

統
計
方
法

統計方法

STATISTICAL METHODS

美國 F. C. MILLS 著

李黃孝貞 譯
陸宗蔚

中華書局印行

譯 本 序

美國密爾斯教授 (Prof. F. C. Mills) 所著統計方法 (Statistical Methods Applied to Economics and Business) 爲研究統計學者最適用之初級教本，出版已十有餘年而聲價勿替，吾國各大學亦多採用之。孝貞宗年來曾數度講授是書，爰爲合譯，以貢其數學相長之一得。

是書着重於經濟及商業統計方法之研討。其優點有二：一、理論精闢，無蕪雜膚淺之弊，並注意於應用所設實例，足爲舉一反三之助；二、對於統計量數 (statistical measures) 之意義及其應用之方法，敘論不厭其詳，而對於各種量數在應用時所受之限制，尤再三致意，良以統計量數多係根據數學公式，應用時必須合於若干條件，並非任何資料皆可適用，且對於結論之解釋，亦常受資料性質上之限制，此在經濟及商業統計爲尤然。英人恆言所謂統計可以證明任何事物者，蓋指統計資料之解釋而言，解釋苟不得當，則數字斑斑，具文而已，何足以語於真理之左證哉。

譯本係根據 1935 年九月版之原本，分任譯述，初稿各自起草，脫稿以後，互爲校勘，以求一貫。行文雖取直譯法，然艱澀語氣亦所力避。其中術語大體以書末所載中國統計學社編訂之統計譯名爲藍本。

李黃孝貞
陸宗蔚

一九四一年九月一日



原 著 序

近十年來經濟界與研究社會科學者，對於數量法 (quantitative methods) 之應用，漸見深切之注意。在昔恃個人之智慧以解答經濟上之問題，或憑未證實之假說以論斷社會問題者，已不復見於今日。社會學家及嚴正之經濟學者亦步自然科學家之後，基於事實，以施其觀察與分析，靡然成風，遠非古昔所能企及。此種觀察既屬於數量性質，則欲整理之，解釋之，自當應用適當方法。本書即為論述前項觀察之各種整理與分析之方法而作，而尤偏重於實際經濟資料之研究。

本書之論述，必欲限於經濟方面者，其故有二：一、普通統計方法之應用雖廣，然任何方面之研究，各有其特殊問題，而尤以在經濟方面所遇之困難與特殊問題為多，於是切應此種特殊需要之方法，應運而生，欲以此種方法筆之於書，則此書所及之範圍，固不得不加以限制。二、方法之研究，如以專門論題為例證，解釋當愈顯明，蓋抽象之論述，每為讀者所厭倦。本書之專為適合經濟界從事數量研究者之需要而作者，職此之由。

本書於方法之闡明，不厭其詳，而於純屬數學方面之演算，則務求簡略。又本書係備學者研究之用，非供斯學已有造詣之教師閱讀，故凡所解釋，概以適切前者之需要為準。此書為統計學之入門，非窮究高深學理之作，對於意義深奧之定理，其解釋每略而不詳。

今日一般所用之數量分析法，乃係積聚多年來各界人士之貢獻，逐漸推演之結晶。對於統計學之發展，曾有相當貢獻者，在本書未遑遍舉，書中若干處，對於前人雖間有徵引，然近世統計學家所享受前人之功業，實非著者數言之所能盡。

本書各種資料之蒐集，備承各方之贊助。第十一、十三、十四、三章內之一部分統計資料，為安得孫(H. E. Anderson)、凱洛(H. B. Killough)

兩氏所供給。哈佛大學經濟研究委員會潘遜斯教授(Prof. Warren M. Persons) 慨允將彼對於物價指數研究所得之結果，刊入本書。第七章內之建築費指數 (the index of building costs) 與第九章內之商情指數 (the index of business conditions) 係美國電話電報公司統計部所編製。此項資料之採入，事前嘗得該部主任安德魯氏(Seymour L. Andrew) 之特許。本書油印初稿中頗有意義晦澀之處，賴加利福尼亞大學毛勃雷教授(Prof. A. H. Mowbray) 之指示，得以修正。他如哥倫比亞大學摩爾(Prof. Henry L. Moore)、布朗 (Prof. Theodore H. Brown) 兩教授暨經濟學院舒爾茲氏 (Henry Schultz) 將原稿一部分為之校訂，均所心感。塔特氏 (Herbert F. Tutt) 擔任統計圖表之繪製，襄助良多。在編印本書期間內之各種工作，如資料之蒐集，統計圖之繪製，以及印刷事務之管理等，隨時得達文波特氏(Donald H. Davenport) 之襄助。對於余妻所給與不斷的助力，更為銘感。

密爾斯(Frederick Cecil Mills)

一九二四年十一月

311.2
701
2

統計方法

目次

譯本序

原著序

統計圖索引

第一章 統計方法與經濟及企業問題I

企業活動之類別——經濟與企業問題之數量性質——
統計方法與企業內部管理問題——統計方法與企業本
身以外之問題

第二章 圖示法6

垂直之縱橫坐標——自變量與倚變量——函數關係——
一直綫——非直綫關係——對數——對數方程式——
雙對數與單對數圖——統計圖之功用——選擇圖形之
條件——表示時間數列之圖解——比率圖之優點——
比較各組次數之圖解——表示成分之圖解——累積圖
——甘梯氏進行圖——製圖之標準規則

第三章 統計資料之整理：次數分配52

未經整理之資料——整列——次數表——編製次數表
之步驟——組距之大小——組限之位置——組限與觀
察之確度——其他要件——統計表之結構——次數分
配之圖示——曲綫之修勻——連續數列與間斷數列——
統計資料之累積排列法——累積次數曲綫——累積

1 4 2 7 8 8

曲綫與普通次數曲綫之關係

第四章 次數分配之表述：平均數82

次數分配之比較——經濟資料之分配——次數分配之一般特徵——表述次數分配之方法——平均數——算術平均數之計算——計算算術平均數之簡捷法——中位數之決定（根據未經分組之資料）——中位數之決定（根據已經分組之資料）——四分位數與十分位數——衆數之決定——由算術平均數及中位數決定衆數之方法——按圖決定衆數中位數四分位數及十分位數之方法——幾何平均數——幾何平均數之特質——幾何平均數表示集中趨勢之討論——倒數平均數——各種平均數之相互關係——各種主要平均數之特質

第五章 次數分配之表述：離散度與偏斜度之量數·124

離散度之性質與意義——絕對離散度之量數——全距——平均差——標準差——四分位差——機誤——各種離散度量數之相互關係——各種離散度量數之特質——相對離散度之測度——離散度係數——偏斜度之量數——峯度

第六章 物價指數143

指數之性質——物價之變動——普通批發物價指數之目的——物價比率之次數分配——物價變動率之平均問題——編製指數之各種方法——符號之意義——簡單物價指數——實價綜合法——比價之算術平均數——比價之簡單平均數含有加權之意味——比價之中位數——比價之幾何平均數——比價之倒數平均數——各種簡單指數之比較——時間顛倒測驗法——指數之

加權問題——實價加權綜合法——比價之加權算術平均數——比價之加權幾何平均數——因子顛倒測驗法——理想公式——各種指數之可靠性——關於編製物價指數之其他問題——指數包含物品之項數問題——物價制度之原素——批發物價之價格分類

美國之批發物價指數：

美國勞工統計局指數——聯邦準備局指數——戰時工業局指數——勃蘭特斯脫批發物價指數——滕氏批發物價指數——費暄氏每週批發物價指數——潘遜氏測度商情循環之物價指數——其他批發物價指數——各指數編製方法之差異——概略：批發物價指數
其他物價指數——零售物價指數——生活費指數——農產品之價格指數及其購買力指數——貨幣工資及真實工資之指數

第七章 時間數列之分析：長期趨勢之測度 ………210

時間數列之初步整理——時間數列之圖示——時間數列所受各種勢力之影響——長期趨勢——週期變動——意外變動——長期趨勢之測度——移動平均數法——移動平均數之特質——配合數學曲綫法——直綫之配合：最小平方法——直綫之配合：特例——遞升冪級數式曲綫之配合——測度商業倒閉家數長期趨勢之實例——配合直綫之演算——配合二次拋物綫之演算——配合三次拋物綫之演算——各種長期趨勢綫之比較——對數應用於曲綫之配合——配合曲綫之簡捷法：平均數法——選點法——長期趨勢綫之選擇——用關聯之統計數列代表長期趨勢之方法——分析時間數列

所需平減價值之手續

- 第八章 時間數列之分析：季節變動及循環變動之測度 264
- 每月項目平均法——環比中位數法——移動平均數法——實際數值對長期趨勢值之比率平均法——各種季節指數之比較——各月常態數值之計算法——循環變動之測度——分析時間數列之要點
- 第九章 物量指數 290
- 戰時工業局之物量指數——台氏之物量指數——修整之物量指數——根據月數字編製之物量指數——聯邦準備局之基本工業生產指數——美國電話電報公司之一般商情指數——貿易數量指數
- 第十章 關係之測度：直綫相關 307
- 繳納個人所得稅人數與汽車登記數之關係——標準誤之計算——估計方法——相關係數——相關量數計算手續之詳述——相關表之編製——相關量數之計算——計算相關係數之積差公式——積差法之應用（根據未分組之資料計算）——積差法之應用（根據已分組之資料計算）——迴歸綫——迴歸方程式之應用——概略——相關量數應用上之限制
- 第十一章 時間數列之相關 348
- 循環變動相關之測度——測度時間數列相關之困難——相關係數及時間次序之測度——股票價格循環變動與一般商情循環變動之關係——移動平均數之應用於計算時間數列循環變動之相關——短期變動之相關
- 第十二章 關係之測度：非直綫相關 367

拋物綫相關——相關指數——相關指數之意義——計算相關指數之簡捷法——相關率——相關率之計算——相關率之校正——相關率與相關係數之關係

第十三章 關係之測度與估量之問題384

估量所含之假定——估量問題之實例——對數表示之估量標準誤——估量標準誤之解釋；估量區域——比率表示之估量標準誤——估量標準誤之應用——對數表示之相關指數——倒數之應用於相關之測度——倒數表示之估量標準誤及相關指數——倒數表示之估量標準誤之意義——由真數對數及倒數求相關量數之比較——選擇相關量數之原則——用真數對數及倒數表示各種量數之比較——估量區域及其意義

第十四章 關係之測度：複相關及淨相關411

玉蜀黍收穫量與溫度之關係；初步分析——由三個自變量估量玉蜀黍收穫量之方法——常態方程式之形式及其解法——估量標準誤之計算——複相關係數——複相關量數之比較——計算複相關方法與直綫相關之關係——計算複相關方法之應用——淨相關之意義——淨相關與單相關之區別——淨相關之測度方法——由迴歸係數與相關係數之關係測度淨相關之方法——常態方程式之解法——淨相關係數之計算——計算淨相關係數之又一法——一次係數之計算——二次係數之計算——測量離散度之量數——測度相關方法應用上之限制

第十五章 機率及差誤常態曲綫437

機率之簡單定理——簡單機率之舉例——機率之相加

——機率之相乘——兩項展開式與機率之測度——實在次數與理論次數之比較——差誤常態曲綫——應用於經濟資料之舉例——次數分配之動差——斷定曲綫形態之標準——常態曲綫之配合；縱坐標表之應用——曲綫配合適度之測驗——理論次數之斷定——抽樣標準誤之應用於測驗曲綫配合之適度——測定曲綫配合適度之 χ^2 測驗法——關於表述次數分配之註釋

第十六章 統計歸納及抽樣問題464

統計表述及統計歸納——根據統計分析之結果作概括論斷之問題——自然現象劃一性之假定——樣本代表性之重要——簡單抽樣之條件——可靠程度量數之意義——主要統計量數之標準誤——可靠程度之量數在應用上之限制

附錄 A 最小平方法在統計上之應用475

常態方程式之求法——常態方程式之組成——常態方程式之標準形式——估量標準誤公式之求法——相關指數公式之求法——特殊之例——組成常態方程式之校核法——他種校核法——常態方程式之解法——杜立特氏之解法

附錄 B 統計符號彙解493

參考書目錄499

統計譯名(中國統計學社審定)505

統計圖索引

圖一	垂直縱橫坐標圖中點之位置	7
圖二	1923年各月 <u>美國</u> 棉花價格	8
圖三	方程式 $y=x$ 之圖解	11
圖四	方程式 $y=3x+2$ 之圖解	12
圖五	拋物綫:方程式 $y=x^2$ 之圖解	14
圖六	等邊雙曲綫:方程式 $y=x^{-1}$ 之圖解	15
圖七	指數曲綫:方程式 $y=2^x$ 之圖解	16
圖八	正弦曲綫:方程式 $y=\sin x$ 之圖解	18
圖九	方程式 $\text{Log } y=2 \log x$ 之圖解	23
圖十	方程式 $y=x^2$ 之圖解(繪於雙對數尺度之圖紙上者)	24
圖十一	複利法則:本金十元複利六厘計息百年間之本利和(繪於算術尺度上者)	25
圖十二	複利法則:本金十元複利六厘計息百年間之本利和(繪於單對數尺度,即比率尺度上者)	26
圖十三	1901—1923年 <u>美國</u> 麵粉輸出數量	28
圖十四	1896—1922年 <u>美國</u> 鋼鐵產量(繪於單對數尺度上者)	29
圖十五	1910—1923年 <u>美國</u> 亞根公司之全國銷貨總額及某數州之銷貨額(繪於算術尺度上者)	30
圖十六	1910—1923年 <u>美國</u> 亞根公司之全國銷貨總額及某數州之銷貨額;附增加減少及比較之尺度(繪於單對數圖紙上者)	31
圖十七	1920年 <u>新</u> 英格蘭各州農場數	34

圖十八	1923年各月某公司生產費用之分析	36
圖十九	1924年 <u>美國斯彼德革爾汽車製造公司</u> 預計累積生產量與 實際累積生產量之比較	37
圖二十	預計生產量與實際生產量之比較： <u>甘悌氏</u> 進行圖（表示 二月二十八日爲止之生產狀態）	39
圖二十一	預計生產量與實際生產量之比較： <u>甘悌氏</u> 進行圖（表 示九月三十日爲止之生產狀態）	39
圖二十二	直方圖：按某週工資數分組之工人 210 人之次數分配 （組距 = \$2.00）	62
圖二十三	直方圖：按某週工資數分組之工人 210 人之次數分配 （組距 = \$1.00）	63
圖二十四	直方圖：按某週工資數分組之工人 210 人之次數分配 （組距 = \$.50）	64
圖二十五	直方圖：按某週工資數分組之工人 210 人之次數分配 （組距 = \$.25）	64
圖二十六	次數多邊圖：按某週工資數分組之工人 210 人之次數 分配（組距 = \$2.00）	65
圖二十七	次數多邊圖：按某週工資數分組之工人 210 人之次數 分配（組距 = \$1.00）	65
圖二十八	次數多邊圖：按某週工資數分組之工人 210 人之次數 分配（組距 = \$.50）	66
圖二十九	直方圖：1918年 <u>美國</u> 個人收入之次數分配；包括收入 之在\$4000以下者（組距 = \$500）	70
圖三十	直方圖：1918年 <u>美國</u> 個人收入之次數分配；包括收入 之在\$4000以下者（組距 = \$200）	70
圖三十一	直方圖：1918年 <u>美國</u> 個人收入之次數分配；包括收入	

	之在\$4000以下者(組距=\$100).....	71
圖三十二	次數曲綫圖:1918年 <u>美國</u> 個人收入之次數分配;包括收入之在\$4000以下者(根據組距為\$100之直方圖繪成).....	71
圖三十三	累積次數曲綫圖:按耐用時期長短分組之電桿木分配(向上累積式).....	77
圖三十四	累積次數曲綫圖:按耐用時期長短分組之電桿木分配(向下累積式).....	77
圖三十五	1920年按勞工費用分組之 <u>美國</u> 鋸木廠次數分配;說明累積次數曲綫與普通次數曲綫構造上之區別.....	80
圖三十六	次數曲綫圖:按身材高度分組之兵士18,780人之次數分配.....	83
圖三十七	次數曲綫圖:天文觀察差誤之分配.....	83
圖三十八	礮彈發射時彈丸之分佈區域;表示彈丸之理想百分分配.....	85
圖三十九	直方圖:一礮發射彈丸千枚之次數分配.....	86
圖四十	次數多邊圖:擲錢試驗中錢幣正面向上枚數之次數分配.....	87
圖四十一	次數多邊圖:某年批發物價比前一年漲落百分率5540項之次數分配.....	88
圖四十二	次數多邊圖: <u>倫敦</u> 對 <u>紐約</u> 匯價之次數分配.....	90
圖四十三	差誤常態曲綫.....	91
圖四十四	根據未經分組之資料決定中位數位置之圖解(七人之個人收入金額).....	102
圖四十五	按某週工資分組之工人次數分配之修勻曲綫;表示算術平均數中位數與衆數三者關係之圖解.....	111
圖四十六	根據某週工資之累積次數曲綫決定中位數及四分位數之圖解.....	113

圖四十七	次數多邊圖：1914年物品 346 種之比價分配（1913年 平均價=100）.....	152
圖四十八	次數多邊圖：1900年物品 222 種之比價分配（1890年 平均價=100）.....	155
圖四十九	次數多邊圖：1918年物品1437種之比價分配（1913年 七月至1914年六月之平均價=100）.....	158
圖 五 十	次數多邊圖：繪於對數尺度上之1918年物品1437種之 比價分配（1913年七月至1914年六月之平均價= 100）.....	159
圖五十一	1910—1923年農產物價五種簡單指數之比較（1910年 =100）.....	171
圖五十二	1910—1923年農產物價四種加權指數之比較（1910年 =100）.....	183
圖五十三	物價制度內各原素間關係之圖解.....	188
圖五十四	1860—1923年 <u>紐約市</u> 各銀行票據交換額及其移動平均 數.....	216
圖五十五	在九點上配合直綫之圖解.....	227
圖五十六	在九點上配合二次拋物綫之圖解.....	236
圖五十七	1897—1921年 <u>美國</u> 商業倒閉家數及其三種長期趨勢綫.....	241
圖五十八	1908—1922年 <u>美國</u> 石油產量及用對數配合之長期趨勢 綫.....	245
圖五十九	1908—1922年 <u>美國</u> 石油產量及用真數配合之長期趨勢 綫.....	247
圖 六 十	1914—1923年各月建築合同之實際價值與平減價值之 比較.....	262
圖六十一	各月蛋價對常態數值百分比之次數分配.....	272

圖六十二	說明每月常態數值計算法之圖解.....	276
圖六十三	全年常態產量每月平均常態產量與每月常態產量之關係.....	278
圖六十四	1913—1923年各月美國烟煤生產量.....	286
圖六十五	1913—1923年美國烟煤產量數字內所含之循環變動及意外變動.....	287
圖六十六	一般商情之組合指數.....	301
圖六十七	表示1921年各州繳納個人所得稅人數與汽車登記數兩者關係之散佈圖及其平均關係綫.....	309
圖六十八	相關表中項目之表列法.....	320
圖六十九	美國聯邦準備銀行與商業銀行貼現率之散佈圖；附相關綫及估計區域.....	325
圖七十	商業銀行貼現率與聯邦準備銀行貼現率之關係（商業銀行貼現率視為倚變量）.....	334
圖七十一	聯邦準備銀行貼現率與商業銀行貼現率之關係（聯邦準備銀行貼現率視為倚變量）.....	336
圖七十二	紐約州十城市工廠工人數與產品價值之關係.....	345
圖七十三	紐約州十一城市工廠工人數與產品價值之關係.....	346
圖七十四	1900—01年度至1922—23年度美國棉花產量及其長期趨勢綫.....	350
圖七十五	1900—01年度至1922—23年度紐約高地米特令棉花價格及其長期趨勢綫.....	352
圖七十六	1903—1923年實業股票價格循環變動與一般商情循環變動之比較.....	359
圖七十七	1903—1914年實業股票價格指數與一般商情指數依各種月份相配法求得之相關係數.....	360

圖七十八	1919—1923年實業股票價格指數與一般商情指數依各種月份相配法求得之相關係數.....	361
圖七十九	表示香草(alfalfa)之收穫與灌溉水量兩者關係之散佈圖及其兩迴歸綫.....	368
圖八十	小麥收穫量與淡氣用量之散佈圖及其迴歸直綫與通過各直行之中點綫.....	377
圖八十一	表示燕麥產量與價格兩者關係之真數迴歸綫及真數估量區域.....	405
圖八十二	表示燕麥產量與價格兩者關係之對數迴歸綫及對數估量區域.....	406
圖八十三	表示燕麥產量與價格兩者關係之對數迴歸綫及對數估量區域(繪於雙對數尺度之紙上者).....	407
圖八十四	表示燕麥產量與價格兩者關係之倒數迴歸綫及倒數估量區域.....	408
圖八十五	擲骰試驗中實在次數分配與理論次數分配之比較.....	444
圖八十六	按通話次數分組之電話用戶次數分配及其所配之常態曲綫.....	454
圖八十七	測度差誤常態曲綫下面積之圖解.....	458

統計方法

第一章 統計方法與經濟及企業問題

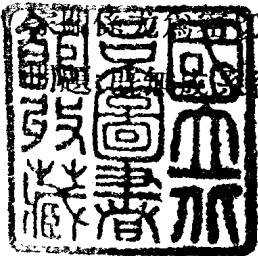
今日普通所用‘企業’(business)之術語,包含人類活動之意義,不一而足。吾人對於處理經濟資料與解決企業問題之方法既欲有所闡明,則對於各種企業之活動,凡可應用統計方法者先當作一簡略之敘述。

