查之結果，於一九一二年發表者卽是。總而言之，調查最後之目的，確可搜得之資料，與夫設問問題之調整，期使各業之僱主樂於確實塡答，三者乃有連鎖之關係，此吾人不可不特別加以注意者也。

（註一）　請參閱第二十六表。

（註二）　見英國人口普查委員會一八九〇年報告書。

（註三）　並請參閱統計學報(Statistical Journal)，一九〇八年號，第四頁九十六頁，及一九二〇年號，第一百三十四頁。

（註四）　參閱白狄雍著統計學初級教本第一百四十六頁。

（註五）　參閱威爾考克斯(Willcox)著『第十一次人口普查下之美國人口與

土地面積』(Area and Population of the United States at the XI. Census)

(註六) 英國一八九一年人口普查總報告。

(註七) 參閱統計學報第四十九卷,布斯氏(Booth)一文。

(註八) 每月最末而屬於常態之一星期。

(註九) 參閱本書原著者所著之『工業生產品之分配』(Division of the Product of Industry)第三十頁,及斯歐(Snow)博士,在統計學報(Statistical Journal),一九一二及一九一三年,第四百七十七頁,所載一文。

(註一〇) 一八九七年工資及工作時間(Salaires et durée du Travail)第十五,十六頁。

(註一一) 辣旦達達(Ratan Tata)基金會所發表,一九一五年版。

(註一二) 以下數節,係就一九一四年以前之情形而言,厥後有何變動,姑不論之。

(註一三) 卽自當年一月一日起,在某港口進口第九百八十艘船隻。

(註一四) 於一九〇四年,爲區別『裝載貨物之地點』(卽上表末行『自何處來』之代替標目,及『委託寄售貨物之地點』(place whence goods consigned)起見,特將表格改變,以後者一項,另添設一欄,作爲新標目。

(註一五) 自一九〇四年,新添一欄,標目定爲『貨物之最終目的地』。

第四章　表列法

空白調查表格，既已討論如上，現請進而研究用表格上所貯積之資料，列成表式之方法。在驟觀之下，以無數萬張之人口普查表格，以如是之多之工資調查表，以若干張數之進口貨物單，製列表格，似僅爲可用機械之力，只須求其準確，而無待於作科學研究之內部工作。確然，製表應用自動力量之處甚多；然以言擬定正確表格形式，選擇足與總計相符之標目（heading），則爲統計主腦人員，煞費心思之工作，亦頗有深密研究之價值者也。

第一節　總論

表列之功能　表列在一統計調查之全盤計劃上，原有頗爲確定之功能；其目的在將調查所得之答案，列成便於利用之表式。例如，今欲查明全國各區之性別及年齡別之人數，則表列之數字，必須示明此等之人數。又如就一較爲籠統之問題而論，設吾人意欲儘力之所能，調查關於工人全年收入之資料。試取工資調查之調查表而研究之，即知所能得來之資料，未必恰應吾人之所需。然則吾人當前之問題，即須用既得之資料，製成適當之表格，以

期表中所得之總數，雖不能完全合乎吾人之要求，確已盡竭力適應之能事，然達此目的之途徑，並非顯明之途徑，稍經試驗者，即可知其困難也。

所得之數字，不特須彙總歸類，以應原定計劃之所需，而於所得之材料，牽涉甚廣，性質多有不同時（一切調查莫不皆然），且須將資料作多方面之研究，並使列成之表，足爲一切學科之學者採用；彼等所需要者雖各有不同，而吾人之表，則皆足以應付之。例如，人口普查，既爲理財家、立法家、商人及商業旅行家所習用；政治經濟學者，以之持爲發展工業之見地，並爲各種商業數目變動之說明；彼對社會問題有興趣者，以之爲研究各區域各職業之年齡與性別分配之根據；社會學家與生物學家，又將視爲準確之資料，因而在人口發展趨勢與年齡分配之變化上，有所發現矣。

玆就特殊方面而言，國外貿易統計表册，可以用查本國與各國貿易發展之情狀：在國外市場，本國尙能保持原有地位耶？抑別有進展耶？本國之原料供給已有窮竭之象耶？抑另有新商品起佔重要地位耶？惟吾人須知，一則原始資料不便於一般人士之需用，必須由原始資料中，理出適於彼等之消息，二則由所有調查表中，雖可提供特殊之資料，然因此所需之工作，乃至極爲繁重，同時若製列表格稍形瑣碎，即足以隔蔽有用之消息也。

表列與特質　表列如何製法，須就前述（見第三章第一節屬性及特質一段）特質之概念論之。某一羣類（group）之各個人物，均具有充分限定之特質，譬如A，B，C，D是。有時各個分子，不只具有一種特質，一方既有E_1，E_2，E_3，……等等特質中之一，二，同時復具有F_1，F_2，F_3……幾種特質中之若干。單項表（a table in single tabulation），用以分別表現E_1，E_2，……等等特質，各該項之總計，已克盡厥職。然有時一表之標目，用直接方法或用備考方法，陳明A，B，C及D四種特質之定義，而另以E表示諸如『地域別』(in each locality)之詞字(假如E爲地域)。於是在第一縱行(column)之各列均謂之E。雙項表式（double tabulation）之作用，在表示E之分類，同時又表示F之分類，每一縱行之標目，即指明F之定義，於是每項（entry）即所以表示具有A，B，C，D，E_2及F_3特質者之人數。橫行之合計表示具有E_1，E_2，……等特質者之總數，縱行之總數，即表示具有F_1，F_2，……等特質者之總數也。

三種表列法　爲簡便起見，列表方法，可分爲三種：甲、僅對於適合某種條件之人或物，爲單簡之陳述者，如一城鎮居住之人數，或由法國進口貨值之總數量；乙、將公具同種性質之若干單位，歸併爲一類，其目的不在回答指定之問題，而實在將資料列爲表格，便於進一步之調查時應用也。例如，依年齡分類之人口

數，依工資高低分類之賃銀勞動者人數；丙、爲鋪敍整個狀態，將非屬於數字之答覆，依適當之羣類，而列成之表格。例如，罷工之起因，及服役狀態等是也。以上三類，甲乙二種並非絕對之區分，分界常時難於確定。

製列表格，應力求讀者之便利。表上之排列，應以醒目爲要，任何總計數字，一望即知。論及此點，實乃屬於排版問題，數字及標目所用活字之號數大小，如何配置適宜，紙張形式及篇幅，如何適得其當，均視排版之技術若何。凡此種種，苟能精當無比，則在一定篇幅之內，可以容納最大量之材料。

表列之分類　第一類：單項表式，只能答一個或數個各自獨立之事，例如：——

職工組合數及其會員數

歲　次	截至年底止之職工組合數	截至年底止之職工組合會員總數
1896 1897 1898	1,317 1,307 1,267	1,493,375 1,611,384 1,644,591

雙項表式，將一總計數再按兩類細分之，玆舉例如下：——

愛爾蘭貧民之分類：一八九二年，至婦女節日止

所賑濟之總數

受賑濟之人數	男	女	總計
十六歲以下者……………………………	44,391	43,648	88,039
滿十六歲及以上與六十五歲以下者	132,370	79,045	211,415
滿六十五歲及在六十五歲以上者	35,121	45,668	80,789
總計	211,882	168,361	380,243

如欲將更多之消息，包括於表內，則應如下式：——

英格蘭及威爾士貧民之分類——一八九二年截至婦女節止受賑濟之總人數

受賑濟之人數	在室內者	室外流浪者	總計	在大都會者	在英格蘭及威爾士其他各地者
不滿十六歲者	111,782	441,805	553,587	100,671	452,916
滿十六歲及以上而不滿六十五歲者………	232,284	385,299	617,583	148,066	469,517
滿六十五歲及以上者………	114,144	287,760	410,904	64,779	337,125
總計	458,210	1,114,864	1,573,074	313,516	1,259,558

三項表(treble tabulation) 之用途，在將總數分列為三類，而計每類之總數。例如下列一表，除依年齡、性別及地域各分為一類外，更介紹一特別之形式，將各類之百分數，一併列入：——

依年齡，性別及地域之貧民分類——一八九二年截至婦女節止，在英格蘭及威爾士受賑濟之總人數

受賑濟之人數	在大都會者				在英格蘭及威爾士其他各地者				在英格蘭及威爾士總人數			
	男	女	總計	各年齡所佔總數百分比	男	女	總計	各年齡所佔總數百分比	男	女	總計	各年齡所佔總數百分比
不滿十六歲者	…	…	100,671	32.1	…	…	452,916	35.9	…	…	553,587	35.2
滿十六歲並以上而不滿六十五歲者……	74,207	73,859	148,066	47.2	202,180	267,337	469,517	37.3	276,387	341,196	617,583	39.3
滿六十五歲及以上者……	27,238	37,541	64,779	20.7	136,392	200,733	337,125	26.8	163,630	238,274	401,904	25.5
性別百分數	42.0	58.0	100	…	40.4	59.6	100	…	40.7	59.3	100	…
總計	…	…	313,516	100	…	…	1,259,558	100	…	…	1,573,074	100

（第八表）

依此方法,更進一步擴張之,可成爲下例依地區、日期、性別、業別、分類排列,並附以附帶消息之四項表式 (Quadruple tabulation);但此並非妥善之辦法,除必須將數類貼近以顯示關係之密切外,與其不斷增長其繁雜性,實不如單用數個以上之表格之爲愈。然形式之變換,苟能運用得當,在一極複雜之表式內,確亦多易收比較之效也。

第二節　布斯氏對於人口普查資料之應用

人口普查所得資料之製法表 現請翻閱前於第三章第二節第一目所舉之人口普查表(第一表),可知表中調查每人之消息,有十三項之多:區域、對戶主之關係、結婚情況、所生兒女、及性別、年齡、職業、所在機關、工業情況、殘疾、出生地點、所屬國籍、房間數。項目如此之多,製表時可以製成七十八個雙項表,二百八十六個三項表,或七百一十五個四項表,是以範圍大小不同者種類甚多,足供選擇之用。

布斯氏製表法 爲求選擇意旨之決定,即以職業爲主體而分類,並就布斯(Booth)先生對於人口普查結果之利用研究之;玆爲單簡起見,只舉其於『人民之生活與勞動』(Life and Labour of the People) 第六卷第一百八十九頁之倫敦印刷工人爲例如下。

第一,布斯先生利用相當於一九一一年人口普查表之,第三、

四及十縱行之資料，作成三項分類——職業，性別及年齡——表：——

一八九一年人口普查分類	女性	男性			總數
	所有各級年齡	-19	20-54	55-	
1. 印刷工人	1,316	9,988	21,784	1,921	35 009
2. 石版工人	809	757	3,037	437	5,040
總計	2,125	10,745	24,821	2,358	40,049

復用人口普查表背面之消息，列成單項表，以示區域與人數：

東區	北區	西區與中區	南區	總計
5,884	9,835	7,577	16,753	40,049

更就第二、三、及第四縱行（性別），第二及第十四縱行（出生地），與第二及第十二縱行（工業情況）等項，製列三種關於戶主之單項表式。茲將其第二種，即用第二及第十兩欄作成者，錄之於下：

	所有有關之人口				
	戶主	其他有職業者	無職業者	僕役	總數
總計	18,048	16,060	47,257	854	82,219
平均每家	1	.89	2.62	.05	4.56

復次所製之表（此處未曾附列），係按房間數與僕役人數，作一簡單之分類，斯乃為調查表資料最為膽大之間接用法。最後一表，為四項表式——職業、工業情況、年齡、性別——之合法應用，茲附列於后：——

僱用狀況（根據人口普查）

一八九一年人口普查分類	僱主		僱員			非僱主亦非為僱員者		總計
			男性		女性			
	男性	女性	未滿二十歲者	滿二十歲者	所有各級年齡	男性	女性	
1. 印刷工人	827	39	9,988	22,565	1,266	313	11	35,009
2 石印，銅印及鋼版印刷工人，	177	2	506	2,571	88	153	6	3,503
地圖印製工人，印刷勞工及售賣者	17	…	49	175	72	62	12	387
票券及簽條編寫人	36	3	202	169	619	114	7	1,150
總計	1057	44	10,745	25,480	2,045	642	36	40,049
	$\underbrace{\qquad}_{1,101}$		$\underbrace{\qquad}_{38,270}$			$\underbrace{\qquad}_{678}$		
			僱主對僱員之比：			1比35		

（第十表）

人口普查資料之製表 欲將無量數之詳細節目，均製列成表，以供特別用途之需者，覓得較佳之例，事乃至難。人口普查當局，在許多之情形下，均未將必需之細節，製出表列，吾人如有所

需，不得不翻查原始表格，始能得出事實。對於此種工作，製表之功用，僅在解答限定之問題。例如人口普查報告，卽表示某地方某門工業依性別與年齡分類之人數各有若干，列入接連許多頁數之四項表中（每頁敍一區域）。此表可以適用於若干不同之用途，每一項目已有一立可採用之總數。人口普查之調查表上，所有各項之記述，欲逐項製列表格，卽爲事實所需要，然爲時間及篇幅所限，乃亦勢所不能；一良好之表式，只供給實際有用之事項而已。僅爲敍述性質之總數，已有若干——如英國各教區依性別及年齡分類之人數是——其主要用途，卽爲供行政目的之需；又有許多堪供經濟學家及社會學家考查工業進步情況，研究各種職業工作人員之年齡，觀察一國各級年齡之變化；更有其他表格，可爲對於特別問題研究之一助者。總之，每一度舉行人口普查，必有新表格發現也。

次要之細目 試展開任何此種數字表之一，任意抽出一數，然後查問；『此一數字，何爲而印出之？所解答者，爲何問題？可供給何人之用？』舉例言之，在『職工組合第八次報告』（第二百五十七頁），可查出聯合造磚工人會，及運磚碼頭勞動者組合，在一八九四年之時，葬喪費一項，耗金二十鎊，平均每一會員，須費三先令七又四分之一便士。此一數字，單獨陳述，引起注意者必少，然此少數之人，卽可據以促成將關於工會之數字，製成表

格，列之普通公家刊物中；如在同一頁上，查得鍋爐製造工人所耗之五千四百八十一鎊，則二十鎊雖小，亦同樣能引人之注意矣。由此觀之，諸如此類之微末節目，是否一併搜羅加入，純為篇幅之問題。如篇幅有限，只得就所關人數較衆之較大數量，斟酌選定之。

原始資料之重要性　然此亦不可一概而論，諸如此類之細目。在另一性質上，又殊有加入付印之理由，蓋此種表格之總計數，均係根據原始資料而來，而原始資料，初學之人，設非藉助該項報告（Report），難期瞭解應用。今編製此種統計者，並不知用資料作研究者，究竟持何觀點。吾人固可將各工會加以分析，並按其各項費用，彙分類別，然後考察其歷史，分為鬭爭組織與互助會性質等類別。一般所需要之表式，事先並不能完全明瞭，故對於資料只能供給其概略，以便有何需要者，各自製列其適當之表格。同時，足供參考之總計數，於此製表法中，即可出現；並於報告之總綱中，將各項重為製表，原始資料，即行取銷，至究應如何整理，則一視編輯人之思索，擇其最為有用者決定之可也。

原始資料之選擇　當發表時，篇幅過狹，不能容納所有各項時，何者當付印，何者不當付印，必須慎為選擇；而選擇之當否，即普通最易引人批評之所在也。

一九一一年，英國人口普查之科文垂（Coventry）郡邑 County Borough)，可作一良好之例證，該邑一百一十五人之事實，詳細情形如下：——

磚、三合土、陶器及玻璃業，男性人數

年齡	10.	13.	14.	15.	16.	17.	18.	19.	20.	25.	35.	45.	55.	65.	總計
製造工人	—	—	2	3	2	3	1	4	12	36	23	17	4	—	107
售賣員	—	—	—	—	—	—	—	—	1	4	—	1	1	1	8

至在腳踏車及汽車業服務之男性——廠主、工頭、有技能工人，粗工工人及僮僕——所敍細目除下列一表外，恐無有再細於此者：

車輛

年歲	10.	13.	14.	15.	16.	17.	18.	19.	20.	25.	35.	45.	55.	65.	總計
腳踏車及汽車製造工人及機械工人	—	1	912	271	303	325	371	372	2,003	3,872	2,488	1,122	379	76	11,775
汽車製造工人及機械工人	—	—	*70*	*108*	*143*	*160*	*219*	*208*	*1,299*	*2,524*	*1,376*	*537*	*158*	*36*	*6,838*
其他	—	—	—	2	1	4	8	7	31	62	49	24	21	4	213

據該書序言第三頁所述，每區較為重要之職業人數，全部節目，均用斜體字標出云。

地位之節省　在此情形之下，有二條極有效之準則，可以實施：第一，將不滿五百之數略去，可以少印一橫行，第二，數在一萬以上者，取至百位爲止，其他數字以比例類推，如此則排印之時，各縱行寬度可以縮小。例如，如取至百位爲止，各區域及各職業之確實數，受僱用者有一千人，則吾人所得之消息，亦頗滿足吾人之需要，且如此製成之表，所佔地位，是否較業已全行包括者爲多，事乃頗堪懷疑。雖然，在許多情形下，原始材料務須保持原狀，不必任意變動。各種表列固各有其優劣也。

第三節　農村工資統計之表列法

一八三三年恤貧律調查結果之表列　吾人講求有效之程序，必須提出某類問題，然後考慮表式之何若，換言之，必何如列表，對於有關之問題，可收明瞭之效果。一八三三年之恤貧律委員會(the Poor Law Commissioners)，自英格蘭及威爾士一千個農村中，收集之消息，主要之問題，只有六點；每一農村勞動者在冬日夏日之工資，麥酒作爲一部工資與否，其全年收入若干，及其妻與兒女之補助工作收入。可以斷言者，該委員會之主旨，在調查勞動者家庭之收入，是否足供其生活之所需，及妻子兒女之工作收入，佔何等比例也。

各郡(County)之貧窮勞動者，可由下列表樣示明之：——

郡名	平均全年收入		
	男子	家屬	共計

各郡郡名，不妨按字母（指西文字——譯者註）次序排列，以便參考，或按地理順序排列，以便將一組各予一平均數（例如，東區：腦富克、沙富克、愛塞克斯；否則可按收入總數多寡次序排列，以便表示最為窮困之勞動者係在何郡。

如按各郡之次序，宜示工作收入低於某一最低限度，或未超過某種範圍之各農村數，可用下列一表：——

男子與家屬之全年工作收入

	全年收入已加平均之農村數							各郡平均全年收入			家屬工入收入佔總數之百分比
	金二十五鎊	已超過二十五而未達三十金鎊	已超過三十而未及三十五鎊	超過三十五而未及四十鎊	超過四十而未及四十五鎊	超過四十五而未及五十鎊	五十鎊以上	男子	家屬	總計	
腦富克	0	1	3	6	4	3	2	£30	£11	41	27
對全部農村之百分比	0	5	16	31½	21	16	10½	…	…	…	…
沙富克	0	3	4	5	3	2	£2	£28	£11	£39	28
對全部農村之百分比	0	16	21	26	16	10½	10½	…	…	…	…

愛塞克斯	1	3	6	7	10	3	1	£28	£10	£38	26
對全部農村之百分比	3	10	19	23	32	10	3	…	…	…	…
東部各郡	1	7	13	18	17	8	5	£28 10	£10 10	£39	27
對全部農村之百分比	1	10	19	26	25	12	7	…	…	…	…

（第十一表）

此表列爲如上之複雜形式，固無不可，但如使變爲簡單化，亦屬可能。關於錢數之分組，初無一定標準，一方須視地位之大小，一方又須視各組所包含之農村數。設列成之表，多數行列，內容只有一或○，最易引人批評。即以上表而論，農村數亦已過少，在百分比上，難期準確也。

表示相關之表列法　此一表格，令人一見，即可信其對全年收入總數所能引起之任何問題，此表幾全可解答之。例如，設吾人欲考查收入總數對家屬補助收入之關係，即須檢閱上表倒數，第二縱行之最小總計數，視其是否與家屬收入所佔之最大百分數相合。如其然也，吾人即將諸郡依此補助收入之多寡次序重新排列，然後查視其能否與以收入總數爲次序而排列者略相符合。此爲用表列法表示相關（correlation）——兩組現象出現之相應性——之一例也。

每週工資與全年工作收入　此種表式論述至此，另外一組之重要問題，必因之引起：每週工資與全年工作收入之關係若何？

就一般情形而論，以實物償付工資，佔工資若干成？至關於家屬補助收入，姑置不論。查農村工資之記錄，最普遍之記述，為『本區工資每週自十先令至十二先令不等』。然因日光不足之時，所需之工作量較少，而工資亦不得不隨之減低，故農村勞動者在冬日勞動之所得，一般均不若在夏日所得之多；職是之故，如每週工資以五十二乘之，得數必較全年收入為多。惟在收獲之時，除額定工資外，多可獲得特別賞錢，即在平時亦有實物償付之事，諸如每日之麥酒，減收租金之房屋場院及其他特別權益等等。是以最善之法，莫過於將此等價值估計於內，而計算其全收入如下：——

	鎊	先令	便士
每週十先令共三十八週	一九	○	○
每週十二先令，共九週（夏季）	五	八	○
割禾一星期	○	一五	○
割麥四星期	五	○	○
麥酒每週一先令	二	一二	○
茅舍及場院	五	○	○
其他犒賞	一	五	
	三九	○	○＝每週十五先令。

由此觀之，全年收入較之普通每週工資，增多百分之五十。類此性質之估計，已故立特路(Little)先生，曾為各郡自一八六

七年起至一八七〇年止，及一八九二年，完全算出。

冬日工資與夏日工資　尚有一問題：冬日工資一般均較夏日工資爲低乎？低幾何耶？可由下列表式——所用資料，乃爲前列諸表所未有——解答之：

郡別	平均每週工資				夏季工資較冬季爲多之農村數					
	夏季		冬季		並不多	多六便士	多一先令	多一先令半	多二先令	多二先令以上
腦富克	先令 11	便士 2	先令 10	便士 3	13	2	3	2	5	3
所包括村數之百分數					46	7	11	7	18	11
沙富克	10	2	9	8	24	0	6	1	2	1
所包括村數之百分數					70	0	18	3	6	3
愛塞克斯	10	9	9	10	22	0	11	0	5	4
所包括村數之百分數					52	0	26	0	12	10
東部各郡	10	6	9	11	59	2	20	3	12	8
所包括村數之百分數					57	2	19	3	12	8

（第十二表）

以上所舉各例，未足以盡表列此等數字之妙用，因未曾對於工資之分佈——按高低不等之定率，給付工資之比較人數——加以分析也。然此乃缺乏個人工資率之調查而僅有各村通行工

資率之故，可見此項調查結果，用作例證，以明表列之功用，實有未見完善之感。

第四節　美國工資統計之表列法

第二類表列　茲以工資分組，作爲第二種表列法之例證。現時吾人並無需要解答之特定疑問，與已討論甚多之方法相同，僅一較爲普遍之問題而已：已有一批資料，吾人須用之列成表式，以期表現有用消息之最高量。原始資料既有萬千，各自不相關涉，故須使之集中，表示特定之意義，並加以整理以爲將來比較之用。

用途不定之統計資料　有數種調查，舉行之目的，既不爲解答任何特定之疑問，亦非用爲研究某項問題之一助，僅爲收集消息而已，此項消息，雖無迫切之需用，然遇有懷有諸多疑難欲舉行調查者，則此項資料，恰能滿足其所需。工資調查，即屬此類。當吾人未能求得充分工資記載之時，工資調查雖爲社會團體測量最重要者之一，而所得消息，不堪應用，於是經濟學家與統計學家，竟以缺乏重要資料之故，工作常受阻撓，然工資調查，並無急切之實際用途，蓋既知其工資之高下，亦無助於吾人對工資數額有所管理也。故對於此種調查，吾人之目的，即在將數字加以分析，並將其分組與平均，以求任何用途皆可適用；吾人之工作

如此，則在進行之時，無形中將渡入一類完全不同之調查事業中矣；種類萬千之詳細節目下，吾人將查出其結構之何若矣；數字在驟觀之下，所呈之混亂，將亦發覺其並不背乎定律矣；此時吾人可以作成現諸文字之綱領，並對於顯然並無特定色彩之大批資料，亦查出其確定之形式矣。

關於此一問題之整個研究，乃屬下章範圍之內，惟表列法無須專門之技術，此時不妨開始討論之。以下所舉之例，概以關於工資者爲主，惟列表方法，仍極爲普通也。

組距之選擇　在美國『一八九一年批發價格，工資及運輸之報告』一文中，曾敍有數約千人之工資詳細情節。玆假設其爲同質之羣類，而討論表列之方法。其表式如下：——

工資表式——美國數字，一八九一年

1. 每日工資 自此數起	不滿此數	2. 人數	3. 自此數起	不滿此數	4. 人數	5. 自此數起	不滿此數	6. 人數	7. 百分數	8. 各組平均工資	9.
	元			元			元				
.25	.35	1	.25	.45	16						
.35	.45	15									
.45	.55	59	.45	.65	144	.25	.75	317	6.2	.26非	.50
.55	.65	85									
.65	.75	157	.65	.85	270						
.75	.85	113									
.85	.95	169	.85	1.05	370						
.95	1.05	201				.75	1.25	1,472	28.7	1.09	1.00
1.05	1.15	304	1.05	1.25	989						
1.15	1.25	685									

1.25	1.35	99	1.25	1.45	557	1.25	1.75	1,297	25.3	1.49	1.50
1.35	1.45	458									
1.45	1.55	466	1.45	1.65	538						
1.55	1.65	72									
1.65	1.75	202	1.65	1.85	531						
1.75	1.85	329				1.75	2.25	970	18.9	1.99	2.00
1.85	1.95	58	1.85	2.05	331						
1.95	2.05	273									
2.05	2.15	45	2.05	2.25	310						
2.15	2.25	265									
2.25	2.35	33	2.25	2.45	134	2.25	2.75	506	9.9	2.53	2.50
2.35	2.45	101									
2.45	2.55	196	2.45	2.65	209						
2.55	2.65	13									
2.65	2.75	163	2.65	2.85	165						
2.75	2.85	2				2.75	3.25	198	3.9	3.04	3.00
2.85	2.95	15	2.85	3.05	144						
2.95	3.05	129									
3.05	3.15	5	3.05	3.25	52						
3.15	3.25	47									
3.25	3.35	12	3.25	3.45	12	3.25	3.75	254	5.0	3.51	3.50
3.35	3.45	0									
3.45	3.55	221	3.45	3.65	226						
3.55	3.65	5									
3.65	3.75	16	3.65	3.85	27						
3.75	3.85	11				3.75	4.25	96	1.9	4.00	4.00
3.85	3.95	0	3.85	4.05	82						
3.95	4.05	82									
4.05	4.15	0	4.05	4.25	3						
4.15	4.25	3									
4.25	4.35	0	4.25	4.45	0	4.25	4.75	4	0	4.50	4.50
4.35	4.45	[illegible]									
4.45	4.55	3	4.45	4.65	4						
4.55	4.65	1									
4.65	4.75	0	4.65	4.85	0						
4.75	4.85	0				4.75	5.25	8	.2	5.00	5.00
4.85	4.95	0	4.85	5.05	8						
4.95	5.05	8									
5.05	5.15	0	5.05	5.25	0						
5.15	5.25	0									
5.25	5.35	1	5.25	5.35	1	在5.35		1		5.35	5.25
總計		5,123			5,123			5,123	100		
平均工資		$1.731							平均工資		$1.70

（第十三表）

原來發表之件，工資尾數至五釐爲止；在上表之第二欄，曾

銀勞動者之人數，按一角爲一組，卽自二角五分至三角四分，三角五分至四角四分，餘類推，至所賺工資恰在分界騎縫中間者，卽視作下一組之數。惟須注意，二元一角五分至二元二角四分一組之平均工資，如賺工資者之人數，以分爲單位分組，必非爲二元二角，而爲二元一角五分，二元一角六……二元二角四分之平均數——卽二元一角九分五釐。

試閱第二欄，數字並無次序，亦無一定之準則；結構如何，未能查出，可知分組過狹，資料不甚適用也。

現請將組距放寬，重將工人分組，卽如第六欄所定，將人數按半元爲一組，乃可見人數順序似尙整齊，且隨第二欄之最高量而出現。返觀較小組距之分組，究竟用何等分組時，最初發現此種整齊性，則於第四欄以二角爲一組者，卽已顯示，雖非絕對，卻亦甚爲可觀之有規則性矣。以三角爲一組者之人數，依次爲75，355，674，1242，740，660，343，310，180，181，233，32，82，3，4，8，1，除三元二角五分至三元五角五分一大組外，幾已全部呈現有規則形態矣。

至於擇用何種分組問題，須視各項項數可以一目瞭然爲定。一角爲一組之五十一組，或二角一組之二十六組，驟觀之下，只有數字一堆，意義已全失去（一切細目固可用圖式正當表明），就中只有五角一組之十一組，頗易使人領會也。

上表第七欄之結果，以文字述之，乃爲賃銀勞動者，工資自二角五分至七角四分，佔百分之六，自七角五分至一元二角四分，佔百分之二十九，其他以此類推。

實際製表法 由原始數字製成表格之實際工作，須用方格紙，順序於各欄之首，分列工資之組距，然後以每項工資之大小，於各該分欄內，劃一直線，隨手以五線或十線爲一束以代加法。

依上文各節所論，可知由二角五分起至五元三角五分止，無須以一分爲一組，而分列一欄記錄之，然爲求得正確平均數起見，組距應分至何種細密程度，仍有略加考慮之必要。

設以一分爲一組之各項如下

\$1.70	\$1.71	\$1.72	\$1.73	\$1.74
11111 1	11111 11111 11111 11	11111 11111 111	11111 11111	11111

如此列入之工資，平均數可立時算出，乃爲一元七角一分八釐。

然若吾人將此五十一項，定爲『一元七角至一元七角四分』，或用較確實之語句言之『自一元七角而不及一元七角五分』，所有各項均爲此組之中心點，換言之，卽一元七角二分也。

如有項數甚多，則假設之平均數，與各組所算出之平均數，

二者必相差甚微。此於第十三表可以見之;第八欄所列爲一角一組各項算出之平均數,而第九欄爲基於一假定　爲平均計,半元各組之數,均作爲在各該組之中點——之平均數。第一及最末之相差爲最大。由第九欄所得之總平均數爲一元七角,卽爲對眞實平均一元七角三分之最近似整略數(round number)。故爲總分組及平均計,僅取半元爲一組之欄十一項卽可。

爲其他目的,不妨作更詳盡之分組;蓋於最低之一組,吾人欲知賺二角五分者有幾人,賺三角者有幾人,三角五分者有幾人,而五分對二角五分而言,其差甚顯而易見也。至在上端,用此法,亦可求知確切工資爲若干也。

高爾頓式(Galtonic)法　表列第二法,對於較詳細之項目,亦有其需要之處。設吾人將工人依其工資之數額排列之,一端爲二角五分,一端爲五元七角五分。則請注意橫行各點之工資,工資最低者爲二角五分;自此以往,十分之一處,第五百一十二個工人之工資,爲自八角五分至九角九分,……進至中途工資爲一元五角。每至十分之一之數字,列表於下。由此吾人對於依工資之分配,可以思過半矣。

此等數字,在按半元分組時,必難確實相合,惟在一角一組時乃可實現,故以一角爲分組之法,殊有採用之價值。吾人於此必須首先決定在何一小組中爲十分之一工人,十分之二工人⋯

十分之一工人之工資	
工人工資	最低工資……………… \$.25
	往上第十分之一組…… .89
	往上第十分之二組……1.12
	往上第十分之三組……1.22
	往上第十分之四組……1.39
	往上第十分之五組……1.49
	往上第十分之六組……1.75
	往上第十分之七組……1.99
	往上第十分之八組……2.36
	往上第十分之九組……2.98
	最高工資………………5.35

平均工資	一千工人之同等數
最低之十分之一…\$.70	.79
第二之十分之一…1.03	1.00
第三之十分之一…1.18	1.24
第四之十分之一…1.28	1.50
第五之十分之一…1.44	1.50
第六之十分之一…1.59	1.88
第七之十分之一…1.86	2.00
第八之十分之一…2.14	2.22
第九之十分之一…2.59	2.58
最高之十分之一…3.51	3.55
總平均數　1.731	1.82

（第十四表）

…所在，然後計算其在更小組之位置。例如設吾人欲求較『八角五分至九角五分』尙爲準確之數字，其程序如下：——自底端之第五百一十二人，爲八角五分至九角五分一組之第八十二人，因工資低於八角五分者，已有四百三十人，而本組共有一百六十九人，如人數分配果甚整齊，每一分各佔十七人，則第八十二人，卽在八角九分至九角一小組之中間。均勻分配之假定，在若干用途之下，頗有充分之正確性，而此法用於以十分之一地位決定工資，亦足供爲確實之工具。算出之結果數字，已列表於上。雖然，設吾人欲確知中間一人之工資，則由一元二角五分至一元七角五分

之半元一組中，可知其乃或在一元四角五分至一元五角五分之內，然後即速查原始之資料，將工資分隔爲一元四角六分，一元四角七分……乃至一元五角五分。（註一）

此法若略加變動，亦甚便利。取最低之五百一十二（或十分之一），其平均數爲七角零五釐；次一十分之一，則爲一元零三分；其他以此類推（見第十四表）。觀此數字，亦頗足予吾人以深刻之印象，且用之以與他類相比較，亦極爲便利也。

以上所提之數字，只爲原來報告材料之一半。適纔所列之表，即係各級十分之一之平均工資。將如此得出之兩組加以比較，可以看出前半部足爲全部之代表。

第五節　工資調查之表列法

一八八六年之工資調查，以如下之款式，搜集得來之資料而列成之表格，可以用爲解釋各種困難之例證。

工　資　率（註二）

絲　織　業

一八八六年十月一日各區絲織業僱工平均工資率

第一區：——且舍爾，斯塔夫，腦次，戴爾具，及窩維克

一八八六年十月一日僱工總人數　　一八八五年所付工資總數

見附表， 五，二七九人　　　　見附表， 一二九，五八八鎊

職位	僱工人數另詳報告	各職位上有報告之人數對全體有報告之人數之百分比	平均每週工資率		工資在平均數上下不出十分之一之有報告人數	每週工資率之平均最高數		平均最高工資率所根據之人數	每週工資率之平均最低數		平均最低工資率所根據之人數
	男工										
	總數一〇八五					每週工作時間五六又二分之一					
絲線部			先令	便士		先令	便士		先令	便士	
監工員及執事員……………時工	79	1.5	23	3	35	31	3	4	20	0	29
監工助理員……………時工	9	.2	17	3	6	19	0	…	15	0	…
紡絲工人…時工	24	.5	11	9	15	13	0	5	9	3	4
搓絲工人 時工	109	2.1	17	7	68	20	11	18	14	0	3
搓絲工人 件工	9	.2	18	8	7	…		…	…		…

（第十五表）

此表格之大部，卽在表明下列事象：

人數　平均工資

紡絲工人——時工：　六　十二先令　五十六小時半

工資調查所用之表列法 如此得來之報告，並非完全確定，因如同一職位同時僱用多人，則工人工資率有多寡不同之可能。卽以六人之項而論，所列之十二先令，則有下列兩種途徑之可能：一爲六人每人各賺十二先令，一爲賺十先令者二人，十二先令者二人，十四先令者二人（平均恰爲十二先令）；如不用平均數，而言一般工資率，則或爲賺十二先令者四人，賺十五先令者一人，

賺十一先令者一人；或爲賺十二先令者五人，十八先令者一人。工資調查之目的，既在對工資爲綜合之研討，以備一切調查之採用，故各業之人數，僱工依年齡性別與區域之分組，各組工資之一般工資率及平均工資率，在在均須示明，且關於各組對平均數之分配，亦須詳細爲充分之敍述，否則僅一平均數，對於特高或特低工資之眞象，往往蒙蔽不見也。

關於原始資料是否較上述形式詳備，又特異之處是否可以遮蔽，經著者向勞工局查詢，查知職位之分組，實際計算上，在一標目之下，工人之工資，已有甚大之變動，故不如將標目分裂爲數個，將各組分別登列，否則卽須在一標目之下，各組應行分列使易辨別；但由其他調查報告觀之，此層並未作到，幸尙有補充之調查，以補其不足，故原始資料，已甚爲詳盡，勿論列成何種精細之表式，均能措置裕如。

然則現在之問題，卽須將一區域內各工廠之調查報告，列簡而明之表式，表示各組與全體之工資分配情形。由此觀之，第十五表例舉工資率調查表式樣，所用方法，謂之盡美盡善，實難令人置信也。

吾人爲求切實明瞭起見，茲假設關於搓絲工人（時工），所根據之詳細內容，係如下所列：

三人賺　十四先令……………『平均最低工資率』

十四人賺	十五先令	
六人賺	十五先令六便士	六十八人在全體平均數，即十七先令七便士，百分之十以內。
二十人賺	十六先令	
一十人賺	十七先令六便士	
二十人賺	十八先令	
八人賺	十八先令	
十人賺	十九先令	
十人賺	二十先令六便士	十八人平均賺二十先令十一便士
八人賺	二十一先令五便士	

各種應有之處置　列表時所用之程序，假設已將一小部賺工資遠較平均數爲低之老弱工人，由全部調查表中剔出，另列爲最低工資一類，同時並將特別優異之有技能工人提列爲最高工資一類。此種辦法，較之僅將各個工資之最高最低額標明，自爲佳妙，蓋最低最高二者，均可有特別之情形，以致付與他人之工資，相差甚遠也。至此等極端項之大小，仍以原來報告爲準。經此移提單列之後，所餘之工資，仍未必即能密接成爲一類，如上例工資自十五先令起，至十九先令止，散佈仍甚廣。關於此種分配情況，可爲解決疑難之線索者，幸以工資在平均數十分之一以內者之人數尙有標明也；僅用一欄即可標列，尙不失爲一最佳妙之途徑，惟十分之一，範圍仍嫌甚廣耳。此外尙有一法，亦可圈定一

範圍，在此範圍之內，凡工人人數十分之一之工資在平均數之上，及人數之十分一在平均數以下者均包括之：即爲十六先令及十八先令是也。

雖然，設每一羣類佔用不及八欄，下列七則，即足供給較多之消息，且即用旣得材料，已足應付一切，雖用途有所不同，亦可隨其需要，適當取材也。

所僱人數	一〇九
每週平均工資	一七先令七便士
工人十分之一人數所賺不過一五先令	
工人四分之一人數不過一六先令	
工人二分之一人數不過一八先令	
工人四分之一人數所得在一九先令以上	
工人十分之一人數所得在二十先令六便士以上	

此即爲一九〇六年工資調查結果發表之報告中所用之法，惟內中並無十分之幾之項而已。

一經讀畢本書第五章及第六章，讀者自然可以不用上文所用之詞句，而用中位數(median)四分位數(quartile)及百分位數(decile)等名詞，然後可以對於用一測量離散度方法，是否較之頃已提及之細節，較爲適用一問題，加以討論矣。

提要　以上所論之表列法，只爲其中之一種，舍此而外，並

非無他法可尋，例如一八八六年（註三）全般工業之工資，亦曾列製成表；該種所用之格式最適於比較用途，茲將原式列下：——

各級工資之僱工人數及百分數

下表示一八八六年十月各業僱工之每週平均工資，及各級工資上之人數及百分數。

		十先令以下	自十先令起在十五先令以下	十五先令起在二十先令以下	二十先令起在二十五先令以下	二十五先令起在三十先令以下	三十先令起在三十五先令以下	三十五先令起在四十先令以下	四十先令以上	總計	平均每人工資
棉織工業	人數	2	370	8,793	8,822	4,525	7,283	1,582	812	32,189	25 3
	百分數	…	1.2	27,3	27.4	14.1	22.6	4.9	2.5	100	
織呢工業	人數	…	146	3,377	5,559	1,725	705	392	344	12,248	23 2
	百分數	…	1.2	27.6	45.4	14.1	5.7	3.2	2.8	100	
織線及織品工業	人數	…	835	1,705	909	2,635	879	28	14	7,005	23 4
	百分數	…	11.9	24.3	13.0	37.6	12.6	0.4	0.2	100	
葛布工業	人數	192	780	2,952	2,070	416	290	39	68	6,807	19 9
	百分數	2.8	11.4	43.4	30.4	6.1	4.3	0.6	1.0	100	
黃蔴工業	人數	…	565	1,038	964	127	53	52	…	2,799	19 4
	百分數	…	20.2	37.1	34.4	4.5	1.9	1.9	…	100	
大蔴工業	人數	…	25	300	581	168	39	94	25	1,232	23 6
	百分數	…	2.0	24.4	47.2	13.6	3.2	7.6	2.0	100	
絲織工業	人數	…	324	881	367	278	121	273	4	2,248	22 3
	百分數	…	14.4	39.2	16.3	12.4	5.4	12.1	0.2	100	
毛毯工業	人數	…	…	130	183	834	100	15	30	1,292	26 7
	百分數	…	…	10.1	14.2	64.5	7.7	1.2	2.3	100	
織襪工業	人數	…	…	296	458	51	190	75	…	1,070	24 5
	百分數	…	…	27.7	42.8	4.8	17.7	7.0	…	100	

（第十六表）

列百分數之各行，應用之處極多。各種工業之工資水準如何，一望而知。例如，棉織工業之平均工資，較毛織工業，高出二先令；且棉織業中，程度甚高之有技能工人，賺工資自三十先令至三十五先令者極多，而毛織業工資接近平均數者幾佔全數之半，大約多在二十至二十五先令之間也。黃蔴及葛布工業中，其平均數大致相同，然前者則有賺十五先令以下者甚多。絲織業中，與棉織業同有其特殊階級，惟絲織業之高級工人人數既少，報酬又無棉織業之高，計工資在三十五先令以上者，只有百分之十二。由此觀之，此一表式，何等緊湊，何等清晰，謂之為集中及明顯之傑作，誠非過言。

第六節　工資變動統計之表列法

工資變動數字之列表法　現請就英國勞工局所搜集與工資率之變動有關數字，討論其表列之方法。下舉一例係用舊日之報告，嗣後表式業已經過多番改革，吾人將其加以研究比較，然後查閱歷年報告案卷，必能獲益不少。據查，舊日之空白表格，所搜集之情節，包括有關工人之職位，及人數，發生變動之日期，及變動前後全時工作星期中之每週工資與工作時間，惟加工時間及工資（加工之意義，與工資調查所用相同）不計之。

一八九四年英國各區普通農村勞動者之工作時

間及工資率變動情形表（根據商業部所收到之報告製成者）摘錄（註四）

區域	夏日工資變動情形（一八九四與一八九三比較）		冬日工資變動情形（一八九四與一八九三比較）		一八九一年，農村勞動者，田莊常工，牧畜人，馬夫，騎士，御夫，駕車夫之人數
	增	減	增	減	
		每週	每週	每週	
林肯協爾					
根茲伯裏	…	…	…	1/6(15/至13/6)	2,466
老斯	…	…	…	1/6(13/6至12/)	3,932
斯皮爾斯拜	…	…	…	1/6(13/6至12/)	3,288
腦富克					
阿爾山姆	…	1/(12/至11/)	…	…	2,576
夫萊哥東區	…	6d.(12/6至12/)	1/(10/-11/)	…	2,487
西區	…	1/(12/至11/)	…	1/(11/至10/)	1,108
佛爾條	…	…	…	1/(11/至10/)	1,448

（第十七表甲）

根據英國商業部所收到報告製成之一八九五年夏英國各區普通農村勞動者工資率變動情形表摘錄（註五）

郡名	一八九一年農村勞動者，田莊常工，牧畜人，馬夫，騎士，御夫，駕車夫之人數	夏日工資變動情形（一八九五年與一八九四年比）	夏日每週工資率			
			一八九四		一八九五	
		每週	先令	便士	先令	便士
得山漢姆						
斯他克頓	437	減六便士	17	6	17	0

提斯得爾（把拿山寨鄉村區）	669	增六便士	17	6	18	0
牛津						
亥丁頓	1,118	減一先令	12	0	11	0
亨累	1,587	減一先令	12	0至	11	0至
（漢佈爾頓鄉村區）			14	0	13	0
腦富克						
夫萊哥東西兩區	1,108	減一先令	11	0	10	0
佛爾鍭	1,448	減一先令	11	0	10	0
亨斯台得	1,504	減一先令	11	0	10	0
密堤佛與耶第斥	3,622	減一先令	11	0	10	0
斯毛爾白爾夫	2,264	減一先令	11	0	10	0
斯濕夫漢姆	1,942	減一先令	11	0	10	0
魏爾	1,535	減一先令	11	0	10	0
卡那文務爾						
卡那文（歸費鄉村區）	1,124+	不供膳者增一先令	19	0	20	0
		供膳者增一先令	11	0	12	0

（第十七表乙）

農村工資變動之表列法　以上兩表，指示吾人第二及第三次工資率變動及工作時間報告中，對於農村工資之變更，製表方法之不同。第十六表甲中，爲『增』『減』各設專欄，以資識別，因而所佔地位甚多，用意原爲甚善，然『所謂一八九四年冬』，究竟係以冬季起耶，抑至冬季止耶，並未明白指示。

第十六表乙，僅言夏日工資一端，各欄排列亦大不相同；增減列於一欄，可以不同字體排印以別之。然至第五次報告中，內容愈形清楚矣，例如：

冬日工資（註六）

區域別	人數	每週工資率				一八九七年每週之增減			
		一八九六年正月		一八九七年正月		增		減	
		先令	便士	先令	便士	先令	便士	先令	便士
吞道令	3,113	10	0	11	0	1	0	…	…

（第十七表丙）

此表係表示冬日之工資，至夏日工資之表列亦同。

所關人數　是項農村表格中之人數欄內，尙有缺點，殊爲美中不足。蓋所調查者如爲其他產業，則報告上之人數，必均爲確實有關者，而此處所報之數，欲求其正確無誤，勢乃有所不能；且在一八九一年之人數，乃指『農村勞動者』而言，然如上表所列，該欄包括各色人等甚雜，非全爲農村勞動者也。夫鄉村之工資一有變動，雖在減低工資時，遇有善良之僱主，未必一律施行，但吾人假定其必爲普遍現象，似並無不可，且在一星期內，工資之變動，雖未必普及於全區，但在此方面大半似不致有甚大之出入。蓋工資更動之時，多在冬季工資改爲夏季工資，或夏季工資改爲冬季工資之際；工資之增減，不必有形式之舉動，應減爲冬季工資，或增爲夏季工資之時，略行拖延數日，則事實上工資已有輕微之增減矣。就一般而言，如認爲每一變動卽足影響全區內所有之成年農村勞動者，自不免產生少許之錯誤，且除工資較普通爲低之老弱工人，情形容有不同外，其他若馬夫若牧人等等之工資，只有依比例之變動，亦大有可能。（註七）調查表格上有『本區強壯勞工之約略數』一問題，惟答案旣無所用，所報告之數，謂其不甚確正，似可爲斷言也。

資料之缺乏　全部表式之目的，在示明全國每週工資之變動，惟以缺乏若干詳細事項之故，致不能作全部之計算。卽以農

村勞動者而論，吾人所需要之資料，除上文所論等項外，尙有額外收入，特別報酬及實物償付數端之變動報告。就全般情形而言，全部工資及變動情形之報告，均應不厭求詳。至關於農村勞動者之資料，英國勞工局年有發表；四季工資，有無變動情形，工會團體六百家中，按年多有報告呈交該局也。

各郡工資率之變動　關於人數一項，調查表格塡答雖欠精確，但此於各郡各鄉工資率變動之計算，並無妨礙，蓋每一區域受工資變動影響之人數，其對於人口普查所報告之人數，所成之比例，必與全郡或全鄉同類農業勞工之人數，對於人口普查所報告之人數比例同。故吾人可藉下章所論加權平均數之原則以解決之。

玆計算第十七表乙得由漢姆一郡一八九四至九五年，夏季工資之變動情形，如下所列：

	變動前之平均數	變動	受影響者之比例數	工資單上變動之總金額
	先令　便士			先令　便士
斯他克頓	一七　六	減六便士	四	共減二
提斯得爾	一七　六	增六便士	七	共增三　六
				一

（第十七表丁）

全郡變動總計，增一先令六便士

佔全郡之比例數，73.

$$對於全郡平均數之影響，\frac{1/6}{73}=\frac{1}{4}便士。$$

玆爲計算簡便起見，受影響之人數，以百人爲準，如此計算，對平均數，不致發生顯而易見之影響。此一簡略方法，只能就原始資料之可能，得出相當確實之得數。將此方法，適當加以修改，即可用以計算其他產業之工資變動。現先將各郡農村工資變動調查結果，總括列表於下：

英格蘭及威爾士數區，一八九六及一八九五年

所付每週現金工資變動純淨效果之比較（註八）

區域別	一八九六年工資與一八九五年者之比較				一八九五年工資與一八九四年者之比較		
	* 總數	工資變動對每週工資之純淨效果 增(十) 減(一)			* 總數	工資變動對每週工資之純淨效果 增(十) 減(一)	
		總額	每人			總額	每人
英格蘭		金鎊	先令	便士		金鎊	便士
北部各郡	5,662	−43	−0	$1\frac{3}{4}$	3,766	+44	$+2\frac{3}{4}$
約克協爾，蘭加協爾及奇協爾	2,897	+100	+0	$8\frac{1}{4}$	3,942	−126	$-7\frac{3}{4}$
東部中部各郡	69,869	+666	+0	$8\frac{1}{4}$	89,576	−2,045	$-5\frac{1}{2}$
南部西部各郡	20,901	−340	−0	4	20,441	−575	$-6\frac{3}{4}$
威爾士	...	...	...		2,165	+73	$+8\frac{1}{4}$
總計	99,329	+383	+0	1	119,890	−2,629	$-5\frac{1}{4}$

* 此數即在一八九一年，各恤貧聯合區（Poor-Law-Union）之男性農村勞動者，田莊常工、牧人、馬夫之總數。

（第十七表戊）

對總括表之批評 上列之總括表，其價值若何，並不顯然。用之可以查知，在一八九六年，增加工資者有五萬八千五百七十八人，減少工資者有四萬零七百五十一人，固不無些須價值，然受工資變動者之總數，並無從表現之必要。表中，恆列有受影響者之總數；設有一人在某一月中增加工資一先令，而在下月又被減去，則此人卽作爲二人計算，惟其對於下一欄（工資變動之純粹效果）之影響則爲零。此一『減四十五鎊』所含意義，可謂爲每人各減一先令者二千人，每人各增三又四分之三便士者三千六百六十二人（同屬一人或不同屬一人），亦可謂爲其他任何之數字，凡足以拼成同一總數者，均有成立之可能。第二欄之每人之變動量，僅爲一算術商數，並無可以形諸文字之具體意義，絕無重要價值可言。但如另以別一商數 $\frac{£43}{n}$ —— n 爲北部各郡農村勞動者之人數——取而代之，則吾人可以查見對於平均工資之效果矣。故在實際上，欲求表列用途較大，應從下式：——

區名	增		減		變動淨數	傭工總人數	變動平均量
	有關之人數	增加總額	有關之人數	減少總額			

（第十七表己）

如用前表所列之數字，乃又蹈原以錯誤數目作成之計算而又加詳細推算之通常覆轍，玆者，僱工總人數，卽爲本已有誤之一端，故如據以演算，必致通盤謬誤也。但在未用平均數前，僱工總人數雖有差誤，尙不致發生有害之後果。如上所述，此處所舉之平均數，可以其他有用之數相代，而此另一數如在限定範圍之內尙可不致錯誤，且足供大多數用途之需也。

抑有進者，有關人數一欄，旣有謬誤，則人數當以千人爲準，不能以一單位論，故與其謂『一八九一年五千六百六十二人屬於一類，而與一八九六年之人數，有鬆弛之連帶關係』，無寧僅言『有關人數五千至六千』之爲有用而正確也。

自英國農村施行最低工資制以來，全部問題已大經修改而單簡多多矣。雖然，上文之研究，仍用表列方法，以適應難辨且不全之資料，兼以敍述約二十年來，英國一般工資變動之記錄，官方辦理之情形也。

以上表列之第一類及第二類均已論及，至第三類非爲數字之答案製表方法，此時尙難討論，須俟將平均數之用途及性質加以研究之後，方能着手也。

（註一）　關於此法，參閱第五章第五節中位數。

（註二）　見『小紡織工業之工資』。

（註三）　詳見一九〇六年工資調查報告。

（註四）　節錄於第二次『工資變動年報』。

（註五）　節錄第三次『工資變動年報』。

（註六）　節錄第五次『工資變動年報』。

（註七）　關於此點，請參閱發克斯(Wilson Fox) 氏『農村勞動者之工資報告』一九〇〇年出版。

（註八）　見英國第四次『工資變動年報』。

第五章 平均數

本書爲專研究統計學之書，對於平均數 (average)，自當予以充分之地位。蓋惟用平均數，複雜之羣類(group) 及極大之數目，用一二有效文字或數字即可表現之；且惟有平均數也，統計學之二種定義——平均數學(Science of Averages) 及大數學(Science of Large Numbers)——乃可符合一致矣。

平均數與均數 有若干著作家企圖劃淸平均數 (average) 與均數(mean)之限界，惟二名詞分途應用之確切意義，迄未得有共同之結論（註一）。依吾人之意，最善之區別，莫過於將平均數定爲純爲算術上之概念，例如言變動人口中之平均壽長 (average length of life) 並非指某一特定羣類，不過爲一算術得數之簡便表明方法而已；至於均數，則於解釋物體數量 (objective quantity)時用之，例如言英國人之身長均數，乃表示一種數量，一切關於身長之量數，均將以之爲確定之分類標準也。如援用此種術語，則下文所論之第一節第二節第三節，大半均屬平均數一類，而第四第五第六等節則爲均數也。

第一節 算術平均數

平均數一詞通俗之用法，本可勿須多論，惟通常用語中，亦有統計學中常用者，茲擇要討論之。第一、平均數有僅爲避免極大數目而用之者。例如，一校之學生平均體重，據云爲一百七十五磅，所以不言十人之體重，共爲一千七百五十磅者，以前者較爲習用，且易使人聯想及於一般之人體重量也。同理，設吾人欲比較十年一期兩期中數種出口商品之價值，必須言在一八七〇至一八七九年一期中，每年平均數爲一千萬金鎊，一八八〇至一八八九一期中。每年平均數爲一千一百萬金鎊，而不云總額爲一萬萬金鎊對一萬一千萬金鎊也。

公分母 由此所論，乃引起普通之第二種用法。設吾人欲比較一八七〇至一八七九之十年，與一八八〇至一八九〇之十一年。並設前者一期總數爲一萬萬金鎊，後者一期總數爲一萬三千二百萬金鎊，在此情形之下，苟非用年數除之，化爲公分母，並求出二期之平均數，一爲一千一百萬金鎊，一爲一千二百萬金鎊，則二期之差別，必不得而知。此種之平均數，在板球戲 (cricket) 中應用最廣，盤數 (run) 及輪值數 (wicket) 之總數，雖有時亦有記錄，惟此乃統計上之奇玩，(curiosity) 對於球員之技巧及運氣，並無作用也。至各季舉行比賽時，用以評判優劣者，爲進門數(inning)除盤數之商數，盤數除輪值數之商數，以及其他等等，如是則所有數量均化爲公分母矣。如此含義之平均數，在機械學

上應用之處極多。例如每方吋之平均壓力，一引擎每分鐘所作平均工作量，一列火車之平均速度，均爲常見之數量。至所謂平均利率(average rate of interest)之用法，與此則完全相同也。

平均數與百分率　百分數者卽平均數之一特別用法也。比較人口或貿易之發展時，僅列其全數，毫無用處可言。以倫敦之人口而有五萬人之增加，實不如以一蕞爾哈羅(Harrow)小村，增加一千爲有意義；欲求其命意令人能以領會，則非以增加之百分數說明，謂一則增加百分之一，一則增加百分之六十不可，如此說明意卽表示每百人中平均增加若干也。基於此種理由，出生數，死亡數，及結婚數，均以百分率或千分率表示之，意卽每千人中各有若干也；且在此種情形之下，乃有雙重平均計算之情形，如言每年每千人中之若干是也。

將此用法，擴而充之，又有人口中每人各有若干之說，是則將數量變爲每人之比率也。此乃完全爲比較之用，其所依據之原理，與公分母相同。如論英國飲酒每年消耗金額，在一八六〇年爲一萬萬金鎊，在一八九〇年爲一萬一千萬金鎊，出言何等乏味；今言一八六〇年每人耗金三鎊半，至一八九〇年時，則每人耗金二鎊又十五先令，於是可以作比較觀矣。但用此類平均數，不可失之於濫，在預備總結比較數字時，是否有採用此種平均之必要，必須有充分之考慮也。

定義 由此觀之，平均數乃純屬算術性質，其定義如何，可由下式表示之：——

平均數×個數＝所關數量總數，

例如，平均體重×人數＝全體人數之總體重

定義之不能適用 雖然，此定義未必卽可完全適用，試解下一問題，可知其詳。一八九二年，英國西南部威爾次、道塞特、戴溫、康恩華爾、及梭梅塞特五地方之平均每週工資，分別爲十先令，十先令，十三先令六便士，十四先令，十一先令。然則英國西南部之平均數如何？

最簡單之答覆，平均數爲

$$\frac{10\text{先令}+10\text{先令}+13\text{先令}6\text{便士}+14\text{先令}+11\text{先令}}{5}$$

$$=\frac{58\text{先令}6\text{便士}}{5}=11\text{先令}8.4\text{便士}。$$

此一答數，已足爲許多用途之需，惟與上列定義，則不能適合耳。何則？蓋『十一先令八.四便士被何數乘？將得何數？』之問題，只有以『十一先令八.四便士用項數乘之，卽等於各項數目之和』一語答復之一途也。

請更進而研究『一部之平均工資』，及『五郡平均工資』之意義。

假設威爾次之平均工資，係由各鄉村之報告編製而成，今各村之工資，爲十二先令，十一先令，九先令，九先令六便士，九先

分，九先令，如將各數相加，然後以村數除之，固可得平均工資矣。無如仍不能適合上列之定義何？且也每村之平均工資爲何？使其能滿足定義上之條件，則所謂各村平均工資者，必爲各該村所付工資之總數，而被工人數除之也。惟事實上，舉行調查時，此一總數，根本無從覓得，有之，乃由觀察或藉猜想而得，非由計算而來也。

每村之平均工資之正確與否，姑置不論，即使其爲正確無誤，且各村報告之人數，亦已查明，則一郡之平均工資，當如下式所列：

$$\frac{12\text{先令}\times200+11\text{先令}\times150+9\text{先令}\times300+9.5\text{先令}\times150}{200+150+300+}$$

$$\frac{+10.5\text{先令}\times400+9\text{先令}\times200+9\text{先令}\times200}{150+400+200+200}=9\text{先令}11.8$$

便士

式中，分母各數爲各村之勞動者人數。如此演算，與假設已知全郡所有工人之工資，一一相加，然後以全人數除之，所得之結果完全相同，且所求得之平均數，與上文之定義，亦能適合也。

此中算術工作，可以化簡，理至顯然，蓋如將人數各以50除之，得數仍不變；猶之各村勞動者，分別有四，三，六……人不等，不必言其有200，150，……也。斯時，只取各村人數對總數之比例即可，不用實在人數，而得數可相同。此一計劃之優點有二：第一，

吾人欲查知勞動者之人數雖不可能，惟欲求其近似之比例數則未爲不可，直言之，人口普查報告關於農業之標目下，可供給之也；第二，查數字原亦無須絕對之準確；例如上文所提之各村工人數，無妨視爲替代213,145,320……之整略數(round numbers)；諸數雖小有相差，然與平均數並無妨礙，理至淺顯，不待言也。總而言之，將各村之人數，權作爲整略數；然後爲求數字計算上簡便起見，各以與此整略數成比例之簡單數字代之。

依此方法得出之平均數，與所舉之定義，未能絕對適合，惟與能滿足條件者之相差，實爲無幾。由此觀之，根據同一原理，英國西南各郡之平均數，不難求得矣。

分組之資料 有一極普通之情形，資料分成若干組(grade)，每組各有若干次數（instances），如下列之表，第一，第三兩欄即是。在此情形之下，欲求其算術平均數，必須對於各組內次數之分配情形，設立某種之假定。在平常情形下——尤以兩極端(extremity)之數漸趨減少之時爲甚——假設各組之數目，均集中於各該組之中心，以便計算平均數，頗能求得極高之確度(degree of accuracy)；就實際言之，各組之平均數，與其謂之接近各該組之中點，毋寧謂之更接近於全類之中心，惟全類中心上下兩邊，各有相當之差誤(error)，恰可互相對銷。爲求工作手續簡便起見，可以組距(breadth of the grade)——下表之組距爲五

年——作單位,然後擇項數最繁之一組(40-45年一組,中心點爲$42\frac{1}{2}$,見下表)中心爲原點以測量之。由此原點(以總件數除第四欄之總數卽得)之平均距離,卽以單位表示平均數距原點之距離,由此單位,平均數可立卽用原來單位算出也。

一九一一年英格蘭威爾士已婚男子之年齡

年齡分組	以$42\frac{1}{2}$歲爲原點起算之各組中心點,單位爲5歲	一千人中之已婚男子數	(2)與(3)之乘積	累計數	
				年齡組限	第五欄入數累計
(1)	(2)	(3)	(4)	(5)	(6)
15－20	－5	0	0	15	1,000
20－25	－4	33	－132	20	1,000
25－30	－3	112	－336	25	967
30－35	－2	152	－304	30	855
35－40	－1	154	－154	35	703
40－45	0	136	0	40	549
45－50	1	118	118	45	413
50－55	2	96	192	50	295
55－60	3	74	222	55	199
60－65	4	54	216	60	125
65－70	5	37	185	65	71
70－75	6	21	126	70	34
75－80	7	9	63	75	13
80－85	8	3	24	80	4
85－90	9	1	9	85	1
90－95	10	0	0	90	0
		1,000	+1,155		
			－926		
			229		

(第十八表)

$$平均數:42\frac{1}{2}+\frac{229}{1000}\times 5=43.645歲。$$

第二節 加權平均數

此項討論,已例示一極爲重要之統計方法,此卽所謂『加權平均』(weighting the average) 者是。茲請更進一步,仍用例釋方法,藉用原來數字,討論英國西南各郡之平均數,應加以何等之權數。旣知各郡之農村勞動者人數,乃可求出其平均數如下:

$$\frac{10\text{先令}\times 20,000+10\text{先令}\times 30,000+\cdots}{20,000+30,000+\cdots}$$;然吾人旣無從得知勞動者之確數,只有根據某種觀點,依照各郡地位之輕重,妥爲規定權數 (weight),如 20,000,30,000……是。至代表此種數量之數,作爲小麥收穫量可,作爲面積亦可,卽作爲人口增加率亦無不可。在此特別情形之下,方法或有謬誤,但其他問題所用,權數之意義,尙無如此明顯。例如,設吾人討論倫敦對四鄉居民之引誘力一問題,並設已知自愛塞克斯,腦富克,沙富克,移居倫敦者有若干,自斯他福特及華塞斯特移來者有若干,於是卽根據此種數字,比較對於農業區與對於工業區,以查引誘力究對何處爲大。然則吾人將用何者爲權數;以有關各郡之居民總數耶,用各該郡距倫敦之距離耶,抑用另由此種數量推出之其他數字耶?

機械學上之例釋 欲求明瞭,宜自研究『權衡』(weight) 字義始。設有一無重 (weightless) 之直桿,桿上分爲一百等分,

自一端起依次分在第四十，第五十，第六十，第七十，及第八十等分上，懸一等重之物；則此桿將在相當於不加權平均數（unweighted average）之點上，現平衡之象，此點即爲第六十等距(interval)。復次，設不用等重之物，而分用七，一，三，四磅之重量，則此桿呈現平衡之點，乃爲加權平均數(weighted average)即第五七·一等距之處也。該項物體移置愈遠，或位置不變，而物體重量變多，則重心（center of gravity）亦必隨之推遠；此即統計問題上各級工資應行加權之理由，乃昭然若揭，不待贅述者也。至靜力學（Statics）上所用之公式，$\bar{x}=\frac{Smx}{Sm}$，與上文所述之算術問題相同，統計學上亦多應用之處焉。

加權之效果並非甚巨　無論上文所論之平均數，抑或其他各種平均數，其所用之加權，在已往之統計論著中，已有極多之討論，實則其所佔篇幅與其重要極不相稱，蓋權數之選擇，並無若何之重要性。統計原理上有一極簡便之事實，即：苟合於某種條件，則不論所採權數爲何，所得結果，必能大致相同。故此假設之數理的研究，可姑置不論，惟對數項代數公式及演算步驟，則頗可論列於下也。

設以$W_1, W_2 \cdots\cdots W_n$爲$M_1, M_2 \cdots\cdots M_n$共有$n$個數量之權數。

則加權平均數爲 $M_w=\frac{W_1M_1+W_2M_2+\cdots}{W_1+W_2+\cdots}$

以$\bar{m}$爲M各數量之平均數，並以 $M_1=\bar{m}+m_1 M_2=\bar{m}+m_2$,

………。

則 $n\bar{m}=M_1+M_2+\cdots\cdots=n\bar{m}+m_1+m_2+\cdots\cdots$，故$m_1+m_2+\cdots\cdots=0$

同理，如以$\bar{w}$爲W各數量之平均數，又$W_1=\bar{w}+w_1$，……餘類推，$n\bar{w}=W_1+W_2+\cdots\cdots$且$w_1+w_2+\cdots\cdots=0$

於是 $(W_1+W_2+\cdots\cdots)M_w=(\bar{w}+w_1)(\bar{m}+m_1)+(\bar{w}+w_2)(\bar{m}+m_2)+\cdots\cdots$,

$\therefore n\bar{w}\cdot M_w=n\bar{w}\bar{m}+\bar{m}(w_1+w_2+\cdots\cdots)+\bar{w}(m_1+m_2+\cdots\cdots)+w_1m_1+w_2m_2+\cdots\cdots$,

$$\therefore M_w-\bar{m}=\frac{1}{n}\left\{\frac{w_1}{\bar{w}}m_1+\frac{w_2}{\bar{w}}m_2+\cdots\cdots\right\}。$$

故加權平均數（M_w），與不加權平均數（$\bar{m}$）之差額，乃視諸如$\frac{w_1}{\bar{w}}\cdot m_1$等各項之平均數而定。$w$各數量之和爲零，$m$各數量之和亦爲零，$w$各數之中若干爲正數，若干爲負數，$m$各數亦然。$m$及$w$兩種數量，同號較異號出現多時，則$M_w$與$\bar{m}$之差額必甚大。

次頁所列關於工資調查之表格中，$\bar{m}$爲二四先令二便士，M_w＝二四先令七便士，$n=38$，$\bar{w}=96^{00}$，玆將權數取至一百爲準，工資至便士爲準，則求得各數值如下，各業仍依次頁表格之次序排列之：

w.	m.	wm.	w.	m.	wm.
+226	+13	+2,938	+431	+ 41	+17,671
+26	−12	− 312	+147	− 41	− 6,027
−26	−10	+ 260	+184	+ 36	+ 6,624
−28	−53	+1,484	− 44	+ 7	− 308
−68	−58	+3,944	− 34	+ 4	− 136
−84	− 8	+ 672	+321	+ 19	+ 6,099
−74	−23	+1,702	+ 11	+ 61	+ 671
−83	+29	−2,407	+ 19	+111	+ 2,109
−85	+ 3	− 255	− 75	+ 1	− 75
−90	+37	−3,330	− 78	+ 65	− 5,070
−69	−48	+3,312	− 91	+ 50	− 4,550
−93	−36	+3,348	− 93	+ 75	− 6,975
+578	−15	−8,670	− 79	+ 28	− 2,212
−46	−90	+4,140	− 67	+ 1	− 67
−66	−10	− 660	− 12	+ 1	− 12
−27	−25	+ 675	− 78	− 46	+ 3,588
−73	−27	+1,971	− 64	− 16	+ 1,024
−56	− 4	+ 224	− 85	− 14	+ 1,190
−91	−66	+6,006	− 74	+ 12	− 888

二十一個正號乘積之和：+69,652。

十七個負號乘積之和：−41,954。

三十八個乘積之和：$17,698=w_1m_1+\cdots\cdots$

$$M_w=\overline{m}+\frac{1}{n\overline{w}}\times 17,698$$

$$=24\text{先令}2\text{便士}+\frac{17,698}{38\times 96}\text{便士}$$

$$=24\text{先令}6.8\text{便士}$$

$$=24\text{先令}7\text{便士}$$（依上表之例，工資取至便士爲準）

工資調查上之例證　次頁所列之表格，可用爲解釋上述原

加權方法不同而平均數變動甚小之例證

根據一八八六年工資調查之資料				人口普查所示各業僱用之人數	武斷規定之權數	均等權數
業別	平均工資（男工）先令	便士	調查得來之人數			
紡織業	25	3	32,189	142	144	1
織呢業	23	2	12,248	54	172	1
毛線業	23	4	7,005	38	219	1
製葛業	19	9	6,807	22	96	1
黃蔴製品業	19	4	2,799	9	23	1
大蔴製品業	23	6	1,232	3	78	1
絲織業	22	3	2,248	10	189	1
毛毯業	26	7	1,292	0	213	1
織襪業	24	5	1,070	8	287	1
花邊業	27	3	593	8	51	1
零星用品業	20	2	2,734	0	225	1
梳毛業	21	2	330	2	200	1
煤鐵礦	22	11	67,429	57	142	1
金屬礦	16	6	5,046	0	190	1
泥板石礦及石蠟油業	25	0	3,021	0	207	1
石板礦	22	1	6,933		232	1
花岡石礦及製造業	21	11	2,315	12	206	1
石廠業	23	10	3,956		34	1
瓷器業	18	8	499	0	39	1
警察	27	7	52,682	58	224	1
築路及溝渠	20	9	24,276	0	29	1
煤氣事業	27	2	27,965	0	40	1
自來水業	24	9	5,187	0	151	1
銑鐵（冶鐵廠）	24	6	6,234	0	128	1
鋼鐵鑄造及機器工程業	25	9	41,658	200	173	1
鋼鐵造船業	29	3	10,661	80	228	1
錫器製造業	33	5	11,514	0	178	1
木板業	24	3	2,088	0	174	1
銅器五金業	29	7	1,838	0	222	1
木質造船業	28	4	454	0	79	1
木桶製造業	30	5	327	0	165	1
車輛製造業	26	6	1,664	0	28	1
靴鞋製造業	24	3	2,902	0	142	1
釀酒業	24	3	8,366	0	46	1
蒸餾業	20	4	1,795	0	129	1
磚瓦製造業	22	10	3,188	0	55	1
化學肥料業	23	0	1,054	0	210	1
火車車輛製造業	25	2	2,239	0	233	1
平均數	先令 便士		先令 便士	先令 便士	先令 便士	先令 便士
	...		24 7	25 3	24 5¼	24 2

（第十九表）

理之例證（註二），頗有詳細研究之價值。着手舉行工資調查之時。先須散發調查表格，凡位址適中之大廠家，均向之作關於工資詳細情節之調查。發出之表格，未必全能收回，故最後作成之報告上，所列之廠數，並非全國各業實有之全數，反之，不過僅爲有表格寄回之數家而已。是以所得之平均數，並非依據前述之平均數定義，而得出之全國各業之算術平均數，反之，乃爲有表格填畢寄回之各業平均工資以各該業人數爲權數之平均數而已；然則假如各業填報表格如數寄回，玆以不完全之資料所算出之平均數，必與所有各業之實在工資平均數，多少有所不同。然就以往之經驗論之，二者相差，並非大相懸殊。下列之表，採用數種加權方法；第一欄爲用調查表格上所列之數作爲權數，得平均數爲二十四先令七便士；第二欄爲根據人口普查用超過某種最低限度以上所求得之各業之數爲權數，其平均數爲二十五先令三便士；第三欄之權數，乃用武斷方法規定者，數字之來源，與工資並無關聯可言，其平均數爲二十四先令五又四分之一便士；最末一欄，爲不加權之平均數，亦卽權數相同之平均數，其平均數得二十四先令二便士。凡此種種明示吾人，雖原來數字，低自十六先令六便士，高至三十先令五便士，變動甚大，而平均數則結果甚近，眞實權數如何，此處雖未指明，惟因此故，吾人置之不問可也。

各種加權制度下平均數始終一致　下列一表，所討論之問

題，在以英格蘭及威爾士於一八六九年九月二十九日密凱爾(Michael)聖誕節，及一八七〇年婦女節所舉行之調查，求出農村平均每週工資，此二年之數字，分列於第一二欄者即是。此項平均數得出之方法甚多，茲擇要述之於下：先求出各郡之夏秋兩季之平均數，列於第三欄，然後計算四十五個數之平均數，而得十二先令七便士。次須假設付出夏季工資之時間，當秋季工資施行日期之二倍，列於第四欄。然後依上項程序求其平均數，得知爲十二先令五又二分之一便士；二數有少許之差額，乃因密凱爾聖誕節日之調查，已包括秋收犒賞在內，故全數較夏日工資爲高也。再次，將各郡依地理順序分爲數組，求各組之簡單平均數(Simple average)——第三欄之數字標註a字，第四欄標b字——然後用第五欄標有c字之數，即各組之農村勞動者人數，將其加權；乃知a列數字用c列爲權數時平均數爲十二先令五便士，b列數字用c列爲權數時平均數爲十二先令四先令。復次，用第五欄爲權數，乘第四欄之各郡工資數。乃爲最淺近之方法，得平均數爲十二先令四便士。更次求各區平均數（a與b）之簡單平均數，換言之，即將此八區各加以均等之權數，則得平均數a爲十二先令四又四分之三便士，b爲十二先令三又四分之一便士。否則以約克協爾及威爾士各作一郡論，計算第三欄之簡單平均數，即十二先令八便士是也。

一八六九至一八七〇年之農村工資——各種加權方法及其結果之例證

	1. 一八六九年密凱爾聖誕節		2. 一八七〇年婦女節		3. 第一，二兩欄之平均		4. 以第二欄作第一欄之兩倍，然後平均之		5. 各郡農村勞動者人數以千人為單位	6. 各組地方之人口全數以千萬人為單位
	先令	便士	先令	便士	先令	便士	先令	便士		
蘇塞克斯	12	3	12	0	12	$1\frac{1}{2}$	12	1	34	...
四瑞	14	0	13	6	13	9	13	8	16	...
肯特	14	6	14	0	14	3	14	2	44	...
韓次	11	0	10	6	10	9	10	8	32	...
布爾克斯	12	0	10	0	11	0	10	8	22	...
平均數	...		...		*a* 12	$4\frac{1}{2}$	*b* 12	3	*c* 148	*d* 22
何爾次	14	7	11	10	13	$2\frac{1}{2}$	12	9	20	...
腦贊次	12	6	11	6	12	0	11	10	23	...
聿次	16	0	11	0	13	6	12	8	9	...
拜得福特	13	0	12	0	12	6	12	4	17	...
劍橋	11	0	12	0	11	6	11	8	24	...
平均數	...		...		*a* 12	6	*b* 12	3	*c* 93	*d* 14
愛塞克斯	12	6	11	0	11	9	11	6	45	...
蘇富克	10	6	11	0	10	9	10	10	41	...
腦富克	11	6	11	6	11	6	11	6	44	...
平均數	...		...		*a* 11	4	*b* 11	3	*c* 130	*d* 12
威爾次	11	0	10	3	10	$7\frac{1}{2}$	10	6	26	...
道塞特	9	6	10	3	9	$10\frac{1}{2}$	10	0	17	...
得旺	10	0	10	3	10	$1\frac{1}{2}$	10	2	34	...
考恩華爾	11	0	11	0	11	0	11	0	17	...
桑摩塞特	11	0	10	6	10	0	10	8	31	...
平均數	...		...		*a* 10	6	*b* 10	6	*c* 125	*d* 19
斯他福特	13	0	13	0	13	0	13	0	19	...
哥老色斯特	11	9	10	9	11	3	11	1	22	...
西爾福特	10	3	10	0	10	$1\frac{1}{2}$	11	1	12	...
沙拉堡	11	0	11	6	11	3	11	4	21	...
華而色斯特	13	6	11	0	12	3	11	6	15	...
華維克	13	6	12	0	12	9	12	8	20	...
平均數	...		...		*a* 11	9	*b* 11	7	*c* 109	*d* 27

雷色斯特	14	0	13	0	13	6	13	4	15	…
如特闌	12	6	12	0	12	3	12	2	3	…
林肯	14	0	13	6	13	9	13	8	49	…
腦次	13	6	13	0	13	3	13	2	16	…
得爾拜	13	6	14	0	13	9	13	10	8	…
平均數	…		…		*a* 13	$3\frac{1}{2}$	*b* 13	3	*c* 91	*d* 14
茄協爾	13	6	13	6	13	6	13	6	18	…
闌斯	15	0	15	0	15	0	15	0	30	…
約克北區	19	0	15	3	17	$1\frac{1}{2}$	16	6	30	…
約克西區	17	4	13	6	15	5	14	$9\frac{1}{2}$	16	…
得由漢姆	16	6	16	0	16	3	16	2	8	…
腦贊倍闌	19	6	16	6	18	0	17	6	12	…
可倍闌	15	0	15	0	15	0	15	0	10	…
外斯他摩闌	16	3	15	6	15	$10\frac{1}{2}$	15	9	3	…
平均數	…		…		*a* 15	9	*b* 15	6	*c* 127	*d* 72
邙毛斯	12	6	13	9	13	$1\frac{1}{2}$	13	4	6	…
哥拉摩根	14	6	14	6	14	6	14	6	5	…
可兒馬贊	12	4	11	6	11	11	11	$9\frac{1}{2}$	4	…
排姆布老克	11	0	10	0	10	6	10	4	4	…
加地幹	9	0	8	6	8	9	8	8	5	…
布來克腦克	12	0	12	0	12	0	12	0	4	…
來得腦爾	10	0	10	0	10	0	10	0	2	…
加那文	12	0	12	0	12	0	12	0	5	…
平均數	…		…		*a* 11	7	*b* 11	7	*c* 35	*d* 14

（第二十表）

除此之外，如欲另組成新羣類，可不用農村勞動者之人數，而用各區之總人口——卽標有d字之數——作權數。但倫敦人口數字龐大且與農村無關，應除去之。惟討論至此，乃又有一新因素發生，緣工業區之人口最大，而內中之農村勞動者，雖不佔重要之地位，但工資則甚高；有此高度之工資，以最多之人口加權，於是以d數爲權數之b數平均數，乃達十三先令一又四分之三便士之多，可見此項加權，頗有不當之處。次將第四欄之數改正近至

一先令爲準，而將第五欄之數，以一萬人爲單位，則加權平均數，必爲十二先令五便士。復次，假如將第三欄之數，以任意抽來與本題無關之數作權數，質言之，卽用對數表中，自二起至四十六止，第三位小數加權，則加權平均數，乃得十二先令十又四分之三便士。讀者亦可試用任何權數，勿論合理與背理，除非選擇權數時先有偏見，或對於少數之郡區，特別加重，否則所得之平均數，並無甚大之影響也。

然則眞實權數，凡得出之平均數，能適合前述之定義者，必與前此所採用者，相去無幾，且用其所得之平均數，卽使與十二先令四便士相差，亦不致比現時所用者爲更遠，故眞實平均數(true average)，與此數相差，不致超出三便士之外，必無可疑。原始各項之全距，旣低自八先令六便士起，高至十九先令止，雖用極端離奇之方法，平均數亦必不致出乎十二先令至十三先令一又四分之三便士限界之外。

權數尙難取消，惟計算權數時之差誤，則可不計　上文討論雖感過繁，然若無如此之論辯，因加權乘錯而生之差誤，大小若何，必無從注意及之。雖然，關於加權問題，基於以上之觀察，未及考查羣類之性質，但經用各種加權方法之後，算得之平均數，視之並無變動，因而卽謂權數可以取消，竟採用不加權之平均數，斯亦未免失之過當。蓋此種情形，僅於數量之巨細，與權數之眞

正大小度(magnitude),並無關聯時,始有之。例如研究城市之工資,將所有各城市合併計算,求出其平均數,如吾人取消權數,各城市以均等視之,則結果之影響,確爲無幾,其故無他,乃因較高之工資,乃出之於大城市也。試觀第六章第二節,工程聯合會一百一十七處分會之認可工資平均數,如將各分會一律看待,則在一八九一年爲三十二先令四便士;但如以各該分會之會員人數爲權數,將各分會之工資施行加權,則同年之工資平均數,乃高達三十三先令四便士矣。雖然,在此種情形之下,權數不得完全取消固矣,惟在內中少數地方,並無特著關係時,則仍不妨爲粗率之計算。以倫敦一城而論,其他之工資,除達蒂福特及恩費爾德滂克二處外,較他區均高,且該會會員幾有全數六分之一,均受此種影響。茲按倫敦所佔之恰當重要性計算之,以各分會會員人數(百人爲單位即取其最近於一百之數爲準)爲權數,則得平均數爲三十三先令六便士,事實上與前所求得之數,乃甚相同也。總而言之,每一羣類,苟欲計算其平均數,須按其本身之情狀,爲決定之標準;在多數情形之下,權數竟可不用;至當羣類所包括之項數甚多時,即使施行加權,其因加權而起之差誤,雖大至相當程度,在一切之情形下,幾全可置諸不問。若欲查視此種差誤之重要,則須檢閱資料也。

此一原則,至爲重要。在若干情形下,眞實權數,乃竟無從計

算，甚且不能加以解釋；然據吾人所知，在某種條件之下，並無計算或釋明之必要；又有時雖有權數，而不能得其確數，然亦無須求其確切也。雖然，如各項原有共同之偏性 (bias)，無論用何種方法，均非加權所能免除。例如工資，實在工資數全部均較調查得來者少一先令，則平均數雖然得出，亦必較實在情形高出一先令。由是吾人可得要訣如下：計算平均數之時，應儘量防範各項之偏性，惟對於加權程序上，則無須斤斤較量確度之何若也。

第三節　統計係數

統計係數 (statistical coefficient) 者，乃一整數或一分數，以之乘一總數(例如人口數)，可得另一有關之數(如嬰兒出生數)者也。如言出生率爲千分之二十八，則統計係數乃爲 .028。統計係數，在普通統計學領域中，應用之處極多，較之差誤律 (Law of error) 在人口學(demography) 上之應用，尤佔極重要之地位。人口有增有減，而關於某某數種數目之係數，變動終屬甚小，必也經過長久之期間，且無非常之因素，足以發生特別之影響時，統計係數乃呈現顯著之變動。欲比較若干國家之統計數字，尤非用統計係數不爲功，又爲推測將來之數字，用此係數，有時亦頗能具有驚人之確實性(accuracy)，然在有特殊之意外事變時，固在例外也。統計係數施用之處甚繁，出生率（各區域中）也，死

亡率（依年齡,職業或死因分類）也，結婚率（分各年齡）也，其他若自殺案件也,犯罪案件也,意外災害也，各種商品之消費也,莫不用之;推而廣之,如能覓得原始資料,則每年經過西敏大橋（West Minster Bridge）者人數若干,紀念塔前臨弔先賢者年有若干人,遺忘車上之傘,以及其他等等,無不可以應用之也。凡文明國家,大多均有重要係數之計算，並發表於統計報告中。至欲舉行統計調查者,此等之係數尤須有成竹在胸也。

雖然,所謂統計係數,無非某種算術平均數之另一說法，事實甚爲顯然。玆請以此爲參考,再將本章第一節所舉之平均數定義,重作概括之討論如下:——

平均數(A)×個數(N)＝所關數量總數(Q)，換言之

$$A=\frac{Q}{N},\quad Q=N\times A。$$

以出生數爲例,A爲統計係數,N卽人口,Q乃出生嬰兒數也。

Q一數在事實可能範圍之內，必受一因素變動之影響而隨之發生變動。例如人口數爲N,苟人口之性別及年齡分配有變化,Q亦必受其影響,至結婚數及結婚年齡,以及多產（fecundity）等,自同爲重要之因素也。關於出生率,結婚率及死生率之分母,爲求便於嚴格比較起見,已訂有一致之方法（用修正因素 correcting factors）（註三）。在未能應用此種方法之前,可請遵用白狄雍（統計學初級教本第九十四頁以下）之準則:效果（Q）

必須與其產生之近因（N）比較；如就結婚數言，吾人可舉問『何人可以結婚?』回答必爲成年未婚者，或已失配偶者，然則凡此人等之總數，即爲N矣。將此準則推而廣之，包羅所有間接有關之人或物；例如論及產煤量，其直接之有關係人數爲掘煤工人，次爲所有煤礦之工人，論及家用煤炭之出產量，則與用煤家數有關（註四）。各個因素，只須保留一個，其餘概行屏除淨盡，分子分母中之各項，必須具有同一性質，同時分子所指人或物，對於分母中人物之潛在關係，必須協和一致。例如，全人口平均每人所攤之出口貨值一言，即與上舉種種條件不合；出口貨所包內容龐雜，而人口又有性別及年齡之不同，所表示者只對國外市場上本國生產力之一部而已。

惟其如此，粗略之係數及平均數，自亦有其相當之用途；如係數或平均數有變，必由於一或數因素業已有變動也，苟能知衆因素之中，除一因素外，其餘幾爲不變之數量，則係數必隨此常變之因素同變。例如人口數爲N，可以結婚之人數爲n，結婚數爲M，則粗率之係數，即爲$C=\frac{M}{N}=\frac{M}{n}\times\frac{n}{N}$；如$\frac{n}{N}$爲常數，則C將隨$\frac{M}{n}$而變動，$\frac{M}{n}$即爲較爲合乎論理之係數也。

第四節　衆數

現所討論之其他二種均數 (mean)，均爲統計學家常用者，

惟普通言談之間,尙少爲有意識之採用,殊爲可惜。然一般常用之習語中,頗有若干字眼(如其有確定之意義),與吾人所謂平均數之意義,頗相近似。

平均人 一般常有『平均職員』(average clerk),『平均工人』之習語,意何所指,解釋大有不同。有時,乃表示各該門類之典型。所謂中常職員者,意或指其所得,與一切職員之平均收入略同,對於必需品及奢侈品之開支,與彼同一階級之平均數相近,如彼之能力平平,年齡中常,則彼之工作任務,亦必極爲普通。須知此一職員,乃爲理想之人物,在任意選來之六人中。求能如此人者,未必得其一;蓋六人必各有其特點,且有數種個性,必與一般有所不同也;『平常之新聞記者』,不能謂之爲生存而生活,實亦爲一合乎理想之人物,必有某種特性存乎其人也。

哥得雷氏之『中常人』 哥得雷(Quetelet)氏所稱之『平均人』(average man)(註五),已爲一般所習知矣;所謂『平均人』者乃一具有平均身長,體重,體力,肺量,而有正常視距及中等色質之眼睛之人也。此一『平均人』,較之『報章上之平均數』,乃爲更爲滿意之模範,其他一般人均與之有不同之點;如將所有之障礙原因排除,出生者將全爲此標準之人物矣。然現有之人類,欲求其與此種種標準吻合,乃絕不可能之事也。

葛得雷氏所指者,旣非算術平均數,亦非中位數(median)

求衆數法

美國一八九三年參議院報告中之工人人數

工資	人數	以二角爲組距之分組		以三角爲組距之分組			以五角爲組距之分組	
自 $.25 至 .34	1	16		75			317	
〃 .35 〃 .44	15		74		159			
〃 .45 〃 .54	59	144				301		
〃 .55 〃 .64	85		242	355				725
〃 .65 〃 .74	157	270			439			
〃 .75 〃 .84	113		282			483	1,472	
〃 .85 〃 .94	169	370		674				
〃 .95 〃 1.04	201		505		1,190			
〃 1.05 〃 1.14	304	989				1,088		2,012
〃 1.15 〃 1.24	685		784	1,242				
〃 1.25 〃 1.34	99	557			1,023		1,297	
〃 1.35 〃 1.44	458		924			996		
〃 1.45 〃 1.54	466	538		740				
〃 1.55 〃 1.64	72		274		603			934
〃 1.65 〃 1.74	202	531				589		
〃 1.75 〃 1.84	329		387	660			970	
〃 1.85 〃 1.94	58	331			376			
〃 1.95 〃 2.04	273		418			583		
〃 2.05 〃 2.14	45	310		343				640
〃 2.15 〃 2.24	265		298		399			
〃 2.25 〃 2.34	33	134				330	506	
〃 2.35 〃 2.44	101		297	310				
〃 2.45 〃 2.54	196	209			372			
〃 2.55 〃 2.64	13		176			178		322
〃 2.65 〃 2.74	163	165		180				
〃 2.75 〃 2.84	2		17		146		198	
〃 2.85 〃 2.94	15	144				149		
〃 2.95 〃 3.04	129		134	181				
〃 3.05 〃 3.14	5	52			64			285
〃 3.15 〃 3.24	47		59			59		
〃 3.25 〃 3.34	12	12		233			254	
〃 3.35 〃 3.44	0		221		226			
〃 3.45 〃 3.54	221	226				242		
〃 3.55 〃 3.64	5		21	32				114
〃 3.65 〃 3.74	16	27			27			
〃 3.75 〃 3.84	11		11			93	96	
〃 3.85 〃 3.94	0	82		82				
〃 3.95 〃 4.04	82		82		85			
〃 4.05 〃 4.14	0	3				3		6
〃 4.15 〃 4.24	3		3	3				
〃 4.25 〃 4.34	0	0			3		4	
〃 4.35 〃 4.44	0		3			4		
〃 4.45 〃 4.54	3	4		4				
〃 4.55 〃 4.64	1		1		1			9
〃 4.65 〃 4.74	0	0				0		
〃 4.75 〃 4.84	0		0	8			8	
〃 4.85 〃 4.94	0	8			8			
〃 4.95 〃 5.04	8		8			8		
〃 5.05 〃 5.14	0	0		1				
〃 5.15 〃 5.24	0		1					

或衆數（mode）——乃爲另一之均數，用之將所有同樣數量依一定之規律彙分類別者；該定律爲何？乃謂宜遵從人體測量(anthropometrical)之數量，卽氏之精彩論文之所論也。

衆數 然則報章上之平均數似爲衆數之意矣。然衆數者密度最大之地位也，其意究何所指？玆請解釋於下；試閱上列之第二十一表，或翻閱前於第四章第四節所列之美國工資統計表（第十三表），可知在第二欄自上而下，數目漸增，至六百八十五（在1.15至1.24一組）之數後，數乃參差不齊漸至於消滅。此六八五在組距爲一角之各組中爲最大數。

一統計羣類（不論爲工資，或身長，或爲其他之可量之數）中，分組之數值，而有最多之次數者，謂之衆數，換言之，亦卽密度最高之地位，或佔最多數之數值。如就以連續曲線代表之羣類言之，則此一數值卽爲最高縱坐標(ordinate)下之橫坐標(abscissa)也。

求衆數法 依修正後之字義，上表第二欄最大數有十四個之多，各數之起伏無定，但在此十四枚衆數之中，一元一角五分至一元二角四分一組之數，最稱顯著。至分組較寬時，在第六欄之五角分組下，衆數只餘三枚，惟如略去五元處數僅爲八之小羣，則所餘僅二枚。具有一四七二者乃爲最大之羣，但此羣之位置，不能確實指定僅可謂之介乎七角五分與一元二角五分之間。

此外，求衆數之近似値，尙有一法，茲舉例釋之如下：一如第二十一表所列美國工資統計中，在一角分組之下，所列之數，何者爲衆數，不能立時決定。在二角分組之下，自.25－.44一組起，其頻數順序爲十六，一百四十四，二百七十，三百七十，九百八十九，五百五十七，五百三十八，五百三十一……等等，則九八九（在$1.05－$1.24一組）之爲衆數，事乃彰然自明；如在二角分組中，自.35-.54一組順數之，則有七四，二四二，二八二，五〇五，七八四，九二四，二七四……等，在1.35-1.54一組之九二四，乃爲衆數；用此複式表列法，可見二角分組不能決定衆數。在三角分組中，自$.55-$.84一組起，數爲三五五，六七四，一二四二（在$1.15-$1.44組內），及七四〇……等；自$.65-$.94一組起，數爲四三九，一一九〇（在$.95-$1.24一組中）及一〇二三……等；如自$.75-$1.04一組起，則有四八三，一〇八八（在$1.05-$1.34一組），及九九六……等數：凡此種種分組，衆數終在$1.15-1.24一組中；反之，此一小組必含有衆數，而衆數果亦即在——或臨近——一元二角也。此例所用數字，極不整齊，惟其如此，乃可特別舉示困難之所在。茲復將全部程序，擇要概述如下：

（1）將分組組距漸次擴展，重行排列數字，一而再，再而三，至數字排列齊整時爲止；

（2）然後再察視組距業已放寬之羣類，並查核在分組之低

限(lower limit)移動時,衆數是否亦隨之轉移;如其移動也,則組距尙須再行放寬;否則,衆數乃在最狹組,如將組距放大,則較寬之組凡包括該狹小之組者,亦必同時爲衆數之所在。至求衆數之圖示法,當在第七章第一節(見第七圖求衆數及中位數圖示法)敍及之。

衆數位置之無定　求衆數之程序,至爲繁難,卽使初步數字,原已呈現整齊狀態,欲確切求得衆數,亦非甚易之事。玆爲指示困難之情形計,特舉一例如下;假設業經測驗多人之身體,得其身長爲

67 吋	455
$67\frac{1}{4}$吋	475
$67\frac{1}{2}$吋	490
$67\frac{3}{4}$吋	500
68 吋	485
$68\frac{1}{4}$吋	467
$68\frac{1}{2}$吋	445

驟視之,衆數似確在六十七又四分之三吋;惟吾人須知,雖極精確之測量,如單位以至一吋之四分之一爲準時,身長六七又四分之三吋零八分之一吋,卽當作六十又四分之三吋論,必至測量更形精確時,乃能用六七又八分之七吋表示之。故就現在而論,

所謂六七又四分之三者，乃實指自六七又八分之五起至六七又八分之七吋也。設上列五〇〇人之身長，在該組中成均勻之分配，則以六七又四分之三爲衆數，已盡確實之能事；惟就數字情形觀之，衆數當尙在此數之下方也。設上列數字，實際測量時，分配情形如下：

自$67\frac{1}{4}$至$67\frac{3}{8}$吋	238	483（身長$67\frac{3}{8}$吋）
自$67\frac{3}{8}$至$67\frac{1}{2}$吋	245	
自$67\frac{1}{2}$至$67\frac{5}{8}$吋	245	495（身長$67\frac{5}{8}$吋）
自$67\frac{5}{8}$至$67\frac{3}{4}$吋	250	
自$67\frac{3}{4}$至$67\frac{7}{8}$吋	250	493（身長$67\frac{7}{8}$吋）
自$67\frac{7}{8}$至68 吋	243	
自68 至$68\frac{1}{8}$吋	242	

並設諸數已列如上行之式，則衆數必在六七又八分之五吋；而上行之排列，衆數乃爲六七又四分之三吋。此種移動之或然性，於原來分類中，已可見之，蓋在六七又二分之一吋者人數較在六八吋者爲多也。總之，就以上所論，可知衆數頗有欠於確定之弊，確定性之深淺，尙須視分組之寬狹，及各組距之確實位置而定。在各組項數甚多之時，如將分組縮小，卽可現露其有規則之形狀，於是求得之衆數，確度必較大也。

用數學方法，求算衆數（見第十章第二節第五款），以上列

身長數字言之，須依照含有衆數之組（$67\frac{5}{8}$至$67\frac{7}{8}$吋）之項數，對其上下各組項數相差之比例，將此含有衆數之組加以配分。例如相差之比例為：500–490；500–485＝10：15，衆數乃為（$67\frac{5}{8}+\frac{10}{10+15}\times\frac{1}{4}$）吋＝$67\frac{29}{40}$吋。用此方法，設如最大者有兩組，其項數完全相同，則衆數卽在二者之中間；又如一最多之組，臨近之上下兩方各一組，項數亦係相等，（假如分配為對稱形式），則衆數必在中間一組之中心。凡此兩者皆如演繹法 (a priori) 之所指示。

所謂『平均人』 討論至此，然則所謂『平均工人』者究何所指：意為每日賺工資一元七角三分（第四章第四節第十三表所列美國工人工資之簡單平均數）之工人乎，抑為賺衆數一元二角工資之工人乎？普通之語法，乃指後者而言。所謂『平均職員』一詞，其意並非指一身具有可以測量之性質恰為同類性質之算術平均數之人，而指具有性質而為同等人物大多數所同有之職員也。職員之階級中，喜讀消閒報紙，較讀荷馬 (Homer) 史詩者為夥焉，涉身游藝場者，較往聆聖樂 (oratorios) 者為多焉，月進百元者較月有五百元收入者為衆焉，住於郊外四里，較住於郊外一里或二十里者為多焉。解釋雖然如此，但『平均人』並非實有其人，蓋秉性各有不同也。平均數為純料抽象之名詞，吾人實際應用統計時，不可不知；前言之美國工人大衆，不能卽

視全賺一元二角之工資，即使賺相同之工資，亦不能認爲其他方面無以異；反之，即得同等之工資，而工作量數乃大有出入也。

衆數之重要性　爲完全表明一羣工人之經濟狀況，無單獨勝任之一數，有之，即衆數是也。吾人所意像之最大幸福之最大多數人可於衆數中求之。算術平均數與中位數(median)——定義見下——所表示者，並非實在，不過數字上之概念而已，而精確表白最常發見之數目者，惟有衆數一端。衆數爲最常得到之普通結果，且有極普通之用途。某段間之火車或公共汽車，求出意欲搭乘者之平均人數，不如得知其最常見人數之爲愈。成件出售之衣服店，與其量得多人胸寬尺寸之平均數，無寧得出衆數爲適宜。一郵局之成立，一商店之開設，應以求得購買匯票金額或茶葉價格之衆數，爲當務之急，其他平均數，需用甚少也。即在集藏之泉幣中，其中受人賞鑒者，每每表示整羣之風氣，勝於各個特色之算術平均數，且就此最後一項事例而論，其衆數頗爲確定焉。

衆數之優點　衆數之一特點，即在完全不受極端項之影響。收解款項，每因發現一千金鎊之一支票，而動搖算術平均數，但衆數乃不因之而受波及。少數之百萬富豪及極大多數貧無立錐之人民，二者收入之算術平均數，或許與一發展平均全爲家給戶足之國家中所得之數相同，然衆數則二者將迥有不同也。吾人論

及一類數字，如極多工人之工資者之歷年變動情形，設以算術平均數爲判別之標準，則工資增加情形，由於工資低者提至水平線耶？抑係由於已有甚高待遇者又急劇增薪耶？吾人竟無從答覆，然衆數則可表明大部人數地位之變動，因而可以答覆之也。布斯(Booth)氏之『倫敦(London)』一書，應用衆數之處甚繁。每一年齡圖，必有一分業別之年齡之衆數；每一工資表，必有各級工資之衆數，固不待言；卽彼對於第五類——現代城市之標準工人——之論述，亦係根據同一原則。至彼對於社會情況之測量，旣以住房間數及僕役人數爲基準，則用衆數（每戶住房四間，並無僕役），當較用其他平均數爲利便。

衆數之劣點　雖然，衆數受人反對之所在，乃因其不能施用於多種之羣類。設有一極不規則之數字羣類，並無一定之典型，例如英國各城市之人口，其衆數極不穩定，對於人口數之重要性，則不能給予吾人以重要之消息。蓋衆數之爲用，主要乃在其能指示一典型之數字，其他數字均與之分歧也。例如在上列第二十一表之工資數字，典型之工資爲一元二角，其他在兩方之工資，則爲技能或機會以或種原由，較正常程度爲高或較低之人所得。如葛得雷氏之著作所示，凡有一種典型之數字，必有衆數能表示之。衆數告吾人者，乃對於每一典型之一事實，故他種平均數，亦不可不恃爲補充之用也。

第五節　中位數

設有一羣人或物於此，其中各有可以測量之屬性，例如身長或工資，吾人爲謀論述全羣類之簡捷起見，頗可擇用某一數量扼要說明之。玆設此種屬性之數量，依數值之大小，由小而大，排爲漸升之序列；由此序列向上察其恰在中途一項之所在，當此中間項目之數量，卽謂之中位數 (median)（註六）。譬如有一羣賃銀勞動者，工資不及二十先令三便士爲二百人，恰爲二十先令三便士者一人，超過二十先令三便士以上有二百人，則二十先令三便士卽爲工資之中位數。蓋在此假設之例中，不及二十先令三便士者，與超過該數者，人數完全相等也。中位數如此，其他以此類推，在序列上四分之一及四分之三處之數量，謂之四分位數 (quartile)（註七）；在序列上十分之一，十分之二，乃至十分之九之數量，謂之十分位數 (decile)；在百分之一，二乃至九十九之數量，謂之百分位數(percentile)。

中位數之位置，較衆數爲穩定。欲求確切之計算，則在項數爲奇數時，中間之數卽爲中位數，在項數爲偶數時，中位數卽在兩中間數之中間，此兩中間數一般多連接甚近，如二數相等時，則二數均爲中位數矣。但若序列中數量不甚準確，而分爲小組時，則以本章末節圖示法（第一圖）對其確實數值，可作一良好之

推算。

中位數之優點，在絕對不受特異項目之牽動；勿論若干人數之百萬富翁之收入，對於全類序列之收入中位數，所給予之影響若何，絕不致較收入在中位數以上之其他相等人數為重。對於此種極端項，在許多目的下，固有較之近於平均數之項目，有特別加以重視之必要；無如算術平均數，對於特別項目，重視已超出正當範圍之外，蓋當此民治之時代，以一百萬之富翁，而與萬千之普通工人，等量齊觀，殊不若中位數較為適當也。除此而外，中位數尚有一易於計算之特長，故不需甚多之算術工作，以吾人對於距平均數甚遠之數，無須逐一檢算之勞，只將臨近平均數之項目，精密查核即可也。

無全部訪查之必要 中位數之應用，尚有頗為重要之優點；蓋對於諸項事物之訪問，若不確實並欠完全時，仍不難將中位數求出也。茲舉一二實例以明之。在舉行工資調查時，或許有工資距平均數甚遠之十萬工人，並未及調查，然則欲查問如此多數工人之工資，對於算術平均數之影響如何，必因缺乏關於彼等收入之消息之故，感受極大之困難；但如求中位數則不然，吾人所需知曉者，只為彼等之人數，不問彼等工資確數為幾何，亦能確定人數之最高額，仍能將中位數查出，不過有極微之出入耳。試取英國『一八八六年工資調查總報告』第四百七十頁之總數，三

十五萬六千人，加上工資在十五先令以下者十萬人，則中位數仍在二十先令至二十五先令一組，此原來之中位數也。惟在最低十分位數及下四分位數，變動則極顯著，然算術平均數，因此降低至少有二先令一便士之多。如就所得問題而論，此說亦有成立之理由；蓋消息既極貧乏，有極多之人，其確實所得數，無從得知，然在多數情形之下，不難斷言其所得乃在某一數額之上，或在某兩限度之間，而此如在低限之下方，卽吾人對於其所有人數所需探知而據以決定中位數之數也。

復次，溯觀十九世紀以來之工資發展史，欲求一正確之平均數而不可得；但吾人既知一大類工人，每週工資在十五先令之下，另有工資在三十先令以上而確實不得而知者極多人，並知工資在十五至二十先令，及自二十五至三十先令者，各有確數若干，則不難求其中位數；爲求中位數，如測知中位數乃在二十至二十五先令一組，應將該組人數加以確實調查；卽使手中缺乏完備之資料，吾人仍不妨發言，謂中位數必不出某兩狹小限度之外焉。

不可量之數量 (incommensurable quantities)　中位數之另一優點，亦卽其較重要之一用途，乃在其對於完全不可測量之數量之應用。此種應用，以賴高爾頓 (Galton) 氏（註八）倡導之力者爲多。玆爲舉例起見，設吾人欲求多數兒童之平均智力。智力並不可以測量，事乃至爲顯然，卽使用精妙之號碼，亦難求

得其算術平均數。然在另一方面，設將一羣二十名之兒童，依其智力之高下，排成序列，不必謂甲童之智力，比乙童高二十分，惟云第十或第十一名兒童，乃爲全組之典型人物，至低限度亦較其他各種之平均數爲佳也。

中位數之劣點 中位數既具有上述之優點，然其所以不能普遍採用之故何在？此無他，一組數量之中位數，或可全然與典型不符，且在事實上或可不近於實在觀察得來之事物中也。例如，設有兩類工資數字，每類各包括一千人之工資，一類之工資爲十五先令至二十五先令，一類在三十五先令至四十五先令，今吾人求得之中位數，竟在二十五至三十五先令之間，而該類中事實上並無一人。由是觀之，中位數之主要，用途只能限於主體之若干事物連接甚爲密切之情形下；至於少數之極端項，倘不致有何影響也。

如以m代表x_1, x_2……x_n共有n個數量之中位數，a代表同數量之算術平均數，以x_1-z,x_2-z……爲x各數量與任一數量z之離中差 (deviation)，則m即爲能使離中差量（全作正數計算）之和成爲最小之z值，而a即爲使離中差之平方之和成最小之數值。

欲明離中差（全作正數計算）之和成爲最小值一語，請比喻如下：設有 $2n+1$ 處地方，在一條直線上，各處各由某盡端數起之第n處地方之接線機，單獨各通一條電話線；電話線之長度，即相當於上言之離中差；今設接線機自第n處，移至 $\overline{n+1}$ 處（或中心點），則必有$\overline{n+1}$條線，因之以相等之距離而縮短，有n條線因之以相等之距離而加長，然則電線之總和，必爲之消減；如地方之數爲偶數，則自任一端數起，凡地位在，或介乎，第n處及第$n+1$處之間者，均可達到使成爲最小數之目的。

關於第二段。相差數之平方和爲最小數，只觀：$\Sigma x=na$, 及 $\Sigma(x-z)^2=\Sigma x^2-na^2+n(a-z)^2$，便知，蓋當$z=a$時，$\Sigma x^2-na^2+n(a-z)^2$，即成爲最小也。

茲有七十二項數量，特用各種平均數論述之如下：

十三歲至十五歲兒童之體格測量

號次	年齡	身長	體重	號次	年齡	身長	體重
	歲 月	呎 吋	呖 嗙		歲 月	呎 吋	呖 嗙
1	14.1	4.11$\frac{1}{2}$	6.0$\frac{3}{4}$	39	14.7	4.11$\frac{1}{2}$	6.3$\frac{1}{2}$
2	14.9	4.10	5.7	40	13.1	4.11$\frac{1}{2}$	5.7
3	14.7	5.5$\frac{1}{4}$	7.5	41	14.3	4.11	6.4$\frac{1}{4}$
4	13.11	5.0	6.3$\frac{3}{4}$	42	13.3	4.4$\frac{1}{2}$	4.11$\frac{1}{2}$
5	14.11	5.3$\frac{3}{4}$	8.0$\frac{1}{2}$	43	14.3	5.3	6.7$\frac{1}{2}$
6	14.7	4.10	5.0	44	13.6	5.1$\frac{1}{2}$	6.13$\frac{1}{4}$
7	14.3	4.10	6.7	45	14.2	4.8$\frac{3}{4}$	6.0$\frac{3}{4}$
8	14.9	5.5	8.5$\frac{1}{4}$	46	13.5	5.2	7.4
9	14.11	4.9$\frac{1}{2}$	5.12$\frac{1}{4}$	47	13.8	5.2$\frac{1}{2}$	6.11
10	14.3	4.11$\frac{3}{4}$	6.11$\frac{3}{4}$	48	14.6	5.4	7.4$\frac{1}{2}$
11	13.4	4.7	5.1$\frac{1}{2}$	49	14.8	5.1$\frac{1}{2}$	6.10
12	14.7	5.3$\frac{3}{4}$	7.8$\frac{3}{4}$	50	13.3	4.8$\frac{1}{4}$	5.0
13	13 8	4.7$\frac{3}{4}$	5.3	51	13.0	5.1$\frac{1}{4}$	6.7
14	14.5	5.2$\frac{1}{2}$	7.8$\frac{3}{4}$	52	13.10	4.11$\frac{1}{2}$	7.3$\frac{1}{4}$
15	14.4	5.0	6.0	53	14.8	4.11$\frac{1}{2}$	6.9$\frac{3}{4}$
16	13.6	4.9	5.6	54	13.8	4.5$\frac{3}{4}$	4.9$\frac{1}{2}$
17	14.0	5.2$\frac{1}{2}$	7.7$\frac{1}{2}$	55	14.8	5.4$\frac{1}{2}$	7.0
18	13.0	4.8$\frac{1}{4}$	5.3	56	14.0	4.10	6.2$\frac{1}{2}$
19	14.7	4.11	6.12$\frac{3}{4}$	57	13.10	4.9	5.5
20	14.10	5.1	6.9	58	13.2	5.0$\frac{1}{2}$	6.4
21	13.9	4.11	5.11	59	13.6	4.7	5.2$\frac{1}{2}$
22	14.10	4.8$\frac{3}{4}$	5.11	60	13.0	4.9	5.9$\frac{3}{4}$
23	13.4	4.9$\frac{1}{2}$	5.8$\frac{3}{4}$	61	13.3	4.8$\frac{3}{4}$	5.5$\frac{1}{4}$
24	13.1	5.2$\frac{1}{4}$	6.1	62	13.5	4.8$\frac{1}{2}$	6.5$\frac{3}{4}$
25	14.0	4.6$\frac{1}{4}$	5.6$\frac{1}{2}$	63	13.10	5.5$\frac{1}{2}$	7.10$\frac{1}{2}$
26	14.6	5.3$\frac{1}{2}$	7.6$\frac{1}{2}$	64	13.1	4.8$\frac{3}{4}$	6.2$\frac{1}{4}$
27	14.3	5.0$\frac{1}{4}$	5.11$\frac{3}{4}$	65	13.10	5.4	7.2
28	13.9	4.9	5.11	66	14.0	4.9	5.0$\frac{1}{2}$
29	13.4	5.1$\frac{1}{4}$	5.9	67	13.3	4.7	5.0
30	14.4	5.1	6.8$\frac{1}{2}$	68	13.8	4.11	6.1$\frac{1}{4}$
31	14.10	4.9	4.7$\frac{1}{4}$	69	13.7	4.41$\frac{3}{4}$	6.4$\frac{1}{4}$
32	13.2	4.9$\frac{1}{2}$	5.13$\frac{1}{4}$	70	13.11	4.8	4.4$\frac{1}{2}$
33	14.1	4.8$\frac{3}{4}$	5.8$\frac{1}{2}$	71	13.11	4.8	4.4$\frac{1}{2}$
34	13.10	5.2$\frac{1}{4}$	6.8$\frac{1}{4}$	72	13.2	4.7$\frac{3}{4}$	4.10
35	14.0	4.11$\frac{1}{2}$	5.7	73	14.0	4.11	6.5
36	14.4	4.11	6.5	74	13.3	4.3$\frac{1}{2}$	4.1$\frac{1}{4}$
37	14.8	4.11	6.0$\frac{3}{4}$	75	13.3	5.0	7.2$\frac{3}{4}$
38	13.7	5.0$\frac{3}{4}$	6.2	76	13.7	4.8$\frac{1}{4}$	5.6

體重表之分析

算術平均數：六呖(註)一又二分之一磅。
算術平均數，當原來重量以呖爲準時；六呖一又五分之一磅。
中位數：六呖一又八分之一磅。
四分位數：六呖九又八分之七磅。五呖六又八分之一磅。
四分位數之平均：六呖一磅。
全部數字之半在中位數九磅距離之內，
衆數：在六呖至六又四分之一呖之間。
十三歲至十三歲半之平均體重：五呖九又二分之一磅。
十三歲半至十四歲之平均體重：五呖十三又八分之一磅。
十四歲至十四歲半之平均體重：六呖三又八分之一磅。
十四歲半至十五歲之平均體重：六呖八又三分之二磅。
身長表之分析從略。

（第二十二表）

身長之數量（以吋爲單位）依次排列如下:——

$51\frac{1}{2},52\frac{1}{2},53\frac{3}{4},54\frac{1}{4},55,55,55,55\frac{3}{4},55\frac{3}{4},56,$

$56,56\frac{1}{4},56\frac{1}{4},56\frac{1}{4},56\frac{1}{2},56\frac{3}{4},56\frac{3}{4},56\frac{3}{4},56\frac{3}{4};$

$56\frac{3}{4},57,57,57,57,57,57\frac{1}{2},57\frac{1}{2},57\frac{1}{2},57\frac{1}{2},$

$58,58,58,58,59,59,59,59,59;$

$59,59,59\frac{1}{2},59\frac{1}{2},59\frac{1}{2},59\frac{1}{2},59\frac{1}{2},59\frac{3}{4},59\frac{3}{4},$

$60,60,60,60\frac{1}{4},60\frac{1}{2},60\frac{3}{4},61,61,61\frac{1}{4},$

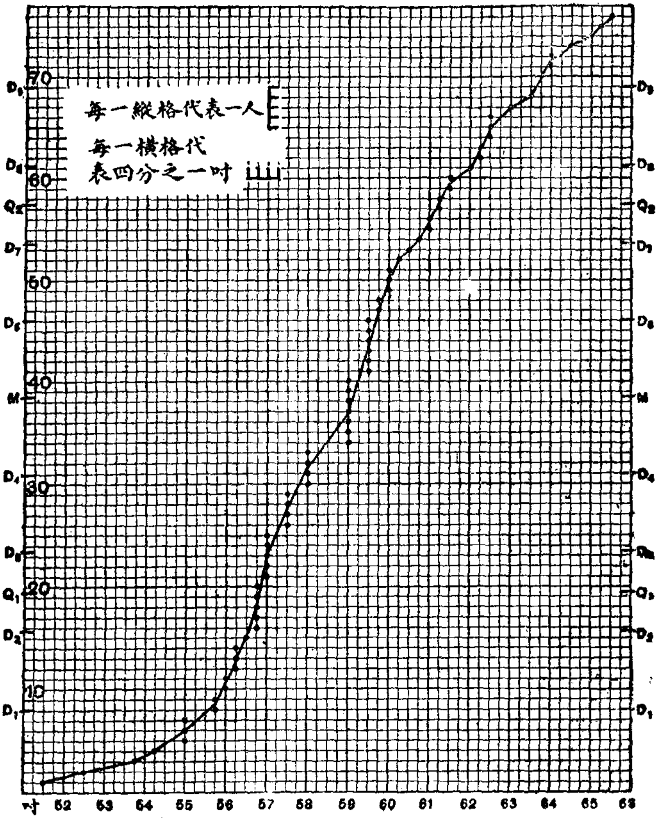
$61\frac{1}{4},61\frac{1}{2},61\frac{1}{2},62,62\frac{1}{4},62\frac{1}{4},62\frac{1}{2},62\frac{1}{2},62\frac{1}{2},63,$

$63\frac{1}{2},63\frac{3}{4},63\frac{3}{4},64,64,64\frac{1}{2},65,65\frac{1}{4},65\frac{1}{2}$。

中位數之圖示法 高爾頓氏在一八八一年英國協會之人體測量委員會 (The Anthropometic Committee of the British Association) 報告中第二百四十七頁，曾用圖示法，求身長數量之中位數，結果頗稱精密，茲爲舉例起見，用將圖式列於第143頁。

在下圖之橫標尺上，劃分爲均等之距離，每一距離代表一單位，此處之單位卽爲一吋。於縱標尺上，亦分爲等距，每一等距代表一項，此處之一項卽爲所量之一人。自底線最低之五一又二分之一吋起，對橫行之身長尺寸，依在該尺寸上之人數，在格上記一點，有若干人卽記若干點（本圖每處只有一人），使每一點卽代表一人。由此記成之點最前一個止，想像中假設劃一直線平行於底邊，直至第二個點所在之縱線上，卽爲五十二又二分之一吋，

求算中位數,四分位數,十分位數之圖示法。
（根據英國協會人體測量委員會高爾頓氏之法）
年齡在十三至十五歲之兒童七十六人之身長



（第　一　圖）

中位數:59$\frac{1}{8}$吋

四分位數: 56.8
61.2

四分位差 half inter-quartile distance)2.2,

十分位數:55.6,56.6,57,57.9,63.6.62,60.7,
59.7。

算術平均數:59.095

最深密度:57或59

修匀曲線上之最深密度:約爲58。

幾何平均數:58.98.

