經濟統計學

鄭仲陶 著

經濟統計學

第一編

統計資料

(Statistical Data)

預定的目的；且其排列，足以表示彼此相互的關係。(註1)

從上面的定義來分析，我們應當注意下列各點：

(1) 統計是事實的綜合，必定是許多事件的集合。所以單獨的事實，不成爲統計，例如濟南慘案，和孫中山先生的逝世，都是一件事，不能說是統計，必定要幾件慘案、或死亡，才能成爲統計。

(2) 統計是表明數字的，所討論的是數量，不是質量。所以須用數字來表明差異。例如敍述某處歷年棉花收穫狀況，以『優』『適中』『劣』等來表明，不足爲統計，(除非能代表相當數量)必定要以每畝收穫的數量來表示才可。

(3) 統計所搜集的材料，不必都是實數，可以估計。例如美國農業經濟局的農產預測統計，都憑估計。不過無論是實際計數，或是估計的資料，假如要得一個合理的結論，而且可以綜合之，平均之，撮要之，那末必須依照相當準確標準。相當二字，全恃統計之目的而定，沒有普遍共同的原則；因爲統計之目的，有時必須很精密，有時祇要概況就夠了。

(4) 假如數量的測量，可以認爲統計，必須用有系統的方法

註1. We shall use the term statistics as meaning aggregates of facts, "affected to a marked extent by a multiplicity of causes," numerically expressed, enumerated, or estimated according to reasonable standards of accuracy, collected in a systematic manner for a predetermined purpose, and placed in relation to each other. Secrist: An Introduction to Statistical Methods P. 10.

第壹編

統計資料

第一章

緒論

統計 (Statistics) 的意義有兩方面：一方面是指統計的資料，而他方面則指統計方法。後者的問題，就是本書要解答的，我們相信讀者讀過此書以後，對於統計方法，自然有一個正確的認識。但在開始敍述以前，為便利討論起見，我們覺得有替統計下個定義的必要。統計的定義很多，各著者有他自己的定義，但依著者的鄙見，認為 Secrist 的定義，含義較廣，特為介紹。

統計的定義 統計為事實之綜合，按照相當準確標準用數字來表示、計數、或估計；用有系統的方法搜集之，以求達到

間、地域、或事物的屬性以表出之。這是因爲事物的比較，必須這些事物有共同的性質，誠然如 Bowley 所說，惟同類始可相比較 (Like can only be compared with like) 。否則，牛溲馬渤，敗鼓之皮，兼收并蓄，豈可算是統計。

統計方法　統計這名詞，他方面可以專指一種科學，這就是平常所說的統計方法。統計方法，包括分析和綜合的科學方法，所以用以搜集統計之資料，或闡明各種單獨或相關的現象。這些方法，是關於：(1)選擇和搜集資料，(2)依共同的特質而分類整理，(3)如何將已分類的資料陳示之 (Presentation)，(4)如何用平均數等來撮總 (Summarizing)、集約 (Abbreviating)，(5)測量相互的關係等方法。將此種統計方法，作一系統的敍述，卽成所謂統計學。

經濟統計學　經濟統計學 (Economic Statistics) 是應用統計學 (Applied Statistics) 的一種。所謂應用統計學，就是拿統計方法，應用於某種具體事實上去。經濟統計學，是特別注重應用於經濟方面的一般方法。我們不可以爲應用於經濟方面的

方法，其他方面決不會用到；更不可以爲在經濟方面，普通一般之方法，決用不到。經濟統計學，僅僅是特別注重統計方法，在經濟問題之應用，所以本書一二兩編，大部分仍爲一般統計方法的範圍，而第三四兩編中，始進於經濟統計的領域。

統計調查之主要目的

tion)變量，故變量之性質，及表示變量之單位

論。

變量 (Variable) 凡是表示數(Number)或量(Magnitude

的差別之任何事物，皆稱變量。變量可爲具體之事物，又可

抽象之比率。變量可由計數(Counting)而得，可由測量 (mea

surement)而得，又可由計算 (Computation)而來。變量之例

隨處皆是。人體之脈搏，一變量也；一國之人口，又一變量

；他如身之長度，體之重量，一物之物價，一室之溫度，火

之延車公里，山芋一畝之平均產量等，均爲變量。變量之意

旣明，試進而討論變量之性質。就變量之性質而論，可分連續

變量 (Continuous Variable) ，和不連續變量 (Discontinuous

Variable) 二種。連續變量佔量表 (Scale) 中兩點間之距離

可細分至無盡數，如測量身長，年齡，船之噸數，皆爲連續變

量。如言身長55寸，此不過言其最相近之度數而已，實則此5

寸之高度，不一定恰在55之一點，或在 54.5 與 55.5 距離之間

；再精密分之，或在54.49與55,49距離之中間。就理論講，

種變量，無論測量至如何精密，總有更加精密之可能，雖則現代科學，還未發達到此程度，在另一方面，事實上亦無需過於精密。不連續變量，乃係量表上之一點，依其所用計量之單位，自然確定者，不可再行細分，如火車所載之旅客，一册書之頁數，一籃中之果子等，皆有一定之整數。有些變量，在性質上，雖屬於不連續的，可以當作連續的看，如在大量觀察時之人口，巨數收入時之金錢是。

單位(Units)統計雖是數的科學，可是這些量數，不是1000或500抽象的數目，乃是1000頭牛，或500個人等有單位的變量。數字若脫離了單位，便毫無意義，因爲單位是表示所研究的事物之特徵。有些單位的習用解釋的差異程度很大，所以我們應將單位先下一個正確的定義。但是定義當根據統計的目的，不顧目的決不能有其定義。而在另一方面，定義模稜不清，有調查事實實際不同者，誤爲相同的弊竇，調查目的，亦無由達到。單位一經定義，在調查進行中，應始終一律適用這定義，這點當然是很重要的。

單位種類——以性質分　統計單位以其性質的不同，可分爲下列數類。

(1) 個體的單位 (Units of the individual case)

(a)自然事物 (Natural Objects) —牛，馬，樹木等。

(b)人爲事物 (Produced Objects) —如桌，椅，房屋，

公民，愛

（2） 度量的單位

（a）體量的單

1，單純的

等。

2，複合的——噸哩，尺磅，車里等。

（b）貨幣的單位（Pecuniary Units）一如圓，鎊，法郎，馬克等。

（c）抽象的單位（Abstract Units）一如物價指數，利率等。

自然事物在性質上具有一般的形式和特質，祗要按事物原有的個體自成單位。他的意義確定，而且完全，在統計調查上，不致發生重大的困難。例如美國農業部作家畜的淸查，在規定定義時，祗提出幼畜除外這一點。

人爲事物單位，在性質上雖是有實體、可感覺的，但無一定之形式，而其定義亦隨人意爲轉移。例如農田（Farm），在普通的定義，是：任何一塊地，專爲農業目的之用，或由田主或由佃農來經營的，都稱農田。而美國 1920 年聯邦淸查處所下定義是：「所有的地，由一人直接耕種的，都叫作一塊田，

不論其農事之經營，是由他一人的勞力，或是由他家屬或傭工來輔助的」。這二定義的差別，很值得注意，前者是表示一般普通的觀念，後者的規定，是要設法在特定情形實際觀察上，避免種種的困難。顯然的，人爲事物單位的定義，卽使是一簡單的事物，亦要考慮用字的確當，和謹愼地顧到實際情形。

社會事物，祗能和政治的經濟的社會的組織，相依而存在。這種單位的意義，更是複雜難定。例如「家庭」有着重經濟關係，凡產業未分或居處衣食相共者，認爲一家；有着重血統觀念，凡在五服之內，均可認爲一家；更有合併二說，在相當程限內，凡血統雖疏，而生活密接，或生活隔離而血統仍親者，亦可視爲一家。更如「犯罪」，稍微觸犯些法律，是否犯罪，還是要案情重大，當眞觸犯了國法，才是犯罪呢。在社會學上的意義，更是不同，若是案情重大，與社會相抵觸者，雖然在法律上不認爲犯罪行爲，也可以說是社會的犯罪。一個「家庭」和「犯罪」就發生了種種困難，要使調查員和被調查者不致誤會起見，怎能不先明白規定一個定義。

體量的單位，因爲各地度量衡制度不一致，在應用上就發生了困難。例如所謂噸，就有美噸，英噸，米突噸的區別。我國的斤，更是複雜。在複合單位更須注意。

貨幣的單位，在經濟統計中，用之很廣。不過這些單位的意義，更是抽象，雖則名義上，各國貨幣的價值，都制定在法

困難。

立場而分爲下列二類

（1） 測量用單位 (Unit of measurement)

料之時，作計數和估計之用，分單純與複合二種。單

，就是指代表種類差別的單位，一噸煤，一張匯票，一公民，一慘案，這都是單純的單位，代表種類的差別，而非程度的差別，所以容易分別。複合的單位，就是指那代表差別的單位，例如銀行匯票、工業慘案、尺磅等，都是這種單位，就是在單純的單位上，加一些限制而已。

（2） 比較用單位 專用以比較時間、地域、情形 (Condition) 的種種統計，這種單位常是比率的或係數的。例如人口生產率或死亡率，一方里之人口密度，流動資本比率等是。此等單位作比較時，尤須留意其係數分子分母之選擇。兩種不相侔的現象，切不可拿來互相比較，例如比較兩地人口之生產率，宜用：生產數$\frac{\text{有產兒能力年齡之已婚女子數}}{1000}$爲單位，切不可用：生產數$\frac{\text{人口數}}{1000}$，或生產數$\frac{\text{產兒年齡之女子數}}{1000}$，因爲兩地社會組織的成分不能相同，甲地的女子占全人口總數之比例，

不一定與乙地相同。即或相同，而兩地之生育年齡的婦女與全人口之比例，亦未必相同。生產年齡的婦女雖屬相同，而兩地之已婚女子，未必相同。上述不過是一個例子，讀者欲知其詳，請參閱 Secrist 之統計方法 89 頁。

（3） 陳示用單位 亦可分時間、地域及情形三種，人的年齡，調查時可以用月份計算，而在陳示時或許是用年。人的高度，可以測量到分，而陳示時可以用寸。地價可以元來計算，可以用每方尺來作比較，而陳示時，可以用方丈。總之，測量、或比較用之單位，不一定就是陳示用的單位。

單位應具之要素 單位應具之要素，其重要者約有五端，茲特簡述如下。

1. 普遍性 (Universality) 統計調查，不是僅僅限於一時或一地之用，所選單位，當以有普遍性者為上。例如我國各地習用之度量衡，尚未統一，在編製統計時，宜折合為一標準的權度。

2. 比較性 (Comparability) 統計作用，原是從比較中以闡發其蘊奧者，故所用單位，宜具有比較性。單獨的數字，無甚價值之可言。而且統計之現象，除計算單位之本身有比較性外，猶須應用於異地異時，作縱橫比較之可能。

3. 可數性 (Countability) 統計事實之現象，固屬比較，而比較之程度，尤須精密，決非大小，優劣，長短等含糊差度

，所能

量現象：

4.穩定性 (Stability) 統

那末這權度的本身，當然

在一次調查中，所應用

變換，以減低統計結果

5.明確 (Accuracy) 單位的

含混，而致誤會，使計算者和調查者發生不正確的報告。例如 Bowley 在研究勞工階級的家庭時，便將「勞工階級」這名詞，給他一個比較清楚些的定義。他把下列三種階級，都不算在勞工階級範圍內：(1)高等職業的人，或依財產收入爲生的人，(2)書記，旅行家，教員，商店經理，小工廠主，(3)商店店員。

第三章

搜集資料

搜集資料的方法，大概可說有兩種：原始調查 (Primary investigation) 和繼起調查 (Secondary Investigation)。原始調查，就是直接去搜集我們所要的資料，繼起調查是根據他人已搜集的成料而利用之爲己有，所以手續較簡，經費較省，不過沒有適用的資料可資利用，亦惟有直接去調查。

原始調查

原始調查有五種不同的方法：(1)親自調查，(2)派員調查，(3)通信調查，(4)登記法，(5)估計法。

親自調查　這種調查很適宜於作範圍較小的精密研究，同時調查者對於所攷察的問題，須有特殊的志趣。在另一方面，因爲標樣 (Sample) 的範圍究屬太少，有難以代表全體的缺憾。再調查者很容易有主觀直覺的見解，影響結論的危險。法教授Le Play 曾經用此法來研究歐洲工人家庭的預算，親自住在工人家裏，甲家住數月，轉至乙家，又轉至丙家，如是者數年，所得資料，才能準確。

派員調查　是任用專門的調查員，施以相當訓練和指導，

很可應用。美

調查事務。不

逅

抽查

應

之成 全恃

的裁制，被調查者

亦不易完全、正確。再調查之目的和調查主幹姓名或機關，必須註明調查表上，否則被調查者之狐疑，與玩忽之劣根性相合，更不願眞實相告，或竟不作覆。惟此種調查，經費節省，故無論公私機關作範圍較廣之調查，多採用之。因爲，有時就收回之調查表，擇其完整者作標樣以代表全體，亦可得相當之精密。定期調查，規定被調查者須按期答覆者，較偶爾一次之調查，所得結果，似乎靠得住些。

登記法　便是使被調查者，至調查機關報告一切調查事項；例如公司創設之註册，社團成立之立案，人口死亡、出生、婚姻、遷移等戶籍登記。此方法之採用，大都爲政府機關，含有強迫登記之性質，否則、不能見效。

估計法　有些事實，常常不能確實測量，祗能用估計法來救濟。其法將應行調查的事實，規定適當問題，特約專門人員估計報告，有時亦有遣派專員或代理人至各地徵集者。美國農業部許多農產收穫統計，我國統計局的農況調查，農產收穫統

計等，都是由估計得來的。各方面的報告，雖然不十分精確，但是他們的錯誤，是有互相補償的，故其結果亦有相當可靠。此法之優點，在乎簡易而經濟。不過須注意，所謂估計，并非憑空揑造，亦須有相當的根據，方有價值。

總之，調查方法務須適合該調查之性質，又須不違背調查地方的情形和習慣，故有時得酌用數法。美國勞工統計局採用的方法，至不一律。例如編工人生活費指數時，食品類之調查，用通信法，衣服則用派員調查。又如上海特別市社會局工資之調查，是兼用通信調查和派員調查的。時間、經費、和結果的精密程度，在決定調查方法時，亦應先加以考慮。

調查步驟

調查方法雖有種種，而調查步驟，則大致相同，可分(1)釐定目的，(2)擬訂表格及(3)表格之訂正三步。

釐定目的　統計方法既是歸納的方法，并不是盲尋事實。統計家着手調查之前，必先有一預定的目的，并且對於問題的各方面，有相當的認識，正確的觀念；那末，才有標準來定出適當的方法，搜集準確有用的資料，例如我們作工資調查，必先認清：還是用以作生產成本的研究呢？還是在研究工人生活的狀況？作第一種研究，則當用貨幣工資率，作第二種研究，則注重實際收入，而同時尤須顧及貨幣購買力之強弱。設預先

並不規定，調查之時，將何所適從？再如經濟學者作物價指數，必先規定還是用以測量物價的升降，貨幣購買力之強弱，測量進出口貨價，還是用以研究一種階級的生活費而作，然後才可以決定要搜集的物價，是批發的，零售的，某地方的，某時間的，某等級的，還是某種類的。而將來編製之方法，亦各不同。許多沒有經驗的統計者，常常得不着美滿的結果，多半因為他們在開始調查時，只知道搜集資料，而對於他們當前的問題，因為事前沒有正確的觀念，結果把他們調查的能力分散了，調查的範圍擴大了，而所得的都是些無目的、無作用的資料。不過在此時，統計家所持的態度，應用客觀的，切不可引起主觀的感情的見解。因為一有成見，只是想勉強得到自己所要的結論，就失了統計的價值。

擬訂表格　依預定之目的，實施調查時，即需調查表格。此項調查表格，和調查之進行，及結果之良窳，大有關係，因表格之不良，而結果失敗者，數見不鮮，惟表格之製訂，屬於技術問題，全恃統計家之經驗。至於一般之原則，可從表格之問題，及表格之形式，兩方面來敘述。

I.表格的問題　問題的措詞，總要使被調查者聽了，不但在消極方面不容易敷衍塞責，而且在積極方面，還要他樂於按實填記。關於私人人格及干涉私權之詞句，切宜避免，以免引起答者的厭惡和反感。如不得已必須問時，最好用他種旁證性的問

題來補救。譬如要調查靑年男子之結婚與否，你與其問「你曾經結婚否？」不如問「你的妻子現在那裏？」所得答語，比較可信。又如要調查女子的年齡，不如從他的兄弟姊妹的年齡來旁證，比較準確些。

問題的答案最好能以是否二字，或數字，或一定時間，一定空間，或以其他確切不移的字句來答復。其最要之點，就在使問題化抽象爲具體，化性質的分量。問題的答復，能使被調查者用符號✓來答復者，更爲迅速。問題的數量，以足夠需要而止。不可貪多，而語句要簡單；否則被調查者將望而生畏，因此不復者，數見不鮮。若被調查者在法律上有必須答復的義務者，問題不妨稍多。再從經費上說，經費是和問題數量成正比例，有時國家人口調查，設多一問題，便要多費數十萬元。

問題的排列，不可過於擁擠，當疏密得宜。重要的問題，可排列在前，以便被調查者，能夠首先答復，卽使其後的問題，不能答覆完備。如爲通信調查，此點更屬重要。

II.表格的形式　關於調查表的形式方面，當注意下列各點。

(1)　表的大小廣狹，以便於攜帶，易於收藏爲原則。普通美國標準調查表的大小，有3×5英寸，5×8 英寸，$8\frac{1}{2}\times11\frac{1}{2}$英寸三種。歸卷制度 (Filing System) 的發達，對於這種卡片的收藏很便利。

(2)　表的質料須堅硬，紙面須平滑，便於調查者持在手上用

種不同的顏色。在調査的時候，不致混淆，而於整理時，亦省目力而便於分類歸卷。

（4） 調查表上為醒目起見，常有使用粗細不同的線，來分別部分的重要，凡此種種，對於調查者的目力，和統計上的整理工作，都有極大的幫助。

表格之訂正　將收回之調查表格每張都審查一次，就是訂正，這是原始調查的最後一步手續。訂正的目的，在使表上的記載，達到最高之正確 (Accuracy), 一貫 (Consistency) ，劃一 (Uniformity), 完備 (Completeness) 。在原則上，調查表的答案，每個都有存在的權利。編者不能隨意改正，除非是發現有兩個或兩個以上的不一致的答案。編者不能按照普通的或然的原則，決定那一個答案較為正確時，可以作第二次之徵詢，或竟棄之。修改之處，亦宜以鉛筆註以符號，以備將來之參考。

調查表間有二個答案不一致，須改成一貫；例如一個人在表裏，已經說明他已經結婚，并且是一家之長，而他的年齡記為八歲，在這情形之下，我們可以假定大概漏了一個二字。

有時同一事項，用了幾個不同的名詞來記載。或是所用單位不同，為便於將來分類和列表起見，宜以藍色鉛筆修改之，

使之前後劃一。

一切消息，在可能範圍內，務求完備。有時寄回之調查表，留有空白未塡的問題，校對者可從這表上的其他答案中，找到相當之答語來補充之。設未答之問題，甚爲重要，惟有作第二次之徵詢。

繼起調查

繼起調查，就是搜集他人現成之資料，供吾人之研究。搜集時所應注意者，即原來搜集此資料之機關，究竟可靠否？主持其事者的能力如何？資料之刊布，爲暫時的？還是永久繼續的？這種現成的資料，或許是憑私人的臆度杜撰，也未可知，假若貿貿然拿來作科學探討的根據。豈不貽誤。此外，還須注意下列各點：

1. 來源的研究　統計資料的來源，有原始的，有次級的。原始資料是指刊布之資料，他的搜集、編製、和刊布都由一機關爲之。次級資料，是指該項材料，是重印原始的，或刊布機關不是原來搜集的機關。以普通的原則來說，搜集資料，總以愈接近原始的來源愈可靠。因爲事實上謄寫重印時的差誤，在所不免；而且次級來源上，往往有將原編者的註釋省去，刊布的資料或許祇有全部中的一部分。雖則有時次級的資料，更較原始資料，格外整齊些，吾人採用時，總宜格外注意。

作工人之調查，所謂工人，還是指工會的工人？還是不問入會與否，都包括在內？還是指工廠的工人？還是包括手工業工人？凡此等等，都應當先明瞭。

3.原來搜集的目的　一種資料之搜集，即有一特定之目的，非盲目從事。所以利用資料的人，應當追究原來搜集的本意何在？與吾人之宗旨相脗合否？

4.方法如何　搜集目的，固有不同，搜集方法，亦各不同。調查的時間和空間的範圍如何？應用全體調查？還是標樣調查？直接調查？還是由估計而得？諸如此類之問題，都當先了解，然後才能判斷他的方法，是否合理。

5.數碼精確之程度　統計數字，沒有絕對精確之可能，所以發布時，常將差誤註出，以供他人研究。吾人在採用時，當特別注意，看他精密之程度如何，是否有偏誤及揑造之處。如係標樣調查，標樣大小如何？代表性如何？都當詳加考察。

如上述諸問題，皆已圓滿解決，考察者才可決定此種資料，是否可用。有時對於同樣的問題，而二種原始資料之推算，因來源不同而異。遇此情形，當極力考察這不相同的數目之來源。如二種來源，都屬可靠，則應折衷之。否則，僅擇其可靠者，而捨去其餘之數字。

第四章

標樣調查 (Sampling Investigation)

全體調查和標樣調查　調查方法，除了從資料來源方面研究之外，更可以從調查的範圍來討論。我們知道有幾種調查非調查事實的全體，是不會知道他的眞相。有的只要調查那事實全體的一部分，就可以知道或估測他全體的眞相。譬如我們要調查上海的人口多少，那最好是將所有上海人口，每個都加以調查，然後總合起來，才得到上海人口的多少。凡是這種把調查範圍內全部分子，都加以調查，叫做全體調查。還有一種調查，就是我們在調查範圍內，只調查全體之一部分，然後根據這一部分資料來代表、或估計全部事實的眞相，這叫做標樣調查。譬如我們要調查上海居民之肺病狀況，如果用全體調查，就要把上海市凡患肺病之人統統調查，然後才可說上海患肺病者共有若干。設用標樣調查，則僅就一部分之居民內調查其患肺病之人數，占該部分居民之百分之若干，根據此點以推測上海居民中患肺病者之百分數。這種標樣調查，是否可靠，係另一問題，在下文中再當詳加研究。

普通統計資料，除國家之人口淸查或範圍狹小之調查外，極少用全體調查法。因爲全體調查，雖於理論上最稱完善，惟

可能。再者，有旷……

斷力，亦無需有全體……

特質之研究，勢不能將……

如研究上海人力車夫之生活狀況……力車夫，

盡行調查，取若干之標樣卽可。

標樣正確的要件　在標樣調查中，有一根本問題——怎樣使局部之事實，代表全體。——有兩點當嚴格顧及。第一，標樣之產生，應絕對根據機遇律（Random Principle）。這就是說，全體之每個個體被選爲標樣之可能，機會均等，而每個選擇皆當絕對獨立，決不可因選擇其一個體之故，而影響於他一個體被選之機會。第二，是性質要相同（Homogenity）不同類的資料，不足以爲標樣。我們知道有工會組織工人之工資和沒有組織工人之工資不同，鋼骨水泥之房屋，火災危害之機會和木材之建築物不同，欲調查工會工人之工資率，當然不能以未入工會之工人充標樣。假使要根據火災危害案件以定保險率，水泥和木材建築物，自不能混爲一談。上述二點要件外，取樣時當顧到標樣的大小，和標樣之可靠。

標樣大小與確度　標樣究竟要多大，尙無一定之法則。總之，被選擇之標樣，須將調查之全體資料之要素，盡行包括。否則，不足爲全體之代表。如某校學生共1000人，其中研究化

學者10人，假使現在要研究學生之志趣如何，選擇50人爲標樣，則研究化學之學生，未必有被選之機會。很明顯的，標樣之最小限度，也不得少於100人。從此可知標樣之最小限度，當爲能代表全體各部之最小數。標樣之適當數目，究應若干，隨所欲研究之材料，及統計家之經驗，與經濟之供給而決定。此外還可藉數學公式來幫助，假如 σ 爲標樣之標準差，N 爲標樣之項數，P.E.爲算術平均數之概差 (Probable error)，卽得下列公式以測量標樣之確度 (Reliability)。(註)

$$P.E._{M} = .6745\frac{\sigma}{\sqrt{N}}$$

從這公式，可知確度與標樣項數之平方根爲正比，是故希望二倍之確度，當四倍其項數，三倍之確度，當九倍其項數。總之，標樣項數愈小，差誤愈大，項數加多，差誤愈減，最後差誤極小，吾人卽可置諸不理。根據這結論，吾人用標樣法所得統計，究竟精密可用否，可以最簡之試驗方法以試驗之。其法卽將研究對象，任取標樣幾組，每組標樣之項數，約略相同，如計算所得結果皆極近似，足徵標樣已足夠可用；否則，增加各組標樣項數，直至各組所示之結果相近似而止。

取樣之方法　取樣之方法，普通有三種。

1,無限取樣法 (Method of Extensive Sampling)　採取標樣，愈多愈妙，漫無限制，不問取得的標樣之特性是否能代表全體

註　關於標樣確度之公式參考第二編第七章

乏眞正之代表性。因此，這種只顧形式，不問實質之取樣法，已大半被人擯棄而不用。

2,機遇取樣法 (Method of Random Sampling) 機遇取樣完全是根據數學上之機遇律。採取標樣，純憑機遇。在使全體中之各分子，於當選爲代表上，均有平等之機會。此法優點，在使選擇標樣者，無絲毫主觀之成見參雜其間。不過不能說無絲毫之危險。本方法之危險，在標樣中包括之個體不多，未能適如其量；再名義上雖爲機遇取樣，而實際仍不能免除主觀之感情作用，故當注意下列三點：

a. 標樣須從全體中選出，并非限於一部分現成或易於搜集之資料。

b. 每一個體被選爲標樣之機會，須完全相等。

c. 每一標樣皆係完全獨立，不受其他個體被選或落選爲標樣之影響。

3,代表取樣法 (Method of Representative Sampling) 在取樣之前，將全部統計資料，先作一詳細之觀察與切實之研究，然後規定一個確切的標準，去選那雖則僅爲全體之一部，却具有代表全體的眞實性之資料。此法在今日統計界中，用之最廣，不過另一方面，他的最大危險在所訂之標準過於主觀，調查者

可以有意的、或無意的採取他所希望之標樣，以便達到預期之結果。惟果能於事前作詳密普遍之觀察，力求客觀之訂定標準，搜集足量之標樣，其結果自當較前二法爲勝。

標樣調査的基礎定律　標樣調査的基礎定律有三，一爲統計常性定律 (Law of Statistical Regularity) 二爲大數惰性定律 (Law of Inertia of Large numbers) 三爲小數恆性定律 (Law of Permanence of Small Numbers) 。此三定律普通稱之統計的定律，實則不如稱之爲標樣的定律，較爲明切。因爲所研究之統計對象，何止千萬，統計家只選取一部分之現象，憑此所下斷語卽敢以爲全體千萬現象之結論者，惟恃此三定律。

1，統計常性定律　所謂統計常性，卽無論何種事物，在其全部內任選一局部觀察之，就平均言，此一部必具有全部之普通現象。例如吾人欲知一樹樹葉之長度，可任採樹葉若干羣，每羣數千張，每張測量之，然後求其平均長度，則各羣之平均長度，必極相近似，或完全相同。如將全樹之樹葉平均長度與之比較，亦必極相近似或完全相同。這就足以證明全部之一部，具有全部之普通現象之實例。 Prof. W. I. King 在其所著之統計方法中所下統計常性定律之定義，謂若從大羣項數中任便取出一些項數相當之小羣，求其平均數，便可代表全體項數，全體項數之特質，亦可由此平均數看出，卽此意也。惟須注意者，由大羣中所選出之項數，不能過小，此點已在前申述，因

本，昔有研究大數之學之稱，其定律自亦以適用於大數爲本。

2. 大數惰性定律　此定律實爲統計常性定律之一分系，所謂大數惰性，卽無論何種現象，就大量觀察之，非有特別情形爲之因，則不變。例如：米麥之產量，就一地看，每年或有所增減，就全國看或全世界看，則無增減。因爲此增彼減，此減彼增，減增之數，互消互償，所以無所增減。King 所謂"在一大羣數目中，一部分之變化向一方向，同時另一相同之部分向相反之方向變化，反正兩變化相消，因之全體之變化極小"其義相同。但此定律只能在平常狀況有效，因爲假使如我國政治未上軌道之各省，農民不勝租稅之繁重，圖種植雅片之利厚，爭將種米麥之田畝，改種雅片，由是種米麥之田地日狹，設長時間的繼續下去，則全國米麥之產量必大減。有此種特別情形，大數惰性律不復再能保持。

3. 小數恆性定律　亦爲常性定律分系之一，卽無論何種事物，其罕見之現象，或具有非常之性質者，常爲少數。Bowley 說"多數事物中，其表現特別形狀者爲數極少，就我們普通經驗而言，此少數變大時至罕，完全消滅時亦至罕，如天災偶然之事，稀奇之物等，載於報章者，皆足證實吾人之經驗。從事於專門職業之耳科醫生，以至於古董商，都恃此少數恆性以爲業，"卽此意也。

第五章

分類及統計數列

分類的意義　甫經搜集之資料，必須加以整理，分類就是整理的第一個階段。所謂分類，在統計上可以說是將一大羣之個體，依特定標準或共同特質分為較小之若干組。他的功用，在乎決定整理調查資料之界限。譬如關於工人工資之調查，設統計資料已經搜集，試問還是從工人所從事之工業為標準來分類而分別整理，還是以工人之職務為標準，還是以工人之工作、技能、年齡、性別、工資收入為標準，還是以籠統去整理？這些問題，都要在製表開始前精密討論。

分類的方法　自然也應得從許多不同的立場來看，通常總括來說可有三種：

1.根據事物的屬性 (Attributes of thing),

2.根據地域 (Geographic Location),

3.根據時間 (Time of Occurrence),

例如：廣東省工會統計以職業為標準而分類，就屬於第一類的分類(表1)，以縣治為標準，就屬於第二類的分類(表2)。如以立案時期之逐年年次為標準而分類，即屬於第三類(表3)。

[illegible]	[illegible]
廣州	38
惠陽	11
潮安	28
南海	35
清遠	11
四會	4
茂名	16
河源	5
汕頭	19
關坡	4
陽江	4
三水	7
台山	12
靈山	4
興甯	11
五華	1
海口	1
梅縣	1
普甯	1
陸豐	1
新興	14
蕉嶺	5
英德	1
總數	234

職業性質別	工會數
運輸業	41
生活供應	29
雜材製造	26
貨品販賣	20
金屬製造	14
木材製造	14
飲食品業	11
紙製品業	11
皮革竹漆	11
建築工程	10
紡織工業	10
土石製造	9
化學工業	5
藝術	5
家事使用	3
通信業	2
原動力業	2
漁業	1
未詳	10
總數	234

表3 歷年廣東工會立案統計

立案時期	工會數
民國9年	4
民國10年	14
民國11年	8
民國12年	3
民國13年	3
民國14年	22
民國15年	74
民國16年	50
民國17年	31
民國18年	4
未詳	21
總數	234

從統計之觀點，第一種分類，更可根據種類 (Kind) 的差別及程度 (Degree) 的差別，而分為二種。前者之例，如人口之以種族分或性別分等屬之；後者之例如人口之以年齡分為老、壯、幼等。再者程度之差別更可細分為品量的 (Qualitative) 及數量的 (Quantitative) 二種。上述年齡分類之分為老、壯、幼者，屬於品量的，如以之分為「16歲以下」，「16歲至44歲」，「45歲以上」，三組則屬於數量的。數量的分類，更可分為等級的和不等級的二類。不等級的分類，和品量分類差不多，如（表 4）；而等級之一類，就是標準的次數分配 (Frequency Distribution) 完全成為統計數列之形式，如（表 5）。

表4　杭州市各種工業資本分配表(民國18年)

資本等級 (元)	廠數
0 — 500	36
500 — 1,000	44
1,000 — 2,000	41
2,000 — 3,000	29
3,000 — 4,000	30
4,000 — 5,000	6
5,000 — 6,000	18
6,000 — 7,000	7
7,000 — 8,000	1
8,000 — 9,000	3
9,000 — 10,000	1
10,000 — 20,000	13
20,000 — 30,000	5
30,000 — 40,000	3
40,000 — 50,000	1
50,000 以上	8
總數	246

50 — 69,9	11
70 — 89,9	14
90 — 109,9	26
110 — 129,9	15
180 — 149,9	13
150 — 169,9	8
170 — 189,9	2
190 — 209,9	5
210 — 229,9	2
230 — 249,9	3
250 — 269,9	1
總 數	100

從另一方面，事物屬性之分類(Attributive Classification)更可從形式之簡單和複雜，而分爲二種。將一大羣中之個體，依單一之性質分類，是爲單純的，設在同時依據數種特質而分類之，是爲複雜的。單純之分類，上舉各例均屬之；複雜之分類，試舉例於表6，表中每一個體，同時根據兩種不同之標準——年齡和高度——來分類。

分類之方法，既已略述如上，至於究竟應如何分類，本無一定之法則，可資遵循，全視資料之性質及其研究之目的而定。惟吾人尚須注意者，因分類和併組之不同，足以影響分析結果之差異。

表6 某大學學生年齡及高度相關表

年齡	高度（英寸）							
	62-63	64-65	66-67	68-69	70-71	72-73	74-75	合計
15歲	1							1
16		2	1		1	2		6
17	2		3	6	5	2	1	19
18	1	1	5	11	8	2	2	30
19			2	4	4	3	1	14
20					1	1		2
21					1			2
總數	4	4	11	21	20	10	4	74

統計數列 (Statistical Series) 將原始資料列成統計數列，亦爲統計重要程序之一。統計數列乃依次排列之若干項數字；一變量之差異，因其他變量而差異。換言之，統計數列，所以示一變量爲他一變量之函數。變量之用以決定他一變量數值之標準者，稱爲自變量（Independent Variable）；而他一變量則稱爲倚變量 (Dependent Variable)。是故，自變量在分析時，乃主要之要素。組成數列時，其數值之決定，多少有些強定的，而倚變量之數值，則由此而決定。自變量和倚變量聯合在一起，則數列之性質，始可顯然。

A Statistical series is an orderly succession of items relating differences in one variable to differences in anot her. Technically, we may say that a statistical series expresses one variable as the function of another. Day: Statistical Analysis, P. 42.

2,時間數列 (Time Series)

3,地域數列 (Geographical or Spatial Series)

第一類數列更可分爲單純與複合二類。單純屬性數列，普通稱之爲次數數列 (Frequency series) (表7)，複合屬性數列，普通稱之爲相關數列(Correlative series)，表8卽其一例，此種資料，表示兩變量之相關，設年齡增加，則收入亦增加。既無說明[illegible]數之意，又無表明時間地域差別之意。

次數數列

表7 美國Cincinnati城報販700人之年齡分配(1918)

年齡	人數
8	8
9	14
10	76
11	150
12	154
13	140
14	84
15	48
16	12
17	14

相關數列

表8 美國Cincinnati城報販700人之每週平均收入(1918)

年齡	每週平均收入
8	.31
9	.57
10	.68
11	.87
12	.95
13	1.11
14	1.62
15	2.14
16	3.10
17	3.68

時間數列

表9　1914-1929年全世界歷年銀產額

年份	銀產量 單位：百萬純盎斯
1914	172.3
1915	173.0
1916	180.8
1917	186.1
1918	203.2
1919	179.8
1920	173.3
1921	171.3
1922	209.8
1923	246.0
1924	239.5
1925	245.2
1926	253.8
1927	254.0
1928	257.3
1929	256.5

地域數列

表10　1929年各市區之人口

市名	戶數
上海	339.782
天津	279.910
北平	267.296
廣州	175.699
漢口	163.557
南京	100.116
青島	66.977

從統計數列性質的立場可分連續數列 (Continuous Series) 及分立數列 (Discrete Series) 兩種。所謂連續數列者，數列之自變量，爲連續變量。分立數列，其自變量爲不連續變量。

平常，時間數列和地域數列之自變量，總是連續的；而倚變量或爲連續，或爲不連續，並不一定。至於屬性數列的自變量和倚變量，有時爲連續，有時爲不連續。

分類和統計數列之關係　分類和統計數列旣已略加討論，玆再述其二者之關係。有時分類之結果，就可以直接形成統計數列，假如分類是根據一變量之等級，例若將商船航行速率，分爲「9 哩以下」，「9—14½哩」，「15—½哩」，「18哩

冰車(c)油車(d)客車[illegible]

不表明二變量之數量關係。

同時，數列不一定皆需要分類，僅將每次之觀察數值，依其先後次序排列之，並不分類，亦可成爲數列。例如作某種物價之研究，祇要將每月特定之一日之物價，依時序而排列，就成一時間數列。有時地域數列，亦可不需經過分類之階段而得之。雖然，通常統計數列之形成，總先經分類，是故分類在統計分析之程序中，仍不失其重要之位置。

第六章

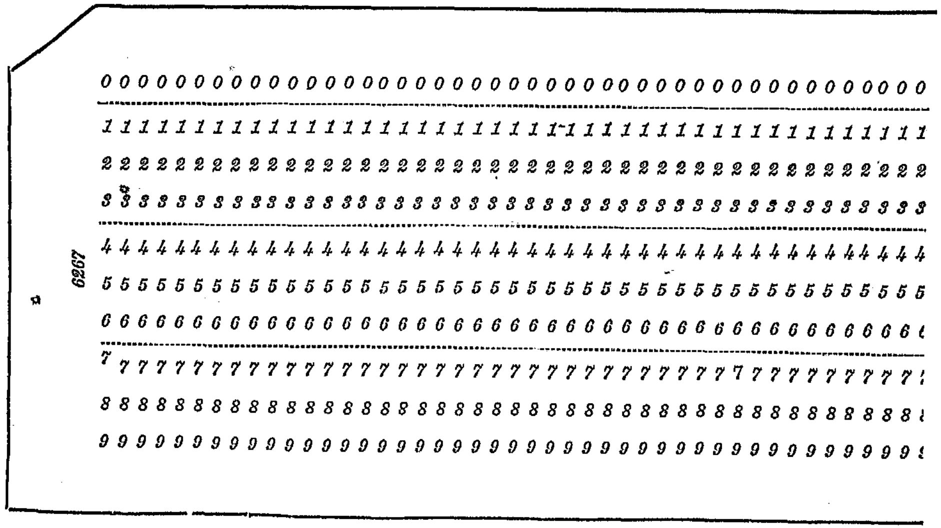
表列 Tabulation

表列之意義　表列就是將調查所得資料，經過分類，或組成數列後，排列成表，使成系統，以便研究。凡資料編成爲表，就可由兩方面 (Dimension) 或數方面讀之。平常未經表列之資料，例如說：上海市十七年罷工統計，一月分案件共九起，參加人數共 20185人，二月分罷工案件共九起，參加人數7865人，三月份罷工案件共十起，參加人數共 67278人………。這樣的敍述，因爲沒有編成表，就沒有列表那樣明顯易讀。

表的功用　資料在未分類和列表前，簡直毫無系統，其全體彼此間的關係，更是茫然。一經分類列表，或以數量之大小爲標準，或以時間之先後爲標準，或以地域的觀念爲標準，然後才有系統之排列，易於作科學之研究，發現其規律之狀態，或全體彼此相互的關係。并且因爲排列有次序，有時可藉聯念的作用，使人有深刻之印象，容易記憶。至於因列表之故，而使資料便於總核，和減少重複之說明等，尙其餘事。

表列的預備手續　在正式分類製表之前，要先經過預備的手續，有使用機械和手工整理二種。如若調查表的數量，比較的少，而且統計事實不多，可以用手工整理。否則，如有大

圖1　編號卡片



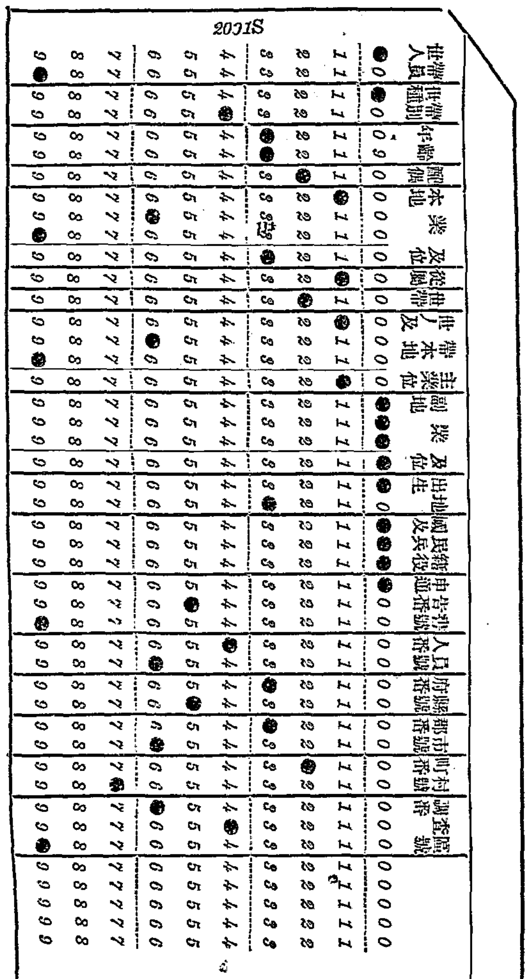


圖2　已穿孔之編號卡片

(1)機械整理　今日大規模之調查，多使用機械來整理。其法先將調查表上之各項資料，代以符號，編成一種符號册，這樣就可以免了完全過錄 (Transcribe) 和[illegible]述之麻煩。編號旣竟，乃用一種特製編號卡片，上面滿印着1,2,3,4,………等數字，將已編號之各種消息排列於上。排列完畢，然後按各種情形的號碼在穿孔機 (Key Punch) 上穿孔。圖 1 圖 2 爲編號卡片之形式，圖 2 爲已穿孔之卡片，每一小孔，即代表一消息。再將已穿孔之卡片，置入歸類機 (Sorting Machine) ，通以電流，將每個同類的現象，各歸一類。歸類機每分鐘可歸檢卡片二百五十張。要證明歸類是否正確，可將穿孔之卡片，向光處一照，看其是否通過各小孔即可。此時將打孔歸類之卡片，置諸電力集計機(The Electric Total Printing and Listing tabulating 上，machine)藉電流之力，每分鐘能傳遞卡片 150張，將各消息之總數，呈現於計數盤上，即可編成原始表格矣。

上述歸機及集計機，在美國最通行者有 Power 及 Hollerith 兩種，日本有川口式。

(2)手工整理　通常就將調查單上的各項資料，依照預定的方式，謄錄在卡片上，或謄錄稿 (Tally Sheet) 上，卡片之大小，在美國平常有 3 × 5 吋和 5 × 8 吋二種。再大些的卡片，有時

亦有應用，但總不及小卡片之歸類正確而迅速。卡片劃成各種界線，以便將調查表上之各種消息謄錄其上。資料既經謄錄，可按其性質，分別歸類，其步驟先將表上之主題歸類，次將子題歸類，再一一計算各類之總數，而成爲原始表格。有時不用卡片，而用謄錄稿。謄錄稿可依製表人之便利，而列成各種形式。茲以表11爲例，聊備一格。原始表既成，可按研究之目的，而製成各種適用之正式表格。

表11　120工會之會員失業百分率之次數分配

失業百分率		報告工會數
0.0 — 0.9	卌	5
1.0 — 1.9	卌 卌 卌 卌 卌 卌 卌 /	36
2.0 — 2.9	卌 卌 卌 卌 卌 卌 卌 //	37
3.0 — 3.9	卌 卌 卌	17
4.0 — 4.9	卌 卌	10
5.0 — 5.9	卌 //	7
6.0 — 6.9	卌	5
7.0 — 7.9	/	1
8.9 — 8.9	卌	1
9.0 — 9.9	/	1
總　數		120

表的種類　通常可分總表、概表二種。

工總表 General table 是將固有之事實，應有盡有地詳加記載，保存其實在情形，以備作詳細研究時之資料，此種表格之編

關之部分，製成簡賅明瞭之記載，以適應研究之[illegible]分析，此種表格之編製，隨作者研究之目的而異。概表可依不同之立場而區分為：

A, 按概表之用途可分為刊布（Presentation or Publication）用表和計算表（Working table）二種。刊布用表為刊布統計之用。一方面固求簡賅明瞭，他方面尤須能顯明關係，和易於比較。計算表之排列，當依運算步驟之便利，而且要簡單。這樣，才能收計算敏捷及正確之效。

B, 按概表標目分類的多少為根據，可分為單項式、二項式、三項式、四項式等。

1.單項式（First order）的縱行和橫列，都只有一種事實。例如表12，在縱行所表之事實，係學生人數，只有一項特質，橫列所表之學院，亦僅一觀點。

表12　某大學註册學生人數，　某年度，

學　院	學生人數
總　數	························
文學院	························
理學院	························
法學院	························
教育學院	························
商學院	························

2.二項式 (Second order)　表之縱行或橫列上所記載之事實，有二種平列的特質。例如表13在學生人數之標目下，更分爲正式生特別生二部。

表13　某大學學生人數學籍統計，　某年度，

學院	學生人數		
	總數	正式生	特別生
總數	....	....	....
文學院	....	....	....
理學院	....	....	....
法學院	....	....	....
教育學院	....	....	....
商學院	....	....	....

3.三項式 (Third Order)　就是在表之縱行或橫列上，包括有三項平列之事實，如表14。

表14　某大學各院學生人數學籍性別統計

學院	學生人數					
	總數		正式生		特別生	
	男	女	男	女	男	女
總數	......	......	......	......	......	......
文學理	......	......	......	......	......	......
理學院	......	......	......	......	......	......
法學院	......	......	......	......	......	......
教育學院	......	......	......	......	......	......
商學院	......	......	......	......	......	......

事實；或在縱行上臚列一事實，橫列上臚列三事實；表15卽其一例。

表15　某大學歷年各院學生人數學籍及性別統計

年度	學院	學生人數					
		總數		正式生		特別生	
		男	女	男	女	男	女
總數	總數	……	……	……	……	……	……
	文學院	……	……	……	……	……	……
	理學院	……	……	……	……	……	……
	法學院	……	……	……	……	……	……
	敎育學院	……	……	……	……	……	……
	商學院	……	……	……	……	……	……
廿一年度	總數	……	……	……	……	……	……
	文學院	……	……	……	……	……	……
	理學院	……	……	……	……	……	……
	法學院	……	……	……	……	……	……
	敎育學院	……	……	……	……	……	……
	商學院	……	……	……	……	……	……
廿二年度	數總	……	……	……	……	……	……
	文學院	……	……	……	……	……	……
	理學院	……	……	……	……	……	……
	法學院	……	……	……	……	……	……
	敎育學院	……	……	……	……	……	……
	商學院	……	……	……	……	……	……

C，概表之實質上看，可分爲二類。

1.總數表　祇解答某項問題之總數，如罷工工人爲若

干，人口若干等，前述各表均屬之。

2.次數表　除解答問題之總數外，更將構成此總數之各分子，依其特質發現之次數多少，臚列之為若干組，如表13。關於次數之編製，當另節討論之。

表16　民國17年清華大學校工工資分配表

每月工資	人數
6.0—9.99元	117
10.0—13.99	15
14.0—17.99	5
18.0—21.99	1
22.0—25.99	1
26.0以上	2
總數	141

D.按槪表讀法方面說，更可分為由二面讀Two Dimension由三面讀 Three Dimension 二種，前者如上述之普通統計表；後者如表6 之相關表。由橫面讀，我們可明瞭15—22歲之人數分配；由縱面讀，可以明瞭62—75英寸高之人數分配：更由斜面讀，可以明瞭年齡和體重是成正比，而且顯出二數列之相互關係。

次數表之編製　吾人搜集各種資料之後，假如不用分類，將各種事實彙集之，則對於此種事實，一定看不清，不能用於分析。次數分配就是一種已分類之事實，此點已於前章略述。

不過怎樣分類，換言之，怎樣製成次數分配，有詳細研究之必要。次數分配之製成，在簡單之一種，亦頗不難。法將各變量依其數量之大小，順序排列之，這種就叫順列 (Array)。再將其最大量數，減去最小量數，所得之數，就是全距 (Range)。求得全距之後，可審度情形，將全體量數分爲相等距離的若干組。組的界限，就是組限 (Class Limits)。兩組限之距離，就名組距 (Class internal)。再計每組所有之次數，將各組排列之，即成一次數分配表。但在分組時，有幾點應加注意，例如(I)組距大小 (Size of Class)，(II)組限，(III)怎樣表示組限(Designating of Limits)，試依次討論之。

(I)組之大小　關於組之大小，第一要注意一律，不能或大或小。因爲組距相等的分配，可以使二變量間之關係，很明白地表現出來。有時次數分配之兩極端用「以上」或「以下」等，這明明是破壞了一律的規律。這類「以上」或「以下」之組距，從列表方面說，是便利的；可是在統計分析上，不足爲訓。

若尚從計算便利上說，自然組距愈大，計算愈簡，若兼從精確程度設想，則組距愈大，精確之度亦愈減。據 Yule 之意見，組距不可過大，亦不可過小，總以使次數分配的組數，在15 組至 25組之間。因爲多則計算麻煩，少則易失精密。

再則，組距之決定，當視分組後所得之次數分配，是否有

規則的現象。這就是說，所編成之次數分配，在近衆數的組，項數最多，而離衆數漸遠之各組，次數亦漸減，至兩極端之次數爲最少，在整個之分配中，無斷裂分立之形態。

(丌)組限　次數分配之採用分組辦法，應有根本之假定，即假定各組量數，均勻分配，而各組之中數值，可作諸變量之代表。如是，變量大於中數值之差誤，和變量小於中數值之差誤，可以互相補償。有時，材料之各變量，有集中全域上幾點之趨向，最好將組限決定後，各組之中數值，適爲該集中之各點。

組限之起訖，和中數值，最好適爲整數，而且是十數系統(Ten system)。對於將來分析時，計算便利及正確。(不過現代統計方法中，簡捷法發明很多，此點已不復如從前之重要)。

(ㄗ)怎樣表示組限　組限既已決定，應說明組限應如何表示。普通之方法，有下列數種。

(a)		(b)		(c)	
0.0—1.0	1	0.0至少於1.0	1	0.0—0.99	1
1.0—2.0	3	1.0 " " " 2.0	3	1.0—1.99	3
2.0—3.0	5	2.0 " " " 3.0	5	2.0—2.99	5
3.0—4.0	3	3.0 " " " 4.0	3	3.0—3.99	3
4.0—5.0	1	4.0 " " " 5.0	1	4.0—4.99	1

在上列幾種方法中，(a)法最令人不滿意，因爲太含混。

例如一量數1.0，就可以任意歸入第一組或第二組。(b)法雖則避免含混之危險，惟書寫時究屬太累事。(c)法才爲明確簡便之方法，普通採用最多。除上述三法外，更有用每組之中數值或下限(Lower limits)來表示。

(d)		(e)	
0.5	1	0.0—	1
1.5	3	1.0—	3
2.5	5	2.0—	5
4.5	3	3.0—	3
4.5	1	4.0—	1

累積次數表 (Cummulative Frequency Table)　上述之次數表，乃簡單之一種。假使將簡單次數表中各連續次數相加，即成累積次數表。其法，先將第一組之次數與第二組之次數相加，即作累積次數之第二組；再將此數加以第三組簡單次數，爲累稱次數之第三組；如是繼續相加，直至最末一組爲止，即得一累積次數表。累積次數表可分二類：(1)從小數量累積起，稱以下累積 (Less than cumulation), (2)從大數量累積起，稱以上累積 (More than Cumulation)，玆示例於下表。

表列的方法　表爲事實之縮影，要使表有價值，應當使表之結構很完善。要達到這目的，則關於表之內容、標題、標目、總數欄、數字、行格界線等各方面都應注意，試逐一討論之。

表 17　　120家工會會員失業百分率之次數分配

july 1, 1920

簡單分配		累積		分配	
		(以下)		(以上)	
失業百分率	報告工會數	失業百分率	報告工會數	失業百分率	報告工會數
0.0—0.9	5	0.0—0.9	5	9.0—9.9	1
1.0—1.9	36	0.0—1.9	41	8.0—9.9	2
2.0—2.9	37	0.0—2.9	78	7.0—9.9	3
3.0—3.9	17	0.0—3.9	95	6.0—9.9	8
4.0—4.9	10	0.0—4.9	105	5.0—9.9	15
5.0—5.9	7	0.0—5.9	112	4.0—9.9	25
6.0—6.9	5	0.0—6.9	117	3.0—9.9	42
7.0—7.9	1	0.0—7.9	118	2.0—9.9	79
8.0—8.9	1	0.0—8.9	119	1.0—9.9	115
9.0—9.9	1	0.0—9.9	120	0.0—9.9	120

1.表之內容　每表應各自為單位，以顯明特質之一點。性質不同之二事實，切不可編入一表，以減低表之價值。例如罷工統計和生活費統計，合在一表，未免不倫不類。在另一方面，所表列之事實，在可能範圍內總以列入一表為佳，這樣有一覽無餘之利。但事實過繁，亦可按其事實性質自然之界限，分列數表，附於一總表之下。

2.標題 Title　標題通常置於表之上端，其字數須簡賅易讀，含義須明晰正確，無雙關之弊。總之使讀者能一覽標題，即可了然，毋事參閱註釋之煩。不得已如無適當之標題，足以概括表之全體，註譯亦不可缺。有時資料之範圍，有關於時間性和空間性，則標題上須將時間地域，分別敍明為要。

3. 標目　縱行之標目謂之縱標目 (Capition)，橫欄之標目，謂之橫標目Stub。標目之排列，雖然無一定之法則，但須依事實性質之輕重，而合乎邏輯者爲上。標目之行數，亦很重要。平常標目愈細分，事實愈精密詳細，在另一方面，其重要意義之所在，反而蒙蔽。所以在通常情形之下與其多用平列之標目，無如在平列之標目下，分列數細目，較爲明瞭。

4. 總數　總數地位，從前多放在極右端或下端。自美國華盛頓統計局首先將總數放在各項之前，統計界遂逐漸仿傚推行。究以何者爲優，則各有所長。第一法，便於計算，和適合讀者心理。第二法，因總數緊接標目，容易醒目。要之，隨作者之目的，而爲取捨。

5. 表之數字　或用絕對數，或用百分數，或兩者兼用，視情形而定。設吾人之志趣，在知各國人口數，當取絕對數；欲比較二地人口變動之狀況，當取百分數，絕對數便毫無意義。設表之編製爲參考之目的，則絕對數百分數都要。絕對數、百分數、平均數當置於接近之地位，以便對照。

數字之排列宜整齊，無數字之格，宜以短線或虛線塡補之，以增美觀。數字之有特殊意義者，可用特種之字體，或在其下畫一直線以顯明之。有時數字之位數太長，而吾人所需要者僅爲大數，可略去尾數若干位，遇此情形，宜於標目下註「以千爲單位」「以百萬爲單位」等。

6, 行格界線　表中每一縱行，應用直線劃分之，并以單雙或粗細線；以區別各標目性質之重要與否。重要者用粗線或雙線，普通者可用單線或細線。表之上下兩端，宜用雙線或粗線，左右兩邊可任其開着，不必劃線。關於行格空間，應以適稱爲原則，重要之欄格，如總數欄，應當稍寬，以引起讀者之注意，他如格中之數位多者，亦宜稍寬，免得擁擠。

第七章

圖 解 Charting

表列法雖則可以使數字釐然有條，但是不便於記憶或比較，而且拿表格來說明事實，又嫌艱深，非有統計智識者不能完全了解。一般人對於數字，總覺枯燥無味，不易引起興趣，所以須有圖來救濟。因為事實一經圖解，他的全部情形，特質所在，和他的前後關係，均能表現於圖形之上，普通讀者亦得了然而得其意義。

圖解之用途 圖解法在統計中有兩種不同的應用。第一種是拿圖來說明統計的結果，第二種是拿圖來作統計分析工作之有效幫助，這點尤其是在分析時間數列時，更顯見其效用。本章討論祇限於第一種圖解。

選擇圖解的標準 選擇圖解可根據兩種標準。第一要看所表現的材料之性質。有時一種統計資料，雖則可用幾種圖解來表現，其中總有一種比較地最稱職，最適用，所以選擇時應先明瞭材料之性質。第二要注意各種圖所表顯的目的。平常一種圖，祇能適用於某種特定目的，或顯出材料的某種特質，或着重在材料的某種關係，决沒有一種圖是萬能的，可以在各方面都適當。

圖解的類別 圖的種類，當然可以有不同的立場來分類，美國統計家 Secrist 在他的大作「統計方法」中將統計圖概括爲形圖 (Pictogram), 地圖 (Cartogram), 曲綫圖 (Graph) 三大類。著者深表同情，本書卽採用其分類法，將三大類之圖解，依次分述之。

形圖

形圖之中，種類很多，勢難列舉，茲就其通用而簡單者略舉一二種，由簡而繁，舉一反三，是在讀者之悟其理而神其用。

1. 直條圖 (Bar Diagram)

A.單式直條圖 這種圖是將應表顯的事實，按數量大小的比例作成長條（或長線）依次排列之；排列的次序，或作縱列式（如圖 3），或作橫列式（如圖 4）；各條起點，應從零點起，而且應同列在一基線上 (Base line)。這樣，才便於比較。量表可放在左端或上端。事實較複雜的圖，條數較多，爲醒目起見，在各條間隔空處可以劃指導線 (Guide line)。

圖3　民國十七年上海特別市各業罷工案件

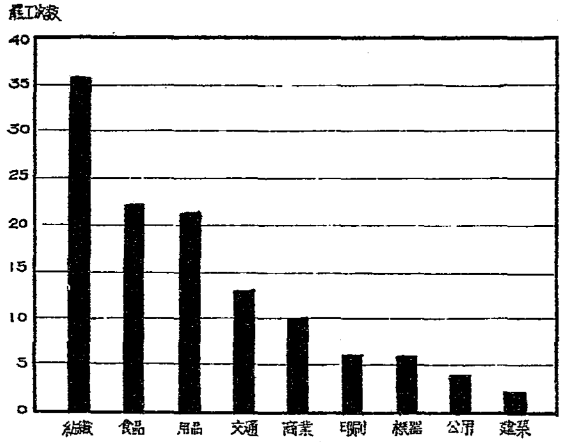


圖4　民國四年至十四年國有鐵道淨收入變動情形比較圖

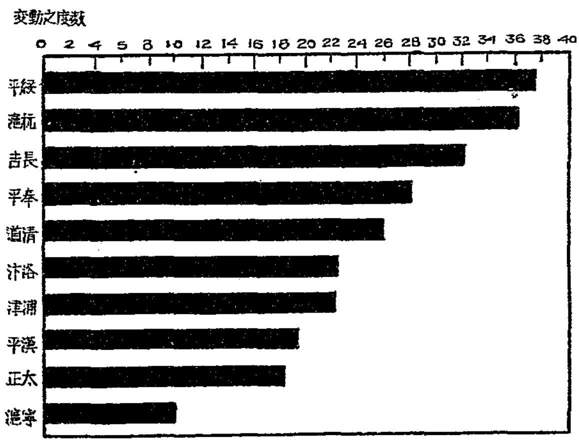
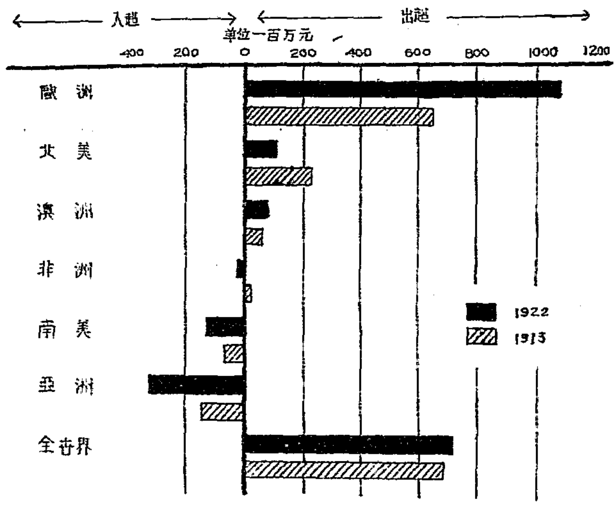


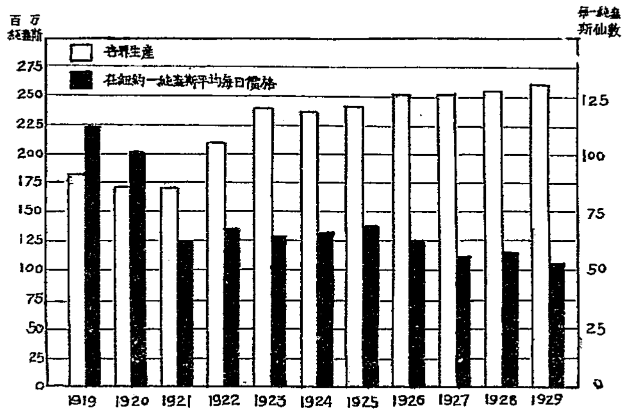
圖5 1913年及1922年美國對外貿易差額之比較



倘事實現象之性質相反，那可以將各條或各線分為上下（或左右）兩截而排列之（如圖5）。條圖為單方面（One Dimension）之圖，無關狹粗細之誤會，所以簡單明確，用以表顯簡單事實，最為相宜。

B.複式直條圖 以二條或二條以上之直條，並立為一組，代表某事物之一大項目；以每組中之一直條，代表大項目中之一小項目；以直條之長短，比較各小項目數值之大小。圖6即其一例，圖中黑條代表紐約銀價，斜線條代表生產量，二條並立，表示各年之生產量及銀價。

圖6　世界銀價及生產量，*1919—1930*.



3. 分段直條圖 (Component bar diagram)　有時我們要同時顯示事實之全部，和他構成的各分子，可用分段直條圖。以一直條代表一事實，再將此直條按構成事實的各部數值，分為數段，每段代表一分子。為醒眉目起見，各部可加以陰影或顏色，以示區別。這種圖在事實較多而複雜時，就不一定有美滿的結果。圖7為簡單分段直條圖，圖8為複合分段直條圖。

圖7 上海市勞資糾紛結果之分析，*1928*.

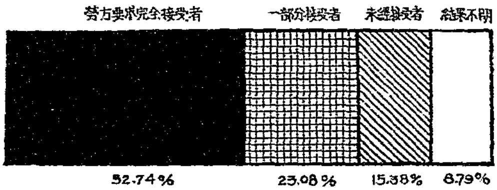
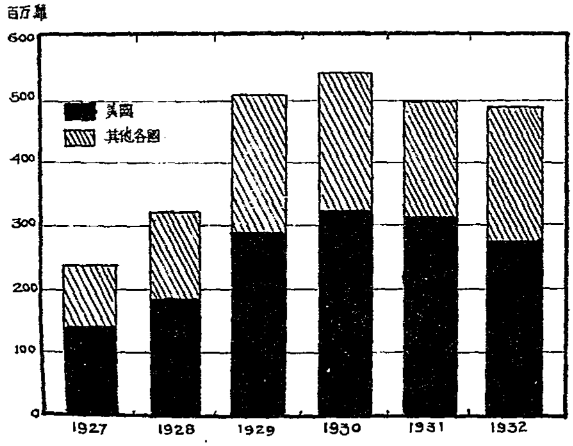


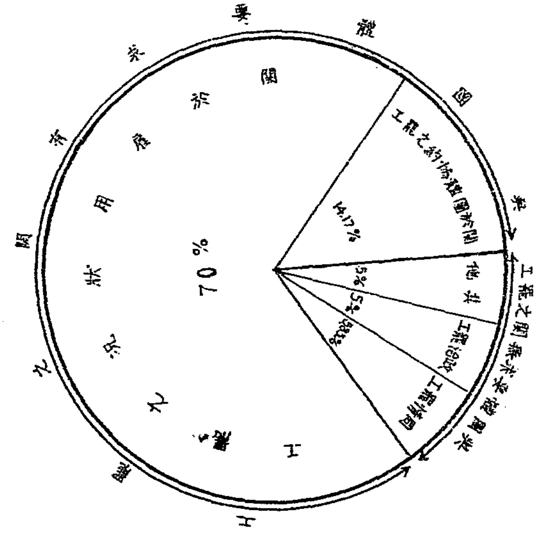
圖8 世界及美國小麥之供給，*1927—1932*.



2. 圓形圖Circle graph

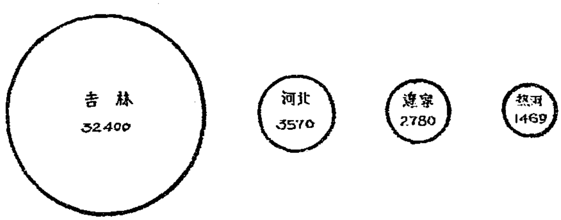
a 簡單圓形圖　係以全圓作百分，代表一事實的全體。再按各分子數值的比例，分全圓爲若干扇形，加以陰影或顏色，以示區別。這種圖的目的，在表明各局部和全體的關係，和分段長條圖一樣。他的長處，在簡單易畫，普通人亦很易了解。不過各扇形大小的差度，雖可辨別，只是究竟差多少，不能很精密明白地表示。所以作這種圖時，要將百分數或實數註明，不可或缺。例見圖9

圖9　上海市工人罷工原因之分析，1928年



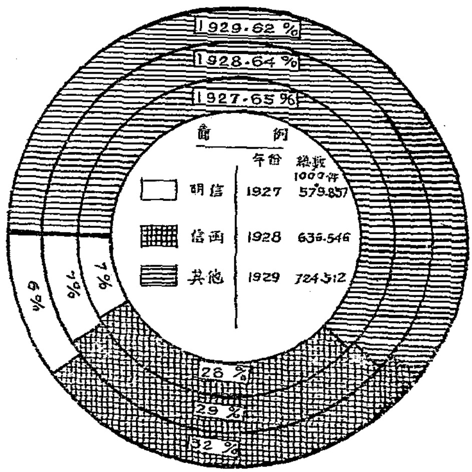
b 多圓圖 是以若干圓形代表若干事實，圖的大小比例，或以直徑爲標準。或以面積爲標準，兩者都不能認爲滿意。因爲以直徑爲標準，常使閱者失之看大，以面積爲標準，失之看小。所以這種圖，採用得很少，假使採用，亦應附以數字之說明，如圖10。

圖10 我國各省銀礦面積圖 1930.
(單位——畝)



c 同心多圓圖 (Concentric circle diagram) 是從一圓心上，作成多圓，每圓代表一事實。這種圖，不易使閱者明瞭，平常人很容易誤會：最外面的圓圖，是代表各小圓的總和。圖11

圖11 1927—1939三年來我國郵件比較圖



Ⅲ方形圖 (Square Diagram) 最普通的是毗連方形圖 Adjent Square 和重疊方形圖 Superimposed Square diagram 二種。這些都是從二面看的(Two Dimension,)所以非藉重數字的說明，不能示人以正確的觀念。應用範圍，亦很有限。圖12,及圖13

12　五特別市之人口統計，1930年1月

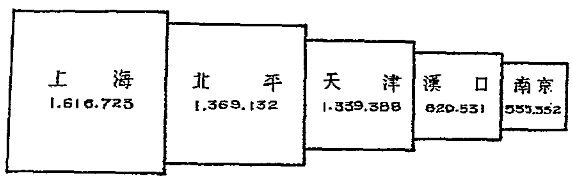
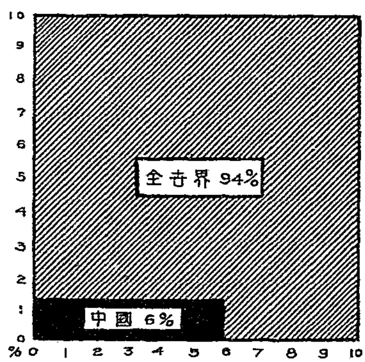


圖13　我國錫產量與全世界之比較，1929.