企業活動之類別

企業家日常所遇之事件,可分為三類。多數企業家所從事活動之範圍,儘有僅限於其中一類者,然企業界應須處理之各問題,每兼三類而有之。

第一類為發生於生產程序方面者,如關於物理、化學、工程學、牧畜學、航海術等之問題。解答此種問題所需之基本專門智識,為吾人經濟生活之基礎。此類活動乃屬於處理原料、控制天然力之區域者也。

第二類為關於企業個體內部組織與管理方面之活動。凡管理關於滿足人類慾望一切有機與無機物質之職務,各有其執行管理之單位,如農場、礦山、工廠、鐵道、百貨商店等是。企業家在組織此種單位,調劑其各部分間之工作,或監督每個組織內各員工之日常工作時,即遇種種新問題。此種問題縱不及前述第一類生產技術問題之緊要,但與一般企業
科學方法對於解決此種問題,應用較難,且關於此種
統之智識,以為引導,又乏有訓練之專家,以託付解答之



職責。

前述兩類之經濟活動，性質較爲集中，且可受人力之控制。如鋼鐵製造家有其熔鍊鋼鐵之技術問題，與其管理上之各種職務；農夫、礦主所遇問題亦同。凡執行此等業務，雖各有其特殊問題，惟此各種問題之原素，多少皆受人力之控制；故縱有困難發生，亦爲普通企業上本身之困難，既非由於問題所含原素之驟變，亦非由於新因素之驟然滲入，蓋可斷言。

至於第三類之經濟活動，其特殊之點，在於其所含之原素非任何個人所能控制。此類活動包括商品之買賣及與物價有關之一切經濟活動。在現時經濟組織之下，此種活動爲企業家所視爲最重要之業務，蓋企業家經濟活動之最終目的，在於銷售其產品以獲取贏利，產品銷售前之一切預備手續，如生產之技術、與夫企業本身之組織與管理，不過爲達到此最終目的之階梯。且企業家在賣買商品之時，恆遇種種問題，其要素非人力所得控制，如購置原料、採集生產必需之其他要素、與銷售製品時，企業家必須與各種市場——商品市場、勞力市場、金融市場等——互相交易，而在此等市場之中，其物價制度 (the system of prices) 實非企業家所能左右。故對於第一二兩類之企業活動，企業家雖常能控制，然對於處理其最後與其最重要部分，即以善價出售其產品之問題時，其控制力漸見薄弱。夫企業活動之原動力，在於獲得金錢之贏利，而欲獲得金錢之贏利，在於商品買賣之得宜，而欲求商品買賣之得宜，須在人力不能控制之物價制度中，攫取優先之地位，此物價制度所以爲近代企業最主要之因素也。

近代企業家實處於物價環境 (an environment of prices) 之中。此處‘環境’之一名詞，非不切之喻；此企業家所服務之物價世界，係各個獨立部分所組成之互相團結、互相關聯之制度 而又爲包括企業上一切活動之制度。此種制度既非人力所能控制，則惟有適應之、瞭解之、以爲活

動之南針，苟對此茫無所知，則企業上之主要問題，將無法解決矣。

經濟與企業問題之數量性質

各種企業問題之屬於上述第一類者，其性質大率為數量的，久為吾人所公認。此種問題須賴自然科學界所應用之精密方法，以求解決，即其他兩類之純屬於經濟及企業之問題，亦有應用數量法之必要，蓋在解答此種問題時，雖常包含品質上與不涉量度之觀察，而此種觀察大抵可以數量為基礎。各種事實經數量上之比較，即可作為處理企業上問題與解釋經濟理論之根據。統計方法實即以數量為主體，而比較事實之方法也。

前節所述之三類問題，其中第二、第三、兩類，屬於本文討論範圍之內。統計方法雖亦可用以解答生產技術問題之一部分，然本書宗旨，不在討論此種問題，惟其他兩類問題——屬於企業個體內部組織與管理之問題及企業家與物價制度發生關係之商品買賣問題——之解答，則統計分析方法，至為適用。

統計方法與企業內部管理問題

一般企業家在其組織之管理方面，必須處理各種數量資料，資料之屬於數量性質者，因其為用各種數量單位以表示者也。彼須計算煤之噸數，煤氣之立方英尺數，工力之仟瓦特小時數 (kilowatt hour)，生鐵產量之噸數，皮鞋出產之雙數，機械小時數 (machine hour) 及人工小時數 (man hour)，以及用貨幣單位所表示之工資、生產費用及商品售價之數。近代企業組織之規模日見擴大，因此處理之資料亦日就繁複，而其真實意義，遂漸難判明。僅憑個人之智力或固定的管理方法，實不能將大量之資料 (masses of data) 作有效的分析，而管理規模較大之企業組織。報償遞減法則 (law of decreasing returns) 之能在企業中

發生作用者，多由於企業管理方面之困難也。

吾人處理大量資料時，首須整理、簡縮、而分析之。吾人須將資料化繁就簡，而後可着手處理；又須將資料加以分析，而後可瞭解問題之要素。統計方法乃為整理、簡縮與分析大量資料之利便而推演者也。

試舉例以明之。設吾人所欲解決之問題為成本之分析 (allocation of costs)，即會計學中所謂成本會計是也。如欲將該問題之各因素，作適當之分析，則非應用統計方法不為功。會計方法祇限於處理金錢方面之事實，對於各項支出之總分析，殊非此法所能盡事，蓋除成本以外，各項支出之本身，亦應作統計分析，以明其變動，比較其今昔，而察其有無浪費之處。又銷貨紀錄 (sales record) 之分析，亦須將資料化繁就簡，以簡單明瞭之形式表出之，而後可斷定其意義。至於市場之分析，進貨紀錄 (purchasing record) 之考查，以及商品之研究，皆須利用數量法，以數量法之應用，不限於任何度量之單位也。在企業內部管理之各方面，統計方法可補會計方法之不逮，增廣管理人員之智識，並可促進企業管理上之效率。

統計方法與企業本身以外之問題

企業家在市場中從事買賣，即與物價制度發生密切關係，故用數量法以分析物價之變動，實為切要之圖。但物價資料至為繁多，欲處理之，必須化繁就簡，且商業具有週期變動之特性，如企業家欲採取與週期變動適應之政策，更宜應用適當之工具，以分析此種現象。此外如關於經濟分配等各種問題，雖與企業家關係較少，而為經濟學家所推重。此種問題，在在與價值、價格等發生關聯，且其解決之途徑亦繫於數量之研究。

然則此種數量法果為何種方法乎？且應用此法以研究之問題，又何以異於他種研究之問題乎？科學之研究，無論採取何種方法，必須經慎

重之觀察、合理之推考、與準確之證實。數量法之異於他法者，即在觀察、推考、證實三方面，皆可以測度(measurement)為依據，故較諸與數量無涉之分析方法，更為確切。一種科學在不能應用數量以前，不論其研究者之悟解力如何靈敏，工作如何勤奮，其觀察與其所得之結論，必缺乏準確性。吾人藉測度法(methods of measurement)始可用精確之單位，分析各問題之原素，實予科學研究以種種利便。數學及其支系之統計與會計，皆為近代經濟學家採用之利器，且因研究機關與研究方法之發展，企業界之實際應用亦漸見推廣矣。

統計家所用之工具僅為切合於特殊研究所用之數學方法。此種特殊研究，在統計方法發達之初期，不屬於經濟方面，而屬於社會、政治、與人體測量方面，而另一方面之發展（關於機率原理 the theory of probabilities 者），則由論理學之領域，而及於賭博。統計方法在昔為用殊狹，今則應用之廣不可同日而語。斯學經相當之演進，應用漸博，獲得良好之結果，在經濟上之應用，僅其一例耳。經濟學家自利用統計方法後，頗能得心應手，而其工作之準確性大見增加；企業方面亦與較為深奧之經濟科學，同受其利。

至於統計方法應用上之限制，本章中可暫置勿論。統計方法之應用固有時而窮，且應用統計方法者每忽視其應用上之限制。關於此點，容俟後文討論，惟此處所當致意者，統計方法僅為一種工具，吾人應善為利用之，由統計分析所獲之結果，應善為解釋之，蓋吾人尚未發明一種完善之自動機，可將混雜之事實投入機械之一端，各種問題之答案即可由另一端得之；換言之，即吾人之理解力與批判力，並不因所用統計研究方法之進步遽可廢棄也。

第二章 圖示法

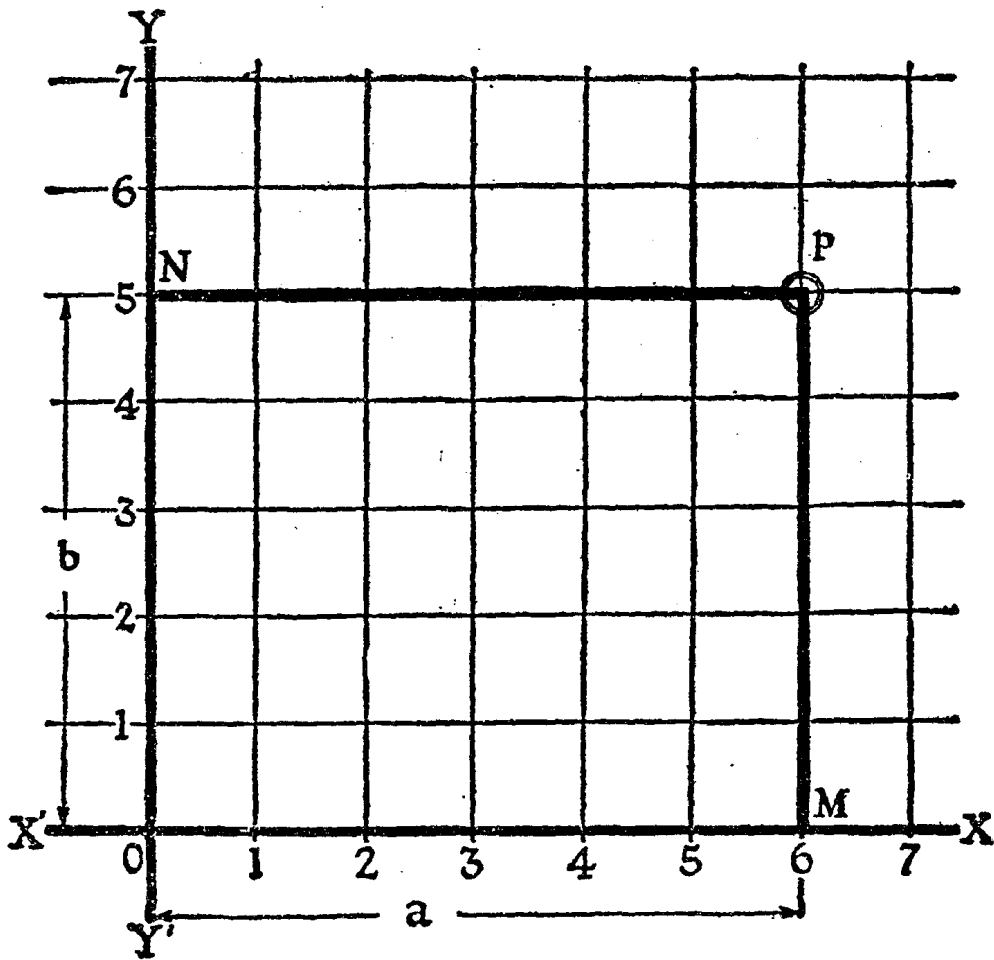
吾人欲將整理、分析及解釋企業與經濟事實之方法作一研究，必先討論若干基本之原則，而此項基本原則涉及統計學之範圍者少，涉及數學之範圍者多。本章爰將若干數學概念予以解釋，後此各章均將時加引證。此種數理雖或為讀者所共喻，在此實有簡略敘述之必要也。

統計分析大抵以基於實際測度而以貨幣或數量單位表示之資料為對象。哲學家笛卡兒氏 (Descartes) 所發明之坐標幾何法 (method of coördinate geometry)，對於前項資料之處理甚切實用。茲將解析幾何學之基本原理，簡述如次。

垂直之縱橫坐標

試於平面上繪一交叉成直角之兩直綫，則在該平面之任何一點，即可依兩綫之交叉點以說明其地位。吾人稱此兩直綫中之一綫 (縱綫) 為 $Y'Y$ ，另一綫 (橫綫) 為 $X'X$ ，其交叉點 (原點 origin) 為 O (參看圖一)。設 P 為平面上之任何一點，試作與 $Y'Y$ 綫平行之 PM 綫，與 $X'X$ 綫交於 M 點；又作與 $X'X$ 綫平行之 PN 綫，與 $Y'Y$ 綫交於 N 點。今假定 OM 之長度為 a 單位， ON 之長度為 b 單位， a 與 b 即為 P 之縱橫坐標 (coördinates)，說明以原點為標準時 P 點之地位。圖中 a 為 6， b 為 5。沿 x 軸 (x -axis) 之距離 a ，稱為 P 點之橫坐標 (abscissa)；沿 y 軸 (y -axis) 之距離 b ，稱為 P 點之縱坐標 (ordinate) (記法應先記橫坐標，後記縱坐標)。在同一平面上其他任何一點之縱橫坐標，均可依此決定；反之，任何兩實數，如以一數作為橫坐標，另一數作為縱坐標，亦可在平面上決定一點之地位。

平面上一點之地位，或在原點 O 之左，或在其右，或在其上，或在其下。通常以在原點右方之橫坐標為正，在原點左方之橫坐標為負；在原



圖一. 垂直縱橫坐標圖中點之位置

點上方之縱坐標為正，在原點下方之縱坐標為負。經濟統計所處理之數值(value)，大率均在圖中右上方之象限(quadrant)內，以該象限內之縱橫坐標均為正也。

此種縱橫坐標之概念為數學之基本原理，亦為統計學上之基本要點。試以表一數字為例以說明此法對於表述商業資料之用途。

此種資料可用垂直縱橫坐標圖表示之。沿 x 軸作為月份，沿 y 軸作為價格，如圖二。定橫坐標時，以一九二三年一月為原點，故一九二三年一月之橫坐標之數值(x -value)為0，二月之橫坐標為1，餘類推。一

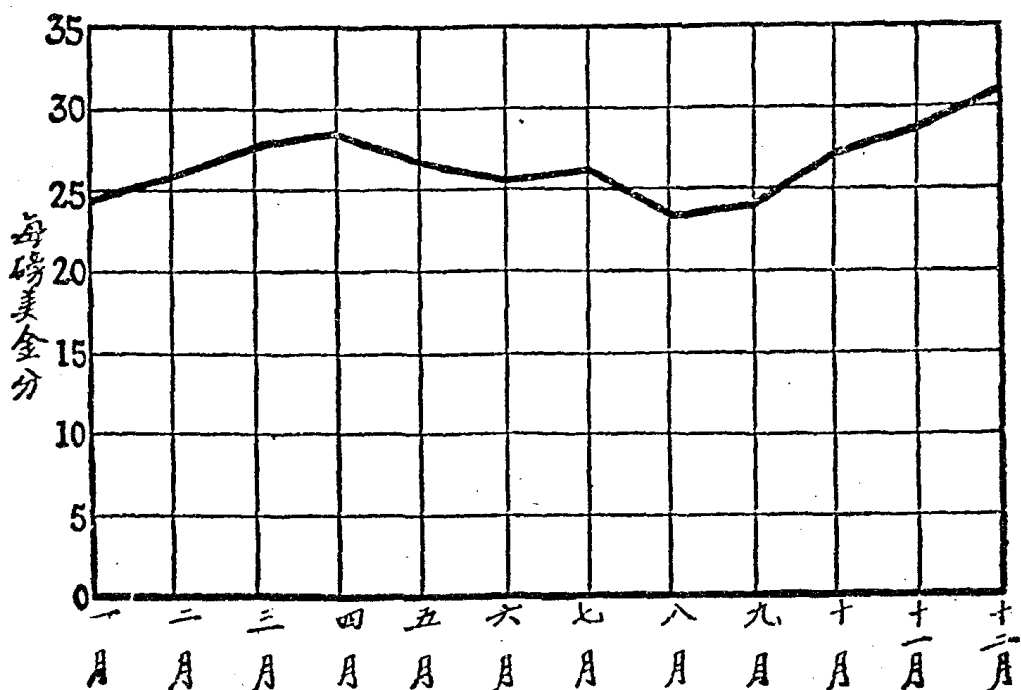
表 一

1923年各月美國棉花價格

(生產者所得之平均價格)

月 份	每磅價格(美金分)
一 月	24.5
二 月	25.9
三 月	27.7
四 月	28.4
五 月	26.9
六 月	25.6
七 月	26.2
八 月	23.5
九 月	24.1
十 月	27.2
十一 月	28.8
十二 月	31.0

(根據 "Weather, Crops and Markets", Dec. 29, 1923)



圖二. 1923年各月美國棉花價格

九二三年一月份棉花價格之橫坐標及縱坐標爲 0, 24.5; 二月份爲 1, 25.9; 十二月份爲 11, 31.0。如用直綫將各點相連, 則對於棉花價格之變動趨勢, 觀圖可易於領悟矣。

自變量與倚變量

依縱橫坐標制, 任何一點之地位必涉及兩數值, 故每點聯合兩個因素而表示其相互之關係。在前例中此兩因素爲月份與棉花價格。棉花價格隨時日之推移而有變動, 其斷續綫 (broken line) 係表明此種變動之方向與大小 (magnitude)。時間與價格皆爲變量 (variable), 蓋其數量均不固定, 而具有變化之特徵。如圖一中橫坐標之值爲 6, 縱坐標之值爲 5, 皆係固定者; 而在圖二中縱橫坐標均不固定, 橫坐標變動於 0 與 11 之間, 縱坐標變動於 23.5 與 31.0 之間。通常以 x 與 y 兩符號表示此兩種變量, 以 x 代表橫軸 (horizontal axis) 上之變量, y 代表縱軸 (vertical axis) 上之變量(註)。

在表示棉花價格與時間關係之圖二中, 時間之變動係任意擇定一個月爲單位, 並假定此時間之因素變動在先, 而後測度逐個時間單位間價格變動之情形。此依定量爲增減之變量—時間—謂之自變量 (independent variable), 恆繪於 x 軸上; 其另一變量, 謂之倚變量 (dependent variable), 恆繪於 y 軸上。相倚之意義, 有真實的, 有形式的。如倚變量之數值可依自變量之數值確定者, 則兩數值之關係爲真實的相倚; 如兩者之間並無此種關係, 則其相倚不過形式的而已。時間之變量, 在製圖時常視爲自變量。

函數關係

(註)英文二十六字母之末數個, 通常用爲代表變量之符號, 字母之前數個則用爲代表常數之符號。常數者謂在某種情形下其值不變之數量也。

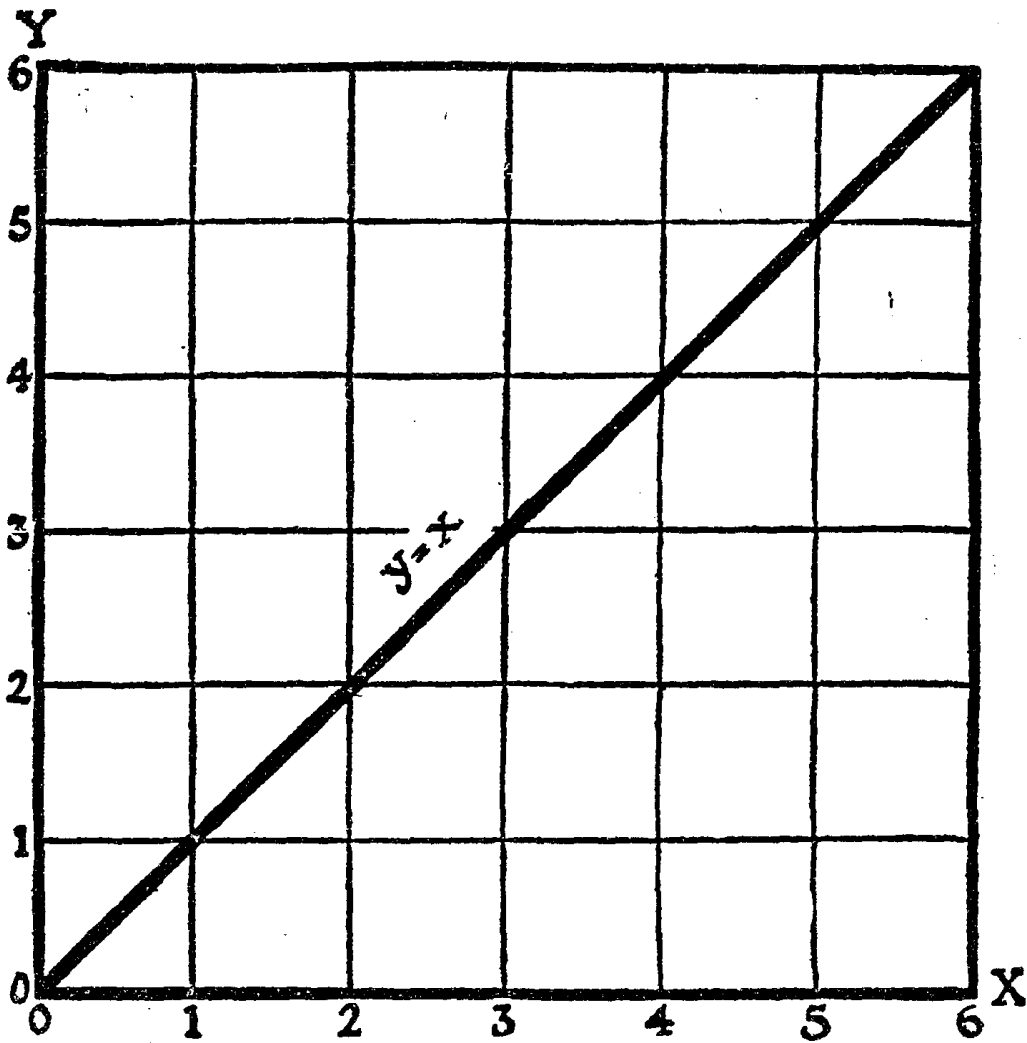
如兩變量之關係完全相倚，即已知 x 值可明確決定 y 值時， y 稱為 x 之函數 (function)。此種關係大率以 $y = f(x)$ 之方程式表示之。例如物體由空中下墜時，其在某時間內之速度為下墜時間之函數；一定容積之氣體之壓力為其熱度之函數；依固定利率計息，其本利和之增加為時間之函數。如將自變量之各數值依直綫圖 (rectilinear chart) 之 x 軸繪之，將其函數 (即倚變量) 之各數值依 y 軸繪之，即得函數之圖解，成一曲綫 (curve) 之形狀(註)。此種屬於函數關係 (functional relationship) 之概念在統計學中甚為重要。茲將較為簡單之各種函數略述之。