然後復自第二點起，仍照前法，依次進行，以尋在五十二又二分之一吋上面之點。次即連結成為一線，自上下成貫各點之中間一點經過，在一數量上同時有數人時，如人數為奇數，則中間一人即作為中心，如為偶數時，則以中間二數之中間為中心。

欲求中位數，四分位數及其他等等，即須將縱行依代表項數之距離，分為二等分，四等分及其他等分，並劃一平行線經過各該分割點，與已劃成之連結線交叉。各線之交叉點，向下直看，在橫標尺上，即為所求之身長數量。

終於此必須有一假定：假定某數量，譬如五十八又四分之三吋上，所量得之人數，實係自五十八又八分之五至五十八又八分之七止成均勻之分配；上列之作圖，所示中位數，四分位數，十分位數……等等之位置，即係依此假定而成也。

如項數過少，四分位數，及十分位數……等之位置不甚顯明時，應作如後之分析：於此有兩種情形，第一種情形，觀察所得之數量，即各項確實之數字，第二種情形，觀察所得之數量，作分組表示，或按組距中點計算。

（1）茲根據火車時間表，將四十五次車，經過某一距離所耗之分數(minutes)，列之如下：

45，46，47，48，48，51，53，54，55，58；61，61，62，65，65，69，69，69，71，76；76，76，77，77，78，80，81，81，82，82；83，83，84，85，85，85，85，87，88，89；90，92，94，101，103。

中位數在第二十三項,卽爲七十七分鐘。

欲求四分位數,須將四十五項,均分爲四等分。假設將前圖以二分之一起,經一又二分之一……以至四十四又二分之一,分爲尺度,縱長之全部分爲四十九個等分,則四分位數分別爲十一又二分之一,及三十三又四分之三。十一又二分之一,乃介乎在十又二分之一之第十一項(六十一),及在十一又二分之一之第十二項(亦爲六十一)之間。故下四分位數,乃爲六十一。三十三又四分之三,乃在第三十四與三十五項之間,二者均爲八十五。如各不相等,可取較近之項(第三十四項)之四分之三,加第三十五項之四分之一卽成。

同理,十分位數乃在縱標尺上之四又二分之一,九,十三又二分之一……等處。最低一數,乃在第五項(四十八分鐘),次一數乃在第九與第十(五十六又二分之一分鐘)之中間,其他以此類推。

上圖之D,Q,M各數之位置,卽係此原理而指出者。

茲將中位數及四分位數之求法,列之如下:

項數	中位數	下四分位數	上四分位數
$4n$	$\frac{1}{2}$(第$2n$項+第$\overline{2n+1}$項)	$\frac{1}{2}(n^{th}+\overline{n+1}^{th})$	$\frac{1}{2}(3n^{th}+\overline{3n+1}^{th})$
$4n+1$	$\overline{2n+1}^{th}$	$\frac{1}{4}n^{th}+\frac{3}{4}\overline{n+1}^{th}$	$\frac{3}{4}\overline{3n+1}^{th}+\frac{1}{4}\overline{3n+2}^{nd}$
$4n+2$	$\frac{1}{2}(\overline{2n+1}^{th}+\overline{2n+2}^{nd})$	$\overline{n+1}^{th}$	$\overline{3n+2}^{nd}$
$4n+3$	$\overline{2n+2}^{nd}$	$\frac{3}{4}\overline{n+1}^{th}+\frac{1}{4}\overline{n+2}^{nd}$	$\frac{1}{4}\overline{3n+2}^{nd}+\frac{3}{4}\overline{3n+3}^{rd}$

(第二十三表)

至計算十分位數之方法,理與上同,茲從略。

(2)各數分組排列(或自五十三至五十四吋,或五十三而以四分之一爲組距,

卽自五二又八分之七至五十三又八分之一），則可假定各組之數分配甚爲均勻，而照第一種情形辦理。

玆以本章第一節所列之已婚男子年齡（第十八表）爲例，釋明此法。該表中，年齡在四十歲以上者五百四十九人，四十五歲以上者四百一十三人，第五百人當屬於四十至五十歲一組中一百三十六人之一，事實上乃爲該組之第四十八或第四十九人也。如該組之分配情形，果係甚爲均勻，則五歲組距，可分爲一百三十六等分，所謂第四十九人，卽在該組之第四十九等分。故中位數卽爲$40+5\times\frac{549-500}{136}=41.80$歲。此數已盡確實之能事。若欲再求進一步之精確，似已無此必要。同理，下四分位數卽爲第七百五十人之年齡，約在三十歲至三十五歲一組內，故可作爲 $30+5\times\frac{855-750}{152}=33.45$歲，而上四分位數爲$50+5\times\frac{295-250}{96}=52.34$歲也。

以此等數量，用圖示法，對於上舉二種情形，甚易求出。

第六節　幾何平均數

如有$a_1, a_2, \cdots\cdots a_n$共$n$個數量，G，幾何平均數(geometric mean)，或名對數平均數(logaritmic mean)，可用下式求之：

$$G=\sqrt[n]{a_1a_2\cdots\cdots a_n}$$

$$\text{及}\ \log\ G=\frac{1}{n}(\log\ a_1+\log\ a_2+\cdots\cdots+\log\ a_n)。$$

同一之若干數量，其幾何平均數，必比算術平均數爲小。若吾人重在兩數量間之比率，而不重視其絕對差數，自以用幾何平均數爲宜。如八與十三之差，與十三與十八之差，有同等之重要

性，則八與十八之平均，即爲十三，於是十三距八與十八，距離乃相等；但八對十二之比率，與十二對十八之比率，若同其重要，則八與十八之平均，乃等於$12=\sqrt{8\times 18}$。

茲爲明瞭起見，更作比喻如下：設有a_1,a_2,a_3,a_4,a_5五數，就中a_1，及a_2比A（五數之算術平均數）及G爲小；a_3,a_4，及a_5則比A及G爲大。

則 $$5A=a_1+a_2+a_3+a_4+a_5$$

$$(A-a_1)+(A-a^2)=(a_3-A)+(a_4-A)+(a_5-A)$$

$$G^5=a_1+a_2+a_3+a_4+a_5$$

$$\frac{G}{a_1}\times\frac{G}{a_2}=\frac{a_3}{G}\times\frac{a_4}{G}\times\frac{a_5}{G}$$

由此觀之，一方，多過平均數之部分之和，等於缺少部分之和；一方，平均數對少於平均數之數之比率之積，等於平均數對多於平均數者之比率之乘積。

此平均數之重要用途，乃對於物價之計算。物價自100升至120，其意義與自120升至144之情形相同，但較自120升至140之高漲程度爲大。翟翁士(Jevons)在其指數（金價之跌落）之第一步論述中所以施用幾何平均數者，或即以此之故。

最後一點，吾人不可不注意者，幾何平均數，較之算術平均數，常將小數之重要性稍擡高，將大數稍抑低。

第七節 總論

平均數之功能 平均數之功能如何，吾人不可不知，其功用在以簡短之單純數字，表明甚繁之羣類。萬千之項目當前，欲吾人立時領會若是其多之數量，勢乃有所不能，必也彙成組類，化爲簡單，然後施以平均。各種平均數，以選用何者爲適宜？此須視其是否能揭示醒目之形態，及顯露該類之主要特質而定。羣類有種種之不同，方法自亦隨之而異，此宜就各個之所長，而善爲利用之者也。凡一良好而適宜之平均數，須具下述之特質：如羣類有一型態存乎其間，平均數須能表明之；對於極端之項目，應給予公平之待遇；應不致輕易受差誤之影響，或因計算方法稍變，而地位移動甚烈之情事，應不易發生；且須易於計算。

各種平均數，不妨同時算出，以比較其位置，而藉以明瞭羣類之全部性質。在對稱之羣類中，算術平均數，中位數及衆數，必相合於一處。如羣類甚小而數目甚大時，算術平均數大多在中位數之上方。如數並不大，則算術平均數，必在中位數之下方，但項數集中時，則算術平均數，仍將位於中位數之下首。應用衆數之要件，必須羣類係爲同一體質之物所組成，否則雖求出衆數，則斯衆數乃亦毫無意義。如羣類甚小，但每項之數目甚大時，衆數多位於算術平均數之下。如羣類之分配，甚爲均勻時，衆數之位

置，甚爲顯明——凡此種種定則，乃就一般之羣類分配而言，如遇有特殊情形，只能用之作爲試驗，但大多均易失其效用也。

（註一）請參閱醫學百科詞典(Dictionaire encyclopédique des Sciences Médicales)白狄雍(Bertillon)博士之平均數(Moyenne)一章。

（註二）原文見統計學報(Statistical Journal)之一八九七年十二月份，茲已加修改。

（註三）參閱本書原著者所作「統計學概要」(Elementary Manual of Statistics)第一百零五至一百零七頁，及統計學報一九〇六年號第三十四至一百四十七頁。

（註四）參閱統計學報一九〇八年號第四百六十三至四百六十八頁，關於討論同質性(homogeneity)，可比較性(comparability)及相對性(relativity)一文。

（註五）見葛得雷氏著社會物理學(Physique Social)及愛基華斯氏(Edgeworth)在統計學報一八九三年號所發表一文。

（註六）此等數量，前於第四章第四節論表列法業已論及，請參考（見第十四表）。

（註七）見高爾頓氏之「自然遺傳性」(Natural Inheritance)第四十七頁。

（譯者註）㕂(stone)等於一百一十二磅。

第六章　離散度與偏斜度之測量及平均數之應用

統計羣類　前於第五章討論平均數時，時以考查統計羣類(statistical group)之中心位置爲言，所謂統計羣類者，乃指一羣人或物，各具有某種一定之性質(一九一一年英格蘭及威爾士挨戶調查之男子人數)，依另一變動之性質(年齡)，彙分類別也。關於此一羣類，表白之法不一，用分組表列法固可，用他種之表列法(如第四章第四節之美國工資表(第十三表))亦可，卽用圖示法(如第七章第一節第二圖工資統計圖)亦無不可，惟爲簡捷，或便於與其他羣類比較起見，頗有將關於該羣類之數量，加以限定並計算之，以表現羣類特質之必要。爲達到此目的，最簡便之法，莫過於選擇(一)足以代表中心位置之平均數(mean)，(二)離散度(dispersion)，卽觀察項數之分散情形，並(三)偏斜度(skewness)。玆請就離散度及偏斜度二者討論之。

第一節　離散度

離中差　一羣類中各項數量與平均數或其他固定點之相差，謂之離中差(deviation)。在下列第二十四表中，所列羣類，內容

包括英格蘭及威爾士各大城市一九〇二年中五十二個星期按週之死亡率合計數。第一欄由上而下,第二欄由下而上，依數值之

英格蘭及威爾士各大城市一九〇二年
中五十二個星期每週之死亡率合計數

每萬人中每週死亡率		求對中位數之平均差		求標準差		求均互差		
a	b	a列少於 173 者	b列多於 173 者	與 173 差額之平方		a與b之差數	乘數	乘積
(一)	(二)	(三)	(四)	(五)	(六)	(七)	(八)	(九)
244	136	71	37	5,041	1,369	108	51	5,508
233	139	60	34	3,600	1,156	94	49	4,606
226	141	53	32	2,809	1,024	85	47	3,995
209	143	36	30	1,296	900	66	45	2,970
206	144	33	29	1,089	841	62	43	2,666
201	145	28	28	784	784	56	41	2,296
196	149	23	24	529	576	47	39	1,833
196	150	23	23	529	529	46	37	1,702
196	151	23	22	529	484	45	35	1,575
191	152	18	21	324	441	39	33	1,287
183	154	10	19	100	361	29	31	899
182	155	9	18	81	324	27	29	783
182	159	9	14	81	196	23	27	621
181	160	8	13	64	169	21	25	525
179	164	6	9	36	81	15	23	345
177	165	4	8	16	64	12	21	252
177	166	4	7	16	49	11	19	209
177	166	4	7	16	49	11	17	187
176	167	3	6	9	36	9	15	135
176	169	3	4	9	16	7	13	91
176	169	3	4	9	16	7	11	77
174	169	1	4	1	16	5	9	45
174	170	1	3	1	9	4	7	28
174	170	1	3	1	9	4	5	20
173	172	0	1	0	1	1	3	3
173	172	0	1	0	1	1	1	1
9,029		434	401	26,491				32,659
		+13	−13					
		$\underbrace{\qquad\qquad}_{835}$						

（第二十四表）

算術平均數：$9029\div52=173.63$

中　位　數：$172\frac{1}{2}$

四 分 位 數：$159\frac{1}{2}, 181\frac{1}{2}$

對於中位數之平均差(mean deviation)：$\eta=835\div52=16.06$。

對於算術平均數之平均差：$\eta=16.11$。

四分位差(quartile deviation)或概誤 (probable error)，$r=\frac{1}{2}(181\frac{1}{2}-159\frac{1}{2})=11$。

全部羣類項數之半，在 $170\frac{1}{2}\pm11$ 以內。

標準差(standard deviation)：$\sigma=\sqrt{\left\{\frac{1}{52}\times26,491-(0.63)^2\right\}}=22.56$。

離中係數(coefficient of variation)$=\frac{100\sigma}{173.63}=13.0$。

均互差(mean difference)：$g=32.659\div\frac{1}{2}\times52\times51=24.63$。

大小度，順序排列。中位數爲 $172\frac{1}{2}$，玆取近於此數之173爲中心而計算離中差，分列於三，四兩欄。據第五章第五節之論述，如以中位數爲中心點，而計算離中差，則各離中差（均作正數計算）之總合與平均，必成爲最小數；玆爲求得若是之離中差，必須將第三欄各項各加 $\frac{1}{2}$，而自第四欄各項各減去 $\frac{1}{2}$，換言之，即將第三欄之總計數加上 13，而由第四欄之總計數內減去 13 也。由此觀之，對於中位數之正離中差之總數爲447，負離中差之總數爲388，全部五十二個離中差之總合爲 885。離中差之平均爲 $835\div52=16.06$，又羣類之算術平均數爲 173.63。欲求由算術平均數計算之離中差之總合，須將小於 174 之二十八個數量對假定中位數 173 之各離中差，各加上 0.63，並由其餘二十四個離中差，各減去 0.63；於是對於算術平均數之離中差之總合，爲 837.52，平

均爲 16.11。

平均差 各數量與其算術平均數相差之平均數(此處爲16.11),謂之該羣類之『平均差』(mean deviation);普通多以希臘字母 η 表示之;『對中位數之平均差』(此處爲 16.06)一詞,亦多有應用之者。平均差用以測量一羣類之離散度,意義甚爲明顯,手續亦極簡便,當數量各自獨立不成組時,算之尤易。

由上表之第三四兩欄,欲求算術平均數,只須作如下之計算:多於 173 者之平均數=第三欄之總數 (434) 減第四欄之總數 (401)被 52 除=33÷52=0.63,即可求得算術平均數=173.63。

標準差 然諸如此類之絕對差,如欲用數學方法,研究統計羣類時,因在代數程序中有須成爲正數者,有須成爲負數者,用之至感不便,且如應用機率(probability)原理,則離中差之所以成爲重要者,乃在於離中差之平方,而不在其第一次方,故吾人乃用對於算術平均數之離中差之平方之平均,即均方差,而不用上述之平均差。此均方差(average of the squares of the deviation)之平方根,卽名之曰該羣類之『標準差』(standard deviation),卽一般所習用而用希臘字母 σ 代表之者也。計算標準差之方法,必須逐一計算各數之離中差,但如照下述之程序,則頗收簡單化之效:

設 x_1, x_2……x_n 代表各數量,x_0 代表最便於計算離中差之

中心數量。以 $d_1=x_1-x_0$, $d_2=x_2-x_0$……，此即為上表所列之離中差。並設 $\bar{x}$ 為該羣類之算術平均數，於是 $n\bar{x}=x_1+x_2+\cdots\cdots+x_n$；又以 $d_0=\bar{x}+x_0$，則 $nd_0=(x_1-x_0)+(x_2-x_0)+\cdots\cdots=d_1+d_2+\cdots\cdots d_n$。

但標準差之定義，則為

$$\sigma^2=\{(x_1-\bar{x})^2+(x_2-\bar{x})^2+\cdots\cdots+(x_n-\bar{x})^2\}\div n$$

$$=\{(d_1-d_0)^2+(d_2-d_0)^2+\cdots\cdots\}\div n$$

$$=\{d_1{}^2+d_2{}^2+\cdots\cdots-2d_0(d_1+d_2+\cdots\cdots)+nd_0{}^2\}\div n$$

$$=\{d_1{}^2+d_2{}^2+\cdots\cdots-nd_0{}^2\}\div n,$$ 因為 $d_1+d_2+\cdots\cdots=nd_0$。

於是 $\sigma=\sqrt{\left\{\frac{Sd^2}{n}-d_0{}^2\right\}}$，至 Sd^2 則代表 $d_1{}^2+d_2{}^2+\cdots\cdots+d_n{}^2$ 也。

上表中，$\bar{x}=173.63$, $x_0=173$, $d_0=0.63$。

$d_1{}^2, d_2{}^2$……等數分列於第五六兩欄，其 $Sd^2=26.491$。

$\therefore \sigma=\sqrt{\{26.491\div 52-0.63^2\}}=22.56$。

標準差之計算，均以算術平均數為基準，與中位數不生關係。

四分位差　測量離散度之法，除此之外，尚有更為簡單者，即採用四分位數。上下兩四分位數之相差，與離散度之大小，關係至為顯然。但如彼一羣類之四分位數，與此一羣類之四分位數相同時，則不論上下兩四分位數之間，數量之分配情形如何，亦不論極端數量距四分位數之遠近，而所得之四分位相差數，終屬

相同，此乃四分位相差數之缺點。故就對於適應資料離散情形之感覺一點而言，四分位相差數遠不若平均差或標準差之靈敏。茲爲補救此一缺憾起見，特用上下兩四分位數間距離之一半，而不用其全部距離，卽以距離之一半名之爲四分位差 (quartile deviation)，普黑多以 r 表示之者是也。r 之數値約略與所有離中差之中位數相等，但就一般情形而論，並非完全相等。在上表中，下四分位數爲 $159\frac{1}{2}$，上四分位數爲 $181\frac{1}{2}$，二者相距爲 22，於是 $r=11$。

中位數之位置，未必卽在上下兩四分位數之中點。卽以上列一表而言，上下兩四分位數之中間，爲$\frac{1}{2}(159\frac{1}{2}+181\frac{1}{2})=170\frac{1}{2}$，而上下兩四分位數，則爲$170\frac{1}{2}\pm 11$。討論至此，請以極簡單之語句，敍述該一羣類之情形如次：其算術平均數爲 173.6(或云中位數爲 173)，全部觀察數量之半，乃在 $170\frac{1}{2}\pm 11$ 範圍之內。

在成對稱(symmetry)形式之羣類，算術平均數與衆數相合於一點時，r 又名爲機誤(probable error)；機誤一詞，有時應用之，頗收簡便之效，惟有時往往有令人誤解之虞也。

分組之資料 資料分組排列時，求離散度之方法，與不分組之資料，略有不同，且所求出之離散度，又不能十分準確。例如第五章第一節所舉之第十八表英格蘭及威爾士已婚男子之年齡，其中位數及四分位數，經在同章第五節之計算，得知中位數爲

41.80歲，下四分位數爲 33.45 歲，上四分位數爲 52.24 歲。然則其四分位差，必爲 $\frac{1}{2}(52.24-33.45)=9.45$ 歲，可知全部項數之半，均不出 42.90 ± 9.45 歲範圍之外。在同表之第二欄，已將各項對於 $42\frac{1}{2}$ 之離中差算出。假定每組之所有各項，均集聚於各該組之中心點，則同表第四欄所列，卽爲各該組離中差與項數相乘之總合離中差，此列總合離中差，如不計其符號之正負，則其總和爲 $1155+926=2081$，亦卽爲全部離中差之總數。於是各項對 $42\frac{1}{2}$ 歲之平均差，大約將等於 $\frac{2081}{1000}\times5$ 歲，卽等於 10.40 歲。茲如計算對於中位數或對於算術平均數 (43.645) 之平均差，須加以少許之修正。惟頗感煩難者，卽假定爲中心而作爲零點之組有一百三十六項，須將其離中差算出合併於離中差總數之內，同時各組之項數，旣已假定均集聚於各組之中心，則斯時亦須將此假定加以修正。在各組分組甚狹（如就同表而論，苟年齡分組以一年爲組距，則可謂之甚狹）時，此項修正，可以作罷，惟在此情形之下，如欲計算分組資料之離散度，則不可不用平均差。在計算平均差之時，須以包含中位數或算術平均數之一組中心爲原點，然後計算對於中位數或算術平均數之離中差，至究以中位數抑以算術平均數爲出發點，則仍須視吾人之決定如何也。

關於上述種種理論，如計算標準差，只須分組甚狹，則原理亦全相同。例證見第二編第一章。

吉尼教授(Corrado Gini)對於差度(variatiation),曾介給一新穎之方法(見 Variabilità e Mutabilità)。彼之主張,以爲在人口統計學上、人類學上、生物學上、或經濟學上對於變量之研究,所引起之問題,乃爲『各個量數彼此差異之程度如何』?並非求其『各個量數對於算術平均數之相差若干』?也。第二問題,對於自然科學方面,頗有其存在之理由,但對於論述羣類時,則又非盡然如此。

均互差 吉尼氏依彼之見地,提出一種測量離散度之方法,在此方法之下,第一須先求出 n 個數量相互間之 $\frac{1}{2}n(n-1)$ 個相差量;第二計算相差量之算術平均數,然後即以此算術平均數,表示離散度。此一離散度,吾人可名之曰均互差 (mean difference),而以字母 g 代表之。均互差一詞,迄未爲一般所採用,此或因計算方法乃爲間接且極繁重之故(極爲單簡之羣類中,自爲例外),然吾人殊不能因此而否認其存在之價值,謂其過於單純且不合論理也。

設有 $a_1, a_2, \cdots\cdots a_n$ 共有 n 個之數量,由小而大依次上升排列之,則

$$g\times\frac{1}{2}n(n-1)=(a_n-a_1)+(a_n-a_2)+\cdots\cdots+(a_n-a_{n-2})+(a_n-a_{n-1})+(a_{n-1}-a_1)+(a_{n-1}-a_2)+\cdots\cdots+(a_{n-1}-a_{n-2})$$

$$\vdots \quad \vdots \quad \vdots \quad \vdots$$

$$+(a_3-a_1)+(a_3-a_2)$$

$$+(a_2-a_1)$$

$$=(n-1)a_n+(n-3)a_{n-1}+(n-5)a_{n-2}+\cdots\cdots$$

$$+(1-n)a_1+(3-n)a_2+(5-n)a_3+\cdots\cdots$$

$$=(n-1)(a_n-a_1)+(n-3)(a_{n-1}-a_2)$$

$$+(n-5)(a_{n-2}-a_3)+\cdots\cdots$$

此種計算，已見於本章本節之每週死亡率即第二十四表之第七,第八,第九三欄;但此爲當 n 爲偶數之例,在 n 爲奇數時,則將中心之數以零乘之。

至於 g（均互差）與 η（平均差）之關係,可示之如次:

設 $a_1,a_2\cdots\cdots a_n$ 各個與中位數之差量,爲 $d_1,d_2,\cdots\cdots d_n$,負號均作正號計,則

$a_n-a_1=d_1+d_n;a_{n-1}-a_2=d_2+d_{n-1}$;其他依此類推,

又
$$g=\frac{1}{n}\left\{2(d_1+d_n)+2\cdot\frac{n-3}{n-1}(d_2+d_{n-1})\right.$$
$$\left.+2\cdot\frac{n-5}{n-1}(d_3+d_{n-2})+\cdots\cdots\right\}$$

$$\eta=\frac{1}{n}\left\{d_1+d_n+d_2+d_{n-2}+d_3+d_{n-3}+\cdots\cdots\right\}。$$

g 對於極端之差度,均予以均等之重視,而 g 又永大於 η。例如,設觀察數量,均分爲均等之間距(interval),並以 k 代表間距,

則可示明 g 約略等於 $\eta\times\frac{4}{3}$；蓋在此情形之下，苟如 $n=2m+1$，則 $g=\frac{2}{3}(m+1)k$，$\eta=\frac{m(m+1)}{2m+1}k$，$g\div\eta=\frac{4}{3}\left(1+\frac{1}{2m}\right)$；同理，$g-\eta$ 亦約略等於 $\frac{1}{6}mk$。

又各數量之次數，如每一數量不只有一個，但在 a_1 處有 y_1 個，在 a_2 處有 y_2 個，……在 a_t 處有 y_t 個，而 $y_1+y_2+y_3+\cdots\cdots y_t$ 相加之和等於 N，則計算手續，必將愈加煩瑣。茲示明之如下：

$$
\begin{aligned}
&g\times\tfrac{1}{2}\mathrm{N}(\mathrm{N}-1)=\\
&y_t d_t(\mathrm{N}-y_t)+y_{t-1}d_{t-1}(\mathrm{N}-2y_t-y_{t-1})\\
&+y_{t-2}d_{t-1}(\mathrm{N}-2y_t-2y_{t-1}-y_{t-2})+\cdots\cdots\\
&+y_1d_1(\mathrm{N}-y_1)+y_2d_2(\mathrm{N}-2y_1-y_2)+y_3d_3(\mathrm{N}-2y_1-2y_2\\
&\qquad -y_3)+\cdots\cdots
\end{aligned}
$$

式中，第一及第二行之 d，乃分別爲位在中位數之上之數量，與在中位數下之數量，相減之差量。各因子之計算，依照上式，頗爲易易，然後列成表格（註一）。

又如各數量之分配，成爲差誤常態曲線（normal curve of error 見第二編第二章）形式時，則可得下述之關係：$\eta=\sigma\sqrt{\frac{2}{\pi}}=\sigma\times\cdot798\cdots\cdots$，$r=\sigma\times\cdot6745$，$g=\eta\sqrt{2}=\eta\times1.414\cdots\cdots$。此種關係，即在非常態之分配中，亦時常得出約略相近之數。例如在

本章本節第二十四表每週死亡率下，$\eta=\cdot 7\sigma$，$g=\eta\times 1.41$，但 r 則僅等於 0.5σ。

如遵照吉尼教授原意，採用所有相互差量平方之平均數之二次方根，則不論分配所現之形式是否合於常態，仍可得 $\sigma\sqrt{2\left(\frac{n}{n-1}\right)}$，即作為 $\sigma\sqrt{2}$，亦極相近。

離散度之數量，既如上述，但離散度數量，如成為名數，而言若干先令，幾歲，或標尺之若干格時，吾人須以之表示對於平均數之關係。例如，在一工資羣類，中位數為四十先令，下四分位數為三十先令，上四分位數為五十先令，四分位差為中位數之四分之一，但在另一羣類，中位數為四十五先令，下四分位數為三十五先令，上四分位數為五十五先令，則四分位差乃為中位數之九分之二；由此可作一合理之論斷：二羣類之四分位差雖相同，但仍可認為第二羣類之離散程度，較之第一羣類為輕。

離中係數　為達到比較目的計，可用下列三種算法：

$$(a)\ \frac{\text{四分位差}}{\text{上下四分位數之平均數}}\ (\text{註二})；$$

$$(b)\ \frac{\text{平均差}}{\text{中位數}}；$$

$$(c)\ \frac{\text{標準差}}{\text{算術平均數}}；$$

但一般所應用者，乃用標準差作為算術平均數之百分比，換言之，即用 $\frac{\text{標準差}}{\text{算術平均數}}\times 100$，此即名為離中係數(coefficient of variation）者是也。如用第二十四表每週死亡率表之數字，算出之離中係數，爲$(22.56 \div 173.63) \times 100 = 13.0$。

第二節 偏斜度

一個分配曲線之偏斜度(skewness)，或名爲不對稱(asymmetry)，可在衆數，中位數與算術平均數三者不能相合於一點時，判定之。但以此爲判斷之標準，尙嫌不能十分準確，欲求更爲確定之表現，則須視對於中位數計算之正號離中差之和，是否與負號之離中差之和，數值完全相等，如其不相等也，則知羣類有偏斜之形狀，而可以作計算偏斜度之根據矣；又當上四分位數與下四分位數，或十分位數與十分位數，與中位數之距離不相等，則亦可視爲羣類有偏斜之表示。所謂偏斜度者，乃指一曲線之形式，非言曲線之大小也，但用一絕對之數量(absolute quantity)——類似一橢圓形之離心率(eccentricity)——測算之，結果亦頗切合，故求用作測量之用，必須覓得一種由兩個具體數量而成之比率。計算之法爲簡單者，乃如下述：

設以 q_2 代表上四分位數超出中位數之部分，以 q_1 代表中位數超出下四分位數之部分，則$s=\frac{q_2-q_1}{q_2+q_1}$卽爲求得之偏斜度(註三)。

曲線如爲對稱形式，則 $q_2=q_1$，而 $s=0$；如 q_2 大於 q_1，則 s 爲正數，如 q_2 小於 q_1，則 s 爲負數。又如 q_1 等於零，則 s 變爲 $+1$，

如 q_2 等於零，則 s 變爲 -1；換言之，如中位數與下四分位數相合於一點，則 s 爲正一，如中位數與上四分位數相合於一點，則 s 爲負一。故 s 一數量，永無大於一之可能，且於等於 $+1$，-1 或等於零時，其意義均有一定也。返觀本章第一節之每週死亡率表，其 $q_2=9$，$q_1=13$，則 $s=\frac{-4}{22}=-0.19$。又查第五章第一節之已婚男子年齡表，其 $q_2=10.54$，$q_1=8.35$，則 $s=0.12$。至 s 之值大小，所含意義若何，非集成經驗，不易領會，惟在 $s=.1$ 時，不妨謂其偏斜度甚微，$s=0.3$ 時，則其偏斜度較大也。

最後吾人須加注意者，一羣類之中心地位，離散度及偏斜度，三種表徵數(characteristics)，只須用中位數及四分位數卽可求出；中位數代表中心之位置，四分位差表示離散度，而上述之數量，卽表明偏斜度者也。

第三節　平均數之應用舉例

平均數之應用　平均數之性質及用途，前已討論及之，如前所研究者倘不完全則已，如其果爲透澈，而平均數之應用範圍果爲甚廣也，玆請應用愼爲選擇具有確定意義之少數數目，進而表明任何之數字羣類。

對於行車時刻之應用　請先作一嚴重之試驗，舉一日常生活事項爲例，卽就郊外商人之觀點，考察兩條火車路綫之優劣，

以便決定在買季票時，究以購買何線者爲宜。

玆將一八九八年英國自萊仔亥德(Leatherhead)至倫敦之火車行車時刻，列表於下：

萊仔亥德至倫敦火車行車時刻
行車所用之分數

華鐵廬站

下行車——60,50,52,48,47,61,50,44,48,53,45,42,45,49,43,48,42,43。

星期日——50,50,47,49,50。

上行車——51,46,51,48,43,44,48,48,64,45,48,47,45,47,46,47。

星期日——48,48,51,51,51。

倫敦橋站

下行車——67,65,65,61,74,51,56,66,65,53,59,41,49,44,58,57,56,67,80。

星期日——67,52,66,68,88,65,65,68,65。

上行車——69,57,53,58,54,41,58,52,42,40,55,67,79,98,69,66,68,64,71。

星期日——72,71,69,70,62,81,73,73。

維多利亞站

下行車——77,65,55,76,77,88,48,53,46,69,89,54,82,71,90。

星期日——92,45,81,84,78,61,85,83,85。

上行車——87,65,69,69,47,48,51,83,101,58,62,61,76,103,

星期日——81,76,80,85,85,82,94。

各站間之必要數量，列表於後：——

	倫敦橋站（分鐘數）	維多里亞站（分鐘數）	華鐵廬站（分鐘數）
四次特別快車之平均數	41	$46\frac{1}{2}$	$42\frac{1}{2}$
最低十分位數	$47\frac{1}{2}$	48	43
中位數	65	77	48
衆數	65	...	48
星期一至六，開車次數	38	29	34
總平均數	63	73	48

（第二十五表）

於此須加聲明者，統計方法之施用，一般只限於一問題之一面，準時行車之問題，固可作統計之研究，但車上設備是否舒適，及沿路風景等問題，均不在吾人研究範圍之內也。

下舉一例，示明一關於社會學之資料，所組成標準羣類之表徵數。

工資調查表列　下表所列，爲一八六二年及一八九一年，英國工程人員聯合會認可之工資數。

英國工程人員聯合會若干分會一八六二年
及一八九一年之每週工資（加工不計）（第一）

	一八六二 先令	一八六二 便士	一八九一 先令	一八九一 便士		一八六二 先令	一八六二 便士	一八九一 先令	一八九一 便士
亞克令頓	27	0	31	0	布來克倍兒恩	27	6	32	0
阿石福特	33	6	30	0	波爾頓	27	6	28 32	0
阿石頓恩得爾林	29	3	34	0	布里治華德	24	6	24	0
拔可程	26	1	28	0	布來頓	24	$8\frac{1}{2}$	29	0
拜柔音份爾奈斯	31	0	34	0	布里斯他爾	31	0	32	0
巴斯	29	0	31	0	伯爾恩累	27	0	30	0
白得福特	27	0	29	0	伯爾吞昂春特	25	0	30	0
必爾斯頓	28	0	30	0	伯瑞	28	3	30 32	0 0
賓累	24	0	29	0	卡地夫	31	0	34	0
倍兒京亥得	29	0	35	6					
伯明漢	32	0	36	0					

英國工程人員聯合會若干分會一八六二年及一八九一年之每週工資（加工不計）（第二）

	一八六二		一八九一	
	先令	便士	先令	便士
卡里斯羅	24	6	30	0
卡布斯頭	30	0	34	0
卡斯特	30	0	32	0
抽奔特	26	0	32	0
寇恩	25	0	31	0
康里頓	24	0	28	0
寇文特利	28	0	34	0
克里由	29	4	30	0
達林頓	25	0	31	6
達特福特	34	0	38	0
達爾文	27	0	32	0
得爾拜	26	0	29	0
當卡斯特	28	6	31	6
賓維爾	35	6	36	0
昂賁爾潯洛克	36	0	40	6
愛塞斯特爾	23	0	28 32	0 0
賁維山	34	0	33	0
佛克斯東	34	0	32	0
佛婁木	24	0	27 30	0 0
根斯巴婁	27	6	28	0
哥老摟波	27	2	32	0
哥老塞斯特	28	0	32	0
哥藍仔木	28	6	30	4
哥里木斯白	28	0	32	0
哈里發克斯	23	1	31	0
漢累	28	3	32	0
哈特爾浦	26	0	34	10
里物得	27	0	30 34	0 0
侯里亥得	32	0	28	0
赫得斯賁爾得	26	0	26	0
赫爾	27	6	34	0
亥得	30 28	0 0	30 28	0 0
伊布斯維治	28	6	28	0
克爾恩累	23	0	27	0
可以得敏士得	28	0	30	0
藍加斯得	25	0	32	0
里茲	25	0	30	0
雷塞斯特	26	0	31	6
雷	27	9	31	6
林肯	26	7	28	6
利物浦	29	0	34	0
倫累	22	0	26	0
麥克爾賁爾德	24	0	29	6
曼塞斯特	29	7	35	0
麥克斯巴婁	25	0	34	0
密得爾頓	29	5	33	0
密爾頓及愛爾斯加	28	0	34	0
尼斯	32	0	30	0
紐阿爾克	25	0	29	0
紐卡斯爾	25	0	35 37	0 0
紐荷蘭	30	8	34	0
紐浦特	30	0	32	0
紐唐（斯他科浦特）	29	0	32	0
牛頓阿巴特	33	0	33	0
腦贊布頓	26	0	32	0
鸞絲夫里特	36	0	36	0
南北歇爾次	26	0	35	0
腦爾維支	32	0	29	0
腦廷漢	27	5	34	0
奧爾得拜瑞	28	0	34	0
奧爾得漢木	29	0	33	0
彼得巴婁	28	6	33	0
浦來毛斯	32	0	33	0
滂堤浦來得	24	0	30	0
浦爾特毛斯	35	0	34	0
浦來斯頓	27	0	32	0
拉得克里夫布里治	27	0	30 32	0 0
里丁	28	0	32 34	0 0
里浦累	26	0	26	6
拉則漢木	27	6	32	0
拉哥白	32	0	28 32	0 0
拉治累	24	11	30	0
聖海倫斯	28	0	34 36	0 0
晒賁爾得	28	0	36	0
西浦累	25	9	28 30	0 0
市盧白瑞	30	6	32	0
斯麥斯維克	28	0	35	0
掃贊頓	32	0	34	6
掃爾拜布里治	24	6	30	0
斯塔里布里治	38	3	32	0

英國工程人員聯合會若干分會一八六二年及一八九一年之每週工資（加工不計）（第三）

	一八六二 先令	一八六二 便士	一八九一 先令	一八九一 便士
斯塔克浦爾特	28	0	{32	0
			{34	0
斯塔福特	28	0	32	0
斯他克呑昂布里治	24	0	36	0
斯顛克昂春特	29	0	32	0
斯特拉德及斯拉浦	26	0	30	0
孫頓	31	6	31	0
桃得毛爾頓	26	0	28	0
維克費爾德	25	0	30	0
瓦林呑	28	0	24	0
瓦特福特	35	0	36	0
溫斯布瑞	26	0	31	0
圖得黑文	25	0	{28	0
			{36	0
維甘	28	0	34	0
維爾費漢普呑	28	0	33	0
維爾費呑	29	2	29	0

	一八六二 先令	一八六二 便士	一八九一 先令	一八九一 便士
華爾塞斯特	31	0	30	0
倍爾蒙得賽	35	4		
布來克華爾	34	0		
鮑	36	0		
哥林維支	34	0		
京斯克拉斯	36	0		
藍白斯	35	8		
倫敦東區	35	0		
倫敦北區	35	10	38	0
倫敦南區	35	0		
倫敦西區	35	6		
馬利爾崩	33	0		
斯特拉得福特	{35	0		
	{33	6		
桃爾漢木來次	36	6		
吳爾維治	36	0		

由此可得簡括之表徵數如下：

	一八六二年（一分會作一計）先令	（一）便士	一八九一年（一分會作一計）先令	（二）便士	一八九一年（按各分會會員計算）先令	（三）便士
最高工資	36	6	40	6	...	
最高十分位數	35	0	38	0	38	0
上四分位數	31	4	34	0	36	0
中位數	28	0	32	0	34	3
算術平均數	28	10	32	4	33	4
衆數	28	0	{30	0	...	
			{32	0		
下四分位數	26	0	30	0	31	6
最低十分位數	24	6	28	6	30	0
最低工資	22	0	24	0	...	
四分位差	2	8	2	0	2	3
偏斜度（用四分位數計算）	.25		0		−.22	

上表第三欄之計算，乃以會員實在所得之工資爲基準，假若各分會會員之工資，並非實在之數，而爲會員實在工資平均數，雖實在工資與其平均數相差不遠，但表中第三欄之數字，必因之而有變動。

以第二欄與第三欄比較，可知不惟所有各種平均數，均有增加，而且因最低十分位數及下四分位數之增加，較最高十分位數及上四分位數之增加爲尤速，可知下半部有逼近上半部之勢。再者，第二欄之工資，較第一欄者爲緊湊。

第四節　第三類表列法

文字答語之表列法　非數字或文字敍述性質之答語，表列之方法，在未將平均數討論完竣之前，不得不暫行從緩。茲既討論至此，特舉例如下，以示用形容詞之答語，所組成之羣類，如用中位數四分位數等論述之，亦頗能收簡捷之效。

一八九一年，英國工程人員聯合會，經向各分會調查『加工工作幾許』？一問案，茲將各分會書記之答語，列表於下：

表格之解釋　將下表加以檢查，卽足示明列表之方法。答語之大部，用想像之尺度，審定其位次，可謂已有十分確定，惟惜數字之答案，如何安插，反覺不甚明顯而已；然此則須視內證(internal evidence)決定之，否則卽須熟知該業之情形也。各分會之答覆，旣由各該分會之書記所塡報，然則各人所用之形容詞，

（第二十六表）

答語	分會數	會員數
絕無	4	140
未有	1	78
極少	23	4.836
微有少許	1	63
偶爾有之，但極少	1	350
對於修繕工作有一點	1	500
有一點	2	73
必要時有二小時	1	80
少有	1	59
有少許	1	16
除修繕工作外少有之	1	66
僅修繕工作有之	2	216
不多	6	1,125
修繕工作有之	1	500
並不多	3	644
並不甚多	2	162
不大有	1	7
不一定何時有	2	43
有損壞或緊急之時	7	606
例爲二小時	1	136
主要爲修繕工作	1	20
偶爾有之	2	90
必要時	1	348
臨時加工	2	142
修繕工作上甚多	1	23
四星期至多十八小時	1	1,000
中常	3	262
生意盛者習以爲常	1	200
平均一星期五小時	1	96
海洋船上頗夥	1	400
船塢工作習以爲常	1	650
大部有之	2	146
習以爲常	1	693
多量	1	263
甚多	1	72
極多	1	550
一星期九小時	1	39
一星期十小時	1	106
一星期十二小時（最高量）	1	700
一星期十四小時（忙時）	1	106
一星期十至十八小時	1	5,000
	88	20,666
未分類者：		
未答	36	5,114
少極	1	250
近來已不多	1	160
機械工廠中每年有六個月	1	60
鋼鐵工廠有之	1	348

在各人心目中，對於數目之意義，未必盡人皆同，然上表所列，所表示之情形，已盡淸晰之能事，故應用中位數及四分位數，亦頗適宜也。茲如以各分會之會員爲單位，除未曾分類者可以不計外，其中位數乃爲『四星期至多十八小時』，或『中常』，下四分位數乃爲『極少』，上四分位數乃爲『忙時每星期十四小時』。如以分會爲單位，中位數爲『不多』，下四分位數爲『極少』，上四分位數爲『必要時』，或『偶爾有之』。

此法結果之精度 (precision)，雖大有出入，但應用者極廣，蓋遇文字答語組成之兩種羣類，非用此法，不足以言比較也。然所謂精度，測量之法何如？此則須在列表時，將序列加以合理之變更，再求出其中位數；然後以此中位數與未變更前之中位數相比。而查二次中位數之位置有無移動，如有移動也，則移動之距離，即爲精度之測算根據矣。

總括敍述 現既有平均數方法，供吾人任意之應用，則請即用之將一數字羣類列成表格並爲總括之敍述。

試舉一例，請就一八八六年英國工商業蕭條問題調查委員所發問題之答案討論之。問題甚多，茲擇四則列下：

(1)入會工人人數；

(2)一八八五年失業者之人數；

(3)一八八五年之每週工資；

區　名 （一）	一八八五年各區人數 （二）	一八八五年失業人數 （三）	一八八五年現行工資 （四）	一八六五與一八八五年間工資之變動 （五）
白爾發斯特	1,100	130	二八先令至三六先令	略有增加
寇文特利	2,500	230	三一先令六便士	定合同之工作,減百分之五十
杜京費爾德	170+	20+	三一先令	略有增加
當第	1,400	45%	有技能工人二五先令 無技能工人一五先令	計時工資——一八六五年二二先令;一八七二年,二四先令;一八八〇年,二六先令;一八八三年,二四先令;一八八五年,二五先令;
哥拉斯告	28,000	4,000	二六先令	計時工資,比一八六四年高百分之五;
哥拉斯告（聖崟拉克斯）	1,600	250	××	一八七二至一八七三年增加百分之十五,一八八五年與一八六五年相同。
哈特爾浦	1,200	400	三一先令六便士	增三先令
哥老搜波	135	10	三二先令	×　×　×　×　×
利物浦	280	38	××	在一八七二至一八七三年，升起百分之七又二分之一，一八八五年與一八六五年相同。
蒙尼費斯	114	18	二一先令	有技能工人——一八六五年二四先令;一八七六年二七先令;一八七八年二五先令;一八八三年，二八先令;一八八五年,二五先令。
腦廷漢	4,000	600	最低三四先令	一八六五年,二八先令;一八八五年三四先令。
奧爾得漢木	1,600	96	平均三三先令	增百分之五。
牛津	45	××	三三先令	×　×　×　×　×
賠斯果	800	××	二八先令六便士	一八六五年二六先令；一八八五年二八先令六便士。
浦來斯頓	630	40	二八先令	無變動
浦來斯頓	900	120	二八先令	無變動
西浦果	201	15	二八先令六便士非會員工人，二四先令	一八六五年二六便士；一八六九至一八七三年三二先令;一八八五年;二八先令六便士。
播爾拜布里治	1,120	43	二八先令	一八六五至一八七五年二五先令六便士;一八七五至一八八五年二八先令。
桑得蘭	3,200	400	三三先令	一八六四年二七先令;一八七四年,三四先令;一八七五年至一八八五年,自三一至三七先令不等。
孫頓	6,050	2	三一先令六便士	×　×　×　×　×
餓爾維爾頓	45	××	三一先令	一八六五年二六先令；一八七五年三一先令。
溫斯布瑞	400	30	三〇先令	增二先令。
寧京頓	170	70	二八至三六先令不等	增百分之三十。

（第二十七表）

（4）一八六五年與一八八五年間工資之變動。

次將英國工程人員聯合會各分會之書記所給之答覆，列如上表。

將表格整理就緒擬成報告之時，須併入下列之提綱表。茲為舉例起見，下列之表，只列一分會之數字，其餘從略。

第一表　僱傭情形提要

分會會名	調查所得之各分會會員人數	失業人數	失業人數之百分數	各分會失業人數百分數之中位數
A.S.E.	55,170	7,142	13	12
O.S.B.				
其他從略				

（第二十八表）

第二表　現行工資

分會會名	各分會之平均工資		各分會工資之四分位數	離散度
	不加權	加權		
A.S.E.	先令 30　便士 0	先令 29　便士 7	先令 28　便士 0　先令 32　便士 0	$\frac{1}{15}$
O.S.B.				
其他從略				

（第二十九表）

第三表（甲）

一八六五至一八八五年間之工資變動

分會會名	各分會數				增加之百分數之中位數	各分會會員所佔之百分數			
	無答覆者	減低者	無變動者	增加者		無答覆者	減低者	無變動者	增加者
A.S.E.	4	1	5	13	10	11	4	6	79
O.S.B.									
其他從略									

（第三十表）

文字提綱——『就全部情形考察，絕大部分均在一八六五至一八八五年間有顯著之增加，約等於全部增高百分之十。因數字不完全確定，故不能求得確實之平均數』。

第三表乙，一八六五與在一八七三年左右最高額之工資變動情形。

第三表丙，一八七三年左右之最高額與一八八五年之工資變動情形。

（註一）　上列公式之計算，與吉尼氏原用之法，微有不同，請參閱吉尼氏同書第三十頁及二十九頁底註。

（註二）　本書原著者，於本書之前數版中，名此數量爲離散度 (dispersion)，以該數有永不大於 1 之便利在也。

（註三）　並請參閱第二編第一章第三節

第七章　圖示法

第一節　總論

平均數與圖式　在初級統計學中，有重要方法兩種，爲一般學生或辦理統計事務之官員，所不可不知，且無數學知識亦易瞭解，但竟爲一般所誤解，或被對統計無興趣或向不通統計之人士所忽略者，厥爲平均法與圖示法。此兩種方法所以相提並論者，乃以平均數及圖式之用途，有幾不可分之連繫在也。設遇有極大極繁之數字羣類當前，雖用表列法已爲明晰之表示，但吾人對於許多數字之整個情形，仍不能完全領略。任何一列數字——各城市之人口，各級年齡上之死亡率，許多工人之工資，若干年數之進口貨值——數列愈長，愈難令人索解。十個數目組成之序列，吾人或可一望而知，二十個數目之序列，雖稍費力，亦不難明瞭；但如一列印就之一百年數字，則欲其予吾人以何等之印象也殊難。一樹易見，成林之樹木則不易辨清矣。關於平均數用途上所有問題之試金石，即視所選用之平均數，是否能給予全部羣類以最佳之總結，使此一總結數字令人可以一目瞭然。在平均數一名詞意義大爲擴張時，即可知同時選用三，四乃至十個適當之數字，

悉能充分表現任一羣類之主要形態。平均數如此，圖示法亦然，圖示法之主要用途，乃在表現大羣類之數字，以便瞭解全部之情形，至一切圖式之是否完善，其試金石乃視所繪之圖，能否托出一組數字之最佳觀感，令人一望卽可領會。惟圖式尚有一用途，爲平均數所不及者，卽欲表現時間數列，唯圖式實優爲之。然就實質言之，圖式又不如平均數之重要，蓋平均數雖由若干數字推出，但離開數字亦能獨立存在，而代表所測量之數量之眞正形態。唯用平均數，乃可以此一羣類，與其他之羣類相比較，而圖式則不然，蓋圖式乃立於輔助之地位，非主要之元素，卽完全取消之，亦無不可，因用之乃在助目力之所不及，且爲閱者節省時間計也。

圖式法及平均數 爲使本章與前一章關係銜接起見，玆請同時用兩種方法，表示同一之數字羣類，卽以多數工人之工資爲例以討論之，資料如下：

工人之工資及人數

工資	人數	
自十五先令至十六先令	200	1000
自十六先令至十七先令	400	
自十七先令至十八先令	100	
自十八先令至十九先令	100	
自十九先令至二十先令	200	

自二十先令至二十一先令	200	
自二十一先令至二十二先令	300	
自二十二先令至二十三先令	300	2,200
自二十三先令至二十四先令	500	
自二十四先令至二十五先令	900	
自二十五先令至二十六先令	1,200	
自二十六先令至二十七先令	800	
自二十七先令至二十八先令	700	3,500
自二十八先令至二十九先令	500	
自二十九先令至三十先令	300	
自三十先令至三十一先令	300	
自三十一先令至三十二先令	400	
自三十二先令至三十三先令	400	2,100
自三十三先令至三十四先令	500	
自三十四先令至三十五先令	500	
自三十五先令至三十六先令	600	
自三十六先令至三十七先令	400	
自三十七先令至三十八先令	100	1,200
自三十八先令至三十九先令	80	
自三十九先令至四十先令	20	

如用平均法，上列羣類可以下列數字代之：

全部之平均數　　二十七先令六便士

最低1000人之平均數一十七先令

最高1000人之平均數三十六先令六便士

中等4000人之平均數二十七先令

或

中位數	二十六先令九便士
四分位數	二十四先令二便士
	三十二先令
十分位數	二十先令
	二十三先令六便士
	二十四先令九便士
	二十五先令八便士
	二十六先令九便士
	二十八先令二便士
	三十一先令
	三十三先令四便士
	三十五先令四便士
衆數	二十五先令三便士
次要衆數	十六先令六便士
	三十六先令

或

工資	人數對全部之百分數
自十五先令至二十先令	10

自二十先令至二十五先令	22
自二十五先令至三十先令	35
自三十先令至三十五先令	21
自三十五先令至四十先令	12

簡單圖式之構造法 此一羣類，以圖式表列之，如第二圖。此為用圖表示兩變數相互關係之一例。與此相類之圖式，可用以表示各級年齡中之結婚率或死亡率，表示身長各異之人數，在各種價格上之需要，或用以表示其他任何同質數量之羣類。卽以此同一之構圖法，表示若干年中任何數值之變動，亦無不可。構圖之法，須在紙之底邊，畫一平行線，在此線上分成相等之距離，以代表依次增加之若干數量，諸如年齡，所得，身長，價格，時間以及其他等等。此線名曰橫軸 (axis of abscissa)，線上某一點至零點之距離，謂之該點之橫坐標 (abscissa)。在紙之側邊，經過零點再畫一線，與橫軸成為直角，則此垂直之線，卽名縱軸 (xis of ordinate)。在此縱軸上，亦分成等距，依次代表含有橫軸上所代表數量之人數或件數。又於橫軸上分別在各點各畫一垂直之線，標明在該一點上有若干人數或件數，此線卽為縱坐標 (ordinate)。請閱下列一圖，橫坐標所表示者，為工資數，縱坐標代表賺各級工資之人數。橫坐標縱坐標既已判明，然後卽在各縱坐標之上端，各連一直線，於是圖式乃告完成。實際上，如用方格

紙，則可不用畫縱坐標，而縱坐標之頂端，亦可標出。

工資圖之解釋 由此一圖，令人一望即知，賃銀勞動者依所賺工資分配之情形。工資在十五至十六先令者人數甚少，在十六至十七先令者則漸多，但在十七至十九先令者人數又大爲減少。自十九先令起，人數依次連續上升；至二十四先令與二十七先令之間，人數乃登峯而造極，其中尤以在二十五先令至二十六先令者爲最多。但自是而後，則每況愈下，直至三十先令一組而後已。此三十先令以前一組人數雖過少，而與十七至十九先令之間者相較，尙非過低。降至三十先令之後，復又以有規則之進展，漸次上升，至三十六先令後，乃又急轉直下，而達於三十九先令一組之最低數。由此觀之，在二十五先令左右集合之大部，乃爲全類之主體，而三十六先令之一組則屬其次，但在十六先令之處，則爲極小而幾別成一部。凡此種種，示與吾人，在三十至四十先令間，有高度技能之工人甚多，至有平常技能之工人大衆，則其工資乃在二十先令至三十先令之間，但有極少之粗笨無技能工人，工資均在十六先令左右。用一圖式形態畢現，如用表列豈可得哉！

於此吾人須加注意者，表上在十五先令至十六先令間之人數，正由在十五先令六便士（該組之中點）之縱坐標表示之。如表列所根據之原來數字，能取至最近之一便士爲止，則縱坐標應畫在十五先令五又二分之一便士處。各組之中點位置，畫時須準

確安置，此一要事也。

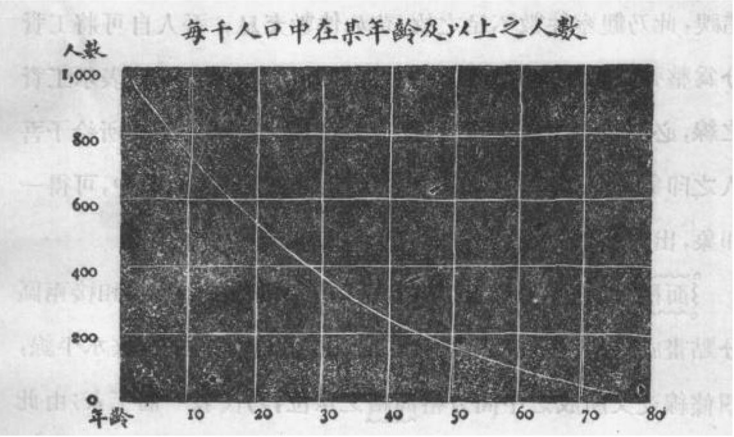
連續性　連結各縱坐標頂端之線，用途有二。第一為便於判斷各相對高度 (relative height)；第二可表示連續性 (continuity)。第一點姑不具論，茲就第二點作圖舉例解明於下：

每千人口中之年齡分配

（根據一八九一年之人口普查）

年齡	人數	年齡	人數
自零歲以上	1000	自十七歲以上	607
自一歲以上	973	自十八歲以上	587
自二歲以上	949	自十九歲以上	567
自三歲以上	925	自二十歲以上	547
自四歲以上	901	自二十一歲以上	528
自五歲以上	877	自二十二歲以上	510
自六歲以上	854	自二十三歲以上	491
自七歲以上	830	自二十四歲以上	474
自八歲以上	807	自二十五歲以上	456
自九歲以上	783	自二十六歲以上	439
自十歲以上	760	自二十七歲以上	423
自十一歲以上	738	自二十八歲以上	407
自十二歲以上	715	自二十九歲以上	391
自十三歲以上	693	自三十歲以上	376
自十四歲以上	671	自三十一歲以上	361
自十五歲以上	649	自三十二歲以上	346
自十六歲以上	628	自三十三歲以上	332

年齡	人數	年齡	人數
自三十四歲以上	318	自五十八歲以上	85
自三十五歲以上	305	自五十九歲以上	79
自三十六歲以上	292	自六十歲以上	73
自三十七歲以上	280	自六十一歲以上	67
自三十八歲以上	268	自六十二歲以上	62
自三十九歲以上	256	自六十三歲以上	57
自四十歲以上	244	自六十四歲以上	52
自四十一歲以上	233	自六十五歲以上	47
自四十二歲以上	222	自六十六歲以上	43
自四十三歲以上	211	自六十七歲以上	38
自四十四歲以上	201	自六十八歲以上	34
自四十五歲以上	191	自六十九歲以上	31
自四十六歲以上	180	自七十歲以上	27
自四十七歲以上	171	自七十一歲以上	24
自四十八歲以上	161	自七十二歲以上	21
自四十九歲以上	152	自七十三歲以上	18
自五十歲以上	143	自七十四歲以上	15
自五十一歲以上	135	自七十五歲以上	13
自五十二歲以上	127	自七十六歲以上	11
自五十三歲以上	119	自七十七歲以上	9
自五十四歲以上	112	自七十八歲以上	8
自五十五歲以上	104	自七十九歲以上	6
自五十六歲以上	98	自八十歲以上	5
自五十七歲以上	91		



（第　三　圖）

上圖，橫軸代表年齡，縱軸代表一八九一年夏末英格蘭及威爾士人口每千人中尙存在及在各級年齡以上之人數估計數。縱坐標乃由代表每歲年齡之中點之橫軸量起；但壽命之長短，不能照幾年幾月幾日計算也。此圖之本意，即爲表明每一整歲尙存人口之比例，爲達此目的計，縱坐標之頂點所連成之線，必須不致出現破斷及折角，而應有絕對之連續性也。

實際上，不滿一整歲之較小年齡組距，無標出其縱坐標之必要，恐令人一見不能領會其詳細情節也。然上圖所畫之線，亦確與人口數爲無限大而年齡分組無限小者所呈現之線之形式相同。

就此所論各點，請考察前繪之工資統計圖（第二圖）。某一年之平均工資收入，計數時未必每一先令，甚至每一便士，均無

錯誤,此乃觀察件數不足之故,苟如件數充足,吾人自可將工資分爲整齊之順序,而以微至一法尋(farthing)爲準,則表示工資之線,必不致有尖銳之折角,而成爲連續之曲線。圖式所給予吾人之印象,卽爲連續性之存在。又吾人於下列第六圖中,可得一印象,出口貨值之線,亦確爲日接一日連續不斷也。

面積 經過一顯然之階段,吾人頗可假定,由橫軸相接兩區分點畫成之兩條垂直線,與由縱軸相接兩點畫成之兩條水平線,四條線交叉所成之中間方格面積之單位,乃代表一個工人;由此假定,可知上以曲線爲界,下以基線爲止,以左右兩條經過代表任何兩種工資額之兩點所成之垂直線爲範圍,所畫成之面積,卽爲工資介乎底線上所代表之該兩種數額之工人總數。

故在第二圖中,經過中位數,M,四分位數,Q_1,Q_3,十分位數,D_1,D_2,D_3,D_4,D_5,D_6,D_7,D_8,D_9,等點所繪成之線,卽將面積 $ABm_1m_2m_3CD$,分別分爲二,四,及十個等面積。該圖之重心,乃在於經過平均工資,V所畫成之垂直線上;至經過最高位置m_1,m_2,m_3 三點之縱坐標,在基線上之基點,卽爲三個衆數之所在。

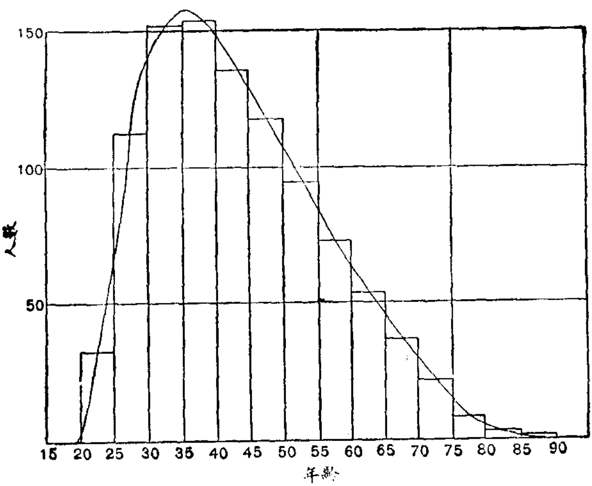
分組資料 在資料之分組甚寬時,吾人以用下列方柱圖(block diagram)之法爲宜。

一九一一年英格蘭及威爾士已婚男子之年齡分配

甲　方柱圖——每千人中每五歲一組之人數

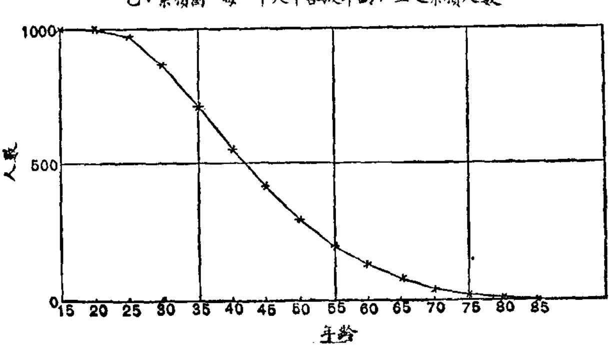
第四圖　一九一一年英格蘭及威爾士已婚男子之年齡分配

甲、方柱圖—每千人中每五歲一組之人數



(第四圖甲)

乙、累積圖—每一千人中各級年齡以上之累積人數



(第四圖乙)

上列之甲，乙，二圖，乃爲前表第十八表已婚男子人數依年齡分配之例釋。在第十八表中，吾人除只知其年在二十歲而不滿二十五歲之比例外，並無別項消息可得。但今用圖式，以代表五歲之組距爲底，以與該組記錄上人數成比例之數爲高，乃成一長方形，所有前表未能宣示之事實，茲均已有精確之表示。根據前第二圖之指示，各組人數乃均在各該組之中心。如就年齡之圖式而論，各歲人數理應連續不斷，於是如完全表示之，卽爲連續曲線，吾人已熟知之矣，然則以五歲組距爲底之面積，必與各該長方形之面積相等。此種曲線雖爲用隨手畫法，在圖上繪成，但若全線之位置，並無動搖不定之處所，則此曲線必足爲事實之代表。

資料亦可用圖乙代表之，在圖乙中，十字號卽爲表上記錄之資料。將此若干十字號各連成直線，若事實現象果爲連續的，則結果所得之線，與曲線乃無二致，卽在本圖之曲線，亦殊難與各點所連成之直線辨別也。

必需之確度 繪圖之詳細技術，標尺尺度之位置，使圖式明晰之辦法，等項可於本章所舉各圖示明之。至於數字究以萬爲單位，以千爲單位，或以一個爲單位，其確實之程度，乃純以目力能以領會爲決定之標準。在本章所舉之圖例，所表示之數字，設在一千個中，有一易其位置，卽判然可見，此乃平常之限度也。若進

而欲求更高之確度，誠亦非事實所需要，蓋圖式不能離表列而獨立，其功能亦僅爲補目力之所不及，特別顯示數列之重要形態而已。

目力　在未討論選擇代表數字之尺度前，圖式必如何乃可予吾人以明顯之印象，應有加以研究之必要。夫吾人之肉眼，所能判斷者有三：一，距離；二，比率；三，角度是也。玆以第六圖之虛線爲例，討論距離，比率及角度三端。

（一）肉眼對於距離之長短，可予以頗爲安全之判斷；兩點對於基線之距離，何者較遠，何者爲近，一見卽知，幾無可疑之點；若用方格紙，則雖有千分之一之差，亦能明察無遺。且眼目對於差額之判斷，最爲神速；例如圖中一八八三年出口貨值超過一八八五年出口貨值之數，較之一八九〇年出口超過一八八三年者爲尤巨，一望可知也。

（二）一八八九年之出口貨值，幾爲一八六二年之兩倍，或一八七八年之出口貨值，約當一八九〇年之四分之三；均不難一望而知。但只恃目力，其測量所得之確度，並不甚高；一八七三年出口貨值與一八七一年出口貨值之比(1.095:1)，較一八八二年出口貨值與一八八〇年出口貨值之比(1.073:1)爲大，實難用目力覺察之；然此並非目力不能判斷之證明，反之，該圖所給予吾人之印象，一部乃由此種性質無意識之計算而完成也。至如欲

得確實之觀察,須用本章第五節所述之方法。於此,吾人須注意,爲表現此等觀察,必須插進基線;且因目力之計算,本爲無意識者,若以一圖表示若干年之運動形勢,而無基線,則其將給予吾人以錯誤之印象也無疑。

(三)設吾人質問:一八八六至一八八七年增加之量大耶?抑一八八七至一八八八年之增額大耶?則欲得迅速之答覆,與其查視二年度之差額,無寧觀察二者上升之角度。後一年度之變動,上升之線,較前一年度上升之線爲峻峭(與水平線成一較大之角);故後一年之變動,必較前一年度者爲大,實際上,後一年度增加有一千二百六十萬鎊之多,而前一年度只增加九百二十萬鎊也。此種目力觀察最有效之練習,莫過於判斷增加率發生變動之年度;例如,一八六二至一八六三年,出口貨值旣有增進,一八六三至一八六四年,增加漸緩,一八六四至一八六五年漲勢仍輕,在一八六五至一八六六年,則驟然上騰;自一八六六至一八六七年,漸漸下降後,復又以加速度接續上升,直至一八七一年而後已,其他年度依此類推,姑不多述。一八七二年至一八七六年之線,與基線成凹面(concave),乃爲急轉直下之形勢;一八七九至一八八二年之凹形,又緩緩上升。總之,如此所示之增加,乃絕對或實際之高漲,並非與各該年度之最初數量成比率而計算也。

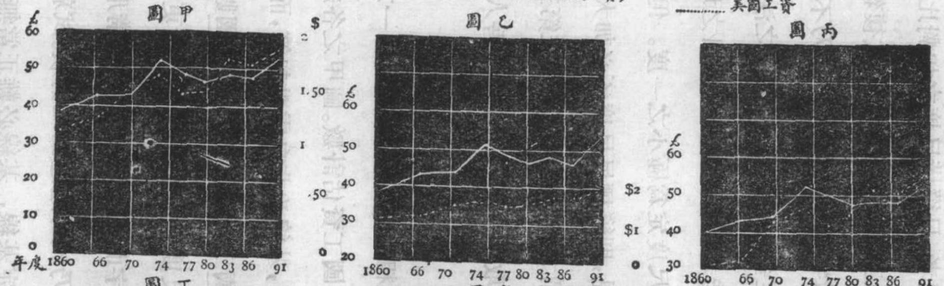
尺度之選擇　尺度爲定數字位置所必需，究應如何選擇之，殊難規定準則。吾人所欲研究者，只縱橫尺度之相對比率一問題。全圖不可過大，以一望可窺及其全豹爲宜；如圖式複雜，所述者爲極多年代，及極多之變動數字，則纖細之確度，乃不得不付之犧牲。設橫標尺尺度，業已定妥，選擇縱標尺尺度之法，必須使一代表增加率最大之線段，能表現對縱軸十分傾斜之形勢爲度，而欲其如此，唯有將縱標尺尺度定爲極小之一度。但若欲使所有數字重要之異動，顯然可見，則縱標尺尺度不必加大。唯能適合此兩種條件之尺度，乃可謂之適得其當。下列之第五圖，例示由故意操縱尺度之大小，及取消基線，所能給予吾人之錯誤印象。

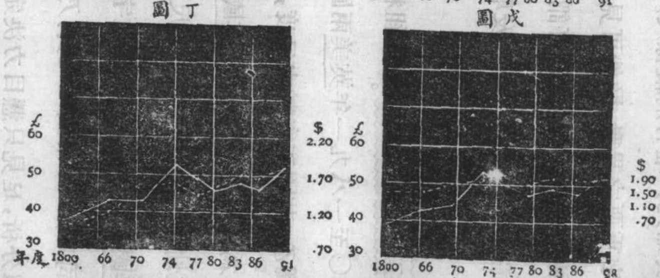
正當底線之必要　此組粗略之圖式，均表示同一之數字：一八六〇至一八九一年英美兩國工資估計數。圖甲之各線，尙爲正當。圖乙，英國工資數之尺度，基線並非自零點起算，而美國數字亦大爲縮減；其結果乃成英國工資變動甚大，而美國數字略有增加之情形。圖丙，丁，戊三圖，尺度大加改竄，基線亦行更易，故圖丙之美國工資，有超出英國數字而上升之勢，而圖戊情形又恰形相反，同時圖丁，兩國之工資，乃成等速運動之情狀。試就以下各圖，加以分析，足見只憑目力決定基線，基線必難正當無誤，若取消基線，又必不能據以估計變動之重要性。又除少數之例外容當提

第五圖　各種尺度及錯誤之基線表示同一數字

(下列各圖左方尺度代表英國工資，右方表示美國工資)

—— 英國工資
······ 美國工資





工資平均

年度	英國 每年	美國 每日
1860	£38.8	$1.10
1866	43.2	1.22
1870	43.9	1.50
1874	52.4	1.60
1877	49.8	1.48
1880	47.6	1.57
1883	48.8	1.75
1886	48.0	1.71
1891	53.2	1.85

及者（註一）外，凡數字圖式因取消基線關係而縮小空間者，其所代表之情形，實難令人信任。

修勻曲線　論及修勻曲線(smoothing curve)，不能不以吉芬爵士(Sir R. Giffen)在英國皇家統計學會，宣讀論文中，所提出之『英國出口貿易之靜止狀態』一問題，爲最妙之例證。

下列第六圖之輕虛線，表示各年出口之貨值，由此線吾人得來之印象，可知最近數年，出口貨值並未增長。但吉芬爵士則提出如下數字：

平均每年出口貨值

1855－57	134,000,000鎊
1865－67	228,000,000鎊
1875－77	264,000,000鎊
1885－87	274,000,000鎊
1895－97	292,000,000鎊

於是彼乃依據此列數字，發生結論謂：各年數字均有增加，惟唯一足爲表示其靜止狀態者，厥爲後半期中增加率，較前半期之增加爲低一點。

英國『星期六評論』(Saturday Review)，因見其所選之數字只爲每十年選出三年之平均數，僅爲一種情形之巧合，乃著文論之曰：『(此一結論實爲極大之誤解)；否則，何以不將一八九八年計入耶？』吾人見此評語，若查閱數字，殊覺無從答覆，但試閱

第六圖，則全部情狀，一覽無餘矣。第六圖上自一八六五年之後，曲線呈現三大波浪式。一八七二年，因物價膨脹，致該年出口價值之最高點極高，但一八九〇年之最高點尤大，前此各年無與匹敵者，而一八八二年之最高額，相形之下，則反較甚低。至於各年之最低點，從圖式觀之，則全部均有增加之勢；一八六八，一八七九及一八八六年之最低點，三年形成有規則之增進，迨一八九一年，忽又一瀉而下。一八九四至一八九六年間，似有另一每十年一循環之現象出現，不期至一八九七年中途又生變化。自此以後，以至一九〇六年，各年累有增加，是又超過一八七二年以前而愈趨上昇矣。

『星期六評論』復進而質問，何以吉芬爵士，不用有如下列每五年一平均之數字：

平均每年出口貨值

1870—74	235,000,000鎊
1880—84	234,000,000鎊
1890—94	234,000,000鎊
1898	233,000,000鎊

而此列數字，恰與吉芬氏所提出之數，大相反對也。

由此觀之，吾人必須採用某種通行方法，使表示數字之形式，不致因擇用特別年度而受影響。玆將全部數字及圖式列下：

據公布英國國產貨物出口眞實值

以一百萬鎊爲單位

		平均數					平均數		
		三年一平均	五年一平均	十年一平均			三年一平均	五年一平均	十年一平均
1855	95.7	...	...	...	1881	234.0	216.2	208.2	221.6
1856	115.8	...	...	...	1882	241.5	232.9	216.7	220.1
1857	122.0	111.2	...	...	1883	239.8	238.4	226.0	218.6
1858	116.6	118.1	...	...	1884	233.0	238.1	234.3	217.9
1859	130.4	123.0	116.1	...	1885	213.1	228.6	232.3	216.9
1860	135.9	127.6	124.1	...	1886	212.7	219.6	228.0	218.1
1861	125.1	130.5	126.0	...	1887	221.9	215.6	224.1	220.4
1862	124.0	128.3	126.4	...	1888	234.5	223.0	223.0	224.5
1863	146.5	131.9	132.4	...	1889	248.9	235.1	226.2	230.2
1864	160.4	143.7	138.4	127.2	1890	263.5	249.0	236.3	234.2
1865	165.8	157.6	144.4	134.3	1891	247.2	253.2	243.2	235.5
1866	188.9	171.7	157.2	141.6	1892	227.1	245.9	244.2	234.1
1867	181.0	178.6	168.7	147.5	1893	218.1	230.8	240.9	231.9
1868	179.7	183.2	175.1	153.8	1894	215.8	220.3	234.3	230.2
1869	190.0	183.6	181.0	159.8	1895	225.9	219.9	226.8	231.4
1870	199.6	189.8	187.8	165.9	1896	240.1	227.3	225.4	234.1
1871	223.1	204.2	194.6	175.7	1897	234.3	233.4	226.8	235.4
1872	256.3	226.3	209.7	188.9	1898	233.4	235.9	229.8	235.3
1873	255.2	244.9	224.8	200.0	1899	255.3*	241.0	237.8	236.1
1874	239.6	250.4	234.7	207.9	1900	283.6*	257.4	249.3	238.1
1875	223.5	239.4	239.6	213.7	1901	270.9*	269.9	255.5	240.5
1876	200.6	221.0	235.1	214.9	1902	277.7*	277.4	264.2	245.5
1877	198.9	207.7	223.7	216.7	1903	286.5*	278.4	274.8	252.3
1878	192.8	197.4	210.9	218.0	1904	296.3*	286.8	283.0	260.4
1879	191.5	194.4	201.4	218.1	1905	324.4*	302.4	291.2	270.2
1880	223.1	202.5	201.3	220.5	1906	367.0*	329.2	310.4	282.9

（註二）　　　　　　（第三十一表）

上圖之輕連續線，表示每三年一平均之各年數字，換言之，卽每年與其前後二年數字之平均，其位置幾與每年數值之虛線相合；此線僅將虛線之折角修勻，對於全部形態，並未波及。十字之線代表每五年一平均之每年數值，卽以其前兩年與後兩年之數值合併加以平均也。圓圈之線，表示每十年一平均之每年數值，每一圓圈所在，卽爲一期間之中心，代表該期間之平均數，例如代表一八七五至一八八四年十年間平均數之圓圈，卽在一八七五與一八八〇之縱分界線上。（註三）

期間之選擇 試觀每五年平均數之曲線，可知『星期六評論』駁論吉芬爵士之點，不爲無見，蓋吉芬氏亦確僅就利於彼之論辯之年份，而取作資料也。何則？所取用之每五年之平均數，恰與一八九八年相反，盡在波形之上半部，代表此種平均數之各符號，均與每五年一平均之線之最高值相近，但一八九八年之數，則非爲最高數也。玆請以或多或少之確度，選用年份如下：

出口貨值每五年之平均數

1865—69	181,000,000鎊
1875—79	201,000,000鎊
1885—89	226,000,000鎊
1898	233,000,000鎊

由此數字從而發言，謂一八九八年之出口貨值，較之所選以前之任何平均數均高。無須用武斷方法選擇相當期間，亦能得事實之證明矣。由此觀之，若欲作一論斷，必須注意商業循環，否則

任何理論,均無成立之餘地。此種週期循環,必至用每十年之平均數後,方能剔除。

圖上特別符號,乃爲每十年之平均數,由此數點,可見在一八七〇年以前,增加之勢甚速,自一八七〇年以後,則增進甚緩,但後又繼以飛躍之擴張。全線所表示之全部情態,與每十年平均數之線所示者相同。

最後,設吾人在可能範圍內,用隨手畫法,儘量將全線修勻,而對於曲折過甚之線段,並不厲行纂改,則修勻後之曲線,仍可代表同一之情況,如上圖之重線,即此是也。修勻時,事先有一假定,認商業循環乃爲十年一期;今在此假定之下,發見在同一之十年期中,竟有兩個最高值者,故圖中此期之平均數,本應爲該期之最低點,現則呈現最高值之地位如一八八七年矣。此一現象,若欲消除之,只有變換平均期間之法,以適應此變動之週期長度(wave length),而所謂變動之週期長度者,乃一頗爲武斷之前提也。然此一困難,解決之法,本不甚難,只須隨吾人之所見加以修正而已,如圖中最後已修勻之重線,即足傳示此正當之印象也。

自一八九九年之五年,去一八九八年曾幾何時,而高下相去懸殊,殊出人意料之外;惟嗣後之又一五年數字,未能先事判斷,故不知次一週期之波形,當吾人將此又一五年之數字加入計算時,可知每十年(如1890–1899,1891–1900……)之平均數,又

達最高之記錄，而自一九〇〇年起至一九〇六年止，各年之數字，均較以往之最高值爲大，亦從可知矣。然則，吉芬爵士所言『各年數字均有增加，惟唯一足爲表示其靜止狀態者，厥爲後半期增加率，較前半期之增加爲低一點』一言，業已完全實現矣。

修勻曲線之意義及趨勢 上圖之修勻曲線，在將偶然及暫時變動(accidental and temporary variations)剔除之後，則所表示者，即爲歷年出口貨值之一般趨勢。從長期運動中，完全剔出短期之變化，即從繼長增高之商業大流中，將商業上潮水之漲落，另行剔出，是否可能，現尚不得而知，如其果爲可能也，則吾人必可求得用此一修勻曲線代表之趨勢矣。蓋較小數字歷年之升降，乃現出變化不規則之曲線，今若將此曲線修勻，則在修勻之線中，即無突然之變動，即增加率亦將以穩健之步驟上升也。

討論至此，試添加以下各年之數字：

	各年平均數	三年一平均	五年一平均	十年一平均
一九〇七 － － －	416.0*	369.1	338.0	301.1
一九〇八 － － －	366.5*	383.2	354.0	314.4
一九〇九 － － －	372.3*	384.9	369.2	326.1
一九一〇 － － －	421 6*	386.8	388.7	339.9
一九一一 － － －	448.5*	414 1	405.0	356.7
一九一二 － － －	480.2*	450.1	417.8	377.9
一九一三 － － －	514.2*	481.0	427.4	400.7

（註四）

至歐戰時之記錄，不便與其他平時之數值相比，故自一九一四後之數字，付之闕如。讀者在查閱上列一九〇七至一九一三年

之數字前，可請先就前列之圖式(第六圖)，加以預測，察其數值，變化及一般趨勢若何，然後再參看實在數值。

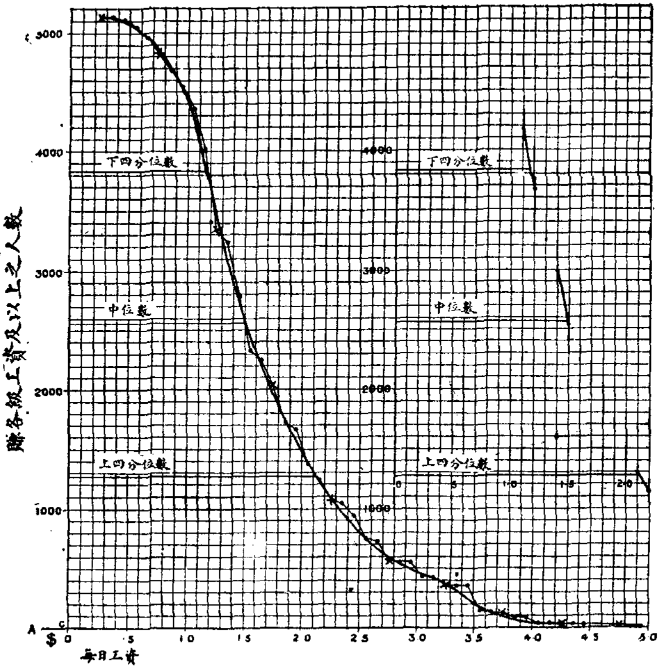
修勻曲線在某一期日行進之方向，謂之數列(series)在該期日之趨勢(trend)。當修勻曲線經過若干年份而仍大概成爲直線形式時，則該曲線行進之方向，卽表示在彼一期間中之趨勢。

對於趨勢之求法，近年來摩爾(Moore)教授，曾創用一特別方法(見一九一九年號統計學報第三七五頁)。彼之方法，乃假定若干年份之一般趨勢，可表之爲下述之方程式：$y=a+bx+cx^2+dx^3$，並以『如某一期日(x_t)之觀察值爲y_t，則$\Sigma(y_t-a-bx_t-cx_t^2-dx_t^3)^2$應爲最小』爲條件，而求 a,b,c 及 d 之值。但貝孫斯(Persons)教授之假定(見美國哈佛大學經濟統計評論創刊號)，則以爲甚爲確實而能使 $\Sigma(y_t-a-bx_t)^2$成爲最小者，乃爲一條直線。二人方法有別，何一可供一般之應用，尙成疑問。惟貝孫斯氏之假定，應用時務須酌量情形，不可普遍實施。實則上文所用之移動平均(moving averages)法，設所論年份甚長，而支配現象之總原因，又屢有顯著之變化時，則此法對於趨勢方向變動之表現，確較他法爲敏銳。

同質羣類之修勻　『修勻』數列之詳細研究，乃屬於第十章內插補法之範圍，不在本章討論之列，惟另有一羣類此時不妨加以考慮者，卽用圖示法從不規則之原始資料中，求出有規則之

形態也。試就前列第十三表之美國工資統計而論，吾人頗可將此五千人之工資，用圖表示之如下：

第七圖　決定中位數與衆數之圖示法



圖中之縱坐標，代表在賺某定額工資或以上之人數。成有尖角之輕線，亦代表工人人數，惟係按一角分組之各組人數登入者。此種圖示法，對於類似此一工資羣類之不規則數列，頗有特殊之效用，因工資自最高額起至最低額止，人數由少而多，工資在某定額以上人數之曲線，自亦由低而高。於是第五章第五節所論求

中位數之圖示法，亦可由上圖實施之矣。

求中位數之圖示法　上圖輕線所呈現之不規則狀態，並非因工資分類之任何法則而起，而係出乎觀察之意外事故而來。設所得之資料，乃認爲係由極大之羣類中抽樣而成，則吾人可以假定：如作挨戶調查，則完全之資料，作成之形式，必與此修勻曲線相去不遠。修勻之法，可用隨手畫法，畫一條距各點愈近愈佳之線，不可現出特別破裂曲折之角，則形式卽如上圖所示。基於上述理由，可用修勻曲線求中位數，四分位數，十分位數乃至百分位數之近似值，求算之法，須在縱軸代表人數二分之一，四分之一，四分之三，……等處，各畫一條水平線，以達於修勻曲線，然後再由各水平線與修勻曲線交叉之各點，向下各畫一條垂直線，以至橫軸爲止，則在橫軸所代表之尺度上，卽得工資中位數，四分位數，以及十，百分位數之工資。

茲將得數列下：

	中位數	下分位數	上四分位數
第十四表（第四章等四節）求得者	$1.49	....	....
用第五章第五節之圖示法，經在上圖求得者	$1.49	$1.16	$2.12
由上圖之修勻曲線求得者	$1.51	$1.15	$2.13
用第十章第二節代數內插補法求得者	$1.536		

雖然，此法所得之結果，精度並不甚大；修勻曲線微有彎曲，則其所得數之差額，必較上表之第二與第三行相差爲尤大。

求衆數之圖示法 用此法求衆數，亦能求出其近似値。惟此法有兩種困難，前已提及，吾人當能記憶，一、衆數之兩邊分配不均，二、表列變更則總數位置亦移。第一種困難，如用圖示法，卽可完全消滅，第二種困難，因衆數位置之移動，全視修勻線之彎曲是否有輕微之變動爲斷，故困難亦可減少。夫衆數之位置，乃爲最大人數之所在，其理至明，吾人已知之，惟今用圖示方法，則衆數者卽波折——或修勻——線最峻峭之處也。在修勻曲線上，最峻峭之處所，卽爲切線 (tangent) 經過曲線之點，在數學上所謂轉向點(point of inflexion) 者卽是。用機械求此點，可用一尺置於圖上與曲線接觸，然後將尺繞曲線轉動，至尺與曲線相交時，則相交之點，卽爲衆數所在。如上圖之一元一角至一元四角一組所出現者卽是。除此以外，尙有較爲繁複之求中位數與衆數法，容當於第十章第二節討論之。

求中位數與衆數而用圖示法，最大之優點有二：一、用於不整齊之人數（例如，工資在三十先令六便士者三十人，三十先令八便士又二分之一者四十人，四十先令一便士者三十五人……等）時，與用於整齊之資料，其造圖之難易相同，而確度亦相同；且如修勻曲線造圖精密，則衆數之個數，一望卽知，並可估計各

個乘數之相對重要性也。即如上圖，基線上自三角起至一元二角止，曲線成凹形；一元二角至三元一角五分，曲線成凸形，其後至三元四角為止，又現凹形，復又繼以凸形，以至於終端而後已。轉向點——或云乘數——即凹形轉為凸形之點。故此處有乘數二，而其中之一，臨近三元四角處者，乃次要之乘數也。

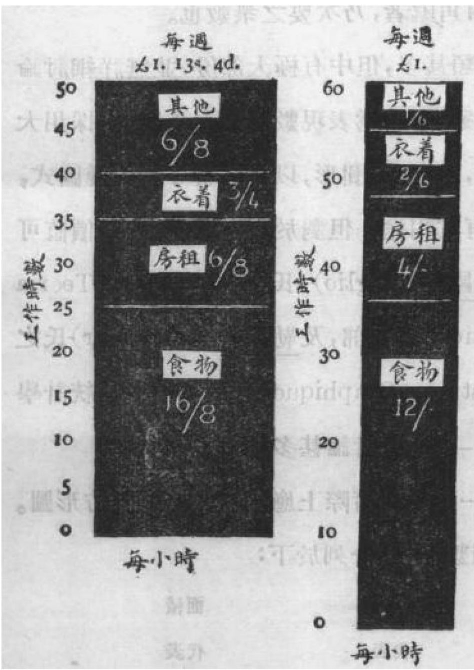
象形圖　圖式之種類甚多，但中有極大部份，並無詳細討論之必要。著作家及講演家，有時為表現數目之大小起見，採用大小不同之點，線，三角形，四方形，圓形，以及圖畫等。此種圖式，對於演說及手册，本各有其用途，但對於數字意義，則無價值可言。此種圖式，哥巴戈理歐(Gabaglio) 氏之統計學原理(Teoria Generale della Statistica)第二部，及勒瓦舍(Levasseur)氏之圖式統計學 (La Statistique Graphique 見於英國皇家統計學會之五十週年紀念刊）一書中，討論甚多，可以參閱也。

此種圖式之中，有一類足為實際上應用者，厥為長方形圖。長方形可用以代表三種數量，玆分列於下：

	側邊代表	底邊代表	面積代表
1.	物價	物量	物值
2.	房間數	每間住人數	人口總數
3.	每週工作時間	每小時平均產量 每小時工資	總產量 每週工資

茲舉一例，以示此法應用之範圍：

用長方形代表三件事實　下圖為一技師與一勞工之家庭預算，分別表示每週各種消費所用之金額，與每一消費金額所需之工作時間。



橫標尺代表每小時便士數

0.125英寸等於一便士

縱標尺代表每週工作時數

0.1英寸等於二小時

面積代表消費金額，

全長方形代表每週工資數

一平方英寸等於十三

先令四便士

（第　八　圖）

英國之圖示法標準聯席委員會(Joint Committee on Standard for Graphic Representation)，自一九一六年後，曾議決

統計圖示之最佳方法，並爲謀圖式之統一，免除錯誤起見，特提供若干有用之建議。

統計地圖　統計地圖之用途，現可簡略討論之。關於人口之數字，不論其爲人口之密度，抑人口之平均所得數，或平均負擔租稅額，均可分區用適當之記號或彩色標示之。在此種方法之中，普通最便於應用者，厥爲以一種彩色（例如藍）代表大於平均數之區域，另以他色（例如紅）代表小於平均數之區域。將區域分成九類，大於平均數者，分爲百分之七以上，百分之五至七，百分之三至五，百分之一至三幾種，以深淺不同之藍色表示之，距平均數最遠者，藍色最深，愈近於平均數其色愈淺。在平均數上下百分之一以內則用白色。在平均數以下者，分別以深淺之紅色表示之，去平均數愈遠色亦愈濃。但用此法時，深淺度之分級，不可過多，此吾人不可不注意者。應用統計地圖之書，舉例言之，有布斯氏之『人民之生活與勞動』附圖，有美國第十一屆人口普查之統計地圖，有『印度統計地圖』，上文所舉之『勒瓦舍氏圖式統計學』一文中，亦附有地圖。勒瓦舍氏之文中，有一省費而極有效之方法，乃僅用黑白兩色，其結果亦甚佳，又白狄雍氏之『統計學初級教本』，第一百三十三頁後圖式法一章，亦有同樣之介紹焉。

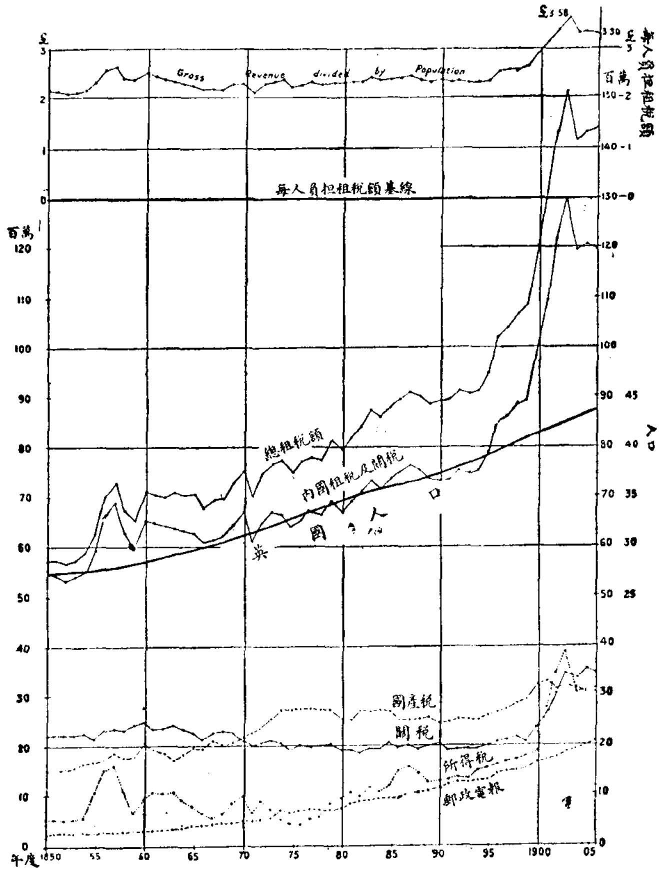
雖然，地圖法自有其缺點，蓋地圖法之記錄，乃以行政分區

爲準，但所欲表示之現象，則殊與此無關。茲舉例以明之。假如一九一一年，英格蘭之各部，用染色方法，標明人口之密度，則可伯蘭(Cumberland)之色度，應大略等於每百英畝二十七人，而腦贊伯蘭(Northumberland)，爲每百英畝五十三人。但若遇有荒漠之處，人烟稀少，則人口密度，用彩色表示之，則若干英里之內必將一片荒涼。故補救之法，須用下列二種途徑：一、略分區域，可以一教區作一單位，各教區均僅用黑色煊染之，色之深淺一隨人口之多寡。二、用點標記，視資料之許可，儘量求其準確，各點大小應完全相等，一點代表一百人，但遇人口稠密之區域，則應加以變更，此類之統計地圖，在塞可利斯特(Secrist)教授所著之統計方法(Statistical Method)一書中(一九一七年版，第一八九頁)，曾加複述。

第二節　歷史圖

數字之比較　圖式之主要用途，大致不外供人瀏覽兩序列事項之相互關係。此中情節，以舉例釋明之爲最佳。最簡單之一種，厥爲比較同以一單位(例如金鎊)表示之兩列數字，而就中尤爲簡單者，莫過於以全體與一部數字作比較。

以租稅爲例　下圖中部第一條線表示英國各年總租稅額(見統計輯要 Statistical Abstract 一九〇六號)；(註五)第二



（第　九　圖）

條線爲內國租稅及關稅，此項與總租稅額相差之數，主要卽爲郵政收入。租稅之內，以關稅，國產稅(excise)，所得稅，及郵政收入爲大宗。玆各以曲線將其各年之數字，分別列如上圖之下部，各

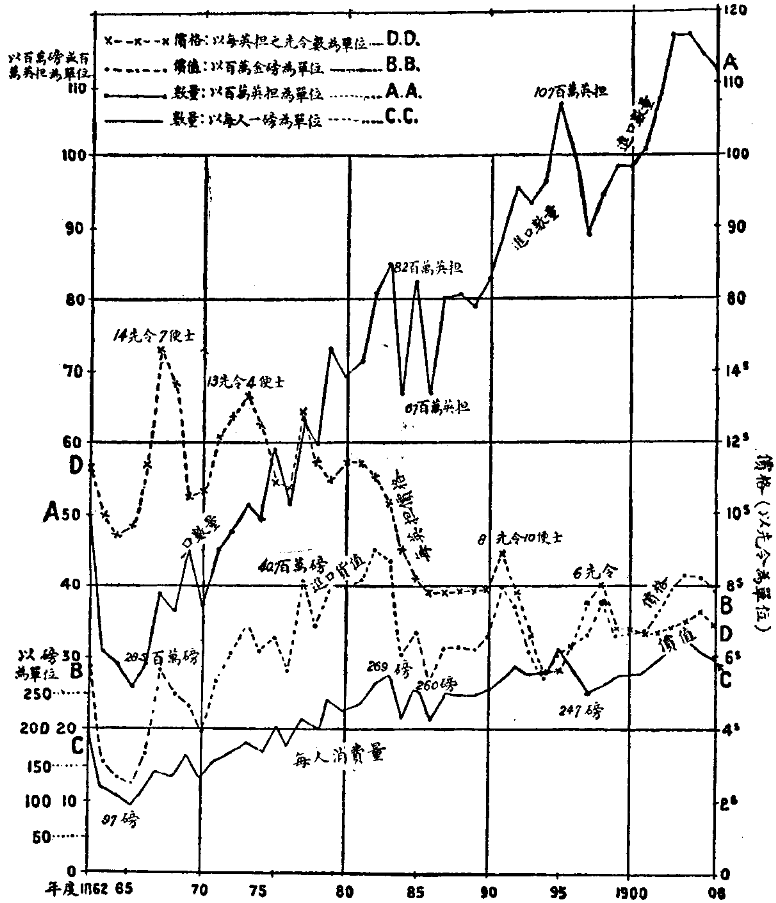
線各自獨立，不相連繫，但標尺尺度及基線均相同。如此畫法，較之繪一線代表總租稅額減去關稅，再繪一線代表總租稅額減去關稅及國產稅，依次減少之畫法，強似多多，因如彼不能令人一望卽可判斷各項之相對變動情況也。就圖觀之，所有關於租稅趨向之主要情勢，歷歷在目：增加雖速，但不規則。一八五四至一八五七年之進展甚急，隨後卽生變化，但一八六〇至一八七〇十年間之數字，終比一八五〇至一八六〇年間之數字爲高。自一八七〇年之後，變化甚大，增進之勢，亦較爲整齊，直至一八八七年，幾始終未生頓挫；但自一八八七年以後，中經一短期之靜止狀態，至一八九五年，突飛猛晉，一八九八至一九〇三年之漲勢亦然。總租稅額之曲線如此，內國稅及關稅之線，大致相同。更就各項租稅而論，其有增加及變動情形者，在一九〇〇年以前，國產稅收入增加最多，郵政收入次之，所得稅又次之，而關稅則漸減。各條曲線各有其各別之形式。郵政收入之增長，大致甚有規則。所得稅變化最劇，總租稅之激烈變動，幾全受此項之影響。尤以一八五六年一八七〇年及一九〇〇至一九〇二年爲甚。國產稅之線，在一八七〇年之前，有緩和之增加趨勢，其後繼以猛烈之突進，至一八七四年而後，後又現遲緩上漲之情勢。然在關稅方面，情形恰與國產稅之線相反，故二者之總數，在一九〇〇年以前，並無急劇之變化。上圖之上部，另有一新基線，所表示者乃爲各年

度之全人口每人負擔總租稅額金鎊數；由圖中可以看出，各年之增加，以一八五三至一八五七年及一八九八至一九〇三年爲最顯著。

第二尺度之選擇　由此觀之，尺度之決定，已不若先時僅有一條線者之難，蓋如上列之大圖，以百萬金鎊作單位，至用一鎊作單位時，則另定一新基線。然如欲將人口之變動，在大圖表現之，則仍可用同一之基線，以供兩種數量之比較。所不同者，第二種數量之尺度，可任以何處爲起點，卽如上圖之縱軸尺度，人口曲線所在之處，乃求其使人便於考察與租稅數成比例之變動情狀也。此一起點，決定之法，最佳莫過於視問題之性質，以求便於比較。例如欲比較一八五〇年以後，租稅之增長及人口之增加，則後者一數之線，應以前者起始之一八五〇年處爲起點，惟其如此，二條曲線之犬牙交錯，乃可顯然示明。雖然，一八五〇年乃假設之年份，並非問題所指定，故仍以使二線在資料中之最近年份相合爲宜，以便與以前各年比較也。試觀上圖，第二條線之位置，適足與內國租稅大部路線相密接也。

數量與價值之比較　尺度之決定，尙有其他困難，下列一圖，可爲例釋。本圖之目的，在說明人口與輸入小麥（數量，價值，及價格）之關係。A線所用之尺度，以求其能將曲線之變化曲折，完全托出爲標準，由此一線，將各人每年負擔之數額算出，則小

第十圖　一八六二年至一九〇六年輸入英國之小麥及麵粉



麥與人口之關係，卽可顯然。如C線乃卽代表每人消費量者，此線之尺度，與他線所用迥異，以免與其他各線交相錯雜，致混淆

一八六二至一九〇六年英國之小麥及麪粉
麪粉已折成小麥計算

年度	A. 進口總數量（單位十萬英擔）	B. 進口總價值（單位十萬金鎊）	C. 英國人口每人消費量（單位鎊）	D. 每英擔麥及麪粉之平均價值（單位先令）
1862	500	286	191 lbs.	11.44
1863	309	155	118 ,,	10.03
1864	288	135	109 ,,	9.37
1865	258	124	97 ,,	9.61
1866	294	168	110 ,,	11.[illegible]3
1867	391	285	144 ,,	14.58
1868	365	249	134 ,,	13.64
1869	444	233	166 ,,	10.50
1870	369	196	132 ,,	10.62
1871	444	268	158 ,,	12.07
1872	476	303	168 ,,	12.73
1873	516	344	186 ,,	13.33
1874	493	309	170 ,,	12.53
1875	595	324	203 ,,	10.89
1876	519	279	176 ,,	10.75
1877	635	407	212 ,,	12.82
1878	597	342	197 ,,	11.46
1879	730	400	239 ,,	10.95
1880	685	393	222 ,,	11.47
1881	713	407	229 ,,	11.42
1882	808	449	257 ,,	11.11
1883	851	438	269 ,,	10.30
1884	669	301	210 ,,	9.00
1885	823	337	256 ,,	8.19
1886	670	261	207 ,,	7.79
1887	802	314	245 ,,	7.82
1888	804	315	244 ,,	7.82
1889	789	311	238 ,,	7.88
1890	824	327	246 ,,	7.94
1891	895	396	265 ,,	8.85
1892	956	371	281 ,,	7.76
1893	938	308	273 ,,	6.57
1894	967	268	277 ,,	5.54
1895	1,073	302	305 ,,	5.63
1896	996	309	279 ,,	6.21
1897	887	330	247 ,,	7.44
1898	944	377	259 ,,	7.99
1899	985	330	267 ,,	6.71
1900	986	334	266 ,,	6.78
1901	1,011	334	270 ,,	6.60
1902	1,079	360	288 ,,	6.67
1903	1,167	397	309 ,,	6.80
1904	1,182	415	310 ,,	7.02
1905	1,142	413	296 ,,	7.23