Ⅳ象形圖(Figure diagram)　有時爲引起閱者興趣起見，所作

之圖，係象事實的本體來作比較，例如畫軍艦以比較各國海軍，畫人體以示各國人口之數量等。在展覽會中，這類圖很多看到，不過很難能示人以正確的差度，和明白的實數，亦一大缺陷。

V體積圖及面積圖 有時繪各種面積圖或體積圖，以解答各種事實，那就要注意：面積圖是以平方計算，所代表的事實，應和平方的比例相合，體積圖是以體積計算，所代表的事實應和立方的比例相合。這種圖的大小，常使人有錯覺，所以不常採用。

地圖 Cartogram

地圖的優點，在使事實與他發生的地域，能同顯於一紙上，令人因此而得聯念作用。用以表顯地域數列，很爲適宜。統計地圖，亦有多種，應用時可按情地之適當而爲採擇之標準。

1.顏色地圖 Colored map 和陰影地圖 Shaded map 是以不同的顏色，或陰影的深淺，來代表數量的大小。用顏色之優點，在乎顯明美觀，合於一般展覽之用；但是印費太貴，爲經濟起見，多以陰影替代之。

2.花線地圖(Cross-hatched map) 係以各種縱橫交叉的直線，來代表密度的不同。上述三種地圖，用以顯示分立數列之事實，甚爲相宜。至於連續數列的表顯，可用下述之點圖。因爲上

述各圖，使人對在每一區域內的分配有絕對一律的觀念；且在不同區域間，有頓成斷裂的現象。

3. 點地圖 Dot map 亦可分爲三種；

a 第一種是以大小不同之點，代表數量多寡。

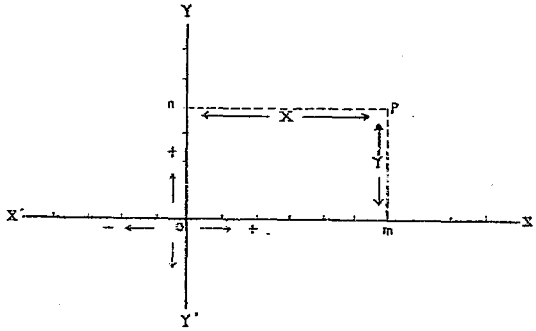
b 第二種是以大小一律之各點，敷以不同之陰影，以表明數值之不同。例如以全黑之點，代表最大數量，以$\frac{3}{4}$陰影之點，代表較小數值，$\frac{2}{4}$陰影之點，再較小之數值。$\frac{1}{4}$陰影之點，代表最小數值等。不過這種圖，不能示人以清晰的現象。

c 第三種是以點的密度厚薄，示現象之多寡，不問點的大小如何。這種圖，由疏薄而至稠密，示人以逐漸變遷的印象，集散的地域及程度，可以一望而知。此種地圖，用以表顯連續數列，甚爲滿意。圖上各區域的疆界，可說無足輕重，所以仍舊劃着者，無非爲便於研究或解釋起見。

4. 針圖 Pin map 假如要同時明瞭異地異時的事實之變化，非用針圖不可。例如：以針圖表示各地銷售數量，可以針之密度代表各地的銷售數量，而針之數目，又可以按實際情形之變化，隨時增減之，更可以不同顏色的針頭，代表不同的售貨員。這樣，一種圖就可以表示三種現象。有伸縮性，確是這圖的特質。

曲線圖

坐標制度 (Coordinate System) 曲線圖的作法，是根據坐標制度。故在討論曲線時，有先將坐標制度說明之必要。其法先在平面上作水平線 X'X，次作一垂直線 YY'，二線相交於 O，(如下圖)。



通常稱 O 點為原點 (Origin)，X'X為橫標軸，或X軸，YY'為縱標軸，或Y 軸。設 P 為此平面上之任何一點，作平行YY'之垂線，與X'X相交於m，再作一線平行X'X，而與YY'相交於n。試量O至m，得數以X表之，再量O至n，得數以Y表之。此二數記為X,Y，在本例中為6,4，此即點P之坐標，X為橫坐標，Y為縱坐標。平面上所有各點，多可以同樣方法，拿他和二軸的距離來決定。設將若干點，以線連綴之，即成曲線圖。

橫坐標在YY'之右者爲正值，在左者爲負值。縱坐標在X'X之上者爲正值，下者爲負值。這樣，分平面爲四象限。平常，統計中的數值，大都在右上方象限內，所有橫縱二坐標，都是正值。

曲線圖的種類　曲線圖可以不同的立場而分類如下：

1. 根據所表示的資料之性質而分類

a 次數曲線圖 Frequency curve　用以表示次數分配之資料，乃以次數所由生之量數爲主體，有累積次數曲線圖及簡單次數曲線圖二種，前者係將顯示各量數所有次數之點，用線連接之，後者係將各量數所有次數遞次累積，以累積之和所代表各點，用線連綴之即得。累積次數有以上累積及以下累積二種，故累積曲線圖亦分「以上」及「以下」二種。

b 時間曲線圖 Historical curve　用以代表時間性之資料，而觀察其變動狀況。所謂時間性的資料，即按時序之先後而發生有連帶關係之資料。此種曲線，分簡單，累積二類。前者代表時間數列每一時期之簡單數量，後者用以代表時間數列之由開始時累積至各時間之數量。

2. 根據量表劃分之標準而分類：

a 算術量表圖 (Arithmetic Scale)——量表劃分，係根據自然數(Natural Number)即以等距代表等量。

b 比率量表圖(Ratio scale)——量表劃分係根據對數，即以等距代表等比。

3. 根據曲線之形式而分類：

a 直線的——即以直線連綴已知各點

b 修勻的 (Smoothed Curve) ——即已修勻之曲線

各類曲線已於本章分別討論之，惟歷史的修勻曲線，則散見於第四編各章中。

次數曲線圖　次數圖總以次數所由生的量數為獨立變量，置於橫標軸，用橫量表表明量數之價值，以所發現的次數為倚變量，用縱量表表明之，然後按各量數之價值，以觀察他的次數變遷的規律。簡單次數圖，有直方圖(Column Histogram)，多邊圖(Polygon) 修勻次數曲線圖 (Smoothed Histogram) 三種，現在依次來說明。

1. 直方圖　製此圖時，可先在各量數的中點地位上，和相當次數所在的位置上，各作一小點，然後從此點左右平伸，得一橫線，長度恰恰一組距，再用垂直線連接各橫線即得。

2. 次數多邊圖　作此圖之一切手續，與直方圖相似。惟在製多邊圖時，縱標上各小點，以直線連接之即得。至於上下兩極端的直線應伸長至底線上最大和最小二量數外各$\frac{1}{2}$組距之地位上，圖15即一次數多邊圖。

圖14　直方圖

120 家工會失業會員百分率之次數分配

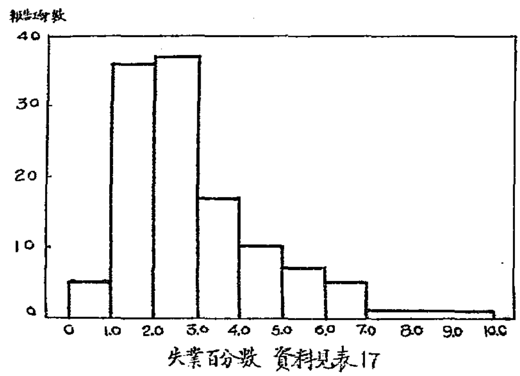
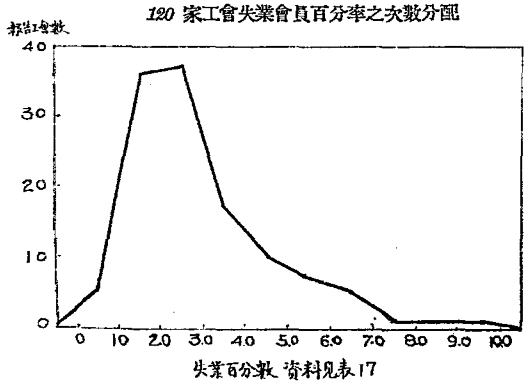


圖15　次數多邊圖

120 家工會失業會員百分率之次數分配



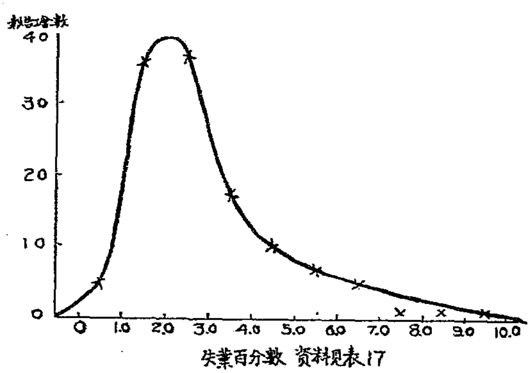
3.修勻次數曲線　簡稱次數曲線(圖16)。圖之製作，仍脫胎於上述二圖。所以可先作一直方圖或多邊圖，再審度形勢，去了棱角不齊的部分。這種修勻方法，可以用數學方程式，或者隨手配合，前者比較麻煩，所以多用後法（數學方程式之配合參，第二編第七章）。作成修勻曲線後之面積，應和直方圖或多邊圖，約略相等。這方法的適用，是根據：直方圖和多邊形雖則可以代表各組的相對大小，惟以組限是強訂的，組距大小或組的地位有變更，足使結果大異。若將次數增加，同時將組距減少，就可成一修勻曲線。事實上，我們的統計資料，不過是全體的標樣，所以我們可以假定那直方圖和多邊圖上所發生之不規則現象，都是偶然的，他的眞正現象的分配，乃是連續不斷的。不過這種假定，只能適用於連續數列之事實，因此修勻曲線，只能適用於連續數列，可是平常爲便利起見，分立數列，亦有用修勻者，雖在理論上不很合邏輯。至於修勻應到那種程度，各種資料的情境不同，程度亦不同。自然現象或機遇現象，常逼近常態曲綫，修勻可自由爲之，至於社會現象或經濟現象，常態曲綫上常有稜角，修勻時，祇能將小稜角除去，大稜角依然存在。

三種圖的個性已經約略討論，究竟怎樣選擇，我們再來總括說一下：直方圖的應用，於組數較少的分立數列，表顯事實最爲明顯。多邊形在簡單分配中，能給人以深刻的印象，若比

較二組或二組以上之數列時，尤爲適宜。在此要附帶提明的一點，就是幾個數列作比較時，最好先將原有分配，化爲百分數。至於修勻曲線的應用，於表顯事實之常態現象更爲滿意，不過要注意，將該項資料修勻之，是否合邏輯。

圖16　次數曲線圖

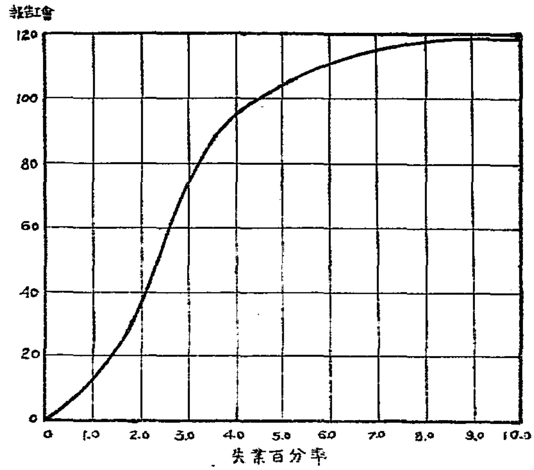
120 家工會失業會員百分率之次數分配



累積次數圖 (Gummulative Frequency Curve) 累積次數分配的圖解，雖則同簡單分配一樣地可有三種，祗修勻之一種，爲用較廣，這就是所謂累積次數曲線，又稱弧形 (Ogive)。作圖法與次數多邊形不同。作次數多邊形，是用各組中點，連線綴成。作弧形，卻用各組上限。此種曲線，有「以上」「以下」二種。平常多用「以上」弧形，但有時「以下」弧形，亦甚有用，全視資料之性質而定。

圖 17　累積次數曲線圖

120家工會失業會員百分率之次數分配



累積次數曲線有一種特點，就是雖則組距不等，次數不同的資料，所作弧形，總是一式。因此對於處置組距不等的困難，大可減少。平常兩個簡單次數圖，除非是用同一的歸類法(Grouping)，不便比較，在弧形就沒有這些限止。從弧形可以決定中位數、四分位數等，這是弧形的大貢獻，關於如何決定，留待下編討論。

時間曲線　時間曲線用以表現時間數列之資料，橫坐標所代表的是時間，縱坐標所代表者爲事實的量數。作此種圖解，縱橫二量表須有相當的比度。否則，不易得一清淅的概念，而

且相對的變化，容易被誤解，圖18是簡單時間曲線，圖19是累積時間曲線。

圖18　世界銀產量與銀價之比較

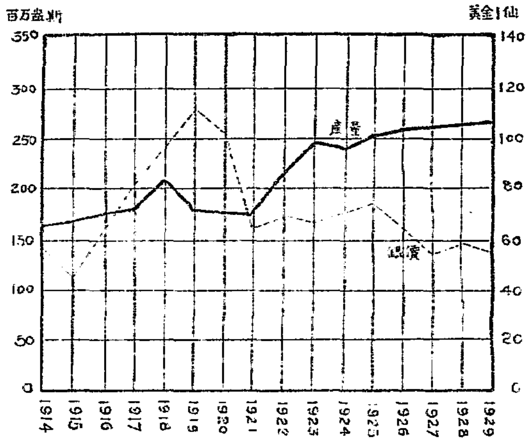
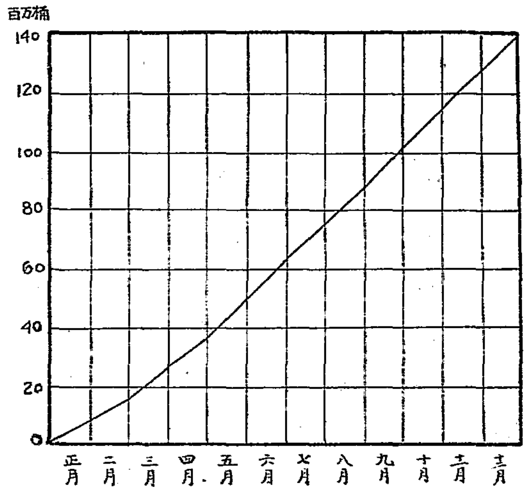


表18　美國朴脫倫水泥每月累積之產額

月分	產額（千桶）	
	每月	每月累積
一月	7990	7990
二月	8219	16200
三月	9880	26080
四月	11359	37439
五月	12910	50349
六月	12382	62731
七月	12620	75351
八月	12967	88318
九月	13109	101427
十月	13350	114777
十一月	12603	127380
十二月	9997	137377
	137377	

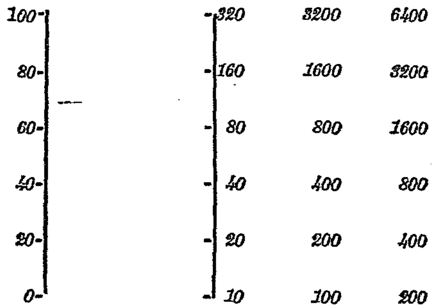
圖19 美國朴說倫水泥每月累積之出產額，1923.



比率圖 上面討論之圖所用量表是絕對數。就是說，在圖上等距的地位足以表示等差變化。有很多事實，要用比率變化，才有意義。這種比率變化，在圖解上的表顯，可有二種方法。一種是將量數的對數在算術量表上作圖，還有一種是用自然數在對數的或比率的量表上作圖。平常多用第二種方法，因爲比較地簡便。

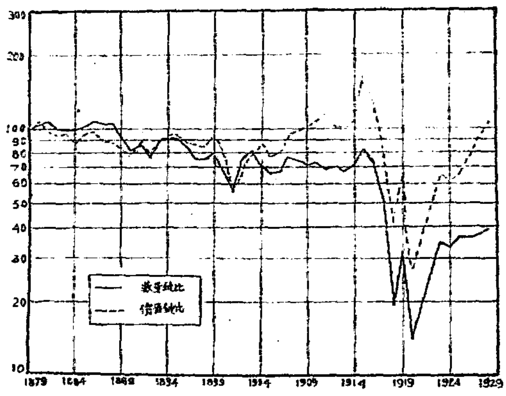
比率圖有三種(1)縱橫二標軸多用對數的量表(2)橫標軸的量表是對數，縱標軸則爲自然數(3)縱標軸是對數量表，橫標軸是自然數量表。三種中以第(3)種爲最有用。對數量表，

圖 20　算術量表與對數量表之比較



可以從一層至五層，圖21則爲一二層比率圖(但不完全)，每一層之距離相等，可以應用於自 10 而 100，或自 100 而 1000，或 1,000,000 而10,000,000。總之，須爲10之自乘數。

圖21　近五十年華茶出洋指數圖



比率圖的特性　他的特性，有下列幾點：

1.比率量表上，不設底線，〇度原點可以不必要。

2.設曲線之上升(或下降)，差不多成一直線，那末他所代表的數之遞增(或遞減)率，都是一致。

3.曲線向橫標軸成凹形，則顯明遞增率的減少。反之，曲線向橫標軸成凸形，則顯明遞增率的增加。

4.曲線斜度相同，足以表示相對變化的相同，所以二數列如同為等比增加或等比減少，所繪曲線為二平行線。要比較二數列之相對變化，只要比較二線的斜度。

5.比率圖上，相等距離的直線，表示相等比例變化，所以從比率圖上，可直接比較比例的增減。不過須注意，比例增加祇能和比例增加相比，比例遞減亦祇能和比例遞減相比。比例增加與比例遞減不能直接比較試用一簡單的例，來說明這點。例如從 100 變為 200 表示增加 100% ，但是從 200 變為 100則表示減少 50%。二者不同之原因，由於前者以 100 為基點，而後者以200 為基點。

美國標準圖示法聯合會 Joint Commitee on Standards for Graphic Presentation 規定之圖示標準法則：

1, 圖的普通排列須從左而右．

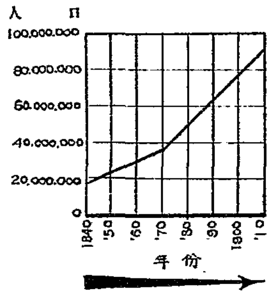


圖 1.

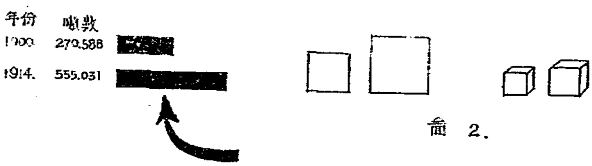


圖 2.

2, 能用直條代表數量最好，因為面積和體積都容易引起誤解．

3, 在曲線圖上，最好能將零度橫坐標表示出來．

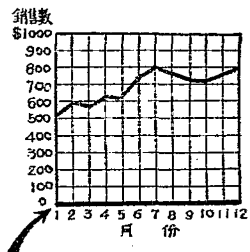


圖 3.

4, 若是零度橫坐標不能照常表示在圖上，那末就應該用一橫斷的裂痕將零度橫坐標表示出來。

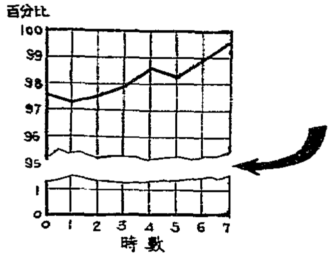


圖 4.

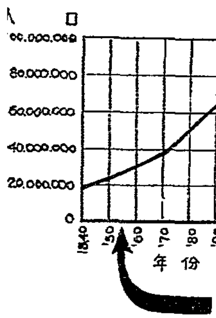


圖 5A.

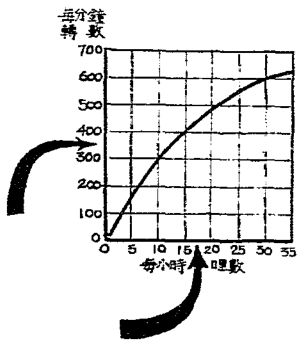


圖 5B.

5, 零度橫坐標應該和其餘的各線有顯明的區別。

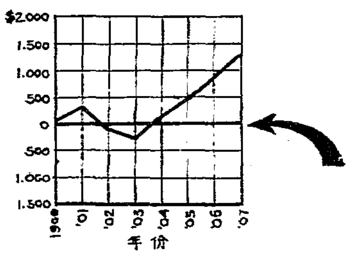


圖 5C.

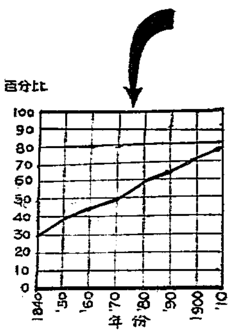


圖 6A.

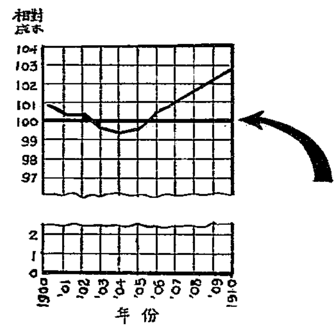


圖 6B.

6, 有些曲線的量表用百分比來表示，凡百分線宜較顯明，其餘用以作比較之線亦宜較為粗大。

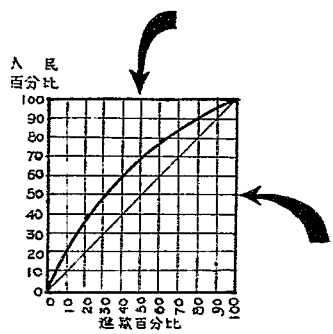


圖 6C.

7, 如圖是表示時間性的，而所表示之時間，又不完全，首線及末線不必特別區別，因為他并不表示時代的始終。

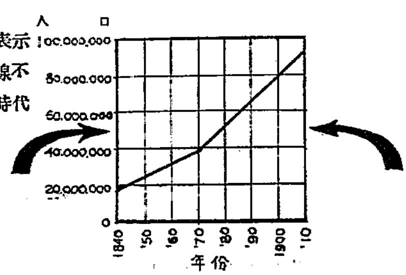


圖 7.

8, 對數格子上作圖，上下沿線應在對數的量表十乘冪上·

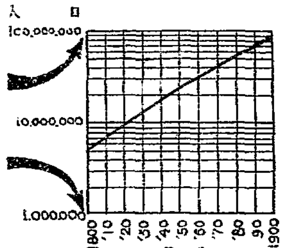


圖. 8

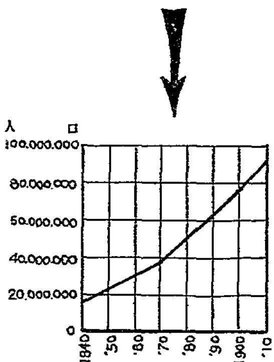


圖. 9A

9, 縱橫線除為醒目之必要外，不宜太多·

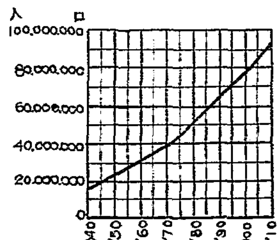


圖. 9B

10, 圖上曲線應與行格有顯明的區別·

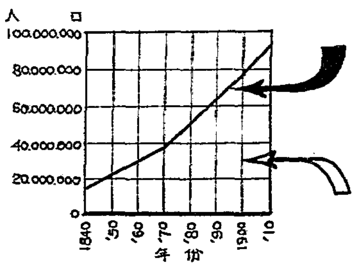


圖. 10

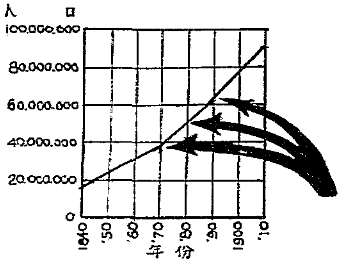


圖.11A

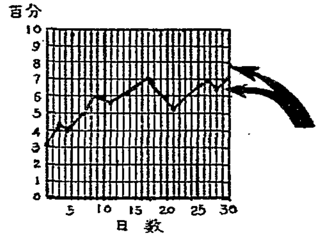


圖.11B

11, 如果曲線是表示連續一系的觀察，最好由圖上各點，切實地將各種觀察分別表明。

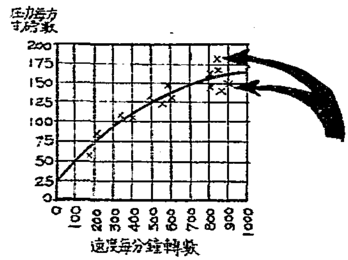


圖.11C

12, 橫量表須自左至右讀之，縱量表須自下而上讀之。

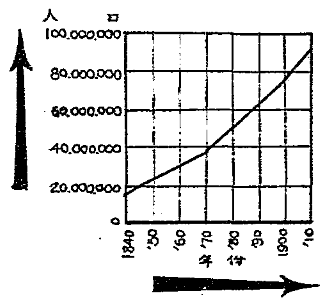


圖.12

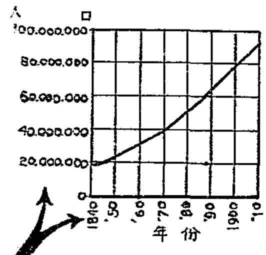


圖.13A

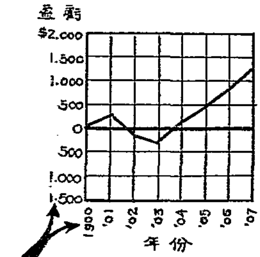


圖.13B

13, 量表上的數字須放在縱坐標之左橫坐標之下或沿列有關係之軸上・

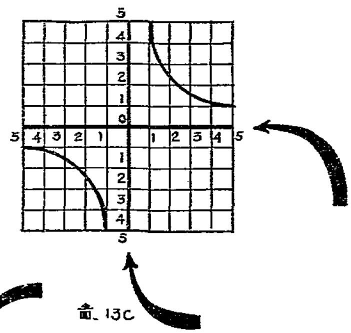


圖.13C

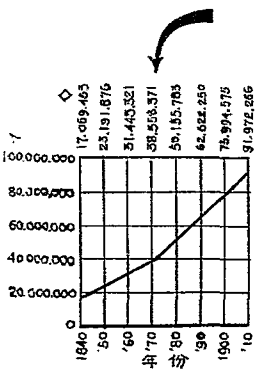


圖.14A

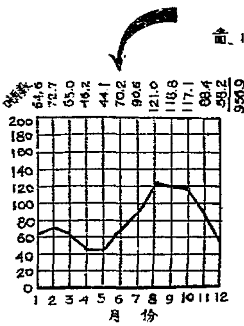


圖.14B

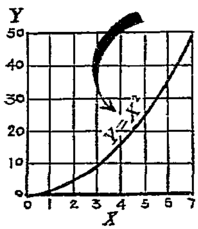


圖.14C

14, 圖線所代表的數量或公式，常需列入圖中・

15，設數字材料不便列入圖中，可另作一表附列之。

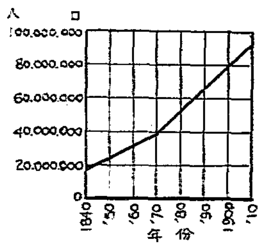
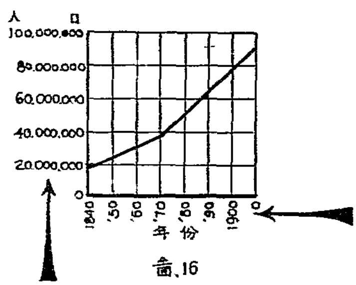


圖 15

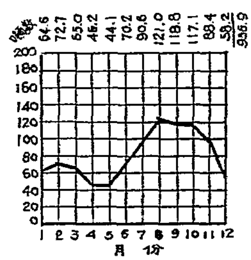
年份	人口
1840	17.069.453
·50	23.191.876
·60	31.443.321
70	38.558.371
·80	50.155.783
'90	62.622.250
1900	75.994.573
.10	91.972.266

16，圖上所有文字及數字，都宜放在下面或右面，以便閱讀。



圖、16

17，圖之標題宜清晰完備，如遇必要，不妨加以標目及說明，務使讀者易明。



第二工廠各月鑄鋁產量，1914.
產量以短噸計.
廢鋁錠數不計在內.

圖.17.

第貳編

一般的分析方法

(General Analytical Methods)

第貳編

一般的分析方法

已經蒐集的統計資料，要想把牠分門別類，互相比較，或尋求其間所存在的統計法則，那便不得不假助於分析。統計分析的主要目的，是在將繁複的資料，用簡明而總括的數字表示之。本編各章，就是敍述這類一般的分析方法。

第一章

平均數(Average)

英國統計學家 Bowley 氏說：統計學是平均數之學，雖則現在的統計學，不單是討論平均數，然而平均數在統計學中，還占着重要的地位。平均數之性質，是用一簡單的數字，以表示大量的統計資料之共通的性質。換言之，這便是使各個的統計資料所有的差異，相互抵消，而總括底將其全體之意義，表

現出來，以作其全體的代表。他的代表，有否意義，完全看他全體事項集中趨勢 (Central tendency) 而定。事項愈集中，代表之職愈相稱；否則，就沒有多大價値了。

平均數的功用

(1)平均數是以一簡單的數目，把測量的全體情形，淸淸楚楚地表現出來。例如：測量某城的工人工資，我們得到各種量數。但是對於這城工人的工資，究爲幾何，仍無明確的觀念。若從各量數求出一個代表數——平均數——便易於了解。

(2)平均數可以拿來作爲比較各組事實的差異。要比較兩個城市工人的工資，我們決不能將每個工人的工資來作比較，但能以甲城工人的平均工資，和乙城工人的平均工資，相互比較之。

(3)平均數可以測量事實一部分所得結果，代表全部事實。例如，要知上海工人的工資，不必把全體工人都測量到，只要選取其中一部分而測量之，作爲標樣，再去求出平均數就夠了。

(4)平均數可用來確切表明各組數量的關係。例如我們說，甲地的工資比乙地高，這未免太含混。較量其平均數，便有確切的意義。

平均數的種類

平均數的種類很多，但通用的有下列五種：

I. 算術平均數，

II. 中位數，

III. 衆數，

IV. 幾何平均數，

V. 倒數平均數，

試逐一討論之。

算術平均數 (Arithmetic mean)(M)

算術平均數是一種最普通的平均數，將量數的總和，除以項數(即次數)，所得商數，就是算術平均數。

設以 $X_1X_2X_3 \cdots\cdots X_n$ 代表各量數，ΣX 爲量數的總和，
的總和
N 爲項數，我們就可依上述的定義，得一計算公式如下：

$$M=\frac{X_1+X_2+X_3+\cdots\cdots\cdots+X_n}{N}=\frac{\Sigma X}{N}\cdots\cdots(1)$$

例如 2, 3, 4, 5, 6 五量數的算術平均數，就得

$$M=\frac{2+3+4+5+6}{5}=\frac{20}{5}=4$$

量數已成次數分配的算法　上述計算法，是指量數未經分組的資料，不過普通在項數較多時，可先編成次數分配而後再計算其平均數。他的算法，是將各組中數乘以次數，相加得總

和，再以總次數除之即得。設以$f_1 f_2 f_3 \cdots\cdots f_n$代表各組的次數，$m_1\ m_2\ m_3\ \cdots\cdots m_n$ 爲各組的中數，上面公式1就變爲：

$$M=\frac{f_1m_1+f_2m_2+f_3m_3+\cdots\cdots+f_nm_n}{f_1+f_2+f_3+\cdots\cdots+f_n}$$
$$=\frac{\Sigma f\ m}{\Sigma f 即 N}\cdots\cdots\cdots\cdots\cdots\cdots 公式(2)$$

從次數表來算M，我們應當有一種假設，就是假定任何一組的分配，是均勻整齊的，而且有集中各組中點的趨向；所以每組的中點，可以當作該組各量數的代表數。顯然的，這種假定不能希望他完全正確；不過在組距不很大的資料，除非組內

表 19　用次數表計算算術平均數

北京西郊第一區漢人家庭之進款(民國六年)

組　距 元 (1)	中　點 m (2)	次　數 f (3)	fm (4)
50— 69.9	60	11	660
70— 89.9	80	14	1,120
90—109.9	100	26	2,600
110—129.9	120	15	1,800
130—149.9	140	13	1,820
150—169.9	160	8	1,280
170—189.9	180	2	360
190—209.9	200	5	1,000
210—229.9	220	2	440
230—249.9	240	3	720
250—269.9	260	1	260
總　數		100	12,060

$$M=\frac{\Sigma(fm)}{N}=\frac{12.060}{100}=120.6$$

的分配太失稱，則對於全體尙不致有大差誤。試舉一例於表19以示算法之一般。

簡捷法　上項方法，仍覺繁瑣，實際計算，常用簡法，以假設平均數爲原點(Origin)，以組距爲單位，末後再改正之，復原至其原用之單位，就得其算術平均數。至於該簡算方法的根據，則有下列理論：從算術平均數的算法，我們可以知道，各量數對平均數的離差總和是0，試證明之：

設　$X_1X_2X_3\cdots\cdots X_n$爲各量數之值，

$d_1d_2d_3\cdots\cdots d_n$爲各量數對M之離差，

因平均數之公式爲：

$$M=\frac{X_1+X_2+X_3+\cdots\cdots\cdots+X_n}{N}$$

$16=\frac{10+16+n}{2}=16$

$16\times2=10+16+n$

$(10-16)+(16-16)+(n-18)$

$=0\ -6\ -\ \ldots$

或

$$NM=X_1+X_2+X_3+\cdots\cdots+X_n,$$

項數旣爲N，故

$$(X_1-M)+(X_2-M)+(X_3-M)+\cdots\cdots+(X_n-M)=0$$

但　$(X_1-M)=d_1$; $(X_2-M)=d_2$; $(X_3-M)=d_3$; $\cdots\cdots\cdots\cdots\cdots\cdots(X_n-M)=d_n$

$\therefore\ \Sigma d=0$，

我們根據這點特質，就可任取一數，爲假設平均數M'，計算各

項對此M′的離差，求其代數和，再以次數除之，所得商數，就是假設平均數與眞正平均數的差額。將這差額，加於假設平均數上，即得其眞正平均數。今試以代數證明之：

設　M′爲假設平均數，　c 爲應改正之數值，

$d'_1\ d'_2\ d'_3\cdots\cdots d'_n$爲各量數對M′之離差，

$M=M'+c$,

$d'_1=X_1-M';\ d'_2=X_2-M';\ d'_3=X_3-M';\cdots\cdots$

$d'_n=X_n-M'.$

於是　$d'_1=d_1+c;\ d'_2=d_2+c;\ d'_3=d_3+c;\cdots\cdots$

$d'_n=d_n+c.$

$\Sigma d'=\Sigma d+Nc,$

而　$\Sigma d=o,$

$\therefore\ \Sigma d'=Nc,$

$c=\frac{\Sigma d'}{N}$

將　$c=\frac{\Sigma d'}{N}$代入$M=M'+c$,即得

$$M=M'+\frac{\Sigma d'}{N}\qquad \text{公式(3)}$$

這公式如以組距爲單位，則計算更形簡便。他的公式是

$$M=M'+\frac{\Sigma fd'}{N}\times i\qquad \text{公式(4)}$$

i是組距。現在試舉一例於表20，以明其運算之方法。

表 20 算術平均數計算

我國郵政職工每月消費之狀況

消費元數 X (1)	職工數 f (2)	d' (3)	fd' − (4)	fd' + (5)	d' (6)	fd' − (7)	fd' + (8)
10—15	1	−3	3		−5	5	
15—20	13	−2	26		−4	52	
20—25	52	−1	52		−3	156	
25—30	61	0			−2	122	
30—35	49	1		49	−1	49	
35—40	53	2		106	0		
40—45	37	3		111	1		37
45—50	12	4		48	2		24
50—55	20	5		100	3		60
55—60	13	6		78	4		52
60—65	6	7		42	5		30
65—70	7	8		56	6		42
70—75	4	9		36	7		28
75—80	4	10		40	8		32
80—85	2	11		22	9		18
85—90	1	12		12	10		10
90—95	—	13		0	11		—
95—100	1	14		14	12		12
100—105	1	15		15	13		13
105—110	2	16		32	14		28
110—115	1	17		17	15		15
115—120	—	18		0	16		—
120—125	—	19		0	17		—
125—130	—	20		0	18		—
130—135	—	21		0	19		—
135—140	1	22		22	20		20
總數	341		−81	800		−384	421
			719			37	

$$C' = \frac{\Sigma f d}{N} = \frac{719}{341} = 2.1085 \text{（組距單位）}$$

$$C = 2.1085 \times 5 = 10.5425$$

$$M = M' + \frac{\Sigma f d'}{N} = 27.5 + 10.5425 = 38.0425$$

或 $C' = \dfrac{\Sigma f d'}{N} = \dfrac{137}{344} = .1085$ （組距單位）

$c=1085\times 5=.5425$

$$M=M'+\frac{\Sigma fd'}{N}=37.5+.5425=38.0425$$

由簡法計算算術平均數的手續，總括之爲下列各步。

1. 將原有資料列成次數分配表。
2. 求出次數的總和。
3. 選擇相當的一組，取他的中點，當作假設平均數。
4. 將組距作單位，以假設平均數爲原點，求各組中點與假設平均數之差(d')，各組中點大於假設平均數者爲正數，小於假設平均數者爲負數。
5. 各差數乘以相當次數。
6. 求fd' 各乘積的代數和。
7. 第6步所得結果以總次數除之，所得商數，就是應行校正之數(c')。〔須注意，此時之 c' 仍爲組距的單位。〕
8. 以組距乘c', 得校正數，此時c已恢復爲原有單位。
9. 將校正數 c, 加於假設平均數上，就得眞正算術平均數。

算術平均數之特質　下列各點，說明算術平均數的特質，同時足以證明他對於數學上處理之便利。

1, 對算術平均數之離差之代數和為零。

2, 一數列 X，包含二小組時，全數列之平均數，可即從二分組之平均數以表出之。例如第一分組之各數值為 X_1，第二分組之數值為 X_2，

$$\Sigma X=\Sigma(X_1)+\Sigma(X_2)$$

假如第一分組有 N_1 次數，第二分組有 N_2 次數，二數列之平均數，各為 M_1, M_2，則

$$NM=N_1M_1+N_2M_2$$

例如 346 個愛爾蘭人之平均身長為 67.78 英寸，741 個威爾斯人之平均身長為 66.62 英寸，因此，全體 1087 人之平均身長由方程式

$$1087M=(346\times67.78)+(741\times66.62)$$

$$\text{得}M=66.99\text{英寸，}$$

這種關係，我們可再用更普遍些的方程式來表示，例如有 r 個分組 $X_1, X_2, X_3, \cdots\cdots X_r$; $M_1, M_2, M_3, \cdots\cdots M_r$, 為各數列之平均數，則全數列之平均數 M，可由下式得之。

$$NM=N_1M_1+N_2M_2+\cdots\cdots+N_rM_r$$

假使各數列，係用同一之假設原點，為便於勘誤起見，我們可仍以 1,2,3⋯⋯r 代表各分組，則

$$\Sigma(f\cdot d')=\Sigma(f_1\cdot d'_1)+\Sigma(f_2\cdot d'_2)+\Sigma(f_3\cdot d'_3)+\cdots\cdots$$
$$+\Sigma(f_r\cdot d'_r)$$

倘使總數符合，則計算正確，毫無疑義。

3, 次數相同之二數列，他們相當各項所有之和或差之平均數，等於二數列平均數之和或差。因

$$X=X_1+X_2$$

$$\Sigma(X)=\Sigma(X_1)+\Sigma(X_2),$$

若M, M_1, M_2,代表各平均數，則

$$M=M_1\pm M_2$$

假如把這公式擴充成為一個普遍的公式，便是

$$X=X_1\pm X_2\pm X_3\cdots\cdots\cdots\cdots\pm X_r$$

$$M=M_1\pm M_2\pm M_3\cdots\cdots\cdots\cdots\pm M_r$$

中位數 (Median) (Md)

當全體量數依大小順序排列時，分全體為相等的二部，中位數是恰居於中的數值，上下兩端，各有量數之一半。

中位數之算法　在未經歸類的資料，中位數的決定，比較地簡易，只要應用定義，將全體量數，依大小順序排列之。若量數的數目是奇數，則其正中間的量數，即第$\frac{n+1}{2}$項，就是中位數。若量數的數目是偶數，則中位數就等於最中二量數的算術數平均數。

如所用資料，係一次數分配之形式，計算中位數第一步之

公式仍爲$\frac{n+1}{2}$，第二步之公式可用補插法。這方法是假定中位數所在一組的量數分配，是整齊均勻的，各數間的相差是等距的。他的公式是

$$Md = L + \left(g - \frac{1}{2}\right)\frac{i}{f} = L + \frac{(2g-1)}{2f} i \text{ (註)} \cdots\cdots\cdots (1)$$

$$\text{或 } Md = U - \left(g - \frac{1}{2}\right)\frac{i}{f} = U - \frac{(2g-1)}{2f} i \cdots\cdots\cdots (2)$$

其算法，先就次數分配表上，決定中位數所在之一組。設全體爲 n 項，那末第$\frac{n+1}{2}$項所在的一組就是。再求此組下限向上（或上限向下，用公式 2；因求中位數，可從小量數數起，亦可從大量數數起。）到中位數的次數 g，於是以 2 乘之，其乘積減以一，再以此結果除以中位數所在一組的次數之 2 倍，所得商數，乘以組距 i 後，加於該組下限 L 上（或從上限 U 減去）即得中位數。

(註) 此公式與 $Md = L + \frac{\frac{n}{2} - F}{f} i$ 之結果相等，因

$$L + \frac{2\left(\frac{n+1}{2} - F\right) - 1}{2f} i = L + \frac{\frac{n}{2} - F + \frac{1}{2} - \frac{1}{2}}{f} i = L + \frac{\frac{n}{2} - F}{f} i$$

第一步中先添$\frac{1}{2}$，第二步中，又減去$\frac{1}{2}$，本書並非故作無謂之麻煩，實以欲使未歸類之資料，與次數數列之資料，其計算方法使其一致，於後者之計算時，僅多一第二步之公式，俾免初學誤會。

試以美國工資之資料求其中位數，示例於下表：

表21　工資分配次數表

工資(元)	簡單次數	累積次數	
	f	向上	向下
2.00—2.49	6	6	336
2.50—2.99	16	22	320
3.00—3.49	34	56	314
3.50—3.99	61	117	280
4.00—4.49	66	183	219
4.50—4.99	57	240	153
5.00—5.49	37	277	69
5.50—5.99	28	305	59
6.00—6.49	9	314	31
6.50—6.99	6	320	22
7.00—7.49	8	328	16
7.50—7.99	8	336	8
總數	336		

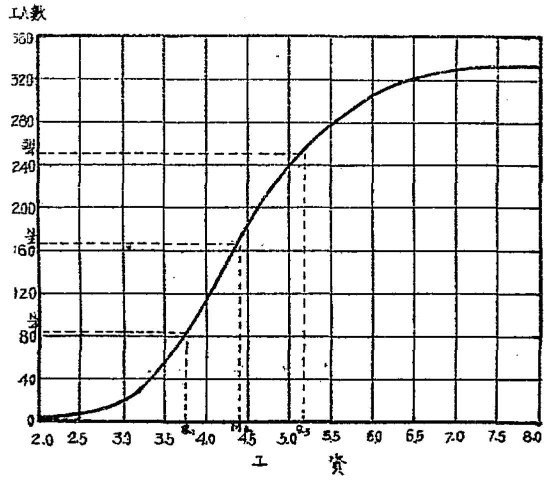
$$Md = 4.00 + \left(\frac{2(51.5)-1}{2(66)}\right).50 = 4.00 + .386 = 4.386$$

或 $$Md = 4.50 - \left(\frac{2(1.55)-1}{2(66)}\right).50 = 4.50 - .114 = 4.386$$

圖解法．我們也可用圖解來決定中位數，先將各量數編成累積次數分配，再繪一修勻之弧形（Ogive），從縱坐標的量表上$\frac{n}{2}$處，作與橫坐標軸平行的一綫，而與弧形相交，這交點的橫坐標，就是中位數。

在圖22給我們一個圖解法的示例，自縱標軸$\frac{N}{2}$之處，作一與底綫平行之直綫，橫交於修勻曲綫，自其交點再作垂直線，交於橫坐標上，則中位數爲4.4。

圖 22　累積次數分配圖（材料根據表21）



四分位數 Quartile 十分位數 Decile 及百分位數 Percentile 有時統計分析上，應用四分位數，十分位數，及百分位數，他的原理，與中位數相同。此等量數連中位數，可統稱之為分割數 (Partition Valve)。分割數的定義是：凡是把全體量數，依小大順序排列，將他分割成若干等分，在分割界上各項的數值，便是分割數。四分位數，便是分割全體量數為相等的四分；十分位數，便是分割全體量數為相等的十分；百分位數，便是分割全體量數為相等的百分。假設把 N 項的順列(Array)，分割為 P 個等分，則第 r 個分割數，便是第 $\left(\frac{rn}{P}+\frac{1}{2}\right)$ 項的數

值，從此公式，可決定任何分割數。例如第一個四分位數(First Quartile)（Q_1）便是第$\left(\frac{n}{4}+\frac{1}{2}\right)$項的數值，第三四分位數(Third Quartile)（Q_3）便在第$\left(\frac{3n}{4}+\frac{1}{2}\right)$項，餘類推。第二個四分位數（$Q_2$），第五個十分位數（$D_5$）及第五十個百分位數($P_{50}$)，卽中位數；因$Q_2$是在第$\left(\frac{2n}{4}+\frac{1}{2}\right)$項，$D_5$是在第$\left(\frac{5n}{10}+\frac{1}{2}\right)$項，$P_{50}$是在第$\left(\frac{50n}{100}+\frac{1}{2}\right)$項，簡化之，卽第$\frac{n+1}{2}$項。設在次數數列，則第二步之公式，與計算中位數者相同。

他們的計算，以表21之資料作例，結果如下：

$$\frac{N}{4}+\frac{1}{2}=\frac{336}{4}+\frac{1}{2}=84.5$$

$$Q_1=L+\frac{2g-1}{f}i=3.50+\frac{2(28.5)-1}{2(61)}\times.50$$

$$=3.50+.23=3.73$$

$$Q_3=U-\frac{2g-1}{f}i=5.50-\frac{2(25.5)-1}{2(37)}\times.50$$

$$=5.50-.34=5.16$$

$$\frac{5N}{10}+\frac{1}{2}=\frac{5(336)}{10}+\frac{1}{2}=168.5$$

$$D_5=L+\frac{2g-1}{2f}\times i=4.00+\frac{2(51.5)-1}{2(66)}\times.50$$

$$=4.00+.39=4.39$$

四分位數等，又可用圖解來決定，方法與中位數同，不再贅。

衆數 Mode (Mo)

衆數可以說是最常有的項。或者說是在量表上項數最密集的地位的量數，從次數分配來講，衆數就是次數曲線下最高縱坐標所在橫標軸上的數值。衆數依他的精粗可分爲二種：近似衆數與理論衆數。近似衆數的數值，常因歸類法的不同而變更，很不固定。理論衆數，雖屬可靠，但以算法難繁，非初學所能涉及。所以初步統計中，所求得的衆數，大致多以逼近的程度爲止。我們解釋他的結果時，應當注意。現在就介紹幾種求近似衆數的方法在下面：

算法1　決定衆數，最好先將資料編成次數分配，取其次數最多一組，該組的中值，就得作爲衆數的近似值。這方法的缺點，在所得結果，隨次數最多一組的地位而不同。所以應當加以改良，最好將最多次數組上下二鄰組的次數，使其補插之，所得結果，較爲精確。其公式表之如下：

設以L爲衆數所在組的下限，或U爲衆數所在組的上限，

f_1爲鄰近衆數下一組的次數

f_2爲鄰近衆數上一組的次數

Md

於是$Mo = L + \frac{f_2}{f_1+f_2} \times i$,……………(1)

或　$Mo = U - \frac{f_1}{f_1+f_2} \times i$,……………(2)

將表21的事實代入公式 (1) 得

$$Mo = 4.00 + \frac{57}{61+57}.50 = 4.00 + .24 = 4.24$$

或用表21的事實代入公式 (2) ，得

$$Mo = 4.50 - \frac{61}{61+57}.50 = 4.50 - .26 = 4.24$$

上列二公式，可任擇其一而用之。在次數分配不很整齊時，可用鄰近衆數所在組每端的二組，或二組以上的次數而均衡之。不過須注意：所分析的事實，如係眞正的多峯衆數分配 (Multimodelity)，這方法不適用。

算法二　批爾生的經驗法則　批爾生 (Pearson) 計算衆數，是根據算術平均數與中位數的數値而決定。因爲在失稱不甚的分配中，M, Md 和 Mo 三者，恆有一定的關係。就是衆數與算術平均數相離最遠，而中位數對算術平均數的距離，約等於算術平均數與衆數的 $\frac{1}{3}$。根據這關係，可得下式。

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

試將表21的資料，求出 M=4.513，Md=4.386，代入公式，得

$$Mo = 4.513 - 3(4.513 - 4.386) = 4.132$$

這公式，只能適用於失稱不甚的次數分配上，若失稱過甚，不能引用。

算法三　繼續歸類法(Regrouping Process)　設次數分配

不甚整齊，衆數究在何組，難於確定時，最好用繼續歸類法。其法先將每兩組相加，得出一列新數，於是將上限移下一組（就是從第二組起），再兩組相加，又得一列新數。再三組相加，得出一列新數。再將上限移下一組，再三組相加；再從第三組起，三組相加。如仍須加時，可每四組相加，直至不規則的現象除去，而最大次數的位置，不致因上限變更而改變爲止。衆數即在歸類後，所得次數最多組之數值。惟此法手續太麻煩，有時所得結果，過於人爲，未始非缺憾也。表22就是這方法的例子。

經表22之繼續歸類後而得一角組之次數最多組爲685 (1.15－1.24)，二角組爲989 (1.05－1.24)，924 (1.35－1.54)，三角組爲1242(1.15－1.44)，1190(.95－1.24)，1088(1.05－1.34)，五角組爲1472(.75－1.24)，2012(1.05－1.54)，試將此結果作一次數表如下，則我人得知

.75— .84	1
.85— .94	1
.95—1.04	2
1.05—1.14	5
1.15—1.24	6
1.25—1.34	3
1.35—1.44	3
1.45—1.54	2

由表觀之，我人可決定衆數在1.15－1.24組之間，因其次

表 22　繼續歸類法決定衆數

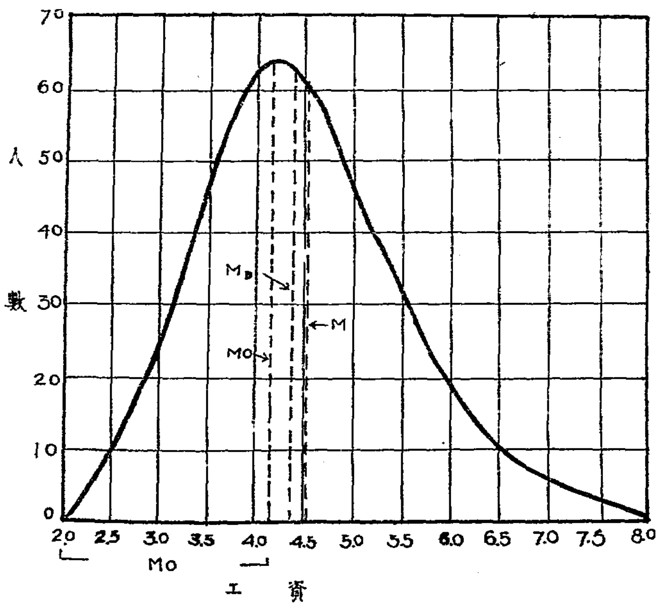
工資率	工人數 一角組	二角組		三角組			五角組	
$.25—.34	1	16		75			317	
.35—.44	15		74		159			
.45—.54	59	144				301		725
.55—.64	85		242	355				
.65—.74	157	270			439			
.75—.84	113		282			483	1472	
.85—.94	169	370		674				
.95—1.04	201		505		1190			2012
1.05—1.14	304	989				1088		
1.15—1.24	685		784	1242				
1.25—1.34	99	557			1023		1297	
1.35—1.44	458		924			996		
1.45—1.54	466	538		740				934
1.55—1.64	72		274		603			
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		640
2.05—2.14	45	310		343				
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			322
2.55—2.64	13		176			178		
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				285
3.05—3.14	5	52			64			
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		114
3.55—3.64	5		21	32				
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			6
4.05—4.14	0	3				3		
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				9
4.55—4.64	1		1		1			
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					
5.25—5.34	1	1						

Bowley's test P. 97.

數最大，而次數分配最爲規則之五角組之最後一行，其最大次數2012，亦跨於此值($1.15—1.24)。

圖解法　衆數的決定亦可以圖解法得之。因爲照衆數的定義，是次數曲線下最高縱坐標所在底線上的量數，所以我們可從簡單次數曲線，取其最高縱坐標所在的數值，就可得到衆數，例如圖23，圖中M,Md 和Mo 三者的關係亦可看出。衆數可以在弧形上約略決定，曲線上最陡而幾成直綫之點，卽爲衆數，惟平常用弧形決定的衆數，無甚價値耳。

圖 23　衆數之圖解



幾何平均數

幾何平均數(Geometric mean) (G)雖不如算術平均的普通，但在統計分析中，有特殊的應用。他對於等比變化，與以同等看待。所以要計算平均比率，或平均遞加率，或遞減率，非此不爲功。自 Jevons 首先用於計算物價指數後，幾何平均數的應用範圍，亦日益見廣。此外，對於人口增加率的計算，亦多利用之。幾何平均數，可以說是n項乘積所開n方的方根。公式是：

$$G=\sqrt[n]{X_1\cdot X_2\cdot X_3\cdot\cdots\cdots\cdots X_n}$$

例如2，3，4，5，6五量數的幾何平均數是

$$G=\sqrt[5]{2\times3\times4\times5\times6}=\sqrt[5]{720}=3.73$$

用算術方法算幾何平均數手續太煩，實際多利用對數，計算較簡。上列公式變爲

$$\log G=\frac{\log X_1+\log X_2+\cdots\cdots+\log X_n}{N}=\frac{\Sigma\log X}{N}$$

所以G的對數，可以說是各量數的對數的算術平均數。設各量數已經歸類成爲次數分配，可以下列公式計算之。

$$G=\sqrt[n]{(X_1)^{f_1}\cdot(X_2)^{f_2}\cdot(X_3)^{f_3}\cdots\cdots(X_n)^{f_n}}$$

如用對數，變爲：

$$\log G=\frac{f_1\log X_1+f_2\log X_2+f_3\log X_3+\cdots+f_n\log X_n}{N}$$

$$=\frac{\Sigma f(\log X)}{N}$$

如用各組距中值則公式爲

$$\log G=\frac{\Sigma f(\log m)}{N}$$

試舉一例以明其計算之程序

表 23　幾何平均數之計算

一九三一年山東山西河北河南陝西湖北安徽七省各縣小麥之價比

(1932年 1月=100)

價比 X (1)	組距中值 m (2)	log m (3)	報告縣數 f (4)	f·log m (5)
50—60	55	1.47036	1	1.74036
60—70	65	1.81291	10	18.12910
70—80	75	1.87506	5	9.37530
80—90	85	1.72942	16	30.87072
90—100	95	1.97772	33	65.26476
100—110	105	2.02119	83	167.75877
110—120	115	2.06070	75	154.55250
120—130	125	2.09691	71	148.88061
130—140	135	2.13033	46	97.99518
140—150	145	2.16137	29	62.67973
150—160	155	2.19033	18	39.42594
160—170	165	2.21748	13	28.82724
170—180	175	2.24304	6	13.45824
180—190	185	2.26717	4	9.06868
總數			410	848.02713

$$\log G=\frac{848.02713}{410}=2.06836$$

$$G=117.05$$

幾何平均數的應用　在經濟統計中，幾何平均數最著的用途，是在物價指數。幾何平均數，是就各項之比例，而求其平均物價，即表示物價漲落的比例，所以很爲適用。因爲物價從

50漲至100，和從100漲至200，同等重要，算術平均數就不能顯出這點。

例如兩種貨物，一種從100元至1000元漲10倍，一種從100元至10元跌至$\frac{1}{10}$，若用算術平均數，則$\frac{1000+10}{2}=505$，而G為$\sqrt{1000\times10}=100$。一漲一跌，兩相抵銷，算術平均數505，不能代表眞實的情形。

此外任何事物，依複利率增加者，求其平均增加率，亦可用幾何平均數。複利率之公式如下

$$p_n=p_o(1+r)^n$$

$$於是\ r=\sqrt[n]{\frac{p_n}{p_o}}-1$$

設本金1000元，複利計息，12年後本利和爲1600是則增加者有60%。若依算術平均數計算，則利率爲5%。但此例之複利，並非5%，其眞正之利率爲

$$r=\sqrt[12]{\frac{1600}{1000}}-1$$

$$=\sqrt[12]{1.60}-1$$

$$=1.04-1=.04=4\%.$$

人口自然之增加，亦依複利率進行。故人口增加率之決定，亦宜用幾何平均數。例如某市人口在1920年的調查，總數爲100,000在1930年的清查總數爲150,000，求1920年後任何一年之人口，即可以下式之增加比率推算之。

因 $r = \sqrt[10]{\dfrac{150.000}{100.000}} - 1 = .04138 = 4.138\%$

此外教育測驗上，計算學生能力之平均增進率，亦可用幾何平均數。

倒數平均數

倒數平均數 (Harmonic mean) (H) 是各量數的倒數的平均數的倒數。公式如下

$$H = \frac{1}{\dfrac{\dfrac{1}{X_1} + \dfrac{1}{X_2} + \dfrac{1}{X_3} + \cdots + \dfrac{1}{X_n}}{N}}$$

$$= \frac{N}{\dfrac{1}{X_1} + \dfrac{1}{X_2} + \dfrac{1}{X_3} + \cdots + \dfrac{1}{X_n}} = \frac{N}{\Sigma(\dfrac{1}{X})}$$

如是，則 2,3,4,5,6, 五量數的倒數平均數，當爲

$$H = \frac{1}{\dfrac{\dfrac{1}{2}+\dfrac{1}{3}+\dfrac{1}{4}+\dfrac{1}{5}+\dfrac{1}{6}}{5}} = \frac{5}{\dfrac{87}{60}} = \frac{300}{87} = 3.45$$

算倒數平均數時，如項數稍多，最好利用倒數表，比較簡便。設從次數資料而求H，則上列公式變爲

$$\frac{1}{H} = \frac{f_1(\dfrac{1}{X_1}) + f_2(\dfrac{1}{X_2}) + f_3(\dfrac{1}{X_3}) + \cdots\cdots + f_n(\dfrac{1}{X_n})}{\Sigma f}$$

$$=\frac{\Sigma f(\frac{1}{X})}{\Sigma f\text{或}N}$$

或　$$H=\frac{\Sigma f\text{或}N}{\Sigma f(\frac{1}{X})}$$

如用各組距中值，則公式變爲

$$H=\frac{\Sigma f}{\Sigma f(\frac{1}{m})}$$

計算方法用表34來舉例說明。

倒數平均數，應用範圍，至爲有限，在求時間速率的平均數，和物價以「每元若干」計算者，求他的平均價格，或其他類似的資料，當用倒數平均數。例如某甲行路，先以每小時行八里的步行速率行4里，次以每小時行十里的速率行4里，求此人的平均速率。如用算術平均數算，$M=\frac{8+10}{2}=9$ 里，則每里需時 $6\frac{2}{3}$ 分，八里需時$53\frac{1}{3}$分，他的結果很不正確。因爲每小時八里的速率步行 4 里，需時 30 分（每里需7.5分，7.5 分×4），每小時行十里的速率步行 4 里需時 24 分（每里需 6 分，6 分×4），所以步行 8 里需時54分(30分+24分)，換言之，他的平均速率是每小時 $8\frac{8}{9}$里。試以H算之，其結果正同。

$$H=\frac{1}{\frac{\frac{1}{8}+\frac{1}{10}}{2}}=\frac{2}{\frac{9}{40}}=\frac{80}{9}=8\frac{8}{9}\text{里}$$

表 24　倒數平均數之計算

在華日本紗廠投資數額(1930)

每廠資本額 (1,000元) X (1)	廠數 f (2)	組距中點 m (3)	$\frac{1}{m}$ (4)	$f\frac{1}{m}$ (5)
500以下	1	250	.0040000	.0040000
500— 1,000	1	750	.0013333	.0013333
1,000— 1,500	5	1250	.0008000	.0040000
1,500— 2,000	3	1750	.0005714	.0017142
2,000— 2,500	8	2250	.0004444	.0035552
2,500— 3,000	7	2750	.0003636	.0029231
3,000— 3,500	3	3250	.0003077	.0002667
3,500— 4,000	1	3750	.0002667	.0004706
4,000— 4,500	2	4250	.0002353	.0006315
4,500— 5,000	3	4750	.0002105	0001905
5,000— 5,500	1	5250	.0001905	.0001739
5,500— 6,000	1	5750	.0001739	.0003200
6,000— 6,500	2	6250	.0001600	.0005924
6,500— 7,000	4	6750	.0001481	.0000930
10,500—11,000	1	10750	.0001379	.0008096
總數	43			.0208096

$$H=\frac{34}{.0208096}=2066.4$$

註　本表材料，用倒數平均數求平均數，不合理論，僅能作計算示例觀之

加權 Weighted

意義　平均數中，尤其是在算術平均數，和幾何平均數，有時有所謂(1)簡單或(2)加權之分。加權的意義，有時像次數分配表各組的次數；但在另一方面，是表示各量數的比較重要性，并非代表次數，意義截然不同。雖則在計算上沒有區別。看下面的例子，就可以明瞭。例如，有每斤6角的酒二斤，

每斤5角的酒三斤，每斤四角的酒四斤，現在混合之，求其平均價格，必須將每種酒的斤數，作權數。在這種情形的權數，與次數分配之次數，毫無軒輊，再如某省調查人口死亡率，假設將各縣的死亡率總合之，除以縣數，决不能代表全省眞實死亡情形。必須拿各縣人口的比例來作權數，去乘各縣的死亡率才可。這時的權數，决與次數不同。在「指數」的加權，就是屬於後者的意義。權數有時有所謂「概權」者，大都由估算決定之。

加權的影響　若是加權的數很少。權數的大小對平均數有很大的影響，若加權的數目很多，因爲有互相補償的作用，影響結果就很小。不過亦要看權數的大小與項形 (item) 的關係，倘使權數大，項形亦大，或權數小，項形亦小，權數稍有錯誤，對於平均數，很有影響。

選擇平均數的標準

五種普通的平均數，已經在上面略加申述，現在要進一步來研究各種的優劣所在，以爲選擇標準。Yule 曾經規定一個好的平均數所應具的六種要素(註1)。現在先介紹在下面，再一種一種來討論。

註1，參Yule: Introduction to the theory of Statistics P. 108

(1)平均數應有切實的數量，不宜由觀察者的主觀推測而定。

(2)平均數應根據全體各項中算出，否則拿平均數來代表全數列，實屬疑問。

(3)平均數應簡單易懂，不過亦有例外，因爲有些平均數，對於平常人，難以了解，但對於特殊的目的，很適用。

(4)平均數應要容易計算，因爲時間與經濟的節省，亦很重要。

(5)平均數要有穩定性，就是說這平均數不致因標樣的變動，而受極大的影響。

(6)平均數要適於更進的代數上的處置，有時求平均數不過分析上的第一步，所以爲便於進一步的分析起見，平均數要適於代數的計算才好。

從第一點來講，算術平均數，幾何平均數，與倒數平均數三者，似乎較勝於二種地位平均數。因爲後二者，尤其是衆數，常因歸類的不同，在同一資料上，可以得出不同的結論。

關於第二點設資料的項數甚少，而且散漫，并無集中傾向時，衆數不適用。因爲他是由一部份的項數求出，例如某級學生成績有 5 人得100分，在次數爲最多，他的衆數當然是100分。假如我們就此說，這級學生的平均成績是 100 分，豈非笑話

。就這點來說，衆數不及中位數，（須注意中位數的量數必須依次排列）算術平均數，倒數平均數，幾何平均數等三者，將全體各項都包括在內，尤無此弊。但是在另一方面，算術平均數等，因為一切項數都包括在內，就受絕端項的影響。設若極端項有錯誤遺漏，平均數就不能精確決定，反不如二種地位平均數。中位數有時若兩端項數稍有增減，還要略受影響，至於衆數，簡直沒有絲毫影響。在三種計算平均數中，就極端的影響論，自然算術平均數為最大，其次是幾何平均數，倒數平均數又次之。

就簡明易曉來講，算術平均數中位數衆數都具此特長，尤其是衆數，他是表示最普通的狀態，所根據的是大多數的事實，有時竟有稱為「平常人的平均數」的雅號。幾何平均數和倒數平均數，在這點上似乎較遜一籌。

就計算難易說，算術平均數只要稍有算學常識的人，就能勝任。中位數衆數雖不及算術平均之簡易，但亦不難。倒數平均數幾何平均數二者，雖則因近代統計技術進步，有特別演算法可以利用，比從前似乎簡易些，但是比較他種平均數，還是麻煩。至於理論的衆數，更是費事。

對於穩定性一點，要算算術平均數為最，中位數次之。

要適於更進一步代數上的計算起見，近似衆數和中位數就落選了。因為二者的本身，是以地位而決定，對於適合代數計

算，本身就不很健全。

上面所討論的結果，似乎算術平均數比較任何其他平均數都勝。誠然，就一般的目的講，算術平均數很爲適宜，但是在特殊情形之下，雖則有些平均數，在有幾點原則上失敗了，仍然不失爲一個有用的平均數。我們不可固執成見，因爲他一二點瑕疵，就將他犧牲了。例如統計資料不能度量者，如心理狀態等，中位數就很適用；再如每小時速率，或每元若干計等同類的事實，非用倒數平均數不可。幾何平均數，亦有特長，要計算含有比率增進或成幾何級數形式的事實，例如人口增加率等，非他莫屬。在算指數時，他還有更多的優點。關於各種平均數特殊的用途，已散見上列各節中，不多贅。

各種平均數的關係

從上述五種平均數，我們可以發現他們間的關係，有下列幾點。

1. 在對稱次數分配中，算術平均數中位數與衆數三者合而爲一。

2. 在失稱不甚之分配中，中位數與算術平均數和衆數的中間，約等於從前者到後者的距離之$\frac{1}{3}$。因此這種次數分配有下列近似的關係。

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

3. 任何量數的算術平均數，必大於他的幾何平均數。

4. 任何量數的幾何平均數，必大於他的倒數平均數。

這兩條法則，惟一的例外，就是一數列的各量數，各各相等的時候。這樣，M,G,H 三者的數值都相等。

5. 任何二數的幾何平均數，等於這二數的算術平均數和倒數平均數的幾何平均數。例如2與8二數的H是等於 $3\frac{1}{5}$，G是4，M是5，但是4就是$3\frac{1}{5}$同5的幾何平均數。但在包括二項以上的數列，這種關係，就不存在了。

6. 統計資料的離差，如遵依算學定律時，衆數與中位數，常與算術平均數相近。如遵依幾何定律，衆數與中位數常與幾何平均數相近。

第 二 章

離差 (Dispersion or Variation)

意義　上章所論的平均數，是表示集中趨勢的量數，可說是量表上足以代表全部分配的一點。例如研究紗廠工人工資而得一每月三十元的平均數。就大體講，工資固然是三十；可是其中有的比三十元多，有的比三十元少。要知道詳細情形，還當研究工資之多於平均數者，與小於平均數者，他們的分配如何，這就該研究其離差。所以說，離差是表示集中量數前後各量數分布的情形，簡單地說，就是表示離中的程度。這種離中程度的測量，在統計學的術語上稱爲離差測量。

平均數表示集中趨勢的意義亦隨離差的小大而深淺，離差數小，則平均數愈合代表一切之資格。離差愈大，則平均數代表之資格，愈不稱職。所以要詳細說明一組或比數數組分配的現象。平均數之外，還要計算其離差，例如研究社會財富，不獨應注意每人的平均進款，尤應考察其進款還是平均分配於組成社會全體的各分子，還是集中於少數人之手。再如，研究人生統計者，不獨要明瞭各城傷寒病的平均死亡率，卽各城差異之程度。亦應注意。教育統計中，考察二級學生的程度，不但要知道他們的平均程度，也應該知道其差異之情形如何。

絕對離差與相對離差(Absolute Variability and Relative Variability) 表示離差的單位，有用原有材料之測量單位，有用抽象的數目字。第一種稱絕對離差，第二種稱相對離差，或離差係數。要說明一個統計數列的離差，絕對離差已夠應用。惟如有兩個數列而欲比較其差異的大小，若僅就各數列的絕對離差相比較，事實上常受單位不同的影響。而且假使單位相同，又每因計算離差所根據的平均數懸殊太大，比較就感困難，所以最好用相對離差。

離差的種類　離差測量種類很多，最通用者有全距，四分位差，平均差，標準差及均互差等五種。

全距 Range

全距是表示離差最簡單的一種方法，同時也是最粗疏的一種，他是最大量數與最小量數間的絕對差數。就次數數列來講，他是最小一組的下限，與最大一組的上限間的距離，表21中資料的全距是：

$$Rg=7.99-2.00=5.99$$ = max − min

全距僅僅根據二絕端項的數值而決定，對於中間量數次數分配的情形，竟漠然不顧。倘其絕端項有一二項增減，即將使全距之值大變。所以他的數值是很不穩定的。

全距在表示金價證券等市場價格的變動時，尙見應用，至

於稍爲精密之統計工作中，即覺不甚適當。

四分位差 (Quartile Deviation; Q.D.)