直 綫

兩變量之關係如其數值始終相等，則其關係顯為 $y = x$ 之形式。樹齡與樹幹橫斷面所表現輪數之關係，即為其一例。樹木生長二十年時，則樹幹之橫斷面即有二十輪。此種關係可用數對假定之 x 與 y 值，以縱橫坐標圖表示之。將此數點繪於圖中，並繪一綫經過各該點，此綫即為經過原點 (假定縱橫軸所用之尺度相同) 而等分直角 XOY 之直綫 (參看圖三)。

任何一次方程式 (即不含 xy 或一次冪以上之 x 或 y 等項者) 亦可同樣以直綫表示。普通直綫之方程式均可化為 $y = a + bx$ 之形式。式中 a 為常數 (constant)，表示原點及直綫與 y 軸交點間之距離； b 亦為常數，表示直綫之斜度 (slope，即直綫與橫軸構成角度之正切 tangent of the angle)。常數 a 稱為 y 截綫 (y -intercept)。自直綫方程式觀之， x 值為 0 時， y 即等於此常數。前例 (圖三) 中 a 值為 0， b 值為 1。綫之地位係由 a ， b 兩數值之大小及其正負號決定之，故欲決定一直綫，應就所用資料中求 a ， b 之數值。此項問題在統計方法中，性質不一，當分別討論之。

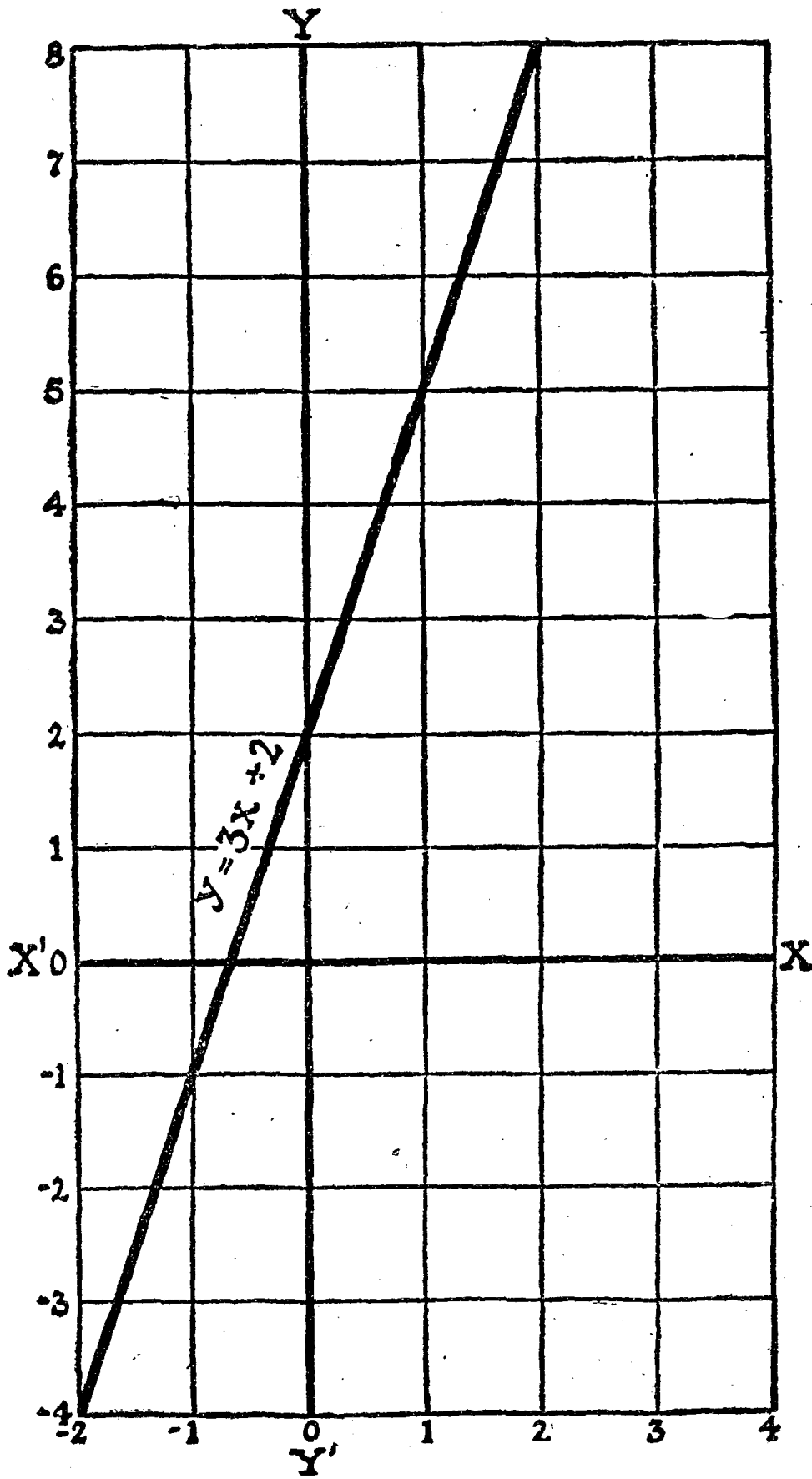
(註)通常所謂“curve”，係指縱橫坐標圖中之任何種綫，或為直綫，或為曲綫。



圖三. 方程式 $y = x$ 之圖解

關於上述各點，可繪一次簡單方程式之圖解以說明之。例如吾人欲作函數 $y = 2 + 3x$ 之圖解，則可假定 x 值為各個不同之數字，以決定各個 y 值。今將假定之各 x 值與由方程式決定之各個 y 值，列表於下：

x	y ($2 + 3x$)
-4	-10
-2	-4
0	2
2	8
4	14



圖四. 方程式 $y = 3x + 2$ 之圖解

將表中各值繪於圖上，更將各點接連，即得圖四所示之綫。此函數既為直綫關係（即圖為一直綫形式），故若已知任何兩點，即可確定綫之位置。圖中 y 截綫為 2，該綫與橫軸構成角度之正切（該綫之斜度）為 3，即 x 之係數（coefficient）。此綫上任何一點之縱橫坐標，均適合於前項方程式，而適合於此式之 x 及 y 值，又必為綫上之一點，足證此綫為表示該方程式之綫。又若一變量依定量遞增時，其另一變量亦依定量遞增，則兩者之關係，可用直綫表示之。例如上表內 x 值依定量 2 遞增時，則 y 值依定量 6 遞增。此種依定量遞增之級數，謂之算術級數（arithmetic series）。

在自然科學方面，兩變量之具有直綫關係者，其例不少概見。在經濟界中，按單利率計息時（即利息並不併入本金生息），其本利和之增加，即為其一例。設 r 表示單利率， x 表示年數， y 為 x 年後本金一元所有之本利和，則表示此種關係之方程式為

$$y = 1 + rx。$$

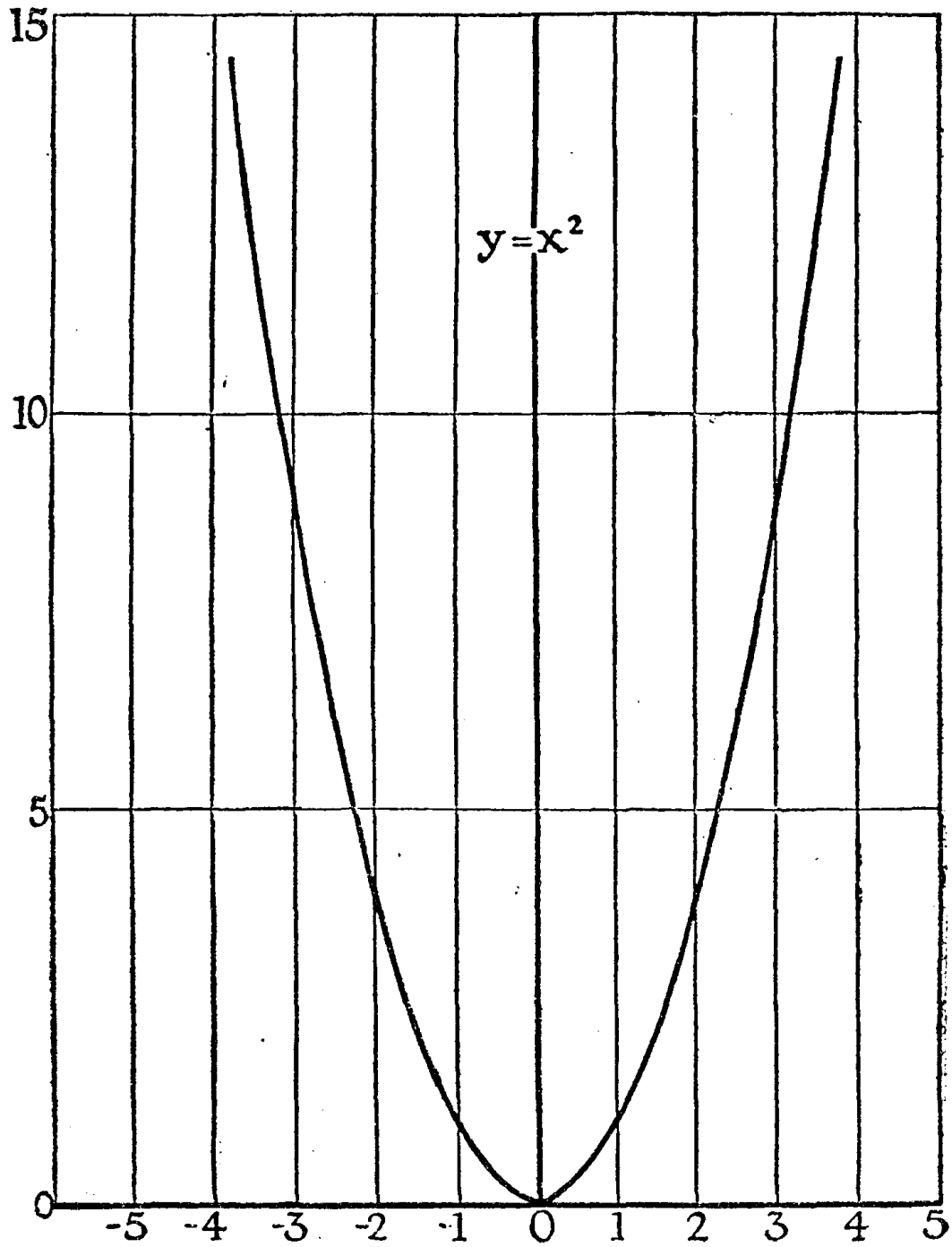
前式中 r 既為常數，故該式為一簡單之直綫方程式。在統計學中，表示此種簡單直綫關係之事實，實屬絕無僅有，惟近似於直綫關係者亦所常見。

非直綫關係

非直綫函數之形式有多種，本章內僅將少數較為普通者論述之。讀者對於各種主要非週期性曲綫（non-periodic curve）之普通性質應熟審之。此項曲綫中最為重要者，有屬於拋物綫（parabola）及雙曲綫（hyperbola）之形式者，亦有屬於指數曲綫（exponential curve）之形式者；其中尤以遞升冪級數（potential series）之形式最為普通，應用亦較廣。在週期函數（periodic functions）中，關於基本形式之正弦曲綫（sine curve）亦將簡述之。

拋物綫或雙曲綫形式之函數關係，在自然科學中甚為普通，此種曲

綫在經濟資料中亦常可採用之。其普通方程式爲 $y = ax^b$ 。該式中指數 b 之符號爲正時，此綫爲拋物綫；符號爲負時，此綫爲雙曲綫。茲舉例題兩則以說明其形式如下：



圖五 拋物綫：方程式 $y = x^2$ 之圖解

例題：試作一函數 $y=x^2$ 之圖解。

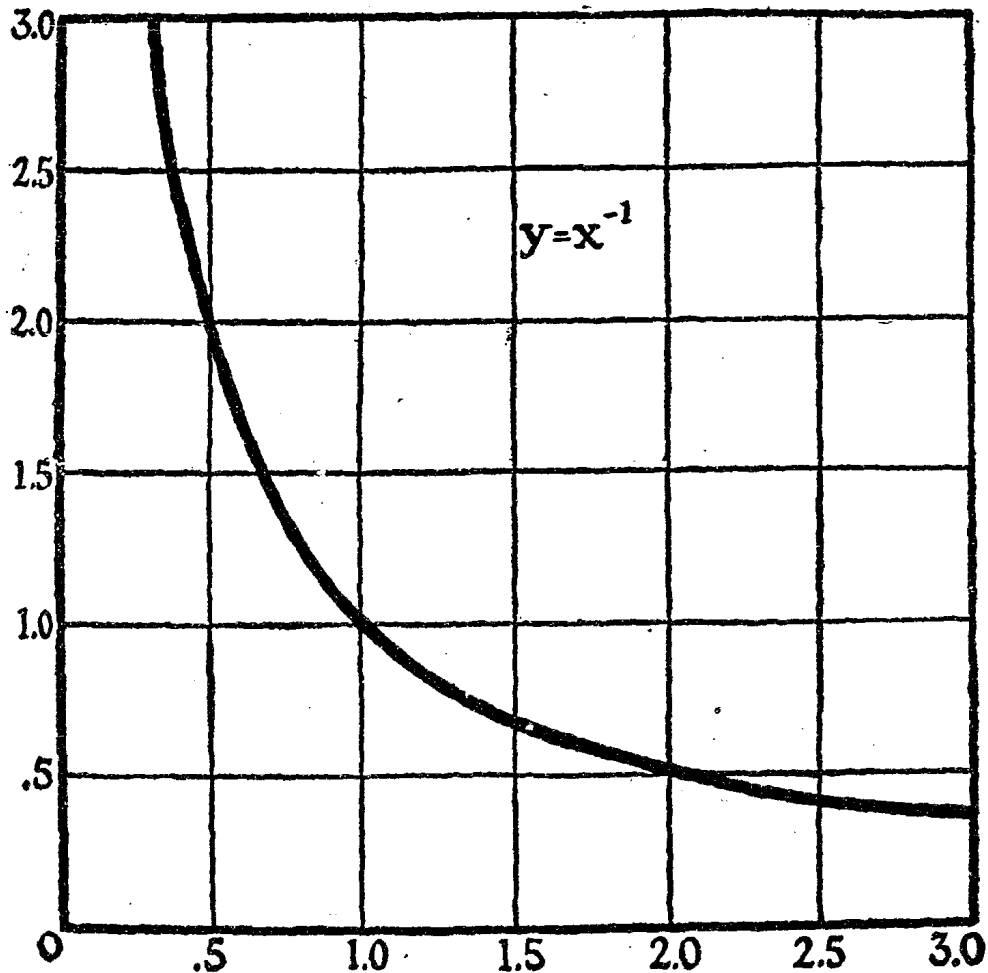
x	-5	-4	-3	-2	-1	0	1	2	3	4	5
$y(=x^2)$	25	16	9	4	1	0	1	4	9	16	25

依題製成之圖解如圖五。

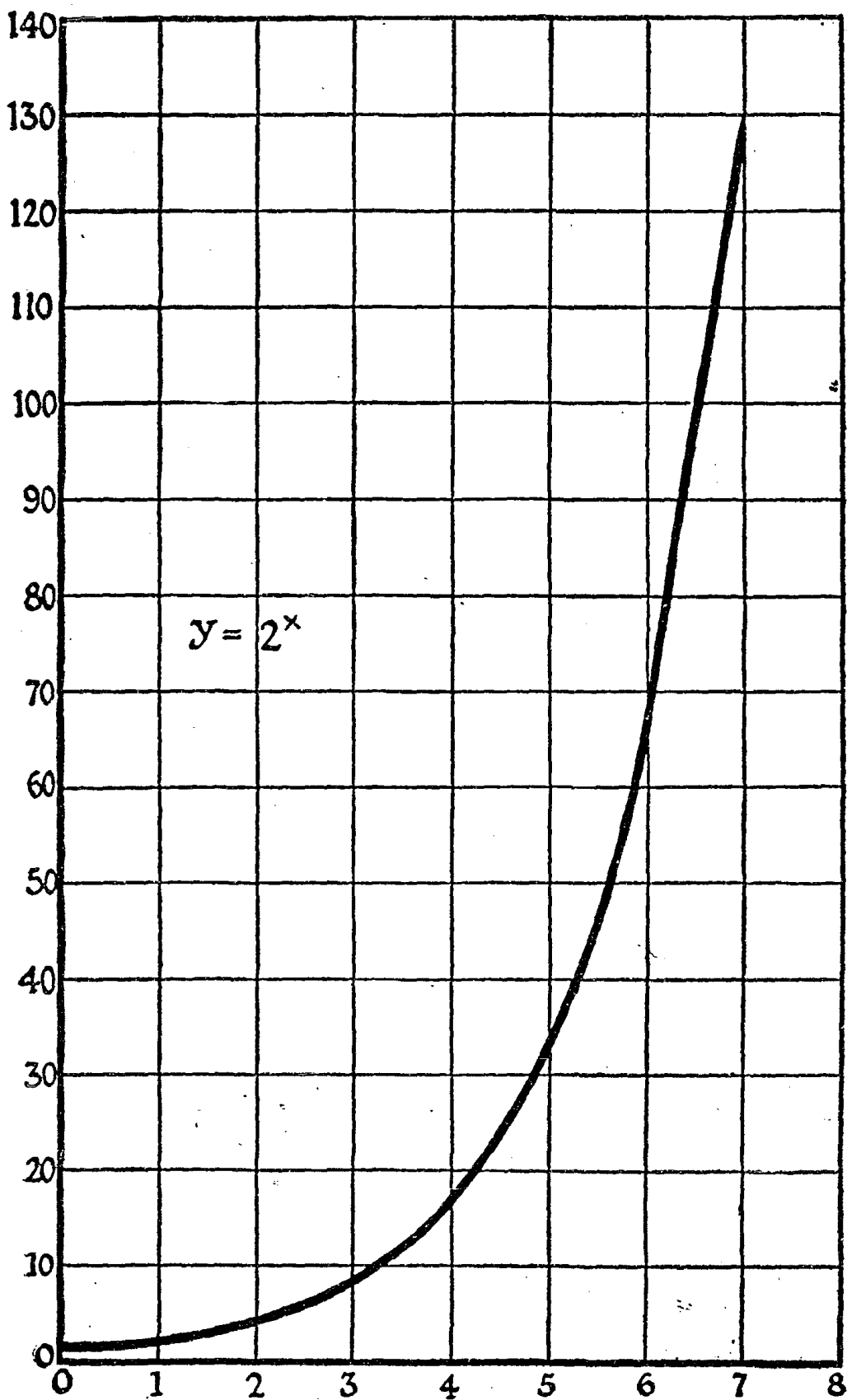
例題：試作一函數 $y=x^{-1}$ 之圖解，式中 x 值假定為正數：

x	$\frac{1}{3}$	$\frac{1}{2}$	1	2	3	4	5
$y(=x^{-1})$	3	2	1	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{4}$	$\frac{1}{5}$

表示此函數之圖解為等邊雙曲綫 (equilateral hyperbola)，如圖六所示。此式亦可寫作 $y = \frac{1}{x}$ ，或 $xy = 1$ 。



圖六。等邊雙曲綫：方程式 $y=x^{-1}$ 之圖解 (x 值為正數)



圖七. 指數曲線: 方程式 $y = 2^x$ 之圖解 (x 值爲正數)

在此種方程式中， x 之值如按幾何級數 (geometric progression) 增加，則 y 之值亦按幾何級數增加。如在前述拋物綫之一例中 ($y=x^2$)，假定各 x 值成一幾何級數 (geometric series) (註)，則由此式求得相對於各 x 值之 y 值，亦具有幾何級數之形式。

x	1	2	4	8	16	32
y	1	4	16	64	256	1024

此外另有一種函數，其形式可以方程式 $y=ab^x$ 代表之。在此式中，其一變量為指數 (exponent)，故表示此式之圖解，稱為指數曲綫 (exponential curve)。茲舉例說明於下：