（第三十三表）

不淸也。如圖式內容過於複雜，不妨援照前第九圖表現每人負擔租稅額之例，如法辦理。

構圖詳解 年度之尺度，必須一致不變，且爲計算簡捷，表示便利起見，每人消費一百鎊，應與一千萬英擔(hundrədweight)，在縱尺度上佔同一之距離。A 線與C 線所表示者均爲數量，故以同線代表之。B線表示價值，茲以斷續線爲代表，此線之尺度，較難決定。以後續列圖式，有時參用特殊方法，以供某項比較之用。但本圖則無用此之必要；茲用一尺度，使A，B，二線發生密切關係，而將B之變動現於圖上，縱尺度定每二十金鎊與二十英擔(cwt)，同一距離，故圖式簡單而明瞭。

D線乃由第三十三表A，B兩欄算出之小麥變動價格。此線之尺度，已決定如上圖，其所以如此者，以恰能與A，B兩線相交也；試觀上圖，該線之曲折變化，瞭如指掌，各數判然可見，蓋每英擔二先令，恰與千萬英擔同隸一縱尺度也。上圖繪製尙非甚精，不然，A與D兩線，在一八七六至一八七七年間，必互相貼合；可見現在已向上下互離，然本圖所示已甚明晰，吾人頗可得一大概印象也。

運動情況需加解釋 上圖各線，已有第三十三表爲之說明，但由此表現之特點及變化，必須請由研究經濟史者加以解釋。輸入英國之小麥，英國每人之消費量，在一八九五年以前三十年間，

歷年均有增加,迨至一八九五年而後,數年之中,乃又漸趨下遊。再輸入數量之線,有猛烈之短期變化,甚爲顯然。至代表價格之線,自一八六二年起至一八七八年左右止,經過劇烈之升降變化後,以後十七年間,竟有每況愈下之勢。現象如此,追原其故,原因極爲複雜:一則由英國人口之增加,二則進口小麥之多寡,須視英國農產之豐歉,而農產之豐歉,須受天時之支配,三則由於全世界收穫量之變動,四則由於政局之變遷,五則由於銀價之低落,六則由於交通運輸之發展,以及其他等等,不一而足。此圖之功能,只在標明各種運動之一般趨勢及發生變動之時期,惟運動之背後原因,則非圖式所能爲力者矣。

各個曲線之符號,究應如何規定?論及此問題,其準則雖夥,但主要者不外將相互交叉各線(除相交成爲銳角者外)均分別標誌清晰一端。至於凡相近數量者以相近符號表之:此乃第二之準則。如遇可用多種顏色時,此一原則,頗易實施也。(註六)

茲爲補充第三十三表賡續完成第十圖起見,又覓得以後各年之數字列下:

年　度	A	B	C(鎊)	D
1906	1.127	395	290	7.01
1907	1.156	440	295	7.61
1908	1.091	450	275	8.32
1909	1.132	516	284	9.12
1910	1.191	497	296	8.35
1911	1.120	442	276	7.89
1912	1.237	520	301	8.41
1913	1.225	502	296	8.20

趨勢與變動 欲檢察時間上數列之一般性質，其途徑有二；一為檢視數列運動之趨勢，一為考察數字變動之性質，茲將時間數列分成以下五類：

一、有方向穩定，或方向逐漸轉變之趨勢，而無忽高忽低之變動者。一國之人口統計數字即屬此類。(註七)

二、變動無定者；換言之，即一種運動，其多年記錄上之數字運動（或上或下）情形，不能據以推測下年度之為上升或下降。例如各年雨量統計即是。

三、具有補償的變動情形者；換言之，即變動情況，在一年成上升運動者，下年一般必有一下降運動以補償之。出生率，死亡率，及結婚率時或呈現此種補償情形。

四、成擺動式者；換言之，在達到最高或恐慌點後，必發生下降運動，如此一年復一年，經過數年之後，乃降低至最低點，其後復繼以上升運動，與年上升，乃又達於最高點。一般物價統計，及所有關於商業循環之極多記錄，均屬於此類也。

五、成循環性者；換言之，即每十年，或十二月，或他種定期間，上升與下降發生之順序，每期乃全相同，且（在數種情形之下）變動之大小度，亦各期相一致。季節變動之例，容在本章第四節討論之。

上舉二，三，四及五四種中，其趨勢可與變動同時存在。且有

一無定形式之變動，或一補償式之變動，發現於擺動式運動及一趨勢之上；所謂昇潮中大浪之微波，即此意也。故如有一時期數列當前，極重要之事項，厥爲考察數列所形成之趨勢及變動，大概性質若何，性質決定之後，乃可推斷最近將來之情形。如變動之情狀爲無定形式或變化出入甚爲驟烈，則設遇有極低之數字，吾人似可無須爲之擾亂，因而相信有亟圖補救之必要也。在補償變動中，如在一高價之後，必隨之出低價，此則可爲預斷者。如變動成擺動現象，則數字在一經高價跌落之後，回後之日必須有待，乃可在吾人意料之中也。

第三節　數列之比較

一、在未研究次一圖式之前，似應先請就用數字作比較研究之目的，加以分析，並考慮作比較研究而堪供應用之方法。

比較之主旨所在　在研究兩組類似數量（例如兩國之各該國貿易或二國人口之各別趨勢）時，吾人欲知者有二端：一則爲一般進展率(rate of progress)（此可用修勻曲線法求得之），二則爲特殊增進之年份，即最高或最低點所在之時期。一言以蔽之，即欲就吾人目所能見者，用增加量，增加率，及增加率發生變動之時期三者，以作比較研究也。爲達此目的，最顯而易見之方法，不外將兩國用同一尺度，同一基線，以表示之。至數量之單位，

兩國亦使成一致；然此法應用之途有時而窮，不能到處皆通也。何則？蓋吾人用此方法，可以判斷兩國各年中之增加及減退，可以考查最高點，最低點，及增加率發生變動之日期，固矣，其奈不能比較增加率何？夫作一大略之臆斷，固屬可能，但比率究非易於判斷之事。例如：設兩國之貿易數量，大小迥然不同，雖二國之絕對增值 (absolute increment) 完全相等，而相對增值 (relative increment) 則大有區別，且此乃又不易使人時時覺察者也。

百分數尺度 此為美中不足，宜如何補救之，唯有變更尺度之排列一途。故除上文所言之圖式之外，應另作一圖，在此第二圖中，尺度之單位必用百分數，而不言貨幣數額：例如，茲以一八五〇年英國國外貿易之百分之一，為英國貿易尺度之單位；而以同年德國貿易之百分之一，為德國貿易尺度之單位。易言之，卽以某年兩國貿易總值之百分數，代表兩國之貿易，而各以一線，分別代表此兩組百分數。此外，在圖旁不妨分列兩個或兩個以上之尺度，以示各國貿易之絕對量。如是，則增加率乃可從事比較，如在一八五〇年，所謂增值相等者，卽表示各國貿易百分數相等；而且，此一國較之彼一國，稍佔優勢之日期，圖上已瞭如指掌。

絕對進展與相對進展 絕對增加率與相對增加率二者，究應以何者為研究之對象？此一問題，乃統計學上極普通之問題。

有時必須求得絕對數量，此尤以吾人如欲估計促進特殊階級福利之社會政策，實行後之影響，或計算某某數國之貿易時爲然；但有時必須求得相對數量，例如考察各種產業之增進量，或探尋將來之繼起競爭者時即是。絕對進展與相對進展，二種研究，雖或代表同一數字，但仍非用兩個不同之圖式表示之不可也。

作圖比較時，須以某一年之數量爲基準，但所謂某一年，究以採用何年爲宜，此乃主要困難之所在；決定之法，必須視論辯之性質，蓋此論辯，即圖式所欲例證者也。

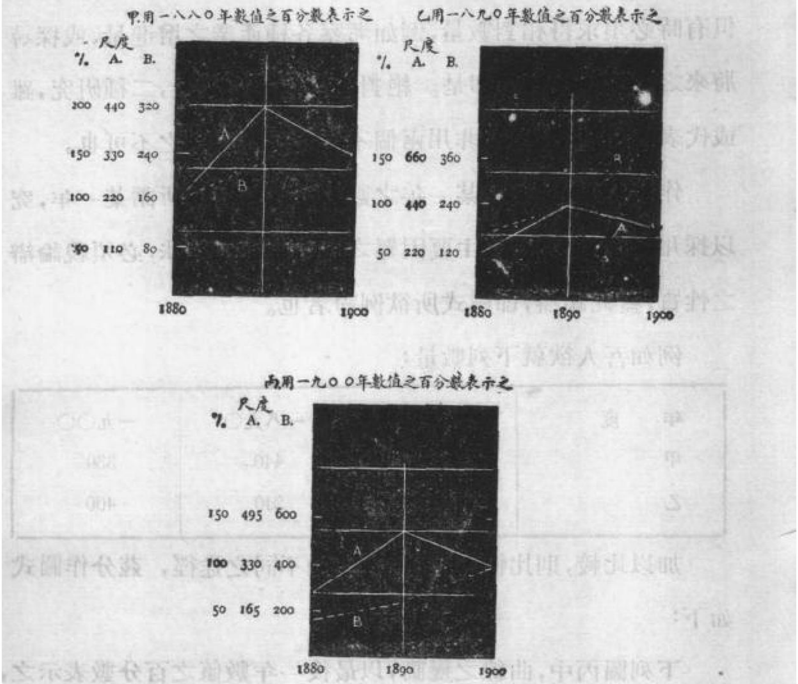
例如吾人欲就下列數量：

年　度	一八八〇	一八九〇	一九〇〇
甲	220	440	330
乙	160	240	400

加以比較，則比較之法，可有三種不同之途徑，玆分作圖式如下：

下列圖丙中，曲線之變動，以最後一年數值之百分數表示之，各年之進展比例，表現情形，較圖甲所示者爲佳。一般情形，多係以最近年度之數量爲基準，故以前各年度之數字雖小，但以現代觀點考察之，則各年變化之比例，必能適當其分。吾人如言一八八〇年數值甲，當近今數值甲百分之四十，一八八〇年數值乙，當一八八〇年數值乙百分之一百五十，必使人易於理解；但此亦

第十一圖



宜隨情形而決定，何年最爲適當，即以何年之數值爲基準，未可一概而論也。

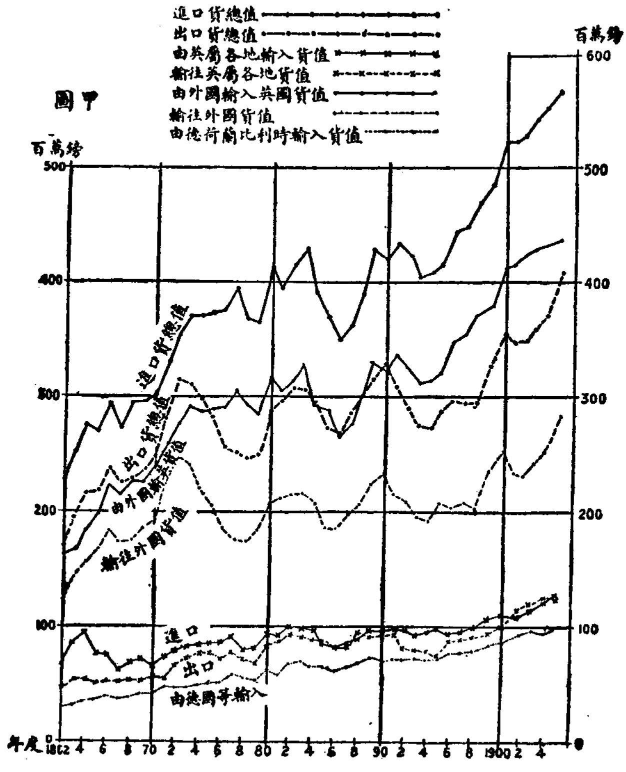
用英德貿易爲例證　上文所論幾點，茲可更以一圖式表明之，下列第十二圖之目的，在分析英國與其屬地及外國間貿易之進展，英德貿易情形，尤當重視焉。

一八六二至一九〇年英國之出入口貨值
以十萬金鎊爲單位

	進口總貨值	出口及復出口總貨值	英國輸往英屬各地之貨值	英國輸往外國之貨值	由英屬各地輸入之貨值	由外國輸入之貨值	由德國，荷蘭，比利時輸入之貨值
1862	2,257	1,662	454	1,207	653	1,604	279
1863	2,489	1,969	550	1,419	847	1,642	283
1864	2,749	2,126	557	1,569	937	1,812	332
1865	2,711	2,188	515	1,673	728	1,982	364
1866	2,953	2,389	572	1,817	722	2,231	388
1867	2,752	2,258	534	1,724	607	2,144	373
1868	2,947	2,278	537	1,741	670	2,277	379
1869	2,955	2,370	519	1,851	704	2,250	405
1870	3,033	2,441	554	1,887	648	2,384	409
1871	3,310	2,836	556	2,280	729	2,581	469
1872	3,547	3,146	656	2,490	794	2,753	455
1873	3,713	3,110	711	2,399	810	2,903	463
1874	3,701	2,977	779	2,197	822	2,879	494
1875	3,739	2,816	767	2,050	844	2,895	515
1876	3,752	2,568	701	1,866	843	2,908	516
1877	3,944	2,523	758	1,766	896	3,049	590
1878	3,688	2,455	720	1,735	779	2,908	575
1879	3,630	2,488	665	1,823	789	2,840	543
1880	4,112	2,864	815	2,049	925	3,187	616
1881	3,970	2,971	867	2,104	915	3,055	582
1882	4,130	3,067	923	2,143	994	3,136	658
1883	4,269	3,054	904	2,150	987	3,282	692
1884	3,900	2,960	883	2,077	958	2,942	646
1885	3,710	2,715	885	1,860	844	2,866	638
1886	3,499	2,690	822	1,867	819	2,680	609
1887	3,622	2,813	823	1,990	838	2,784	646
1888	3,876	2,986	917	2,068	869	3,007	684
1889	4,276	3,156	908	2,248	973	3,304	715
1890	4,207	3,283	945	2,337	962	3,245	694
1891	4,354	3,091	933	2,158	995	3,360	716
1892	4,238	2,916	812	2,104	979	3,259	715
1893	4,047	2,771	786	1,986	919	3,128	720
1894	4,083	2,738	786	1,952	940	3,143	716
1895	4,167	2,858	761	2,098	957	3,210	729
1896	4,418	2,964	907	2,057	933	3,485	761
1897	4,510	2,9[illegible]1	871	2,071	941	3,569	760
1898	4,705	2,9[illegible]0	901	2,038	998	3,708	786
1899	4,850	3,295	943	2,352	1,069	3,781	834
1900	5,231	3,544	1,021	2,523	1,096	4,134	861
1901	5,220	3,479	1,132	2,347	1,057	4,163	897
1902	5,284	3,492	1,176	2,317	1,069	4,215	950
1903	5,426	3,604	1,195	2,409	1,137	4,289	973
1904	5,510	3,710	1,208	2,502	1,200	4,310	962
1905	5,650	4,076	1,227	2,849	1,279	4,372	990

（第三十四表）

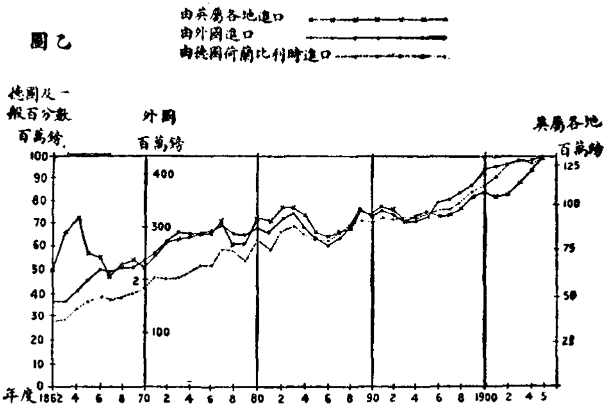
英國與其屬地及外國間之貿易價值



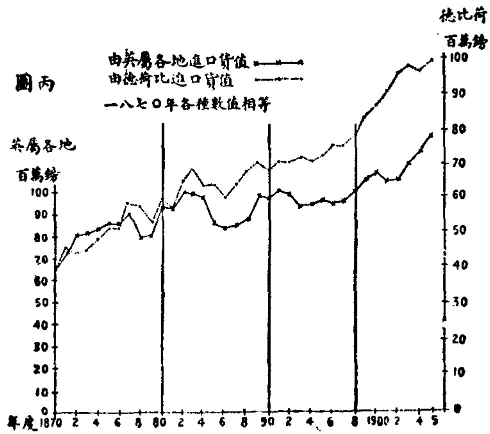
（第十二圖）

進口貨值對一九〇五年進口總值之百分數

圖乙



圖丙



（續第十二圖）

第十二圖圖甲所表示者，爲英國之出入口貨總值，及分別與英屬各地及外國之出入口貨值，各線之尺度，均以百萬金鎊爲單位，由德國，荷蘭，比利時輸入英國之進口貨值，另以一線表示之；此三國之數值，乃合併計算而得，蓋截至一九〇四年止，德國之製品，運輸之時，未能與荷，比二國之貨分淸，故統計數字，亦無從單列也。觀察圖甲，英國由外國或屬地進口，試問何者進展最速，並無明白表示。故在圖乙，另將三條線重新造圖，而用百分數爲尺度，意卽以一九〇五年各該數值之百分數，分別代表三種數值也。於是在圖乙中，自其他外國及英屬各地與印度輸入英國之貨物，除在棉產荒歉期間外，餘時均有並駕齊驅之勢，惟自德荷比三國入口，則進展甚急，遠非其他外國及英屬各地與印度所能及耳。吾人苟以一八六二年爲基準，而使此年各數量全然相等，則代表德荷比三國進口之曲線，至一九〇五年，必超越其他曲線甚多；但由此所得之印象，如就絕對數量觀之，則不免陷於錯誤，蓋由德荷比三國入口增加七千一百一十萬金鎊，其餘外國輸入英國，則增加二萬七千七百萬金鎊也。

圖丙表示德荷比三國及英屬各地，自一八七〇年以來，輸英貨值之相對增加率。

國際統計局(International Institute of Statistics) 爲求作比較用之歷史圖標準化起見，經多番考慮，結果於一九一一年

開會時議決以一九〇一至一九一〇年間數字之平均數爲標準，並議決此十年數字之平均數，須以縱長尺度代表之，此縱長之高度，應與代表三十年之橫尺度相等，用此標準化之尺度，作成之圖，乃可任與代表何種數量之圖，相互比較。此種議案之意，並非不可作別種之比較（例如本章第二節第十圖，一八六二至一九〇六年輸入英國之小麥及麪粉），更非不能作以同一單位（金鎊或噸數）表示並以同一自然單位作圖而代表之時間數列之圖式。反之，本議決案之用意，乃以此標準作無他法可用時之唯一標準形式，如有理由必不能用此標準時，則此標準只供參考云耳，非不用不可也。總而言之，凡作一比較——尤以國際統計爲尤甚——如遵循上述準則；便利甚大焉。

因果關係　二、關於數列之比較，爲探討或例證其因果之關係(casual relation)，時常採用地圖表示法。在地圖顯示法之下，吾人不惟研究增長率，如上列第十二圖圖丙所示，且更須就全部期間中，考察增加率有無任何相似之徵象，出現最高量及最低量之日期，或研求發生變動時同時發生之事象。然用此種種方法以作比較，其事甚難，非經精謹審慎之分析不爲功。例如，設吾人欲考慮室外貧民救濟費之增加，是否與貧民加多有關：則須以一線代表款額，一線代表人數，二者並無共同單位；吾人於此無須計算百分數，但款額之尺度，旣經加以規定，則人數之尺度，卽可以

最簡單之方法排定,於是二種尺度可以任一年度爲基準,二種數量在基準之年乃彼此相等。然後吾人所欲知者有二事,一查款額或增或減,是否與人數增減同時發生,或款額之增減,恰在人數增減之前;二查款額增加多者人數是否亦隨之俱多。爲求顯示直接關聯起見,構圖之時,二線應使相愈距近爲愈佳。

構圖法 構圖之法,第一步先作一粗略圖,二線之尺度,任意規定之;由此粗略圖式,即可作一試驗,而查明二線相似之情況。請注意,在何期間所生變動爲最大;此一變動最大之期間,即一般用作比較之日期,蓋惟在此期間,因果關係乃最爲顯露也。反之,如因特殊理由,採用其他期間,亦須明瞭此點,否則必引人批評,否認所採期間爲唯一之關係明顯之期間也。如此,吾人從任何兩條曲線中,查出同時發生變化之短短期間,事必不致甚難。爲達此目的,可求所採用期間之款額平均數及人數平均數,並重新再作一圖,在此第二圖中,人數尺度之決定,以使人數平均數與款額平均數相等爲標準。如此,兩條曲線苟有任何適應性,即可一望而知矣。

負相關 在許多情形之下,常有一種數量,爲吾人視爲原因者,其變動乃與另一種數量而爲吾人視爲結果者之變動,成相反之方向。例如,設吾人以貿易進展與失業人數作比較觀察,而按上文所言之構圖法顯示之,則第一條曲線之最高點,必與第

二曲線之最低點,同時出現。但如此尚非十分明顯,如欲求更大之明晰,可以一線倒置之;換言之,即用一線代表就業之人數,而不使表示失業之人數,則二條曲線之變動必亦步亦趨完全一致矣。

較複雜之關係　依照上文所論之構圖,設一種數量之變動,恰與他一數量之變動恰成比例,詳言之,如以室外貧民救濟費爲例,代表款額之線,超出平均數百分之十,貧民人數亦必增多百分之十,則在所有期間中,無論兩線之變動如何,一線必始終在另一線之上,此種關係,乃爲極罕見之簡單關係;吾人一般所見者,大多最高點最低點在同時出現,兩組數量之變動,自始至終大多同一方向,且一線變動較大者,他線變動亦較烈也。

圖式表示相關之用途　圖式常用以表現兩組數字之相關(correlation),而表示兩數字之相關,亦確爲圖式之主要優點之一也。又圖式可在論辯某問題時作爲例證之用,惟圖式之效用,如就此點而論,誠亦甚微,蓋圖式所表示者,只足爲簡略之例釋,非極詳備之證明也。夫因果關係之建立,事本甚難,在理論付諸試驗之際,採用之原始數字,非經過極嚴密之考量不可也。

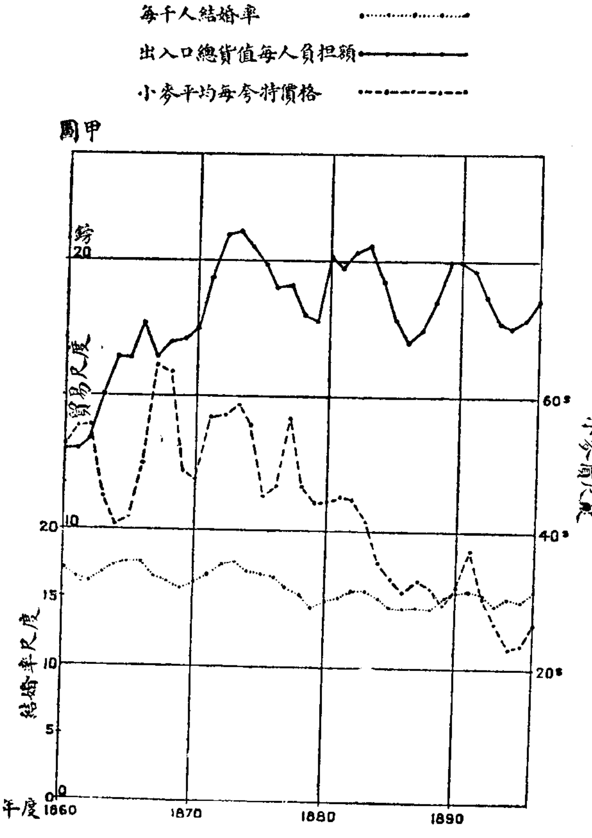
更爲確切之方法 爲作比較而用之圖式，其用途並非止此而已，尙有較爲確切者在焉。然下文所論方法，採用之時，須特別謹愼。設吾人意欲確定一金鎊多買一布舍爾(bushel)小麥，是否與結婚率增加千分之一·五，或其他任何之嚴格數字比例相適應。必須繪一圖代表小麥數量，將已擇定實行比較之期間中之數量加以平均，然後以平均數爲零點，由平均數向上，分成一，二，三……布舍爾，或自平均數而下，亦分成一，二，三……布舍爾等尺度。如此並無基線。現請更繪一曲線，代表結婚率，在擇定期間中，超過結婚率平均數之部分，或低於平均數之部分，此一條線之尺度，必須每超過平均數有千分之一·五，其距離卽與一布舍爾小麥數量所佔之縱標尺距離相同。然則此兩條線，接近之程度，卽可測驗理論是否可以成立，如能成立，卽可測驗成立之限度爲何如矣。惟此一方法，用時每有危險，蓋既無基線之設，自不能以總計數爲比較，而鑑定變動之總額也。雖然，如畫兩條基線，固屬合乎需要，然一有基線，裨益雖非甚大，而因之反有易致紛擾之虞也。

執此而論，吾人苟能善自採用尺度及基線，則兩日期中之各點，必能與若干代表數列精確畫成之曲線相合。玆舉一圖爲例以釋明之。

例釋　英國之結婚率，出入口總貨值平均每人負擔額，與小麥每夸特(quarter)平均價格

年份	結婚率	出入口總貨值平均每人負擔額 金鎊	先令	便士	小麥每夸特平均價格 先令	便士
1860	17.1	13	0	8	53	3
1861	16.3	13	0	3	55	4
1862	16.1	13	8	0	55	5
1863	16.8	15	2	7	44	9
1864	17.2	16	8	7	40	2
1865	17.5	16	7	5	41	10
1866	17.5	17	14	5	49	11
1867	16.5	16	9	6	64	5
1868	16.1	17	0	6	63	9
1869	15.9	17	3	9	48	2
1870	16.1	17	10	3	46	10
1871	16.7	19	9	6	56	8
1872	17.4	21	0	0	57	0
1873	17.6	21	4	2	58	8
1874	17.0	20	11	0	55	8
1875	16.7	19	19	4	45	2
1876	16.5	19	0	10	46	2
1877	15.7	19	5	5	56	9
1878	15.2	18	2	1	46	5
1879	14.4	17	16	10	43	10
1880	14.9	20	3	3	44	4
1881	15.1	19	17	5	45	4
1882	15.5	20	8	10	45	1
1883	15.5	20	13	2	41	7
1884	15.1	19	4	1	35	8
1885	14.5	17	16	9	32	10
1886	14.2	17	0	10	31	0
1887	14.4	18	11	7	32	6
1888	14.4	18	12	1	31	10
1889	15.0	19	19	9	29	9
1890	15.5	19	19	7	31	11
1891	15.6	19	14	0	37	0
1892	15.4	18	15	6	30	3
1893	14.7	17	14	9	26	4
1894	15.1	17	11	9	22	10
1895	15.0	17	19	3	23	1
1896	15.8	18	14	1	26	2

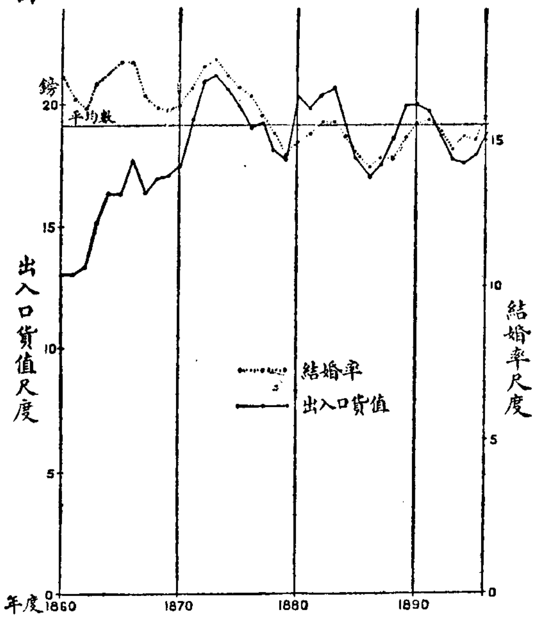
（第三十五表）



（第 十 三 圖 ）

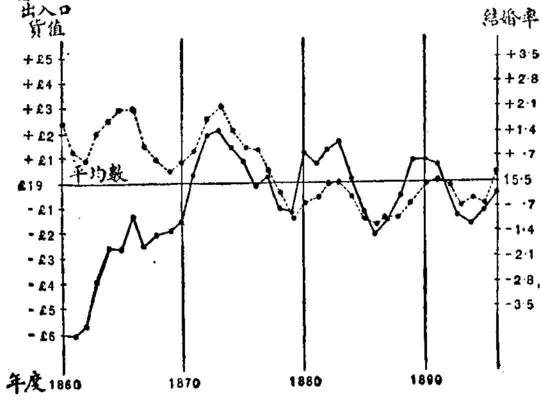
在上列第十三圖甲圖中，有代表小麥每夸特價格先令數之線，有代表出入口總貨值除以英國人口之線，並有代表每千人中結婚率之線。尺度之擇定，有二簡單之原則，一則最便於應用，二

圖乙



英國人口平均每人負担出入口總貨值與結婚率

圖丙　一八六九至一八九六年平均數均在同一線上
一八六九至一八九六年間對平均數之平均差均相等



（續第十三圖）

則將各曲線刻畫盡致,起伏狀況明顯。各個標尺尺度上諸點,凡屬於同一年份者,必相互重疊,但各點各有其尺度,所代表之單位,彼此並不相同。至基線則無相合之必要也。

貿易額與結婚率 試觀上圖,各種數字之進行曲折,是否有類似之情狀,此則不問可知。先就貿易額與結婚率二者言之,自一八七〇年以來——稍前數年或亦同然——貿易額之變動,與結婚率之起浮,大概有相應之狀態,此乃可爲斷言者。又圖上,小麥價格及貿易總值,相似之點甚多;一八七〇至一八七三年,二者相率上昇,一八七三至一八七五年,又相繼下游;一八七六至一八七七年,復同時俱起,其後兩年又同時俱落,至一八七九年後,乃復同趨上昇;自一八八一年,二線又偕以俱降,直至一八八六年爲止,迨一八八七年,乃又齊趨上昇矣。雖然,相同之點固多,行動相左之點,亦所在多有,例如一八六二至一八六四年,及一八八七至一八八九年,乃特別顯著者也。

結婚率與小麥價格 現請再檢視小麥價格與結婚率之線,可見在該世紀前數十年中,關係甚爲密切,此漲彼落,似頗顯然,惟細加考察,二條曲線,既不甚爲相似,復非恰相反對。在一八六〇至一八六二年,小麥價格上漲,結婚率則降低,在一八六二至一八六四年,小麥價格下落,而結婚率則上漲;一八六五至一八六七年,小麥價格騰起,而結婚率初則堅挺不動,繼則微形頹弱;

一八六八至一八七〇年，小麥價格雖日趨低下，結婚率亦每況愈低；一八七〇至一八八〇年，結婚率出現一極大之變動，而小麥價格，則有兩個短促之變動，迨至一八八〇年以後，小麥價格連年低降時，結婚率一般形勢亦始終衰弱不振，小麥價格上昇之日，結婚率亦趨上昇。

二者之聯環關係　此二現象，或有聯環之關係，吾人不妨略一討論之。小麥爲勞動階級生活費用中之主要項目，故小麥價格之漲落，實爲勞動羣衆最爲關心之主要事項，然則小麥價漲，而結婚率降，非無故矣。如就另一方面言之，當麥價廉，工資高之時，麵包價格偶有變動，則重視之者，僅爲少數人而已；而此乃國家一般興盛之象，由對外貿易情況可知，是結婚率擡高，有由來矣。

在出口入口貨值增加之時，貿易大形發達，預備結婚者，雖當此物價高昂之際，猶充滿樂觀空氣，以爲繁榮可常留，物價終當降落也；但物價一落，利潤大減，國民所得自亦降低，於是預備結婚者，乃不得不稍持穩重。然則吾人基於此種理由，而謂結婚率與國外貿易二曲線，彼此有一致之趨向，要非過言矣。

結婚率之增加，乃隨貿易發達以俱來，而貿易發達，又出現於一般物價上漲之際，故僅就小麥價格一項而論，麥價與一般物價有連帶關係，國外貿易發達，則麥價漲，對外貿易衰落，則麥價亦隨之俱低；於是麥價漲，結婚率亦隨之而昇，麥價落，結婚率亦

隨之俱降。然麥價之漲落，原因非止一端，除對外貿易一因素外，尙有其他特殊之原因在，故麥價並非永隨貿易狀況而變動，與結婚率尤少相關，此十九世紀中前期麥價之趨勢，所以恰與結婚率成相反之變動也。

由此而論，結婚率與出入口貿易二曲線，有相應之趨勢，麥價與貿易二曲線，略爲一致，但並非有甚密切之傾向，麥價與結婚率有雙重之趨勢，本不足爲奇。試觀上列第十三圖，結婚率與出入口貿易相應之情形，已躍然紙上。小麥價格與貿易相應情狀，仍不難用同一方法釋明。至結婚率與麥價之關係，應另作比較。比較之法，須以不同之計劃，重作二次之比較：第一先求其正相應(direct correspondence)，次則以上端作麥價曲線之基線，另構一圖，以求二者之負相應 (inverse correspondence)。

圖式之構造 比較貿易及結婚率二曲線之程序，有如下述：吾人檢查上列第十三圖圖甲，知二條曲線，在一八六九年以前，並無任何相似之關係，故自一八六九年起之各部曲線，應使之顯露密切之相應。一八六九至一八九六年，結婚率平均數爲一五・五，平均每人負擔出入口貨值額爲一九金鎊。茲以平常方法將結婚率曲線繪成，次藉用滑動尺度(sliding scale)，將貿易曲線畫上，如此則二者之基線相同，一九鎊與一五・五兩平均數同在一條線上，詳見第十三圖圖乙。

經此構圖結果,吾人由圖中可以看出,二條曲線之升降,均發現於同一日期,惟變動程度不同耳;蓋自一八七三至一八七九年,二曲線幾成平行,二線達到最高點之後,下落之程度,約略相等,但至一八七九年而後,貿易及結婚率二條曲線,在平均數線上下擺動,貿易線尤較結婚率線爲甚焉。

最終之比較　一切程序至此,即須用作圖測驗法,考察各次變動,是否互成比例。此時,可令出入口貨值對一八六九至一八九六年平均數之平均差(1.04鎊),與結婚率對同期平均數之平均差(0.72)相等,而成一尺度方程式;不然即用粗略方法,以等距離之縱尺度,一方代表一鎊入口貨值,一方又代表結婚率千分之〇·七,如此乃成立一假定;凡出入口總貨值每人負擔額,有一鎊之變動,則每千人中結婚率必同時有〇·七之變動。上列第十三圖圖丙之尺度,即依此辦法而決定者,已在圖中公共平均數線之上下標明矣。

在丙圖中,吾人一望即知自一八七〇年以後,二曲線之變動,合攏甚爲密切,但此一密切關係,惟犧牲一八七〇年以前之部分,始能得之。在一八七九至一八九三年一短促之期間中,二條曲線愈形貼合;然此既經用特殊之選擇而繪成,則以此作爲一般論辯之基礎,必將有誤入歧途之危險。

吾人根據構圖之結果,可得一結論:自一八七〇年以後,支

配英國國外貿易之原因，亦同樣爲支配結婚率之因素，使二者發生變動既在同一時期，變動方向亦相互一致；且一者受影響之程度愈大，他一曲線受牽動之程度亦愈甚；惟二者之間，並無成簡單比例之一定法則耳。

在比較曲線之時，可不必比較對一期間平均數之離中差，正當而有利之方法，莫過於計算對修勻曲線之離中差，至計算對修勻曲線之離中差，由移動平均數求之固可，否則即用其他方法求之亦無不可也。總之，算出對修勻曲線之離中差後，吾人可忽略有逐漸而永久效果之原因，從事比較短期之變動矣。此節容以後再申論之（見第二編第六章）。

附註——上列第十三圖圖乙，所欲測驗之關係，可以一方程式$\frac{x}{a}=\frac{y}{b}$表示之，丙圖所欲測驗者，可以方程式$\frac{x-a}{y-b}=c$（c爲常數）代表之。式中之x爲出入口貨值，y爲結婚率，a爲出入口貨值在某一期間中之平均數，b爲結婚率在一期間中之平均數，選擇c值之法，須能使兩組數量之平均變動相等。如用最小二乘法，則擇定之c，必能使二條曲線之相應性，較用本書算出之數值所求得者爲密切也。

第四節　循環數字

循環數字　現當論及循環數字（periodic figures）矣。所謂

循環數字者何？乃即在某一週期（例如數字報告爲每月一次時週期即爲一年）中，依一定之時間，達到最高點及最低點，並表示連續期間中呈有規律之變動之數字也。此種數字之例證甚多，在自然現象中，有如日之升起，每日之記錄相同，均可代表現象，年復一年，幾無變化。舉潮汐之例而論，季節現象之較爲呆板之各年曲線，與社會統計中稍欠明顯之週期，二者之間，乃確有其相連之關係；惟潮汐往往受諸種不同之影響，因其週期乃有二十四小時一循環，及其他長短不同之週期循環，且因素既衆，影響力量復有強弱之異，一因素之效果，往往爲他因素之效果所掩蔽耳。即在英格蘭銀行之每週記錄中，翟翁士(Jevons)氏，（註八）亦曾發現有按月，按季，按年，種種不同之循環焉。

在社會及產業統計中，吾人慣常發現有按年之循環，同時有遲滯之向上向下之一般運動，攙雜其間，另外復因有循環之商業繁盛及蕭條關係，呈現大約每十年一循環之不規則週期。此三種運動對於數字之影響，吾人不難考察而得，茲就英國鑄鐵工人協進會失業人數之每月調查數字爲例，作一完全之討論，以謀徹底明瞭研究循環數字之一般方法。翟翁士氏論文『秋收對貨幣市場之頻繁壓力』（註九）可請讀者參考作爲第二例釋之用；如欲作爲練習之用，可請採用英國統計月報中之小麥價格，在此數字中，各年數字圖，形式有逐漸變化之象，此種變化之經過，可參照

全球各處秋收影響日愈強烈之情形察視之。

循環數字之一般形態 此種循環數字，如用圖示法，表明兩重之循環，有特別適用之效，考察每年急促變動及較長時期之一般運動，相互發生之影響，尤爲適宜。請閱下列第三十六表：

循環數字

鑄鐵工人失業人數，以英國鑄鐵工人協進會會員總數估計數之百分數按月份分別表示之：計算所根據之數字，爲鑄鐵工人協進會一八九四年之常年報告。

年份	一月	二月	三月	四月	五月	六月	七月	八月	九月	十月	十一月	十二月	一年平均
一八五五	11.1	14.1	14.0	12.5	10.0	9.9	8.7	8.7	6.8	7.7	8.8	12.0	10.4
一八五六	10.9	12.6	12.2	10.0	9.4	7.5	6.9	7.3	6.9	8.1	8.7	9.9	9.2
一八五七	10.1	9.5	8.7	8.7	8.1	7.3	6.8	6.9	6.2	8.0	14.0	17.7	9.3
一八五八	20.2	20.6	20.9	19.8	20.3	17.8	15.9	14.3	13.1	11.9	11.5	11.2	16.5
一八五九	10.6	8.8	6.5	5.2	4.0	4.4	3.2	3.6	3.4	3.8	4.6	5.1	5.3
一八六〇	4.0	3.2	2.6	2.2	1.6	1.7	2.3	2.6	2.6	2.9	3.7	5.6	2.9
一八六一	6.0	6.9	6.5	7.9	7.8	8.4	6.9	7.9	9.5	10.7	12.4	13.8	8.7
一八六二	14.5	14.0	14.0	14.6	14.4	13.7	13.3	12.9	12.2	13.5	14.9	16.0	14.0
一八六三	15.5	13.9	13.6	11.6	10.4	9.3	8.1	7.8	7.4	6.6	5.3	5.0	9.5
一八六四	6.0	7.1	6.6	5.1	4.4	3.3	2.8	2.8	2.6	3.3	4.2	8.1	4.7
一八六五	5.4	5.3	5.5	4.6	3.4	2.9	2.6	3.1	2.7	2.6	2.3	4.9	3.8
一八六六	4.2	5.4	5.1	3.6	5.1	6.5	5.9	6.5	6.9	7.4	9.3	13.8	6.7
一八六七	12.4	13.2	15.4	16.7	14.9	14.6	14.2	13.9	15.7	16.3	18.9	22.6	15.7
一八六八	22.1	20.9	19.8	18.6	16.7	15.8	14.9	14.7	14.2	14.1	15.6	17.4	17.1
一八六九	17.3	17.1	16.8	15.6	15.2	13.6	13.3	11.8	13.1	13.6	14.8	15.3	14.8
一八七〇	14.5	10.9	8.7	7.2	5.0	4.5	3.7	4.5	4.9	5.0	5.6	8.3	6.9
一八七一	7.2	5.6	3.6	2.8	1.6	1.5	1.6	1.2	.9	1.4	1.1	2.2	2.6
一八七二	1.1	1.1	.9	.8	1.2	.7	.9	1.0	1.3	1.8	2.6	4.1	1.5
一八七三	3.3	2.8	2.7	2.5	2.1	2.0	3.0	4.9	4.5	3.3	3.3	5.1	3.3

平均數一八五五至七三	**10.3**	**10.2**	**9.7**	**8.9**	**8.2**	**7.7**	**7.1**	**7.2**	**7.1**	**7.5**	**8.5**	**10.4**	**8.6**
一八七四	4.9	3.9	3.9	3.5	4.9	3.9	3.8	3.4	3.5	3.7	3.9	5.0	4.0
一八七五	4.6	3.4	3.5	2.8	2.8	2.8	3.3	3.4	3.6	4.1	4.1	5.0	3.6
一八七六	4.9	4.9	4.9	5.4	4.8	5.2	5.7	5.8	6.4	6.4	6.2	10.3	5.9
一八七七	7.7	7.4	7.0	6.9	8.4	7.6	7.4	7.8	9.6	10.9	12.3	16.3	9.1
一八七八	14.0	14.3	13.5	15.3	13.3	14.6	13.6	13.2	13.3	14.0	15.7	21.0	14.7
一八七九	23.2	23.8	24.7	25.5	22.3	23.4	21.5	22.6	22.5	21.1	18.0	16.6	22.1
一八八〇	15.2	12.9	11.1	10.0	10.0	9.7	9.8	10.0	10.0	9.2	9.2	10.2	10.6
一八八一	11.5	10.8	10.1	10.1	7.6	7.5	6.5	5.8	5.6	5.4	5.0	6.6	7.7
一八八二	5.5	5.2	5.3	4.5	3.6	3.8	3.2	3.4	3.6	4.1	4.4	6.0	4.4
一八八三	3.6	4.8	5.2	4.3	4.2	3.6	3.9	4.3	4.3	4.2	4.0	6.6	4.4
一八八四	6.1	6.2	5.9	6.5	6.5	6.9	6.5	7.6	8.1	7.8	9.8	10.9	7.4
一八八五	10.2	11.1	10.0	10.1	9.8	9.1	9.8	10.7	11.8	11.6	12.7	13.6	10.9
一八八六	14.1	15.0	15.2	15.5	13.4	13.1	12.1	12.7	13.6	13.9	12.7	12.9	13.7
一八八七	12.4	11.6	10.2	9.1	9.2	10.6	9.2	8.8	9.6	9.4	9.4	9.1	9.9
一八八八	7.8	7.5	6.4	6.4	5.9	5.2	5.7	5.0	5.1	4.8	3.2	3.5	5.5
一八八九	3.1	3.3	2.4	2.1	1.7	1.6	1.7	1.7	1.6	1.5	1.2	1.4	1.9
一八九〇	1.3	1.5	3.2	3.1	2.8	2.4	2.4	2.7	2.7	2.7	2.7	2.7	2.5
一八九一	3.9	3.5	4.2	4.2	4.6	4.0	4.5	4.8	5.4	5.6	5.7	6.3	4.7
一八九二	7.0	7.2	7.9	8.1	7.9	7.9	7.7	7.6	9.3	11.4	10.9	12.0	8.7
一八九三	11.5	11.2	10.1	7.7	9.6	8.5	8.3	9.2	11.7	11.9	11.5	11.5	10.2
平均數一八七四至九三	**8.6**	**8.5**	**8.2**	**8.1**	**7.7**	**7.6**	**7.3**	**7.5**	**8.1**	**8.2**	**8.1**	**9.4**	**8.1**
平均數一八五五至九三	**9.4**	**9.3**	**8.9**	**8.5**	**7.9**	**7.6**	**7.2**	**7.4**	**7.6**	**7.9**	**8.3**	**9.9**	**8.3**

第三十六表

於諸年各橫列中，可見在每年中間，必有降低之現象。再查縱行每月份之下，各年增減並無通盤顯著之趨勢，因前數年與末數年，均見高低不同之數字也。惟最引人注意者，此種數字之主要

形態，卽爲成組之低數字年份與成組之高數字年份，有循環返復之趨勢。每月失業人數佔會員總數之百分數，在百分之十以上者，有一八五七至一八五八年，一八六一至一八八一年，一八八四至一八八七年及一八九二至一八九三年數組。茲僅就一八六六至一八七〇年一期研究之。一八六六年一月份之數字，比以前各年同月份之數字均低；一八六六年二月，三月，及四月份之數字亦然；惟自五月起至九月止，此五個月份之數字，則比一八六五或一八六四年同期之五個月份數字爲高；一八六六年自十月至十二月之數字，則比一八六三，一八六四或一八六五年之數爲高；至一八六七年十二月乃愈大，以前各年之數字無與倫比者。一八六八年之數字，多半均比以前九年爲高；惟自一八六八年九月以來，各年同月數字每況愈下，必較上一十二個月份之數字爲低，此種情形，直至一八七二年九月爲止，始形緩和。由此可見，此一失業狂瀾，波濤澎湃，自一八六六年五月起，延至一八七二年九月，有六年餘之久焉。

季節影響　現請再考察季節之影響。一八六六年，除在四月一月外，夏季並無低落之現象，至十二月反有急速之昇進。一八六七年，自五月至八月一落之後，自九月至十二月，卽繼以猛烈之擡高。一八六七年十二月起，又形式微，直至一八六八年九月，始告一段落，而同年十月十一月及十二月，反甚高漲；由此

觀之,失業數字之擡高,在一般情形之下,非至八月後,不能出現,可知一般低落現象,並未使季節影響延長甚多也。復查次年,即一八六九年,在八月數字降落甚低,打破以往最低點之記錄,惟在十二月,數字升起程度甚微,再次一年,數字降低迅速,至八月而極,然因季節而應有之上昇,仍及時而至,並未稍見延擱也。於斯可見,除在開始一年即一八六六年尙未見有低落現象外,其餘各年,在全局之變動過程中,均受季節之影響,且各年之擡起,均見於秋季,似有一致之情形;此種情形,在一八七五年八月起至一八八一年五月止一期中均相同。然吾人試問:一八六七至一八七〇年,就業情狀,以何月最爲最惡劣?則一八六七年十二月之百分之二二·六,乃爲最大之數字,似爲十二月矣,然十二月之失業,向來比其他各月爲甚,而一月至六月共六個月中,任何一月之數字,頗有較爲反常之可能,反不若十二月情形之一致,故欲查就業狀況最惡劣之確定日期,非用圖示法,必不能得出最優良之表明。下列圖式爲英國鑄鐵工人協進會前任書記海君(Mr. Hey)所繪,上文所言各點,即多爲海君所提示,吾人於此不妨一併聲明。

圖式說明　上列第十四圖,內中包含之事實,如下所述。重線之作用,在表示每年之平均百分數,吾人試察該線,必見一八五七年以前之步步趨跌趨勢,至一八五七年,驟然上騰,一八五八至一八六〇年,又突然下降,嗣後復經過兩年之上昇,乃達於原來

之水準，至一八六五年，竟又回復至最低點；次一波浪形，自一八六五年起至一八七二年止，經過七年之久，在此一波浪形中，最令人注意者，爲一八六七年之異常挺進，及一八七二年之重現極低新記錄。一八七二年之特異貿易狀態，並未保持甚久，即行漸漸擡起，直至一八七六年爲止，於是另有一個六年期之波浪形開始，而成另一商業循環矣。此波浪形之升進，已非如前一波形運動之峻峭，惟經過時間之長久，昇進最高點之高度，則有過之無不及。反之，在降落之時，下降角度，幾與上昇所成角度相同，且在一八八二年所現之最低點，亦與在一八六五年之最低點大致相同。次一波浪形，本尙未至理應開始之年，乃竟提前實現，且波浪形之經過，有七年之久，亦非復爲六年，惟上升下降情形，均極緩和，特上昇尙較下降爲陡峭耳。一八八九年所呈現之最低點，並未維持甚久，即行逐漸上昇，其後圖雖中止，但吾人得知其最高點，將見於一八九四年，惟高度尙爲緩和而已，又如支配以前各商業蕭條時期之原因依然存在，則下一最低點可望於一八九八或一八九九年實現也。實際上，據英國商業部之調查，就有報告寄還之各職工組合綜合觀之，最高月份爲一八九二年之十二月，最高年份爲一八九三年；嗣後即逐步降落，至一八九七年而達於最低點，一八九八年一度輕微漲起之後，又於一八九九年重現極低之數字焉。(註一〇)

圖戊所表示者爲就業人數所佔之百分數，故此圖式恰爲以上各圖之倒置，而篇幅則大爲縮小。在一八七六至一八八二年一期之間，如以兩條縱線劃開，則依圖所示，在此數年中，英國全國損失之工作量，何等之大，鑄鐵工人協進會會員損失之工資，爲數何等之鉅，讀者必可一目瞭然。由此數字，指示吾人在鑄鐵工人協進會中，因特殊原因而失業之百分數，猶大於其他職工組合（只就有同期數字可資比較者而言）失業人數百分數之兩倍。

圖戊中，每年平均數用本章第一節所論修勻曲線法將之修勻，而以每七年一平均之平均數，(註一一) 適應一般之波長(wave length)。由此可知，在此三十九年中，運動或上或下，並無顯著之趨勢可言，且已修勻之曲線，與就業人數之總平均數，卽91.7，相距尙不甚遠也。

此一圖式，最妙莫過於與前列第六圖（英國國產貨物出口總值）作比較觀。茲將二圖比較之結果，擇要揭露如下：

日期		日期	
出口貨值最低點	失業人數最高點	出口貨值最高點	失業人數最低點
1862	1858及1862	1866	1865
1868	1868	1872	1872
1879	1879	1882	1882或1883
1886	1886	1890	1889
1894	1893		

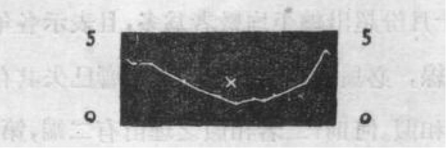
此外，此兩種數字，如用前節或下節所述方法，用圖比較之亦可。

季節影響之測算 在一八五五至一八七三年中十九個一月，十九個二月，十九個三月……等月份之平均數，在一八七四至一八九三年二十個一月，二十個二月……等月份之平均數，以及自一八五五至一八九三年全期中同月份之平均數，均已見於前列第三十六表，並列於第十四圖之乙，丙，丁。當吾人計算前論之各年平均數時，旣已用上述方法，將季節變動剔除；今用各月之平均數，吾人乃又剔除特定年份之影響。例如，設吾人從一數列完全不受季節影響之數字（假如果有此種數字）中，將所有十一月份之數字提出，而以此十一月份之數字，與所有各月之總平均數作比較，如就通盤計算，則在此平均數之上者，與在此平均數之下者，二者出現之次數，必完全相同。然各月數字，果已受季節之影響，則非在在平均數上之數字，較在平均數下之數字甚多，必在平均數上者，較在平均數下者過少。此過多過少之數，卽視季節影響而定，凡季節影響愈大者，則相差之數亦愈大。今用此方法，將數字施以平均，必可剔除非屬於季節之影響。設非用平均方法，勢必難以剔除，否則必須從根本就實際計算而來。季節原因之影響如使十一月超出總平均數之數，較使十月份超出總平均數之數爲大，則各年十一月份數字之平均數，超出總平均數之

數，亦必比十月份超出總平均數者爲多，且表示各年某一月份之平均數之曲線，必與因非屬於季節之影響已失其存在而得出之平均數曲線相似。何則？二者相似之理由有二端，第一，因在比較短期而爲吾人認爲足够應用之數年中，必有一極有力之非屬於季節之影響，對此一平均數發生永遠顯著之作用，如在前列第三十六表所表示者即是；第二，因爲季節原因以及非屬於季節之原因，時常互有影響，並非各自獨立；例如，商業蕭條時期，每因酷寒之冬日，而加強其深刻之程度；又如商業不振之年，復遇生意淸淡之月，則失業人數必驟然大增，反之在繁榮之年，加臨生氣蓬勃之夏季，失業人必大爲減少，甚可幾至於零。故在一方面，各種不同原因之相互作用，可以擡高季節之最高度，復可壓低季節之最低度，但在另一方面，二者相互爲用，因生補償作用，反可緩和尖銳之程度也。

在前列第十四圖乙，丙，丁三圖中，代表後半年之曲線，置於歷年後半年曲線之前，此乃因每年波浪形之特性，必由最低而最高，始最爲顯明之故。試觀圖丙之波浪形，形式稍欠穩定，而漲落則較圖乙所示前一期中之漲落爲衰弱；此無他，季節已失其影響故也。

每年之波浪形　每年一度之波浪形，如甚穩定，如圖丁所示者，則類似下圖形式：



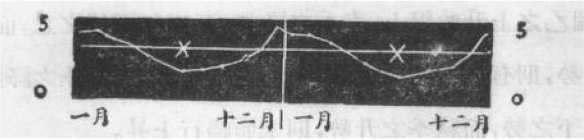
（第十五圖）

之數字，每年必與第十四圖甲相同；一八六四年，一八八二年及其他數年，卽其最顯著者也。就所有各年觀察，絕大部份，每年最高度，非現於十二月，卽在一月；在一八五八年之末，並無最高點出現，但因下降過速，致現折斷之象；一八六〇年之末，曲線居然上升，但一升之後，又發生春季之瀉落，終以一般上升趨勢甚強，瀉落未久，卽行轉而向上邁進；自一八六一年以後，尙有劇烈變動甚多，可依此類推，姑不一一敍述。總之，沿線之突出點，恰在平均數線之上，而各點相距大致相等，則無可疑。

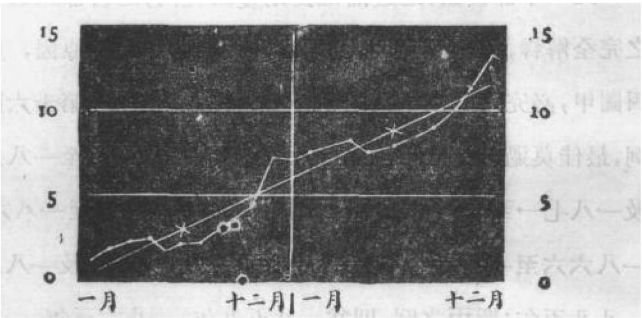
各最低點，尙欠十分明顯，此係因各點連結而成之形式不明之故，是以每以些須原因，各點卽趨近修勻曲線，且各最低點每隨一般降落趨勢而被遮蔽，或因一般上昇趨勢而致消失。雖然，在一八六一年，雖同時有強烈之上升趨勢，而最低點仍甚鮮明；在一八六五至一八七〇年變動行程中，各最低點亦頗顯著；至於自一八五九至一八八八年間，除一八七六，一八八〇，及一八八一年三年爲例外外，其餘各年，遇有最低點出現時，亦尙易於辨認也。

茲爲示明每年平均數之靜止，上升，及下落等傾向，對於季節波浪形之影響，分作三圖於下：

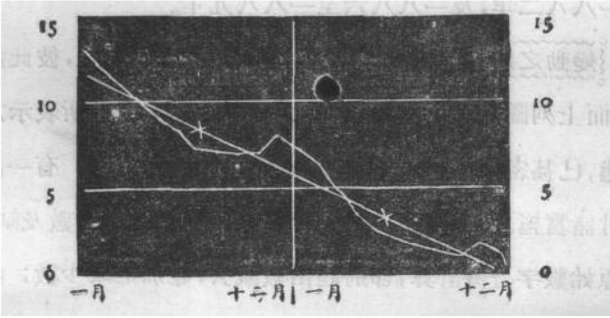
甲、平均數線在靜止狀態下之季節波浪形



乙、平均數線在上升狀態下之季節波浪形



丙、平均數線在下降狀態下之季節波浪形



（第十六圖）

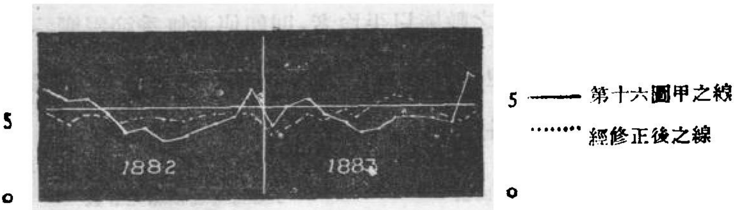
此三圖式,係用每月對總平均數之平均差額(詳言之,即

一月	二月	三月	四月	五月	六月	七月	八月	九月	十月	十一月	十二月
+1.1	+1.0	+.6	+.2	-.4	-.7	-1.1	-.9	-.7	-.4	0	+1.6

)按月加上連結每年平均數之直線上之距離,或由該直線上之距離減去之。圖乙之上升曲線上,春季之降勢,有趨向平行之意,而秋季上升之勢,則有更趨陡峭之意;圖丙下降曲線上,春季之降勢,有急轉直下之勢,而秋季之升勢,則未能暢行上昇。

此一季節波動,連遲滯之長期變動在內,是否即爲數字變動之完全解釋,尙難確定,如其果爲數字變動之唯一原因,則第十四圖甲,必完全爲第十六圖甲,乙,丙三圖所合成。第十六圖甲之例,最佳莫過於一八五五至一八五七年,一八六四至一八六五年,及一八七一至一八七三年;圖乙之例,爲一八六〇至一八六一年,一八六六至一八六七年,一八七七至一八七八年,及一八八三至一八八五年;圖丙之例,則爲一八五九年,一八六三年,一八八〇至一八八二年,及一八八六至一八八九年。

變動之剔除 依上文所述,變動之原因有兩組,彼此並非獨立,而上列圖式之再繪製,又非十分準確;然由此所表示之相合形態,已甚密切,故以下所論之剔除季節變動方法,有一部份頗可付諸實施。將以上所得之各月對總平均數之超出數及缺少數,與原始數字合併計算,即將超出數減去,並加上缺少數;如此則可產生一直線如下圖:



（第　十　七　圖）

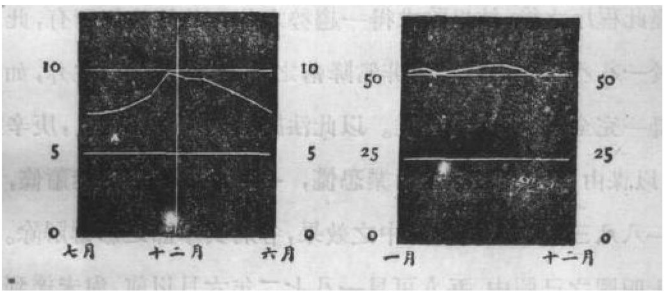
但經此程序之後，結果除求得一趨勢之外，其餘並無所有，此乃由於一八八三年一月曾有非常降落之故，然除此一期之外，如欲尋得一完全之例，其事甚難。以此法施於第十四圖之己，庚辛三圖，以謀由一八七二年之商業恐慌，一八七九年之商業蕭條，以及一八八三年之潮流轉向中之效果，各將其季節之影響解除。在第十四圖之己圖中，吾人可見一八七二年六月以前，尚未達到絕對最低點，但該年之一月乃爲比較最佳之月份。由此可知一八七二年一月，乃爲商業全盛時代之轉向點，惟此月份爲期則較其他一般轉向期爲早。但一八七九年最高點之日期，經此法修正之後，並未有何改變，至一八八九年最高點之日期，亦不過移過一月而已。

週期存在之判準　討論至此，關於週期存在之判別標準，不得不加以考慮。在第十四圖甲圖中，即憑吾人目力，已足判明每年一度之週期，然用一代表小麥價格之圖式，是否可以引起每年一度之變動，則不無可疑。總而言之，任何調查材料之每月數字，

凡就若干年將同月份之數施以平均者，則卽使並無季節影響，一月，二月……以及其他各月之平均數，必不能準確相等。下列第十八圖各式，卽表示各種平均數者也：

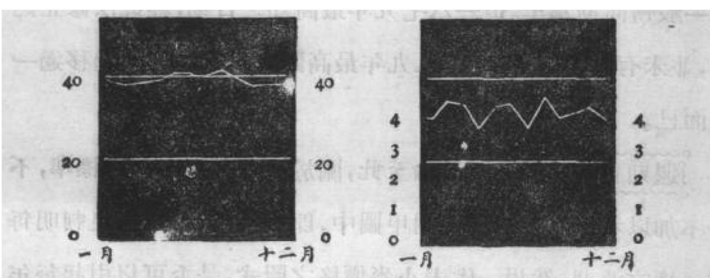
鐵工失業人數同前　　　　小麥每夸特價格先令數

一八六二至一八七六



小麥價格每夸特先令數　　　　各月第一星期日平均日期

一八七七至一八九一　　　　一八八一至一九〇〇



（第　十　八　圖）

在此四種圖式中，前三者可謂之爲有季節性，末一圖，旣係表示二十年中各月第一個星期日所在日期之平均數，自非屬於季節性質。

此點，可另作一簡單測驗於下，以決定之。如有一種週期循環，勿論任何種形式，凡與季節有關聯者，則此一週期，必與普通天氣如溫度及其他等圖，多少相適應，蓋天氣圖，不論溫度與雨量，每平均僅有一最高點及最低點也。受天氣影響之現象，自亦應表現唯一最高點，此一最高點，大致必與最高溫度或最低溫度相合。例如，失業人數最高度，卽與晝長最低度相合，而在最低溫度之前。在幾種情形之下，除一種主要効果之外，同時又有一第二附帶發生之最高度出現，例如，一逾額之死亡率，或許由於炎寒或酷暑之故；然卽就此例而論，如再經進一步之研究，必可證明，炎寒之最高度，足致老年人之死命，酷暑之最高度，足促成少年人之傷亡。溫度之例如此，小麥之價格，亦何莫不然，地球有南北兩半球，收穫時間不同，因而小麥價格，乃出現兩個最低點。第十八圖之星期日曲線，最高點有四個之多，故與季節無關。最高點不止一個而有兩個或兩個以上之多，卽足證明並無週期性存乎其間，否則必須別有理由證明確有週期性存在也。

機率測驗　第二種測驗法，卽考查時間序列圖，而分析其在同一月中最高點出現次數；非屬於週期性之原因，動輒遮蔽最高

點，但經長期序列之觀察，必有一月特別顯露其優越性。第十四圖甲圖中，三月及四月出現之最高點，每月各有二次，二月則有三次，一月則有十一次，而在十二月則有二十一次。惟此處之最低點，則尚欠鮮明。

將此分析引而申之，茲作一表如下；

	三十九年中出現之次數
十二月份之百分數，大於同年十一月份之百分數者	33
十二月份之百分數，大於次年一月之百分數者	28
十二月份之百分數，大於同年七月份之百分數者	33
十二月份之百分數，大於次年七月份之百分數者	30

設季節並無影響，十二月份已佔絕大之優勢，若反乎此種優勢而出現之機率，必分別爲六萬五千次對一次，一百六十次對一次，六萬五千次對一次，及一千二百次對一次。依此方法，可將所有各月份逐一加以測驗。惟此種測驗方法，並不能將驗證發揮無餘，因吾人以上僅以兩月作比較，而討論兩月中以何月份之百分數爲較大，但此月果大於彼月，究有幾何，並未論及。關於此點，讀者可請參閱，英國皇家統計學會五十週紀念刊第二百零六頁，愛基華斯教授(Edgeworth)之統計方法論(On Method of Statistics)一文；雖然，吾人尚須在第二編將數學方法研究旣畢之時，方可對此加以討論也。

第五節　對數曲線

以前各節所述圖示法之嚴重缺陷,乃爲:當吾人討論一列漸增數字時,即使各年之總計數,或許爲漸增數字,因爲方法所限,亦只能以相等之縱距離,以代表各年漸增總計數之相等增加量。譬如,一年之總計數爲二〇鎊,另一年之總計數,則爲二〇〇〇鎊,現二年各增加二〇鎊,則各年增加量,只有以同一縱距離表示之之一途。試以下列第十九圖爲例,該圖表示英國十九世紀進出口貿易之進展;先就代表出口貨值之曲線而論,在一八一五至一八一六年,出口貨值由五二,〇〇〇,〇〇〇鎊,一降而爲四二,〇〇〇,〇〇〇鎊,雖降落已達百分之二十,但由圖觀之,並不甚引人注意,然當此之時英國工業區正在極度恐慌中也。至於一八八三年,由三〇五,〇〇〇,〇〇〇鎊,至一八八六年,降爲二六九,〇〇〇,〇〇〇鎊,雖僅落下百分之十二,竟已立即引人注意。又如一八四八至一八五〇年,出口貨值增加有百分之三十四,然以與一八七〇至一八七二年僅增加百分之二十九者比較,似反無甚重要可言。故吾人在研究因果關係之時,與其求得實在增加額,究不若計算比例增加率之爲愈。吾人在討論一國對外貿易之

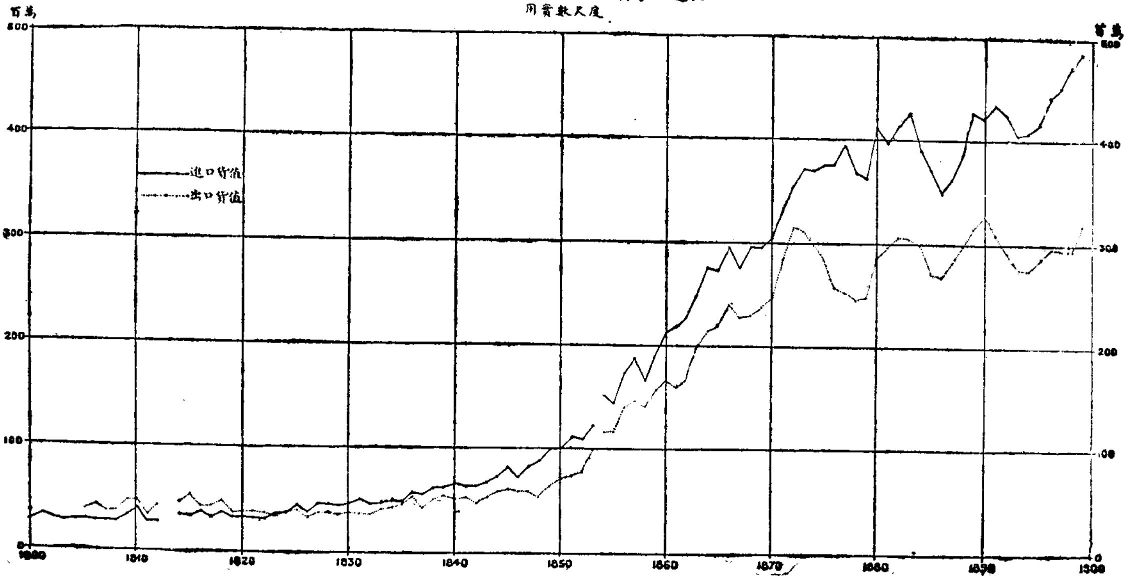
逐漸進展，或比較兩國貿易之發達時，如以實數尺度作圖如下列第十九圖甲圖者，往往對於早期之掙扎時代發生極荒謬之印象。爲今之計，吾人所需要者，非爲表示數量之圖，乃爲表示比率之圖，在此表示比率之圖上，相等之縱距離所表示者，已非爲相等之絕對增加量，而爲相等之比例增加量，換言之，卽相等之增加率也。今應用對數，乃得作成普遍一致之尺度，於是吾人之目的乃可於以完成。不諳數學之學生，對於如此構成之圖，甚易養成應用之習慣，只須（一）研究記載實在數之圖式，（二）並不論所用尺度爲何一部分，均須留心分別以同一之縱距離，代表雙倍，半倍，或百分之二十……以及其他等等。

對數圖之構造法　用此一尺度構圖，其程序有如下述：(一)將所欲代表之序列數字記下；（二）在各數字之側；查出各該數之對數，記下；（三）在紙上分成相等之正方形，並沿縱線，以相等之間距，將數字按有規則之級數，依照漸次上升次序，分列於上，但須足以包括所有查得之對數；（四）在橫線上，分別標列日期；（五）在如此作成之尺度上，卽將對數記上，不必再用原始數字。下列第三十七圖：

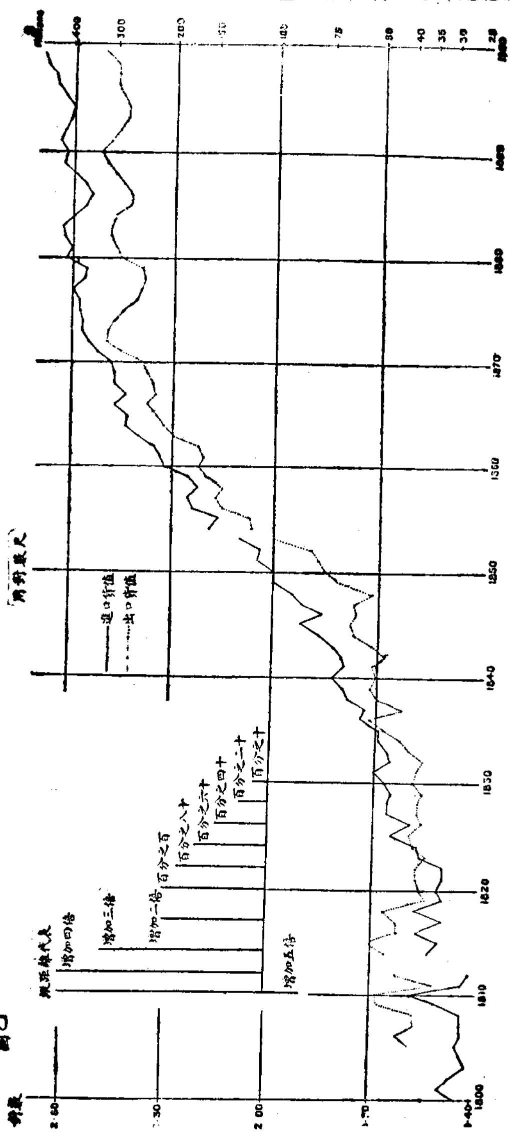
觀下列第十九圖

圖甲

英國十九世紀間進出口貿易之進展
用實數尺度



（第　十　九　圖）


圖乙
進口貨值
出口貨值
增加五倍
增加四倍
增加三倍
增加二倍
百分之百
百分之八十
百分之六十
百分之四十
百分之二十
百分之十
1800
1810
1820
1830
1840
1850
1860
1870
1880
1890
1900

(續第十九圖)

所表示之數字，卽爲英國之出入口貨值，而用對數尺度作成者。第十九圖乙圖之右方，列有絕對數字；左方卽爲相當於絕對數字之對數。某一縱距離，例如一英寸，爲對數尺度0.301之距離；吾人如以此數加在任一數之對數上，則得該數之兩倍之對數，因爲 $\log a+.301=\log a+\log 2=\log 2a$;，舉例言之，設吾人欲將代表三十鎊之距離，增加長度一英寸，則結果位置必在代表六十鎊之距離。反之，設吾人將一英寸之1.59（在對數尺度上代表0.477亦卽 $\log 3$）加在$2a$之對數上，則必得$6a$，故

$$\log 6a=.477+\log 2a=.477+.301+\log a,\text{理由如上述}$$
$$=.778+\log a=\log 6+\log a;$$

換言之，吾人在此尺度上求得同一之位置，其途徑有二，一爲用兩個分立之比率，一爲僅一複比率。依此原則繪成之圖式，必能滿足兩種必需之條件；一則相等之縱距離，不論所用爲何一部尺度，均須代表同一之比率，二則不論點數若干，均可記錄於圖上，而不致發生矛盾。本章本節之末，附有修正至小數點三位止自一至一千之對數表，此表對於上述之用途，已足應付裕如。

對數圖之應用舉例　試觀上列第十九圖，可見英國之進口貨值，自一八一一年，至一八三六年，貨值已變爲兩倍，至一八五三年，復變爲一八三九年之兩倍，至一八六六年又變爲一八五五年之兩倍，而在一八九九年之進口貨值，則較之一八八六年已增

加百分之四十。由此圖上，同時可加注意者，尙有兩點，一則在一八五〇及一八八〇年，英國進口貨值超出出口貨值之數，恰爲出口貨值之百分之四十，二則一八九九年之進口貨值，當一八六〇年出口貨值之三倍。

假如吾人目力果已經過精密之訓練，能以了解諸如此類之圖式，又假如吾人深知此一圖式乃爲一比率圖而非數量圖，而將此事實深印心中，則此圖必完全適合圖示法之宗旨；換言之，即使吾人對於緊複事件，立可產生眞實之印象。反之，設并此眞實印象而不可得，此係由於不能使心理成正當狀態，則只能採用實數尺度之圖式，然吾人不可不知，此種僅用實數尺度之圖式，每予吾人以背謬之比率印象也。(註一二)

速度及加速度 尙有一點，吾人務須注意者，即在此類圖式中，不須再畫基線，否則必予人以一錯誤之印象。此外，尤須注意者，在相等之縱距離，代表圖上任何部分彼此相互之相等比率，而非代表在實數尺度上之相等增加量時，相等之斜度(degree of slope) 實即代表相等之增加率（相等加速度），而非代表在實數尺度上相等時間之相等增加量（相等速度）。在用對數尺度時，一上升之線，如與橫線成凸形，乃表示增加率之擴展，如一八三〇至一八五三年之進口貨值（設此曲線已修勻過）。反之，如一線與橫線成凹形，即表示增加率之遲滯，如一八五四至一八七

三年之進口貨値是。但若在實數尺度圖上，則在一八三〇至一八五三年，及一八五四至一八七三年兩期中，曲線幾全呈凸形，此表示實在增加量無時不在增加也。

對數尺度對於指數應用之有利　如果篇幅許可，頗可多列數圖，兼用兩種尺度；蓋在極多之數列中，用此兩種方法後，顯示出極大之差異，而惟此差異，乃可爲吾人之教訓。今有一種情形，可以明白指示，卽對數尺度所以須特別重視者，卽在原始數字所表示者爲比率而非實在數字之時。例如孫巴克(Sauerbeck)氏有名之圖式，乃用實數尺度繪成，該圖之作用，乃在表示彼之物價指數，在此圖式中，所列之數字，乃爲某若干年中之物價百分數。玆設三年之指數，爲一〇〇，八〇，六〇，如用實數尺度，則逐年之減少量，必將以相等之距離表示之，因而似有相等之情勢。然金價之變動，在兩時期絕無完全相等之理。第一期，由一〇〇落至八〇，乃降落百分之二十；在前年價値一金鎊之貨物，去年乃以十六先令購得之。第二期，由八〇落至六〇，降落達百分之二十五；在去年價値一金鎊之貨物，今年竟以十五先令購得之。故就對於物價指數之應用而論，比率最爲重要，作圖所應表示者，亦唯比率一端也。

用對數尺度作比較　對數尺度對於數列之比較，有其特別之用途，專論對數尺度之節目中，所討論之方法，可以立卽採用

之。在比較性質不同之數量，以擇定單位爲最困難，但此困難，唯在以比率相比較時，乃可消滅；故從此對於百分數方法，不致再感煩難矣。夫在調查因果關係之時，在比率中所發見之密切關聯，遠比在實在數量中所見者爲多。蓋一組現象，與另一組現象，如無關聯則已，如有關聯，則二者之關係，大半以比例關係（例如，某一現象可量性質每增加百分之十，另一現象之可量性質，則增加百分之八）時爲多，而爲絕對數量之關係（例如，一種物價，不論原來位置如何，凡每增加二先令，則買主必減少一百人）時則較少。今用對數尺度，設兩曲線如有類似狀態，則其意義即足表示兩曲線，以比例的變動相適應，但如用實數尺度，則兩曲線之類似狀態，不過表示絕對的變動相應和而已。

不特此也，在此新方法之下，如欲令兩平均數相等，亦不若前此之煩難。蓋一經取用兩組數列之對數，此兩數列在對數尺度上之地位，高度幾何，並不重要；高度之變換，不過意即以一常數，乘所有各項(item) 而已，並未變換各項之形態，或變動之比例也。用此方法，其程序可列之如下：（一）在對數尺度上，將代表兩組數列之曲線畫出；（二）然後將下面之一曲線，縱直向上移動，使接近上面之曲線，使二曲線相疊至二曲線可能之相應狀態，最爲接近爲止；（三）即在此地位，將此下一曲線繪成，於是兩組數列，可以從事比較矣。

變動方程式　茲為明瞭此一方法起見，特以英國職工組合有業人數之百分數，與結婚率比較為例，列如第三十八，三十九兩表，及第二十圖。

結婚率與有業人數

每千人中結婚率

年度	最高點	最低點	差額	對最高額之百分數
一八六九	…	15.9		
			1.7	10
一八七三	17.6	…		
			3.2	18
一八七九	…	14.4		
			1.1	7
一八八二至八三	15.5	…		
			1.3	8
一八八六	…	14.2		
			1.4	9
一八九一	15.6	…		
			.9	6
一八九三	…	14.7		
				9.7

有業人數所佔之百分數

年度	最高點	最低點	差額	對最高額之百分數
一八六八	…	91.5		
			7.4	7.5
一八七二	98.9	…		
			11.4	11.5
一八七九	…	87.5		
			10.6	10.8
一八八二	98.1	…		
			7.6	7.8
一八八六	…	90.5		
			7.4	7.6
一八八九至九〇	97.9	…		
一八九三	…	92.5	5.4	5.5
				8.4

有業人數在總人數中所佔之平均百分數在一八六五至一八九三年，為百分之九十五·一；九五·一之百分之八·四，即為八二·四之百分之九·七。

（第三十八表）

下列圖甲，數字用實數尺度表示；圖乙，二十九年間之各平均數，均令其完全相等，數字則用對數尺度表示之。此時可依照本章第十三節所論之構圖法辦理；惟方法不妨別加改良，如第三十八表，列有各時期中之最高及最低點，並將由最高點降為最低點時之

年　度	結婚率	對　數	有業人數百分數	減去12.7	對　數
一八六五	17.5	1.243	98.0	85.3	1.931
一八六六	17.5	1.243	96.9	84.1	1.925
一八六七	16.5	1.217	92.7	80.0	1.903
一八六八	16.1	1.207	91.5	78.8	1.896
一八六九	15.9	1.201	92.6	79.9	1.902
一八七〇	16.1	1.207	95.7	83.0	1.919
一八七一	16.7	1.223	98.2	85.5	1.932
一八七二	17.4	1.240	98.9	86.2	1.935
一八七三	17.6	1.245	98.7	86.0	1.934
一八七四	17.0	1.230	98.2	85.5	1.932
一八七五	16.7	1.223	97.5	84.8	1.928
一八七六	16.5	1.217	96.4	83.7	1.923
一八七七	15.7	1.196	95.6	82.9	1.919
一八七八	15.2	1.182	93.7	81.0	1.908
一八七九	14.4	1.158	87.5	74.8	1.874
一八八〇	14.9	1.173	94.1	81.4	1.911
一八八一	15.1	1.179	96.5	83.8	1.923
一八八二	15.5	1.190	98.1	85.4	1.931
一八八三	15.5	1.190	97.8	85.1	1.930
一八八四	15.1	1.179	92.6	79.9	1.902
一八八五	14.5	1.161	91.0	78.3	1.894
一八八六	14.2	1.152	90.45	77.7	1.890
一八八七	14.4	1.158	92.6	79.9	1.902
一八八八	14.4	1.158	95.2	82.5	1.916
一八八九	15.0	1.176	97.9	85.2	1.930
一八九〇	15.5	1.190	97.9	85.2	1.930
一八九一	15.6	1.193	96.5	83.8	1.923
一八九二	15.4	1.187	93.7	81.0	1.908
一八九三	14.7	1.167	92.5	79.8	1.902
		平均數 1.196			平均數 1.916

（第三十九表）

變動之平均數(表上列有對最高額之百分數，意卽指此)算出。據查在有調查表可稽之職工組合中，有業人數之變動，爲百分之八·四，恰與結婚率百分之九·七之變動相當。(註一三) 如欲求更密切

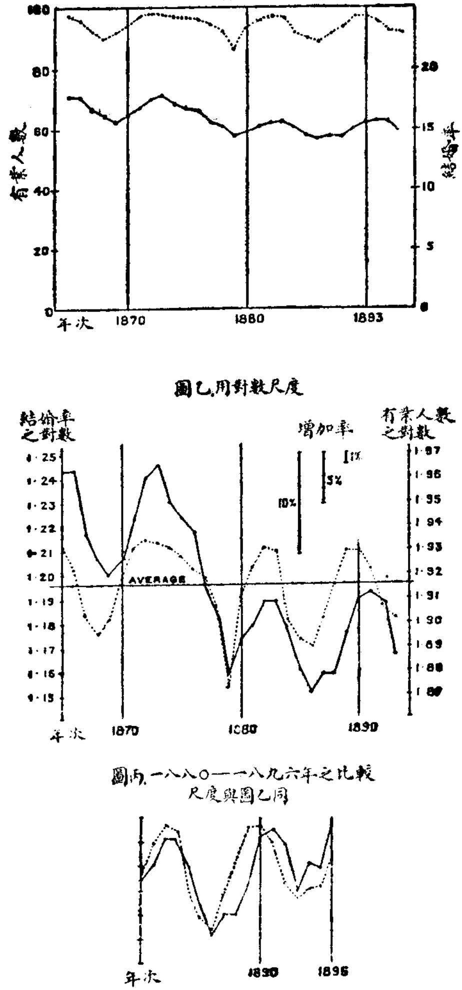
一八六五至一八九三年之比較

圖甲　用實數尺度

結婚率與有業人數

一八六五—一九三年之比較

圖甲用實數尺度



結婚率　　有業人數

(第二十圖)

之相應,可假定有業人數之一部,並不致影響結婚率,並求出在此總數之百分之八·四,成為餘額之百分之九·七以前,究應扣除幾何;在擇定之期間,職工組合會員有業人數之平均百分數,為九五·一,此九五·一之百分之八·四,為七·九九,而此七·九九即為八二·四之百分之九·七。例如,九五·一與八二·四之差為一二·七,此一二·七可視為與本問題無關之數,故在未查出對數以前,即從各數中將此一二·七扣除。用實數尺度時,此種程序可令兩組數列之平均數相等以代替之,並在一基線之下,另畫一基線,使與上一基線,距相當之距離,以便兩數列之同一縱距離,乃可代表平均之變動;此一程序純與前在本章第三節末段所述者完全相同。如以代數表示之,可得下列方程式:

$$\log(y-c)-\log x=k,\ k\text{為一常數},$$

在此式中,c 及 k 為常數,此二常數,如何擇定,完全視能否使配合最為密切為斷,至 x 及 y,乃即欲作比較之二數量也。

在上列第二十圖甲,數字用實數尺度表示之,圖乙則用對數尺度表示之,並已令兩組變動平均百分數相等矣;圖丙,自一八八〇至一八九六年,期間較短,方法恰與圖乙所用者相同。至實在數字及其對數,則列於上列第三十九表。

費許 (Irving Fisher) 教授,於一九一七年六月,在美國統計學會之季刊中,曾發表一文,論述對數曲線,極稱對數曲線之妙用。茲特列恰足供吾人應用之三位小數對數表於下:

對數表

數字	對數	數字	對數	數字	對數	數字	對數	數字	對數	數字	對數	數字	對數
1	0	34	1.531	67	1.826	101	2.004	134	2.127	167	2.223	201	2.303
2	.301	35	1.544	68	1.833	102	2.009	135	2.130	168	2.225	202	2.305
3	.477	36	1.556	69	1.839	103	2.013	136	2.134	169	2.228	203	2.307
4	.602	37	1.568	70	1.845	104	2.017	137	2.137	170	2.230	204	2.310
5	.699	38	1.580	71	1.851	105	2.021	138	2.140	171	2.233	205	2.312
6	.778	39	1.591	72	1.857	106	2.025	139	2.143	172	2.236	206	2.314
7	.845	40	1.602	73	1.863	107	2.029	140	2.146	173	2.238	207	2.316
8	.903	41	1.613	74	1.869	108	2.033	141	2.149	174	2.241	208	2.318
9	.954	42	1.623	75	1.875	109	2.037	142	2.152	175	2.243	209	2.320
10	1.	43	1.633	76	1.881	110	2.041	143	2.155	176	2.246	210	2.322
11	1.041	44	1.643	77	1.886	111	2.045	144	2.158	177	2.248	211	2.324
12	1.079	45	1.653	78	1.892	112	2.049	145	2.161	178	2.250	212	2.326
13	1.114	46	1.663	79	1.898	113	2.053	146	2.164	179	2.253	213	2.328
14	1.146	47	1.672	80	1.903	114	2.057	147	2.167	180	2.255	214	2.330
15	1.176	48	1.681	81	1.908	115	2.061	148	2.170	181	2.258	215	2.332
16	1.204	49	1.690	82	1.914	116	2.064	149	2.173	182	2.260	216	2.334
17	1.230	50	1.699	83	1.919	117	2.068	150	2.176	183	2.262	217	2.336
18	1.255	51	1.708	84	1.924	118	2.072	151	2.179	184	2.265	218	2.338
19	1.279	52	1.716	85	1.929	119	2.076	152	2.182	185	2.267	219	2.340
20	1.301	53	1.724	86	1.934	120	2.079	153	2.185	186	2.270	220	2.342
21	1.322	54	1.732	87	1.940	121	2.083	154	2.188	187	2.272	221	2.344
22	1.342	55	1.740	88	1.944	122	2.086	155	2.190	188	2.274	222	2.346
23	1.362	56	1.748	89	1.949	123	2.090	156	2.193	189	2.276	223	2.348
24	1.380	57	1.756	90	1.954	124	2.093	157	2.196	190	2.279	224	2.350
25	1.398	58	1.763	91	1.959	125	2.097	158	2.199	191	2.281	225	2.352
26	1.415	59	1.771	92	1.964	126	2.100	159	2.201	192	2.283	226	2.354
27	1.431	60	1.778	93	1.968	127	2.104	160	2.204	193	2.286	227	2.356
28	1.447	61	1.785	94	1.973	128	2.107	161	2.207	194	2.288	228	2.358
29	1.462	62	1.792	95	1.978	129	2.111	162	2.210	195	2.290	229	2.360
30	1.477	63	1.799	96	1.982	130	2.114	163	2.212	196	2.292	230	2.362
31	1.491	64	1.806	97	1.987	131	2.117	164	2.215	197	9.294	231	2.364
32	1.505	65	1.813	98	1.991	132	2.121	165	2.217	198	2.297	232	2.365
33	1.519	66	1.820	99	1 996	133	2.124	166	2.220	199	2.299	233	2.367
				100	2.000					200	2.301		

數字	對數	數字	對數	數字	對數	數字	對數	數字	對數	數字	對數
234	2.369	267	2.427	301	2.479	335	2.525	401	2.603	467	2.669
235	2.371	268	2.428	302	2.480	337	2.528	403	2.605	469	2.671
236	2.373	269	2.430	303	2.481	339	2.530	405	2.607	471	2.673
237	2.375	270	2.431	304	2.483	341	2.533	407	2.610	473	2.675
238	2.377	271	2.433	305	2.484	343	2.535	409	2.612	475	2.677
239	2.378	272	2.435	306	2.486	345	2.538	411	2.614	477	2.679
240	2.380	273	2.436	307	2.487	347	2.540	413	2.616	479	2.680
241	2.382	274	2.438	308	2.489	349	2.543	415	2.618	481	2.682
242	2.384	275	2.439	309	2.490	351	2.545	417	2.620	483	2.684
243	2.386	276	2.441	310	2.491	353	2.548	419	2.622	485	2.686
244	2.387	277	2.442	311	2.493	355	2.550	421	2.624	487	2.688
245	2.389	278	2.444	312	2.494	357	2.553	423	2.626	489	2.689
246	2.391	279	2.446	313	2.495	359	2.555	425	2.628	491	2.691
247	2.393	280	2.447	314	2.497	361	2.558	427	2.630	493	2.693
248	2.394	281	2.449	315	2.498	363	2.560	429	2.632	495	2.695
249	2.396	282	2.450	316	2.500	365	2.562	431	2.634	497	2.696
250	2.398	283	2.452	317	2.501	367	2.565	433	2.636	499	2.698
251	2.400	284	2.453	318	2.502	369	2.567	435	2.638	501	2.700
252	2.401	285	2.455	319	2.504	371	2.569	437	2.640	503	2.702
253	2.403	286	2.456	320	2.505	373	2.572	439	2.642	505	2.703
254	2.405	287	2.458	321	2.507	375	2.574	441	2.644	507	2.705
255	2.407	288	2.459	322	2.508	377	2.576	443	2.646	509	2.707
256	2.408	289	2.461	323	2.509	379	2.579	445	2.648	511	2.708
257	2.410	290	2.462	324	2.511	381	2.581	447	2.650	513	2.710
258	2.412	291	2.464	325	2.512	383	2.583	449	2.652	515	2.712
259	2.413	292	2.465	326	2.513	385	2.585	451	2.654	517	2.713
260	2.415	293	2.467	327	2.515	387	2.588	453	2.656	519	2.715
261	2.417	294	2.468	328	2.516	389	2.590	455	2.658	521	2.717
262	2.418	295	2.470	329	2.517	391	2.592	457	2.660	523	2.719
263	2.420	296	2.471	330	2.519	393	2.594	459	2.662	525	2.720
264	2.422	297	2.473	331	2.520	395	2.597	461	2.664	527	2.722
265	2.423	298	2.474	332	2.521	397	2.599	463	2.666	529	2.723
266	2.425	299	2.476	333	2.522	399	2.601	465	2.667	531	2.725
		300	2.477							533	2.727

數字	對數	數字	對數	數字	對數	數字	對數	數字	對數	數字	對數
535	2.728	601	2.779	667	2.824	735	2.866	801	2.904	900	2.954
537	2.730	603	2.780	669	2.825	737	2.867	804	2.905	903	2.956
539	2.732	605	2.282	671	2.827	739	2.869	807	2.907	906	2.957
541	2.733	607	2.783	673	2.828	741	2.870	810	2.908	909	2.959
543	2.735	609	2.785	675	2.829	743	2.871	813	2.910	912	2.960
545	2.736	611	2.786	677	2.831	745	2.872	816	2.912	915	2.961
547	2.738	613	2.787	679	2.832	747	2.873	819	2.913	918	2.963
549	2.740	615	2.789	681	2.833	749	2.874	822	2.915	921	2.964
551	2.741	617	2.790	683	2.834	751	2.876	825	2.916	924	2.966
553	2.743	619	2.792	685	2.836	753	2.877	828	2.918	927	2.967
555	2.744	621	2.793	687	2.837	755	2.878	831	2.920	930	2.968
557	2.746	623	2.794	689	2.838	757	2.879	834	2.921	933	2.970
559	2.747	625	2.796	691	2.839	759	2.880	837	2.923	936	2.971
561	2.749	627	2.797	693	2.841	761	2.881	840	2.924	939	2.973
563	2.751	629	2.799	695	2.842	763	2.883	843	2.926	942	2.974
565	2.752	631	2.800	697	2.843	765	2.884	846	2.927	945	2.975
567	2.754	633	2.801	699	2.844	767	2.885	849	2.929	948	2.977
569	2.755	635	2.803	701	2.846	769	2.886	852	2.930	951	2.978
571	2.757	637	2.804	703	2.847	771	2.887	855	2.932	954	2.980
573	2.758	639	2.806	705	2.848	773	2.888	858	2.933	957	2.981
575	2.760	641	2.807	707	2.849	775	2.889	861	2.935	960	2.982
577	2.761	643	2.808	709	2.851	777	2.890	864	2.937	963	2.984
579	2.763	645	2.810	711	2.852	779	2.892	867	2.938	966	2.985
581	2.764	647	2.811	713	2.853	781	2.893	870	2.940	969	2.986
583	2.766	649	2.812	715	2.854	783	2.894	873	2.941	972	2.988
585	2.767	651	2.814	717	2.856	785	2.895	876	2.942	975	2.989
587	2.769	653	2.815	719	2.857	787	2.896	879	2.944	978	2.990
589	2.770	655	2.816	721	2.858	789	2.897	882	2.945	981	2.992
591	2.772	657	2.818	723	2.859	791	2.898	885	2.947	984	2.993
593	2.773	659	2.819	725	2.860	793	2.899	888	2.948	987	2.994
595	2.775	661	2.820	727	2.862	795	2.900	891	2.950	990	2.996
597	2.776	663	2.821	729	2.863	797	2.901	894	2.951	993	2.997
599	2.777	665	2.823	731	2.864	799	2.903	897	2.953	996	2.998
				733	2.865					1,000	3.000

（第四十表）

（註一） 見本章第三節及第五節

（註二） 表中有符號表示數不包括出口船隻價值

（註三） 在各種平均數曲線中，表示平均數之符號，位在表示所平均之三，五，或十五年數字之數符號之重心。

（註四） 表中有 * 號者表示該數並未將船隻出口貨值包括在內。

（註五） 因英帝國及地方租稅記載變更，後此各年度之圖式，不能繼續畫出。

（註六） 請參閱本書原著者之『十九世紀之工資』(Wages in the Nineteenth century)一書第九十頁附圖。

（註七） 此外尚有一種數列，其數字經數年始一變，在此數年中，數字完全相等，經過數年之後，忽一躍而至另一水準，既至另一水準之後，又保持不動者許久。此種數列之例證，為標準計時工資率 (Standard time-rates)卽是。

（註八） 見翟翁士氏著對於通貨及金融之調查(Investigation in Currency and Finance)。

（註九） 見翟翁士氏著對於通貨及金融之調查。

（註一〇） 請參閱英國一八九五年勞工統計簡報第七十三頁，該處對於類似上文所論各種數字之處置方法，列舉甚多。

（註一一） 關於曲線修勻方法及週期曲線之硏究，請參閱一八八四年布安廷教授(Poynting)及一九一九年摩爾教授 (Moore) 同在統計學報所發表之文。

（註一二） 馬夏爾(Marshall)教授)在彼『統計圖示法』(On the Graphic Method of Statistics) 一文中，建議糾正此一錯誤印象之簡單方法。該文刊於英國皇家統計學會學報五十週年紀念特刊，第二百五十七頁。

（註一三） 在第三十九表第二，四兩欄之數字，乃自務德 (G. H. Wood)先生在一九〇〇年英國統計學報發表之『一八六〇年關於工人階級發展之幾種統計』一文，並有一極有價值之對數表，可為本節若干點之例證。

第八章　確度

第一節　引論

度量之性質　無論在物體上，或在經濟上，均無完全準確度量之存在，正與無完全直線或完全流體之存在同。經濟度量之性質，最好可由物體度量性質考察而解明之。權衡物質而不致有一公分之差，其事甚易。若有精確之天平，吾人卽可因器械進步之故，在權衡上不致有一公毫，一公絲，乃至一公絲之十分之一之差。但欲獲得超過此等程限之確度，則無此天平矣。於角度之度量亦然。肉眼能辨別對角一度之三十分之一之物體。藉六分儀之助，則度量可精確至弧之十五秒。格林尼治(Greenwich)天文學家所作觀察，可精確至一秒之百分之一，逾此則不復可得。

在此等情形下，度量結果謂已精確至不致有一公絲之差，或多高程度。同樣，於錢財之計算上，吾人可以一鎊不差之錢幣估計數稱之。

物體與統計度量　太陽視差之度量，與某種統計的估計之工作，頗有相似之處。此種度量，係決定太陽與地球之距離者。十八世紀天文學家，估計此距離爲十秒，卽等於九千六百萬英里。

以觀察方法及工具改良之故，觀察者遂開始公認秒之整數爲八，但對於第一位小數之計算則異其詞。自一八六五年以來，計算上不以 8 爲此數(8.8″)之最近似值者爲數極少。而較爲晚近之觀察，則咸認此種視差爲八秒七六至八秒七八之間。故吾人可認爲此一距離，吾人所能知者，現精確至四百分之一。討論及此，吾人須加注意者，第一，早日之觀察須時加改正；第二，歷時愈久，愈能得出較佳之一致意見；第三，時至今日，絕對一致之意見，及絕對不移之確度，尙未獲得。物體度量如此，於統計度量亦然。例如代表平均壽命曲線之逐漸穩定，物價跌落之度量，與工資統計之發展是也。

精確之可能度 復次，在物體度量中，有時固可達到極高之確度，例如一立方尺水之重量，吾人可確知其一百萬分之一而無疑，但在其他情形下，吾人如能量至十分之一即已引爲滿意，例如離吾人最近之恆星距離，據測量所得，大約即爲三百四十至三百七十億英里之間也。故在統計上，如知一八八五年英國之資本總額爲七十五億至一百億金鎊之間，或知一八八六年全時工作工人之平均每週工資，爲二十一至二十七先令之間，於願足矣。此種言論之缺點即在，往往一估計作成之後，雖知其並不確實，但吾人不能估計差誤之限度。於此吾人殊不若『近代旅行家』之鑿確，彼

『……絲毫不差的，

知道天氣，

知道經度和緯度。』

而吾人則不敢具言『吾人之估計數，爲二十四先令五便士；此一估計數，大概不致有三便士之相差，但差誤而達六便士乃不可能』；惟在物體之度量中，吾人往往可精確至所用工具之一最小分度也。

一般所需要之確度　就另一方面論之，吾人雖不能獲得絲毫之確度，但在許多情形下，吾人所估計之數字，可達到一種確度，已足供實際應用之所需。在普通用途上，僅須有合乎慣例之確度即足。試舉雜例數則，例如一田產之面積，係以英畝，路得(rood 面積度量名，等於一千二百一十平方碼 — 譯者註)，及波爾(pole，面積度量名，等於三〇七又四分之一平方碼 — 譯者)計算之，非精確至一方碼不差也。股票之市場價格，變動至少以十六分之一爲起碼。吾人之生辰，記其日而不記其時。鐵路時刻表，並不註明秒數。海洋船隻，規定在某點鐘起碇，乃不言某點之某分鐘。身長之度量，求其精確到一吋之十分之一而已。百碼競走以十分之一秒而計時。物體度量如此，統計度量又何獨不然，吾人所得之答數，鮮有求精確至千分之一者，甚至精確至百分之一者，亦罕見之。一工作星期之千分之一，不及三分鐘；一週

工資之百分之一爲六便士。倫敦一城之人口，一百以內者，吾人鮮注意之，英國財政部之支出，不及一千鎊者，吾人殊不注意之，平均壽長不滿一日者，吾人亦不注意及之。總之，惟在此種限度以內，求達到實用之確度，乃往往可能。

差誤之定義——爲實施度量計，吾人可採用下述定義：——一估計數中之相對差誤者，即估計數與眞值相差對估計數之比率也。當眞值超過估計數時，則其差誤即爲正差誤（positive error）。

例如，若農村勞動者之平均每週工資，實在爲十四先令，而吾人之估計數，僅爲十三先令，則吾人差誤，即爲$\frac{14-13}{13}=\frac{1}{13}$，或百分之七・七。設吾人之估計數，爲十五先令，則差誤乃爲$\frac{14-15}{15}=-\frac{1}{15}$，或百分之—6.6。

用代數記號述之，設某一數量之度量爲 u，其眞值爲 u^1，則$\frac{u^1-u}{u}$即爲估計數中之差誤。$\frac{u^1-u}{u}$吾人可稱之爲 e，故 $e=\frac{u^1-u}{u}$而 $u^1=u(1+e)$（註一）。所謂 e，即爲相對差誤，至 ue 則係絕對差誤。

差誤之說明 就事理言之，當吾人計算差誤之時，差誤之量無從得知；吾人所能知者，最多不過差誤大概及或有之限度耳。例如，吾人可估計某一年內之失業百分數爲四・五，吾人由所有

之報告(由工資單據之研究,或救濟機關之報告得來)觀之,吾人認爲此一差誤,乃在失業百分數之二分之一以內。於是吾人應書爲 4.5±.5,意即估計數中之差誤(定義已如上述),大概不致超過$\frac{.5}{4.5}=\frac{1}{9}$,或百分之十一,而相當之絕對差誤,乃爲二分之一。在此種場合,吾人亦能指明確定之限度。失業百分數必須在零與一百之間。設吾人實際上能列舉百分之一之勞工階級爲失業工人,百分之九十二爲有業工人,則可知所求之數,即在百分之一,與百分之八之間,而吾人估計四・五中之差誤最大限,即爲$\frac{3.5}{4.5}=\frac{7}{9}$,或百分之七十八。此種差誤雖大,然較之原來說明『失業百分數爲四・五,差誤巨細則非所知』,猶爲精確也。吾人如再加研究,或可使各差誤之限度愈益接近,然後斷定所求之百分數實際上確爲三・五與四・五之間。於是應謂失業數係勞工階級之〇・〇四…,此一估計數,精確至所舉之最後一位數字。此一說明,與下一說明,係屬同一性質:『某實體之重量爲十五磅三盎斯——正確至一盎斯爲止』。

雖然,在一方面,甚爲明顯者,吾人對於差誤往往不能求得狹小之一定限度,在另一方面,吾人卻常知一總數中有若干數字幾乎一定無誤,而其他數字則幾乎一定有誤。例如,由英國人事登記總監之報告,吾人可知一八九五年英國人口爲三九,一二四,四九六,此一估計數字乃依一八九一年之戶口普查作成,而

自一八九一年至一八九五年之增加數，乃係根據一八八一年起之增加數計算而得。故其最後兩位或三位數字，一定僅爲一種猜測；而前二位或三位數字則正確無誤。因此之故，此一說明應爲：人口三千九百一十萬，或三九，一二四，〇〇〇±五，〇〇〇。此後一數，究爲何數，應視吾人對於人口之變動增加率之分析而決定。如此說法，較之前項說明，實際上較爲精確。

細數之删略　在許多種估計中，習常須將數字舉至極端細微之數。此在官方發表之文告上，可謂正則；因有司之責任，卽係收集答案而列爲表格，示明此種答案之如何得來，及由何處得來；而此種報告之確度如何，則爲經濟學家或統計學家之工作。但在總綱之敍述與記載上，以及在科學的估計上，則非特不需要列舉最後數字（一則因其並非準確，二則因其對於議論上不關重要，或對於讀者亦無意義可言），且其本身又純屬不正確。避免不確實，最易之法，卽將總數以千爲單位而說明之（例如地球直徑爲八千英里）；反之，如以某種理由，而需要較確切之度量（如當吾人欲比較赤道直徑與經過兩極之較短直徑時），則依科學方法，必須將數字舉至業已計算明確之處，並指陳其精確之程度。

第二節　計算相對差誤之效果之定則

吾人現請舉出若干定則，以說明一繁複之估計數之差誤，與

作成此估計數各成分中差誤之關係。

總和中之差誤　一、當每一部分數字之差誤，乘以該部分對總和之比率時，則某一總估計數中之差誤，卽等於各部分差誤之總和。

設吾人估計n數量爲$u_1, u_2\cdots\cdots u_n$，而其總和爲u，則$u=u_1+u_2+\cdots\cdots u_n$，而各數量之差誤，爲$e_1, e_2\cdots\cdots e_n$，總和中之差誤爲$e$：於是此總和之眞值，爲$u(1+e)$，而各部分之眞值，爲$u_1(1+e_1)$，$u_2(1+e_2)\cdots\cdots$，故

$$u(1+e)=u_1(1+e_1)+u_2(1+e_2)++$$

但　$$u=u_1\qquad+u_2\qquad++$$

二式相減，則$uc=u_1e_1\qquad+u_2e_2\qquad++$

且　$$e=e_1\times\frac{u_1}{u}\qquad+e_2\frac{u_2}{u}\qquad++$$

各部分之中，如有若干可以減去時，則此公式易於施用。

試取一算術實例以說明之。假如勞工階級之平均食物費在一九一四年爲二十五先令，衣着費用爲五先令六便士，房租費用爲六先令六便士，而眞實平均數，食物費爲二十七先令，衣着費爲四先令六便士，房租爲六先令；則差誤爲$-\frac{2}{25}$，$+\frac{2}{11}$，$+\frac{1}{13}$，於是三數總和中之差誤爲

$$-\left(\frac{2}{25}\right)\times\left(\frac{25}{37}\right)+\left(\frac{2}{11}\right)\times\left(\frac{5.5}{37}\right)+\left(\frac{1}{13}\right)\times\left(\frac{6.5}{37}\right)=-.054+.027$$
$$+.0133=-.0135\text{或}-1\tfrac{1}{3}\%。$$

在一重要事例，當吾人能以頗高之確度估計某一總和之大部分，而其一小部分爲吾人所不知時，即可應用此一定理。例如，吾人可自若干職工組合中搜集若干報告，得悉失業者有三萬三千六百五十人，則雖有較小組合未作報告，吾人亦有理由相信此中差誤，不致超過百分之一。吾人可爲此較小之職工組合，作一估計，姑謂其會員中失業者有一千人，並且假設此中差誤甚大，姑謂有三分之二或百分之六十七之多；於是此一總數中之差誤，乃小於

$$\frac{33650}{34650}之\frac{1}{100}+\frac{1000}{34650}之\tfrac{2}{3}=.029$$

或小於百分之三。此一差誤，以大組合報告之差誤，與小組合報告之差誤相比，則與前者相近遠較後者爲大。在前言，吾人謂『小於』者，以吾人假定已爲該一小部分之差誤，設置一最高限度也。

〔平均數中之差誤〕 二、當各個估計數之差誤，乘以各該項估計數對所有估計數總和之比率時，則各項估計數之算術平均數中之差誤，即爲此等估計數之諸差誤之總和。

因設以 m_1，m_2，……m_n 爲某某等數量之估計數，而此等數量之眞值，則爲 $m_1(1+e_1)$，$m_2(1+e_2)$，……則各估計之算術平均數爲 $\frac{m_1+m_2+\cdots\cdots m_n}{n}$，而各眞值之算術平均數，爲 $\frac{m_1(1+e_1)+m_2(1+e_2)+\cdots\cdots+m_n(1+e_n)}{n}$，於是算術平均數中