(附十分位差等)

與全距相似，同取二項間距離的遠近，以定離差的大小者有四分位差十分位差等。關於四分位數，已於前章略述，是分全體量數爲相等的四分。就 Q_1 Q_3 之距離，亦足表示次數分配的離差程度。距離愈近，則分配愈集中，反之，距離愈遠，則集中愈弛。其公式如下：

$$Q.D.=(Q_3-Md)+(Md-Q_1)=\frac{Q_3-Q_1}{2}$$

將表 21 的事實代入公式，即得

$$Q.D.=\frac{5.16-3.73}{2}=\frac{1.43}{2}=.715$$

四分位差係數的公式，是

$$V.=\frac{\frac{Q_3-Q_1}{2}\times 100}{\frac{Q_3+Q_1}{2}}=\frac{(Q_3-Q_1)\times 100}{Q_3+Q_1}$$

與四分位差同樣的理論和方法，我們可以用十分位差及百分位差等，來量離差，Kelley 主張用第十百分位數 P_{10} 與第九十百分位數 P_{90} 間之距離，他的公式是 $D_p=P_{90}-P_{10}$ (註)

四分位差及其係數的特質　四分位差的特長，就是簡單易

(註)　請參 Kelley's Text P.75

算，明白易曉。不過他是根據Q_1 Q_3二者的地位而決定，並非從特定的平均數而算出，有時幾種完全不同的次數分配，或許他們的四分位數，偶爾相合，就說他們的離差是一樣，豈非大謬。對於Q_1與Q_3距離以外的量數，不生影響，所以欲以極端數加權，四分位差就完全無用武之地。四分位差同全距一樣，都不能再作數學上的處置，亦一憾事也。

平均差(Average Deviation; A.D.)

以上二法，嚴格地說，并非是離差，不過用間接的方法，來觀察離差的程度。平均差和標準差，才是眞正根據各量數而得的離差。

平均差乃各量數與其平均數(M,Md或Mo)的絕對差的算術平均數。在差數相加時，可不問他的正負號。因爲假若要按正負號來總計各差數，則從平均數所求得的各差數，他們正負總和，或適相抵銷等於零，或者爲數甚微，我們就不易得一正確的差數。因此，計算平均差當取各絕對差之總和而除以次數之總數。惟各絕對差的計算，可以根據衆數、中位數、或算術平均數。但由中位數算出之平均差，數值最小，能符合所謂最小差數之原則；故論理當取中位數，不過亦有從算術平均數或衆數計算者。

計算平均差之公式：

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

式中 d 爲各量數與平均數的絕對差數，其計算如下表。

表25　平均差之計算

x	f	d
2	1	2
3	1	1
4	1	0
5	1	1
6	1	2
	5	6

中位數＝4

$$A.D.=\frac{\Sigma d}{N}=\frac{6}{5}=1.2$$

假如統計資料，已經形成次數分配，求平均差的公式，當爲

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

式中 fd ，是次數與差數相乘的積數，d 是每組中值對平均數的差數。在算平均數時，我們假定每組的中値，可以代表每組各項，現在算平均差，亦當有同樣的假定。表26是計算平均差的一個例子，先求出總次數 202 ，中位數 53.9，再求出各組中值對此中位數的差數，得21.4, 16.4,……等，將每個差數與各相當的次數相乘，得 fd 欄各數，將 fd 欄各數，不問正負號總加之，得差數的總和1068.8 ，然後以總次數除之 $\frac{1068.8}{202}=5.34$，結果就是平均差。

表26　平均差之計算

上海市立小學五年級女生T智力分數(18年度)

分數 X (1)	中數 m (2)	次數 f (3)	差數(Md＝53.9) d (4)	fd (5)
30—34.9	32.5	1	21.4	21.4
35—39.9	37.5	4	16.4	65.6
40—44.9	42.5	7	11.4	79.8
45—49.9	47.5	35	6.4	224.0
50—54.9	52.5	69	1.4	96.6
55—59.9	57.5	53	3.6	190.8
60—64.9	62.5	24	8.6	206.4
65—69.9	67.5	7	13.6	95.2
70—74.9	72.5	1	18.6	18.6
75—79.9	77.5	1	23.6	23.6
總計		202		1068.8

$$\text{A.D.}=\frac{1068.8}{202}=5.34$$

用上述方法來求平均差，每感繁複，我們可以利用簡捷法來補救，簡捷公式爲$\text{A.D.}=\dfrac{\Sigma fd'+c(\text{N}_s-\text{N}_l)}{\text{N}}$。如次數數列之組距爲等距的，更有再簡之方法，其法先以平均數所在這組的中值作假定平均數，以組距爲單位，然後改正之，以資還元。他所根據的理論，與計算算術平均數簡法的理論相似，不再復述。他的公式是：

$$\text{A.D.}=\frac{\Sigma(fd')+c'(\text{N}_s-\text{N}_l)}{\text{N}}\times i$$

式中d'爲各組對假定平均數的差數。

c'是校正數，就是眞正平均數與假設平均數之差，而以組

距爲單位者，

N_l 是比眞正平均數爲大的量數，

N_s 是比眞正平均數爲小的量數，

試以表 27 爲例說明之（須注意在本例中所有差數，是根據中位數算出）。

表27　平均差之計算

各國在華紗廠之投資數額（1930）

每廠資本數 (100元) (1)	廠數 (2)	d' (3)	fd' (4)	計算
0—50	17	4	68	$Md=217$ $Md'=225$ $c'=\frac{217-225}{50}=-.16$ $A.D.=\frac{\Sigma fd'+c'(N_s-N_l)}{N}$ $=\frac{324+[-16(53-67)]}{120}$ $=2.7195$組距單位 $\therefore A.D.=2.7195\times 50$ $=135.975$原有單位
50—100	15	3	45	
100—150	13	2	26	
150—200	8	1	8	
200—250	22	0	0	
250—300	17	1	17	
300—350	5	2	10	
350—400	2	3	6	
400—450	5	4	20	
450—500	4	5	20	
500—550	1	6	6	
550—600	3	7	21	
600—650	3	8	24	
650—700	4	9	36	
700—750	0	10	0	
750—800	0	11	0	
800—850	0	12	0	
850—900	0	13	0	
900—950	0	14	0	
950—1000	0	16	0	
1000—1050	0	15	0	
1050—1100	1	17	17	
總數	120		324	

計算步驟

1. 先求出總次數和中位數。
2. 因眞正中位數在 200—250 組距間，所以就以這組的中值(225,)當作假定中位數。
3. 以組距爲單位，假定各組量數相差爲一，求假定中位數與各組的差數，得 d' 欄各數，再以次數乘之，得 fd' 欄各數，然後求出絕對差的總和，得 $\Sigma fd'=324$
4. 將眞正中位數，減以假定中位數，得校正數 c，化爲組距單位得 $c'=\dfrac{217-225}{50}=-.16$
5. 全體各項之中，對於用假定中位數而得之離差，有大於對眞正平均數的離差，是爲多算；有小於對眞正中位數的離差，是爲少算；多算各項爲 N_l 得 67，少算各項爲 N_s 得 53，分別算清後，將二者相減，再以校正數乘之，得 $c'(N_s-N_l)=-.16(53-67)=2.24$，這才是應行改正的總數。
6. 將總改正數加於總差數之上得 $324+2.24=326.24$
7. 將總次數除上面所得結果，得平均差 2.7195；不過這結果還是組距的單位，所以應當乘以組距，才能復原得 135.975，爲原有單位的平均差。

平均差係數　將絕對平均差，除以相當之平均數，就可得一係數。

平均差係數之公式：

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

或 $$V_{A.D.}=\frac{A.D.}{Md,M\text{或}Mo}\times 100$$

Thorndike 氏則主張用平均數的方根，來除離差，而不主張直接用平均數去除離差。他的公式是：

$$V_{A.D.}=\frac{100_{A.D.}}{\sqrt{Md}}$$

式中乘以 100，用意在使之化爲百分數。上列二公式，所得結果，不相一致。據一般統計家之經驗，若欲比較兩組分配大概的差異，用第一式較爲適宜。還有一點，就是第一式的係數結果，較爲固定些。

平均差的特質　平均差計算容易，使人亦易於明瞭。他是各項對平均數的絕對差的平均數，所以是根據全體量數求出，與各量數均能發生影響。　平均差是依離差的大小與以權重。極端離差所得的權數，比小離差所得權數爲大；但是並不因此而失掉他們的比率，各權數還是成比例的。

他的短處，在於應用絕對差數，不便代數的處置。

據 King 氏的意見，平均差用以研究經濟問題，很爲合宜。假如要推算一國的個人財富分配，用這係數最好，因爲他對於最富的人，和最貧的人，都注意到。

標準差（Standard deviation;）

平均差計算雖然便利，可是在數學原理上，有很大的弱點，因為他不問各項的差數為正為負，一一相加，標準差便沒有這種弊病。標準差是差數平方的平均數的方根，（Root mean spuare deviation），計算的時候，將一切離差，先一一自乘，則依代數原理，任何一數，不問正負，自乘以後，結果必為正數。但初既自乘，不得不開方以資還元。計算標準差，當以算術平均數為中心；因為要求精密的離差數，須使離差的平方和為最小。卽須符合所謂"最小二乘之原則"，要求最小的差數平方和，須從算術平均數來求離差。

算標準差的公式(1)　將未歸類的資料，計算標準差，可從上面的定義，直接得一公式如下。

$$\sigma=\sqrt{\frac{\Sigma d^2}{N}}\cdots\cdots(1)$$

試舉一簡單的例子來說明算法，

表23　標準差之計算

m	f	d	d^2	
3	1	−6	36	$M=9$
6	1	−3	9	$\sigma=\sqrt{\frac{90}{5}}$
9	1	0	0	$=\sqrt{18}$
12	1	+3	9	$=4.24$
15	1	+6	36	
			90	

公式(2)　設資料已經歸類而成次數表，他的公式，應稍加變換如下：

$$\sigma=\sqrt{\frac{\Sigma f d^2}{N}} \text{ 或 } \sigma^2=\frac{\Sigma f d^2}{N} \quad \cdots\cdots\cdots\cdots(2)$$

試舉一例以示其算法，

表29　標準差之計算

m	f	d	d^2	fd^2	
3	1	−6.3	39.7	39.7	$M=9.3$
6	2	−3.3	10.9	21.8	$\sigma=\sqrt{\frac{116.3}{11}}$
9	4	−0.3	.1	.4	$=\sqrt{10.6}$
12	3	+2.7	7.3	21.9	$=3.3$
15	1	+5.7	32.5	32.5	
	11			116.3	

簡捷法　用上列公式計算標準差，手續麻煩；因爲每次對平均數的差額，常常有小數，自乘的時候，位數既多且易錯誤；所以實際多用簡捷法。簡捷法的公式有二。

簡捷公式 1

$$\sigma=\sqrt{\frac{\Sigma f(d')^2}{N}-\left(\frac{\Sigma f d'}{N}\right)^2} \text{(註)} \quad \cdots\cdots\cdots\cdots\cdots\cdots(3)$$

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

因　$M=M'-C$

式中$\sqrt{\frac{\Sigma f(d')^2}{N}}$係對假定平均數而算出之標準差，通常以$\sqrt{s^2}$表之。依據數理，標準差之値，以從眞正算術平均數而計算者爲最小，故尙須從$\sqrt{\frac{\Sigma f(d')^2}{N}}$値中，減去一改正數。這改正數，就是$\sqrt{\left(\frac{\Sigma f d'}{N}\right)^2}$常以$\sqrt{c^2}$表之。用這公式來計算，亦可用組距爲單位，然後校正之。試以表30來作例，說明這公式的運用。

(接前註)　$d'=d+c$

$$\Sigma f d' = \Sigma f(d+c)$$

$$\Sigma f(d')^2=\Sigma f(d+c)^2$$

$$\Sigma f(d')^2=\Sigma f d^2+2\Sigma f d\cdot\Sigma f c+Nc^2$$

但　$\Sigma d=0$

$$\therefore\quad \Sigma f(d')^2=\Sigma f d^2+Nc^2$$

$$\frac{\Sigma f(d')^2}{N}=\frac{\Sigma f d^2}{N}+c^2$$

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

但　$c^2=\left(\frac{\Sigma f d'}{N}\right)^2$

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

$$\therefore\quad \sigma^2=\frac{\Sigma f(d')^2}{N}-\left(\frac{\Sigma f d'}{N}\right)^2$$

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

表30　標準差之計算

我國郵政職工每月消費之狀況

消費元數 X (1)	職工人數 (2)	對假設原點之離差 d' (3)	fd' (4)	$f(d')^2$ (5)	$(d'+1)^2$ (6)	$f(d'+1)^2$ (7)
10—15	1	−5	−5	25	16	16
15—20	13	−4	−52	208	9	117
20—25	52	−3	−156	468	4	208
25—30	61	−2	−122	244	1	61
30—35	49	−1	−49	49	0	0
35—40	53	0	0	0	1	53
40—45	37	1	37	37	4	148
45—50	12	2	24	48	9	108
50—55	20	3	60	180	16	320
55—60	13	4	52	208	25	325
60—65	6	5	30	150	36	216
65—70	7	6	42	252	49	343
70—75	4	7	28	196	64	256
75—80	4	8	32	256	81	324
80—85	2	9	18	162	100	200
85—90	1	10	10	100	121	121
90—95	0	11	0	0	144	0
95—100	1	12	12	144	169	169
100—105	1	13	13	169	196	196
105—110	2	14	28	392	225	450
110—115	1	15	15	225	256	256
115—120	0	16	0	0	289	0
120—125	0	17	0	0	324	0
125—130	0	18	0	0	361	0
130—135	0	19	0	0	400	0
135—140	1	20	20	400	441	441
總　數	341		37	3913		4328

N＝341

組距＝5元

$$C\ (\text{組距單位})=\frac{37}{341}=.109$$

C^2(組距單位)＝.612

$$s^2(\text{組距單位})=\frac{\Sigma f(d')^2}{N}=\frac{3913}{341}=11.475$$

$$\sigma^2(\text{組距單位})=s^2-c^2=11.475-.012$$

$$=11.463$$

$$\sigma\ (\text{組距單位})=3.386$$

$$\sigma\ (\text{原有單位})=3.386\times 5=16.93$$

表中(6)(7)二欄，係爲勘誤而設。假使另外找一個原點，比原用之原點低一組，就是$(d+1)$，我們將這數值平方之，如欄(6)，再各乘以相當之次數如欄(7)，其總和與計算，標準差所得之各數值，有一定之關係。因爲

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

$$=\Sigma f(d')^2+2\Sigma fd'+\Sigma f$$

或 $\Sigma f(d'+1)^2=\Sigma f(d')^2+2\Sigma fd'+N$ ……………(4)

如將表中相當數值代入最後一式，卽可用以勘誤。此法爲Charlier所發明，(註)故稱 Charlier check

$$4328=3913+2(37)+341$$

$$=4328$$

註：Vorlesungen über Die grundzünge Der Mathematischen Statistik, C. V. L. Charlier, *19*.

計算步驟，可槪括如下：

1. 就次數分配中選取一組（最好要是接近眞正平均數的一組），而以他的組距中値，爲假定平均數。

2. 以組距爲單位，求各組與假定平均數所在組之差，得 d' 各數。

3. 將各組差數，與相當的次數相乘，得 fd' 欄各數。

4. 求 fd' 欄各數的代數和，得 $\Sigma fd'$，以總次數N除之，卽得改正數 c，將改正數平方之，得 $\left(\frac{\Sigma fd'}{N}\right)^2$，卽 c^2。

5. 將 fd' 欄各數，再乘以相當的 d'，得 fd'^2 欄各數，加之得 $\Sigma f(d')^2$。

6. 以總次數除 $\Sigma f(d')^2$ 得 $\frac{\Sigma f(d')^2}{N}$（卽 s^2）這就是由假定平均數求出的標準差的平方。

7. 旣得 $\frac{\Sigma f(d')^2}{N}$ 與 $\left(\frac{\Sigma fd'}{N}\right)^2$ 然後代入公式，開方卽得 σ 之値。不過這 σ，是以組距爲單位，當以組距乘之，以資還元。

簡捷公式 2　又可稱爲計算機適用之公式，因爲在計算機上計算，用此公式，最爲便捷。

$$\sigma=\sqrt{\frac{N\Sigma fX^2-(\Sigma fX)^2}{N^2}} \quad \cdots\cdots\cdots\cdots\cdots\cdots(5)$$

這個公式也可從 $\sigma=\sqrt{\frac{\Sigma fd^2}{N}}$ 演繹而來，其演法如下：

因 $d=X-M$

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

$$=\frac{\Sigma fX^2}{N}-2\left[\frac{\Sigma f(XM)}{N}\right]+\frac{N(M)^2}{N}$$

但 $M=\frac{\Sigma fX}{N}$

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

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

$$\therefore \sigma=\sqrt{\frac{N\cdot\Sigma fX^2-(\Sigma fX)^2}{N^2}}$$

如原有資料不爲次數分配之形式，則其公式爲：

$$\sigma=\sqrt{\frac{N\cdot\Sigma X^2-(\Sigma X)^2}{N^2}}\cdots\cdots\cdots\cdots(6)$$

茲舉例於表31. 及表32，計算步驟，可概括之如下：

1. 求總次數N，及總次數的平方N^2。
2. 求(ΣfX)與(ΣfX^2)，或(ΣX)與(ΣX^2)。
3. 求每個量數的平方(X^2)，再求各量數平方的總和。
4. 求$(\Sigma X^2\times N)$，或$(\Sigma fX^2\times N)$。
5. 求$\frac{N\Sigma X^2-(\Sigma X)^2}{N^2}$卽$\sigma^2$，或$\frac{N\Sigma fX^2-(\Sigma fX)^2}{N^2}$。
6. 將上項結果開方，卽得σ。

簡捷公式2，計算時便利處很多，特爲介紹，如下：

1. 沒有求算術平均數之必要，

2. 沒有把原來的量數，順序排列之必要，
3. 不必算差數 d 或 d'，而直接根據於原來的量數，
4. 假使X是整數，則除了末一步之外，所有的數目。都是整數，
5. 標準差之數，除了末一位外，隨便求到多少位，都是實數，不是近似數，
6. 可以直接在計算機上演算，
7. 省時間，
8. 如有 Barlow's table，算 X^2 更可由表中查得。

標準差係數　Pearson 氏求標準差係數（亦名變量係數）的公式，為

$$V\sigma=\frac{\sigma}{M}\text{或}\frac{\sigma}{M}\times100 \cdots\cdots\cdots\cdots\cdots\cdots(7)$$

茲用表30的事實為例，以求其離差係數。

$M=38.04$

$\sigma=16.93$

代入公式，得 $V\sigma=100\times\frac{16.93}{38.04}=44.5\%$

觀上例，可知這種係數，是求出標準差與算術平均數之百分比率也。

標準差的特點　標準差之計算，係根據全體的量數，所以最為公允。至於用自乘再開平方的方法，以取消負號，沒有不

合數學上邏輯的缺點。他對於標樣(Sample)變動的影響，亦很小。而且可以有切實的數量來表示，和適合於更進的數學處置。求 Pearson 相關係數時，標準差的計算方法，大有幫助。就標準差求得之離差，精確程度，較上述各方法，都要高些，所以 Yule 氏曾極力鼓吹用標準差。他說，除非有特別原因，測量離差，總當用標準差。

表31　簡捷公式之計算（次數資料）

衛生部職員年齡之分配(18年2月)

年齡 (1)	組距中點 X (2)	f (3)	fX (4)	fX^2 (5)
16—20	13	2	36	648
21—25	23	20	460	10580
26—30	28	34	952	26656
31—35	33	34	1122	37026
36—40	-38	22	836	31768
41—45	43	9	386	16641
46—40	48	5	240	11520
51—45	53	2	107	5618
總數		128	4139	140457

$N^2 = 16384$

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

$$= \sqrt{\frac{128(140457) - 4136)^2}{16384}}$$

$$= \sqrt{51.7} = 7.2$$

表32　簡捷公式之計算(分立資料)

測驗分數 X	X^2
6	36
15	225
88	7744
93	8649
43	1849
44	1936
74	5476
76	5776
56	3136
56	3136
58	3364
65	4225
18	324
12	144
97	9409
95	9025
86	7396
82	6744
74	5476
99	9801
1237	93971

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

$$=\sqrt{\frac{20(93971)-(1237)^2}{(20)^2}}$$

$$=\sqrt{\frac{1879420-1530169}{(20)^2}}$$

$$=\sqrt{\frac{349251}{400}}$$

$$=\sqrt{\frac{873.1275}{29.54}}$$

$$=29.54$$

求標準差之雜例　和討論平均數性質時一樣，我們來繼續討論從分組的標準差，如何以求全數列之標準差。在上章曾經提起：設全數列之平均數爲 M, 二分組之平均數爲 M_1, M_2, 而

$$N\cdot M=N_1\cdot M_1+N_2\cdot M,$$

N_1, N_2 爲各分組之次數，則全體次數，當爲 $N=N_1+N_2$。同法，全數列之 σ，亦可從其各分組之 σ 求出之。

設　$M_1-M=d_1$

$M_1-M=d_2$

故分組各量數，對平均數M之均方差 Mean square Deviation，由公式 $\sigma^2=s^2-c^2$，應為$(\sigma_1{}^2+c_1{}^2)$及$(\sigma_2{}^2+c_2{}^2)$，故全數列之標準差，為

$$N\sigma^2=N_1(\sigma_1{}^2+c_1{}^2)+N_2(\sigma_2{}^2+c_2{}^2)\cdot\cdots\cdots\cdots\cdots(8)$$

設二分組之次數相等，而平均數亦在同一點上，對於此種特例之公式，更為簡單。即

$$\sigma^2=\tfrac{1}{2}(\sigma_1{}^2+\sigma_2{}^2)\cdot\cdots\cdots\cdots\cdots(9)$$

是即全數列之 σ 之平方，等於各分組之 σ 之平方之算術平均數。設將公式 $N\sigma^2=N_1(\sigma_1{}^2+c_1{}^2)+N_2(\sigma_2{}^2+c_2{}^2)$寫得普遍些，便是

$$N\sigma^2=\Sigma(N_m\cdot\sigma_m{}^2)+\Sigma(N_m\cdot\sigma_m{}^2)\cdot\cdots\cdots\cdots\cdots(10)$$

計算各組之 σ，如用同一之假設原點，可用下列公式為勘誤之用，

$$\Sigma(fd'^2)=\Sigma(f_1d'_1{}^2)+\Sigma(f_2d'_2{}^2)+\cdots\cdots+\Sigma(f_rd'_r{}^2)$$
$$\cdots\cdots\cdots\cdots\cdots\cdots(11)$$

以上討論，由各分組以推算全數列之範圍，限於選用同一之假設平均數以計算之，現在進而討論各組所用之假設平均數，各不相同，當如何以求全數列之 σ，便是各組的 d' 不相等。我們用 d'_1代表 A 數列之 d'，用 d'_2 代表 B 數列之 d'，在這情境之下，我們應先把不同之假設平均數，換算為同一假設平均數，然後應用上面所示之公式，推演之。

設A數列之假設平均數，與B數列假設平均數之差為h，則：

$$d'_2=d'_1+h$$

由A數列之假設原點，改為B數列相同之假設原點，便是將d'_2改為d'_1+h。我們推演等式如下：

$$\begin{aligned}\Sigma f_1(d'_1+h)^2&=\Sigma f_1(d'_1)^2+2\Sigma f_1d'_1h+\Sigma f_1h^2\\&=\Sigma f_1(d'_1)^2+2h\Sigma f_1d'_1+N_1h^3\\&=\Sigma f_1(d'_1)^2+h(2\Sigma f_1d'_1+N_1h)\cdots\ (12)\end{aligned}$$

此公式之第一項$\Sigma f_1(d'_1)^2$為已知數，後一段$h(2\Sigma f+N_1h)$便是改換假設原點之改正數。

求從1,2,3,……N個自然數之標準差，更有一簡單之公式。因1－N之平均數，為$\frac{N+1}{2}$，而1－N自然數之平方之和為：

$$\frac{N(N+1)(2N+1)}{6}$$

故標準差可以下列方程式求之。

$$\sigma^2=\frac{1}{6}(N+1)(2N+1)-\frac{1}{4}(N+1)^2,$$

約簡之為

$$\sigma^2=\frac{1}{12}(N^2-1)\ \cdot\ \cdots\cdots\cdots\cdots\cdots\cdots\ (13)$$

此公式求標準差之方法，最適用於某種資料之以等級計者，故有N個個體，即有N等級。

均互差 (Mean difference)

均互差，爲意大利統計家借尼教授 Prof. Corrado Gini 所首倡。(參Variabilità e Muabilità; Fascicolo I; Bologna, 1912, pp. 19 seq.) 其法乃將各項之數値，一一相互比較，而得其各項間相互之差數，此差數再平均之，即得所謂均互差，符號爲 g。前述諸離差測量祇將各量數與某一標準數值作比較而求其差數，本方法異於是。本方法之特點，則在將各項之數量互相比較，故亦有其特殊之應用。例如：研究財富之分配，吾人不但當知其富有者或貧窮者與全體平均數之差異如何；更當研究其最窮與最富之間之差異，及各部分相互間之差異。類此問題，如土地分配，工資分配，所得分配等，研究其差異，固非前述各種離差測量所能勝任愉快也。本方法之應用，在今日尚未普遍，這也許是計算方面較爲間接及麻煩之故，但是他的觀念之簡單，和合乎邏輯二點，則不容埋沒。

計算公式1. 未經列成次數分配之計算——設 $X_1, X_2, X_3, \cdots\cdots X_{n-1}, X_n$ 爲由小而大之順序排列之量數，今將各項互相比較，即當求 X_n 與 $X_{n-1}, X_{n-2}, X_{n-3} \cdots\cdots X_2, X_1$，之差數，又 X_{n-1} 與 $X_{n-2}, X_{n-3}, \cdots\cdots X_2, X_1$，之差數，……，直至 X_2 與 X_1 之差數爲止；是故有 N 個量數之數列，即可得 $\frac{N(N-1)}{2}$ 相差數。

假定S爲各相差數之總和，則計算均互差之公式，可表之如下。

$$g=\frac{S}{\frac{1}{2}N(N-1)}=\frac{2}{N(N-1)}S\cdots\cdots\cdots\cdots\cdots\cdots(1)$$

觀此公式，欲求 g 當先知其差數之總和 S，S 之計算爲：

$$\begin{aligned}S=&(X_n-X_1)+(X_n-X_2)+\cdots\cdots+(X_n-X_{n-2})+(X_n-X_{n-1})\\&+(X_{n-1}-X_1)+(X_{n-1}-X_2)+\cdots\cdots+(X_{n-1}-X_{n-2})\\&\cdots\cdots\cdots\cdots\cdots\cdots\\&\cdots\cdots\cdots\cdots\cdots\cdots\\&+(X_3-X_1)+(X_3-X_2)\\&+(X_2-X_1)\end{aligned}$$

若將此式右項第一行第一列各差數相加，得$(N-1)(X_n-X_1)$，復將其次行次列之各差數相加，得$(N-3)(X_{n-1}-X_2)$，…………，依次行之，即得下式：

$$\begin{aligned}S=&(N-1)X_n+(N-3)X_{n-1}+(N-5)X_{n-2}+\cdots\cdots\\&+(1-N)X_1+(3-N)X_2+(5-N)X_3+\cdots\cdots\\=&(N-1)(X_n-X_1)+(N-3)(X_{n-1}-X_2)+(N-5)(X_{n-2}-X_3)+\cdots\cdots+\cdots\cdots\end{aligned}$$

計算方法舉例於表 33 (錄自Bowley：Elements of Statistics P. 111)

表33　1902年英格蘭及威爾士每人口10,000按週死亡率

項 次	x_a	項 次	x_b	x_a-x_b	λ	$\lambda(x_a-x_b)$
52	244	1	136	108	51	5508
51	233	2	139	94	49	4606
50	226	3	141	85	47	3995
49	209	4	143	66	45	2970
48	206	5	144	62	43	2666
47	201	6	145	56	41	2296
46	196	7	149	47	39	1833
45	196	8	150	46	37	1702
44	196	9	151	45	35	1575
43	191	10	152	39	33	1287
42	183	11	154	29	31	899
41	182	12	155	27	29	783
40	182	13	159	23	27	621
39	181	14	160	21	25	525
38	179	15	164	15	23	345
37	177	16	165	12	21	252
36	177	17	166	11	19	209
35	177	18	166	11	17	187
34	176	19	167	9	15	135
33	176	20	169	7	13	91
32	176	21	169	7	11	77
31	174	22	169	5	9	45
30	174	23	170	4	7	28
29	174	24	170	4	5	20
28	173	25	172	1	3	3
27	133	26	172	1	1	1
總 數						32659

由上表　即得 S 之數值等於 32,659，則令 N=52，故其均互差 g 即爲

$$g=\frac{2}{52\times 51}\times 32,659=\frac{32,659}{1326}=24\cdot 63$$

表中之 N 爲偶數，設爲奇數時，則中央一項，可以 0 乘之。

2. 次數數列之計算　設爲次數數列之資料，則 f_1+f_2+

$\cdots\cdots + f_i + \cdots\cdots + f_k = N$，計算公式爲

$$g = \frac{2}{N(N-1)} S$$

依定義 S 之計算，當爲：

$$S = f_k[f_1(X_k - X_1) + f_2(X_k - X_2) + \cdots\cdots + f_i(X_k - X_i) + f_{i+1} + (X_k - X_{i+1}) + \cdots\cdots\cdots + f_{k-2}(X_k - X_{k-2}) + f_{k-1}(X_k - X_{k-1})]$$

$$+ f_{k-1}[f_1(X_{k-1} - X_1) + f_2(X_{k-1} - X_2) + \cdots\cdots + f_i(X_{k-1} - X_i) + f_{i+1}(X_{k-1} - X_{i+1}) + \cdots\cdots + f_{k-2}(X_{k-1} - X_{k-2})]$$

$\cdots\cdots\cdots\cdots$

$\cdots\cdots\cdots\cdots$

$$+ f_{i+1}[f_1(X_{i+1} - X_1) + f_2(X_{i+1} - X_2) + \cdots\cdots\cdots + f_i(X_{i+1} - X_i)]$$

$$+ f_i[f_1(X_i - X_1) + f_2(X_i - X_2) + \cdots\cdots + f_{i-1}(X_i - X_{i-1})]$$

$\cdots\cdots\cdots\cdots$

$\cdots\cdots\cdots\cdots$

$$+ f_3[f_1(X_3 - X_1) + f_2(X_3 - X_2)]$$

$$+ f_2[f_1(X_2 - X_1)]$$

依此計算太覺繁複，可變更之爲：

$$S = f_k X_k[f_1 + f_2 + \cdots\cdots + f_{k-1}] + f_{k-1} X_{k-1}[f_1 + f_2$$

$$+\cdots\cdots+f_{k-2}-f_k]+\cdots\cdots+\cdots\cdots+f_{i+1}X_{i+1}[f_1+f_2+\cdots\cdots+f_i-f_{i+2}-f_{i+3}-\cdots\cdots-f_{k-1}-f_k]+f_iX_i[f_1+f_2+\cdots\cdots f_{i-1}-f_{i+1}-f_{i+2}\cdots\cdots-f_{k-1}-f_k]+\cdots\cdots+\cdots\cdots+f_3-X_3[f_1+f_2-f_4-f_5\cdots\cdots-f_k]+f_2X_2[f_1-f_3-f_4-\cdots\cdots-f_k]+f_1X_1[-f_2-f_3\cdots\cdots-f_k]$$

再變之，得：

$$S=f_kX_k(N-f_k)+f_{k-1}X_{k-1}(N-2f_k-f_{k-1})+f_{k-2}X_{k-2}(N-2f_k-2f_{k-1}-f_{k-2})+\cdots\cdots+f_{i+1}X_{i+1}(N-2f_k-2f_{k-1}-\cdots\cdots-2f_{i+2}-f_{i+1})-f_1X_1(N-f_1)-f_2X_2(N-2f_1-f_2)-f_3X_3(N-2f_1-2f_2-f_3)-\cdots\cdots-f_iX_i(N-2f_1-2f_2-\cdots\cdots-2f_{i-1}-f_i)$$

因此，卽得均互差之公式如下：

$$g=\frac{2}{N(N-1)}[f_kX_k(N-f_k)+f_{k-1}X_{k-1}(N-2f_k-f_{k-1})+f_{k-2}X_{k-3}(N-2f_k-2f_{k-1}-f_{k-2})+\cdots\cdots+f_{i+1}X_{i+2}(N-2f_k-2f_{k-1}-\cdots\cdots-2f_{i+2}-f_{i+1})-f_1X_1(N-f_1)-f_2X_2(N-2f_1-f_2)-f_3X_3(N-2f_1-2f_2-f_3)-\cdots\cdots-f_iX_i(N-2f_1-2f_2-\cdots\cdots-2f_{i-1}-f_i)]\cdots\cdots\cdots\cdots\text{公式A}$$

玆以無錫工人家庭人口年齡分配之資料，示例於表 34。

表34　無錫工人家庭人口年齡統計表(註)

年齡	X	f	fX	λ	λfX	
					−	+
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1—10	5.5	479	2634.5	−1760	4.636.720	
11—20	15.5	444	6882 0	− 837	5.760.234	
21—30	25.5	392	9996.0	− 1	9.996	
31—40	35.5	415	14732.5	806		11.874.395
41—50	45.5	203	9236.5	1424		13.152.776
51—60	55.5	193	10711.5	1820		19.494.930
61—70	65.5	76	4978.0	2089		10.399.042
71—80	75.5	30	2265.0	2195		4.971.675
81—90	85.5	7	598.5	2232		1.335.852
總計		2239			10.106.950	61.228.670

$$g=\frac{2}{N(N-1)}S$$

$$=\frac{61228670-10406950}{2505441}$$

$$=20.3$$

均互差之計算，亦可間接求得之，其法先求各數量與中位數之絕對差數 d，不計負號，而均互差卽得變成一 d 之函數。

$$d_k=X_k-Md\ d_{k-1}=X_{k-1}-Md;\cdots\cdots\cdots d_{i+1}=X_{i+1}-Md.$$

$$d_i=Md-X_i;\cdots\cdots\cdots d_2=Md-X_2;\ d_1=Md-X_1.$$

而

$$X_k-X_1=d_x+d_1;X_{k-1}-X_2=d_{x-1}+d_2;\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

(註)　此表計算轉錄褚一飛論文

故上述之定義公式，可變爲：

$$S=f_k[f_1(d_k+d_1)+f_2(d_k+d_2)+\cdots\cdots+f_i(d_k+d_i)+f_{i+1}(d_k-d_{i+1})+\cdots\cdots+f_{k-2}(d_k-d_{k-2})+f_{k-1}(d_k-d_{k-1})]$$

$$+f_{k-1}[f_1(d_{k-1}+d_1)+f_2(d_{k-1}+d_2)+\cdots\cdots+f_i(d_{k-1}+d_i)+f_{i+1}(d_{k-1}-d_{i+1})+\cdots\cdots+f_{k-2}(d_{k-1}-d_{k-2})]$$

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

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

$$+f_{i+1}[f_1(d_{i+1}+d_1)+f_2(d_{i+1}+d_2)+\cdots\cdots\cdots+f_i(d_{i+1}+d_i)]$$

$$+f_i[f_1(-d_i+d_1)+f_2(-d_i+d_2)+\cdots\cdots+f_{i-1}(-d_i+d_{i-1})$$

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

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

$$+f_3[f_1(-d_3+d_1)+f_2(-d_3+d_2)]$$

$$+f_2[f_1(-d_2+d_1)$$

上式可變爲：

$$S=f_kd_k(f_1+f_2+\cdots\cdots+f_{k-1})+f_{k-1}d_{k-1}(f_1+f_2+\cdots\cdots+f_{k-2}-f_k)+\cdots\cdots+\cdots\cdots+f_{i+1}d_{i+1}(f_1+f_2+f_3+\cdots\cdots+f_i-f_{i+2}-f_{i+3}-\cdots\cdots-f_k)$$

$$+f_id_i(-f_1-f_2-f_{i-1}+f_{i+1}+f_{i+2}+\cdots\cdots+f_k)+$$

$\cdots\cdots + \cdots\cdots + f_3 d_3(-f_1 - f_2 + f_4 + f_5 + \cdots\cdots +$

$f_k) + f_2 d_2(-f_1 + f_3 + f_4 + \cdots\cdots + f_k) + f_1 d_1(f_2 +$

$f_3 + \cdots\cdots + f_k)$

再變之爲：

$$S = f_k d_k(N - f_k) + f_{k-1} d_{k-1}(N - 2f_k - f_{k-1}) + f_{k-2} d_{k-2}(N - 2f_k - 2f_{k-1} - f_{k-2}) + \cdots\cdots + f_{i+1} d_{i+1}(N - 2f_k - 2f_{k-1} - \cdots\cdots - 2f_{i+2} - f_{i+1}) + f_1 d_1(N - f_1) + f_2 d_2 (N - 2f_1 - f_2) + f_3 d_3(N - 2f_1 - 2f_2 - f_3) + \cdots\cdots + f_i d_i(N - 2f_1 - \cdots\cdots - 2f_{i-1} - f_i) \cdot \cdots\cdots \text{公式 B}$$

特舉一例於表35，以示運算之一般。

表35　無錫工人家庭人口年齡統計表　(註見135頁)

年齡 Col. 1	x Col. 2	f Col. 3	d Col. 4	fd Col. 5	λ Col. 6	λfd Col. 7
1—10	5.5	479	20	9580	—1760	16860800
11—20	15.5	444	10	4440	— 837	3716280
21—30	25.5	392	0	0	— 1	0
31—40	35.5	415	10	4150	806	3344900
41—50	45.5	203	20	4060	1424	5781440
51—60	55.5	193	30	5790	1820	10537800
61—70	65.5	76	40	3040	2089	6350560
71—80	75.5	30	50	1500	2195	3292500
81—90	85.5	7	60	420	2232	937440
總計		2239		33080		50821720

$$g = \frac{2}{N(N-1)} S$$

$$= \frac{2}{2239 \times 2238} \times 50821720 = \frac{50821720}{2505441} = 20,3$$

上述二種公式之取捨，全觀吾人之目的而定，設吾人同時欲求平均差，公式B較爲便利，因計算平均差時，亦須求fd之乘積，故可利用此已成之計算以推求均互差。設吾人同時計算算術平均數或標準差，取公式A較爲便利，因計算算術平均數時，亦須求fX之乘積，故亦可利用fX已成之計算，以推求均互差。

依照偕尼教授之意見，將所有差數之平方之平均數開方之，吾人可得下列之數量，

$$\sigma\sqrt{2\left(\frac{N}{N-1}\right)}$$

或與$\sigma\sqrt{2}$很相近。

均互差係數　欲求均互差之係數，可以平均數除之，至採用何種平均數，一般統計家之意見，以爲中位數或算術平均數，均可適用。

各種離差數之關係

各種離差的重要意義，前已略加討論，現在要將各種離差相互間的關係，再總結一下。

1. 全距是橫坐標上包括全體量數的距離。
2. 四分位差亦一距離問題。從第一四分位數與第三四分位數的中點，左右展開，各取一四分位差之距離，當包括全體

量數之一半。取九倍四分位差之距離，約含全體99%

3. 平均差在完全對稱，或稍失稱的次數分配中，約等於標準差之$\frac{4}{5}$。平均差七倍半之距離，約包括全體數量 99%，據經驗的結果，在常態分配，或稍微失稱的分配中，從算術平均數向左右伸展，各取一標準差的距離，約可包含全體量數之$\frac{2}{3}$（在常態分配中確含全體量數 68.26%）。若各取 2σ 之距離，約包括 95%。若各取 3σ 之距離，約含有 99%。

4. 各種離差數在完全對稱的分配中，據 Chaddock 的意見，可得下列的比率。

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

$$\sigma = 1.4825 \quad Q$$

$$\text{A.D.} = .7979 \quad \sigma$$

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

$$Q = .6745 \quad \sigma$$

$$Q = .8453 \quad \text{A.D.}$$

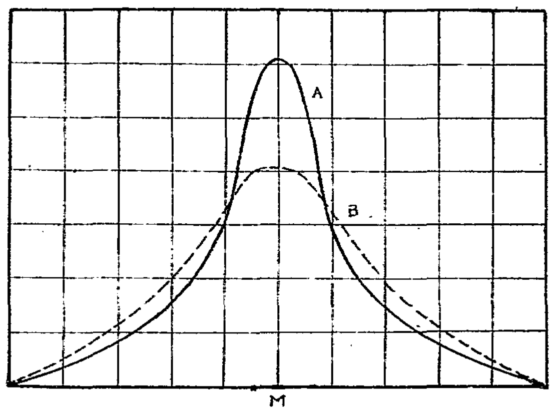
至於均互差，設各量數之距離爲相等時，g 恆大於 Md 約等於 Md $\times\frac{4}{3}$。

離差圖解

研究離差，有時可用圖解法，雖不能如離差數之給我們一個具體的數字，亦有相當之優點，略述於下。

I. 次數圖　次數圖有時亦能明示吾人以離差的概況。例如圖24，A,B,二線，代表二種分配，其平均之數值相同，然而看了圖，就很明白知道，A線所代表的分配離差較B線爲小

圖24　示平均數相等離差不等之二次數分配

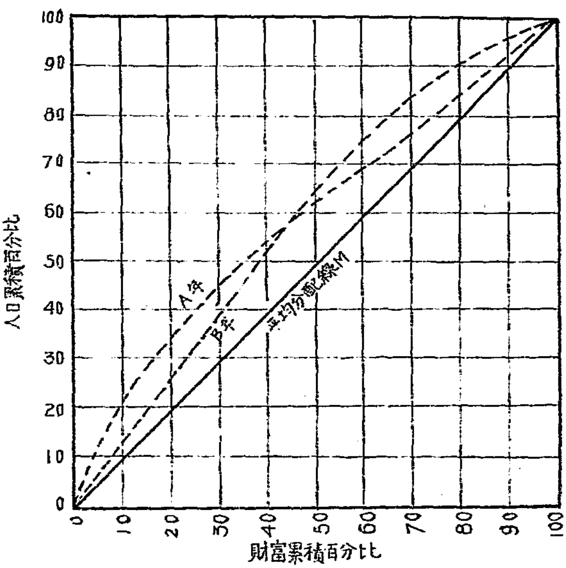


II. 羅倫曲線 Lorenz Curve　是羅倫博士 Dr. Max .O Lorenz 所發明，用以研究財富分配等問題，最爲適宜。他的長處，在乎能夠顯明各部分人口所得財富分配之量，不單是以一個平均數籠統地代表全體。圖25爲一羅倫曲線，假如財富的分配完全平均，可畫一直線 m 表之。可是實際上，人民之財富，不能完全平均，所作之線多如 A,B。凡曲線離相等分配直線愈近，則財富之分配愈平均。否則離直線愈遠，則表示財富集

中於少數富豪，而多數之人，多爲貧困者。

羅倫曲線對於土地分配，工資分配，及所得 Income 分配等問題之研究，亦極有用。

圖25　以羅倫線顯示A年與B年之財富分配

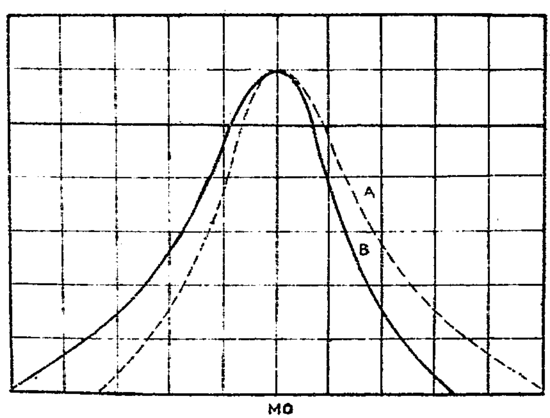


第三章

失稱 (Skewness)

失稱的意義　次數數列的第三種重要特質，是失稱。所謂失稱，是對稱的相對名詞。衆數兩邊的次數，假如相等，就是對稱。否則衆數的兩邊的次數不相等，稱爲失稱。在研究二組次數分配時，有時他們的平均數和離差數，或許相同，但在另一方面，他們平均數上下兩端的次數分配，或許完全不同，這時失稱測量，就甚爲重要。試以下圖爲例來說明。

圖26　衆數相同離差數相同而失稱測量不同之次數分配



這圖的二曲線的衆數相符，而他們的標準差，亦復相等，

但是平均數二邊的次數分布，則顯然不同，A 線是在衆數的上端(量數大於衆數)者逐漸低平 (Taper off)，而 B 線是在衆數的下端(量數小於衆數者)，逐漸低平。這種不同，祇要經過失稱測量，就可明明白白地顯示出來。失稱有正負之分，圖中線 B 的失稱是正失稱，線 A 是負失稱。(註：Pearson and Eldertón 之解釋適爲其相反)

社會現象的統計事實，很少有完全對稱，往往有很高度的失稱。例如財富或所得 (Income) 分配之簡單次數曲線，集中於一邊，衆數近於一極端之處。公債價格的次數曲線，及新婦年齡曲線，多是向左或向右傾於一方面的。其他分配爲 U 形曲線者，如圖27之死亡率曲線，分配之兩端，次數最多。盲人失明年齡之次數曲線，亦作 U 形。有時失稱之狀態成 J 形曲線，如圖22，乃一理想的 J 形曲線，表36,37之事實已差近J 形矣。不過統計界，對於失稱的應用，除了生物學家，其餘不多用。

表36　美國1887—1966年離婚案件

已結婚年數	離婚案件
合　計	900,584
5年以下	255,082
5—9	282,904
10—14	162,407
15—19	91,176
20—24	54,578
25—29	29,245
30—34	15,035
35—39	6,555
40—44	2,507
45—49	805
50以上	287

U.S. Statistical Abstract 1918 P.85

表37　美國六十七區徵稅費對總稅收之百分率分配

分百比	區　數
0— 2	29
2— 4	24
4— 6	4
6— 8	4
8—10	4
10—12	0
12—14	1
14—16	1

圖27　U形曲線　一九二〇年紐約人口各年齡死亡率分配

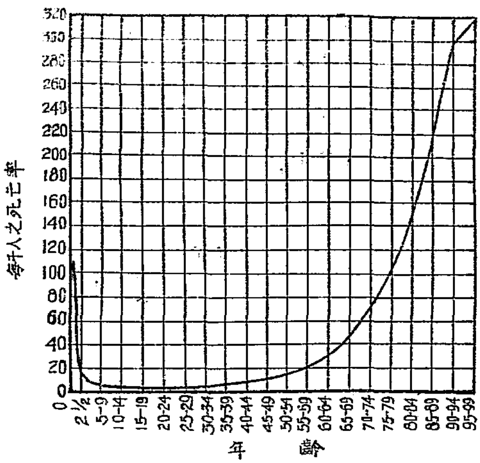
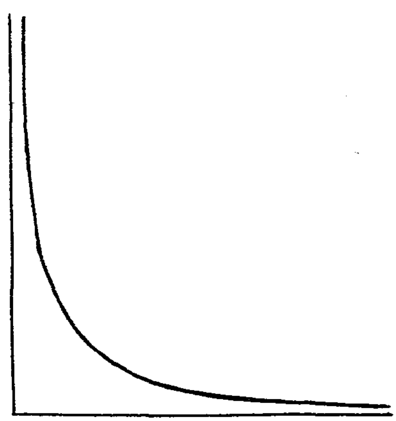


圖28　理想的J形次數分配(極度失稱)



測量失稱的方法　失稱測量，有三種。第一是批爾生 Pearson 測量，第二是尤爾 Yule 測量，第三是根據三次希望差（Third Moment）測量。失稱測量在平常應用時，各除以相當的離差數，而變為失稱係數。因為事實上，失稱的可能，完全依靠分配中的離差數，所以用離差數來做分母，以求係數，是很合邏輯的。

批爾生測量　是根據次數分配在對稱時，Mo, Md 和 M 三者數值相等，否則三者各不相同。失稱愈甚，三者數值相差亦愈遠，而 M 與 Mo 的距離為最大，根據這點，就可以測量失稱。他的公式是：

$$Sk = -(M-Mo) \quad \cdots\cdots\cdots\cdots\cdots\cdots(1)$$

如以 σ 除之，即得係數，

$$Sk = -\frac{M-Mo}{\sigma} \quad \cdots\cdots\cdots\cdots\cdots\cdots(2)$$

不過所用之離差數，不僅限於 σ，從乘數而算得之平均差，亦可應用。

上列公式，乃一理想的公式，因為理論衆數的計算太麻煩不易求得。但在失稱不甚的分配中，$Mo = M-3(M-Md)$，以之代入上式，即得

$$Sk = -\frac{3(M-Md)}{\sigma} \quad \cdots\cdots\cdots\cdots\cdots\cdots(3)$$

以表 17 之資料求其失稱係數，則得下列之數值。

$$Sk=-\frac{3(4.513-4.386)}{1.15}=-.33$$

尤爾測量　是根據四分位數的關係，因爲在次數分配對稱時，Q_1 與 Q_3 離中位數的距離適相等，設爲失稱，則不復相等。故(Q_3-Md)與$(Md-Q_1)$之相差，卽得爲一失稱測量。設以抽象的數值來表示，則須除以四分位差。下面就是他的公式：

$$Sk=-\frac{(Q_3-Md)-(Md-Q_1)}{Q.\ D.}$$

$$=\frac{-Q_1+Q_3-2Md}{Q.D.} \quad \cdots\cdots\cdots\cdots\cdots\cdots(4)$$

這公式求出之結果，最小的數值爲 0，卽當次數分配完全對稱之時，因爲在完全對稱時，$(Q_3-Md)=(Md-Q_1)$，則結果爲 0，其最大之數值，可假定不致超出 +2 或 −2，實際上少有過於 ±1 者。表 17 之資料，其失稱係數爲：

$$Sk=-\frac{3.73+5.16-2(4.39)}{.72}=-.15$$

批爾生係數，感應最靈，不過他的缺點，在理論衆數不易計算。設用 $3(M-Md)$ 之公式，亦限於失稱不甚的分配，設失稱太著，亦不適用。尤爾係數，簡單易算，而實際亦有相當精密。其缺點在失稱不甚之時，所得結果，不甚有效。再者，這種係數，把兩極端項的大小，漠視了，亦是遺憾。再精密的測量，則有應用三次希望差爲失稱係數。

應用三次希望差測量失稱係數　是根據三次希望差(Third Moment) 幷根據正數的立方爲正，負數立方爲負的道理。在分

配對稱之次數數列，對算術平均數的離差之立方和，必等於零，而在不對稱的次數數列，對算術平均數的離差之立方的總和，卽不等於零，三次希望差卽各項對算術平均數的離差之立方的平均數。三次希望差之絕對値愈小，次數數列之分配卽愈形對稱，否則，卽不對稱，故該三次希望差，可用以測量其失稱的程度，玆爲避免單位的影響，應以 σ^3 除之，其公式爲：

$$Sk=\frac{m_3}{\sigma^3}=\frac{1}{N}\Sigma f\left(\frac{d}{\sigma}\right)^3 \cdots\cdots\cdots\cdots\cdots(5)$$

但亦有用 $Sk=\frac{\sqrt[3]{m_3}}{\sigma}$ 者……………………………(6)

係數所用分母，不一定用 σ，不過因爲 σ 是很注意極端項，所以用爲分母，最爲相宜。就上式直接來計算，手續太繁，故可利用簡法，簡法的公式爲：

$$Sk=\frac{1}{\sigma}\sqrt[3]{\frac{\Sigma f d'^3}{N}-3c'\sigma^2-c'^3}^{(註)}\cdots\cdots\cdots\cdots(7)$$

其計算方法，請參表38之例。

註 $Sk=\frac{1}{\sigma}\cdot\sqrt[3]{\frac{\Sigma f d^3}{N}}$

$=\frac{1}{\sigma}\cdot\sqrt[3]{\frac{1}{N}\Sigma f(x-m_1')^3}$　　m_1'＝算術平均數 M

$=\frac{1}{\sigma}\cdot\sqrt[3]{\frac{1}{N}\Sigma f(x^3-3x^2m'_1+3xm'^2_1-m_1'^3)}$

$=\frac{1}{\sigma}\cdot\sqrt[3]{\frac{1}{N}(\Sigma fx^3-3m'_1\Sigma fx^2+3m'^2_1\Sigma fx-\Sigma fm_1'^3}$

表38　失稱測量之計算——用三次希望差

美國 Alabama *82* 家商業銀行之存款 Sept, *6*, *1921*,

存　款 (\$1,000) x	銀行數 f	d'	fd' −	fd' +	fd'^2	fd'^3 −	fd'^3 +
0—50	11	−3	33		99	−297	
50—100	19	−2	38		76	−152	
100—150	21	−1	21		21	−21	
150—200	9	0			0		
200—250	6	1		6	6		6
250—300	5	2		10	20		40
300—350	4	3		12	36		108
350—400	0	4		0	0		0
400—450	0	5		0	0		0
450—500	1	6		6	36		216
500—550	2	7		14	98		686
500—600	0	8		0	0		0
650—650	2	9		18	162		1458
600—700	1	10		10	100		1000
700—700	0	11		0	0		0
750—850	1	12		12	144		1728
合　計	82		−92	88		−470	5242
			−4			4772	

$c = -.05$(組距)

$\sigma = 3.12$(組距)

$$\mathrm{Sk} = \frac{1}{3.12}\sqrt[3]{\frac{4772}{82} - 3(-.05)(3.12)^2 - (-.05)^3} = -1.25$$

(接前註)

$$= \frac{1}{\sigma} \cdot \sqrt[3]{m'_3 - 3m'_1 m'_2 + 3m'^2_1 m'_1 - m'^3_1}$$

$$= \frac{1}{\sigma} \cdot \sqrt[3]{m'_3 - 3m'_1 m'_2 + 2m'^3_1}$$

因　$m'_2 = m_2 + m'^2_1$

而　$m_2 = \sigma^2$

故　$$= \frac{1}{\sigma} \cdot \sqrt[3]{m'_3 - 3m'_1 \sigma^2 - m'^3_1}$$

$$= \frac{1}{\sigma} \cdot \sqrt[3]{\frac{\Sigma f d'^3}{N} - 3c'\sigma^2 - c'^3}$$

第四章

相　關 (Correlation)

相關的意義

前數章所述如平均數離差數失稱等，都是從一種事實上研究其特性。但有時幾種事實，彼此互生關係。例如雨量與米穀之收穫，溫度與農作物的發育，他若棉花的產量和他的價格，一公司的營業總數和他的盈餘率等。或此長彼消，成此消彼長，或消長相共。凡此種種，皆有一定的關係，值得人們研究的。研究此類相互關係的方法，就是本章所要討論的。相關的意義，依 Bowley 氏的界說，譬如有二種現象的數量，其一種現象的各數量增加或減少時，其他一種現象各數量，亦隨之而增減，且當一種現象的各量數差異漸大時，其他一種現象的各量數的差度，亦隨之而漸大，這二種現象就有相關了。

相關的種類

相關可以分為兩類：正相關和負相關。假如兩種變量，同為消長，這種關係，就是正相關。假如兩種變量，性質相反，一種變量增加，他種反而減少；或一種減少，他種反而增加，這便是負相關。熱度與物質的漲大性，貧窮和犯罪數等，都是

正相關。物價和商品的供給，利率和公債價格等，都是負相關。有時兩種變量，各自增減，毫不相關，是謂無相關。

相關之程度

二變量可成正相關或負相關，但不必是絕對的相關。當甲變量增加之時，其乙變量並非必然的應當增加或減少，換句話說，就是甲乙兩變量的增減，不必個個相同(或相反)，但祇須多數相同(或相反)。如體高與體重，大概論之，體高者固比體低者較重，體高與體重，成正相關。但其間的相關，也祇是相對的，而非絕對的。

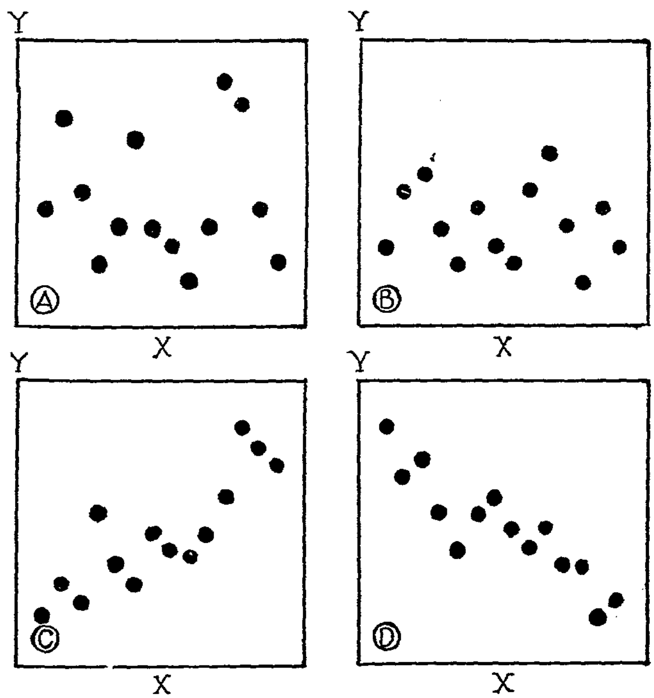
用分佈圖表示相關

兩種事實的各數量，兩兩相對者，稱爲兩相配對的變量。這些變量，可用分布圖來表示。因爲每一配對，就有一點，將這些點，都排列好，就可以看出二種變量，是否有相關的存在。假使有，到什麼程度。假如各點的分布，毫無規則且無體系的趨勢，如圖 29A，或各點聚沿於一水平線上，如圖 B，這種現象，是表示無相關。如各點聚沿於一線，作向上的趨勢，如圖C ，則爲正相關。如各點聚沿於一線，作向右而下的趨勢，如圖 D，則爲負相關。總之，各點分布愈向一線集中，則相關愈大，分布愈散漫，相關愈小，所以我們可以從分布圖而知相關的梗概。

相關又有直線曲線等形態，我們在分布圖上，用隨手配合法來配合一線，就可以看出事實的相關，究爲若何形態。

不過這種方法來表顯相關，未免太粗陋，我們不能不依賴數理的處置，作更精密的推算。

圖29　假設各縱行平均數之分配，以示相關之有無



相關係數

最常用而簡捷的相關測量，是 Pearsonian coefficient，此法爲英人批爾生所首倡，故名。這方法的應用，只限於兩個變量的關係，可以一直線配合者。換言之，即有一直線方程式的關係。他的最普通的公式，是

$$r=\frac{\Sigma(xy)}{N\sigma_x\sigma_y}=\frac{\Sigma(xy)}{\sqrt{\Sigma x^2}\sqrt{\Sigma y^2}} \qquad \text{即等於} \quad r=\frac{1}{N}\Sigma\left(\frac{x}{\sigma_x}\cdot\frac{y}{\sigma_y}\right)$$

這公式是以平均數爲原點，x 與 y 是對這原點的離差，σ_x σ_y 是對這原點而算出的標準差。現在試以表 39 爲例，以示其算法。

表39 相關係數之計算

田地	每畝施肥數量(百磅) X	每畝水果產量(千磅) Y	對 M_x 之離差 x	對 M_y 之離差 y	x^2	y^2	xy
A	60	10	−144.3	−5.4	20.736	29.16	779.2
B	100	14	−104.3	−1.4	10.816	1.96	146.0
C	150	12	− 54.3	−3.4	2.949	11.56	168.4
D	2[illegible]0	16	15.7	.6	.247	.36	9.4
E	240	19	35.7	3.6	1.236	12.96	128.5
F	310	17	105.7	1.6	11.236	2.56	169.1
G	350	20	145.7	4.6	21.346	21.16	670.2
					68.567	79.72	2087.0

$$r=\frac{\Sigma(xy)}{N\sigma_x\sigma_y}=\frac{\Sigma xy}{\sqrt{\Sigma x^2}\sqrt{\Sigma y^2}}=\frac{2087}{\sqrt{68,567}\sqrt{79.72}}$$

$$=\frac{2087}{261.85\times 8.94}=.89$$

批氏係數的特質，很可以推演而得。假如X變量的 $\frac{x_1}{\sigma_x}$，$\frac{x^2}{\sigma_x}$，…………，$\frac{x_n}{\sigma_x}$ 各項，同Y變量 $\frac{y_1}{\sigma_y}$，$\frac{y_2}{\sigma_y}$，…………，$\frac{y_n}{\sigma_y}$ 各配對項，其值相等，所冠符號，亦都相同，則X與Y的相關可證明他是絕對完全。至於r之值，當視二變量之以標準差爲單位的數值 Reduced value 而定。設二變量之標準差單位的各數值常爲相等，則各 $\frac{x}{\sigma_x}$ 可用以代替各 $\frac{y}{\sigma_y}$，此時r之值爲＋1。同理，假如二變量之以標準差爲單位的各數值(Rdeuced value)雖同，但是所冠符號不同，則r之值爲−1。＋1 或−1，都是表示完全相關。＋1 是完全正相關，−1 是完全負相關。

設在另一方面，二變量間絕對沒有關係，則X變量的任何量數，和Y變量大小不一律的量數相配對，xy 的正負值，差可相消，因此，他的代數和 (Σxy) 變爲0。

再依數學理論，如二數的和不變更，他的乘積的數值最大時，必定二數是相等的。所以 X,Y 的相當量數的 Reduced value 相等時，r的值必爲最大。X,Y 的 Reduced value 漸差漸遠，r的值漸變漸小。不過，X,Y 的 Reduced value 相等，必定要在完全相關的情形之下，所以1是係數的最大值，而r值總在±1至0的限域內。

簡捷法　上列方法，係由平均數的原點，推演而得，按步就班，手續太繁，故實際計算多用簡法，先用一假定平均數爲原點，然後設法校正之。他的公式，是

$$r=\frac{\frac{1}{N}\Sigma(x'y')-c_xc_y}{\sigma_x\ \sigma_y}\quad\text{(註)}$$

在此式內，x' 爲 X 行任何量數的假設平均數的差數，y' 爲Y行任何量數與假設平均數的差數。對假定平均數求標準差

註　相關係數簡單公式之證明 Cheddock's Text, p. 268,

證 $r=\dfrac{\frac{1}{N}\Sigma(x'y')-c_xc_y}{\sigma_x\ \sigma_y}$

設　x' 爲假定平均數 M'_x 之差

　　y' 爲假定平均數 M'_y 之差

於是 $x'_1=x_1+c;\ x'_2=x_2+c;\cdots\cdots\cdots\cdots;$

　　$y'_1=y_1+c;\ y'_2=y_2+c;\cdots\cdots\cdots\cdots;$

再　$x'_1y'_1=(x_1+c_x)(y_1+c_y)$

$$=x_1y_1+x_1c_y+y_1c_x+c_xc_y$$

$$x'_2y'_2=x_2y_2+x_2c_y+y_2c_x+c_xc_y$$

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

加之得 $\Sigma x'y'=\Sigma xy+c_y\Sigma x+c_x\Sigma y+\Sigma c_xc_y$

但　$\Sigma x=0$ 及 $\Sigma y=0$ 而 $\Sigma c_xc_y=Nc_xc_y$

代入卽得 $\Sigma x'y'=\Sigma xy+Nc_xc_y$

移項 $\Sigma xy=\Sigma x'y'-Nc_xc_y$

兩邊除以 $N\sigma_x\sigma_y$ 卽得

$$\frac{\Sigma xy}{N\ \sigma_x\ \sigma_y}=\frac{\Sigma x'y'-Nc_xc_y}{N\sigma_x\sigma_y}=\frac{\frac{\Sigma x'y'}{N}-\frac{Nc_xc_y}{N}}{\sigma_x\sigma_y}$$

$$=\frac{\frac{1}{N}\Sigma(x'y)-c_xc_y}{\sigma_x\sigma_y}$$

的算法，前面已經講過，現在代入上列公式，即得

$$r=\frac{\frac{1}{N}(x'y')-c_xc_y}{\sqrt{\frac{\Sigma(x')^2}{N}-c_x^2}\times\sqrt{\frac{\Sigma(y')^2}{N}-c_y^2}}$$