例題：試作一函數 $y=2^x$ 之圖解，式中 x 值為正數。

x	0	1	2	3	4	5	6
$y(=2^x)$	1	2	4	8	16	32	64

依題製成之圖解，如圖七所示。

兩變量之相互關係，如各按定量遞增 (組成算術級數 arithmetic series) 時，可以直綫表示之；如各按幾何級數遞增時，可以拋物綫或雙曲綫表示之，已如前述。指數曲綫則為前兩種之混合形式，因其所表示兩變量之相互關係，其一變量按算術級數遞增時，另一變量則按幾何級數遞增也。上列之數字即係表示此種關係。

表示下列關係之曲綫，在統計研究方面應用頗廣。

$$y = a + bx + cx^2 + dx^3 + \dots$$

所謂遞升器級數 (potential series) 者，即指此種形式之方程式而言。此種曲綫雖非圓錐截面形之拋物綫，但如該曲綫方程式右邊之行列係至 x 之二次冪為止者，則此曲綫通常亦稱為二次拋物綫 (a second degree parabola)；如至 x 之三次冪為止者，稱為三次拋物綫 (a third degree parabola)，餘類推。按此種級數並無劃一或單純之形式，後文當詳論之。

(註) 某一級數如其中每項數字，係可由前一項數字乘以常數得之者，謂之幾何級數。

週期函數 (periodic functions) 為另一種曲綫, 常用以表示電學與氣象學中之各種關係者, 惟其範圍並不以此為限。此種關係之特徵, 厥為在自變量變動之固定時期內, 倚變量之數值重覆不變。茲將代表週期函數基本形式之正弦曲綫 (sine curve) 舉例說明於下:

例題: 試作一函數 $y = \sin x$ 之圖解。

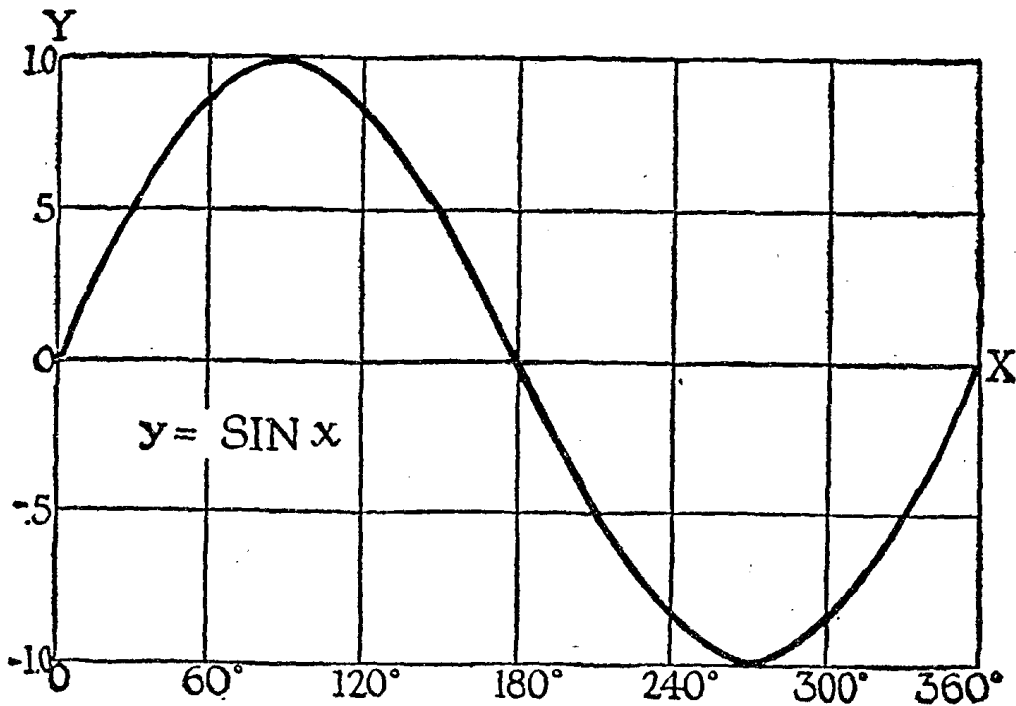
x (角度)

0° 30° 60° 90° 120° 150° 180° 210° 240° 270° 300° 330° 360° 390°....

$y (= \sin x)$

.000 .500 .866 1.000 .866 .500 .000 —.500 —.866 —1.000 —.866 —.500 .000 .500....

依題製成之圖解, 如圖八所示。



圖八. 正弦曲綫: 方程式 $y = \sin x$ 之圖解

求數學方程式以表示兩變量之相互關係, 對於統計工作上之重要性, 在未詳論本題之前, 尚不能有所表現, 惟其基本目的在於根據觀察所得之現象, 以決定自然或經濟的法則。其另一種更較實用之目的, 則在求一公式, 俾藉此公式, 由已知一變量之值, 以推求另一變量之值。達

到此種目的之方法，在本書各節中將詳論之(註)。

對 數

對數 (logarithms) 之爲用，在普通數學演算方面，固占重要地位，而於統計資料之處理，其重要程度正復相等。茲將關於對數之性質及其便利計算手續之方法簡述之，惟本文所論者，祇限於以 10 爲基數 (base) 之普通對數。

任何正數均可以 10 之冪數 (power) 表示之。如

$$1,000 = 10 \times 10 \times 10 = 10^3$$

$$10,000 = 10 \times 10 \times 10 \times 10 = 10^4$$

前式中 10 之冪數 (即 10 之右上角較小之數字)，表示因數 10 之自乘次數。眞數化爲 10 之自乘次數如成爲整數時，其冪數亦爲整數，其他數字則其冪數必含分數矣。如 100 等於 10 之二次冪，即 10^2 ；110 等於 10 之 2.04139 次冪，即 $10^{2.04139}$ 。

此 10 之冪數，即係將 10 提升至與某數相等時乘冪之指數，稱爲某數之對數。如 100 之對數爲 2；110 之對數爲 2.04139；998 之對數爲 2.99913 等是。此項對數均以 10 爲基數，惟對數法中任何數字均可作爲基數。通常如

$$a = b^c$$

$$\text{則} \quad \log_b a = c$$

此式可讀如“以 b 爲基數之 a 的對數爲 c ”。在用普通對數時，眞數、基數、與對數三者之關係，可憑下式記憶之。

$$100 = 10^2$$

$$\log_{10} 100 = 2$$

一數之對數，常具整數與小數兩部分，其整數稱爲首數 (character-

(註)關於各種曲綫之形式，當於下文時間數列之分析一節中詳論之。

istic), 其小數稱為尾數 (mantissa)。在求對數時, 首數可由觀察斷定之, 尾數則須查閱對數表得之。首數須視真數之小數點位次而有變更, 尾數則否, 凡數字排列相同之真數, 不論其小數點位次之先後, 其尾數常屬相同。此點可以下列數字舉證之。

$$\begin{aligned} \log 8450 &= 3.92686 \\ \log 845 &= 2.92686 \\ \log 84.5 &= 1.92686 \\ \log 8.45 &= .92686 \\ \log .845 &= 9.92686 - 10 \\ \log .0845 &= 8.92686 - 10 \end{aligned}$$

由對數求真數時 (此真數 natural number 謂之反對數 anti-logarithm), 可由尾數決定各位數字之排列, 由首數決定小數點之位置。例如欲求 2.17609 之反對數, 可將其小數點以後之尾數 .17609, 查閱對數表得相對於該尾數之真數為 1500。因其首數為 2, 故真數必在 100 至 1,000 之間, 而可斷定此數為 150。

下列各項數字, 為 10 之乘冪及相對於各乘冪之真數, 讀者略加研究, 即可記憶 10 之倍數與其對數之關係, 則求對數時其首數自易決定矣。

$$\begin{array}{cccccccccc} .0001 & .001 & .01 & .1 & 1 & 10 & 100 & 1,000 & 10,000 & \\ 10^{-4} & 10^{-3} & 10^{-2} & 10^{-1} & 10^0 & 10^1 & 10^2 & 10^3 & 10^4 & \end{array}$$

上列兩行數字中, 其下行內 10 之冪數, 即係上行內各數之對數。

自 0 起至 1 止之一切數字, 其對數皆為負數; 例如 .845 之對數為 $-1 + .92686$, 可寫作 $9.92686 - 10$ 。自 0 起至無限大數間之一切正數, 其對數由負而正, 包括一切正負數值。由是知負數既無正對數, 亦無負對數。

以 10 之冪數表示各種數字, 其優點在於乘, 除, 求乘冪, 或開方根時, 可使計算手續便捷不少。

求乘積用對數相加 各因數的對數之和，等於其乘積之對數。

其公式爲：

$$a^b \times a^c = a^{(b+c)}$$

例如：

$$\begin{aligned} 10^2 \times 10^3 &= (10 \times 10) \times (10 \times 10 \times 10) = 10^5 = 100,000 \\ 100 \times 1,000 &= 100,000 \end{aligned}$$

求商數用對數相減 如以甲數除乙數，可用乙數之對數，減去甲數之對數，即得所求商數之對數。

其公式爲：

$$a^b \div a^c = a^{(b-c)}$$

例如：

$$\begin{aligned} 10^5 \div 10^2 &= \frac{10 \times 10 \times 10 \times 10 \times 10}{10 \times 10} = 10^3 = 1,000 \\ 100,000 \div 100 &= 1,000 \end{aligned}$$

求乘冪用對數相乘 欲求某數之乘冪，可將該數之對數乘其冪數，其乘積即爲所求乘冪之對數。

其公式爲：

$$(a^b)^c = a^{bc}$$

例如：

$$\begin{aligned} (10^3)^2 &= (10 \times 10 \times 10) \times (10 \times 10 \times 10) = 10^6 = 1,000,000 \\ 1000^2 &= 1,000,000 \end{aligned}$$

開方根用對數相除 欲開某數之方根，可將該數之對數除以方根之指數，其商數即爲所求方根之對數。

其公式爲：

$$\sqrt[b]{a^c} = a^{(\frac{c}{b})}$$

例如：

$$\begin{aligned} \sqrt[3]{10^6} &= 10^{\frac{6}{3}} = 10^2 = 100 \\ \sqrt[3]{1,000,000} &= 100 \end{aligned}$$

茲將各式綜述於下：

$$\log(a \times b) = \log a + \log b$$

$$\log(a \div b) = \log a - \log b$$

$$\log a^b = b \times \log a$$

$$\log \sqrt[b]{a} = \log a \div b$$

計算尺 (slide rule) 爲利用對數之優點所製成之工具，所以便利計算手續者也。研究統計學者，應明瞭其使用法。

對數方程式

採用垂直縱橫坐標制作圖之方法及其優點，已如前述。作圖時有須用對數以代替真數者，蓋以對數作圖常可使各種重要關係明白表現，而又可將處理資料之手續化繁爲簡；且利用對數時，可將複雜之曲綫化爲直綫，在處理與解釋方面較爲簡易，其所獲之利便爲尤著焉。

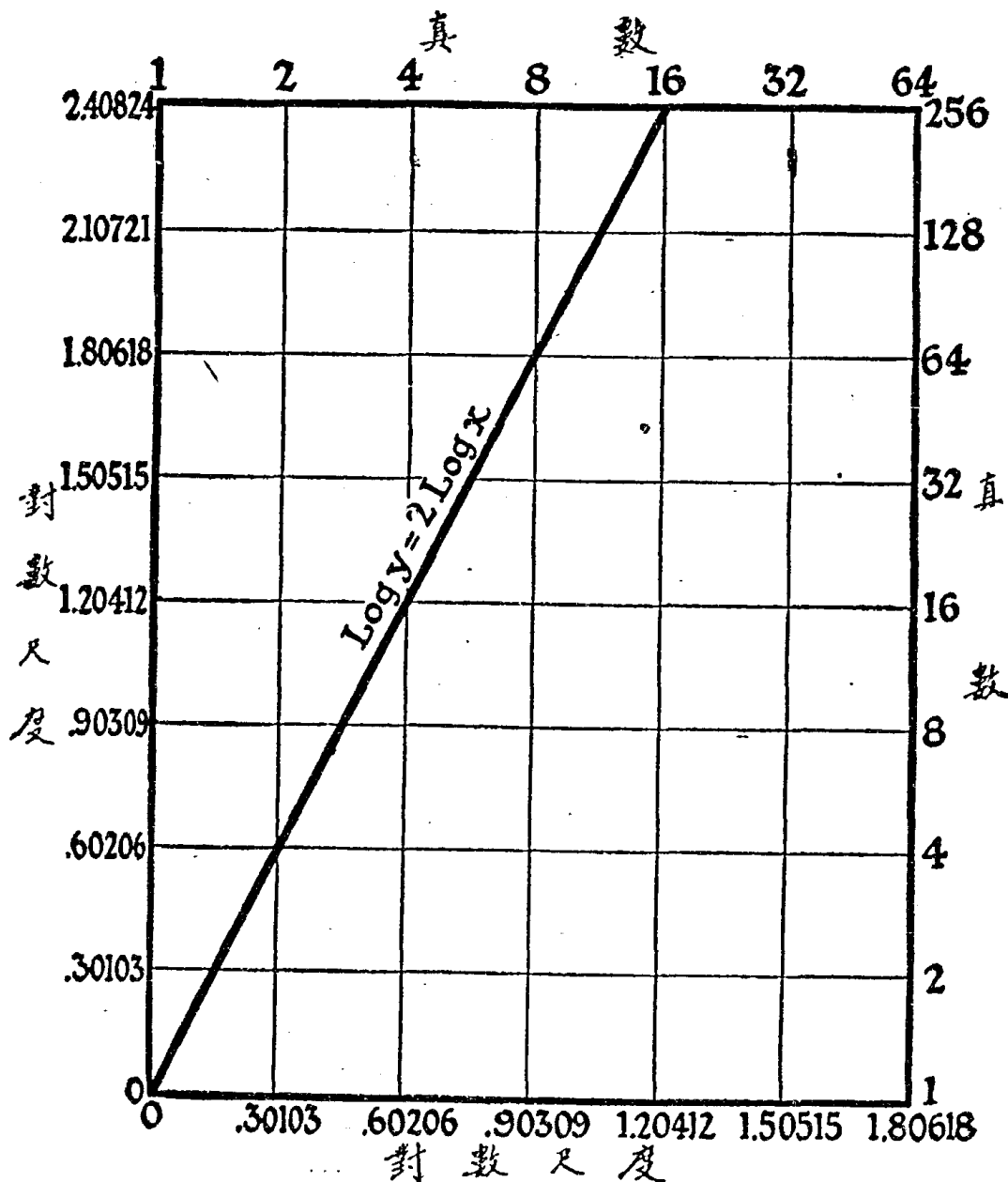
直綫方程式之普通形式爲 $y = a + bx$ ， a 與 b 皆爲常數，爲表示某綫與縱軸相交之截綫與其斜度。用對數將方程式化簡之法，係將 x 或 y ，或同時將 x 與 y 之真數各代以 $\log x$ 或 $\log y$ ，如是則可將高次方程式化爲較簡之形式。

此法可以方程式 $y = x^2$ 說明之。如以此式繪入垂直縱橫坐標圖中，即得拋物綫式之曲綫（參看圖五）；若將此式化爲對數之形式，即成爲 $\log y = 2 \log x$ 。此用 $\log y$ 與 $\log x$ 爲變量之方程式，即屬於直綫之形式。假定 $\log x$ 之值爲正，則所得圖解如圖九。

爲便於表示此中關係起見，可同時將相對於各對數之真數，分別註明於圖之右方及上方之尺度上。真數尺度上之各數成爲幾何級數，其對數則爲算術級數。惟須注意者，圖上相等之縱距離在對數尺度上係表示相等之絕對增量，而在真數尺度上則係表示相等之增加百分率。

方程式 $y = 5x^3$ 可依此化爲直綫形式，如 $\log y = \log 5 + 3 \log x$ 。其

他屬於 $y = ax^b$ 形式之一切方程式，即一切單純拋物綫與雙曲綫，亦皆可同樣化為直綫形式，如 $\log y = \log a + b \log x$ 。在圖解方面，此即以 y 之對數與 x 之對數作圖也。



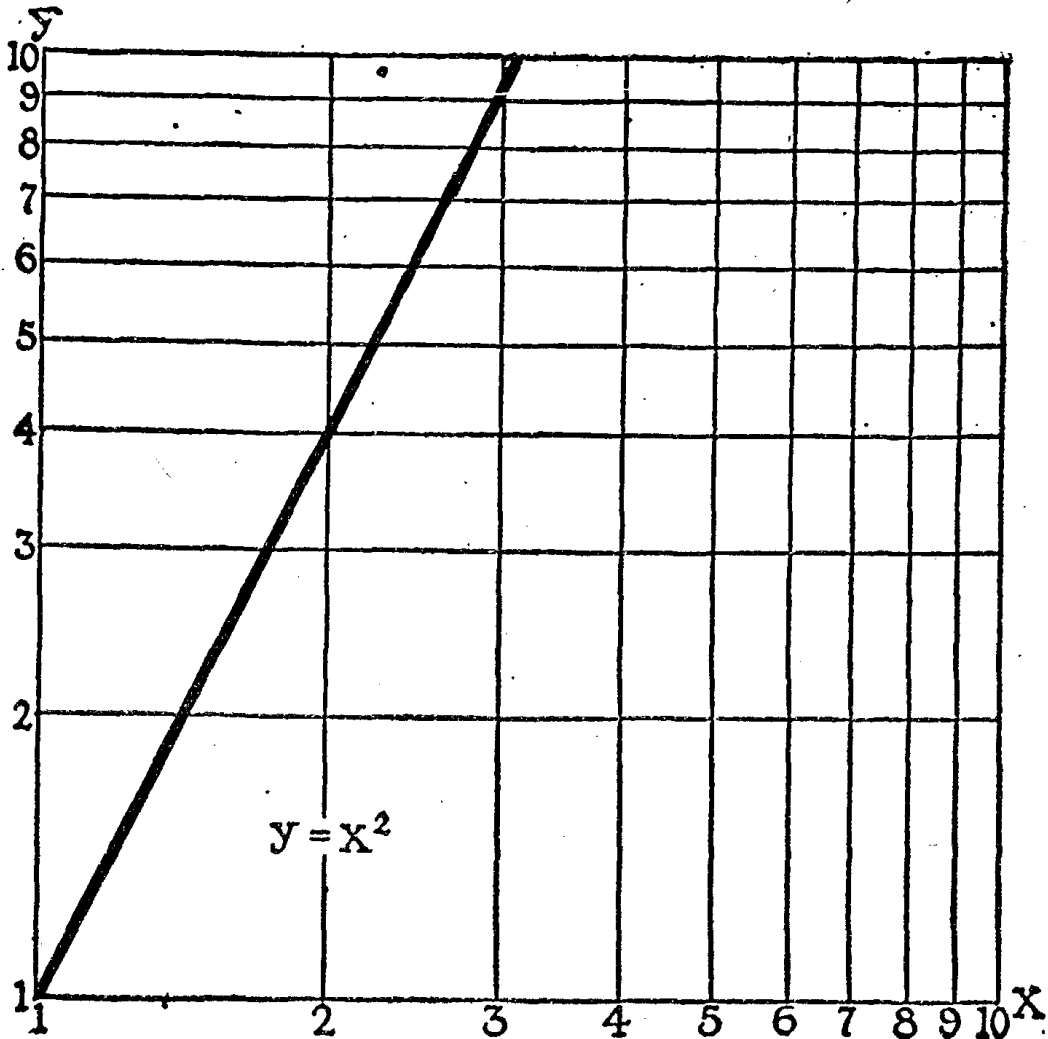
圖九. 方程式 $\log y = 2 \log x$ 之圖解(方程式 $y = x^2$ 之對數形式)

至於指數曲綫 $y = ab^x$ 形式之方程式，其變化方式與前式略異。此式以對數形式表示之為 $\log y = \log a + x \log b$ ，亦屬直綫形式，其常數為

$\log a$ 與 $\log b$, 其變量為 x 與 $\log y$ 。依此方程式將 x 之真數與 y 之對數作圖, 即得一直綫。茲將此種形式之曲綫說明於下。