之差誤，乃爲

$$\frac{\dfrac{m_1(1+e_1)+m_2(1+e_2)++}{n}-\dfrac{m_1+m_2++}{n}}{\dfrac{m_1+m_2++}{n}}=\frac{e_1m_1+e_2m_2++}{m_1+m_2++}$$

$$=e_1\times\frac{m_1}{S.m}+e_2\times\frac{m_2}{S.m}+\cdots\cdots$$

此處之S，指示 m 各數之總和。

吾人甚易見到，一則個別之差誤，對於得數不能發生極大影響，二則假如各別差誤，若非甚不相等，並若非全爲正數，或全爲負數，則算術平均數中之差誤，必幾可與各別差誤中之一數同其大小；三則如依一般情形所實現，假如各別差誤中，有若干爲正數，有若干爲負數（此點吾人卽將討論及之），則差誤必大形削減。

加權平均數中之差誤　三、一加權平均數中之差誤，爲以下兩種差誤之總和：（一）爲由於諸數量中諸差誤之差誤，此種差誤與一不加權平均數中之差誤相同；（二）爲由於諸權數中衆差誤之差誤，當原來數量幾全相等時，此種差誤必變爲甚小。

令 $W_1, W_2\cdots\cdots W_n$ 爲權數，加於 n 個估計數 $M_1, M_2\cdots\cdots M_n$ 上；設權數之眞值，爲 $W_1(1+\epsilon_1), W_2(1+\epsilon_2)\cdots\cdots$ 1數量之眞值爲 $M_1(1+e_1), M_2(1+e_2)\cdots\cdots$。

以 $M_w=\frac{SWM}{SW}$，於是則 M_w 卽爲估計得來之加權平均數，今令 $M_w(1+E)$ 爲此加權平均數之眞值。

然則 $M_w.E=\frac{SW(1+\epsilon)M(1+\epsilon)}{SW(1+\epsilon)}-\frac{SWM}{SW}$

$=[SW.S\{W_tM_t(1+\epsilon_t)(1+e_t)\}-SWM.S\{W_t(1+\epsilon_t)\}]$

$\div SW.SW(1+\epsilon)$,

此處之附數 t,表示任一選來之數量,等等。

於是

$$E.SWM.SW(1+\epsilon)=SW.SW_tM_te_t+SW.SW_tM_t(\epsilon_t+e_t\epsilon_t)$$
$$-SWM.SW_t\epsilon_t。$$

現在假設 E, e_t, ϵ_t 甚小,等於.1;乘積之等於.01者,略去之。

$$E.SWM.SW=SW.SW_tM_te_t+S\{W_t(M_tSW-SWM)\epsilon_t\}$$

$$\therefore E=\frac{SW_tM_te_t}{SWM}+\frac{S\{W_t(M_t.SW-SWM)\epsilon_t}{SWM.SW}$$

如以 W_1M_1 代 m_1,……其他依此類推,則包含 e_t 之項,卽與它則二之項相同,至 e_t 乃卽一數量中之差誤。

ϵ_t 之係數,尙須進一步加以硏究。

但因 $SWM=M_w.SW, M_tSW-SWM=SW(M_t-M_w)=m^1_tSW$,(此處之 m^1_t,係一數量超出加權平均數之數)。

$$\therefore E=S\frac{W_tM_t}{SWM}.e_t+S\frac{W_tm^1_t}{SWM}\epsilon_t。$$

由此觀之,起於各數量諸差誤之差誤,包括 M_1, M_2……等數量,而起於各權數諸差誤之差誤,則僅包括此等數量對其加權平均數之離中差。此等數量對其平均數之離散度,如與該平均數相

較爲小，則此等離中差各個亦甚小。復次，係數$W_t m^1_t$之總和，等於$SW_t M_t - M_u S W = 0$。各權數中之差誤，如完全相等，則所得之平均數中差誤必爲零，此可演繹知之。又如正差誤一般並非與正離中差(m^1_t)並存，負差誤非與負離中差相存，且如巨大差誤一般非與大權數並存（反之亦然），則$W_t m^1_t \epsilon_t$諸項之總和，漸趨於零。

故權數中之差誤，有一種效果，此種效果減小之原因，與使數量中差誤減小者相同。且此種效果，使各係數有互相消除之強烈傾向；惟若諸差誤，諸數量，及諸權數，以特殊形式相結合，則不致有此傾向矣。設數量甚多時，若欲在平均數中有甚大之差誤，則權數中必須有巨大之差誤不可。總之，事實上當權數一經合理估定，而諸數量並非甚不相等時，則數量中之差誤，必有一種影響，遠比權數中差誤之影響爲大，故權數中之差誤，吾人可間或忽略之。關於此一原則之數字實例，已見於前第五章第二節（加權平均數）內矣。

乘積中之差誤　四、一乘積中之差誤，約等於各因數中各差誤之總和，但正負號務須加以充分注意。

因如估計得來之各因數爲$f_1, f_2, \cdots\cdots f_n$，而眞值爲$f_1(1+e_1)$，$f_2(1+e_2)$，……，則乘積中之差誤，等於

$$\frac{f_1(1+e_1)f_2(1+e_2)\cdots\cdots f_1 f_1\cdots\cdots}{f_1 f_2\cdots\cdots} = (1+e_1)\cdot(1+e_2)\cdots\cdots - 1$$

如吾人删略二個或二個以上之e之乘積，則等於$e_1+e_2+e_3++e_n$。

e之各值之爲正或爲負，原各有均等之機會。如兩個e值之正負號各異，則有相互對銷之傾向。設各因數之所有差誤符號相同，則卽使諸因數之差誤，各個均爲甚小，而乘積中之差誤，亦可爲甚大。

舉例言之，設吾人估計結果，平均各賺二十五先令者有一百人，但實際上乃有一百零五人各賺二十六先令，則依公式之規定，估計之所賺總數中之差誤，應爲：$\frac{5}{100}+\frac{1}{25}=.09$。

如仍用此同一估計數，設眞實人數爲一百零五，工作收入爲二十四先令，則乘積中之差誤，應爲$\frac{5}{100}-\frac{1}{25}=.01$。

比率中之差誤　五、一比率中之差誤，約等於其二項差誤之相差，但正負號務予以充分注意。

蓋若估計得來之二項爲u_1u_2，而其眞值乃爲$u_1(1+e_1)$及$u_2(1+e_2)$，則此比率中之差誤，必爲——

$$\frac{\frac{u_1(1+e_1)}{u_2(1+e_2)}-\frac{u_1}{u_2}}{\frac{u_1}{u_2}}=\frac{1+e_1}{1+e_2}-1=\frac{e_1-e_2}{1+e_2}$$

$$=(e_1-e_1)(1-e_2+e_2^2-e_2^3+-)$$

吾人如將e有二次方之各項删略，則等於e_1-e_2。

如二項中之差誤，均爲正或均爲負，則有相互抵消之傾向。若二差誤亦幾乎相等，則比率中差誤必變爲甚小。

在不同日估計得來類似數量之兩個平均數相比中之差誤，吾人可應用定則五以計算之。

仍依用定則二及定則三下之符號，以 m_1, e_1, ϵ 指一日期所估計之數量，以 m^1, e^1, ϵ^1, 指另一日期所估計之數量，則 $m_1{}^1$, $m_2{}^1$, ……之簡單平均數，對 m_1, m_2,……之簡單平均數之比率中之差誤，爲

$$\mathrm{S}\left\{e^1\left(\frac{m^1}{\mathrm{S}m^1}\right)\right\}-\mathrm{S}\left\{e\left(\frac{m}{\mathrm{S}m}\right)\right\}$$
$$=\left(e_1{}^1\cdot\frac{m_1{}^1}{\mathrm{S}m^1}-e_1\cdot\frac{m^1}{\mathrm{S}m}\right)+\left(e_2{}^1\cdot\frac{m_2{}^1}{\mathrm{S}m^1}-e_2\frac{m_2}{\mathrm{S}m}\right)++$$

設在二次觀察間之期內，各數量無大變遷，則 $\frac{m^1}{\mathrm{S}m^1}$ 一分數與 $\frac{m}{\mathrm{S}m}$ 相差無幾，餘類推。

此等相差數，與數量本身相比，吾人殊可將相差數删略不計，此爲吾人計算差誤之約略影響時所不可少之正當手續，惟如此則

$$兩簡單平均數比率之差誤=\mathrm{S}\left\{\frac{m_1{}^1}{\mathrm{S}m^1}(e_1{}^1-e_1)\right\}。$$

設兩估計數，乃在幾乎相同之環境下作成，因而招致差誤之機會亦相同，則 $e_1{}^1$ 及 e_1 多半非特具有同一符號，而且幾乎相等。

試以 d_1 代替$(e_1{}^1-e_1)$，以 d_2 代替$(e_2{}^1-e_2)$，……餘類推，則

$$差誤=\mathrm{S}.\left\{d_1\cdot\left(\frac{m_1{}^1}{\mathrm{S}m^1}\right)\right\}$$

（此處 d 之各值，可爲甚小）。

兩簡單平均數相比比率之差誤，已討論如上，如更對於兩加

權平均數比率之差誤，亦加以相當分析，則因其過於複雜，此處不便予以檢討（註二）。但若遵用『權數之差誤，不若數量中差誤之重要』一原則，而於應用時酌加修改，則吾人不妨採用上述在求算對兩加權平均數比率中差誤之第一近似值時所舉之公式。茲將此一公式用文字述之如下文。

平均數相比之差誤 六、在不同期間所估計之兩類似數列之平均數之比率中之差誤，約等於此兩數列各相當項中差誤諸差數之總和，但每一相差數，務須各乘以其相當項之後項，對於後期所有各項總和之比率。

此一定則異常重要，故殊有舉一實例以解明之之價値，在此實例中將更介紹一新數量。

設在兩年中每一年吾人均可估計一總數之一部，而較另一部分之估計爲確實，如在定則一下所舉實例所云然，則吾人可採用下列公式：

	第一年		第二年	
估計數或估計權數	w;	差誤ϵ;	w^1;	差誤ϵ^1
平均收入估計數	m_1;	差誤e_1;	m_1^1;	差誤e_1^1
估計數，確度較低	rw;r之差誤，p;		r^1w^1;r^1之差誤，p^1	
收入估計數	m_2;	差誤e_2;	m_2^1;	差誤e_2^1

依假定，e_1及e_1^1小於e_2及e_2^1。

第一年平均數之差課

$$\frac{\dfrac{w(1+\epsilon).m_1(1+e_1)+r(1+\rho).w(1+\epsilon).m_2(1+e_2)}{w(1+\epsilon)+r(1+\rho)w(1+\epsilon)}-\dfrac{wm_1+rwm_2}{w+rw}}{\dfrac{wm_1+rwm_2}{w+rw}}$$

$$=e_1\frac{m_1}{m_1+rm_2}+e_2\frac{rm_2}{m_1+rm_2}+\rho\frac{r}{1+r}\cdot\frac{m_2-m_1}{m_1+rm_2}$$

設吾人略去 e 及 ρ 之各乘積時。

由此觀之,與確度較低之部分有關之差課,e_2 及 ρ,各乘以r(r 爲確度較低部之權數,對確度較高部分權數之比率),ρ 亦乘以 m_2-m_1,m_2-m_1 在多數情況下,值均爲甚小,至於其餘之差課,卽爲 e_1,依假定亦爲甚小。

爲討論簡便計,吾人假設未知部分對全部之比率(並非估計時之差課)不變,並設兩部分之平均收入估計數之比率,亦未變更,則相比中之差課爲:

$$(e_1^1-e_1)\cdot\frac{m_1}{m_1+rm_2}+(e_2^1-e_2)\cdot\frac{rm_2}{m_1+rm_2}+(\rho^1-\rho)\frac{r}{1-r}\cdot\frac{m_2-m_1}{m_1+rm_2}。$$

例如,在估計蘇格蘭農村勞動者平均工資之變動時,吾人得有與下列類似之數字:

	一八六七年			一八九二年		
	已婚農夫					
估計數	1,000	平均收入	36鎊	1,200	平均收入	49鎊
假定眞正數	1,010	平均收入	35鎊	1,220	平均收入	48鎊
	農場傭工					
估計數	200	平均收入		240	平均收入	
		貨幣	21鎊		貨幣	27鎊5先令
		膳費估計值	23鎊		膳費估計值	24鎊
		合計	34鎊		合計	41鎊5先令
假定眞正數	200	收入總數	37鎊	240	收入總數	47鎊

由此所見 $w=1000$，$m_1=36$，$r=\frac{1}{5}$，$m_2=34$，$w^1=1200$，$m_1{}^1=49$，$r^1=\frac{1}{5}$，$m_2{}^1=41\frac{1}{4}$，$\epsilon=\frac{1}{100}$，$e_1=-\frac{1}{36}$，$\rho=\frac{9}{101}$，$e_2=\frac{3}{34}$，$\epsilon^1=\frac{1}{60}$，$e_1{}^1=-\frac{1}{49}$，$\rho^1=-\frac{1}{61}$，$e_2{}^1=\frac{23}{165}$。

以上二例吾人對於已婚農夫之收入估計過高，而對於農場傭工之收入估計則過低。玆假設（事實亦確如此）農場傭工之膳宿及其他賞賜之價值，估計難以精確；又此兩階級之比例數，亦不能確知。

代入上列公式，則可知在二年中二階級平均收入總數之估計數比率中之差誤，爲

$+\cdot 0062$，屬於農夫收入估計數之差誤。

$+\cdot 0081$，屬於農場傭工收入估計數之差誤。

$-\cdot 0008$，屬於二階級中人數比率估計數中之差誤。

此最末一差誤，卽屬於權數之差誤，其值甚小；第二差誤，乃因吾人對於膳宿所知不詳之故，現以傭工人數甚少，故此一差誤亦減至與第一差誤相去無幾。

職是之故，經用公式求得，整個差誤爲$+\cdot 0135$。試查實際數字，可知第二差誤對第一差誤之估計數比率；爲一・三三七六對一之比，而假定之眞正數比率，爲一・三五二九與一之比；換言之，差誤乃爲$\frac{.153}{1.3376}=+.011$ 也。

此兩種計算方法之差別，卽在删略次要各項之不同。

所須注意者，二個數量相比之差誤，與吾人所或欲計算之差誤（即百分率增加量中之差誤）不同。例如在上述之例中，估計之百分率增加量爲三三・八，眞正之百分率增加量爲三五・三，而百分率增加量中之相對差誤，則僅爲．〇四五。故在此種計算中如求確實，則依定則六之公式，所算出之差誤，非爲甚小不可也。

如就一工人家庭預算，而計算其衣着費之相對重要性，實有人所共知之困難，玆更擧一例以明之。

下列估計數，即一九一八年英國生活費委員會報告中所採用之數字（總號八九八〇，第七，第十八及第二十三頁）。

有技能工人，平均每週費用

	一九一四年	一九一八年	比率
食物	27先令	49先令10便士	1·84
衣着	7先令	13先令 9便士	1·96
總計	34先令	63先令 7便士	1·864

玆以 $w=27$，$r=\frac{7}{27}$，$m_1=1.84$，$m_2=1.96$。

假設 r 應作爲 $\frac{1}{3}$，m_1 應作爲 1.90；而 m_2 應作爲 2.10。

則 $e_1=\frac{8}{92}=\cdot0326$，$e_2=\frac{1}{14}=.0714$，而 $\rho=\frac{2}{7}=.286$。

依公式得差誤爲

$+.0256$，屬於兩期食物費相比之差誤。

+.0155,屬於兩期衣着費相比之差誤。

+.0030,屬於衣着費與食物費兩期相比之差誤。

至整個之相對差誤,則爲 .044。

各差誤之影響,與差誤之大小,成反比例。衣着費比率之巨大差誤,僅僅影響得數中之第二位小數而已。

雖然,若 m_2-m_1 較大,換言之,卽若衣着費之估計增加數,遠比食物費之估計增加數爲大,則此一差誤之效力,必亦依比例而增大。

討論至此,吾人現又提及相對差誤之整個問題,而此卽本書第二編第四章之機率原理所解明者也。

第三節　偏誤及非偏誤

偏誤及非偏誤　在平均或比較時,提及所有差誤,有兩類差誤,不可不加以區別,一類卽偏誤 (biassed error),一類卽非偏誤 (unbiassed error) 也。此中差別可設例以闡明之。設派遣多人,分往各地,調查各種工業狀況,以證明工資之果然爲高,工作條件之果甚無損於健康,以及其他等等,則彼輩多半將僅就經營最完善之工廠,而加以考察,並僅就技能較高工作有定之工人,而記錄其工資。於是所產生每一城市之平均數,無乃過高。反之,

如各人心中並無成見，僅事公正不偏之調查，則調查人員在某某城市所得之平均數，即屬過高，而在他數城市所得者，反乃過低，此一隨調查員之個人特性及環境情勢而定者也。然則在前一情節之下，差誤係屬偏誤，各差誤均在同一方向，均有使平均數增長之傾向，故此平均數之差誤，乃即等於各城市之平均差誤。在後一情節下，差誤乃為非偏誤，超過平均數或較平均數為低，二者各有均等之機會，是以所作估計數愈多，而結果之差誤亦愈微。茲以下列數解明之：

	事實	有偏誤之估計數	有非偏誤之估計數
	先令	先令	先令
甲地之平均工資	24	25	24
乙地之平均工資	23	25	25
丙地之平均工資	26	27	25
丁地之平均工資	27	28	28
戊地之平均工資	28	30	27
平均數	25.6	27	25.8
差誤		5.2%	1%

沿一具有哩程碑之道路，度量腳踏車騎行之距離，吾人可知連續各哩程碑間之距離，並不準確，乃或有五十碼至百碼之相差。但超過一里或少於一哩之差誤，二者各幾有均等之機會，故騎行經過之距離愈長，差誤亦愈少，此點上文業已闡明之。此差誤即

非偏誤也。反之，若乘脚踏車者信任彼之迴輪計，則將不免於偏誤，蓋彼之迴輪計，未必準確貼合彼之車輪，車每行一哩，迴輪計輒指一千八百碼也。在此情形之下，偏誤(bias)可加以度量而以結果中減去之，至於非偏誤則任其自行抵消。總而言之，偏誤往往由於工具級度劃分錯誤，非偏誤則往往由於各別度量有錯也。

偏誤及非偏誤之相對重要性　在人口普查報告表中，有一事實，即許多婦女報告之年齡，較其出生報告書所載者爲小，惟有此事實，乃致引起人口平均年齡之一種偏誤。又有一事實，即人民往往將其年齡報爲最相近之整略數，而惟此一事實，乃致引起非偏誤，但大體上對於平均數並無甚影響。即在一九〇六年之工資調查中，亦未嘗無僅由某種工業中經營方法較爲寬大之工廠採取材料之傾向，故對於所得之平均數中，乃招致一種偏誤。由此種種例證，吾人現可提出另一極一極爲重要之原則。非偏誤與簡單估計數中之偏誤相比，非偏誤實不甚重要；但如取用二個類似估計數之比率時，則偏誤亦減小。

蓋在數個數量之平均數中，而此數個數量具有偏誤（η_1，η_2……）及非偏誤（e_1，e_2……）時，吾人依據定則二之規定，即知結果所得之差誤，可寫成：

$$S\left(e_1\cdot\frac{m_1}{S.m}\right)+S\left(\eta\cdot\frac{m_1}{Sm}\right)。$$

第一項中之差誤，爲非偏誤，此等差誤之中，有許多爲正，有

許多爲負，二者有相互銷除之傾向。實際上，E若爲 $e_1, e_2 \cdots\cdots e_n$ 諸差誤之代表，則求平均數中諸差誤所發生之差誤之近似值，其第一近似值必爲$\frac{2E^*}{3\sqrt{n}}$。（*註三）

偏誤之巨大效果 例如，玆有數量百個，每一數量之非偏誤大約爲$\frac{1}{10}$，則此一百數量之平均數中，所得之差誤未必大於$\frac{2}{3}\times\frac{1}{10}\div\sqrt{100}$，卽未必大於$\frac{1}{150}$。反之，則無互趨平衡之趨勢。假如每一數量，各超過百分之十，則平均數亦超出百分之十。爲求確實計，吾人之原則，僅注意於鎊，而未兼及便士。當吾人之數量爲巨大之偏誤所貫穿時，則減少非偏誤以增進吾人數量之確度，實爲徒勞。吾人苟不知實際上彼充滿吾人估計數之偏誤之存在，則無可挽救。設吾人熟知之，如欲獲得較高之確度，則與其完全删略之，實不若加以最差誤之修正之爲愈。蓋當吾人爲多數偏誤，加以不偏不倚之修正時，是吾人卽已將其化爲非偏誤矣。於是在吾人之平均數中，包括之項數愈多，則結果之差誤亦愈小矣。例如，設吾人得知英國全國農村勞動者之平均每週工資，爲十三先令，並假設得出此一平均數之調査報告共有一千份，玆就此一千份報告之情況而考慮之結果，吾人有理由假設一先令之差誤，當爲調査報告中非偏誤之代表，則結果在平均數中之差誤，吾人以意料之，當爲$\frac{1}{\sqrt{1000}}$之$\frac{2}{3}$，換言之，卽僅四分之一便士之差誤也。但此處吾人乃有一完全虛假之確度。蓋農村勞動之收入，吾人未曾計入

者，尚有甚多，如割禾時及收穫時之償付，計件工作之利益，便宜之房租及地租，以及零星賞金；此一部分收入，吾人殊不能加以準確計算。吾人如將此等收入完全棄之不顧，則在平均數中之差，將有二先令左右之差誤。反之，吾人就此一千份之全數報告中，將此等增加數，作成個別之估計數，如其中並無偏性，則每一估計數中，雖或有二先令之出入，而結果在平均數中之差誤，可望為$\frac{2}{\sqrt{1000}}$之$\frac{2}{3}$先令，即僅二分之一便士：於是吾人整個差誤，現可少於一便士，而非一先令矣。在計算已發表之平均數之確度時，務應將此等原則牢記心中，而偏誤之可能性，尤須常常加以考慮也。

相比之確度 當吾人計算一比率中之差誤時，情形即有不同。一比率之差誤，約等於該比率各項差誤間之差額。設 n, n^1 為各項中之偏誤，e, e^1 為各項中之非偏誤，則依定則五之規定，$(n^1-n)+(e^1-e)$ 即為此比率中之差誤。至非偏誤 (e^1-e) 之大小度，大致幾與 e 或 e^1 之大小度相同（註四）。又如上例所述，設 e 及 e^1 不致遠大於 $\frac{2}{3}$，則 (e^1-e) 不致遠大於 $\frac{2}{3}$。但此比率中各項之偏性，意味如為相同（全為正，或全為負），則偏誤結果之 (n^1-n)，將小於原來之差誤。設吾人以恰為相同方法，為兩項各作一估計數，設吾人曾對同類人發同一之問題，而在兩處取捨之細節亦全相同，則在兩估計數中，吾人將已造成幾乎相同之偏性差。試就

以前所舉之例證言之，設吾人在兩期間調查一農村勞動者之收入，除其平均每週工資外，其餘各節竟顯然置之不顧，則此比率中所生之唯一差誤，必由於此等特別收益對通常工資之比例之變遷；而此種變遷在短期間內大致必甚輕微也。不然，設吾人乃以兩期之夏季工資作爲該年之平均工資，則比率中之差誤，將僅隨夏季工資對該平均數關係之變遷而定。是故對於一變遷遲緩之數量，在不同日期所作兩估計數，如估計數作成之方法相同，則此兩估計數之比率之差誤，往往小於任一單獨估計數中之差誤；蓋非偏誤並未增大，而更重要之偏誤，則大爲減小也。此時，偏誤之是否存在，吾人無須求知，偏誤卽將自行消失。如吾人果知有偏誤之存在，並有爲彼作成甚爲精確之估計數之任何方法，則有加以估計之價值；但若僅改正其一年之偏性，而另一年之偏性，又不予改正，則實爲大錯。爲供比較之用，而使後期之估計數，較前期之估計數爲精確，往往遇巨大困難，非止常無大用而已也。總之，由非偏誤而生之差誤，固能略形減消，但由更重要之偏誤，而生之差誤，則只有增大之一途。

定期報告結構上對於統一性之需要　一切編纂每年報告之政府官員及其他人士，在此均有進退兩難之苦。苟欲使每年報告表册本身精確無疵，則永須努力精益求精，且恆須察視所量數量之變動，而使所用方法及列表法，與此等變動相適合。但爲求各

年報告,可相互比較,則必須絕對保守成法,前人卽有錯誤亦應一仍其舊,惟須注意,無使錯誤加多,或另有新遺漏也。雖然,此種進退兩難之苦,並非絕不可免。蓋引用一改進方法之時,有時可爲新舊兩法各作數年之表格;於是經此改動而起之區別可見,而早日數字,不難使其達到後期數字之較大精確程度矣。例如,英國商業部,自一八九八年起,始於出口貨值表中,添列裝載貨物駛離英國海口而連貨物一併售與外人之船隻。玆列一表於下:

	一八九九	一八九八
國產品出口貨值(售與外人之船隻除外)	255,465,000鎊	233,359,000鎊
國產及殖民地商品之復出口貨值	65,020,000鎊	60,655,000鎊
總計	320,485,000鎊	294,014,000鎊
出口之新船隻貨值	9,195,000鎊	未詳
新總計	329,680,000鎊	

由此足見在材料之搜集及列表時,微細更動之忽略,實爲統計上許多錯誤之原因。

結論 玆將本章之主要結論,摘要述之如下:

增進確度之方法有二,一爲平均法,平均法可減少非偏誤;一爲相比法,相比法可減少偏誤。權數中之差誤,往往不若估計數中其他差誤之重要。一答數中之差誤,固然不能算出,但能以該答數所自出之各項中之差誤說明之。吾人不能達到定而不可

移之程度，但吾人可指出減小差誤之方法，而藉數學之助，吾人並能度量此減縮之範圍。當吾人計算類似及估計方法相同之數量之加權平均數之比率時，原起之差誤，即能減去大部。吾人在下章所討論之指數，即為此類之例證。

由抽樣方法而得之確度，必需多加以數理之討論，容當於本書第二編第二章及第四章論及之。

（註一）　有時書如 $u=u^1(1+\epsilon)$，與眞值相對而計算錯誤，則更為便利。如此則 ϵ 約等於 $-e+e^2$；當 e 不及百分之十時，則可令 $\epsilon=-e$。

（註二）　見統計學報，一九一一年，自第八十五頁起，及本書第二編附錄七，但算式已略變動。

（註三）　能否得到如是其大之差誤，尚在兩可之間。請參閱本書第二編第四章。

（註四）　設 E 為 e 或 e^1 之機誤，則 $E.\sqrt{2}$ 為二者相差之機誤；見本書第二編第三章。

第九章　指數

指數(Index number)之討論,可作爲上章所述原理異常良好之解明,而指數之本身,又異常重要,故吾人本意,雖欲避免特殊問題而不談,但究値特闢一章以研究之。

指數之功能　指數者，用以測量吾人不能直接觀察之某種數量之變動者也。蓋此種數量,吾人不能直接觀察之，但知其對於能以直接觀察之其他數量,乃有其確定之影響,若上漲即有全部上漲之傾向,若下落即有全部下落之趨向,然此一影響反爲對於各個數量發生種種不同影響之多宗原因之作用所隱蔽，因而並不可見也。例如,就三種應用指數之數量而論,一、貴金屬與其購買力之關係之變動,即足波及所有各種商品之價格;但同時尙有其他無數原因亦在發生作用，足以影響各類商品之價格;二、有平常技能工人之每週工資,即有促其上漲之一般原因,但此一般之增加，已爲無數對於上下各級勞動有大小不同之影響之原因所遮蔽矣;三、勞動及其他階級消費量之變動，乃爲十分穩定之數量,但此一數量,只能用間接方法，就各個物品消費量之變動,觀察而測量之。

雖然,指數之功用,並非止此而已,實則指數之用途,乃幾與

統計學之領域，同其廣被焉；蓋統計學一詞，吾人既只限於對繁複羣類及其變動之測量，統計學之目的，既在測量支配一異質羣類(heterogeneous group)之一般定律之作用，且一般勢力所產生之變動，依例只能就其對於各個事件之效力而測量之；是則指數方法，自可立卽用以自專對於各別物件有特殊影響之變動原因中，分解對於全體羣類均有公共影響之力量也。

指數之性質 更嚴格論之，一列指數，實卽一列加權平均數，而此加權平均數，乃按期算出，所平均之數量，又彼此相似(如價格或工資)，至所用之權數，必須以能將每一次測量所涉及之全部羣類施以實在之平均者爲限。如就其廣義論之，一列指數，乃卽一組數列，而能將所論及之某種數量之運動，在其趨勢及變動中反映出來者也。如諸數量及其權數，均能確切得知，則指數方法，乃卽直捷表現純粹算學得數之簡便方法，此一單簡性，於編製出口貨價指數時，庶幾可以顯現。如諸數量乃爲自一範圍甚寬之羣類中抽選得來之樣本，而又無決定各數量相對重要性之顯明方法者，則此指數對於有一定之意義及可量之性質之現象之運動，必不能有十分直接之關係；此乃大多數之物價指數之性質，亦卽幾種工資指數之性質也。至如諸數量，並非由吾人所欲研究之現象中，得來樣本之直接度量，而爲相聯現象之抽樣數量時，則此列指數與現象之關係，必爲間接者；事實上大多數工資指數

及就業狀況指數，均屬於此類也（註一）。

最普通之編製指數法，乃介乎極正確與間接關係兩極端之間。例如英國商業部勞工局之工資率變動指數，其對象乃即照例定之計劃按年查算指數，而此指數之比率，即與英國一定工廠中，工人平均每週工資率之比率相同；最低限度，一般所用之指數之意義，乃不外如此，至英國勞工統計輯要（Abstract of Labour Statistics）中標題所云，乃為『英國工資之一般動向』（如第十六期輯要，總號第七一三一號，第八十二頁）也。此一指數之算法，須先擇出經承認之計時或計件工資率數百個，作為一九〇〇年工資額之百分數以表示之，然後將此得數按年平均之。此平均數中之權數，擇用之法乃為間接者；凡建築業，煤礦業，工程業，紡織業，及農業五類，每類均以同等重要視之，雖建築業一類，包含七十四件，農業包含一百一十五件，其他各類件數，亦多寡不同云。孫巴克（Sauerbeck）氏之物價指數，求得之法，選擇若干代表商品之價格，對於較重要商品則用雙重價碼（duplicating quotations）方法以加權之。然則在此處以及其他各處，凡論及指數時，吾人須擇用一種『數量』（quantity），此種數量或熟籌而後用之，或偶然用及之，並須決定一種『權數』（weight），此種權數或為直接或為間接求得之。故不論吾人之對象，為平均工資，抑為平均物價平準，凡吾人擬加以測量者，則此指數之運動，

必期望其能與現象成正比例也。

在此種情形之下，須加考慮者有三點：一，此羣類之性質與範圍，以及此羣類所具特質（此特質之一般變動，卽吾人欲行研究者）之本性；二，抽選樣本之方法；三，權數之效果。

（一）就孫巴克氏之物價指數而論，其羣類乃為英國之躉售物價；就其他物價指數而言，所謂羣類乃卽出口貨物之價格，進口貨物之價格，以及其他種種之價格。英國勞工局之工資指數中，其羣類為雙重，乃包括（甲）每週計時工資率，（乙）計件工資率，而其結果自為混合物。羣類之範圍，及將施以度量之本質或屬性，規定二者之定義，甚為重要。夫本質之意義，有時甚為費解，如『貨幣之購買力』，或『失業之總量』是，在此情形之下，吾人只得就有關之屬性，給以定義，而度量之，如一般物價平準，或依照失業之某種定義，而計算之失業人數是。

（二）在抽樣之時，依一般遵循之準則，必須就有適合定義而度量可以確實者抽選之，卽在最佳之有名指數中，選擇之標準，異常嚴苛，所有之價格，必須適合準則乃能採納之。當此之時，吾人往往須將羣類之定義重行考慮並限制其範圍。例如吾人進行度量一般物價，必為定義中之要件所限制，因而只能就市價甚有規則之躉售物價度量之；至於工資方面，英國勞工局亦只有採用合於標準之工資率（農業工資為例外）之一途。但為求製成之

指數，適於差誤律之分析，吾人所抽之樣本，務須隨機而得，且其變動必須與一般運動無涉；苟求此獨立性而不可得，則樣本之數，必須加多，以期達到某種確度。隨機性(randomness)，或可因不期而遇之事件，能以保持，蓋惟有不期而遇之事件，乃使所取之樣本可以被選用也；但此大多以在躉售物價爲然，至於工資，則不可能。抽選時，如發生偏性(bias)，吾人有時可將羣類之定義，重行加以限制，卽可期於安全。

（三）關於權數之效果，如互爲獨立之數量，個數甚多，則任何合理之加權辦法，多能得到良好之結果，與爲該問題之條件所許可而得者相同也。

假設支配一羣類數量之變動者，有一般勢力一，有其他勢力若干；此一般勢力，以同一意味而活動，換言之，卽如漲則全部漲，落則全部落；其他數種勢力，各對於一數量或多數量發生作用，且其中有使數量趨於上升者，同時亦有使數量減退者；故總計諸項特殊勢力，使平均數趨於增加者有之，使趨於降落者亦有之，而此一般勢力，則始終保持一種累積效力，使各數量全部增加，或全部漸減。然如此諸端特殊勢力爲數甚多，而其效力則甚輕微，則此諸勢力對於平均數之影響，必有趨於中和之傾向；如此則平均數之變動，必將僅顯現此一般原因之影響矣。如用上章之用語言之，此諸種特殊勢力雖產生多數偏性(biassed)變動，但此種

偏性變動，對於平均數之影響，乃甚輕微，較之一般勢力所產生之偏性變動，殊可略去不計也。

吾人經將多數之普通應用指數加以檢討以後，可知實際上施以度量之數量，並非吾人意欲知其一般運動之數量。夫躉售之物價，絕不依照任何簡單定律——不論定差(constant difference)或定比(constant ratio)——而與零售物價，同其運動；標準工資與平均工資，大不相同，但何以不同，人皆不知；計件工資率與工作收入之關係，變化靡常，且不可知。可加度量之數量，與實在欲加調查之本質，二者之間，並無有如 $y=x$，或 $y=kx$，或 $y=a+bx$ 之簡單關係，實則乃爲 $y=f(x)$ 之關係，而此函數之形式如何，吾人又非所知也。如求作成之指數，令人易於瞭解，$y=a+bx$ 在 x 之普通全距(range)中必須爲一良好之近似值，因具有高級乘冪之 x 項之極端值，有變爲甚大之可能，致結果作成之指數，反不可靠也。在編製指數之過程中，a 即行消滅，但 b 所度量者，爲 y 之變動對 x 之變動之比率，決定 b 時，往往甚難。

製成某一年之指數，其定義可述之如下：設有 n 個數量，吾人欲研究其一般運動，茲以 x_1，$x_2 \cdots\cdots x_n$ 代表此種數量。設另有一種已加度量之數量，與前一數量用一方程式聯繫之，此另一種數量，以 $y_1, y_2 \cdots\cdots y_n$ 代表之，聯繫 x 與 y 之方程式，以 $y_r-100=b_r(x_r-100)$ 爲其良好近似值（註二）。設已規定一種適當之權

數，如 w_1，w_2……等，並以 J 代表$\frac{w_1x_1+w_2x_2+\cdots\cdots}{w_1+w_2+\cdots\cdots\cdots}$，以 I 代表$\frac{w_1y_1+w_2y_2+\cdots\cdots}{w_1+w_2+\cdots\cdots\cdots}$。此 J 即爲一指數，此指數之變動，即足表示運動，但此指數乃理論上之指數，無從加以計算，至實際可以計算而且付諸應用之指數，乃爲 I。

$$\mathrm{I}-100=\frac{\Sigma w(y-100)}{\Sigma w}=\frac{\Sigma wb(x-100)}{\Sigma w}。$$

設 $b_1=k+d_1, b_2=k+d_2, \cdots\cdots, k$ 乃選來（如可能時）作爲 b 各值之平均數，而此平均數能在 x 各值之普通全距間使

$$\frac{\Sigma wd(x-100)}{\Sigma w}(=\mathrm{F})$$

變小者。

於是 $\mathrm{I}-100=k\frac{\Sigma w(x-100)}{\Sigma w}+\frac{\Sigma wd(x-100)}{\Sigma w}=k(\mathrm{J}-100)+\mathrm{F}$。

僅就一般情形言之，如 x 各值數值之範圍並不甚寬，如 b 各值幾爲相等，而 b 之極端值，並不與 w 之極端值相合爲一時，則 F 之值必甚小，而其各年間之變量必甚微，乃可以略去不計之。

在此情形之下，I 與 J 之關係，在標準年 (standard year) 當 J（及各個 x 與 y）爲 100 時，I 亦等於 100，且 I 值之變動，乃幾爲 J 值變數之 k 倍，至 k 乃爲 b 各值之平均數，而各 b 值則爲度量各個 y 值之變動對相當之各 x 值變動之比率者也。

吾人如企圖藉由躉售物價作成零售物價指數，b 之各值，並

不可知，且各個 b 值，由一商品至另一商品，原即大不相同，由此時至彼時，亦迥然有異，又當物價特高或特低時，b各值變化靡常，人甚難測定。故一般零售物價與躉售物價之關聯，並非異常密切，吾人殊不能發言，謂此之變動，與彼之變動，成正比例也。試就英國勞工局之工資指數而論，計時工資之變動，與計件工資之變動，對於工作收入之關係，彼此並不相同，且二者對於工作收入之關係，亦無從得知；換言之，b 之各值，乃爲未知且不相等也。僅就計件工資率而論，當工資率上漲時，工作收入往往上升較速（b小於１），而工資率下落時，工作收入下降亦往往較緩（ b 大於１）；換言之，b 之各值並非一定，而上列公式中之Ｆ，乃爲未知且不能刪略也。

如 b 之各值果爲相等，則Ｆ必爲零，k 之值乃可在將兩年間之數值，加以特別研究後，而求得之。如然，則Ｉ之運動，必將緊隨Ｊ之運動，而在一已知尺度上現明顯之反應。

無如實際之關係，並非顯然，設有一 x 值超出其平均數百分之四，吾人不能卽謂 y 亦大於其平均數百分之四，只能謂 y 之離中差爲百分之四×b，而 b 幾爲常數也。b 之各值，與某種（加權）平均值（k）相差，而此等相差數之影響，在取用此平均數時，幾可全然消滅，且此平均值 k，在各年幾皆一定不變，此亦可加以假定者也。總之，對於 b 之各值，及結果所得之 k 值，無不可以設立

各種之假定，於是指數之變動，乃可解明矣。

於此有一重要之條件，即當 x 經過一變動後復回至一數值時，則相當之 y 值，亦必還歸其原來數值，否則，即有任何相差數額，至少，必爲極微且無偏性。但如用躉售物價以度量零售物價之變動，此一條件必爲之破壞，蓋躉售物價與零售物價之關係，已逐漸變更，果如吾人所預料也。執此而論，英國勞工局之工資一般動向指數，因標準工資或計件工資之變動，對於平均工資之變動，關係時在變更，則此一條件自必爲之破壞矣。

英國商業部之指數　英國現存之躉售物價指數甚多，吾人可擇要略一檢閱之。商業部按年有出入口貨值貨量記載發表，則出入口貨之平均價格，必可加以計算。所舉之商品，乃自一定之整個期間之答案中均曾出現者選擇而來。商品選定之後，即擇定某一年作爲基年 (base year)；然後依基年之物價，分別估量其他所有各年之各種價格；任何一年估得價格之總數，即爲該年貨物價格未曾變更之總價值；此一數值對於記錄上實在所有價值之比率，即爲在基年之平均物價對其他某一年之平均價格(假如平均一詞乃用其廣義之意義）之比率；又如此一比率之前項使與 100 相等，則比率之後項，即爲所求之某一年之指數，而所謂指數，乃即對基年之指數之百分數。然則，吾人現所論者，顯已論及加權平均數矣。

權數方式 設 p_1, p_2, p_3…… 爲基年中選定各單位貨物之價格，以 r_1p_1, r_2p_2, r_3p_3……爲吾人欲求算指數之某年物價：則 r_1, r_2, r_3……即爲各個商品價格變動之度量，而此種所有之 r 值，即爲吾人抽來欲從中查出物價一般變動之樣本也。以上所述之方法中，所用之權數，求得之法，不外：設 b_1, b_2, b_3……爲某一年貨物單位數，但該年貨物總價值，以當年之物價計之，爲 $(b_1r_1p_1 + b_2r_2p_2 + \cdots\cdots)$，以基年之物價計之，則爲 $(b_1p_1 + b_2p_2 + \cdots\cdots)$；二者之比率，爲 $\Sigma brp : \Sigma bp$，於是該年之指數，必爲

$$100 \times \frac{\Sigma brp}{\Sigma bp} = 100 \times \Sigma\left(r \cdot \frac{bp}{\Sigma bp}\right)。$$

由此觀之，加於 r 各值上之權數，即爲該年相當貨物按基年價格計算之價值。吾人於此可見標準年份之選擇，對於權數甚有影響，因以物價甚高之年爲基年，竟對於某一商品加以特別之權數，而爲尋覓一『正常』年份，業已煞費周折矣；然各別商品之權數，雖受其影響，但平均數未必即隨之變更，且依上文所述之原則，即有變更，亦必極爲輕微也。實際上茲有下列之數字（註三）：

一八八六年指數與一八八三年指數之比較

權數 / 年份	進口貨				出口貨			
	以一八七三年價格計算之價值	以一八八三年價格計算之價值	以一八六一年價格計算之價值	以一八八一年價格計算之價值	以一八七三年價格計算之價值	以一八八三年價格計算之價值	以一八六 年價格計算之價值	以一八八一年價格計算之價值
1883	100	100	100	100	100	100	100	100
1886	81.7	82.1	82.9	82.3	$88\frac{1}{2}$	88	87	89

吾人本可作成一種數字，以示因掉換基年而生之參差，但此種工作，已有人用抽選利於特別論辯之樣本而完成之矣。

然觀乎上表，權數變換如彼之大，結果相差如此之微，是以吾人甚至是否需要以進口貨數量（即上列公式中之 b 值）爲權數，實頗有考慮之價值。玆列一表於下，（註四）以示無甚影響之權數：

以一八八一年指數爲100，用各種加權方式求得之一八九五年指數

	物價之比率（r_1, r_2……）					$\frac{1}{r_1}$，$\frac{1}{r_2}$，…等算術平均數之倒數	英國經濟雜誌之數字
	以一八八一年物價，一八九五年物量之物值加權	以一八八一年公佈物值加權	算術平均數	中位數	幾何平均數		
進口貨	$67\frac{1}{2}$	69	$73\frac{1}{2}$	$72\frac{1}{2}$	$72\frac{1}{2}$	69	71
出口貨	83	87	82	81	$78\frac{1}{2}$	75	

設 b_1，b_2……爲一八八一年之物量

p_1，p_2……爲一八八一年之物價

c_1，c_2……爲一八九五年之物量

r_1p_1，r_2p_2…爲一八九五年之物價

由第一欄，可得

$$100\times\frac{\text{以一八九五年物價與一八九五年物量計算之各種物值總和}}{\text{以一八八一年物價與一八九五年物量計算之各種物值總和}}$$

$=\dfrac{100Sc_1p_1r_1}{Sc_1p_1}$，至加於 r 各值之權數，乃爲一八八一年公佈之物值。

第三欄爲 r 各值之算術平均數，第四欄爲 r 各值之中位數，第五欄爲 r 各值之幾何平均數。倒數第二欄，爲 $\frac{1}{r_1}, \frac{1}{r_2}$……各值之算術平均數，換言之，卽一八八一年物價對一八九五年物價之比率之算術平均數；此算術平均數對 100 之比率，卽等於 100 對另一新指數之比率，至於此新指數，乃卽相當於前一算術平均數而將一八八一年與一八九五年互爲更代者也。最末一欄之數，乃根據倫敦經濟雜誌(Economist)所登載之資料而算出；每年之出口入口貨，均以其前一年之價格計算其物值，是以所得之每年比率，與同表第一欄之比率相同；一八八一年之指數 100，須用此每年比率按年乘之，至一八九五年爲止，則得數乃爲 71。[如用代數表示之，此一指數爲 $100 \times \Sigma\left(r \cdot \frac{bp}{\Sigma bp}\right) \times \Sigma\left(r^1 \cdot \frac{b^1p^1}{\Sigma b^1p^1}\right) \times \times$]

試將上表所列數字，加以較完全之分析，並就出口貨價指數最低75與最高87之歧異考察其原因，卽可證明種種方法究以採用何者爲宜矣。然僅就上表而論，不加權平均數爲 82，第一個加權平均數爲 83，二者相去無幾，吾人大可滿意矣。

處理此種權數之其他方法，容當在本章零售物價指數一節『加權方法』一段內討論之。

客觀度量 以英國商業部基礎而編製之指數，最大之優點，卽在能約略度量一客觀數量，而度量之結果，又能以可訴諸非統計學家之平常人士之用語陳述之，例如『一八九五年之進口貨，

僅值一八八一年價格之一半』；雖然如此，此一指數對於意義欠確定之數量如『進口貨價之跌落』者，未必即為一最佳之度量，蓋所謂『進口貨價之跌落』，其中必有一般原因支配此類商品，但此種商品之行動，乃已被其他部分原因所左右矣。

基年之選擇　選擇一正常年份或一期間之平均數為基年，其事甚為重要，蓋年份之選擇，足以影響嗣後作比較之有效權數也。茲採用下列之符號：

所選權數	基年中之價格	第二年之價格	第三年之價格	第三年價格與第二年價格之比
w_1	100	$100r_1$	$100r_1{}^1$	$R_1=r_1{}^1:r_1$
w_2	100	$100r_2$	$100r_2{}^1$	$R_2-r_2{}^1:r_1$
⋮	⋮	⋮	⋮	⋮

並以 100，I_1，I_2 分別代表此三年之指數，則得

$$\frac{Sw100}{Sw}:\frac{Sw100r}{Sw}:\frac{Sw100r^1}{Sw}=100:I_1:I_2$$

$$\therefore I_2=I_1\times\frac{Swr^1}{Swr}=I_1\times\frac{S(wr\cdot R)}{Swr}。$$

但吾人如以第二年所有之價格均作為 100，則 $I_2=I_1\times\frac{SwR}{Sw}$。

假如此平均數為未加權之平均數，吾人必仍不能免此困難，蓋如此則一者之值為$\frac{SrR}{Sr}$，而他者之值乃為$\frac{1}{n}Sr$也。

但因權數中之差誤，在平常情形之下，影響僅為甚微，故若非所擇之基年，果為極端反常之年份，或物價之運動，極不規則，

則此點並無顧慮之必要。

幾何平均數 據愛基華斯教授之指示，吾人如用幾何平均數，則在不加權平均數中所遇之困難，即可完全避免。試用同一符號：

$$100:I_1:I_2=100:100\sqrt[n]{r_1r_2\cdots\cdots r_n}:100\sqrt[n]{r_1{}^1r_2{}^1\cdots\cdots r_n{}^1}$$

$$\therefore I_2=I_1\times\sqrt[n]{\frac{r_1{}^1r_2{}^1\cdots\cdots r_n{}^1}{r_1r_2\cdots\cdots r_n}}=I_1\times\sqrt[n]{R_1R_2\cdots\cdots R_n},$$

於是不論用何年為基年，若欲比較兩年之指數，得數均為相同矣（註五）。

其他指數 孫巴克先生及倫敦經濟雜誌(Economist),為求在商品之消費相對重要性將各別比率加權時，避免其一部分之困難起見,特自價格最準確之商品中，多用消費廣遍之物品,如小麥,而少用次要之商品,如亞麻仁。孫巴克先生在其一年一度發表於英國皇家統計學會學報(Journal of the Royal Statistical Society)之文中,曾對於彼四十五個比率之不加權平均數,與已用各種原則加權之同一比率之平均數,證明其相合之點（註六）。

選樣正當之重要 如所得比率為數甚多，而所用特別權數之選擇無關重要時,則欲加研究之數量之選擇,對於得數必有極大之影響。例如外國之原料及製品輸入我國之數字,據上列一表所示,勿論權數如何,而所受影響雖然甚微，但與英國製品之出口價格未必即得同一之指數；且此兩種指數無一可與孫巴克氏

指數或經濟雜誌之指數切合，實則卽孫巴克氏與經濟雜誌二者之指數，亦並非完全融和也。此四種指數所根據之樣本，乃由羣類不同之商品選來，且據此等指數所示，同一之勢力，對於此等各羣類之影響，亦各不相同也。總之，吾人對於選來之樣本，無論用何數乘之，但如再用某數除之之時，對於所得到之指數，並不發生影響，則吾人可以期望吾人之得數，必能代表所需之度量矣（註七）。

中位數之絕大優點　試就一八六〇至一八七〇年一期中，以經濟雜誌之指數，與孫巴克氏指數作一比較，吾人可見前者在棉花歉收期中，表現之增加度，較後者強烈多多。一種指數，凡因在一羣商品中之變動，以致出入甚大者，此必係缺乏物價平準之一般度量應有之數種主要原質無疑。此一難點，亦卽加權上所有困難之所在，有一簡單之辦法，可以避免之，此一方法，乃以某年所有物價比率之中位數，作爲該年之指數，如包括項目果然甚多，則理論上示明其他各種平均數，對於必需條件之滿足，較中位數爲佳，乃爲不可能之事，且中位數在實際上爲最便於計算之平均數，則無可疑。

經人提議之標準　在另一方面，資料如甚稀少，採用權數乃爲必需，且公衆感覺需要一種具體度量，以便對加權之便利，明白表示時，最好之標準，恐無過於英國協會之委員會所提議之編

製指數標準，蓋此一製作指數之標準，可作爲商業上關於將來償付交易之基礎也。玆將此標準，述之於下：

一八八八年英國協會經濟組所組委員會介紹之指數基礎

品名	每種物品每年耗用金額（單位百萬金鎊）	故指定之權數爲		價格之根據
小麥	60	5	20	勞工公報之平均數，英國小麥
大麥	30	5		勞工公報之平均數，英國大麥
雀麥	50	5		勞工公報之平均數，英國雀麥
番薯，稻米及其他	50	5		進口價格平均數，番薯
肉	100	10	20	市價，生肉
魚	20	2½		商業部調查表；上陸每英擔(Cwt)平均價格
乳酪，黃油，牛奶	60	7½		乳酪及牛奶，進口平均價格
糖	30	2½	20	平均進口價格，精製糖
茶	20	2½		平均進口價格，茶
啤酒	100	9		平均出口價格，啤酒
酒精	40	2½		平均進口價格，酒精
酒	10	1		平均進口價格，酒
煙草	10	2½	10	平均進口價格，煙草
棉花	20	2½		平均進口價格，棉花
羊毛	30	2½		平均進口價格，羊毛
絲	20	2½		平均進口價格，生絲
皮	10	2½		平均進口價格，獸皮
煤	100	10	20	平均出口價格，煤
鐵	50	5		市價，蘇格蘭銑鐵
紫銅	25	2½		平均進口價格，紫銅礦砂
鉛，鋅，錫	25	2½		平均進口價格，鉛礦砂
木料	30	3	10	平均進口價格
煤油	5	1		平均進口價格
靛青	5	1		平均進口價格
亞麻及亞麻仁	10	3		平均進口價格
棕櫚油	5	1		平均進口價格
硬橡皮	5	1		平均進口價格

（第四十一表）

美國之統計學家，爲製成指數，曾採用一總數比較法(method of comparing totals)，而不用加權或不加權之價格比率(price ratio)。『如此作法,可以克服兩點困難:第一點爲選擇基年問題,因如此則實在物價可以不必化成價比(relative),第二點爲依價比之適宜平均數而決定之問題』(註八)。實則此法固有令人明瞭與作法簡單之優點,但其中並無新原理。茲請述之如下:某年,卽如一九一四年,每一物品之價格,須以在上一舉行調査之年,如一九〇九之成交數量乘之;另一年卽如一九一二年之物價,亦以同數量乘之。以一九一四年之綜合數(aggregate)爲基數(base),卽爲 100,然後以一九一二年之綜合數與一九一四年之綜合數相比,卽得一九一二年之指數。假如一九一四年之物量爲 w_1, w_2……,物價爲 P_1, P_2……,一九一二年之物價爲 p_1, p_2……,則一九一四年之綜合數,卽爲 SwP,而一九一二年之綜合數爲 SwP,於是得指數爲 $100\frac{SwP}{SwP}=100\frac{S(wP.R)}{SwP}$ 式中,R_1, R_2……卽一九一四對一九一二年之價格比率也。此種指數與上述之英國商業部指數相同,而未敢云有特別確度者也。

零售物價指數　對於躉售物價,吾人所得者,旣僅爲粗略之相應性,則度量零售物價,自不能希望有極大之精確。蓋據上章所述,一平均數中之差誤,與在作成此平均數之各個項目中,所出現之差誤,有其一定之關係;如各項中之差誤,通盤均有雙倍,則在平均數中,及在兩平均數之比率中之差誤,大半亦必有兩倍之多,於是吾人不得不更需要四倍之樣本(註九),以期恢復精確

度。然計算零售物價指數,資料之不完全,較計算躉售物價指數所用,有過之無不及,且因包括之物品種類更少,而種種項目如麵包及房租等,實佔極重要之地位,故加權問題,乃愈形重要。

特別困難點 吾人在以一指數表明特殊階級之人士之貨幣購買力時,有若干點,為計算躉售物價指數時在所不計者,此時亦不得不注意及之。在同一時期之不同階級,與在不同時期之同一階級,對於彼之所得,開銷之比例不同,對象亦復有異。吾人果能蒐集若干充分之確實樣本,此一事實卽無十分關係,無如一般對於減價之商品旣有多購之趨勢,此事仍有某種之重要性。旣然如此,吾人似有為每一階級每一區域各作一指數之必要矣。資料不充分與不確實之困難,迄今猶未能克服;惟吾人旣有於將來覓得一定之零售物價記錄甚多足以完成其確實性之可能,則吾人不妨略一討論其他各點矣。

加權法 如為某一階級,作一指數,必須有該階級人民在所論及之所有日期中,對於所得開支方法之記錄,記錄並須甚多,足以得到些須之確度,以應加權之所需。在吾人旣得尙為良好之零售物價記錄後,乃有數種加權法當前(註一〇),而此所有各法,大致多能得出同一之得數。至於加權之必要,及其方法,最佳須以數字例證解明之(註一一)。

不論生活費之定義若何,對於生活費變動之度量,所用之資

料，性質均係相同，內容包括有各層社會階級之代表，在兩期間或兩地方，所購各種商品之數量，及付出物價之記錄。例如，茲有資料如下：

商　　品	地方或日期（甲）					（乙）				
	數　量		價　格		費　用	數　量		價　格		費　用
1	Q_1	×	P_1	=	E_1	q_1	×	p_1	=	e_1
2	Q_2	×	P_2	=	E_2	q_2	×	p_2	=	e_2
3	Q_3	×	P_3	=	E_3	q_3	×	p_3	=	e_3
⋮	⋮		⋮		⋮	⋮	⋮	⋮		⋮
n	Q_n	×	P_n	=	E_n	q_n	×	p_n	=	e_n

在下列第四十二表，卽依此式，表明一九一九年英國生活費委員會報告所用之家庭預算。

如以第一年之物價，計算第二年之預算，則得生活費爲二二五・五便士，而非四五五・五便士。以此爲基礎，而計算零售物價指數，乃爲 $100\times\frac{455.5}{225.5}$，或爲 $100\frac{Sqp}{SqP}=202.0$。加於價格比率 p: P 之權數，爲 qP。指數等於 $100\frac{Se}{Se\frac{P}{p}}$ ········(a)

用第二年物價計算第一年預算，則生活費必爲五二一・六便士，而非二四六・五便士。以此基礎得指數爲

$$100\times\frac{521.6}{246.5}，或爲100\frac{SQp}{SQP}=211.6\cdots\cdots(b)$$

加於價格比率 p: P 之權數，乃爲 Q P，或 E。此卽勞工公報(Labour Gazette)用以度量『零售物價之平均增加量』所用

之法也。

城市勞動階級家庭預算

標準家庭之生活費

		一九一四年			一九一八年，六月			
		Q 數量	P 價格 (便士)	E 費用 (便士)	q 數量	p 價格 (便士)	e 費用 (便士)	p/P 價比
(1)麵包及麪粉	磅	33.5	1.51	50.5	34.5	2.36	81.5	1.56
(2)肉	磅	6.8	8.6	58.5	4.4	18.6	82.0	2.15
(3)鹹肉	磅	1.2	11.7	14.0	2.55	26.1	66.5	2.24
(4)豬油,牛羊油及其他	磅	1.0	7.5	7.5	.78	17.9	14.0	2.29
(5)雞蛋	枚	13	1.0	13.0	9.1	4.0	36.5	4.00
(6)鮮牛奶	瓶	9.2	1.8	16.5	11.7	3.0	35.5	1.69
(7)煉乳	罐頭	.25	6.0	1.5	.59	14.5	8.5	1.42
(8)乳酪	磅	.84	8.9	7.5	.41	20.7	8.5	2.32
(9)黃油	磅	1.70	14.4	24.5	.79	29.7	23.5	2.07
(10)人造牛奶油	磅	.42	6.0	2.5	.91	12.1	11.0	2.01
(11)番薯	磅	15.6	.7	11.0	20	1.25	25.0	1.78
(12)稻米及太皮歐加(tapioca)	磅	1.4	3.2	4.5	1.3	5.8	7.5	1.82
(13)雀麥粉	磅	1.3	1.9	2.5	1.4	4.3	6.0	2.24
(14)茶	磅	.68	21.3	14.5	.57	33.3	19.0	1.56
(15)咖啡茶	磅	.09	16.7	1.5	.12	25.0	3.0	1.50
(16)可可茶	磅	.18	19.4	3.5	.23	32.6	7.5	1.69
(17)糖	磅	5.9	2.2	13.0	2.83	7.07	20.0	3.21
合　計				246.5			455.5	
其他食物				52.5			111.5	
總　計				299.0			567.0	

S.QP=246.5　S qp=455.5　S.Qp=521.6　S.qP=225.5

S.e÷S.E=1.90　S.Qp÷S.QP=2.12　S.qp÷S.qP=2.02

（第四十二表）

在若干種情形之下，上述二式，或用(a)，或用(b)，均各有其理由。如其不然，吾人卽採用二者之平均數，於理亦無不合；算術平均數爲 206.8，幾何平均數爲 206.74，倒數平均數爲 206.69，各數相差無幾，吾人究用何數，並無關係。否則不妨介紹一含義淺顯之法，卽依次計算各數量之平均數（$\frac{1}{2}(Q_1+q_1)$，$\frac{1}{2}(Q_2+q_2)$，……），並求各年之費用，而比較各年費用之總數。於是得出

$\frac{S(Q+q)p}{S(Q+q)P}\times 100=203.7$。至加於價格比率之權數，則爲

$$(Q+q)P \cdots\cdots\cdots\cdots (c)$$

另有一法，乃以每項物品兩期中費用之平均數，作爲該項價格比率之權數，乃得

$$\frac{S(E+e)\frac{p}{P}}{S(E+e)}=198.6 \cdots\cdots\cdots\cdots (d)$$

雖然，此式在分子中，含有 p^2/P 一數量，因此乃對於某一商品價格之異常運動，給以不正當之加重。

如不知數量之多寡時，價格比率之簡單平均數

$$\frac{1}{n}S\frac{p}{P}\times 100=209.1 \cdots\cdots\cdots\cdots (e)$$

有時亦可用之；但在此問題而忽略權數，雖不必求得極大之精度，究非妥當之辦法。

最後一法，手續繁複多多，蓋在此法之下，必須假設第二年之費用總數，其中各項概依第一年之比例而耗用，且第二年所購各項物品之數量，須以第一年之價格而計算其價值。第一年實在全部費用（以 100 乘之），對依上法算出之費用之比率，等於

$$\frac{100\mathrm{S}e}{\mathrm{S}\left\{\mathrm{E}_1\frac{\mathrm{S}e}{\mathrm{SE}}\times\frac{\mathrm{P}_1}{p_1}+\cdots\cdots\right\}}=100\frac{\mathrm{SE}}{\mathrm{SE}\frac{\mathrm{P}}{p}}=196.4\cdots\cdots\cdots\cdots(f)$$

此處加於價格比率$\frac{p}{\mathrm{P}}$之權數，乃爲 $\mathrm{QP}^2\div p$，此一權數，與第(d)例同，對於特別價格，亦給與以不當之權數。並且吾人殊無理由可以假設各年費用中各項物品始終保持定比也。

以上所述度量零售物價六法，究以何法爲最佳？此一問題迄今尚無一致公認之結論；但對於(d)，(e)，及(f)三者，則有基於理論之嚴重反對論。(a)與(b)之間，一般無甚差別，但爲達到此目的，此一年既有併入計算之必要，彼一年又何獨不然，故吾人不若用二得數之平均數之爲愈。在此各種辦法中，平均諸數量之方法，以(c)爲最敏捷，計算極簡易，亦即基於一切理由堪以介紹於吾人之前者也(註一二)。

度量零售物價之運動一問題，一般多與用兩期間之同一項目以度量一標準(此標準代表最低生計(minimum subsistence)或代表有效生計(efficiency subsistence)之費用之變動一問題，混爲一談。本書對於此項度量法之細節，本不預備多所討論，惟吾人不可不知者，各種商品之供給上及價格上之變動，乃爲連續不斷者。對於每一份家庭預算，吾人必須假定在每一日期以最經濟的購買，各得到同一之滋養物（以更廣泛之語言之，各得同一之滿足），如是則價格上漲最少，下落最多之各種食物數量必增

加，而價漲多落少之食物數量則減少，故上升運動必較用(a)法所量得者爲小，而下降運動必較用(a)法所量得爲大也(註一三)。

其他之難點　此外尚有妨礙本問題不能作完全解決之困難兩點，不能不加以考慮。在所有之家庭預算中，房租乃一重要之項目，而增加房租與設備改良(連同與房租一併繳納之捐稅而享受公共支出所給予之利益在內)之關係，現似無有良好估計之期望。再者，假如吾人考慮者，非問錢如何用去，而爲應如何用法，則吾人不得不提出一較爲普通之因子；蓋在必需業已滿足後尚留之邊際，其購買力增漲必甚速，以機製產品種類日繁，而價格日低也；在躉售物價中已算出之降落，或可卽足作爲此一增長之優秀度量焉。

消費指數　姑置此一極難問題而不論，吾人可請仍就對於較爲特色之指數之數量度量法，略一探討之(註一四)。吾人如須度量一一原因之作用，而此原因所支配之數量，並無公同之度量，則仍可用指數以度量之。進口貨物之消費量，已有一般之增加，故如能對此種消費量之增加，加以度量，但並不使其因物價之變動，而受何影響，則吾人卽可用此消費增加量，以批評實際工資之運動之任一度量法矣。麵包，葡萄乾，牛奶酪，肉，以及其他等物品之實在價值，唯一之普通度量，厥爲各該物品之價格，至其重量在此處乃毫無用處可言；故不得不另用其他方法。如將此種物

品若干按年消費之數量一一記下,而以任一年(不必均用一年)消費量之百分數表示之,乃卽得一列僅須加權卽成所求指數之數字。在此情形之下,吾人可以證明:關於權數任何合理之採擇方法,只須以物品之價值或各項物品假定之重要性爲根據,甚或卽用隨機權數方式,其所得之指數,無不與用簡單算術平均數所得者相同。實際上,吾人如有十分良好之一組樣本,吾人幾可完全脫離權數而獨立運用之也。如事實果然如此,則吾人可具有把握而謂所求之數卽在由各種權數方式而來之一羣附近,於是卽擇其似最合理之方式,以作爲吾人所採用之估計數可矣。在務德(Wood)先生一八九九年英國統計學報所載之『勞工階級進步之數種統計』一文中,僅有十四項商品,所用權數有五種不同之方式,而所得答數,在一八七三至一八九六年期間中,消費之增加,均在百分之一三・八,與百分之二〇・一之間。

工資指數 指數在工資統計上之應用,其中並不含有任何新原則。惟在編製工資指數之時,對於權數之變動,不可置之不顧;非然者,工資上漲者人數增加之一般傾向,必無從得知。在各別平均數之中,極有陷於偏誤之危險;蓋加工之工資,特高之計件工資,人類極多而無聯絡之低技能或待遇惡劣工人之工資,往往不見於工資記錄中也。雖然,此等偏誤在比較時乃有趨於消滅之傾向;故由此足見吾人頗有製成具有頗佳確度之工資指數之

可能(註一五)。

(註一)　本文一部分乃係由一九一二年統計學報第七百九十一至七百九十五頁摘錄而來。

(註二)　如重新排列之後，此一方程式可卽由 $y=a+bx$ 得之。

(註三)　見一八九七年六月份之英國經濟學報(Economic Journal)及統計學報。

(註四)　見經濟學報，但對於權數之說明，則已加修正。

(註五)　關於此點，以及本章其他各處，請參閱帕爾格拉夫(Palgrave)之經濟學辭典(Dictionary)中『指數』一條。

(註六)　讀者可參閱一九〇〇年統計學報第九十七及九十八頁。

(註七)　孫巴克氏之指數，見彼在統計學報中每年發表一次之文；另外並有一表示自一八二〇年以來各年指數之圖式，係 P. S. King and Son 書店所印行。

(註八)　請參閱塞可利斯特(Secrist)著，統計方法概論 (An Introduction to Statistical Methods)，一九一七年版，第三百二十九頁三百三十九頁，及三百四十頁。並參閱美國勞工局統計公報 (Bulletin of the U. S. Bureau of Labour) 總號一八一，一九一五年十月號。

(註九)　見本書第二編第四章。

(註一〇)　請參閱帕爾格拉夫氏編經濟學辭典第六百四十至六百四十一頁『名義工資與實際工資』一條。

(註一一)　此處及上下文之一部，均摘錄於統計學報一九一九年第三百四十三頁起之『對於生活費變動之度量』一文。

(註一二)　此意見已異乎本書以前各版所持之論調。爲求更進一步之參考，可閱帕爾格拉夫氏編經濟學辭典，關於工人家庭預算(Workingmen's Budget)之參考書目。

(註一三)　關於此種問題之討論，請參閱一九一九年五月份統計學報，『生活費』一文。

(註一四)　下舉例解，乃以務德(G. H. Wood)先生載於一八九九年英國統計學報『勞工階級進步之數種統計』一文爲根據。

(註一五)　關於此種方法及各項因子，若欲求完全之例解，請參閱『英國工資統

計，第十四編，工程及造船」，見於一九〇六年三月份統計學報，第一百五十四頁起，尤以一百六十六，一百六十八，及一百八十五諸頁爲重要。

第十章　插補法

第一節　總論

插補之必要　在實用統計中，吾人對於時間數字往往不能隨將來進一步研究之需要，作成頻繁之時間數列，或將羣類作一極詳盡之鋪敍。例如一國之人口，只能每十年舉行一次；但吾人需要以按月或按年之記錄——如出生率、死亡率、貿易報告……等等——與現有人口數，聯成密切之關係，又如國家之概算，賦稅之收入，尤須以當年納稅人數之估計數為基準：故在不舉行人口普查年間而估計人口數，誠有插補(interpolate) 之必要也。再者，作依年齡分配之人口數報告，乃為保險精算工作及社會學研究所必需，而此亦非用插補法推算不可。英國人事登記舉辦之戶主報告表，在名義上原確為當年之年齡報告，無如實際上任人皆知其並不確實，以其每有填報接近之整略年齡之情勢也；比較言之，填報年齡在三十五至四十五歲之間者，尙屬較為正確，蓋報告年齡為四十歲之人，其年齡或許不致超出四十歲上下五歲之外也。實則英國在一九一一年以前之原始報告，錯誤之處尙不止此，故歷次登記，迄未實行發表，其發表者亦不過以十歲為一組

之分類人數而已。然則由此分類之人數，自非對各歲年齡之人數，加以估計不可矣。且也，英國一八八六至一八九一年舉辦之工資調查，編製人員計算賺各級工資之人數，乃按『十五先令以上而不滿二十先令者』，『二十先令以上而不滿二十五先令者』……等組分類，並非以一先令爲工資組距之人數也。然在與工資有關之問題中，常時有須詳加推算之必要；且在吾人欲以英國工資，而與法國工資羣類作比較時，必須籌定一種計劃，使二法郎之分組，可與五先令之分組相比較，而此則非用一適當之插補法不可者也。

插補法之需要，當吾人欲比較性質相同而排列方式不同之羣類時，尤極常見。例如，兩國舉行人口普查之日期不同者卽是。一國之人口數，年齡以十五歲爲一組，而另一國則以十歲爲一組；一國以不滿二十一歲者爲未成年人，而另一國則以十八歲以下者爲未成年。至於不定期之估計數，兩國之日期，鮮有能適合者；例如法國之工資統計已舉辦者，有一八四〇，一八五〇，及一八九二年，而英國則爲一八六六，一八八五，一八八六，及一八九一年。在以甲國與乙國比較時，必可求出相同之差度；在此情形之下，決定平均數之方法，必須加以討論，而此一討論，卽將爲幾個初步插補法問題之例釋焉。

初級例證 下列一表，假設表中之正體數字，爲某三區域每

週工資之確實調查數字，茲請就三區中求其整個之平均變動。

	一八六〇		一八六二		一八六四		一八六六		一八七〇		一八七一		一八七五		一八七八		一八八〇		一八八一	
	先令	便士	先令	便士	先令	便士	先令	便士	先令	便士	先令	便士	先令	便士	先令	便士	先令	便士	先令	便士
甲區	12	6	15	0	*15*	*0*	15	0	*15*	*0*	14	6	18	0	18	0	17	6	17	0
乙區	18	0	*19*	*0*	19	0	*20*	*0*	20	0	*19*	*6*	*21*	*0*	21	0	*20*	*6*	20	0
丙區	10	0	*11*	*0*	11	0	12	0	*12*	*0*	*12*	*0*	15	0	*15*	*0*	15	0	*14*	*6*
平均數	*13*	*6*	*15*	*0*	*15*	*0*	*15*	*8*	*15*	*8*	*15*	*4*	*18*	*0*	*18*	*0*	*17*	*8*	*17*	*2*

試觀此表，可知吾人必須由此資料中，求出關於工資一般行程之事實，惟此表所示與吾人者，並不明鮮。上列表中之斜體數字，乃有其自然之成立理由，甲區在一八六二至一八六六年間，並無何種變動，故一八六四年之工資，必爲十五先令無疑。再由乙區判斷，甲區一八七〇年之數字，似不致比一八六四年之數字爲低，故甲區一八七〇年之數字，可寫爲十五先令。甲區既告完全，然後吾人觀察甲區中之初度上升，終止於一八六二年，假定乙區情形與甲區同，則乙區一八六二年之數，必爲十九先令。丙區中之數字，自一八六四年上升，至一八六六年停止，而在甲區，一八六六至一八七〇年間，並無任何變動；假設吾人以乙區一八六六年之數字爲二十先令，則乙區必與甲丙兩區相呼應。設乙區中之數字，一八七一年爲十九先令六便士，一八七五年爲二十一先令，一八八〇年爲二十先令六便士，則一八六六至一八八一年

間，甲乙兩區必成密切之相應。基於同一理由，吾人可將丙區之數字補插而得。於是逐年之不加權平均數，吾人乃可求得，但若用當時日期則不能直接算出也。此一平均數對於原始數字之變動，均能發生反應，惟對任何部分，則並無特殊之加重。由此觀之，此插補完成之數列，可視為以所有資料為基礎之最或然之數列也。

事先之假定 在施用此法之前，不言而喻，必有數種假定，茲就此種假定檢討之。第一項假定，假設各數字間並無猛烈之升降，假設一八六四年甲區，絕不許有二十先令之存在；此一假定，必須有兩個先決條件，能滿足此二條件，假定乃可認為正當而無誤，第一條件，必須熟知支配工資率之一般原因，第二條件，兩期之間，數字必須確知並無激烈之波動。由此論之，當美國南北戰役時，吾人對於棉織工業之工資，即不能施用此一假定，對年代甚長之數列亦然。

第二假定，假設如果同時並無反面的證據，則數字之上升與下降，必係協和運動。例如，乙區中一八七八至一八八一年間，一八八〇之工資，乃即吾人假定為在一八七八至一八八一年間者；假如在甲區各工資分點之間，並無恰在中途之證明，則可謂其乃在時間行程上之三分之二，故如不用整略數，則一八七九年之數，必須以二十先令八便士插補之，而在一八八〇年，則須以二十先

令四便士插補之。

第三，假定三個區域之工資運動行程，彼此相同。例如甲區在一八六〇至一八六二年間，有上升之趨勢，但在一八六六年以前，迄無任何之進展；故吾人假定乙丙兩區記錄上在一八六四年以前之上升運動，確於一八六二年以前實現。又在一八七〇至一八七五年間，甲區工資一直瀉落以至一八七一年爲止，乃又儘情上升而達一八七五年之高度，其後直至一八七八年爲止，竟無任何變動；故乙區工資率之運動，吾人假定在一八七五年之工資，乃與一八七八年者相等，並且一八七八年之下降，乃爲事實所許可，因此一降落，乃愈促成一八七一至一八七五年上升之尖銳也。至在丙區中，一八七一年之十二先令，是否不應爲十一先令六便士，則頗堪致疑。反對此一判斷之理由無他，乃以一則低工資上之增加，不若在高度工資上之增加之易於消滅也；由十二先令中減落六便士，其降落程度，就比例言之，必比由十五先令中減去六便士者爲大；二則在一八七一至一八七五年顯現之三先令六便士之增加，較之甲區或乙區中之上升，比例較大多多；三則一八七〇至一八七一年之降落之存在，乃純由在一八六六至一八七一年間有下降趨勢之證據之故。

數字僅有數個時，必須依上所述詳加分析而求出其最或然之數字；且如此辨理往往頗易補塡對於現存證據適合十分密切

之數字。然於此乃立時發生一疑問：此種數量依假定既爲未知數，則謂此種數量在實際上無論遠近終必近於在表面上爲最或然之數字，吾人究有何種把握耶？

測驗法　在數種插補問題如現在所論者之下，此一問題，吾人可以一數學機率用語答覆之。所謂數學機率用語，例如，與某一數字相差六便士之機率，以二對一而失敗，與某一數字相差一先令之機率，以三〇對一而失敗，與一數字相差二先令六便士之機率，以一〇〇〇對一而失敗……等等卽是。惟在調查時最常偶然出現之數字中，求得確然如此之機率，乃不可能。玆有一簡略但甚有效之方法，可用以測驗上舉一例中插補結果之確度，仍請舉例以明之。現欲測驗者，爲吾人所算出之一八七〇年平均數，在不致十分防害吾人對於本問題之常識範圍內，究爲多少。將甲丙二區在此數十年間儘量擴大；因乙區在一八六四至一八七〇年間，曾有一上升現象，吾人或可由此假設在一八六六年之上，有一先令之上升。如熟知決定此數十年間工資率之原因，則吾人總不能假設，一八七〇年之數字，與一八七五至一八七八年同其高度，亦不能假設在此一年中工資大爲降落，計有二先令之多。假設甲區之最高工資，爲十六先令六便士，丙區之最高工資，爲十三先令六便士，則吾人所得之平均數，乃爲十六先令八便士，而非十五先令八便士。依此同理，吾人或可以十四先令爲一八七

〇年甲區之最低工資，以十一先令爲一八七〇年丙區之最低工資；如是，則平均數必爲十五先令。假設吾人對於在此數十年間事象之一般趨勢，所知甚詳，足以依此方法決定數字大小之範圍，則吾人可以斷定：一八七〇年之平均工資，最小少於十五先令，最大多於十六先令八便士之事，或不易見，而且依證據所示，平均工資爲十五先令八便士一語，勢乃有所難能。

由此觀之，吾人所行插補結果之確度，乃視下列兩件而決定：（一）爲吾人對於數字之可能變動之知識，此種知識之取得，必須就數字所出現之期間，觀察數字之一般變動；（二）爲吾人對於事象（此種事象，卽與數字關聯）運動過程之知識。

數字計算舉例　茲爲例釋數字計算方法起見，特取用同類資料（註一），作爲第二例如下：

英國北部各郡	農村每週工資			
	一八六七至一八六九		一八六九至一八七〇	
	先令	便士	先令	便士
且協爾	*13*	*1*	13	6
耶加協爾	15	0	15	0
約克協爾西區	14	6	16	5
東區	14	6	*14*	*11*
北區	14	6	15	4
得由漢姆	16	6	16	0
腦贊伯蘭	16	6	16	7
可伯蘭	*14*	*4*	14	9
外斯特摩蘭	*15*	*7*	16	1

上表資料就兩期原有數字之五區而論，一八六七至一八六九年一期之五區工資平均數，爲一五先令四・八便士，一八六九至一八七〇年之平均工資，爲一五先令一〇・四便士，換言之，卽爲三三與三四之比。吾人如假定，除此五郡之外，其他各郡之工資，乃受同一原因之支配，並依同一之比率增加，則吾人得列如上表之插補數字。經此補插之後，於是英國北部各郡工資之不加權平均數，在一八六七至一八六九年，爲一四先令一一便士，在一八六九至一八七〇年，爲一五先令五便士，而非原來僅有五區數字之平均數，前期一五先令三便士，後期一五先令五便士矣。如就英國全國以前期與後期作一般比較，吾人只得删略前後兩期中無數字可查之各郡，以免無端掣低一般之平均數，蓋近年來此等各郡之工資，雖較北部各郡平均工資爲低，但終比英國全國總平均數爲高也。在同時，吾人必將北部各郡顯然可見之平均數，作不公正之擡高，且吾人必已失去特別數郡在前期（唯前期乃爲較安全之基礎）之或然數字；蓋北部各郡在最後五十年中，幾全維持於同一狀態之下。由此觀之，可見此類工資，並不若非由補插而來之工資之確實，故吾人務宜留心以此等數字爲基礎之論辯，察其應用之插補數字，究有幾何也。

此種方法，與用以爲在一校中上課時缺席學生判分所用者，頗相近似；給分時一方固須注意該生在同班中之一般地位，一方

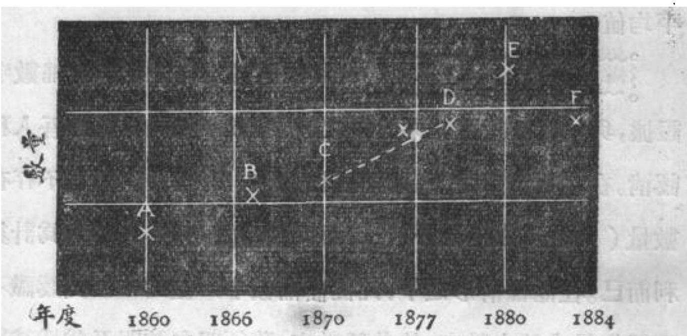
尤須注意缺席學生中除該生外班中其餘之缺席學生所得分數之平均值。

辨別插補數字之必要　此法雖已甚爲完全，但插補數字之證據，與用直接證據結果而得之數字，大不相同，此點吾人務須認清。在某種情形之下，此種插補數字，有時乃代表並不存在之數量（如上述之學校分數），而此數量之用途，亦不過爲計算便利而已。在他種情形之下，此種插補數字，僅爲一種因認識不足而課作最或然之數字。故此種數字，務須明白註明乃爲插補數字；最佳更須敍明求得之方法，如有任何附帶資料（此種資料有時可視爲證明確度之直接證據），亦須敍及，並且如實際上爲可能時，此種數字，不妨作爲並不確切之資料，而視爲有某種範圍中之伸縮性；如且協爾郡之插補數字，可以寫作一二先令六便士至一三先令六便士，不必竟謂爲一三先令一便士也。

論及插補法，其問題有種種不同，其中有須以代數解決者，容在下節討論之，其他數種可以數字例題解釋之如下。

用圖插補法　用圖插補法——設已知各別位置上之數值，如年齡二五至三五歲，三五至四五歲……之人口數；如一八七一年，一八八一年，一八九一年……之人口數；如一八六〇年，一八七〇年，一八七三年……之工資；如工資在一五先令至二〇先令，二〇先令至二五先令……之人數，則吾人可用如下圖式表示事

實：



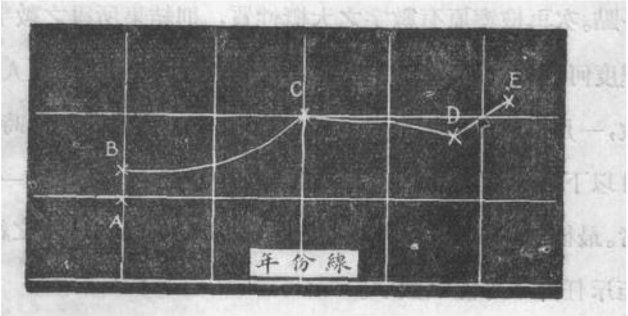
（第二十一圖甲）

設吾人需要一八七五年之數值。如圖上僅有C，D兩點，則一最簡單亦即無反面證明始生效之假定，乃為C與D間之數量，乃依同樣的步調逐漸增加；於是CD一條直線，即代表此種增加趨勢，則x點之高度，必代表一八七五年之數量。

如另外已有一個E點，則CD，DE兩條直線所代表之假定，必不能成立，蓋此假定乃假設在一致趨勢中間，忽於一八七七年之D點，驟然折斷，而所謂一八七七年D點之折斷，並無證據存在也。然則吾人不可不就所有各已知點通盤計議，即經過此等所有各點，畫一條線，此畫成應之線，應儘量使其均勻，曲折力求減少，蓋若無反對方面之證據時，吾人即假定在此種數量中間，並無猛烈變動也。此一曲線之構成，可以用數學原理爲基礎，不然即用隨手畫法（freehand method），亦無不可；如果採用隨手畫

法，則所畫之線，必時常隨所持論辯之許可，十分接近事實。

此一方法，只可用於連續之數量，如各級年齡之人數，各年份之人口數，一極大工資羣類中，各級工資上之人數均是。例如，英國全國之平均工資，變動必甚遲緩，但倫敦一市建築工人之工資，乃竟因某日罷工議和之結果，驟起猛烈之變動。在此情形之下，繪曲線必須盡力之所能以求切合證據；例如下列第二十一圖乙：



（第二十一圖乙）

圖中 AB 線代表驟然之昇進；BC 線代表因商業發達之逐漸加速度的增加，CD 線為工資到達 C 點後之遲緩下降勢，至 DE 線，則為一堅挺而蒸蒸向上以謀恢復已往頹勢之力量。

循環數字　吾人如知每一年為一週期之數字之每年平均數，並設有若干每月平均數，足以用第七章第四節所述方法，推測此種數字之循環變動，則任何月份，苟有關調查材料，雖甚確

實可信，但倘殘缺不全時，吾人均可插補之。例如就英國勞工公報所載之失業人數而論，在所有各月份中，週期雖不甚顯著，但吾人仍能查出其週期，蓋在春季必有普遍之降勢，晚秋必有一般之升勢，而六月份則大概均爲最低之月份。於是吾人頗可應用前在第七章第四節所列之第十七圖甲、乙、丙三圖，先將所用之資料，在圖上標示各點，然後隨平均數線之或爲上升，或爲靜止不動，抑或下降，畫出波浪曲線，則各變動之曲線，必經過已有標記之各點。次再檢察原有數字之大概性質，則結果所得之數字，確實程度何若，吾人卽可知之；並由該公報查得：一則失業人數百分數，一月中變動向無有過於兩單位者，二則變動向無延時在三四月以下者；三則失業人數百分數，向無有在一以下或在一〇以上者。最後吾人可察視某日之商業盛衰史，並由此所得之結果，可拒斥任何不可能之數字。

補助曲線之利用　如吾人採用第七章第三節所述之構圖法，及同章第五節所述之變動方程式，而能查出兩組數列間之密切關聯，則可擇其較完全之一組，以便補插另一組缺失之數字。當此之時，第一必先詳細檢閱兩組數列在均有完全數字時之相應密切度及相應性質。次復依照類似前在第七章第三節所列第十三圖之形式，另行作圖，而圖上各線之中，留有一不完全之線。後完成此一殘缺之線，使其隨原有各點之地位，儘量與已完成

之線相切合，於是吾人對於缺失之數字，即得最或然之數值矣。至於由此所得之結果，確度如何，測驗之法，與前論相同。此一方法用以插補數字，用途甚多：如用一財源之收入，以插補另一財源之收入；如用進口貨值以插補出口貨值；如用對外貿易數值以插補結婚率；如用一區域之工資，以插補另一區域之工資；如用食物消費量之變動，以插補失業之人數；如當吾人知全部人口之變動情形時，以插補一部人口之變動狀況，以及其他甚多之數列等均是。

第二節　代數方法

插補問題，最值得吾人注意者，可述之如下：當有一數量呈現連續性之有規則變動，另有一數量，其變動與前一數量發生聯結關係，且吾人已知此第二數量之某某僅有數個之不連續數值，或能直接推算出來時，則吾人之任務，即在為此第二數量與第一數量相當之某項數值，求出其或然值。例如，已有至一五歲，二〇歲，二五歲……等等年齡後之壽長平均數 吾人須求出中間年齡之壽長平均數；或如已有某國在一八七一，一八八一，一八九一，及一九〇一年之人口數，須求其在中間年份之人口數。於此有唯一為情理所許之假設數則：其一，假設數量必連續變動，換言之，即任何中間數，絕無破裂之處；其二，假設數量之變動率，亦為連

續不斷，換言之，即代表變動之線乃爲平滑均勻者，而非爲多角形。

研究插補法，有系統之探討，只能用代數上之有限相差(finite difference) 一法，惟欲應用此法，不可不自符號之定義，及若干基本公式之導來始。

(一)設 y 爲 x 之連續函數，並設 y_0，y_1，y_2，……爲 y 之各值，同時 x 之各值則爲 x_0，x_1，x_2……。

於是吾人排成一表如下：

x之值	y之值	第一相差	第二相差	第三相差
x_0	y_0			
x_1	y_1	Δ_0^1	Δ_0^2	
x_2	y_2	Δ_1^1	Δ_1^2	Δ_0^3
x_3	y_3	Δ_2^1	Δ_2^2	Δ_1^3
x_4	y_4	Δ_3^1	Δ_2^3	Δ_2^3
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

此處每一 Δ，乃由前一欄恰在其下之項，減去前欄恰在其上之項；例如，$\Delta_0^1=y_1-y_0$，$\Delta_1^1=y_2-y_1$，……。$\Delta_0^2=\Delta_1^1-\Delta_0^1$，……。$\Delta_0^3=\Delta_1^2-\Delta_0^2$……。此表假設之可以向下並向右無限繼續。

於是吾人可得

$$\Delta_0^2=\Delta_1^1-\Delta_0^1=(y_2-y_1)-(y_1-y_0)=y_2-2y_1+y_0$$

$\Delta_t^2=y_{2+t}-2y_{1+t}+y_t$，式中 t 爲任何整數。

$$\Delta_0^3=(y_3-2y_2+y_1)-(y_2-2y_1+y_0)=y_3-3y_2+3y_1-y_0$$

$$\Delta_t^3=y_{3+t}-3y_{2+t}+3y_{1+t}-y_t$$

且就一般情形論之，如用普通證明二項定理常用之歸納法，並均含同一係數，則

$$\Delta_0^r = y_r - r\cdot y_{r-1} + \frac{r(r-1)}{1\cdot 2}y_{r-2} - \frac{r(r-1)(r-2)}{1\cdot 2\cdot 3}y_{r-3} + \cdots\cdots$$

直至 $r+1$ 項爲止$\cdots\cdots(\alpha)$

$$\Delta_t^r = y_{r+t} - r\cdot y_{r+t-1} + \frac{r(r-1)}{1\cdot 2}y_{r+t-2} + \cdots\cdots$$ 直至 $r+1$ 項爲止$\cdots\cdots(\beta)$

式中，r 爲任何整數。

又得

$y_1 = y_0 + \Delta_0^1$，及 $y_2 = y_1 + \Delta_1^1 = (y_0 + \Delta_0^1) + (\Delta_0^1 + \Delta_0^2) = y_0 + 2\Delta_0^1 + \Delta_0^2$,

同理，　$\Delta_1^1 = \Delta_0^1 + \Delta_0^2$,

且　$\Delta_2^1 = \Delta_1^1 + \Delta_1^2 = (\Delta_0^1 + \Delta_0^2) + (\Delta_0^2 + \Delta_0^3) = \Delta_0^1 + 2\Delta_0^2 + \Delta_0^3$。

$$\therefore\quad y_3 = y_2 + \Delta_2^1 = y_0 + 3\Delta_0^1 + 3\Delta_0^2 + \Delta_0^3,$$

且依同理　$\Delta_3^1 = \Delta_0^1 + 3\Delta_0^2 + 3\Delta_0^3 + \Delta_0^4$。

依此程序,繼續向下推演,吾人可重得兩項係數(the Binomial Coefficients),則

$$y_r = y_0 + r\cdot\Delta_0^1 + \frac{r(r-1)}{1\cdot 2}\Delta_0^2 + \cdots\cdots \text{直至} r+1 \text{項爲止} \cdots\cdots\cdots\cdots(\gamma)$$

$$\Delta_r^t = \Delta_0^t + r\cdot\Delta_0^{t+1} + \frac{r(r-1)}{1\cdot 2}\Delta_0^{t+2} + \cdots \text{直至} r+1 \text{項爲止} \cdots(\delta)$$

更往下推演之,則

$$y_{r+s} = y_s + r\cdot\Delta_s^1 + \frac{r(r-1)}{1\cdot 2}\Delta_s^2 + \cdots\cdots \text{直至} r+1 \text{項爲止} \cdots\cdots(\epsilon)$$

式中 s 爲任一整數。

例如,設 $y=x^4$,並設 x 之各值爲 $0, h, 2h, 3h,\cdots\cdots$則

x之值	y之值	相差 第一級	第二級	第三級	第四級	第五級
0	0	h^4				
h	h^4	$15h^4$	$14h^4$			
$2h$	$16h^4$	$65h^4$	$50h^4$	$36h^4$		
$3h$	$81h^4$	$175h^4$	$110h^4$	$60h^4$	$24h^4$	
$4h$	$256h^4$	$369h^4$	$194h^4$	$84h^4$	$24h^4$	0
$5h$	$625h^4$	$671h^4$	$302h^4$	$108h^4$	$24h^4$	0
$6h$	$1296h^4$					

由公式(α)得

$$\Delta_0^4=(256-4\times81+6\times16-4\times1+0)h^4=24h^4,$$此處r等於4。

由公式(β)得

$$\Delta_0^5=(7^4-5\times6^4+10\times5^4-10\times4^4+5\times3^4-2^4)h^4=0,$$

式中 $r=5, t=2$。

由公式(γ)得

$$(5h)^4=(0+5+10\times14+10\times36+5\times24+0)h^4=625h^4,$$

式中 $r=5$。

由公式(δ)得

$$\Delta_2^3=(36+2\times24+0)h^4=84h^4,$$式中$r=2, t=3$,

又由公式(ϵ)得

$$(5h)^4=(16+3\times65+3\times110+84)h^4=625h^4,$$式中$r=3$,

$s=2$。

（二）如 y 與 x 之關係，成爲下列形式:

$$y=a_0+a_1x+a_2x^2+\cdots\cdots a_nx^n,$$

x之值又係成爲算術級數，卽依次爲$x_0, x_0+h, \cdots\cdots x_0+(n-1)h$,

則由此可知$\Delta_0{}^n = a_n . h^n \underline{n}$，且除此之外，更無較高級之相差。

因為$\Delta_0{}^1 = a_0 - a_0 + a_1(x_0 + h - x_0) + \cdots\cdots + a_n\{(x_0+h)^n - x_0{}^n\}$

$= ha_1 + \cdots + a_n\{nhx_0{}^{n-1} + x_0$ 之次級乘方$\}$

$\Delta_1{}^1 = ha_1 + \cdots + a_n\{nh(x_0+h)^{n-1} + (x_0+h)$之次級乘方$\}$

$\Delta_0{}^2 = 2h^2a_2 + \cdots + a_n\{n(n-1)h^2x_0{}^{n-2} + x_0$ 之次級乘方$\}$

由此可知 $\Delta_0{}^1$ 之乘方，最高無過於 $x_0{}^{n-1}$；$\Delta_0{}^2$ 之乘方，最高無過於 $x_0{}^{n-2}$。

繼續向下推演——

$$\Delta_0{}^n = a_n n(n-1)\cdots\cdots 3\cdot 2\cdot 1 h^n = a_n h^n n! \cdots\cdots (\zeta)$$

於是$\Delta_0{}^{n+1}$ 及更高級相差，乃完全消滅。

在上一例中，$y = x^4$，$a_n = 1$，$n = 4$，故

$$\Delta_0{}^4 = 1\cdot h^4 \cdot 4! = 24h^4，又\Delta_0{}^5 = 0。$$

反之，假設相差至第n級為最高，其上更無相差，則如下列附註所證明，y 與 x 之方程式，乃為下一形式：$y = a_0 + a_1x + \cdots\cdots + a_nx^n$。

附註——相差 (Difference) 與引申函數 (Derived Functions) 或各微係數 (Differential Coefficients) 之關係，在相差原理上，佔有極重要之地位，茲可以連續微分法 (Method of Operators) 簡括表明之。

援用微積分常用符號，依照台洛爾氏定理 (Taylor's Theorem)，得公式如下：

$$f(x+h) = f(x) + hf'(x) + \frac{1}{2!}h^2f''(x) + \cdots\cdots = e^{hD}\cdot f(x),$$

式中 D代表連續微分，至e^{hD}則須展開為$1+hD+\frac{1}{2!}h^2D^2+\cdots\cdots$，然後用之將$f(x)$各項依次展開。D之所以用為一代數符號，純因有$D\{Df(x)\}=D^2f(x)$，$D^m\{D^nf(x)\}=D^{m+n}f(x)$，$aD(f(x))=D(af(x))$，$\cdots\cdots$之關係。

但 $\Delta f(x)=f(x+h)-f(x)=(e^{hD}-1)f(x)$。

$\Delta\{af(x)\}=a\Delta f(x)$， $\Delta\{\Delta f(x)\}=\Delta^2 f(x)$，

$\Delta^m(\Delta^n f(x))=\Delta^{m+n}f(x)$，$\Delta$亦作一代數符號用之。

故 $\Delta\equiv e^{hD}-1$

$\Delta^n\equiv(e^{hD}-1)^n=(hD+\frac{1}{2!}h^2D^2+\cdots\cdots\cdots)^n=h^nD^n(1+\frac{1}{2}hD+\frac{1}{6}h^2D^2+\cdots\cdots)^n=h^nD^n(1+\frac{n}{2}hD+\frac{n(n+1)}{24}h^2D^2+\cdots\cdots)\cdots\cdots(1)$

且 $hD\equiv\log(1+\Delta)$

$h^nD^n\equiv\{\log(1+\Delta)\}^n=(\Delta-\frac{1}{2}\Delta^2+\frac{1}{3}\Delta^3\cdots\cdots)^n$

$=\Delta^n(1-\frac{n}{2}\Delta+n\frac{(3n+5)}{24}\Delta^2+\cdots\cdots)\cdots\cdots(2)$

但如 $f(x)=a_0+a_1x+\cdots\cdots+a_nx^n$，

$D^nf(x)=a_n\cdot n!$，

及 $D^{n+1}f(x)=0=D^{n+2}f(x)\cdots\cdots$

$\therefore$ $\Delta^nf(x)=h^na_nn!$，又依公式(1)，則

$\Delta^{n+1}f(x)=h^{n+1}D^{n+1}(1+\cdots\cdots)f(x)=0$，如上正文。

反之，如$\Delta^{n+1}f(x)=0=\Delta^{n+2}f(x)=\cdots\cdots$

則由公式(2)，$D^{n+1}f(x)=0\cdots\cdots D^nf(x)=$常數$=c_n$，

$D^{n-1}f(x)=c_nx+c_{n-1}$，

$$D^{n-2}f(x)=\frac{1}{2}c_nx^2+c_{n-1}x+c_{n-2},$$

及
$$f(x)=\frac{1}{n!}c_nx^n+\cdots\cdots+c_1x+c_0\text{。}$$

故第 n 級相差如爲常數，則此函數爲有理的，整數的，並爲第 n 次方。

牛頓(Newton)氏插補公式，如下(K)所述，用連續微分法，可簡捷求得如下：

$$y=f(x_0+k)=e^{kD}f(x_0)=(1+\Delta)^{\frac{k}{h}}f(x_0),\text{惟因 } e^{hD}=1+\Delta,\text{故}$$
$$=f(x_0)+\frac{k}{h}\Delta f(x_0)+\frac{1}{2}\cdot\frac{k}{h}\left(\frac{k}{h}-1\right)\cdot\Delta^2f(x_0)+\cdots\cdots$$
$$=y_0+\frac{x-x_0}{h}\Delta_0^1+\frac{x-x_0}{h}\cdot\frac{x-x_0-h}{2h}\Delta_0^2+\cdots\cdots$$

在此式中，$x=x_0+k$。

當第 n 級相差（或曰第 n 次引申函數）爲零時，依公式(β)所示，應

$$y_{n+t}-ny_{n-1+t}+\frac{n(n-1)}{2}y_{n-2+t}\cdots\cdots\pm y_t=0\cdots\cdots(\eta)$$

不論 t 之值爲何。

（三）普通之插補公式，全以一假定爲基礎，在此假定之下，假設對於所求值所在地之鄰近地段之觀察數，可以一連續函數，$y=f(x)$代表之。

據假定，此函數可以展開爲 x 之一個冪級數，如一般對於連續函數（註二）然，則吾人可得下式

$$y=a_0+a_1x+a_2x^2+\cdots\cdots+a_nx^n\cdots\cdots(\theta)$$

式中 n 爲 x 之最高乘冪之指數，其值如何，尚待加以決定。如a_0,

$a_1, \cdots\cdots a_n$ 各値選用得當，則此方程式可以若干$(n+1)$對之 x 及 y 値滿足之。例如，對於 $y=a_0+a_1x$ 一條直線，可以選定兩點（或云兩對數値），對於 $y=a_0+a_1x+a_2x^2$ 一條拋物線，可以選出三點，其他以此類推。

最簡單之形式，厥爲 $y=a_0+a_1x$；吾人採用此式，必須假定用成比例之部分（此法在用對數表，三角函數表及其他數學表時多施用之），以作插補，其結果尙爲確實。在此情形之下，此第一級相差，與第一次引申函數（或名線之斜度），必爲常數。

拋物線有三個數値，用此拋物線，必須假定斜度之變動，爲均勻一致，其第二級相差及其第二次引申函數，乃爲常數。

因更高級相差之差量，必須將項數加多，而函數之展開，至第 n 項卽行截止，必能與第 n 級相差之恆性相適應。

設吾人之問題，爲於一已知數學函數中之插補問題，則删略第 n 級相差之差量，對於計算上發生何等之影響，吾人可以測驗之。例如在一七位數對數表中，作下列之數字：

自然數	對數	相差 第一級	第二級	第三級	第四級	第五級
20	1.3010300	.0211893	−.0009859	+.0000876	−.0000110	+.0000015
21	1.3222193	.0202034	−.0008983	+.0000766	−.0000095	+.0000015
22	1.3424227	.0193051	−.0008217	+.0000671	−.0000080	+.0000016
23	1.3617278	.0184834	−.0007546	+.0000591	−.0000064	
24	1.3802112	.0177288	−.0006955	+.0000527		
25	1.3979400	.0170333	−.0006428			
26	1.4149733	.0163905				
27	1.4313638					

由此觀之,連續各次相差,依次漸減,甚有規則,至第六級,相差已絕不過.0000001矣。

此中原理,在統計上應用時,一般並不知有函數之存在,吾人只得假定,確有此函數存在,並可展開成為一種收斂甚速之級數,惟其收斂甚速,吾人頗可删略第五項(譬如說)後所有之各項;否則,如以稍欠準確之語言之,吾人假定此產生總計數之原因,乃有由此一點至彼一點依次逐漸變動之效果,故此種變動之差量,僅在一小部分間,甚屬輕微。

(四)設y之各值為$y_0, y_1 \cdots\cdots y_n$,與此等數值相當者,則有$x$之距離均等各數值,如$x_0$, x_0+h, x_0+2h, $\cdots\cdots x_0+nh$。由此吾人可求得公式(θ)之係數,特其中經過需用算術工作極其煩難,而較便應用之算式,尚有用相差法解算之一途也。

試請就下列一方程式而考慮之

$$y=y_0+\frac{x-x_0}{h}\Delta_0{}^1+\frac{x-x_0}{h}\cdot\frac{x-x_0-h}{2h}\Delta_0{}^2+\frac{x-x_0}{h}\cdot\frac{x-x_0-h}{2h}\cdot\frac{x-x_0-2h}{3h}+\cdots\cdots+\text{直至 } n+1 \text{ 項為止}\cdots\cdots(\kappa)$$

(牛頓公式)。

如$x=x_0$,則$y=y_0$。

如$x=x_0+h$,則$y=y_0+2\Delta_0{}^1=y_1$。

如$x=x_0+2h$,則$y=y_0+2\Delta_0{}^1+\Delta_0{}^2=y_2$。

如 $x=x_0+rh$，則 $y=y_0+r\Delta_0{}^1+\frac{r(r-1)}{1\cdot 2}\Delta_0{}^2+$ 至 $r+1$ 項止，後面數項已消失，故依公式(γ)之規定，$y=y_r$。

然則(κ)顯而易見必爲 n 次方，而且可以上述 n 對數值滿足之。

例如，用上列一表，求 $y=\log 20.5$。

$x_0=20, h=1, x=x_0+\cdot 5, y_0=1.3010300, \Delta_0{}^1=.0211893$，餘從略。

$y_0=1.3010300+.0211893\times .5+\frac{1}{2}(.5)(-.5)(-.0009859)+$

$\frac{1}{6}(.5)(-.5)(-1.5)(.0000876)+\frac{1}{24}(.5)(-.5)(-1.5)$

$(-2.5)(-.0000110)+\frac{1}{120}(.5)(-.5)(-1.5)(-2.5)$

$(-3.5)(.0000015)$。

取用爲首二項，可得 $y=1.3116247$

取用爲首三項，可得 $y=1.3117479$

取用爲首四項，可得 $y=1.3117534$

取用爲首五項，可得 $y=1.3117538$

取用全數各項，可得 $y=1.3117538$

至眞值則爲 1.3117539。

此點對於統計上之應用，容在第十目討論之。

（五）反之，吾人如已知 y，即可得一求 x 之方程式，而此乃可以亨納(Horner)氏方法或他法解決之。

例如，吾人僅有四個觀察值，而欲求其中位數，則解算程序

如下。設有 y_0 人工資在 x_0 以下，y_1 人工資在 x_0+h 以下，y_2 人工資在 x_0+2h 以下，並有 y_3 人工資在 x_0+3h 以下，並設總數共有 $(2y_m-1)$ 人，則 x_m（為 x 各值之一，與 y_m 相當）即為中位數。

於是

$$y_m=y_0+\frac{x_m-x_0}{h}\Delta_0{}^1+\frac{x_m-x_0}{h}\cdot\frac{x_m-x_0-h}{2h}\Delta_0{}^2$$

$$+\frac{x_m-x_0}{h}\cdot\frac{x_m-x_0-h}{2h}\cdot\frac{x_m-x_0-2h}{3h}\Delta_0{}^3,$$

此為求 x_m 之三次方程式。

吾人有隨意以 x_0 為任何一組開始之自由固矣，惟分組之決定，應以中位數恰在插補之中間組為宜。例如，吾人如用上述三次方程式，則包含中位數在內者，厥為 x_0+h 至 x_0+2h 一組。

前在第五章第五節所述求中位數公式之計算，即删略其第二級以上之相差，並以 x_0 至 x_0+h 一組為含有中位數之組。然則 $y_m=y_0+\frac{x_m-x_0}{h}(y_1-y_0)$，故

$$x_m=x_0+\frac{y_m-y_0}{y_1-y_0}\cdot h.$$

如求衆數，吾人仍當以 y 為達於 x 值之累積數。據吾人所知，衆數乃為最簡單之一種，且就一般而論，只用四個觀察值即足，如此則衆數恰在第二與第三組之間。然後吾人只用公式(κ)之前四項，並求曲線最陡峭時，亦即橫坐標每一單位所佔次數為最多時，x 之值為何。然 D_xy 即為最大，則 $D_x{}^2y$ 必為零無疑。

$$0=D_x{}^2y=\frac{1}{h^2}\Delta_0{}^2+\frac{x-x_0-h}{h^3}\Delta_0{}^3.$$

$$\text{故}\quad x=x_0+h-\frac{h\Delta_0{}^2}{\Delta_0{}^3}=x_0+h+\frac{(u_2-u_1)h}{(u_2-u_1)+(u_2-u_3)},$$

在此式中,u_1 代表 y_1-y_0, u_2 代表 y_2-y_1, u_3 代表 y_3-y_2,而 u_1 卽爲 x_0 至 x_0+h 之間之次數, u_2 爲 x_0+h 至 x_0+2h 間之次數, u_3 爲 x_0+2h 至 x_0+3h 間之次數。如衆數在第二組,則 $u_2>u_1$, 而 $u_2>u_3$。該一公式,指示吾人, x_0+h 至x_0+2h 間之組距,必須如何除之,以便求得衆數所在之位置。(請參閱第五章第四節『求衆數法』末段)。

(六)中間相差數——在插補法中,吾人往往只能應用y之若干值,而在此若干 y 值中, 吾人欲以確定之某項數值之部位,乃在全部之中央,於是僅用公式 (θ) 以求之,有時必不能順利達到吾人之目的。當此之時,於是乃有同等公式產生, 以避免偏態情形,此公式卽應用所謂『中間相差數』(central differences)者也。但此公式並不含有新的原理,實乃由公式(θ)轉變而成。至此爲止所用之相差數,爲別於其他之相差起見,可名之曰『升級相差』(ascending differences)。

其適當之符號,乃如下列:

$$\begin{array}{llllll}
x_{-2}=x_0-2h & y_{-2} & & & & \\
 & & \delta_{-\frac{3}{2}} & & & \\
x_{-1}=x_0-h & y_{-1} & & \delta^2{}_{-1} & & \\
 & & \delta_{-\frac{1}{2}} & & \delta^3{}_{-\frac{1}{2}} & \\
x_0 & y_0 & & \delta^2{}_0 & & \delta^4{}_0 \\
 & & \delta_{\frac{1}{2}} & & \delta^3{}_{+\frac{1}{2}} & \\
x_1=x_0+h & y_1 & & \delta^2{}_1 & & \delta^4{}_1 \\
 & & \delta_{\frac{3}{2}} & & \delta^3{}_{+\frac{3}{2}} & \\
x_2=x_0+2h & y_2 & & \delta^2{}_2 & & \\
 & & \delta_{\frac{5}{2}} & & & \\
x_3=x_0+3h & y_3 & & & &
\end{array}$$

此處 $\delta_{\frac{1}{2}}=y_1-y_0$；$\delta^2{}_0=\delta_{\frac{1}{2}}-\delta_{-\frac{1}{2}}=y_1-2y_0+y_{-1}$；$\delta_0{}^4=y_2-4y_1+6y_0-4y_{-1}+y_{-2}$，其他以此類推。

有一 x 值，同時必有一 y 值，今以 x 值除 x_0 至 x_1 之組距，使成 $p:q$ 之比率，則 $x=x_0+ph=x_1-qh$，而 $p+q=1$。

於是再用替換法，可知下一公式

$$y=py_1+qy_0-\tfrac{1}{6}pq\{(p+1)\delta^2{}_1+(q+1)\delta_0{}^2\}$$
$$+\tfrac{1}{120}pq(p+1)(q+1)\{(p+2)\delta^4{}_1+(q+2)\delta_0{}^4\}\cdots\cdots(\lambda)$$

（此式——以 $q=1-p$——乃爲一有理整數函數而爲 p 之五次方，同時亦爲 q 之五次方）必可以 $(x_{-2}y_{-2})(x_{-1}y_{-1})\cdots\cdots(x_3y_3)$ 六對數值滿足之；但若將含有第四級相差之一項删略，只用由 $(x_{-1}y_{-1})\cdots\cdots$ 至 (x_2y_2) 四項，亦能滿足之。