將上式的分子分母各乘以N，則變爲

$$r=\frac{\Sigma(x'y')-Nc_xc_y}{\sqrt{\Sigma(x')^2-Nc_x^2}\times\sqrt{\Sigma(y')^2-Nc_y^2}}$$

根據這公式來計算 r，非常簡便。如應用分組法，用組距爲單位，不用原有單位，更爲便利。試以表39的事實，用簡法計算，示例於表40

表40　相關係數之計算(簡便法)

田地	每畝施肥數量(百磅) X	每畝水果產量(千磅) Y	對假定原點M'x之離差(M'x=200) x'	對假定原點M'y之離差(M'y=15) y'	$(x')^2$	$(y')^2$	$x'y'$
A	60	10	−140	−5	19,600	25	700
B	100	14	−100	−1	10,000	1	100
C	150	12	−50	−3	2,500	9	150
D	220	16	20	1	400	1	20
E	240	19	40	4	1,600	16	160
F	310	17	110	2	12,100	4	220
G	350	20	150	5	22,500	25	750
總數			+30	+3	68,700	81	2100

$$c_x=\frac{30}{7}=4.3\qquad c_y=\frac{3}{7}=.4$$

$$r=\frac{\Sigma(x'y')-Nc_xc_y}{\sqrt{\Sigma(x')^2-Nc_x^2}\ \sqrt{\Sigma(y')^2-Nc_y^2}}$$

$$=\frac{2100-7(4.3)(.4)}{\sqrt{68700-7(4.3)^2}\ \sqrt{81-7(.4)^2}}=\frac{2088}{2330}=.89$$

Ayres公式　L. P. Ayres 更計劃一種公式，可完全用 X,Y 的原來數目，不用差數法來計算相關係數，公式是

$$r=\frac{\Sigma(XY)-\left(\frac{\Sigma X}{N}\right)\Sigma Y}{\sqrt{\Sigma X^2-\left(\frac{\Sigma X}{N}\right)\Sigma X}\sqrt{\Sigma Y^2-\left(\frac{\Sigma Y}{N}\right)\Sigma Y}}$$

$$r=\frac{\Sigma XY-\frac{\Sigma X\times\Sigma Y}{N}}{\sqrt{\Sigma X^2-\frac{(\Sigma X)^2}{N}}\sqrt{\Sigma Y^2-\frac{(\Sigma Y)^2}{N}}}\quad\text{(註)}$$

此公式中，無需用平均數及假定原點等，并免除了差數的正負號之麻煩。不過數字太大，非借助於計算機及乘方表不可。此公式在計算上，頗有特殊之利益。

相關表法　我們在研究一單獨事實的時候，應用次數分配表，便利很多。同理。在研究二組數列時，可用相關表。相關表可以說是縱橫二次數表組合而成， X 軸為 X 變量，代表一事實， Y 軸為 Y 變量，代表另一事實。今試以美國聯合準備銀行貼現率，與商業銀行貼現率之相關材料，用上列簡法公式。示例於表41。

(註)　又可等於

$$r=\frac{\frac{1}{N}\Sigma XY-M_xM_y}{\left(\sqrt{\frac{\Sigma X^2}{N}}-M_x\right)\cdot\left(\sqrt{\frac{\Sigma Y^2}{N}}-M_y\right)}$$

表41　相關之計算(相關表法)

X　聯邦準備銀行貼現率(百分數)

Y　商業銀行貼現率(百分數)

	3.75—4.24	7.25—4.74	4.75—5.24	5.25—5.74	5.75—6.24	6.25—6.74	6.75—7.24	f_y	y'	$f_y(y')$	$f_y(y')^2$	S_s	$S_s y$
7.75—8.24					1	1	2	4	3	12	36	9	27
7.25—7.74					7	9	1	17	2	34	68	28	56
6.75—7.24			5	4	63	9	36	117	1	117	117	184	184
6.25—6.74		2	9	10	22	1	3	47	0	0	0	20	0
5.75—6.24	1	90	29	6	30			156	−1	−156	156	−182	182
5.25—5.74	11	110	5					126	−2	−252	504	−255	510
4.75—5.24	10	24						34	−3	−102	306	−78	234
4.25—4.74	2	1						3	−4	−12	48	−8	32
f_x	24	227	48	20	123	20	42	504		−359	1235		1225
x'	−3	−2	−1	0	1	2	3						
$f_x(x')$	−72	−454	−48	0	123	40	126	−285					
$f_x(x')^2$	216	908	48	0	123	80	378	1253					

$$C_y=\frac{-359}{504}=-.712$$

$$\sigma_y{}^2=\frac{1235}{504}-C_y{}^2=1.943 \quad \sigma_y=1.394$$

$$C_x=\frac{-285}{504}=-.565$$

$$\sigma_x{}^2=\frac{1753}{504}-C_x{}^2=3.159 \quad \sigma_x=1.777$$

$$\Sigma x'y'=\frac{1225}{504}=2.4305$$

$$r=\frac{\Sigma x'y'-NC_xC_y}{N\sigma_x\sigma_y}=.82$$

表中符號，有待說明者，祇 S_s 一欄。欄中之各項爲各列 Row (或各行 Colocm) 之離差代數和，此代數和係以組距單位表示者。例如表中 S_s 欄第一項(9)爲 $1\times1, 1\times2$, 及 2×3 之和，而此乘數

1,2,3，各爲x'欄中之離差；第四項(20)，爲$2\times(-2)$，$9\times(-1)$，10×0，22×1，1×2，及3×3之和。餘可類推。

對角線法　計算r的方法，假如用對角線法，更可以簡單。所謂對角線法者，係將相關表之各格，自左而右，向上作45度之對角線。此表中之次數分配數列，共計有三，(1)x變量之次數數列，由各縱行之次數總合而得。(2)y變量之次數數列，由各橫列之次數總合而得。(3)$y-x$之次數數列，稱之曰z，由各對角線間之次數總合而得。此三數列之次數，各置於f行f列及右上角之邊線上。各數列之次數分配旣已求得，乃從fx，fx^2；fy，fy^2：及fz，fz^2：各欄，得其總和，而各求其標準差，然後代入下列公式，卽得r之値。

$$r=\frac{\sigma_x^2+\sigma_y^2-\sigma_z^2}{2\sigma_x\sigma_y}\quad(\text{註})$$

計算方法，示例於表42。

如次數分配之總數很大，採用此法，省便之處，可以立見。苟將表再作向下45度之對角線，我們更可得校勘之法，可以試計算之正確與否。如是$z=x+y$，而

$$r=\frac{\sigma_z^2-\sigma_x^2-\sigma_y^2}{2\sigma_x\sigma_y}$$，設計算無訛，二者之結果，必相符合。

(註) $\sigma_x^2=\frac{1}{N}\Sigma z^2$

$=\frac{1}{N}\Sigma(y-x^2)$

表42　相關之計算(對角線法)

Fy²	Fy	f	y																							z	F	fz	fz²
																1		1	2	2	3		7	4	7				
169	13	1	13																					1	2				
			12																										
			11																							8	1	8	64
200	20	2	10									1		1											1	7			
81	9	1	9																	1						6	1	6	36
			8																							5	2	10	50
			7																							4	2	8	32
72	12	2	6								1				1											3	3	9	27
			5																							2			
			4																							1	7	7	7
9	3	1	3										1													0	4		
4	2	1	2								1															-1	7	-7	7
3	3	3	1			1	1					1														-2	2	-4	8
		2	0				1		1																	-3			
6	-6	6	-1		1	2		1	1					1												-4			
8	-4	2	-2				1		1																	-5	1	-5	25
			-3																										
112	28	7	-4		2	1	2	2																					
50	-10	2	-5		1	1																							
714	+14	30	X		-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15			30	32	256
$C_y = .47$			F		4	5	5	3	3		2	2	1	2	1					1				1	30	$C_z = 1.07$			
$\sigma_y = 4.86$			FX		-16	-15	-10	-3			4	6	4	10	6					11				15	+12	$C_x = .40$	$\sigma_z = 2.72$		
$r = .836$			FX²		64	45	20	3			8	18	16	50	36					121				225	606	$\sigma_x = 4.46$			

$$r=\frac{\sigma_x{}^2+\sigma_y{}^2-\sigma_z{}^2}{2\sigma_x\sigma_y}=\frac{(4.48)^2+(4.86)^2-(2.72)^2}{2(4.48\times 4.86)}=.832$$

(接前註)

$$=\frac{1}{N}\Sigma y^2-\frac{2}{N}\Sigma xy+\frac{1}{N}\Sigma x^2$$

$$=\sigma_y{}^2-2\sigma_x\sigma_y r+\sigma_x{}^2$$

因此　$r=\dfrac{\sigma_x{}^2+\sigma_y{}^2-\sigma_z{}^2}{2\sigma_x\sigma_y}$

回歸線及回歸係數

(Lines of Regrssion and Coefficient of Regression)

相關度之算法旣明，苟二數列之相關，確有存正，則第二步之分析，當爲決定此關係之最最或然之數值。由此最最或然之數值，我們就可依一事實的變化，而預測他一事實的變化，由已知推未知，由已往推未來，這是科學方法的特質，亦是統計學精髓所在。二數列之關係，苟爲直綫形式，則最能表示兩數列間關係之線，卽可從批爾生係數得之。設通過二數列各平均數而配合最適當之直線，以斜度形 $y=bx$ 表之，則此方程式 b 之值，可作 $r\left(\frac{\sigma_y}{\sigma_x}\right)$ 之式。換言之，$y=r\frac{\sigma_y}{\sigma_x}x$，乃表示最最或然數值(Most probable value)之方程式。此方程式，在統計上，稱爲回歸方程式 (Regression equaton)。此方程式所配合之線，爲回歸線(Line of regression)，乃表示二數列平均關係之直線也。上式以 y 爲 x 之函數，易言之，x 爲自變量，(Independent variable) y 爲倚變量。(Dependent variable)。$r\frac{\sigma_y}{\sigma_x}$ 之值，爲此回歸線之斜度，卽所謂回歸係數者是也。此斜度之表示，卽謂 x 有一單位之變化，y 卽須平均有若干單位之變化也，以貼現率相關之例所得各值，代入公式，得

$$y=.82\frac{.697}{.8885}x=.64x$$ (註見162頁)

回歸直線之方程式，在任何情形之下，必有二式。一為Y對X之回歸線，即上述之式，其他一式為X對Y之回歸線，以Y為自變量，X為倚變量，而表示其平均關係者也。試以各值代入公式，得

$$x=.82\ \frac{.8885}{.697}y=1.045y$$

總之，Y對X之回歸線，如圖30，Y行各平均數對此線之平方差之和必為最小。X對Y之回歸線，如圖31，X列各平均數對此直線之平方差之和，亦必最小。此定義之由來，完全應用最小自乘法之理論。

圖30　表示商業銀行貼現率與聯邦準備銀行貼現率之關係

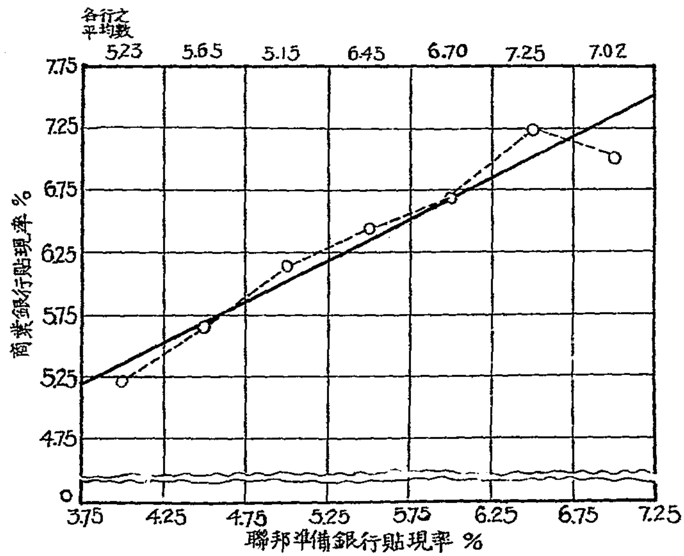
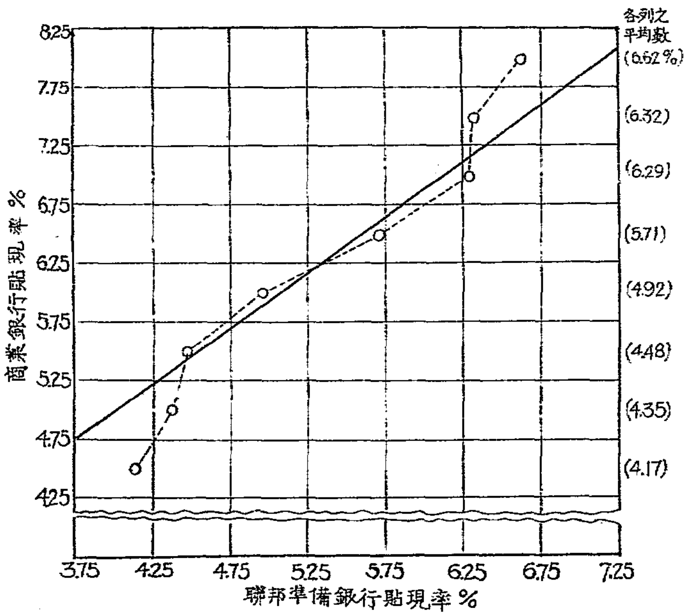


圖31 表示聯邦準備銀行貼現率與商業銀行貼現率之關係



此時，更有一點須注意者，卽相關度 r 苟爲 1，二回歸線合而爲一。設 r 爲1. 而 σ_x 與 σ_y 之値，亦復相等，則此回歸線適在平分縱橫二軸所成直角之地位。如所作之圖，縱橫二軸，爲標準差單位，則此方程式變爲 $y=rx$，而回歸線之斜度，則與 r 之値相等。

註 160 頁 計算 r 之時，分子分母均以組距爲單位，故可不必改爲原有單位，若計算回歸線時，設 x 與 y 之距組不相同，則式中之 σ_x 與 σ_y，均當改爲原有單位，然後計算；在本例，二者單位相同，可不必更改。

回歸係數我們常以 b 表之。在簡單相關問題中，此項係數共有二，即表示此二回歸線之斜度，其式如下，

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

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

（式中 b 下標註二字，所以表示此二變量之關係，第一字爲倚變量，第二字爲自變量。）

二公式中，均有係數 r ，由此可知 r 之值，亦可由回歸係數計算而得，因爲

$$r=\sqrt{b_{yx}\times b_{xy}}=\sqrt{r\frac{\sigma_y}{\sigma_x}\cdot r\frac{\sigma_x}{\sigma_y}}=\sqrt{r^2}$$

設我們已知二回歸線之斜度，即可求得 r ，將本例各值代入公式，即得

$$r=\sqrt{.64248\times 1.045}=.819$$

我們尙須注意，在上面計算時，用批爾生公式計算，原點(Orign)已由原有數值轉移至 x，y 之公共平均數（General mean)之點，故所有結果，必須再行換算，至原有量表。其法令 $\overline{X}$ 爲 X 之平均數，$\overline{Y}$ 爲 Y 之平均數，將 $y=r\frac{\sigma_y}{\sigma_x}x$ 略改，則爲

$$Y-\overline{Y}=r\frac{\sigma_y}{\sigma_x}(X-\overline{X})$$

代入上例，則得

$$Y-6.144=.64(X-5.218)=.64X-3.340$$

$$Y = .64X + 2.804$$

同理，$x = r\frac{\sigma_x}{\sigma_y}y$ 可改為 $X-\overline{X} = r\frac{\sigma_x}{\sigma_y}(Y-\overline{Y})$

此時原點之地位，已由平均數之點，移至 X, Y 原有量表之上，我人推測之時，可由一變量而知他一變量之值矣。

推測之標準差誤 (Standard Error of Estimate)

回歸線之爲用，可從一變量而推知他一變量之值，前已略述之矣。不過此種推測，除極少數之例外，兩變量之實在數值，與從回歸線推測之數值，不盡相同。比較實際數值與推測數值所得之差數名曰推測差誤。此差誤之大小足以定此回歸推測之正確程度如何。

設回歸爲完全直線，各行之算術平均數，適落於此直線上，則各行之推測差誤爲 0 。否則，如回歸非屬直線，大多數或全部之平均數，並不通過此直線，則其推測差誤決不爲零，各行平均數對回歸之離差，或者毫無規則 (Erratic) ，如圖32，或者依據規則，如圖33。前者之回歸，爲近似直線之形，而後者則爲曲線形。二者對各行平均數所作之曲線，皆非適當之配合。

配合適當之回歸，須吻合二種條件，(一)各行之平均數，須落於此回歸線上，或其附近之處。(二)每行所有各點之分佈，均作集中該平均數之傾向。第一種條件，涉及以直綫關係代表相關是否適當，第二種條件涉及分配之全體是否一致。

圖32　各縱行平均數對回歸線之離差(假設之分配)

不規則之例

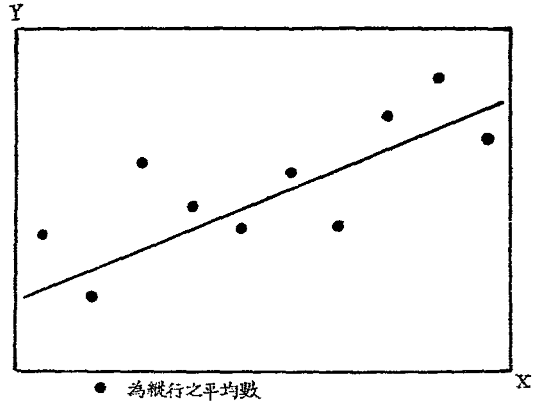
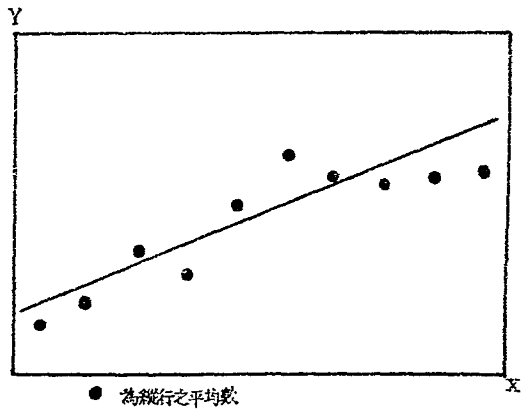


圖33　各縱行平均數對回歸綫之離差(假設之分配)

規則之例



推測差誤之計算，固可由各實在數值，與推測數值之差數，一一計算而得之。設我們已求出回歸，則可將各值代入下式而得之，

$$S_y^2=\sigma_y^2(1-r^2)$$

S_y 卽推測差誤之符號，一名標準差誤，S_y 之數值，卽 Y 變量各行對回歸綫之標準差也。

S_y 之意義，與 σ 同。假令各項對此直線之分配，近於常態曲線之形態，則全體項數 68% 當在 ±S 之範圍，95% 當在 ±2S 之範圍內，99.7% 當在 ±3S 之範圍。如圖中各點，盡在此直綫上，則 S 之值爲零，且 Y 之値可依 X 之值推算而得，非常正確。對此直線之差量愈小，則 S 之値亦愈小。否則，差量愈大，則 S 之值亦愈大。故 S 之值，可作此直線表示二變量之關係是否有價值之象徵。以聯邦準備銀行貼現率及商業銀行貼現率之例中各值代入公式，卽得

$$S_y^2=1.943(1-82^2)$$

$$=.6257$$

$$S_y=.791$$

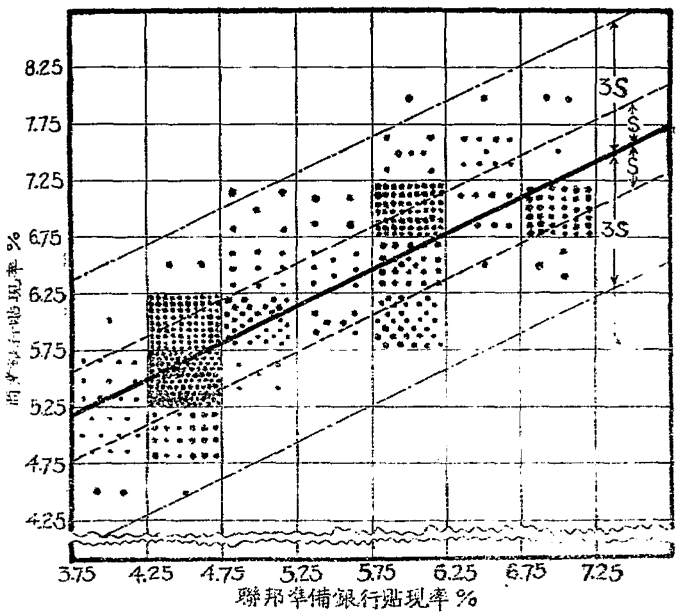
此 S_y 之值，爲組距之單位，Y 變數之組距，旣爲 .5%，故其原有單位之數值，當爲

$$S_y=.791\times.5\%$$

$$=.4\%,$$

標準差誤之圖解，請參圖34。

圖34　分佈圖—表示聯邦準備銀行及商業銀行貼現率，及其平均關係線及推測區



第五章

相關（續）

相關率（The Corrlation Ratio）

相關係數適用的範圍，僅限於 X,Y 的關係，是一直線。假設 X,Y 的關係不是直線，而是曲線，根據批爾生係數以求 r 的數値，决不能得二變量相關的眞實情形，而陷於方法的錯誤。例如圖 35 ，圖中各點，很顯然的有集中的趨勢，雖則各集中點連接所成的線，是曲線的形態，而非直線。如用批爾生係數來求相關度，得 .793 用適當的方法來求，則得實際相關度 .965 ，因此可以明瞭前者的不正確。在這種情形之下，我們可用相關比率來補救。這方法亦批爾生所倡導，其理由就在求各行（或各列）平均數的標準差，與全體標準差的比例，如用各行之平均數則用 σ_y，各列則用 σ_x。他的公式是

$$\eta_{yx}=\frac{\sigma_{my}}{\sigma_y}\text{(註)}$$

註 η 之另一式為 $\eta_{xy}=\frac{\sigma_{mx}}{\sigma_x}$，此為 X 對 Y 之回歸節，Y 為自變量，X 為倚變量，此處之 σ_{mx}，乃 X 各列平均數對 X 全體平均數的標準差，η 與 r 不同，在 r 可不問 X 或 Y 為自變量，其值不變，但在 η 則 η_{xy} 與 η_{yx}，未必相同，除非 r 與 η 之値，二者相同時，但平常二比率的數値還接近，所以祇取其一，平常總以 η_{yx} 為當選。

η 旁標註小字，所以表明 Y 對 X 之回歸。換言之，即 X 為自變量，Y 為倚變量，σ_{my} 為各行平均數對 Y 全體的平均數的標準差。

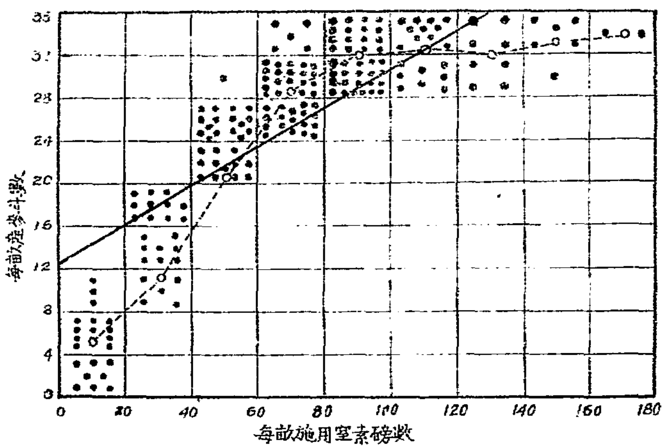
σ_y 的算法，我們已經熟習，毋容再贅，所以現在要詳為解釋的，就是 σ_{my}，從何算得。他的演算，可分為下列數步。

1. 先就 Y 變量的各行，求其算術平均數(或就 X 的各列)。
2. 就各行平均數，而求其對 Y 全體平均數的離差(或 X 全體)。
3. 將此離差平方之，更乘以各行項數 N_x (或 N_y)，得其總和。
4. 再以總項數除之，即得。

表 43 每畝產麥量與施用窒素之相關

Y — 每畝產麥斗數	X — 每畝施用窒素磅數										
	0—19.9	20—39.9	40—59.9	60—79.9	80—99.9	100—119.9	120—139.9	140—159.9	160—179.9	總計	各列平均數
32—35.9				5	16	12	4	5	2	44	107.27
28—31.9			1	20	21	8	4	1		55	88.91
24—27.9			16	19						35	60.86
20—23.9			13							13	50.0
16—19.9		12								12	30.0
12—15.9		8								8	30.00
9—11.9	3	5								8	22.5
4—7.9	10									10	10
0—3.9	8									8	10
總計	21	25	30	44	37	20	8	6	2	193	
各行之平均數	5.05	15.12	24.4	28.75	31.78	32.4	32.0	33.33	34.0		

圖35　分布圖表示麥產與施用窒素之關係



這樣，$\eta_{yx}=\frac{\sigma_{my}}{\sigma_y}$ 的公式，可依他的計算程序，轉變而爲下式。

$$\eta_{yx}=\frac{\sqrt{\frac{\Sigma[N_x(\overline{y}_x-\overline{y})^2]}{N}}}{\sigma_y}$$

式中 N_x，爲各行中所有的項數。

$\overline{y}_x$ 爲各行之平均數。

$\overline{y}$ 爲表中 y 全體之平均數。

N 爲全體項數，

σ_y 爲Y變量全體之標準差，

今試以每畝麥之產額，與所施窒素之相關爲例，用上列公式以求相關率於表44

表44　相關率之計算

X 各組中點	$\overline{y}_x$	$\overline{y}_x-\overline{y}$ (25005)	$(\overline{y}_x-\overline{y})^2$	N_x	$N_x(\overline{y}_x-\overline{y})^2$
10	5.05	−19.955	398.202	21	8,362.242
30	15.12	− 9.885	97.713	25	1,442.825
50	24.40	− .605	.366	30	10.980
70	28.73	+ 3.725	13.876	44	610.544
90	31.73	+ 6.725	45.226	37	1,673.362
110	32.40	+ 7.395	54.686	20	1,093.720
130	32.00	+ 6.995	48.930	8	391.440
150	33.33	+ 8.325	69.306	6	415.836
170	34.00	+ 8.995	80.910	2	161.820
總　數				193	15,162.769

$$\sigma_{my}=\sqrt{\frac{\Sigma N_x(\overline{y}_x-\overline{y})^2}{N}}=\sqrt{\frac{15,162.769}{193}}=8.864$$

吾人求得 σ_y 之值爲 9.188，σ_{my} 之值 8.864 代入公式卽得

$$\eta_{yx}=\frac{\sigma_{my}}{\sigma_y}=\frac{8.864}{9.188}=.965$$

上項計算，手續麻煩，不切實用，我們可以用相關表來幫助，祇要將求批爾生係數之方法，略爲增加幾行卽可。試另舉一簡單的例於表45.

相關率，旣爲兩種標準差的比率，故常爲正而不爲負，而其數值常在1－0之間，在計算 r 時，己知 x,y 的標準差，則回歸方程式代入卽得。但在 η 則不同，雖已求得 η 之數值，若欲對於二變量的關係，求其數理的關係，不能不再配合一曲線，以求其方程式。此點讀者宜注意及之。

表15　用相關表計算相關率與相關係數

Monthly Pig-iron Tonnage (in millions of tons)	Monthly Price Index 7.50 TO 8.00	8.00 TO 8.50	8.50 TO 9.00	9.00 TO 9.50	9.50 TO 10.00	f_y	y'	$f_y y'$	$f_y(y')^2$	S_x'	$S_x' y'$	$(S_x')^2$	$\frac{(S_x')^2}{f_y}$
28 TO 32				4	4	8	3	24	72	12	36	144	18.00
24 TO 28			4	16	1	21	2	42	84	18	36	324	15.43
20 TO 24		9	26	7	5	47	1	47	47	8	8	64	1.36
16 TO 20	8	21	14	3	1	47	0			−32		1024	21.79
12 TO 16	10	8	2	2	1	23	−1	−23	23	−24	24	576	25.04
8 TO 12	7	3				10	−2	−20	40	−17	34	289	28.90
						156		70	266		138		110.52

						Total
f_x	25	41	46	32	12	
x'	−2	−1	0	1	2	
$f_x x'$	−50	−41		32	24	−35
$f_x(x')^2$	100	41		32	48	221
S_y'	−24	−5	32	19	18	
$S_y' x'$	48	5		49	36	138
$(S_y')^2$	576	25	1024	2401	324	
$\frac{(S_y')^2}{f_x}$	23.04	0.61	22.26	75.03	27.00	147.94

$$c_y = \frac{70}{156} = .4487 \text{ (intervals)}$$

$$\sigma_y^2 = \frac{266}{156} - c_y^2 = 1.5038$$

$$\sigma_y = 1.2263 \text{ (intervals)}$$

$$c_x = \frac{-35}{156} = -.2244 \text{ (intervals)}$$

$$\sigma_x^2 = \frac{221}{156} - c_x^2 = 1.3663$$

$$\sigma_x = 1.1689 \text{ (intervals)}$$

$$\frac{\Sigma(x' \cdot y')}{N} = \frac{138}{156} = .8846 \text{ (intervals)}$$

$$r = \frac{1}{\sigma_x \sigma_y}\left\{\frac{\Sigma(x' \cdot y')}{N} - c_x c_y\right\} = .687$$

$$\frac{\Sigma\frac{(S_x')^2}{f_y}}{N} = \frac{110.52}{156} = .7085$$

$$\eta^2_{xy} = \frac{1}{\sigma_x^2}\left\{\frac{\Sigma\frac{(S_x')^2}{f_y}}{N} - c_x^2\right\}$$

$$\eta_{xy} = .694$$

$$\frac{\Sigma\frac{(S_y')^2}{f_x}}{N} = \frac{147.94}{156} = .9483$$

$$\eta^2_{yx} = \frac{1}{\sigma_y^2}\left\{\frac{\Sigma\frac{(S_y')^2}{f_x}}{N} - c_y^2\right\}$$

$$\eta_{yx} = .705$$

相關率之校正　在應用 η 的時候，最好要資料較多，而又可組成相關表之形式。假如包含之項數很有限，排成相關表後，每格的項數不過一二項，那末 σ_{my} 與 σ_y 的測量，結果相符。而 η 之值當然是 1。我們要明白，從項數很少的資料，編成組數很多，來算 η 是沒有意識的。

Karl Pearson 曾經規劃一個校正法，公式如下

$$校正之\eta^2=\frac{\eta^2-\frac{(K-3)}{N}}{1-\frac{(K-3)}{N}}$$

式中K，代表行數，應用這校正法將上例的結果代入，得

$$校正之\eta^2=\frac{(.965)^2-\frac{(9-3)}{193}}{1-\frac{(9-3)}{193}}$$

$$=.929$$

$$校正之\eta=.964$$

在本例中校正之數極微，但如N很小，K很大，則其校正之數，大有可觀。

相關率與相關係數之關係　當二變量間的關係，絕對成一直線時，通過各行平均數之線，當然與算相關係數所根據之線相合。當此情形，η與r數值相同。設二變量間之關係，漸離直綫之形態，η與r之值亦異，η總要比r大些。這由於對連結各行平均數一綫的離差，總要比對就各點所配合的直線的離差少些，除非那直線是通過各平均數的。而且對表示平均關係線的離差愈少，相關測量的數值亦愈大。在表43之例，η為.965，r為+.793，相差很著。由於η與r的關係，可給我們一種測驗以驗相關是否成直線。他的公式，是$\zeta=\eta^2-r^2$。(註)

註ζ讀作 Zeta

此外關於相關是否成直線的測驗，還有一標準，足以參考，因爲相關而成直線者，則

$$\frac{\sqrt{N}}{.67449}\cdot\frac{1}{2}\sqrt{\eta^2-r^2}$$

必小於 2.5。表 43 之例代入，得

$$\frac{\sqrt{193}}{.67449}\times\sqrt{(.865)^2-(.793)^2}=5.646$$

其結果 5.646＞2.5，是可證明必非直線之相關矣。

等級相關（Rank Correlation）

上述之係數和比率法，計算相關，都是從各變量的絕對數值，與位置去求出，所得結果，自較精密。有時，材料數字之表示，並不完全，祇能以各項之大小，定其等次。例如欲估計某城市各區之地價，勢難正確，祇能定某地價值最貴，某地次之，用連續數字1,2,3,4,……………來表示價值的等次，這就是等級。同理，各地建築物之價值，欲估計精確，亦不可能，亦祇能以其價值之大小，依等次排列，假使二數列的各相當變量之等級，俱屬相同，則相關屬於完全。但實際兩變量的等級，焉能盡同。我們要推算此種材料二變量之相關如何，惟有應用等級相關法，他的公式，是

$$\rho=1-\frac{6\Sigma(R_x-R_y)^2}{N(N^2-1)}$$

式中之 ρ 讀作Rho，爲由等級差法，求出之相關係數，R_x

表46　等級相關之計算

美國三十五城之 Clearings 及 Debits 之相關 Oct, 6.1923

單位 $1,000,000

城名	Clearings X	Debits V	R_x	R_y	R_x-R_y	$(R_x-R_y)^2$
Baltimore	104	102	30	30	……	……
Birmingham	27	27	19	10	9	81
Buffalo	51	87	26	28	−2	4
Columbus, Ohio	16	37	11	18	−7	49
Dallas	47	57	25	24	1	1
Des Moines	12	18	5	6	−1	1
Ft. Worth	15	19	10	9	1	1
Galveston	13	33	6	14	−8	64
Indianapolis	20	34	15	16	−1	1
Louisville	28	38	20	19	1	1
Memphis	26	31	18	13	5	25
Milwaukee	39	65	22	25	−3	9
Minneapolis	80	80	29	27	2	4
Nashville	21	18	16	7	9	81
New Orleans	56	67	28	26	2	4
Norfolk	9	19	4	8	−4	16
Oakland	17	28	13	11	2	4
Oklahoma	22	17	17	4	13	169
Omaha	38	46	21	23	−2	4
Portland, Oregon	41	44	24	22	2	4
Providence	14	33	8	15	−7	49
Richmond	58	34	27	17	10	100
Rochester	14	42	9	21	−12	144
St. Joseph	8	39	3	12	−9	81
St. Paul	17	40	14	20	−6	36
Salt Lake City	17	15	12	3	9	81
Seattle	40	88	23	29	−6	36
Sioux City	7	18	1	5	−4	16
Spokane	14	14	7	2	5	25
Wichita	8	10	2	1	1	1
總數	874	11,84	465	465	0	1,092

$$\rho=1-\frac{6\Sigma(R_x-R_y)^2}{N(N^2-1)}=1-\frac{6552}{26970}=1-.24294=.75706$$

$$\frac{\pi}{6}\rho=30°(.75706)=22.°7118=22°42'42''$$

$$\sin\frac{\pi}{6}\rho=.38609$$

$$r=2\sin\frac{\pi}{6}\rho=.772$$

為X各變量之等次，R_y為Y各變量之等次，N為項數。

依Pearson之方法，這ρ尚須依下列公式換算成r。

$$r = 2\sin\left(\frac{\pi}{6}\rho\right)$$

r與ρ之關係，有現成的表可查，所以祇要按ρ的數值，可從表上查得r的相當數值，不必實際照公式去計算，不過r與ρ，二數值之相差，登峯造極，亦不會超過.018。此種相差，不妨忽視。所以為便利計，求得ρ後，就看作為相關係數，亦無不可。計算方法，示例於表46。

第六章

相 關(再續)

複相關，局部相關

(Multiple Correlation, Partial Correlation)

上面討論之相關，只計及二個變量，一爲自變量，一爲倚變量。但是社會經濟現象中的事物，彼此關係，决不如此單純，因爲一種事物常受多種事物交互之影響，其情形甚爲複雜。例如社會學家研究城市房租的問題，人口增減，固然是其中重要的原因，不過除出人口增減之外，和交通，地勢，商業中心等，亦多有關係。他如與物價有關之原因，有供給和需要等；農產收成之豐歉，受雨量氣候等之影響。可見簡單的兩種現象的相關，不是普通的情形。如遇此等場合，簡單的兩個變量的相關測量，就不能支配自如。幸而統計學中，除簡單的兩個變量的相關測量之外，更有所謂複相關及局部相關者。

統計學裏，一種相關度，可以測量一倚變量 X_1 和其他幾個自變量，如 $X_2, X_3, X_4, \cdots\cdots X_n$ 之相關者，就稱複相關。相關之爲測量一倚變量 X_1，與一自變量 X_3, X_5，或 X_n 的關係，假定其餘的變量 $X_2, X_4, X_6, \cdots\cdots X_n$ 等不變 (Held constant)，是爲局部相關，或稱淨相關 (Net correlation)。這種分析

法，在自然科學中常用之，社會學家經濟學家研究社會經濟現象，在合理的範圍內，亦可適用。

表47　康色司州之玉蜀黍產量及氣溫　1890－1022

(1) 年份	(2) 每畝實際所產斗數	(3) 趨向值每畝所產斗數	(4) 實際產量對趨向值之百分比 X_1	(5) 六月之平均氣溫 X_2	(6) 七月之平均氣溫 X_3	(7) 八月之平均氣溫 X_4
1890	15.6	22.4	69.6	77.6	80.1	76.1
1891	26.7	22.2	120.3	70.7	74.0	75.1
1892	24.5	22.1	110.9	73.4	77.5	76.5
1893	21.3	21.9	97.3	74.7	79.5	73.8
1894	11.2	21.8	51.4	74.2	77.8	78.0
1895	24.3	21.6	112.5	71.7	74.9	76.0
1896	28.0	21.5	130.2	74.1	78.1	78.7
1897	18.0	21.3	84.5	76.6	80.2	76.0
1898	16.0	21.2	75.5	75.0	77.7	78.2
1899	27.0	21.0	128.6	73.9	76.2	80.6
1900	19.0	20.9	90.9	74.9	77.9	81.0
1901	7.8	20.7	37.7	77.3	85.0	79.1
1902	29.9	20.6	145.1	70.9	76.8	78.2
1903	25.6	20.4	125.5	67.2	78.3	75.3
1904	20.9	20.3	103.0	70.4	75.6	74.6
1905	27.7	20.1	137.8	75.5	74.5	78.7
1906	28.9	20.0	144.5	71.8	73.8	76.3
1907	22.1	19.8	111.6	72.0	78.4	78.1
1908	22.0	19.7	111.7	72.1	75.8	76.2
1909	19.9	19.5	102.1	73.1	78.1	80.1
1910	19.0	19.4	97.9	72.2	70.5	75.7
1911	14.5	19.2	75.5	80.5	78.6	76.4
1912	23.0	19.1	120.4	69.3	70.9	77.4
1913	3.2	18.9	16.9	74.2	82.1	84.2
1914	18.5	18.8	98.4	78.2	79.9	78.2
1915	31.0	18.6	165.7	69.2	74.0	70.1
1916	10.0	18.5	54.1	70.3	81.2	79.6
1917	13.0	18.3	71.0	72.8	80.8	73.4
1913	7.1	18.2	39.0	78.4	78.3	82.3
1919	15.2	18.0	84.4	72.3	80.2	78.3
1920	26.5	17.9	148.0	72.8	77.6	72.9
1921	22.2	17.7	125.4	74.4	79.2	78.6
1922	19.3	17.6	109.7	75.2	77.0	80.1

轉載 Mills' text P. 486

照理論講，一倚變量，固然可以和任何多個自變量，以研究其相關，不過在事實上，所用之自變量，至多亦不過三四組

。本章以美國玉蜀黍生產量與六月七月及八月之氣溫相關材料爲例，以說明複相關及局部相關之應用。

複相關(Multiple Correlation)

複相關的普通公式，是

$$X_1 = a + b_{12.34}X_2 + b_{13.24}X_3 + b_{14.23}X_4$$

式中所用符號，須略加詮釋，$b_{12\cdot 34}$是代表X_2對X_1之淨回歸係數，符號旁標註之小字，以點爲界，分爲二部。在點左者，名爲基本標註 Primary subscripts，第一字表示倚變量，第二字表示自變量。在點右之一部，名爲次級標註 (Secondary subscripts)，表示假定不變的自變量。

上項方程式之解法，亦須先有標準方程式。假使有四個變量，就有四個標準式，如下，

1　$\Sigma(X_1) = Na + b_{12\cdot 34}\Sigma(X_2) + b_{13\cdot 24}\Sigma(X_3) + b_{14\cdot 23}\Sigma(X_4)$

2　$\Sigma(X_1X_2) = a\Sigma(X_2) + b_{12\cdot 34}\Sigma(X_2^2) + b_{13\cdot 24}\Sigma(X_2X_3) + b_{14\cdot 23}\Sigma(X_2X_4)$

3　$\Sigma(X_1X_3) = a\Sigma(X_3) + b_{12\cdot 34}\Sigma(X_2X_3) + b_{13\cdot 24}\Sigma(X_3^2) + b_{14\cdot 23}\Sigma(X_3X_4)$

4　$\Sigma(X_1X_4) = a\Sigma(X_4) + b_{12\cdot 34}\Sigma(X_2X_4) + b_{13\cdot 34}\Sigma(X_3X_4) + b_{14\cdot 23}\Sigma(X_4^2)$

以相當各值代入此標準式，固然卽可解之而得四常數，不

過這樣計算，手續太煩。爲計算便利計，我們可將四標準式併爲一式，法用每個變量對平均數的離差，以代替其絕對值。這樣，原有方程式之常數項 a 已除去。

設以 A_1, A_2, A_3，等代表各組變量之算術平均數，x_1, x_2, x_3，等代表對算術平均數之離差，我們以相等數 x_1+A_1，x_2+A_2，x_3+A_3，…………等，代絕對數 X_1, X_2, X_3，…………等，則上列四標準式中之第一式消去，其餘三式縮減如下式。

$$\frac{\Sigma(x_1x_2)}{N}=\frac{\Sigma(x_2{}^2)}{N}b_{12\cdot34}+\frac{\Sigma(x_2x_3)}{N}b_{13\cdot24}$$
$$+\frac{\Sigma(x_2x_4)}{N}b_{14\cdot23}$$
$$\frac{\Sigma(x_1x_3)}{N}=\frac{\Sigma(x_2x_3)}{N}b_{12\cdot34}+\frac{\Sigma(x_3{}^2)}{N}b_{13\cdot24}$$
$$+\frac{\Sigma(x_3x_4)}{N}b_{14\cdot23}$$
$$\frac{\Sigma(x_1x_4)}{N}=\frac{\Sigma(x_2x_4)}{N}b_{12\cdot34}+\frac{\Sigma(x_3x_4)}{N}b_{13\cdot24}$$
$$+\frac{\Sigma(x_4{}^2)}{N}b_{14\cdot23}$$

上列方程式之各變量，既是表示對各算術平均數之離差，所以 $\frac{\Sigma(x_1x_2)}{N}$ 僅爲變量 x_1 和 x_2 的平均數之乘積，而 $\frac{\Sigma(x_2{}^2)}{N}$ 爲 $\sigma_2{}^2$ 其餘類推。設以 p_{12}, p_{13}，等符號代表各平均數之乘積，并以各標準差平方之符號代入上列各式，我們可得下列各標準方程式。

$$p_{12}=\sigma_2{}^2 b_{12\cdot34}+p_{23}b_{13\cdot24}+p_{24}b_{14\cdot23}$$

$$p_{13}=p_{23}b_{12\cdot34}+\sigma_3{}^2 b_{13\cdot24}+p_{34}b_{14.23}$$

$$p_{14}=p_{24}b_{12\cdot34}+p_{34}b_{13\cdot24}+\sigma_4{}^2 b_{14\cdot23}$$

依此形式而解標準式，最爲簡便，從表47之材料可得下列各數值。

$\Sigma(X_1)=3298.1$　　$\Sigma(X_1{}^2)=368846.67$

$\Sigma(X_2)=2426.9$　　$\Sigma(X_2{}^2)=178755.75$

$\Sigma(X_3)=2581.5$　　$\Sigma(X_3{}^2)=202163.79$

$\Sigma(X_4)=2553.8$　　$\Sigma(X_4{}^2)=197890.32$

$$\Sigma(X_1X_2)=240967.22$$

$$\Sigma(X_1X_3)=255954.11$$

$$\Sigma(X_1X_4)=253664.85$$

$$\Sigma(X_2X_3)=189941.83$$

$$\Sigma(X_2X_4)=187909.38$$

$$\Sigma(X_3X_4)=199845.00$$

$$c_1=\frac{\Sigma(X_1)}{N}$$

$=99.9424$　　$c_1{}^2=9988.4833$

$c_2=73.5424$　　$c_2{}^2=5408.4846$

$c_3=78.2273$　　$c_3{}^2=6119.5105$

$c_4=77.3879$　　$c_4{}^2=5988.8871$

由上列各值，算成標準方程式內之各項數值如下

$$\sigma_1{}^2=\frac{\Sigma(X_1{}^2)}{N}-c_1{}^2$$

$$=\frac{368,846.67}{33}-9988.4833=1188.688$$

$$\sigma_2{}^2=\frac{178,755.75}{33}-5408.4846=8.3564$$

$$\sigma_3{}^2=\frac{202,163.79}{33}-6119.5105=6.6645$$

$$\sigma_4{}^2=\frac{197,890.32}{33}-5988.8871=7.7893$$

$$p_{12}=\frac{\Sigma(X_1X_2)}{N}-c_1c_2$$

$$=\frac{240,967.22}{33}-(99.9424\times73.5424)=-47.967$$

$$p_{13}=\frac{255,954.11}{33}-(99.9424\times78.2273)=-62.037$$

$$p_{14}=\frac{253,664.85}{33}-(99.9424\times77.3879)=-47.519$$

$$p_{23}=\frac{189,941.83}{33}(73.5424\times78.2273)=2.790$$

$$p_{24}=\frac{187,909.38}{33}-(73.5424\times77.3879)=2.932$$

$$p_{34}=\frac{199,845.00}{33}-(78.2273\times77.3879)=2.063$$

代入各標準式卽得

$$-47.967=8.3564b_{12\cdot34}+2.790b_{13\cdot24}+2.932b_{14\cdot23}$$

$$-62.039 = 2.790b_{12\cdot34} + 6.6645b_{13.24} + 2.063b_{14.23}$$

$$-47.519 = 2.932b_{12\cdot34} + 2.063b_{13\cdot24} + 7.7893b_{14\cdot23}$$

以解聯立方程式之法解之，卽得下列各常數之值：

$$b_{12\cdot34} = -2.095$$

$$b_{13\cdot24} = -7.394$$

$$b_{14\cdot23} = -3.354$$

因此，得方程式

$$x_1 = -2.095x_2 - 7.394x_3 - 3.354x_4$$

此方程式爲 X_1 對 X_2 X_3 X_4 之淨囘歸，設以三組自變量(六月，七月，及八月之氣溫)之任何數值代入此方程式，卽可決定倚變量(每畝玉蜀黍產量)之最最或然之數值矣。不過我們須注意：在此方程式內，所有變量，係以對算術平均數之離差爲單位，爲實際運用便利計，當以原有單位表之。換言之，卽將原點之地位，從平均數一點，轉移至原有量表0點上。欲達到此目的，須將常數 a，仍舊再行加入。a 之數值，可用下列方程式求出之。

$$A_1 = a + A_2b_{12\cdot34} + A_3b_{13\cdot24} + A_4b_{14\cdot23}$$

式中 A_1, A_2, A_3, A_4，代表各算術平均數，以相當值代入，得

$$99.9424 = a + (73.5424 \times 2.094816) + (78.2273 \times -7.39374) + (77.3879 \times -3.353796)$$

解之，得 $a = 1091.94$

因此，以原有數值爲單位之回歸方程式，爲

$$X_1 = 1091.94 - 2.095X_2 - 7.394X_3 - 3.354X_4$$

標準推測差誤之計算　標準推測差誤之公式，爲

$$S^2_{1\cdot 234} = \sigma_1{}^2 - b_{12\cdot 34}p_{12} - b_{13\cdot 24}p_{13} - b_{14\cdot 23}p_{14}$$

$S_{1\cdot 234}$ 代表標準推測差誤，S 旁標註之小字，表示一倚變量 X_1，及三自變量 X_2，X_3 X_4。以相當各值代入公式，卽得

$$S^2_{1\cdot 234} = 1188.688 - 100.482 - 458.700 - 159.369 = 470.137$$

$$S_{1\cdot 234} = 21.68$$

此 S_1 之意義，與吾人在簡單相關時之標準推測差誤相同，不再贅。

複相關係數　離差之測量，最好不以原有單位，而以一抽象的數值表之，這就是相關係數。此項係數之價值，全恃 S 與 σ 之關係而決定，在複相關之係數公式爲

$$R^2_{1\cdot 234} = 1 - \frac{S^2_{1\cdot 234}}{\sigma_1{}^2}$$

$R_{1\cdot 234}$…………係代表一單獨倚變量，與多個自變量之複相關係數。標註之字，在點左者表示倚變量，點右者表示自變量。　以 $S^2_{1\cdot 234}$ 之相等值代入此公式，卽得

$$R^2_{1\cdot 234} = 1 - \frac{\sigma^2_1 - b_{12\cdot 34}p_{12} - b_{13\cdot 24}p_{13} - b_{14\cdot 23}p_{14}}{\sigma_1{}^2}$$

可縮減爲

$$R^2_{1\cdot 234} = \frac{b_{12\cdot 34}p_{12} + b_{13\cdot 24}p_{13} + b_{14\cdot 23}p_{14}}{\sigma_1{}^2}$$

以相當數值代入，即得

$$R^2_{1\cdot234}=\frac{100.482+458.700+159.369}{1138.688}$$

$$R_{1\cdot234}=.778$$

此複相關係數，乃表示一單獨倚變量，與多個自變量聯合相關程度之指標。設將所有自變量，當作一單獨之獨立數列視之，則此相關係數之意義，甚爲明顯。通常計算複相關數時，R 之值，常常不用正負符號以表之。在本例中，各自變量與倚變量之關係各爲負相關，固不妨以負號表之。但有時各自變量與倚變量之相關，有爲正者，有爲負者，至不一律，因此多不以正負號表之。

直線關係與複相關　在討論複相關時，有一點當須注意。複相關之測量，能否合乎邏輯，當視每對變量間是否有直線關係之存在。在本例中，計有四變量，故可列爲六對 (Pairing)。設在此六對中之任何一對之關係，不爲直線，則結果之價值，勢必減低。用以爲推測之標準，勢必不甚可靠。

推測方程式應用之示例　推測方程式之應用，試舉一例以明之。設 1922 年 Kansas 城六月氣温之平均爲華氏 75.2 度，

註　複相關係數亦可從一般公式中演繹而得公式如下

$$R_{1\cdot234\cdots n}=\frac{a\Sigma(X_1)+b_{12\cdot34\cdots n}\Sigma(x_1x_2)+b_{13\cdot24\cdots n}\Sigma(x_1x_3)+b_{14\cdot23\cdots n}\Sigma(x_1x_4)+\cdots-NC_1^2}{\Sigma(x_1)^2-NC_1^2}$$

七月之平均氣溫爲77度，八月之平均氣溫爲80.1度，試問每畝玉蜀黍之最最或然之產量如何。我們可將上述各數代入公式中之X_2X_3及X_4

$$X_1=1091.94-2.095X_2-7.394X_3-3.354X_4$$

即得 $$X_1=1091.94-(2.095\times75.2)-(7.394\times77.0)-(3354\times8.01)=96.4$$

此結果表示實際產量作恆差值之百分比。表47第3欄玉蜀黍產量恆差線之方程式，爲$Y=19.967-.1502X$（式中Y代表產量，X代表時間而以1906年爲原點）；從此方程式中得1922年之恆差值爲每畝17.6斗，則估計產量爲此數之96.4%，或作17斗。

此處僅舉一年以爲例，設吾人用以預測將來之一年，仍可仿用此法，惟須視恆差線之伸引（Projection）是否合理。

局部相關（Partial correlation）

相關係數與回歸係數之關係 在討論簡單相關時，方程式之表示二變量間之關係者，可有二式，以Y爲倚變量，一以X爲倚變量，這二方程式，設以其對各平均數之離差表出之，則爲下列之形式。

$$y=r\frac{\sigma_y}{\sigma_x}x \qquad x=r\frac{\sigma_x}{\sigma_y}y$$

此方程式中 $r\frac{\sigma_x}{\sigma_y}$及 $r\frac{\sigma_y}{\sigma_x}$ 名爲回歸係數，用以測量二回歸

線之斜度。相關係數亦可等於此二回歸係數相乘之平方根，即

$$r=\sqrt{r\frac{\sigma_y}{\sigma_x}\cdot r\frac{\sigma_x}{\sigma_y}}$$

設以回歸係數常用之符號 b 表之，即得

$$r_{yx}=\sqrt{b_{yx}\cdot b_{xy}}$$

此 b_{yx}，即為 Y 對 X 之回歸係數，b_{xy} 為 X 對 Y 之回歸係數，設以符號 X_1 及 X_2 代替 X 及 Y，則標註之小字，應加更改，而得

$$r_{12}=\sqrt{b_{12}\cdot b_{21}}$$

無論 x_1 或 x_2 為自變量，r 之值總是相同。就是說，$r_{12}=r_{21}$。簡單相關時之情形如是，在局部相關時，亦復如是。局部相關，是求二變量之淨相關，假定其餘的原因，是不變的。在研究複相關時，我們雖已求出淨相關之係數，不過那回歸方程式，是敍述一組倚變量，和幾組自變量之關係。其式為

$$X_1=a+b_{12\cdot34}X_2+b_{13.24}X_3+b_{14.23}X_4$$

$b_{12\cdot34}$ $b_{13\cdot24}$ 及 $b_{14\cdot23}$ 乃淨回歸之係數，淨相關之係數，亦可由此法得之，祇須加添幾個數值。係數 r_{yx}，既為 b_{yx} 及 b_{xy} 之方根，係數 r_{13} 既為 b_{13} 及 b_{31} 之平方根，同理，則測量 X_1(玉蜀黍產量)及 X_3(七月氣溫)相關程度之係數，設 X_2(六月氣溫) X_4(八月氣溫)為不變時，亦可由下列公式得之。

$$r_{13\cdot24}=\sqrt{b_{13\cdot24}\cdot b_{31\cdot24}}$$

在此公式中，$b_{13\cdot24}$ 為 X_1 對 X_3 之淨回歸係數，$b_{31\cdot24}$ 為 X_3 對 X_1 之淨回歸係數。而 X_1 及 X_3 兩變量間淨回歸係數之

普遍公式，是

$$r_{13\cdot25\cdots\cdots n}=\sqrt{b_{13\cdot245\cdots\cdots n}\cdot b_{31\cdot245\cdots\cdots n}}$$

係數 $b_{13\cdot24}$ 之數量，已於上例中求得，而係數 $b_{31\cdot24}$ 則爲下列方程式之一常數。

$$X_3=a+b_{31\cdot24}X_1+b_{32\cdot14}X_2+b_{34\cdot21}X_4$$

此方程式，所以表示七月氣溫(倚變量)，與六月氣溫，八月氣溫，及玉蜀黍產量(自變量)之平均關係。

標準方程式之解法　方程式 $X_3=a+b_{31\cdot24}X_1+b_{32\cdot14}X_2+b_{34\cdot21}X_4$ 之各値，亦可仿方程式 $X_1=a+b_{12\cdot34}X_2+b_{13\cdot24}X_3+b_{14\cdot23}X_4$，之解法解之。其標準方程式有四，可縮減爲三式。即

$$p_{13}=\sigma_1^2 b_{31.24}+p_{12}b_{32.14}+p_{14}b_{34.21}$$

$$p_{23}=p_{12}b_{31\cdot24}+\sigma_2^2 b_{32.14}+p_{24}b_{34.21}$$

$$p_{34}=p_{14}b_{31\cdot24}+p_{24}b_{32\cdot14}+\sigma_4^2 b_{34\cdot21}$$

以相當各値，(除出 σ_1^2 一項外其餘各値均與上例相同) 代入方程式得

$$-62.039=1188.688\,b_{31.24}-47.967\,b_{32.14}-47.519\,b_{34\cdot21}$$

$$2.790=-47.967b_{31.24}+8.3564b_{32\cdot14}+2.932b_{34.21}$$

$$2.063=-47.519b_{31\cdot24}+2.932b_{32.14}+7.7893b_{34.21}$$

解之，得下列各數値。

$$b_{31\cdot24}=-.0531$$

$$b_{32\cdot14}=+.0574$$

$b_{34\cdot 21}=-.0807$

淨相關係數之計算　上例中 $b_{13\cdot 24}=-7.394$; $b_{31.24}=-.0531$, $b_{31\cdot 24}$ 與 $b_{13.24}$ 之值，既已求得，代入公式

$$r_{13.24}=\sqrt{b_{13\cdot 24}\cdot b_{31\cdot 24}}$$

卽得

$$r_{13.24}=\sqrt{-7.394\times -.0531}$$
$$=-.6266$$

(r 之符號視各回歸係數之符號爲轉移，設二回歸係數俱爲正號，r 爲正值，設二回歸俱爲負號，則 r 爲負值。)

此 r 之數值，乃 Kansas 玉蜀黍產量和七月氣溫間之淨相關係數，所以測量此二變量之關係程度，假定六月氣溫及八月氣溫爲不變時。

計算局部相關之又一法　在平常簡單相關係數，僅及二變量者，可稱之爲0次之係數。此等係數，以符號 r_{12},r_{24} 等表之。二變量之淨相關係數，設有另一變量爲不變時，則稱爲一次之係數，以符號 $r_{12\cdot 3}$ $r_{24\cdot 3}$ 等表之。同理，設不變之變量爲 2,3,4,5,……………n 個時，則得 2,3,4,5,…………n 次之係數。

一次局部相關之公式，爲

$$r_{12\cdot 3}=\frac{r_{12}-r_{13}\cdot r_{23}}{\sqrt{1-r^2_{13}}\sqrt{1-r^2_{23}}}$$

二次局部相關之公式，爲

$$r_{12\cdot34}=\frac{r_{12\cdot3}-r_{14\cdot3}\cdot r_{24\cdot3}}{\sqrt{1-r^2_{14\cdot3}}\ \sqrt{1-r^2_{24\cdot3}}}$$

局部相關之普遍公式，爲

$$r_{12\cdot34\cdots\cdots n}=\frac{r_{12.34\cdots\cdots(n-1)}-r_{1n.345\cdots\cdots(n-1)}r_{2n.345\cdots(n-1)}}{\sqrt{1-r^2_{1n\cdot345\cdots\cdots(n-1)}}\sqrt{1-r^2_{2n.345\cdots\cdots(n-1)}}}$$

高次局部相關之係數，可由各低次之局部相關求之。此方法之計算，雖有系統，但手續麻煩，如能有製就之表(註)可查，可將計算手續減至最低限度。仍以上例之材料，舉例說明之。

在本例中當有三係數；$r_{12\cdot34}$；$r_{13\cdot24}$ 及 $r_{14\cdot23}$；其公式如下

$$r_{12.34}=\frac{r_{12.4}-r_{13.4}\ r_{23.4}}{(1-r^2_{13.4})^{1/2}(1-r^2_{23.4})^{1/2}}$$

$$r_{13.24}=\frac{r_{13.4}-r_{12.4}\cdot r_{32.4}}{(1-r^2_{12.4})^{1/2}(1-r_{32.4})^{1/2}}$$

$$r_{14.23}=\frac{r_{14.3}-r_{12.3}\cdot r_{42.3}}{(1-r^2_{12.3})^{1/2}(1-r^2_{42.3})^{1/2}}$$

一次係數之計算　計算第二次係數之前，當先從事於求得一次之係數，一次係數之公式，爲

(註)　Miner, J.R. Tables of $\sqrt{1-r^2}$ and $1-r^2$ for use in partial correlation and in trigonometry, Johns Hopkins press, Baltimore, Md., *1922*.

$$r_{12.3}=\frac{r_{12}-r_{13}\cdot r_{23}}{(1-r^2{}_{13})^{\frac{1}{2}}(1-r^2{}_{23})^{\frac{1}{2}}}$$

計算方法，可不從普通公式以求之，祇須依法排列如表48，這樣計算較有系統及省便。一次係數之計算，須先求出三個

表48　局部相關之一次係數計算法之示例

美國康省玉蜀黍產量與氣温

r_0次		$(1-r^2)^{\frac{1}{2}}$	分子之	分子之	分　母	r_1次	
符號	係　數		乘積項	全　部		符號	係　數
12	−.4314		−.2604	−.2210	.6653	12.3	−.3322
13	−.6968	.7173					
23	+.3737	.9275					
14	−.4937		−.1994	−.2943	.6873	14.3	−.4282
13	−.6968	.7173					
43	+.2862	.9582					
24	+.3633		+.1070	+.2563	.8887	24.3	+.2884
23	+.3737	.9275					
43	+.2862	.9582					
13	−.6968		−.1799	−.5169	.8130	13.2	−.6358
12	−.4814	.8765					
32	+.3737	.9275					
14	−.4937		−.1749	−.3188	.8166	14.2	−.3904
12	−.4814	.8765					
42	+.3633	.9317					
34	+.2852		+.1358	+.1504	.8642	34.2	+.1740
32	+.3737	.9275					
42	+.3633	.9317					
12	−.4814		−.1794	−.3020	.8102	12.4	−.3727
14	−.4937	.8696					
24	+.3633	.9317					
13	−.6968		−.1413	−.5555	.8333	13.4	−.6666
14	−.4937	.8696					
34	+.2862	.9582					
23	+.3737		+.1040	+.2697	.8928	23.4	+.3021
24	+.3633	.9317					
34	+.2862	.9582					
14	−.4937		−.1994	−.2943	.6873	14.3	−.4282
13	−.6968	.7173					
43	+.2862	.9582					
12	−.4814		−.2604	−.2210	.6653	12.3	−.3322
13	−.6968	.7173					
23	+.3737	.9275					
42	+.3633		+.1070	+.2563	.8887	42.3	+.2884
43	+.2762	.9532					
23	+.3837	.9275					

0 次之係數。

係數 $r_{23.4}$ 之值，當然和 $r_{32.4}$ 之值相符，$r_{34.2}$ 當然和 $r_{43.2}$ 相同，餘類推，是故此等量數，無需重行計算也。

二次係數之計算　三個一次係數既已求得，則二次係數之計算，卽可求得。計算方法，列表如次。

表49　局部相關之二次係數計算之示例

康省玉蜀黍產量及氣溫

r_1次		$(1-r^2)^{1/2}$	分子之乘積項	分子之全部	分母	r_2次	
符號	係數					符號	係數
12.3	−.3322		−.1225	−.2087	.8353	12.34	−.2412
14.3	−.4282	.9037					
24.3	+.2884	.9575					
13.2	−.6358		−.0679	−.5579	.9065	13.24	−.6265
14.2	−.3904	.9206					
34.2	+.1704	.9847					
14.2	−.3904		−.1106	−.2[illegible]8	.7601	14.23	−.3681
13.2	−.6358	.7719					
43.2	+.1704	.9847					
12.4	−.3727		−.2014	−.1713	.7706	12.34	−.2411
13.4	−.6666	.7454					
23.4	+.3021	.9533					
13.4	−.6666		−.1126	−.5540	.8877	13.24	−.6262
12.4	−.3727	.9280					
32.4	+.3021	.9533					
14.3	−.4282		−.0958	−.3324	.9031	14.23	−.3681
12.3	−.3322	.9432					
42.3	+.2884	.9575					

第 七 章

機率,常態曲線及確度測量

機率的意義　設有一事，試驗$a+b$次，其中a次成功，b次失敗，若每次試驗之方法一樣 Equally like, 則其成功之機率爲$\frac{a}{a+b}$，其失敗之機率爲$\frac{b}{a+b}$。換言之，機率就是成敗之機遇的比率。例有壹袋，內盛黑白球二種，爲數相同，則由袋內任取一球，取得黑球之機會與取得白球之機會完全相等，故黑球與白球取得之機率各爲½，

凡事僅有成功與失敗二途，故成功之機率和失敗之機率，二者之總和必爲一，一是表示確定性 Certainty 的符號。假設成功之機率爲p，則由機率之定義知其失敗之機率爲$1-p$;何則，因成功之機率爲$\frac{a}{a+b}=p$,則失敗之機率爲$\frac{b}{a+b}=1-\frac{a}{a+b}=1-p$。$1-p$常以$q$表之。

解釋機率應注意之點　在運用機率及解釋機率時，尤須首先明白吾人對於機率應有之觀念，否則觀念不明，遑論其他，例若我們所謂擲銅元得背的機率是½，并不是說在二擲之中，我們準能擲得其背面一次，也不是說在任何次擲中，得背得字的次數，須準準相等。一連擲得幾個背，是可能的，如欲一連擲得十個

背的機率，是$\frac{1}{1024}$，幷不是零。不過要使連續擲得的次數愈多，其機率愈微而已，然每次試驗之初，其欲得背得面之機率應各等於½，而並不隨其前次試驗之結果如何而或有變更，例如有背已連來了三四次，再下一次試驗之初求其來背的機率，仍是½。已經來過的次數再多些，多到任何次，下次的機率，仍是這樣。以前擲得怎樣，和現在或將來是沒有關係的。如謂背已經接連來過好幾次，而就謂下一次，來字的機率比較的大些，那就完全把機率的意義誤解了。

不能並立之事實 (mutually exclusive events) 設有甲乙兩種現象，若甲現象實現，而他一現象乙必不能同時發現者，則甲乙爲兩不能並立之現象。如骰有六面，擲時僅現出一面，其於擲出 1 點時，其他 2, 3, 4, 5, 6 卽不能同時擲出。如果二件事，是不能並立的，則得到這個或那個的機率，就是二個單獨發現的機率的和數。例如擲一個銅幣，得着背的機率是½，得着字的機率，也是½，得着背或字的機率，就是二者之和 (½+½=1) 擲骰一粒得着 5 的機率是⅙，得着 6 的機率也是⅙，得着 5 或 6 的機率，就是⅙+⅙=⅓。

假如用代數符號來表示，設有三件不同的事，他們各自成功之機率，以同分母之分數表之，爲$\frac{a_1}{\omega}$，$\frac{a_2}{d}$，$\frac{a_3}{d}$，則在 d 次內，三事成功之機率爲 a_1, a_2, a_3。惟各事原不能並立，故知 d 次內三事任何一事成功之機率爲 $a_1+a_2+a_3$，故三事任何一事