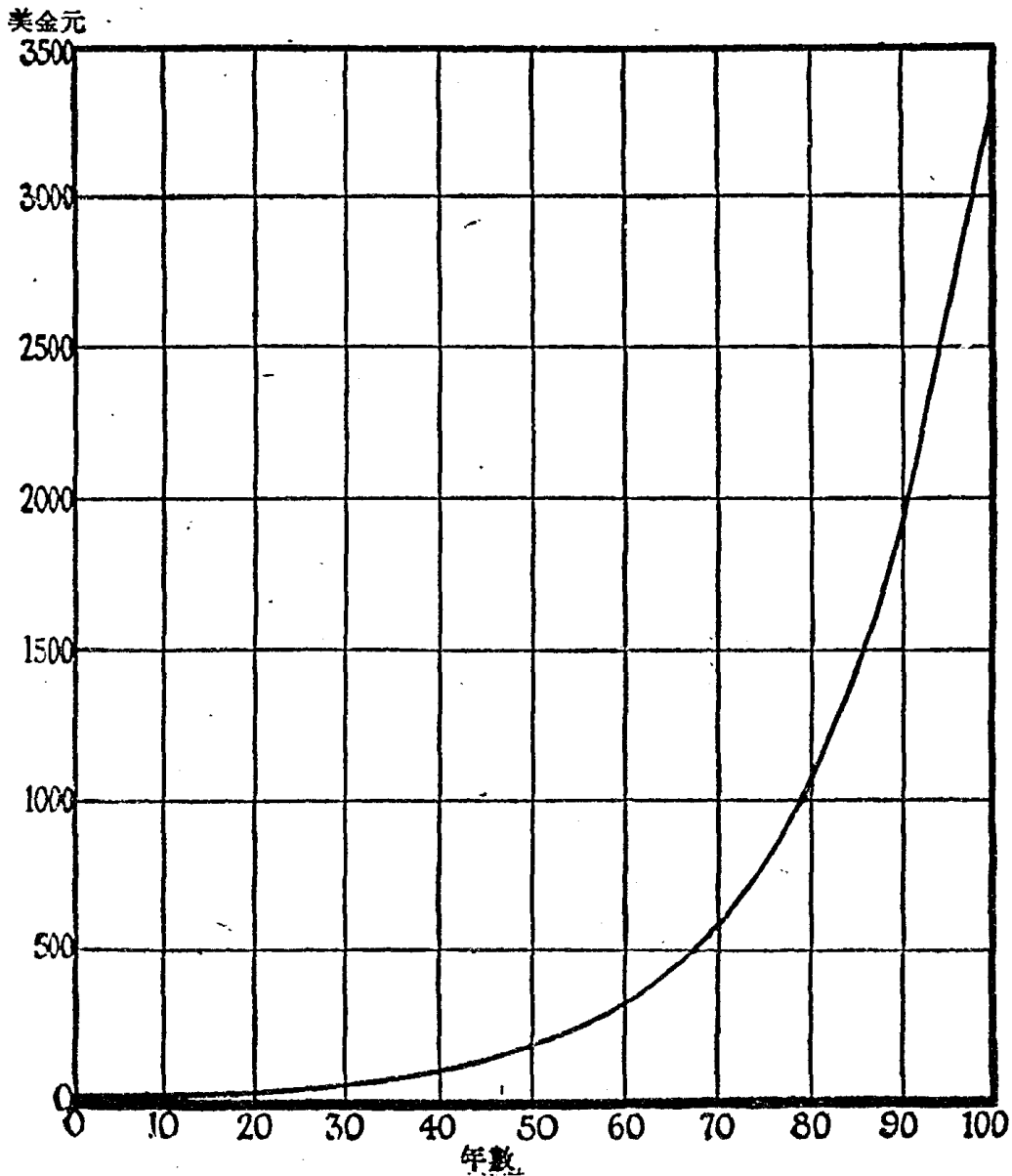
雙對數與單對數圖

以對數作圖, 未必有利而無弊。如點數過多時, 則查閱對數必甚煩瑣, 而吾人所注意之原值, 圖中反無載明。欲避免此種困難, 可於製圖時用對數定尺度, 而用真數註明之。此種方法與製造計算尺所採取者完全相同, 蓋計算尺在尺度上所註明者雖為真數, 其距離則仍以各數之對數為比例也。圖十即為此種圖解, 係表示方程式 $y = x^2$ 者。



圖十. 方程式 $y = x^2$ 之圖解(繪於雙對數尺度之圖紙上者)

如將此種圖解之形式略加變化，以真數為橫軸之尺度，以對數為縱軸之尺度，即成統計工作上重要之圖解。在此種混合尺度上所作之圖，自與按照 x 之真數與 y 之對數所繪製者相同。上述指數形式之曲線化為直線時，即須用混合尺度之圖制繪之。此種單對數式（或稱比率式）之圖紙，可按照計算尺或對數表繪製，即製就之圖紙亦可購得。在經濟統計製圖方面，此種圖制尤有特殊之價值，蓋在經濟統計中，時間每為變



圖十一．複利法則：本金十元、複利六厘計息、百年間之本利和（繪於算術尺度上者）

量之一，而此變量自以採用真數尺度繪製，較為適當也。

表示複利法則之方程式，即屬此類曲綫。設以 r 表示利率， p 表示本金， y 表示 x 年後之本利和（依複利計息，每年結息一次），則所得方程式應為：

$$y = p(1+r)^x$$

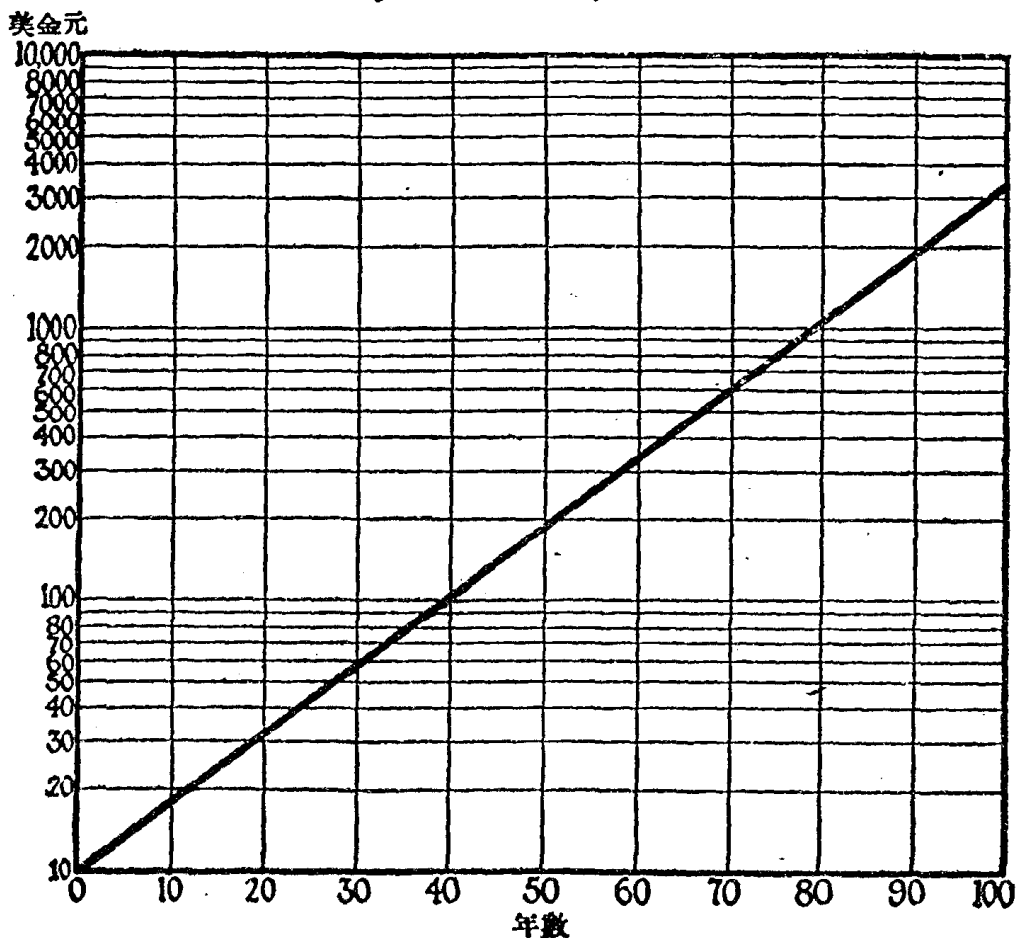
以對數形式表示之，其式為：

$$\log y = \log p + x \log(1+r)$$

即變為直綫之方程式矣。

圖十一係繪於真數尺度上，表示本金十元、複利六厘計息、本利和增加之曲綫。此為指數方程式

$$y = 10(1+.06)^x$$



圖十二 複利法則：本金十元、複利六厘計息、百年間之本利和
（繪於單對數尺度，即比率尺度上者）

之圖解， y 表示第 x 年年末之本利和。圖十二係用同一資料繪於單對數尺度之圖紙上者，其指數曲綫已化爲直綫。

單對數圖紙 (semi-logarithmic paper) 之用途，並不限於由指數曲綫化爲直綫，蓋有多種資料如用單對數圖紙作圖，其意義最能明白顯示也。其優點將於下文詳述之。

統計圖之功用

根據觀察或統計調查所得之數字結果，加以分析及解釋時，應先作圖，以明其大概，此爲其初步工作之一。此種手續不特具有科學上之價值，可爲更進一步研究之張本，且各項數字可賴圖示而益明瞭；蓋僅依邏輯之順序以研究事物，失之抽象，而藉具體之圖形以刺戟視官，則爲達到理解與意想之捷徑。例如吾人欲解釋一行列之數字，或爲困難之事，而將同一資料示以圖形，則其意義可由圖上窺見一斑。此圖示法之所以不特在實驗室與製圖室中占極重要之地位，卽在日常商業活動上，其重要程度亦復相若也。

至於現時工程師與統計家所用之圖，種類繁多，因不在本書範圍以內，故不備述；惟關於圖示法較重要之原理，應略加詮釋，日常通用之各種主要圖形，亦將予以例示。其他例證，當於本書後數章內舉示之。

選擇圖形之條件

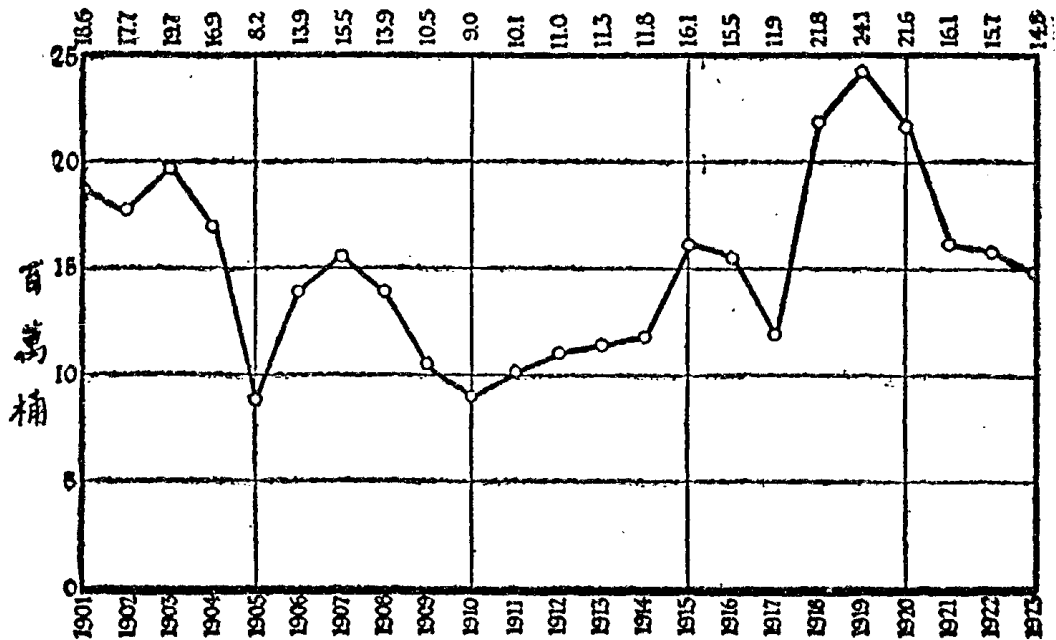
在應用圖示法時，選擇圖形之條件有二：其一，爲關於資料之性質者。一種資料通常雖可用各種不同之圖形表示，然其中大抵祇有一種圖形爲最適於該項資料之用，亦有某種圖形絕不適用於某種資料者，是則選擇圖形時應先參酌資料之性質。

其二，爲關於圖解所欲表示之目的者，則較第一條件更爲重要。普通所用之各種圖形各有其特殊之功用。圖解之目的，或爲表明某種特

徵，或為揭示某種關係，未有一種圖形能合於各種目的者，故在作圖目的未曾確定以前，選擇圖形殆無標準可言。茲將少數標準圖形縷述於下，以供選擇圖形之參考。

表示時間數列之圖解

時間數列之圖解，其注重之點在於表示各種數值在時間上之變動與其長期趨勢，以及長期趨勢綫兩邊數值之變動情形。如圖示之目的注重於絕對數值之變動(absolute variation)，即注重於各值實際單位之變化時，則同一簡單圖形如圖十三者，即可達此目的。此圖為一九〇一年至一九二三年間美國麵粉每年對外輸出數量之圖解。圖之縱橫軸均係用算術尺度。圖中所示各點表示各年之數量，惟為便於解釋起見，各點之間均用直綫連之。此圖可說明逐年變動之簡單情形，如一九〇七年至一九一〇年之向下趨勢，一九一一年至一九一九年之向上趨勢，及一九一九年後之下降趨勢，均可於圖中見之。關於製圖方法，應注意下列各點。



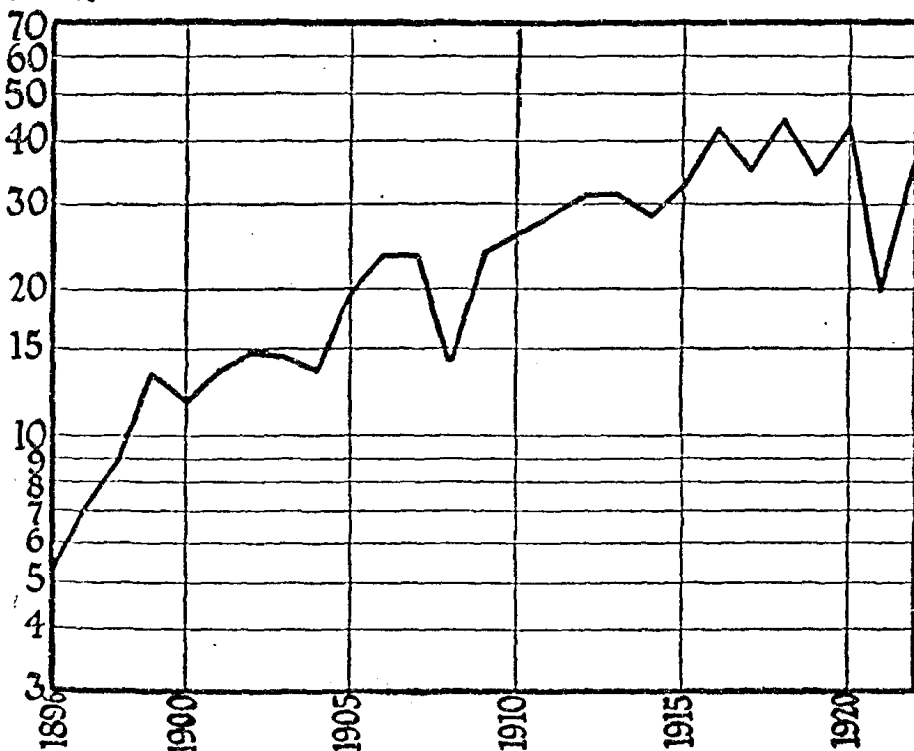
圖十三. 1901—1923年美國麵粉輸出數量

1. 圖之標題應載明圖中所示之資料及其包含之時期。
2. 縱尺度應自零綫 (zero line) 起, 使讀者對於變量之大小, 可得一明確之印象。
3. 零綫及連接各點之綫, 應較縱橫尺度之綫稍粗, 以求醒目。
4. 表示尺度之數字, 應置於圖之左邊及下邊, 惟縱尺度可重複置於圖之右邊, 以求讀者之便利。數字之排列, 應自下而上, 以基綫作底綫; 或自右而左, 以圖之右邊綫作底綫。
5. 各點之 y 值可分別註明於圖頂。此種辦法雖有裨助, 但非必要, 因圖外常附有載明數字之表也。

比率圖之優點

作圖之目的, 如不欲注重絕對變動 (absolute variation), 而注重

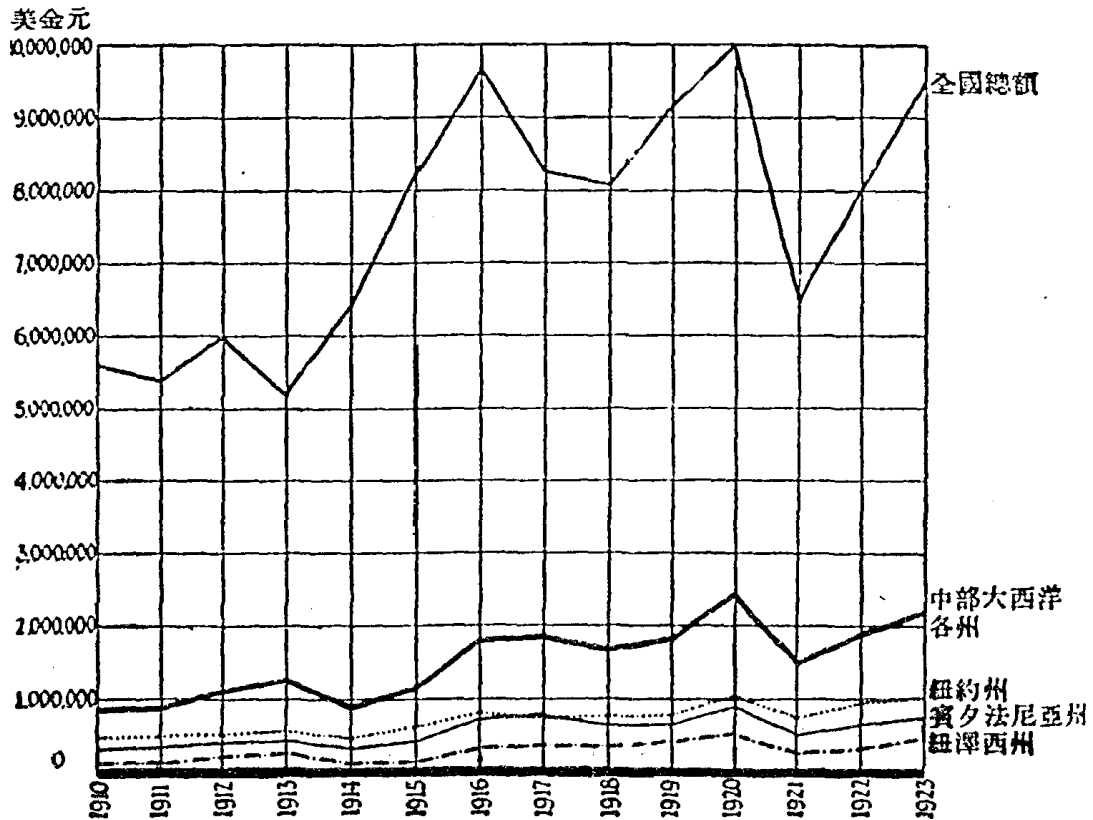
百萬英頓



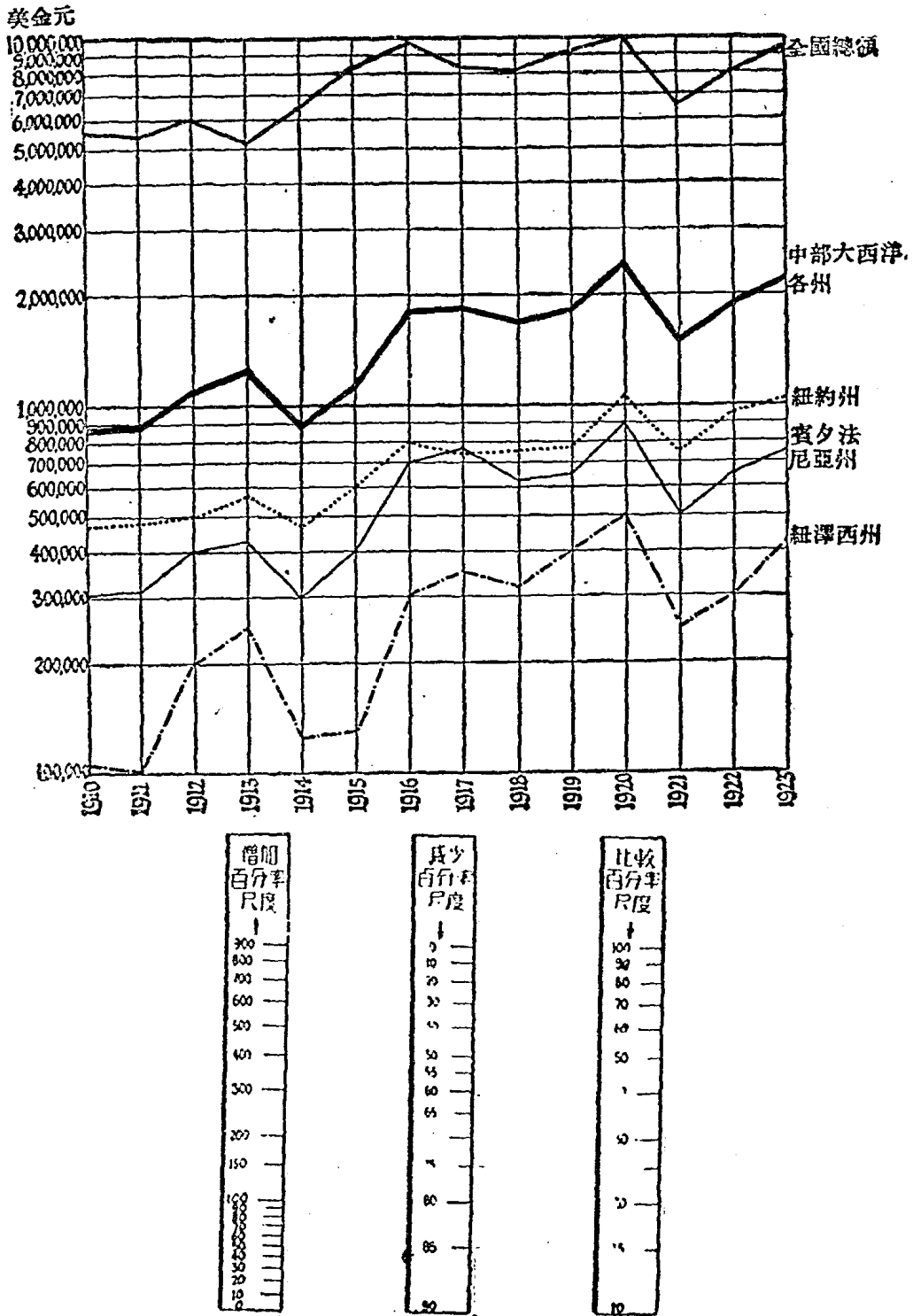
圖十四. 1896—1922年美國鋼鐵產量(繪於單對數尺度上者)