茲爲舉例釋明應用符號起見，吾人可以 $y_1=\log 23$，已見於本節第三目之對數函數相差表中，並以 $p=.2$，然後計算 $\log 22.2$。

$$\log\ 22.2=.2\ \log\ 23+.8\ \log\ 22$$
$$-\frac{.16}{6}\{1.2\text{ of }(-.0008217)+1.8\text{ of }(-.0008923)\}$$
$$+\frac{.16\times1.2\times1.8}{120}\{2.2\text{ of }(-.000095)+2.8\text{ of }(-.0000110)\}$$
$$=1.3462837+.0000694-.0000002=1.3463529。$$

而眞值則爲 1.3463530。

此一公式之重要，在並無一般代數函數而欲僅由鄰近數項

插施行補時，愈爲顯然。

（七）拉格郎支氏公式——上論(ζ)(η)(θ)(κ)及(λ)數則公式，所論情形，必須 x 之觀察值，彼此有相等之距離。至各觀察值距離並不相等時，迄無如此簡單之插補方法。惟拉格郎支(Lagrange)氏，曾得出一方程式，係爲第n次方者，並能滿足(x_0y_0)，$(x_1y_1)\cdots\cdots(x_ny_n)$共有$(n+1)$對數值，不論$x$各值間之關係若何也，玆請述之如下：

$$y=y_0\frac{(x-x_1)(x-x_2)\cdots(x-x_n)}{(x_0-x_1)(x_0-x_2)\cdots(x_0-x_n)}+y_1\frac{(x-x_0)(x-x_2)\cdots(x-x_n)}{(x_1-x_0)(x_1-x_2)\cdots(x_1-x_n)}$$

$$+\cdots\cdots+y_n\frac{(x-x_0)(x-x_1)\cdots(x-x_{n-1})}{(x_n-x_0)(x_n-x_1)\cdots(x_n-x_{n-1})}\cdots\cdots(\mu)$$

式中分子，可爲任何分數，譬如卽爲 y_t 之乘數亦可，求此分子，只須乘 $(x-x_0)(x-x_1)\cdots\cdots(x-x_n)$ 各因子而略去 $x-x_t$ 卽可；至於分母，則以 x_t 代 x，由分子卽可得之。

理由最爲明顯者，當 $x=x_t$ 時，除 y_t 之乘數而外，其餘各分數，均爲零，蓋 y_t 之乘數，乃爲整一，故 $y=y_t$。

（八）前以公式(θ)表示 y 與 x 相互關係時，所下之假定，現頗有再加考慮之必要。

假如 y 與 x 用一函數法則發生聯結關係，換言之，卽假如對於 x 之所有已知各數值，y 亦各有確定之數值（此爲一假定，無此假定，多數之插補問題，必將失其意義），則 y 可作爲 x 之一函

數而表示之，譬如示如 $y=f(x)$ 者卽是。如該函數及其引申函數，係爲連續的，則依麥克老介氏定理，

$$y=f(0)+xf^1(0)+\frac{x^2}{2!}f^2(0)+\frac{x^3}{3!}f^3(0)+\cdots\cdots$$ 以下連續至於無窮。

如 $f^{n+1}(0)$ 及隨後之係數甚小，且 x 永不大，則自第 $n+2$ 項以上之各項，較之以前各項，所值甚微，大可删略不計，故僅以在前之 $n+1$ 項，已能約略決定 y 之值；但依本節第二目附註中之第(1)，(2)兩公式所規定，當 Δ^{n+1}，$\Delta^{n+2}\cdots\cdots$ 等等均小時，f^{n+1} 乃甚大，Δ^{n+1}，$\Delta^{n+2}\cdots\cdots$ 均甚大時，則反是。由此觀之，吾人可得一結論於下：y 與 x 間任何之函數關係，必均可化爲 n 次元之拋物線方程式（見公式 θ），如高於第 n 級之相差消失時，且如此種高於第 n 級之相差雖未消失但爲甚小時，方程式(θ)終爲表示此函數關係之約略公式。

但如經過所示各點，畫成之曲線，其確度爲連續而變動緩和時，則可證明：臨近各點之第二級相差，必不甚大，蓋縱坐標增加率變動過急，則曲度變化必速也。又如吾人更作一第二曲線，橫坐標仍如前，而以第一級相差，爲其縱坐標，則甚小之第三級相差，卽爲第一級相差並無激烈變動之表示，其他以此類推。但如在此點之外，則爲插補法背景之假定，與各次連續相差之減少量，二者之關係，必不易看出。雖然，在反對方面則較爲明顯；假如在任一數字級數中，依試驗所示，知連續各項相差，有趨消滅之傾

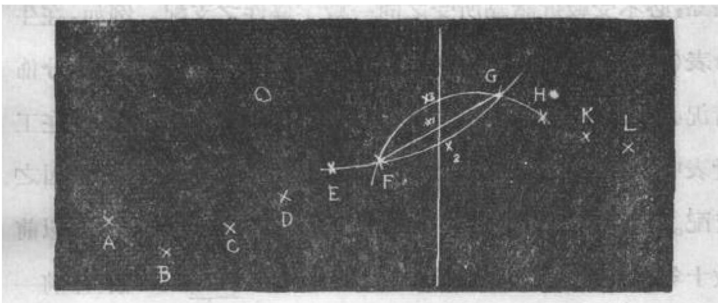
向，則不論任何曲線，凡經過此諸點者，均可由一拋物線方程式，約略表示之。德摩剛(De Morgan)氏，述此結論如下:『假如吾人取定之 n 點，彼此相距不遠，其橫坐標，爲一算術級數，而有甚小——至少須不太大——之公相差，至其縱坐標，彼此相距並不過於懸殊時，則一有 $n-1$ 次方之拋物線，必將與同樣大概形式之任何有規則曲線，大致相合，至少當以在同點與同點間爲然也』。布爾氏(Boole)對此之解釋，爲:『依慣例，吾人須假定所提數值之一般方程式，爲 x 之一有理且爲整數之函數，然後根據所給條件，以決定常數。此一假定之基礎，建築於一「假設連續各次之相差減退甚速」之假定上，而此假設，確已在一切函數(註三)表中證實之矣』。

依照本節第二目附註中公式(1)之規定，當 h 甚小時，連續各級相差之級數愈高，則連續各級相差對於任何曲線之影響亦愈微，然則吾人以假定更高級相差均消滅爲基礎，建立任何函數數值之一級數，自不失爲一合法之處置。

假如隨手畫成之曲線，畫時確能經過選就之固定點，且此曲線之曲度，變化儘量求其緩和，則必可得一曲線，與用公式(θ)所求得之曲線，相合甚近。此種曲線，絕似騎腳踏車者，故意經過數點，或避免幾處障礙物時，所過之軌跡。

(九)由以上之研究，可知隨意經過若干點，吾人卽可畫成

一調勻而連接之曲線；蓋用拋物線方程式(θ)，當 x 連接變動時，$y, \frac{dy}{dx}$，或$\frac{d^2y}{dx^2}$之各值，絕不致現出驟然之跳動；而且吾人可隨常數之多寡，如數求出若干一次方程（唯一次方程，乃永必有眞值），而方法甚簡單，只須就原來方程式中之 n，即作爲固定點之點數即可。



（第二十二圖）

並設吾人意欲在F與G之間一條已定直線上找出一個點，則僅就F與G而論，將此兩點畫成一條直線，然後找出 x_1 一點固可；如就E，F與G，或就F，G，與H而論，畫成一條拋物線，因而求得 x_2 或 x_3 亦可；即或不然僅就E，F，G及H畫成三次之拋物線，亦無不可。由此三次拋物線上，亦能於近於下點處找出一轉向點(point of inflexion)。此一條線，必約略與騎腳踏車者所經路線相合，假如此騎腳踏車者，由E點起行，騎向臨近之點H，然後趨往下及G。但如吾人連D及K兩點合併在內（假如，騎腳踏車

者由D點起行，經過E，F，G及H四點，而達於K點)，則全部曲線，必將稍有改變。同理，吾人如併入各點愈多，則FG之路綫，將不斷微受影響。再者假如吾人併入所有較近各點，能使FG線與一最終位置愈益約略相近，同時如更將較遠各點合併在內，卽能使FG一線與最終位置離開，則吾人可得一結論，斷定此較遠諸點，必不受較近諸點所受之同一數字條件之支配。例如，在生命表(table of survivals)中，年齡在五歲以下者之數字，其分佈情況必不與用年齡較長者之數字所繪成之曲線相合。又如在工資表中，可見工資甚高之數字與工資低者，並不受同一原因之支配。在他一方面，每次舉行人口普查所得之數字，乃須視以前數十年之數字而定。故如插補一八七六年之英國人口數，僅將一八五一，一八六一，一八七一，一八八一，及一八九一年，或並將一九〇一年之數字合併計算在內，吾人必將得若干不同之數字。此種處置方法，並不足奇，蓋插補一八七六年之人口，果有錯誤，若非連查二十五年之數字，必不能免除也。由此可見，距欲插補之數字，所在之期間，甚遠之諸點，必不若相距甚近之諸點，所生之影響之大，且據試驗所示，此一條件，在上述之方法中，可以完全滿足也。不特此也，在公式(κ)之級數中，連續各項之係數，自第r項(此時$x<x_0+(2r-3)h$)起，卽行逐漸減少，換言之，卽當x在x_0與x_0+h之間時，第一級相差之係數，必開始漸減也。

在此吾人可以注意，此曲線之突升突降，已爲一條件所限制，在此條件之下，有 $n-1$ 次方之曲線，其轉向點必不致超過 $n-3$ 個，蓋$\frac{d^2y}{dx^2}$並無一項之方次，能大於 x^{n-3} 也。

仍就上舉之例而論，由 F 起至 G 止中間諸點，可用 D，E，F，G，H，或用 E，F，G，H，K 五點求得之。此兩條曲線，在 F 與 G 之間，可以治合爲一。臨近 F 之諸點，以用前五點求得爲確實，而在此五點中，F 恰在中央；至接近 G 之點，可用後五點求之，G 亦在其中央。兩條曲線在 F 與 G 間治合爲一之線段，與第一曲線在 F 點相接，與第二曲線接觸於 G 點。故爲簡便起見，此可以用正弦 (sine) 曲線求之。以意度之，英國人事登記總監處所用者，不外乎此也。

雖然，現在所論之統計插補法理論，吾人殊不能謂之有完全滿意之基礎（註四）。蓋一則支配此一理論之原則，並未善爲說明，二則將原理與事實發生關係，必須有方法，而此種方法之數理的研究，並未完全也。然既有較爲密切之方法，如再苦心另求別法，恐亦無此必要，蓋一則除非吾人確知支配此種數字之法則之代數公式外，僅恃插補所得之數字，求其完全精確，事乃絕不可能，二則玆所討論之方法，依經驗所示，已能滿足條件，若再求進一步之精密，亦不過產生些須之修正而已。

（十）公式在數字上應用之例釋

（1）玆有工資以五先令爲一組之工人人數，試估計工資在二四先令以上而不滿二五先令之人數。

		每千人中之工人人數（成年男子）	相差			
			第一級	第二級	第三級	第四級
工資在10先令以上而不滿	15先令	39				
			257			
	20先令	296		46		
			303		−144	
	25先令	599		−98		151
			205		7	
	30先令	804		−91		18
			114		25	
	35先令	918		−66		
			48			
	40先令	966				

上表中工資在一五先令以下之人數之漸增相差數，可删略之。

用公式（κ），$x_0=20$（先令），$h=5$，$y_0=296$，$\Delta_0^1=303$，$\Delta_0^2=-98$，$\Delta_0^3=7$，$\Delta_0^4=18$。

工資不滿二五先令之組，$y=599$，見上表。

工資在二四先令之處，$x=24$，

$$y=296+\frac{4}{5}\times303+\frac{4}{5}\cdot\frac{-1}{10}\times(-98)$$
$$+\frac{4}{5}\left(\frac{-1}{10}\right)\cdot\frac{-6}{15}\times7$$
$$+\frac{4}{5}\cdot\frac{-1}{10}\cdot\frac{-6}{15}\cdot\frac{-11}{20}\cdot18$$
$$=296+242.4+7.84+.224$$
$$+.3168=547\text{（約略）}。$$

故所求數爲 $599-547=52$。

再在二三先令處，$x=x_0+3$，$y=489$，故工資在二三先令以上而不滿二四先令之人數爲 58。

（2）設一八一三年之進口貨值，因記錄已燬於火，茲請作一估計以補充之。

各年進口貨值如下：

一八一〇…………£39,202,000…………y_1。

一八一一………… 26,510,000…………y_2。

一八一二………… 26,163,000…………y_3。

一八一三………… 26,163,000…………y_4。

一八一四………… 33,755,000…………y_5。

一八一五………… 32,987,000…………y_6。

一八一六………… 27,431,000…………y_7。

用公式 (η)，僅取 y_3 及 y_5 二數字，並假定第二級相差業已消滅，

$$y_5-2y_4+y_3=0,\ 則\ y_4=29,959。$$

用公式 (η)，並加用 y_2 及 y_6 二值，並假定第四級相差業已消滅，

$$y_6+y_2-4(y_5+y_3)+6y_4=0,\ 則\ y_4=30,029。$$

依公式 (η)，並加用 y_1 及 y_7 二值，並假定第六級相差業已

消滅，

$$y_7+y_1-6(y_6+y_2)+15(y_5+y_3)-20y_4=0,\text{則} y_4=30,421。$$

觀上各值，第一，第二相距甚近，而第三值，則大不相同，故吾人可即採用 £30,000,000 爲所求值。

（3）在布斯（Boothe）先生之『人民之生活與勞動』第五卷第四十六頁，載有表示各等級年齡分配之圖式數則，用處甚多，茲將其所用數字列下：

年　齡	在十歲至八十歲總數一萬人中所佔之比例	各組年齡中每歲平均數
10—15歲	193.5	38.7
15—20歲	880	176
20—25歲	933	188.6
25—35歲	1636	163.6
35—45歲	1201	120.1
45—55歲	830	83
55—65歲	434	43.4
65—80歲	192.5	12.8

布斯先生之圖式，即用最末一欄畫成，此最末欄之數，即爲各該年齡分組中點之縱坐標。既將各點求出，然後即連以一條直線。如此辦理，如就其原來目的而言，確已盡確實之能事，惟吾人

如欲從中間年齡中求出較爲詳盡之數字，則布斯氏之方法，可供吾人爲研究插補問題之有趣舉例。

年　齡	年齡在 x 以下者在總數一萬人中所佔之比例
$15=x_1$	$193.5=y_1$
$20=x_2$	$1073.5=y_2$
$25=x_3$	$2006.5=y_3$
$35=x_4$	$3642.5=y_4$
$45=x_5$	$4843.5=y_5$
$55=x_6$	$5673.4=y_6$
$65=x_7$	$6107.5=y_7$
$80=x_8$	$6300\quad=y_8$

請用拉格郎支氏公式 (μ)，以求年齡在三十歲以下之人數，至年齡在五十五歲以上之人數，則略去之。如此則 $x=30$。

$$y=193.5\times\frac{10.5(-5)(-15)(-25)}{(-5)(-10)(-20)(-30)(-40)}$$

$$+1073.5\times\frac{15.5(-5)(-15)(-25)}{5(-5)(-15)(-25)(-35)}$$

$$+2006.5\times\frac{15.10(-5)(-15)(-25)}{10.5(-10)(-20)(-30)}$$

$$+3642.5\times\frac{15\cdot10\cdot5(-15)(-25)}{20\cdot15\cdot10(-10)(-20)}$$

$$+4843.5\times\frac{15\cdot10\cdot5(-5)(-25)}{30\cdot25\cdot20\cdot10(-10)}$$

$$+5673.5\times\frac{15\cdot10\cdot5(-5)(-15)}{40\cdot35\cdot30\cdot20\cdot10}=2879。$$

此數據布斯先生圖式之計算，乃爲 2824.5，而布斯氏所用者，僅爲 y_3 及 y_4 二值。

在此公式中，如僅用y_2,y_3,y_4,y_5四值，則y據計算乃爲2869。

以上所用之拉格郎支氏公式，意與假定第六級相差消滅而年齡分佈甚爲均勻相等。茲以三十歲之 y 值爲 a，四十歲之 y 值爲 b，五十歲之 y 值爲 c。

然後用 y_1，y_2，y_3，a，y_4，b，y_5，依公式(β)或公式(η)，則得

$$y_1-6y_2+15y_3-20a+15y_4-6b+y_5=0,$$

且依同理，

$$y_2-6y_3+15a-20y_4+15b-6y_5+c=0$$

又

$$y_3-6a+15y_4-20b+15y_5-6c+yb=0$$

由此直進而計算之，則 $a=2879$ 如上。此一方法，應用之時，較用拉格郎支氏公式，計算尤爲簡單。

（4）爲例釋求中位數及衆數方法起見，吾人可用已於第四章第四節第十三表所用之數字，茲列表於下：

工資		x	y	相差			
在$.25	以上	−1	0				
				317			
在 .75	以上	0	317		1157		
				1472		−1332	
在 1.25	以上	1	1789		−175		
				1279		−152	
在 1.75	以上	2	3086		−327		+15
				970		−137	
在 2.25	以上	3	4056		−464		
				506			
在 2.75	以上	4	4562				

工人總數爲 5123，爲求中位數，以 $y=2562$，並用自 $x=0$ 至 $x=4$ 之各項。如求至第四級相差爲止時，則

$$2562=317+1472x-\frac{1}{2}\cdot175x(x-1)-\frac{1}{6}\cdot152x(x-1)(x-2)$$
$$+\frac{1}{24}\cdot15x(x-1)(x-2)(x-3)$$

$$\therefore 61488=7608+36122x-111x^2-698x^3+15x^4$$

解此方程式，如用亨納(Horner)氏解法，得 $x=1.5715$。

故中位數乃在 \$.75+1.5715×.50=\$1.536。

此外另有一法，乃假設 x 爲 y 之函數（註五），如用拉格郎支公式，則

$$x=\frac{(y-y_1)(y-y_2)(y-y_3)}{(y_0-y_1)(y_0-y_2)(y_0-y_3)}x_0+++$$

吾人如僅用上列一表之四項，則得

$$x=\frac{(2562-1789)(2562-3086)(2562-4056)}{-1472\times-2769\times-3739}(\times0)+++$$

由此得 $x=1\cdot5624$，於是中位數爲 1·531。

此法適用計算機計算之。

爲求衆數，請用自 $x=-1$ 至 $x=-2$ 之各項。

本節第六目公式之相差符號表中，如用以求第二及第三級相差，此處之第二相差爲 1157，第三相差爲 −1332

所求値爲 $\$.75+\frac{1157}{1332}\times.50=\$1.18$，

其實其他各種方法，無不可以應用，惟結果則略有不同。事

實上，上表分組既如是之寬，而更高級相差並不漸趨於零，求出之總數，自不能十分精確。

此一方法，應用之處甚多，如人口數，結婚率，出生率，死亡率及其他等等，在何期間，其增加率為最速，及在何年年齡，死亡之機率增加最烈，種種問題，均可應用之（註六）。

（十一）當一種原始調查報告，尚須加以修正時，例如在人口普查所得原始材料中，以決定依年齡之人口分佈時，乃又有一類極重要之插補問題焉。

現先就經過極多點之附近，但不必經過任何一點，以繪一修勻曲線之問題而論。在此之時，必須有一假定，此假定一則假設所得報告甚少不敷獨立應用，或其確度不足使人完全信賴，二則假設據此種報告所示，年齡確有有規則之分佈，而此即調查報告有表現之原來使命者也。

（1）第一法，須假定極大羣類中之平均數均為確實。然後採用上文所論之任一方法，以插補此中平均數。

（2）第二法在討論修勻曲時（見前第七章），業已用過。茲可重述如下：取二點，或三點，或四……十點之連續羣類，依一定之橫坐標再三由不同之縱坐標起畫之。然後求每一羣類之重心；換言之，即立定一縱坐標，使與在此羣類之外部縱坐標之橫坐標兩端之中間一點上之羣類各縱坐標之平均數相等。經過求得之

各點，書一條線。則吾人可以察見，此一條線必能與已定之條件相合。此一方法之例釋，已列如在第七章第一節之第六圖。

(3)另外一法，乃須將原始數字之曲線，加以修勻，直至第四級，或第五級，或更高級相差消失時爲止；然後再應用普通插補公式。

例如，援用本章本節第十目之第一例，可重作一表於下：

工　資 在十五先令以上	修勻後之人數	修正相差數 第一級	第二級	第三級
至 20 先令爲止	296			
		$303+a$		
至 25 先令爲止	$599+a$		$-98-a+b$	
		$205+b$		$7-3b$
至 30 先令爲止	$804+a+b$		$-91-a-2b$	
		$114-a-b$		$25+2a+3b$
至 35 先令爲止	918		$-66+a+b$	
		48		
至 40 先令爲止	966			

吾人如以 $b=2\frac{1}{3}$，則 $a=-16$，而第三級相差乃消滅，於是 $\Delta_0^1=287$，$\Delta_0^2=-79\frac{2}{3}$，$\Delta_0^3=\Delta_0^4=0$；當 $x=25$ 時，則 $y=583$，又當 $x=24$ 時，則

$$y=296+\tfrac{4}{5}\cdot 287-\tfrac{2}{25}\cdot\left(-79\tfrac{2}{3}\right)=531.97$$

然則工資在二十四先令而不滿二十五先令者之人數，現經求得爲 51，並非 52。

任何原始數字，均可加以修正。

吾人現時尙須多解一方程式，以便完成自工資二十先令至三十先令之表格。

當 $x=23$ 時，$y=296+\frac{3}{5}\cdot287+\frac{3}{25}\cdot79\frac{2}{3}$。此值與 y 之值之差額，在 $x=24$ 時，爲 $\frac{1}{5}\cdot287-\frac{1}{25}\cdot79\frac{2}{3}=54\cdot21$。

於是吾人乃得出下列一表，在此表中，斜體字之數字，乃舊已算出者，至用正體字之數字，則爲在第三級相差爲零之假定下加入者。

工資	人數	相差數		
		第一級	第二級	第三級
至 20 先令爲止	296		…	
		63.75		…
至 21 先令爲止	360		3.18	
		60.57		0
至 22 先令爲止	420		3.18	
		57.39		0
至 23 先令爲止	478		3.18	
		54.21		0
至 24 先令爲止	532		3.18	
		51.03		0
至 25 先令爲止	583		3.18	
		47.85		0
至 26 先令爲止	631		3.18	
		44.67		0
至 27 先令爲止	676		3.18	
		41.49		0
至 28 先令爲止	717		3.18	
		38.31		0
至 29 先令爲止	755		3.18	
		35.13		…
至 30 先令爲止	790		…	

吾人計算第二級相差，如果更爲確切，則上表最末一數，自應爲 $804+a+b=790\frac{1}{3}$。

在此法之下,一經發現其重要差(Signipicant differences),則許多數字,卽可隨手找出,但在原來資料,已甚確切時,此法亦能作普遍之應用。

(4)此外另有一法,所用數理較深,本應在將差誤律(the Law of Error)研究之後,再行討論,方爲適宜;惟此時不妨先作一簡短之解釋,順便介紹一有用之公式。

假設有相連之點五個:$(-2,y_{-2})$,$(-1,y_{-1})$$(0,y)$,$(1,y_1)$,$(2,y_2)$,均爲已知。

經過此五點,可畫成一具有四次方之拋物線,但此線卻有兩個轉向點。如臨近所有五點,可畫成三次方之拋物線,爲數極多,各線並無轉向點,但亦能滿足普通之插補法之條件。

茲借用最小二乘法(註七)之一原理,假設拋物線

$$y=a+bx+cx^2+dx^3$$

之係數,選定時足使此數量

$$\Sigma(a+bx+cx^2+dx^3)^2$$

(式中總和號一直施用於 x 與 y 之五年數値)成爲最小,則如此求得之拋物線,卽爲最適於其目的。

關於此點,其必不可少之數理的研究,吾人可就達爾文(Darwin)教授之『難免錯誤之度量』(On Fallible Measures)(註八)一文,加以探討,蓋上一方法,卽根據該文而來也。

年次	英國人口平均每人負擔進口之小麥數量（磅）	相差數				修勻後數字
		第一級	第二級	第三級	第四級	
一八九〇	226					
		18				
一八九一	244		−17			
		1		19		
一八九二	245		2		−16	$245+\frac{3}{35}$ of $16=246\frac{1}{2}$
		3		3		
一八九三	248		5		13	$248-\frac{3}{35}$ of $13=247$
		8		16		
一八九四	256		21		−94	$256+\frac{3}{35}$ of $94=264$
		29		−78		
一八九五	285		−57		134	$285-\frac{3}{35}$ of $134=263\frac{1}{2}$
		−28		56		
一八九六	257		−1		−16	$257+\frac{3}{35}$ of $16=258\frac{1}{2}$
		−29		40		
一八九七	228		39			
		+10				
一八九八	238					

每年年底之小麥存貨，年有不同，且無記錄可查，消費量之統計，並無確實可言。故此列數字，應以倘待修正，並須削去過大之出入者視之，方爲合理。

（5）除此而外，尚有一較爲普遍之插補問題，乃須求出一代數公式，與前時所用，目的在表現全數列或全羣類之拋物線方程式，迥有不同。關於此公式，本書第二編第五章，曾有一簡短之介紹焉。

附註——公式(λ)乃愛爲來特(Everett)教授所建立，通用名詞及公式之證明，均爲彼之貢獻（見理論與實用數學季刊，一九〇一年，第一百二十八號，公式G）。

茲覓得其證明如次：

假如$f(x)=\cosh\left(2q\ \sinh^{-1}\frac{x}{2}\right)$，則立可證明$f^{n+2}(0)=(q^2-\frac{1}{4}n^2)\ f^n(0)$，故依麥克老令氏定理，$f(x)$之展開式，爲

$$1+\frac{1}{2!}q^2x^2+\frac{1}{4!}q^2(q^2-1)x^4+\frac{1}{6!}q^2(q^2-1^2)(q^2-2^2)x^6+\cdots=\cosh\left(2q\ \sinh^{-1}\frac{x}{2}\right)。$$

經微分並以qx除之後，則得

$$q+\frac{1}{3!}q(q^2-1)\,x^2+\frac{1}{5!}q(q^2-1^2)\,(q^2-2^2)x^4+\cdots=\frac{2}{x\sqrt{4+x^2}}\sinh\left(2q\ \sinh^{-1}\frac{x}{2}\right)$$

$$=\frac{\sinh(qh\mathrm{D})}{\sinh(h\mathrm{D})},$$

在此最後一式中，$x=2\sinh\left(\frac{h\mathrm{D}}{2}\right)$。

據照本章本節第六目之符號，$\delta_0{}^2=(e^{h\mathrm{D}}-2+e^{-h\mathrm{D}})\,y_0=\left(e^{\frac{h\mathrm{D}}{2}}-e^{-\frac{h\mathrm{D}}{2}}\right)^2y_0$，如是，則$\delta=2\sinh\left(\frac{h\mathrm{D}}{2}\right)$。

$$y_{ph}=e^{ph\mathrm{D}}(y_0)，又y_1=e^{h\mathrm{D}}(y_0)。$$

$$e^{ph\mathrm{D}}=\{e^{(p+1)h\mathrm{D}}-e^{(p-1)h\mathrm{D}}\}\div(e^{h\mathrm{D}}-e^{-h\mathrm{D}})$$

$$=\{e^{qh\mathrm{D}}-e^{-qh\mathrm{D}}+(e^{ph\mathrm{D}}-e^{-ph\mathrm{D}})e^{h\mathrm{D}}\}\div(e^{h\mathrm{D}}-e^{-h\mathrm{D}})$$

因 $p+q=1$，$$=\frac{\sinh(qh\mathrm{D})}{\sinh(h\mathrm{D})}+\frac{\sinh(ph\mathrm{D})}{\sinh(h\mathrm{D})}e^{h\mathrm{D}}。$$

$$\therefore y_{ph}=\frac{\sinh(qh\mathrm{D})}{\sinh(h\mathrm{D})}y_0+\frac{\sinh(ph\mathrm{D})}{\sinh(h\mathrm{D})}y_1。$$

在上級數中，x與δ乃爲一個，且如用先此級數表示對於y_0之微分，次（經以p代q後）表示對於y_1之微分，則

$$y_{ph}=qy_0+\frac{1}{3}q(q^2-1)\delta_0{}^2+\frac{1}{5!}q(q^2-1^2)(q^2-2^2)\delta_0{}^4$$

$$+\frac{1}{7!}q(q^2-1^2)(q^2-2^2)(q^2-3^2)\delta_0{}^6+\cdots$$

$$+py_1+\frac{1}{3!}p(p^2-1)\delta_1{}^2+\frac{1}{5!}p(p^2-1^2)(p^2-2^2)\delta_1{}^4$$

$$+\frac{1}{7!}p(p^2-1^2)(p^2-2^2)(p^2-3^2)\delta_1{}^6+\cdots$$

此卽通則化之公式(λ)也。

關於插補法之其他書籍，讀者可請參考：(1)發爾 (Farr) 博士之生命表(第三表)，一八六四年版；(2)布爾 (Boole) 先生之『有限相差』(Finite Difference)，『保險精算學院課本』(Text-Book of Institute of Actuaries) 第二本，第四百二十頁起；(3)來斯 (Rice) 氏之『插補法之理論與實線』(Theory and Practice of Interpolation) 一八九九年版；(4)麥瑞費爾德(Merrifield)氏之『求方法與插補法』(On Quadratures and Interpolation)，見一八八〇年英國協會報告；(5)蕭文內(Chauvenet)氏之『球面及實用天文學』(Spherical and Practical Astronomy)；(6)吳爾好斯(Woolhouse)在保險雜誌第十一，第十二卷發表之文；(7)愛爲來特(J. D. E erett) 氏之『相差表之代數論』(On the Algebra of Difference Tables)，見一九〇〇年第一百二十四期數學季刊 (Quarterly Journal of Mathematics)；及其『中間相差插補公式』，見一九〇〇年英國協會報告，及一九〇一年一月份保險精算學院月報：(7)薛伯 (W. F. Sheppard) 博士之『中間相差公式』(On Central Difference Formulae)，見倫敦數理學會會報第三十一卷，第七百零七至七百一十期)，及『補助曲線在連續差量統計中之應用』(On the Use of Auxiliary Curves in Statistics of ContinuonsVariation)，見一九〇〇年九月份統計學報。在以上文獻中，並可查得其他參考書籍。

(註一) 此項資料取自一八九八年十二月份統計學報中本書原著者之『英格蘭之農村工資』一文。表中數字正體字爲原有，斜體字爲插補而來。

(註二) 此言尤以連續之函數，及其引申函數爲連續並非爲無限之函數時爲恰當。

(註三) 卽爲數學函數，如 $\int_0^x e^{-x^2}dx$，非統計近似數也。

(註四) 對於數學函數數值之插補，不在此範圍之內。

(註五) 請參閱英國一八九八年號統計學報第六百九十八頁愛基華斯 (Edgeworth)一文。

(註六) 參閱一八九九年號統計學報第三百八十一頁所載愛基華斯氏一文，其他參考書籍，該文另有介紹。

(註七) 參閱本書第二編附十。

(註八) 參閱一八七七年七月份英國哲學雜誌(Phil. Mag. and Journal) 如此，吾人可將第七章第二節所列第十圖中之C線，加以修勻。

第二編　數理統計之部

第一章　頻數曲線

第一節　導言

數學方法，在統計領域中，極多部分，均甚重要。本書第一編，曾用代數法，將算術結果歸納於通則，較爲簡單之插補法，亦多用之。但非借重高深數理不能解決之問題，厥類甚繁，本編之作，卽欲對此方面，略加討論者也。然統計學中，應用數理之範圍，本至廣擴，非本編篇幅，所能盡行包羅，玆僅擇其方法之重要者，及與經濟學或其相關科學有直接關係之問題，討論之。實則醫學，生物學及其他科學中之統計問題，原亦須用同一方法，此事於相當之期刊見之。玆爲討論方便起見，本編通例，以認定由經濟社會調查而生之問題爲討論對象，應用例證，亦儘量於此限定範圍內取材。

數學方法之應用，有種種不同，然大要分之，不外三類：

（1）各羣類之有系統的敍述(The systematic description of groups)；

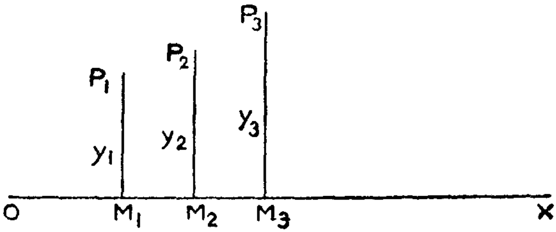
(2)現象間相互關係之測算 (The measurement of relationship between phenomena);

(3)抽樣方法所得結果精度之測算(The measurement of the precision of results obtained by a process of sampling)。

各種分析之基礎,乃爲機率原理(Theory of chance)(註一)其高深僅若干數學專家造詣及之,然吾人殊不能假定讀者之所通曉已超過代數機率之較淺理論;而欲介紹參考書籍,英文中又乏熟識教本;不得已,只得佔用許多篇幅,專論純粹數學原理;惟對於曾受數學訓練而非精於此道者,仍力求顧到,以期其能以領會。故於可能範圍內,能不用微積分而能到到證明者,必不用之;各項問題之結論,儘量用文字明白闡述,用算術例題證明;最簡單之例題,首先加以討論,以說明各種程序及結果,而較爲普通之研究,則舉其概略,而以曾有透徹研究之書文,列爲參考。至不諳數學之讀者,如將書中小體字,略去不讀,亦未爲不可。書末附有附錄,凡他處不易得到證明之定理,擇要搜羅之,而分析過於繁難之部分,在教本中不便多所發揮者,亦歸併附錄之內。

第二節 頻數羣類及曲線(Frequency groups and curves)

以下專就頻數羣類之有系統的測量討論



設有任何羣類之數量於此，Ox軸卽代表此數量，在此軸上，分列尺度，具有數量x_1者有y_1個，x_2有y_2個，其他以此類推；如此，圖上$OM_1=x_1$，$M_1P_1=y_1$其他以此類推，則該羣類，乃如上圖所示。

M_1M_2，M_2M_3，……各級(grade)不必相等。

假如該類之總數爲n，使

$$n=y_1+y_2+\cdots$$

如此，在x_1，x_2…等處觀察之『頻數』(frequency)爲

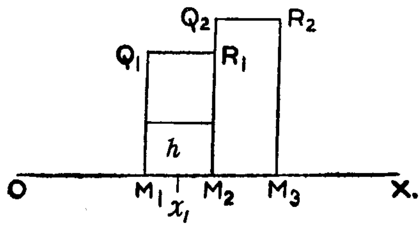
$$\frac{y_1}{n},\frac{y_2}{n}\ldots$$

設P_1，P_2，P_3……諸點，可視爲在一連續曲線(continuous curve)上，則其軌跡(locus)卽成頻數曲線(Frequency curve)。

又設量數並不分級，亦不按特定數值而分爲小羣，而每一次觀察，卽得一個量數，則此一羣類，卽成爲一滿佈星點之横軸，一點卽是一項目，

以後所論之公式，大部可以適用於此滿佈星點之綫，及頻數曲綫亦同。

一個羣類之各個量數，往往聚集一起，分別形成等級（譬如20-25，25-30……歲）或者在着手時，即已採用最狹之單位（譬如 55-56， 56-57……吋）。在此情形之下，各級個數，約略與長方形相等（譬如在 M_2M_3 一級之 $M_2M_3R_2Q_2$）。



以h代表各級之寬，x_1x_2……爲各級中點之橫坐標，y_1，y_2……爲長方形之高，而y_1h，y_2h……爲各級之個數。則

$$n=y_1h+y_2h\cdots\cdots,$$

如以h作單位，則

$$n=y_1+y_2+\cdots\cdots.$$

各級之頻數爲

$$\frac{y_1h}{n},\frac{y_2h}{n}\cdots\cdots.$$

如果一個連續曲線之條件可以決定,然後構成圖式,而使建在 $M_1M_2, M_2'M_3'$……上面之諸部分曲線面積,與 y_1h, y_2h……成比例,則此曲線,卽爲該類之頻數曲綫。

變動參差爲自然界之通律,人事變幻,多所不免;頻數羣類,於是乃由大規模觀察而生。羣類可分爲四種:

(a)一羣類中各個分子,皆予測查,如某業成年男工工資是;

(b)從羣類中選樣觀察,如由一五萬戶之小城市,抽樣一千戶,而查其兒童人數若干,或測量某一種樹之樹葉,均屬此類;

(c)物體數量之重複測量(如重複測量某星位之赤緯),此種數量之變量,乃由於工具差誤(Instrumental "error");

(d)各種數目實現之數學機率(mathematical probability),(如擲錢五十個,擲得一,二,三……個表面向上之機率是);或因未知之複雜原因而決定之事件次數。

無論現象屬於何類,普通所用方法,完全相同。在此法之下,須選出x值及y值之某種代數函數,並須以數值代表x各值及y各值之代數函數,以表示該羣類。其實,羣類之敍述乃係:

(1)決定其中心位置,

(2)測量觀察對此中心之離散度,

(3)測量其去中心之偏斜度,

(4)根據代表一羣類之圖形,作其他之測量。

中心位置之決定，可用算術平均數，中位數，或衆數，有時用幾何平均數亦可。算術平均數，其他計算上需用最廣，故應以之為一切之慣常出發點。中位數無助於一般代數上之演算；精確數值又不可常得，如非別有用途，無計算之必要。至於衆數，普通甚難從觀察中確切決定，欲介紹其近似值，當此初步計算時又非所宜；然如有一代表羣類之確定代數公式，衆數可確實得出，則亦未嘗不可重視。

測量離散度，可以採用『機誤』(probable error)，卽四分位差(Half-intequartile range)，或用平均差(mean deviation)或均方差 (deviation of mean square) 在此數者之中，機誤一者與中位數同，只能求得其近似值，對於進一步之測量，甚難為有系統之運用。平均差，其據以測算之原點位置、含混不清，姑置不論，且因其根本量數，無視符號之正負，將來應用，必感極大困難。惟均方差，此種困難，可以完全避免。所謂均方差，乃各量數對平均數離差平方之平均數方根。故不僅易用代數推演，且為多種演算所必需。此之謂標準差，數理統計上普遍應用者卽此是也(見第一編第六章)。

曲線不對稱，中位數，衆數及算術平均數，數者必不能相合於一點，而上下四分位數對中位數之距離，亦必不能相等。若羣類為對稱時，此等數量，必等於零；故此種數量，可為測量之根據；

但中位數，衆數及四分位數，只能得其近似值，作成之測量，必受觀察不充分之影響而難期完善，且變更觀察量，亦難得若何效果，除非將其轉移而逾過中位數或四分位數也。

測量必須與每一數值之位置感應銳敏。用在平均數下數值與平均數上數值兩個平均差之餘數，固無不可，惟此餘額，不易列成公式，以與其他有系統的測量，相互爲用。故爲免除困難起見，特採用平均立方差（各數值與平均數離中差三次方之平均數，符號或正或負，仍依其原來狀態），且立方差對偏態（skewness）之感應，甚爲靈敏。在度量差量時，自然而慣常之方法，須用名數表示之，如吋若干，磅幾何之類，標準差，均方差，及機誤之表示亦然，惟偏斜度之測量，苦無明確名數單位，故不得不使測量，脫離所用單位；而此方法，卽以離中差作爲標準差之倍數；例如，以x代表某種測量，$\bar{x}$代表平均數，σ爲標準差，則平均後之離中差，爲$\left(\frac{x-\bar{x}}{\sigma}\right)^3$，而偏斜度，乃可由$\frac{1}{n}\left\{所有\left(\frac{x-\bar{x}}{\sigma}\right)^3各值之和\right\}$而得。由此測量之結果，感應最爲敏銳，只惜缺乏明鮮單位耳。至其意義，是否可以明瞭，須視對曲線形狀，及實際測得之偏斜度，有無經驗而定也。

此外倘有用平均數之四次，五次乃至更高次方，以資測量者。皮爾生（Karl Pearson）教授，於其動差（moment）論，曾推論之。第一，第二，第三乃至更高級之動差，卽爲離中差之一次，

二次,三次乃至更高次方之平均。離中差可隨任何一點起算，而測得之動差,即爲對該點而言,惟一般應用，均以平均數爲測量之中心,而由其他各點得來之動差,不過圖演算之利便計耳。

第一編第五章,曾論平均數,謂其爲敍述一羣類之捷徑，在比較兩類材料時,此言尤信。然現時此概念應行擴大，現須另尋敍述主要表徵數之有系統的方法,而此方法，即用三數種符號,以測量平均數,標準差,偏斜度及其他類似數量也。若此等測量之意義及尺度,一經辨識清楚,原始材料，便無效用(除非留作參考或製圖);羣類之表示,完全用扼要方法;計算各羣類間之相互關係,即以此種數量爲根據,並特作數理研究之基礎。

第三節　動差所用之標號

玆將動差所用之標號及術語,述之如下:——

$$m'_t=\frac{1}{n}(x_1{}^ty_1+x_2{}^ty_2+\cdots\cdots)=\mathrm{S}(x^ty)\div n\cdots\cdots\cdots\cdots(1)$$

此之謂一羣類對其原點而言之第 t 級動差。

$$n=\mathrm{S}y\cdots\cdots\cdots\cdots(2)$$

$$m_1'=\bar{x}=\mathrm{S}xy\div\mathrm{S}y\cdots\cdots\cdots\cdots(3)$$

乃爲該羣類之平均數。

$$m_t=\mathrm{S}\ (x-\bar{x})\ ^ty\div n\cdots\cdots\cdots\cdots(4)$$

爲對平均數而言之第 t 級動差。

於是

$$nm_2=\mathrm{S}(x-\bar{x})^2y=\mathrm{S}x^2y-2\bar{x}\mathrm{S}xy+\bar{x}^2\mathrm{S}y=nm_2'-2\bar{x}\cdot n\bar{x}+n\bar{x}^2$$

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

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

此之謂標準差，定義如上所述。

$$nm_3=\mathrm{S}x^3y-3\bar{x}\mathrm{S}x^2y+3\bar{x}^2\mathrm{S}xy-n\bar{x}^3$$

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

m_3，在對稱曲線，其值爲零。求算偏斜度，簡便之測量法，乃以橫坐標定爲標準差之倍數，如此則測量所用具體單位之煩，可以免去也。

例如

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

卽偏斜度之測量法。

同理

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

爲第四級動差，而

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

爲不用單位之測量法。

標準差旣然求得，κ_2 之大小要看該類各份子自中心向外離散之程度而定，離散愈大，k_2 值亦愈大。在差誤常態曲線之特種

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

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

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

$\frac{m_4}{c^4}-\frac{3}{4}=\frac{1}{4}\left(\frac{m_4}{\sigma^4}-3\right)=\frac{1}{4}(k_2-3)$。於是在差誤常態曲線中，$i$之值必爲零。

第四節　動差算法舉例

$\bar{x}$, σ, k, k^2 四種基本測量之算法，舉例如下。

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

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

例)散佈若屬對稱,其體重理宜表示正偏斜度,實際上亦確爲$k=\cdot 643$。惟內中兒童一人,體格迥異常人(身長五英尺四吋,體重十四碼)(註三),在計算動差時,曾將其剔除。因$k-3$只餘·457,可知該曲線離常態不遠。

紐約市僱傭特許證上十四至十五歲兒童之體重

重量(磅)	尺度 x	人數 y	乘積 xy	x^2y	x^3y	x^4y
65–	−7	3	−21	147	−1,029	7,203
70–	−6	9	−54	324	−1,944	11,664
75–	−5	142	−710	3,550	−17,750	88,750
80–	−4	301	−1,204	4,816	−19,264	77,056
85–	−3	289	−867	2,601	−7,803	23,409
90–	−2	380	−760	1,520	−3,040	6,080
95–	−1	416	−416	416	−416	416
100–	0	404	—	0	—	0
105–	1	315	+315	315	+315	315
110–	2	320	+640	1,280	+2,560	5,120
115–	3	262	+786	2,358	+7,074	21,222
120–	4	221	+884	3,536	+14,144	56,576
125–	5	131	+655	3,275	+16,375	81,875
130–	6	76	+456	2,736	+16,416	98,496
135–	7	52	+364	2,548	+17,836	124,852
140–	8	20	+160	1,280	+10,240	81,920
145–	9	29	+261	2,349	+21,141	190,269
150–	10	14	+140	1,400	+14,000	140,000
155–	11	10	+110	1,210	+13,310	146,410
160–	12	2	+24	288	+3,456	41,472
165–	13	2	+26	338	+4,394	57,122
170–	14	5	+70	980	+13,720	192,080
175–	15	1	+15	225	+3,375	50,625
		3,404	+4,906	37,492	+158,356	1,602,932
			−4,032		−51,246	
			+874		+107,110	

(第一表)

以102.5爲原點，單位爲五磅。

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

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

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

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

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

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

$\kappa=\frac{m_3}{\sigma^3}=.643=\sqrt{\beta_1}$, $\kappa_2=\frac{m_4}{m_2^2}=3.457=\beta_2$

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

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

孫巴克氏一九一六，四十五種商品之指數

指數	x	x^2	x^3	指數	x	x^2	x^3
68	−68	4,624	−314,432	138	+2	4	8
71	−65	4,225	−274,625	148	+12	144	1,728
84	−52	2,704	−140,608	148	+12	144	1,728
86	−50	2,500	−125,000	153	+17	289	4,913
93	−43	1,849	−79,507	154	+18	324	5,832
96	−40	1,600	−64,000	154	+18	324	5,832
100	−36	1,296	−46,656	157	+21	441	9,261
100	−36	1,296	−46,656	159	+23	529	12,167
101	−35	1,225	−42,875	159	+23	529	12,167
104	−32	1,024	−32,768	160	+24	576	13,824
104	−32	1,024	−32,768	161	+25	625	15,625
107	−29	841	−24,389	163	+27	729	19,683
114	−22	484	−10,648	163	+27	729	19,683
114	−22	484	−10,648	166	+30	900	27,000
119	−17	289	−4,913	168	+32	1,024	32,768
121	−15	225	−3,375	169	+33	1,089	35,937
125	−11	121	−1,331	172	+36	1,296	46,656
128	−8	64	−512	173	+37	1,369	50,653
128	−8	64	−512	174	+38	1,444	54,872
131	−5	25	−125	183	+47	2,209	103,823
132	−4	16	−64	197	+61	3,721	226,981
135	−1	1	−1	202	+66	4,356	287,496
135	−1	1	−1				
				22	629	22,795	988,637
23	−632	25,982	−1,256,414	23	−632	25,982	−1,256,414
				45	−3	48,777	−267,777

（第二表）

原點在 136。

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

$m_2'=\frac{48777}{45}=1083.933$　$m_2=m_2'-\bar{x}^2=1083.929$，$\sigma=\sqrt{m_2}=32.9$

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

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

三、北極星赤經度之觀察(註五)。

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

$x=-.170$

$m_2=5.390-.029=5.361$。$\sigma=2.3$

$m_3=-5.234-3(-.170)\times5.390+2(-.170)^3=-2.49$　　$\kappa=-.2$

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

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

四、下舉一例，示明機率表(Table of Chances)之可以作頻數羣類看待。所選出者爲一不對稱之事實，卽從十二骰子中擲出六點之機率是。例如，三個六點之機率爲

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

擲出六點次數	十二個骰子中之機率	
x	y	
0	244,140,625÷6^{12}	$\bar{x}=2$
1	585,937,500 ,,	$m_2=1\frac{2}{3}$
2	644,531,250 ,,	
3	429,687,500 ,,	$m_3=1\frac{1}{9}$
4	193,359,375 ,,	
5	61,875,000 ,,	$m_4=8\frac{11}{18}$
6	14,437,500 ,,	
7	2,475,000 ,,	$\sigma=1.29$
8	309,375 ,,	
9	27,500 ,,	$\kappa=.516$
10	1,650 ,,	
11	60 ,,	$\kappa_2=3.1$
12	1 ,,	
	2,176,782,336 ,,	

五、於整數中，任意抽樣，其平均數，總在4·5左右。下表係從七位對數表之末一數字中，以二十五爲一類，抽出四百類之結果。此種分佈，不甚對稱，且$k_2>3$。

二十五個數字之和，已用五約過

與22.5之差	次數	
Over 9	1	
8 to 9	5	
7 ,, 8	9	
6 ,, 7	5	
5 ,, 6	12	
4 ,, 5	10	
3 ,, 4	15	以23爲原點，
2 ,, 3	36	$\bar{x}=-.2575$；平均數，22.7425
1 ,, 2	48	$m_2=8.8662$；修正後，8.783
0 ,, 1	57	$\sigma=2.964$
0 ,, −1	62	$m_3=13.584$；$\kappa=.522$
−1 ,, −2	58	$m_4=274.24$；修正後，269.8
−2 ,, −3	39	$\kappa_2=3.50$
−3 ,, −4	17	
−4 ,, −5	13	
−5 ,, −6	10	
−6 ,, −7	2	
−7 ,, −8	1	
	400	

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

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

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

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

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

$${}_0S_t=\text{觀察數}=n$$

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

式中 $\bar{x}$（爲平均數）$=nm_1'$。

$$\begin{aligned}{}_2S_t&=(1+2+\cdots\cdots+t)\,y_t+(1+2+\cdots\cdots+\overline{t-1})\,y_t-1+\cdots\cdots\\&+(1+2)\,y_2+y_1=\frac{t(t+1)}{2}y_t+\frac{(t-1)t}{2}y_{t-1}+\cdots\cdots+\frac{1\cdot2}{2}y_1\\&=\frac{n}{2}(m_2'+m_1'),\end{aligned}$$

式中 $m_1', m_2' \cdots\cdots$ 爲由原點計算之動差。

$$\begin{aligned}{}_3S_t&=\tfrac{1}{2}\{1\cdot2+2\cdot3+\cdots\cdots t\,(t+1)\}y_t+\tfrac{1}{2}\{1\cdot2+\cdots\cdots(t-1)\,t\}y_t\\&-1+\cdots\cdots=\tfrac{1}{6}\{t(t+1)\,(t+2)\,y_t+(t-1)\,t\,(t+1)\,y_{t-1}+\cdots\\&\cdots+1\cdot2\cdot3y_1\}\\&=\frac{n}{6}(m_4'+6m_3'+11m_2'+6m_1')。\end{aligned}$$

則依公式 5, 7, 9，得

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

$$m_3=\frac{6}{n}{}_3S_t-3m_2(1+\bar{x})-\bar{x}(1+\bar{x})(2+\bar{x})$$

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

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

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

數字之和÷5	x	vx'	${}_0Sx'$	${}_1Sx'$	${}_2Sx'$	${}_3Sx'$
31.5以上	18	1	1	1	1	1
30.5	17	5	6	7	8	9
29.5	16	9	15	22	30	39
28.5	15	5	20	42	72	111
27.5	14	12	32	74	146	257
26.5	13	10	42	116	262	519
25.5	12	15	57	173	435	954
24.5	11	36	93	266	701	1,655
23.5	10	48	141	407	1,108	2,763
22.5	9	57	198	605	1,713	4,476
21.5	8	62	260	865	2,578	7,054
20.5	7	58	318	1,183	3,761	10,815
19.5	6	39	357	1,540	5,301	16,116
18.5	5	17	374	1,914	7,215	23,331
17.5	4	13	387	2,301	9,516	32,847
16.5	3	10	397	2,698	12,214	45,061
15.5	2	2	399	3,097	15,311	60,372
14.5	1	1	400	3,497	18,808	79,180
總　計		400 $={}_0S_{18}$	3,497 $={}_1S_{18}$	18,808 $={}_2S_{18}$	79,180 $={}_3S_{18}$	285,560 $={}_4S_{18}$

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

$$m_2=\frac{2}{400}\times18808-8.7425\times9.7425=8.8662$$

$$m_3=\frac{6}{400}\times76130-3\times8.8662\times9.7425-8.7425\times9.7425\times10.7425=13.584$$

$$m_4=\frac{24}{400}\times285560-2\times13.584\times20.485-8.8662\times626.95-10744.15=274.24$$

（註一） 爲初步研究用，可參閱 Whitworth 著：Choice and Chance。

（註二） 本書爲便利計不用前此求偏斜度之標號，而用K。

（註三） 譯者註：啞(Stones)，重量相當十四磅。

（註四） 薛伯氏校正(Sheppard's Correction)參閱附錄五。上表中，其他計算亦同，各級人數（即各組次數），可視作恰當各該組之中心，此種解釋，設非分組得當，第二，四動差，顯有誇大之虞，惟如兩極端各級之人數極少，則第一，三動差，所受影響尚微。以 h 代各級之寬度，且不爲一，則經修正後之動差爲$m_2-\frac{1}{12}h^2$，及$m_4-\frac{1}{2}h^2m_2+\frac{7}{240}h^4$。

（註五） 見Quetelet著：機率論(Lettres sur la théorie des probabilités)第一百二十八頁。

（註六） 『頻數曲線與相關』第十九至二十三頁。在二十三頁，艾得頓氏示吾人近於中心取用原點之法，以省數字計算之煩。可參考哈地氏著：死亡率表編製原理(Hardy, The Theory of the Construction of Tables of Mortality)第五十九頁起。

第二章 代數機率與差誤常態曲線

第一節 初步原理

代數機率之方法與基本定理,可概述如下:

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

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

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

為失敗之機率；p 乃對於 q 而成功，q 對於 p 而失敗。

第二節 機率乘法

假設以 p_1 p_2 為兩個各自獨立實驗中之成功機率，則$p_1\times p_2$，可如下文所示，為二者均成功之機率。

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

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

在此 $n_1\times n_2$ 個事件中，有 $m_1\times m_2$ 個，為二者均成功。

$m_1\times(n_2-m_2)$……前者成功後者失敗

$(n_1-m_1)\times m_2$……前者失敗而後者成功

$(n_1-m_1)\times(n_2-m_2)$……二者均失敗

n_1n_2 個機會均等之事件內，只有 m_1m_2 個係二者均成功，餘則不能。故二者均成功之機率，p

$$=\frac{m_1m_2}{n_1n_2}=p_1\times p_2$$

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

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

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

但如在一副牌內已抽出一張後，再抽第二張，則有下列幾種可能：

（1）可供抽出之事件共有51×52。

（2）如先抽出之牌，為一A點，則在下餘51張中，尚有三個A點。

（3）二者均成功之途徑，有4×3個。

（4）前者成功，後者失敗之途徑，有4×48個。

（5）前者失敗，後者成功之途徑，有48×4個。

（6）二者均失敗之途徑，為48×47個。

（7）所以二者均成功之機率$\frac{4}{52}\times\frac{3}{51}=\frac{1}{221}$。

此問題寫作下式亦可：

（1）每副紙牌中，共可抽出${}_{52}C_2=\dfrac{52\cdot51}{1\cdot2}$對。

（2）在此${}_{52}C_2=\dfrac{52\cdot51}{1\cdot2}$對中，抽得兩個么點，只佔${}_4C_2=\dfrac{4\cdot3}{1\cdot2}$對。任何一對可被抽出之機會，彼此完全相同。所以，在抽兩個么點時，不論一齊抽出或挨次抽出，其抽得之機率，為$\dfrac{{}_4C_2}{{}_{52}C_2}=\dfrac{4\cdot3}{52\cdot51}$。

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

$$\frac{{}_{13}C_8 \times {}_{39}C_5}{{}_{52}C_{13}} = \frac{13\cdot12\cdot11\cdot10\cdot9\cdot8\cdot7\cdot6\cdot39\cdot38\cdot37\cdot36\cdot35\cdot(13!)}{52\cdot51\cdot50\cdot49\cdot48\cdot47\cdot46\cdot45\cdot44\cdot43\cdot42\cdot41\cdot40\cdot(8!)(5!)}$$

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

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

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

第三節 機率加法

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

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

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

$$p = p_1p_1' + p_2p_2' + \cdots\cdots$$

第四節 差誤常態律之演繹

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

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

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

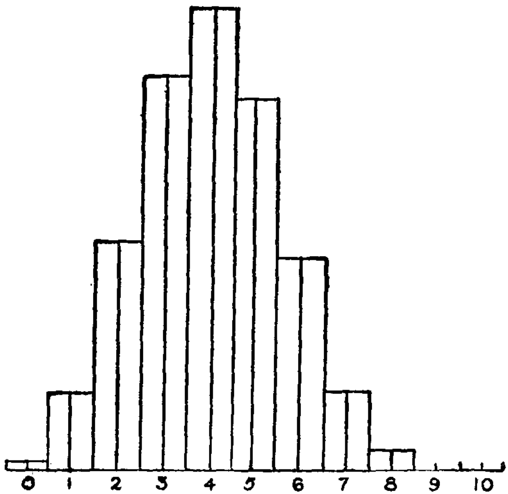
$p\times p\times\cdots\cdots$到 r 個因子 $\times q\times q\times\cdots\cdots$到 $\overline{n-r}$ 個因子 $=p^r\times q^{n-r}$；且不拘其他何種順序，其機率終屬相同。在一組之 n 次實驗中，選定任一 r 爲成功次數，即可得該種順序，即成功有 $n\mathrm{C}r$ 種方法是。是以，其全體機率乃係 $n\mathrm{C}_r.p^r=q^{n-r}$。

成功 $0,1,2,\cdots\cdots n$ 次之機率乃自然形成二項式之連接項：

$$1=(q+p)^n=q^n+n.q^{n-1}p+\cdots+{}_n\mathrm{C}r\cdot q^{n-r}p^r+\cdots+nqp^{n-1}+p^n$$

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

r	${}_n\mathrm{C}_r$	p^rq^{n-r}		${}_n\mathrm{C}_r\cdot p^rq^{n-r}$
0	1	$3^{10}\div5^{10}$		·006,046,617,6
1	10	2×3^9	,,	·040,310,784,0
2	45	$2^2\times3^8$	,,	·120,932,352,0
3	120	$2^3\times3^7$	,,	·214,990,848,0
4	210	$2^4\times3^6$	,,	·250,822,656,0
5	252	$2^5\times3^5$	,,	·200,658,124,8
6	210	$2^6\times3^4$	,,	·111,476,736,0
7	120	$2^7\times3^3$	,,	·042,467,328,0
8	45	$2^8\times3^2$	,,	·010,616,832,0
9	10	$2^9\times3^1$	,,	·001,572,864,0
10	1	2^{10}	,,	·000,104,857,6
				1·000,000,000,0



此縱標尺擴大百倍，故此圖面積爲一百個單位的方格。

上圖例釋成功次數多寡不同之相對機率，並以之作成一『頻數羣』。

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

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

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

$\bar{x}=m_1'$，由原點之第一動差

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

$$=n\cdot q^{n-1}p+n(n-1)/2\cdot q^{n-2}p^2\times 2+\cdots\cdots+{}_nC_r q^{n-r}p^r$$
$$\times r+\cdots\cdots+p^n\times n$$
$$=np(q+p)^{n-1}=np\cdots\cdots(11)$$
$$m_2'=(y_0\times 0^2+y_1\times 1^2+\cdots\cdots+y_r\times r^2+\cdots\cdots+y_n\times n^2)\div \mathrm{N}$$
$$=\Sigma\,{}_0 r^2\cdot {}_nC_r\cdot q^{n-r}p^r=\Sigma\,\{r(r-1)+r\}\frac{(n)_r}{r!}q^{n-r}p^r$$
$$=n(n-1)p^2\Sigma\,\frac{(n-2)_{r-2}}{(r-2)!}q^{n-r}p^{r-2}+np\Sigma\,\frac{(n-1)_{r-1}}{(r-1)!}$$
$$q^{n-r}p^{r-1}$$
$$=n(n-1)p^2(q+p)^{n-2}+np(q+p)^{n-1}=n(n-1)p^2+np$$
$$=n^2p^2+np\,(1-p)=\bar{x}^2+npq\cdots\cdots(12)$$

而以平均數爲原點之第二動差，m_2

$$=m_2'-\bar{x}^2=npq=np(1-p)\cdots\cdots(13)$$

同樣

$$m_3'=\Sigma_0^n r^3\cdot {}_nC_r\cdot q^{n-r}p^r=n(n-1)(n-2)p^3+3n(n-1)p^2+np\cdots(14)$$

至以平均爲原點之第三動差，m_3

$$=m_3'-3m_2'\bar{x}+2\bar{x}^3$$
$$=n(n-1)(n-2)\,p^3+3n\,(n-1)\,p^2+np-3n^3p^3-3n^2p^2$$
$$(1-p)+2n^3p^3$$
$$=np(2p^2-3p+1)=np(1-p)(1+2p)=npq(q-p)\cdots\cdots(15)$$

$m_4'=\Sigma_0^n r^4\cdot {}_nC_r\cdot q^{n-r}p^r$，而$m_4$亦可證明等於

$$3(pqn)^2+pqn(1-6pq)。$$

故依第一章第三節之公式，$\sigma=\sqrt{pqn}$，$\beta_1=\dfrac{(q-p)^2}{pqn}$，$k_2=\beta_2=3+\dfrac{1-6pq}{pqn}$，$C=\sqrt{2pqn}$，$k=\dfrac{q-p}{\sqrt{pqn}}$，$i=\dfrac{1-6pq}{4pqn}$。

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

復次，請就成功r次之機率，及n增加時之圖形討論。

第一例：當$p=q=\frac{1}{2}$，n為偶數$=2n'$時。

設p_x為$n'+x$次成功之機率，亦卽$(n'-x)$次失敗之機率。

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

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

$P_0=\dfrac{(2n')!}{2^{2n'}\cdot n'!n'!}=\dfrac{1}{\sqrt{\pi n'}}$，根據瓦立斯氏(Wallis)定理，核正至$\dfrac{1}{n'}$止(附錄一，公式132)。

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

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

$$=-2\left(\frac{1}{n'}+\frac{1}{3n'^3}+\cdots\right)-2\left(\frac{2}{n'}+\frac{2^3}{3n'^3}+\cdots\right)\cdots\cdots$$

$$-2\left(\frac{x}{n'}+\frac{x^3}{3n'^3}+\cdots\right)-\log\left(1-\frac{x}{n'}\right)$$

$$=-2\frac{1+2+\cdots+x}{n'}-\frac{2}{3}\cdot\frac{1^3+2^3+\cdots+x^3}{n'^3}\cdots\frac{2}{2t+1}$$

$$\cdot\frac{1^{2t+1}+2^{2t+1}+\cdots+x^{2t+1}}{n'^{2t+1}}\cdots-\log\left(1-\frac{x}{n'}\right)$$

t 爲任一整數

$$=-\frac{x(x+1)}{n'}-\frac{2}{3}\frac{{}_2x^2(x+1)^2}{4n'^3}-\cdots$$

$$-\frac{2}{2t+1}\cdot\frac{x^{2t+2}+\cdots+\cdots}{(2t+2)n'^{2t+1}}\text{（註一）}-\cdots+\left(\frac{x}{n'}+\frac{x^2}{2n'^2}+\cdots\right)$$

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

$c^2=2pqn=2\cdot\frac{1}{2}\cdot\frac{1}{2}\cdot2n'=n'$。

$$\log(P_x.c\sqrt{\pi})=-\tau^2-\frac{\tau}{\sqrt{n'}}-\frac{2}{3n'}\frac{\tau^2}{4}\left(\tau+\frac{1}{\sqrt{n'}}\right)^2-\cdots$$

$$-\frac{2}{(2t+1)(2t+2)}\left(\frac{\tau^{2t+2}}{n'_t}+\cdots\right)-\cdots+\left(\frac{\tau}{\sqrt{n'}}+\frac{\tau^2}{2n'}+\cdots\right)$$

$$=-\tau^2+\text{包含}\frac{1}{\sqrt{n'}}\text{各項}$$

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

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

因爲標準差，$\sigma=c/\sqrt{2}$ 。

又因 $c^2=\frac{n}{2}$，$\therefore \mathrm{P}_x=\frac{\sqrt{2}}{\sqrt{\pi n}}e^{-\frac{2x^2}{n}}$

第二例：在 p 與 q 不相等時。

設P_x爲$pn+x$個成功之機率（註二），亦卽$qn-x$失敗之機率。

$$\mathrm{P}_x=\frac{n!}{(pn+x)!(qn-x)!}\cdot p^{qn+x}q^{qn-x}$$

$$=\frac{1}{(pn)!(qn)!}p^{pn}q^{qn}\cdot\frac{qn(qn-1)\cdots(qn-x+1)}{(pn+1)(pn+2)\cdots(pn+x)}\cdot\frac{p^x}{q^x}$$

$$=\frac{\left(1-\frac{1}{qn}\right)\left(1-\frac{2}{qn}\right)\cdots\left(1-\frac{x-1}{qn}\right)}{\left(1+\frac{1}{pn}\right)\left(1+\frac{2}{pn}\right)\cdots\left(1+\frac{x}{pn}\right)}$$

$$\log(\mathrm{P}_x/\mathrm{P}_0)=\sum_{s=1}^{s=x}\log\left(1-\frac{s}{qn}\right)-\sum_{s=1}^{s=x}\log\left(1+\frac{s}{pn}\right)-\log\left(1-\frac{x}{qn}\right)$$

$$=-\sum_1^x\left(\frac{s}{qn}+\frac{s}{pn}\right)-\sum_1^x\frac{1}{2}\left(\frac{s^2}{q^2n^2}-\frac{s^2}{p^2n^2}\right)-\cdots\cdots$$

$$-\sum_1^x\frac{1}{t}\left(\frac{s^t}{q^tn^t}\pm\frac{s^t}{p^tn^t}\right)-\cdots-\log\left(1-\frac{x}{qn}\right)$$

$$=-\frac{x(x+1)}{2}\cdot\frac{p+q}{pqn}-\frac{x(x+1)(2x+1)}{6}\cdot\frac{p^2-q^2}{2p^2q^2n^2}$$

$$-\frac{x^2(x+1)^2}{4}\cdot\frac{p^3+q^3}{3p^3q^3n^3}-\cdots$$

$$-\frac{1}{t}\cdot\frac{x^{t+1}+\cdots}{t+1}\cdot\frac{p^t\pm q^t}{p^tq^tn^t}-\cdots+\left(\frac{x}{qn}+\frac{x^2}{2q^2n^2}+\cdots\right)$$

以 $x=\tau c$，而 $c^2=2pqn=2\sigma^2$

$$\log(\mathrm{P}_x/\mathrm{P}_0)=-\frac{\tau^2c^2+\tau c}{c^2}+\frac{2\tau^3c^3+3\tau^2c^2+\tau c}{3c^4}(q-p)$$

$$-\frac{\tau^4c^4+2\tau^3c^3+\tau^2c^2}{\frac{3}{2}\cdot c^6}\cdot(1-3pq)-\cdots$$

$$-\frac{\tau^{t+1}c^{t+1}+\cdots}{t(t+1)2^{-t}c^{2t}}(p^t\pm q^t)-\cdots+\frac{2\tau cp}{c^2}+\frac{2\tau^2c^2p^2}{c^4}+\cdots,$$

因 $\mathrm{P}+q=1$,

$$=-\tau^2+\frac{\tau}{c}\left\{-1+\frac{2\tau^2}{3}(q+p)+2p\right\}$$

$$+\frac{\tau^2}{c^2}\left\{(q-p)-\frac{2\tau^2}{3}(q-p)+2p\right\}$$

$+$包含$\frac{1}{c^3}$之項

以 τ 爲有限數論，換言之，在 x 之衆數値中，只就其可與 $\sqrt{pqn}$ 比較者討論之。

如删略$\frac{1}{c}$（卽删略$\frac{1}{\sqrt{n}}$），則得

$$\mathrm{P}_x=\mathrm{P}_0e^{-\tau^2}=\mathrm{P}_0e^{-\frac{x^2}{c^2}}\cdots\cdots\cdots(17)$$

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

$$\mathrm{P}_x=\mathrm{P}_0e^{-\tau^2}\cdot e^{-\frac{\tau(q-p)(1-\frac{2}{3}\tau^2)}{c}}$$

$$=\mathrm{P}_0e^{-\tau^2}\left\{1-\frac{q-p}{c}\left(\tau-\tfrac{2}{3}\tau^3\right)\right\}$$，因$\frac{1}{c^2}$已略去

$$=\mathrm{P}_0 e^{-\frac{x^2}{c^2}}\left\{1-\frac{q-p}{c}\left(\frac{x}{c}-\frac{2}{3}\cdot\frac{x^3}{c^3}\right)\right\}$$

$$=\mathrm{P}_0 e^{-\frac{x^2}{2\sigma^2}}\left\{1-\frac{k}{2}\left(\frac{x}{\sigma}-\frac{x^3}{3\sigma^3}\right)\right\},\cdots\cdots\cdots\cdots(18)$$

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

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

$$\mathrm{P}_0=\frac{n!}{(pn)!(pn)!}p^{pn}q^{qn}$$

$$=\frac{n^n}{(pn)^{pn}(qn)^{qn}}\sqrt{\frac{2\pi n}{2\pi pn\cdot 2\pi qn}}\cdot e^{-n+pn+qn}\cdot p^{pn}q^{qn}$$

略去 $\frac{1}{pn},\frac{1}{n}\cdots\cdots$，

$$=\frac{1}{\sqrt{2\pi pqn}}\text{（因}p+q=1\text{）},=\frac{1}{c\sqrt{\pi}}=\frac{1}{\sigma\sqrt{2\pi}}。$$

以 y 代 P_x，則得方程式

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

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

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

如上所述，當舉行包有 n 種事件之實驗時，成功之次數總比

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

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

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

n若大,任一成功次數之實際機率必小,舉例言之,假如$P=\frac{1}{2}$ $n=1000$,成功恰當 500 次（最大概然數）時之機率,僅約達 $\frac{1}{40}$ 而已。吾人所需用者並非特定縱坐標之量數,而在各數值全距之機率總和數量;所謂各數值之全距者,例如由 x_1 至 x_2, 即是,此處之 x_2-x_1, 與 σ $(=\sqrt{pqn})$,爲同一方次。

藉一極著盛名之定理（註四）吾人可以從縱坐標之總和，而達面積之積分,成功之次數,大小與 $pn+x_1$同,而未必比$Pn+x_2$大，則其全部機率必爲$\int_{x1}^{x2}ydx$, 此處之 $y=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{x^2}{2\sigma^2}}$，並已將含有$\frac{1}{\sqrt{n}}$各項略去。

以 z 代 $\frac{x}{\sigma}$ 則得$\int_{x1}^{x2}ydx=\frac{1}{\sqrt{2\pi}}\int_{z1}^{z2}e^{-\frac{1}{2}z^2}dz$, 此函數數值表，容當附及（即第三表）

上文所提及之函數,其重要常數,得出之法如下:

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

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

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

以$m_s=\int_{-\infty}^{\infty}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{x^2}{2\sigma^2}}\cdot x^s dx$，爲對平均數之第S次動差，此乃以平均數爲原點，面積爲1。

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

$$m_2=\frac{1}{2\sqrt{2\pi}}\int_{-\infty}^{\infty}x^2e^{-\frac{x^2}{2\sigma^2}}dx$$

$$=\left[-\frac{\sigma}{\sqrt{2\pi}}xe^{-\frac{x^2}{2\sigma^2}}\right]_{-\infty}^{\infty}+\frac{\sigma}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{x^2}{2\sigma^2}}dx$$

$$=0+\sigma^2=\sigma^2 \quad\cdots\cdots\cdots\cdots (21)$$

此在公式13已論及矣。

$$m_{2t}=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}x^{2t}e^{-\frac{x^2}{2\sigma}}dx$$

$$=\left[-\frac{\sigma}{\sqrt{2\pi}}x^{2t-1}e^{-\frac{x^2}{2\sigma^2}}\right]_{-\infty}^{\infty}+\frac{(2t-1)\sigma^2}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}x^{2t-2}e^{-\frac{x^2}{2\sigma^2}}dx$$

$$=0+(2t-1)\,\sigma^2m2t-2 \quad\cdots\cdots\cdots\cdots (22)$$

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

$$k_2=\beta_2=\frac{m^4}{m_2{}^2}=3, i=\beta,_2-3=0$$

此數依公式 16 亦可求得，當 n 爲無限時。

$$m_{2t}=(2t-1)(2t-3)\cdots\cdots3.1\sigma^{2t}, \text{用歸納法},$$

$$=\frac{(2t)!}{2^t!t!}\sigma^{2t} \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (23)$$

例如 $m_6=15\sigma^6$, $m_8=105\sigma^8$。

平均差，n，因面積爲一，故

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

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

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

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

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

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

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

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

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

曲線自底線。至 σ 之一段，當等於表上數值之

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

照樣查表，則可立得下列各近似值：

各段底線之曲線面積比例

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

（註六）

所謂機誤爲自中心位置向左向右佔有全觀察數量，恰當一半之距離。

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

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

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

第五節　代數機率與經驗

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

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

粗率之實驗，用極簡單之用具，卽可舉行。例如，取樸克牌若

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

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

$\frac{1}{10^4}(x+x^2+\cdots\cdots x^{10})^4$ 即 $\frac{x^4}{10^4}\cdot\left(\frac{1-x^0}{1-x}\right)^4$ 一式內 x^r 之係數,此機率與逐次實驗之結果,可表列如下:——

r	機率合計	×90=希望數	實驗得數
4 至 9	.0126	1.134	0
10 ,, 14	·0871	7.839	7
15 ,, 19	.2375	21.375	27
20 ,, 24	.3256	29.304	25
25 ,, 29	.2375	21.375	22
30 ,, 34	.0871	7.839	9
35 ,, 40	.0126	1.134	0
	1.0000	90.000	90

九十組(每四張爲一組)所有各點總數爲 1956,平均數每張合 5.43。各副所有各牌之平均數爲 5.5。

試驗結果與希望數相合,至少亦相逼近。

第六節　白諾立氏定律

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

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

之機率爲$(q+p)^{90}$之第$\overline{i+1}$項。實現機會最多之數，非爲21卽爲22，而其各種可能的成功數之標準差爲

$\sqrt{pqn}$，其中$n=90$，故在4左右。

在此重疊舉行多次之繁複實驗中，從17至26一類內，任何成功數之機率，用前表查出，大約等於$\frac{2}{3}$；求得一大與27相等之數（如上試驗表中所列），其機率約爲$\frac{1}{6}$。與21有三個標準差之分歧，未必卽有此事；換言之，實現次數多過33，少於9，殊屬罕見之事也。

此種方法，概而言之，已引到白諾立氏定律（Bernoulli's law），玆將此定律解釋如下。假如成功機率爲p之一實驗，舉行n次，以$p'n$代表成功次數，則若n增加，p'將漸近於p。比$p\sim p'$較大之離中差之實現機率，爲$2\int_z^\infty \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}dz$，

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

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

基於一般之經驗及多次實驗之結果，白氏定律確合於事實。

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

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

第七節 例證

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

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

每次抽取五十字之試驗，舉行三百次

實在頻數與理論頻數，二者之相合程度，果如理論所示（參閱第十章）。標準差理論上應爲$\sqrt{pqn}=3.535$。援用第一章第四節所得第二級動差之平方根，亦可求得觀察之實在標準差。其平均數爲25.043。以25爲原點而得之第二動差，爲$(1\times 11^2+0\times 10^2+3\times 9^2+\cdots\cdots 1\times 10^2)\div 300=11.30$，而以平均數爲原點之第二動差，則爲$11.300-.043^2=11.298$。其平方根爲3.361，與

r	z	$F(z)$	差額（註八）	$\times 300=$	實現次數（註九）期望數	實在數		
13.5	−3.2522	.4994	.0008	.2	0	1	在	14
14.5	−2.9694	.4986	.0020	.6	0或1	0	,,	15
15.5	−2.6866	.4966	.0047	1.4	1或2	3	,,	16
16.5	−2.4038	.4919	.0088	2.6	2或3	2	,,	17
17.5	−2.1210	.4831	.0161	4.8	5	3	,,	18
18.5	−1.8382	.4670	.0270	8.1	8	7	,,	19
19.5	−1.5554	.4400	.0416	12.5	12或13	9	,,	20
20.5	−1.2726	.3984	.0595	17.85	18	18	,,	21
21.5	− .9898	.3389	.0787	23.6	24	26	,,	22
22.5	− .7070	.2602	.0959	28.8	29	21	,,	23
23.5	− .4242	.1643	.1082	32.5	32或33	32	,,	24
24.5	− .1414	.0561	.1122	33.7	39	42	,,	25
25.5	+ .1414	.0561	.1082	32.5	32或33	36	,,	26
26.5	+ .4242	.1643	.0959	28.8	29	30	,,	27
27.5	+ .7070	.2602	.0787	23.6	24	28	,,	28
28.5	+ .9898	.3389	.0595	17.85	18	15	,,	29
29.5	1.2726	.3984	.0416	12.5	12或13	16	,,	30
30.5	1.5554	.4400	.0270	8.1	8	5	,,	31
31.5	1.8382	.4670	.0161	4.8	5	2	,,	32
32.5	2.1210	.4831	.0088	2.6	2或3	2	,,	33
33.5	2.4038	.4919	.0047	1.4	1或2	1	,,	34
34.5	2.6866	.4966	.0020	.6	0或0	1	,,	35
35.5	2.9694	.4986	.0008	.2	0	0	,,	36
36.5	3.2522	.4994						
				299.6				

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

II. 如不求各個數值之希望數，而照下例所示之法，仍可測驗其分佈情況。

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

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

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

理論之標準差爲 $\sqrt{pqn}=2.913=\sigma$。

其實現次數如下：

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

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

		差額×100	實現數	
$\bar{x}-3\sigma=4.43$	$F(-3)=.499$	2.3	$7\frac{1}{2}$以下	1
$\bar{x}-2\sigma=7.34$	$F(-2)=.477$	13.6	$7\frac{1}{2}$ 至 $10\frac{1}{2}$	17
$\bar{x}-\sigma=10.26$	$F(-1)=.341$	34.1	$10\frac{1}{2}$,, $13\frac{1}{2}$	40
$\bar{x}=13.17$	$F(0)=.0$	34.1	$13\frac{1}{2}$,, $16\frac{1}{2}$	30
$\bar{x}+\sigma=16.08$	$F(1)=.341$	13.6	$16\frac{1}{2}$,, $19\frac{1}{2}$	9
$\bar{x}+2\sigma=19.00$	$F(2)=.477$	2.3	$19\frac{1}{2}$,, $22\frac{1}{2}$	8
$\bar{x}+3\sigma=21\ 91$	$F(3)=.499$	100.0		100

觀察之量數，非爲整數不可者，不易使各級成爲σ之倍數。但

若觀察之分級狹小，或其量數係繼續不斷者，則此法——以標準差，σ之相等的次倍數(submultiple)爲準——甚爲簡捷，且在測驗未應用之前，分級卽可決定，故可供爲一種良好而簡單測驗之用。

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

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

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

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

第八節 對於抽樣法之應用

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

第九節　抽樣方法舉例

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

各區人數

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

由此觀之——以上表第一行爲例——人口在一百人以下者，二百五十區中佔三十五區

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

據此推測，每千區之數，必爲 1000 之.14＝140。

p' 之標準差爲$\sqrt{\frac{p'(1-p')}{250}}=.022$（見上節），然則$1000p'$，即推測值，140，之標準差必爲22。實際上英格蘭及威爾士各區人口在一百人以下者，爲千分之一百五十二。推測與實數相差約半個標準差。（見統計學報 Statistical Journal 一九一二至一九一三年號第一八二頁）

（2）自三千八百七十八家公司股息率單中，擇出四百家，列表於下：

股息百分率

	三鎊以下	三鎊	四鎊	五鎊	六鎊	八鎊
樣本所取公司數…………	34	180	117	60	48	33
$1000p'$ …………………	85	270	$292\frac{1}{2}$	150	120	$82\frac{1}{2}$
$1000\sqrt{\left(\frac{p'q'}{400}\right)}$ …………	14	22	23	18	16	14
單中全數每1000中公司數	75	272	311	177	180	57

（見統計學報 Statistical Journal 第一九〇六年號第五五二頁）

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

南北緯度

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

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

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

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

第十節 範圍實非爲無限或選擇未能獨立之例

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

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

種種不同之抽取方法，共數有${}_N C_n$種。

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

$$\text{故}P_x=\frac{{}_{pn}C_{pn+x}\times{}_{qn}C_{qn-x}}{{}_nC_n}$$

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

此處之 $M=N-n$。

援照斯德令氏（Sterling）定理，應用到階乘（Factorial）上面，將$\frac{1}{pn}$，$\frac{1}{n}$及其他較小之數量省略。（見附錄三公式134）

$$P_0=(pN)^{pN}(qN)^{qN}(n)^nM^M(p_n)^{-pn}(pM)^{-pM}(qn)^{-qn}(qM)^{-qM}N^{-N}$$

$$.(2\pi)^{\frac{4}{2}-\frac{5}{2}},\ e^0.\left(\frac{pNqNnM}{pnpMqnqMN}\right)^{\frac{1}{2}},$$

e 之指數為

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

$=0$，因$p+q=1$。

當將此諸指數集攏於一起時，則得

$$P_0=\left(\frac{N}{2\pi pqnM}\right)^{\frac{1}{2}}\text{……………………}(27)$$

$$\frac{P_x}{P_0}=\frac{(pn)!(pM)!(qn)!(qM)!}{(pn+x)!(pM-x)!(qn-x)!(qM+x)!}$$

$$=\frac{(pn)^{pn}(pM)^{pM}(qn)^{qn}(qM)^{qM}.(2\pi)^0.e^0.(pn.pM.qn.qM)^{\frac{1}{2}}}{(pn+x)^{pn+x+\frac{1}{2}}(pM-x)^{pM-x+\frac{1}{2}}(qn-x)^{qn-x+\frac{1}{2}}(qM+x)^{qM+x+\frac{1}{2}}}$$

$$\therefore\frac{P_0}{P_x}=\left(1+\frac{x}{pn}\right)^{pn+x+\frac{1}{2}}.\left(1-\frac{x}{pM}\right)^{pM-x+\frac{1}{2}}.\left(1-\frac{x}{qn}\right)^{qn-x+\frac{1}{2}}$$

$$.\left(1+\frac{x}{qM}\right)^{qM+x+\frac{1}{2}}$$

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

$$-\left(pM-x+\tfrac{1}{2}\right)\log\left(1-\frac{x}{pM}\right)-\left(qM+x+\tfrac{1}{2}\right)\log\left(1+\frac{x}{qM}\right)$$

$$=-\left(pn+x+\tfrac{1}{2}\right)\left(\frac{x}{pn}-\frac{x^2}{2p^2n^2}+\cdots\right)$$

$$+\left(qn-x+\tfrac{1}{2}\right)\left(\frac{x}{qn}+\frac{x^2}{2q^2n^2}+\cdots\right)$$

$$+\left(pM-x+\tfrac{1}{2}\right)\left(\frac{x}{pM}+\frac{x^2}{2p^2M^2}+\cdots\right)$$

$$-\left(qM+x+\tfrac{1}{2}\right)\left(\frac{x}{qM}+\frac{x^2}{2q^2M^2}+\cdots\right)$$

$$=\frac{x}{2}\left(-\frac{1}{pn}+\frac{1}{qn}+\frac{1}{pM}+\frac{1}{qM}\right)-\frac{x^2}{2}\left(\frac{1}{pn}+\frac{1}{qn}+\frac{1}{pM}+\frac{1}{qM}\right)$$

$$+\frac{x^2}{4}\left(\frac{1}{p^2n^2}+\frac{1}{q^2n^2}+\frac{1}{p^2M^2}+\frac{1}{q^2M^2}\right)+\cdots\cdots\cdots$$

n 當然小於 N，且可認爲小於 $\frac{1}{2}$N(而不背乎通理 [generality])故 n 必小於M。

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

若以$\frac{x^2}{n}$爲可與整一 (unity) 相比敵之數量，則如第四節例二所言，$\frac{x}{n}$屬於$\frac{1}{\sqrt{n}}$次(order)，$\frac{x^2}{n^2}$屬於$\frac{1}{n}$次。

於是將$\frac{1}{\sqrt{n}}$次或更高次各項省略，則得

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

以 σ^2 代$\frac{pqnM}{N}$。

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

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

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

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

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

第十一節　小數律

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

設 $u=pn$，且爲一極小之有限數。$p=\frac{u}{n}$，$q=1-\frac{u}{n}$。

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

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

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

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

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

且於$\frac{r}{n}$達到0時,漸及於1。

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

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

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

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

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

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

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

於$\frac{r}{u}$歷經各整數值時，則p_r以迅速之速率變小，例如假以 $u=6$，而 $r=3u$，則 $P_{3u}=.00004$。

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

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

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

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

r	$e^{-u} \cdot u^r/r!$		預測數	實在数
0	.00006	×20	=.001	0
1 至 4	.0343		=.7	3
5 ,, 9	.4564		=9.1	6
10 ,, 14	.4408		=8.8	8
15 ,, 19	.0683		=1.4	3
20 ,,	小		—	0

（註一） 見附錄二，公式133。

（註二） 下文爲簡便起見，假定 pn 爲整數，故 P_0 爲最大項；而 n 旣大，則 $\frac{1}{n}$ 之乘方乃可於最後刪略之，但與算式證明無干。

（註三） 參閱 Edgeworth 氏所著大英百科全書第十二部之『機率』一文，自第三百九十一頁起。

（註四） 附錄四。

（註五） 因 $m^2t+1 = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} x2t+1\, e^{-\frac{x^2}{2\sigma^2}}dx = \int_{-\infty}^{\infty}\phi(x)dx$,

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

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

（註七） 注意，$p \sim p'$ 爲比例（proportion）之離差。實際結果，離中差應爲 $pn \sim p'n$，故 $z = \frac{pn \sim p'n}{\sqrt{p(1-p)n}}$。如是，則 $\sqrt{n}$ 增加，機率亦隨之增加。

（註八） 例如 $r = 13.5$，及 14.5，則 $F(z) = .4994$ 及 .4986。此差額，.0008 ×300，卽爲 $r = 14$ 時之希望數（Expectated number）。

（註九） 前欄最靠近之全數。

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

（註一一）例如，$N=10,000$，$n=500$，則可取七位數字表按頁取其各數目之末四位數字，以至取得自 0 至 10,001 間之數，五百個爲止。然後再查明各號碼所指之事物，而加以調查。此法於調查各區人數時用之。見下節。

（註一二）例如從一副五十二張之牌，每抽取十三張爲一把，則每把中抽到三個A點之機率，爲 $P_2={}_4C_3\times{}_{48}C_{10}\div{}_{52}C_{13}$，由此看來$N=52$，$n=13$，$p=\frac{1}{13}$，$x=2$，$Px$乃約略等於.041。

（註一三）如u等於10時，此式與$\frac{1}{\sqrt{2\pi u}}$相差，不過百分之一。

第三章　大數律(普遍的差誤律)

在上文中,吾人旣已以差誤常態曲線,作爲二項式$(q+p)^n$之極限,並已表明其整數在p有確定意義時之應用。但現有同一之方程式 $y=\frac{1}{\sigma\sqrt{\pi 2}}e^{-\frac{x^2}{2\sigma^2}}$,乃爲十分廣泛之假說下之得數,本章之主要目的,卽欲對此種種假說而加以發揮也。

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

第一節　平均及總和之標準差及均立方差誤

(1)以 $u_1,u_2\cdots\cdots u_t\cdots\cdots u_{m1}$ 爲一頻數羣類之 m_1 個量數,而以 $\bar{u}$ 爲其平均數,σ_u 爲其標準差。

假以 $u_t=\bar{u}+u'_t$。

則　$m_1\bar{u}=\mathrm{S}u_t\quad\therefore\mathrm{S}u_t'=0$

而 $m_1\sigma_u^2 = Su_t'^2 = S(u_t-\bar{u})^2 = Su_t^2 - 2\bar{u}Su_t + m_1\bar{u}^2$

$$= Su_t^2 - 2\bar{u}\cdot m_1\bar{u} + m_1\bar{u}^2$$

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

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

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

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

然則 $H_2 \times m_1 \times m_2$

$$= (u_1+v_1) + \cdots\cdots + (u_1+v_t) + \cdots\cdots + (u_1+v_{m_2})$$

$$+ (u_2+v_1) + \cdots + (u_2+v_t) + \cdots + (u_2+v_{m_2}) \cdots\cdots$$

$\cdots\cdots$ 共有 m_1 行，每行各有 m_2 項。

$$+ (u_{m_1}+v_1) + \cdots + (u_{m_1}+v_t) + \cdots (u_{m_1}+v_{m_2})$$

$$= m_2 \cdot Su_t + m_1 \cdot Sv_t = m_2 m_1 \bar{u} + m_1 m_2 \bar{v}$$

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

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

$$=(u_1+v_1)^2+\cdots\cdots+(u_1+v_{m_2})^2+(u_2+v_1)^2+\cdots\cdots+(u_2+v_{m_2})^2+\cdots\cdots$$

$$=m_2 Su_t^2+m_1 Sv_t^2+2\cdot Su_t\cdot Sv_t$$

$$=m_2 m_1(\sigma_u^2+\bar{u}^2)+m_1 m_2(\sigma_v^2+\bar{v}^2)+2m_1\bar{u}\cdot m_2\bar{v}$$

$$=m_1 m_2\{(\sigma_u^2+\sigma_v^2)+(\bar{u}+\bar{v})^2\}$$

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

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

(4)復次以從三類中抽出之和（或差），另成一類，其三類之平均數爲$\bar{u},\bar{v},\bar{w}$，標準差爲σ_u，σ_v,σ_w，而和或差爲H_3,S_3。

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

其理顯而易見。

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

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

此公式，用歸納方法，可引申於任何多少羣類也。

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

如此總和乃由此等羣類n個所組成，而其標準差爲S，則

$$S^2=S_n^2=\sigma^2+\sigma^2+\cdots\cdots\text{直到 } n \text{ 項爲止}=n\sigma^2$$

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

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

$$\therefore \qquad \sigma_a=\frac{S}{n}=\frac{\sigma}{\sqrt{n}} \quad\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots(38)$$

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

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

u_s 與 v_t 相加之均立方(mean cube)。爲 $\frac{1}{m_1m_2}S\,(u_s+v_t)^3$

$$=\frac{1}{m_1m_2}\{m_2Su^3+m_1Sv^3+3\cdot SvSu^2+3\cdot SuSv^2\}=\frac{1}{m_1}Su^3+\frac{1}{m_2}Sv^3$$

$={}_u\mu_3+{}_v\mu_3$，即爲各類以平均數爲中心而得之第三動差之總和。

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

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

(K′ 爲該類之"K"值，其定義已見公式(8))

不論，n 個項目之平均數，抑或爲其總和，K 值終爲相同。

第二節 差誤曲線之發生

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

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

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

$$\frac{1}{\sigma_u\sqrt{2\pi}}e^{-\frac{u^2t}{2\sigma u^2}}\times\frac{1}{\sigma_v\sqrt{2\pi}}e^{-\frac{v^2t}{2\sigma v^2}}$$

$$=\frac{1}{2\pi\sigma_u\sigma_v}e^{-\frac{1}{2}\left(\frac{ut^2}{\sigma u^2}+\frac{(xt+ut)^2}{\sigma v^2}\right)}。$$

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

$$\frac{1}{2\pi\sigma_u\sigma_v}\int_{-\infty}^{\infty}e^{-\frac{\sigma u^2+\sigma v^2}{2\sigma u\ \sigma v^2}\left(u-\frac{\sigma u^2xt}{\sigma u^2+\sigma v^2}\right)^2}e^{-\frac{xt^2}{2(\sigma u^2+\sigma v^2)}}du\cdot\delta x$$

$$=\left(以u'代u-\frac{\sigma_u{}^2x_t}{\sigma_u{}^2+\sigma_v{}^2}\right),$$

$$\frac{1}{2\pi\sigma_u\sigma_v}\cdot e^{-\frac{xt^2}{2(\sigma u^2+\sigma v^2)}}\cdot\delta x\cdot\int_{-\infty}^{\infty}e^{-\frac{\sigma u^2+\sigma v^2}{2\sigma u^2\sigma v^2}}du',$$ 又依第二章第四節公式 20 至21

$$=\frac{1}{S_2\sqrt{2\pi}}e^{-\frac{xt^2}{2s_2}}\cdot\delta x,\text{式中 } S_2{}^2=\sigma_u{}^2+\sigma_v{}^2。$$

故 x 值之機率爲 $\frac{1}{S_2\sqrt{2\pi}}e^{-\frac{x^2}{2s_2{}^2}}$ ………………………………… (40)

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

均中，求得 x 之機率，分別爲

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

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

第三節　用多項式定理證明之

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

設有 n 個元始羣類，各有 $m_1,m_2\cdots\cdots m_n$ 個可量之事物；在任何一類中，譬如第 t 類，以 $\bar{u}_t$ 爲其平均數，σ_t 爲其標準差，${}_t\mu_2$，${}_t\mu_3\cdots\cdots$爲其以平均數爲中心之動差，並設各項爲 $\bar{u}_t+{}_tu_1,\bar{u}_t+{}_tu_2,\cdots\cdots\bar{u}_t+{}_tu_s\cdots\cdots$

則，${}_tu_1+{}_tu_2+\cdots\cdots+{}_tu_s\cdots\cdots=0$。

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

在第 S 次抽選時，其相加之和爲 $\mathrm{H}+\mathrm{E}_s$，（此處之 $\mathrm{H}=\bar{u}_1$

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

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

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

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

$$M_2=S^2=\frac{1}{N}\left\{\frac{N}{m_1}S_1u^3+\cdots\frac{N}{m^2}S_2u^3+\cdots\right\}+\frac{3}{N}\left\{\frac{N}{m_1m_2}S_1u\cdot S_2u+\cdots\right\}$$

$$=\sigma_1^2+\sigma_2^2+\cdots\cdots\sigma_n^2\text{，因爲}0=S_1u=S_2u=\cdots\cdots\text{依同理，}$$

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

$$=\frac{1}{N}\left\{\frac{N}{m_1}S_1u^3+\frac{N}{m_2}S_2u^3+\cdots\cdots\right\}+\frac{3}{N}\left\{\frac{N}{m_1m_2}S_1uS_2u^2+\cdots\right\}$$

$$={}_1\mu_3+{}_2\mu_3+\cdots\cdots+{}_n\mu_3\text{，因爲}0=S_1u\ \cdots\cdots\cdots\cdots\cdots\cdots(42)$$

$$\text{且}M_4=\frac{1}{N}\left\{\frac{N}{m_1}\cdot S_1u^4+\cdots\cdots\right\}+\frac{4}{N}\left\{\frac{N}{m_1m_2}S_1u^3S_2u+\cdots\cdots\right\}$$

$$+\frac{12}{N}\left\{\frac{N}{m_1m_2m_3}S_1u^2S_2uS_3u+\cdots\cdots\right\}$$

$$+\frac{6}{N}\left\{\frac{N}{m_1m_2}\cdot S_1u^2\cdot S_2u^2+\cdots\cdots\right\}$$

$$+\frac{24}{N}\left\{\frac{N}{m_1m_2m_3m_4}S_1uS_2uS_3uS_4u+\cdots\cdots\right\}$$

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

$$\therefore M_4-3s^4=S_t\mu_4+6(\sigma_1^2\sigma_2^2+\cdots)-3(\sigma_1^2+\sigma_2^2+\cdots)^2=S({}_t\mu_4-3\sigma_t^4)$$

若幾類元始曲線之標準差相同，動差亦一樣，如是，$\sigma_1=\sigma_2=\cdots=\sigma$，${}_1\mu_3={}_2\mu_3=\cdots\cdots=\mu_3$，其他以此類推，則S(總和)$=\sigma\sqrt{n}$，

$$\sigma_a(\text{平均數})=\frac{\sigma}{\sqrt{n}}\cdots\cdots\cdots\cdots\cdots\cdots(44)$$

$$K(\text{總和或平均})=\frac{M_3}{S^3}=\frac{n\mu_3}{n^{\frac{3}{2}}\sigma^3}=\frac{K'}{\sqrt{n}}\cdots\cdots\cdots\cdots(45)$$

（K′爲各元始曲線中之“K”）。

$$K_2=\frac{M_4}{S^4}-3=\frac{n(\mu_4-3\sigma^4)}{n^2\sigma^4}=\frac{1}{n}\left(\frac{\mu_4}{\sigma^4}-3\right)$$

$$=\frac{\text{各元始曲線中之“K”值之和}}{n}\cdots\cdots\cdots\cdots(46)$$

故當$\sqrt{n}$變大時，K漸近於零，而於$\frac{1}{n}$可以略去時，K_2亦可作爲零。

欲求更高級動差，須以數值表示m_t而以t爲任何之整數；換言之，卽爲$({}_1u+{}_2u+\cdots\cdots{}_nu)^t$之平均值，至$({}_1u+{}_2u+{}_nu)^t$式（依多項式定理（註一）），在$n_1+n_2+\cdots\cdots=t$條件之下，所有各可能值加在一起時，乃卽

$$S\frac{t!}{n_1!n_2!\cdots\cdots}\cdot{}_1u^{n1}\cdot{}_1u^{n2}\cdots\cdots$$

之平均值。

第一就t爲偶數時之例言之。

$$M_{2t}=\frac{(2t)!}{n_1!\,n_2!\cdots\cdots}{}_1u^{n1}\cdot{}_2u^{n2}\cdots\cdots\text{各項平均數之和}$$

當$n_1+n_2+\cdots\cdots=2t$

$M_{2t}=\frac{(2t)!}{n_1!\,n_2!}\cdot {}_1u^{n_1}$平均值$\times {}_2u^{n_2}$平均值$\times\cdots\cdots$各項之和

因自各元素曲線中之抽取係獨立的，故每一${}_1u$與每一${}_2u\cdots\cdots$之實現，各有均等之次數也。

$$\therefore M_{2t}=\frac{(2t)!}{n_1!\,n_2!\cdots\cdots}\cdot {}_1\mu_{n1}\cdot {}_2\mu_{n2}\cdots\cdots$$

其次再就一種情形——各元素曲線之標準差及動差均相等——而論；在此情形之下，${}_1\mu_{n1}={}_2\mu_{n1}=\cdots\cdots=\mu_{n1}$，以此類推。

在任一抽出項中，設有F個因子${}_1\mu_{n1},{}_2\mu_{n1}\cdots\cdots$則此一項，以各種形式而出現，有${}_nCf$次之多：

$${}_1\mu_{n1}\times{}_2\mu_{n2}\times{}_3\mu_{n3},\cdots{}_1\mu_{n2}\times{}_2\mu_{n1}\times{}_3\mu_{n3}\cdots,{}_1\mu_{n3}\times{}_2\mu_{n2}\times{}_3\mu_{n1}\cdots,$$

式中之每一項均與$\mu_{n1}\times\mu_{n2}\times\mu_{n3}\cdots\cdots$相脗合。

故$M_{2t}={}_nCf\frac{2t!}{n_1!\,n_2!\cdots}\mu_{n1}\times\mu_{n2}\times\cdots\cdots$各項之和，而此各項均適合$n_1+n_2+\cdots\cdots=2t$之條件。

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

$$\therefore\frac{M_{2t}}{s^{2t}}=\frac{(2t)!}{n_1!\,n_2!\cdots}\cdot\frac{\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{f-1}{n}\right)}{f!}\cdot\frac{n^f}{n^t}\cdot\frac{u_{n1}}{\sigma^{n1}}\cdot\frac{\mu_n}{\sigma^{n2}}$$

$\cdots\cdots$各項之和。

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

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

因μ_1等於0，n_1, n_2……之每一項在不爲零之各項，其值必不下於2；故因f項n_1, n_2,……等項之和等於$2t$，則各項之最大可能數乃爲t，而且$f \not> t$。

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

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

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

因爲$\mu_2=\sigma^2$且$\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{f-1}{n}\right)$在1與$1-\frac{f(f-1)}{2n}$之間。當略去含有$\frac{1}{n}$之各項時，

是以$M_{2t}=S^{2t}\cdot\frac{(2t)!}{2^t t!}=1\cdot3\cdot5(2t-1)S^{2t}$……………………(47)

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

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

然則 $\frac{M_{2t+1}}{S^{2t+1}}=f\times\frac{(2t+1)!}{2^{t-1}3!}\cdot\frac{1}{t!\sqrt{n}}\cdot\frac{\mu_2{}^{t-1}\cdot\mu_3}{\sigma^{2t+1}}$

$$=\frac{t}{3}\cdot1\cdot3\cdot5\cdots\cdots\overline{2t+1}\cdot\frac{\mu_3}{\sqrt{n}\,\sigma^3}\cdots\cdots\cdots\cdots\cdots\cdots \quad (48)$$

如將包含$\frac{1}{\sqrt{n}}$之項略去，

$$\therefore M_2t+1=0,$$

且 $$M2_{t+1}=\frac{t}{3}\cdot1\cdot2\cdot3\cdot\overline{2t+1\cdot}M_3\cdot S^{2t-2}\cdots \quad (49)$$

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

如此得來之動差，與(參閱公式 23，及附錄六)由曲線（已將含$\frac{1}{\sqrt{n}}$項者删除）$y=\frac{1}{S\sqrt{2\pi}}\cdot e^{-\frac{x^2}{2s^2}}$，

得來者，毫釐不差。再如將有$\frac{1}{\sqrt{n}}$之項保留，而删除其$\frac{1}{n}$項，則可由曲線 $y=\frac{1}{S\sqrt{2\pi}}\left[1-\frac{k}{2}\left(\frac{x}{s}-\frac{1}{3}\cdot\frac{x^3}{s^3}\right)\right]e^{-\frac{x^2}{2s^2}}$

得來，惟式中所用之$K=\frac{M_3}{S^3}$。

是故，如以標準差相合，動差亦一致，卽認爲曲線亦脗合。則上述之方程式，便爲該頻數曲線之第一，二近似值。

第四節　愛基華斯氏之證明

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

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

援用上述之標號及條件,

設 $E_s={}_1u_s+{}_2u_s+\cdots\cdots{}_nu_s$。

設 a 爲任一固定小數量,僅用以抽選同元(dimensions)之各項,

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

$e^{a\cdot{}_1u_s}$之平均值,換言之,即

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

從各元素曲線之抽選既係獨立,則

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

$\therefore 1+a\cdot M_1+\frac{a^2}{2}M_2+\frac{a^3}{3!}M_3+\cdots\cdots=$因子 n 個之乘積,例如以下之因子:$\left(1+\frac{a^2}{2}\cdot{}_t\mu_2+\frac{a^3}{3!}\cdot{}_t\mu_3+\cdots\cdots\right)$

$$\therefore \log\left(1-a\mathrm{M}_1+\frac{a^2}{2}\mathrm{M}_2+\cdots\cdots\right)$$

$$=\mathrm{S}_{t=1}^{t=n}\log\left(1+\frac{a^2}{2}\cdot{}_t\mu_2+\frac{a^3}{3!}\cdot{}_t\mu^3+\cdots\cdots\right)$$

七 $$=\frac{a^2}{2}\mathrm{S}_t\mu_2+\frac{a^3}{6}\cdot\mathrm{S}_t\mu_3+\frac{a^4}{24}\mathrm{S}_t\mu_4+\cdots-\tfrac{1}{2}\left(\frac{a^2}{2}\mathrm{S}({}_t\mu_2)+\cdots\right)^2$$

$$\therefore 1+a\mathrm{M}_1+\frac{a^2}{2}\mathrm{M}_2+\cdots\cdots$$

$$=e^{\frac{a^2}{2}\cdot\mathrm{S}_t\mu_2}\cdot e^{\frac{a^3}{6}\mathrm{S}_t\mu_3}\cdot e^{\frac{a^4}{24}(\mathrm{S}_t\mu_4-3\mathrm{S}({}_t\mu_2)^2)}$$

$$=\left(1+\frac{a^2}{2}\mathrm{S}_t\mu_2+\cdots\cdots+\frac{1}{p!}\left(\frac{a^2}{2}\mathrm{S}_t\mu_2\right)^p+\cdots\right)\cdot\left(1+\frac{a^3}{6}\mathrm{S}_t\mu_3+\cdots\cdots\right)$$

五 $$\cdot\left(1+\frac{a^4}{24}\left\{\mathrm{S}_t\mu_4-3\mathrm{S}({}_t\mu_2)^2\right\}+\cdots\cdots\right)\cdots\cdots$$

以 a^4 爲準，令各係數相等。

$$\mathrm{M}_1=0$$

$\mathrm{S}^2=\mathrm{M}_2=\mathrm{S}_t\mu_2=\mathrm{S}\sigma_t^2=n\sigma^2$，設 σ^2 爲 σ_1^2，σ_2^2 ……之平均

$\mathrm{M}_3=\mathrm{S}_t\mu_3=n\mu_3$，設 μ_3 爲 ${}_1\mu_3$，${}_2\mu_3$ ……………之平均

$$\frac{\mathrm{M}_4}{24}=\frac{1}{2}\cdot\left(\frac{1}{2}\cdot\mathrm{S}_t\mu_2\right)+\frac{1}{24}\left((\mathrm{S}_t\mu_4-3\mathrm{S}({}_t\mu_2)^2\right)$$

$$\therefore \mathrm{M}_4-3\mathrm{S}^4=\mathrm{S}({}_t\mu_4-3\sigma_t^4)$$

$k_2=\dfrac{\mathrm{M}_4}{\mathrm{S}^4}-3=\dfrac{1}{n^2\sigma^4}\mathrm{S}\left(\dfrac{{}_t\mu_4}{\sigma_t^4}-3\right)\sigma_t^4=\dfrac{1}{n}k_2'$，設 k_2' 爲

$\left(\dfrac{{}_t\mu_4}{\sigma_t^4}-3\right)\left(\dfrac{\sigma_t}{\sigma}\right)^4$ 之平均，

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

$$\therefore 1+\frac{a^2}{2}S^2+\frac{a^3}{3!}M_3+\cdots\cdots+\frac{a^t}{t!}M_t\cdots\cdots$$

$$=e^{\frac{1}{2}a^2s^2}\left(1+\frac{1}{6}a^3\cdot s^3\cdot\frac{1}{\sqrt{n}}k'+\cdots\right)\left(1+\frac{1}{24}a^4s^4\cdot\frac{1}{n}k_2'+\cdots\right)\cdots$$

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

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

略去$\frac{1}{n^{\frac{1}{2}}}$及其以上之乘羃

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

$\therefore$每奇數級動差，M_{2t+1}，$=0$

至偶數級動差，$M_{2t}=(2t)!\frac{s^{2t}}{t!\,2^t}=1\cdot3\cdots(2t-1)\cdot s^{2t}\cdots$ (50)

如在差誤常態曲線（公式23）中。

現存其$\frac{1}{\sqrt{n}}$，而删其$\frac{1}{n}$。

M_{2t} 如前述。

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

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

s^{2t-2} 換言之，卽下列曲線之第$(2t+1)$級動差

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

（參閱附錄六）

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

更進一步之近似值，愛氏雖均曾作出，然此乃純粹理論之研究，本書不取也。

第五節　普遍的差誤律或大數律之說明

上文所證之定理，玆可概述於下，其有效條件，仍敍及並爲擴大焉。

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

從各類中，各取出一事物，n 個事物之總計爲H，則自此類抽選，不至 —— 或至微 —— 影響其他各類之抽取；照此手續重複

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

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

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

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

式中之s^2，爲軌距之第二動差，ks^3爲軌距之第三動差。

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

元素曲線必須適合之條件，爲：如up爲第p次之動差，σ爲其中任一羣類之標準差，則不論p值若干，$\frac{up}{\sigma p}$終爲一極小有限數（註三）（此值如乘以$\frac{1}{n^{\frac{p}{2}-1}}$，則變爲極微可以省略）。欲達

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

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

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

第六節　範圍有限制之例

在第一節吾人曾有假定，謂一個項目之選擇，不至影響其他抽選之機會。

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

假設一羣類含有n個事物，此羣類乃自含有N個事物之羣類

中抽選而來，而此N個事物之量數爲 $\bar{u}+u_1, \bar{u}+u_2 \cdots\cdots, \bar{u}+u_N$，在此式中 $\bar{u}$ 爲其平均數，而 $\mathrm{S}_1^N u_t=0$。茲以 H+E 爲所選出 n 個事物之總和，而 $\mathrm{H}+n\bar{u}$。

E之均等或然值，有 $_N\mathrm{C}_n$ 個，例如

$$u_1+u_2+u_3+\cdots\cdots+u_n$$

$$u_1+u_3+u_4+\cdots\cdots+u_N$$

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

以上數值之和，必等於零，此乃顯而易見之事實，故E之平均值亦等於零。

設 s^2 爲E之標準差。

則 $_N\mathrm{C}_n \cdot s^2 = {_N\mathrm{C}_n}$ 個平方之和，例如$(u_1+u_2+\cdots\cdots+u_n)^2$等，每個內包含 n 項即是。

在此總和中每一平方，例如 u_t^2，出現$\dfrac{n}{\mathrm{N}}\times {_N\mathrm{C}n}$次，而每一乘積$\cdot 2u_s u_t$ 出現$\dfrac{1}{_N\mathrm{C}_2}\times\dfrac{n(n-1)}{2}\times {_N\mathrm{C}n}$次，因共有 $n\times {_N\mathrm{C}n}$ 個平方及$\dfrac{n(n-1)}{2}\times {_N\mathrm{C}_n}$ 個乘積也。

$$\therefore {_N\mathrm{C}n}\cdot \mathrm{S}^2=\frac{n}{\mathrm{N}}\times {_N\mathrm{C}n}\cdot \mathrm{S}_1{}^N u_{t^2}+2\cdot\frac{n(n-1)}{\mathrm{N}(\mathrm{N}-1)}\cdot {_N\mathrm{C}n}\cdot \mathrm{S}u_s u_t$$

$$\mathrm{S}_2=\frac{n}{\mathrm{N}}\cdot \mathrm{N}\sigma_2+\frac{n(n-1)}{\mathrm{N}(\mathrm{N}-1)}\left\{(\mathrm{S}u_t)^2-\mathrm{S}u_t{}^2\right\},$$

（σ 爲所從抽選之『範圍』之標準差）

$$=n\sigma^2-\frac{n(n-1)}{N-1}\sigma^2, \qquad \therefore Su_t=0$$

假如$\frac{1}{N}$可以删略，

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

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

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

假如N爲無限大，可依前文而得$\frac{\sigma}{\sqrt{n}}$（公式38）。至$\frac{n}{N}$一值，如果删除之則標準差必有張大之虞。

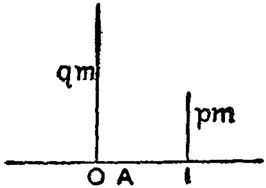
據證明(Isserlis，統計學報 Stat. Journal，一九一八年號第七十五頁下同)，總和或平均所得之頻數，酷近常態，設N爲極大（事實上亦多如此）。

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

附註——前論大數律係由$(p+q)^n$之極限得來一言，現可證明乃爲一般分析之特殊情形。

設每一元素羣類，各具有 qm 個零，pm 個單位，而 $p+q=1$。

此類之常數爲$\bar{x}=\dfrac{qm\times 0+pm\times 1}{qm+pm}=p$。



pm 與平均數，A，之距離爲$+q$，qm 與 A之距離爲$-p$。

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

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

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

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

第七節 例證

大數律，如取其積分形式，以第二近似值而論，乃爲

$$P(\pm x)=\frac{1}{\sigma\sqrt{2\pi}}\int_0^x e^{-\frac{x^2}{2\sigma^2}}dx\mp\frac{k}{6\sqrt{2\pi}}\left\{1-\left(1-\frac{x^2}{\sigma^2}\right)e^{-\frac{x^2}{2\sigma^2}}\right\}\text{（註四）}$$

$$=\frac{1}{\sqrt{2\pi}}\int_0^z e^{-\frac{1}{2}z^2}dz\mp\frac{k}{6\sqrt{2\pi}}\left\{1-(-z^2)e^{-\frac{1}{2}z^2}\right\}$$

$$=F(z)\mp kf(z)\ \cdots\cdots\cdots\cdots\ (54)$$

式中 $P(x)$ 為對平均數之正標準差不超過 x 之機率，$z=\frac{x}{\sigma}F(z)$ 已列表見二章第四節第三表，至 $f(z)=\frac{1}{6\sqrt{2\pi}}\left\{1-(1-z^2)e^{-\frac{1}{2}z^2}\right\}$ 之表則列於次頁（第四表）。

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

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

以下各例，僅將第一則全部作出，其他則未也。

(1)茲有冗長之書一本，(A)依次取其 10,000 行之行首第一整字，數各該字之字母若干；(B)以十行之行首字為一批，各批字母相加，然後計算 1000 批之字母總數若干；(C)再以 100 行行首字為一批，各批字母相加，便得 100 個總數。

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

足應用矣。

A——10,000個字依其所含字母數之分佈情況

字 母 數	x	觀察數 y	xy	x^2y	x^3y	z	(註五) $F(z)$	差額 ×10,000
1 或 .5 至 1.5	−7	127	− 889	6,223	− 43561	−1.62	0.447	490
2 ,, 1.5 ,, 2.5	−6	1,792	−10752	64,512	−387072	−1.27	.396	770
3 ,, 2.5 ,, 3.5	−5	1,984	− 9920	49,600	−248000	− .92	.321	1,020
4 ,, 3.5 ,, 4.5	−4	1,240	− 4960	19,840	− 79360	− .58	.219	1,280
5 ,, 4.5 ,, 5.5	−3	968	− 2904	8,712	− 26136	− .23	.091	1,390
6 ,, 5.5 ,, 6.5	−2	812	− 1624	3,248	− 6496	+ .12	.048	1,330
7 ,, 6.5 ,, 7.5	−1	893	− 893	893	− 893	+ .47	.181	1,130
8 ,, 7.5 ,, 8.5	0	634	0	0	0	+ .82	.294	850
9 ,, 8.5 ,, 9.5	1	602	+ 602	602	+ 602	+1.17	.379	570
10 ,, 9.5 ,, 10.5	2	460	+ 920	1,840	+ 3680	+1.52	.436	330
11 ,, 10.5 ,, 11.5	3	260	+ 780	2,340	+ 7020	+1.87	.469	180
12 ,, 11.5 ,, 12.5	4	116	+ 464	1,856	+ 7424	+2.22	.487	80
13 ,, 12.5 ,, 13.5	5	69	+ 345	1,725	+ 8625	+2.57	.495	30
14 ,, 13.5 ,, 14.5	6	21	+ 126	756	+ 4536	+2.92	.498	10
15 ,, 14.5 ,, 15.5	7	18	+ 126	882	+ 6174	+3.27	.499	10
16 ,, 15.5 ,, 16.5	8	4	+ 32	256	+ 2048	+3.62	.500	0
		10,000	−31942	163,285	−791518			
			+ 3395		40109			
			−28547		−751409			

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

$\mu_2=16.3285-\bar{x}^2=8.1792$，$\sigma=2.860$

$\mu_3=-75.1409-3(-2.8547)(16.3285)+2(-2.8547)^3=18.1704$。

$\alpha=\frac{\mu_3}{\mu_2^{\frac{3}{2}}}=\cdot 78$。

爲計算動差，即以 8 爲假定原點。

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

B——十字爲一批，千批字母總和之分佈情況

字母數	z	F(z)	差額 ×1000	觀察數	F(z)∓Kf(z) (註七)		差額 ×1000
			4	0			0
26.5	−2.650	.496			.078	.528	
			13	8			8
31.5	−2.119	.483			.091 (註六)	.520	
			39	38			37
36.5	−1.588	.444			.095	.483	
			89	97			99
41.5	−1.057	.355			.071	.384	
			154	155			173
46.5	− .526	.201			.025	.211	
			203	227			213
51.5	+ .005	.002			.000	.200	
			202	202			191
56.5	+ .536	.204			.026	.193	
			153	154			135
61.5	+1.067	.357			.072	.328	
			88	76			78
66.5	+1.598	.445			.095	.406	
			38	37			40
71.5	+2.129	.483			.091	.446	
			13	13			18
76.5	+2.660	.496			.078	.464	
			3	9			7
81.5	+3.191	.499			.069	.471	
			1	3			2
86.5	+3.722	.500			.067	.473	
			0	1			0

關於此 1000 批總和，平均數爲 51.453，$\sigma=9.4155$，$k=.4093$。字母數之和，最大不過 87，最小亦在 26 之上。第七八九三欄關於 z，$F(z)$ 及差額之計算與 A 相同。此者對於常態曲線之配合情形，良好多多，而在其全距 31.5 至 76.5，卽平均數士 2σ 之間，其爲佳妙，可謂無以復加，但公式應用結果，在 31.5 以下過多，而在 76.5 上者甚少，此頗嫌美中不足耳。

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

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

但 $2.860\times\sqrt{10}=9.04.81\div\sqrt{10}=.25$，而吾人由 B 直接求得之標準差爲 9.42，K 爲 .41，相差如此之鉅，可知乃係每十字一批，每批字母攏集時，未能完全獨立之故；迨經分析，乃知連續所抽十字之間，果有相當關聯也。實際上，如照 C 式連續將百字字母相加，則得 $\sigma\doteq33.311$ 非復 $2.86\times\sqrt{100}$ 矣，然若將各批字母數次序變更重爲排列，使以便將全書各部份之行首字盡行包括於每次一百字之各批中，則 σ 爲 28.87，則與理論相合。

C——百字爲一批，共百批字母總數之分佈

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

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

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

(2)一花園中,行人路側有邊界用磚堆壘而成,(並非實砌),兩端縱長相間連接,乃經時日既久,風吹雨打,遊人雜踏,歷經消磨,迄今已非原狀。玆以四塊磚爲一列,量其 143 列之長,量法務求精密,以至一英寸之十六分之一爲止。但其變量必須顧及，究

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

四磚爲一列，各列長度之分佈

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

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

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

頂可(Dinka)族頭蓋骨及體高之測量

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

除兩極端項註有 * 號者外，全距係爲常態形，且以上取才之

例如此之少，與常態曲線相差並不如所預料之多。

(4)北海漁業調查所，五百五十四條鰈魚之體長測量，結果如下：

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

平均數 31.778；$\sigma=1.335$.