出現之機率，為$\frac{a_1+a_2+a_3}{d}$即$\frac{a_1}{d}+\frac{a_2}{d}+\frac{a_3}{d}$。

不相關之事實 (Independent events)　設有二事互不相關，而可以同時並立者，則二事成功之機遇，等於各事成功機遇之乘積。如由甲囊中取出一赤球之機率為$\frac{2}{5}$由乙囊取出一赤球之機率為$\frac{3}{4}$，則由甲乙二囊，同時各取出一赤球之機率，為$\frac{2}{5}\times\frac{3}{4}=\frac{6}{20}$。

設第一事於a_1+b_1次試驗中，其成功有a_1次，失敗有b_1次，又假使第二事於a_2+b_2次試驗中其成功有a_2次，失敗有b_2次，惟(a_1+b_1)與(a_2+b_2)兩事組合之總數為$(a_1+b_1)(a_2+b_2)$，而其中祇有a_1a_2為二事同時成功者，故其機率為$\frac{a_1a_2}{(a_1+b_1)(a_2+b_2)}$即$\left(\frac{a_1}{a_1+b_1}\right)\times\left(\frac{a_2}{a_2+b_2}\right)$

試推論之，設以二件不相關的偶然事件，其成就之機率，各為p_1及p_2，則欲令其同時失敗之機率為$(1-p_1)(1-p_2)$，又第一事成，第二事失敗之機率為$p_1(1-p_2)$。第一事敗第二事成之機率為$p_2(1-p_1)$。此三者相加，為$(1-p_1)(1-p_2)+p_1(1-p_2)+p_2(1-p_1)=1-p_1p_2$，由是知不相關之二事，其同時成功之機率為$p_1p_2$，而其非同時成功之機率為$1-p_1p_2$同法不相關之若干事，若其各自成功之機率為$p_1$，$p_2$，$p_3$，………，則其同時並成之機率為$p_1p_2p_3$………，其同時失

敗之機率爲。$(1-p_1)(1-p_2)(1-p_3)\cdots\cdots\cdots\cdots$，

多次試驗之機率與二項展開式　既知1次試驗時成就之機率，則於n次試驗時，其間有一次二次三次迄n次成就之機率，亦可推算而得。何則，設一事成就之機率爲p_1則其不成之機率爲$1-p=q$，於n次中有r次成功，卽有$n-r$次不成，而其機率爲$p^r q^{n-r}$。但試驗n次，其中僅有r次成之組合爲C_n^r，而此等組合之機率相等，且爲不能並立（互拒）者。由是知於n次試驗中，內有r次成就之機遇，爲$C_n^r p^r q^{n-r}$

試將$(p+q)^n$依二項式定理展開之，則亦得謂該展開式之第一，第二，$\cdots\cdots\cdots\cdots$第$n+1$項，卽等於試驗n次中，有n，$n-1$，$\cdots\cdots 1,0$，成功之機率

$$(p+q)^n=p^n+C_n^1 p^{n-1} q+C_n^2 p^{n-2} q^2+\cdots\cdots\cdots\cdots\cdots\cdots$$

$$+C_n^r p^r q^{n-r}+\cdots\cdots\cdots\cdots+q^n$$

其連次之項，如p^n爲於n次中有n次成就之機率，$C_n^1 p^{n-1} q$爲於n次中有$n-1$次之成就機率，$C_n^2 p^{n-2} q^2$爲於n次中有$n-2$次成就之機率，以下類推。C_n^r爲n個字母取r字一組之組合數目，應等於

$$C_n^r=\frac{n(n-1)(n-2)\cdots\cdots\cdots(n-r+1)}{1\cdot 2\cdot 3\cdots\cdots\cdots\cdots\cdots\cdots\cdots r}$$

$$=\frac{n!}{n!(n-r)!}$$

$n!\ r!$ 爲 n,r 之累乘數之符號

一個擲幣的試驗　以上所論，皆就理論而言，實際上試驗的結果，即觀察的機率，也和理論所得的機率，極相近似。這是很可注意的。玆且引衛斯脫衛 F.W. Westaway 擲幣試驗的結果爲例。衛氏取十個銅幣，作兩組拋擲，每組拋 1024 次，合計之爲 2048 次，其結果表列之如下。

表50　擲幣之試驗

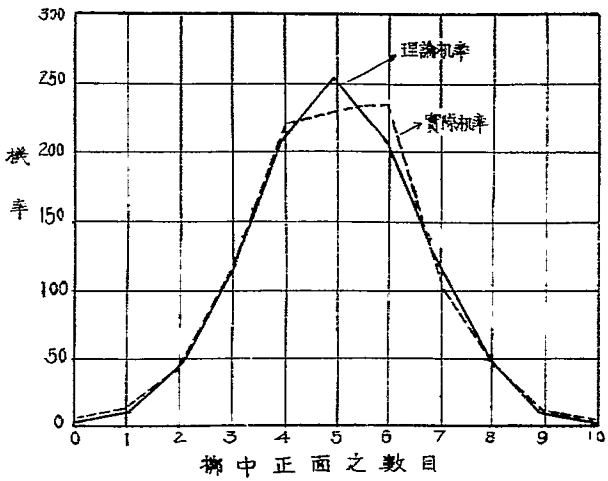
拋擲的性質	理論的次數	第一組	第二組	平均	差異
10面　0底	$C_{10}^{0}=1$	4	0	2	+1
9,,　1,,	$C_{10}^{1}=10$	20	6	13	+3
8,,　2,,	$C_{10}^{2}=45$	40	40	40	−5
7,,　3,,	$C_{10}^{3}=120$	83	150	116½	−3½
6,,　4,,	$C_{10}^{4}=210$	224	222	223	+13
5,,　5,,	$C_{10}^{5}=252$	250	209	229½	−22½
4,,　6,,	$C_{10}^{6}=210$	242	222	232	+22
3,,　7,,	$C_{10}^{7}=120$	115	107	111	−9
2,,　8,,	$C_{10}^{8}=45$	28	60	44	−1
1,,　9,,	$C_{10}^{9}=10$	14	6	10	0
0,,　10,,	$C_{10}^{10}=1$	4	2	3	+2

表中理論的次數，就是依二項式展開而求得。在這試驗中，銅幣單次拋擲的全數爲 $2048\times10=20480$，按理論其半數應爲幣面，而實際試驗所得爲 10234（第一組 5098，第二組

5136）可見每一試驗實際的結果，與理論極爲近似。

現在將理論的次數，與實際的次數，繪圖於下，

圖36　理論機率與實際機率之比較



常態曲線（註）　圖36卽所以表示擲幣試驗的理論次數，乃一完全對稱之十邊形（底線除出），所有邊數，等於不相關事項之數。設爲六件事項，則爲六邊形，設爲二十件不相關事項，

（註）　名爲常態差誤曲線，因爲：設將變量，如體重樹葉之長等，測量多次，則每個測量，必相差若干，有相差很大者，有相差很小者，若案此種差誤之大小，將他們製一曲線，卽成一常態差誤曲線。又名鐘形曲線（Bell shaped Curve）象其形也，又名機率曲線 Normal Probability Curve 因與機率現象相合也，亦稱 Gaussian Curve 所以紀念發現此線之數學家 Causs 氏也

則爲二十邊形，餘可類推。很明顯的，n 愈增加，邊數亦相依而增加，而所得之曲線，用二項展開式之連縮各項代表者，亦愈覺修勻。設 n 爲無窮大，可得一完全修勻之曲線。這就是常態差誤曲線 Normal curve of error，簡稱常態曲線。此種曲線，在從前許多科學家的觀念中，以爲是形容各種數量的分配的基本法則，不過現在已經發見那些觀念，是不正確的。常態曲線，在現在，亦不過看作曲線中的一種，可以形容次數分配。這種曲線，仍然是很重要的一種，對於他的特質如何，研究統計的人，不能不有相當的了解。此種曲線，雖則可以從二項式的展開來配合，不過實際上因爲那樣算法，手續麻煩，可根據此曲線之積分表以決定其理論次數。常態曲線普通之公式爲

$$y = y_o e^{-\frac{x^2}{2\sigma^2}}$$

在此式內，x 爲量數與平均數之差數之用標準差表示者，y 爲縱距并以表所希冀之次數，σ 爲標準差，e 爲常數2.7182818，即納氏Napier 對數之底數，而 y_o 爲立於平均數上之最高縱距。

根據此方程式，則知任何一差數後，即可決定其相當縱線。平均數上之縱線，可用下式表出之。

$$y_o = \frac{N}{\sigma\sqrt{2\pi}}$$

在此式內，N爲分配之總次數，σ 爲標準差，π 爲常數3.1416。根據此方程式，可以決定平均數上之次數，此次數可用實得

之次數，及標準差爲單位者表出之。故機率曲線較詳之方程式，爲

$$y_o=\frac{N}{\sigma\sqrt{2\pi}}e^{\frac{-x^2}{2\sigma^2}}$$

常態曲線之配合　常態曲綫之配合，手續頗繁。因公式中之 σ 及 y_o 必須有指定之價值，曲線之各縱線(即在底線上由平均數至標準差分數部分之距離向上豎立之各縱線)，常爲 y_o 之分數，而有一定之比例，我們如將 x 之指定各值，與曲線下相當縱線 (y) 各值，一一求出，列爲一表。則以後計算，非常便利。這種表，現在已經有現成製就的，見 Pearson: Table for Statistician and Biometrician，附錄中亦備一格。我們配合時，祗要檢表即可，例如由平均數至 1σ 向上引出之縱距，等於由平均數所在之點，向上引出之縱距 y_o 高度之60653 。至 2.8σ 引出之縱線，等於 y_o 高度之 .01984 y_o 等。惟有必須注意者，該表中各數之計算，x 係以 σ 爲單位，y 以 y_o 爲單位，而 σ 及 y_o 均假定其等於 1；x 及 y 之各值，均爲 σ 及 y_o 之分數部分。各縱線之高，如已求得，乃將各頂端用線連綴，即得一多邊之常態次數圖。設所採 σ 分數部分愈小，則此曲綫愈覺修勻。

實際次數與常態次數之比較　繪作理論常態曲線圖的方法，已如上述，假使拿實際事實的分配，與理論的分配，兩相比較，則可拿二者的圖形，重疊起來，看他們是否湊合。繪作這

種比較用的圖，有三個重要的條件。第1,理論分配的平均數地位，與實際分配的平均數點，在同一地位上。而二圖的標準差距離之測量，都從這點算起。第2,二圖之標準差，須用同一距離代表之。第3,理論的分配之平均縱線y_0，可用$y_0=\frac{N}{\sigma\sqrt{2\pi}}$公式求出，因爲N, σ及π都是已知數，而$y$之各數值，可由$y_0$之數值乘附表所載$\frac{x}{\sigma}$之數值而得之。

茲根據表51所載之事實，先繪一次數多邊圖，再配合一常態曲線，如第37圖。

表51　常態曲線之縱距計算法

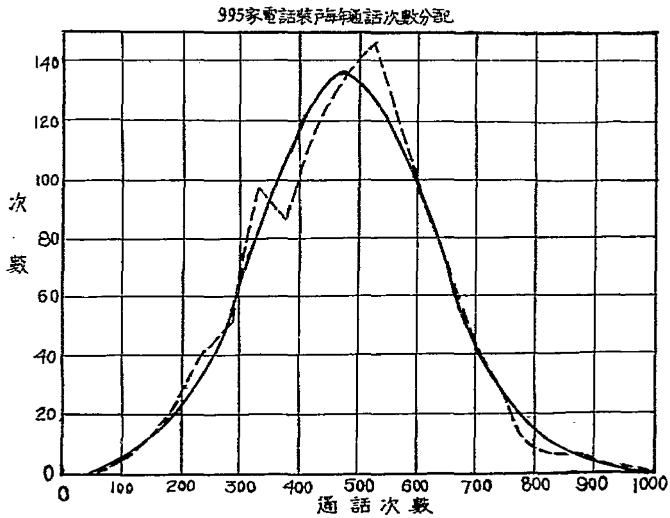
美國995家電話裝戶每年通話次數分配

通話數目 m	次數 f	對假定原點之離差 d'	fd'	$f(d')^2$	離M之對差(組距) x	$\frac{x}{\sigma}$	$\frac{y}{y_0}$	縱距 y_c
25	0	−10	0	0	−9.0392	−3.06	.00153	.21
75	1	− 9	− 9	81	−8.0392	−2.72	.02474	3.33
125	9	− 8	− 72	576	−7.0392	−2.38	.05888	7.91
175	19	− 7	−133	931	−6.0392	−2.05	.12230	16.44
225	38	− 6	−228	1,368	−5.0392	−1.71	.23176	31.15
275	50	− 5	−250	1,250	−4.0392	−1.37	.39123	52.59
325	95	− 4	−380	1,520	−3.0392	−1.03	.58834	79.08
375	85	− 3	−255	765	−2.0392	− .69	.78817	105.95
425	115	− 2	−230	460	−1.0392	− .35	.94055	126.43
475	132	− 1	−132	132	− .0392	− .01	.99995	134.41
525	144	0	0	0	.9608	.32	.94856	127.51
575	116	1	116	116	1.9608	.66	.80428	108.11
625	79	2	158	316	2.9608	1.00	.60653	81.53
675	54	3	162	486	3.9608	1.34	.40747	54.77
725	31	4	124	496	4.9608	1.68	.24385	32.78
775	11	5	55	275	5.9608	2.02	.13000	17,47
825	5	6	30	180	6.9608	2.36	.06174	8.30
875	6	7	42	294	7 9908	2.70	.02612	3.51
925	2	8	16	128	8.9608	3.03	.01015	1.36
975	1	9	9	81	9.9608	3.37	.00342	.46
1025	1	10	10	100	10.9608	3.71	.00103	.14
1075	1	11	11	121	11.9608	4.05	.00027	.04
	995		−956	9676				993.48

$$\frac{\Sigma f(d')}{N}=\frac{-956}{995}=.9608 \qquad \left(\frac{\Sigma fd'}{N}\right)^2=.9231$$

$$\sigma^2=\frac{9676}{995}-.9231=8.8015 \qquad \sigma=2.953$$

圖37 常態曲線之配合



理論次數決定之又一法 在圖37所繪的實際分配和常態曲線，就大致觀察，他們的配合，還很適當，雖則還有幾處有很顯著的差異。既有差異，我們就須問那種差異的原因何在。他們的原因，不外乎二端，或者是由於取樣之差誤，或者是由於統計事項的分配本身不合於常態之形式。我們拿實際次數與理論次數，作一比較，將二者差異最大之各點而研究之，即可斷定

。不過在比較之先，必先有更精密的理論次數之計算。常態曲線下之縱線，測量次數，不甚精密，惟由曲線下之面積，以決定理論次數，才算比較精密。

一曲線下的全面積，可以拿來代表次數的總數。設能知某一部分對於全面積的比例，則此部分所代表之次數，即易求得。各部分對全面積之比例，可從按積分法則製就之表上檢得之。這種表，已經算學家算好，Dr. W. F. Sheppard 所製之群表，在美國很通用，在 Pearson: Table for Statistician and Biometrician 書中亦有此表。本書附錄中，亦備一格。常態曲線對最高縱線之二邊，既爲對稱，故表中各值，對於正值負值，都可適用。應用此表時，宜取其從算術平均數之離差，以標準差爲單位。在任何$\frac{x}{\sigma}$上立一縱線 y，則 y_o 與 y 間曲線下之面積多少，檢表即得。例如：在最高縱線和在 $+1\sigma$ 地位所立之縱線間，在常態分配中，對於全體事項所占之比例如何，我們可以在$\frac{x}{\sigma}$欄中檢視至 1.0，而得其相對數爲 .34134。這是一個比例，設以百分數表之，即得 34.134%，圖 38 陰影部分 A，即表示此面積與曲線下全面積之關係。

上面所討論者，乃平均點之縱線，與任何一$\frac{x}{\sigma}$點上之縱線間之次數比例，在橫線上任何二點所立直線間之面積，亦可如法推求。例如：1.4σ 與 $2.\sigma$ 間之次數之百分數，從附錄表先檢出 y_o 與 1.4σ 間之次數百分數，次檢出 y_o 與 2σ 之次數百分

數，於是以前者求得之數減後者，即得 1.4σ 與 2σ 間之次數比例。其結果如下：

$$
\begin{array}{rl}
0 \text{ 至 } 2.0\,\sigma = & 47.73\% \\
0 \text{ 至 } 1.4\,\sigma = & 41.92\% \\
\hline
1.4\sigma \text{ 至 } 2.0\,\sigma = & 5.81\%
\end{array}
$$

此項次數百分數，以任何特定之次數分配之總次數乘之，即可得該分配之各理論次數。圖38陰影部分B，代表上項結果之面積。

計算理論次數之手續，既如上例說明，茲為讀者明瞭起見

(205頁註) $\Sigma f = \mathrm{N} = (q+p)^n = 1$

$$\mathrm{M} = \frac{\Sigma f x}{\mathrm{N}} = \Sigma f x$$

$$= np\left[q^{n-1} + (n-1)q^{n-2}p + \frac{(n-1)(n-2)}{1\cdot 2}q^{n-3}p^2 + \cdots\cdots\right]$$

$$= np(q+p)^{n-1} = np$$

$$\sigma^2 = \frac{\Sigma f x^2}{\mathrm{N}} - \left(\frac{\Sigma f x}{\mathrm{N}}\right)^2 = \Sigma f x^2 - n^2p^2$$

$$= np\left[q^{n-1} + 2(n-1)q^{n-2}p + 3\,\frac{(n-1)(n-2)}{1\cdot 2}q^{n-3}p^2 + \cdots\cdots\right] - n^2p^2$$

$$= np\left[(q+p)^{n-1} + (n-1)p(q+p)^{n-2}\right] - n^2p^2$$

$$= np[1+(n-1)p] - n^2p^2)$$

$$= np - np^2 = np(1-p) = npq$$

$$\therefore \sigma = \sqrt{\mathrm{N}pq}$$

，不殫其煩，再以電話裝戶通話次數之資料，示例於表52

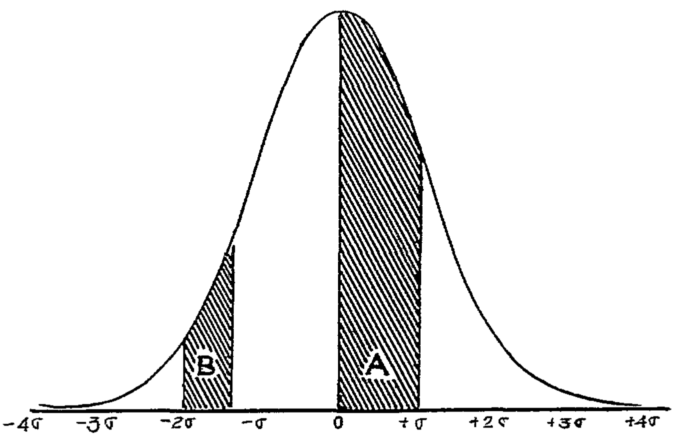
「標樣之標準差誤」之測驗　較爲精密之理論次數，旣已求得，我們就可測驗常態曲線配合之結果，是否適合，先擇取實際次數與理論次數最大差異之各點，然後決定其差異之意義何在。假如機率已經知道，標準差卽可從公式 $\sigma=\sqrt{Npq}$（註）計算。式中 p 爲成功之機率，q 爲失敗之機率，在次數分配研

表52　面積表計算理論次數之示例

組限	對平均數之離差 $\frac{x}{\sigma}$	yo與y間面積之比例 $\frac{x}{\sigma}$ 單位	yo與y間包括之項數 $\frac{x}{\sigma}$ 單位	理論次數	
0	−3.23	.4993810	496.88		
50	−2.89	.4980738	495.58	0—50	1.92*
100	−2.55	.4946130	492.14	50—100	3.44
150	−2.22	.4867906	484.36	110—150	7.78
200	−1.88	.4699450	467.60	150—200	16.76
250	−1.54	.3482198	476.03	200—250	31.57
300	−1.20	.3849303	383.01	250—300	53.02
350	−.86	.3051055	303.58	300—350	78.63
400	−.52	.1984582	197.48	350—400	106.90
450	−.18	.0714237	71.07	400—450	126.41
500	+.16	.0635595	63.24	450—500	124.31
550	+.495	.1896931	188.74	500—550	125.50
600	+.83	.2967376	295.25	550—600	107.51
650	+1.17	.3789995	377.10	600—650	81.85
700	+1.51	.4344783	432.31	650—700	55.21
750	+1.85	.4678432	465.50	700—750	33.19
800	+2.19	.4857379	483.31	750—800	17.81
850	+2.53	.4942979	491.83	800—850	8.52
900	+2.87	.4979476	495.46	850—900	3.63
950	+3.20	.4993129	496.82	900—950	1.36
1000	+3.54	.4997999	497.30	950—1000	.48
1050	+3.88	.4999478	497.45	1000—1050	.15
1100	+4.22	.4999878	497.49		
				1050以上	.05
					995.00

* 在 -3.23σ 以下之理論分配爲.62，惟在本例中，則無甚意義，故以之加於 0—50 之理論次數上。

圖38　常態曲線下之面積測量



究中，我們可以$\frac{f}{N}$代p（此f表示在x軸上選定之一點之理論次數，N爲事項之總數，）於是q當爲$\frac{n-f}{n}$將各值代入上項公式，卽得

$$\sigma_S=\sqrt{N\times\frac{f}{N}\times\left(\frac{N-f}{N}\right)}$$

$$=\sqrt{\frac{f(N-f)}{N}}$$

此種測量名爲標樣之標準差誤。

此公式之意義如何，請解釋如下：若理論次數，與實際次數之差，較三倍標樣之標準差爲小，則此差誤之發現，乃由於

標樣之變動，可視之爲無足重輕，而實際分配亦可以常態視之，若理論次數與實際次數之差，較三倍標樣之標準差爲大，則此種差誤，必非由於標樣之變動，而由於他種原因之影響。那種資料之實際分配，就不能以常態分配之法則來規律之。

電話裝戶之例，理論次數與觀察次數最大差異之各點，如下：

(理論次數以 f 表之，觀察次數以 f_o 表之)

m	f_o	f	f_o-f
325	95	78.63	+16.37
375	85	106.90	−21.90
525	144	125.50	+18.50
775	11	17.81	− 6.81

其第 1 點之'標樣之標準差誤'，爲

$$\sigma_S=\sqrt{\frac{78.63\ (995-78.63)}{995}}=8.51$$

其餘各點，以同樣方法計算之，得下列各值。

中點 325 一組之 $\sigma_S=8.51$

中點 375 一組之 $\sigma_S=9.73$

中點 525 一組之 $\sigma_S=10.47$

中點 775 一組之 $\sigma_S=4.18$

第一點之差數爲 16.37，尚不及 σ_S 之三倍爲大 (因 $8.51\times3=25.53$)，在最大一點 −21.90 用同法測驗之，亦較 σ_S 之三倍

為小(因 $21.90<9.73\times 3=29.19$)。

確度測量　常態曲線最重要之用途，在乎評定結果之確度。我們假使在全體事實中，取若干標樣，而求其平均數，此平均數，除出偶爾湊巧，決不能與眞正平均數完全相同，祇能得近似之值。所謂「眞正」平均數，卽由全體事實求出之平均數。從標樣求出之平均數，稱之爲「實得」平均數。其他各種量數，亦有「實得」「眞正」之分。「實得」量數與「眞正」量數相差如何，卽統計上所稱確度差量。

「實得」量數之確度，全恃原有資料的離差數，和標樣的大小而決定。因爲原有資料的離差愈小，則「眞正」與「實得」量數之差亦愈小，就是確度增高；標樣愈大，則「實得」數與「眞正」相近，卽二者之差度小，而確度高。

確度測量有二種方法，一以 σ 表之，一以概差表之。用表示之算術平均數確度公式，爲

$$\sigma_m=\frac{\sigma_{dis}}{\sqrt{N}}$$

式中之 σ_m 代表算術平均數之標準差，N 代表次數之總和，σ_{dis} 爲標樣之標準差，卽實得標準差。

通常確度之以概差表示者，較爲常用。在常態分配中，概差與 σ 之關係，爲 .6745，故以概差表示之算術平均數確度之公式，爲

$$P.E._m = .6745\frac{\sigma_{dis}}{\sqrt{N}}$$

試舉一例以明其意義，假設：某城有工廠工人100,000,求其平均每週工資。將全體工人來研究，事實上有所不許。所以選取 900 工人作標樣，得其平均工資為$27.50試問其確度如何？設資料之標準差為$2.00.則

$$P.E._m = .6745\frac{\$2.00}{\sqrt{900}}$$
$$= \$.045$$

因此，這城工人的平均工資，便在$27.50±.045 之間。這就是說，假使再取同樣大的標樣，所得結果，在 ±.045 之間者，機率為 1 與 1（因±1 P.E 包含全體量數之 $\frac{1}{2}$）。所得結果在 ± .090 者，其機率為4.5 與 1 之比（因$\frac{82.26}{100}$量數在±2P.E之內）。茲將 C. B. Davenport 所列之概差之機率表轉錄如下：

±1P.E,=	機率相等
±2P.E,=	4.5 與 1 之比
±3P.E,=	21 與 1 之比
±4P.E,=	142 與 1 之比
±5P.E,=	1,310 與 1 之比
±6P.E,=	19,200 與 1 之比
±7P.E,=	420,000 與 1 之比
±8P.E,=	17,000,000 與 1 之比
±9P.E,=	100,000,000 與 1 之比

至於確度之以 σ 表示者，其解釋與算法，和概差表示者相似。其不同者，卽 σ 包括全體之$\frac{2}{3}$而概差則包括全體之$\frac{1}{2}$上述之解釋，可適用於任何確度測量，各量數結果之確度公式，總錄於下：

1.中位數

$$\sigma_{md}=1.25331\frac{\sigma}{\sqrt{N}} \qquad P.E._{md}=.84535\frac{\sigma}{\sqrt{N}}$$

2.第一四分位數(Q_3之數量同此)

$$\sigma_{Q_1}=1.36263\frac{\sigma}{\sqrt{N}} \qquad P.E._{Q_1}=.91908\frac{\sigma}{\sqrt{N}}$$

3.標準差

$$\sigma_{\sigma}=\frac{\sigma}{\sqrt{2N}} \qquad P.E._{\sigma}=.6745\frac{\sigma}{\sqrt{2N}}$$

4.相關係數(相關比率亦可適用)

$$\sigma_{r}=\frac{1-r^2}{\sqrt{N}} \qquad P.E._{r}=.6745\frac{1-r^2}{\sqrt{N}}$$

5.回歸係數之標準差誤

$$\sigma_{b_{12}}=\frac{\sigma_1\sqrt{1-r^2}}{\sigma_2\sqrt{N}}$$

6.ζ 之標準差誤

$$\sigma_{\zeta}=2\sqrt{\frac{\zeta}{N}}\sqrt{(1-\eta^2)^2-(1-r^2)^2+1}$$

上列確度測量之公式，亦有相當之限制。假使標樣的項數少於 15，求標準差誤，卽不能適用。在相關係數之公式，以25項爲最少限度云。

經濟統計學

第三編

指數之編製

(The Making of Index Numbers)

第三編

指數之編製

第一章

通論

指數之意義　指數是一種統計的數字，所以指示一羣變量之平均變動，以便比較某現象在各地域或各時期的情形，或在同一時間同一地域比較不同情形之現象。例若：柴米油鹽等各物之價格日有變動，是爲一羣變量；甲乙丙丁等各工人工資，月有變動，又爲一羣變量。取柴米油鹽等各物價格之變動而求得一平均變動，就得一物價指數。取甲乙丙丁等各工人工資之變動，而求其其平均變動，就得一工資指數。

指數之應用　在三四十年前，指數之應用，只限於物價一類，如今却差不多一切社會經濟現象，凡可用數字以表明之，

而又與時間有關係者，都可用指數整理之，以表明其相對狀况。所以除吾人習見之批發物價指數，零售物價指數，生活費指數以外，又有工資，生產，利潤，雇傭，工業活動 (Industrial Activity)，公用事業，國外貿易，國外匯兌，海陸運費率，股票證劵等，莫不有指數。他如指數之用於特種產業者，如阿盤沙之建築費指數。亦有用於一特定公司者，如美國造紙公司之造紙費指數。甚有用於一特定公司之一部分者，如霍爾脫Holt公司之教科書指數。指數應用之廣，可以想見。是故統計家謂：『我們人類現在是生於指數之時代，』誠係實情，並非驚人之談。指數之編製，不但可供學理研究上之資助，且可為人事舉措設施之憑依，如證劵價格之指數，為投資者及銀行家所切需，以覘各種證劵價格變動之趨勢，而定其投資及營業之方針。又如生活費指數，可為伸縮工資之標準，為解決勞資糾紛之工具。單就英國而言，已有三百萬以上的工人，他們的工資，是隨物價指數的升降而為增減，這就是所謂升降率制度 Sliding Scale System是。

簡單公式之六類

指數公式可分為六大類：綜合類，算術平均類，倒數平均類，幾何平均數，中位數類，及乘數類六種。每類之中，又可分為簡單和加權二種。乘數類實際不能應用，可略而不述，其

餘五種簡單的先分述於下，至於加權的公式，於下節敍述之：

簡單綜合法　編製簡單綜合指數，是以本期之物價相加，再以各期所得之結果，與之比較，以測量物價變化。其公式爲

$$\frac{P_1}{P_0}=\frac{p_1{}'+p_1{}''+p_1{}'''+\cdots\cdots\cdots\cdots}{p_0{}'+p_0{}''+p_0{}'''+\cdots\cdots\cdots\cdots}\times 100$$

$$\frac{P_1}{P_0}=\frac{\Sigma p_1}{\Sigma p_0}\times 100$$

式中符號，須待說明。

p_0爲基期物價水準，

p_1爲計算期物價水準，

$p_0{}'$，$p_0{}''$，$p_0{}'''$，…………爲基期各物價格，

$p_1{}'$，$p_1{}''$，$p_1{}'''$，…………爲計算期各物價格，

表53　物價表

類別	項　目	單位	1926之物價 (p_0)	1930之物價 (p_1)
原料品	棉花	担	14.20	11.40
	小麥	担	6.00	4.90
	黃豆	担	4.60	4.60
	鋼鐵	担	3.32	3.33
	羊毛	担	30.00	30.85
	鷄蛋	千個	25.00	21.10
生產品	棉紗	包	78.00	90.60
	桐油	担	10.15	13.20
	紙	令	9.75	11.25
	水泥	桶	5.60	7.60
消費品	啤酒	箱	12.00	14.40
	肥皂	箱	2.50	4.18
	麵粉	包	4.40	3.40
	紙烟	箱	37.50	47.00

試以表53假定物價表之資料，用綜合法以編指數，可得各類

指數及總指數於下表。

	Σp (1926)	Σp (1930)	指數 (1926=100)
原料品	83.12	75.53	91
生產品	103.50	122.05	118
消費品	56.40	68.93	122
總指數	243.02	266.51	110

應用此種方法，計算指數，顯然有好幾處弱點，第一點，這種指數，將不同單位之物品，多加在一起，似乎不合算學的理論。第二點，此公式求得之結果，依物價所採之單位而轉移。這種現象，在統計學的術語上可稱爲無定性 (Indeterminate)。例若物價表中，雞蛋價格以千個爲單位，假如改以個爲單位，此種改變，應與指數完全無關影響。因爲蛋千個值 25 元，和蛋一個值 .025 元，於價格上仍屬相等。然用本法以求指數，得食品類指數爲 94，總指數 113，與上表之結果頗有出入。第三點，這方法既爲無定性，公式在名義上是簡單公式，但實際上却無異加權，而所加權數，又幷無邏輯的根據。每種貨物影響於結果，全視其交易的單位價格。單位定得高，無異是加以權重。有人的意見，以爲使一切物品，化爲同一之單位，然後再行總合而求指數。不過這方法，難以實行，卽使能實行，仍是以暴易暴，將值昂的物品，大加其權重，依然是不合邏輯。

價比算術平均法　係將基年各項物價，除本年各項物價而得價比，再求這些價比的算術平均數，卽得指數。公式爲

$$\frac{\dfrac{p_1{}'}{p_0{}'}+\dfrac{p_1{}''}{p_0{}''}+\dfrac{p_1{}'''}{p_0{}'''}+\cdots\cdots\cdots}{N}=\frac{\Sigma\left(\dfrac{p_1}{p_0}\right)}{N}$$

$\dfrac{p_1{}'}{p_0{}'}$，$\dfrac{p_1{}''}{p_0{}''}$，$\dfrac{p_1{}'''}{p_0{}'''}$，………為計算期對基期之各物價比

仍用表53之材料，以 1926 年為基期，計算 1930 年之物價指數，示例於表 54 。

表54　價比算術平均法之計算

類別	項目	單位	p_0	p_1	$\dfrac{p_1}{p_0}$
原料品	棉花	担	14.20	11.40	80.3
	小麥	担	6.00	4.90	81.7
	黃豆	担	4.60	4.00	87.0
	鋼鐵	担	3.32	3.33	100.3
	羊毛	担	30.00	30.85	102.8
	鷄蛋	千個	25.00	21.10	84.4
	小計				536.5
生產品	棉紗	包	78.00	90.00	115.4
	桐油	担	10.15	13.20	130.0
	紙	令	9.75	11.25	115.4
	水泥	桶	5.60	7.60	135.7
	小計				496.8
消費品	啤酒	箱	12.00	14.40	120.0
	肥皂	箱	2.50	4.18	167.2
	麵粉	包	4.40	3.40	77.3
	紙烟	箱	37.50	47.00	125.3
	小計				489.8
總計					1522.8

原料品指數：$\dfrac{\Sigma\left(\dfrac{p_1}{p_0}\right)}{N}=\dfrac{536.5}{6}=89.4$

生產品指數：$\dfrac{\Sigma\left(\dfrac{p_1}{p_0}\right)}{N}=\dfrac{496.5}{4}=124.1$

消費品指數：$\dfrac{\Sigma\left(\dfrac{p_1}{p_0}\right)}{N}=\dfrac{489.8}{4}=122.5$

總指數：$\dfrac{\Sigma\left(\dfrac{p_1}{p_0}\right)}{N}=\dfrac{1522.8}{14}=108.8$

用上述方法，計算指數，驟然一看，既是簡單指數，當然無權數之意味。其實是隱含着不合理之權數，每項貨品所加權數，就是在基期（1926）每種物品每洋百元所販賣之數量。上例所用之權數，即爲下列各數量。

棉花	7.0 担
小麥	16.7 担
黃豆	21.7 担
生鉄	30.1 担
羊毛	3.3 担
鷄蛋	4.0千個
棉紗	1.3 包
桐油	9.9 担
紙	10.3 令
水泥	17.9 桶
啤酒	8.3 箱
肥皂	40 0 箱
麵粉	22.8 包
紙烟	2.7 箱

價比的倒數平均法　倒數平均數之特質，已於討論平均數時論及，我們還記得，倒數平均數是每個變量的倒數的算術平均數之倒數。在現在，各個變量是物價之價比，作$\dfrac{p_1}{p_0}$之形式。此價比之倒數，當爲$\dfrac{p_0}{p_1}$，故N 個價比之倒數平均數之公

式，爲

$$\frac{1}{\mathrm{H}}=\frac{\frac{p_0{}'}{p_1{}'}+\frac{p_0{}''}{p_1{}''}+\frac{p_0{}'''}{p_1{}'''}+\cdots\cdots\cdots\cdots}{\mathrm{N}}$$

或

$$\mathrm{H}=\frac{\mathrm{N}}{\Sigma\left(\frac{p_0}{p_1}\right)}$$

計算方法，示例於表55。

表55 倒數平均法之計算

類別	項目	單位	p_0	p_1	$\frac{p_0}{p_1}$
原料品	棉花	担	14.20	11.40	1.2456
	小麥	担	6.00	4.90	1.2245
	黃豆	担	4.60	4.00	1.1500
	鋼鐵	担	3.32	3.33	.9970
	羊毛	担	30.00	30.85	.9724
	鷄蛋	千個	25.00	21.10	1.1848
	小計				6.7743
生產品	棉紗	包	78.00	90.00	.8667
	桐油	担	10.15	13.20	.7689
	紙	令	9.75	11.25	.8667
	水泥	桶	5.60	7.60	.7368
	小計				3.2391
消費品	啤酒	箱	12.00	14.40	.8333
	肥皂	箱	2.50	4.18	.5981
	麵粉	包	4.40	3.40	1.2941
	紙烟	箱	37.50	47.00	.7979
	小計				3.5234
總計					13.5368

原料品指數：$\frac{\mathrm{N}}{\Sigma\left(\frac{p_0}{p_1}\right)}=\frac{6}{6.7743}=88.6$

生產品指數：$\frac{\mathrm{N}}{\Sigma\left(\frac{p_0}{p_1}\right)}=\frac{4}{3.2391}=123.5$

消費品指數：$\dfrac{N}{\Sigma\left(\dfrac{p_0}{p_1}\right)}=\dfrac{4}{3.5234}=113.5$

總指數：$\dfrac{N}{\Sigma\left(\dfrac{p_0}{p_1}\right)}=\dfrac{14}{13.5368}=103.4$

價比的幾何平均法　單獨之價比爲$\dfrac{p_1}{p_0}$，如有N 個價比，則幾何平均的公式，應爲

$$G=\sqrt[N]{\frac{p_1'}{p_0'}\times\frac{p_1''}{p_0''}\times\frac{p_1'''}{p_0'''}\times\cdots\cdots\cdots\cdots}$$

通常幾何平均數之計算，多借助於對數，而其計算公式，應爲

$$\log G=\frac{\log\left(\frac{p_1'}{p_0'}\right)+\log\left(\frac{p_1''}{p_0''}\right)+\log\left(\frac{p_1'''}{p_0'''}\right)+\cdots\cdots}{N}$$

玆仍用表 53 之資料，以示幾何平均數之計算於表 56 。

表56　比價幾何平均法之計算

類別	項目	$\frac{p_1}{p_0}$(1926=100)	$\log\frac{p_1}{p_0}$
原料品	棉花	80.3	1.90472
	小麥	81.7	1.91222
	黃豆	87.0	1.93952
	鋼鐵	100.3	2.00130
	羊毛	102.8	2.01199
	鷄蛋	84.4	1.92634
	小計		11.69609
生產品	棉紗	115.4	2.06221
	桐油	130.0	2.11394
	紙	115.4	2.06221
	水泥	135.7	2.13258
	小計		8.37094
消費品	啤酒	120.0	2.07918
	肥皂	167.2	2.22324
	麵粉	77.3	1.88818
	紙烟	125.3	2.09795
	小計		8.28855
總計			28.35558

原料品指數　$\log G=\frac{11.69609}{6}=1.94935$

$G=89$

生產品指數　$\log G=\frac{8.37094}{4}=2.09274$

$G=123.8$

消費品指數　$\log G=\frac{8.28855}{4}=2.07214$

$G=118.1$

總指數　$\log G=\frac{2.35558}{12}=2.02520$

$G=106.1$

價比的簡單中位數法　即將各價比依大小順序排列，而取其最中一價比爲指數。上例之價比可排列於下。

77.3
80.3
81.7
84.4
87.0
100.3
102.8
115.4
115.4
120.0
125.3
130.0
135.7
167.2

項數爲偶數，求中位數，須取中央二項之算術平均數$\frac{102.8+115.4}{2}$

$= 109.1$，此中位數，即 1930 年之總指數。各類指數，因例中所包括之項數太少，用中位數法，容有未妥，從略。

五種價比的簡單平均法之比較　上述之各種價比平均法，都要牽連到平均數問題。所以我們祗要知道各種平均方法之優劣，對於各種公式之優劣，便不難立判，五種價比平均數公式，除了乘數及倒數平均數很少用於指數外，其餘三種，都有人採用。

將算術平均法幾何平均法，及倒數平均法之結果，一加比較，我們可發見一種現象，就是算術平均數比幾何平均數大，而倒數平均數却比幾何平均數小，統計學上稱前者之現象爲偏上(Upward bias)，後者爲偏下 (Downward bias)，而幾何平均數居中，不偏不倚。關於偏性，還待後面討論。

加權和加權公式

加權　上節所討論之指數，爲簡單指數，便是不問指數裏物品重要與否，一視同仁，使各種物價之變化，在指數中有同等之影響。不過各種物品因性質之不同，對人生需要之程度和交易數量之多少，便各有不同。吾人於編製指數時，可按其輕

(221頁註)　據愛奇渥試驗，權數的錯誤，影響於指數結果，只不過$\frac{1}{20}$。又據密乞爾之試驗，簡易和加權指數的差數，不到$\frac{1}{10}$。

重，分別加權，使他們在指數中，得到相當的勢力。雖則，有人主張不用權數，他們的理由有二。第一，因爲權數的材料，不易搜集，更不易精確。第二，因加權和簡單指數的結果，相差極微。(註) 不過費暄等輩，仍舊覺得加權有所必要。

假使認爲加權有相當之必要，試問以什麼爲權數。在編製一般物價的指數，平常可於生產數量，消費數量，和交易數量中，選取一種或兼用數種爲權數。依密乞爾之意見，以爲交易數量，作爲權數，較爲優越；但亦未必盡然。

權數是否應隨時變更？ 各方意見不一，主張用固定權數的理由，以爲要觀測物價之變化，應該摒去其他一切變化之影響。如果指數之權數，時時變換，物價之變化和權數之變化，混在一起，不易指出物價變動程度之眞相。擁護變動權數的理由，以爲物品本身就有時代性，從前風行一時，銷路很大的物品，曾幾何時，竟至需要很少。反之，有些物品，在從前視爲不足輕重，現在或許非常重要。假如權數不變更，指數的結果，就要令人懷疑。通常折衷二法，對於權數材料，恆取若干年之平均數，同時每隔十年修改一次，使合乎當時情形。不但免去二種變化之混合，又不致有每年修改之麻煩。

加權方法 加權方法，要看指數公式之爲價比平均，還是實價總合，才能決定。在價比平均法所用權數，當爲物值，不過這種物值權數之資料，應該取材於基期？還是計算期？還是

其他時期？費喧氏概括歸納於下列四種方法：

加權Ⅰ. p_0q_0 （基期之量乘基期之價）

加權Ⅱ. p_0q_1 （本期之量乘基期之價）

加權Ⅲ. p_1q_0 （基期之量乘本期之價）

加權Ⅳ. p_1q_1 （本期之量乘本期之價）

至於實價總合之公式$\frac{\Sigma p_1}{\Sigma p_0}$，為二時期物價之和之比率。今欲將此公式加權，便不能用物值；因為物值（pq），本來已含有價（p），不能再以價乘之，故應用量為權數；然量有 q_0,q_1 二種，

表57. 勒氏公式之計算法

類 別	項目	單位	1926之物價 p_0	1926之物量 q_0	$p_0\ q_0$	1930之物價 p_1	$p_1\ q_0$
原料品	棉花	担	14.20	340	4,823	11.40	3,876.0
	小麥	担	6.00	250	1,500	4.90	1,225.0
	黃豆	担	4.60	110	506	4.00	440.0
	生鐵	担	3.32	1,250	4,150	3.33	4,162.5
	羊毛	担	30.00	120	3,600	30.85	3,702.0
	鷄蛋	千個	25.00	1,000	25,000	21.10	21,100.0
生產品	棉紗	包	78.00	12	936	90.00	1,080.0
	桐油	担	10.15	45	456.75	13.20	594.0
	紙	令	9.15	125	1,218.75	11.25	1,406.3
	水泥	桶	5.60	50	280	7.60	330.0
消費品	啤酒	箱	12.00	1,240	14,880	14.40	17,856.0
	肥皂	箱	2.50	120	300	4.18	501.6
	麵粉	包	4.40	100	440	3.40	340.0
	烟紙	箱	37.50	5	187.5	47.00	235.0
總 計					58,283		56,898.4

$$\frac{\Sigma p_1q_0}{\Sigma p_0q_0}=\frac{56,898.4}{58283}\times 100=97.6$$

故綜合加權法，亦有二種。有時指數之權數，係以指定之數字爲權數，這就稱爲概權 (Arbitrary Constant of weights)。當然這種加權，太近於武斷。

加權總合法　上節所述，綜合加權有二種，第一種是以 q 作權數，公式爲 $\frac{\Sigma p_1q_0}{\Sigma p_0q_0}$，此式爲 Laspeyres 所倡導，故稱勒氏公式。美國勞工統計局，即用此法以編製指數。本公式之計算方法，示例於表 57。

表58　派許公式之計算法

類　別	項目	單位	1926 之物價 p_0	1930 之物量 q_1	$p_0\ q_1$	1930 之物價 p_1	$p_1\ q_1$
原料品	棉花	担	14.20	420	5.964	11.40	4,788.0
	小麥	担	6 00	280	1.680	4.90	1,372.0
	黃豆	担	4.60	140	644	4.00	560.0
	生鐵	担	3.32	1.200	3.984	3.33	3,996.0
	羊毛	担	30.00	122	3.660	30.85	3,763.7
	鷄蛋	千個	25.00	1.200	30 000	21.10	25,320.0
生產品	棉紗	包	78.00	10	780	90.00	900.0
	桐油	担	10.15	40	406	13.20	528.0
	紙	令	9.75	110	1.072.5	11.25	1,237.5
	水泥	桶	5.60	40	224	7.60	304.0
消費品	啤酒	箱	12.00	1.300	15.600	14.40	18,720.0
	肥皂	箱	2.50	120	300	4.18	501.6
	麵粉	包	4.40	120	528	3.40	408.0
	紙烟	箱	37.50	6	225	47.00	282.0
總　計					65.067.5		62680.8

$$\frac{\Sigma p_1q_1}{\Sigma p_0q_1}=\frac{62,680.8}{65,067.5}\times 100=96.3$$

第二種是以 q_1 作權數，公式爲 $\frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}$，此公式爲派許 (Paasche) 所首創，故亦稱派許公式。本公式與上述公式所異者，本公式不用基期數量，而用計算期數量以替代之。至其計算方法，亦無大異，參表 58 。

加權算術平均法　計算方法，係以每一價比乘相當之權數，所得乘積之和，即以總權數除之。

算術 1. 即算術類加權法指數之簡稱，以下各式簡稱同此。

表59　算術工之計算

類別	項目	單位	價比(1926)	$p_0\ q_0$	價比×權重	價比(1930)	價比×權重
原料品	棉花	担	100	4,828	482 800	80.3	387,688.40
	小麥	担	100	1,500	150.000	81.7	122,550.00
	黃豆	担	100	506	50.600	87.0	44,022.00
	生鐵	担	100	4,150	415.000	100.3	416,245.00
	羊毛	担	100	3,600	360.000	102.8	370,080.00
	鷄蛋	千個	100	25,000	2,500.600	84.4	2,110,000.00
生產品	棉紗	包	100	936	93 600	115.4	95,041.44
	桐油	担	100	456 75	45 675	130.0	59,377.50
	紙	令	100	1,218.75	121.875	115.4	140,643.75
	水泥	桶	100	280	28.000	135.7	37,996.00
	小計						
消費品	啤酒	箱	100	14,880	1,488 000	120.0	1,785,600.00
	肥皂	箱	100	300	30.000	167.2	50,160.00
	麵粉	包	100	440	44 000	77.3	34,012.00
	烟紙	箱	100	187.5	18 750	125.3	23,493.75
總計				58283	5,828.300		5,676,909.84

$$\frac{\Sigma p_0 q_0 \frac{p_1}{p_0}}{\Sigma p_0 q_0} = \frac{5,676,909.84}{5,828,300} = 97.4$$

公式：$$\frac{\Sigma p_0 q_0 \frac{p_1}{p_0}}{\Sigma p_0 q_0}$$ (註)

計算步驟，舉於表 59

算術 II. $$\frac{\Sigma p_0 q_1 \frac{p_1}{p_0}}{\Sigma p_0 q_1}$$

算術 III. $$\frac{\Sigma p_1 q_0 \frac{p_1}{p_0}}{\Sigma p_1 q_0}$$

算術 IV. $$\frac{\Sigma p_1 q_1 \frac{p_1}{p_0}}{\Sigma p_1 q_1}$$

加權幾何平均法　計算方法與簡單的相似，每個價比的對數，應各以權數乘之。其乘積之總和，再以權數總和除之。其商數，即指數之對數，化爲眞數，即得指數。公式：

$$\frac{\Sigma p_0 q_0 \log \frac{p_1}{p_0}}{\Sigma p_0 q_0}$$

註　算術 I 與總合 I 結果相同，試驗證之。

$$\frac{\frac{p_1}{p_0} p_0 q_0 + \frac{p_1'}{p_0'} p_0' q_0' + \frac{p_1''}{p_0''} p_0'' q_0'' + \cdots\cdots\cdots\cdots}{p_0 q_0 + p_0' q_0' + p_0'' q_0'' + \cdots\cdots\cdots\cdots}$$

相消得 $\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0}$

同理 $\frac{\Sigma p_0 q_0 \frac{p_1}{p_0}}{\Sigma p_0 q_0}$ 之結果，等於 $\frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}$，證明同前。

上項公式爲幾何工之算法，至於幾何 工，ㄒ，亚，之公式，與上式相似，僅權數不同而已，計算方法見表 60

表60　幾何類工之計數

類別	項目	單位	1930之價比 $\frac{p_1}{p_0}$	價比之對數 $\log\frac{p_1}{p_1}$	權數 $p_0\ q_0$	價比之對數×權數 $p_0\,q_0\log\frac{p_1}{p_0}$
原料品	棉花	担	80.3	1.90472	4,838	9,195.95816
	小麥	担	81.7	1.91222	1,500	2,868.33000
	黃豆	担	87.0	1.93952	596	981.39712
	生鉄	担	100.3	2.00130	4,150	8,305.39500
	羊毛	担	102.8	2.01199	3,600	7,243.16400
	鷄蛋	千個	84.4	1.92634	25,000	48,158.50000
生產品	棉紗	包	115.4	2.06221	936	1,930.22856
	桐油	担	130.0	2.11394	456.75	965.54210
	紙	令	115.4	2.06221	1,216.75	2,513.31854
	水泥	桶	135.7	2.13258	280	597.12240
消費品	啤酒	箱	120.0	2.07918	14,880	30,938.10840
	肥皂	箱	167.2	2.22324	300	666.97200
	麵粉	包	77.3	1.88818	440	830.79920
	紙烟	箱	125.3	2.09795	187.5	393.36563
總計					58,283	115,588.32101

$$logG=\frac{115{,}588.32101}{58\ 283}=1.98323$$

$$G=92.2 \qquad \text{(卽1930之指數)}$$

加權倒數平均法　計算方法，亦與簡單倒數平均法相同，不過每一價比，應各乘以權數，而N之值，須改爲權數總和。倒數工之公式爲：

$$\frac{\Sigma p_0 q_0}{\Sigma p_0 q_0\left(\frac{p_0}{p_1}\right)}$$

表61即此公式計算之實例，

表61　倒數類Ⅰ之計算

類別	項目	單位	價比之倒數 $(\frac{p_0}{p_1})$	p_0q_0	$p_0q_0\frac{p_0}{p_1}$
原料品	棉花	担	1.2456	5,964	7,428.7584
	小麥	担	1.2345	1,660	2,057.1509
	黃豆	担	1.1500	644	740.6000
	生鐵	担	.9970	3,984	3,972.0480
	羊毛	担	.9724	3,660	3,558.9840
	鷄蛋	千個	1.1848	30,000	35,544.0000
生產品	棉紗	包	.8667	780	676.0260
	桐油	担	.7689	406	312.1734
	紙	令	.8667	1,072.5	929.5358
	水泥	担	.7368	244	179.7792
消費品	啤酒	箱	.8333	15,600	12,999.4800
	肥皂	箱	.5981	300	179.4300
	麵粉	包	1.2941	528	683.2848
	紙烟	箱	.7979	225	179.5275
總計				65,067.5	69,440.7871

$$H=\frac{65,067.5}{69,440.7871}=93.7$$

此項結果，即爲1930 年之物價指數。倒數Ⅱ，Ⅲ，Ⅳ.之公式及算法倣此。

指數之二大測驗

計算指數之公式很多，上述數種乃基本之公式，要判斷一公式是否正確，有二種方法，就是時間倒置測驗 (Time reversal test)，及因子倒置測驗 (Factor reversal test)。

時間倒置測驗　假如一九三〇年之麥價一倍於一九二六年

，那末反轉來說，一九二六年之麥價，祗及一九三〇年之半。假設甲地棉花較乙地貴三倍，那末乙地的棉價自然祇有甲地棉價三分之一，此點顯而易明，毋煩喋喋。時間倒置測驗，卽基於此理。此測驗之意義，乃以兩時間(或兩地域)輪流爲基數，求另一時間(或地域)之指數，則該二指數應互爲倒數。依照數學理論，凡互爲倒數之二數相乘，其積必等於1.故互爲倒數之二指數相乘，亦應等於 1。可是事實上，許多公式計算出來之結果，不能適合此原則，顯見得公式的本身有差誤之處。在六類簡單指數中，測驗施行之結果，只有綜合法，幾何平均法，中位數法，衆數法，合於時間測驗。其餘如算術平均及倒數平均法，都不適合測驗。今試以時間倒置測驗，施於幾何平均法及算術平均法，以驗其結果，藉示方法之一般。

簡單幾何平均法：

設以 P_{01} 代表以 '0' 爲基期，'1' 爲本期之任何公式所得之指數，P_{10} 代表以 '1' 爲基期，'0' 爲本期同一公式所得之指數。

$$\therefore\ P_{01}=\sqrt[n]{\frac{p_1}{p_0}\times\frac{p_1'}{p_0'}\times\frac{p_1''}{p_0''}\times\cdots\cdots\cdots\cdots}$$

$$P_{10}=\sqrt[n]{\frac{p_0}{p_1}\times\frac{p_0'}{p_1'}\times\frac{p_0''}{p_1''}\times\cdots\cdots\cdots\cdots}$$

$$P_{01}\times P_{10}=\sqrt[n]{\frac{p_1}{p_0}\times\frac{p_0}{p_1}\times\frac{p_1'}{p_0'}\times\frac{p_0'}{p_1'}\times\frac{p_1''}{p_0''}\times\frac{p_0''}{p_1''}\times\cdots\cdots}$$

$$=\sqrt[n]{1\times1\times1\times\cdots\cdots\cdots\cdots}$$

$$=1$$

再以實例來證明，依表 53 之資料，1926 年爲基期之1930年幾何平均指數，爲 106.1，以 1930 年爲基期之1926年物價指數，爲94.3，二指數相乘，得1.000523，差近於 1，是則證明幾何平均法之適合於時間倒置測驗。

簡單算術平均法：

$$P_{01}=\frac{\frac{p_1}{p_0}+\frac{p_1'}{p_0'}+\frac{p_1''}{p_0''}+\cdots\cdots\cdots}{N}$$

$$P_{10}=\frac{\frac{p_0}{p_1}+\frac{p_0'}{p_1'}+\frac{p_0''}{p_1''}+\cdots\cdots\cdots}{N}$$

$P_{10}\times P_{10}\neq 1$

再表35之資料，1926基年之 1930 年之指數爲108.8，而其時間倒置之指數爲 96.7，二指數相乘得 1.052096，其積不等於1，而向上偏，足證此公式不適於時間倒置測驗。

因子倒置測驗　每種物品總值，卽含有二因子，一爲物價，一爲物量。換言之，物價與物量相乘，等於物值，以代數式表之爲 p_1q_1，而與他一年價值之關係，應爲$\frac{p_1q_1}{p_0q_0}=\frac{p_1}{p_0}\times\frac{p_1}{q_0}$，若今年之米價和數量比去年增加一倍，則價比爲 200，量比爲200，比值應爲 400。此種關係，在單獨一項物品，既然如此，

若照邏輯來推論，物價指數乘物量指數，應等於物值指數。不過事實上將許多物品計算指數，則價之指數爲許多價比之平均數，量之指數爲許多量比之平均數。經平均之手續，其結果是否仍合此原則，則成問題。而公式之準確與否，卽藉此以證明。以一公式計算物價指數及物量指數，二指數之乘積，若等於物值指數，是爲適合於因子倒置測驗。否則，不合於因子倒置測驗。不合於因子倒置測驗之公式，以之計算指數，所得結果，便不正確。

以代數式表之則因子倒置測驗爲：

$$P_{01} \times Q_{01} = V_{01}$$

此式之 P_{01}，代表一公式計算之物價指數，Q_{01} 代表同一公式計算之物量指數，V_{01} 代表物值指數，卽 $\frac{\Sigma p_1 q_1}{\Sigma p_0 q_0}$

今以算術1之公式驗之，

$$P_{01} = \frac{\Sigma p_0 q_0 \frac{p_1}{p_0}}{\Sigma p_0 q_0}$$

$$Q_{01} = \frac{\Sigma q_0 p_0 \frac{q_1}{q_0}}{\Sigma q_0 p_0}$$

而 $$\frac{\Sigma p_1 q_1}{\Sigma p_0 q_0} \neq \frac{\Sigma p_0 q_0 \frac{p_1}{p_0}}{\Sigma p_0 q_0} \times \frac{\Sigma p_0 q_0 \frac{q_1}{q_0}}{\Sigma q_0 p_0}$$

故　$P_{01} \times Q_{01} \neq V_{01}$ 是爲不合於因子倒置之測驗。

設令表59之資料，以實值代入公式，試驗其結果，更易

顯明。

1930之物價指數

$$\frac{\Sigma p_0 q_0 \dfrac{p_1}{p_0}}{\Sigma p_0 q_0} = 97.4$$

1930之物量指數

$$\frac{\Sigma q_0 p_0 \dfrac{q_1}{q_0}}{\Sigma q_0 p_0} = 111.6$$

二者之乘積

$$.974 \times 1.116 = 1.086984$$

這就是說物價減少2.6%，物量增加11.6%，總值增加8.7%，但總值之比，爲

$$\frac{\Sigma p_1 q_1}{\Sigma p_0 q_0} = 107.5$$

二者之結果，一作比較，相差108.9−107.5=1.4%，是爲不合因子倒置測驗之鐵證，

其他公式，可以同樣方法測驗之，不再贅述。總之，上述諸公式，無一能適合此測驗。

指數之偏性(Bias)

指數之偏性　上節所述二大測驗，皆爲測驗指數之有效方法。一公式計算二時期（或二地域）之指數，如向前指數乘向後指數，乘積不能爲一，或物價指數乘物量指數，不能與物値

指數相等，這就是公式偏上或偏下之現象，統計學上稱爲偏性。指數偏誤有二種，類偏 Type bias 和權偏 Weighting bias，分述於下。

類偏 簡單公式中，算術類有偏上性，倒數類有偏下性。算術類之簡單公式爲$\dfrac{\Sigma\frac{p_1}{p_0}}{N}$顚倒基期之公式爲$\dfrac{\Sigma\left(\frac{p_0}{p_1}\right)}{N}$，二式相乘爲$\dfrac{\Sigma\left(\frac{p_0}{p_1}\right)}{N}\cdot\dfrac{\Sigma\left(\frac{p_1}{p_0}\right)}{N}$吾人欲證明者，乃此乘積大於1。

試將$\dfrac{\Sigma\left(\frac{p_1}{p_0}\right)}{N}$，$\dfrac{\Sigma\left(\frac{p_0}{p_1}\right)}{N}$先展開之，爲$\left(\frac{p_1'}{p_0'}+\frac{p_1''}{p_0''}+\frac{p_1'''}{p_0'''}+\cdots+\frac{p_1^n}{p_0^n}\right)$ $\left(\frac{p_0'}{p_1'}+\frac{p_0''}{p_1''}+\frac{p_0'''}{p_1'''}+\cdots\cdots\cdots\cdots+\frac{p_0^n}{p_1^n}\right)$此二因數相乘結果，當得 N^2 項，此 N^2 項中，有$\frac{p_1'}{p_0'}\times\frac{p_0'}{p_1'}$，$\frac{p_1''}{p_0''}\times\frac{p_0''}{p_1''}$ ，…………共 N 項，每項皆等於1。其餘 (N^2-N) 項，每兩項互爲倒數，如$\frac{p_1}{p_0}\times\frac{p_0'}{p_1'}$與$\frac{p_0}{p_1}\times\frac{p_1'}{p_0'}$，$\frac{p_1}{p_0}\times\frac{p_0''}{p_1''}$與$\frac{p_0}{p_1}\times\frac{p_1''}{p_0''}$ ，…………。然依數學之原理，一數與其倒數之和大於2，故此(N^2-N)項之和大於(N^2-N)。如此 N^2 項之中，有 N 項每項等於1，其餘，(N^2-N)項之和大於 N^2-N，故 N^2 項之和大於 N^2。於是在$\dfrac{\Sigma\left(\frac{p_1}{p_0}\right)\Sigma\left(\frac{p_0}{p_1}\right)}{N^2}$，分子大於 N^2，分母等於N^2，故分數大於1，可證明算術類有偏上性。倒數法證明可以同樣之方法行之，不贅。

免除偏性最有效之方法，是將二種偏性相反方向的指數接合之(Crossing)。　接合之方法，不過取其適當平均數。譬如二種指數，表示計算期與基期之關係，假設一種有偏上性，一種有偏下性，那末，那二者平均數之指數，其偏誤可比任何一種少，或竟無之。接合法所用平均數，當合乎邏輯，普通以幾何平均數最爲適當。

算術平均數　$\dfrac{\Sigma\left(\dfrac{p_1}{p_0}\right)}{N}$

倒數平均數　$\dfrac{N}{\Sigma\left(\dfrac{p_0}{p_1}\right)}$

今將此二公式接合之，得

$$P_{01}=\sqrt{\frac{\Sigma\left(\dfrac{p_1}{p_0}\right)}{N}\times\frac{N}{\Sigma\left(\dfrac{p_0}{p_1}\right)}}$$

茲將此公式試驗之，以視其偏性若何，F_{01} 之倒置基期公式，爲

$$P_{10}=\sqrt{\frac{\Sigma\left(\dfrac{p_0}{p_1}\right)}{N}\times\frac{N}{\Sigma\left(\dfrac{p_1}{p_0}\right)}}$$

$$=\frac{1}{P_{01}}$$

$$\text{故}\quad P_{01}\times\frac{1}{P_{01}}=\sqrt{\frac{\Sigma\left(\dfrac{p_1}{p_0}\right)}{N}\times\frac{N}{\Sigma\left(\dfrac{p_0}{p_1}\right)}}\ \sqrt{\frac{\Sigma\left(\dfrac{p_0}{p_1}\right)}{N}\times\frac{N}{\Sigma\left(\dfrac{p_1}{p_0}\right)}}$$

$$=\sqrt{\frac{\Sigma\left(\frac{p_1}{p_0}\right)}{N}\times\frac{N}{\Sigma\left(\frac{p_1}{p_0}\right)}\times\frac{\Sigma\left(\frac{p_0}{p_1}\right)}{N}\times\frac{N}{\Sigma\left(\frac{p_0}{p_1}\right)}}$$

$$=1$$

是可證明此方法之結果，並無偏性。

權偏　就上面所講，加權公式中，除總合法外，都有權偏可能。加權方法有四種，即

Ⅰ. p_0q_0

Ⅱ. p_0q_1

Ⅲ. p_1q_0

Ⅳ. p_1q_1

四方法中，Ⅰ.Ⅱ.兩種常偏於向下，Ⅲ與Ⅳ，常偏於向上今請先證明加權Ⅲ之結果，必大於Ⅰ，加權Ⅳ，必大於Ⅱ。

加權Ⅰ之權數為 $p_0'q_0'$，$p_0''q_0''$，$p_0'''q_0'''$，…………，加權Ⅲ.之權數為 $p_1'q_0'$，$p_1''q_0''$，$p_1'''q_0'''$，…………今$\frac{p_1'q_0'}{p_0'q_0'}=\frac{p_1'}{p_0'}$，$\frac{p_1''q_0''}{p_0''q_0''}=\frac{p_1''}{p_0''}$，$\frac{p_1'''q_0'''}{p_0'''q_0'''}=\frac{p_1'''}{p_0'''}$，…………是以ⅠⅢ二種權數之比，即一價比。加權Ⅲ之各項 $\left(p_1'q_0'\frac{p_1'}{p_0'}，p_1''q_0''\frac{p_1''}{p_0''}\cdots\cdots\right)$ 即等於以價比$\left(\frac{p_1'}{p_0'}，\frac{p_1''}{p_0''}\cdots\cdots\cdots\right)$各乘加權Ⅰ之各項$\left(p_0'q_0'\frac{p_1'}{p_0'}，p_0''q_0''\frac{p_1''}{p_0''}，\cdots\cdots\cdots\right)$。但是加權法Ⅰ之各項，既已含有價比，再以相同之價比乘之，其結果價比大者，所得權數大，價

比小者所得權數小，則最後之平均數(卽指數)必漲高，是卽加權Ⅲ之指數，必大於加權Ⅰ。加權Ⅳ之指數大於加權Ⅱ，可以同理證明之，不贅。

由是推論，算術類加權法Ⅲ，Ⅳ，之公式以類論，爲偏上，以加權論，亦爲偏上，所得指數應有二重偏上。又算術 Ⅰ. Ⅱ. 之公式，於類爲偏上，於加權爲偏下，二者相抵，偏性自可稍殺，亦接合法之一種。

理想公式(The Ideal Formula)

對於公式之偏性，旣如上述，可用接合法來糾正，因得擬一理想公式如次：

$$P_{01}=\sqrt{\frac{\Sigma p\ q_0}{\Sigma p_0 q_0}\ \frac{\Sigma p\ q_1}{\Sigma p_0 q_1}}$$

吾人可知此公式，僅爲勒氏公式和派許公式之幾何平均而已。他不但適合時間倒置測驗，又適合因子倒置測驗，試以代數證明之。

1. 時間倒置測驗：

$$P_{01}=\sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_1 q_0}\times\frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}}$$

$$P_{10}=\sqrt{\frac{\Sigma p_0 q_0}{\Sigma p_1 q_0}\times\frac{\Sigma p_0 q_1}{\Sigma p_1 q_1}}$$

$$P_{01}\times P_{10}=\sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0}\times\frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}}\sqrt{\frac{\Sigma p_0 q_0}{\Sigma p_1 q_0}\times\frac{\Sigma p_0 q_1}{\Sigma p_1 q_1}}$$

$$=\sqrt{\frac{\Sigma p_1q_0}{\Sigma p_0q_0}\times\frac{\Sigma p_1q_1}{\Sigma p_0q_1}\times\frac{\Sigma p_0q_0}{\Sigma p_1q_0}\times\frac{\Sigma p_0q_1}{\Sigma p_1q_1}}=1$$