於相對變動 (relative variation) 時, 則所用之圖以取 y 軸用對數尺度, x 軸用算術尺度之單對數圖形最為相宜。在普通算術尺度之圖形中, 相等之縱距離表示相等之絕對增減量; 而在單對數之圖形中, 則相等之縱距離表示相等之變動百分率。吾人所以用單對數圖 (semi-logarithmic chart, 或稱比率圖 ratio chart) 表示時間數列者, 因數量之變動必須與其所由變動之基數相比, 始有意義; 例如由基數 100 增加 100, 與由 10,000 增加 10,000, 其重要程度彼此相等, 蓋兩數各增一倍也。然依絕對增量而言, 則後者適百倍於前者, 用算術尺度所表示兩者之變動, 亦適成此比例。若用單對數尺度表示, 則此兩種變動之重要程度彼此相等。

此種單對數圖形如圖十四所示。該圖為自一八九六年至一九二二年美國鋼鐵產量之圖解。繪曲綫時雖用絕對數量, 然因其縱軸為比率尺度, 故該圖可表示逐年產量之變動, 與其變動百分率成正比例。



圖十五. 1910—1923年美國亞根公司之全國銷貨總額及某數州之銷貨額
(繪於算術尺度上者)



圖十六. 1910—1923年美國亞根公司之全國銷貨總額及某數州之銷貨額
附增加、減少及比較之尺度(繪於單對數圖紙上者)

試以圖十五與圖十六相比較，則比率尺度(即單對數尺度)之優點，即可判明。此兩圖所表述之資料，列如下表：

表 二
1910—1923 年美國亞根公司之銷貨額

	賓夕法尼亞州	紐澤西州	紐約州	中部大西洋各州	全國總額
1910.....	\$305,000	\$105,000	\$465,000	\$ 875,000	\$5,600,000
1911.....	310,000	100,000	480,000	890,000	5,400,000
1912.....	400,000	200,000	500,000	1,100,000	6,000,000
1913.....	425,000	250,000	575,000	1,250,000	5,200,000
1914.....	300,000	125,000	465,000	890,000	6,400,000
1915.....	400,000	130,000	600,000	1,130,000	8,200,000
1916.....	700,000	300,000	800,000	1,800,000	9,700,000
1917.....	760,000	350,000	740,000	1,850,000	8,300,000
1918.....	630,000	320,000	750,000	1,700,000	8,100,000
1919.....	650,000	400,000	775,000	1,825,000	9,200,000
1920.....	900,000	500,000	1,050,000	2,450,000	10,000,000
1921.....	500,000	250,000	750,000	1,500,000	6,500,000
1922.....	650,000	300,000	950,000	1,900,000	8,000,000
1923.....	750,000	425,000	1,025,000	2,200,000	9,500,000

如將表內五個數列，一併繪入一算術尺度之圖形中，則因數列中既有一千萬元之一項，所定尺度勢須收縮，其結果則較小之數量在圖中遂失去其重要性。觀於圖十五所示，各年間全國銷貨總額之變動似甚劇烈，中部大西洋各州銷貨總額之變動似較和緩，紐約等三州之各州銷貨額幾無變動。此種圖形最易使人誤解，其實際情形當如圖十六所示。該圖係以同一資料繪入單對數尺度之圖紙上者。依該圖所示，則紐約等三州銷貨額之變動均較全國銷貨總額為劇烈，故欲以數個數列互為比較，而各個數列所含之數值相差又甚鉅時，算術尺度之圖形頗不適用，以其不能表示真實意義，徒足使人曲解，而在比率尺度之圖形中，則可作合理之比較也。

在圖十六下方所刊之三種尺度，係欲將對數尺度之某種功用為更明顯之說明。其中增加尺度 (scale of increase) 可用以測量一個數列內前後兩期之增量。前嘗說明，在圖中任何地位，一定之縱距離係表示一定之增加百分率，例如沿縱尺度自十萬元至二十萬元之距離，等於

自二百萬元至四百萬元之距離。吾人可量任何兩點之縱距離，即將此距離合於增加尺度上，由下至上讀之，即得增加百分率。例如欲定紐澤西州一九一三年銷貨額較一九一二年之增加百分率，可先量此兩年間之縱距離，將此縱距離合於增加尺度上可得25%之增加率。

減少尺度之用法與增加尺度相似。吾人可量任何兩點之縱距離，將此距離合於減少尺度上，由上至下讀之，即得減少百分率。各尺度上所繪之箭形，所以表示讀尺度時之方向也。

比較尺度係用以決定在任何時期兩個數列之百分比者。例如吾人欲知一九一二年中部大西洋各州之銷貨總額占同年全國銷貨總額之百分比時，則可量此兩點間之縱距離，將此縱距離合於比較尺度上，由上至下讀之，可得此百分比約為18%。

繪上述三種尺度時可將尺度上應用之各數，依比率尺度繪之。吾人如欲採用標準形式之單對數圖紙，作若干圖解，可用預製之尺，以供繪製此種尺度之用，則尤為便利矣。

用單對數尺度作圖，其主要優點略如下述：

1. 指數形式之曲綫，如繪於單對數圖上，即成一直綫。例如依複利計息所有本利和之增加，如繪入此圖中，即為一直綫。
2. 任何數列中，如其增加率或減少率固定不變，則在單對數尺度之圖中為一直綫。
3. 在此種圖形中，表示相等的變動百分率之數綫，其斜度必相等，故兩種數列如其增加率或減少率彼此相等，則代表此種數列之兩綫彼此平行。
4. 比較兩個或數個數列之變動率時，可由比較各綫之斜度得之。
5. 單對數尺度之圖形，既可用絕對數值作圖，同時又可作相對變動之比較。
6. 在比較若干數列，而各數列所含之數量相差又極鉅時，則繪於單

對數圖上較為相宜。

7. 變動百分率及各項數值間之百分比，皆可直接由圖中得之

比較各組次數之圖解

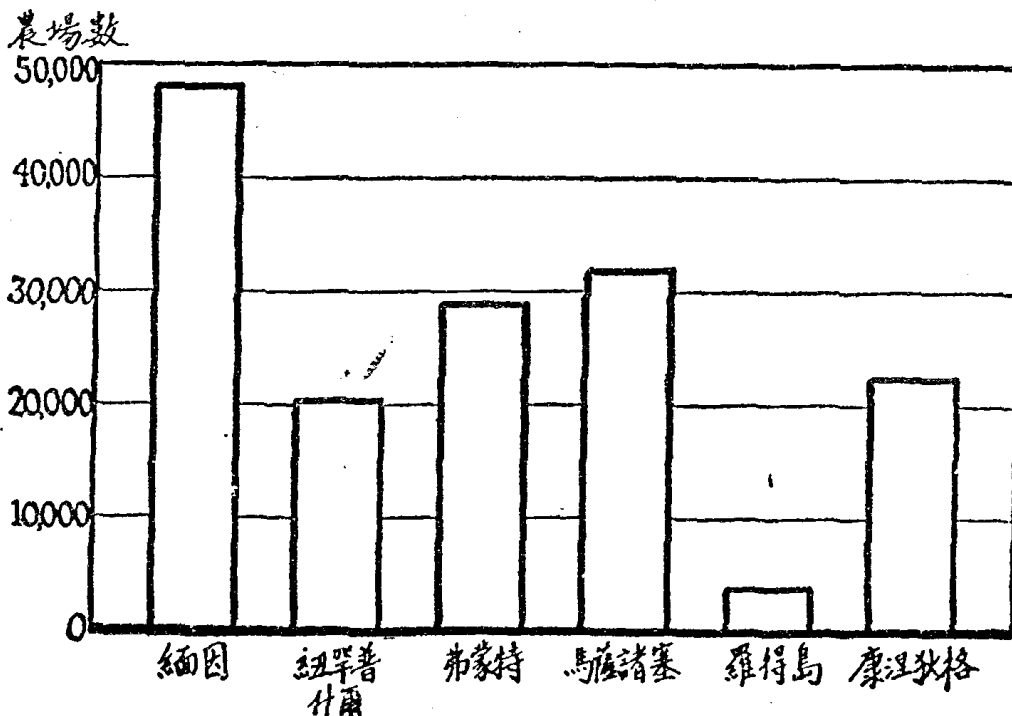
圖示之目的，如為比較各組次數 (frequency)，則當用另一種圖解以表示之。次數者，即各類事物發現之次數也。試以下列數字為例說明之。

表 三

1920年新英格蘭各州農場數

州 別	農場數
緬因(Maine)	48,227
紐罕普什爾(New Hampshire)	20,523
弗蒙特(Vermont)	29,075
馬薩諸塞(Massachusetts)	32,001
羅得島(Rhode Island)	4,083
康涅狄格(Connecticut)	22,655

如欲比較上述六州一九二〇年之農場數，可用圖十七所示之條形



圖十七 1920年新英格蘭各州農場數

圖(bar diagram)表示之。此種圖形雖簡，然頗合此項比較之用。

此種表示次數分配 (frequency distribution) 之圖形，容於下章內再行例示，並將說明處理某種資料時，由單純條形圖化為次數多邊圖 (frequency polygon) 或次數曲綫 (frequency curve) 之方法。此種次數曲綫乃圖形中之極重要者，另於後文內詳述之。

表示成分之圖解

在製圖時，有須將總量依其成分 (component parts) 各別圖示者，如是則吾人不僅可明瞭總量之變動，且可明瞭各個成分之變動。下表數字為屬於此類資料之一簡例。

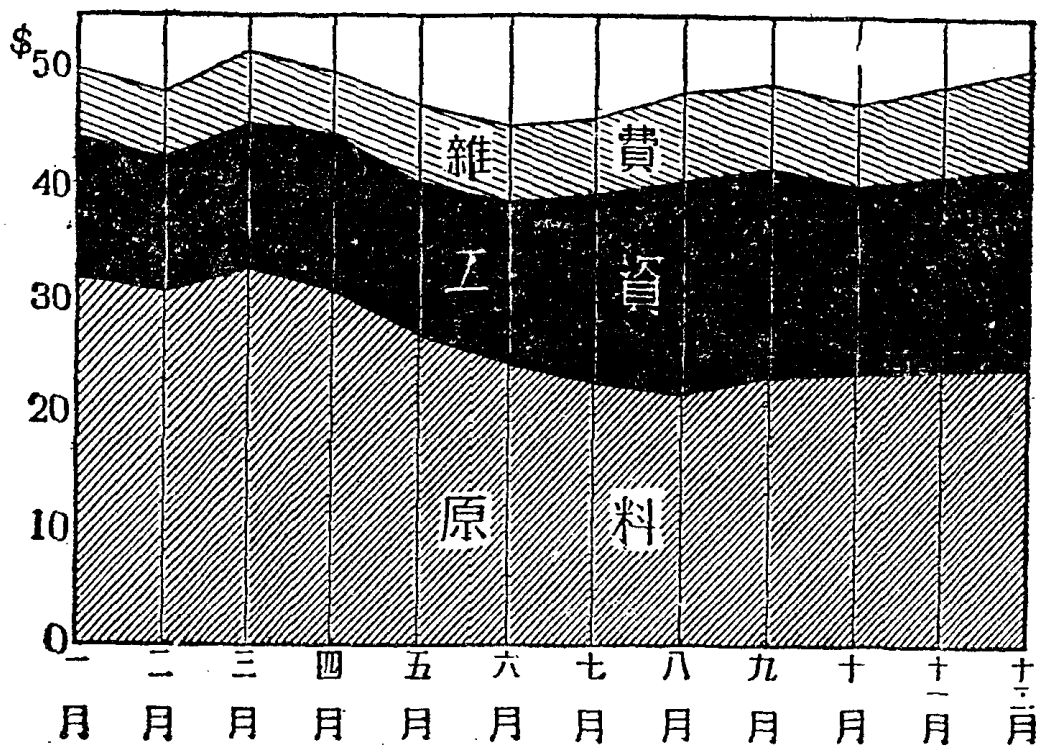
表 四

1923年某公司之生產費

(此項數字係表示製造每一單位之物品所需之各項生產費及其總數)

月 份	原料費	工 資	雜 費	生產費總計
一 月	\$32.00	\$12.00	\$6.00	\$50.00
二 月	31.00	11.00	6.00	48.00
三 月	33.00	12.00	6.50	51.50
四 月	31.00	13.00	6.00	50.00
五 月	27.00	13.00	7.00	47.00
六 月	24.50	13.50	7.00	45.00
七 月	23.00	16.00	7.00	46.00
八 月	22.00	18.50	8.00	48.50
九 月	23.50	18.00	7.50	49.00
十 月	24.00	13.00	7.50	47.50
十一月	24.00	17.00	8.00	49.00
十二月	24.50	17.50	8.50	50.50

前項數字繪入成分圖中，當如圖十八所示。由圖可知各月生產費總數之變動較為穩定，但其各個成分頗有變動甚烈者。是以吾人但知生產費之總數，猶未能窺其全豹，而須同時明瞭各項費用之變動。此種圖形不特表示總數之變動，且表示其各個成分之變動。



圖十八. 1923年各月某公司生產費用之分析

累積圖

吾人於展開一個數列時，往往不注重於數列中各個別項目之數值，而注意於其累積之數值。吾人在擬定每年生產數量之預算時，即有如此情形。在此種情形之下，吾人欲知實際生產量之總數與預計生產量總數之比，此則須藉圖形以資比較。下列數字即屬於此類資料而可適用此種圖形者。

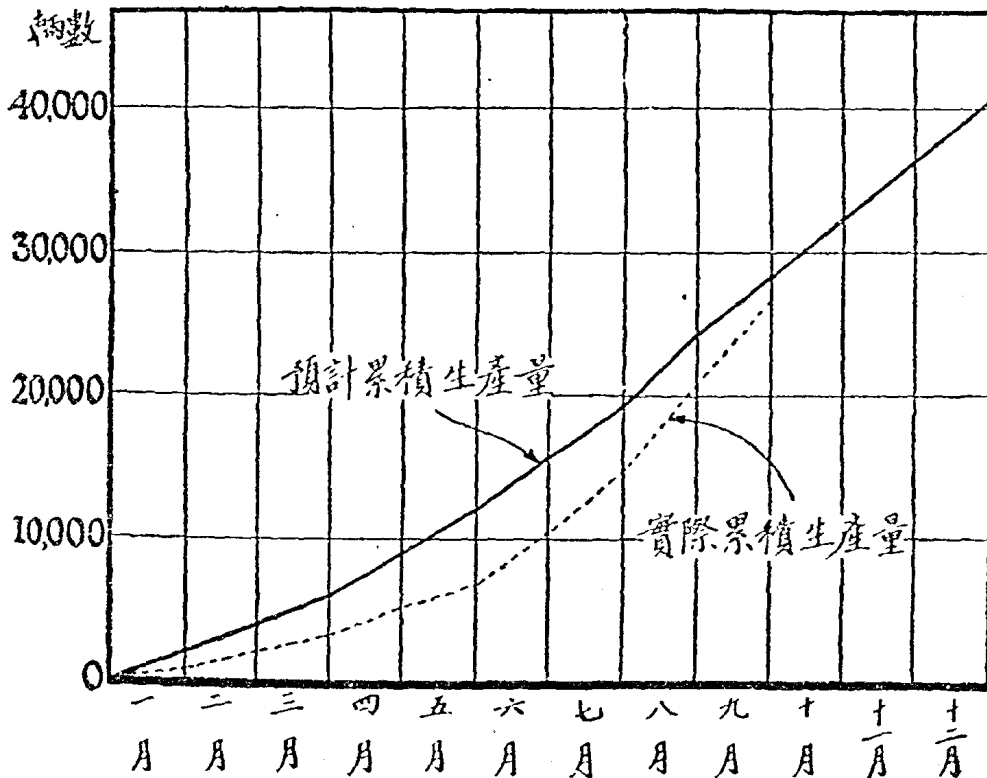
表 五

1924年美國斯彼德韋爾汽車製造公司預計累積生產量與實際
累積生產量之比較

月 份	預計生產量 (輛數)	預 計 累積生產量 (輛數)	實際生產量 (輛數)	實 際 累積生產量 (輛數)
一 月	2,000	2,000	750	750
二 月	2,000	4,000	1,250	2,000
三 月	2,000	6,000	1,250	3,250
四 月	3,000	9,000	2,000	5,250
五 月	3,000	12,000	1,500	6,750
六 月	3,750	15,750	3,750	10,500
七 月	3,750	19,500	4,250	14,750
八 月	5,000	24,500	6,000	20,750
九 月	4,000	28,500	5,750	26,500
十 月	4,000	32,500		
十一 月	4,000	36,500		
十二 月	4,000	40,500		

上表係假定為表示九月末之生產狀態者。

圖十九中之兩曲綫，係累積次數曲綫 (cumulative frequency



圖十九. 1924年美國斯彼德韋爾汽車製造公司預計累積生產量與實際累積生產量之比較

curve)。觀圖可知每月月末之實際生產量與預計生產量之關係，而實際生產量低於預計生產量者約為若干，亦可自尺度上推定之。按統計圖表相依為用，故兩者恆同時並列，閱者欲知差異之確實數量，可參閱附表中數字。此種累積次數圖應用頗廣，容於下章內說明之。

甘悌氏進行圖

前表內之統計資料，亦可應用甘悌氏（H. L. Gantt）所發明之圖形以表示之。關於此圖之構造及其種種用途，以本書篇幅無多，未能詳述，茲僅將其特點略加說明於下。

生產量預計表一經編竣，即可用甘悌氏進行圖（Gantt progress chart）以核對實際生產數量與預計生產數量之差別。例如表五內所示之生產量預計表擬就後，可將每月與每年之預計數量錄入如圖二十之圖形中。在每月地位左方之登錄，係表示各該月份之預計生產數量；在其右方之登錄，則表示各該月末之預計累積生產量。在該圖中年首兩月之工作結果均已表明。其較粗之綫係表示在該時期內之實際累積生產量為二千輛。在一月份與二月份欄內上方較細之綫係表示各該月份之實際生產量。倘該兩月份之每月實際生產量適與預計生產量相等，則其較細之綫當延長至各該月地位之全部。如某月實際生產量超過該月之預計生產量時，則用細綫二條表示之。

於此所應注意者，圖中所劃分之每月地位，雖代表相等之時間，然所代表之生產量則不相同。如一月份全距離五分之一之地位，係代表四百輛之生產量（一月份預計生產量為二千輛）；八月份全距離五分之一之地位，係代表一千輛之生產量（八月份預計生產量為五千輛）。閱此圖時，如欲知生產量之實際數字，則須參看表中每月之數額。

圖二十一表示九月末之生產狀態。圖上方所示之箭形，表示工作已完之時日。此時之實際累積生產量較預計生產量落後半個月，此可由圖

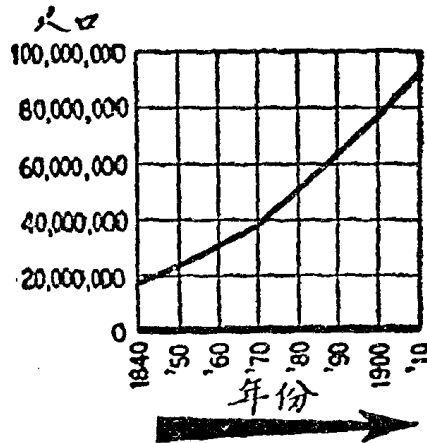
中箭形與粗綫之比較而得之。七、八、九、三個月生產量之細綫，係表明各該月實際生產量已超過預計生產量之情形。

甘梯氏圖在政府機關與商業組織方面為用頗廣。用此種圖形能將各地或各部分之進展狀況併列於一圖，故地位頗為經濟。此圖為表示工作進行狀態及比較實際工作與預計工作時最簡單最有效之圖示法，故應用時頗可促進行政管理之效率焉。

製圖之標準規則

在自然科學與社會科學方面以及企業界中，圖示法之應用已極普遍，惟因其用途廣泛，作圖之例，極不一致。為補救此項缺憾起見，在美國嘗有標準圖示法討論會之召集，對此論題有關之各團體均派代表出席。該會卒編就一報告書，介紹各種標準圖示法以供採用。是項報告係一種簡明圖示之範本，嘗發表於美國各種刊物內，惜未能廣為流傳耳。該會所提出之建議如下(註)：

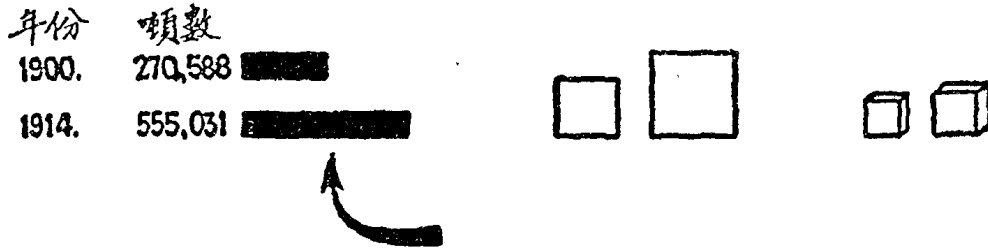
1. 圖之普通排列應自左而右。



圖例一

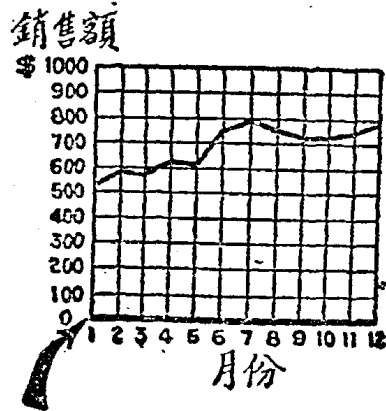
(註)此項報告書係 The Joint Committee on Standards for Graphic Representation 所編製者，該會主席為 Willard C. Brinton 氏。全文曾刊載於一九一五年出版之 “Quarterly Publications of the American Statistical Association” 第 14 卷第 790—797 頁。本書之轉載，事前嘗得 Brinton 氏之同意。紐約 American Society of Mechanical Engineers 備有印本，供各界參考之用。