兩極端各項未能與常態相緊合。

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

下列之表，將所得資料與大數律之第一第二近似值作一比較。

年齡	學童人數	用近似值求得之人數	
		第一近似值	第二近似值
10	26	39	27
11	201	207	204
12	673	630	670
13	1,001	983	995
14	739	785	746
15	310	323	307
16	80	67	79
17	13	9	15
18	1	0	0

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

第一近似值，在平均數左右 2σ 以內，尚能配合。

第二近似值，則與觀察值密切配合（參閱附錄六另圖）。

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

平均速度每秒，1.5846 公尺。$\sigma=.2179$ 公尺。

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

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

(7)採用 1918 年勞工生活費委員會所搜集之材料，將在城市之家庭九百七十家，每星期食物費，用『等成年』(equivalent adults 一卽以兒童作爲成年人之幾分之幾計算——)除之。平均數爲 10.75 先令；$\sigma=3.156, k=.84$。

每週每一單位之食物費	家數 實際數	家數 用第二近似値求得之數	家數 由皮爾生氏第三形態求得數（註八）
5.5 先令以下	18	22	7
5.5先令	107	123	122
7.5	255	233	252
9.5	245	248	250
11.5	173	168	172
13.5	101	89	95
15.5	38	51	45
17.5	17	22	19
19.5	9 } 33	11 } 35	7 } 27
21.5以上	7	1	1

$\beta_1=.708$。$\beta_2=4.035$。

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

每磅價 2.629 分,而 $\sigma = .3334$。

依麵粉價格對各城市之分類

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

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

（註一）　多項式定理乃爲二項式定理之引申；玆將其證明，扼要述之如下：$(a_1+b_1+c_1+\cdots)(a_2+b_2+c_2+\cdots)(a_t+b_t+c_t+\cdots)$ t 個因子之乘積 = 各可能項,諸如 $a_1a_2b_3c_4d_5b_6\cdots k_t$（每一附數 (suffix) 各出現一次）之總和。a 出現 n_1 次,b 出現 n_2 次……諸如此類之項數,乃

卽每次取 t 個事物而有 n_1 相同，n_2 個相同……，意卽 $\frac{t!}{n_1!n_2!\cdots}$，之排列(permutation)數。以 $a_1=a_2\cdots={}_1u$，$b_1=b_2=\cdots={}_2u$，……以此類推，則得結果如正文。

（註二） 設自一基本羣類抽選，不能彼此獨立，而各數量乃成批出現，則須多得 H 之數值，當所得數值甚多時。方能求出其最終顯數形式之已知近似值，

（註三） 以準確之語言之，此數非他，卽此比率與在被包含之常態曲線中之相當比率之相差也。

（註四） 見附錄六，積分部分。

（註五） 見第二章第四節第三表。

（註六） 十，z 爲負數時用之。

（註七） 見第三章第七節 $f(z)=\frac{1}{6\sqrt{2\pi}}\left\{1-(1-z^2)e^{-\frac{1}{2}z^2}\right\}$ 數值表

（註八） 見第五章第一節。

第四章　差誤律之應用

第一節　平均數及總和數之精度

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

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

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

次舉二例（數字及緯度），表明若干平均數之分佈，與差誤常態曲線相合之情形。

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

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

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

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

誤，必係一常態之頻數。

第二節　平均數之精度

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

註册區		五十抽一之抽樣 住宅數	人數	每宅人數	全部各區 住宅數	每家住宅人數	標準差
白斯諾爾	東北區	277	1.224	4.42	13.850	4.35	.14
	西南區	278	1.261	4.54	13.905	4.60	.14
愛爾地支	南區	187	792	4.24	9.331	4.26	.18
	西區	152	693	4.56	7.623	4.34	.19
	東北區	156	653	4.19	7.847	4.39	.19
斯皮特菲爾		130	637	4.90	6.476	4.79	.21
懷德加白		117	519	4.44	5.914	4.72	.22
聖喬治		187	924	4.93	9.374	4.88	.18
色得外爾		95	387	4.07	4.800	4.37	.25
萊母好斯		133	611	4.59	6.655	4.54	.21
邁爾恩德	西南區	267	1.211	4.54	13.366	4.71	.15
	東北區	207	839	4.05	10.364	4.40	.17

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

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

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

第三節 平均數之常態分配

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

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

下列一表,將一千個總數之分配與差誤常態曲線,二者作一比較。

在某種限度內1000個總和之次數

與平均數之距離	算出之次數	標準差	觀察數	差額
$\frac{5}{2}\sigma$以上	6	2.4	8	+2
2σ	17	4.1	17	0
$\frac{3}{2}\sigma$	44	6.5	47	+3
σ	92	9.1	75	−17
$\frac{1}{2}\sigma$	150	11.3	157	+7
0至$\frac{1}{2}\sigma$	191	12.4	197	+6
0至$-\frac{1}{2}\sigma$	191	12.4	201	+10
$-\frac{1}{2}\sigma$	150	11.3	148	−2
$-\sigma$	92	9.1	77	−15
$-\frac{3}{2}\sigma$	44	6.5	50	+6
-2σ	17	4.1	20	+3
$-\frac{5}{2}\sigma$以下	6	2.4	3	−3

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

將觀察值排列起來,以9.205作爲σ。

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

由此可知，常態曲線實爲該一羣類之適宜代表。

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

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

總平均數爲 35.0°，而此80個平均數所成之羣類，標準差爲1.628°。

下列之表，依上例之計劃，將觀察之分配，以與差誤常態曲線比較。

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

八件之差額在一標準差以下，餘二件之差額則略超過之。

平均數35.0°，既從 $80\times25=2000$ 緯度樣本中得出，標準差為$\frac{1.628}{\sqrt{80}}$度 $=.18$ 度，故平均數之第一位小數，並不能確切求得。

第四節　加權總和及加權平均之絕對差誤

前於第三章第一節曾論及，如 H + E，係從 n 個頻數羣類，獨立抽出 n 個數量之總和，（在此 n 個頻數羣類中，平均數為${}_1\bar{u}$，${}_2\bar{u}$，………${}_nu$，標準差為 σ_1, σ_2，………σ_n，而 $H={}_1\bar{u}+{}_2\bar{u}+\cdots\cdots+{}_n\bar{u}$），則$H+E_t$——任一次抽選之總和——必等於 ${}_1u_t+{}_2u_t+\cdots\cdots{}_nu_t$，而其標準差為 s $(s^2=\sigma_1^2+\sigma_2^2+\cdots\cdots+\sigma_n^2)$。

將上文重為排比，可知如用加權總和(weighted sum) $H+E_t=W_1\cdot{}_1u_t+W_2\cdot{}_2u_t+\cdots\cdots+W_n\cdot{}_nu_t$（$W_1, W_2$……為常數），標準差變為

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

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

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

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

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

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

$$s^2=\sigma^2 S(W_t^2)$$

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

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

於是

$$S(W_t^2)=S(\bar{w}^2+2\bar{w}w_t+w_t^2)=n\bar{w}^2+2\bar{w}Sw_t+Sw_t^2=n(\bar{w}^2+\sigma w^2)\tag{58}$$

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

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

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

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

如此平均數未曾加權，則

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

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

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

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

$$=\text{平均值}\{N_te'_t+I_te_t\}^2$$

如將 e 之乘積省略，則

$$=N_t^2s'^2_t+I_t^2s_t^2。$$

N_tI_t 之總和之標準差必爲 s，而

$$s^2=S(N_t^2s'^2_t+I_t^2s_t^2)。$$

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

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

在調查中 $S(I_t^2s_t^2)=315\times10^5$，$S(N_t^2s'^2_t)=4\times10^8$，故N內

之差誤，不甚重要。$S(N_t)=4023$，$S(N_tI_t)=284,700$，至於一九一一年不付所得稅者中，除工人外，其平均收入，係七十一鎊，其標準差爲五鎊。

第五節　相對差誤

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

設 x 爲某數量之觀察值，該數量之眞值或平均數則爲 x'，且 $x=x'(1+e)$，於是 $e=\frac{x-x'}{x'}$ 即爲相對差誤或離中差。（註二）

(1)乘積及商數

假如 F_1，F_2 兩因子，互爲獨立，但測量時，因有差誤，故成爲 $F_1(1+e_1)$，$F_2(1+e_2)$。結果 e 即爲乘積 P 之相對差誤，因而 $P(1+e)=F_1(1+e_1)\cdot F_2(1+e_2)$，式中 $P=F_1F_2$……………(62)

如果 e 之乘積可以略去時，$e=e_1+e_2+e_1e_2=e_1+e_2$。

是以，如 σ，σ_1，σ_2 爲 P，F_1，F_2 之標準差，依照公式(34)，$\sigma^2=\sigma_1^2+\sigma_2^2$。

此一得數，可以引申使因子數達於任何之有限數，於是

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

x^4之差誤（n爲有限數），如用 $x^n(He)=\{x(1+e_1)\}^n$ 算來，

（至 e_1，乃爲 x 之差誤），

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

當删除其平方時，如以 σ 爲 x 之標準差，x^n之標準差爲$n\sigma$。

如 n 爲分數，此得數卽爲眞實。例如，立方根中之差誤，卽爲該數量差誤之三分一。如此，假如 1006 一數，誤書爲1000（其相對差誤爲.006），則立方根中之相對差誤，卽爲.002，但此乃以10爲根，並非$10.02=10(1+.002)$（約略數）也。

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

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

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

$\sigma_q{}^2=\sigma_1{}^2+\sigma_2{}^2$，此處 σ_q 乃爲 e 之標準差 $\cdots\cdots(66)$

設 e 爲乘冪，a^x，之差誤，如 a 爲已知，e_1 爲 x 中之差誤，

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

$$e=a^{xe_1}-1 \qquad \text{如將 } e_1^2 \text{ 略去，則}$$

$$=e_1.x\log a \cdots\cdots(67)$$

就一般情形言之，e 爲一函數，$f(x)$，中之差誤，則

$$f(x)\times(1+e)=f\{x(1+e_1)\}=f(x)+e_1xf'(x)+\cdots\cdots$$

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

(2) 平均數中之差誤

設 m 爲 n 個數量 M_1，M_2，$\cdots\cdots M_t\cdots\cdots M_n$ 之未加權平均

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

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

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

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

如 Sa，σ_t 爲 e，e_t 之標準差，則以公式(55)，

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

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

$$sa^2=\sigma^2.\frac{S(\bar{m}+m_t)^2}{n^2\bar{m}^2}=\sigma^2.\frac{\bar{m}^2+\sigma m^2}{n\bar{m}^2},(\because Sm_t=0)$$

同時 $$sa=\frac{\sigma}{\sqrt{n}}\sqrt{\left(1+\frac{\sigma m^2}{\bar{m}^2}\right)}\quad\cdots\cdots(70)$$

此公式與前論（公式60）加權平均數之絕對差誤時所用公式，完全相同。

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

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

(3) 加權平均數中之相對差誤

〔以統計學報(Statistical Journal), 1911-12號,第八十一至八十八頁一文爲根據〕

設 $\bar{m}_w=\dfrac{S(W_tM_t)}{SW_t}$, 而 M_t (以及 $\bar{m}, \sigma_m$ 等)意義同前, W_1, $W_2 \cdots\cdots W_t \cdots\cdots W_n$ 爲權數。

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

於是 $S(W_tM_t)=n\bar{w}\bar{m}_w$。

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

設 M 中之差誤,如故,而以 e 爲 $\bar{m}_w$ 中之差誤。

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

$$\therefore\ e=\frac{S\{W_t(1+\eta_t)M_t(1+e_t)\}\cdot SW_t-S(W_tM_t)\cdot SW_t(1+\eta_t)}{S(W_tM_t)\cdot S\{W_t(1+\eta_t)\}}$$

$$=\frac{S(W_tM_te_t)\cdot SW_t+S(W_tM_t\eta_t)SW_t-S(W_t\eta_t)\cdot S(W_tM_t)}{S(W_tM_t)\cdot SW_t},$$

略去 $e\eta$ 及 η^2

$$=\frac{S(W_tM_te_t)}{S(W_tM_t)}+\frac{S\{(W_tM_t\cdot n\bar{w}-W_t\cdot n\bar{w}\bar{m}_w)\eta_t\}}{S(W_tM_t)n\bar{w}}$$

$$=\frac{S(W_tM_te_t)}{n\bar{w}\bar{m}_w}+\frac{S\{W_t(m+\bar{m}-\bar{m}_w)\eta_t\}}{n\bar{w}\bar{m}_w}\cdots\cdots\cdots\cdots(71)$$

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

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

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

$e=\dfrac{S(W_tM_te_t)}{n\bar{w}\bar{m}_w}+\dfrac{S(W_tm_t\eta_t)}{n\bar{w}\bar{m}_w}$,此爲約略之數,且須將$\bar{m}-\bar{m}_w$之差删卻,如將全式寫出,則至第二項之分子爲$S\{W_t(M_t-\bar{m}_w)\eta_t\}$。

設 ∂, σ_t, σ_t' 爲 e, e_t, η_t 之標準差。

於是 $\partial^2=\dfrac{1}{(n\bar{w}\bar{m}_w)}\{S(W_tM_t)^2+S[W_t(M_t-\bar{m}_w)\sigma_t']^2\}$ (73)

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

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

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

其餘有關數量,可由觀察算得。在平常情形下,求得此一得數之良好近似值,乃爲

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

此近似值,求法如下:——

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

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

$$+Sw_t^2(m_t^2-\sigma_m^2)+4\bar{m}\bar{w}Sw_tm_t+2\bar{w}Sw_tm_t^2+2\bar{m}Sm_tw_t^2$$

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

$$+\frac{2\sigma_w{}^2\sigma_m r_{12}}{\overline{w}^2\overline{m}}$$

式中

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

$$S\{W_t(M_t-\overline{m}_w)\}^2=S\{W_t(m_t+\overline{m}-\overline{m}w)\}^2$$

$$=SW_t{}^2m_t{}^2+2(\overline{m}-\overline{m}_w)SW_t{}^2m_t+(\overline{m}-\overline{m}_w)^2SW_t{}^2$$

$$=\overline{w}^2\cdot n\sigma_m{}^2+2\overline{w}Sw_tm_t{}^2+Sw_t{}^2m_t{}^2-\frac{2}{\overline{w}}Sw_tm_t(2\overline{w}Sw_tm_t+$$

$$Sw_t{}^2m_t)$$

$$+\left(\frac{Sw_tm_t}{n\overline{w}}\right)^2n(\overline{w}^2+\sigma_w{}^2)$$

$$\therefore \frac{S\{W_t(M_t-\overline{m}_w)\}^2}{n(\overline{w}\overline{m})^2}=\frac{\sigma_m{}^2}{\overline{m}^2}+\frac{2\sigma_w\sigma_m{}^2}{\overline{w}^2\overline{m}^2}\cdot r_{12}$$

$$+\frac{\sigma_w{}^2\sigma_m{}^2}{\overline{w}^2\overline{m}^2}\left(R_{22}+1\right)-4\frac{\sigma_w{}^2\sigma_m{}^2}{\overline{w}^2\overline{m}^2}r^2-2\frac{\sigma_w{}^3\sigma_m{}^2}{\overline{w}^3\overline{m}^2}r\cdot r21+\frac{\sigma_w{}^2\sigma_m{}^2}{\overline{w}^2\overline{m}^2}\left(1+\frac{\sigma m^2}{\overline{w}^2}\right)r^2$$

$$\therefore \bar{\sigma}^2=\frac{1}{n}\left(\frac{\overline{m}}{\overline{m}_w}\right)^2(l_1{}^2\sigma^2+l_2{}^2\sigma'^2)$$

式中

$$l_1{}^2=\left(1+\frac{\sigma w^2}{\overline{w}^2}\right)\left(1+\frac{\sigma m^2}{\overline{m}^2}\right)+4\frac{\sigma w}{\overline{w}}\cdot\frac{\sigma m}{\overline{m}}r+2\cdot\frac{\sigma w}{\overline{w}}\cdot\frac{\sigma m^2}{\overline{m}^2}r_{12}+2\frac{\sigma w^2}{\overline{w}^2}\cdot\frac{\sigma m}{\overline{m}}r_{21}+\frac{\sigma w^2}{\overline{w}^2}\cdot\frac{\sigma m^2}{\overline{m}^2}R_{22}$$

$$2^2=\frac{\sigma_m{}^2}{\overline{m}^2}\left\{\left(1+\frac{\sigma_w{}^2}{\overline{w}^2}\right)+\left(\frac{\sigma_{w4}}{\overline{w}_4}-3\frac{\sigma_{w2}}{\overline{w}_2}\right)r^2+2\frac{\sigma_w}{\overline{w}}r_{12}-2\frac{\sigma_{w3}}{\overline{w}_3}r\cdot r_{21}+\frac{\sigma_w{}^2}{\overline{w}^2}R_{22}\right\}$$

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

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

第六節　例證(一)

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

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

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

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

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

五金業

（工程及造船業除外）

一九〇六年

門類	僱工人數 W 以千爲單位	男工平均收入 M 先令	便士
生鐵業	14	34	4
鋼鐵業	54	39	1
馬口鐵業	11	42	0
鐵道車輛業	46	30	9
鑄鐵業	12	31	4
電氣器料業	15	34	7
金屬線業	8	35	7
銅業	8	31	9
金銀業	8	36	6
寶石業	3	38	0
利刃（刀剪）業	3	31	2
献鑄業	8	31	5
自轉車業	7	34	4
水管業	7	28	3
製釘業	5	31	0
牀架業	2	36	3
蹄鐵業	2	27	9
科學用具業	2	36	10
針業	2	31	9
鏈環業	1	35	4
鎖　業	1	28	0
鐘表業	1	32	7
活字鑄造業	1	33	3
雜類	45	32	5
總計	266		

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

二十四項收入之算術平均數，$\bar{m}=33$先令$.6\frac{1}{8}$便士$=33.511$先令。

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

$$r=\frac{Swm}{n\sigma w\sigma_m}=\cdot150,\ r_{12}=\frac{Swm^2}{n\sigma w\sigma m}=\cdot095,\ r_{21}=\frac{Sw^2m}{n\sigma w^2\sigma_m}=\cdot280,$$

$$R_{22}=\frac{Sm^2w^2}{n\sigma m^2\sigma w^2}-1=\cdot264。\ \left(\frac{\sigma m}{\bar{m}}\right)^2=\cdot011,\ \left(\frac{\sigma w}{\bar{w}}\right)^2=1\cdot77,$$

$$\frac{\sigma m}{\bar{m}}=\cdot104,\quad \frac{\sigma w}{\bar{w}}=1\cdot33。$$

以各行業之人數爲權數，則收入平均數，$\bar{m}_w=34$ 先令$2\frac{1}{2}$便士，$=\bar{m}\left(1+\frac{r\sigma m\sigma w}{\bar{m}\bar{w}}\right)$；$\therefore\left(\frac{\bar{m}}{\bar{m}\bar{w}}\right)^2=\cdot959$。

現依前節求近似值（小體字）之標號，

$$\therefore\ l_1^2=2\cdot77\times1\cdot011+4\times1\cdot33\times\cdot104\times\cdot150+2\times1.33$$
$$\times\cdot011\times\cdot095+2\times1\cdot77\times\cdot103\times\cdot280+1\cdot77\times\cdot011$$
$$\times\cdot264=2\cdot80+083+\cdot003+\cdot102+\cdot005=2\cdot99。$$

$l_1^2\times\left(\frac{\bar{m}}{\bar{m}w}\right)^2=2\cdot87$。依求近似值公式，得2.80。

$$l_2^2=\cdot011\{2\cdot77+(3\cdot13-3\cdot99)\times\cdot0225+2\times1\cdot33\times\cdot095$$
$$-2\times2\cdot35\times\cdot150\times\cdot280+1\cdot77\times\cdot264\}$$
$$=\cdot011\{2\cdot77-\cdot020+\cdot253-\cdot198+\cdot467\}=\cdot011(2.77$$
$$+\cdot50)=\cdot036。$$

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

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

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

至權數上之差誤，因旣以男女工人總數爲權數，而非男工人數，於理顯有不合，故必甚大。

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

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

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

(2) 加權平均最重要之用途，莫過於物價指數。

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

取孫巴克氏之一九〇〇及一九一一各年商品物價指數，再改用一九〇〇年爲基期。例如英麥價格，以一八六七至一八七七年爲一〇〇，則一九〇〇年爲四九，一九一一年爲五八。如以一九〇〇年爲一〇〇，則一九一一年爲一一八。數目共有四十五個，其算術平均數爲107.82，但孫巴克所得平均指數，在一九〇〇爲

75.07,在一九一一年爲79.69,成100:106·16之比。

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

$$I=100\frac{sp'}{sp}=100\frac{sp\cdot\frac{p'}{p}}{sp};$$

詳言之,卽各項變動情形之比率,用一九〇〇年各項指數作爲權數也。

現請就孫氏指數,研究其平均數之確度。以p寫作w,是爲權數，以$\frac{p'}{p}$書爲m,是爲量數。玆將有關數量列下:——

$\bar{w}=75\cdot07$,　$\sigma_w=20\cdot67$,　$\frac{\sigma_w}{\bar{w}}=\cdot275$,

$\bar{m}=107\cdot82$,　$\sigma_m=20\cdot03$　$\frac{\sigma_m}{\bar{m}}=\cdot186$,

$r=-\cdot2944$,　$\bar{m}_w=106\cdot2$,　$r_{12}=\cdot506$,

$r_{21}=\cdot347$,　$R_{22}=\cdot936$。

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

$$\sigma^2=\frac{\sigma^2}{45}(1\cdot076)(1\cdot035)+\frac{\sigma_1^{2'}}{45}(1\cdot076)\times(\cdot186)^2=\sigma^2\times\cdot025+\sigma^2_1\times.00083。$$

如包括在內，則 $\bar{\sigma}^2=\sigma^2\times\cdot024+\sigma'^2+\cdot0012$。

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

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

$$\sigma^2\times\cdot024=(\cdot0005)^2,$$

如此之微，雖省略亦無傷也。

w之差誤，如商品輕重，有一定標準，則亦可以算出。非然者，卽假定本應加同等權數，如上交替演算之例是也。於是$\frac{\sigma_w}{w}=\cdot275$，可將實在權數距假定眞權數之離散度求出，而 $\sigma'^2\times\cdot0012=(\cdot275\times\cdot034)^2=(\cdot0093)^2$。

是以$\sigma^2=(\cdot0005)^2+(\cdot0093)^2=(\cdot0093)^2$(約略數)，至於指數：可書如下式：

$$106\cdot2(1\pm\cdot0093)=106\cdot2\pm1,$$

由此式可示吾人用指數時，應有之伸縮範圍。

實際上，根據兩假定算出指數之差，爲

$$107\cdot8-106\cdot2=1\cdot6。$$

第七節　平均數之比較

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

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

玆取一簡單商數 $Q=F_1\div F_2$ 爲例。

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

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

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

加權與不加權平均數比率之必要分析，可參閱附錄七及附錄八。

求近似值之公式如下，符號同上節。

如以 s_r 代表二不加權平均數比率之標準差，

$$s_r^2=\frac{1}{n}\sigma d^2\left(1+\frac{\sigma m^2}{\bar{m}^2}\right) \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (76)$$

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

然若以 s_r 代表二加權平均數比率之標準差，則在某種條件下，其近似值爲

$$s_r^2=\frac{1}{n}\left(1+\frac{\sigma w^2}{\bar{w}^2}\right)\left\{\sigma d^2+\left(\frac{\sigma u}{1+\bar{u}}\right)^2\left(\sigma^2+\sigma'^2\right)\right\} \cdots\cdots\cdots\cdots (77)$$

式中 σd 仍舊，σ 爲數量中差誤之標準差，σ' 係權數中差誤之標準差，又

$$M_t'.=(1+\bar{u}+u_t)M_t$$

但 $Su_t=0$，故 $1+u$ 足以測量數量平均增長率，而 σu 爲 u 之標準差，可用以測量增長率之擴散度。

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

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

低值也。

第八節　例證（二）

英國產業工人平均每週工資變動情況

	1880 人數 單位一千人 W	1880 工資 單位先令 M	1900 人數 單位千人 W′	1900 工資 單位先令 M′	M之增加率 $1+\bar{u}+u$
農業					
英格蘭及威爾士……	135	15	120	16.2	1.08
蘇格蘭……………	24	18	20	21.2	1.18
愛爾蘭……………	98	9	86	10.4	1.16
建築業……………	84	27	123	31.0	1.15
印刷業……………	8	31	13	32.9	1.06
造船業……………	7	28.5	13	34.8	1.22
工程業……………	72	25	106	30.5	1·22
煤礦業……………	44	23	75	34.3	1.49
泥水業……………	9	31	11	38.1	1.23
棉　業……………	52	16	54	19.5	1.22
紡毛業……………	12	14	12	13.6	.97
毛線業……………	12	14	12	14.4	1.03
煤氣業……………	3	27	8	31.0	1.15
木器業……………	12	23	18	24.8	1.08
	572		671		

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

材料算得。至各比率之確度較 M 確度，尤當重視。

$$n=14,\ \bar{m}=21\cdot54,\ \bar{m}'=25\cdot20,\ \bar{m}_w=18\cdot69,\ \bar{m}'_w=24\cdot09$$

$$\frac{\sigma_m}{\bar{m}}=\cdot319,\frac{\sigma_m'}{\bar{m}'}=\cdot351,\frac{\sigma_w}{\bar{w}}=1\cdot00,\frac{\sigma_w}{\bar{w}'}=\cdot90,\bar{u}'=\cdot160,\sigma_u=\cdot12$$

$$r=-\cdot42,\ r_{21}=\cdot44,\ r_{21}=-\cdot42,\ R_{29}=.25。$$

不加權平均數之比率爲$\frac{\bar{m}'}{\bar{m}}=1\cdot170$；加權平均數之比率 $=\frac{\bar{m}_w'}{\bar{m}_w}=1\cdot288$。

務德氏得出之不加權與加權平均數（用各種權數）比率爲 $\frac{100}{86}=1\cdot163,\frac{100}{82}=1\cdot219$。

$$sr^2=\frac{1}{14}\left(1+1\cdot00^2\right)\left\{\sigma_d^2+\left(\frac{\cdot12}{1.160}\right)^2\left(\sigma^2+\sigma'^2\right)\right\}$$

$$=\cdot143\sigma_d^2+\cdot0015(\sigma^2+\sigma'^2),$$

根據求近似值公式(77)，暨求完全值公式(148)。得出約略數如下：

$$sr^2=\cdot145\sigma_d^2+\cdot022\sigma^2+\cdot0035\sigma'^2+\cdot016\sigma'_d{}^2,$$

σ'_d爲兩期權數中差誤相減餘額之標準差。

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

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

然則$s_r^2 = \cdot 000091 + \cdot 000055 + \cdot 000035 + \cdot 000040 = \cdot 00022$。

$$s_r = \cdot 015$$

平均數之比率,可寫作下式:

$$\frac{\overline{m}'_w}{\overline{m}_w}(1 \pm \sigma_r) = 1.288 \pm \cdot 020$$

詳言之,增加百分數,並非爲二九,乃在二七至三一之間也。

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

第九節　平均數與平均數間差額之重要

實用統計上時常發生一極端重要之問題, 即決定兩相似羣類平均數之差額之問題也。此相差之額, 或由於觀察上之差誤——尤其在隨機抽樣所包個體過狹時 —— 抑完全歸因於特質之不同,皆應加以決定。例如,有兩類觀察得來之死亡率,一爲千分之14·7,一爲千分之 14·3, 吾人何以處之,將謂第一類之死亡率高耶,抑將認爲實有 ·4 相差,而逕將人口妄分兩類耶?

設觀察得來者,較在機會均等之抽選中所期望者,差額甚大,則知此一差額,甚爲重要,換言之,現象間之眞正差額,關係頗爲

重要也。

一般分析方法，不外下述：設有兩類，一類有事物 n_1 個，一類有事物 n_2 個，平均數一爲 $\bar{x}_1$，一爲 $\bar{x}_2$。自大範圍中任意抽出 n_1 及 n_2 個事物，將二者之平均數分別算出，卽由此平均數之相減差額，組成頻數曲線，然後計算此曲線之標準差，設已算得之標準差爲 σ。

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

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

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

例A．—— 具有特質之事物在範圍中所佔之比例

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

將 n 個隨機抽選出來，設其中 $p'n$ 個，爲具有該種特質者。

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

例如——取一骰連擲一千二百次，六點出現有一百八十次，

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

$$\frac{p-p'}{\sigma}=\frac{\cdot 0167}{\cdot 0108}=1\cdot 6,$$

於此不能不聲明者，骰子之六面，並無有欠一律之證明。

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

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

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

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

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

例如——從一小城市選出 1000 家，其中二百(n_1)家，家主爲技師，八百(n_2)家，家主爲勞動者。有學齡兒童者在第一類($p_1=.4$)中有八十家，在第二類($p_2=\cdot525$)中有四百二十家（家數係假定的）。

$$p\times1000=80+420\quad\therefore p=\tfrac{1}{2}=q$$

$$p_1\sim p_2\text{ 之標準差}=\sqrt{\left\{\frac{1}{4}\left(\frac{1}{200}+\frac{1}{800}\right)\right\}}=\cdot04$$

$$\frac{p_2-p_1}{\sigma}=\frac{\cdot525-\cdot4}{\cdot04}=3\text{（約略數）。}$$

此差額卽爲重要。

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

例如——從兩國共抽出一千人，其中有藍眼睛者，一國爲三百人，一國爲二百五十人。

此次抽選，$p_1=\frac{3}{10}$，其$\sigma=\sqrt{\left(\frac{3\times7}{10^5}\right)}=\cdot014$，而該國（設抽選時不問該國種族氣候如何）全國碧睛人數，約近於 p_1。同理，另一國家之數，約略爲 $p_2=\frac{1}{4}$。

p_1-p_2 之標準差，兩獨立羣類相減餘額之標準差，卽：

$$\sqrt{\sigma_1^2+\sigma_2^2}=\sqrt{\left(\frac{p_1q_1}{n_1}+\frac{p_2q_2}{n_2}\right)}=\cdot 02 \cdots\cdots\cdots\cdots (79)$$

$$\frac{p_1-p_2}{\sigma}=\frac{3-\cdot 25}{\cdot 02}=2\cdot 5。$$

此法一般於比較兩類職業工人（如礦工及瓦匠）之死亡率時多用之。

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

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

此種礦工，假定係從所有礦工（之範圍中）任意抽樣而來，瓦匠亦然。則

$$r_1-r_2\text{ 之標準差}=1000\sqrt{\frac{p_1q_1}{n_1}+\frac{p_2q_2}{n_2}}=1000\sqrt{\left\{\frac{m_1(n_1-m_1)}{n_1^3}+\frac{m_2(n_2-m_2)}{n_2^3}\right\}}。$$

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

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

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

設一範圍在 t 層中有事物 n_1，n_2，……n_t 個，並設具有某種特質者，在 t 層中，爲 p_1, n_1, p_2n_2，……p_tn_t。

$N=n_1+n_2+\cdots\cdots n_t$，並以 $p_1, n_1, +p_2n_2+=PN$。

設 t 層中，所檢查之事物爲 kn_1，$kn_2\cdots\cdots kn_t$，總之，$KN=n$。

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

樣本中 p_1 之標準差爲 $\sqrt{\frac{p_1q_1}{kn_1}}$，$p_2$ 之標準差及其他均以此類推。

如此，設 σ 爲樣本中 P 之標準差，則依公式 55，

$$\sigma^2=\left(\frac{n_1}{N}\right)^2\frac{p_1q_1}{kn_1}+\left(\frac{n_2}{N}\right)^2\frac{p_2q_2}{kn_2}+\cdots=\frac{1}{kN^2}(n_1p_1q_1+n_2p_2q_2+\cdots)$$

$$\therefore kN^2\sigma^2=n_1p_1(1-p_1)+\cdots\cdots$$

$$=S(np)-S(np^2)=NP-S\{n(P+d)^2\}$$

$$=NP-NP^2-2P\cdot Snd-Snd^2=NPQ-N\sigma_p^2$$

$$\sigma_p^2=\frac{1}{N}(n_1d_1^2+n_2d_2^2+\cdots)$$

$$\therefore \sigma^2=\frac{PQ}{n}-\frac{\sigma_p^2}{n}\cdots\cdots(80)$$

所謂 σ，乃觀察結果之標準差，P 爲在範圍中所估之實在比

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

如從範圍中任意抽取數目，則差誤之標準差，將為 σ_0，而 $\sigma_0^2=\frac{PQ}{n}$。

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

茲欲調查四城鎮之經濟狀況（見生計與貧乏一書），每街每二十戶中，提出一家，而不用將所有各戶排號任意從每二十號中各取其一之辦法。因吾人如此辦理，不致有某區未被代表之弊，而此弊則隨機抽樣乃不可免者也。且用此辦法，上列公式所提之優點，可以完全得到，因各街市之社會情形，多有某種相似之處也。茲設十個相等區域共有一萬六千戶，且設各區在某種標準之下之比例為 ·02, ·06, ·10……·38。於是 $N=16000$; $n_1=n_2=\cdots=1600$; $p_1=\cdot02$, $p_2=\cdot06\cdots P=\cdot2$; $d_1=\cdot18$, $d_2=\cdot14\cdots$;

$$\sigma_p^2=\frac{1}{10}(\cdot18^2+\cdot14^2+),\sigma_p=\cdot115$$

茲設每區各擇八十戶而調查之，$n=800$, $k=\frac{1}{20}$, $\sigma^2=\frac{\cdot2\times\cdot8}{800}-\frac{\cdot0132}{800}$, $\sigma=\cdot0136$，於是得數乃為 ·20±·0136，或百分之 20±1·36。

以上乃就分層而言，如抽選方法，並不依各層排列抽選，則

應得$\sigma=\cdot0141$。確度之增加，爲數無幾，然分層抽取方法，乃基於常識，苟非有何困難，當以用此法爲是。

例B.——一範圍中含有若干可量物件之例

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

從中任意抽出n個，其平均數爲$\bar{x}_1$。

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

實例——某城市有住宅10,000所，每所平均住人4·5，標準差爲2。一千所工人住宅之平均數爲4·7。是則$N=10,000$，$n=1000$，$\bar{x}=4\cdot5$，$\bar{x}_1=4\cdot7$，$s=2$。

$$\sigma=2\sqrt{\left(\frac{1}{1000}-\frac{1}{10000}\right)}=\cdot06,\ \frac{\bar{x}_1-\bar{x}}{\sigma}=\frac{\cdot2}{\cdot06}=3\cdot3。$$

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

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

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

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

自原點算出之動差，則得

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

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

但 $\bar{x}\sim\bar{x}_1$ 對 $\bar{x}_1\sim\bar{x}_2$ 之比，乃爲常數，卽等於$\frac{n_2}{n}$。

$\therefore$ $\bar{x}\sim\bar{x}_1$ 之標準差爲 σ，而

$$\sigma^2=\frac{n_2^2}{n^2}\left(\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right)=\frac{\sigma_1^2}{n_1}+\frac{s^2-2\sigma_1^2}{n}-\frac{n_1}{nn_2}(\bar{x}-\bar{x}_1)^2,$$

將 σ_2 及 x_2 消去，卽得出。

設 n_1 小於 n_2，而 n 及 n^2 爲甚大。

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

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

（見 Biometrika，卷五，第一百八十二頁）。

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

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

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

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

實例——蘇格蘭人口普查

結婚總數 $n=133,960$。

$\bar{x}$,每對夫婦所生子女數,$=5 \cdot 82$,其 $s=3.099$。

鍋匠之生育,$n_1=923$, $\bar{x}_1=6 \cdot 00$, $\sigma_1=3 \cdot 069$。

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

$\dfrac{\bar{x}_1-\bar{x}}{\sigma}=\dfrac{\cdot 18}{\cdot 10}=1 \cdot 8$,差額顯然重要。

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

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

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

$\dfrac{\bar{x}_1-\bar{x}}{\sigma}=\dfrac{\cdot 123}{\cdot 018}=7$ (約略數),可知北大西洋各州麵粉價格確比美國全國之平均爲高。

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

$\bar{x}_1 \sim \bar{x}_2$ 之 σ 乃爲

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

此係兩個獨立觀察之差額,此兩個獨立觀察各由例(1)而來。

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

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

此外尚有他種變量(variant)，亦可用此原則解決之。

實例——生活費指數中之食物費用，與第三章第七節所列者相似，在全部羣類中其 $\bar{x}=10\cdot3$（先令），$s=3\cdot3$。

有技能工人五百六十六人之家庭，$\bar{x}_1$ 爲 $10\cdot9$，無技能工人二百六十六人家庭之 $\bar{x}_2$，爲$9\cdot3$。

此兩類之標準差，並未算得，但大概與 s 相近似。

$$\bar{x}_1-\bar{x}_2 \text{ 之 } \sigma \text{ 爲 } 3\cdot3\sqrt{\left(\frac{1}{566}+\frac{1}{266}\right)}=\cdot25。$$

$\dfrac{\bar{x}_1-\bar{x}_2}{\sigma}=\dfrac{1\cdot6}{\cdot25}=6$（約略數），差額重要可知。

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

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

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

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

於是　$Snd=0$

$$Ns^2=n_1(s_1^2+d_1^2)+n_2(s_2^2+d_2^2)+\cdots\cdots$$

s_1s_2……之標準差之平方爲

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

以 σ 爲 D 之標準差。

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

$$\therefore \quad D = \frac{n_1}{N}\delta_1 + \frac{n_2}{N}\delta + \cdots\cdots$$

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

依公式55，
$$= \frac{1}{kN^2}(n_1 s_1^2 + n_2 s_2^2 + \cdots\cdots)$$

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

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

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

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

$$於是 \quad \sigma^2 = \sigma_0^0 - \frac{\sigma_m^2}{n} \cdots\cdots\cdots\cdots\cdots\cdots \quad (83)$$

平均數之確度，公式80所不能精確者，玆用成層排列法可以推進矣。

例如本章第二節之例，以 N 爲最末七條街（自斯皮特菲爾街以上）之住宅總數。在此七條街每家住宅之平均住人數，$\bar{x}$，在普查報告所得爲 4·64，s 爲 2,750 而住宅抽樣，係五十取一，

故 $k=\frac{1}{50}$;$N=57000, n=1140$。

	住宅數	每宅人數 x
斯皮特菲爾街	$n_1=65^{00}$	$+.15=d_1$
懷德加白爾街	$n_2=59$	$+.08=d_2$
聖喬治街	$n_3=94$	$+.24=d_3$
色得外爾街	$n_4=48$	$-.27=d_4$
萊毋好斯街	$n_5=66$	$-.10=d_5$
邁爾恩德街　西南區	$n_6=134$	$+.07=d_6$
邁爾恩德　東北區	$n_7=104$	$-.24=d_7$
	570	

$$S(nd2)=1806 \qquad \sigma m^2=\frac{1806}{57000}=.0317$$

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

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

以各區爲層,從七層中,用抽樣方法所得之數,雖略有進步,但厥值甚微。

第十節　趨勢之存在

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

此種方法及其難點,舉二例以明之:

(1) 一八五〇至一八九九年賽馬時間記錄如后

	分,秒			分,秒			分,秒			分,秒			分,秒	
1850	2	56	1860	2	56	1870	2	52	1880	2	49	1890	2	$40\frac{4}{5}$
1851	2	52	1861	2	44	1871	2	51	1881	2	46	1891	2	$54\frac{3}{5}$
1852	3	0	1862	2	49	1872	2	52	1882	2	49	1892	2	$43\frac{1}{5}$
1853	2	52	1863	2	54	1873	2	$50\frac{3}{4}$	1883	2	53	1893	2	$44\frac{4}{5}$
1854	3	0	1864	2	47	1874	2	$48\frac{1}{5}$	1884	2	49	1894	2	50
1855	2	58	1865	2	51	1875	2	$49\frac{1}{2}$	1885	2	$43\frac{3}{5}$	1895	2	$48\frac{3}{5}$
1856	3	4	1866	2	53	1876	2	50	1886	2	$54\frac{2}{5}$	1896	2	$45\frac{1}{5}$
1857	2	50	1867	2	54	1877	2	$54\frac{1}{2}$	1887	2	$50\frac{3}{5}$	1897	2	45
1858	2	$53\frac{1}{2}$	1868	2	$47\frac{1}{2}$	1878	2	54	1888	2	$42\frac{4}{5}$	1898	2	$45\frac{4}{5}$
1859	2	55	1869	2	59	1879	3	2	1889	2	45	1899	2	44
十年平均數	2	56•05		2	51•45		2	52•395		2	48.22		2	46.26

以上數字恰配合常態曲線,其平均數爲2分50•87秒,標準差爲5•20秒。兩記錄差額之標準差,乃爲$5\cdot2\sqrt{2}=7\cdot4$秒。超過此額者,數年間只有十一次,而差額絕無超出兩倍此數者;足見每兩次賽馬之間,並無猛烈變動之事。但先年與最末幾年,時間相差,有時卻達20秒。如以十年爲一期,則兩期平均數相差額之標準差,爲$52\sqrt{\left(\frac{1}{10}+\frac{1}{10}\right)}=2\cdot33$秒。一八五〇至一八五九年之平均數及一八九〇至一八九九年之平均數,相差將及十秒之多,與一八五〇至一八五九年及一八八〇至一八八九年之平均數之差額相同,差額均甚爲重要。至中間數年之相差,尙不爲巨。由此可知自一八五〇至一八八〇年之間,必有幾種原因,乃使賽馬逐漸加快也。

(2) 英格蘭威爾士一八六〇至一九〇九年間之結婚率

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

此五十年間平均數爲15·80，五十個記錄之標準差爲·89，且如不計次序，幾成常態形式。各年間，並無相差甚鉅之情形。取十年記錄，作一平均，兩平均數間差額之標準差爲0.4，可知一八七〇至一八七九至一八八〇至一八八九年之下降，及一八八九後之上升，甚爲重要。

第一個二十五年，較第二個二十五年間之變動爲劇，玆作更精密之測驗如下：

1860－1884 $\sigma=\cdot894$ $\sigma\sqrt{\frac{2}{5}}=\cdot566$

	平均數
1860－4	16·70
1865－9	16·70
1870－4	16·96
1875－9	15·70
1880－4	15·22

1885－1909 $\sigma=\cdot613$ $\sigma\sqrt{\frac{1}{2}}=\cdot388$

	平均數
1885－9	14·50
1890－4	15·24
1895－9	15·88
1900－4	15·76
1905－9	15·34

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

論辯必須以一圖式解明之，蓋惟有圖式乃表明在何期間始可施用測驗也。

第十一節　週期性

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

取一旣知距度(interval)爲例，譬有按月之記錄，共一年，試查正月份二月份之相差，是否與隨機觀察出現（不計時間先後）者相同。茲設記錄每年十二次積至 t 年之久，則應共有記錄 $12\times t$ 之多，又設其平均數爲 $\bar{x}$，如以 x 代表任何觀察，則標準差（依平均數算來）$\sigma=\sqrt{\left(\frac{s(x-\bar{x})^2}{12t}\right)}$。

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

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

如此記錄作爲一羣類而幾近常態形式，或與第三章第五節所舉條件相符，則超過上列標準差之機率，卽可由第二章第四節（第三表）常態機率表查出。

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

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

雖然，此一方法不易盡行包括所有之證據。例如查閱前編第七章第四節之一百八十個失業記錄，可知平均數為 4·269，標準差為1·924。以15個作一平均，則兩平均數差額之標準差為1·924 $\sqrt{\dfrac{2}{15}}=\cdot 70$。以正月或十二月份之平均數，與四月；五月，六月，或七月之各該平均數比較，雖超過·70，但差額尚不算大；即隨機抽樣得來者，亦不過如此。但此外尚有累積證據(cumulative evidence)，甚難加以測量。例如，自十二月經由正月二月三月（如將一九一二年之反常記錄取消）四月以至五月之平均數之低落，以及自五月以後至十月，按月之平均數之上升是。此乃波浪形運動(wave motion)，非此法所能測量者也。

另有一法，係比較某月對其下月之下降上升數，然欲求其精確，卻亦甚難。

	正月 至 二月	二月 至 三月	三月 至 四月	四月 至 五月	五月 至 六月	六月 至 七月	七月 至 八月	八月 至 九月	九月 至 十月	十月 至 十一月	十一月 至 十二月	十二月 至 正月
低落數	12	13	10	10	$5\frac{1}{2}$	7	2	$7\frac{1}{2}$	$8\frac{1}{2}$	10	15	10
上升數	3	2	5	5	$9\frac{1}{2}$	8	13	$7\frac{1}{2}$	$6\frac{1}{2}$	5	0	4

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

驗中，各次之正負，機會一律均等，取得 10 以上而具同符號之機率約為 $\frac{1}{9}$，足見三月至四月，四月至五月，十月至十一月之運動，即在隨機抽樣中，亦不見得即無實現之機會。但12以上具有同號者，機率僅有$\frac{1}{28}$，則自正月至二月，二月至三月，七月至八月，十一月至十二月，如不受季節影響，該種運動實現必甚難。

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

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

取一月之平均數，與全部之平均數相比，二者差額之標準差為 $108\sqrt{\left(\frac{1}{18}+\frac{1}{216}\right)}$，即約為 26·5，而一月之平均數與另一月之平均數，相減差額之標準差，為 $108\sqrt{\frac{2}{18}}=36$，以上兩種試驗，均以隨機抽選及無季節影響為前提。

記錄上之平均數

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

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

於平均數約比兩倍 27 猶多，六月及十一月大於平均數只在一27之內。於此得一結論；七月至十月一季，比十二月至四月之季爲佳。

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

試從原始材料中，得

	一月至二月	二月至三月	三月至四月	四月至五月	五月至六月	六月至七月	七月至八月	八月至九月	九月至十月	十月至十一月	十一月至十二月	十二月至一月
下降數	$5\frac{1}{2}$	10	16	4	4	7	4	11	1	11	15	8
上升數	$12\frac{1}{2}$	8	2	14	14	11	14	7	7	7	3	9

由此可知多於11或少於 7 之數，頗爲重要。

附註：——

(1)因以未知眞實平均數爲中心算得之離中差，與由觀察之平均數算出者不同，故平均數之標準差時常寫作$\frac{\sigma}{\sqrt{n-1}}$，而不用第三章第一節之$\frac{\sigma}{\sqrt{n}}$。

設以 x_0 爲一羣類之眞實平均數，其標準差爲 σ_0，又設從中抽出 n 個事物，而得平均數 x，至 n 個事物各別分爲 X_1, X_2，……而其標準差爲 σ。

以$x=x_0+d$

X_1, X_2……對眞實平均數之離中差，爲 X_1-x_0, X_2-x_0……，其標準差依假設當爲σ_0。

故　$\sigma_0^2=$平均 $(X-x_0)^2=$平均 $(X-x+d)^2=$平均$(X-x)^2+d^2$

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

故 $=\sigma^2+\dfrac{\sigma_0^2}{n}$

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

而且 $\dfrac{\sigma_0}{\sqrt{n}}=\dfrac{\sigma}{\sqrt{n-1}}=\sqrt{\left\{\dfrac{S(X-\bar{x})^2}{n(n-1)}\right\}}$

此處 σ 須以 $\sqrt{n-1}$ 除而非以 $\sqrt{n}$ 除。

此種改正只有原理上的重要性，因其差額只於 n 爲極小時方能察覺，而且 σ 與此差額，無論何時，均易陷于同次之差誤。

(2) 前論及時間數列時，曾言以 t 年之平均數爲中心，測量時間數列之觀察離中差，必須注意平均時有無陷於差誤之危險（此差誤用 $\dfrac{\sigma}{\sqrt{t}}$ 測量之，而 σ 卽爲鄰近各年份之觀察數値之標準差）。觀察與此一平均數之差額之標準差，非爲 σ，乃係 $\sqrt{\left(\sigma^2+\dfrac{\sigma^2}{t}\right)}=\sigma\sqrt{\dfrac{t+1}{t}}$。因 t 小，差誤易於察覺，前編第七章第三節，結婚率與平均每人擔負進出口貨值之圖，所示之離中差，估計頗欠完善，於此可知，本編第七章第一節第六實例，所算相關，亦欠精確。

（註一） 此種抽選，未能完全滿足獨立條件，參閱 Statistical Journal, 1912–13, p. 702 nixon。

（註二） 在第一編第八章，卽以上文之 $\dfrac{x'-x}{x}$ 作差誤，此差誤如名之爲 e_1，則 e 與 e_1 二者之關係爲 $e_1=-e+e^2-e^3\cdots\cdots$，但普通大都假定，$e^2$ 値甚微，不妨略去，故 $e_1=-e$。

第五章　經驗頻數方程

所有頻數羣類，未必全能用大數律表現之。蓋方程式背後之特種繁複獨立原因，未可一概認爲適用於所有之觀察羣類也。常態曲線之主要用途，爲對於平均數及其他函數（其作成方法，業已知曉）之應用。至於對於人體測量的 (anthropometrical) 或生物測量的 (Biometrical) 羣類之應用，則須按量數類別，逐一加以檢證，而心理上及倫理上之性質，是否依常態形式而分佈，則須作專門之調查。雖然，吾人可得而認定者，在極多類別中，各類之中心部分（自中心計上下各一二標準差之間），亦頗合常態分配之形式，且觀察去平均數二標準差以上之機率，並不算大，故常態頻數表即在非常態情形下，亦足供吾人之參考。

爲求羣類之完全敍述，或用一較有伸縮性之辦法，以包進比差誤曲線較寬之各類，或用基於經驗得出之方程式，以配合特別觀察羣類。茲於本章擇其適於達到此（及其他）目的之方程式，略一討論之。

一般所用方法，係用一含有二，三，或四個未知常數之數學方程式。擇定常數之時，須方程所代表之曲線，能與由觀察構成之圖相配合；圖上點數既比常數數目爲多，於是方程式亦多於未

知數，乃不得不決定適當解法矣。普通解此難題方法，係用最小二乘法（見附錄九），惟遇觀察得來之頻數曲線時，一般多用皮爾生法；此法係將由曲線方程式中得出之動差，用數學推算方法，使與由觀察得來(本編第一章第四節)之動差相等。其實，前論由觀察材料之前三級動差，以求平均數，標準差，及偏斜度($\bar{x}, \sigma, k$)時，即已曾用此法，下節所論曲線系 (system of curve)，尤惟此是賴。此外尚有其他方法，然亦不外二途：一求得可以滿足條件——觀察資料係從隨機抽樣而有最小反機率者得來——之常數；二，或從所選點——而為方程式確切相合之點——中取用少數幾點。

第一節　皮爾生氏曲線系

皮爾氏曲線系，因其符號為統計調查一般所應用，吾人頗有注意之必要，與現所論者，亦復有關係，故不惜一說明之。惟若為精詳討論，則不妨參閱艾得頓先生所著『頻數曲線與相關』(Elderton: Frequency Curves and correlation)，及哈地先生著『死亡表編製原理』(Hardy: Theory of the Construction of Table of Mortality)。

μ_0，μ_1，………μ_t………表示一頻數曲線之連續各級動差，μ_0 代表面積，係作整一視之。設曲線之重心點落於縱坐標上，μ_1 之值

等於零。在此情形下，標準差，σ，即爲$\sqrt{\mu_2}$。β_1代表$\frac{\mu_3^2}{\mu_2^3}$，β_2代表$\frac{\mu_4}{\mu_2^2}$。

此方程式

$$D_x y=\frac{(x+a)y}{b_0+b_1x+b_2x^2}\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots(84)$$

即爲解析之基礎。

此方程式甚合一條件 — 當 $y=0$ 時，曲線必經過縱軸，並同時在另一位置成水平線，如在 $x=-a$ 時便是。換言之，該曲線只有一衆數。

實際上，分母即以 b_2x^2 爲止，以下無須多求。

此方程式，經積分而得下列三個交替之一般形式：

$$y=y_0\left(1+\frac{x}{a_1}\right)^{va1}\left(1-\frac{x}{a_2}\right)^{va2},\ y=y_0\left(1+\frac{x^2}{a^2}\right)^{-m}e^{-v\tan^{-1}\frac{x}{a}},$$

及
$$y=y_0(x-a)^{2x-1},$$

在上三式中，y_0 及各組之三常數(v,a_1,a_2)，(m,a,v)，(a,q_2,q_1)均可用動差在基本方程式中之 a,b_0,b_1,b_2，求出。哈地氏曾用較簡之符號告吾人以另一解析方法。

當 a,b_0,b_1,b_2 各有特定值，或相互間有特定之關係時，僅含有兩個（甚至僅有一個）常數之簡單方程式，不難求得。關於曲線，茲共分爲主要七型，艾得頓氏曾示吾人，各型配合相當觀察頻數彙類之法。惟所用代數及算術甚繁耳。此法應用於生活費之

食物費用之結果，在本編第三章第七節，可見一斑。

差誤常態曲線方程式，可書如下列形式：

$$D_x y = -\frac{xy}{\sigma^2} \cdots\cdots\cdots (85)$$

此乃特型之一。

求得一般差誤曲線之第二近似值爲

$$D_x y = -\frac{\left(x+\frac{\kappa\sigma}{2}\right)y}{\sigma^2+\frac{\kappa\sigma x}{2}} \cdots\cdots\cdots (86)$$

此式已將 k^2 略去，亦一特例也。

據皮爾生氏及其同仁所知，單峯(unimodal)觀察頻數羣類，一般均可用此公式之一二變量表現之。故平均數，σ，β_1 及 β_2 之計算，乃成爲用四個淺顯量數以表示羣類之一般而有效之方法，依此程序進行，普通卽以一平均數代表一羣類。用此四數，代表羣類之曲線方程，卽可成爲適當之形式，復可據以用插補法求相當任何 x 值之 y 值，不論觀察資料如何分級也。

現所欲討論者，並非此等方程式如何用於機率問題之問題，亦非此方程式如何由經驗得來之問題，更非欲討論與機率發生之假定有如何之關係也。

第二節　愛基華斯氏法

在一羣類偏斜度過大，不能適用普遍化之差誤律之第二近

似值時，則代表此羣類之差誤常態曲線，加以變形，卽可得一公式，此爲愛基華斯氏所採用。但爲羣類之敍述，插補或其他用途上，便利至何程度，尙未有充分之試驗，現時自不能輕予決定。（參閱統計學報，一八九八年，十二月份及一九一六年七月份所發表二文）。

第三節　巴里多氏方程

設 $b_0=0$，$b_2=\frac{b_1}{a}=-\frac{1}{m}$，則由上述曲線系，可得一方程式 $D_xy=-\frac{my}{x}$。此方程式所表示者，乃爲當 m 爲正數時，不拘 x 爲何值，總歸向右偏斜之曲線。

如積分之，則方程式變爲 $\log y=-m\log x+$ 常數，換言之，卽爲 $y=Cx^{-m}$。

曲線自 x 至∞之面積爲

$$z=\int_x^\infty Cx^{-m}dx=\left\{\frac{Cx^{1-m}}{1-m}\right\}_x^\infty$$

$$=\frac{C}{(m-1)x^{m-1}}\text{（設 }m>1\text{）}$$

以 a 代 $m-1$，A 代$\frac{C}{m-1}$，則得

$$y=\frac{Aa}{x^{a-1}},\ z=\frac{A}{x^a}\cdots\cdots\cdots\cdots(87)$$

最末一方程式，係爲所得問題所用巴里多氏律（Pareto's

Law)之最簡式。A與 a 為常數,z 為所得在 x 磅(或 x 佛郎及其他等等)左右之總人數。

其他羣類與此一公式相合者甚多,例如年租等級不同之房屋,此為年租之變量數及次數全距可為極寬者,並可用雙對數尺繪製成圖者,即是。

仍以所得問題而論,所得總額自 £x_1 至 £x_2 者,為

$$\int_{x_1}^{x_2} xydx=\frac{Aa}{a-1}\left(\frac{1}{x_1{}^{a-1}}-\frac{1}{x_2{}^{a-1}}\right)。$$

全距中所得數為 $A\left(\frac{1}{x_1{}^{a}}-\frac{1}{x_2{}^{a}}\right)$。

此定律對於極低或極高之所得,一般均以為不便應用。但如用於最高所得,則得

$$在\ £x\ 或以上之所得總額=\frac{Aa}{(a-1)x^{a-1}},$$

$$在\ £x\ 或以上所得數=\frac{A}{x^a}=N,$$

故 $$£x\ 或以上之平均所得=\frac{a}{a-1}x,$$

用此數方程式,由原始資料中,即可求出A,a。

巴里多氏方程對於一九一一至一九一二年須繳所得附加稅之所得統計,在全距 £5,000—£55,000 之間,配合情形良佳,惟在 £55,000 以上之所得數,算得者高於實在數耳。

$a=1.5$, $\log A=9.618$ 能切實配合。

所得全距	所得次數 計算數	記錄數
£5 to £10	7,546	7,411
10 ,, 15	1,890	2.029
15 ,, 20	790	787
20 ,, 25	424	438
25 ,, 35	411	382
35 ,, 45	199	186
45 ,, 55	103	107
55 ,, 65	70	56
65 ,, 75	50	37
75 ,, 100	118	55
100 and o er	83	66
Totals	11,700	11,554

所得總額:計算額 £166,000,000;記錄額 £145,000,000。

如繪一雙對數尺圖,全距—全距上劃一直線,則直線卽爲恰當之近似值—可以一目瞭然,而其斜度,卽可取爲 a 之試驗值。此值可選用 x 之兩值,例如 x_1 與 x_2,測驗出來,而 x_1 與 x_2 可由隔近經驗線之點,譬如 N_1 與 N_2,而得N之值。

於是 $$a=\frac{\log N_1-\log N_2}{\log x_2-\log x_1}。$$

如上表,以 $x_1=5,000$, $x_2=45,000$,則得 $N_1=11,554$,$N_2=321$;故 $a=1.63$。

雖然,此法係假設最高所得之次數,乃與定律相合;但一般情形,則未必盡然;實際上仍以三值 x_1, x_2, x_3 爲宜。

於是,方程式

$$f(a)=\left(\frac{1}{x_1{}^a}-\frac{1}{x_2{}^a}\right)\div\left(\frac{1}{x_1{}^a}-\frac{1}{x_3{}^2}\right)=\frac{\text{自 } x_1 \text{ 至 } x_2 \text{ 所得次數}}{\text{自 } x_1 \text{ 至 } x_3 \text{ 所得次數}}=k(\text{已知}$$

數）足可得出 a 值。設 $a=1.60$ 爲試驗值；依次以 $a=1.5$，$a=1.55$，$a=1.60$，$a=1.65$，以解算 $f(a)-k$，再用插補法求 a 之值，使 $f(a)=k$ 愈近愈佳。然後就此得數，測驗其他各值。a 之值既已求得，求 A 不難矣。

此外尙有一法，應用資料或可較爲完全，其法係用方程式

x_1 至 x_2 全距中之平均所得

$$=\frac{a}{a-1}\left(\frac{1}{x_1{}^{a-1}}-\frac{1}{x_2{}^{a-1}}\right)\div\left(\frac{1}{x_1{}^a}-\frac{1}{x_2{}^a}\right)。$$

關於此種公式之應用，可參閱英國下議院所得稅委員會之報告（一九〇六年三六五號，第二二〇至二三〇頁，第二四〇至二四一頁，及第二四五至二四六頁）。

第四節　梅克漢氏公式

由方程式 $-\frac{1}{y}\cdot\frac{dy}{dx}=a+bc^x$，可引出一實際應用上甚爲重要之公式，

爲簡便起見，用積分法，以 $a=-\log s$，$b=-\log c\times\log g$，則 $\log y=x\log s+c^x.\ \log g+$常數

$$\therefore y=ks^x。(g)^{c^x}。(k,s,g,c,\text{ 均爲常數})\cdots\cdots\cdots\cdots(88)$$

此卽梅克漢 (Makeham) 氏公式。在此公式中，y 爲某一時

活至 x 歲之人數，並寫作 l_x。

某一期間，δ_x，死亡人數，對該期開始時生存人數之比，再用期間除之，當期間無限消滅時，則爲

$$\frac{l_x - l_{x+\delta x}}{l_x \cdot \delta_x} = -\frac{1}{l_x} \cdot \frac{dl_x}{d_x} \quad \cdots\cdots (89)$$

此式多以 μ_x 代表之，所謂『死亡勢』("force of mortality")者即是。

用此公式之微分方程，則得

$$\mu_x = a + bc^x \quad \cdots\cdots (90)$$

此係以死亡勢爲二數之和，二數者一爲常數 a，另一(即 $bc^x = \mu'_x$)則因 $\frac{D_x \mu'_x}{\mu'_x} = \log c$，而依一定之幾何級數而增加者也。

哈地氏(同前書第八十八頁)用 $\overline{a + a'x}$ 代 a，其式較爲繁複，英國議會之管理國營保險法案一九一二至一九一三年報告，第一部第五百八十五頁亦如之。

哈地氏又用雙曲線方程式爲分度(graduation)之用，

$$z = \log \frac{y}{N-y} = k + \frac{m}{a+x} + \frac{n}{b+x}$$

y 爲年齡在 x 歲以下之丈夫人數，N 爲丈夫總人數(見死亡表編製法第五十頁至第五十一頁及 CD. 6907. P. 595)。

第六章　相關論

第一節　導言

各現象相互間是否彼此了不相關，各自獨立？此爲統計學重要問題之一。如答覆係爲反面，則請測量其依賴性如何。

本章討論對象，以此問題爲主要；提及此種問題，必聯帶論及二三變量，而各種變量之變動，又或有共同之原因。

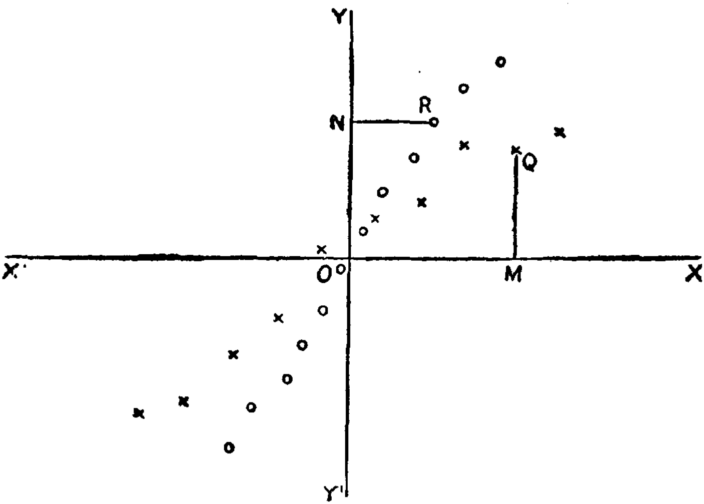
設有一對之觀察，如人之身長及指距，兩弟兄之身高，一家庭之所得及房租均是。假以各對量數爲(X_1, Y_1)，(X_2, Y_2)……，並設X，Y各成一頻數分類，平均數分別爲$\bar{x}, \bar{y}$。如X與Y彼此完全獨立，則知X之値，未必卽知Y之相對値；對於$\bar{y}$之某種離中差，有無一定之機率，乃視Y之頻數曲線而定；但如X，Y變動，有共同之原因，存乎其間，則一知X値爲何，相對之Y，離中差之機率，必得而預測矣。

X與Y可在一方程內，發生密切之關係，此爲當然之事實，例如X磅與Y公斤，乃爲同一體重之不同說法，依此而論，$X=2.204Y$，Y/X必爲常數也。然有多種事實，吾人今後欲加研究者，關係並不如此簡單；雖知X爲若干，Y値竟無從決定，在此情

形下，惟有藉重成列量數（如身高），以求隨此X而變之Y值。

如Y之各項，各與一X相聯，其頻數曲線平均數（或形式），與在未依X值而分類時之Y各值平均數（或形式）不同，則在X，Y兩數量之間，必有共同性質存在，此謂之X與Y相關。

解析之法，當以將Y之觀察值列爲整列（array）爲第一，每一整列同時將與X相同之各值全行列入，如列普通交叉表然。當$X = X_t$，Y整列之平均數$\bar{Y}_t$，如具有完全獨立性而觀察數甚多，則有漸近於所有各Y之平均數，y，之傾向；苟該平均數與y相差，駕乎隨機抽樣所得之上，則可知Y值並未與X值脫離牽涉，故不能互爲獨立也，



上圖，設O代表X，及Y二列之平均數。以x_t，y_t爲X_t，Y_t超過平均數$\bar{x}$，$\bar{y}$，上之部分，並以OM爲任一值，x_t，MQ爲整列中各y之平均數；如此，則$MQ=\bar{Y}_t-\bar{y}$；×號表示Q之各處位置。

於是，如Y與X各爲獨立，則×××只於觀察次數不足以求得眞實平均數時，距XX'方爲甚遠。如Y不能脫離X之影響，則Q必形成一定軌跡，而此軌跡卽爲經過各點隨手劃成之線。在此情形下，可以$\bar{Y}_t-\bar{y}=f(X_t-\bar{x})$，故當$X_t$值已知，$Y_t$之實在值，雖不得而知；但整列之平均數，$\bar{Y}_t$，卻可見於疊次抽樣中，故亦可約略決定之。

同理，如取相當y_t (ON) 任一值之X整列，則此等整列之平均數—— 0000標記之各R值卽是——有與曲線$\bar{X}_t-\bar{x}=f_1(Y_t-\bar{y})$相合之趨勢（此處，$f_1$與$f$同，須注意）。

Q之軌跡，名曰Y對X之迴歸線，R之軌跡，亦名X對Y之迴歸線。

此線曲線多大略爲直線，尤以鄰近O點時爲最，故$\bar{Y}_t-\bar{y}$約略等於kx_t（當x_t甚小時，k爲常數）。

此線之斜度，以k表示之，約略與$\frac{MQ}{OM}$相等（不論OM之值如何），求之方法，只須將$\frac{MQ}{OM}$之各值，加以平均卽可。以後當再論之（本章第三節及第七節）。

茲另從各方解決此問題：

設有n個量數X_1，X_2，……，並有n個量數Y_1，Y_2，……。

隨機抽出一X，再獨立抽得一Y，乃得一乘積XY。據經驗所知，一特定X被抽選之後，Y之各值將以相等之頻數而出現；且就一般而論，每一n^2個乘積X_1Y_1，X_1Y_2……X_2Y_1……X_nY_n必以相等之頻數而發生。

極大一數，N個，乘積之和$=S(XY)=S(\bar{x}+x)(\bar{y}+y)=N\bar{x}\bar{y}+\bar{x}\cdot Sy+\bar{y}Sx+Sxy$。

此處Sy漸近於$\frac{N}{n}(y_1+y_2+\cdots\cdots y_n)=0$，$S_x$亦漸近於0。

$S(xy)$漸近於$\frac{N}{n^2}(x_1y_1+x_1y_2+\cdots\cdots+x_2y_1+\cdots\cdots+x_ny_n)=\frac{N}{n^2}\cdot S_x\cdot Sy=0$。

故$S(XY)$漸近於$N\bar{x}\bar{y}$，乘積XY之平均值漸近於X之平均數與Y之平均數之乘積。

但如抽取Y時，不能與X之抽取脫離關涉，則n^2個乘積x_1y_1……，即不能以相同頻數而出現，故，平均$XY=\bar{x}\bar{y}+$平均xy。

又如$\bar{m}$為M_1M_2……之未加權平均數，依第四章第五節所用之符號，$M_t=\bar{m}+m_t$，而$\bar{m}_w$為加權平均數，至w_t——M_t之權數——則等於$\bar{w}+w_t$，又$\bar{w}$乃權數之平均數也。

$$\bar{m}_w=\frac{S(\bar{w}+w_t)(\bar{m}+m_t)}{n\bar{w}}=\frac{1}{n\bar{w}}(n\bar{w}\bar{m}+Sw_tm_t)$$

$$=\bar{m}\left(\mathrm{I}+\frac{\mathrm{I}}{n}\mathrm{S}\frac{w_t}{\bar{w}}\cdot\frac{m_t}{\bar{m}}\right)\cdots\cdots\cdots\cdots(91)$$

至於$\bar{m}_w=\bar{m}$之條件有三：

(1)只於$\mathrm{S}w_tm_t=0$時，有之

(2)只於n甚大時，有之

(3)只於大權數與大數量比與小數量難於發現時，有之。反是，則不等。

依第三章第一節所論，如用上述符號，$(x+y)^2$之平均值爲$\sigma_x^2+\sigma_y^2$，所設σ_x與σ_y，抽取X與Y時，在完全獨立情形下，乃爲X，Y之標準差。

由此可見，如不能彼此獨立時，則上式應當變爲

$$\mathrm{S}^2=\text{平均值}(x+y)^2=\sigma_x^2+\sigma_y^2+2\cdot\text{平均值}xy\cdots\cdots\cdots(92)$$

第二節　相關係數

由此觀之，X與Y未能彼此獨立時，『平均值xy』一數量，算式應用者甚多，且由此一數量，可藉明相關之存在及其總量。然其數值乃以測量x,y之單位爲單位，並無相當天然尺度以名之，故下述一數量，實爲最優之名稱。

設有$X_1Y_1,X_2Y_2\cdots\cdots X_tY_t\cdots\cdots X_nX_n$若干對量數，X及Y之平均數及標準差爲$\bar{x},\sigma_x,\bar{y},\sigma_y$，則X，Y間之相關係數，乃爲

r_{xy}。r_{xy}之求法，如下：

$$r_{xy}=\frac{S\{(X_t-\bar{x})(Y_t-\bar{y})\}}{n\sigma_x\sigma_y}=\frac{1}{n}S\left(\frac{x_t}{\sigma_x}\cdot\frac{y_t}{\sigma_y}\right)$$

(以$X_t=\bar{x}+x_t$，$Y_t=\bar{y}+y_t$)

$$=\frac{1}{n\sigma_x\sigma_y}\left\{S(X_tY_t)-\bar{x}SY_t-\bar{v}SX_t+n\bar{x}\bar{y}\right\}$$

$$=\frac{1}{n\sigma_x\sigma_y}\left\{S(X_tY_t)-n\bar{x}\bar{v}\right\}\cdots\cdots\cdots\cdots\cdots\cdots(93)$$

(因$SY_t=n\bar{y}$，$SX_t=n\bar{x}$)

依上例

$$\text{平均值}XY=\bar{x}\bar{y}+r_{xy}\sigma_x\sigma_y$$

$$\bar{m}_w=\bar{m}\left((I+r_{mw}\frac{\sigma_m}{\bar{m}}\cdot\frac{\sigma_w}{\bar{w}}\right)$$

$$s^2=\sigma_x^2+\sigma_y^2+2r_{xy}\sigma_x\sigma_y$$

以r代r_{xy}·可知r永不$>I$，或<-1。

因爲 $n^2r^2\sigma_x^2\sigma_y^2=(Sx_ty_t)^2$；

但 $n^2\sigma_x^2\sigma_y^2-(Sx_ty_t)^2=(x_1^2+x_2^2+\cdots\cdots)(y_1^2+y_2^2+\cdots\cdots)-(x_1y_1+x_2y_2+\cdots\cdots)^2$

(因爲 $n\sigma_x^2=x_1^2+x_2^2+\cdots\cdots$，且$n\sigma_y^2=y_1^2+y_2^2+\cdots\cdots$)

$$=(x_1y_2-x_2y_1)^2+(x_1y_3-x_3y_1)^2+\cdots\cdots+(x_2y_3-x_3y_2)^2+\cdots\cdots+(x_ny_{n-1}-x_{n-1}y_n)^2$$

$n^2\sigma_x^2\sigma_y^2>0$(除非$x_1y_2-x_2y_1=0=x_1y_3-x_3y_1=\cdots\cdots$)

$$\therefore\quad \frac{x_1}{y_1}=\frac{x_2}{y_2}=\frac{x_3}{y_3}=\cdots\cdots=\frac{x_n}{y_n}=\pm\frac{\sigma_x}{\sigma_y},$$

在此情形下，該式等於零，於是$r=\pm\mathrm{I}$。

$$\therefore\quad r^2=\frac{(\mathrm{S}_{xt}y_t)^2}{n^2\sigma x^2\sigma_y^2}<\mathrm{I},$$

$\therefore\quad \mathrm{I}>r>-\mathrm{I}$（$y$隨$x$順變時爲例外）

於是　$r=+\mathrm{I}$或$-\mathrm{I}$ ……………………………………………………(94)

r之性質，可歸納於下：

(1)r爲一隨所有觀察而起變化之數量，

(2)當抽選X，Y各值完全獨立，而平均值$xy=0$時，r之值爲零，

(3)與測量X，Y所用之單位無關，

(4)凡遇一正x_t與正y_t同時出現，或遇一負x_t與負y_t同來時，r之值必增加，

(5)但r增加僅能達到$+\mathrm{I}$，而不能超過$+\mathrm{I}$，當x,y之關係爲方程式『$y=x\times$常數』時。

(6)如正x，與負y同來，或負x與正y同出時，r則由0漸變直至$-\mathrm{I}$爲止。

故r測算相關量(amount of corretation)甚爲敏捷。

爲與他種量數區別起見，此相關量數，謂之積和相關係數(product-sum coefficient of correlation)。

如以各對分爲整列，例如$x_{s1}y_s$；$x_{s2}y_s$……，而以$\bar{y}_s$爲n_t個數

量${}_1y_s,{}_2y_s$……之平均數，則$Sxy=Sx_s\cdot n_s\bar{y}_s$，而$\frac{\sigma_y}{\sigma_x}\cdot r=S\left(\frac{n_sx_s^2}{n\sigma_x^2}\cdot\frac{\bar{y}_s}{x_s}\right)$，(因$Sn_sx_s^2=n\sigma_x^2$)；故$\frac{\sigma_y}{\sigma_x}r$及爲本章第一節$\frac{MQ}{OM}$各值之加權平均數，而$y=r\frac{\sigma_y}{\sigma_x}\cdot x$爲Q之軌跡之近似值。

茲於下節討論(1)檢察r得各種數值之原因，(2)研究X，Y依各種假定之分佈，(3)求迴歸線方程式

第三節 r之特性

以X，Y爲二個變量，隨其他變量U，V，W，而起變動，其變動之情形如下：

$$X_t={}_1U_t+{}_2U_t+\cdots\cdots+{}_pU_t+{}_1V_t+{}_2V_t+\cdots\cdots+{}_qV_t,$$

$$Y_t={}_1U_t+{}_2U_t+\cdots\cdots+{}_pU_t+{}_1W_t+{}_2W_t+\cdots\cdots+{}_qW_t,$$

${}_1U_t$係由一以${}_1\bar{u}$爲平均數，${}_1\sigma_u$爲標準差之任何形式之頻數羣類隨機抽來，${}_2U_t$又另從其他羣類獨立抽來，其餘U，V，W各類亦然。p與q爲任何整數。

以${}_1U_t={}_1\bar{u}+{}_1u_t$……，以$X_t=\bar{x}+x_t$，$Y_t=\bar{y}+y_t$，$\bar{x}$與$\bar{y}$爲X，Y各可能值之平均數。以$\sigma_x$，$\sigma_y$爲X，Y之標準差。於是

$$x_t={}_1u_t+{}_2u_t+\cdots\cdots+{}_pu_t+{}_1V_t+{}_2V_t+\cdots\cdots+{}_qV_t,$$

$$y_t={}_1u_t+{}_2u_t+\cdots\cdots+{}_pu_t+{}_1W_t+{}_2W_t+\cdots\cdots+{}_qW_t,$$

$$\sigma_x{}^2={}_1\sigma_u{}^2+\cdots\cdots+{}_p\sigma_u{}^2+{}_1\sigma_v{}^2+\cdots\cdots+{}_q\sigma_x{}^2,$$

$$\sigma_y^2 = {}_1\sigma_u^2 + \cdots\cdots + {}_p\sigma_u^2 + {}_1\sigma_w^2 + \cdots\cdots + {}_q\sigma_w^2,$$

以上各u, v, w項，依假說之規定，彼此各係獨立（見第三章第一節）。

如${}_1\sigma_u = {}_2\sigma_u = \cdots\cdots = \sigma_u$，${}_1\sigma_v = {}_2\sigma_v = \cdots\cdots = \sigma_v$，且${}_1\sigma_w = {}_2\sigma_w \cdots\cdots = \sigma_w$，或以$\sigma_u^2, \sigma_v^2, \sigma_w^2$為U，V，W各標準差之平均值，則

$$\sigma_x^2 = p\sigma_u^2 + q\sigma_v^2, \quad \sigma_y^2 = p\sigma_u^2 + q\sigma_v^2。$$

又　平均值$x_t y_t$＝平均值${}_1u_t^2$＋平均值${}_2u_t^2$＋……＋平均值，

$${}_1u_t \cdot {}_1w_t + \cdots\cdots + \text{平均值}\,{}_1v_t \cdot u_t + \cdots\cdots = p\sigma_u^2,$$

因依假說之規定，U，V，W各項之抽取，彼此完全獨立，則在長久經驗之後，${}_1u_t \cdot {}_1w_t$等各項當等於零。

$$\text{故}\ r = \text{平均值}\,x_t y_t / \sigma_x \cdot \sigma_y = \frac{p\sigma_u^2}{\sqrt{\{(p\sigma_u^2 + q\sigma_v^2)(p\sigma_u^2 + q\sigma_w^2)\}}} \quad (95)$$

又在特別情形之下，如$\sigma_u = \sigma_v = \sigma_w$，則

$$r = \frac{p}{p+q} \quad \cdots\cdots\cdots\cdots(96)$$

此為r數值之最簡單概念；如以文字說明之，此式之意義，乃指：相關係數漸近於兩變量產生時之共同原因數，對各變量所依之獨立原因總數之比。

如加常數a, b, c, d，使

$$\left.\begin{aligned} x_t &= a_1 \cdot {}_1u_t + \cdots\cdots + a_p \cdot {}_pu_t + b_1 \cdot {}_1v_t + \cdots\cdots + b_q \cdot {}_qu_t \\ y_t &= c_1 \cdot {}_1u_t + \cdots\cdots + c_p \cdot {}_pu_t + d_1 \cdot {}_1w_t + \cdots\cdots + d_q \cdot {}_qw_t \end{aligned}\right\} \cdots(97)$$

於是　$\sigma_x^2 = \mathrm{S} \cdot a^2\sigma_n^2 + \mathrm{S} \cdot b^2\sigma_v^2$，$\sigma_y^2 = \mathrm{S} \cdot c^2 b_u^2 + \mathrm{S} \cdot d^2\sigma_w^2$，

平均值$x_t y_t = Sac\sigma_u^2$,

求r之方程式，於此可得矣。

第四節 相關面

試就U，V，及W各頻數曲線係常態之情形而論，先請考查當$x_t = u_t + v_t$，$y_t = u_t + w_t$時X，Y之分類。

抽得u_t，v_t，w_t之機率爲

$$\frac{1}{\sigma_u\sqrt{2\pi}}e^{-\frac{u_t^2}{2\sigma_u^2}} \times \frac{1}{\sigma_v\sqrt{2\pi}}e^{-\frac{v_t^2}{2\sigma_u^2}} \times \frac{1}{\sigma_w\sqrt{2\pi}}e^{-\frac{w_t^2}{2\sigma_w^2}}$$

消去v_t及w_t。

於從U類中抽得u_t（以至$u_t+\delta_u$爲止）時，取得

$$x_t(至x_t+\delta_x), y_t(至y_t+\delta_y)$$

之機率爲：

$$\frac{1}{\sigma_u\sigma_v\sigma_w(2\pi)^{\frac{3}{2}}}e^{-\frac{1}{2}\left\{\frac{u_t^2}{\sigma_u^2}+\frac{(u_t-x_t)^2}{\sigma_v^2}+\frac{(u_t-y_t)^2}{\sigma_w^2}\right\}}\cdot\delta x\delta y\delta u$$

$$=\frac{1}{\sigma_u\sigma_v\sigma_w(2\pi)^{\frac{3}{2}}}\cdot e^{-\frac{1}{2}k\{u_t-lx_t-my_t\}^2-\frac{1}{2}(ax_t^2+2hx_ty_t+by_t^2)}\cdot\delta x\delta y\delta u$$

式中 $k=\frac{1}{\sigma_u^2}+\frac{1}{\sigma_v^2}+\frac{1}{\sigma_w^2}, kl=\frac{1}{\sigma_v^2}, km=\frac{1}{\sigma_w^2}$

$$a=\frac{1}{\sigma_v^2}-kl^2, b=\frac{1}{\sigma_w^2}-km^2, h=-klm。$$

從u之任何值，求得x_t，y_t之機率(譬如P_{xy})，將從u之指定值求得x_t，y_t之機率，相加卽得。

故，以x,y代x_t,y_t，

$$P_{xy}=\frac{1}{\sigma_u\sigma_v\sigma_w\cdot 2\pi}e^{-\frac{1}{2}(ax^2+2hxy+by^2)}\cdot\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}k(u_t-lx_t-my_t)^2}\cdot du$$

$$=\frac{1}{\sigma_u\sigma_v\sigma_w 2\pi\sqrt{k}}e^{-\frac{1}{2}(ax^2+2hxy+by^2)}$$

但　$\sigma_x{}^2=\sigma_u{}^2+\sigma_v{}^2,\ \sigma_y{}^2=\sigma_u{}^2+\sigma_w{}^2,\ r\sigma_x\sigma_y=\sigma_u{}^2$

$$\therefore k=\frac{(\sigma_u{}^2+\sigma_v{}^2)(\sigma_u{}^2+\sigma_w{}^2)-\sigma_u{}^4}{\sigma_u{}^2\sigma_v{}^2\sigma_w{}^2}=\frac{\sigma_x{}^2\sigma_y{}^2(1-r^2)}{\sigma_u{}^2\sigma_v{}^2\sigma_w{}^2}$$

$$a=kl-kl^2=l(k-kl)=l\left(\frac{1}{\sigma_u{}^2}+\frac{1}{\sigma_w{}^2}\right)=\frac{1}{k\sigma_v{}^2}\cdot\frac{\sigma_y{}^2}{\sigma_u{}^2\sigma_w{}^2}$$

$$=\frac{1}{\sigma_x{}^2(1-r^2)}$$

同理　$b=\dfrac{1}{\sigma_y{}^2(1-r^2)}$

$$-h=\frac{1}{k}\cdot kl\cdot km=\frac{1}{k\sigma_v{}^2\sigma_w{}^2}=\frac{\sigma_u{}^2}{\sigma_x{}^2\sigma_y{}^2(1-r^2)}=\frac{r}{\sigma_x\sigma_y(1-r^2)}$$

$$\therefore P_{xy}=\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-r^2}}e^{-\frac{1}{2(1-r^2)}\left(\frac{x^2}{\sigma_x{}^2}+\frac{y^2}{\sigma_y{}^2}-\frac{2rxy}{\sigma_x\sigma_y}\right)}\cdots\cdots\cdots(98)$$

將此法擴張，可知（見艾得頓著：『頻數曲線』第一百零九頁，該文係根據 Pearson，在皇家學會報，第一百八十七卷（一八九六年），A 175號之文）如x_t,y_t係由若干均爲常態頻數（如公式98所示）之變量加權總和(weighted sum)而成，則 P_{xy} 卽成$ke^{-(ax^2+2hxy+by^2)}$形式，上文已述及；又當以曲線係整體爲條件時，標準差σ_x；σ_y及$r\sigma_x\sigma_y$之平均乘積用積分法表示，則k,a,h,b之值，與已論過之單純情形所得者相同。

此法頗有意趣，且此爲皮爾生氏用積和公式(the product-sum formula)求算相關，在現代統計學中之一貢獻，但此乃基於元素頻數曲線係爲常態之假定，由此假定，積和方法之價值大爲消失，蓋觀察資料未必盡合常態也。玆爲免除此種假定起見，爰就愛基華斯氏『大數律』一文（該本上已介紹），重爲分析推論於下。

第五節　愛基華斯氏法

假以$x_t={}_1u_t+{}_2u_t+\cdots\cdots+{}_pu_t\cdots\cdots+{}_nu_t$

$$y_t={}_1v_t+{}_2v_t+\cdots\cdots+{}_pv_t\cdots\cdots+{}_nv_t$$

${}_1u_t$爲自平均數之離中差，換言之，即自以${}_1\sigma_n$爲標準差之頻數曲線中抽得一數量之平均數而來之離中差，${}_2u_t\cdots\cdots{}_nu_t$，${}_1v_t\cdots\cdots{}_nv_t$亦如之。設抽取各個$u_t$值，彼此完全獨立，則平均值${}_1u_t\cdot{}_2u_t$漸近於零，$v_t$各值亦爲獨立，則亦同然；但設有數$v_t$值不能離$u_t$值而獨立，則平均值${}_1u_t\cdot{}_1v_t$平均值${}_2u_t\cdot{}_2v_t\cdots\cdots$平均值${}_nu_t\cdot{}_nv_t$，並不全漸近於零。至平均值${}_1u_t\cdot{}_2v_t$一數量，則視爲近於零。

設n甚大，$\frac{1}{\sqrt{n}}$可以略去，且與第三章第五節所舉之其他條件相合，則x，y之各別頻數曲線必爲差誤常態曲線。

於是　$\sigma_x^2=S(_p\sigma_u^2)$，$\sigma_y^2=S(_p\sigma_v^2)$，

而　　平均值$xy=S($平均值${}_pu_t,{}_pv_t)\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots(99)$

式中p，爲1與n間之一整數。

再將測量x與y之軸轉換，經過θ角，使$2\theta=\frac{2r\sigma_x\sigma_y}{\sigma_x^2-\sigma_y^2}$，$r=\frac{\text{平均值}xy}{\sigma_x\sigma_y}$即爲$x$與$y$間之相關係數。

試以$X_t=x_t\cos\theta+y_t\sin\theta$，

$$Y_t=x_t\sin\theta-y_t\cos\theta,$$

再以${}_pU_t={}_pu_t\cos\theta+{}_pv_t\sin\theta$，

$${}_pv_t={}_pu_t\sin\theta-{}_pv_t\cos\theta\text{（註一）}$$

於是$X_t=S_pU_t$，

且　$Y_t=S_pV_t$

對$\tan 2\theta$之值而言，

$$\text{平均值}X_tY_t=\sin\theta\cos\theta(\sigma^2_x-\sigma_y^2)-\cos2\theta\cdot(\text{平均值}xy)=0。$$

$$\sigma_x^2=\cos^2\theta\cdot\sigma_x^2+\sin^2\theta\cdot\sigma_y^2+\sin2\theta\cdot r\sigma_x\sigma_y=S(\text{平均值}{}_pU_t^2)$$

$$\sigma_y^2=\sin^2\theta\cdot\sigma_x^2+\cos^2\theta\cdot\sigma_y^2-\sin2\theta\cdot r\sigma_x\sigma_y=S(\text{平均值}{}_pV_t^2)$$

$$\therefore\sigma_x^2+\sigma_y^2=\sigma x^2+\sigma_y^2$$

$$\sigma_x^2-\sigma_y^2=(\sigma_x^2-\sigma_y^2)\cos2\theta+2r\sigma_x\sigma_y\sin2\theta=(\sigma_x^2-\sigma_y^2)\sec2\theta$$

$$=\sqrt{\{(\sigma_x^2-\sigma_y^2)^2+4r^2\sigma_x^2\sigma_y^2\}}$$

$$4\sigma_x^2\sigma_y^2=4\sigma_x^2\sigma_y^2(1-r^2)$$

$$\text{平均值}{}_pU_t\cdot{}_pV_t=\sin\theta\cos\theta({}_p\sigma_u^2-{}_p\sigma_v^2)-\cos2\theta$$

$$(\text{平均值}{}_pu_t\cdot{}_pv_t)。$$

$$\therefore S(\text{平均值}{}_pV_t\cdot{}_pV_t)=\text{平均值}X_tY_t=0\quad\cdots\cdots\cdots\cdots(100)$$

現依第三章第四節所用計算X，Y之動差方法。

設α，β爲任何小之常數，茲用以收集相似之項。

$$e^{\alpha X_t+\beta Y_t}=e^{\alpha_1 U_t+\beta_1 V_t}\times e^{\alpha_2 U_t+\beta_2 V_t}\times\cdots\cdots$$

將此指數函數展開，得出t個可能值，而取其平均值，吾人依前述原理，已知一和之平均值，即爲各項之平均總和，又知獨立因子（如上式右方之因子）積數之平均值，即爲各因子之平均值積數，且知一乘冪之平均，乃爲零。

如以k，l爲任何之整數，

$$\left\{1+\frac{\alpha^2}{2}\sigma_X{}^2+\cdots\cdots+\frac{\alpha^k}{k!}(\text{平均數}X^k)\cdots\cdots\right\}$$

$$\times\left\{1+\frac{\beta^2}{2}\sigma_Y{}^2+\cdots\cdots+\frac{\beta^l}{l!}(\text{平均值}Y^l)\cdots\cdots\right\}$$

$=n$個因子之乘積

$$\{1+\tfrac{1}{2}\alpha^2(\text{平均值}_pU_t{}^2)+\tfrac{1}{2}\beta^2(\text{平均值}_pV_t{}^2)+\alpha\beta(\text{平均值}_pU_{tp}V_t)+\cdots\cdots\}$$

$$\therefore\quad \log\left\{1+\cdots\cdots+\frac{\alpha^k\beta^l}{k!\,l!}(\text{平均值}X^kY^l)+\cdots\cdots\right\},$$

（其k或l或可等於零），

$$=S[\log\{1+\tfrac{1}{2}\alpha^2(\text{平均值}_pU_t{}^2)+\tfrac{1}{2}\beta^2(\text{平均值}_pV_t{}^2)+\alpha\beta(\text{平均值}_pU_{tp}V_t)+\cdots\cdots\}]$$

＝（用對數級數將其展開，並歸併各項）

$$\tfrac{1}{2}\alpha^2s(\text{平均值}_pU_t{}^2)+\tfrac{1}{2}\beta^2S(\text{平均值}_pV_t{}^2)+\alpha\beta S(\text{平均值}_pU_{tp}V_t)+\cdots\cdots$$

$=\frac{1}{2}\alpha^2\cdot\sigma_X{}^2+\frac{1}{2}\beta^2\cdot\sigma_Y{}^2+\alpha\beta\times0$（註二）＋含有$\alpha^3,\alpha^2\beta\cdots\cdots$之項。

依第三章第四節末段之理論，有α^3之項係屬$\frac{1}{\sqrt{n}}$次，α^2項則不然，故如將$\frac{1}{\sqrt{n}}$略去，則

$$1+\cdots\cdots+\frac{\alpha^k\beta^l}{k!\,l!}(\text{平均值}X^kY^l)+\cdots\cdots=e^{\frac{1}{2}\alpha^2\sigma_X{}^2}\times e^{\frac{1}{2}\beta^2\sigma_Y{}^2}$$

$$=\left\{1+\frac{1}{2}\alpha^2\sigma_X{}^2+\cdots\cdots+\frac{1}{k!}\left(\frac{1}{2}\alpha^2\sigma_X{}^2\right)^k\cdots\cdots\right\}$$

$$\left\{1+\frac{1}{2}\beta^2\sigma_Y{}^2+\cdots\cdots+\frac{1}{l!}\left(\frac{1}{2}\beta^2\sigma_Y{}^2\right)^l\cdots\cdots\right\}$$

取此方程式兩方各項係數令其相等，則在$l=0$時，

$$\text{平均值}X^{2k+1}=0,\ \text{平均值}X^{2k}=\frac{(2k)!}{2^kk!}=\sigma_X{}^{2k}$$

與在常態曲線同，當Y爲零時，Y亦同然。

凡與X或Y之奇數乘冪（odd power）有關各平均值均等於零。

$$\text{平均值}(X^{2k}Y^{2l})=\frac{(2k)!}{2^kk!}\sigma_X{}^{2k}\cdot\frac{(2l)!}{2^2l!}\sigma_Y{}^{2l}\cdots\cdots\cdots\cdots(101)$$

此等恰爲將下列之平面積分，求得之平均乘冪(mean power)

$$z=\frac{1}{2\pi\sigma_X\sigma_Y}e^{-\frac{1}{2}\left(\frac{X^2}{\sigma_X{}^2}+\frac{Y^2}{\sigma_Y{}^2}\right)}=\frac{1}{\sigma_X\sqrt{2\pi}}e^{-\frac{1}{2}\frac{X^2}{\sigma_X{}^2}}\cdot\frac{1}{\sigma_Y\sqrt{2\pi}}e^{-\frac{1}{2}\frac{Y^2}{\sigma_Y{}^2}},$$

其X係與Y獨立。故如第三章第四節末段之理論，可以此方程式，用以求X，Y之頻數。

現當仍轉換爲原軸。

前已示明$\sigma X\sigma Y=\sigma_x\sigma_y\sqrt{1-r^2}$,

$$X^2\sigma Y^2+Y^2\sigma X^2=(x\cos\theta+y\sin\theta)^2\sigma Y^2+(x\sin\theta-y\cos\theta)^2\sigma X^2$$

$$=x^2(\cos^2\theta\sigma Y^2+\sin^2\theta\sigma X^2)+y^2(\sin^2\theta\sigma Y^2+\cos^2\theta\sigma X^2)$$

$$-xy\sin2\theta(\sigma X^2-\sigma Y^2)$$

於是x^2及y^2係數之和$=\sigma X^2+\sigma Y^2=\sigma X^2+\sigma Y^2$

x^2及y^2係數之差$=\cos2\theta(\sigma Y^2-\sigma X^2)=\sigma_y^2-\sigma_x^2$

故其係數爲σ_y^2及σ_x^2

且$\sin2\theta(\sigma X^2-\sigma Y^2)=\tan2\theta(\sigma_x^2-\sigma_y^2)=2r\sigma_x\sigma_y$。

故$\dfrac{X^2}{\sigma X^2}+\dfrac{Y^2}{\sigma Y^2}=\dfrac{1}{\sigma_x^2\sigma_y^2(1-r^2)}\{x^2\sigma_y^2+y^2\sigma_x^2-2r\sigma_x\sigma_y\cdot xy\}$

而平面之方程式爲

$$z=\frac{1}{2\pi\sigma_x\sigma_y\sqrt{(1-r^2)}}\cdot e^{-\frac{1}{2(1+r^2)}\left\{\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}-\frac{2rxy}{\sigma_x\sigma_y}\right\}}\text{……}(102)$$

如上（公式98）所言，在假定——元素屬類係常態曲線——之下,理應如是。

此方程式$\sigma_x^2=S\sigma_u^2$, $\sigma_y^2=S\sigma_v^2$, $r\sigma_x\sigma_y=S(r_p\cdot{}_p\sigma_u\cdot{}_p\sigma_v)$,內,係從公式99而來,其$r_p$卽爲$_pu$與$_pv$之相關係數。

如將常數加進原來方程式內,則$x={}_1a_1u+{}_2a_2u+$……且$y={}_1b_1v+{}_2b_2v+$……而與分析程序無妨。

第六節 常態相關面之性質

以x與y變數之平均數爲中心。

$$容積=\int\int z\,dx\,dy$$

$$=\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-r^2}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{-\frac{1}{2(1-r^2)}\left(\frac{x}{\sigma_x}-r\frac{y}{\sigma_y}\right)^2}\cdot e^{-\frac{1}{2}\frac{y^2}{\sigma_y^2}}\cdot dx\,dy=1$$

$$\cdots\cdots\cdots\cdots\cdots\cdots(103)$$

$$y之第二動差=\int\int zy^2dx\,dy$$

$$=\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-r^2}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{-\frac{1}{2(1-r^2)}\cdot\frac{x'^2}{\sigma_x^2}}\cdot y^2\cdot e^{-\frac{1}{2}\frac{y^2}{\sigma_y^2}}\cdot dx'dy,$$

其$x'=x-r\frac{\sigma_x}{\sigma_y}\cdot y$；然後對$x'$而積分，則該式（即$y$之第二動差）

$$=\frac{1}{\sigma_y\sqrt{2\pi}}\int_{-\infty}^{\infty}y^2e^{-\frac{1}{2}\frac{y^2}{\sigma_y^2}}dy=\sigma_y^2\text{（依公式21之規定）}\cdots\cdots(104)$$

同理x之第二動差爲σ_x^2。

$$xy之平均積數=\int\int 2xy\,dx\,dy\text{（積分之極限爲}\pm\infty\text{）}$$

$$=\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-r^2}}\int\int\left(x'+r\frac{\sigma_x}{\sigma_y}y\right)e^{-\frac{1}{2(1-r^2)}\cdot\frac{x'^2}{\sigma_x^2}}\cdot y\,e^{-\frac{1}{2}\frac{y^2}{\sigma_y^2}}dx'dy$$

$$=\frac{1}{\sigma_y\sqrt{2\pi}}\int r\frac{\sigma_x}{\sigma_y}y^2e^{-\frac{1}{2}\frac{y^2}{\sigma_y^2}}dy=r\sigma_x\sigma_y\cdots\cdots\cdots\cdots(105)$$

$$\int\int zyx^3dx\,dy=3r\sigma_x^3\sigma_y,\ \int\int zx^2y^2dx\,dy=(2r^2+1)\sigma_x^2\sigma_y^2\cdots(106)$$

由每一與XOZ，YOZ平行之平面，所構成之部分，係一常態

曲線。例如，以$x=x_1$，

$$z=\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-r^2}}\cdot e^{-\frac{1}{2(1-r^2)\sigma_y^2}\left(y-r\frac{\sigma_y}{\sigma_x}x_1\right)^2}\cdot e^{-\frac{1}{2}\frac{x_1^2}{\sigma_x^2}} \quad\cdots\cdots\cdots\cdots(107)$$