2. 因子倒置測驗：

$$p_{01}\times q_{01}=V_{01}$$

$$\frac{\Sigma p_1q_1}{\Sigma p_0q_0}=\sqrt{\frac{\Sigma p_1q_0}{\Sigma p_0q_0}\times\frac{\Sigma p_1q_1}{\Sigma p_0q_1}}\sqrt{\frac{\Sigma q_1p_0}{\Sigma q_0p_0}\times\frac{\Sigma q_1p_1}{\Sigma q_0p_1}}$$

$$=\sqrt{\frac{\Sigma p_1q_0}{\Sigma p_0q_0}\times\frac{\Sigma p_1q_1}{\Sigma p_0q_1}\times\frac{\Sigma q_1p_0}{\Sigma q_0p_0}\times\frac{\Sigma q_1p_1}{\Sigma q_0p_1}}$$

$$=\frac{\Sigma p_1q_1}{\Sigma p_0q_0}$$

試以表 57 及 58 之資料作例，可得同一之結果如下：

1. 時間倒置測驗

$$P_{01}=\sqrt{\frac{56898.4}{58283}\times\frac{62680.8}{65067.5}}=.9698$$

$$P_{10}=\sqrt{\frac{58283}{56898.35}\times\frac{65067.5}{62680.8}}=1.0312$$

$$P_{01}\times\frac{1}{P_{01}}=.9698\times1.0312=1.00005776$$

2. 因子倒置測驗

$$V_{01}=\frac{62680.8}{58283}=107.55$$

$$Q_{01}=\sqrt{\frac{65067.5}{58283}\times\frac{62680.8}{56898.4}}=110.98$$

證　　$V_{01}=P_{01}\times Q_{01}$　？

$P_{01} \times Q_{01} = .9698 \times 1.109 = 1.0754082$

1.0755與1.0754.082　相減

相差不滿 .0001，可證實合於因子倒置測驗。此項極微細之差數，由於計算時小數取捨之關係。

理想公式之批評　理想公式，有幾點確能滿足吾人之理想，例如

1. 以交易數量爲加權的材料，極合理論。

2. 適合時間倒置及因子倒置二大測驗。

3. 無偏性。

4. 不致因物品單位之大小，影響結果之危險。

5. 在複式公式中，此式尙稱比較簡單。

6. 代表市價的淨趨勢最精確。

但在另一方面，未始毫無遺憾，據密乞爾之批評，有下列缺點。

1. 所用權數每年更動，因此物價之變化及權數之變化，混在一起，不能分淸究竟是那種變化之影響。若在連環基期制，基期權數亦每年更改，權數之變化更大。

2. 如在連環基期制，則與固定基期之結果，相差甚遠，不適於長期之比較。

理想公式的替用公式　理想公式還有一個替用的公式，這公式因爲是愛奇華士(Edgeworth)和馬雪爾(Marshall)首先倡

用，所以又名(Edgeworth and Marshall Aggregative Method)。他的公式，是

$$\frac{\Sigma(q_1+q_0)p_1}{\Sigma(q_1+q_0)p_0}$$

費暄(Fisher) 以爲從準確，迅速，簡單，差誤少，四點立場，此公式認爲最切實，而且到處可用的公式 (The best Practical all-around Formula)。他的結果與理想公式之結果，相差不及$\frac{1}{4}\%$。計算方法，示例於表 62。

表62 愛氏馬氏公式之計算

類別	項目	單位	p_{1926}	q_0+q_1	$(q_0+q_1)p_{1926}$	p_{1930}	$(q_0+q_1)p_{1930}$
原料品	棉花	担	14.20	760	10,792	11.40	8,664
	小麥	担	6.00	530	3,180	4.90	2,597
	黃豆	担	4.60	250	1,150	4.00	1,000
	生鐵	担	3.32	2,450	8,134	3.33	8,158.5
	羊毛	担	30.00	242	7,260	30.85	7,465.7
	鷄蛋	千個	25.00	2,200	55,000	21.00	46,420
生產品	棉紗	包	78.00	22	1,716	90.00	1,980
	桐油	担	10.15	85	862.76	13.26	1,122
	紙	令	9.75	235	2,291.25	11.25	2,643.75
	水泥	桶	5 60	90	504	7.60	684
消費品	啤酒	箱	12 00	2,540	30,480	14.40	36,576
	肥皂	箱	2.50	240	600	4.18	1,003.2
	麵粉	包	4.40	220	968	3.40	748
	紙烟	箱	37.50	11	412.5	47.00	517
總計					123,350.5		119,579.15

$$\frac{\Sigma(q_0+q_1)p_1}{\Sigma(q_0+q_1)p_0}=\frac{119{,}579.15}{123{,}350.5}=96.96$$ (以1926爲基期之1930年指數)

基期（Base Period）

固定基期(Fixed base)與連環基期 (Chain system) 固定基期，係擇定某時期之物價爲基數作100，而以其他各時期同樣之物價與之相較，以觀察一般物價漸次之升降。所選定之基期，在指數未修改以前，賡續不變。連環基期即以各時期物價，迭爲其下一時期之基數。換言之，求每個時期之指數時，即以上一時期爲基期，故便於二年之比較。

基期移動之簡捷法 研究物價的人，有時或許要將所有各年物價，和基期以外另一時期之物價作比較，這樣勢必要拿基期移動。移動基期，有一簡捷法，可毋事重編指數，祇將原有數列之各年指數，各除以新基期，所得結果乘以 100 即得。例如下列之指數，原來以 1913 年爲基期，現在要移動以 1926 年爲基期，使各年物價，可與 1926 年比較，故各年各以 1926 年指數 203.6 除之，其商數各乘以 100，即得一列以 1926 年爲基期之新指數。

	1913	1925	1926	1927	1928
203.6	100	206.2	203.6	239.9	269.7
	49.1	101.3	100	117.8	132.5

此種基期移動之簡捷法，不是能普遍地應用，對於算術平均法所編之指數，所得結果，常有出入，惟幾何平均法及綜合法之指數，則非常正確，此點讀者宜注意之。

連鎖指數　以連環基期制求得之指數，對於鄰近兩時期之比較，既很便利，就是要和任何一期比較，亦無不可，惟須求出連鎖指數。連鎖指數之算法，卽以基期之連鎖指數仍作100.用於連環指數，其餘各期之連環指數，乘以上期之連鎖指數，連乘推遞而得之。

年　分	定基指數（1926=100）	連環指數	連鎖指數（1926=100）
1926	100.0		100.0
1927	76.7	76.7	76.7=(100×76.7%)
1928	63.5	82.7	63.4=(76.7×82.7%)
1929	101.8	160.4	101.7=(63.4×160.4%)
1930	132.1	129.8	132.0=(101.7×129.8%)

觀上表，可知連鎖指數，實由連環基期之指數，恢復爲固定基期之還原。在本例中定基指數與連環指數，相差極微，此由於數列較短，設經過長久之時期，則其累積於乘積之差誤，亦頗可觀，未容忽略。

固定基期制與連環基期制之比較　採用固定基期制，各年指數祇能和基期作比較，雖則有基期移動簡捷法，可以救濟，不過亦不能普遍地可以應用於各公式之指數。再設基期相距太遠，則於物價變化之測定，大感困難。普通經過時期愈長，價比的離差愈大。採用連環基期制，則便於與鄰近時期作比較。再連環制物價漲落之程度，較爲集中，測定物價的變化，也比較容易而精密。若貨物須增減時亦非常便利；不過連鎖指數又有累積差誤的危險。所以許多學者主張編製指數時，最好編成

二種，一爲固定基期指數，一爲連環指數，以便隨時考察其差異，而免發生誤解。

定基基期之選擇　基期須擇定適當之時期，所謂適當之時期，就是這時期中之物價很平穩，沒有多大的變動，很合乎常軌。要測驗物價是否平穩，可利用離差數，看那一年的離差最小，卽知該年物價之是否平穩。選擇基期，又有長短之問題，或爲一年，或爲數年，或一月，或一星期，當須考慮。通常一短時期之物價變動，較爲劇烈，時期稍長，升降之趨勢互銷，結果便較爲平穩。所以就大體而論，基期時期長勝於短時期。基期須取與吾人研究有關之時期，決非隨隨便便選一很遠之時期作基期，失了指數之價値。所以大戰後各國指數，多取1913年爲基期，以便比較戰前戰後物價之趨勢。迨大戰後數年，各國以爲情勢變遷，1913不很適用，多改用1926。而1926年國際統計協會之決議，又以爲各國應一致以1930年爲基期。

第二章

批發物價指數

I　編製程序

物價指數編製之程序，概括之可分下列六端。

I. 確定目的　編製指數的第一件事，是確定目的，因編製方法和指數本身之目的，關係極大。物價指數，有為測量貨幣購買力而作，有為測量進出口貨價而作，有為推測商情變化而作。為貨幣購買力而作的指數，包含物品的多寡及性質，當然和為預測商情變化而作者不同。普通言之，前者包含物品較多，而後者所選物品，當以變化最早，感應最靈者為合格。為進出口貨價而作之物價指數，又有其特殊之目的，當然和上二者又不同。目的不同，方法因之而異，故編製指數時，當首先確定目的。

II. 選擇項目　編製指數之目的，既已確定，即須着手選擇項目，為達到該目的所必需之資料。惟項目之選擇包括二問題——物品數目和物品性質。

III. 物價取材　一種物品，在市場上即有一價格，此即常人

註　本文所謂批發物價指數，僅及為一般物價之水準而作之指數，亦即測量一般貨幣購買力之指數

所謂之物價。不知物價之種類甚多，內容複雜。我們對於項目雖已决定，採用物價，究以何種爲準？編製指數者當先明物價之類別，然後决定採用一種物價，以適合其所編之指數。

Ⅳ. 基期　物價指數，是相對的測量物價，一定要有一時間和現在比較，才可知道現在物價之爲升爲降，此等用以作比較標準之時期，便稱基期。基期并非隨意亂定，應注意三點(1)基期種類(2)基期長短(3)基期之决定。此三問題，已於通論章中討論之，不再贅。

Ⅴ. 權數之决定　各種物品應相等重要，抑或有輕重軒輊之不同，而應用權數。若用權數，則權數之資料，如何搜集，其方法如何，凡此種種問題，請參通論章加權節。

Ⅵ. 公式之决定　指數之資料，基期，權數等，均已確定，則當决定公式，以便計算指數。公式之種類，據費暄教授之分類，約有百數十種，採用何種，始爲適當，本書已擇其最重要者數十種，詳論於通論中，讀者互參可也。

Ⅱ 物品數目

物價是編製指數的原料，可是世界上的物品，何止千萬種，當然不能每種都採入，祇能應用取樣法，就全體中，依照我們所編指數之目的，酌量採取幾種作代表。現在歐美各國著名的指數中，物品多少，各不一律。如滕氏(Duns)指數有300項

，加拿大政府指數有271項，美國勞工統計局所編指數爲328，而戰時工業局指數，竟有1474項。此外數目在50種以下的指數，也很多，例如美國沙貝克(Sanerbeck)指數僅包45種，歷時旣久，成績斐然。繼起而倣效者，如法國政府指數，紐絲蘭指數，奥國馮瓊確維(Von Jan Kovick)指數，印度Atkinson指數，都含45種。更有較少物品之指數，例如德國許密志(S-chmitz)指數祇29項，奇勃生(Gibson)指數僅22種，紐約聯合準備銀行月刊所公布之指數祇20種(從前僅12種)，紐約紀年雜誌所公布者亦祇25種。此外私人所編製者，較此更少，例如Pearson指數僅10項，可謂最少。

指數所含物品之多少，差別旣若上述，究竟他們的結果，是否相同？還是物品多好，還是少好？費暄教授嘗作種種之計算以試驗之。選商品200, 100, 50, 25, 12, 6, 3, 種，分別計算指數，以200種之指數爲標準，則3或6種物品之指數，與之相差，爲百分之三或四，其條之差爲百分之一或二。由此可見物品之多寡，與指數雖有關係，惟所關不大。

密乞爾(Wesley C. Mitchell)亦曾取美國1890—1913年之物價編成六種不同之指數，作種種之試驗，以研究大指數與小指數之異同。現在將他研究之結果，介紹於表63，以作吾人之參攷。

表63　1890—1913美國六種物價指數比較表

年分	[illegible]4[illegible]品至261品	14[illegible]種	50種	40種	25種甲	25種
1890	113	114	114	113	115	113
1891	112	113	114	114	112	118
1892	106	106	105	105	103	112
1893	106	105	105	101	103	107
1894	96	96	94	93	92	96
1895	94	93	94	95	95	93
1896	90	89	87	83	88	85
1897	90	89	89	89	90	84
1898	93	93	95	95	96	90
1899	102	103	103	108	107	103
1900	111	111	112	115	113	109
1901	109	110	109	116	111	107
1902	113	114	116	122	116	117
1903	114	114	115	118	118	117
1904	113	114	116	118	122	110
1905	116	116	118	122	123	115
1906	123	122	123	128	130	122
1907	130	130	132	138	132	132
1908	122	121	125	129	124	122
1909	125	124	132	135	133	128
1910	130	131	135	141	133	134
1911	126	130	129	135	129	131
1912	130	134	138	142	140	138
1913	130	131	138	139	142	133
1890—1899平均數	100	100	100	100	100	100
1900—1910平均數	118	118	120	124	122	118
1910—1913平均數	129	132	135	139	136	134
1890—1896漲落數	−23	−25	−27	−35	−27	−28
1896—1907漲落數	+40	+41	+45	+55	+44	+47
1907—1908漲落數	− 8	− 9	− 7	− 9	− 8	−10
1908—1913漲落數	+ 8	+13	+13	+13	+ 6	+16
價比最高最低差	40	45	51	54	54	54
逐年平均變化	4.0	4.1	4.9	5.5	5.0	6.2

表中第一行取材於美國勞動統計局的物價編成，第二行物品與第一行相同，不過將一種物品有幾種價格者，只取其平均數，所以物品較少。第三行是根據奇卜生(Gibson)指數。第四行指數包含物品20種，每種有原料價及製品價二種，故有項目40。第五第六二行，則係任意從一般重要物品中，各選出

25種。

觀察上表，可見大指數與小指數漲落之趨勢，頗有相似之處，如：1890—1896 各指數之趨向均下落，1896—1907 各指數增高很快，1908年之大跌，1909 又復原狀，1910—1913 則漲落無定，六種指數之變動，無多大顯著之差異。可見包含25種之指數，和包括十數倍於此之指數，無多大之區別。由此足證，祇要物品選擇得當，數目多少，尙屬次要問題。國際統計學會在比京開會時，法國中央統計局長 March(前任)和Huber二人，曾經有以50項爲適足數目的主張。普通而言，指數中物品數目如不足20種，代表性一定很弱，反之，如物品太多，計算手續和經費，多要增加。再者，物品少之指數，變動較顯著，大指數變動較和緩。普通之現象如此，但須視物品本身之性質如何。

四 物品性質

指數所含物品多少，與物品本身之性質如何，極有關係，本節試將物品之性質問題，略加研究。密乞爾之主張，以爲物品可分原料品與製造品，原料品又可復分農產，林產，畜產，礦產品，製造品又可復分爲消費品和生產品。

原料品與製造品　原料品之價格，常影響製造品。例如棉花爲原料，棉織物是製品，棉花價格的變動，影響棉織物之行

表64　原料品物價指數與製造品物價指數比較表(1890—1899年平均=100)

年份	原料 49品	製品 183—193品	原料 [illegible]品	製品 20品
1890	115	112	113	112
1891	116	111	114	114
1892	108	106	104	105
1893	104	106	99	103
1894	93	97	91	94
1895	92	94	94	96
1896	84	92	85	92
1897	83	90	88	89
1898	94	93	98	92
1899	106	101	114	103
1900	113	110	118	111
1901	111	108	120	113
1902	122	111	127	118
1903	123	112	122	114
1904	120	111	123	113
1905	121	115	127	117
1906	127	122	135	120
1907	133	129	146	131
1908	124	121	135	124
1909	131	123	143	127
1910	135	129	149	132
1911	135	124	144	127
1912	145	127	151	132
1913	139	128	149	128
1890—1899年平均	100	100	100	100
1900—1909年平均	122	116	130	119
1901—1919年平均	139	127	143	130
1890—1896漲(+)落(—)	— 31	— 29	— 28	— 20
1896—1907漲(+)落(—)	+ 49	+ 37	+ 61	+ 39
1907—1908漲(+)落(—)	— 9	— 8	— 11	— 7
1908—1913漲(+)落(—)	+ 15	+ 7	+ 14	+ 4
價比最高最低差額	61	39	66	43
逐年平均變動	5.5	40	6.4	4.9

情，所以在選擇原料品和製造品時，須維持一相當均勢。平常，選擇物品要各物本身之性質，愈遠愈妙。倘性質相同之原料品及製造品，採用品數太多，指數之結果，不免謬誤。此種誤謬指數，不乏先例。如倫敦經濟週報(London Economist)1860—1865年指數，採用物品計22項，其中棉花和棉製品，却各占四項。在美國南北戰爭時，南部海港被封鎖，英國棉花來源

驟減，而價因之大漲，倫敦週報之指數，便大受影響，而得一不可思議之謬誤。試以沙貝克指數與之兩相比較，就可發見當時謬誤之程度。

年分	經濟週報指數 (1860=100)	沙貝克指數 (1860=100)
1860	100	100
1861	103	100
1862	109	106
1863	136	109
1864	145	112
1865	136	106

原料品和製造品價格之變動，雖然形影相隨，可是他們影響的程度，不一定和原料品一樣。平常原料品價格之變動，較製造品為劇烈，製造品變動和緩而範圍又較小。密乞爾曾有試驗，請參表 64 。

農產，林產，畜產，礦產， 四類物品之變動，亦各不相同。農產品因受天然氣候之支配，每年收穫之多寡，毫無一定，所以價格變動亦最烈。有時竟和一般物價之趨勢，背道而馳。畜產品對於天時之影響較差，不過就變動激烈來講，雖較遜於農業品，亦難捉摸。林產品價格變動又較和緩，但亦難以常理繩之。惟礦產物之升降起落，常與一般商情之盛衰，行動一致，亦步亦趨。這點值得我們注意的。

消費品與生產品 Consumer's goods and Producers goods

消費品和生產品之區別，在他們加工的程度。二種物品變動之

程度，也各不相同。平常，消費品之物價較爲穩定，因爲他所受商情變化之影響很小。生產品之需求緩急，往往以一般商情盛衰爲轉移，漲落比較劇烈。

選擇物品種類之原則　總上所述吾人可得下列之原則：

1. 採入指數之物品，其本身之性質彼此愈遠愈妙，和未採入之物品，彼此性質，愈近愈好。

2. 採入指數之物品，其價格之變動，彼此愈無關聯愈妙，和未經入選之物品，愈有關聯愈好。

四 批發物價之取材

在編製一般批發物價指數時，物價之取材，有標準四。

1. 市場價格　大都直接得自大工廠，販賣代理店，或批發商所宣布之行情，亦有得自物品交易所之記載，或雜誌新聞紙之報告。美國之指數，多取材於此，我國之批發物價指數亦用此法。

2. 輸出入物價　是從進出口貿易公司所刊布之價格，或海關報告中得之。歐洲各國之指數多採用此法。如德國沙密皮爾指數，1921 以前之英國商務局指數，1911 年前之法國政府指數等。

3. 社團價格　是公共機關如學校，醫院，兵站等購買物品之價格，意大利政府指數，法國拉瓦色 (Leuasseur) 指數等

，多用此法。

4.合同價格 是買賣雙方在貨物未上市或未出廠前所定之價格，如期貨價格。採用此法者，祇美國戰時工業局物價部所披露之指數。

四種標準之比較 四種之中以合同價格爲最劣，因爲(1)合同價格，難於調查，(2)物價之種類不多，(3)成交時與訂購時之價格因時境之變遷，不能相合。社團價格亦不適用，因公共機關所購物品，數量雖大，但種類不能普遍，品質又不齊一，所以測驗代表一般的物價，不克勝任。

市場價格用以測驗全市批發物價之平均變化最爲優越。上列二種，與之相較，固自嘆勿如，卽和輸出入價格比較，也有一得之長。因爲輸出入價格，將物價的變動，和買賣品質的變動，混在一起，不易看出物價變動的程度，究爲何如。

輸出入物價，雖有包含二種變動的遺憾，但在另一方面，也有勝過市場價格之處。(1)因爲進出口貨價，常能顯出一國輸出入重要商品的實際所支付或承受之價格，而市場價格，祇能表示特種等級之貨物。(2)輸出入貨價之變動，較之市場價格爲平穩。

V 我國現有之批發物價指數

I. 上海躉售物價指數 始於民國八年九月，初由財政部

駐滬貨價調查處主編，現由稅則委員會繼續辦理，結果在上海物價月報，及貨價季刊中發表之。民國二十年修正之後所取基期，改用民國15年，以其全年平均物價爲基價。所用貨品，分爲八類，包含物品及項目數，列表如下：

類別	物品	項目
糧食類	9	22
其他食物類	27	31
紡織品及其原料類	24	38
金屬類	12	12
燃料類	9	13
建築材料類	11	11
化學品類	10	10
雜類	18	18
合計	119	155 註

計算公式爲簡單幾何平均數。

Ⅱ. 修正華北批發物價指數　由天津南開大學社會經濟委員會主編，民二至十六年按年編製，取每年每月一日及十五日市價之平均數，十七年一月至三月間，按月編製。該年四月以降，按週編製，取天津北平二地每日批發市價之平均數，而以每週星期三爲終算期，按期在天津大公報之統計週刊中刊布之。共包括貨品 106 種，分按工業分類、按製造分類，分別編製，按工業之分類如下：

註　21年1月後因抵制日貨關係，市價從缺，另以他貨替代，無法替代者，即刪去，故商品項目減爲154,項

類別	物品
食物	43
布疋及其原料	19
金屬	15
建築材料	12
燃料	12
雜類	5
	106

計算公式，用簡單幾何平均數。

五. 廣州批發物價指數　初由廣東省政府農工廳編製，現由建設廳繼續進行，自民國元年至十三年按年編製，十四年一月以後，改為每月編製，以民國二年全年為基期，包括貨品205 種。計算公式初用簡單算術平均法求各類指數，然後用加權算術平均法求總指數，現已改用簡單幾何平均法。貨品分類如下：

類別	物品
米類	20
其他食品類	65
衣料	43
燃料	14
金屬及建築材料	41
雜項	22
	205

除上述三種最重要之批發物價指數外，尚有漢口，青島，南京之批發物價指數，都由前工商部現實業部調製。此種指數，多以民國19年1月為基期。計算公式為簡單幾何平均法，每種包括之貨品種類列表如下：

漢口批發物價指數：

類別	物品
食料	54
衣料	20
燃料	15
金屬及建築材料	18
雜項	11
	118

青島批發物價指數：

類別	物品
食料	50
衣料	20
燃料	10
金屬	14
建築材料	10
雜項	13
	122

南京批發物價指數：

類別	物品
食料	43
衣料	16
燃料	10
金屬及電器	17
建築	11
雜項	9
	106

第三章

生活費指數

I 生活費，生活程度，生活費指數

生活費 人生生活所需之費用，謂之生活費。凡食品，衣服，房租，燃料，交通，教育，娛樂等費，統稱之爲生活費。

生活費與生活程度 社會上各級之生活所需，互有不同，所以各人各家各級的生活費，也有差異，我們想研究生活費可以把全社會分做幾個生活平面，生活平面，就是普通所謂生活程度。他的造成大概根據於社會習慣和經濟情形，所以，因地而異，因人而異，因時而異，因社會之地位而異。某種物品的需要，某種役務的享受，在一階級認爲必需，而另一階級則認爲奢侈，也許是在期望着而不可得。所以要明瞭某種階級之生活費，當先研究其生活程度。

生活費指數 生活費一方面固依生活程度而定，他方面又受物價變動之影響。設一家今年之生活，與去年完全一致，是爲生活程度之不變，然若物價改變，則貨幣購買力改變，則生活費亦必變更。設今年與去年之物價不變，然若一家之生活程度改變，則生活費亦必變。故生活程度與物價，實爲構成生活費之二要素。任何一要素變動，生活費卽隨之而變動。生活費

指數既爲表示各時期實際生活所需費用增減之比率，則二因子作同時之研究，實爲編製生活費指數必要之工作。不過生活程度之變動，常緩而微，可數年或十數年，舉行調查一次，而物價之變動，有時劇而速，非繼續不息調查不可。所以平常生活費指數所測量者，僅生活費中物價之一因子，而假定（事實亦復如是）其能代表全部變動。生活費指數編製後經過若干年，覺生活程度之變更已大，指數去實際之生活費已遠，則須重行調查新生活程度，作編製新指數之根據；如是，生活費指數才足以代表實際生活費之變動。

II 編製之目的

生活費指數既是測量某種階級的生活費，所以指數之計算，必須根據某種生活程度，則此生活程度爲勞工階級的，抑爲買辦階級的；爲一般勞動者的，抑爲某一業的勞動者；爲全國工人的，抑爲某一地工人的。此屬於指數之目的，當然視編製指數之用意而定。如指數之用意爲解決一地之勞資糾紛，則以一地工人之生活程度爲標準。如欲表示全社會一般生活費之變動，則以全國平均之生活程度爲標準。目的既定，然後才可作實際生活程度之研究，而得着手編製指數也。

III 實際消費之研究

實際消費之研究，亦即基本預算之研究，就是討論實際消

費之狀況，求出生活費中究包括何種物品之費用，及各物品費用彼此相互輕重之比例如何，或各物消費之數量如何。然生活費包括之物品，何止千百種，不得不擇其重要者幾項，各附以消費之數量，或費用上彼此相互之比例，列成一表，稱爲基本預算表。因之，一方面可以確定消費物品之種類，以便調查其零售價格，一方面可以決定各物品費用彼此輕重之比例，定爲權數，以便計算指數。這種研究，實爲編製指數之第一步驟。

預算表中所列物品及役務 (Services)，通常分爲五大類，食品，衣服，房租，燃料與燈光，及雜項是也。英，美，法，意，比，挪威，丹麥，芬蘭，波蘭，瑞典諸國，及我國已編之生活費指數，皆採用之。間有删去一類或二類者，如德奧與孟買之生活費指數，不列雜項，埃及不列燃料，坎拿大不列衣着及雜項。

預算表製定之方法，最要者有三種。

1. 總合消費法 The Aggregative Expenditure Method 綜合消費推算之法，乃就一年或某一時期內國內生產之量，加以進口之數量，而減去出口之數量，作爲全國之總消費。如不用量而用價值，其法亦同。預算表中各物之權數，可任用消費量或消費值。故全國之消費額，或以量計，或以值計，均無不可。設縮小其範圍以求一城市之消費總額，亦可應用，即將一城市之貨物輸入，加出產，減輸出，即得該地之消費總額，可以

代表當地一般之生活程度。

此法在生產與輸出入統計完備時，極爲扼要，但亦有缺點。

a.物品之消費量旣與輸出入有關，則國外市場的供給需要及經濟狀況有變動時，難免受其影響。

b.物品之消費量，旣和全國生產量有關係，則生產狀況變更時，亦受其感應。

c.此法雖可代表全社會，但不能代表某特定階級。

d.社會上消費習慣，時常有變遷，生活程度，並非一成不變，這消費數量，不能爲長期之比較。

e.房租一項，須另行調查。

f.雜項中屬於役務之事項如看電影，聽戲，遊覽之類，無法於全國消費總額中求出。

因此採用此法，至少有一部分須採用實際調查來補充。

2.標準家庭預算法(The Standard Family Budget method)一稱生計調查法。(註1)這方法先決定調查範圍，再就調查範圍內選取家庭人數與經濟能力約略相仿之若干家庭，作爲標樣，然後用問題表或記賬法，求其各家之平均各項消費額，及其平均消費總額(註2)以平均消費總額除各項平均消費額，卽得各項

(註1) 詳參拙著家庭生計調查之研究。

(註2) 此項平均消費，又可以家庭組織之單位數除之，而得每單位之消費，俾免家庭組織不同之影響。

消費額對於總額之百分數；然後以此爲選擇物品之標準，及計算指數之權數。問題表與記賬法比較，問題表雖則簡單易行，但其弱點在難得詳盡之資料。因生活消費之物項甚多，問題表之篇幅有限，且問題表所得物項消費之數量，乃由被調查者回憶或估計而答覆之，自難十分準確。至於記賬法，係將家用日記簿分發被調查者之家庭，囑其填寫或派員逐日至其家而爲之代記，所得材料當然精確。但其弱點即在費財力費時間，難得大量之家數，普通可兼用二法，相輔爲用，則能互補其短而得各盡其長。

標準預算法，決定權數之資料皆直接從有關係的家庭調查之，所選資料，必取其足以代表全國或某區域某階級某職業之人，例如欲爲礦工或農業工人編製一種生活費指數，則所選家庭當然限於這類工人，又如欲觀察小學教師或商店職員之生活費，所選家庭亦當限於這類人員，本方法各國採用最多，我國北平上海南京天津之生活費指數均用此法。

3. 理論的家庭預算法（The Theoretical Family Budget Method）上列二法須在貨物之消費，無甚變遷之時，方裨實用；反之，凡在消費習慣或貨物項目有劇烈變遷時，則不適用，於是有採用所謂理論家庭預算法。本法即將預算表中之各種食物消費量，根據食物化學分析之學理而決定，使熱量及各種滋養之成分，各有適當之分量。所謂適當分量，隨人之性別，年

齡等而異。故計算一家庭之食品，滋養成分，須將男女老幼各人，合成同一單位，稱爲等成年男子數（Equivalent Adult Male）。然後求其等成年男子數，以決定平均家庭應需某種食品各若干。若以實際家庭消費者相參訂，乃爲最後之確定。

一成年男子每日所需之各種滋養數量，各學者之標準互異，其通常之標準，約略如下數：

熱量	3000 卡羅里 (Calories)
蛋白質	90 格蘭姆
脂肪	100 ,,
水炭素	440 ,,
鈣	.68 ,,
燐	1.32 ,,
鐵	.015 ,,
維他命 A.B.C.D.E	數量研究尚未準確

實則各種滋養數量之成分，又隨人之體重及工作而不同，故上項普通標準亦僅一概括之數量。

以上祗就食品類而言，其他各類之消費，都要應用科學方法以研究之，於是困難叢生，本方法之採用不廣。嚴格講，理論預算表并非全憑學理，往往參酌實際之生活情形，此點不可不加注意。德國聯邦統計局所編生活費指數。係採用此法。

就方法論，理論預算表既注重於滋養之成分，只須滋養成分不變，物品可以替換，故爲用甚爲活動便當；不過代入新物品，使指數之內容變更，卽係生活程度之改變，失生活費指數之本意，未始非缺點也。

四　測量生活程度之量表

家庭中各分子之消費，往往因其年齡性別之不同而異，吾人研究家庭每種平均消費，應將一切人口合成共同單位 Common Unit 表示之，以便比較。此種單位，通常又稱量表Scale；可分三類如下：

1. 等成年男子(Adult male equivalent)，

2. 成年男子維持費(The Ammain)，

3. 消費單位(The Cost Consumption Unit)。

等成年男子之量表(食品類)

年齡	Engels quet system		Atwater		Lusk		Amsterdam		美國制	
1歲以下	100		30		50		15		15	
1及 2歲以下	110		30		50		20		15	
2及 3歲以下	120		40		50		30		15	
3及 4歲以下	130		40		50		35		15	
4及 5歲以下	140		40		50		40		40	
5及 6歲以下	150		40		50		45		40	
6及 7歲以下	160		50		70		50		40	
7及 8歲以下	170		50		70		55		75	
8及 9歲以下	180		50		70		60		75	
9及10歲以下	190		50		70		65		75	
10及11歲以下	200		60		83		70		75	
11及12歲以下	210		60		83		75		90	
12及13歲以下	220		70	60	83		80		90	
13及14歲以下	230		80	70	83		85		90	
14及15歲以下	240		80	70	100	83	90		90	
15及16歲以下	250		90	80			100	90	100	90
16及17歲以下	260		90	80						
17及18歲以下	270		100	80						
18及19歲以下	280									
19及20歲以下	290									
20及21歲以下	300									
21及22歲以下	310	300								
22及23歲以下	320	300								
23及24歲以下	330	300								
24及25歲以下	340	300								
25歲及　以下	350	300								

註：　表中兩行字，大者表示男子之消費，小者表示女子之消費。

等成年男子數 係將各家庭之食品消費，以成年男子爲標準，其他人口之消費，折合於成年男子數。此種主要量表，有 Engel's quet 制，Atwater 制，Lusk 制，Amsterdam 制及美國制等。

就中以 Engel 量表爲最早而著名。該表又稱 Quet 制，所以紀念統計學泰斗 Quetlet，蓋 Engel 所用之資料，爲Quetlet 研究之結果。氏以爲食品消費和一人之生育發達有關係，即食品消費，和身長身高成正比例，氏乃統計人之平均長度及高度，爲計算食品消費之根據。沿用此量表者，一九二一年瑞士及比利時之調查。

Atwater 量表以一八九五——一八九六年紐約食品調查之材料爲根據，該調查由 Atwater 教授及 Wood 担任，但後經修正。此量表最爲通用，我國上海天津北平等所舉辦之調查，亦皆採用此量表。

Lusk 量表爲美國 Cornell 大學教授 Lusk 氏所制定，以人之生理上所需要之食品量爲根據，沿用此制者有英國孟買埃及等調查。

Amsterdam 量表，以食品作根據，爲一九一七年三月該市調查所定，該市以後各次調查，多以此爲標準，海牙市政統計局，亦採用此量表。

美國制爲勞工統計局所訂。採用此制者，如瑞士一九一九

，一九一七，一九一八年等調查，瑞士一九二〇年政府農業調查，丹麥政府一九一五，一九一六，一九二二年調查，挪威政府一九一八，一九一九年調查，南非一九二五年調查等。

上列各量表，互有不同，其原因或由編製方法之不同，或由各國氣候及生活程度之互異，其詳簡亦各不同，例如Quet制較 Lusk 制詳備。有幾種量表，對於男女二性消費量之區別，僅限於成年男女，但大多數之量表，於未成年男女之消費，不列差別。人之消費以年齡幾歲達到最高限度？此問題之答案，各量表不一， Quet 量表以男子二五歲，女子二〇歲爲消費最高度。其他各量表，年齡還要小些。表中各量表，對於老年人消費量，應減少一點，并未規定；大概因爲此種減少之規定，於家計調查研究不足重輕。

成年男維持費　食品消費之標準，未必適用於其他各種物品與役務之消費，當另有其標準，成年男維持費卽由此而作。此種著名量表有二，一爲德國修正制，一爲澳大利亞制。

德制由德國統計局於一九〇七年作家庭預算調查時所定，不僅限於食品一類，凡生活費內重要項目，多包括在內。採用此制而略加修正者，有 Stockholm 調查（一九〇七——一九〇八），波蘭一九二二年調查，捷克迭次舉辦之調查等。

澳大利亞制，由該國淸查統計局於一九一三年作家計調查所定。

成年男子維持費之量表

年齡	澳大利亞制		德國修正制	
1歲以下	20		10	
1—2歲以下	20		10	
2—3歲以下	35		10	
3—4歲以下	35		10	
4—5歲以下	35		20	
5—6歲以下	35		20	
6—7歲以下	50		20	
7—8歲以下	50		30	
8—9歲以下	50		30	
9—10歲以下	50		30	
10—11歲以下	65		40	
11—12歲以下	65		40	
12—13歲以下	65		40	
13—14歲以下	80	70	50	
14—15歲以下	80	70	50	
15—16歲以下	80	70	70	60
16—17歲以下	80	70	70	60
17—18歲以下	100	80	90	70
18—19歲以下			90	70
19歲或以上			100	80

(表內兩種數字，大者表示男子消費，小者表示女子之消費。)

此外又有美人 Edgar Sydenstricker 及 Wilford I. King所制定之一種，以男子年在二三——二六歲之消費力作標準，認當此年齡之消費力爲最高限度，各年歲各性別人口之消費，當折合於此數，所得比例，卽爲各人口之相對消費。其量表載 U. S. Public Health Service, November 26, 1930. 作者手邊無此書可參，祇得從略。其計算似較上二法爲精密，分級情形，大約如下：

1. 男子年齡在「二三至二六歲」維持費作一，
2. 男子年齡自「一歲」長至「二三至二六歲」相對消費，自·二二漸增至一，

消費單位量表 (E. L. Kirkpatrick)

食品：性別及年齡	第一人	第二人
男子：		
19歲及以上	1.0	.9
15歲至18歲以內	.8	.7
女子：		
19歲及以上	.9	.8
15歲至18歲以內	.7	.6
男或女：		
12歲至14歲以內	.6	.5
6歲至11歲以內	.4	.3
5歲或以下	.3	.2

衣服	
家主	1.0
主婦	1.0
其他人口：	
24歲以上	1.4
19歲至24歲	1.7
15歲至18歲	1.3
12歲至14歲	1.0
6歲至11歲	.6
1歲至 5歲	.4

傢具 壽險及健康保險	
家主	1.0
主婦	1.0
其他人口：	
第一人	.4
第二人	.3
第三人	.2
第四人	.1
第五人及以上	.0

房租	
家主	1.0
主婦	1.0
其他人口：	
第一人，男，15歲或以上	.2
第一人，女，15歲或以上	.2
第二人，男，15歲或以上	.0
第二人，女，15歲或以上	.0
第三人，男，15歲或以上	.2
第三人，女，15歲或以上	.2
以上類推	
第一人，男孩，6—14歲	.1
第一人，女孩，6—14歲	.1
第二人，男孩，6—14歲	.0
第二人，女孩，6—14歲	.0
第三人，男孩，6—14歲	.1
第三人，女孩，6—14歲	.1
以上類推	
五歲以下之人口	.0

醫藥衛生	
家主	1.0
主婦	1.0
子女：	
24歲以上	.4
6—24歲以上	.2
6歲以下	.6

娛樂	
家主	1.0
主婦	1.0
子女：	
19歲以上	.5
15歲—18歲	.3
6—14歲	.1
6歲以下	.0

個人性質之支出	
家主	1.0
主婦	.6
子女：	
子，19歲及以上	1.0
子，15—18歲	.5
子，6—14歲	.4
女，19歲及以上	.6
女，15—18歲	.5
女，6—14	.3
5歲以內之子女	.2

3. 男子年齡自「二三至二六歲」長至「八〇歲」相對消費自一起漸降至·七四，

4. 女子年齡自「一歲」長至「二一至二五歲」，相對消費自·二二漸次增加至·七九，

5. 女子年齡，由「二一至二五歲」長至八〇歲，相對消費自·七九漸次減低至·六二。

消費單位之量表　264 頁之表為消費單位量表之一例，由美國於一九二一——一九二二年為調查紐約 Livingston 地方四〇〇家農民家庭而定，主其事者為E. L. Kirkpatrick 氏。此量表承認主婦對於各種家庭生活所需之費用，除食品及個人性質之費用外，與成年男子相等。從實際數字上，發現衣服費用之支出，主婦與家主差不多；至於住屋器具設備衞生等費用，亦復大約相等。此量表具有二優點：(一)，分別各項目而各求其消費標準，(二)規定家庭中增添人口之相對消費。

此種種之量表：各國根據其己國之國情，人民之生理而制定，吾人研究中國人之生活程度，最好根據中國人之情形及生理，從醫學及經濟方面，製成中國之量表。所惜國內統計事業，方在萌芽，覺付缺如。

V 消費項目

生活費指數之分類，各國雖互有不同，大抵多採五類，已

述於前，惟各類之中，究應包含物品若干，各國指數亦多少不一。如維也納指數，食物類僅16品其中占咖啡三項，而挪威指數之食物類多至55品。雜類中包含物品數目之多寡，尤各不同，德奧指數不列雜類，而在美國傢具及家用品列至12種之多，并另設一類列入租稅，醫藥，旅費，娛樂費等43目。以衣着論：有竟不列者；有僅列數品者，如奧之三品是；亦有分別季節，列入很多，如美國採用常年服用品40，夏令服用品14，冬令17是。可見生活費指數，除燃料及房租二類外，其除各類，所採物品，往往多寡不等。再據實際調查所得結果，品目繁多，製定基本預算時，若盡取一切消費項目而列入之，則調查物價，計算指數，將不勝其繁重，事實上幾為不可能，亦無需於此。蓋生活費指數所測量者，乃生活費之變動，而非在各時期維持最低生活或安適生活所需用之一切物品之總共費用。設能愼選物品，使其物價之變動，足為未經選入同類各品之代表，則删去比較不關重要之物品，於指數之結果，毫無影響。

惟生活費指數之品目，雖則不必將實際消費品目，都搜羅在內，但為適合實際消費狀況起見，各類品目之消費值，照理當與其全消費值保持相當之比例，方能與實際預算所得之結果相符。玆將各國生活費指數所包含之物品，及其各類消費值之百分比，分列二表，以便參攷。

表65　各國生活費指數食物類，燃料燈火類，衣着類包含項目表

國別	食物	燃料燈火	衣着
南非洲	20	3	未明
澳洲	41	2	未明
比利時	33	7	25
加拿大	29	5	未明
西班牙	19	8	無
美國	43	6	54—57
英國	20	5	8
印度			
工人	17	3	3
歐人	30	4	22
盧森堡	13	4	2
挪威	55	未明	未明
荷蘭	13	未明	未明
瑞士	50	9	25
德國	17	4	13
奧國	16	4	3
布加利亞	36	6	5
埃及(開羅)	20	2	無
愛爾蘭	36	6	23
芬蘭	14	1	2
匈牙利	10	2	5
意大利	18	未明	16
麻省	37	4	17
紐西蘭	55	7	無
波蘭(華沙)	16	2—3	7
捷克斯拉夫	42	10	無

VI 搜集物價

物價爲編製指數之原料，欲求指數之精確，第一要件，即在搜集精確之物價。物價有批發零售二種，生活費指數既爲消費者而編製，原則上自當採用零售物價，但亦有因調查零售物價困難，而用變通辦法，以批發物價替代之。

零售物價，缺乏標準，同在一城，各處不同，一日中亦有數變，欲求準確，自宜取材較廣，而用其平均價。至於品質牌號等，較批發物價，尤爲紛雜，在衣服類及雜類更甚。調查時

表66　各國生活費指數各類消費値百分數

國別	調查時期	食物	房租	燃料及燈火	衣着	雜類
奥(維也納)	1921	65	1.6	14.4	19.0	——
比利時						
第一級工人家庭	1921	68.13	7.68	6.70	18.84	5.65
第二級工人家庭		66.80	7.76	6.08	13.41	5.95
第三級工人家庭		63.94	6.20	5.03	14.57	10.26
第四級工人家庭		67.88	4.77	4.09	18.83	11.68
中上階級之家庭		64.	9.65	4.82	13.81	7.72
加拿大	1913	52.6	33.9	13.6	——	——
丹麥	1922	41.8	12.4	4.0	11.3	30.5
埃及	1920	51.9	11.7	——	16.7	19.7
愛爾蘭	1922	57.1	5.4	7.0	17.5	13
芬蘭	1908—9	55	11.8	4.1	11.6	17.5
美國	1918	38.2	13.4	5.3	16.6	26.4
法(巴黎)	1914	60	12	5	15	8
英國	1909及1912	60	16	8	12	4
印度孟買						
工人指數	1909—1914	81.7	9.1	4.9	4.3	——
歐人指數	1923	11.3	18.5	4.3	6.8	59.1
意大利	1913	62.09	11.4	4.5	12	10
挪威	1912—13	47.9	15.65	5.2	12.65	18.55
荷蘭(阿姆斯特丹)	1920	49.4	7.5	4.4	17.9	27.8
波蘭(華莎)	戰後	59.5	6.1	8.7	20.2	14.6
瑞士	1913—14	48	15	4	12	26

，總要使品質等級劃一。有標準牌號者，必註明牌號；無牌號者，亦當附帶樣品，或填列上期價格，庶報告者可認清品質，不會誤報。

搜集方法，可有下列數種。

1.根據報章及行情單所刊布之物價，

2.由職業調查所或消費合作社，代為搜集，

3.製成表格，分送零售商，令其填報，

4.房租一項，有由產業聯合會報告，或由商會或地方政府供給材料，

5.根據政府機關或各團體之調查，

6.派員直接至各區調查。

搜集區域　搜集區域，要看統計之目的而異。如爲代表全國之生活費指數，則調査之區域，當遍及全國；如指數之目的僅在測度一地方生活費之變動，則調査範圍可以一地方爲限；如爲某種工人特編之生活費指數，用作改訂工資之標準者，則物價自以此種工人集居地域之零售市場直接取得者，較爲適當。

搜集時期及次數　搜集時期及次數之多少，當看物品的性質與價格變動之劇緩而定。變動和緩之物價，如油，鹽，布疋等類，每月調査一二次已夠。變動較劇或具有季節性之物價，如魚，肉，蔬菜之類，應多調査幾次，而求其平均。房租一項，變動最少，調査之時間，相隔不妨稍久。

Ⅶ 指數之計算

計算指數，有基期及公式二大問題。

1. 基期　生活費指數所用之基本預算表，比較地固定，少變動，少有更動項目之麻煩，所以生活費指數，常用固定基期。至基期之選定，當在物價平穩時期，1925 年國際勞工統計專家會議，主張用 1930 年爲基期，幷希望各國一致採用，以便作國際間之比較研究。基期之長短，普通言之，長時期勝於短時期，此等原理，已於討論基期時敍述之，讀者互參可也。

2 公式　指數公式，雖有百數十種，然在生活費指數中，普通祇用二種加權公式，(1)加權綜合法，(2)加權算術平均法，尤以採用第一法爲多，試分述於下。至於簡單公式，大都無人問津。

a.加權綜合法

$$\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0}$$

q_0爲基期之物量，p_0爲基期之物價，p_1爲計算期之物價。應用此式，各物量之調查，須與基期同一時期。但基本預算表在基期之前後，或有一部分在基期之內者，則公式可略加變通如下：

$$\frac{\Sigma p_1 q_m}{\Sigma p_0 q_m}$$

q_m爲任何指定時期各物之消費量。

b.加權算術平均法　公式如下

$$\frac{\Sigma p_0 q_0 \frac{p_1}{p_0}}{\Sigma p_0 q_0} \text{ 或 } \frac{\Sigma p_m q_m \frac{q_1}{q_0}}{\Sigma q_m p_m}$$

$p_0 q_0$爲基期各物之消費值，$p_m q_m$爲任何指定時期之消費值。

上述二式，形異而實同。第二式若消去p_0即爲第一式，故由指數準確之立足點看，無分軒輊。至於計算方面，則第一法較爲簡便，因爲省去計算價比一步手續。且意義明確，便於改

換基期，又其所長。但第二法亦有長處，就是在於活動便當。在總合法中，係由物價與具體之消費量相乘，所用者為實價，在價比法，係以價比與權數相乘，所用者為價比，不同之物品，實價不能平均，而價比則可平均。

計算生活費指數，大都須分別計算分類指數及總指數。求分類指數，無特別之點；計算總指數，因有權數關係，方法有三：

1 直接加權法　若將求得之各類指數分別加權并計算其平均數卽得總生活費指數，此法可不必計及各類指數係用何種公式求得，均可應用，試以民國十九年一月分天津工人生活費之分類指數為例，示其計算如下

	各類指數	權數
食物類	119.25	62
服用品類	101.62	6
燃料類	118.74	13
房租類	123.41	14
雜項類	142.51	5
		100

總指數(假設)為：

$$\frac{119.25\times62+101.62\times6+118.74\times13+123.41\times14+142.51\times5}{100}$$

$=119.87$

2 間接加權法　是在製定基本預算時先將各類所含物品之多寡選擇適當，使各類代表之費用，成適合之比例，此法祇在各類指數均用總合法計算之場合，始能應用。我國北平生活費指數卽用此法，玆假設一例於下。

設　食品類指數：　$\frac{159.50}{145}\times100=110.0$

衣服類指數：　$\frac{15.60}{14}\times100=111.4$

燃料類指數：　$\frac{26.47}{23}\times100=115.1$

房租類指數：　$\frac{12.47}{15}\times100=83.1$

雜項類指數：　$\frac{6.82}{6}\times100=113.6$

則　總指數：

$$\frac{159.50+15.60+26.47+12.47+6.82}{145+14+23+15+6}\times100$$

$$=\frac{220.86}{203}\times100=108\text{(19年1月)}$$

3 增補權數法　生活費指數之品目，雖不必將預算表中之品目網羅無遺，惟爲適合實際消費之狀況起見，則各類品目之消費值，在理當與其總消費值保持相當之比例，方能與家計調查所得之結果相符，欲達到此目的，在計算總指數時，可採用增補權數。例如編製上海生活費指數，家計調查之結果，平均每家全年消費值共$390，各類分配爲：食品類$218.5，衣服類 $36.7

；燃料類$29.3；雜項類$80.40；房租類$25.1。而實際選入指數之物品消費值爲：食品類$198.8；衣着類$22.4；燃類$29.3；房租類$25.1；雜項類$14.6。

計算總指數，應將食品，衣着，雜項三種，採用一種增補權數。玆以民國十九年一月份上海生活費指數爲例，該月之各類指數及總指數如下

食品類	106.0
衣着類	99.3
房租類	103.9
燃料類	121.6
雜項類	141.1
總指類	113.6

此總數之求得，由於

$$P_{01}=[(106.0\times198.8+106.\times19.7)+(99.3\times22.4+99.3\times14.3)+(103.9\times25.1)+(121.6\times29.3)+(141.1\times14.6+141.1\times65.8)]/390$$

$$=113.6$$

四 我國之生活費指數

1. 北平生活費指數 北平生活費指數係社會調查所爲北平一般工人所編製。民國17年底初次發表，由民國15年一月起

逐月編製。而在平津各報中發表之。

表67　編製北平生活費指數選定之物品及其消費量

物　　品	單位	數量	物　　品	單位	數量
食品					
米麵類			肉類		
伏地白麵	斤	391	羊肉	斤	7
玉米麵	斤	482	豬肉	斤	3
小米麵	斤	370	牛肉	斤	3
白米	斤	67	燃料		
小米	斤	80	煤球	斤	1862
蕎麥麵	斤	71	煤油	斤	26
蔬菜類			劈柴	斤	37
			木炭	斤	46
醃蘿蔔(全年)	斤	21			
水咯噠(全年)	斤	11	衣料		
葱(全年)	斤	13	月白市布	尺	13
豆芽(全年)	斤	16	青市布	尺	7
大白菜(冬春)	斤	142	灰粗洋布	尺	22
白薯(冬春)	斤	50	毛藍粗洋布	尺	27
菠菜(冬春)	斤	18	白粗洋布	尺	25
小白菜(夏秋)	斤	76	愛國布	尺	21
韭菜(夏秋)	斤	47	棉花	斤	5
茄子(夏秋)	斤	48	住房		
黃瓜(夏秋)	斤	61	房租	間	1
油及調和類			雜類		
香油	斤	11	茶葉	斤	2.0
鹽	斤	26	洋碱	斤	1.4
黃稀醬	斤	16	肥皂	塊	4.4

家計調查　採用標準預算法。該會於民國 15 年 11 月至次年 4 月，選定北平工人四十八家，派員逐日爲之記賬，根據此四十八家記賬之資料，加以補充和修正，編成上列之預算表(表67)。

此項預算表中共計 38 項，其中食品 23 項，燃料 4 項，衣料 7 項，雜項 3 項，房租 1 項。上項物量，以調查時之物價計算，共值 $160，與實際費用比較(平均每家全年 $203)約佔80%。

物價調查　自民國17年6月起，每月於二日及十六日調查二次，每物之價，係用六家店鋪之價求其平均。物價之供給，除房租外，其餘各物，按售賣之店鋪，分爲七類，共42家中得之。房租則由拉房縴者每月報告之，間有採數人之報告，以資校對。

指數之計算　各分類之計算公式，係用綜合加權法，總指數之計算，則用間接加權法。所取基期爲1927，因該年之物價較爲平穩也。此指數發表之時，有一特別之點須注意者，卽除刊布五類分指數及一總指數外，復加一行生活總費按銅元物價計算之指數。因工人之收支，多有以銅元計者，我國幣制不良，銅元本身，漲落甚巨。以銅元爲收支之單位者，其生活費之漲高，常較用銀元之工人爲速；因不但受物價之漲高，且受銅元兌換跌落之影響，故添此一行，以明此類工人之生活費之變動。

2. 天津工人生活費指數　爲南開大學社會經濟研究委員會主編。

家計調查　自民國16年9月起，在天津舉行工人家庭生活調查，爲期十月，選擇工人家庭240家。所選多爲手藝工人之家庭，每日派員持日用流水簿，分至各家調查而登記之。根據此十月調查之結果，求得各家平均每次消費物品額，與其平均消費物品總額之百分比，以爲選擇指數中所應包含之物品之標

準。其結果得物品40種，見表68。

加權　所用權數有二(1.)爲家庭生活調查所得之平均每家各該項物品之消費數量，見表。(2.)爲各類物品之權數，計食物類爲六二，衣服類爲六，燃料類爲一三，房租類爲一四，雜項類爲五。

搜集物價　搜集方法，係於每星期三派員調查之。每月物價，爲該月各星期三之平均價格。搜集地點有四，即河東郭莊子，南關大街，西頭如意巷，河北小王莊，均爲工人住居之所。

表68　天津工人家庭平均每家消費物品之消費數量

物品	單位	消費量	物品	單位	消費量
食物類			服用品類		
上海大米	升	96 25	棉花(北河)	斤	4.81
美國白麵(次)	斤	221.94	高陽布(八斤重)	尺	4.55
伏地白麵	··	1.38	本地市布(花鳥)	··	51.15
北河玉米麵	··	467.06	西洋市漂布(十八子)	··	46.66
海下小米	升	27.56	東洋花標布(眼鏡)	··	10.55
北河綠豆	··	12.88	東洋絲綸(美人)	··	4.68
黃豆菜	斤	31.13	西洋冲禮服呢(人檬)	··	1.10
綠豆菜	··	35.63	東洋直工呢(寒山寺)	··	3.04
土豆	··	47.25			
豆腐	··	17.25			
葱	··	31.94	燃料類		
蒜	··	31.19	煤球	十斤	47.86
大白菜	··	167.56	木炭	斤	18.50
伏地菠菜	··	12.56	柴禾	十斤	82.69
小白菜	··	25,88	煤油(幸福油)	斤	45.06
韮菜	··	30.13			
醬油(中等)	··	17.69	房租類		
醋(次)	··	8.13	灰房	間	1.40
香油(次)	··	13.31	土房	間	1.20
食鹽	··	28.25			
麵醬(次)	··	5.50	雜項類		
猪肉	··	24.19	紙烟紅屋牌	小盒	19.00
牛肉	··	3.13	自來水	担	132.90
羊肉	··	8.25			

計算公式與基期　公式係用加權綜合法，而以基期各物品之消費量爲權數，基年爲1926年。

3. 上海生活費指數　爲國定稅則委員會主編。

家計調查　採用標準家計調查法，於民國16年11月，開始調查，至17年10月止，計歷時12個月，家數爲230。在上海各工業中，以棉紡織業最居重要，故所選家庭，多數爲紗廠工

表69　上海生活費指數選入指數之物品及其權數

物　品	權　數（即平均等成年男子數3.7之工人家庭每家全年消費值）單位：銀幣大洋一角	物　品	權　數（即平均等成年男子數3.7之工人家庭每家全年消費值）單位：銀幣大洋一角
食物類		衣着類	
白米	998	棉花	14
麵粉	114	粗布	33
綫粉	15	細布	49
青菜	108	花標	12
菠菜	11	細斜紋布	37
鹹菜	25	線呢	30
豆芽	28	棉直貢呢	24
葱	16	長統線襪	25
蕃芋	14	合計	224
蘿蔔	27	房租	
黃豆	24	單幢住房	251
豆腐	89	燃料類	
花生	5	木柴	218
豆油	146	煤球	24
醬油	47	煤油	47
白糖	15	火柴	4
鹽	26	合計	293
猪肉	120	雜類	
牛肉	28	香烟	78
鮮魚	78	紹酒	37
鹹魚	12	茶葉	10
鷄	15	肥皂	14
鴨蛋	24	牙粉	1
鷄蛋	3	毛巾	6
合計	1988	合計	146
		總計	2902

人。預算表所選物品共43種，所選物品及權數見表69。

表中食物類所選物品計24種，占家計調查本類消費值91%；衣着類計8種，占其消費值61%；房租，燃料各占其消費值之全部；雜類計6種，占是類消費值18.2%。其全年消費值爲290.2元，在平均等成年男子數3.78之工人家庭每家每年開支中(390元)計占74.4%

物價　係用派員調查法，除房租外，每月一日及十五日各調查一次。食物類中肉，魚，蔬菜之零售價格，係由該市八個小菜場調查而取其平均。房租擬每季或每半年調查一次。

基期及公式　基年取民國15年全年平均。公式爲加權算術平均法。

$$\frac{\Sigma W\frac{p_1}{p_0}}{\Sigma W}$$

p_0 爲基期物價，p_1 爲計算期物價，W爲平均等成年男子數3.78每家之全年消費值。計算總指數時，關於食物，衣着，雜項三類，更採用一種增補權數，以補足各類權數與其消費值之差額。

4. 上海市工人生活費指數　爲上海市社會局所主編，自民國十五年一月起。

家計調查　採用標準家庭預算法，於民國十八年四月開始調查至次年三月止，歷時十二月，選定人口在三人至六人，和

每月收入大約在二十元左右的工人家庭，結果得三百零五家，共得賬簿三千六百六十本。根據此種賬簿，來算每家每種物品之平均消費額，以觀察工人階級實際之消費狀況。幷根據之選得指數中應包括之物品，共六十品：計食物31，衣著11，房租3，燃料8，雜項7。

加權　各種物品所用權數就是調查所得之平均每家各該種物品的消費量。計算總指數，並不再用分類權數。

物價　物價調查之區域均為工人集居之地，定南京路為全市中心，分東，南，西，北，及浦東五區，依區域大小，每區選查代表商鋪若干家，幷調查菜市九處。調查方法，由該局按期派員持調查表分赴各特約零售商鋪指導填寫。所選商鋪，多為歷史較久，規模較大，交易興盛，花色齊全，而為工人所樂於光顧者。

計算公式及基期　公式係用加權總合法，而以家計調查每種物品的消費量，作為固定權數。基年為民國十五年。

上列四種生活費指數，各主編機關，均印編製說明書，讀者欲知其詳，參閱可也。

第四章

工資指數

工　編製目的

工資統計之編製，可隨編者意旨之不同而異，故在敍述編製工資統計之目的，可從編者之立場敍述。

1 僱主　工資在僱主的立場，僅爲生產費用(Cost of Production) 之一部分。

2 勞工主管機關及社會改革家　工資問題，幾爲現代勞資糾紛之焦點。勞工主管機關，爲着手調解勞資糾紛起見，希望憑公正的數字，作調解之根據。至於社會改革家，不忍見幾千萬勞動者的生活，日趨困苦，思利用平和手段以改進工人生活，所以對於工人藉以爲生之工資，更有統計研究之必要。

3 工人團體　工人團體，爲謀他們所代表之勞工羣衆自身之福利起見，便不能不編製工資統計，去作經濟鬪爭的論據，美國勞動總同盟之編製各種統計，卽其一例。

4 研究工資學說的科學家　工資問題，在今日之嚴重，已是不可諱言，因此，早引起一般科學家之注意。研究工資學說之科學家，往往利用工資統計之資料，作爲研究學說，論斷的

證據。例如：研究工資與人口之相關，以檢討生存工資學說(Subsistence Theory,)研究工資與僱用人數的相關，以檢討工資基金學說(Wage Fund Theory)。

就上述四種不同之編著立場，綜合來說，亦不過二種目的。第一，是指數之目的，在表示工資爲生產費用之一部，第(1)類屬之。第二，是指數之目的，在測量工人之生活程度，及貨幣購買力之消長，表示工人經濟狀況之變遷，第(2)(3)(4)類屬之。

II　收集資料

工資指數資料之來源，可由下列各方面收集之。

1. 僱主工資簿，
2. 僱主與工人所訂合同，
3. 勞工團體之記錄，
4. 地方代理人及工廠監督員之報告，
5. 職業介紹所之報告，
6. 疾病及意外保險公司之報告，
7. 政府及其他團體之直接調查及報告，
8. 工廠呈報官署之工人名册，
9. 勞資爭議協調決定書。

5. 6. 9. 三處來源，祇限於局部少數特殊之工人，故應用最

狹。而且在我國，職業介紹所之組織，非常簡陋，無甚報告。疾病及意外保險公司尚無此種組織，地方代理人及工廠監督員現尙缺如。僱主與工人商定之工資，多屬口頭說定，具書面合同者幾爲少數。因之，祇有(1)(3)(7)(8)四款，可算我國工人工資資料，較爲妥善的來源。如無適當之材料，可資利用，則惟有直接調查矣。

Ⅲ 工資分類

工資材料搜集後，必須適當分類，工資分類，是勞工統計中最麻煩之一事。普通分類方法，有下列二原則。

1. 編製指數之目的，如在表示工資爲生產費之一種，比較勞工費用以增進僱主之利益，且視其異時異地之升降，則工資分類當先依工業分類，然後再將各業依照工人職務分類。

2. 如目的在計算眞實工資或社會工資，以顯示勞工底經濟實況，則工資分類應先依職務分類，然後每類中再分爲性別年齡技能等。

Ⅳ. 工資指數之種類

工資指數，總括說可分二類：工資率 (Wage rate)和實際所得(Actual Earning) 。

1. 工資率指數是表示僱主所定購買工人勞動力的價格之標準。就工資率的形式講，有計件計時二種。計時工資率以工作

時間之長短爲計算工資之標準，有按時，按日，按週，按月，甚有按年計算者。計件工資率以出品單位爲計算工資之標準。二種工資率，在形式上雖有區別，實質上互相接近。計件工資率之標準，還是依照計時工資規定的，不過在計時的標準之下，在同一時期內，各勞動者如果職務相同，則所得工資有相同之可能，而在計件標準之下，雖在同一時期內，各勞動者所得工資，未必即能相等，技能之高下，體力之強弱，工具之優劣，工作之勤惰，在在足以影響其所得。總之，工資率指數，是根據每個工人在一固定期內應得之額定工資數而編製。工資率指數的計量單位，有爲每週的，有爲每日，有爲每時。近數年來因科學管理盛行，各國工資率統計，都有爭以每時爲單位之趨勢，這不消說是因爲計量單位愈細密，愈能表現出產品的勞工費用，愈能適合僱主目的之故。

2. 實際所得指數　實際所得，是指工人於一定時間內，實際收入之數，實際所得指數，是於額定工資數之外，還要加入米貼，房貼，升工，賞工，紅利，年底加薪，或其他種種額外進款；并且把那些因請假而扣除的工資，因違犯廠規，損壞機件，出品不良等等而罰扣的工資減除之後，編製而成指數。實際所得指數之目的，在表現勞工的經濟狀況，所以用較長的時間單位，較爲適宜，而且事實上各種額外收入的發給都是一週一次，一月一次，甚或一年一次。故實際所得以全年所得爲合

宜，每月或旬週者次之。但近世各國，劝有以每時爲實際所得統計之計量單位，若然，工人之平均旬週工作時數，當並列於統計表中。

V 指數之計算

計算指數之重要點，有二：公式之取決與基期之確定。

1. 公式有加權簡單之分，而每種又有數類，已詳於通論章，不贅。至於編製工資指數究應採用何種公式，查閱各國所編指數，採用之公式，各不一致。例如南非合衆國與美國，採用加權算術平均法，德奥等國，採用簡單算術平均法，上海特別市十八年工資指數之試編中，擬用加權總合法，而加權之數爲基期某業或某職或某種工人之工人數。總之，加權公式，雖優於簡單公式，惟取材困難。究竟應用何式，最爲相宜，學者意見紛紛，迄無定論。玆將各國工資指數之編製法，列於表70，以備讀者之參攷。

2. 基期 編製工資指數，當採固定基期，若沿用太久，認爲不適時，可每隔若干年，變換一次。基期時間，亦宜以包括較長之時間爲較優。至於時期，歐美各國多用1913爲基年，該時期雖能代表戰前之情形，惟距今太遠，似有更換之必要，故最近趨勢，大有擺脫1913之勢。總之，基期之選定，當以該時期之工資甚爲平穩，而無劇烈變動者爲上。再者，工資調

表 70 各國工資指數編製方法一覽表

國名	編製機關	範圍	材料性質	材料來源	分類法	計算公式
南非合衆國	Office of Census Statstics	少數區域	工資率	工團及僱主報告	先以工業分再以職務分	加權算術平均法(權數爲工人人數)
德國	Statistisches Reichsamt	,,	,,	工資合同	,,	簡單算術平均法
,, ,,	,, ,,	,,	實際所得(礦工)	僱主工資簿	以職業分	,,
澳洲	Common wealth Bureau of Census and Statistics	全洲	工資率	工團報告	先以工業分再以職務分	,,
奥國	Bundesamt Fuir Sratistik	全國	,,	,,	,,	,,
加拿大	Department of Labour	少數區域	,,	,,	,,	,,
丹麥	Department of Statistics	全國	每時實際所得	僱主之公所	以職務分	,,
美國	Bureau of Labor Statistics	,,	工資率	派人赴工廠調查	以工業分	加權算術平均法
,, ,,	,, ,,	,,	實際所得	,,	以職務分	,,
意大利	Ministry of Labour and of Social statistics	少數區域	工資率	失業及意外保險公司	以工業分	簡單算術平均法
英國	Ministry of Labour	,,	,,	工團報告	,,	,,
,, ,,	,, ,,	,,	實際所得	僱主報告	以職務分	,,

查之最初期，不可選爲基期，以初期調查，常難準確。據國際勞工局工資統計會議之意見，關於測量一般工人經濟生活狀況之工資指數，計算指數，各國似有採用同一基期之必要，并以種種原因，以大戰後年度爲佳。

Ⅴ 眞實工資指數

上述之工資指數限於貨幣形式之工資，茲進而討論眞實工資。貨幣之機能，不過是商品交換之媒介，他自身不能滿足人類最終之慾望，他不過是用以作媒介物去換取衣食燃料……等生活資料，生活資料之價值，由貨幣之形式表示，便是生活費。社會是動的，一切事物皆隨社會之進化，時刻地在變動。工人勞動力之價值，是在變動，各項物品以及金銀之價值，同時也在變動。此二項價值變動之結果，勢必使工資和物價時起漲落，從另一方面說，即貨幣之購買力時漲時落。而工資和物價漲落之方向及程度，又往往不能一致，物價升高時，也許工資低落。工資升高一倍，也許物價已漲二倍。所以要明瞭工人之眞實經濟狀況，非計算眞實工資不可。所謂眞實工資，直捷說，便是以工人之生活費爲標準與之比較，而求得之工資。其最大效用，在將工人勞動所得之眞實代價，從迷離彷彿之貨幣數字中，提示出來，使吾人得以看淸勞工羣衆之經濟生活之實況。再者，統計重乎比較，在同時同地之場合以比較工資，貨幣