2. 數量宜用綫之長度表示,因面積圖與體積圖最易引起誤解也。



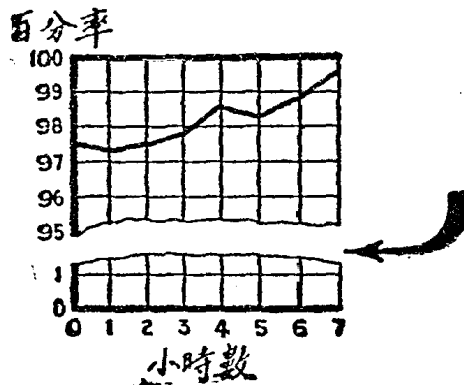
圖例二

3. 在選定綫圖之縱尺度時,宜將零綫繪入圖中。



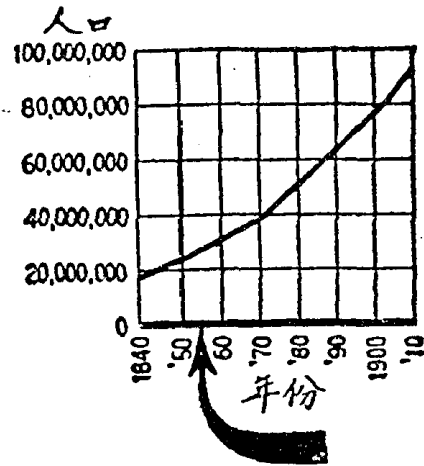
圖例三

4. 如綫圖中縱尺度上之零綫,不易繪入,可將該圖繪成橫截狀,使零綫仍能繪入圖中。

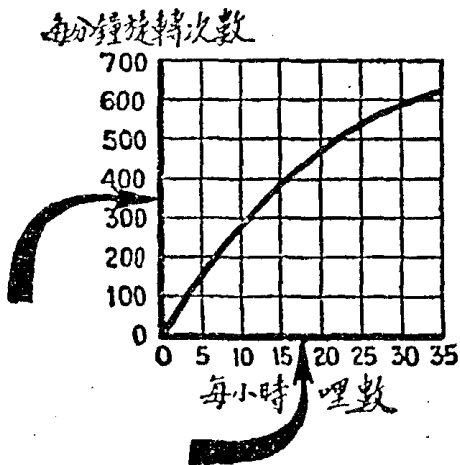


圖例四

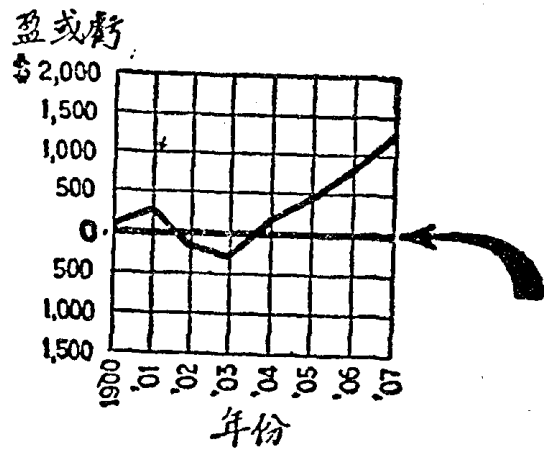
5. 尺度上之零綫須較行格為粗。



圖例五(甲)

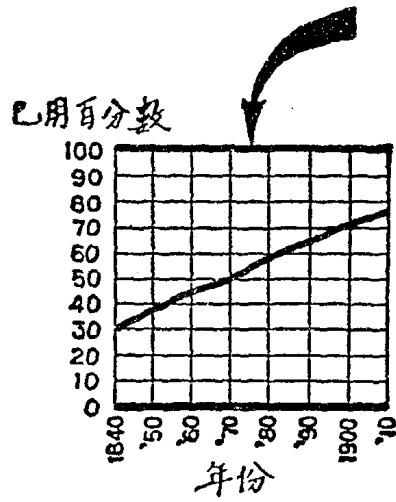


圖例五(乙)

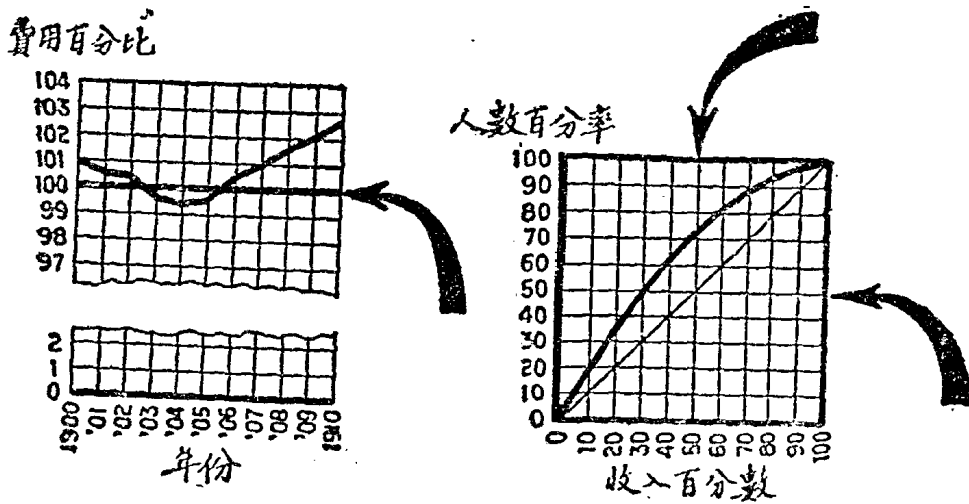


圖例五(丙)

6. 綫圖尺度如表示百分數，則代表百分之綫或其他用作比較標準之綫，例須較粗，俾便識別。



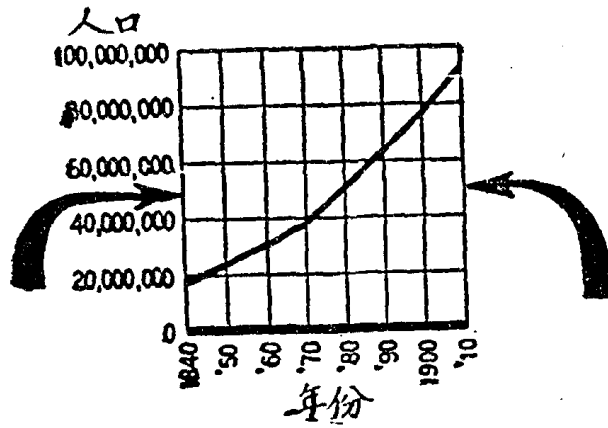
圖例六(甲)



圖例六(乙)

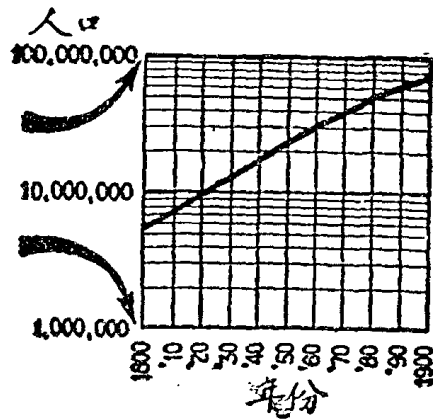
圖例六(丙)

7. 如圖之尺度表示時期，但所表示之時期並不完整時，則該尺度之起訖兩邊綫，不必加粗，因此兩邊綫實際並不表示時間之起訖也。



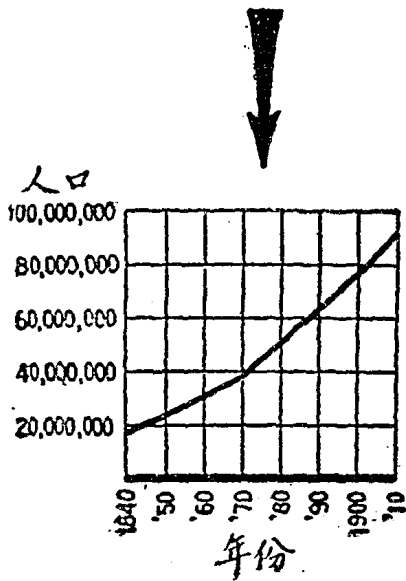
圖例七

8. 曲綫圖紙如用對數尺度，則其上下邊綫應取10之自乘數。

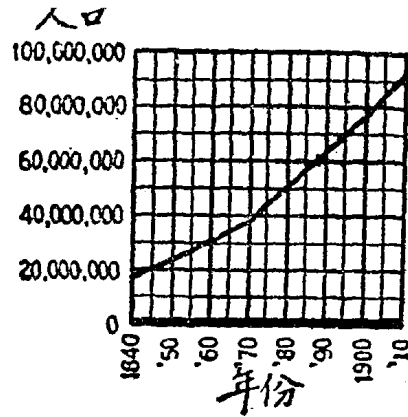


圖例八

9. 圖中行格之疎密，須求適當，不宜過密，以免眉目不清。

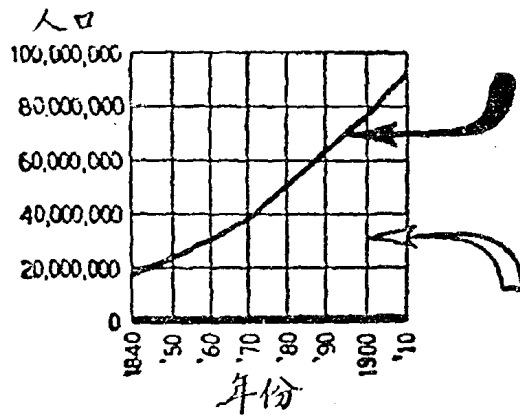


圖例九(甲)



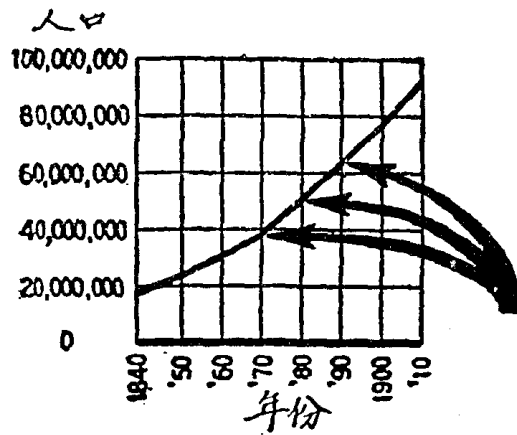
圖例九(乙)

10. 圖中曲綫應較粗於行格。

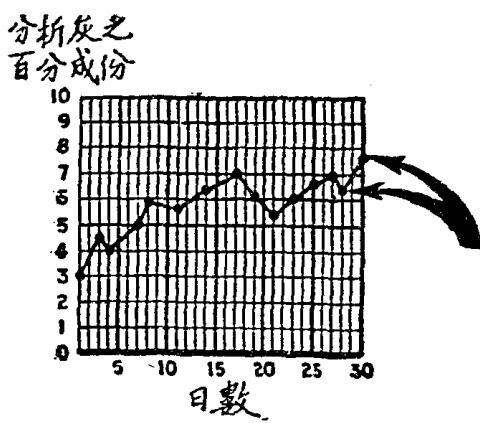


圖例十

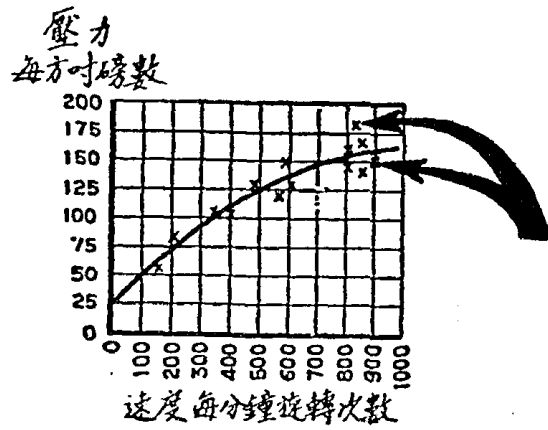
11. 如一曲綫係表示數個觀察數值時，應將圖中代表每個觀察之各點分別標明之。



圖例十一(甲)

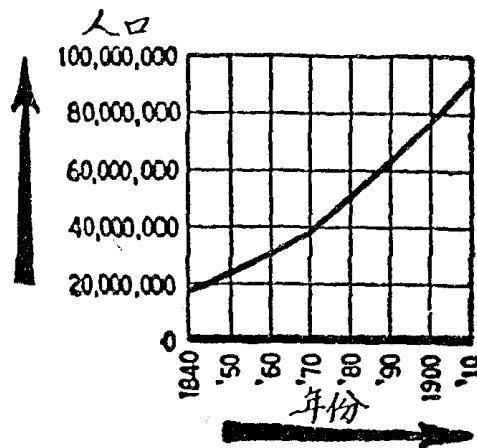


圖例十一(乙)



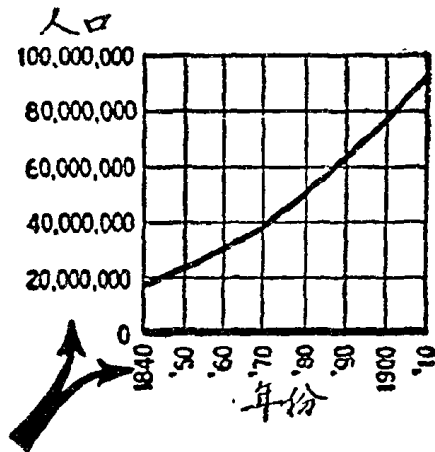
圖例十一(丙)

12. 綫圖之橫尺度,其讀法應自左而右,縱尺度應自下而上。

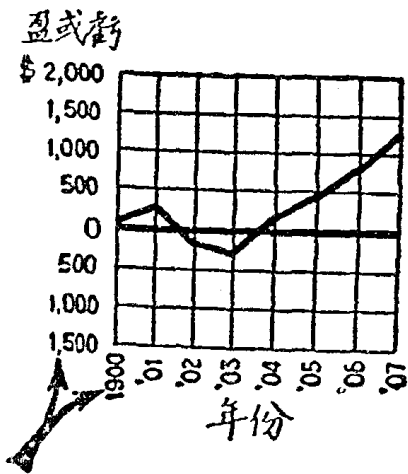


圖例十二

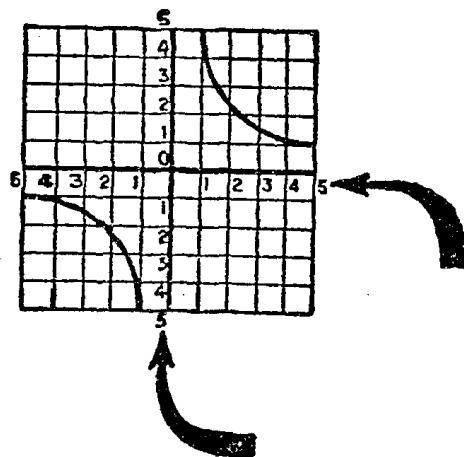
13. 圖中表示尺度之數字，須置於圖左及圖下，或沿縱橫軸排列。



圖例十三(甲)

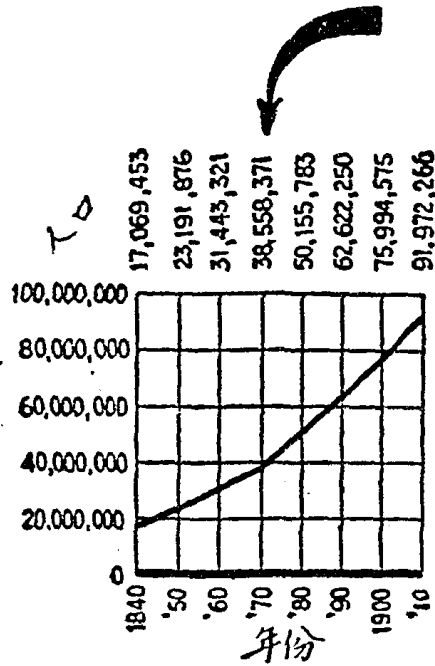


圖例十三(乙)

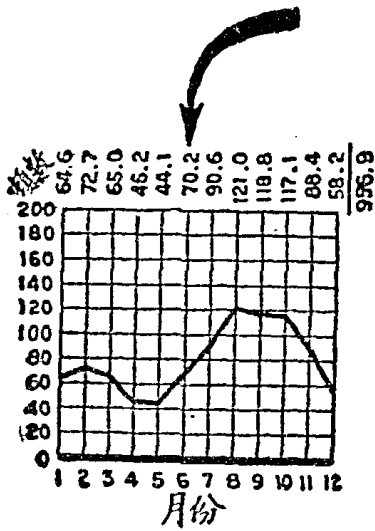


圖例十三(丙)

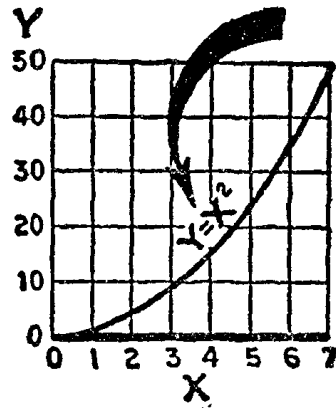
14. 曲綫所代表之數字資料或其公式，有時亦須列入圖中。



圖例十四(甲)

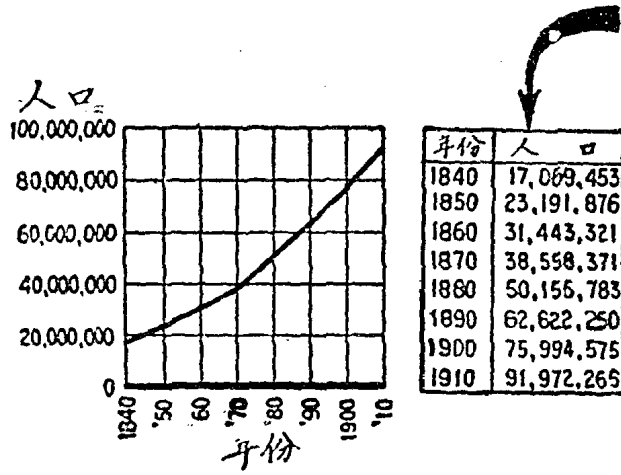


圖例十四(乙)



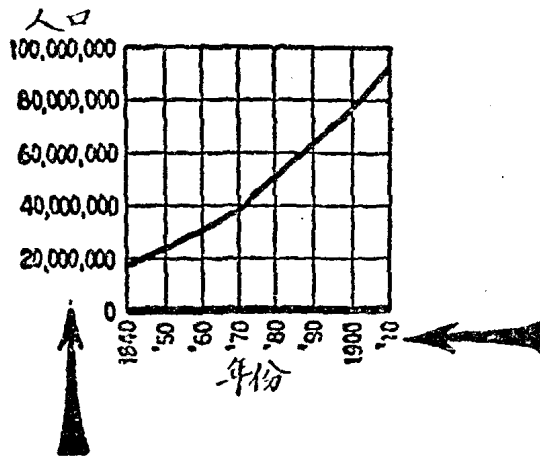
圖例十四(丙)

15. 如圖中不便列入數字時，應將此項數字製為表式，與圖並列。



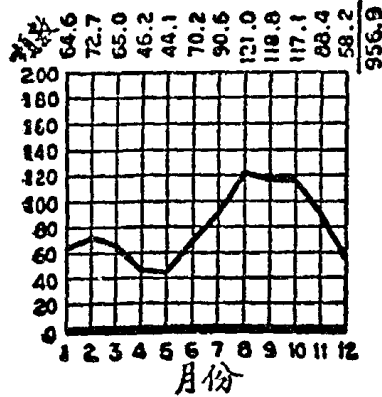
圖例十五

16. 圖上所有一切標識及數字，務須置於適當地位，使閱者以圖之基綫作底，或以圖之右邊綫作底閱看，不感困難。



圖例十六

17. 圖之標題應力求簡明完備，遇必要時，可用副標題或附註以補充其意義。



1914年各月第二廠之鑄塊產量

產量以公噸計，

碎塊鋁之銷售額不在內。

圖例十七

參考書

1. 關於數學概念者

關於圖示法與對數用法之基本數學概念，在代數學與解析幾何學之任何種標準教科書中均有述及。在下列各書內論述頗詳，足供研究統計學者之參考。

Griffin, F. L. "Introduction to Mathematical Analysis."

Karsten, Karl G. "Charts and Graphs".

Lipka, Joseph. "Graphical and Mechanical Computation"

Mellor, J. W. "Higher Mathematics for Students of Chemistry and Physics"

Schultze, Arthur. "Graphic Algebra".

Steinmetz, C. P. "Engineering Mathematics".

Whitehead, A. N. "An Introduction to Mathematics".

2. 關於圖示法者

Brinton, W. C. "Graphic Methods for Presenting Facts"

Clark, Wallace. "The Gantt Chart".