此卽爲一常態曲線，中心在$y=r\frac{\sigma_y}{\sigma_x}x_1$，標準差爲$\sigma_y\sqrt{1-r^2}$，又最大縱坐標爲

$$\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-r^2}}e^{-\frac{1}{2}\frac{x_1^2}{\sigma_x^2}}$$

相當一指定x值，各個y值組成之頻數羣類，自爲常態形式，其標準差亦脫離x之關係。羣類之平均數(及其衆數與中位數)，卽在$y=r\frac{\sigma_y}{\sigma_x}x$線上，（不論$x$值爲何），此線卽謂之迴歸線（本章第一節）。

$r\frac{\sigma_y}{\sigma_x}$卽爲$y$對於$x$之迴歸係數(coefficient of regression)

同理，對一已知之y值，x_1之頻數羣類亦屬常態，其標準差爲$\sigma_x\sqrt{1-r^2}$，其平均數係在$x=r\frac{\sigma_x}{\sigma_y}y$線上，例如$y=x\tan\phi_2$。

$r\frac{\sigma_y}{\sigma_x}$爲$x$對於$y$之迴歸係數。

兩個迴歸係數之幾何平均爲r。

橫面部分均爲相似橢圓形。例如以$z=z_1$，

$$\frac{x^2}{\sigma x^2}+\frac{y^2}{\sigma y^2}-2r\frac{xy}{\sigma_x\sigma_y}=-2(1-r^2)\log(z_1 2\pi\sigma_x\sigma_y\sqrt{1-r^2})\cdots\cdots(108)$$

此等橢圓(以$\sigma_x>\sigma_y$)，任一之長軸，均與x軸，成一θ角，而$\tan 2\theta$

$$=\frac{2r\sigma_x\sigma_y}{\sigma_x^2-2\sigma_y^2}。$$

則$\tan 2\theta=\dfrac{2r\sigma_x\sigma_y}{\sigma_x^2-r^2\sigma_y^2}$，又$\tan 2\phi_2=\dfrac{2r\sigma_x\sigma_y}{r^2\sigma_x^2-\sigma_y^2}$。

如$r=\pm1$，則$\theta=\phi_1=\phi_2$，該面乃降爲平面$\dfrac{y}{\sigma_y}=\dfrac{x}{\sigma_x}$。反之，如$|r|<1$，則$\phi_2>\theta>\phi_1$，而迴歸線乃落於該橢圓長軸平面兩邊之任一邊。如$\sigma_x=\sigma_y$，則$\theta=\dfrac{\pi}{4}$，而兩迴歸線便與包含數橢圓形主軸之平面成爲均等之角。

吾人於此應加注意，一面純視五數量而定，五數量者何？兩平均數，兩標準差，及一相關係數也。

第七節　直線相關

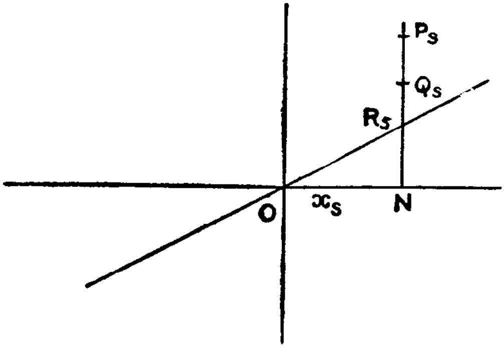
在某種情形下，性質單純而具複雜原因之迴歸線——即對一變量(x)之已知值，另一變量(y)之各列平均數之軌跡——係爲一經過多平均數所代表位置之直線，關於此點，前已論及。

但如條件未能嚴格滿足，卻只能大略適應，則須假定迴歸之直線性(retilinearity)，亦能大略實現也。

然即使變量x，y並非常態分配，迴歸依然大略合於直線；雖分佈平面決不能用r值而決定，而方程式$y=r\dfrac{\sigma_x}{\sigma_y}x$仍可爲迴歸方程。

設相當x之一值x_s，整列中共有n_s個y值，並設其平均數爲$\bar{y}_s$。以$m_s=\dfrac{\bar{y}_s}{x_s}$，則$m_s$乃爲僅由一羣類決定之迴歸線之斜度。

依本章第二節所論，$\tan\phi_1 = r\frac{\sigma y}{\sigma x}$，爲$m_1, m_2,$ ……… m_s ………之加權平均數，至其權數則爲$n_s x_s^2$………也。



設 $ON = x_s$，NP_s 爲隨x_s而來之任何y值，又以 $NQ_s = \bar{y}_s$爲n_s此等數值之平均數。設一條線$y = ax + b$與NP_s相遇於R_s點。

然則據猶爾氏（統計學報第一八九七年號第八百一十七至八百一十八頁）證明，經過x_s之所有各值後，$(R_sP_s)^2$——卽$S[y_s - (ax_s + b)]^2$，——所有各值之和，當於$b = 0$，及$a = r\cdot\frac{\sigma y}{\sigma x}$時，其值爲最小。此法根據最小二乘法，可參照附錄十。

此$y = r\cdot\frac{\sigma y}{\sigma x}x$一線經過各觀察值，經過時確使由各觀察值平行於$y$軸到該線之距離之平方之和成爲最小值。如是，無論分佈情況如何，該線始終爲迴歸之唯一良好代表。

請再進一步討論，試假定任一第s個整列內，y之離散度，求爲σ_2且與x_s無關。因如各平均數有在一條直線之傾向，除在觀察

量不充足，以致平均數不能落在直線上爲例外外，則對平均數之離中差，如R_sQ_s，必形成頻數曲線$Ke^{-\frac{(R_sQ_s)^2}{2\sigma^2}}\left(\sigma^2=\frac{\sigma_2^2}{n_s}\right)$。

故離中差$(R_1Q_1, R_2Q_2\cdots\cdots$之連帶機率(joint probability)爲$K'e^{-\frac{1}{2\sigma^2}Sn_s(R_sQ_s)^2}$，在$Sn_s(R_sQ_s)^2$爲最小時，此機率則爲最大。

$$Sn_s(R_sQ_s)^2=S[n_s\{\bar{y}_s-(ax_s+b)\}^2]$$
$$=S(n_s\bar{y}_s^2)+a^2s(n_sx_s^2)$$
$$+Nb^2-2aS(n_s\bar{y}_sx_s)-2bSn_s\bar{y}_s+2abSn_sx_s$$

$(N=n_1+\cdots\cdots+n_s+\cdots\cdots)$

但$Sn_s\bar{y}_s=0=Sn_sx_s$，因原點爲雙重平均數。

$$S(n_sx_s)^2=Sx^2=N\sigma_x^2, S(n_s\bar{y}_sx_s)=Sxy=Nr\sigma_x\sigma_y$$

於是該式等於$Sn_s\bar{y}_s^2+N(a\sigma_x-r\sigma_y)^2+Nb^2-Nr^2\sigma_y^2$。

當$b=0$，$a=r\frac{\sigma_y}{\sigma_x}$與前同，則此爲最小。

如假定一整列中之離中差，與x之相當值間，保持有直線性及獨立性，而距直線之離中差係由觀察太少而來，則此$y=r\frac{\sigma_y}{\sigma_x}x$線，爲迴歸線之最或能之軌跡。

從此線計算得來之$Sn_s(R_sQ_s)^2$值，爲$Sn_s\bar{y}_s^2-Nr^2\sigma_y^2$。

雖然，如測量之開始，或其結局，不能證明直線性之假定，則r爲分析中有用之參考雖有餘，但爲原因公同性度量之標準則不足。

第八節 相關率

為不受關於觀察值分佈狀況假定之拘束，皮爾生氏制定一種量數，謂之相關率 (correlation ratio) —— (見 Draper's Company Research Memoirs, Biometric Series, II. 1905)。

假以${}_s\sigma_y$為第s個整列之標準差，使$n_s\cdot{}_s\sigma_y{}^2=S(y_s-\bar{y}_s)^2$，又以$\sigma_a{}^2$為${}_1\sigma_y{}^2, {}_2\sigma_y{}^2$ ······ 之加權平均值，使$N\sigma_a{}^2=S(n_s\cdot{}_s\sigma_y{}^2)=SS(y_s-\bar{y}_s)^2$，裏面求和表示一整列內之各離中差相加，外面求和表示所有各整列之總和。

以$N\sigma_m{}^2=S(n_s\bar{y}_s{}^2)$，使 $\sigma_m{}^2$ 為各整列平均數之加權平均二次方。

於是$N\sigma_y{}^2=S(y^2)$，其求和普及於y之所有值，而且$N\sigma_a{}^2=S\{S(y_s{}^2)-2\bar{y}_sSy_s+n_s\bar{y}_s{}^2\}=S(Sy_s{}^2-n_s\bar{y}_s{}^2)=S(y^2)-S(n_s\bar{y}_s{}^2)$。

$\therefore \sigma_y{}^2=\sigma_m{}^2+\sigma_a{}^2$(如就別方面而論，其理甚明)。

於是 $$\eta=\frac{\sigma_m}{\sigma_y}=\sqrt{\left(1-\frac{\sigma_a{}^2}{\sigma_y{}^2}\right)}\cdots\cdots(109)$$

η名曰相關率。所謂相關率，乃即若干整列之平均數散佈度，對於未編成整列之羣類散佈度之比率。

若$\sigma_m=0$，因而每一$\bar{y}_s=0$時，則$\eta=0$。換言之，即於每一整列之平均數，與該羣類之總平均數相合時，則其相關為0。

若$\sigma_a=0$，則$\eta=1$。換言之，如每一${}_s\sigma_y=0$，且每一整列之各項，

均集中於一點,$\bar{y}_s$,時,則相關率為正一。

反是,則η小於1,而大於零,$1>\eta>0$。

在常態相關之下,每一${}_s\sigma_y=\sigma_y\sqrt{(1-r^2)}$, 如公式(107),則$\sigma_a{}^2=\sigma_y{}^2(1-r^2)$,而$\eta^2=r^2$。

又　$Sn_s(R_sQ_s)^2=N\sigma_m{}^2-Nr^2\sigma_y{}^2=N(\eta^2-r)\sigma_y{}^2$

$$\eta^2=r^2+\frac{Sn_s(R_sQ_s)^2}{N\sigma_y{}^2}, \quad\cdots\cdots\cdots\cdots(110)$$

且若非每一R_sQ_s均為0,而各整列之各該平均數,均在$y=r\dfrac{\sigma_y}{\sigma_x}x$線上時,則$|\eta|>|r|$。

相關問題,討論至此,已可作一結論:

如(x,y)為一對測量——自平均數量起——兩個變量(關於時間,空間,或關於事物,或關於有機體)之量數,且當x已知為正(或負)數時,則假定y為正(或負),或假定y為負(或正);於是該兩個變量,乃謂之相關。在此情形下,當n增加時,$\dfrac{1}{n}Sxy$不漸趨於零,而達一極限$r\sigma_x\sigma_y$。所謂$r=0$,$r=1$,$r=-1$者,各含有特定之意義,而r對x與y間,所有各種關係,感應均甚敏銳。依普通情形觀察,σ_a(各整列內平均散佈度)愈小,則r亦愈大。如x與y各為$(p+q)$個獨立元素——在此若干元素中,只有p個為x與y所共有——之總和,則在各元素之標準差相等時,r等於$p/(p+q)$。假如x與y由許多複雜獨立原因——內有若干為x,y所

共有——作成直線，則(一)乃說明r各對之全部頻數分配，(二)迴歸線爲直線，(三)迴歸方程式，爲$y=r\frac{\sigma y}{\sigma x}\cdot x$及$x=r\frac{\sigma x}{\sigma y}\cdot y$。如常態頻數曲線，不能決定，但迴歸線卻爲直線，則此同一方程，亦足爲迴歸之經驗的良好說明。但如關於x及y之分配，或各整列之各該平均數，並無若何規定，則r數值，必毫無意義可言（η在不等於0或1時，卽係如此）。雖然，就一般而論，r未嘗不可測量x，y共同原因系統之數額也。

第九節　未分級變量之相關

(Correlation between Ungraded Variables)

依照上述方法，測算相關，必須觀察有適當之詳密。未達到此詳細程度時，必發生有絕大興趣之問題。

頭髮顏色

		父母 淺	父母 深	總計
兒童	深	a	b	n_1
	淺	c	d	n_2
	總計	m_1	m_2	N

以毛髮之深淺，將父母兒童各加區分；淺色髮之父母，所生m_1個兒童，髮爲淺色者c人，爲深色者有a人；同時深色髮之父母，所生兒童共有m_2個，其髮爲淺色者有d人，深色者有b人。設$a+b$

$=n_1, c+d=n_2$，又$n_1+n_2=\mathrm{N}=m_1+m_2$。

見此資料，吾人須審定，父母兒童之間，髮色深淺有無相關存乎其中，如果有相關，則請測算之。

在此情形下，如變量（譬如色質數額）果爲常態分配，並有常態相關，則此問題卽可確定。蓋m_2對於N之比，n_1對於N之比，卽得（用第二章常態機率表，反查卽得）色質尺度之橫坐標，而此色質尺度，又依髮色之深淺，而劃分爲二部；不論r爲何值，由此等橫坐所成平面之一部分相關面，旣爲已知，且b/N對於此者之方程式，亦卽反求r之方程式也。

二 皮爾生氏(Phil. Trans. A. Vol CXCV，第一頁起)，艾得頓氏（『頻數曲線』第七章），均曾作應有之分析，而終局得一求r之方程式，由此方程式r可約略求出也。

如其資料吾人可得而支配，並能將兩變量由中位數各分爲兩部，則可得一簡易之解法。

設學算術及學代數之智力，係依常態而分配。將一大羣之學生，依其學算術之智力（以分數爲準，或用其他方法）爲次第，列爲一類，復依學代數之智力，另分列爲一類；假以b爲二者均在中位數以上之學生，c爲均在中位數以下者，其學算術在中位數之上，學代數在中位數之下者，有d人，學代數在中位數上，學算術在中位數下者，有a人。然此並非謂智力確經測量，不過僅排列

其次第而已。

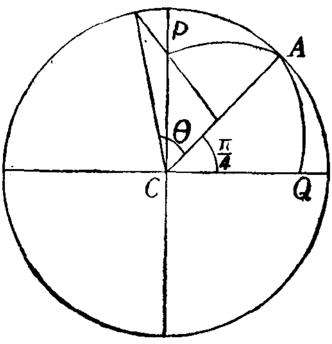
$$於是 a+b=\frac{N}{2}=c+d=a+c=b+d,$$

$$\therefore a=d=(\tfrac{1}{4}-q)N,\ b=c=(\tfrac{1}{4}+q)N.$$

$$r=\sin 2\pi q$$

其理容下再述。定每個尺度之標準差爲一。設所求之平面爲

$\dfrac{1}{2\pi\sqrt{1-r^2}}e^{-\frac{1}{2(1-r^2)}(x^2+y^2-2rxy)}$。則該平面之主軸，與 x 軸形成 $\frac{\pi}{4}$。$(\frac{1}{4}+q)$ 等於 x 與 y 均爲正數之象限內之容積，該象限乃由平面 $y=0, x=0$ 所成，至此等平面又將相似橢圓形之水平部分，各切去其面積之 $(\frac{1}{4}+q)$。蓋有橢圓形 $x^2+y^2-2rxy=1$。



在上圖，橢圓形面積 CPQ＝橢圓形面積之 $(\frac{1}{4}+q)$

以θ爲 P 之偏心角(eccentric angle)。

以其主軸而言之橢圓形爲

$$(1-r)x^2+(1-r)y^2=1。$$

$$\frac{\tan\theta}{\tan\frac{\pi}{4}}=\frac{長軸}{短軸}=\sqrt{\left(\frac{1+r}{1-r}\right)}。$$

$$\therefore \quad r=-\cos 2\theta$$

$$\tfrac{1}{4}+q=\frac{\text{CPA之兩倍面積}}{\text{橢圓形面積}}=\frac{2\theta}{2\pi}$$

$$\therefore \quad 2\pi q=2\theta-\frac{\pi}{2}$$

$$\sin 2\pi q=-\cos 2\theta=r$$

例如兒童爲百分之四十，在兩者中均在中位數以上，

$$\tfrac{1}{4}+q=4,\ q=\cdot 15,\ r=\sin\tfrac{3}{10}\pi=\cdot 81。$$

如$q=0$，$r=0$，如$q=\frac{1}{4}$，$r=1$。

據布魯昂(W. Brown)之實驗(見 Biometrika, Vol. VII, p. 366)，八十三名兒童中，代數智力高於中位數，而算術智力低於中位數者，佔一十一名。然則$q=\frac{1}{4}-\frac{11}{83}=\cdot 12$，故$r=\cdot 68$。布氏用全列次第——不只用中位數——計算，得知爲·65，又用分數得數·79。至於兒童在校中之地位及年齡亦有關係，不可不加以修正，布氏亦曾注意及之。

但如此兩種品質，分配不能形成常態，則相關量之測算問題，即不能得以確定，縱經用多數方法，亦不能達到目的也。

第十節　相聯

如另無因果關係存乎其間，深色髮之父而有深色髮之子，其父親人數當為$\frac{m_2}{N}\times\frac{n_1}{N}\times N=\beta$，即從N個事件中，深色髮者，父當有$m_2$人子有$n_1$人。

前於第九節所用符號仍舊，並以α,γ,δ代表a,c,d各區分內或然性最大之人數，則

$$a+b=\alpha+\beta=n_1,\ a+c=\alpha+\gamma=m_1\cdots\cdots\text{餘類推。}$$

$$\alpha-a=b-\beta=c-\gamma=\delta-d=b-\frac{m_2m_1}{N}$$

$$=\frac{b(a+b+c+d)-(b+d)(a+b)}{N}=\frac{bc-ad}{N}=qN.$$

q即為相聯(association)之量數，其意義除在特別情形之外不能確定。

猶爾氏不用q而用$Q=\frac{bc-ad}{bc+ad}$，即相聯係數(The "coefficient of association")或$w=\frac{\sqrt{bc}-\sqrt{ad}}{\sqrt{bc}+\sqrt{ad}}$，即相連係數(The "coefficient of colligation")為量數。(參閱 Introduction to Theory of Statistics, p. 37, 及 Statistical Journal, 1912, p. 593)。當$bc=ad$，$q=0$時，$Q=w=0$，此為不相聯之例；如a或d為零，$Q=w=1$；b或c為零，$Q=w=1$，此為相聯之最高量。

此等係數，在極端情形下，其意義乃可確定，但如以$Q=\frac{2}{3}$，其命義如何，非檢視若干試驗不能知，且最後，不能以Q之大小而斷定相聯量之大小，因相聯一詞並無確定之量數意義也。

第十一節　相依

如吾人不以求得相聯量爲止，更進而求其存在之證據，換言之，卽在品質彼此獨立時，求其觀察是否仍能發生，則理論基礎當愈爲鞏固矣。

如$p:1-p$，爲兒童深色髮者對淺色髮者之比率，則$p=\frac{n_1}{N}$爲自觀察中所得之最佳值。

故如將N名兒童武斷分爲兩類（例如以其基督教名自A至K爲一類，自L至Z爲一類），使分別佔有m_1及m_2人，則在第一類找到a之機率（已知第二章第十節所述），乃爲

$$\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{x_2}{2\sigma_2}},\left(\frac{1}{\sqrt{m_1}}\text{業已略去}\right)$$

但　$x=a-pm_1=a-\frac{n_1m_1}{N}=a-a=qN$（註三）

$$\sigma^2=p(1-p)m_1\left(1-\frac{m_1}{N}\right)=\frac{n_1}{N}\cdot\frac{n_2}{N}m_1\cdot\frac{m_2}{N}。$$

離中差，或正或負，其等於$(a-a)$時，出現之機率，爲

$$2\int_z^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}dz,\left(z=\frac{x}{\sigma}\right)。$$

請注意$\frac{x^2}{\sigma^2}=\frac{q^2N^5}{n_1n_2m_1m_2}=q^2N^3\cdot\frac{(n_1+n_2)(m_1+m_2)}{n_1n_2m_1m_2}$

$$=q^2N^3\left(\frac{1}{n_1m_1}+\frac{1}{n_1m_2}+\frac{1}{n_2m_1}+\frac{1}{n_2m_2}\right)$$

$$=\frac{(a+\alpha)^2}{\alpha}+\frac{(b-\beta)^2}{\beta}+\frac{(c-\gamma)^2}{\gamma}+\frac{(d-\delta)}{\delta}=\mathrm{X}^2。$$

例如，在分配$\frac{65}{35}\Big|\frac{235}{165}$內

$n_1=300, n_2=200, m_1=100, m_2=400, \mathrm{N}=500,$

$$\sigma^2=19\cdot2,\ \alpha=60,\ x=q\mathrm{N}=5,$$

$\frac{x}{\sigma}=1\cdot14$, $\mathrm{F}(1\cdot14)=\cdot373$（常態機率表），$2\{\frac{1}{2}-\mathrm{F}(1\cdot14)\}$ $=\cdot254$。

從500中抽出100個，出現者在65以上或55以下者之機率，爲•746對·254之比，卽約爲三對一之比。

n_1, m_1, N及a爲已知，其餘b, c, d, n_2, m_2亦已知，則所得之機率，恰爲均等之機率，足使b, c, d任何一數之出現，彼此完全獨立。此不能謂之爲全部分配之機率；欲求此機率，吾人一則須知，機率p係由較廣之宇宙間而來，二則須知非如$\frac{n_1}{\mathrm{N}}$於有限之觀察個數卽可決定者，三則須知$\frac{m_1}{\mathrm{N}}$卽爲對總機率之近似值。

爲例釋此難點，請就上已論過之問題，再考慮之。

	未種牛痘	已種牛痘	總計
痊愈者	a	b	n_1
死亡者	c	d	n_2
總計	m_1	m_2	N

天花流行病患者，共有N人，就中種痘者爲m_2人，死亡者有

n_2人，其他範疇俱見上表。

痊愈率，如該項全部統計所述，乃為$\frac{n_1}{N}$，又如種牛痘（無論直接影響，或其他連帶關係）與病之痊愈無關，則種痘及未種痘患者痊愈之機率，必為$\frac{n_1}{N}$，而在m_2個種痘患者中，b人業已痊愈之機率為

$$\int_{b-\beta}^{\infty}\frac{N^{\frac{3}{2}}}{\sqrt{(2\pi n_1 n_2 m_1 m_2)}}\cdot e^{-\frac{1}{2}\cdot\frac{x^2N^3}{n_1n_2m_1m_2}}\cdot dx,$$

$(qN=b-\beta=x)$；如此數甚低，則種痘（或環境使然）與痊愈果有關係，乃有證明。

然$\frac{n_1}{N}$一比率，須受$\sqrt{\frac{n_1n_2}{N}}$之標準差之影響，如此數不能略除，則其及於計算結果之影響，必須加以測驗。

如欲測量種痘（不問其中數種影響之存在有無證據）之利益（或弊點），即可比較$\frac{b}{m_2}$及$\frac{a}{m_1}$二痊愈率而得，但此法必須具有此等痊愈率之說明及其標準差，舍此二者，無他法可以直接求得。

相聯之存在及其測算之問題，如每種品質之交替數，並不如是之簡單，則變為複雜之問題。在複雜之情形下，吾人乃有若干不同之種類，如父子髮之色質有多種類別是。

皮爾生教授，為此種事實之測算，特設『相依係數』(coefficient of contingency)。(參閱 Drapers' Company Research

Memoir, Biometric Series I, 1904;及艾得頓書之第十章)。

觀 察 次 數

第 一 品 質 分 類

第二品質分類				
	a_1	a_2	…………	n_1
	b_1	b_2	…………	n_2
	c_1	c_2	…………	n_3
	…	…	…………	…
	…	…	…………	…
	m_1	m_2	…………	N

設n_1, n_2……爲各行 (line) 之總數，以m_1, m_2……爲各列(column)之總數，並以N爲觀察總次數，如上表所示。

如是，設$n_1/N, m_1/N$爲每一品質之第一類對總數之確實比例，則在無『相聯』時，a_1處出現之最或然數，將爲$a_1=\frac{n_1}{N}\times\frac{m_1}{N}\times N$。同樣，$\alpha_2$……，$\beta_1, \beta_2$……，$\gamma_1\gamma_2$……，亦可爲其他地位算出。

關於$a_1-\alpha_1$, $a_2-\alpha_2$,……$b_1-\beta_1$……各等項分歧，可供爲相聯之測算。因超過之數，與不足之數，二者機會均等，故以用二次方，$(a_1-a_1)^2$……較直線數量爲便。

以上論四項範疇之例爲準，得函數如下:

$$X^2=\frac{(a_1-\alpha_1)^2}{\alpha}+\frac{(a_2-\alpha_2)^2}{\alpha_2}+\cdots\cdots+\frac{(b_1-\beta_1)^2}{\beta_1}+\cdots\cdots+\cdots(111)$$

此函數用於測算，公式代表已知觀察之適切度，亦頗佔相當地位。

相依係數 (coefficient of contingency) 之定義，乃如下述

$$C=\frac{1}{\sqrt{\left(1+\frac{N}{X^2}\right)}}\cdots\cdots\cdots\cdots(112)$$

$X=0$時，則無相聯(associetion)，而$C=0$。

$$\frac{X^2}{N}=\frac{(a_1-\alpha_1)^2}{n_1m_1}+\frac{(a_2-\alpha_2)^2}{n_1m_2}+\cdots\cdots,$$

只隨比率而變，與所測之全數無關。

據艾得頓氏 (Elderton 書第一百四十七頁) 證明，如a_1，a_2……，b_1……等數，出現於常態相關面之相當區分內一項亦不下於一小整數，則C將漸近於r，即相關係數。此種關係乃形成X^2之函數，而C之定義乃由X^2而出。

所得之C值，因所用之區分數而有不同，惟其如此，故其可用為測算之效用大減；幸現已有新法，足以克服此困難矣 (見 Biometrika, Vol. IX pp. 116-139)。

至於得出某一C值，其意義若何，非具有極多試驗之經驗，不能領略也。

抑有進者，C以及前節所論之法，羣類有限制之事例而無可以測量之性質者，用之甚佳。

第十二節　時間數列之相關

前文所論之相關，均為關於同時之兩統計羣類；關於兩個數

列，以$x_1, y_1 \cdots\cdots x_p, y_p \cdots\cdots x_t, y_t$各對爲連續時間中數量之測量，吾人不得不測驗兩數列關係之法。但現所提出者，乃爲x之一值，與其以往及後來之數本有關涉，不能互相獨立，而x與y之關係，僅能在時間上影響及於一般定期之進度，並無密切之聯系。時間上所有數列，幾均有一趨勢(trend)，此等趨勢，無論其依等速進行與否，抑或方向如何，均可有高度之相關係數，甚至數量非彼此獨立時亦有之。例如，介乎

$$1,\ 2,\ 3 \cdots\cdots\cdots\cdots 20$$

$$100, 98, 96 \cdots\cdots\cdots\cdots 62$$

$(x_t = t, y_t = 100 - 2(t-1))$二者之相關係數爲1。

求離中差之相關，必須將時間因素消除。

此外尙有一法(註四)，此法，一則可將每一種數量之修勻線求出；二則可計算每年之觀察值，與修勻線所得數值之差額；三則即以此種差額爲據以測算相關之數量；換言之，即測量此種數量與圖式(見上編第七章第三節)所代表數值之相關。

如該數列特別有週期性，則其結果必至得出屬於週期性之相關，而此以用倒數分析法(harmonic analysis)爲佳。又如此種數列係爲『補償的』(compensated)，則於一正離中差之後，必隨一副離中差，其兆徵即可用相關反應出來。

惟在反覆擺動無一定標準之時，則一年之量數，並無一定之

趨勢，且與臨近各年之量數無關，於是從修勻之線，算出離中差間之相關係數，所測量之『關係』，與前論各羣類間之相關者，完全相同。

以x_p，x_p，爲第i年爲其中心，m爲奇數。則欲求之相關係數，乃在表示$x_p-\bar{x}_p$與$y_p-\bar{y}_p$之相關。

如線在修勻時，係用斐孫方法（Professor Persons, The Review of Economic Statistics, No. I, 1919, Harvard University Press），此法亦可應用。在此法之下，須將m年之平均數，$\bar{y}$，求出（m年之趨勢，須在同一方向，而斜度須無顯著之變動，且如t爲距期間中心之年數，修勻之線須成爲$y-\bar{y}=kt$之形式）。至k之決定，係由於一條件——對此線之離差平方之和必爲一最小值，換言之，即$S\{y_t-(\bar{y}+k_t)\}^2$爲一最低數（y_t爲自中心t年之觀察）——而得。故$k=\frac{Sty_2}{St^2}$。

如$k=2n+1$，$St^2=2(1^2+\cdots\cdots+n^2)=\frac{n(n+1)(2n+1)}{3}$。

此種方法，當趨勢繼續凹下（或凸出），而平均數常大（或小）於觀察值時，用於以移動平均法求出之線上，亦可戰勝一部困難。

更有一法，係克孚女士（F. E. Cave，文見 Royal Society's Proceedings, Vol LXXIV, p. 407, 1904）及虎克（Hooker），見 Statistical Journal, 1905, 自六百九十六頁起）所採用，近

更經皮爾生教授，克斈女士及其他等人，加以發揮。此法並非依觀察值求其相關，而係以相接觀察之差額以求相關，有$m+1$年之時期於此，各年觀察值分別爲$x_0, x_1 \cdots\cdots x_m$及$y_0, y_1 \cdots\cdots y_m$，相關係數卽由以下各對：$x_1-x_0$，$y_1-y_0$，$x_2-x_1$，$y_2-y_1$，$\cdots\cdots x_m-x_{m-1}$，$y_m-y_{m-1}$，算出

因x_1-x_0，$x_2-x\cdots\cdots$各等數量之平均數，爲$\frac{1}{m}(x_m-x_0)$；求r之公式，必須用對平均數之離中差；而此離中差，卽爲某年增值超過平均增值之部分，換言之，卽對平均數之每年變量。如x與y之修勻線，顯著爲凹形（或凸形），相關必由此兆徵而受支配，但如觀察值對一直線反覆擺動無一定標準時，則吾人可得測量相關，而不受時間因素之影響。

爲免除凸凹形之困難，另有一較爲詳密之方法，此卽皮爾生教授所謂變差相關（variate difference correlation）者是也（註五）。此法之基礎，係假定x_p可示爲$x_p=X_p+bt_p+ct_p{}^2+\cdots\cdots$，（$X_p$不受時間影響），而時間影響，以拋物線函數表示之，同理$y_p=Y_p+b't_p+\cdots\cdots$

$$x_{p+1}-x_p=X_{p-1}-X_p+b+c(2t_p+1)+\cdots\cdots$$

虎克氏方法，删略c及其他常數。

用第二級差額，則得

$$x_{p+1}-2x_p+x_{p-1}=X_{p+1}-2X_{p-1}+2c+6+dt_p\cdots\cdots$$

用時復略卻d以及其他常數，使成一嚴格之拋物線形。

由此可知，如將時間因素消除，任何差額間之相關卽等於X_p與Y_p之相關。當進而更高級差額時，相關係數不至受其影響，此種程序便告完成。

雖然，差額往往因觀察時之精度及小數之有效數字關係，除爲首三級尙可用外，其他差額應用此法，不免招致絕大困難。此中關係可由2·6，2·7………等數之二次方，仍取其至小數點後一位時知之。

	Δ	Δ^2
6·8		
7·3	·5	0
7·8	·5	·1
8·4	·6	0
9·0	·6	0
9·6	·6	0
10·2	·6	·1
10·9	·7	

第二級差額，如書全式，當全爲·02。實則此法未免過於詳密，普通統計之觀察用之，似不適宜。茲將各種方法之差別，示之如下：

x量既已發生相關，如取用第四級差額，當爲

$$6\left\{x_0+\tfrac{1}{6}(x_2+x_{-2})-\tfrac{2}{3}(x_1+x_{-1})\right\},$$

式中附數表示自中心向左向右之距離。此極端項，將該式增大許多。

以第五項爲基礎，用移動平均數法，則量數當爲

$$x_0-\frac{1}{5}(x_{-2}+x_{-1}+x_0+x_1+x_2)=\frac{4}{5}\{x_0-\frac{1}{4}(x_2+x_{-2})-\frac{1}{4}(x_1+x_{-1})\}$$

而其極端項又將該式減少許多。

用第八級差額，則得

$$x_0-\frac{4}{5}(x_1+x_{-1})+\frac{2}{5}(x_2+x_{-2})-\frac{4}{35}(x_3+x_{-3})+\frac{1}{70}(x_4+x_{-4})$$

用移動平均數，則爲

$$x_0-\frac{1}{8}(x_1+x_{-1})-\frac{1}{8}(x_2+x_{-2})-\frac{1}{8}(x_3+x_{-3})-\frac{1}{8}(x_4-x_{-4})$$

但第二級差額爲

$$x_0-\frac{1}{3}(x_1+x_0+x_{-1})$$

此與僅用三項之移動平均法所得相同，可知不免於極大之機率差誤(chance error)。

各種方法必須加以詳細研究，並須多用以長經驗。據吾人所知，移動平均法對極端各項，頗有看輕之勢，而差額法則每因觀察之粗率，感受嚴重之影響。總而言之，兩種方法，相關之測量，均以假設爲基礎，非如測量羣類相關時之意義鮮明也。

第十三節 時間數列之圖式比較法

除決定相關量數外，尙有一問題，卽如何以圖式表示其關係。方法可示如下：

以$x_1, x_2 \cdots\cdots x_n, y_1, y_2 \cdots\cdots y_n$爲對移動平均數之離中差（如

第七章第六例之表），或使——在無趨勢時　代表實在量數，復設$\bar{x}$與$\bar{y}$爲其平均數。

構造一圖，表示y值，尺度之決定一以便利爲主，時間作爲橫軸。於是x值在此圖上，以任何尺度，任何原點，均無不可。

以b爲x之原點，以一單位x相當c個單位y。決定b與c時，使距代表x_1、y_1，各對之點，垂直距離平方之總和，須爲最小爲便。（見附錄十）。

換言之，$S\{c(x+b)-y\}^2$，其值爲最小。

對b與c而微分之，得$c=\dfrac{S(x-\bar{x})(y-\bar{y})}{n\sigma_1^2}$，又$c(\bar{x}+b)=\bar{y}$，（$\sigma_1$爲$x$之標準差）。

各離中差之平均數，必在縱標尺上之同一點，且與x之平均數之差，必乘以$\dfrac{r\sigma_2}{\sigma_1}$然後在$y$軸上以平均數爲中心，向上向下分別測量（$\sigma_2$爲$y$之標準差，$r$爲相關係數。

關於此法，1912年出版之統計學報（Statistical Journal）第799-800頁，曾用一例，可以參考。

（註一）　X，Y，U，V之意義，自與本章第三節所用者不同。

（註二）　見公式100。

（註三）　q之意義與前數節所用者相同，並非指第二章之$1-p$也。

（註四）　請參閱統計學報一九〇一年號，第四百八十五頁起，虎克（Hooker）一文。

（註五）　見 Biometrika, Vol. X. 自第一百九十七起，及自三百四十頁起。

第七章　相關例證

茲爲例證上述理論並示明測算步驟起見，特就數種實驗及觀察之結果，一一縷列於下。

在未開始舉例之前，有一先決問題，必須釋明者，卽r之理論值，僅可於無量數觀察值中得之。前論機誤時，曾言 r 由 n 對中算來，去其眞值，必有若干相差，此相差數額之標準差，如自差誤常態尺度(normal scale of error)算出，當爲$\frac{1-r^2}{\sqrt{n}}$。如在例一，相關係數爲$\cdot6$；其變量共有二十四對，則其去眞值之差額當在$\cdot6$之$\frac{1-6^2}{\sqrt{24}}=\cdot13$ 以內；但若差額達$\cdot13$ 之三倍，乃幾爲不可能之事實。反之，如吾人不知眞實係數爲若干，務須以計算值$\pm\frac{1-r^2}{\sqrt{n}}$爲宜。

下舉例證數則之目的，僅在示知從觀察中，求r之計算方法。其他各例，則釋明當觀察量數甚多時，求得各整列之平均數，並與方程式 $y-\bar{y}=r\frac{\sigma_2}{\sigma_1}(x-\bar{x})$（在迴歸線爲直線時，此爲衆平均報之軌跡）比較之法。至於最後一則，以 1000 對之分配，逐一與理論相關面之分配比照。一般言之，x 與 y 之計算，並非由其平均數，而係自一武斷之原點，故 $r=\left(\frac{\mathrm{S}xy}{n}-\bar{x}\bar{y}\right)\div\sigma_1\sigma_2$，見公式 93。

例一：爲求得一實例，以解相關起見，可在所有環境均爲已知，而相關係數可用由果推因方法以事推出時，用一數學表從中任意抽出若干數字。以五個數字之和作爲 x_t，以任一五個數字之和作爲 y_t，但後一五個數字之中，有三個業已被包括於前一五個數字之內，其餘二個數字，則爲不同，如此共湊成二十對 (x_1y_1) ………(x_ty_t)………。此等各對之相關係數，爲 $\frac{3}{5}$(公式96)。在一僅取二十四對之實例中，相關係數爲 0·537；相關係數 $\frac{3}{5}$ 之標準差爲 $\frac{1-\cdot 36}{\sqrt{24}}=\cdot 13$，故對若是其小之數目，所差並不甚巨。

計算方法，可以下列一表示明之。

x	y	x^2	y^2	xy
22	32	484	1,024	704
27	27	729	729	729
12	19	144	361	228
21	30	441	900	630
21	26	441	676	546
27	26	729	676	702
23	25	529	625	575
17	22	289	484	374
25	23	625	529	575
11	9	121	81	99
16	24	256	576	384
20	28	400	784	560
37	29	1,369	841	1,073
33	25	1,089	625	825
18	20	324	400	360
24	26	576	676	624
22	17	484	289	374
17	16	289	256	272
32	27	1,024	729	864
29	29	841	841	841
26	17	676	289	442
27	20	729	400	540
26	26	676	676	676
21	17	441	289	357
554	560	13,706	13,756	13,354

$n=24 \quad \bar{x}=23\frac{1}{12} \quad \bar{y}=23\frac{1}{3}$

$24\sigma_1^2=13706-24(23\frac{1}{12})^2$

$\sigma_1=6.18$

$\sigma_2=5.36$

$$r=\frac{13354-24\bar{x}\bar{y}}{24\sigma_1\sigma_2}$$

$=.537$

如 x 與 y 各以23爲原點而計算，其計算手續必較簡易。

例二：吾人如有稀少而散見之觀察值，則將全部完全計算出來亦頗簡易。舉例言之，玆將二十六個城市之嬰兒死亡率，列如次表，以與此二十六個城市之人口（以一千爲準）比較。然則 r 僅當其標準差之兩倍，因而其準確數值亦欠一定，但吾人可得而證明者，城市愈大者，死亡率亦愈高。如欲切實硏究嬰兒死亡原因問題，必須多舉事例，並考慮除人口略數以外之其他種因素。

二十六個城市之人口與嬰兒死亡率

	人口 x	死亡率 y	xy
	000.		
	55	162	8,910
	39	201	7,839
	36	241	8,676
	35	162	5,670
	31	179	5,549
	30	174	5,220
	27	176	4,752
	24	208	4,992
	24	163	3,912
	23	206	4,738
	22	172	3,784
	20	200	4,000
	19	218	4,142
	19	198	3,762
	19	132	2,508
	16	155	2,480
	15	148	2,220
	15	220	3,300
	15	141	2,115
	12	169	2,028
	7	155	1,085
	6	129	774
	6	167	1,002
	5	150	750
	5	171	855
	4	161	644
總　計	.529	4,558	95,707
平均數	20.354	175.31	—
	$\sigma_1=12.1$	$\sigma_2=27.9$	

$$r=\frac{Sxy-n\bar{x}\bar{y}}{26\sigma_1\sigma_2}$$

$$=\frac{95707-26\times20.354\times175.31}{26\times12.1\times27.9}=.34$$

$$r\text{ 之標準差}=\frac{1-.34^2}{\sqrt{26}}=.17$$

例三：北海漁業調查所(North Sea Fisheries Investigation)，之青魚統計，可爲一良好例證。該種青魚身上具有圈輪 (ring)，每年生一輪，故可爲年齡之表明，此項統計即以魚之體量與圈輪之關係爲題材。

原始曲線雖有斜度；但各整列之平均數，均甚臨近理論上之迴歸直線。

表上，y 軸代表魚之體量，原點爲 31 cm. 單位爲 1 cm. 圈輪數以 x 軸爲代表，以七輪爲原點，以一輪爲單位，出現次數則填於方格表內。

n_2 爲 y 某一種値之次數，最末二列記 n_2y 與 n_2y^2 及其總和，以便計算圈輪之平均數及標準差。同樣，以 n_1 爲某一 x 整列之總次數，由 n_1x 與 n_1x^2 之各該總和，算出青魚體量之平均數及標準差。

最後一行，列明每一整列之平均數，此乃在每一 x 整列中，用其出現次數乘相當之 y 值而得。

在每一次數之下，在括弧中，註明相當之 $x\times y$ 值；例如，在 $x=-1-$ 列下，$y=3$ 一排 (row)共有四件，而 $xy=-3$，故四件與 xy 之總值，相乘而得 $4\times-3=-12$。如此乘得之各項，分別彙類下列四象限內。

至於原點之決定，以使 Sxy 之各零項，愈多愈妙（請參照猶

爾氏著：統計原理概論第一百八十三頁）。

青魚之圈輪數及體量（長度單位為公分）

圈輪數	x	4 −3	5 −2	6 −1	7 0	8 1	9 2	10 3	11 4	12 5	13 6	總計 n_2	$n2y$	$n2y2$
Size. cm.	y													
35	4	—	—	1 (−4)	—	1 (4)	2 (8)	2 (12)	—	—	—	6	24	96
34	3	—	1 (−6)	4 (−3)	4	15 (3)	14 (6)	7 (9)	3 (12)	1 (15)	1 (18)	50	150	450
33	2	1 (−6)	1 (−4)	11 (−2)	26	26 (2)	22 (4)	11 (6)	3 (8)	3 (10)	1 (12)	105	210	420
32	1	1 (−3)	24 (−2)	49 (−1)	53	26 (1)	7 (2)	5 (3)	1 (4)	— —	— —	166	166	166
31	0	—	28	43	45	21	6	2	—	—	—	145	0	0
30	−1	1 (3)	15 (2)	21 (1)	16	7 (−1)	1 (−2)	—	—	—	—	61	−61	61
29	−2	2 (6)	3 (4)	5 (2)	—	—	—	—	—	—	—	10	−20	40
28	−3	1 (9)	1 (6)	3 (3)	2	—	—	—	—	—	—	7	−21	63
27	−4	—	3 (8)	—	—	—	—	—	—	—	—	3	−12	48
26	−5	—	1 (10)	—	—	—	—	—	—	—	—	1	−5	25
總計 n_1		6	77	137	146	96	52	27	7	4	2	554	431	1,369
	n_1x	−18	−154	−137	0	96	104	81	28	20	12	$Sn_1x=$	32	
	n_1x^2	54	308	137	0	96	208	243	112	100	72	$Sn_1x_2=$	1330	
各整列之平均數		30.17	30.85	31.34	31.65	32.25	32.92	33.07	33.3	—	—			

$\bar{x}=32\div554=\cdot0578$。平均數，$7\cdot0578$ 輪。

$\sigma_1^2=1330\div554-\cdot0578^2-\cdot083$（註一）$\sigma_1=1\cdot521$。

$\bar{y}=431\div554=\cdot778$。平均數，$31\cdot778$ 公分。

$\sigma_2^2-1369\div554-\cdot778^2-\cdot083$（註二）$\sigma_2=1\cdot335$。

Sxy ++		−−	+−	−+
4	52	3	7	4
16	88	30	2	6
24	66	21	—	12
45	24	12	9	6
84	30	12		4
63	12	10		22
36	26	9		3
15	14	6		48
18	15	9		49
	4	24		
		10		
636		146		154

$Sxy=636+146-9-154=619$

$r=\dfrac{Sxy-554\bar{x}\bar{y}}{554\sigma_1\sigma_2}=.528$

r 之標準差 $=.026$

$$\frac{\text{體長}-31.778\text{公分}}{1.335\text{公分}}=.528\frac{\text{輪數}-7.0578}{1.521}$$

輪數	從方程式推算之體長公分	整列之平均數(公分)
4	30.36	30.17
5	30.82	30.85
6	31.29	31.34
7	31.75	31.65
8	32.21	32.25
9	32.68	32.92
10	33.14	33.07
11	33.60	33.3

例四：下舉一例，釋明相關甚小時，求 r 所得之值。以七位對數表，2500–2549，及 2600–2649，二組中每一數之末位字爲 x；以比 x 大 50 之對數末一位字爲 y，換言之，卽以 2550–2599 及 2650–2699 之每數末一字爲 y。求得 r 爲 .086，此值在一百對中乃小於其標準差。

茲又採用一算術計算法，此法較之本章之其他方法均爲簡便，且可卽時算出相關比率。

成對數字之出現數

x	y 0	1	2	3	4	5	6	7	8	9	n_s	Sy	xSy	$n_s\bar{y}_s^2$
0	—	—	1	—	1	1	1	—	—	4	8	53	0	351
1	3	—	—	1	1	—	—	1	1	1	8	31	31	120
2	—	2	1	—	3	1	—	—	—	1	8	30	60	112
3	2	—	—	1	—	1	—	2	2	3	11	65	195	384
4	1	2	2	—	—	2	1	3	1	—	12	51	204	217
5	1	1	—	—	1	1	4	1	—	—	9	41	205	187
6	1	1	1	2	—	1	—	2	1	—	9	36	216	144
7	3	3	2	1	1	3	1	—	3	1	18	68	476	257
8	—	2	—	3	2	1	—	—	2	1	11	49	392	218
9	—	—	—	1	—	—	—	1	1	3	6	45	405	338
	11	11	7	9	9	11	7	10	11	14	100	469	2,184	2,328

$\bar{x}=4.72$, $\sigma x=2.69$, $\bar{y}=4.69$　$\sigma y=3.03$, $n=100$,

$S(x-\bar{x})(y-\bar{y})=Sxy-100\,\bar{x}\bar{y}=S(xSy)-100\,\bar{x}\bar{y}=2184-2114=70$

$r=\dfrac{70}{100\sigma x\sigma y}=.086$。　r 之標準差爲.1。

茲以 n_s 爲各個 x 數 0, 1……出現之次數。Sy 爲相當之 y 值總和;例如,第一行爲 $2+4+5+6+9\times4=53$。$\bar{y}_s$ 爲對某一 x 值之 y 值平均數,等於 $Sy\div n_s$,

又　$$n_s\bar{y}_s^2=(Sy)^2\div n_s。$$

相關比率可由最末一列算出(見第六章第八節)。

$$100\sigma_m^2=Sn_s(\bar{v}_s-\bar{y})^2\text{(註三)}=Sn_s\bar{y}_s^2-2\bar{y}Sn_s\bar{y}_s+n\bar{y}^2$$

$$=Sn_s\bar{y}_s^2-n\bar{y}^2=2328-2200=128。$$

$\eta=\dfrac{\sigma_m}{\sigma_y}=\dfrac{1.13}{3.03}=.37$,相關係數之值雖小,而相關比率卻甚大。

例五:下表資料,係根據『紐約市兒童之身長體重報告』一文,共有兒童 3,405 人,年齡在 14–15 之間。

身長	人數	報告中之平均體重	—	—	由方程式算出來之體重	—
原點 61 吋 x	ns	原點 100 磅 $\bar{y}s$	$ns\bar{y}$	$x.ns\bar{y}s$	原點 100 磅 —	$ns\bar{y}s^2$
−12	1	−12	− 12	144	−49·6	144
− 9	1	−20	− 20	180	−36·6	400
− 7	13	−18	− 234	1,638	−27·9	4,212
− 6	59	−19	−1,121	6,726	−23·6	21,299
− 5	96	−17	−1,632	8,160	−19·2	27,744
− 4	190	−14	−2,660	10,640	−14·9	37,240
− 3	283	−12	−3,396	10,188	−10·5	40,752
− 2	349	− 8	−2,792	5,584	− 6·2	22,336
− 1	440	− 3	−1,320	1,320	− 1·9	3,960
0	434	+ 2	+ 868	0	+ 2·5	1,736
+ 1	400	+ 7	+2,800	2,800	+ 6·8	19,600
+ 2	355	+11	+3,905	7,810	+11·2	42,955
+ 3	307	+17	+5,219	15,657	+15·5	88,723
+ 4	200	+20	+4,000	16,000	+20·0	80.000
+ 5	137	+24	+3,288	16,440	+24·2	78,912
+ 6	78	+30	+2,340	14,040	+28·6	70,200
+ 7	34	+35	+1,190	8,330	+32·9	41,650
+ 8	15	+34	+ 510	4,080	+37·2	17,340
+ 9	6	+42	+ 252	2,268	+41·6	10,584
+10	7	+42	+ 294	2,940	+45·9	12,348
總計	3,405	—	11,479	134,945	—	622,135

$\bar{x}=\cdot2270$, $\sigma_1=2\cdot99$。　　平均數 61.227 吋。

$\bar{y}=3\cdot371$, $\sigma_2=16\cdot3$（註四）　　平均數 103·37 磅。

$$r=\frac{Sxy-n\bar{x}\bar{y}}{n\sigma_1\sigma_2}=\left(\frac{134945}{3405}-\cdot7652\right)\div\sigma_1\sigma_2=\cdot797$$

迴歸方程式為$\dfrac{\text{體重}-103\cdot37}{16\cdot3}=\cdot797\times\dfrac{\text{身長}-61\cdot227}{2\cdot99}$，　或

體重$=103\cdot4+4\cdot345$（身長$-61\cdot23$）。

由此方程式求得之體重，列於上表之第六列，須用以與相當第三列各項身長之平均體重對照。公式所得之值與實在相合甚近，約自 57 吋至 70 吋；但在 57 吋以下，實在體重並不如公式之急轉直下。迴歸線實際上對低等身材者，並非爲一直線。

$$3405\sigma_m{}^2 = \mathrm{S}n_s\bar{y}_s{}^2 - 3405\bar{y}^2,$$

$$\sigma_m = 13 \cdot 1$$

$$\eta = \sigma_m \div \sigma_2 = \cdot 81。$$

此處相關率與相關係數，實係一數。

例六：第六章第十二節所論之時間數列相關，玆以英國聯合王國之按年平均每人進口貨值，與英格蘭與威爾士之各年結婚率爲例以明之。

以 x 爲各年超過五年平均數之值，每五年以平均數爲中心。由結婚率，計算 y 值亦然。

$$\bar{x} = -\cdot 62,\quad \bar{y} = \cdot 3,\quad \sigma_1 = 36\cdot 9,\quad \sigma_2 = 3\cdot 61,\quad \mathrm{S}xy = 4309,$$

$$n = 50,\qquad r = \frac{4309 \div 50 - \cdot 3 \times \cdot 62}{36\cdot 9 \times 3\cdot 61} = \cdot 65,$$ 其標準差爲

$$\frac{1 - \cdot 65^2}{\sqrt{50}} = \cdot 09。$$

此處相關之計算，卽係用移動平均法。

爲比較差額以求相關起見，玆將最初五年差額列表於下，其他各行則從略。

		進口貨值			結婚率			
	X	DX	D^2X	Y	DY	D^2Y	DX.DY	$D^2X.D^2Y$
1845	330	−15	—	172	0	—	0	—
1846	315	+6	+21	172	−14	−14	−84	−294
1847	321	−30	−36	158	+1	+15	−30	−540
1848	291		+91	159		+2		+182

	DX	D^2X	DY	D^2Y
平均數	15.74	1	−.19	.04
標準差	57.13	80	5.3	6.78

DX.DY 之和 =8902。$D^2X.D^2Y$ 之和 =12076。

故從第一級差額得 r 爲 .60，從第二級差額得 r 爲 .45。

例七：依第三章第七節第一例所舉之實驗，以十字字母之數相加，共有 1,000 個總和。茲以前五字之字總和爲A，B爲次五字字母之和，於是 $x=A+B$。在每十字取出之後，更取出五字，將五字字母之和定爲C；於是 y 乃爲 $B+C$。如此共得 1,000 對，此千對之相關係數必爲 $\frac{1}{2}$，其標準差爲 ·024。

但實際上相關係數乃爲 ·553，超過所期值有兩標準差以上。此中原因，或可以歸咎於缺乏完全獨立性爲解釋。以 250 總和爲一類（其標準差爲·047），則四類之相關係數爲·56，·50，·58，·59。

迴歸線在中心部分自 $x=40$ 至 $x=61$，幾爲一條直線；此數之外，每一 x 值出現不下於 20 次；而一盤列平均數之標準差大於 2，故比較方法，不無相當價值。y_x 值之所有標準差，爲自 1·3

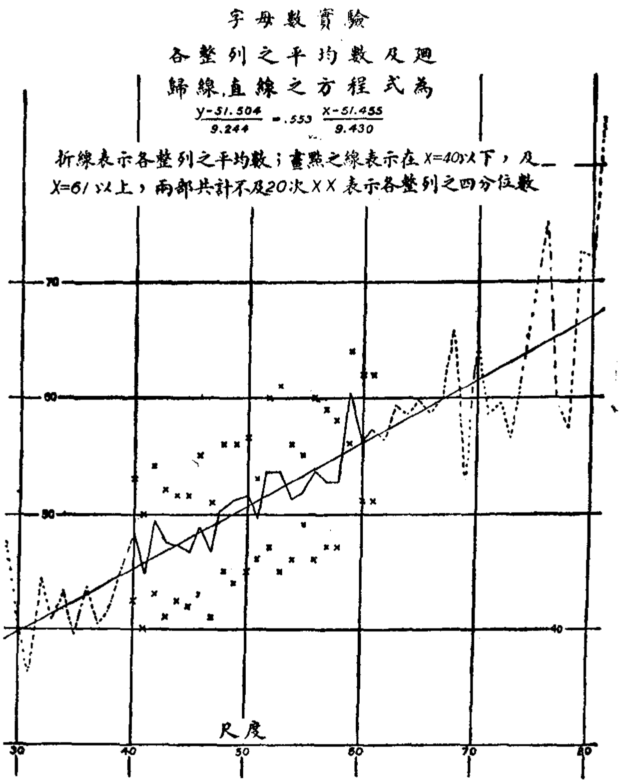
至 2·0。

x 值	相當之 y 值平均數	y 之四分位差	由方程式算出之 $\bar{y}s$
40	48·3	$10\frac{1}{2}$	45·3
41	44·8	10	45·8
42	49·1	11	46·4
43	47·4	11	46·9
44	47·1	9	47·5
45	46·4	$9\frac{1}{2}$	48·0
46	48·9	12	48·5
47	46·4	10	49·1
48	50·2	11	49·6
49	51·1	12	50·2
50	51.5	$11\frac{1}{2}$	50·7
51	49·6	7	51·3
52	53·6	13	51·8
53	53·6	16	52·3
54	51·1	10	52·9
55	51·9	6	53·4
56	53·4	14	54·0
57	52·7	12	54·5
58	52·7	11	55·1
59	60·2	8	55·6
60	56	11	56·1
61	57·2	11	56·7

四分位差，從理論得來（$2\sigma\sqrt{1-r^2}$ 之 ·67），依公式 26及107，其值爲 105；而觀察得來之四分位差，平均爲 10·75。此似與 x 值無關，果如理論（公式 107）預期。

此等數之適應性，圖上表示無遺，至圖上之迴歸方程式乃為

$$\frac{y-51\cdot50}{9\cdot24}=\cdot553\frac{x-51\cdot46}{9\cdot43}\text{。}$$



（第一圖）

如分級時，將羣類分為較寬之級，以便將由抽樣而來之差誤

消滅。則迴歸之數字表示,可以較爲明鮮。如:

y 之分級	次　數	y 之平均數	相當之 x 平均數	方程式得來之 x
30－39	85	36·3	43·5	42·9
40－49	348	44·7	47·7	47·6
50－59	360	54·3	52·7	52·0
60－69	173	63·0	57·7	58·0
70－79	30	72·7	64·1	63·4

此處係指 x 對 y 之迴歸而言,而上圖則爲 y 對 x 之迴歸。

在 30 以下及 80 以上尙有二數未計。

比較觀察分配,與常態相關面之分配,其方法甚多, 茲擇其單簡者介紹如下:

設眞 $r=\frac{1}{2}$, x 與 y 標準差之平均值爲 $\sigma=9{\cdot}32$。則相關面方程式爲

$$z=\frac{1}{2\pi\sigma^2\sqrt{1-r^2}}e^{-\frac{1}{2(1-r^2)\sigma^2}(x^2+y^2-xy)}\cdot$$

但　$x=\dfrac{X-Y}{\sqrt{2}}, y=\dfrac{X+Y}{\sqrt{2}}$。

方程式變爲 $z=\dfrac{1}{\pi\sigma^2\sqrt{3}}e^{-\frac{X^2}{3\sigma^2}}\cdot e^{-\frac{Y^2}{\sigma^2}}$,

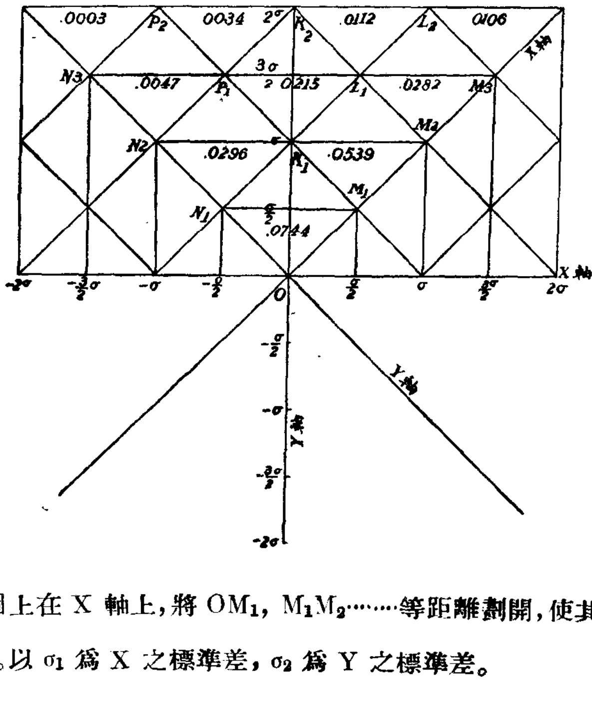
此代表對於主軸之平面,對中心軸斜角有 45°。

以 $X=X_1$, $X=X_2$, $Y=Y_2$, $Y=Y_2$ 爲限界之面積上,其

容積爲

$$\frac{1}{\sqrt{2\pi}\cdot\sigma\sqrt{\frac{2}{3}}}\cdot\int_{X_1}^{X_2} e^{-\frac{X^2}{2(\sigma\sqrt{\frac{3}{2}})^2}}dx\cdot\frac{1}{\sqrt{2\pi}\cdot\sigma\sqrt{\frac{1}{2}}}\int_{Y_1}^{Y_2} e^{\frac{Y^2}{2\left(\frac{\sigma}{\sqrt{2}}\right)^2}}dY,$$

如用第二章之常態機率表,不難立卽求得。



圖上在 X 軸上,將 OM_1, M_1M_2……等距離劃開,使其各等於 $\frac{\sigma}{\sqrt{2}}$。以 σ_1 爲 X 之標準差,σ_2 爲 Y 之標準差。

$$\sigma_1=\sigma\sqrt{\tfrac{3}{2}},\qquad \sigma_2=\sigma/\sqrt{2}。$$

於是 $OM_1=M_1M_2=\cdots\cdots=\sigma_1/\sqrt{3}=.577\sigma_1$。

對X軸垂直之平面,構成一立體形,此立體形之各部分容積,

經 OM_1 爲 $F(\cdot577)$，經 M_1M_2 爲 $F(1\cdot155)-F(\cdot577)$，其他以此類推，此在常態機率表中查出，乃爲·2180，·1580……。

次復將Y軸劃爲 ON_1，N_1N_2……等距離，使各等於 $\frac{\sigma}{\sqrt{2}}$，即使等於 σ_2 也。

以對 Y 軸垂直之平面爲限界，所構成之容積，在 ON_1 爲 $F(1)$，在 ON_2 爲 $F(2)-F(1)$，餘類推，即等於·3413，·1359……。

因方程式中，X與Y之積分，各自獨立，垂直於X軸之各部，必在 N_1，N_2……等處，被平面割爲相等部分。故組成下表，表明在第一線之平方 $ON_1K_1M_1$，$M_1K_1L_1M_2$，$M_2L_1L_2M_3$，……上，第二線平方 $N_1N_2P_1K_1$，$K_1P_1K_2L_1$……（其他各線以此類推），常態平面之容積各部分。

常態頻數平面之分配

X/σ_1		·577	1·155	1·732	2·31	2·89	3·46
$F(X/\sigma_1)$（差額）		·2180	·1580	·0824	·0311	·0085	·0017
Y/σ_2	$F(Y/\sigma_2)$ 差額	差額乘積					
1	·3413	·0744	·0539	·0282	·0106	·0028	·0006
2	·1359	·0296	·0215	·0112	·0042	·0012	·0002
3	·0214	·0047	·0034	·0017	·0007	·0002	·0000
4	·0014	·0003	·0002	·0001	·0001	·0000	·0000

OX 與 OY 二軸所成之四象限,各個分配狀況全同。表中之小數以 1000 乘之,卽得 1000 對數之理論上的分配,假如置偏斜度於不論。

用方格紙將觀察值逐一記下,並將在X,Y各平方內出現次數註明。結果書如次表,表中各排之第一行,重記已知之理論數,第三行為觀察數。

理論數與觀察數,在中心左右三平方內,中心上下兩平方內,頗能脗合,以在 $\pm 1.7\sigma_1$ 與 $\pm\sigma_2$ 之內也。隨機抽樣中分歧若此之機率,約為 $\frac{3}{5}$ (見第十章例釋)。

自此方格再向左,情形乃全不同 (觀察數為 31, 而期望數為 41),向右則又多出許多 (觀察數 54,期望數41)。但十二個方格自中心向左,略有盈餘,向右則有不足。此恰與吾人對於原始曲線 (見第三章第七節第一例) 之偏斜度所預料之現象相彷。茲將偏斜度之影響,另於本章附註中討論之,而修正後之結果,卽列入下表各排之第二行。改進之效果甚為顯著;例如最末三列向之期望數,現為 $33\frac{1}{2}$ (觀察值為 3·1), 最末三行向右,則為 49 (觀察值為 54)。

1000 對字母之和

觀察值分配與常態及與偏態頻數之比較。

中心之縱橫直線,並非坐標軸,實成 45° 斜角之對稱軸也。

第一行　常態分配………（例如 29·6）

第二行　第二近似值……（例如 28·7）

第三行　觀察值…… ……（例如 35　）

0 0 0	0 0 0	·1 ? 0	·1 ? 0	·2 0 0	·3 ·2 1	·3 ·4 0	·2 ·4 1	·1 ? 0	·1 ? 0	0 0 0	0 0 0
0 0 0	·2 ? 0	·7 0 0	1·7 ·2 1	3·4 1·9 1	4·7 4·3 1	4·7 5·1 5	3·4 4·9 6	1·7 3·2 4	·7 1·7 0	·2 ? 0	0 0 0
·2 ? 0	1·2 ·2 0	4·2 2·8 5	11·2 10·8 11	21·5 22·3 20	29·6 30·5 33	29·6 28·7 35	21·5 20·7 20	11·2 11·6 11	4·2 5·6 5	1·2 2·2 0	·2 ? 0
·6 0 0	2·8 2·7 2	10·6 11·1 12	28·2 33·3 34	53·9 63·0 47	74·4 79·2 77	74·4 69·6 72	53·9 44·8 61	28·2 23·1 24	10·6 10·1 8	2·8 2·9 5	·6 1·6 1
·6 0 0	2·8 2·7 1	10·6 11·1 9	28·2 33·3 36	53·9 63·0 75	74·4 79·2 76	74·4 69·6 64	53·9 44·8 46	28·2 23·1 26	10·6 10·1 15	2·8 2·9 5	·6 1·6 4
·2 ? 0	1·2 ·2 1	4·2 2·8 1	11·2 10·8 8	21·5 22·3 21	29·6 30·5 32	29·6 28·7 26	21·5 20·7 18	11·2 11·6 9	4·2 5·6 8	1·2 2·2 0	·2 ? 2
0 0 0	·2 ? 0	·7 0 0	1·7 ·2 0	3·4 1·9 3	4·7 4·3 2	4·7 5·1 2	3·4 4·9 0	1·7 3·2 3	·7 1·7 1	·2 ? 0	0 — 0
0 0 0	0 0 0	·1 ? 0	·1 ? 0	·2 0 0	·3 ·2 0	·3 ·4 2	·2 ·4 1	·1 ? 0	·1 ? 0	0 0 0	0 0 0

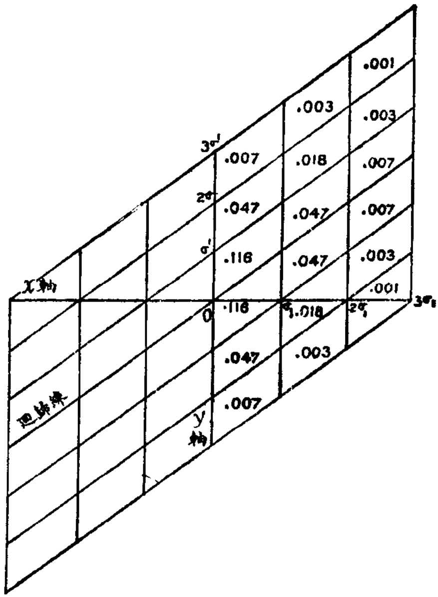
與整個期望値分歧之機率，據測驗結果（見第十年例釋），約等於$\frac{3}{5}$；換言之，如此之實驗，五次只有二次，可望其密切相合（註五），反之，如分配果爲常態及對稱，向左向右所得如是之分歧，幾全爲不可能之事；爲補充此理論，必須求第二近似值也。

關於觀察値之分配，與常態平面下之分配，是否能以一致之問題，簡單之測驗方法，可由分析x一整列之分配而得，無須如上段之須變爲對稱軸也。

$$\iint \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} e^{-\frac{1}{2(1-r^2)}\left(\frac{x^2}{\sigma_1^2}+\frac{y^2}{\sigma_2^2}-\frac{2rxy}{\sigma_1\sigma_2}\right)} dxdy$$

$$= \int \frac{1}{\sigma'\sqrt{2\pi}} e^{-\frac{y'^2}{2\sigma'^2}} dy' \times \int \frac{1}{\sigma_1\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma_1^2}} dx,$$

式中所用之 $\sigma'=\sigma_2\sqrt{1-r^2}$, $y'=y-\frac{rx\sigma_2}{\sigma_1}$,換言之,$y'$ 係自迴歸線上與 y 軸平行,如公式 107 所論。



	F(z) 0–1	F(z) 1–2	F(z) 2–3
	•341	•136	•022

(Fz)		乘積		
0–1	•341	•116	•047	•007
1–2	•136	•047	•018	•003
2–3	•021	•007	•003	•001

以標準差作區分之單位，示如上表及圖。就此爲基準，將字母實驗之結果，列表如下：

		-3σ	-2σ	$-\sigma$	0	σ	2σ	3σ	
	0	0	0	0	0	0	0	0	
	0	0	0	0	0	0	0	0	
$3\sigma'$									
	0	1	3	7	7	3	1	0	
	0	0	0	4	8	2	0	0	
$2\sigma'$									
	0	3	18	47	47	18	3	0	
	0	1	17	63	56	18	4	1	
σ'									
	1	7	47	116	116	47	7	1	
	0	1	57	131	114	42	6	0	
0									迴歸線
	1	7	47	116	116	47	7	1	
	0	7	49	98	103	48	9	1	
$-\sigma'$									
	0	3	18	47	47	18	3	0	
	1	2	13	47	50	16	0	0	
$-2\sigma'$									
	0	1	3	7	7	3	1	0	
	0	2	3	10	7	3	0	0	
$-3\sigma'$									
	0	0	0	0	0	0	0	0	
	0	0	1	0	3	1	0	1	

縱列示 $x-$ 整列，可與本章例七之詳細表對照參看。每一間隔，上面一數爲計算得來之數值，而下面之數乃係觀察之值。

對於偏斜度之修正，可以在同一方向改進適應程度，與以前之表同。

附註：——求相關面之第二近似值

當 $\frac{1}{\sqrt{n}}$ 次之各項，依一般大數律，加以保留時，一含有差誤之平均立方（mean cube of error）之項，必現諸方程式上（見第三章第五節）。同理，據愛基華斯教授之證明（註六）相關面之方程式，在同樣條件下，應書如

$$z=z_0-\frac{1}{6}\left(k_{30}\frac{\partial^3}{\partial x^3}z_0+3k_{21}\frac{\partial^3}{\partial x^2\partial y}z_0+3k_{12}\frac{\partial^3}{\partial y\partial x^2}z_0+k_{03}\frac{\partial^3}{\partial y^3}z_0\right)\cdots(113)$$

此處 $$z_0=\frac{1}{2\pi\sqrt{1-r^2}}e^{-\frac{1}{2(1-r^2)}(x^2+y^2-2rxy)},$$

x 與 y 爲觀察值與平均數（被其數準差除後）之差額。

$k_{30}=$ 平均值 x^3，$k_{21}=$ 平均值 x^2y，

$k_{12}=$ 平均值 xy^2，$k_{03}=$ 平均值 y^3。

如本章七例，$x\sigma=A+B$，$y\sigma=B+C(\sigma=9.44$（約略數）。

平均值 $xy\sigma=$ 平均值 $B^2=\frac{1}{2}\sigma^2$，又 $r=\frac{1}{2}$（在實驗中得 $r=.55$）。

平均值 $x^3\sigma^3=k_{30}=k_{03}-='k'$（見第一章第三節）$=.409$。

平均值 $x^2y\sigma_3=k_{21}=$ 平均值 $B^3=\frac{1}{2}$ 平均值 $(A+B)_3=\frac{1}{2}'k'=k_{12}$。

如將此等數值微分時，則得

$$z=\frac{1}{\pi\sqrt{3}}e^{-\frac{2}{3}(x^2+y^2-xy)}\left\{1-.01(x+y)(18+11xy-8x^2-8y^2)\right\}=z_0(1-w)。$$

z_0w 一式不易加以積分，較爲簡易之步驟，係對適當之面積上，將 z_0 積分，然後約略加以修正。西普生氏定則(Simpson's rule)所用之方法，如

z_0，k，z_0，$-k$，zh，0，$z-h$，0，

爲長方形上平面四角之縱坐標，而長方形之對角線爲

$2h$，$2k$，

z_{00} 爲長方形中心之縱坐標；則平均縱坐標爲

$\frac{1}{6}(2z_{00}+z_0k+z_0-k+zh_0+z-h_0)$。

故如將 z_0w 算出以求前節常態頻數面平方之分配表各容積之中心及四角，則須用一量數 z_0w'（w' 爲中心值兩倍及四角度之平均數）使每一容積復原。如此通盤施行，而將所得之值，加上或從常態曲線得來之數減去，以計算前段之修正值。

吾人須加注意者，即當平面係指其主軸而言，因而以 $x+y=\sqrt{2}X$，$-x+y=\sqrt{2}Y$，則 w 在 Y 軸上變爲對稱，但非在 X 軸。

（註一）
（註二）薛伯氏校正，見附錄五。

（註三）第六章第八節，y 各值由其平均數計算，此處須通盤減去 $\bar{y}$。

（註四）計算根據之資料，此處未曾書明。

（註五）　表中重線，僅爲表示應用第十章測驗方法之區分。

（註六）　差誤律 (Camb. Phil. Trans., Vol. XX, 1905, Part II, § 6)，及統計學報 (Statistical Journal, 1917，第 268 頁起。愛基華基氏，在教本中不用模差(modulus)，而用標準差作單位。

第八章　淨相關與複相關

第一節　淨相關

依第七章所論之調查，可知一量數之變化，足以影響另一量數之變化。然情況尚不止此而已，一量數之變動，往往波及於其他多數之量，因亦隨之而變動。故頻數分配不能再以三元(dimension) 之平面爲代表矣。關於此者，須求一類似之函數，前論者不過爲其簡單之情形耳。

討論至此，前言之迴歸方程，已非直線或曲線之方程式可比，此時之方程，乃一變量，與其他若干變量發生關係者也。斯時可將現有之許多變量，用類似『局部微分』(partial differentiation) 方法，剔除其中之一，使其不至再生影響。如此一一提出，然後求其所餘兩個變量之相互關係，此即極爲重要之淨相關 (partial correlation) 法也。

下文討論三個變量之例，詳細解剖，而將其較爲普遍之解法，提綱舉要，加以說明。

設有 x, y, z, 三個變量於此，各由其平均數算來，並設其彼此互有相關。假如三者關係如此：$z=ax+by+c$，此爲一理論之

平面，由此平面所得 z 之平均值，與 x, y 一一對之值相應。

現求出 a, b, c, 求時以使 z 之觀察值 (observed value) 對由方程式求得值之觀察離中差，具有最低之反機率(improbability)爲妙。

設 $\bar{z}_s$ 爲 k_s 個觀察之平均數，每種觀察，x 有 x_s (到 $x_s+\delta_x$) 個成員，y 有 y_s (到 $y_s+\delta_y$) 個成員。

以 η_s 代 $\bar{z}_s-(ax_s+by_s+c)$, 此即第 s 個觀察羣類平均數對其理想值(ideal value)之離中差。

以 σ_x, σ_y, σ_z 爲 x, y, z 各頻數羣類之標準差。

於是如在長久經驗之後，z 羣類之標準差，與 x, y 之值無關，則 η_s 之標準差必爲 $\frac{\sigma_z}{\sqrt{k_s}}$ (公式38)，而 η_s (到 $\eta_s+\delta\eta$) 出現之機率，爲 $ke^{-\frac{k_s n_s^2}{2\sigma z^2}}.\delta\eta$

設有 n 對數值，如 x_s, y_s, 而共有觀察N個，則

$$N=k_1++k_s+\cdots\cdots+k_n。$$

$\eta_1\cdots\cdots\eta_s\cdots\cdots\eta_n$出現之機率，爲$Ce^{-\frac{1}{2\sigma_2^2}\phi}$, 式中 $\phi=k_1\eta_1^2+\cdots\cdots k_s\eta_s^2+\cdots\cdots+k_n\eta_n^2$，而 C 爲常數。

$\phi_s=S_1^n k_s[z_s-(ax_s+by_s+c)]^2$爲最小時，機率必最大，茲求出 a, b, c, 以達此目的。

$$\phi=S(k_s\bar{z}_s^2)+a^2S(k_sx_s^2)+b^2S(k_sy_s^2)+c^2Sk_s-2aS(k_sx_s\bar{z}_s)-2bS(k_sy_s\bar{z}_s)-2cS(k_s\bar{z}_s)+2abS(k_sx_sy_s)+2acSx_s+2bcSy_s。$$

但 $Sx_s=0=Sy_s$。$Sk_s\bar{z}_s=z$ 所有各值之和$=0$。

$Sk_sx_s^2=N\sigma_x^2$，x_2 在全羣類重現 k_s 次，又 $Sk_sy_s^2=N\sigma_y^2$ 同。

$Sk_sx_s\bar{z}_s=Sxz$，$\therefore k_s\bar{z}_s=$ 在第 S 羣類中 z 所有值之和。

又 $sk_sy_s\bar{z}_s=Syz$。 又 $Sk_sx_sy_s=Sxy$。

$\therefore\ \phi=S(k^s\bar{z}_s^2)+Na^2\sigma_x^2+Nb^2\sigma_y^2-2aSxz-2bSyz+2abSxy$
$+Nc^2$。

以 $Sxy=Nr_{xy}\sigma_x\sigma_y$，$Sxz=Nr_{xz}\sigma_x\sigma_z$，$Syz=Nr_{yz}\sigma_y\sigma_z$。

然則 ϕ 必爲最小，如$\frac{\partial\phi}{\partial a},\frac{\partial\phi}{\partial b},\frac{\partial\phi}{\partial c}$各爲零（註一），換言之，如

$$0=\frac{\partial\phi}{\partial c}=2Nc\quad\therefore\ c=0,$$

$$0=\frac{\partial\phi}{\partial a}=2N(a\sigma_x^2+b\sigma_x\sigma_yr_{xy}-\sigma_x\sigma_zr_{xz})=0,$$

$$0=\frac{\partial\phi}{\partial b}=2N(a\sigma_x\sigma_yr_{xy}+b\sigma_y^2-\sigma_y\sigma_zr_{yz})=0。$$

故 $$a\sigma_x+\sigma_yr_{xy}=\sigma_zr_{xz}$$

$$a\sigma_xr_{xy}+b\sigma_y=\sigma_zr_{yz}。$$

$$\therefore\quad \frac{a\sigma_x}{r_{xz}-r_{xy}r_{yz}}=\frac{b\sigma_y}{r_{yz}-r_{xz}r_{xy}}=\frac{\sigma_z}{1-r_{xy}^2}\cdots\cdots\ (114)$$

方程式 $z=ax+by+c$ 變爲

$$\frac{z}{\sigma_z}=R_x\cdot\frac{x}{\sigma_x}+R_y\cdot\frac{y}{\sigma_y}\cdots\cdots(115)$$

式中，$R_x=\frac{r_{xz}-r_{xy}r_{yz}}{1-r_{xy}^2}$，$R_y=\frac{r_{yz}-r_{xy}r_{xz}}{1-r_{xy}^2}$。

R_x, R_y 爲 z, x 及 z, y 間之淨迴歸係數(partial regression coefficient)；對某一 y 值，$z=R_x\frac{\sigma_z}{\sigma_x}x+$ 常數，對某一 x 值，$z=R_y\frac{\sigma_x}{\sigma_y}\cdot y+$ 常數，與 $y=r\frac{\sigma_y}{\sigma_x}\cdot x$ 相對照之公式，已詳第六章第五節。

類似此式之方程，當 x 以 y, z 表示，或當 y 以 z, z 表示時，自亦可照樣解出。

x 與 z（y 爲不變）淨相關係數(partial correlation coefficient)之定義，以僅有兩變量之例爲比喻，可解爲當 z 用 x 與 y 表示（如公式115），及 x 用 z 與 y 表示時，分別求得淨相關係數之幾何平均數。故淨相關係數爲

$$\frac{r_{xz}-r_{xy}r_{yz}}{\sqrt{1-r_{xy}^2}\ \sqrt{1-r_{yz}^2}}。$$

以上之研究，均根據猶爾(Yule)先生之論文（統計學報1897年號，自八百三十一頁起）及其所著之書，蓋此科目大部均爲猶爾先生之貢獻也。惟此地所論者，與猶爾氏略有不同，本書重要之考慮，一則係由於差誤律之通用（見第三章第五節），一則係假定 z 之標準差與 x 及 y 之各值無關，但此不合通例之事也。至猶爾先生，則不用此假定，而用最小二乘法(least square)，此法因其原理根基之難點，本書，除極少爲例外外，不取之。

z 對於 x 及 y 之方程式，與猶爾先生所用者相同，而與由常態複相關原理所得者，亦無少異。

例一：生活費調查委員會，於一九一八年，用家計調查法，搜

集多數工人家庭之每週食物費用資料（參閱第三章第七節第七例）。就中有三百九十家，係有技能工人。今將此 390 家資料，依各家人數，分爲十四歲以上以下二類（參閱統計學報1919年號，第三百六十頁）。

茲將其標號及數量列下：

	食物費用	十四歲以上人數	十四歲以下人數
平均數 對平均差離差 標準差	51先令 $z\times5$先令 $\sigma z=3\cdot03\times5$先令	$2\cdot48$ x $\sigma x=\cdot836$	$3\cdot56$ y $\sigma y=1\cdot40$

$r_{xy}=-\cdot0525$，$r_{zx}=\cdot504$，$r_{xy}=\cdot315$。

由此得方程式$z/\sigma_z=\cdot52x/\sigma_x+\cdot35y/\sigma_y$，更由此式得公式：—食物費用（先令）$=14\cdot5\times9\cdot4\times$十四歲以上人數$+3\cdot7\times$十四歲以下人數，並組成下表：

家庭食物費用

十四歲以上之人數	照公式算得				實在數之平均數			
	兒童數額				兒童數額			
	2	3	4	5	2	3	4	5
2 ………	40·7	44·4	48·1	51·8	40·5(74)	45·2(74)	47·1(53)	52·9(25)
3 ………	50·1	53·8	57·5	61·2	54·8(21)	51·2(17)	58·2(16)	64·9(17)
4 ………	59·5	63·2	66·9	70·6	58·0(10)	60·2(10)	78·1(6)	—(0)

括弧中數目係平均實在數

就標準差大而事件次數小者而論，經驗所得與公式所得，二者適應情形甚佳。

由此實例觀之，淨相關法與平常所用之交叉表法，至爲接近；但一經用此公式，即將此兩得數，使生始終一貫之關係。現得一結論：（一）平均計算，每添一成年人（家庭收入普通均恃成年人）家庭食物費用，應增九先令五便士，而添一兒童應加三先令八便士；（二）兒童人數愈多生活程度愈低，因一兒童之養育費約當成年人三分之二也（此處所謂成年人，係指在十四歲以上者而言）。

例二：下列資料係自倫敦郡之一九一一年人口普查中摘錄而來：

$z+3\cdot7$ 爲一住宅之房間數

$x+4\cdot15$ 爲一家庭之人數

$y+\cdot86$ 爲一家庭十歲以下之兒童人數

$3\cdot7, 4\cdot15, \cdot86$，爲倫敦郡之各該平均數。

$r_{xy}=\cdot57$, $r_{xz}=\cdot44$, $r_{yz}=-\cdot03$, $\sigma_z=2\cdot59$, $\sigma_x=2\cdot32$, $\sigma_y=1\cdot24$, $R_x=\cdot676$, $R_y=\cdot402$。

關於 1,023,951 家之數字，求至三位有效數字，堪稱確實。

$$z=x\times\frac{\sigma_z}{\sigma_x}\times\cdot676-y\times\frac{\sigma_z}{\sigma_y}\times\cdot402=x\times\cdot754-y\times\cdot840$$

或（房間數—$3\cdot7$）$=\cdot754$(人數—$4\cdot15$)$-\cdot84$(兒童數—$\cdot86$)。

兒童人數增加時，某一定等級之家庭，房間數必急劇變少。

以下式記之亦可：

房間數$=1\cdot29+\cdot68$人數$-\cdot84$兒童

$=1\cdot29+\cdot68$成人（十歲以上）$-\cdot16$兒童（十歲以下）

例三：據『生計與貧乏』(Livelihood and Poverty) 一書之記載，五百六十六家之社會研究，曾將收入，房租及家庭組織逐一調查。收入增加與有酬報人數多者，房租必較多，但若收入數與有酬報之人數不變，則兒童人數愈多，則可供爲房租之用者愈少。

房租：$z+6\cdot075$ 先令，式中 $6\cdot075$ 先令爲平均數。

相等成年人數：$x+3\cdot287$，式中 $3\cdot287$ 爲平均數。

收入：$y+31\cdot712$ 先令，式中 $31\cdot712$ 先令爲平均數。

相等成年人數，係依成人與兒童需用之房屋爲根據，將成人，兒童照武斷之標準，加以分類；五歲下兒童，作爲成人之四分一，五歲至十四歲，作爲二分之一，男童十四至十八歲，女童十四至十六歲，均爲四分之三，自此以上則作爲一。

房租對房間數之相關甚近，故房租可用以推算房間數。

$\sigma_z=1\cdot33$，$\sigma_x=1\cdot22$，$\sigma_y=13\cdot0$，$r_{xy}=\cdot543$，

$r_{yz}=\cdot458$，$R_x=-\cdot136$ $R_y=\cdot532$。

故 $$\frac{z}{\sigma_z}=-\cdot136\frac{x}{\sigma_x}+\cdot532\frac{y}{\sigma_y}$$

或 $$z=-\cdot148x+\cdot0544y。$$

於是，如收入不變更，家庭變大則房間數減少。

以上三例，證明有兒童之家庭，比無兒童者，漸有減少每人食物費用及房間數之傾向，且於相當程度內測量其損失。

吾人仍須研究三變量之三原理論上分配，以與具有兩變量之常態相關面相呼應。下文特就簡單情形將分析及結論示明之，吾人將見所有證明方法，與具有兩變量之例，並無不同也。

節二第　複相關

下述分析，須假設各元素具有常態頻數，方能成立。

以X，Y，Z 為因變量，U，V_1，V_2，V_3，為自變量，二種變量之關係為 $X=U+V_1$，$Y=U+V_2$，$Z=U+V_3$。

設 U，V_1，V_2，V_3，係從常態羣類中隨機抽出，該羣類之平均數為 $\bar{u},\bar{v}$……，標準差為 σ_u，σ_v……。

以$\bar{X}$，$\bar{Y}$，$\bar{Z}$，為 X，Y，Z 之平均數；σ_x，σ_y，σ_z 為其標準差，並以$X=\bar{X}+x$……，$U=\bar{U}+u$……

則至終局，$\bar{X}=\bar{U}+\bar{V}$等等諸如此類，且

$$\therefore\quad x=u+v,\ y=u+v_2,\ z=u+v_3。$$

設 u, v_1, u_2, v_3,各個彼此完全獨立,則

平均值 $uv_1=0=$平均值 $v_1\ v_2$ 其他以此類推。

茲以 r_{xy} 爲 x 對 y 之相關係數。

則 $\sigma_x{}^2=\sigma_u{}^2+\sigma_v{}^2\cdots\cdots$,

又 $\sigma_x\sigma_y r_{xy}=$平均值$(u+v_1)(u+v_2)=\sigma_u{}^2$

$$=\sigma_y\sigma_z r_{yz}=\sigma_z\sigma_x r_{zx}。$$

x_1, y_1, z_1, 抽出之值對特定值 u, v_1, v_2, v_3, 而生之連帶機率爲

$$\frac{1}{\sigma_u\cdot\sqrt{2\pi}}e^{-\frac{1}{2}\frac{u^2}{\sigma u^2}}\times\frac{1}{\sigma_1\sqrt{2\pi}}e^{-\frac{1}{2}\frac{v_1{}^2}{\sigma v_1{}^2}}\times\quad\times\quad=\mathrm{P}_u,$$

受下列條件之支配:$x_1=u+v_1$,$y_1=u+v_2$,$z_1=u+v_3$, 消去 v_1, v_2, v_3。

x_1, y_1, z_1,（因某一 u 值而生）之連機率,可以下式得之: $-2\log(\mathrm{P}_u\cdot4\pi^2\sigma_u\sigma_{v1}\sigma_{v2}\sigma_{v3})$

$$=\frac{u^2}{\sigma_u{}^2}+\frac{(u+x_1)^2}{\sigma v_1{}^2}+\frac{(u+y_1)^2}{\sigma_{v2}{}^2}+\frac{(u+z_1)^2}{\sigma_{v3}{}^2}=a(u-b)^2+c,$$

式中 $a=\frac{1}{\sigma_u{}^2}+\frac{1}{\sigma_{v1}{}^2}+\frac{1}{\sigma_{v2}{}^2}+\frac{1}{\sigma_{v3}{}^2}$

$$ab=\frac{x_1}{\sigma_{v1}{}^2}+\frac{y_1}{\sigma_{v2}{}^2}+\frac{z^1}{\sigma_{v3}{}^2}$$

$$c=\frac{x_1{}^2}{\sigma_{v1}{}^2}+\frac{y_1{}^2}{\sigma_{v2}{}^2}+\frac{z_1{}^2}{\sigma_{v3}{}^2}-ab^2。$$

對某一 u 值而抽來之 x_1, y_1, z_1,其全部機率,爲

$$P=\int_{-\infty}^{\infty}Pudu,$$

如再以x_1，y_1，z_1 為常數 則

$$=\frac{e^{-\frac{c}{2}}}{(2\pi)^{\frac{3}{2}}\sigma u\sigma v_1\sigma v_2\sigma v_3}\int_{-\infty}^{\infty}\frac{e^{-\frac{a}{2}(u-b)^2}}{\sqrt{2\pi}}du=\frac{1}{(2\pi)^{\frac{3}{2}}\sigma u\sigma v_1\sigma v_2\sigma v^3}\cdot$$

$$\frac{1}{\sqrt{a}}e^{-\frac{c}{2}}。$$

以 x, y, z 代 x_1, y_1, z_1。

$$ac=\left(\frac{1}{\sigma u^2}+\frac{1}{\sigma v_1{}^2}+\cdots\right)\left(\frac{x^2}{\sigma^2}+\cdots\right)-\left(\frac{x}{\sigma^2}+\cdots\right)^2。$$

但 $$\frac{1}{\sigma u^2}+\frac{1}{\sigma v_2{}^2}+\frac{1}{\sigma v_3{}^2}=\frac{1}{\sigma u^2}+\frac{1}{\sigma y^2-\sigma u^2}+\frac{1}{\sigma z^2-\sigma u^2}=\frac{\sigma y^2\sigma z^2-\sigma u^4}{\sigma u^2\sigma v_2{}^2\sigma v_3{}^2}$$

且 $$a=\frac{1}{\sigma u^2}+\frac{1}{\sigma x^2-\sigma u^2}+\frac{1}{\sigma y^2-\sigma u^2}+\frac{1}{\sigma z^2-\sigma u^2}$$

$$=\frac{\sigma x^2\sigma y^2\sigma z^2-\sigma u^4(\sigma x^2+\sigma y^2+\sigma z^2)+2\sigma u^6}{\sigma u^2\sigma v_1{}^2\sigma v_2{}^2\sigma v\ {}^2}$$

$$\therefore\ c\{\sigma x^2\sigma y^2\sigma z^2-\sigma u^4(\sigma x^2+\sigma y^2+\sigma z^2)+2\sigma u^6\}$$

$$=x^2(\sigma y^2\sigma z^2-\sigma u^4)++\ -2xy\sigma u^2\sigma v_3{}^2-\ -。$$

以 $R\sigma x^2\sigma y^2\sigma z^2$ 代 $\sigma x^2\sigma y^2\sigma z^2-\sigma u^4(\sigma x^2+\quad+)+2\sigma u^6$,

故 $$R=1+2r_{xy}r_{yz}r_{zx}-r_{xy}{}^2-r_{yz}{}^2-r_{zx}{}^2。$$

於是 x, y, z 同時而起之機率為

$$P=\frac{1}{(2\pi)^{\frac{3}{2}}\sigma x\sigma y\sigma zR^{\frac{1}{2}}}\cdot e^{-\frac{1}{2R}\left\{\frac{x^2}{\sigma x^2}(1-ryz^2)++-\frac{2xy}{\sigma x\sigma y}(rxy-rxzryz)--\right\}}$$

$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$(116)

因 $\dfrac{r_{xy}-r_{xz}r_{yz}}{\sigma_x\sigma_y}=\dfrac{\sigma_x\sigma_y r_{xy}\sigma_z^2-\sigma_x\sigma_z\cdot r_{xy}\sigma_y\sigma_z\cdot r_{yz}}{\sigma_x^2\sigma_y^2\sigma_z^2}$

$$=\frac{\sigma_u^2\sigma_z^2-\sigma_u^2\cdot\sigma_u^2}{\sigma_x^2\sigma_y^2\sigma_z^2}=\frac{\sigma_u^2\sigma_{v3}^2}{\sigma_x^2\sigma_y^2\sigma_z^2}\text{。}$$

在特別情形之下

$$\sigma_u=\sigma_{v1}=\sigma_{v2}=\sigma_{v3},\quad \sigma_x^2=2\sigma_u^2=\sigma^2,$$

$$r_{xy}=\tfrac{1}{2}=r_{yz}=r_{zx},\quad \mathrm{R}=\tfrac{1}{2}\text{。}$$

則機率爲

$$\frac{1}{2\sigma^3\pi^{\frac{3}{2}}}e^{-\frac{1}{4\sigma^2}\left\{3(x^2+y^2+z^2)-2(xy+yz+zy)\right\}}\text{。}$$

對 x, y 之某值，z 之最或然值，用 $\dfrac{\partial \mathrm{P}}{\partial z}=0$ 計算，乃爲（如公式 115）

$$\frac{z}{\sigma_z}(1-r_{xy}^2)=\frac{x}{\sigma_x}(r_{xz}-r_{xy}r_{yz})+\frac{y}{\sigma_y}(r_{yz}-r_{xy}r_{xz})\text{。}$$

在特別情形下，此又變爲 $z=\frac{1}{3}(x+y)$。

據艾得頓 (Elderton) 氏——追隨皮爾生氏——之證明，如 x, y 與 z, 爲任何有限變量數（如上文之 u, v ……）之和，全係常態頻數，並且 (x,y), (y,z) 或 (z,x) 各對，均有共同性，餘則僅有一數時，則 P 之形式，乃

爲　$\mathrm{K}e^{-(ax^2+by^2+cz^2+2fyz+2gzx+2hxy)}$

式中 a, b, c, f, g, h 均爲待決之常數。

以機率之總數爲一。

設A, B, C, F, G, H 爲行列式 $\Delta=\begin{vmatrix} a & h & g\\ h & b & f\\ g & f & c\end{vmatrix}$,

之子行列式，則 $\mathrm{A}=bc-f^2$, $\mathrm{F}=hg-af$……,

$BC-F^2=a\Delta\cdots\cdots$

$$於是 -\log\frac{P}{K}=a\left(x+\frac{h}{a}y+\frac{g}{a}z\right)^2+\frac{c}{a}\left(y-\frac{F}{C}z\right)^2+\frac{\Delta}{C}z^2$$

$$1=\iiint P\cdot dxdydz=(\sqrt{\pi})^2\cdot K\frac{1}{\sqrt{a}}\cdot\sqrt{\frac{a}{C}}\sqrt{\frac{C}{\Delta}}, 而 K\pi^{\frac{3}{2}}=\Delta^{\frac{1}{2}}$$

$$\sigma z^2=\iiint Pz^2dxdydz=K\pi\cdot\frac{1}{\sqrt{a}}\cdot\sqrt{\frac{a}{C}}\int z^2e^{-\frac{\Delta}{C}z^2}dz=K\pi^{\frac{3}{2}}\frac{1}{\sqrt{C}}\cdot\frac{1}{2}\left(\frac{C}{\Delta}\right)^{\frac{3}{2}}$$

$$=\frac{C}{2\Delta}$$

同樣 $\sigma y^2=\frac{B}{2\Delta}$, $\sigma x^2=\frac{A}{2\Delta}$

$$\sigma y\sigma zryz=\iiint Pzydxdydz=K\sqrt{\pi a}^{-\frac{1}{2}}\int\int zye^{-\frac{c}{a}\left(y-\frac{F}{C}z\right)^2}e^{-\frac{\Delta}{C}z^2}dydz$$

$$=K\sqrt{\pi a}^{-\frac{1}{2}}\int\int z\left(y'+\frac{F}{C}z\right)e^{-\frac{C}{a}y'^2}e^{-\frac{\Delta}{C}z^2}dy'dz,$$

式中,$y'=y-\frac{F}{C}z$,又積分之極限爲$\pm\infty$

$$=K\pi\cdot a^{-\frac{1}{2}}\sqrt{\frac{a}{C}}\cdot\frac{F}{C}\int z^2e^{-\frac{\Delta}{C}z^2}dz=\frac{1}{2}\frac{F}{\Delta}。$$

同理 $\sigma x\sigma yrxy=\frac{1}{2}\frac{H}{\Delta}$, 又 $\sigma x\sigma zrxz=\frac{1}{2}\frac{G}{\Delta}$。

$$\therefore\quad a\Delta=BC-F^2=4\sigma y^2\sigma z^2(1-ryz^2)\Delta^2$$

$$f\Delta=GH-AF=4\sigma x^2\sigma y\sigma z(rxyrxz-ryz)\Delta^2$$

$$\Delta^2=ABC+2FGH-AF^2-BG^2-CH^2$$

$$=8\Delta^3\sigma x^2\sigma y^2\sigma z^2(1+2rxyryzrzx-ryz^2-rzx^2-rxy^2)。$$

以 R 代括弧內之數量。$R=\begin{vmatrix}1 & rxy & rxz\\ rxy & 1 & ryz\\ rxz & ryz & 1\end{vmatrix}$

於是 $\Delta=\dfrac{1}{8R\sigma x^2\sigma y^2\sigma z^2}$, $a=\dfrac{1-ryz^2}{2R\sigma x^2}$, $f=\dfrac{rxyrxz-ryz}{2\sigma y\sigma zR}$。

故　$$P=\frac{1}{\sqrt{R}(2\pi)^{\frac{3}{2}}\sigma x\sigma y\sigma z}e^{-\frac{1}{2R}\left\{\frac{x_2}{\sigma x^2}(1-ryz^2)++\frac{2yz}{\sigma y\sigma z}(ryz-rxyrxz)--\right\}}$$

與上述之特別情形相同。

如不用 x, y, z,而用 n 個數量 x^1, x^2, ……, 同一方法(應歸功於皮爾生教授)之更概括證明,可得

$$P=\frac{1}{\sqrt{R}(2\pi)^{\frac{n}{2}}\cdot\sigma_1\sigma_2\cdots\sigma n}e^{-\frac{1}{2R}\left\{\frac{x_1^2}{\sigma_1^2}R_{11}+++2\frac{x_1x_2}{\sigma_1\sigma_2}R_{12}++\right\}}$$

式中 $R=\begin{vmatrix} 1 & r_{12} & r_{13}\cdots r_{1n} \\ r_{21} & 1 & r_{23}\cdots r_{2n} \\ \cdots & \cdots & \cdots \\ rn_1 & rn_2 & rn_3\cdots rnn \end{vmatrix}$

r_{st} 與 r_{ts} 數量相同,而 Rst 為自第 s 列第 t 排斜交之子行列式,原來正負符號仍舊。

證明之步驟如下:

如前述,$P=Ke-\phi$, 式中 $\phi=a_{11}x_1^2+\cdots+2a_{12}x_1x_2+\cdots$, $a_{11}\cdots$, $a_{12}\cdots$ 為常數。

設$\Delta n=\begin{vmatrix} a_{11} & a_{12}\cdots & a_{1n} \\ a_{21} & a_{22}\cdots & a_{2n} \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ an_1 & an_2\cdots & ann \end{vmatrix}$

式中,$a_{st}=a_{ts}$,並以 A_{st} 為第 s 列第 t 排交叉之子行列式

$$A_{nn}=\Delta_{n-1}。$$

然後可將 ϕ 引申,使 x_1 僅出在第一項內,x_2 僅現於第二項內,其他以此類推,則

$$\phi=a_{11}\left(x_1+\frac{a_{12}}{a_{11}}x_2+\cdots\right)^2+\frac{a_{22}a_{11}-a_1^2}{a_{11}}\left(x_2+\cdots\right)^2+\cdots+\frac{\Delta n-1}{\Delta n-2}\left(xn-1-\frac{An\cdot n-1}{\Delta n-1}\right)^2$$

$$+\frac{\Delta n}{\Delta n-1}xn^2$$

此可用歸法納證明，但手續極爲繁雜。

$$1=\iint\cdots P\cdot dx_1dx_2\cdots dxn=K\pi\frac{n}{2}\cdot\left(\frac{1}{a^{11}}\right)^2\left(\frac{\Delta_1}{\Delta_2}\right)^{\frac{1}{2}}\cdots\left(\frac{\Delta_n-1}{\Delta_n}\right)^{\frac{1}{2}}=\frac{K\pi^{\frac{n}{2}}}{(\Delta n)^{\frac{1}{2}}}$$

$$\sigma_n{}^2=\iint\cdots Pxn^2dx_1dx_2\cdots dxn=1\cdot\frac{1}{2}\frac{\Delta n,-1}{\Delta n}=\frac{Ann}{2\Delta n}$$

同樣，將次數改變，$\sigma_t{}^2=\frac{Att}{2\Delta n}$

$$\sigma n\cdot\sigma n-1\cdot rn,\ n-1=\iint\cdots Pxnzn-1dx_1\cdots dxn$$

$$=\left(\frac{\Delta n}{\Delta n-2}\right)^{\frac{1}{2}}\iint xnxn-1e^{-\frac{\Delta n-1}{\Delta n-2}\left(xn+1-\frac{An,n-1}{\Delta n-1}\right)^2-\frac{\Delta n}{\Delta n-1}xn^2}dxn-$$

$$1dxn=\frac{An,n-1}{2\Delta n}。$$

同樣，將次數改變，

$$\sigma_s\sigma_t r_{st}=\frac{Ast}{2\Delta n}$$

在求R之行列式中，以此等數值代r_{st}……，則得

$$R=\frac{1}{(2\Delta_n)n\sigma_1{}^2\cdots\sigma_n{}^2}\begin{vmatrix}A_{11} & A_{12}\cdots A_1n\\ A_{21} & A_{22}\cdots A_2n\\ \cdots\cdots & \cdots\cdots\\ \cdots\cdots & \cdots\cdots\\ An_1 & An_2\cdots Ann\end{vmatrix}=\frac{\Delta n^n-1}{(2\Delta n)n\sigma_1{}^2\cdots\sigma_n{}^2},$$

此係根據一著名之行列式定理。

$$\therefore\quad \frac{1}{K}=(2\pi)^{\frac{n}{2}}\sigma_1\cdots\sigma n\cdot\sqrt{R}。$$

$$R_{11}=\frac{1}{(2\Delta n^{(n-1}\sigma_2{}^2\cdots\sigma_n{}^2}\begin{vmatrix}A_{22}\cdots\cdots A_2n\\ \cdots\cdots\\ \cdots\cdots\\ An\cdots\cdots Ann\end{vmatrix}=\frac{a_{11}\Delta_n{}^{n-2}}{(2\Delta_n)^{n-1}\sigma_2{}^2\cdots\sigma_n{}^2},$$

此又係一著名之定理。

$$\therefore \qquad a_{11}=\frac{R_{11}}{2R\sigma_1^2}。$$

同樣，$a_{12}=\frac{R_{12}}{2R\sigma_1\sigma_2}$，又將次數改變，$ast=\frac{R_{st}}{2R\sigma_s\sigma_t}$，

對於 x_2，$x_3\cdots\cdots x_n$ 之已知值，x之最或然值，乃可於下列之式求出：

$$\frac{x_1}{\sigma_1}R_{11}=-\frac{x_2}{\sigma_2}R_{12}-\frac{x_3}{\sigma_3}R_{13}\cdots\cdots-\frac{x_n}{\sigma_n}\cdot R_{1n}。$$

（註一）　a，b，c各值，如以 ϕ 當作各平方之和不用微分，亦可求得。

五

第九章　平均數動差及相關等測量之精度（註一）

第一節　逆機率

以前數章，討論抽樣方法所生差誤問題，均以宇宙——樣本所自取之大範圍——爲觀點，並非樣本之本身；換言之，以前之問題，乃係樣本究能如何代表其範圍，代表至何程度之問題。但實際問題，乃適得其反：吾人由將何以樣本以推斷範圍？與言及此，乃不能不提出繁難費解之逆機率（inverse probability）學說矣。逆機率爲何？其形式如是：原已抽出樣本之各範圍，何者究係產生該樣本者耶？

爲求議論清楚起見，似以少論逆機率原理爲宜；故舉數例於下，以解釋問題並求解答焉。

袋中有金鎊一枚，先令二枚，今遺失其一。迨將所餘二幣，取出一枚，視之乃爲先令。請問金鎊遺失之機率，爲何？

金鎊遺失之原來機率，爲 $p'_1=\frac{1}{3}$，先令遺失之機率，原爲 $p'_2=\frac{2}{3}$，假設一種錢幣遺失之機會，與另一錢幣之遺失機會，完全均等。

如金鎊果已遺失，則取出先令之機率，爲$p_1=1$，因袋中除先令外已無他物也。

如所遺失者爲金鎊，所取出者爲先令，其原來機率，爲 $p'_1p_1=\frac{1}{3}$。

如所遺失者爲先令，所取者亦爲先令，則其原來機率，爲 $p'_2p_2=\frac{2}{3}\times\frac{1}{2}=\frac{1}{3}$。

假定所給，只言此等機會均等之繁複事件之一，業已出現，但所遺失者爲何，則不得而知。

故袋中所餘之第三個錢幣，其爲金鎊，抑爲先令，機會正復相等。

玆將此假設，照下列方法，概而論之：設有各種可能事件，其實現之機率，分別爲 p'_1, p'_2……p'_t……，於此等事件中，只知其中之一，業已出現。吾人可另外得一結果，如事件之第一，第二……第 t……已經實現，則其各該機率，當分別爲 p_1, p_2, p_t……。

第 t 個事件實現並發生此結果之原來機率，爲 $p'_t\times p_t$。

第一序列之事件，實現並產生此結果之原來機率，當爲下列比率：

$$p'_1p_1:p_2'p_2:\cdots\cdots:p'_tp_t\cdots\cdots=P_1:P_2:\cdots\cdots$$

但吾人現知此等之任一個，確已出現，且雖多得此項消息，亦並不影響相對量數P_1,P_2……，不過使其總數提升成爲如此比

率,K,因而使其等於一,一卽代表代數機率尺度之一定(certainty)。故 K.S $P_t=1$,而實現者果爲第一序列第 t 個事件之機率,爲

$$KP_t=\frac{p_t'p_t}{p_1'p_1\cdots\cdots+\cdots\cdots+p_t'p_t+\cdots\cdots}$$

一囊中,有類似黑白二種球共六個。玆取出一個白球。吾人將何以推斷囊中之白色原有若干耶?