工資，尚可敷衍，設異時同地，或異地同時之各貨幣工資相比較，因貨幣購買力及經濟狀況，各不相同，簡直毫無意義，我們非用眞實工資指數不可。

眞實工資，旣以工人生活費爲標準，故計算眞實工資指數之先，必編工人生活費指數。工人眞實工資指數之計算，亦復簡易，卽以工人生活費指數除貨幣工資指數，再以 100 乘之卽得。公式爲

$$\text{眞實工資指數}=\frac{\text{貨幣工資指數}}{\text{生活費指數}}\times 100$$

普通，工資指數與生活費指數，須同一基期，計算才方便；設貨幣工資指數：

1920—100

1930—120

生活費指數：

1920—100

1930—150

則眞實工資指數：

1920—100

1930—80

IV　國際間工資指數之比較

因爲各國工人生活程度之不同，各國貨幣購買力之相異，

而國際匯兌行市，又漲落無定，所以工資指數在國際間比較時，困難叢生，至今尙無澈底妥善之解決。現在將三種通用方法，敍述於下。

I. 先搜集各國固定工人之工資及生活費之材料，乃擇定一國爲標準，以其他各國之工資及生活費合成該國貨幣，分編指數，以相比較。例如民國 14 年甲乙二國工資及生活費指數爲

	甲國	乙國
工資指數	100	80
生活費指數	100	120

設以甲國爲標準，則二國之眞實工資指數，當爲

甲國	乙國
100	$\frac{80}{120}\times 100=66.6$

II. 先求得各國生活費內所包含之物品，再求各國工人在各該國應作若干時間工作，方能購買生活費內所包含之物品，設甲國工人每人每月須十元方能維持生活，而乙國則須二十元。吾人再調查二國工人須工作若干時，始能各得此數。設調查結果，甲國工人須工作 20 小時，方得十元，乙國工人須工作 15小時，方得二十元，則甲乙二國眞實工資之比，爲：

$$15:20=100^{(乙國)}:75^{(甲國)}$$

III. 此法須先得 4 種材料：

a. 各國某種工人之工資，

b. 各國工人生活費指數(同一基期)，

c. 在基期時各國國內同單位貨幣購買力之關係

d. 在基期時各國國外匯兌之行市。

以此材料，按步計算眞實工資，以作比較，舉例如下。

1. 民國 14 年，某種工人之工資，在中國每時爲銀元 4 角，在美國每小時美金 8 角

2. 同年中國此種工人之生活費指數爲 156，美國爲 165 (基期同爲民國二年)。

3. 民國二年(基期)中美二國生活費之比率爲 100 : 70。

4. 民國二年(基期)中美匯兌率，中國銀幣一元，合美金45元。

如是，則先求民國十四年美金 8 角，於民國二年時值若干，即

$$\frac{80\times100}{165}=.48$$

台得美金，48元，照民二匯價，應合中國銀幣 1.07 元。計算式，爲

$$\frac{.48}{.45}=1.07$$

但民二中美二國生活費比率爲 100 : 70，卽美國比中國低 30%。故民二在美國銀幣 1.07 元等於在中國 1.52 元。計算式爲

$$\frac{1.07}{70}\times 100=1.52$$

由此可知民二工人1.52元之貨幣購買力，可抵民14年2.37元（$1.52\times\frac{156}{100}=2.37$元）。

而中美二國民十四年實際工資之比，卽爲銀幣4與2.37之比，或$\frac{40}{159}\times 100:1.52$ 卽 100:592。

四 社會工資指數

社會工資這名詞，似乎耳生得很，其理論起源於社會主義者鼻祖馬克斯。馬氏之 Lohnarbeit und Kapital 一書中，所論相對工資(Relative Wages)，卽此處所謂社會工資。至於實際作統計之研究，則始自美國勞動總同盟。(註)

社會工資指數卽工人之眞實工資，和勞動生產力作比較之比數。所以社會工資如果逐年減低，則不但足以證明工人之經濟生活，相對地日趨貧困，而且足以證明資本階級與勞動階級的貧富懸殊，日形顯著，資本主義社會之危機，日現危迫。指數不消說，是最能指示吾人以社會現象變動之一般趨勢的工具。

至於社會工資指數之計算方法，亦極簡單，卽以生產力指

(註) 該總同盟所編之 Wages and Labors' Share, Wages in Manufacturing Industries *1899—1927* 等書均有社會工資統計刊布

數 (Index of Productivity) 除眞實工資指數，再乘以 100 卽得。公式：

$$社會工資指數=\frac{眞實工資指數}{生產力指數}\times 100$$

假使 1920 到 1930 年工人之生產力增加 33.3%，卽每小時工人能多生產 33.3倍。例如礦工於 1920 年每小時平均產煤3噸，1930 年平均產煤 4 噸，則生產力指數，爲

1920—100

1930—133.3

如二年之眞實工資指數爲

1920—100

1930—80

則二年之社會工資指數當爲

1920—100

1930—60

IX　我國之工資指數

1. 廣州市工資指數　前由廣東農工廳調製，係按年編製，結果見該廳出版之統計彙刊第三期工資指數號。材料分類爲飲食，茶煙，藥材，織染，衣着，建築，木料，金屬，玻璃，運輸，機器店員，皮業，文具，雕琢，衞生，娛樂用品，雜類，交通等20類。

材料之來源，由各工會調查所得之報告，而用其普通平均工資率。計算公式，係用簡單幾何平均法，每類各編一分指數，并綜合編一總指數，基期為1913。

2. 上海特別市工資指數(民國18年)　由該市社會局調製，民國十七年起，已着手調查，其內容要點如下。

材料分類　係參酌市內工業之情形，選定重要工業三十餘種。所謂重要工業之標準，係指一業中工廠規模較大，和工人數多者而言，一業工人數如超過1500人者，即可入選。所選入之廠家，以工人人數在三十人以上者為合格。工業分類問題原係按調查所得之三十種工業的業務性質，工資等級或工作情形分為五門，後來決定採用國際勞工局的業務分類法改為十一門，但三十業仍舊即：繅絲，棉紡，絲織，針織，毛織，造紙，皂燭，火柴，油漆，製革，玻璃，搪瓷，化裝品，漂染，機器，電機，翻砂，造船，水泥，磚瓦，鋸木，麵粉，榨油，製蛋，調味罐頭，冷飲食品，煙草，自來水，電氣，印刷。

至其職務之分類，係依製造程序為原則。

徵集材料　據該市17年5月調查，全市各業工廠凡1504家，其中選取能代表該業各廠普通規模情形者若干家，為標準工廠，約定按月詳細填報所發工資調查表格，并隨時派員，至各廠指導填寫，以期準確詳盡。至一業標準工廠工人數之和，以能達該業工人數三分之一為度，各廠所填表格寄回之後，分

業登記，更將各廠男工女工及童工所得工資獎金，分紅等項，分別計算，各自編製指數。

公式　指數公式，原擬採用加權總合法，但因全市工人調查和估計，迄未舉辦，只有十八年八月三十業工廠及工人調查和估計，較為可靠。故工資指數，即以該調查結果之工人數作權數。公式為

$$\frac{\Sigma W\frac{p_1}{p_0}}{\Sigma W}$$

基期　原擬以一九三〇年為基期，後以該年我國經濟狀況有變動，改用一九二九年。

第四編

時間數列之分析

(The Analysis of Time Series)

第四編

時間數列之分析

第一章

時間數列之性質

時間數列之研究亦是統計學的一部分，尤其是在經濟統計中，這部分很占重要，所以特闢數章來討論。

時間數列之初步整理　時間數列的資料，初步整理工作，不很需要，比較次數數列，簡便很多。在時間數列，無論資料是原始的，或次級的，都即刻可供分析之用。不過我們在選擇時，亦得謹愼將事。

數字所指示的日期，必須明白易曉，確實規定。按月的資料，或爲每月初之數字，（如美國Bradstreet 之批發物價指數）或爲每月平均數（如美國勞工統計局之物價指數），或爲每月總

數，有時或爲每月之累積數。假如是平均數，不論是每月的或每年的，所用的平均方法是那一種，如何算得，應當明瞭。再者，在任何一時間數列中，各時期之資料，尤須絕對有比較之可能。分析性質不同之一數列，往往足以引起誤解，而使結果毫無價值。

時間數列之圖解　平常在分析時間數列時，第一步手續就是將資料作圖。一經圖解，則其升降趨勢，特質所在，就很明白地顯示出來。所作圖解，或用普通算術量表，或用單對數紙 Semi-logarithmic Paper，更有少數繪於雙對數紙上者，全視資料之性質和研究之目的，而爲採擇之標準。如吾人之志趣在變動之絕對數，或幾組數列之絕對比較，可採用算術量表圖。設吾人之志趣，在變動之百分數，或比較其相對值，則單對數量表之用，似乎較勝。總之，如習於後者之應用，對於許多經濟資料之分析，單對數圖很可以提倡。有時圖如製就，則分析之工作，即算完成，長期趨向，季節變動，和短期變動，從圖上已可知其梗概，不精密之比較，亦可於圖上行之。不過我們須注意，所有觀察之結果，缺乏數量之表示，而其比較亦至概況而已。故適當精密之方法，還須從長討論。

時間數列之因子　一時間數列之變動，决非單純的，乃許多變動之總結果。Edmund E. Day 在他的 Statistical Analysis 書中，規定時間數列之變動，可根據(a)變動所佔之週期（The length of period in which the movement completes its course) (b)變動特具之形勢 (Characteristic Configuration of the movement)，及(c)變動所表現之規律(The Regularity with which the movement appears)各點，而將變動分爲五類：

1. 長期趨向(Euolutionary Movements or Secular trends)。
2. 週期變動 (Periodic Movements)
3. 循環變動 (Cyclical Movements)
4. 特異變動 (Episodic Movements due to specific causes)。
5. 偶然變動 (Fortuitous or accidental movements) 。

凡此種種變動，當須詳加考察，玆先將各種變動之性質，略加敍述，至於分析方法，詳於後列各章。

表71 1790—1920年美國之人口

清查年度	人 口
1790	3,929,214
1800	5,308,483
1810	7,239,881
1820	9,638,453
1830	12,866,020
1840	17,069,453
1850	23,191,876
1860	31,443,321
1870	38,558,371
1880	50,155,783
1890	62,947,714
1900	75,994,575
1910	91,972,266
1920	105,710,620

1.長期趨向　是指長時期繼續的一種趨向，或爲向上或爲向下。以表示增加或減少。其形式可假定種種，或爲直線型，或作其他簡單的，或複雜的曲線型。試舉一實例於表71,圖39。

圖39　1790—1920年美國之人口

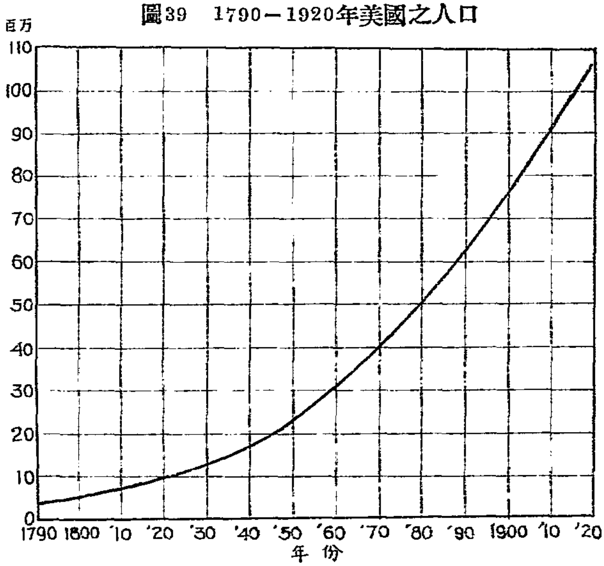
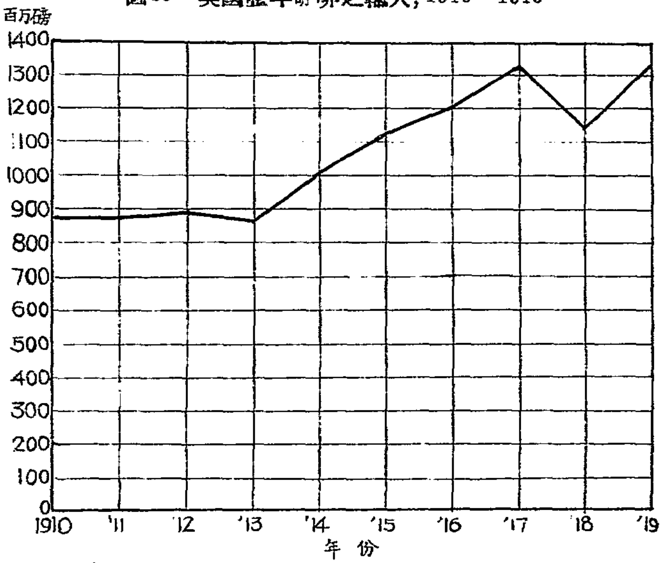


表72　美國歷年咖啡之輸入

年　分	咖啡(單位1,000,000)
1910	871
1911	875
1912	885
1913	863
1914	1002
1915	1119
1916	1201
1917	1320
1918	1144
1919	1334

圖40　美國歷年咖啡之輸入, 1910—1919



2. 週期變動　爲另一之經濟現象，乃一種週而復始的變動。他的種類很多，而以季節變動 (Seasonal Fluctuation) 爲最著。所謂季節變動者，依隨季節月令而變動，有若干月在常態之上，有若干月在常態以下，其週而復始所包括之時期，却爲十二月(一年)。圖41所示爲美國 Public Service Electric Company of New Jersey, 1910—1923 各年電燈用戶用電統計，此例中呈極顯著之季節變動。季節變動並非是唯一之週期變動，其他週期較短之變動，其週而復始之時間，有爲一星期，有爲一日者；前者之例如工廠工人一星期內各日工作效率之變動

，後者之例，如表73

圖41 公共電氣公司電燈用戶用電基羅瓦特時數，1910—1923

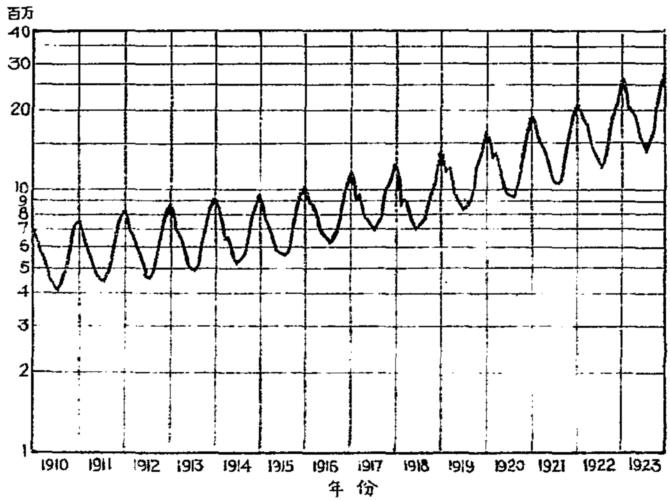


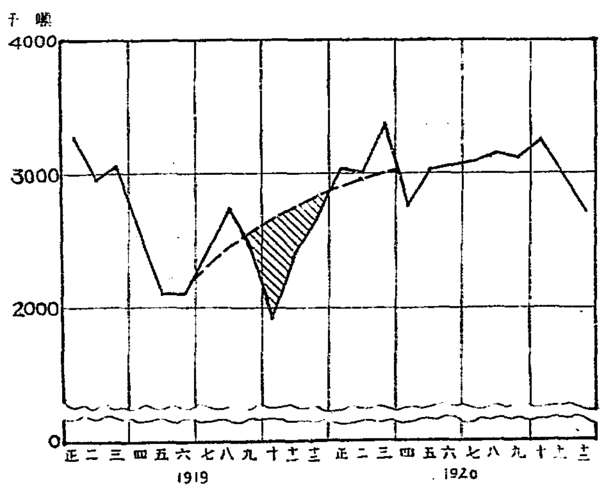
表73 紐約市某街行車之變動，上午7時至下午7時

時間	車輛數	
	北行	南行
上午 7—8	140	345
8—9	425	1105
9—10	785	1415
10—11	860	1120
11—12	995	1025
下午 12—1	960	1510
1—2	910	960
2—3	815	1055
3—4	710	950
4—5	840	835
5—6	1175	740
6—7	1305	710

3 循環變動　循環變動有些像週期變動，因為他亦是週而復始，循環往復的變動，但各循環所佔時間，長短不一致，未能如週期變動之固定。在幾個重要農業國家五穀生產變動的循環，大約八年或十一年。

4 特異變動　特異變動為發生特種意外事件而起之變動，如受地震，戰爭，水火災害，同盟大罷工等，其結果足以影響經濟現象，使發生短時間之劇變，因其性質，事事不同，影響亦至不一律，須逐事逐項為個別之研究，殊難用任何數學函數以表示之，故本書暫置不論。至於普通欲得近似之結果，可用隨手修勻曲線法(Free-hand Smoothing of the curve)，如圖42。

圖42　美國一九一九年罷工影響於每月生鐵之生產(1919—1920)



5 偶然變動　偶然變動如不規則干擾 (Distubance) 之結果，大多數之時間數列中，皆有此種變動之存在，不過因爲他們是潛伏在他種變動之中，不甚顯著耳。

各種變動，極少有單獨行動者，往往二種或三四種之變動，混合一起，潛伏於一數列。試以表74圖43，1913—1919年美國每月生鐵產量之統計作例，一加觀察。其曲線爲向上之形勢，已確切不移，易言之，卽有肯定之長期趨向。且也，曲線每至冬令則增高，春季則下降，每年一升一降，表示其有季節變動之存在。又圖中曲線，繼續上升若干年，達至最高點，便繼續下降，直至達最低點，又復繼續上升，如是循環往復之趨勢，可認爲受一般商情影響之循環變動。在1918年之初及1919年之末，無疑地有特異變動——由於1918年冬之鐵路運輸擁擠 Railroad congestion，及1919年秋之鋼鐵罷工。意外變動，亦有存在，惟起伏甚微，不易觀察耳。就此數列全體而論，可爲此種種變動混合之總結果。

時間數列之統計分析工作，可分爲二點（1）各種變動務須剔除 (Isolation)，卽使其分解，各自分立，於是各各對總結果之影響，得以決定，及（2）各單獨變動之特質 (Character) 須精密推求之。

下列各章首先討論長期趨向。次及季節變動，循環變動，循環變動之相關，最末章殿以預測，作本編之結束。

表74 美國每月生鐵之出產，1903—1919年(單位：1000噸)

	1903	1904	1905	1906	1907	1908	1909	1910	1911	1912	1913	1914	1915	1916	1917	1918	1919
正 月	1472	921	1781	2068	2205	1045	1801	2608	1759	2057	2795	1885	1601	3185	3151	2412	3302
二 月	1390	1205	1597	1904	2045	1077	1703	2397	1794	2100	2585	1888	1675	3087	2645	2319	2940
三 月	1590	1447	1936	2155	2226	1228	1832	2617	2188	2405	2763	2348	2064	3338	3251	3213	3090
四 月	1608	1555	1922	2073	2216	1149	1758	2483	2065	2375	2752	2270	2116	3228	3335	3288	2478
五 月	1713	1534	1963	2098	2295	1165	1880	2390	1893	2512	2822	2093	2263	3351	3417	3446	2108
六 月	1673	1292	1793	1976	2234	1092	1929	2365	1787	2440	2628	1918	2381	3212	3270	3324	2115
七 月	1540	1106	1741	2013	2255	1218	2101	2148	1793	2410	2560	1958	2563	3220	3342	3421	2429
八 月	1571	1167	1843	1926	2250	1848	2246	2106	1926	2512	2543	1995	2780	3204	3248	3390	2743
九 月	1553	1352	1899	1960	2183	1418	2385	2056	1977	2463	2505	1883	2858	3202	3184	3418	2488
十 月	1425	1450	2033	2196	2336	1563	2600	2093	2102	2689	2546	1778	3125	3509	3303	3487	1864
十一月	1039	1486	2014	2187	1828	1577	2547	1909	1999	2630	2233	1518	3087	3312	3206	3354	2392
十二月	846	1616	2045	2235	1234	1740	2635	1777	2043	2782	1983	1516	3203	3171	2883	3434	2633

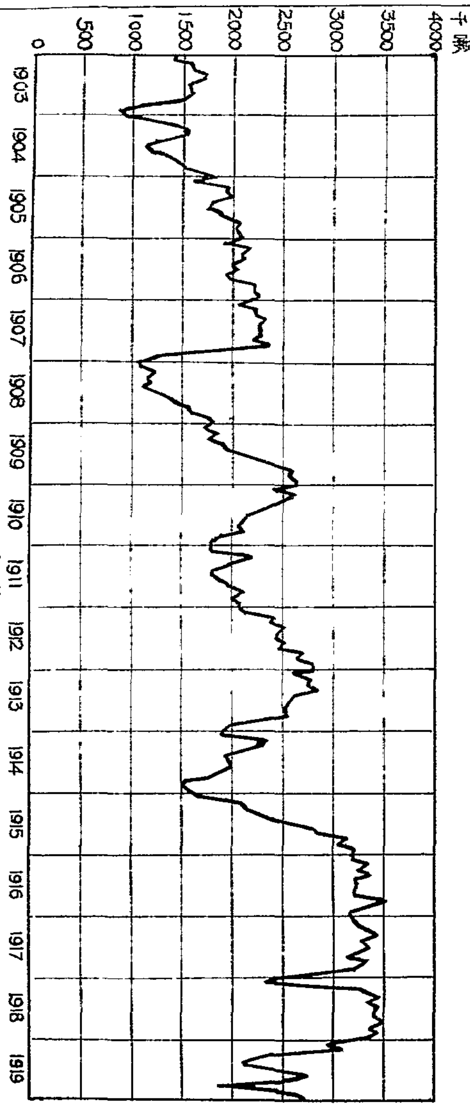


圖43 美國每月生鐵之出產,1903—1919,

第二章

長期趨向 Secular Trend

長期趨向，簡稱恆差，是一種向上或向下的恆常繼續的趨向。其逐年之增減，或爲百分率，或爲絕對數，要皆歷久不甚大變，繼續地經過很長時期。分析時間數列時，一時間之恆差值。常擬爲該時間之標準值 (Normal Value)。

研究之時，第一步，即將原有數列全部繪之於圖。此種圖解可不必太詳細，取每年平均數或總數即可，至多亦不得短於每月之數。因爲時間太短，徒使趨向之觀察，混淆不清。製圖之作用，所以示長期趨向之是否存在，如已證明長期趨向確有存在，則第二步當決定其變動之形態如何，須愈精密愈好，決定之方法，有四種。

1. 隨手配合法(Free hand Fitting)，
2. 移動平均法 (Moving Average)，
3. 數學方程式配合法 Representation of Secular Trend by Mathmatical Curves)
4. 平均數法(The method of Averages)(附半數平均法)。

I. 隨手配合法

隨手配合法，乃憑作者之經驗，將資料繪於相當圖上，審度形勢，作一直線或修勻之曲線，以代表長期趨向線，作線之時，或信手作成，或藉助曲線板均可。線上各點與線下各點，對此恆差線之離差，務使約略相等。此法之利在乎簡捷，不過無數字之表示，且各人所得結果，各執其一，勢必各異，未免遺憾。

Ⅱ. 移動平均數法

移動平均法是採取繼續平均的方法，將他種變動的因素除去，單留長期趨向；結果所得之各平均數，即代表此長期趨向。至於平均方法，普通都採用算術平均數。求移動平均數，當先決定採用若干年的平均數。設求五年的平均，第一先將數列之前五年各項相加，再以 5 除之，得一平均數，置於第三項相對之處。再從第二項至第六項相加，以 5 除之，得平均數，置於第四項相對之處。以下同樣計算，直至最後一年為止。如表 75 中第 5 欄，係五年移動平均數，與 1903 年相對之第一項係從$\frac{1282+1435+1452+1344+1882}{5}=1479$得來。與1904年相對之第二項；係從$\frac{1435+1452+1344+1882+2066}{5}=1636$得來。其餘 3 年 7 年之活動平均數俱可如法泡製，茲不多贅。設移平動均數所取年數，如為偶數，例如表75之第3欄，即為四年移動平均數。計算之時，稍稍繁複，先將四項相加，除以 4 ，所

得平均數，置於中央二項之間，如此求出之各平均數（如第3欄）須再取二年之移動平均數始可。因第一次各平均數之地位，介於中央二項之間，設欲以之與原有材料繪圖比較，不便殊甚；且欲以原有數字減各平均數，亦勢所難能，因為二者的時間，已前後相差半年，必須再取二年之移動平均數，始能恢復其地位（第4欄）。算移動平均數，手續太繁，可求助於計算機，或採用簡法，將每次去掉之一項，與加入之一項，分別從前一個總和上減去及加上，再求其平均數，手續較簡。例如上例之第一個五年總和為7395，迨求第二個五年總和時，我們當加入第六年，減去第一年，加入者為2066，減去者為1282，，在第一次總和上，分別加減，得8179，是為第二次五年之總

表75　移動平均數之計算

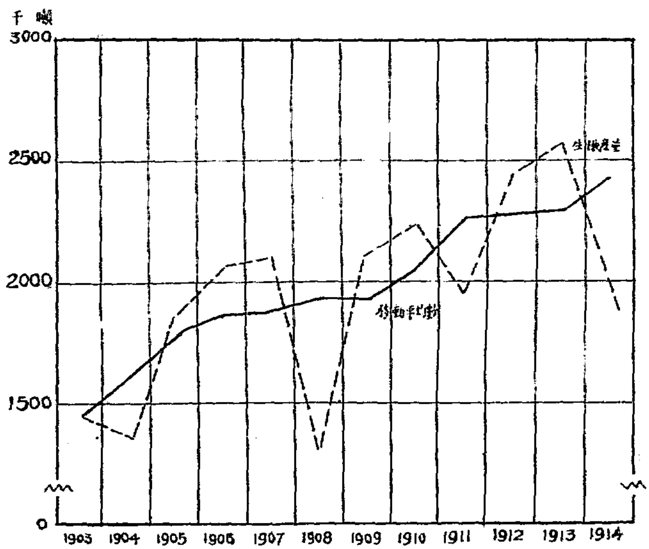
年份	生鐵產額　單位：1000 噸				
	每月平均產量 (1)	三年移動平均 (2)	四年移動平均 (3)	修正四年移動平均(4)	五年移動平均 (5)
1901	1282				
1902	1435	1390	1378		
1903	1452	1410	1528	1453	1479
1904	1344	1559	1686	1607	1636
1905	1882	1764	1850	1768	1771
1906	2066	2019	1840	1845	1741
1907	2109	1826	1898	1869	1895
1908	1302	1842	1941	1920	1966
1909	2116	1885	1900	1920	1942
1910	2237	2099	2186	2043	2009
1911	1944	2210	2297	2242	2261
1912	2448	2317	2218	2257	2222
1913	2860	2310	2359	2284	2269
1914	1921	2318	2551	2451	1531
1915	2472	2215			
1916	3252				

和，再以 5 除之，即得其平均數1636。如是繼續加減，直至最後之一總和爲止。

移動平均數之計算方法既明，當進而討論如何應用移動平均法以決定長期趨向。欲解決此問題，當先從圖上研究原有材料之波長(Wave Length)的性質。波長乃循環一起一伏所需之時間，即所謂循環週期。波長決定之後，始可決定用多長之時間，以求移動平均數。假如一數列所有循環起伏之時間和變動之

圖　美國生鐵之每月平均產量及其四年之

移動平均數 (1903—1914)



廣度(Amplitude)多相同，則所得結果，繪成一線，必爲直線之形式。惟事實上週期之長短和變動之廣度，至不一律，故須找出週期的約略平均時期，以之求移動平均數，平常總取較平均週期稍長之時期；其結果，起伏雖不能完全免除，而起伏之程度，總可小些。

移動平均數之特質　用移動平均數來求傾差，很有伸縮性。故一數列中某一時期雖情形有所改變，亦可適用。不過他的短處，亦有四點：

1.易受極端項的影響。

2.因循環起伏之週期，至不規律，欲決定相當時期之移動平均數，亦頗不易。

3.算出之移動平均數較原來資料，首尾二端少去幾年，要知最近年分之資料，非常重要。此種缺點，雖有變通方法，足資補救，或用圖解外延(Graphic Extxapolation)至極端項，或將最先及最後之一項分別重加。例如表 75 最末一項3252,繼續重加二次，亦可得出極端項之數；不過這些方法，亦不能認爲滿意。

4.假使我們的目的，在計算二數列或幾組數列之相關，因爲對移動平均數之離差總和，其值不爲0,所以需要多一步改正手續。此點，頗與處理次數分配中之假設平均數相似。再者，除非有充分的理由，認爲傾差是成直線之形態，用移動平均數時，還當謹慎從事，注意不純相關Spurious Correlation 之混入。

設一恆差線實際上成爲拋物線之形式，用移動平均法不但將資料修勻 Smoothed ，而且有改變趨向線形式之可能。

累進移動平均數 (Ppogressive mean) 簡單移動平均數，有時稍加修改，將各項加以權數。其權數中央之項最大，前後各項依次遞減。權數係依二項展開式之係數以定之，例如上述之五年平均數，其權數應爲 1,4,6,4,1。此種加權移動平均數，通常稱之爲累進平均數。此法在分析中，有時亦有用之者，惟在時間數列中，用以和緩意外變動，或定期變動，非適當之方法。

III. 數學方程式配合法

有時資料之恆差，其升降依一定之比率的或絕對值的方式，頗與數學方程式相脗合，故可利用數學方程式，以分析解釋此種資料之恆差。數學方程式求恆差，可分二點來討論。第一，就是選擇相當的線來代表恆差。這就是說，還是直線，還是曲線；如爲曲線，那末，是那種形態，複利率線，拋物線，還是其他形態的線；關於這點，尙無確定之標準，足資根據，全憑統計家審視圖線(指原有資料)，參以經驗來判斷。第二點是選擇一適當之時期，就此時期，以便配合特定之線。欲解決這問題，當注意下列各點。

1,關於時間一層，平常總以愈長久愈妙，至少亦須包括一

個循環一起一伏所需之時間，有時以十年爲最低限度者。

2.所選時期始末二端之地位，須有均稱的現象 Symmetrical Phases，假如始點呈升起之勢，則終點當呈下降之勢；如始點爲一高峯 (Peak)，終點之地位，亦當大約在高峯相平之地位。

3.所選時期，須在常態情形中，意外大事發生之變態時期，不足取。

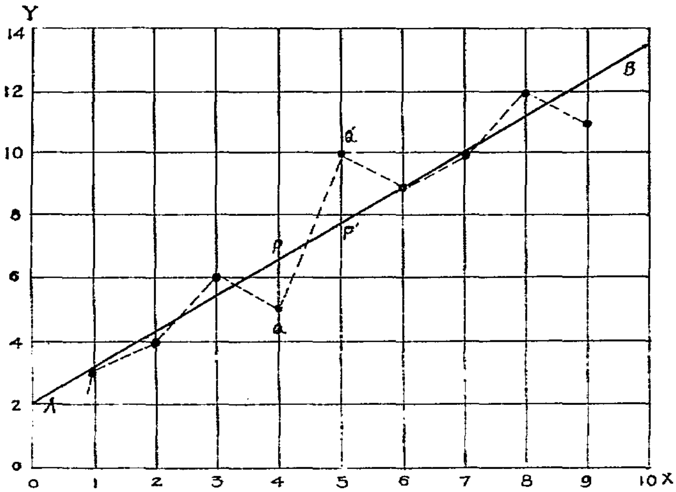
4.在此時期中之恆差，應合乎可以一單獨的趨向線來表示之假定，故趨向不同之時期，應分爲數期，個別研究之。

曲線的配合　曲線的形態，勢難盡量討論，本書就其最重要者數種，縷述於後。

I.直線　在分析時間數列時，設恆差逐年增加數(或減少數)，約略相同者，可假定爲直線方程式之形式。配合方法根據最小二乘方(Least Square)之原理。其意謂：設觀察一物多次，而得若干不同之結果：就此不同之結果中，求其最最或然之數值；所謂最最或然之數值，卽各項乘方差之和，對此爲最小之謂也。就趨向線而言，其理正同，乃謂就直線方程式而配合一線，以代表恆差，則原有資料上之各點，對此線平方差之和，就其縱距言之，當爲最小，例如下圖 PQ, P'Q'等之平方和爲最小也。直線方程式之普通公式，爲

$$y=a+bx$$

圖45　配合直線之示例



式中之y爲倚變量，x爲自變量，b爲直線之斜度，而a爲直線與y縱距之交切點。式中僅a與b爲二未知數，設a與b之值，已經求得，則直線可不勞而獲。故吾人之關鍵，在求得a與b之值。今請以上圖爲例，以說明其求得之方法。圖中九點，代表x,y九對數值。將各數值代入直線方程式$y=a+bx$，則得下列各觀察方程式(Observation equation)

$$3=a+1b$$
$$4=a+2b$$
$$4=a+3b$$
$$5=a+4b$$

$$10=a+5b$$

$$9=a+6b$$

$$10=a+7b$$

$$12=a+8b$$

$$11=a+9b$$

我們可任取二式，依解聯立方程式之法解之，即可得a與b之值。但此數值，只能適合此二方程式，對於其餘七式，未必卽能適合。所以我們的問題是將此九觀察方程式，合併而爲二標準式(Normal Equation)，然後用聯立方程式解之，可得a與b最最或然之數值矣。第一標準方程式，係將第一未知數a之係數，乘各觀察方程式之各項，然後各式相加卽得。第二標準方程式，係將第二未知數b之係數，乘各觀察方程式之各項，然後各項相加而得之。二標準方程式之形式爲

$$\Sigma(y)=Na+b\Sigma(x)$$

$$\Sigma(xy)=a\Sigma(x)+b\Sigma(x^2)\cdots\cdots\cdots\cdots\cdots(1)$$

以相當各値代入，解之，卽得a與b之數值矣，試舉例於表76。設x之各數值，爲連續數(Consecutive Number)，可取中央一項爲原點，則Σx爲0，而上述二標準方程式變爲

$$\Sigma y=Na$$

$$\Sigma xy=b\Sigma(x^2)\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots(2)$$

移項，得

$$a=\frac{\Sigma y}{N}$$

$$b=\frac{\Sigma xy}{\Sigma(x^2)}\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots(3)$$

從此二式以求a,b,較爲簡捷，在經濟統計中之時間數列，x變量(時間)常爲連續之數字，可應用此法。惟須注意者，應用此公式，若x變量之項數(年數)爲奇數，固以中央一項爲原點，此中央之年與前後各年之距離，以一年爲單位；若年數爲偶數，則其原點介乎中央二年之間，以半年爲單位而計算之，在決定每年增加率時，須將$\Sigma(x^2)$除以2，試用1914年以來世界白銀產量之資料將公式(1)之計算，示例於表76，公式(3)之計算，示例於77及表78。

表76　恆差直綫之計算　(公式1)

1914—1929年世界銀產量(單位百萬純盎斯)

年份	x	銀產量 y	xy	x^2	恆差坐標
1914	1	172.3	172.3	1	161.6
1915	2	173.0	346.0	4	168.4
1916	3	180.8	542.4	9	175.2
1917	4	186.1	744.4	16	182.0
1918	5	203.2	1016.0	25	188.8
1919	6	179.8	1078.8	36	195.6
1920	7	173.3	1213.1	49	202.4
1921	8	171.3	1370.4	64	209.2
1922	9	209.8	1888.2	81	216.0
1923	10	246.0	2460.0	100	222.8
1924	11	239.5	2634.5	121	229.6
1925	12	245.2	2942.4	144	236.4
1926	13	253.8	3299.4	169	243.2
1927	14	254.0	3556.0	196	250.0
1928	15	257.3	3859.5	225	256.8
1929	16	256.5	4104.0	256	263.6
	136	3401.9	31227.4	1496	

將各值代入二標準方程式，則

$$3401.9 = 16a + 136b$$

$$31227.4 = 136a + 1496b$$

解之，得

$$a = 154.836$$

$$b = 6.798$$

直線方程式，為

$$y = 154836 + 6798x$$

表77　配合直綫恆差之計算（公式 3 ；x 為偶數）

年份	x	銀產量 y	xy	x^2	恆差坐標
1914	−15	172.3	−2584.5	225	161.6
1915	−13	173.0	−2249.0	169	168.4
1916	−11	180.8	−1988.8	121	175.2
1917	− 9	186.1	−1674.9	81	182.0
1918	− 7	203.2	−1422.4	49	188.8
1919	− 5	179.8	− 899.0	25	195.6
1920	− 3	173.3	− 519.9	9	202.4
1921	− 1	171.3	− 171.3	1	209.2
1922	1	209.8	209.8	1	216.0
1923	3	246.0	738.0	9	222.8
1624	5	239 5	1197.5	25	229.6
1925	7	245.2	1716.4	49	236.4
1926	9	253.8	2284.2	81	243.2
1927	11	254.0	2794.0	121	250.0
1928	13	257.3	3344.9	169	256.8
1929	15	256 5	3847.5	225	263.6
		3401.9		1360	

$$b = \frac{\Sigma xy}{\frac{\Sigma x^2}{2}} = \frac{46225}{\frac{1360}{2}} = 6.798$$

$$a = \frac{\Sigma y}{N} = \frac{3401.9}{16} = 212.619$$

表78　配合直線恆差之計算（公式3；x 爲奇數）

年份	x	銀產量 y	xy	x^2	恆差坐標
1914	−7	172.3	−1206.1	49	164.6
1915	−6	173.0	−1038.0	36	171.1
1916	−5	180.8	−904.0	25	177.5
1917	−4	186.1	−744.4	16	183.9
1918	−3	203.2	−609.6	9	190.4
1919	−2	179.8	−359.6	4	196.8
1920	−1	173.3	−173.3	1	203.3
1921	0	171.3		0	209.7
1922	1	209.8	209.8	1	216.1
1923	2	246.0	492.0	4	222.6
1924	3	239.5	718.5	9	229.0
1925	4	245.2	980.8	16	235.5
1926	5	253.8	1269.0	25	241.9
1927	6	254.0	1524.0	36	248.3
1928	7	257.3	1801.1	49	254.8
		3145.4	1960.2	280	

$$b=\frac{\Sigma xy}{\Sigma x^2}=\frac{1960.2}{208}=6.44$$

$$a=\frac{\Sigma y}{n}=\frac{3145.4}{15}=209.9$$

上項計算方法尙覺煩瑣，茲特再行介紹較捷之方法二種，一爲折半求差法，一爲累積總合法，依次說明之。

1. 折半求差法　就選定之資料，將其 y 軸按時期中點爲標準，分爲兩部，成爲 y_1 及 y_2，然後計算 y_1 與 y_2 各項之差，再與相當時間單位差(x)相乘而後總和之，代入公式以求常數 a,b，。原有求常數 a 之公式，$a=\frac{\Sigma y}{N}$，則改變爲

$$a=\frac{\Sigma y_1+\Sigma y_2}{N}$$

至於求 b 之公式，可分求 Σxy 及 Σx^2 二部，而因時期(年分)爲偶數或奇數，亦有不同。設時期爲奇數，$t=\frac{n-1}{2}$，則

$$\Sigma x^2=\frac{n(n-1)(n+1)}{12}$$

$$\Sigma xy=t(y_n-y_1)+(t-1)(y_{n-1}-y_2)+(t-2)$$
$$(y_{n-2}-y_3)+\cdots+2(y_{t+3}-y_{t-1})+(y_{t-2}-y_t).$$

設時期爲偶數，$t=\frac{n}{2}$，則

$$\Sigma x^2=\frac{n(n-1)(n+1)}{6}$$

$$\Sigma xy=(2t-1)(y_n-y_1)+(2t-3)(y_{n-1}-y_2)+(2t-5)$$
$$(y_{n-2}-y_3)+\cdots\cdots+3(y_{t-2}-y_{t-1})+(y_{t+2}-y_t)$$

仍以世界銀產量之資料，舉二例於表79，表80。

表79　配合恆差直線之計算(用折中求差法；x 爲偶數)

年分	y_1	年分	y_2	y_1-y_2	x	$x(y_1-y_2)$
1929	256.5	1914	172.8	84.2	15	1263.0
1928	257.3	1915	173.0	84.3	13	1095.9
1927	254.0	1916	180.8	73.2	11	805.2
1926	253.8	1917	185.1	67.7	9	609.3
1925	245.2	1918	203.2	42.0	7	294.0
1924	239.5	1919	179.8	59.7	5	298.5
1923	246.0	1920	173.3	72.7	3	218.1
1922	209.8	1921	171.3	38.5	1	38.5
	1962.1		1439.8			4622.5

$$b=\frac{\Sigma x(y_1-y_2)}{n(n-1)(n+1)\frac{1}{6}}=\frac{4622.5}{(16\times15\times17)\frac{1}{6}}=6.797$$

$$a=\frac{\Sigma y_1+\Sigma y_2}{N}=\frac{1962.1+1439.8}{16}=212.62$$

表80　配合恆差直線之計算(用折中求差法；x 為奇數)

年　分	y_1	年　分	y_2	y_1-y_2	x	$x(y_1-y_2)$
1928	257.3	1914	172.3	85.0	7	595.0
1927	254.0	1915	173.0	81.0	6	486.0
1926	253.8	1916	180.8	73.0	5	365.0
1925	245.2	1917	186.1	69.1	4	236.4
1924	239.5	1918	203.2	36.3	3	108.9
1923	246.0	1919	179.8	66.2	2	132.4
1922	209.8	1920	173.3	36.5	1	36.5
1921	171.3	……	……	……	0	……
	1876.9		1268.5			1960.2

$$b=\frac{\Sigma x(y_1-y_2)}{n(n-1)(n+1)\frac{1}{12}}=\frac{1960.2}{(15\times14\times16)\frac{1}{12}}=6.44$$

$$a=\frac{\Sigma y_1+\Sigma y_2}{N}=\frac{1876.9+1268.5}{15}=209.69$$

2. 累積總合法　常數 b 之值，可由一次累積數之總合以求之，此法最適宜於用加法機上計算之。其法先將各時期之時序倒置，卽將最近之時期置於第一位，次近之時期，置於第二位，以下類推，直至最遠之年置於末位，總之使 x 變量成為由大而小之排列，然後將各相當 y 值排列，再依次累積之。累積之工作，可於加法機上完成之，法將 y 每加一項，求一小計(Sub-total)(表81末欄各數，卽轉錄加法機紙條上之小計)，再將此各小計連第一項總合之，卽得 ${}_1S_x$。然後代入下列公式計算之。

$$b=\frac{{}_1S_x}{n(n+1)(2n+1)\frac{1}{6}}$$

$$a=\frac{\Sigma y}{n}$$

此法便捷，勝於上述各法，且其公式，不因時期爲偶數或奇數而異，尤非他法所能及，計算方法示例於表81。

表81　配合恆差直線之計算(累積總合法)

年份	x	y	${}_0S_x'$
1929	16	256.5	256.5
1928	15	257.3	513.8
1927	14	254.0	767.8
1926	13	253.8	1021.6
1925	12	245.2	1266 8
1924	11	239.5	1506.3
1923	10	246.0	1752.3
1922	9	209.8	1962.1
1921	8	171.3	2133.4
1920	7	173.3	2306.7
1919	6	179.8	2486.5
1918	5	203.2	2689.7
1917	4	186.1	2875.8
1916	3	180 8	3056.6
1915	2	163.0	3229.6
1914	1	162.3	3401.9
		3401.9	31227.4

$$b=\frac{{}_1S_x}{\frac{N(N+1)(2N+1)}{6}}=\frac{31227.4}{1496}=6.798$$

$$a=\frac{\Sigma y}{N}=\frac{3401\ 9}{16}$$

$$=212.62$$

<u>恆差之縱距 (Ordinates of trend)</u>　配合恆差線之方程式，如已求得，則恆差之各縱距，都可由其常數之值，推算而得。

任何時期之恆差縱距，即為該時期恆差線上之y'值。該項數值，常擬為該時期之標準值(Normal Value)，其意謂除恆差外，設無他種變動時，變量之數值，必為此標準值。在直線方程式，年數如為偶數，則以中央二年之交為原點；如為奇數，則在中央一年之中點，此點之地位，平常稱為中心縱距(Central Ordinates)。表78之例，中心縱距為209.7，表77之中心縱距為212.6，將此數加以或減去每單位時間之增加率(即斜度)，即得各時期之縱距。

丁.拋物線　上述之範圍，限於直線之恆差，則直線配合之最為適當。有時恆差之形態，非屬直線，則拋物線 (Potential Type) 亦甚有用。此種曲線之普通方程式，為 $Y=a+bX+cX^2+\cdots\cdots$。此方程式固非代表拋物線，惟平常習用拋物線之一字，恆以用於 Potential Series。設其為 X 之二次方，則名之曰二次拋物線 (Second degree Parabola)。設為三次方，則名之三次拋物綫 (Third degree Parabola)。平常應用，總不外二次三次兩種。因二次卽有三未知數，三次卽有四未知數，次數愈高，未知數愈多，不獨計算繁瑣，且缺乏外延性 (Extrapolation)。二次拋物線之方程式，為 $y=a+bx+cx^2$，其算法與前述直線方程式相仿，二次拋物線，須有標準方程式三，如下：

$$\Sigma y=Na+b\Sigma x+c\Sigma(x^2)$$

$$\Sigma xy = a\Sigma(x) + b\Sigma(x^2) + c\Sigma(x^3)$$

$$\Sigma x^2y = a\Sigma(x^2) + b\Sigma(x^3) + c\Sigma(x^4)$$

表82　配合恆差線之計算

美國 1897—1921 年商業失敗統計

年份	x	y (失敗數)	xy	x^2y	x^3y
1897	−12	13,083	−156,996	1,883,552	−22,607,424
1898	−11	11,615	−127,765	1,405,415	−15,459,565
1899	−10	9,642	−96,420	964,200	−9,642,000
1900	−9	9,912	−89,208	802,872	−7,225,848
1901	−8	10,648	−85,184	681,472	−5,451,776
1902	−7	9,973	−69,811	688,677	−3,420,739
1903	−6	9,775	−58,650	351,900	−2,111,400
1904	−5	10,417	−52,085	260,425	−1,302,125
1905	−4	9,967	−39,868	159,472	−637,888
1906	−3	9,385	−28,155	84,465	−253,395
1907	−2	10,274	−20,548	41,096	−82,192
1908	−1	14,066	−14,066	14,066	−14,066
1909	0	11,872	…………	…………	…………
1910	1	11,588	11,588	11,588	11,588
1911	2	12,679	25,358	50,716	101,432
1912	3	13,832	41,496	124,488	373,464
1913	4	14,553	58,212	232,848	931,392
1914	5	16,780	83,900	419,500	2,097,500
1915	6	19,035	114,210	685,260	4,111,560
1916	7	16,498	115,486	808,402	5,658,814
1917	8	13,073	104,584	836,672	6,693,376
1918	9	9,331	83,979	755,811	6,802,299
1919	10	5,515	55,150	551,500	5,515,000
1920	11	8,463	93,093	1,024,023	11,264,253
1921	12	19,982	239,784	2,877,408	34,528,896
		301,958	188,084	15,516,228	9,881,156

$N = 25$

$\Sigma(x) = 0$

$\Sigma(y) = 301,958$

$\Sigma(xy) = 188,084$

$\Sigma(x^2y) = 15,516,228$

$\Sigma(x^2) = 1300$

$\Sigma(x^3) = 0$

$\Sigma(x^3y) = 9,881,156$

$\Sigma(x^4) = 121,420$

$\Sigma(x^5) = 0$

$\Sigma(x^6) = 13,471,900$

設 x 為連續數，計算時可取中央一項為原點，則 $\Sigma(x)$，$\Sigma(x^3)$ 為零，標準方程式如下：

$$\Sigma y = Na + c\Sigma(x^2)$$

$$\Sigma xy = b\Sigma(x^2)$$

$$\Sigma x^2 y = a\Sigma(x^2) + c\Sigma(x^4)$$

設以表82之事實代入，即得下列方程式。

$$301{,}958 = 25a + 1300c$$

$$188{,}084 = 1300b$$

$$15{,}516{,}228 = 1300a + 121{,}420c$$

解之，得各常數之值。

$$a = +12{,}258$$

$$b = +144.7$$

$$c = -3.45$$

代入，即得方程式

$$y = 12{,}258 + 144.7x - 3.45x^2$$

三次拋物線之方程式，為 $y = a + bx + cx^2 + dx^3$ 含有四未知數，而標準式亦有四式。

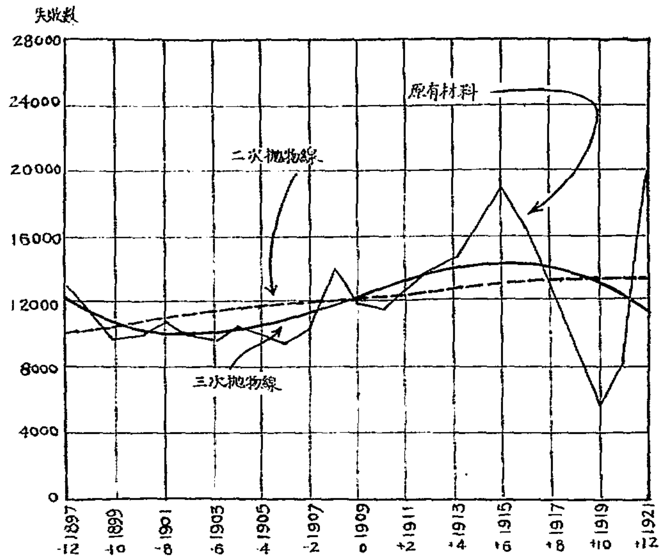
$$\Sigma(y) = Na + b\Sigma(x) + c\Sigma(x^2) + d\Sigma(x^3)$$

$$\Sigma(xy) = a\Sigma(x) + b\Sigma(x^2) + c\Sigma(x^3) + d\Sigma(x^4)$$

$$\Sigma(x^2Y) = a\Sigma(x^2) + b\Sigma(x^3) + c\Sigma(x^4) + d\Sigma(x^5)$$

$$\Sigma(x^3y) = a\Sigma(x^3) + b\Sigma(x^4) + c\Sigma(x^5) + d\Sigma(x^6)$$

圖 46　美國1897—1921年商業失敗數及其恆差線



設 x 爲連續數，可簡縮爲

$$\Sigma(y)=\mathrm{N}a+c\Sigma(x^2)$$

$$\Sigma(xy)=b\Sigma(x^2)+d\Sigma(x^4)$$

$$\Sigma(x^2y)=a\Sigma(x^2)+c\Sigma(x^4)$$

$$\Sigma(x^3y)=b\Sigma(x^4)+d\Sigma(x^6)$$

試以表 82 之事實爲根據，將相當各值代入，卽得下列各方程式。

$$301,958=25a+1,300c$$

$$188,084=1,300b+121,420d$$

$$15,516,228=1,300a+121,420c$$

$$9,881\ 156=121,420b+13,471,900d$$

解之，得各常數之值。

$$a=+12,258$$

$$b=+482$$

$$c=-3.45$$

$$d=-3.61$$

將各值代入，得代表恆差之方程式如下

$$y=12,258+482(x)-3.45(x^2)-3.61(x^3)$$

對數配合曲線　在分析時間數列時，單對數曲線 (Curves of a Semi-logarithmic type)，亦很有用。其優點在表示數量間的比率之眞象。此種曲線，在經濟資料之分析，尤著重要。各線之方程式，仍照前述之形式，惟此時所有 y，已變爲 $\log y$。故直線方程式爲 $\log y=a+bx$；拋物線之普遍公式爲 $\log y=a+bx+cx^2+dx^3+\cdots\cdots$。此種曲線，或繪於算術量表圖上，則 x 爲自然數，y 爲對數；或繪於單對數紙上，y 及 x 皆爲自然數。採用後者之圖解，較爲簡便。試以美國1908－1922 年煤油出產之資料爲例，用 $\log y=a+bx+cx^3$ 之方程式配合恆差線，計算手續，參表83。

表83　美國煤油出產，1908—1922

年份	x	y 產額 (單位一百萬桶)	$\log y$	$x\cdot\log y$	$x^2\cdot\log y$
1908	−7	178.5	2.25164	−15.76148	110.33036
1909	−6	183.2	2.26269	−13.57614	81.45384
1910	−5	209.6	2.32139	−11.60695	58.03475
1911	−4	220.4	2.34321	−9.37284	37.49136
1912	−3	222.9	2.34811	−7.04433	21.13299
1913	−2	248.4	2.39515	−4.79030	9.58060
1914	−1	265.8	2.42456	−2.42456	2.42456
1915	0	281.1	2.44886	....	....
1916	1	300.8	2.47828	2.47828	2.47828
1917	2	335.3	2.52543	5.05086	10.10172
1918	3	355.9	2.55133	7.65399	22.96197
1919	4	378.4	2.57795	10.31180	41.24720
1920	5	442.9	2.64631	13.23155	66.15775
1921	6	472.2	2.67413	16.04478	96.26868
1922	7	557.5	2.74624	19.22368	134.56576
總數			36.99528	9.41834	694.23282

$$N=15 \qquad \Sigma(x)=0$$

$$\Sigma(\log y)=36.99528 \qquad \Sigma(x^2)=280$$

$$\Sigma(x\cdot\log y)=9.41834 \qquad \Sigma(x^3)=0$$

$$\Sigma(x^2\cdot\log y)=694.23282 \qquad \Sigma(x^4)=9352$$

三標準式，

$$\Sigma(\log y)=\mathrm{N}a+b\Sigma x+c\Sigma x^2$$

$$\Sigma(x\cdot\log y)=a\Sigma x+b\Sigma x^2+c\Sigma x^3$$

$$\Sigma(x^2\log y)=a\Sigma x^2+b\Sigma x^3+c\Sigma x^4$$

將各相當值代入，得

$$36.99528=15a+280c$$

$$9.41834=280b$$

$$694.23282 = 280a + 9352C$$

解之，得三未知數之值爲

$$a = +2.450508$$

$$b = +.033637$$

$$c = +.0008488$$

故配合之恆差方程式，應爲

$$\log x = 2.450508 + .033637x + .0008488x^2$$

以任何年之 x 值代入此方程式，即可求得恆差值之對數。將各對數化爲眞數，則得恆差之自然數。上例之各年恆差值，及恆差對實際材料之百分比，列於表84。

表84　對數配合之恆差

美國煤油出產之恆差值與原有數值之比較

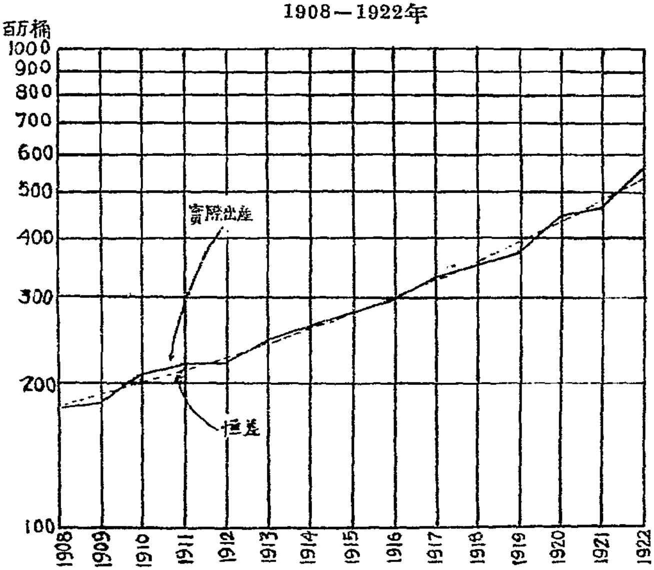
年分	x	y(實際)出產額(單位一百萬桶)	恆差之對數	y(計算)恆差值(單位一百萬桶)	實際數與恆差值之百分比
1908	−7	178.5	2.256639	180.6	98.8
1909	−6	182.2	2.279242	190.2	96.3
1900	−5	209.6	2.303543	201.2	104.2
1911	−4	220.4	2.329540	213.6	103.2
1912	−3	222.9	2.357236	227.6	97.9
1913	−2	248.4	2.386629	243.6	109.0
1914	−1	265.8	2.417720	261.6	101.6
1915	0	281.1	2.450508	282.2	99.6
1916	1	300.8	2.484994	305.5	98.5
1917	2	335.3	2.521177	332.0	101.0
1918	3	355.9	2.559058	362.3	98.2
1919	4	378.4	2.598636	396.9	95.3
1920	5	442.9	2.639913	436.4	101.5
1921	6	472.2	2.682886	481.8	98.0
1922	7	557.5	2.727557	534.0	104.4

代表實際出產額之繪定各點及其恆差線，見圖47。從圖上觀察

，在本例所配合之曲線，代表恆差，頗爲適合。

圖47　美國煤油之出產及其對數配合之恆差，

1908—1922年



對數配合曲線之優點，我們試另行配合同一之曲線（二次拋物線）於自然數上，與本例之結果，一加比較，卽可明顯。自然數之方程式，爲

$$y=279.387+24.262x+1.650x^2$$

此方程式以一九一五年原點。試求其各年之恆差値，及恆差與實際生產之比率，列於表84。此曲線及其實際之生產數字，作圖於圖48。

表84　配合於自然數之恆差

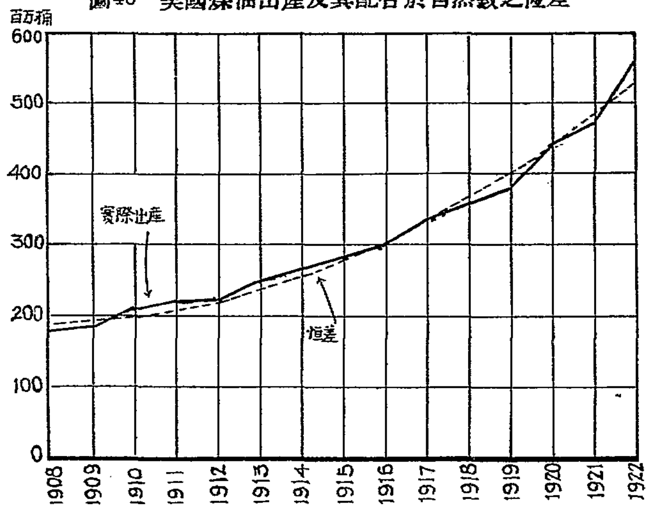
美國煤油出產之恆差值與原有數值之比較

年份	x	y(實際) 出產額 (單位：百萬桶)	y(計算) 恆差值 (單位：百萬值)	眞際數與恆差值之 %
1908	−7	178.5	1[illegible]0.403	93.7
1909	−6	183.2	193.215	94.8
1900	−5	207.6	199.327	105.2
1911	−4	220.4	208.739	105.6
1912	−3	222.9	221.451	100.7
1913	−2	248.4	237.463	104.6
1914	−1	265.8	256.775	103.5
1915	0	281.1	279.387	100.6
1916	1	300.8	305.299	98.5
1917	2	335.3	334.511	100.2
2918	3	355.9	367.023	97.0
1919	4	378.4	402.835	93.9
1920	5	442.9	441.947	100.2
1921	6	472.2	484.359	97.5
1922	7	557.5	530.071	105.2

上述二方程式，皆含有三常數。若將其結果作一比較，尙無不合理。試觀察二者之圖解，（圖47.圖48）即可斷定對數配合之曲線，勝於自然數。更精密之測量，可以二者之均方差比較之。均方差之求法，與次數分配之材料，同一方法；惟差異之測量，不從算術平均數而從恆差值求之。在自然數所配合之曲線，均方差爲12.46；至於對數所配合之曲線，均方差爲9.44(二者所用單位均爲百萬桶）。足證對後者之離差，小於對自然數之曲線。故本例，對數之用，勝於自然數，更無疑義。

其他各種曲線　其餘的曲線，雖不如上述二種之普通，但在時間數列中，有時亦常應用，玆特列舉其重要者數種於下。

圖48　美國煤油出產及其配合於自然數之恆差



1. 普通之拋物線形 (Ordinary Parabolic type) 方程式爲 $y=ax^b$，如用對數則爲：

$$\log y=\log a+b\log x$$

配合此線之二標準式爲：

$$\Sigma\log y=N\log a+b\Sigma(\log x)$$

$$\Sigma(\log x\cdot\log y)=\log a\Sigma(\log x)+b\Sigma(\log x)^2$$

此二方程式之解法，與解普通直線方程式者相同。

2. 簡單指數曲線 Simple Exponential Curve 方程式爲 $y=ab^x$，對數之形式爲

$$\log y=\log a+\log b(x)$$

配合此線之二標準方程式，為：

$$\Sigma(\log y)=N\log a+\log b\Sigma(x)$$

$$\Sigma(x\cdot\log y)=\log a\Sigma(x)+\log b\Sigma(x^2)$$

設以 x 之中點為原點，則 $\Sigma(x)=o$，二標準方程式變為

$$\Sigma(\log y)=N\log a$$

$$\Sigma(x\cdot\log y=\log b\,\Sigma(x^2)$$

移項，得

$$\log a=\frac{\Sigma(\log y)}{N}$$

$$\log b=\frac{\Sigma(x\cdot\log y)}{\Sigma(x^2)}$$

將相當數值代入二式解之，再求其二常數之眞數，則所得之方程式，卽作指數之形式。此種曲線，又稱複利率曲線。

選擇曲線的幾點參考　曲線之種類及其各種之配合方法，已約略討論。關於選擇，雖如上文所述並無一定之原則，大部分由統計家個人之判斷和經驗而決定之，不過尚有數點，可資吾人之參考者，特再申述之。

I. 選擇曲線之第一步，卽將資料繪之於圖。每種資料，可作下列不同之四種圖解，在經濟資料上，(a)(b)二種，更為重要；然後由觀察而決定其適當之形式。

(a)　x, y 俱為自然數(就是用普通算術量表圖)，

(b)　x 為自然數，y 為對數(用單對數圖，縱標軸為對

數量表)。

(c)　x 爲對數，y 爲自然數(用單對數圖，橫標軸爲對數量表)。

(d)　x, y 俱爲對數(繪於複對數圖上)。

II. 適當之曲線，更可由研究二變量 x、y 之關係而決定之。設在簡單之情形，可發見下列之關係：

(a)設 x 值，列成算術級數，其相當之 y 值，成爲幾何級數時，則其關係爲指數曲線。公式爲

$$y=ab^x$$

(b)設 x 值，列成幾何級數，其相對之 y 值，亦爲幾何級數時，則其曲線爲簡單拋物線或雙曲線(Hyperbola)之關係。公式爲

$$y=ax^b \text{ (註)}$$

(c)設 x 值列成算術級數，其相當 y 值之一次差 (First Difference)，爲等差時，兩者之關係，成爲直綫之形式。公式爲

$$y=a+bx$$

(d)設 x 值，列成算術級數，相當 y 值之某次差爲等差時，二變數之關係，可以某次拋物綫表示之，公式

註；b值爲正，則爲拋物線；爲負，則爲雙曲線

$$y=a+bx+cx^2+dx^3+\cdots\cdots\cdots\cdots+qx^n$$

例如下列之資料，y 之三次差，爲一等差，故 x，y 之關係，可以$y=a+bx+cx^2+dx^3$表出之。

x	y	Δy	$\Delta^2 y$	$\Delta^3 y$
1	11	29	32	12
2	40	61	44	12
3	101	105	56	12
4	206	161	68	12
5	367	229	80	12
6	596	309	92	12
7	905	401	104	12
8	1,306	505	116	
9	1,811	621		
10	2,432			

同樣之測驗，如欲作更進之研究，請參閱Running and Lipka: Emperical Formula 。我們還要聲明，除非選一曲綫，能通過給定各點，在分析經濟資料時，所作曲線，未能完全適合上項之測驗，祇能大致類似某種曲線之形式，作我們選擇之參考而已。

丁. 若將原有材料作一度觀察之後，尙未能得一確定之結論時，可配合線多種，而一一比較其結果。（此項比較方法上文已將美國商業失敗統計之資料作例）設二方程式所包函之常數數目相等，可將其對曲線之均方差一一比較，即可測驗配合之是否適當。計算均方差，可由下列之關係以得之。

$$\Sigma(d^2)=\Sigma(y^2)-a\Sigma(y)-b\Sigma(xy)-c\Sigma(x^2y)-\cdots\cdots\cdots$$

（式中之 $\Sigma(d^2)$ 爲對恆差線之離差平方和。）

設常數之數不同，此種試驗，不甚合理，比較之舉，祗能由觀察而得之，統計家個人之判斷，實爲決定曲線之惟一根據。我們還須注意，所配合之曲線，限於資料之內，雖能適合，尙非惟一之標準。因爲我們可以得一方程式，所有常數之數等於其點數，由此而配合之曲線，必通過給定各點。不過此種曲線，未必卽能代表恆差。蓋恆差之觀念，乃一規則的修勻的變動，所以表示一數列之長期趨向。是以須爲簡單之形式，庶無背乎恆差之觀念。這亦並非是說，簡單之曲線，並不適合所給之各點，亦能代表一複雜之恆差。

Ⅳ. 設所配合之曲線，僅用於資料之限止內者，（卽爲補插 Interpolation 用者）其需要之條件，固與曲線之需外延 (Extrapolation)，用以決定將來標準值之根據者，有所不同。前者之惟一需要，是相當的適合 (Reasonable Fit)；而在後者，更要注意所外延之恆差，當合乎邏輯，與過去之事實不相矛盾。因爲外延，本來是代表推度 (Guess)，這種推度，是否合理，當視所配合之曲線，是否適當。而且在過去影響數列之情形，是否在將來仍然相同。今如情形變化，有一新原素混入，可以使外延毫無價值。是故外延之恆差線，所得結論，常起謬誤。平常，外延之時期愈久遠，則謬誤之範圍，

較短近時期爲大。如有新資料，可資利用，恆差線當時加修正。常數較少之簡單曲線，總要勝乎複雜之曲線。用 y 對數所配合之簡單曲線，在外延時，所得結果，較用自然數所配合者，更爲可靠。

V. 假使配合恆差之目的，在乎研究與商情循環有關之離差 (Study of deviation Connected With Business Cycles)，我們還有一種測驗，可以試驗配合之是否正確。其法以循環離差 (Deviation Cycles) 和實際商情的統計來對照。總之，所得恆差線之循環，能與一般商情循環步驟一致，不相矛盾，卽屬可取。

VI. 在某一個數列，要在他全時期中，配合一線，有時實在不可能。這個原因，或許由於在這時期中的情形有所變化，因此，他的趨向亦要改變。例如美國批發物價指數從南北戰爭以降，到1896年，趨向是下降的，可以一直線來代表從1896年以迄歐戰開始時，趨向向上，可以二次拋物線來表示。這類的材料，我們祇有把整個時期，分爲數段，每段分別來配合一適當的趨向線。不過這種方法有時或許會不合理，因爲恆差是一長期漸進的變動。分段的辦法，對於恆差整個的觀念，不免有所矛盾。有時除非情形實在有變動，配合恆差，在全期中，總要用一單獨的線來代表。