欲答覆此問題,須視關於黑白球分配之原有機率之假定爲何。如原來,一球之爲黑爲白,須有均等之機會,而 p'_t 爲 t 個球白色之機率,則 $p_t'=\frac{1}{2^6}\cdot {}_6C_t$。

不論假定云何,$p_t=\frac{t}{6}$。

$$P_0:P_1:\cdots\cdots: P_6=\frac{1}{6\times 2^6}(0:6:30:60:60:30:6)。$$

$$SP_t=\tfrac{1}{2}\text{又}K=2。$$

囊中白球有 0, 1, 2, 3, 4, 5, 6 個之機率,當分別爲 0, $\frac{1}{32}$, $\frac{5}{32}$, $\frac{10}{32}$, $\frac{10}{32}$, $\frac{5}{32}$, $\frac{1}{32}$。

但如囊中白球數額,係用擲骰而決定,骰子上面之點卽用以定白球之數,則

$$p'_1=p'_2=\cdots=p_6'=\frac{1}{6};\ P_t=\frac{1}{6},\frac{t}{6};$$

$$SP_t=\frac{21}{36},\ K=\frac{12}{7},\text{又}\ KP_t=\frac{t}{21}。$$

就一般而論，如囊中共有球 n 個。就中白球若干，數額轉螺旋盤決定之，在圓盤之圓周，以相等之空間，記 0，1……n 之數字，螺旋盤特立軸旋轉，俟其停止，察其圓周對一固定針最近之數，即以此數作爲白球之數，則

$$p'_t=\frac{1}{n+1},p_t=\frac{t}{n},\mathrm{SP}t=\frac{1}{2},\ \mathrm{KP}_t=\frac{2t}{n(n+1)}。$$

原始白球數額，在自 t 個以下之總和機率，爲

$$\mathrm{K}(\mathrm{P}_0+\mathrm{P}_1+\mathrm{P}_2+\cdots+\mathrm{P}_t)=\frac{t(t+1)}{n(n+1)}=f(t)。$$

如 $f(t)=\frac{1}{2}$，則囊中有白色 t 個，或無白球 t 個，各有一半之機會；又如 n 甚大，則 $t=\frac{n}{\sqrt{2}}$一式約略可以適合上方程式。

故當 n 甚大時，則白球對全體球佔$\frac{1}{\sqrt{2}}=.7\cdots$之比例，機會各佔一半。比例在$\frac{1}{2}$與$\frac{\sqrt{3}}{2}$之間，機率約爲$\frac{1}{2}$。

此一實例，至爲重要，一則示吾人所有結果，乃全繫對於未知事件（即吾人所抽取者）相對原來機率之假定如何，二則指示吾人，如將各項機率總集起來，較之各個計算，可得較易理解之結果。

第二節　某類在範圍中所佔比例，P，之精度

玆請應用逆機率原理，以樣本決定其全範圍。設有N個事物

之範圍，具有某種特質者爲p，範圍中具有該種特質者，有pN個，玆由範圍中抽取n個事物，經查具有該項特質者爲$p'n$個。

由已知之p中，發現$p'n$個之機率，爲${}_nC_{p'n}p^{p'n}q^{q'n}$（見第二章第四節）式中$q=1-p$，$q'=1-p'$。

如p自0至1所有各值，原來卽係機會均等，則自p各值中，從p'至x，發現$p'n$之機率，爲由各特定值而生各機率之總和，卽$={}_nC_{p'n}\int_{p'}^{x}x^{p'n}(1-x)^{q'n}dx$，於是依前節第一例之定理，$p$原始各值，自$p'$至$x$，之機率，爲

$$\mathrm{P}_x=\frac{{}_nC_{p'n}\cdot\int_{p'}^{x}x^{p'n}(1-x)^{q'n}dx}{{}_nC_{p'n}\int_{0}^{1}x^{p'n}(1-x)^{q'n}dx},$$

而此式如將$\frac{1}{\sqrt{n}}$略去，可照下述，化爲差誤常態曲線之形式。

以$x=p'+z\sigma$，其$\sigma^2n=p'q'$，而$1-x=q'-z\sigma$。至σ則與$\frac{1}{\sqrt{n}}$同次(order)。

如是

$$\mathrm{P}_x=\frac{\int_0^z\left(1+\frac{z\sigma}{p'}\right)^{p'n}\left(1-\frac{z\sigma}{q'}\right)^{q'n}dz}{\int_{-\infty}^{\infty}\left(1+\frac{z\sigma}{p'}\right)^{p'n}\left(1-\frac{z\sigma}{q'}\right)^{q'n}dz}=\frac{\int_0^z f(z)dz}{\int_{-\infty}^{\infty}f(z)dz},$$

因爲，如$x=1$，則$z=\sqrt{\frac{nq'}{p'}}$，又如$x=0$，則$z=-\sqrt{\frac{np'}{q'}}$，至此則於n甚大時而漸至於$\pm\infty$。

於是

$$\log f(z)=p'n\log\left(1+\frac{z\sigma}{p'}\right)+q'n\log\left(1-\frac{z\sigma}{q'}\right)=z\times 0-\frac{z^2\sigma^2 n}{2}\left(\frac{1}{p'}+\frac{1}{q'}\right)$$

+含有$\sigma^3 n$之各項，

卽$+\frac{1}{\sqrt{n}}$次各項，

$=-\frac{1}{2}z^2$，當$\frac{1}{\sqrt{n}}$略去時，因 $p'+q'=1$ 也。

$$\therefore \mathrm{P}_x=\frac{\int_0^z e^{-\frac{1}{2}z^2}dz}{\int_{-\infty}^{\infty}e^{-\frac{1}{2}z^2}dz}=\int_0^z\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}dz \quad\cdots\cdots\cdots\cdots(117)$$

故一範圍中，成 p' 與 $p'+p_1$ 之比例，則對宇宙之觀察之機率，為（以$\frac{p_1}{\sigma}$代z）$\int_0^{p_1}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{u^2}{2\sigma^2}}$式中$\sigma=\sqrt{\frac{p'(1-p')}{n}}$。

以上分析均係根據投韓德（Todhunter）氏之機率學史(History of the Theory of probability)，第五五四頁起。

從一已知之範圍中，抽出之樣本內，自 p 至 $p+p_1$ 中間抽得一值之機率為$\int_0^{p_1}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{u^2}{2\sigma^2}}du$，（式中，$\sigma=\sqrt{\frac{p(1-p)}{n}}$），此一定理之反面(converse)，卽逆機率也。

上文分析之難點，乃在一假定——自 0 至 1 所有 p 之各值，原始卽有均等之機會。茲將此假定說明於下：

設$n=100, p'=.1$。

如觀察係從一以 $p=.07$ 之範圍得來，則$\sigma^2=\frac{.07\times.93}{100}$，$\sigma=.0255$，$\frac{p'-p}{\sigma}=1.18=z$。自$p=.06$ 至 $p=.08$之總和機，約略等於 $p=.07$ 之機率，此可以縱坐標$\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}$，以.02爲橫坐標（以標準差之倍數爲單位）求得之，換言之，$.02\div.0255=.78$，該總和機率等於$\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(1.18)^2}$之 $.78=.157$。如此繼續逐一計算，乃得出下表：

p之值	求得$p'=.1$之約略機率
.00—.02	.000
.02—.04	.002
.04—.06	.029
.06—.08	.157
.08—.10	.262
.10—.12	.242
.12—.14	.159
.14—.16	.084
.16—.18	.039
.18—.20	.014
.20—.22	.005
.22—.24	.001
.24—.26	.000
	.994

由觀察看來，p可以等於.1，標準差爲$\sqrt{\frac{.1\times.9}{100}}=.03$，且正數偏斜度頗大。

p 之值去 .1 多於兩個標準差，不論 p 值原來是否有均等之機會，其機率終爲甚小。

所舉此例已釋明該項定理：除中心值附近爲例外外，p 值原始機率之分配，如何假定，並無關係。在雖小而甚重要之中心區域，無論實在定律如何，一假定——在某一區域之p之原始機率，與該區別成比例——頗可作一適當之第一近似值（參閱：愛基華斯氏在統計學報一九〇八年號第三八七頁一文，該文並列有參考書）

吾人旣假定，所求數量重要數值，於影響於此分析之數值在全距之任一點上，原始卽有均等之機會，則由此以上項或類似之調查，吾人可免陷於重大之差誤。

現可將用樣本求 p 之結果，總述於下。範圍中之最或然值，卽爲觀察而來之 p'。對觀察值之離中差——大小相當於 p_1——之機率，可由具有標準差

$$\sqrt{\frac{p_1(1-p_1)}{n}},\ \text{或}\sqrt{\frac{p(1-p_1)}{n}\left(1-\frac{n}{\mathrm{N}}\right)}$$

（式中N爲範圍中之事物總數，且 $\frac{n}{\mathrm{N}}$ 不可略去）之常態差誤函數，求出其近似值。

一量數之精度，可用差誤——量數所易陷入之差誤——之標準差之倒數，測算之。

第三節 通用方法

玆更以概括之語論之，設未知範圍之函數，爲X；n個樣本，由範圍中隨機抽選而來，設 X′ 爲樣本之任何已知函數；並設 X $=\mathrm{X}'+x$。

吾人如能證明，當範圍中之值爲X時，求得X′值之機率爲 $\mathrm{P}_x=\mathrm{P}_0e^{-\frac{1}{2}\frac{x^2}{\sigma^2}}$（式中，$\mathrm{P}_0$ 爲最大機率，係當 $\mathrm{X}=\mathrm{X}'$ 時所得，又 σ 爲常數，$x=\mathrm{X}-\mathrm{X}'$)，則可有理由，確定：樣本所示之證據——一則問題中函數之最或然值乃爲X′，二則X′離中差之機率，可由具有標準差σ之常態函數求得之。上例中，p'爲X′，p爲X，x爲$z\sigma$，而$n\sigma^2=p(1-p')$。較爲普通之情形，倒轉之法，並不若是之直接（註二）。

欲決定以樣本爲基礎之量數之精度，其法不外三種步驟：

（1）求眞值與觀察值相差之標準差，

（2）求某一指定離中差發生之機率，

（3）應用逆機率原理。

第四節　算術平均數之精度

在第三章，已證明，如n個量數，由一頻數羣類（須與某種條件相合）中隨機而彼此獨立抽選而來，n個量數之平均數，與範圍各值之平均數，相差僅有x之大，則此種機率爲

$$2\int_x^{\infty}\frac{1}{\sigma a\sqrt{2\pi}}e^{-\frac{x^2}{2\sigma a^2}}dx,$$

式中，$\sigma_a=\dfrac{\sigma}{\sqrt{n}}$，$\sigma$ 爲範圍中之標準差。

由此可知，從樣本中觀察之標準差，σ'，與 σ 相差，可與 $\dfrac{\sigma}{\sqrt{n}}$ 等量齊觀，故如 n 甚大，σ' 可以作爲與 σ 相等也。

現可完成前此之議論，如在有 n 個事物之樣本中，而此樣本又係從一離中差不至多過兩個標準差之羣類抽選而來，樣本之平均數爲 $\bar{x}$，標準差爲 σ，則範圍中平均數與 $\bar{x}$ 相差等於 x 之機率，當 n 甚大時，爲

$$2\int_{x}^{\infty}\frac{\sqrt{n}}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\frac{x^2}{\left(\frac{\sigma}{\sqrt{n}}\right)^2}}\,dx,\quad\cdots\cdots\cdots\cdots\cdots\cdots\cdots,(118)$$

第五節　標準差之精度

次復引申此定理，以測從樣本決定之標準差及第二級動差之精度。

設 $\bar{x}$，σ，μ_2……爲宇宙之未知常數，而以 $\bar{x}+\bar{x}'\sigma'$，μ_2……爲由樣本中算得之相應值。

設以 $\bar{x}+x_t$ 爲任何觀察，並以 $x'_t=x_t-\bar{x}'$。

$x^2{}_t$ 之頻數曲線，以 μ_2 爲其平均數，而其標準差，則可由 $\sigma_x{}^2=$ 平均值 $(x_t{}^2-\mu_2)^2=\mu_4-\mu_2{}^2$ 求來。

其第四級動差爲，平均值 $(x_t{}^2-\mu_2)^2=\mu_8-4\mu_6\mu_2+6\mu_4\mu_2{}^2-$

$3\mu_2^4=M_4$。

在範圍間，$\frac{\mu_s}{\sigma^s}$依假定爲有限數，不論 s 何值，故 $\frac{M_4}{\sigma d^4}=\left(\frac{\mu_8}{\sigma^8}-4\frac{\mu_6}{\sigma_6}+6\frac{\mu_4}{\sigma^4}-3\right)\div\left(\frac{\mu_4}{\sigma^4}-1\right)^2$爲有限數。

同理，求 x_t^2 之任何級動差，$M_s\div\sigma d^s$爲有限數。

故依第四章第一節及本章所論之定理，樣本中各項數量，x_t^2，之平均數，與μ_2相差，有一具有常態頻數及標準差

$\frac{\sigma d}{\sqrt{n}}=\sqrt{\frac{\mu_4-\mu_2^2}{n}}$之『差誤』。

此平均數，m_2，$=\frac{Sx_t^2}{n}=\frac{1}{n}S(x'_t+\bar{x}')=\frac{1}{n}Sx_t'^2+\bar{x}'^2$

因 $Sx'_t=0$，故　　$=\mu'_2+\bar{x}^2$。

但$\bar{x}'^2$屬於$\frac{1}{n}$次（見公式118），而m_2適已證明屬於$\frac{1}{\sqrt{n}}$次。故$\bar{x}'^2$可以略去，而以 μ'_2 代 m_2。

由此可見，觀察得來之μ'_2，與在範圍間之μ_2，相差有一『差誤』，此差誤乃具有常態頻數，其標準差爲

$$\sqrt{\frac{\mu_4-\mu_2^2}{n}}.\qquad\qquad(119)$$

惟 $\sigma^2=\mu_2$，$\therefore\delta\sigma=\frac{\delta\mu_2}{2\sqrt{\mu_2}}$；故$\sigma'$與$\sigma$相差，有一具有常態頻數，而以

$$\sqrt{\frac{\mu_4-\mu_2^2}{4n\mu_2}} \cdots\cdots\cdots\cdots(120)$$

爲標準差之『差誤』。

如範圍爲常態，$\mu_4=3\mu_2^2$，觀察得來標準差及第二級動差之標準差，分別變爲

$$\frac{\sigma}{\sqrt{2n}} \text{ 及 } \sigma\sqrt{\frac{2}{n}} \cdots\cdots\cdots\cdots(121)$$

用同樣方法，$m_4=\frac{Sx_t^4}{n}$之標準差爲$\sqrt{\frac{\mu_8-\mu_4^2}{n}}$，

式中，$m_4=\frac{1}{n}S(x'_t-\bar{x}')^4=\mu'_4-4\bar{x}'\mu_3+6\bar{x}'^2\mu'_2-3\bar{x}'^4$。

故 μ'_4，差誤，其方次之低，有如 m_4 之差誤，換言之，卽屬於$\frac{1}{\sqrt{n}}$次也。

是以計算 μ_2 與 σ 之標準差時，吾人可以已知之 μ'_4 與 μ'_2，代替未知之 μ_4 與 μ_2，並於計算平均數之標準差時，吾人可以 σ' 代σ，以調正之。

高級動差或相關係數之標準差及差誤頻數曲線，無論如何，不能用此法（註三）算出，玆特於下文將全部基礎重新建立，將一切論辯改弦更張，至所根據之書籍，已介紹於本章之首矣。

第六節 平均數之標準差（悶計及逆機率）

設一範圍共有可量之物件N個，對量數x_1共有$N\times y_1$個，對量數x_2共有$N\times y_2$個，其他以此類推。並設隨機抽出n個物件，$\frac{n}{N}$甚小，故求一x值之機率，不至因前所抽過而受影響。

$$N=N\times y_1+N\times y_2+\cdots\cdots,\ \therefore y_1+y_2+\cdots=1。$$

以$\bar{x}'$，σ'，μ'_2為樣本之平均數，標準差，及第二級動差。茲請決定此等數值之精度，以為範圍中平均數，標準差，第二動差之代表。

設$x_1, x_2\cdots\cdots$從範圍之（未知）平均數算出，故

$$\mu_1=S(x_t y_t)=0。$$

以μ為範圍之第二動差，故$\mu_2=S.x^2_t y_t$，復以$\mu_2=\sigma^2$。

樣本不能於x_1處有$x\times y_1$個，於x_2處有$N\times y_2$個，（餘類推）。

設實際求得者，在$x_1\cdots$有(y_1+e_1)個，在$x_t\cdots$有(y_t+e_t)個。

於是$e_1+e_2+\cdots+e_t+\cdots=0$。

$x_1, x_2\cdots\cdots$本為常數。

y_t既為在x_t求一物件之機率，且實驗舉行n次之多，e_t自為常態頻數，而標準差為

$$\sqrt{\left\{\frac{y_t(1-y_t)}{n}\right\}}\text{（見第二章第八節）。}$$

所以e_t^2所有各值之平均為$\frac{y_t(1-y_t)}{n}$，而e_t係屬於$\frac{1}{\sqrt{n}}$一

次。

假以E爲除第s及第t區分外其餘所有之總和差誤。以Y代$1-y_s-y_t$。

於是 $e_s+e_t+\mathrm{E}=0$

$\therefore 2e_se_t=\mathrm{E}^2-e_s^2-e$

$\therefore$ 平均值$e_se_t=\frac{1}{2}$平均值$\mathrm{E}^2-\frac{1}{2}$平均值$e_s^2-\frac{1}{2}$平均值e_t^2

$$=\frac{1}{2n}\{\mathrm{Y}(1-\mathrm{Y})-y_s(1-y_s)-y_t(1-y_t)\}$$

$$=\frac{1}{2n}\{(1-y_s-y_t)(y_s+y_t)-y_s(1-y_s)-y_t(1-y_t)\}$$

$$=-\frac{y_sy_t}{n}。$$

現以F爲y_1, y_2……之任何直線函數，則

$$\mathrm{F}=a_1y_1+a_2y_2+\cdots\cdots$$

式中，a_1, a_2……爲已知常數。

以$\mathrm{F}+f=a_1(y_1+e_1)+\cdots\cdots+a_t(y_t+e_t)+\cdots\cdots$

則 $f=a_1e_1+\cdots\cdots+a_te_t+\cdots\cdots$

$f^2=\mathrm{S}a_t^2e_t^2+2\mathrm{S}a_sa_te_se_t$。

推算至此，當e_1，e_2……之一切可能值，已查知確爲適當之比例，以σf^2代f^2之平均值，則

$$\sigma_f^2=\mathrm{S}a_t^2(\text{平均值}e_t^2)+2\mathrm{S}a_sa_t(\text{平均值}e_se_t)$$

$$=\frac{1}{n}\{Sa_t^2y_t(1-y_t)-2Sa_sa_ty_sy_t\}$$

$$=\frac{1}{n}\{Sa_t^2y_t-F^2\}\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots(122)$$

以 $a_1=x_1\cdots\cdots,a_t=x_t\cdots\cdots$

$F=Sx_ty_t=0$

$\bar{x}'=F+f=Sx_te_t$

$\therefore\sigma_{\bar{x}}^2$（$=\bar{x}'^2$之平均值）$=\frac{1}{n}Sx_t^2y_t$（依公式122）$=\frac{\mu_2}{n}\cdots(123)$

故$\bar{x}'$屬於與$\frac{1}{\sqrt{n}}$同次。

復次，以$a_1=x_1^2\cdots\cdots a_t=x_t^2$

$$\mu_2=F=Sx_t^2y_t,\ \mu'_2=S(x_t-\bar{x}')^2(y_t+e_t)$$

$\mu_2'-\mu_2=f=Sx_t^2e_t-2\bar{x}'Sx_ty_t$ ＋包含 $\bar{x}'e_t$ 及 $\bar{x}'^2$ 之各項，而此等乃屬於$\frac{1}{\sqrt{n}}$一次。

$\therefore\mu'_2=\mu_2=Sx_t^2e_t$，因為 $Sx_ty_t=0$ 當屬於$\frac{1}{n}$次各項略去時。

$\therefore$依公式122 $\sigma\mu_2{}^2=(\mu'_2-\mu_2)^2$ 之平均值

$$=\frac{1}{n}\{S(x_t^4y_t)-\mu_2^2\}=\frac{\mu_4-\mu_2^2}{n}\ \cdots\ (124)$$

現在$\sigma^2=\mu_2$

所以增加數$\delta\sigma$，σ之$\delta\mu$，及μ_2，用下方程式，連結起來

$$2\sigma\delta\sigma=\delta\mu_2,\ 或\ \delta\sigma=\frac{\delta\mu_2}{2\sqrt{\mu_2}}\cdots\cdots\cdots\cdots\cdots\cdots\cdots(125)$$

故　　$\sigma\sigma^2=$平均值$(\delta\sigma)^2=$平均值$\frac{(\delta\mu_2)_2}{4\mu_2}$。

∴　　$\sigma_\sigma{}^2=\frac{\mu_4-\mu_2{}^2}{4\mu_2 n}$（依公式124）。

類似之分析,得出概括之結果

$$\sigma_{\mu p}{}^2=\frac{1}{n}(\mu_{2p}-2p\mu_p+1\mu_{p-1}+p^2\mu_{p-1}\mu_2-\mu_p{}^2)\cdots(126)$$

故 $\bar{x}$, σ 與所有各級動差之標準差,均雜有$\frac{1}{\sqrt{n}}$一項,且如 n 甚大時,明顯量數與眞實量數之相差,必屬於$\frac{1}{\sqrt{n}}$一次,於含有此等相差之公式中,均不妨刪略之。所以,以數值表示公式 123 至 126 時,動差 μ'_2……等之計算值,可用以代替未可知之眞值也。

請注意,各級動差之標準差,隨兩倍於該次之動差而定,而此較高級動差,當方次增高時,亦立卽隨之而變大。實際上,據吾人所知,n 之值如爲普通常見者,高過四級之動差,精度必難保持,卽此故也。

如範圍之頻數曲線,係爲常態,則 $\mu_4=3\mu_2{}^2$,且

$$\sigma=\frac{\sigma}{\sqrt{n}},\ \sigma_{\mu_2}=\sigma^2\sqrt{\frac{2}{n}},\ \sigma_\sigma=\frac{\sigma}{\sqrt{2n}},$$

$$\sigma_{\mu_3}=\sigma^3\sqrt{\frac{6}{n}},\ \sigma_{\mu_4}=\sqrt{\frac{6}{n}}\cdots\cdots\cdots\cdots(127)$$

爲首兩個得數,如在常態曲線下,可用下法,直接得出。設有一羣類,其頻數曲線爲

$$y=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-x_0)^2}{2\sigma^2}},$$

（式中，x_0 與 σ 爲未知數）。今由此羣類內，隨機抽出事物n個，茲以$X_1, X_2\cdots\cdots X_n$代表之。

此等特定之 n 個事物被抽選之機率，P，爲

$$\frac{1}{\sigma\sqrt{\pi}}e^{-\frac{(X_1-x_0)^2}{2\sigma^2}}\times\frac{1}{\sigma\sqrt{\pi}}e^{-\frac{(X_2-x_0)^2}{\sigma^2}}\times\cdots=\frac{1}{\sigma n\pi\frac{n}{2}}e^{-s\frac{(X_1-x_0)^2}{2\sigma^2}}.$$

設 $\bar{x}$ 爲X各值之平均數，$Xt=\bar{x}+xt$，則 $Sxt=0$。

$$\log Px=-n\log\sigma-\frac{n}{2}\log\pi-\frac{1}{2\sigma^2}S\{\bar{x}-x_0+xt\}^2$$

$$=-n\log\sigma-\frac{n}{2}\log\pi-\frac{1}{2\sigma^2}\{n(\bar{x}-x_0)^2+ns^2\},$$

式中 s 爲X各值之標準差。

$\bar{x}$及 s 爲已知數，據以決定 σ 及 x_0。

$\bar{x}-x_0$ 爲最小時，不論 σ 之值如何，Px 必爲最大。

設 x_0 爲 $\bar{x}$。

於是Px 爲最大，當$\frac{dPx}{d\sigma}$爲零，卽當

$$0=-\frac{n}{\sigma}+\frac{1}{\sigma^3}\cdot ns^2,\text{又 } \sigma=s.$$

以 $\sigma=s+\gamma$，$x_0=\bar{x}+\delta$，並以 P_0 爲 Px 於 $r=0=\delta$ 時之值。則

$$\log Px-\log P_0=-n\log\frac{\sigma}{s}-\frac{n}{2\sigma^2}(\delta^2+s^2)+\frac{n}{2s_2}.S^2$$

$$\frac{1}{n}\log\frac{Px}{P_0}=-\log\left(1+\frac{\gamma}{s}\right)-\frac{1}{2s^2}\left(1+\frac{\gamma}{s}\right)^{-2}(\delta^2+s\ )+\frac{1}{2}.$$

$$=-\frac{\gamma}{s}+\frac{\gamma^2}{2s^2}-\cdots-\frac{\delta^2}{2s^2}+\frac{2\gamma}{2s^3}\cdot s^2-\frac{3\gamma^2}{2s^4}\cdot s^2,$$

刪略 γ^3，$\gamma\sigma^2\cdots\cdots$，

$$=-\frac{\gamma^2}{s^2}-\frac{\delta^2}{2s^2}.$$

$$\therefore \qquad P_x = P_0 e^{-\frac{1}{2}\cdot\frac{\delta^2}{\left(\frac{s}{\sqrt{n}}\right)^2}} \cdot e^{-\frac{1}{2}\frac{\gamma^2}{\left(\frac{s}{\sqrt{2n}}\right)^2}} \quad \cdots\cdots(128)$$

故$\bar{x}$及s之差誤，彼此互不相干，且此等差誤均成常態之頻數，而其標準差分別爲$\frac{s}{\sqrt{n}}$，$\frac{s}{\sqrt{2n}}$，（s 爲樣本之標準差，此可以範圍之標準差，σ，代表之，當 n 甚大時，上文解答已得，不待詞費矣）。

第七節　相關係數之標準差

相關係數每易發生之差誤之標準差，可以下法得之（下文以薛伯博士(Dr. Sheppard)之法爲依據）。

設有數值二對，係從平均數測算之離中差，如 x_t，y_t 是。該兩平均數之標準差爲 σ_1，σ_2，第二動差爲λ，μ。設共有對數N個，並設 z_tN之坐標，係在$(\bar{x}+x_t, \bar{y}+y_t)$處，然則 $z_1+\cdots+z_t\cdots=1$。

同時，$S z_t x_t=0=S z_t y_t$。

經由各對取其 $x^s y^l$，而以 M_{sl} 代表 $x^s y^l$ 之平均值。於是 $M_{sl}=S_z x^s y^l$。復以M代M_{11}。

如是，如 r 爲N對之相關係數，

$$r=\frac{M}{\sigma_1\sigma_2}, \log r=\log M-\tfrac{1}{2}\log\lambda-\tfrac{1}{2}\log\mu。$$

現設隨機抽出n對，並設於 x_t，y_t處，選得之數，爲$(z_t+e_t)n$。

假如 x'，y' 爲求得之平均數

$x'=S(z_t+e_t)x_t=S x_t e_t$，又 $y'=S y_t e_t$。

樣本各值與範圍（有r，M，λ，μ）各值之離差，將方程式對 $\log r$ 微分，可成下列聯系之條件。

$$\frac{\delta r}{r}=\frac{\delta M}{M}-\frac{\delta\lambda}{2\lambda}-\frac{\delta\mu}{2\mu}。$$

當微量如e_t，x'，y'（均屬於$\frac{1}{\sqrt{n}}$ 次）任何兩數之乘積被删略時，

$$\delta M=S(z_t+e_t)(x_t-x')(y_t-y')-Sz_tx_ty_t=Sx_ty_te_t-x'.Sz_ty_t-y'.Sz_tx_t$$

$$\therefore \delta M=Sx_ty_te_t$$

如上（第六節公式122－126）所述，

$$\delta\lambda=Sx_t^2e_t \text{ 而 } \delta\mu=Sy_t^2e_t。$$

$$\therefore\qquad \frac{\delta r}{r}=S\left(\frac{x_ty_t}{M}-\frac{x_t^2}{2\lambda}-\frac{y_t^2}{2\mu}\right)e_t。$$

故依概括公式 122，如 σ_r 爲 r 之差誤之標準差，

$$\sigma_r^2=\frac{r^2}{n}\left[S\left(\frac{x_ty_t}{M}-\frac{x_t^2}{2\lambda}-\frac{x_t^2}{2\mu}\right)z_t-S\left\{\left(\frac{x_ty_t}{M}-\frac{x_t^2}{2\lambda}-\frac{y_t^2}{2\mu}\right)z_t\right\}^2\right]$$

$$=\frac{r^2}{n}\left\{\frac{M_{22}}{M^2}+\frac{\lambda_4}{2\lambda^2}+\frac{\mu_4}{4\mu^2}-\frac{M_{31}}{\lambda M}+\frac{M_{13}}{uM}-\frac{M_{22}}{2\lambda\mu}\right\},$$

因　$S\left(\frac{x_ty_t}{M}-\frac{x_t^2}{2\lambda}-\frac{y_t^2}{2\mu}\right)zt=1-\frac{1}{2}-\frac{1}{2}=0$，

式中，λ_4 與 μ_4爲X

如原始分配，與常態相關面相合，$\lambda_4=3\lambda^2$，$\mu_4=3\mu^2$，$M=r\sigma_1$

σ_2, $M_{22}=(1+2r^2)\sigma_1^2\sigma_2^2$, $M_{31}=3r\sigma_1^3\sigma_2$, $M_{13}=3r\sigma_1\sigma_2^3$,（見公式 106），又

$$\sigma r^2=\frac{r^2}{n}\left\{\frac{1+2r^2}{r^2}+\frac{3}{4}+\frac{3}{4}-3-3+\frac{1+2r^2}{2}\right\}=\frac{(1-r^2)^2}{n},$$

而 $$\sigma_r=\frac{1-r^2}{\sqrt{n}}\cdots\cdots(129)$$

此爲一般通用者，但係以分配近於常態爲假設。

迴歸係數，y 對 x 而言，爲 $r\frac{\sigma y}{\sigma_x}=\rho$。用此符號，$\rho=\frac{M}{\lambda}$，以與上述類似之法，仍可得之，

$$\sigma\rho^2=\frac{\rho^2}{n}\left\{\frac{M_{22}}{M^2}+\frac{\lambda_4}{\lambda^2}\quad\frac{2M_{31}}{\lambda M}\right\}$$

在任何種分配情形下。

在常態分配下

$$\sigma\rho^2=\frac{\rho^2}{n}\left\{\frac{1+2r^2}{r^2}+3-6\right\}=\frac{\rho^2}{n}\cdot\frac{1-r^2}{r^2}$$

$$\therefore\sigma\rho=\frac{1}{\sqrt{n}}\cdot\frac{\sigma y}{\sigma_x}\cdot\sqrt{1-r^2}。$$

在常態分配下，用下列方法（茲根據皮爾生教授方法）亦可得之：

假設 $x_1, y_1\cdots\cdots x_n, y_n$ 各對，從一平面抽選而來，該平面之未知中心爲 x_0, y_0，標準差爲 σ_1, σ_2，平均積爲 $r\sigma_1\sigma_2$。

設 $\bar{x}$, $\bar{y}$, s_1, s_2, r'，係由樣本中，算出。

n 對實現之機率爲

$$P=\frac{1}{(2\pi\sigma_1\sigma_2\sqrt{1-r^2})^n}e^{-\frac{1}{2(1-r^2)}\cdot S\left\{\frac{(xt-x_0)^2}{\sigma_1^2}+\frac{(yt-y_0)^2}{\sigma_2^2}-\frac{2r(xt-x_0)(yt-y_0)}{\sigma_1\sigma_2}\right\}}$$

$$\therefore \log Pz = -n\log 2\pi\sigma_1\sigma_2 - \frac{n}{2}\log(1-r^2)$$

$$-\frac{n}{2(1-r^2)}\left\{\frac{s_1^2+d_1^2}{\sigma_1^2}+\frac{s_2^2+d_2^2}{\sigma_2^2}-\frac{2r\cdot(r's_1s_2+d_1d_2)}{\sigma_1\sigma_2}\right\}$$

式中，$d_1=x_0-\bar{x}$，$d_2=y_0=\bar{y}$。

現在 r，σ_1，σ_2，d_1，d_2 爲未知，r'，s_1，s_2爲已知。

述明$\dfrac{\partial P}{\partial d_1}=0=\dfrac{\partial P}{\partial d_2}=\dfrac{\partial P}{\partial \sigma_1}=\dfrac{\partial P}{\partial \sigma_2}=\dfrac{\partial P}{\partial r}$，

等幾種條件後，吾人乃得此五個未知數，蓋此五個未知數，乃使 Px 爲最大也。

$$\frac{d_1}{\partial_1^2}-\frac{rd^2}{\sigma_1\sigma_2}=0=\frac{d_2}{\sigma_2^2}-\frac{rd_1}{\sigma_1\sigma_2},$$

由此 $d_1=d_2=0$，除非 $r^2=1$ 時則反是。

知此，以d_1 及 d_2 作零，

$$\frac{1}{\sigma_1}-\frac{s_1^2}{(1-r^2)\sigma_1^3}+\frac{rr's_1s_2}{(1-r^2)\sigma_1^2\sigma_2}=0=\frac{1}{\sigma_2}-\frac{s_2^2}{(1-r^2)\sigma_2^2}+\frac{rr's_1s_2}{(1-r^2)\sigma_1\sigma_2^2}$$

$$\therefore \qquad (1-r^2)\sigma_1^2\sigma_2=s_1^2\sigma_2-rr'\ s_1s_2\sigma$$

並且 $$(1-r^2)_1\sigma_2^2=s_2^2\sigma_1-rr's_1s_2\sigma_2,$$

由此 $$\frac{s_1}{\sigma_1}=\frac{s_2}{\sigma_2}k，又 1-r^2=k^2(1-rr')$$

又 $$\frac{r}{1-r^2}-\frac{r}{(1-r^2)^2}\left\{\frac{s_1^2}{\sigma_1^2}+\frac{s_2^2}{\sigma_2^2}-\frac{2rr's_1s_2}{\sigma_1\sigma_2}\right\}+\frac{r's_1s_2}{(1-r^2)\sigma_1\sigma_2}=0。$$

$$\therefore \qquad r(1-r^2)-2rk^2(1-rr')+r'(1-r^2)k^2=0。$$

故　$r=r'$，又$k=1$

Pz 故爲最大值，當標本中各值，作爲平面中各值時。卽以求得之 Pz，作爲 P_0。

現以 $\sigma_1=s_1+\gamma_1$，$\sigma_2=s_2+\gamma_2$，並 $r=r'+\rho$，次將有微量 d_1，d_2，γ_1，γ_2，ρ 之乘方之一切函數展開，而删略去其第三乘方。則得

$$\frac{1}{n}\log\frac{P}{P_0}=-\frac{1}{2(1-r^2)}\left(\frac{d_1^2}{s_1^2}+\frac{d_2^2}{s_2^2}-\frac{2r'd_1d_2}{s_1s_2}\right)+\frac{r}{2-r'^2}\left(\frac{\rho\gamma_1}{s_1}+\frac{\rho\gamma_2}{s_2}\right)$$

$$+\frac{r'^2}{1-r'^2}\cdot\frac{\gamma_1\gamma_2}{s_1s_2}-\frac{2-r'^2}{2(1-r'^2)}\left(\frac{\gamma_1^2}{s_1^2}+\frac{\gamma_2^2}{s_2^2}\right)-\frac{1+r'^2}{2(1-r'^2)}\rho^2$$

$$=-\frac{1}{2(1-r'^2)}\left(\frac{d_1}{s_1}-\frac{r'd_2}{s_2}\right)^2-\frac{d_2^2}{2s_2^2}-\frac{2-r'^2}{2(1-r'^2)}\left(\frac{\gamma_1}{s_1}-\frac{r'^2}{2-\gamma'^2}\cdot\frac{\gamma_2}{s_2}-\frac{r'}{2-r'^2}\rho\right)^2$$

$$-\frac{2}{2-r'^2}\left(\frac{\gamma_2}{s_2}-\frac{r'\rho}{2(1-r'^2)}\right)^2-\frac{\rho^2}{2(1-r'^2)^2}。$$

對兩極限依次積分

對 d_1 積分，以 d_2, γ_1, γ_2, ρ 作常數，

對 d_2 積分，以 γ_1, γ_2, ρ 作常數，

對 γ_1 積分，以 γ_2, ρ 作常數，

對 γ_2 積分，以 ρ 作常數。

於是求得：不論 x_0, y_0, σ_1, σ_2 之值爲何，凡從 $r'+\rho$ 一值而來之觀察，其全部機率爲

$$P=Ke^{-\frac{1}{2}\frac{\rho^2}{\left(\frac{1-r'^2}{\sqrt{n}}\right)^2}}。$$

換言之，此分配係爲常態，其 r' 之標準差爲 $\frac{1-r'^2}{\sqrt{n}}$。

第六節末及本節之末，用小體字之推算，指示吾人，如樣本所自產生之頻數羣類爲常態，則求得樣本中之 $\bar{x}$, σ, μ 及 r 內各種差誤之機率，可由常態機率函數 (the normal probability function) 得來；反之，範圍各數量對樣本相當各數量之離差，其機率亦可用之求得。惟於他種條件下，亦可求得同一解答，茲倘

待加以證明。

無論何種情形，該種數量，係列如下式：

$$F+f=a_1(y_1+e_1)+a_2(y_2+e_2)+\cdots\cdots,$$

式中，$e_1+e_2+\cdots=0$，又於$0=e_1=e_2=\cdots\cdots$時，$f=0$。

同理$y_1+y_2+\cdots\cdots=1$。

樣本所取個數，n，爲甚大時（本章第六節），e_1，$e_2\cdots\cdots$之頻數曲線，必爲常態。

如 $e_1,e_2\cdots\cdots$彼此互不相干，或 $x_1,x_2\cdots\cdots$各值，數目之多堪供吾人認爲果係彼此獨立，則吾人可立卽用第三章第四節之定理，而決定f之頻數係爲常態。

如將全部分析（見附錄九）作完，吾人可得一結論：如一範圍（樣本所自出）之條件相同，則在此條件之下，吾人可認定爲常態，由此常態性，卽達到平均數之常態性（the normality of average）。換言之，該範圍所受之限制，爲無論 t 值爲何，$\frac{\mu_t}{\sigma_t}$終爲有限數也。

（註一）　參閱統計學報一九〇八年號第三百八十一頁起愛基華氏一文；猶爾著統計原理概論之最後一章；皇家學會會報一八九八年，第一九一卷(A. 220)皮爾生及費郎(Féon)一文；第一九二卷(A. 229)薛伯(Sheppard)一文；(Biometrika)第二卷，第三部，第二八〇頁。

（註二）　如X之所有各值，原卽有均等之機會，則從範圍觀察（當X之值在X′

士x極限內)之機率爲$2\int_0^x \mathrm{P}x.dx$,如x甚小時;又用逆機率,宇宙中之值在此等極限內之機率,爲

$$2\int_0^x \mathbf{P}x dx \div \int_{-\infty}^{\infty} \mathbf{P}x dx = \frac{2}{\sigma\sqrt{2\pi}}\int_0^x e^{-\frac{1}{2}\frac{x^2}{\sigma^2}}.dx。$$

(註三) 此法係根據愛基華斯給著者之通信。

第十章　資料與公式適應之測驗

第一節　測驗方法

關於以數學公式，表示觀察資料，其通用方法，前已屢見不鮮，然公式究竟恰當與否，此爲必有之疑問，爲解答此疑問，於是測驗乃不可少矣。以常用語言之，此卽配合完好與否之測驗問題也。

茲舉例論之，前述（第三章第七節第七例）九百七十家每『單位』每週食物費用，列表於下。

食物費用	m' 實在件數	m 計算件數	$e=m\sim m'$ 相差額	標準差	$\frac{e^2}{m}$
5·5先令以下	18	22	4	4·6	·7
5·5…………	107	123	16	10·4	2·1
7·5…………	255	234	21	13·3	1·9
9·5…………	245	249	4	13·6	·1
11·5…………	173	168	5	11·8	·1
13·5…………	101	89	12	9·0	1·6
15·5…………	38	51	13	7·0	3·3
17·5…………	17	22	5	4·6	1·1
19·5…………	9	11	2	3·3	·4
21·5以上	7	1	6	?	36·0
總　計	970	970	88	——	47·3

計算得來之件數，係由大數律之第二近似值而得。以前所用之相率方法，係將計算數與各區分觀察數之差額相加，不計符號之正負，而以其總計，作爲總件數之百分數。如此算得之不相配合之百分數(percentage misfit)，卽爲 $88\div9.70=9.1\%$ 也。

此法缺點，在不能用任何之機率量數，令人一見不知其是否完善。綜觀二種公式，當視其求出之不相配合百分數，以最低者爲較完善；又當有數組之類似觀察時，吾人可用此法，查出對公式之配合爲最近，且可知何組觀察之分配最爲整齊也。

但如將區分拼合，則不相配合之百分數，大致可以消滅也。

爲測驗各個區分(compartment)之內容，尙有一簡法。如觀察共有N個，m_t 爲一區分內之計算數，則於此區分內，求出觀察 m_t+e_t 個之機率，爲

$$\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\cdot\frac{e_t^2}{\sigma^2}}\text{（公式19）},$$

式中 $\sigma^2=\frac{m_t}{N}\left(1-\frac{m_t}{N}\right)N$，至於超過標準差某一倍數或次倍數(sub-multiple)之機率，則已見於常態機率表（第二章第四節）矣。上例每級，除最末一級外，其標準差，已述如前，由此可知，九個差額中，少於標準差，σ，者，佔其四，在 σ 與 $\frac{3\sigma}{2}$ 之間者，只有其二，下餘三者，則無過於 2σ。各量數均有其或然之機會，故全部彙類，可以認爲有或然性，至最後一項，在 21.5 先令以上有七件，

乃其例外耳。

極端各級與中間各級之必不能連貫，乃一般觀察多類所不免者。

雖然，離中差並非爲獨立，蓋其總數必爲零；故卽使某一區分內之離中差，以其本身觀之，爲數之大，幾令人不能相信，但就全盤區分而論，則亦並非不可能之事。適於此種變更之測量，皮爾生教授曾爲之制定，玆將此種之一部分析，用簡單化之形式，並將其計算得數，列成簡表後，述及其應用方法，於下文討論之（請參閱 the Philosophical Magazine，第 302 號 157–175頁，一九〇〇年七月出版）。

設有一公式，假設其果能代表觀察之分配；玆由此公式，得出 $m_1, m_2, \cdots\cdots m_n$ 個觀察，分成 n 級或區分，而觀察總數，$N = m_1 + m_2 + \cdots\cdots m_n$。

作一實驗，或一類觀察，設各區分爲 $(m_1+e_1)\cdots(m_t+e_t)\cdots(m_n+e_n)$，使 $e_1+\cdots\cdots+e_t+\cdots\cdots+e_n=0$。

以 $p_1 = \dfrac{m_1}{N} \cdots\cdots p_t = \dfrac{m_t}{N} \cdots\cdots$

設一羣類頗能適合公式，並從此羣類內作一觀察，則此一觀察係落在第 t 級內之機率爲 p_t。

從無限大之範圍中，隨機抽取N個，則 m_t+e_t 落於此級之機率，爲

$$\frac{1}{\sigma_t\sqrt{2\pi}}e^{-\frac{e_t^2}{2\sigma_t^2}},$$

式中，$\sigma_t^2=p_t(1-p_t)N=p_tq_tN$，而 $q_t=1-p_t$。

至於所指各種差誤之連帶機率(joint chance)，爲

$$Ke^{-\frac{1}{2}X^2}，式中\ x^2=S\cdot\frac{e_t^2}{m_t}，又\ Se_t=0,$$

而K則爲一常數。

因爲，如共有兩個區分，$e_1+e_2=0$，且其連帶機率，等於每個之機率。

於是 $p=\frac{m_1}{N}, q=\frac{m_2}{N}, m_1+m_2=N$。

機率爲

$$\frac{N^{\frac{1}{2}}}{\sqrt{2\pi m_1m_2}}e^{-\frac{1}{2}\left(\frac{e_1^2}{m_1}+\frac{e_2^2}{m_2}\right)}，因\ \frac{e_1^2N}{m_1m_2}=\frac{e_1^2(m_2+m_1)}{m_1m_2}，且\ e_1^2=e_2^2.$$

如共有三個區分，$e_1+e_2+e_3=0$，

$m_1+m_2+m_3=N$，$\sigma_1^2=\frac{m_1}{N}\cdot\frac{m_2+m_3}{N}\cdot N$，

對 σ_2^2，σ_3^2，亦然。

$2e_1e_2=e_3^2-e_1^2-e_2^2$，

$$r\sigma_1\sigma_2=平均值\ e_1e_2=\tfrac{1}{2}(\sigma_3^2-\sigma_1^2-\sigma_2^2)$$

$$=\frac{1}{2N}\{m_3(m_1+m_2)-m_1(m_2+m_3)-m_2(m_1+m_3)\}$$

$$=-\frac{m_1m_2}{N}.$$（請與第九章第六節比照）

e_1 與 e_2 ——亦即 e_3　併發之機率，依常態相關面，乃為

$$\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}}e^{-\frac{1}{2(1-r^2)}\left(\frac{e_1^2}{\sigma_1^2}+\frac{e_2^2}{\sigma_2^2}-\frac{2re_1e_2}{\sigma_1\sigma^2}\right)}.$$

現在

$$\sigma_1^2\sigma_2^2(1-r^2)=\frac{m_1m_2(m_2+m_3)(m_1+m_2)}{N^2}-\frac{m_1^2m_2^2}{N^2}$$

$$=\frac{m_1m_2m_3}{N}$$

因 $m_1+m_2+m_3=N$ 也。

故 e 之指數為

$$-\frac{N}{2m_1m_2m_3}(e_1^2\sigma_2^2+e_2^2\sigma_1^2-2r\sigma_1\sigma_2e_1e_2)$$

$$=-\frac{N}{2m_1m_2m_3}\left\{\frac{e_1^2m_2(m_1+m_2)}{N}+\frac{e_2^2m_1(m_2+m_3)}{N}+\frac{2e_1e_2m_1m_2}{N}\right\}$$

$$=-\frac{1}{2m_1m_2m_3}\{(e_1+e_2)^2m_1m_2+e_1^2m_2m_3+e_2^2m_1m_3\}$$

$$=-\tfrac{1}{2}\left(\frac{e_1^2}{m_1}+\frac{e_2^2}{m_2}+\frac{e_3^2}{m_3}\right),\text{ 因 } e_1+e_2=-e_3。$$

現設第二第三個區分，業已合併為一，因而內中包 M+E 個觀察（其$M=m_2+m_3$，$E=e_2+e_3$），則機率為

$$K_1\cdot e^{-\frac{1}{2}\left(\frac{e_1^2}{m_1}+\frac{E^2}{M}\right)}$$

式中 K_1 為一常數。

由此觀之，只除第二區分，而不牽動第一區分，其效果，使常數變換，且在指數以$\frac{e_2^2}{m_2}+\frac{e_3^2}{m_3}$代替$\frac{E^2}{M}$。

同理，只有二區分，僅除其第三，而不動其第一第二，其效果必變換常數並在指數中以$\frac{e_3^2}{m_3}+\frac{e_4^2}{m_4}$代替$\frac{e_3^2}{m_3}$。其他以此類推。

故對於n個區分，差誤$e_1, e_2 \cdots\cdots e_n$機率，P，為

$$Ke^{-\frac{1}{2}X^2} \text{式中，} x^2=\frac{e_1^2}{m_1}+\frac{e_2^2}{m_2}+\cdots\cdots+\frac{e_n^2}{m_n},$$

且 $$e_1+e_2+\cdots\cdots e_n=0 \cdots\cdots\cdots\cdots\cdots\cdots (130)$$

於此請注意，x^2之式，與用以求相依係數者相同。

[皮爾生氏，曾用複相關方程——而不用上述歸納法——證明此公式。]

如各個區分之抽取，能以獨立，並不受$e_1+e_2+\cdots=0$一條件之拘束，則機率乃為

$$Ke^{-\frac{1}{2}X^2}\times e^{-\frac{1}{2}\left(\frac{e_1^2}{N-m}+\frac{e_2^2}{N-m}+\cdots\right)}$$

而其指數，乃為

$$-\frac{1}{2}\left(\frac{e_1^2N}{m_1(N-m_1)}+\cdots\cdots\right)=-\frac{1}{2}\left(\frac{e_1^2}{m_1}+\frac{e_1^2}{N-m_1}+\cdots\cdots\right)。$$

設有區分甚多，最大之分數$\frac{m_t}{N}$卻仍甚小，則指數之後段，與前段比較，相形之下，大可删略，於是兩式漸至相等，且相關之效

驗甚弱也。

如中間並無相關，則最末一因子——如不删略——既小於1，實現之機率，必較有相關者爲小。（進一步之推算中，須將常數消除）。故不相關機率之總和，雖比現用方法爲簡，但對公式之適切性，頗予吾人以不當不利之認識也。

差誤之每一系，凡得出 x^2 之某一值者，其機率必相同。論到對常態頻數平均數之離中差之機率，慣常之測算法，向左或向右如是其大之離中差，發現之機率，爲

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

依此同理，於測算此差誤系——或一或然性較小之系——發生之機率，只須將下式代以數值卽得：

$$2\int\int\cdots\cdots K_e^{-\frac{1}{2}X^2}\,d_x,$$

式中 d_x 係代表 $de_1\ de_2\cdots\cdots de_{n-1}$，積分有 $n-1$ 次，自 x 展至 ∞，其條件爲 $e_1+e_2+\cdots+e_n=0$，至 K 如何選擇，須視 $\int_{-\infty}^{\infty}Ke^{-\frac{1}{2}X^2}$ $d=1$ 而定。

因有此條件，致積分甚爲繁雜，故仍以皮爾生氏推算之原本爲參考爲便。

至其得數爲（當 n 爲偶數時）

$$P=\sqrt{\frac{2}{\pi}}^{\infty}\int_{z}e^{-\frac{1}{2}X^2}\cdot d_x+\sqrt{\frac{2}{\pi}}e^{-\frac{1}{2}X^2}\left(\frac{x}{1}+\frac{x_3}{1.3}+\cdots+\frac{x^{n-3}}{1\cdot3\cdot5\cdots\overline{n-3}}\right)$$

當 n 爲奇數時，爲

$$P=e^{-\frac{1}{2}X^2}\left(1+\frac{X^2}{2}+\cdots+\frac{X^{n-3}}{2\cdot 4\cdots \overline{n-3}}\right)\cdots\cdots(131)$$

X^2 與 n 之各值，求出之P值，於第一卷，第 155 頁起之 Biometrika，曾列一表。茲爲決定一公式，是否能以充分代表一觀察之羣類，以簡短形式，而建立演算法則。吾人方法，只須取出 X^2 之值，至 X^2 則對某一 n 值，使 $P=\frac{1}{2}$ 強，或照查常態機率表，使 P=.0455 弱，如此，乃相當於常態曲線中之兩個標準差。

n.	X.	P.	X.²	P.
3	1	·61	6	·050
4	2	·57	8	·046
5	3	·56	10	·040
6	4	·55	12	·035
7	5	·54	13	·043
8	6	·54	15	·036
9	7	·54	16	·042
10	8	·53	18	·035
11	9	·53	19	·040
12	10	·53	20	·045
13	11	·53	22	·038
14	12	·53	23	·042
15	13	·53	24	·046
16	14	·526	26	·038
17	15	·525	27	·041
18	16	·524	28	·045
19	17	·523	30	·037
20	18	·522		
25	23	·520		
30	28	·518		

如 $x^2<n-2$,則觀察是否現於公式所代表之羣類，其得失成敗之數乃相等。

如 $x^2>2n$,則反機率甚大。

嚴格言之,測驗應用時,應以觀察原有之區分爲區分，因區分之歸併,足以波及結果之P值。然事實上，降格而用未分級之觀察,時常遭遇困難,且在連續變量之例,如身長，原始之分級，乃隨測量之可能,而細密至無以復加也。

但此尙爲較小之難點,其更爲嚴重者,在任一區分內，觀察之 m_t+e_t,必須爲整數,而一般 m_t 則不能,且 e_t 之值,有幾個乃於代表最完好之分配中,方能出現。故其結果，數目最少之區分內,期望之數額,亦應有合理的大,否則對於 x^2，將不免有虛僞之分攤。此在事實上規定詳細之極端區分;且不論其或拒絕或融合,武斷之元素,終不能免,而詳密之測算,亦不可能矣。

反之,根據愛基華斯氏假說（第三章第五節）,測驗差誤常態曲線,或大數律之應用時,在橫坐標於標準差小量倍數——對於測量之獨立元素,數目愈低,標準差亦愈少——之外而希望其密切配合,絕無此理,故測驗僅可用於數目甚多之中間區分；然專顧及施用測量之範圍,又將失去此法之詳密性矣。

故用此法之結果，僅可得一寬泛，但常能充分確定之得數。

第二節 例證

如吾人將第三章第七節第七例之極端一級删去，則$x^2=11.3$，$n=9$，$P=\cdot18$，且卽用公式第二近似值已足。

如用皮爾生氏公式（在同頁）$x^2=21.4$，$n=9$，$P=.006$，但如將最低最高極端級除外，則 $x^2=4.1$，$n=8$，$P=.77$；故此公式，除該類之極端級外，對於中間八級之表現，實優爲之。

如僅分別計算各級之標準差，則所得結論亦同。

又第三章第七節第五例之學童年齡表，$n=8$。由常態曲線得：$x^2=16.7$，$P=.02$，竟不能適合。惟用第二近似值，$x^2=.47$，則P與1乃不能分開矣。

關於字母數之實驗（見第三章第七節第一例），十字之和，以5個字母爲一級，得$n=13$，而由常態曲線，得 $x^2=33$，$P=.001$，如取消二最高最低兩極端級，$n=10$，$x^2=6.1$，$P=.73$。惟第二近似值，雖盡包所有各級，仍得 $x^2=8.4$，$P=.74$。

100 字之和，以 20 字母爲一級，由常態曲線得 $n=10$，$x^2=2.96$，$P=.975$，已不能再用更高級近似值。

玆爲證實抽樣法則，或上述方法之適當起見，特舉一各種不同之例，以抽樣所成之分配，與樣本所自來之羣類比較。

以各種定率付股息之公司數

	樣本中數額 m'	所有各公司之相對數額 m	標準差	$\frac{e^2}{m}$
百分之三以下	34	30	5·3	·50
百分之三	108	108·8	8·9	0
百分之四	117	124·4	9·3	·44
百分之五	60	70·8	7·4	1·65
百分之六至百分之八	48	43·2	6·2	·53
百分之八	33	22·8	4·6	4·57
	400	400		7·72

由此，$n=6$，$\chi^2=7.72$，$P=.185$。結果良佳，只惜最高級多有不合耳。

此種測驗，應用於二次元(dimension)之分配，如第七章之字每總和一千對之例，即是。

從中心向左右各三，向上下各二，在理論上原有十一以上之觀察者，茲共得二十四塊方格，分別作爲區分。外面方格，用武斷方法，以重線分九部，以使各格貼近，而期在第二近似值，至少可得九個期望之觀察數。玆將結果列下；

	常態曲線		第二近似值	
	χ^2.	P.	χ^2.	P.
24個中間方格	20·8	·59	17·5	·79
9個外部方格	27·8		10·1	
共三十個區域	48·6	·035	27·6	·59

用第二近似值，外部之改進，甚爲顯著。

附 錄

數學摘錄

一、求 π 值之瓦立斯(Wallis)氏定理

經用簡單作圖之研究,吾人可見:當 n 爲一正整數時

$$\int_0^{\frac{\pi}{2}}\sin^{2n+1}x\cdot dx<\int_0^{\frac{\pi}{2}}\sin^{2n}x\cdot dx<\int_0^{\frac{\pi}{2}}\sin^{2n-1}x\cdot dx$$

$$\therefore\ \frac{2\cdot4\cdot6\cdots2n}{3\cdot5\cdot7\cdots(2n+1)}<\frac{1\cdot3\cdot5\cdots(2n-1)}{2\cdot4\cdot6\cdots2n}\cdot\frac{\pi}{2}<\frac{2\cdot4\cdot6\cdots(2n-2)}{3\cdot5\cdot7\cdots(2n-1)}。$$

$$\therefore\qquad \frac{2^{2n}(n!)^2}{(2n+1)!}<\frac{(2n)!}{2^{2n}(n!)^2}\ \frac{\pi}{2}<\frac{2^{2n}(n!)^2}{(2n)!}\cdot\frac{1}{2n}。$$

$$\therefore\qquad \frac{2^{2n}(n!)^2}{(2n)!\sqrt{2n+1}}<\sqrt{\frac{\pi}{2}}<\frac{2^{2n}(n!)^2}{(2n)!\sqrt{2n}}。$$

$$\therefore\qquad \sqrt{\frac{\pi}{2}}=\frac{2^{2n}(n!)^2}{(2\pi)!\sqrt{2\pi}},至\frac{1}{n}爲有效\cdots\cdots\cdots\cdots(132)$$

請參閱,吉卜生著:微積分論 (Gibson: Treatise on the Calculus, 1896, EX. XXVI. 22)

二、整數之乘羃總和

如假定

$$S_m=\sum_{t=1}^{t=m} t^r=am^{r+1}+bm^r+cm^{r-1}+\cdots\cdots,$$

用歸納法可求 a，b，c。

因 $(m+1)^r=S_{m+1}-S_m=a\{(m+1)^{r+1}-m^{r+1}\}$

$$+b\{(m+1)^r-m^r\}+c\{(m+1)^{r-1}-m^{r-1}\}\cdots$$

令 m^r，及 m^{r-1} 之係數相等，則得

$$1=a(r+1)$$

$$r=a\frac{(r+1)r}{2}+br,\text{ 又 } b=\tfrac{1}{2}$$

$$\frac{r(r-1)}{2}=a\frac{(r+1)\cdot(r-1)r}{6}+b\frac{r(r-1)}{2}+c(r-1),\ c=\frac{r}{12}\text{等等。}$$

$$\therefore\frac{1^r+2^r+\cdots+m^r}{m^{r+1}}=\frac{1}{r+1}+\frac{1}{2m}+\frac{r}{12m^2}+\cdots$$

如將 $\frac{1}{m}$ 略去，$\qquad=\frac{1}{r+1}$，

如將 $\frac{1}{m^2}$ 略去，$\qquad=\frac{1}{r+1}+\frac{1}{2m}$，……………………………(133)

三、求 $m!$ 之斯德令(Stirling)公式

用瓦立斯氏定理，可求得此公式之第一近似值。

以

$$z=\frac{(2m)!}{(2m)^{m+1}(m-1)!}=(2m-1)(2m-2)\cdots(2m-m)$$

$$\div(2m)^m。$$

$$\log z=\sum_{t=1}^{t=m}\log\left(1-\frac{t}{2m}\right)$$

$$-\log z=\frac{1+2+\quad+m}{2m}+\cdots\cdots+\frac{1^r+2^r+\quad+m^r}{r(2m)^r}+\cdots\cdots$$

$$=\sum_{r=1}^{r=\infty}\left\{\frac{1}{r\cdot 2^r}\cdot\left(\frac{m}{r+1}+\frac{1}{2}+\frac{r}{12m}\right)\right\},$$

此係根據附錄二，當删除$\frac{1}{m}$之較高級乘冪時。

$$但\quad\sum\frac{2^{-r}}{r(r+1)}=\sum\frac{2^{-r}}{r}-2\sum\frac{2^{-r-1}}{r+1}$$

$$=-\log(1-\tfrac{1}{2})+2\{\log(1+\tfrac{1}{2})+\tfrac{1}{2}\}=1-\log 2=\log\left(\frac{e}{2}\right)。$$

$$故-\ \log z=m\log\left(\frac{e}{2}\right)-\tfrac{1}{2}\log(1-\tfrac{1}{2})+\frac{1}{12m}$$

$$z=\left(\frac{2}{e}\right)^m\cdot 2^{-\frac{1}{2}}\cdot e^{-\frac{1}{12m}}=2^{m-\frac{1}{2}}\cdot e^{-m}\cdot\left(1-\frac{1}{12m}+\cdots\right)$$

將$\frac{1}{m}$略去　　$=2^{m-\frac{1}{2}}\cdot e^{-m}$，

惟依瓦立斯氏定理，

$$m!=\frac{(2m)!\,(2m)^{\frac{1}{2}}}{2^{2m}\cdot m!}\cdot\left(\frac{\pi}{2}\right)^{\frac{1}{2}}，至\frac{1}{m}爲有效，$$

$$= e \times \frac{(m\pi)^{\frac{1}{2}} \cdot m^m}{2^{m-1}}$$

$$\therefore m! = m^m \cdot \sqrt{2\pi m} \cdot e^{-m} \quad \cdots\cdots\cdots \quad (134)$$

此公式求出對於 10! 值在百分之一以內之差誤，且如 m 增加，則確度亦增高。

如書其全式，乃為

$$m! = m^m \cdot \sqrt{2\pi m} \cdot e^{-m+\frac{1}{12m}-\frac{1}{360m^3}+}$$

參閱，克里斯圖氏(Chrystal)『代數』第三十章。

四、猶勒麥克老令定理(Euler-Maclaurin) ——用求和表示積分

設以 $f(a), f(a+h), \cdots\cdots f(a+mh)$ 為對於 $\overline{m+1}$ 個 x 連續值之 $f(x)$ 值。

於是依泰樂爾(Taylor)氏展開法

$$f(a+h) = f(a) + hf'(a) + \frac{h^2}{2}f''(a) + \cdots\cdots$$

$$f(a+2h) = f(a+h) + hf'(a+h) + \frac{h^2}{2}f'(a+h) + \cdots\cdots$$

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

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

$$f(a+mh) = f(a+\overline{m-1}h) + hf'(a+\overline{m-1}h) + \frac{h^2}{2}f''(a+\overline{m-1}h) +$$

以$d=a+\overline{m+1}h$，$c=a+mh$，並加之。

$$f(b)-f(a)=h\sum_a^d F(x)+\frac{h^2}{2}\cdot\sum_a^d F'(x)+\frac{h^3}{3!}\sum_a^d F''(x)+\cdots,$$

式中，$F(x)=f'(x)$，又$\int F(x)\cdot dx=f(x)+$常數。

$$\therefore h\sum_a^d F(x)=\int_a^b F(x)\cdot dx-\frac{h^2}{2}\sum_a^d F'(x)-\frac{h^3}{3!}\sum_a^d F''(x)-$$

同理

$$h\sum_a^d F'(x)=\int_a^b F'(x)dx-\frac{h^2}{2}\sum_a^d F''(x)-\cdots$$

$$h\sum_a^d F''(x)=\int_a^b F''(x)dx-\cdots\cdots$$

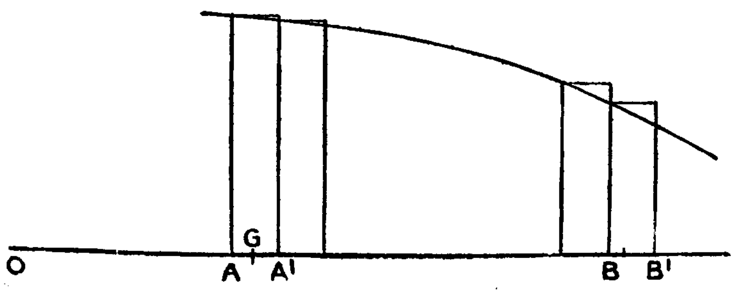
將上數方程式歸併，則得

$$h\sum_a^d F(x)=\int_a^b F(x)\cdot dx-\frac{h}{2}\{F(b)-F(a)\}+\frac{h^2}{12}\{F'(b)-F'(a)\}$$

$+$含有h^4之各項 ……(135)

$$\therefore h\sum_a^b F(x)=\int_a^b F(x)\cdot dx+\frac{h}{2}\{F(b)+F(a)\}+\frac{h^2}{12}\{F'(b)-F'(a)\}$$

$+$含有h^4之各項……(135)



上圖，以OA代表a，OB代b，AA′及BB′代h。

$AB=mh$。

$hE(a)$，$hF(a)$爲AA′，BB′上之長方形面積。$h\sum_a^d F(x)$爲AB上各長方形面積之總和。

$\int_a^b F(x)\cdot dx$ 爲AB上之曲線面積；$\frac{h}{2}[F(a)-F(b)]$一項，即爲曲線面積與直線面積缺陷部分之第一近似值。

此定理應用於差誤曲線，頗有幾許困難。

茲 $F(x)=\frac{1}{2\sqrt{2\pi}}e^{-\frac{x^2}{2\sigma^2}}$，且當$x$以$\sigma$爲比尙不甚大時，必爲一有限垂直之長度。

$\frac{1}{\sigma}=\frac{1}{\sqrt{pqn}}$，又$\therefore\frac{1}{\sqrt{n}}$，必爲垂直之有限數。

橫距離$AB=mb$必須爲有限。據第二章第四節例二之分析，得知在pn上下之成功數，必屬於$\sqrt{n}$次；故m屬於$\sqrt{n}$次，而h（單位級）屬於$\frac{1}{\sqrt{n}}$次。換言之，造圖時，長方形必作爲甚狹，以使許多長方形之和，以與$\sqrt{n}$比，終不得超過一有限之寬度。

依方程式(135)，h屬於$\frac{1}{\sqrt{n}}$次，$\sum_a^b F(x)$每一有限數均含有$\sqrt{n}$項，故$h\sum_a^b F(x)$屬於$F(x)$次，因此等於$\int_a^b F(x)\cdot dx$。

右邊各項，依次屬於$\frac{1}{\sqrt{n}}$，$\frac{1}{n}$……各次。（$F'(b)$ 自爲一簡單數字比率）。

現以 h 之值作一，則對於差誤常態曲線（其中屬於$\frac{1}{\sqrt{n}}$次者均删之），自 $pn+x_1$ 至 $pn+x_2$ 之總合成功機率爲

$$\int_{x_1}^{x_2} F(x) \cdot dx \quad \cdots\cdots (136)$$

更求進一級近似值，得

$$P_x = P_0 \cdot e^{-\frac{x^2}{2\sigma^2}} \cdot \left\{1 - \frac{k}{2}\left(\frac{x}{\sigma} - \frac{x^3}{3\sigma^3}\right)\right\},$$

式中，$\frac{1}{\sqrt{n}}$各項保留之，$\frac{1}{n}$各項删略之。公式 135 之 h 項亦須保留之。

以自 pn 至 $pn+x$ 成功機率之總合，求出得數，最爲簡便；故於上圖以A爲由G向左之半個單位，而 $OG=pn$，至G乃爲曲線重心之橫坐標。然後以 $GB=x$。

自G至B機率總和＝自A至B－$\frac{1}{2}\cdot P_0$之總和

$$= \int_0^x P_x \cdot dx + \tfrac{1}{2}(P_x + P_0) - \tfrac{1}{2}P_0.$$

以 $x = z\sigma$。

故自0至 $z\sigma$ 機率總和，當將$\frac{1}{\sigma^2}$删略時，等於

$$\frac{1}{\sqrt{2\pi}}\int_0^z e^{-\frac{1}{2}z^2}\left[1 - \frac{k}{2}\left(z - \frac{1}{3}z^3\right)\right] dz + \frac{1}{2\sigma\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \quad \cdots (137)$$

但 $\sigma=\sqrt{pqn}\,,k=\dfrac{q-p}{\sqrt{p.n}}$。

參閱投韓德(Todhunter)著,機率學史,第九百九十三節。

五、薛伯(Sheppard)氏對於頻數曲數動差之修正

(參閱 Biometrika 第三卷,第三百零八頁起)。

設以 $y=f(x)$ 爲一連續頻數曲線之方程式,曲線面積爲一。

設 $\mathrm{A}p$ 爲以 $xp\pm\frac{h}{2}$ 爲基線之面積,p 爲整數,並設不論 p 爲何值,$\mathrm{A}p$ 之一切值,由觀察均可知之.

由曲線方程算其第 t 次動差,如 m_t,乃等於 $\int_a^b x^t\cdot f(x)dx$,式中 a 與 b 爲 x 之極端值。

如每個面積認爲均集中於各級之中心,則從觀察求出第 t 級動差,爲 μ_t,等於 $\sum_a^b\cdot x_p{}^t\cdot\mathrm{A}p$。

但對於 μ_t 應加以如何之修正,始能得 m_t,茲請求之如下:

$$\mathrm{A}_p=\int_{xp-\frac{h}{2}}^{xp+\frac{h}{2}}f(x)\cdot dx=\int_{-\frac{h}{2}}^{\frac{h}{2}}f(x_p+x)\cdot dx$$

$$=\int_{-\frac{h}{2}}^{\frac{h}{2}}\{f(x_p)+xf'(x_p)+\frac{x^2}{2!}f^2(x_p)+\cdots\cdots\}\cdot dx$$

$$=hf(x_p)+\frac{h^3}{24}f^2(x_p)+\frac{h^5}{1920}f^4(x_p)+\cdots\cdots$$

所以

$$\mu_t=\sum_a^b hx_p^t f(xp)+\sum_a^b \frac{h^3}{24}x_p^t f^2(x_p)+\sum_a^b \frac{h^5}{1920}x_p^t f^4$$

$(x_p)+\cdots\cdots$

依猶勒麥克老令定理（公式135）

$$\int^b = \int_a^b x^t f(x)\cdot dx+\frac{h}{2}\{b^t f(b)+a^t f(a)\}+\frac{h^2}{12}[D(x^t f(x))]_a^b$$

$$-\frac{h^4}{720}[D^2(x^t f(x))]_a^b+\frac{h^2}{24}\int_a^b x^t f^2(x)\cdot dx+\frac{h^3}{48}\{b^t f^2(b)$$

$$+a^t f^2(a)\}+\frac{h^4}{288}[D(x^t f^2(x))]_a^b+\frac{h^4}{1920}\int_a^b x^t f^4(x)\cdot dx+$$

含有 h^5 之諸項。

玆有一調查，其曲線爲零而兩極端與坐標軸接觸，

則　　$f(a)=0=f(b)=f'(a)=f'(b)$,

並設接觸緊密，以致

$$h^2f^2(a)=h^2f^2(b)=0,\text{ 且 }h^4f^3(a)=h^4f^3(b)=0,$$

在此等情形下，又設乘數如 a^t, b^t, 與零並無重要之相差。

該式變爲 $\mu_t=\int_a^b x^t f(x)\cdot dx+\frac{h^2}{24}\int_a^b x^t f^2(a)dx+\frac{h^4}{1920}$

$\int_a^b x^t f^4(x)dx+$ 十含有 h^5 之諸項。

於是

$$\int_a^b x^t f^2(x)dx = [x^t f'(x)]_a^b - t\int_a^b x^{t-1} f'(x)dx = t(t-1)m_{t-2},$$

用連續局部積分方法，並根據前述條件，則

$$\int_a^b x^t f^4(x)dx = t(t-1)(t-2)(t-3)m_{t-4},$$

因與動差相比 h 普通均爲甚小，故涉及 h^3 之項，均可删略。

$$\therefore \mu_t = m_t \frac{h^2}{24} t(t-1)m_{t-2} + \frac{h^4}{1920} t(t-1)(t-2)(t-3)m_t - 4$$

此爲近似之數。

依次以 t 之值爲 1, 2, 3, 4, 則

$\mu_0 = m_0 =$ 曲線面積 $= 1$。

$\mu_1 = m_1 = 0$ 如方程式平均數經過垂直線時。

$$\mu_2 = m_2 + \frac{h^2}{12}$$

$\mu_3 = m_3 + \frac{h^2}{4} m_1 = m_3$ 如 $m_1 = 0$。

$$\mu_4 = m_4 + \frac{h^2}{2} m_2 + \frac{h^4}{80} m_0 = m_4 + \frac{h^2}{2}\left(\mu_2 - \frac{h^2}{12}\right) + \frac{h^4}{80}。$$

$$\therefore \qquad m_2 = \mu_2 - \frac{h^2}{12} \cdots\cdots (138)$$

$$m_4 = \mu_4 - \frac{h^2}{2}\mu_2 + \frac{7h^4}{240} \cdots\cdots (139)$$

至動差由平均數經過垂直線時，$m_1 = \mu_1$，$m_3 = \mu_3$。

六、對於普遍的差誤曲線第二近似值之動差及

常數曲線方程爲

$$y=\frac{1}{s\sqrt{2\pi}}e^{-\frac{x^2}{2s^2}}\left\{1-\frac{\kappa}{2}\left(\frac{x}{s}-\frac{1}{3}\cdot\frac{x^3}{s^3}\right)\right\}。$$

以 m_p 代 $\int_{-\infty}^{\infty}\frac{1}{s\sqrt{2\pi}}e^{-\frac{x^2}{2s^2}}\cdot x^p dx$。

於是 $m_0=1, m_1=m_3=\cdots\cdots=m_{2p+1}=\cdots\cdots=0, m_2=s^2,$

$m_2=s^2, m_{2p}=1\cdot3\cdot5\cdots\cdots(2p-1)\cdot S^{2p}$（見公式23）。

以 M_p 代第二近似值之第 p 級動。則

於是

$$M_p=\int_{-\infty}^{\infty}\frac{1}{s\sqrt{2\pi}}e^{-\frac{x^2}{2s^2}}\cdot x^p dx-\frac{\kappa}{2s}\int_{-\infty}^{\infty}\frac{1}{s\sqrt{2\pi}}e^{-\frac{x^2}{2s^2}}\cdot x^{p+1}dx$$
$$+\frac{\kappa}{6s^3}\int_{-\infty}^{\infty}\frac{1}{s\sqrt{2\pi}}e^{-\frac{x^2}{2s^2}}\cdot x^{p+3}dx$$

$$=m_p-\frac{\kappa}{2s}m_{p+1}+\frac{\kappa}{6s^3}\cdot m_{p+3},$$

$\therefore \qquad M_{2p}=m_{2p}$，因 $m_{2p+1}=0=m_{2p+3}$，

故偶數級動差不至因併入 κ 項，而受影響。

$$M_2=s^2\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots(140)$$

$M_{2p+1}=-\frac{\kappa}{2s}m_{2p+2}+\frac{\kappa}{6s^3}\cdot m_{2+4}$，因 $m_{2p+1}=0$

$$M_1=-\frac{\kappa}{2s}\left(m_2-\frac{1}{3s^2}m_4\right)=-\frac{\kappa}{2s}\left(s^2-\frac{1}{3s^2}\cdot 3s^4\right)=0$$

$$M_3=-\frac{\kappa}{2s}\left(3s^4-\frac{1}{3s^2}\cdot 15s^6\right)=\kappa\cdot s^3 \cdots\cdots\cdots\cdots(141)$$

$$M_{2p+1}=-\frac{\kappa}{2s}\cdot 1\cdot 3\cdot 5\cdots(2p+1)\cdot s^{2p+2}\left\{1-\frac{1}{3s^2}(2p+3)s^2\right\}$$

$$=\frac{p}{3}\cdot 1\cdot 3\cdot 5\cdots(2p+1)\cdot s^{2p-2}\cdot M_3 \cdots\cdots\cdots\cdots(142)$$

因 $M_1=0$，故原點爲曲線之平均數。

欲求衆數(mode)，須令$\frac{dy}{dx}$等於零。

因κ^2爲屬於$\frac{1}{n}$次，且在第三章第三節之分析，業將其删略，故

$$\log(ys\sqrt{2\pi})=-\frac{x^2}{2s^2}+\log\left\{1-\frac{\kappa}{2}\left(\frac{x}{s}-\frac{x^3}{3s^3}\right)\right\}=-\frac{x^2}{2s^2}$$

$$-\frac{\kappa}{2}\left(\frac{x}{s}-\frac{x^3}{3s^3}\right),$$

以 $x=-\frac{1}{2}\kappa s$，删略 κ^3，故

$$\therefore \frac{\text{衆數向右距平均數之距離}}{s}=\frac{1}{2}\kappa \cdots\cdots\cdots\cdots(144)$$

於是在底線ON上，曲線之面積(ON$=x=zs$)，如下得之：

$$Y_{z_0}=\frac{1}{\sqrt{2\pi}}\int_0^z e^{-\frac{1}{2}z^2}\left\{1-\frac{\kappa}{2}\left(z-\frac{z^3}{3}\right)\right\}dz$$

$$=\mathrm{F}(z)-\frac{\kappa}{6\sqrt{2\pi}}\{1-(1-z^2)e^{-\frac{1}{2}z^2}\}$$

$$=\mathrm{F}(z)-\kappa f(z),$$

式中 $$\mathrm{F}(z)=\frac{1}{\sqrt{2\pi}}\int_0^z e^{-\frac{1}{2}z^2}dz$$

$$f(z)=\frac{1}{6\sqrt{2\pi}}\{1-(1-z^2)e^{-\frac{1}{2}z^2}\}。$$

此等函數，在第二章第四節，及第三章第七節所列之表卽是。

$$\mathrm{Y}^0_{-z}=\mathrm{F}(z)+\kappa f(z)$$

且自 $-z$ 至 $+z$，卽 Y^z_{-z}，之全部機率，爲 $2\mathrm{F}(z)$，如在常態曲線者然（註一）。

如 M 爲中位數位置，則

$$\tfrac{1}{2}+\mathrm{MO}\text{上面積}=\mathrm{Y}^0_{-\infty},\ \tfrac{1}{2}-\mathrm{MO}\text{上面積}=\mathrm{Y}^\infty_0$$

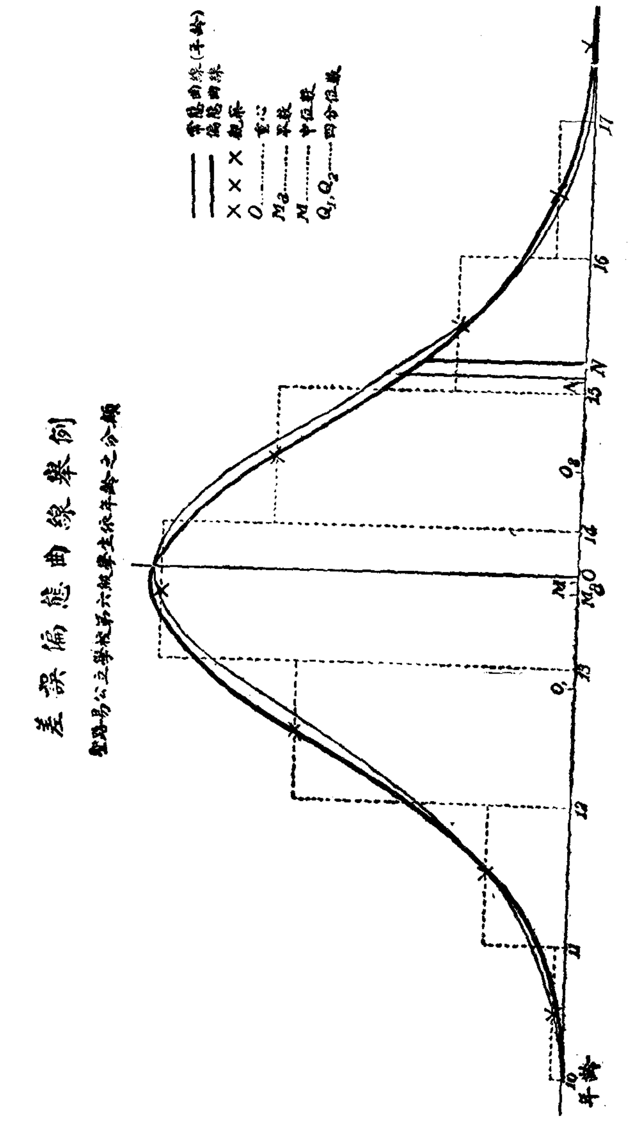
$$\therefore \mathrm{MO}\text{上面積}2\text{倍}=\mathrm{Y}^0_{-\infty}-\mathrm{Y}^\infty_0=\kappa\{f(-\infty)+f(\infty)\}=\frac{\kappa}{3\sqrt{2\pi}}。$$

在底線 OM 上之縱坐標，與 0 之縱坐標，卽 $\frac{1}{s\sqrt{2\pi}}$，相差只爲含有 κ 之項。

$$\therefore 2\mathrm{MO}\times\frac{1}{s\sqrt{2\pi}}=\frac{\kappa}{3\sqrt{2\pi}}，\text{當 }\kappa^2\text{ 可以删略時。}$$

又　$\mathrm{MO}=\frac{1}{6}\kappa s=\frac{1}{3}$（衆數與平均數之距離）……………………(145)

設底線 MN 上之面積——M爲中位數，N 爲任何一點——


差誤偏態曲線舉例
聖路易公立學校第六級學生依年齡之分類
常態曲線(年齡)
偏態曲線
××× 觀察
O……重心
Ma……衆數
M……中位數
Q_1, Q_3……四分位數
年齡
10
11
12
13
Q_1
M
Ma
O
14
Q_3
15
N
16
17

(第二圖)

等於 ON_1 上常態曲線之面積，卽 $F\left(\frac{x_1}{s}\right)$，式中 $x_1=ON_1$，並設 $NN_1=v$（v 值甚小，且屬於 κ 次，於下式可以看出）。

$$ON=x_1-v。$$

則　$F\left(\frac{x_1}{s}\right)=MO$ 上面積$+(x_1-v)$上面積

$$=\frac{\kappa}{6\sqrt{2\pi}}+\frac{1}{s\sqrt{2\pi}}\int_0^{x_1-v} e^{-\frac{x^2}{2s^2}}dx$$

$$-\frac{\kappa}{6\sqrt{2\pi}}\left\{1-\left(1-\frac{(x_1-v)^2}{s^2}\right)e^{-\frac{(x_1-v)^2}{2s^2}}\right\}$$

$$=\frac{\kappa}{6\sqrt{2\pi}}+F\left(\frac{x_1}{s}\right)-\frac{v}{s\sqrt{2\pi}}e^{-\frac{x_1^2}{2s^2}}-\frac{\kappa}{6\sqrt{2\pi}}+\frac{\kappa}{6\sqrt{2\pi}}\left(1-\frac{x_1^2}{s^2}\right)e^{-\frac{x_1^2}{2s^2}},$$

式中，具有 v^2 及 $v\kappa$ 各項，業已删略矣。

$$\therefore\qquad v=\frac{\kappa s}{6}\left(1-\frac{x_1^2}{s^2}\right)。$$

如吾人能得知，橫軸上從最低至三種位置之相對觀察數，且能斷定頻數曲線方程式，卽爲上文所論之式時，則平均數及 s 與 κ，不難求得。此法可以數字例釋之。第三章第七節第五例之觀察相對數如下：

年齡極限	學童人數
0 至13歲	3044之·296
0 至15歲	3044之·867

0 至16歲　　　　3044之·969

設 m 爲年齡之中位數，s 歲爲標準差，κs^3＝第三動差：全爲未知數。

在上列數字表中，設M代表其年齡中位數，N代表十五歲。

則MN上面積，爲 $.867-.500=F(1.112)$，（見第二章常態機率表）。

故　　$ON_1=1\cdot112=\frac{x_1}{s}=z_1$

$15-m=MN=MO+ON_1-NN_1=\frac{1}{6}\kappa s+x_1-\frac{1}{6}\kappa s\left(1-\frac{x_1{}^2}{s^2}\right)$

$15-m=z_1s+\frac{1}{6}\kappa s z_1{}^2$（式中 $z_1=1\cdot112$）。

同理

$16-m=z_2s+\frac{1}{6}\kappa s z_2{}^2$，（式中 $z_2=1\cdot866$）

又 $m-13=z_3s-\frac{1}{6}\kappa s z_3{}^2$，（式中 $F(z_3)=\cdot204$，而 $z_3=\cdot536$）。

N如在M之左方時，必須取作負號，略經思索，便知其當然。

欲決定 m，s，及 κ 之值，現共有三方程式。

$$\tfrac{1}{6}\kappa s(z_1-z_3)+s=\frac{2}{z_1+z_3}$$

$$\tfrac{1}{6}\kappa s(z_2-z_3)+s=\frac{3}{z_2+z_3}$$

$\therefore$　$\kappa s=.278$　$s=1.187$　$\kappa=.234$　$m=13.623$

平均數$=m+\frac{1}{6}\kappa s=13.669$。

（請參閱統計學報一九〇二年號，自339至348頁）

由九級全部算得之動差上，可求出 $s=1.190$，$=\cdot 206$，且平均數$=13.665$。

在已知平均數或中位數，或曲線已知爲常態，而 $\kappa=0$ 時，兩種觀察值已足決定其他各值矣。

惟吾人須注意，代表第一近似值之曲線，與代表第二近似值者，必在 $x=s\sqrt{3}$ 處交叉。

MN上偏態曲線之面積＝ON 上當 x 等於$\pm s$時，常態曲線之面積。

偏態曲線，在 ON 之右，超過常態曲線之距離＝在ON 之左，所短缺之距離。

七、未加權平均數之比率

設$M_1, M_2 \cdots M_n$ 爲某一時間內n個數量之眞值，而以M_1', M_2' ……爲另外一時間內之值。

設$n\bar{m}=SM_t$，$M_t=\bar{m}+m_t$，$Sm_t=0$，$n\bar{m}'=SM'_t$，$M'_t=\bar{m}'+m'_t$，$Sm'_t=0$，$n\sigma_m{}^2=Sm_t{}^2$，$n\sigma_{m'}{}^2=Sm'_t{}^2$。

設$\bar{m}'=m(1+\rho)$，$M_t'=(1+\bar{u}+u_t)M_t$ 式中$Su_t=0$，

$$\therefore\qquad \rho=\bar{u}+\frac{Sm_tu_t}{n\bar{m}}。$$

$\bar{u}$ 爲上述數量增加率之平均值，ρ 爲數量平均值之增加率。假如二者之較大者，增加率不大，或較小之量，增加率較另一數量爲速，則 $\bar{u}$，與 ρ 漸趨相等。

假設此等數量因有錯誤，致成 $M_t(1+e_t)$ 及 $M'_t(1+e'_t)$……等數。則依公式70之規定，m 及 m' 中差誤之標準差，爲 $\frac{\sigma}{\sqrt{n}}\sqrt{\left(1+\frac{\sigma m^2}{\bar{m}^2}\right)}$ 及 $\frac{\sigma_1}{\sqrt{n}}\sqrt{\left(1+\frac{\sigma m'^2}{\bar{m}'^2}\right)}$ —— 式中之 σ 與 σ_1，分別表示 e_t 與 e'_t。

如兩組中之差誤，彼此係屬獨立，則依公式63之規定，

$$S_r^2=\frac{1}{n}\left\{\sigma^2\left(1+\frac{\sigma m^2}{\bar{m}^2}\right)+\sigma_1^2\left(1+\frac{\sigma \bar{m}'^2}{\bar{m}'^2}\right)\right\},$$

式中之 S_r，爲 $\frac{\bar{m}'}{\bar{m}}$，卽 $1+\rho$ 中差誤之標準差。

但 M_t' 各量數之差誤 e'_t，與 M_t 各量數之差誤 e_t，相差常屬無幾，且正負號相同。

以 $d_t=e'_t-e_t$。

則由平均數之比率所生之差誤，爲 e

$$\frac{\bar{m}'}{\bar{m}}(1+e)=\frac{S\{M_t'(1+e'_t)\}}{S\{M_t(1+e_t)\}}$$

$$=\frac{S\{M_t'(1+e_t')\}\cdot SM_t-S\{M_t(1+e_t)\}\cdot SM_t'}{S\{M_t(1+e_t)\}\cdot SM'_t}$$

$$=\frac{\bar{m}S(M_t'e_t')-\bar{m}'\cdot S(M_te_t)}{\bar{m}'S\{M_t(1+e_t)\}}$$ 將 e_t^2 與 e_te_t' 略去，

則 $$=\frac{\bar{m}S(M_t'd_t)+S\{(\bar{m}M_t'-\bar{m}'M_t)e_t\}}{n\bar{m}\bar{m}'}$$

如將 $\bar{u}-\rho$ 删略，

$$=\frac{S(M_t'd_t)}{n\bar{m}'}+\frac{S\{(u_t+\bar{u}-\rho)M_te_t\}}{n\bar{m}'}=\frac{S(M_t'd_t)}{n\bar{m}'}+\frac{SM_tu_te_t}{n\bar{m}'},$$

是故，如 S_r 爲 e 之標準差，σ_d 爲 d_t 之標準差，σ 爲 e_t 之標準差，或爲於非由一致之頻數曲線而來此等數值之加權標準差。

$$S_r^2=\frac{1}{n^2\bar{m}'^2}\sigma_d^2\cdot S(M_t'^2)+\frac{1}{n^2\bar{m}'^2}\cdot\sigma^2\cdot S(M_t^2u_t^2),\text{(見公式55)}$$

$$S(M'_t)^2=S(\bar{m}'+m'_t)^2=n(\bar{m}'^2+\sigma_m'^2),$$

$$S(M_t^2u_t^2)=n\bar{m}^2\sigma_u^2+n\sigma_m^2\sigma_u^2+Su_t^2(m_t^2-\sigma_m^2)+2\bar{m}$$

$$S(m_tu_t^2),$$

式中 $n\sigma_u^2=Su_t^2$

$=n\sigma_u^2(\bar{m}^2+\sigma_m^2)+$可以删略之諸項。

$$\therefore\quad S_r^2=\frac{1}{n}\sigma_d^2\left(1+\frac{\sigma_m'^2}{\bar{m}'^2}\right)+\frac{1}{n}\cdot\sigma^2\cdot\left(1+\frac{\sigma_m^2}{\bar{m}^2}\right)\cdot\sigma_u^2\cdot\frac{1}{(1-\rho)^2}$$

………………………………………………………………………… (146)

如 e_t 與 e_t' 彼此係互爲獨立，σ_d^2 必等於 $\sigma^2+\sigma_1^2$，但如 $e_t=e'_t$，……餘類推，σ_d 將爲零。故 σ_d 可謂爲在 0 與 $\sigma\sqrt{2}$ 之間。

第二項之大小，視 σ_u 而定，蓋 σ_u 表示各種數量增加率之變動，且可由觀察而求得也。

故兩時期之觀察數量，如各具有類似之差誤，而此兩種數量之增加率，又相差無幾，則所算平均數之比率，其差誤必然甚小——n若大時，則尤小矣。

八、加權平均數之比率

論至加權平均數，其公式愈趨複雜。

設 $W_t=\bar{w}_t+w_t$, $W_t'=\bar{w}'+w_t'$ 爲兩時期之兩個權數，而 $\bar{w}$, $\bar{w}'$ 爲權數之平均數。以 $n\sigma_w^2=Sw_t^2$, $n\sigma_w'^2=Sw'_t{}^2$。

設以 $W'_t=W_t(1+\bar{v}+v_t)$，式中 $Sv_t=0$，並以 $n\sigma_v^2=Sv_t^2$。

假設 W_t, W'_t 之值，訛爲 $W_t(1+\eta_t)$ 及 $W_t'(1+\eta_t')$，再以 σ' 爲 η_t 之標準差。其他字母代表，與附錄七同。

設 $\bar{m}_w=\dfrac{S(W_tM_t)}{SW_t}$, $\bar{m}'_w=\dfrac{S(W_t'M_t')}{SW_t'}$。

求$\dfrac{\bar{m}_w{}'}{\bar{m}_w}$之差誤，$e$。

$$\frac{\bar{m}'_w}{\bar{m}_w}(1+e)=\frac{S\{W_t'(1+\eta_t')M_t'(1+e_t')\}}{S\{W_t(1+\eta_t)M_t(1+e_t)\}}\cdot\frac{S\{W_t(1+\eta_t)\}}{S\{W_t'(1+\eta_t')\}},$$

由此觀之，如再加以推演，而將 η_t, e_t 之乘積及平方取銷，則

$$e=\frac{S(W_t'M_t'e_t')}{n\bar{w}'\bar{m}_w'}-\frac{S(W_tM_te_t)}{n\bar{w}\bar{m}_w}+\frac{S\{W_t(\bar{m}_w-M_t)\eta_t\}}{n\bar{w}\bar{m}_w}$$

$$-\frac{S\{W_t'(\bar{m}_w'-M_t')\eta_t'\}}{n\bar{w}'\bar{m}_w}\cdots\cdots\cdots\cdots(147)$$

欲求近似值，只須將各乘積所有之總和取消，蓋某一因子之總和，其值爲零也。由此乃得$\bar{m}_w=m$，$\bar{m}_w'=\bar{m}'$，$\bar{u}=\rho$，$\bar{w}'=(1+\bar{v})\bar{w}$，隨推演之進步，愈可化簡矣。

以$d_t'=\eta_t'-\eta_t$，以σ_d'爲其標準差。

於是
$$e=\frac{S(W_t'M_t'd_t)}{n\bar{w}'\bar{m}'}+\frac{S(W'_tM_{tu t})}{n\bar{w}'\bar{m}'}e_t+\frac{S(W_tM_tv_t)}{n\bar{w}'\bar{m}}e_t$$
$$+\frac{S(W_t'm_t'd_t')}{n\bar{w}'\bar{m}}+\frac{S(W_t'M_tu_t)}{n\bar{w}'\bar{m}'}\eta_t+\frac{S(W_tm_tv_t)}{n\bar{w}'\bar{m}}\eta_t。$$

所以，如用約略數，

$$S_r^2=\frac{1}{n}\left(1+\frac{\sigma_w'^2}{\bar{w}'^2}\right)\left(1+\frac{\sigma_m'^2}{\bar{m}'^2}\right)\sigma_d^2+\frac{1}{n}\left(1+\frac{\sigma_w'^2}{\bar{w}'^2}\right)\left(\frac{\sigma_m'^2}{\bar{m}'^2}\right)\sigma_d^2$$
$$+\frac{1}{n}\left(\frac{\sigma_u}{1+\bar{u}}+1\right)^2\left(\frac{\sigma_u'^2}{\bar{u}'^2}\right)\left(1+\frac{\sigma_m^2}{\bar{m}^2}\right)\left(\sigma^2+\sigma'^2\right)$$
$$+\frac{1}{n}\left(\frac{\sigma_v}{1+\bar{v}}\right)^2\left(1+\frac{\sigma_w'^2}{\bar{w}^2}\right)\left\{\left(1+\frac{\sigma_m^2}{\bar{m}^2}\right)\sigma^2+\frac{\sigma_m^2}{\bar{m}^2}\sigma'^2\right\}\quad(148)$$

涉及σ^2，σ_d^2——測量M_t數量之差誤者——之項，與平均數未加權者相似，但有一因子（大於1而普通小於2）涉及權數，且有一項涉及一小因子σ_v^2（測量權數變動之離散度者），則爲例外也。

於涉及σ'^2，$\sigma_{d}^{2\prime}$（測量權數之差誤者）之三項中，第一項包含因子$\left(\frac{\sigma_m'}{\bar{m}'}\right)^2$，而第三項包括因子$\left(\frac{\sigma_m}{\bar{m}}\right)^2$，此兩因子隨M數量之

離散度而變，離散輕者，該兩因子必小；至第二項，則含有$\left(\frac{\sigma u}{1+\bar{u}}\right)^2$，而此因子，當數量增加率幾近相等時，其值必小。

σ，σ'，σd，$\sigma d'$ 係數之實在值，可於觀察中得之，而其相對重要性(relative importance)亦於以查出；但於離散不劇之數量，增加率與相等不至過甚時，則權數之差誤，與數量之等量差誤相比，必不甚大。

在此情形之下，第一近似值乃得

$$S_r=\frac{\sigma d}{\sqrt{n}}\sqrt{\left(1+\left(\frac{\sigma w}{\bar{w}}\right)^2\right)} \quad \cdots\cdots\cdots\cdots\cdots\cdots (149)$$

惟若$\frac{\sigma u}{1+u}$不爲甚小，可得更近之近似值

$$S_r=\frac{1}{\sqrt{n}}\cdot\sqrt{\left[\left(1+\left(\frac{\sigma w}{w}\right)^2\right)\left\{\sigma d^2+\left(\frac{\sigma u}{1+\bar{u}}\right)^2\left(\sigma^2+\sigma'^2\right)\right\}\right]} \cdots (150)$$

σd——測量數差誤相差量者——雖然多半小於$\sqrt{2}\,\sigma$，然與差誤之一相比，其值已甚大。

故最好須於未删除某項時，即由觀察中概略測驗其係數；並於看出所删略之乘積並不甚小時，或任一差誤，或許特別甚大時，應以用公式147之完全式爲佳。

（參閱統計學報 1911–12 年，第八一至八八頁，『平均數確度之測量』一文）。

九、動差……等中差誤之標準差之常態性

[此則係根據薛伯氏著：差誤理論之應用 (Application of the theory of Error) 一文，見皇家學會會報，第一九二卷，一八九八年版，A二二九，第一一七至一二八頁；但標號及論述，則已加變更]。

在一『宇宙』中，共有事物N個，x_1 處有 p_1N 個，x_2 處有 p_2N 個，其他以此類推，$p_1+p_2+\cdots=1$，$F=a_1p_1+a_2p_2+\cdots$ 而 a_1，a_2……乃為常數。

茲就中抽出 n 個事物來，x_1 處有 n_1 個，x_2 處有 n_2 個，餘類推，$n_1+n_2+\cdots=n$。

以 $F+f=a_1\dfrac{n_1}{n}+a_2\dfrac{n_2}{n}+\cdots\cdots$

$$f=a_1\frac{n_1}{n}+a_2\frac{n_2}{n}+-F\left(\frac{n_1}{n}+\frac{n_2}{n}+\cdots\right)=b_1\frac{n_1}{n}+b_2\frac{n_2}{n}+\cdots\cdots,$$

式中 $b_1=a_1-F$，其他以此類推。

於是 $Sb_tp_t=Sa_tp_t-F\cdot Sp_t=F-F=0$

$$Sb_t^2p_t=Sa_t^2p_t-2FSa_tp_t+F^2\cdot Sp_t=Sa_t^2p_t-F^2。$$

現請求 M_s（＝平均值f^s），並證明其對於 M_2（＝平均值f^2）之關係，乃即於差誤常態曲線中所發生之觀象。

將此式 $$E=\left(p_1e^{b_1\frac{a}{n}}\right)^{n_1}\left(p_2e^{b_2\frac{a}{n}}\right)^{n_2}\cdots\cdots,$$

用多項式定理展開，則得各項之和：

$$\frac{n!}{n_1!\,n_2!\cdots\cdots}\left(p_1e^{b_1\frac{a}{n}}\right)^{n_1}\left(p_2e^{b_2\frac{a}{n}}\right)^{n_2}=\mathrm{P}.e^{fa}\text{ 各項之和},$$

但此係以下爲條件：$n_1+n_2+\cdots=n$，且式中

$$f=b_1\frac{n_1}{n}+b_2\frac{n_2}{n}+\cdots\cdots,$$

$$\mathrm{P}=\frac{n!}{n_1!\,n_2!\cdots\cdots}p_1^{n_1}p_2^{n_2}\cdots\cdots$$

P 乃爲於 x_1 抽出 n_1，於 x_2 抽出 n_2……之全部機率，將下多項式展開卽明：

$$(p_1+p_2+\cdots\cdots)^n$$

$$\therefore\ \mathrm{E}=\mathrm{P}\left(1+af+\frac{a^2}{2}f^2+\cdots\cdots+\frac{a^s}{s!}f^s+\cdots\cdots\right)\text{之和}$$

$$=\mathrm{M}_0+a\mathrm{M}_1+\frac{a^2}{2}\mathrm{M}_2+\cdots\cdots+\frac{a^s}{s!}\mathrm{M}^s+\cdots\cdots$$

$$\text{又 }\mathrm{E}=\left(\mathrm{S}p_t+\frac{a}{n}\mathrm{S}b_tp_t+\frac{a^2}{2n^2}\mathrm{S}b_t^2p_t+\frac{a^3}{6n^3}\mathrm{C}_3+\frac{a^4}{24n^4}\mathrm{C}_4+\cdots\right)^n$$

將 $e^{b_1\frac{a}{n}}$……各項展開必如此，但

$$\mathrm{C}_3=\mathrm{S}b_t^3p_t,\ \mathrm{C}_4=\mathrm{S}b_t^4p_t\cdots\cdots,\ \mathrm{S}p_t=1,\ \mathrm{S}b_tp_t=0。$$

$$\therefore\qquad \mathrm{E}=\left(1+\frac{a^2}{2n^2}\mathrm{S}b_t^2p_t+\cdots\cdots\right)^n。$$

求E之算式中，令為首三個係數相等。

$$M_0=1,M_1=0,M_2=\frac{1}{n}Sb_t{}^2p_t。$$

$$\therefore\ 1+\frac{a^2}{2}M_2+\cdots+\frac{a^s}{s!}M_s+\cdots=\left(1+\frac{a^2}{2n}M_2+\frac{a^3}{6n^3}C_3+\cdots\right)^n。$$

現知，n 若大，$M_2,\frac{C_3}{n^{\frac{3}{2}}},\frac{C_4}{n^2},\frac{C_5}{n^{\frac{5}{2}}}\cdots\cdots$必為有限數，故如將$\frac{1}{\sqrt{n}}$刪略，則得

$$1+\frac{a^2}{2}M_2+\cdots\cdots+\frac{a^s}{s!}M_s+\cdots\cdots=\left(1+\frac{a^2}{2n}M_2\right)^n=e^{\frac{a^2}{2}M_2}$$

$$=1+\frac{a^2}{2}M_2+\cdots\cdots+\frac{a^{2t}}{t!\,2^t}(M_2)^t+\cdots\cdots$$

故，在此情形下，s 若為奇數，$M_s=0$，$M_{2t}=\frac{(2t)!}{t!\,2^t}(M_2)^t$，如在差誤常態曲線中然。

此條件——$\frac{C_3}{n^{\frac{3}{2}}}\cdots\cdots$為有限數——與第三章第四節愛基華斯氏之證明，原理相同，但又不可一概而論，該項定理應用，應就每一情形，而加以討論。

例如第九章第六節所論，$f=x_1{}^2e_1+x_2{}^2e_2+\cdots\cdots$（式中之 e_1 為$\frac{n_1}{n}-p_1$），$F=\mu_2$（樣本所自抽出之『宇宙』中第二動差），及 $b_t=x_t{}^2-\mu_2$。

$$M_2=\frac{1}{n}S(x_t^2-\mu_2)^2p_t=\frac{1}{n}(\mu_4-\mu_2^2)$$

$$C_3=S(x_t^2-\mu_2)^3p_t=(\mu_6-3\mu_4\mu_2+2\mu_2^3)$$

$$\frac{C_3}{n^{\frac{3}{2}}}=M_2^{\frac{3}{2}}\cdot\frac{\mu_6-3\mu_4\mu_2+2\mu_2^3}{(\mu_4-\mu_2^2)^{\frac{3}{2}}}。$$

同理

$$\frac{C}{n^2}=(M_2)^2\frac{\mu_8-{}_4\mu_6\mu_2+{}_6\mu_4\mu_2^2-3\mu_2^4}{(\mu_4-\mu_2^2)}\cdots\cdots$$

由此推論，$\frac{\mu_4}{\sigma^4},\frac{\mu_6}{\sigma^6},\frac{\mu_8}{\sigma^8}$……各比率若爲有限數（而$\sigma^2=\mu_2$），則$\frac{C_3}{n^{\frac{3}{2}}},\frac{C_4}{n^2}$……爲有限數，果不出所料。

是以，『宇宙』之頻數曲線，若與此等條件適合，在平均數左右果有合理之集中現象，而在超過σ之小量倍數以外，又無偏重情形，則對於第二動差——亦卽標準差——之差誤，頻數曲線，必爲常態形狀。

類似而較爲簡短之證明，更可示吾人以平均數之差誤，亦具有常態之頻數($f=x_1e_1\cdots\cdots$, $F=0$, $b_t=x_t$)。

就相關係數（見第九章第七節）之分析而論，

$$b_t=r\left(\frac{x_ty_t}{M}-\frac{x_t^2}{2\lambda}-\frac{y_t^2}{2\mu}-1+\tfrac{1}{2}+\tfrac{1}{2}\right),p_t=z_t。$$

$$M_2=\frac{1}{n}S_{r^2}\left(\frac{x_ty_t}{M}-\frac{x_t^2}{2\lambda}-\frac{y_t^2}{2\mu}\right)^2z_t$$

$$=\frac{r^2}{n}\left(\frac{M_{22}}{M^2}+\frac{\lambda_4}{4\lambda}+\frac{\mu_4}{4\mu^2}-\frac{M_{13}}{M\lambda}-\frac{M_{13}}{M\mu}+\frac{M_{22}}{4\lambda\mu}\right)$$

$$C_3=S_r{}^3\left(\frac{x_ty_t}{M}-\frac{x_t{}^2}{2\lambda}-\frac{y_t{}^2}{2\mu}\right)^3z_t=r^3\cdot\left(\frac{M_{33}}{M^3}+\frac{\lambda_6}{8\lambda^3}+\cdots\cdots\right)。$$

以$\lambda=\sigma_1{}^2$,$\mu=\sigma_2{}^2$,則$\frac{\lambda_t}{\sigma_1{}^t}$,$\frac{\mu_t}{\sigma_2{}^t}$,$\frac{M_{st}}{\sigma_1{}^s\sigma_1{}^t}$若爲有限數,不論 s 及 t 爲任何值,$\frac{C_3}{n^{\frac{3}{2}}}$必等於$M_2{}^{\frac{3}{2}}\times$ 有限數,且依此辦法,更高級之項,亦可同樣處置也。

故如二次元(two-dimensional)頻數分配之動差及乘積,果與前述之條件相合,則相關係數之差誤曲線,必爲常態形式。

十、最小二乘法

此法之由來甚久,於若干不確切之量數中,指示所抽出之多個數值時,用之。

設一數量 z,因方程式 $z=u_1x_1+u_2x_2+\cdots\cdot+u_kx_k$,(式中,$u_1$,$u_2$,$\cdots\cdots$爲可以由觀察得來之數量)而與 k 個未知之常數,x_1,$x_2\cdots\cdots x_k$,有關;並設由 n 組觀察可得

$$1^z={}_1u_1x_1+{}_1u_2x_2+\cdots\cdots+{}_1u_kx_k$$

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

$$n^z={}_nu_1x_1+{}_nu_2x_2+\cdots\cdots+{}_nu_kx_k$$

式中之 z 及 u,均爲已知數。

n 若等於 k, x 各值卽可確實求出。若 $n<k$, 則解答之式將有無限之多,而方程式乃無從決定。

若 $n>k$,則方程式必有矛盾,於是問題乃變爲如何決定 x_1, x_2……各值之問題,蓋必決定 x_1, x_2……之值, 以使矛盾性縮爲最小,至矛盾之來,則假定其由於測量 u 時,量數之不完全也。

以 d_1, d_2……代表 ${}_1z, {}_2z$……與由 x_1, x_2……眞值所得之值,卽 X_1, X_2……,二者間之相差額。

則
$$
\begin{aligned}
&{}_1u_1X_1+{}_1u_2X_2+\cdots\cdots+{}_1u_kX_k-{}_1z=d_1\\
&{}_2u_1X_1+{}_2u_2X_2+\cdots\cdots+{}_2u_kX_k-{}_2z=d_2\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots
\end{aligned}
$$

假定 d_1, d_2……卽爲差誤,而此差誤之機率,可由一常態曲線 $P=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{d^2}{2\sigma^2}}$ 得來。此假定則多係根據理證(demonstrations)而得,蓋在某一關於偶然差誤 (accidental error) 之假定下,此常態形式可應運而出也。不論對於物體或測地(geodetical)之量數,假定之效度如何,均不便承認此等假定對於統計或生物測量能以適用,至爲對平均數之離中差量,抑或由於抽樣方法之差誤,則概所不問也。

欲得解答,須先求 X_1, X_2……之各值,由此各值可使 d_1, d_2……全體出現之機率爲最大,換言之,卽使

$$d_1^2+d_2^2+\cdots\cdots$$

之和爲最小。茲以 $f(d_1,d_2\cdots\cdots)$ 代表此總和。

成爲最小之條件，爲 $\frac{\partial f}{\partial X_1}=0=\frac{\partial f}{\partial X_2}=\cdots\cdots$

由此得

$$\frac{1}{2}\cdot\frac{\partial f}{\partial X_1}=\mathrm{S}_{t=1}^{t=n}\,{}_tu_1({}_tu_1X_1+{}_tu_2X_2+\cdots\cdots-{}_tz)=0$$

$$\frac{1}{2}\cdot\frac{\partial f}{\partial X_2}=\mathrm{S}_{t=1}^{t=n}\,{}_tu_2({}_tu_1X_2+{}_tu_2X_2+\cdots\cdots-{}_tz)=0$$

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

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

書如下式亦可：

$$X_1\cdot Su_1^2+X_2\cdot Su_1u_2+\cdots\cdots=Su_1z$$

$$X_1Su_1u_2+X_2\cdot Su_2^2+\cdots\cdots=Su_2z$$

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

$$X_1Su_1u_k+X_2Su_2u_k+\cdots\cdots=Su_kz,$$

以上共有 k 個方程式，由此可求出 k 個量數 X_1，$X_2\cdots\cdots$

此種方法於實用上，頗能求出大致甚佳之未知數之經驗值。其簡單形式，於第一編第十章第二節最後一法，曾一用之，而此法之效度，可以測驗者，本編第六章第七節曾採用之。

〔參閱麥理曼(Merriman)著：最小二乘法(Method of Least

Squares),及外爾德(Weld)著:差誤理論及最小二乘方(Theory of Errors and Least Squares)〕。

(註一) 應用狄勒麥克老令定理,依照 p, q, n, 之假說,其相當公式,爲 $2F(z)+\frac{e^{-\frac{1}{2}z^2}}{S\sqrt{2\pi}}$; 但若於資料爲連續時,其最末一項,必消滅。