IV 平均數法

如趨向爲直線之形式，則平均數法，亦可應用，惟結果不如最小平方法之精密。此法配合直線，須將全數列分爲相等，或約略相等之二部，然後求得下列形式之二方程式

$$\Sigma y = na + b\Sigma(x)$$

依聯立方程式之解法解之，即得 a 和 b 之値。此方法之意義，即將數列之各半部，各求一點，而此點坐標之所在即等於 x's 和 y's 之算術平均數之地位，連接此二點、即得恆差線。試以表 78 銀產量之材料作例，以明其方法。因該數列共有 15 年，分爲兩部，第一部作 7 年，第二部作 8 年，因得二方程式，

$$1268.5 = 7a - 28b$$

$$1876.9 = 8a + 28b$$

解之得　$a = 212.62$

$$b = 6.5$$

恆差線之方程式爲

$$y = 212.62 + 6.5x$$

由最小平方線所得結果，爲

$$y = 212.62 + 6.44x$$

本例，由平均數法之結果，與最小平方線之結果尙稱接近。

與平均數法相似之方法，更有半數平均法。其法亦分全時期爲相等之二部，每部求得一平均數，即爲二時期中點之縱距，然後繪一通通此二點之直線，以代表恆差線。

每月標準值之決定

將根據每年材料求得之方程式，化爲每月之方程式，可不必重行配合，祇將方式程稍加變換卽可。因二方程式之常數間，互有一定之關係存在，每月之數，平均說是每年總數之$\frac{1}{12}$，故每年方程式除以12，卽得每月之方程式。這樣，測量單位，已經縮小至原有單位之$\frac{1}{12}$。惟每月增加率，尙須除以12；因爲按月方程式增加之數，祇及按年方程式時間之$\frac{1}{12}$。現在試以下列假擬之數字爲例，以明其步驟之一般。

年份	商品甲每年之標準產額
1927	186
1928	330
1929	474
1930	618

這些數字表示各年之標準產額，以1927年爲原點，得其趨向線之方程式，爲

$$y=186+144x$$

各年之數字，旣爲總數，可當作集中於該時期之中點(就是7月1日)。而所繪之線，亦必通過各中點。變爲每月之方程式，可分二步手續。第一步先將186和144，各除以12，而得

$$y=15.5+12x$$

此方程式以1927年7月1日爲原點。斜度12，表示一年之某月

至其前一年該月之增加數。故 x 乃以一年爲單位者，尙須除以12，始得每月增加數。得方程式如下

$$y=15.5+x$$

此方程式之原點爲1927年7月1日。原點之地位，最好能使之集中於每月之月中。本例1927年7月1日之縱距爲15.5,每月增加數爲1,則一月十五日之縱距當爲10,如是已將方程式之原點，移至1927年之1月15日，而得方程式

$$y=10+x$$

從此方程式，即可求得任何一月之標準額。不過尙須注意，每年之數值，須爲十二月之總和，上項方法始可適用。否則，每年數字如爲十二月數字之平均數，求每月增加數時，第一步之手續，可以從略，祗將每年增加數除以12即可。

恆差之剔除The Elimination of Trend

恆差剔除之方法通常可有二種：第一種方法的應用，是假定對恆差的離差，爲絕對數，祗要將各恆差縱距，從原有各項減去即可。(見表85第4欄)第二種是假定對恆差的離差，是相對的。剔除之法，即將各原有事項除以相當的恆差值，其結果以百分率表之。(見表85第5欄)有時，更可再進一步，將其結果減以一百，在100以上者，以(+)號表之，在100以下者，以(−)號表之，見表中第6欄

表85 恆差之剔除

年 份	(2) 銀產量 y	(3) 恆差坐標 y'	(4) $(y-y')$	(5) $(y\div y')\times 100$	(6)
1914	172.3	161.6	10.7	106.6	6.6
1915	173.0	168.4	4.6	102.7	2.7
1916	180.8	175.2	5.6	103.2	3.2
1917	186.1	182.0	4.1	102.2	2.2
1918	203.2	188.8	14.4	107.6	7.6
1919	179.8	195.6	−15.8	91.9	− 8.1
1920	173.3	202.4	−29.1	85.6	−14.4
1921	171.3	209.2	−37.9	81.9	−18.1
1922	209.8	216.0	− 6.2	97.1	− 2.9
1923	246.0	222.8	23.2	110.4	10.4
1924	239.5	229.6	9.9	104.3	4.3
1925	245.2	236.4	8.8	103.7	3.7
1926	253.8	243.2	10.6	104.3	4.3
1927	254.0	250.0	4.0	101.6	1.6
1928	257.3	256.8	.5	100.2	.2
1929	256.5	263.6	− 7.1	97.3	− 2.7

第三章

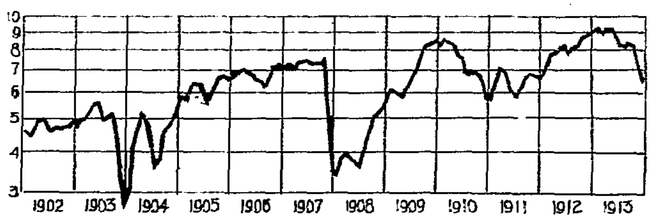
季節變動

季節變動者，依隨季節月令而變動，有若干月在常態之上，有若干月在常態之下，其週而復始所包括之時期，却為十二月(一年)。週期變動，在經濟統計中，既以季節變動為最重要，所以本書的討論，亦限於季節變動。季節變動，簡稱月差。

月差之圖解　研究月差的第一步，就是用圖解來觀察。假設在圖上很顯明地表示每年某幾月是在常態以上，某幾月在常態以下；而且這種現象，每年多大致彷彿，這就可以假定有季節變動的存在，我們再以數學方法，推求他變動的程度。

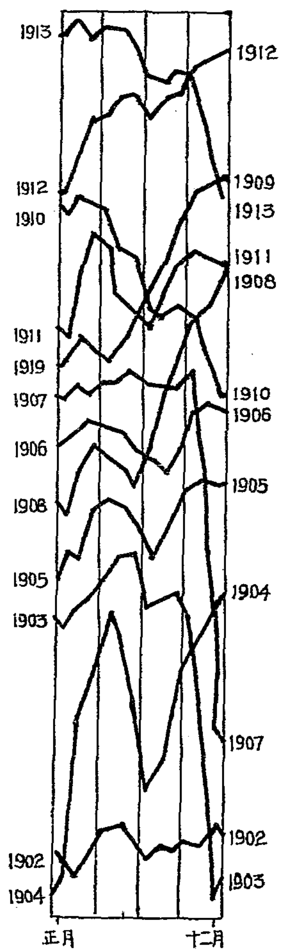
圖解所用之圖，最好用比率圖；因為事實上，同樣的季節變動影響，發生同樣的相對變化；在比率圖上，相對的變化，才能

圖 49　美國生鐵出產統計（1902—1913年）



與以同等的待遇。有些資料的月差，從圖上不易決定，如圖49，需要有更精密的方法。將每年的線，繪在透明紙上，然後重疊而比較之；或在同一圖上，用一公共基線，將每年資料，都繪線其上，如圖50。這種圖對於觀察上大有幫助。

圖50 美國生鐵之出產，1902—1913，



月差之計算　計算月差的方法，有下列四種，試逐一說明之。

1. 同月算術平均法
(Arithmetic Averages of monthly terms)

2. 環比法
(Link relative method)

3. 移動平均法
(Moving Average method)

4. 平均比率法
(Average ratios to Trend)

I.同月算術平均法

同月算術平均法，比較地要算是簡單而又直捷的一種方法。現在以美國鷄蛋價格的資料作例，來說明他的計算方法。

表86　美國鴨蛋每打之平均價

	1910	1911	1912	1913	1914	1915	1916	1917	1918	1919	1920	1921
一月	30.5	30.4	29.5	26.8	30.7	31.6	30.6	37.7	46.3	57.2	64.8	61.1
二月	28.9	22.1	29.1	22.8	28.4	29.2	26.8	35.8	49.4	48.3	56.9	49.6
三月	22.9	16.5	24.5	19.4	24.2	21.3	21.2	33.8	40.4	33.1	46.6	29.2
四月	18.6	14.9	17.8	16.4	17.6	16.6	17.9	25.9	31.2	34.3	38.8	20.4
五月	18.6	14.7	17.1	16.1	16.8	17.1	18.1	30.0	31.0	36.8	37.4	20.2
六月	18.3	14.5	16.7	16.9	17.3	16.6	19.0	31.1	29.8	38.6	37.0	19.4
七月	18.2	14.2	16.7	17.0	17.6	16.8	19.7	28.3	30.7	36.8	36.7	22.0
八月	17.6	15.5	17.4	17.2	18.2	17.0	20.7	29.8	34.4	39.3	40.0	26.6
九月	19.4	17.4	19.1	19.5	21.0	18.7	23.3	33.2	36.4	41.0	44.2	30.4
十月	22.4	20.0	22.0	23.4	23.5	22.3	28.1	37.4	41.6	44.7	50.1	34.2
十一月	25.3	23.5	25.9	27.4	25.3	26.3	32.2	39.4	47.2	54.0	56.9	44.2
十二月	29.0	28.7	29.7	33.0	29.7	30.6	38.1	43.3	55.0	61.9	65.0	51.1
平均	22.5	19.4	22.1	21.3	22.5	22.0	24.6	33.8	39.5	43.8	47.9	34.0

例中資料，包括 12 年，故一月有 12 項：二月有 12 項，其他各月各有12項。先求其平均數，十二月共得十二平均數，表87第 2 欄。若此時期中，歷年之鷄蛋價格，約略相同，祇須將每月之平均數，改爲其全年平均數之百分數，所得結果，即爲月差指數。不過實際上，從 1910 至 1921年，蛋價增高很多，顯然有一種恆差的影響，所以單求月差之時，應當將這些恆差的因子剔除之。

1901—1921年蛋價增高，旣由於恆差之影響，我們拿 1910—1921年之蛋價，配合一直線，求出每年增加數2. 32分(每打)，每

月增加數當爲$\frac{2.32}{12}$卽.193,這就是表示每月蛋價平均要比前一月增高.193分。要剔除此恆差之影響，可任選一月爲基期，將其餘各月平均數改正之。假如用正月爲基期，則二月分之平均數應減去.193。三月分應減去2(.193)，餘類推。設以十二月份爲基期，則十一月分平均數，應加上.193，十月分應加上2(.193)餘類推。不過在本例中，最好取一年中鄰近中央的月分爲基期。所以我們定六月分爲基期，五月分平均數上，須加上.193，四月分加2(.193) …………而七八等月須各減去應減之數。各月

表87　月差指數之計算

同月算術平均法

(1) 月　份	(2) 每月蛋價之算術平均(分)	(3) 改正後之算術平均(分)	(4) 月差指數
一　月	39.77	40.73	138.8
二　月	35.61	36.38	123.9
三　月	27.76	28.34	96.5
四　月	22.53	22.92	78.1
五　月	22.82	23.01	78.4
六　月	22.93	22.93	78.1
七　月	22.89	22.70	77.3
八　月	24.47	24.08	82.0
九　月	26.97	26.39	89.9
十　月	30.81	30.04	102.3
十一月	35.63	34.66	118.1
十二月	41.26	40.10	136.6
總　數		29.357	100.0

平均數，完全改正之後，得表87第3欄各數。第3欄之數字，卽表示季節變動單獨之影響。爲便利進一步分析起見，最好取各月平均數29.357，算作100，將各數化爲百分數，卽成指數，如表87第4欄。

II.環比法

算月差之又一法爲 Pearson 所倡導之環比法。哈佛大學出版之經濟統計雜誌中常用之，其中以用於批爾生所編商情指數爲最著名。

表88　美國鴨蛋每打之平均價之環比

	1910	1911	1912	1913	1914	1915	1916	1917	1918	1919	1920	1921
一　月	107.4	104.8	102.8	90.2	93.0	106.4	100.0	99.0	106.9	104.0	104.7	94.0
二　月	94.8	72.7	98.6	85.1	92.5	92.4	87.6	95.0	106.7	84.4	87.8	81.2
三　月	79.2	74.7	84.2	85.1	85.2	72.9	79.1	94.4	81.8	68.5	81.9	58.9
四　月	81.2	90.3	72.7	84.5	72.7	77.9	84.4	76.6	77.2	103.6	83.3	69.9
五　月	100.0	98.7	96.1	98.2	95.5	103.0	101.1	115.8	99.4	107.3	96.4	99.0
六　月	98.4	98.6	97.7	105.0	103.0	97.1	105.0	103.7	96.1	104.9	98.9	96.0
七　月	99.5	97.9	100.0	101.0	101.7	101.2	103.7	91.0	103.0	95.3	199.2	113.4
八　月	96.7	109.2	104.2	101.2	103.4	101.2	105.1	105.3	112.1	106.8	119.0	120.9
九　月	110.2	112.3	109.8	113.4	115.4	110.0	112.6	111.4	105.8	104.3	110.5	114.3
十　月	115.5	114.9	115.2	120.0	111.9	119.3	120.6	112.6	114.3	109.0	113.3	112.5
十一月	112.9	117.5	117.7	117.1	107.7	117.9	114.6	105.3	113.5	120.8	113.6	129.2
十二月	114.6	122.1	114.7	120.4	117.4	116.3	118.3	109.9	116.5	114.6	114.2	115.6

計算之第一步就是算環比　法將上年十二月分之數字當作100，除本年正月之數，而得一百分比，以一月除二月，二月

表89　月差指數之計算——環比法

(1) 月　份	(2) 每月蛋價之環比中數	(3) 鎖　比	(4) 改正之鎖比	(5) 月差指數
一　月	103.4	100.0	100.0	139.5
二　月	90.1	90.1	89.7	125.1
三　月	80.5	72.5	71.9	100.3
四　月	79.5	57.4	56.7	79.1
五　月	99.2	56.9	56.0	78.1
六　月	98.7	56.2	55.1	76.9
七　月	100.5	56.5	55.2	77.0
八　月	105.2	59.4	57.8	80.6
九　月	111.0	65.9	63.9	89.1
十　月	114.6	75.5	72.9	101.7
十一月	115.8	87.4	84.0	117.2
十二月	116.0	101.4	97.1	135.4
平　均			71.9	100.0

除三月，將此時期中之各月數字，各以前一月除後一月，而得各月對於前月之比例。此等比例，相銜若環，故名環比。

第二步為求環比中數 (Median Link Relatives) 在本例中，一月之環比，共十二個，列為一組，二月分亦有十二個，列成一組。全期共得十二組，乃將每組中之十二環比，依大小順序排列之，而求其中位數，卽得環比中數。組中環比，既為十二，乃偶數，求中位數可取中央二項之平均數，結果如表89第2欄。

第三步算鎖比(Chain Relatives) 上項環比，乃由連環基期而求得，所以示各月與其上一月之關係。環比中數既定，乃須換算之為固定基期，以示各月之環比中數與正月之關係。其法以正月為基期，而求其他各月與此基期之比例，此比例卽為鎖比。既以正月為基期，則正月之鎖比當為100，二月之鎖比仍同於二月之環比；三月之鎖比，等於以三月之環比乘二月之鎖比，除以100，；其餘各月之鎖比，以同法求得之。

第四步為校正　設資料無恆差之影響，則十二月分之鎖比乘一月之環比中數，在理論上當仍為 100。但本例中，其積為104.8,(101.4×103.4)，是則，其間之差數，必由於恆差之影響，當校正之。此項指數，既由環比各項累次相乘而得，此項差數，亦係累積而上，頗與複利率相似，故複利率之公式，亦可應用。其式

$$p_n=p_0(1+r)^n$$

增加率為 $r=\sqrt[n]{\frac{p_n}{p_0}}-1$

代入本例，得

$$104.8=100(1+r)^{12}$$

$$r=\sqrt[12]{\frac{104.8}{100}}-1$$

$$=1.004-1=.004$$

每月因恆差影響而增加之比率為 .004，欲剔除此項影響，二月鎖比當除以 $(1+r)$，卽 1.004，三月鎖比，當以 $(1+r)^2$ 卽 1.008 除之，如是類推，至十二月之鎖比，則當以 $(1+r)^{11}$ 卽1.044除之，是為各月之改正鎖比，表 89 第 4 欄。(此項校正手續，可利用對數，計算較為便利)

最後一步，為求得月差指數　改正之數，為各月對於一月之百分數。於是求此十二個月百分數之平均數，以之除各月百分數，所得結果，為月差指數。

III 移動平均法

季節變動的循環，恆有一定的時期，就是十二個月。所以用移動平均數來剔除恆差，那就很靠得住；而同時季節變動的升降，依舊可以表現出來。計算方法，第一步就是繼續求出十二個月的移動平均數。因為十二是偶數，所以應當再將這結果，求兩個月的移動平均，才可以和原有數字比較，表90就是求出

的移動平均數。

表90　蛋價之移動平均數，1910—1921

（十二月之移動平均數再用二月移動平均改正）

	1910	1911	1912	1913	1914	1915	1916	1917	1918	1919	1920	1921
一　月		20.25	21.26	20.78	22.74	22.29	22.24	30.06	36.72	41.39	46.60	40.47
二　月		19.99	21.45	20.79	22.80	22.20	22.51	30.80	37.01	41.86	46.63	39.30
三　月		19.82	21.60	20.7[illegible]	22.91	22.06	22.86	31.59	37.33	42.25	46.80	38.16
四　月		19.64	21.75	20.87	22.97	21.91	23.29	32.39	37.64	42.57	47.15	36.93
五　月		19.40	21.94	20.93	22.89	21.90	23.77	33.08	38.14	42.98	47.50	35.74
六　月		19.38	22.08	21.19	22.66	21.98	24.33	33.59	38.96	43.56	47.75	34.63
七　月	22.47	19.33	22.00	21.49	22.57	21.98	24.94	34.17	39.90	44.17	47.72	
八　月	22.18	19.58	21.6[illegible]	21.88	22.64	21.94	25.60	35.10	40.31	44.84	47.26	
九　月	21.63	20.20	21.16	22.3[illegible]	22.55	21.74	26.51	35.93	32.96	45.76	46.24	
十　月	21.21	20.66	20.89	22.57	22.39	21.78	27.36	36.43	39.79	46.50	44.74	
十一月	20.89	20.88	20.79	22.65	22.36	21.88	28.20	36.69	40.17	46.71	43.26	
十二月	20.57	21.07	20.76	22.69	22.35	22.0[illegible]	29.20	36.68	40.77	46.67	41.81	

第二步將移動平均數作100,算作標準值，每月實際價格，除以相當的移動平均數，得表91之各百分數。因為各月的百分數，

表91　蛋價十二個月移動平均數對實在價格之百分比

	1910	1911	1912	1913	1914	1915	1916	1917	1918	1919	1920	1921
一　月		150.1	138.8	129.0	135.0	141.8	137.6	125.4	126.1	138.2	139.1	151.0
二　月		110.6	135.7	109.7	124.6	131.5	119.1	116.2	133.5	115.4	122.0	126.2
三　月		83.2	113.4	93.3	105.6	96.6	92.7	107.0	108.2	78.3	99.6	76.5
四　月		75.9	81.8	78 6	76.6	75.8	76.9	80.0	82.9	80.6	82.3	55.2
五　月		75.5	77.9	76.7	73.4	78.1	76.1	90.7	81.3	85.6	78.7	56.5
六　月		74.8	75.6	79.8	76.3	75.5	78.1	92.6	76.5	88.6	77.5	56.0
七　月	81.0	73.5	75.9	73.1	78.0	76.4	79.0	82.8	76.9	83.3	76.9	
八　月	79.4	79.2	80.4	78.6	80.4	77.8	80.9	84.9	85.3	87.6	84.6	
九　月	89.7	86.1	90.3	87.4	93.1	86.0	87.9	92.4	91.1	89.6	95.6	
十　月	105.6	96.8	105.3	103.7	105.0	102.4	102.7	102.7	104.5	96.1	112.0	
十一月	121.1	112.5	124.6	121.0	113.1	120.2	114.2	107.4	117.5	115.6	131.5	
十二月	141.0	136.2	143.1	145.4	132.9	139.0	130.5	118.1	134.9	132.6	155.5	

每年不同，例如一月份的百分數，最大為150,最小為 125.4%,相差很大，須求其各月所有百分數的平均，才得月差指數。平均方法，或用算術平均，或用中位數，可隨情勢而定，表92中

表92　月差指數之計算——移動平均法

(1) 月份	(2) 算術平均數 (未改正)	(3) 中位數 (未改正)	(4) 算術平均數 (已改正)	(5) 中位數 (已改正)
一月	137.4	138.2	137.9	138.7
二月	122.2	122.0	122.6	122.5
三月	95.9	96.6	96.2	97.0
四月	77.0	78.6	77.2	78.9
五月	77.1	77.9	77.6	78.2
六月	77.4	76.5	77.7	76.8
七月	78.4	78.0	78.7	78,3
八月	81.7	80.4	82.0	80.7
九月	89.9	83.7	90.2	90.1
十月	103.3	103.7	103.7	104.1
十一月	118.1	117.5	118.5	118.0
十二月	137.2	136.2	137.7	136.7
平均	99.65	99.608	100.0	100.0

，兩例俱備。最後將上面十二個平均數，表92第(3)(4)二欄，更求出一平均數，算作基數，將十二月的平均數改正之。其結果列於表92第(4)(5)二欄，已爲改正之月差指數。

IV 對恒差線之平均比率法

此法最初爲 Helen D. Falkner 所倡導，亦有其相當之優點。應用時，必須先將資料配合一適當之趨向線，不論曲線的形態如何，求出每月標準值，於是以此標準值作100，除每月實際之數值，而得其百分率。其原理正與上述移動平均法相同。幷用一複次數表(Multiple Frequency table)(表93)的幫助，以觀察季節變動的性質，和決定平均數之選擇。平均數決定後，求出各月份的平均指數。再將此平均指數之總平均數作基數(100)以改正各月平均數。

仍以美國鷄蛋價格之資料爲例，以說明其計算方法。先將

表86配合一直線(以1916年一月一日爲原點)，得方程式爲

$$y=29.45+2.32x$$

根據上式，計算每月之標準値。再以此標準値爲爲基數，求出每月之十二個百分數，然後將上項百分數，編成複次數表。從表

表93　每月蛋價對恆差坐標之百分比率之次數分配

	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
170 and Over	/	/										
165 – 169.9	//											
160 – 164.9												//
155 – 159.9	/											/
150 – 154.9												/
145 – 149.9	//	/										
140 – 144.9		//	/								//	/
135 – 139.9	/										/	
130 – 134.9		/									/	//
125 – 129.9										/		/
120 – 124.9	/	//								/		/
115 – 119.9	///		///							//	///	/
110 – 114.9		//		/	/	/			/	/	/	
105 – 109.9		/					/		//			/
100 – 104.9	/	/	/			/		//	/	/	//	/
95 – 99.9			/	/	/		/	//	/	//		
90 – 94.9		/		/	//	//	/				//	
85 – 89.9			//	//	/		//	/	//	//		
80 – 84.9			/	/	/	/		//	/			
75 – 79.9			/	/	/	//	/		/	//		
70 – 74.9			//	/		/	//	/	//			
65 – 69.9				/	//	/	/	//	/			
60 – 64.9					/	/	/	/				
55 – 59.9				//	/	/	/	/				
50 – 54.9							/					
Under 50				/	/	/						

中我們可以看出冬季各月之價格，總要比春夏二季高，這就可以證明有很顯著的月差。幷由於表上的陳示，見到資料之分配，不甚集中。在選擇平均數時，中位數因爲要受一二項增減而改變其數値的危險；同時，算術平均數也因爲受極端項的影響

太大，有時代表的資格，不很健全；所以用一個折中的辦法，是就拿當中的幾項來求算術平均數。假如在表上觀察，還不能確定時，最好算出幾種平均數，再就其結果觀察之，然後才可決定。表 94 是取中央 2 項、4 項、6 項、8 項、 10 項 12 項的算術平均數。

表94　蛋價之月差指數

	中間各月平均數，未改正之指數						中間各月平均數，已改正之指數					
	2月	4月	6月	8月	10月	12月	2月	4月	6月	8月	10月	12月
一　月	140.5	138.2	138.0	[illegible]	133.8	141.4	142.[illegible]	139.4	138.4	129.1	139.6	140.5
二　月	120.8	121.6	123.4	123.8	123.6	125.7	122.[illegible]	122.6	123.8	123.7	123.5	124.9
三　月	92.8	95.0	96.[illegible]	97.1	96.7	98.3	93.[illegible]	95.8	97.2	97.1	96.6	97.7
四　月	80.0	79.0	79.1	78.1	78.5	78.[illegible]	80.9	79.7	79.3	78.1	78.2	78.4
五　月	79.8	79.[illegible]	79.1	78.8	78.8	79.2	80.7	80.0	79.3	78.8	78.7	78.7
六　月	77.6	77.9	78.0	78.6	79.0	78.9	78.5	78.6	78.8	78.6	78.9	78.4
七　月	76.1	77.3	77.3	77.[illegible]	77.7	78.1	77.0	77.9	77.5	77.5	77.6	77.6
八　月	80.3	80.5	81.4	81.8	82.1	82.0	81.2	81.2	81.6	81.8	82.0	81.5
九　月	88.3	89.2	89.7	90.[illegible]	90.2	90.0	89.3	89.9	90.6	90.1	90.[illegible]	89.5
十　月	00.7	102.2	102.4	102.4	102.3	102.4	101.8	102.1	102.7	102.4	102.2	101.8
十一月	116.6	115.7	116.0	117.[illegible]	117.3	117.3	117.[illegible]	116.7	116.3	117.5	117.2	116.6
十二月	33.0	134.0	134.7	135.0	135.5	135.[illegible]	134.5	135.1	135.[illegible]	135.6	135.4	134.4

在本例，逐年季節變動很顯著而有規則，所以表94各欄的結果，沒有多大區別，其中以取中央四項求得之算術平均，較爲優越。總之複次數表上，次數分配愈集中，所取項數，可以愈少。

四種方法的比較

上文對於四種方法，已分別討論，究竟優劣如何，尙無絕對之標準。不過總結來說，算術平均法，計算方法，較爲簡捷，只是設有絕端差異時，或是某時期中資料之數值，有重大改

變時，這方法就不適用。環比法之計算太繁，在分析性質相同，及很有規則之數列時，這方法亦不見得有超越其他簡法之處。據Mills 之意見，最後二法，大概是最有用和最普通的方法。二者的計算，還不很麻煩。尤其是對算術恆差線的平均比率法，用比率的平均數，不用絕對值，在理論上，更有充分之根據。下表是根據上述四種方法算出之四種月差指數，合成一表，(表95) 以便比較。

表95　鷄蛋價格之月差指數

月分	移動平均中位數	算術平均法	環比法	對恆差之比率法
一月	138.7	138.8	139.5	139.4
二月	122.5	123.9	125.1	122.6
三月	97.0	96.5	100.3	95.8
四月	78.9	78.1	79.1	79.7
五月	78.2	78,4	78.1	80.0
六月	76.8	78.1	76.9	78.6
七月	78.3	77.3	77.0	77.9
八月	80.7	82.0	80 6	81.2
九月	90.1	89.9	89.1	89.9
十月	104.1	102.3	101.7	103.1
十一月	118.0	118.1	117.2	116.7
十二月	136.7	136.6	135.4	135.1
平均數	100.0	100.0	100.0	100.0

第四章
循環變動

恆差與月差旣明，乃進而討論循環變動。經濟狀況，常有盛衰交替而來之現象，盛極則衰，衰極則復盛，是爲經濟循環。本書之立場，爲統計方法之應用，故本章敍述之範圍，限於分析循環作用之統計方法；至於循環之理論學說等，祇能從略。

循環之一字，表示繼續起伏之運動，不過不必有週期變動的意思。循環所需時間之久暫和其起伏之廣度，每個循環各不相同。美國商情循環之時期，大約平均爲 $3\frac{1}{2}$年，或42月，有時祇不過 28 月，有時竟長至 48 月。總之，循環變動在現在還沒有一個明明白白可以指定的週期。循環的廣度，又是各不相同，他們變動的全距，不但在單一數列中，各各循環互有不同，就是若干數列，亦彼此參差互異。試觀圖51二曲線之循環變動，可見批發物價變動之全距較股票價格之變動爲小。

時間數列之因素，有恆差月差意外變動及循環變動等。欲明循環作用，不可不先將各變化之原素分析之，然後剔除恆差月差，意外變動亦當除去，所剩餘者祇循環之作用。（當然尙有其他變動，但有時此種其他變動可忽視，）恆差剔除之方法有二：由實際數內減去趨向値，或以趨向値除實際數。玆以表

表96　循環變動之計算　（美國1906—1909生鐵之產量）

月　分	原有資料	趨向縱距	(a) 對 (b) 之相對數	月差指數	差異	異差之乘方	以σ爲單位之差異
	(a)	(b)	(c)	(d)	(e)	(f)	(g)
1906 一 月	2068	1799	115	101	14	195	.4
二 月	1904	1804	105	95	10	100	.6
三 月	2155	1810	119	107	12	144	.5
四 月	2073	1815	114	103	11	121	.5
五 月	2038	1821	115	103	12	144	.5
六 月	1976	1826	108	95	13	169	.6
七 月	2013	1832	110	94	16	256	.7
八 月	1926	1837	105	97	8	64	.4
九 月	1960	1842	106	98	8	64	.4
十 月	2196	1848	119	104	15	225	.7
十一月	2187	1853	118	101	17	289	.7
十二月	2235	1859	120	102	18	324	.8
1907 一 月	2205	1864	118	101	17	289	.7
二 月	2045	1870	109	95	14	196	.6
三 月	2226	1875	119	107	12	144	.5
四 月	2216	1881	118	103	15	225	.7
五 月	2295	1886	122	103	19	361	.8
六 月	2234	1892	118	95	23	529	1.0
七 月	2255	1897	119	94	25	625	1.1
八 月	2250	1902	118	97	21	441	.9
九 月	2183	1908	115	98	17	289	.7
十 月	2336	1913	122	104	18	324	.8
十一月	1828	1919	95	101	− 6	56	− .3
十二月	1234	1924	64	102	−38	1444	−1.7
1908 一 月	1045	1930	54	101	−47	2209	−2.1
二 月	1077	1935	56	95	−39	1521	−1.7
三 月	1228	1941	63	107	−44	1936	−1.9
四 月	1149	1946	59	103	−44	1936	−1.9
五 月	1165	1951	60	103	−43	1849	−1.9
六 月	1092	1957	56	95	−39	1521	−1.7
七 月	1218	1962	62	94	−32	1024	−1.4
八 月	1348	1968	60	97	−28	784	−1.2
九 月	1418	1973	72	98	−26	676	−1.1
十 月	1563	1979	79	104	−25	625	−1.1
十一月	1577	1984	79	101	−22	484	1.0
十二月	1740	1990	87	102	−15	225	.7
1909 一 月	1801	1995	90	101	−11	121	.5
二 月	1703	2001	85	95	−10	100	.4
三 月	1832	2006	91	107	−16	256	.7
四 月	1738	2011	86	103	−17	289	.7
五 月	1880	2017	93	103	−10	100	.4
六 月	1929	2023	95	95	0	0	.0
七 月	2101	2028	104	94	10	100	.4
八 月	2246	2033	111	97	14	196	.6
九 月	2385	2039	117	98	19	361	.8
十 月	2600	2044	127	104	23	529	1.0
十一月	2547	2050	124	101	23	529	1.0
十二月	2635	2055	128	102	26	676	1.1
						25076	

$\Sigma d^2 = 25076$

$$\sigma = \sqrt{\frac{25076}{48}} = 22.8$$

96美國生鐵出產之資料爲例，以說明之。先將生鐵出產之各月統計，與恆差線之各月縱距並列。(表中 (a) (b) 二欄) 然後以各月恆差線之縱距除各月之出產，得二者相比之百分數(c 欄) 此時恆差之影響已去，再以此百分數，減以月差指數(d 欄)，所得結果，或正或負(e 欄)，乃除去恆差月差後所餘之循環現象。如資料爲年數字，則無月差之因素，祇須將恆差除去，卽可適用。

我們假如要將幾種循環作用作比較，就不能以上節所得結果爲滿足。因爲各數列之升降高低之程度不同，譬如圖51批發物價之百分離差，變動較小，從＋9至－9%；而股票價格之百分離差，則變動甚大，從＋34至－36%；二者之變動，意義迥乎不同，當以其各數列之標準差除百分離差，這樣調節之後，乃互相接近，便於比較。若測量離差，欲減少較大離差之影響，則可用平均差爲除數，以代標準差。

第五章

時間數列之相關

第二編所論之相關，限於次數分配之資料，時間數列之相關，並未提及，雖方法上大致相同，而性質上有其特殊之點，故須分別討論。時間數列既是幾種因素變動的總結果，所以計算相關時，亦須將各因素分別剔除，始合個別研究之用，否則，所得結果駁雜不純，易起誤解。

研究二時間數列之相關，吾人當注意其是否有恆差和月差之存在。因爲恆差是長時間變動的結果，而月差又爲季節自然的人爲的變遷，各變動多受共同之影響，有時二數列之恆差月差，雖屬相同，但不一定是表示其間有相互關係。在此情形之下，計算相關，應將恆差月差先行剔除，單留循環作用來比較。假使材料是用年數字，季節變動之剔除，就不成問題。

函數相關 Functional Relationship

求相關有圖解和批氏係數二方法，圖解當然不很精密，祇能作觀察上的幫助而已。

假使從循環變動之數列計算其相關，因爲恆差月差都不存在，則手續極簡；尤其是在循環變動以標準差爲單位之場合，

更爲簡單。此時 X, Y 二數列之各項，爲 $\frac{x}{\sigma_x}$ 及 $\frac{y}{\sigma_y}$ 之形式，以批爾生之相關方法計算，可用公式

$$r=\frac{\Sigma\left(\frac{x}{\sigma_x}\cdot\frac{y}{\sigma_y}\right)}{N}$$

示例於表 97。

表97　循環變動相關之計算

(紐約市外銀行之清算與生鐵產量,1904—1907)

月　份		市外銀行清算數之循環，σ單位	生鐵產量之循環，σ單位	$\left(\frac{x}{\sigma_x}\cdot\frac{y}{\sigma_y}\right)$
1904	一　月	−.6	−2.0	1.20
	二　月	−.1	−.8	.08
	三　月	−.5	−.6	.30
	四　月	−.6	−.1	.06
	五　月	−.7	−.3	.21
	六　月	−.5	−.8	.40
	七　月	−.7	−1.4	.98
	八　月	−.4	−1.3	.52
	九　月	−.2	−.6	.12
	十　月	−.4	−.7	.28
	十一月	.4	−.3	−.12
	十二月	.4	.1	.04
1905	一　月	−.5	.6	−.30
	二　月	−.1	.3	−.03
	三　月	.4	.7	.28
	四　月	.1	.8	.08
	五　月	.8	.8	.64
	六　月	.5	.6	.30
	七　月	.1	.5	.05
	八　月	.6	.7	.42
	九　月	.5	.8	.40
	十　月	.4	.9	.36
	十一月	.7	1.0	.70
	十二月	.6	1.1	.66
1906	一　月	1.4	1.2	1.68
	二　月	.8	.9	.72
	三　月	.7	1.0	.70
	四　月	.2	.9	.18
	五　月	.7	.9	.63
	六　月	.6	.8	.48
	七　月	.4	.9	.36
	八　月	.8	.6	.48

接97表

月　份		市外銀行清算數之循環，σ單位	生鉄產量之循環，σ單位	$\left(\frac{x}{\sigma_x}\cdot\frac{y}{\sigma_y}\right)$
1906	九　月	.1	.7	.07
	十　月	1.0	1.0	1.00
	十一月	.9	1.3	1.17
	十二月	.6	1.3	.78
1907	一　月	1.5	1.3	1.95
	二　月	.9	1.0	.90
	三　月	1.2	.9	1.08
	四　月	1.0	1.0	1.00
	五　月	1.5	1.1	1.65
	六　月	.7	1.3	.91
	七　月	1.3	1.4	1.82
	八　月	1.0	1.2	1.20
	九　月	.4	1.0	.40
	十　月	1.3	1.1	1.43
	十一月	−2.0	−.1	.20
	十二月	−2.7	−1.8	4 86
				33.28

$$r=\frac{\Sigma\left(\frac{x}{\sigma_x}\cdot\frac{y}{\sigma_y}\right)}{N}=\frac{33.28}{48}=6.9$$

設原有資料，不爲循環之形式，擬就原有資料計算相關，可用下列公式。

$$y=\frac{\Sigma(X-A)(Y-B)-N(M_x-A)(M_y-B)}{\sqrt{\Sigma(X-A)^2-N(M_x-A)^2}\sqrt{\Sigma(Y-B)^2-N(M_y-B)^2}}$$

X, Y 是二數列之各變量

M_x M_y 是算術平均數

* 此公式已在第二編相關章中論及，設以他種符號表之，即爲

$$r=\frac{\Sigma x'y'-nc_xc_y}{\sqrt{\Sigma(x')^2-nc_x^2}\ \sqrt{\Sigma(y')^2-nc_y^2}}$$

N 是項數

A 是 X 數列之假設平均數，B 是 Y 數列的假設平均數。

從表98得下列各數值

$$\Sigma(X-A)(Y-B)=6556$$

$$M_x-A=\frac{\Sigma X-NA}{N}=\frac{-269}{26}=-7.5$$

$$M_y-B=\frac{\Sigma Y-NB}{N}=\frac{-129}{36}=-3.5$$

$$\Sigma(X-A)^2=16509$$

$$\Sigma(Y-B)^2=3083$$

$$r=\frac{6556-36(-7.5)(-3.5)}{\sqrt{16509-(36\times 56.25)}\ \sqrt{3083-(36\times 12.25)}}$$

$$=\frac{6556-945}{\sqrt{16509-2025}\ \sqrt{3083-441}}$$

$$=\frac{5611}{\sqrt{14484}\ \sqrt{2642}}$$

$$=\frac{5611}{6120}$$

$$=.91$$

如有計算機，可用 Ayres.之公式，爲 Leonard P. Ayres 所倡導

$$r=\frac{\Sigma XY-\left(\frac{\Sigma X}{N}\right)\Sigma Y}{\sqrt{\left[\Sigma X^2-\left(\frac{\Sigma X}{N}\right)\Sigma X\right]\left[\Sigma Y^2-\left(\frac{\Sigma Y}{N}\right)\Sigma Y\right]}}$$

此公式之計算，免去正負號差數之麻煩，及眞假平均數之改正手續，而且可以從平方表上找出各數的平方，用計算機加法機以求積數及總和，所以很便利。現在以99爲例，以示其計算方法。

表98　時間數列之相關係數計算法

月分	1 鋼鐵指數 X	2 化學油漆指數 Y	3 X−A (A=100)	4 X−B (B=100)	5 (X−A) (X−B)	6 (X−A)²	7 (X−B)²
1919一月	118	104	18	4	72	324	16
二月	112	102	2	2	24	144	4
三月	108	99	8	−1	−8	64	1
四月	105	98	5	−2	−10	25	4
五月	100	97	0	−3	0	0	9
六月	100	96	0	−4	0	0	16
七月	103	99	3	−1	−3	9	1
八月	105	100	5	0	0	25	0
九月	106	102	6	2	12	36	4
十月	70	101	−30	1	−30	900	1
十一月	81	103	−19	3	−57	361	9
十二月	98	102	−2	2	−4	4	4
1921一月	110	104	10	4	40	100	16
二月	112	104	12	4	48	144	16
三月	113	105	13	5	65	169	25
四月	115	104	15	4	60	225	16
五月	106	102	6	2	12	36	4
六月	112	103	12	3	36	144	9
七月	111	106	11	6	66	121	36
八月	108	106	8	6	48	64	36
九月	111	106	11	6	66	121	36
十月	110	106	10	6	60	100	36
十一月	107	102	−7	2	14	49	4
十二月	98	96	−2	−4	8	4	16
1921一月	81	92	−19	−8	152	361	64
二月	81	91	−19	−9	171	361	81
三月	77	90	−23	−10	230	529	100
四月	72	88	−28	−12	336	784	144
五月	69	87	−31	−13	403	961	169
六月	66	84	−34	−16	544	1156	256
七月	57	83	−43	−17	731	1849	289
八月	58	82	−42	−18	756	1764	324
九月	59	82	−41	−18	738	1681	324
十月	62	83	−38	−17	646	1444	289
十一月	65	82	−35	−18	630	1225	324
十二月	65	80	−35	−20	700	1225	460
			−269	−129	6556	16,509	3083

表99　時間數列之相關係數計算法

月分	1 鋼鐵指數 X	2 化學油漆指數 Y	3 X^2	4 Y^2	5 XY
1919一月	118	104	13,924	10,816	12.272
二月	112	102	12,544	10,404	11.424
三月	108	99	11,664	9,801	10.692
四月	105	98	11,025	9,604	10.290
五月	100	97	10,000	9,409	9.700
六月	100	96	10,000	9,216	9.600
七月	103	99	10,609	9,801	10.197
八月	105	100	11,025	10,000	10.500
九月	106	102	11,236	10,404	10.812
十月	70	101	4,900	10,201	7.070
十一月	81	103	6,561	10,609	8.343
十二月	98	102	9,604	10,404	9.996
1920一月	110	104	12,100	10,816	11.440
二月	112	104	12,544	10,816	11.648
三月	113	105	12,769	11,025	11.865
四月	115	104	13,225	10,816	11.960
五月	106	102	11,236	10,404	10.812
六月	112	103	12,544	10,609	11.536
七月	111	106	12,321	11,236	11.766
八月	108	106	11,664	11,236	11.448
九月	111	106	12,321	11,236	11.766
十月	110	106	12,100	11,236	11.660
十一月	107	102	11,449	10,404	10.914
十二月	98	96	9,604	9,216	9.408
1921一月	81	92	6,561	8,464	7.452
二月	81	91	6,561	8,281	7.371
三月	77	90	5,929	8,100	6.930
四月	72	88	5,184	7,744	6.336
五月	69	87	4,761	7,569	6.003
六月	66	84	4,356	7,056	5.544
七月	57	83	3,249	6,889	4.731
八月	58	82	3,364	6,724	4.756
九月	59	82	3,481	6,724	4.838
十月	62	83	3,844	6,889	5.146
十一月	65	82	4,225	6,724	5.330
十二月	65	80	4,225	6,400	5.200
	3,331	3,471	322,709	337,283	326.756

$$r=\frac{326,756-\left(\frac{3331}{36}\right)3471}{\sqrt{\left[322,709-\left(\frac{3331}{36}\right)3331\right]\left[337,283-\left(\frac{3471}{36}\right)3471\right]}}=.91$$

本例 r 之值爲.91,顯然是表示有高度之相關。設將上列之各相當值，代入回歸方程式，我們可得一推測方程式如下：

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

$$y=.91\frac{\sqrt{2642}}{\sqrt{14484}}x$$

$$=.91\frac{51.4}{120.3}x$$

$$\therefore\ y=3.665\ x$$

標準差誤 S_y 所以測量對回歸線分佈之情形，其值亦可由下列公式得之：

$$S_y=\sigma_y\sqrt{1-r^2}$$

在本例中，

$$S_y=51.4\sqrt{1-(91)^2}$$

$$=15.42$$

此數值之意義若何，已在第二編中說明，玆不再贅。

時間數列計算相關之困難　我們須注意，在上面計算時間數列之相關，有一個人爲的原素Personal Element參雜着。這人爲的原素，在從前討論相關時(次數分配之材料)不致發生。因爲現在的離差，是從長期趨向而測量，不從算術平均數測量。要知趨向線是由個人的私意選擇，採用不同之趨向線，可以

使結果大異。有時材料之趨向線，可以用種種的方程式來配合，假使我們能肯定那一種配合最好，那末，整個的問題，都可以解決。可是現在的統計學，還沒有一個客觀的標準。所以計算時間數列的相關，最重要的問題，不在機械的計算方法，而在趨向線的選擇。假如從圖上觀察，或從各方面旁證，可以證明所選之趨向線，確能代表該數列的恆差，所求得之 r ，始有價值。否則，r之值雖有高度相關，仍屬無用。

前引和落後Lead and Lag

上面討論的相關，是研究數列的函數關係，現在我們更從另一立場來討論，就是時間相關(Temporal Correlationship)說明這時間數列之相關，我們可以拿美國實業股票指數和一般商情指數的材料(1903年1月至1914年6月)作例(表100,表101)。先將二數列製圖如圖 51, 就可看出二數列的變動，顯然有一種相關，但是他們起伏的時間，並不同時。這種現象，大概由於從原因到結果，須經過相當之時間。其第一線落後第二線所經過的時期，稱曰落後 Lag。第二線居先第一線之時期，稱曰前引(Lead)。這種數列間變動之關係，稱為前引相關。由於這前引相關測量，使經濟預測成為科學化。要決定從原因到結果，究竟要經過多久的時間，我們可將兩組循環變動，用同一量表，以 σ 為單位，各繪於一透明紙上，將一圖影於他圖之下，前

表100　實業股票價格之循環(1903—1914)

(數字爲對標準値之差數而以標準差爲單位)

	1903	1904	1905	1906	1907	1908	1909	1910	1911	1912	1913	1914
一　月	−.2	−2.6	−.1	2.3	1.6	−1.3	.5	1.0	−.1	−.4	−.2	−.7
二　月	−.1	−2.1	.2	2.2	1.4	−1.5	.3	.5	.1	−.4	−.5	−.6
三　月	−.3	−2.1	.6	1.9	.6	−1.1	.3	.8	−.1	−.1	−.7	−.6
四　月	−.5	−2.0	.7	1.7	.6	−.8	.5	.5	−.1	.3	−.6	−.8
五　月	−.5	−2.1	.2	1.4	.4	−.5	.8	.4	.0	.2	−.7	−.8
六　月	−.9	−2.1	.3	1.5	.2	−.5	.9	.1	.1	.2	−1.1	−.7
七　月	−1.4	−1.8	.7	1.3	.3	−.2	1.1	−.4	−.1	.2	−.9	
八　月	−1.7	−1.6	.8	1.7	−.3	.3	1.4	−.3	−.3	.3	−.7	
九　月	−1.9	−1.3	.7	1.7	−.5	.1	1.4	−.7	−.7	.4	−.5	
十　月	−2.5	−.9	.8	1.7	−1.3	.2	1.4	.0	.7	.3	−.8	
十一月	−2.4	−.3	1.7	1.7	−1.6	.5	1.4	−.1	−.5	.2	−.5	
十二月	−2.1	−.1	1.8	1.7	−1.6	.5	1.3	−.2	−.4	.0	−.4	

表101　一般商情指數之循環(1903—1914)

(數字爲對標準値之差數而以百分比爲單位)

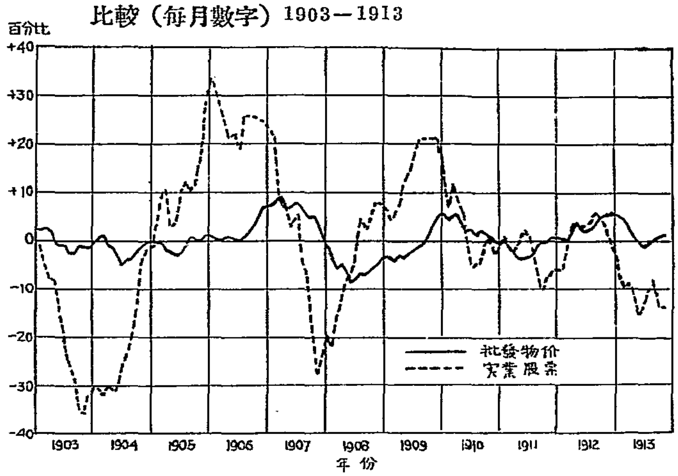
	1903	1904	1905	1906	1907	1908	1909	1910	1911	1912	1913	1914
一　月	3	10	−5	10	17	−13	−5	12	0	−2	10	−0
二　月	1	5	−5	9	14	−15	−5	10	−1	3	8	−2
三　月	2	8	−2	7	15	−15	−5	12	0	3	3	−0
四　月	4	6	−1	5	14	−14	−5	11	−3	6	5	−1
五　月	3	8	0	8	18	−17	−3	6	−2	5	5	−5
六　月	2	10	0	7	16	−17	−3	7	−1	3	4	−4
七　月	3	12	−1	7	18	−16	1	4	−2	5	5	
八　月	−1	12	0	8	16	−14	2	4	−1	6	2	
九　月	−2	8	2	5	12	−11	5	4	−1	5	3	
十　月	−6	9	2	10	13	−11	8	4	−1	8	5	
十一月	−11	4	4	11	−4	−9	9	2	−1	7	−1	
十二月	−14	5	7	13	−12	−8	10	1	−2	7	−2	

*　編製者　美國電報電話公司

後移動，以比較之。設二線之起伏形式大致脗合，則其前引所經過之時期，亦可約略決定。

不過這種觀察，有一點困難，就是前引的時期，並非一律不變，圖上一部分或許時間長些，有幾部分或許短些，所以不易決定其前引，究爲若干時。測量相關的第二種方法，是用相關係

圖51　批發物價 (Bradstreet Index)與十二種實業股票循環之比較 (每月數字) 1903—1913



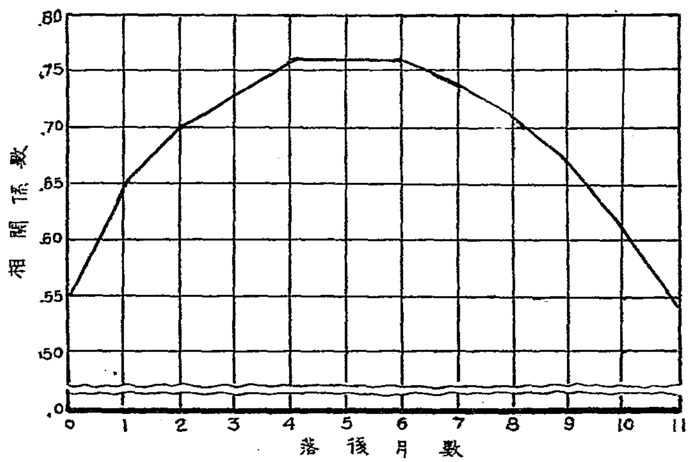
數。先假定前引的時間爲若干月，一一算出他的相關度，當相關度之值爲最大時，卽決定爲前引之時期。在本例中，我們先計算二數列同期相關係數得+.55。然後假定前引爲一月，計算每前一月的股票指數和本月的一般商情指數的相關，就是以1903年1月股票價格乘以1903年2月之一般商情指數，以2月之股票價格乘以3月之商情指數，餘類推。在此計算中，僅包括137個月的數值，因爲1903年1月之商情指數，和1914年6月之股票價格之數字，並不列入計算中。因此 c_x，c_y 之值及二標準差之值，與上一次相關計算，稍稍不同。從此計算，所得相關係數爲+.65。然後再假定前引爲3，4，5，6，……… 月，以同

表102　實業股票價格與一般商情指數之前移相關

前引月數	計算中包括之月數	相關係數
0	138	.55
1	137	.65
2	136	.70
3	135	.73
4	134	.76
5	133	.76
6	132	.76
7	131	.74
8	130	.71
9	129	.67
10	128	.61
11	127	.54

樣方法，一一計算其相關，所得結果，總括於表 102。表中結果可製成圖 52。從圖上觀察係數之值，最大爲 .76，此時之前

圖52　實業股票價格與一般商情指數之前引相關



引爲4，5，6，各月，因取其平均時期——5月——，即定爲前引之時期。

第六章

經濟預測

統計方法之應用於經濟方面，集中在預測上，這一句話並非張大其詞。回歸方程式，標準差誤，相關係數等測量之有價值，亦正因此種量數，可用於實際問題上，以決定最最或然的生產，最最或然的物價，和最最或然的經濟變動等，統計技術之價值，亦卽建築在預測之能實際應用。雖則尙有很多學者對預測特反對論者，但是大都由於技術上之質難，不是預測的本身，預測本身自有其不可泯滅的價值。

統計學中經濟預測之方法甚多，研究現象之長期趨向，季節變動與循環變動，均可作預測之根據。此外更完善而有效之預測方法，係根據現象間已往之關係，作預測將來之探求，此法可分兩類。

一、混合指數預測法

二、回歸公式預測法

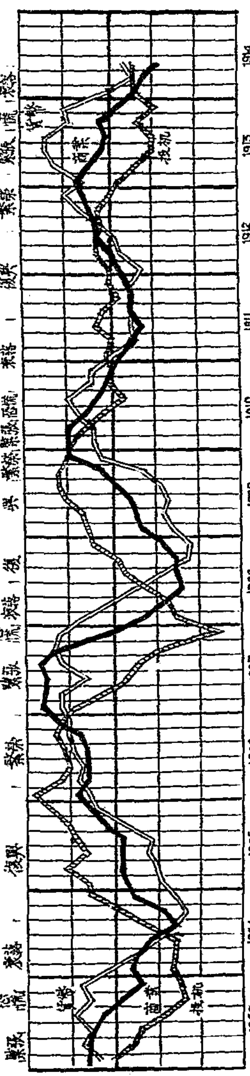
第一類方法根據各種經濟現象變動方向之關係，故祗能示人以變動方向。第二類方法根據現象的數量大小之關係，是故除示人以變動方向外，且能示人以變動之範圍；故優於第一法

Indeterminate	無定性
Index number	指數
Inertia of large number	大數惰性
Interpolation	補插法
Inverse correlation	負相關
J shaped curve	J 形曲線
Lag	落後
Lag correlation	落後相關
Lead	前引
"Less than" cummulation	以下累積
Linear correlation	直線相關
Line of regression	回歸線
Line of means	平均數線
Link relative	環比
Long time change	長期變動
Lorenz curve	羅倫曲線
Lower limit	下限
Magnitude	數量
Mean difference	均互差
Mean deviation	平均差
Mean square deviation	均方差
Measure	數量
Median	中位數
Mode	衆數

Moment	希望差
"More than" cummulation	以上累積
Most probable value	最最或然之數値
Moving average	移動平均數
Multimodality	多峯性
Multiple correlation	複相關
Multiple frequency table	複次數表
Mutually exclusive event	不能並立之事實
Net correlation	淨相關
Normal curve of error	常態差誤曲線
Normal equation	標準方程式
Observed variable	觀察之變量
Ogive	弧形
Ordinate	縱線,縱距
Origin	原點
Parabola	拋物線
Partial correlation	局部相關
Partition value	分割數
Pecuniary units	貨幣單位
Percentile	百分位數
Periodic movement	週期變動
Permanence of small number	小數恆性
Physical unit	體量單位
Pictogram	形圖

圖53　哈佛一般商情指數，1903—1914

試驗時期之指數（雙月）



A 曲 線（投機事業）
1 紐約各銀行之清算，
2 公債、實業股票、鐵路股票之價格，

B 曲 線（商 業）
1 生鐵產額、及紐約市外銀行之清算，
2 潑來特斯屈利報(Bradstreet)批發物價指數、及美國勞工統計局之批發物價指數，

C 曲 線（金 融）
1 紐約銀行放款及存款額，
2 四個月至六個月及六十天至九十天商業票據之利率，

在1919年有幾種材料曾經放棄，如生鐵產額及放款等，其餘的亦改變方式，於是其配合指數有如下列之構造。

A 曲 線
1. 紐約銀行之清算
2. 紐約市之動產售價
3. 實業股票之價格

B 曲 線
1. 紐約市外銀行之清算
2. 潑來特斯屈利報物價指數

C 曲 線
1. 四個月至六個月商業票據之利率
2. 六十天至九十天商業票據之利率

1919年的規定一直維持到1923五月，以後又改如下：

A 曲 線 第1.及第2.類歸併成一類，即關於紐約各銀行之欠項。

B 曲 線 紐約市外各銀行之清算，改成紐約市外140

銀行之欠項。放棄潑來特斯屈利報簡單物價指數。而用十種感應最靈敏之貨物之價格指数。但在1926年又改爲美國勞工統計局之價格指數。

A.B.二曲線以後經過無數之改變，尤其在1927年。

A曲線：　紐約各銀行之欠項取消，而代以四十種實業股票，及三十種鐵路股票Barron之加權價格指數。

到1928年，B曲線又有更動，原有的紐約市外140家銀行之欠項，只存133家了。

此種預測雖經變遷，然於1928年九月二十二日 Weekly Letter 認爲三大類測驗標準其各種不同材料的彙集，從不能完備，因美國的工商業組織太複雜，所以使各月的指數難於應用。故近來經濟恐慌，絲毫未能測定，而束手無策。

由1928十月 A曲線按照1928八月之訂定加以長期趨勢的修正，用1920年一月至1926年十二月爲基期，代替以前的1919七月至1925六月的基期，以使A的水準下降。

其現況如次，其曲線情形見圖54。

A曲線　股票平均價格

Dow-Jones　20種鐵路股票

Dow-Jones　30種實業股票

紐約證券交易所指數第二號

債票價格：Dow-Jones 40種債票

商業票據 紐約交易所公債

B 曲線 紐約市外各銀行債項 已修正

美國勞工統計局物價指數

感應敏銳之物價指數

車輛運輸 總計 已修正

商品 已修正

雜項 已修正

電力生產 已修正

商業倒閉 Dun Bradstreet 滂來特斯屈利指數 已修正

建築合同 Dodge 已修正

房屋建築准許數 滂來特斯屈利指數 已修正

百貨商店銷售數 聯邦準備局修正數

總輸出 商品 未修正

已修正

普通輸入 商品 未修正

已修正

製造品 聯邦準備修正指數

鋼錠

織品

自動車

礦　　聯邦準備局修正指數

石油　煤

生石油

工廠　僱傭　聯邦準備局修正指數

C 曲線　聯邦準備銀行

欠單及抵押品

承兌票據 Acceptance holdings

貼現及再貼現

已有抵押品

其他會員銀行

有抵押品之借款

商業借款　未修正

已修正

已有抵押品

總庫存

紐約市外之活期存款

黃金流動

貨幣流動　已修正

證券發行

銀行認支之款項

商業票據利率

短期借款(九十日及四個月者)

銀行認支款項利率

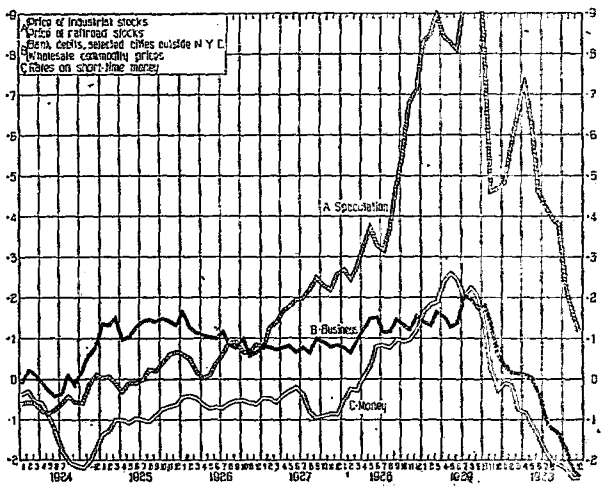
即期借款續訂率

紐約匯兌率

法國(1934年1月30日後　面價6.633)

英國(1934年1月30日後　面價8.240)

圖54　哈佛一般商情指數，1925—1930



哈佛委員會根據1903—1914年之循環研究，連續的ABC，ABC………將無止境，並將經濟循環分爲下列五期。

1. 衰落期

A 曲線(證劵市價)漸漸提高

B 曲線(商品價格)繼續低落

C 曲線(平均利率)作同一之趨勢

2. 復興期

A 曲線　繼續增漲

B 曲線　開始增漲

C 曲線　在復興期末始提高

3. 繁榮期

A 曲線　倘稍稍提高，但不久卽停止

B、C 曲線　繼續增漲

4. 緊張期

A 曲線　下落

B 曲線　極難增漲

C 曲線　特別的高昂

5 恐慌期

A 曲線　下落而達極低限度

B 曲線　開始下落

C 曲線　已達極高度

由此證明，主要預告指數是A曲線(證劵市價)；但是我們

當明瞭下列二點：

1. 據已往的經驗，由交易所來預測，可得正確之結果。

2. 再次如曲線A之趨勢，可以預告B曲線，這不僅由於交易所投機之推進，而在證劵先期的低落，已呈現經濟循環之變遷，尤其金融地位已爲循環所左右。故在繁榮期末，證劵市價不會太高，即受貨幣膨脹之影響，其情形爲：

a. 銀行之流動金已極有限，他們就將商業票據利率漸漸的提高，這是市價低落之一種原因。

b. 銀行因需要流動金而出售其所有之證劵，于是更使轉向低落之趨勢。

c. 投機家之湊成速轉急下之趨勢，他們的投機全靠信用及貸借，如果銀行向他們加重利息，他們就站不住，又是新的降落。在美國如遇特別情形，銀行便限制對經紀人 (Brokers) 的放款。

d. 再次實業家他們也是在繁榮期需要資本，也出賣一切的證劵，又助成阻止市價的提高。

e. 末了是工商業之繁榮，對經濟市場之影響，尤其在紐約市以外之各銀行。在繁榮期這些銀行需要流動金，向紐約銀行提取所存之款，紐約各銀行流動金減少而限制放款，投機減小，證劵行市下跌過剩，而向紐約各銀行存儲。轉而入於投機市場，證劵市價漲高。

在這幾點實可以預測某項工業趨勢，預測的操持者、不僅是投機家，大概是一般熟悉業務之人，如企業家，實業家，這般人完全知道何時繁榮期已達末日，所以常常在價格尚增漲未已，一般民衆多想其增漲還將延長些時日，可是業務場中人，已覺處于傾覆的地步，以趨利遠害。

據哈佛委員會之意見，我們研究須注意 A, B, C 三曲線之整個關係，由曲線 A 固能測知 B，而曲線 C 亦能測知曲線 A。因三曲線之次序為 ABC，ABC，ABC，…………永久是蟬聯賡續，毫無止境。設曲線 C 下降，我們就可預知卽將有一新循環開始，而曲線 A 卽將上昇。不過單知曲線C 下落，還未可斷定曲線 A於若干時後上昇，曲線 C 上昇時，亦復如此，未免缺點耳。一九二二年一月 Person 教授曾在經濟統計雜誌發表意見，由曲線 C 之下降與下期曲線A之上升，我們不能單注意圖上兩線相隔的橫坐標距離，更應注意其昇降之高度，才能預測曲線 A 變動方向之時期。當貼現率由最高點下落 1 ¾單位，我們可推知曲線A 不久卽將上昇或正在上昇；反之若貼現率已從最低點上昇 1 ¾單位時，可推知曲線不久卽將或正在下降。設投機者於曲線 C業已從最高點下落 1 ¾時，購入證券，至曲線 C 上昇至1 ¾時售出，必可獲利。但此種關係，以 1884—1913 年之事實證明之，固合符節，當其發現之時，已在事後，此種關係，不復成立。

回歸公式預測法

簡單相關及複相關之回歸方程式已於第二編相關各章中，詳加討論，其式如下

$$Y=a+bX$$

$$X_1=a+bX_2+bX_3+\cdots\cdots\cdots\cdots+bX_n$$

要使預測具有眞正的價値，務取各自變量之狀態，先於各倚變量之狀態，如 $X_2X_3\cdots\cdots X_n$，的狀態先於 X_1 的狀態，玆如舉行預測物價狀態，必根據收穫；或收穫狀態，必根據雨量及成長期幾月之溫度等。

莫爾教授H.L. Moore所著的"Economic Cycles"(1914)，爲首先引入注意到這類的預測的作家，莫氏根據收穫狀況計算玉蜀黍之價格，造成回歸公式如下：

$$Y=-0.89X+7.8$$

上式中 Y 代表價格而 X 爲收穫，所得相關甚高 $r=0.79$，於是莫氏以爲可藉用這種方程式，以行預測。其中 X 及 Y 是代表玉蜀黍本年之增減之百分比。

例如在1911年美國玉蜀黍的產量爲2,531百萬籮 Bushels，而1911年12月1日之價格爲0.618元。在1912年的生產爲3,125百萬籮Bushels。

依據回歸方程式預測1912年十二月之價格，則得$\frac{3125-2531}{2531}$

$=\frac{594}{2531}=23.44\%$表示增加百分之23.44。以之代入公式

$$Y=-23.44\times0.89+7.8$$

$$Y=13.06\%$$

故在1911年十二月一日之價格爲0.618元，應在其中除去13.06%，其價爲0.537元，卽1912年十二月一日之價。

$$0.618\times\frac{100-13.09}{100}=0.618\times\frac{86.91}{100}=0.537\text{元}$$

實際上當時之市價爲0.487元，其與實際之差爲5分。就標準差而言，在此例中之差誤爲15.92% 距理論差誤甚遠。

於是根據收穫狀況預知價格，略含有某種的機誤。然據理論上在相當對稱分配，兩倍標準差的機誤有百分之九五（95%）。

在"Economic Cycles" 以後出版之 "Forecasting in the Yield and the Price of Cotton (1919)" 書中Moore 教授說明美國爲產棉之國，應重視棉花之生產，而將此業施行預測。

Moore 教授認爲農業生產，與溫度及雨量有相關，故其棉花產量預測中，選取溫度及雨量爲自變量，棉花產量爲倚變量，於是他使用三變量的三元複回歸方程式。但研究此類的事件，公式上很少採用原始之數字；統計數列常預先經過某種的手

績。Moore 教授之研究，將溫度，雨量，收穫，計算爲百分比的數列，卽該年之數字，對前三年的平均數的百分比。於是如三年1892—1893—1894的平均收成爲每畝 150 公磅，而1895年的收成爲 152 磅，則1895年的數字，爲$\frac{152}{150}=101.3\%$

Moore 教授曾計算1892—1914期間的數字，但他的調查只及於某種棉產區，其計算每畝收穫量與六月間溫度的相關係數爲 $r=+0.55$，因此認定棉花亦如玉蜀黍對於六月所需的溫度爲高溫度。

另一方計算每畝收成與五月之雨量的相關係數爲 $r=-0.41$，那麼五月的雨多，可損壞玉蜀黍的收成。

玉蜀黍要有好收穫，必須五月一個月的乾燥及六月一個月的炎熱。

在 Moore 教授的方程式中，X_1 代表倚變量（每畝收穫）X_2 爲五月的雨量及 X_3 六月的溫度。（應用公式時 X_1，X_2 及X_3 均須依照上列方法化作百分比）

於是方程式爲：$X_1=a+bX_2+bX_3$

計算的結果　　$X_1=-95.12-45X_2+2.003X_3$

Moore 教授對於六月所求出之結果，勝過於美國農業部同時期之報告，他差不多可合於官方後兩個月所求得者，卽九月之報告。

其複相關係數爲 $R=0.58$。爲改善他的公式，Moore 教授

改用四元方程式，即加取有關之八月的溫度，他得到複相關係數爲 $R=0.73$ 更加滿意了。此種預測只能行於八月，即在收穫之前，而其結果比較官方八月的預測，就是九月的預測，都要好許多。

Moore 教授同樣從事求預測棉花之價格。施行他的預測先據一自變量，即生產，其回歸方程式有兩變量，即計算本年與上年增減之百分比而得方程式 $Y=8.81-1.08X$ 關於棉花價格及其收穫之相關係數爲 $r=-0.82$。

準此計算之曲線，雖不密合於實際之曲線，但很接近。

爲改良此結果，Moore 教授對於生產上加一新的因子：一般物價指數，而計算棉花，一方面根據棉產的收穫，一方面根據一般的物價，於是方程式有三個變量了，其複相關係數成爲 0.86 較之兩變量的 $R=0.82$，比較進步。

此類方法應用於其他新範圍，現已很多，總之，如能應用合法，誠有無上之價值。