參考書(續)

Field, J. H. "Some Advantages of the Logarithmic Scale in Statistical Diagrams". *Journal of Political Economy*, October, 1917.

Fisher, Irving. "The 'Ratio' Chart". *Quarterly Publications of the American Statistical Association*, June, 1917.

Maskell, A. C. "Graphic Charts in Business".

Maskell, A. C. "How to Make and Use Graphic Charts".

Karsten, Karl. "Charts and Graphs".

(上列各書之出版書局及年月,詳載於書末之圖書目錄中。)

第三章 統計資料之整理：次數分配

研究商業與經濟之統計學家，其所任之工作，為整理、分析與解釋關於商業事務與經濟狀況之數量資料。此外尚須加以蒐集資料之工作，惟通常所用之資料每由已經採集之原始資料(primary data)或次級資料(secondary data)編製而得者。

惟時間數列(time series)及其他與時間無關之資料，其所用分析方法並不相同。研究時間數列之主要目的，在於測度與分析變量之數值在時間上所發生之變動；例如吾人欲研究若干年內銷售額之變動，白煤產量之變動，或一般物價水準之變動。但若研究某時期內收入之分配，則研究之手續與前者迥異。此時吾人所欲知者，為各種收入階級中所包含之人數，故此處之整理問題，在於斷定變量之各值所重覆發現之次數及其分配狀態。此種資料一經整理後，即成一次數數列(frequency series)，而與時間數列迥不相同。研究此兩種數列所用之分析方法互異，故有分述之必要。本章所論述者，為不因時間而有變動之各種資料之整理與其初步分析之方法。

未經整理之資料

吾人所處理之數量資料，在未經整理以前，其原始狀態率皆凌亂，而乏整齊之形式與結構。此種資料或為企業方面所具備之生產或銷售之紀錄，或為搜集所得之各種物價。即在應用別種機關所搜集之資料時，此項資料或已列成表式，但其排列與格式每不合於考查人(investigator)研究之用。是以統計之初步工作，在於將數字編成一適當形式以求合用，俾與性質相似之資料比較，且可為進一步分析之準備。科學方法包括觀察、推演與證實三步驟，前已言之；數字資料乃係觀察之所得，必須先行整理，使成為整齊之形式，始可進而推演。

下列數字係代表某製造廠工人210人在某週內按件計工之工資數，此為資料之未經整理者之一例。

工人 210 人在某週內每人所得工資數

\$26.25	\$28.70	\$24.15	\$29.75	\$29.20	\$30.60	\$23.40	\$24.75
26.70	24.35	25.75	27.20	28.30	25.25	27.75	27.60
28.20	27.30	27.80	26.35	27.40	28.30	26.60	25.75
27.70	28.60	25.30	27.80	26.40	27.30	28.35	27.00
24.30	27.80	27.60	26.30	27.40	23.50	29.60	27.80
27.60	25.35	27.55	29.00	24.10	27.00	24.50	27.25
26.15	29.30	23.10	27.10	28.50	27.45	26.15	28.35
27.95	25.55	27.55	26.60	24.25	30.00	28.55	28.00
27.30	27.90	25.25	24.10	27.45	24.55	26.55	27.55
26.75	31.00	24.00	25.35	26.50	28.30	27.95	26.55
30.25	28.55	26.75	24.60	25.75	26.55	27.80	28.90
29.55	30.00	24.60	25.75	26.30	27.00	28.25	25.25
25.75	26.25	26.30	26.75	27.90	28.30	25.70	26.30
26.60	27.00	30.75	28.60	28.10	23.50	24.75	25.15
26.30	27.25	28.15	29.10	30.10	29.90	28.55	27.30
26.55	27.55	23.00	24.50	22.85	26.55	27.55	28.10
30.70	28.60	27.90	26.80	24.10	25.25	26.30	27.90
26.90	25.30	25.80	28.85	27.55	27.30	25.00	26.00
26.55	27.80	28.60	30.55	29.50	24.10	25.15	27.15
28.10	26.30	27.10	24.60	27.80	26.30	27.90	29.80
24.10	25.15	27.50	24.25	25.70	26.80	30.15	29.30
28.15	28.65	24.55	25.85	26.10	27.00	26.80	27.55
29.00	23.00	28.60	29.30	28.55	28.80	27.55	23.60
26.10	27.15	25.75	26.80	27.15	26.30	28.55	25.80
24.55	25.80	26.75	27.30	27.55	28.25	25.60	26.30
26.85	27.30	28.10	32.00	28.15	26.30	27.75	26.25
28.60	26.00						

整列

前項數字如依大小次序排列，即可成一整齊連貫之形式，其所包含

之最高數與最低數間之全距(range),及在此全距內各項數字分配之概況,可因排列而易於明瞭,作為進一步整理之準備。此項數字依大小次序所排成之整列(array)如下:

整列:工人 210 人在某週內每人所得工資數

\$22.85	\$25.15	\$26.15	\$26.75	\$27.45	\$27.95	\$28.60
23.00	25.15	26.15	26.75	27.45	25.95	28.65
23.00	25.15	26.25	26.80	27.50	28.00	28.70
23.10	25.25	26.25	26.80	27.55	28.10	28.80
23.40	25.25	26.25	26.80	27.55	28.10	28.85
23.50	25.25	26.30	26.80	27.55	28.10	28.90
23.50	25.25	26.30	26.85	27.55	28.10	29.00
23.60	25.30	26.30	26.90	27.55	28.15	29.00
24.00	25.30	26.30	27.00	27.55	28.15	29.10
24.10	25.35	26.30	27.00	27.55	28.15	29.20
24.10	25.35	26.30	27.00	27.55	28.20	29.30
24.10	25.55	26.30	27.00	27.55	28.25	29.30
24.10	25.55	26.30	27.00	27.60	28.25	29.30
24.10	25.60	26.30	27.10	27.60	28.30	29.50
24.15	25.70	26.30	27.10	27.60	28.30	29.55
24.25	25.70	26.30	27.15	27.70	28.30	29.60
24.25	25.75	26.35	27.15	27.75	28.30	29.75
24.30	25.75	26.40	27.15	27.75	28.35	29.80
24.35	25.75	26.50	27.20	27.80	28.35	29.90
24.50	25.75	26.55	27.25	27.80	28.50	30.00
24.50	25.75	26.55	27.25	27.80	28.55	30.00
24.55	25.75	26.55	27.30	27.80	28.55	30.10
24.55	25.80	26.55	27.30	27.80	28.55	30.15
24.55	25.80	26.55	27.30	27.80	28.55	30.25
24.60	25.80	26.60	27.30	27.80	28.55	30.55
24.60	25.85	26.60	27.30	27.90	28.60	30.60
24.60	26.00	26.60	27.30	27.90	28.60	30.70
24.75	26.00	26.70	27.30	27.90	28.60	30.75
24.75	26.10	26.75	27.40	27.90	28.60	31.00
25.00	26.10	26.75	27.40	27.90	28.60	32.00

次數表

上項整列所表現數字之形式，雖較未整理前整齊明瞭，但欲使讀者對於該項資料所包含之意義完全領會，猶須作進一步之整理。工廠經理雖可由此整列而知某週內工人所得最低工資數為 \$22.85，最高工資數為 \$32.00，及大部分工人所得工資數在 \$25.00 與 \$29.00 之間，但所得僅此，而對於資料之真義，猶有失之含混之處，如用分組手續，將工資數分為數組 (class)，而以各項數字分別歸納於各組中，即可得一簡潔之工資分配表。茲將分組後所得之表列下。表中每組之距離為美金二元，此距離謂之組距 (class-interval)。

表 六

按某週工資分組之工人次數分配(組距 = \$2)

某週工資數	獲得左列工資數之人數 (次數)
\$22.00 to \$23.99	8
24.00 to 25.99	48
26.00 to 27.99	96
28.00 to 29.99	47
30.00 to 31.99	10
32.00 to 33.99	1
	210

此表為原來數字之縮影，不特明示吾人以工資之全距，且表述該全距內工人 210 人所得工資數之分配狀況。此表已將該資料所有繁瑣之事實略去，吾人所可得者為在某週內工資數之在 \$24.00 至 \$25.99 之間者為 48 人，但此 48 人在該組內之分配情形，則無從知悉。就表中數字觀之，此 48 人所得之工資數或皆為 \$24.00。但欲將資料分組簡縮，則失却繁瑣與詳盡之事實，要所不免。

如將組距縮小，則失却詳盡之程度可以減少，然同時組數必增多，而表之形式亦必趨於繁瑣，反覺眉目不清。下列各表係用同一資料，以

一元,五角,二角五分爲組距而分組所得之結果。

按某週工資分組之工人次數分配

表 七 (組距=\$1)		表 八 (組距=\$0.50)		表 九 (組距=\$0.25)	
某週工資數	次數	某週工資數	次數	某週工資數	次數
\$22.00—22.99	1	\$22.50—22.99	1	\$22.75—22.99	1
23.00—23.99	7	23.00—23.49	4	23.00—23.24	3
24.00—24.99	21	23.50—23.99	3	23.25—23.49	1
25.00—25.99	27	24.00—24.49	11	23.50—23.74	3
26.00—26.99	42	24.50—24.99	10	23.75—23.99	0
27.00—27.99	54	25.00—25.49	12	24.00—24.24	7
28.00—28.99	34	25.50—25.99	15	24.25—24.49	4
29.00—29.99	13	26.00—26.49	22	24.50—24.74	8
30.00—30.99	9	26.50—26.99	20	24.75—24.99	2
31.00—31.99	1	27.00—27.49	24	25.00—25.24	4
32.00—32.99	1	27.50—27.99	30	25.25—25.49	8
	210	28.00—28.49	17	25.50—25.74	5
		28.50—28.99	17	25.75—25.99	10
		29.00—29.49	7	26.00—26.24	6
		29.50—29.99	6	26.25—26.49	16
		30.00—30.49	5	26.50—26.74	10
		30.50—30.99	4	26.75—26.99	10
		31.00—31.49	1	27.00—27.24	11
		31.50—31.99	0	27.25—27.49	13
		32.00—32.49	1	27.50—27.74	14
				27.75—27.99	16
			210	28.00—28.24	9
				28.25—28.49	8
				28.50—28.74	14
				28.75—28.99	3
				29.00—29.24	4
				29.25—29.49	3
				29.50—29.74	3
			29.75—29.99	3	
			30.00—30.24	4	
			30.25—30.49	1	
			30.50—30.74	3	
			30.75—30.99	1	
			31.00—31.24	1	
			31.25—31.49	0	
			31.50—31.74	0	
			31.75—31.99	0	
			32.00—32.24	1	
				210	

上列四表(表六至表九)係根據同一資料,依四種簡縮之程度整理而成者。表六、表七及表八具有一共同之特點:即在首尾各組內次數(frequency)較小,愈近中間則各組次數愈見增多。組數愈多,次數分配不規則之程度亦愈著。表九之組距爲二角五分,計有 38 組,在其全距內

次數之分配，極無規則，毫不合於對稱(symmetry)狀態；其他各表之結構，較為整齊有序，與對稱狀態亦較接近。各表所示者皆為工資資料之簡縮形式，閱表即可知該工廠工人在某週內所得工資之大小及其分配之情形，比較未經整理以前紛亂無章之數字明瞭多矣。蒐集之資料經此整理後，謂之次數分配 (frequency distribution)，其目的如其定義所述，在於用簡縮之形式表示一變量之各數值在其全距內之分配情形。編製次數分配表為整理與分析此種數量資料之初步手續。

編製次數表之步驟

次數表之編製方法，前節已述其概要，惟關於編製方面之重要各點，尙未述及，此處應作簡略之論述。上例中第一步手續係將各項數字依大小排成整列，惟實際製表時，此項排列手續並非必要。吾人可先就資料查明其上下兩極限，並決定分組之數目，將各組距填入適當之空白紙上，而後將各項數值一一歸納於各組內。各項數值既已歸入各組，即可進而計算各組之次數，而後製成上列表式。此項手續雖屬簡單，然牽涉數要點，不可不分別討論之。

組距之大小

決定組距之大小時(亦即決定組數時)，所應注意之基本要點，在使每組內項目之分配，勿與勻整分配相差過遠，蓋在解釋次數表與以後根據此表從事計算時，例須用每組之中點 (mid-value) 以代表該組內各項目之數值也。例如根據表八計算時，在 \$26.00—\$26.50 一組內 22 項之數值，均須用該組之中點 \$26.25 代表之。此係根據各組內項目分配勻整之一種假定，與實際情形固不能全相符合，試閱表八之原來數值，即知此項假定之不準確，然欲求絕對準確，則非將原來資料中每個不同之數值各成一組不可，但事實上不得不將資料簡縮，故簡縮時務須在不

違反其他條件下，使此種差誤儘量減少。如表六所定之組距即嫌過大，與該資料不合。

欲求適合上項假定，必須將組數增多，已如上述，然分組之第二條件常與此抵觸。此第二條件為分組時各組次數之分配必須整齊有序。如組數過多，組距過小，則其次數分配必欠整齊而乏結構，表九即不合於此項條件；且若組數較少，處理資料當必較易，而其意義亦易於領會。通常組距之決定，宜使組數在 10 與 25 之間。至於某表之組數應為若干，仍須視資料之性質而定。前列表八中，組距定為美金五角，似與上述兩條件最能適合。

組限之位置

組限 (class limit) 位置之決定亦頗關緊要，如選擇得宜，則製表手續較為簡單，而此後計算亦可省便。如所定組限與組距皆為整數，製表時尤為便利。每組中點如為整數，則計算平均數與其他統計量數時可減省手續。欲選擇適當之組限與中點，莫如將組距之數值定為 5 或其倍數，惟此係就一般而言，在應用時仍須斟酌資料之性質而定也。

數字資料之分配，常有集中於若干數值之傾向。茲以下列數字為例，將此點加以說明。下表為一九二一年會員銀行向聯邦準備銀行貼現所用商業票據之件數，依貼現率之高下分配而成之次數表之一部。

貼現率(百分率)	票據件數
6	18,970
6 $\frac{1}{4}$	697
6 $\frac{1}{2}$	4,616
6 $\frac{3}{4}$	135
7	17,362
7 $\frac{1}{4}$	10

上表之次數頗有密集於整數之趨勢，其次則含有 $\frac{1}{4}$ % 之各個百分率，其集中之次數亦多，而介乎鄰接兩個百分率間，則表中並一項而無

之。如遇此種資料，分組時各組之中點應選取項數密集之數值，選定組距時，亦應注意及此，蓋根據次數表計算時每組內之各項數值均以各該組之中點作為代表也。上述資料之組距如定為 $\frac{1}{2}\%$ ，則各組之排列應為 $5\frac{3}{4} - 6\frac{1}{4}$ ， $6\frac{1}{4} - 6\frac{3}{4}$ ，……，而非 $6 - 6\frac{1}{2}$ ， $6\frac{1}{2} - 7$ ……矣。

組限與觀察之確度

製表時組限之定義務須明確表示，使每組之全距無含混之弊，而各項數值歸入各組時，亦可不致發生困難。如下列之表式，為吾人所偶見。

組 距	次 數
0—10	3
10—20	8
20—30	15
30—40	6
40—50	2

上表內之組限如不附以解釋，則遇某項數值適為10時，其應歸入第一組抑第二組，不免發生問題。故每組之組限務必確切標明如下式，使其定義無含混不清之弊。

組 距	次 數
0— 9.9	3
10—19.9	8
20—29.9	15
30—39.9	6
40—49.9	2

惟組限之定義，與觀察之確度有關，如觀察之確度至十分之一單位為止者，則組限依上式標明後，前項困難當可避免；但若觀察之確度僅至單位為止者（如10之數值係介於 9.5 與 10.5 之間），則遇有與上項組限相同之數值，歸組時仍有困難。在此情形下，可將該項數值之次數，平分為二，各以半數列入鄰接兩組之中。

關於組限問題，猶爾氏(Yule)^(註)曾定一法則，謂組限之數值應較

^(註)見 “An Introduction to the Theory of Statistics” 第 81 頁。

紀錄之數值展後一位小數。試以前述數字為例，如觀察之確度至十分之一單位為止者，則紀錄時作為 9.9 之一項數值，其實際數值係介乎 9.85 與 9.95 之間，故組距應定為 0—9.95, 9.95—19.95……，各組始有準確之意義。依此分組，則各組之中點當為 4.95, 14.95……。編製與應用次數表時，皆應注意及此。無論何時，組限之數值應視觀察之確度而定。

製表時如將各組之上下兩限併列表內，則製表工作較為簡單；如僅列各組之下限 (lower limit)，或各組之中點，則歸組時每易發生錯誤。又若用表中數字計算時，更須於表內加列一欄，以記明各組中點之數值。

其他要件

表中各組組距之大小，必須一致，以便彼此比較。然組距大小不一之表式，亦偶有所見。表中之組距如不一致，或為 5，或為 10，則因組距之大小各異，以致各組不能相互比較。惟此種表式或因注重某種距離內次數之分配情形，故有此需要，但遇此情形可製兩表，其主要表中組距仍須一致，而另附一輔助表以說明某一距離內次數分配之情形可矣。

表中組距除須一致外，其意義更須肯定。前例按件計算工資之工人，其工資數在三十元及以上者，如歸入於標明為“\$30及以上”之一組內，則該組之上限 (upper limit) 究為幾何，無從斷定；且根據此表計算時，此種錯誤關係尤鉅。惟若表中包含數項極端之數值，以致組距不能不如此標明時，則須於表末另加附註，載明此數項之實際數值，俾讀者得以了解。

前兩節內所述之差誤，可舉例(參看表十)說明於下：

在此例中，其最小組下限不明，最大組上限不明，故其組距無由斷定，而中間各組之組距，又不一致，組距為 \$1.00 者兩組，\$1.50 者一組，\$2.50 者一組。此種表式殊無價值之可言。

表 十

按每日所得工資數分組之 ABC 公司工人之次數分配

工資組	每組工人數 (次數)
\$3.00 以下	15
3.00—3.99	30
4.00—4.99	40
5.00—6.49	30
6.50—8.99	10
9.00 及以上	5
	<hr/> 130

統計表之結構

上節所述，為關於編製次數表時所發生技術方面之種種問題，對於全表之格式，縱行(column)、橫行(row)之排列，表之標題(title)及其標識(notation)等均未涉及，排列表式之各種普通原則亦未論述。列表細則多不勝舉，此處所欲論述者，僅為關於統計表組織方面所應考慮之各點。

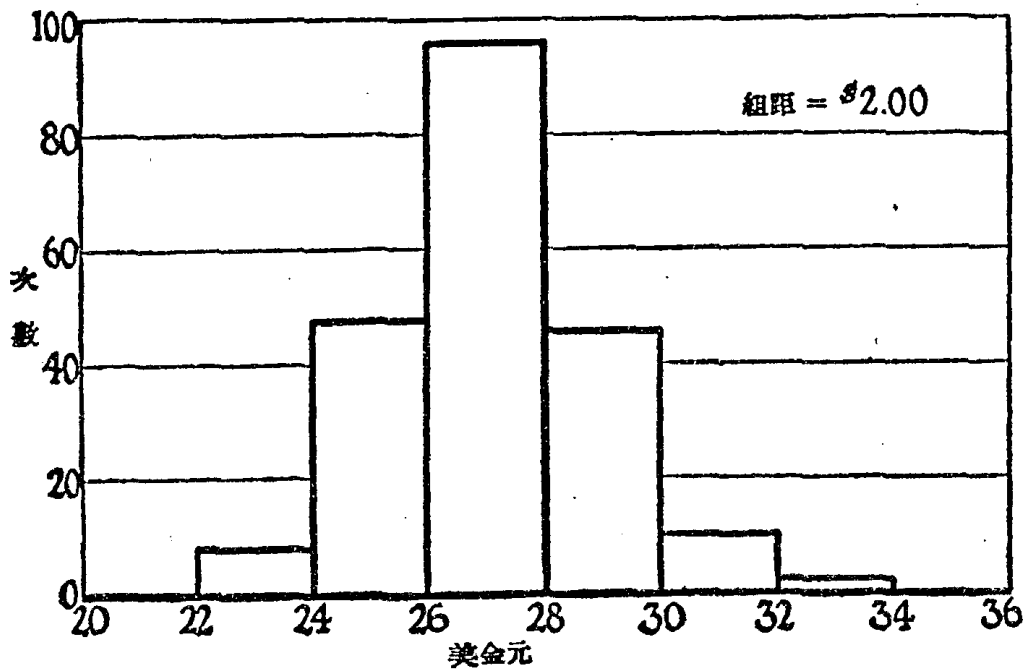
統計表為以簡略之形式表述一羣數量資料之一種方法，故其形式必須簡潔明瞭而有意義，且須易於解釋，否則即不能具備製表與分類之功用。雜亂無章之統計表與缺乏條理之論文無異。編製統計表必有其目的，排列表式時應將此目的明白表現。排列之方法，雖因資料之性質而異，然對於標準規律務須嚴格遵守。下列各項普通原則，可為選擇統計表之形式與排列時之一助。

1. 表之標題應以明瞭簡潔之文字，將表內所列資料詳細載明。
2. 縱行橫行之標目務求簡明，不宜含糊。
3. 變量之尺度應自左而右或自下而上逐漸增加。
4. 縱行橫行可編列號次，以便引證。

5. 表中所用之度量衡單位務宜明白詮釋。
6. 資料之來源務須一一註明。
7. 每一表式應自成一單位，所列資料之意義，須不費解釋而自明。如有解釋之必要時，應將註釋列入表內，或於表末加註。

次數分配之圖示

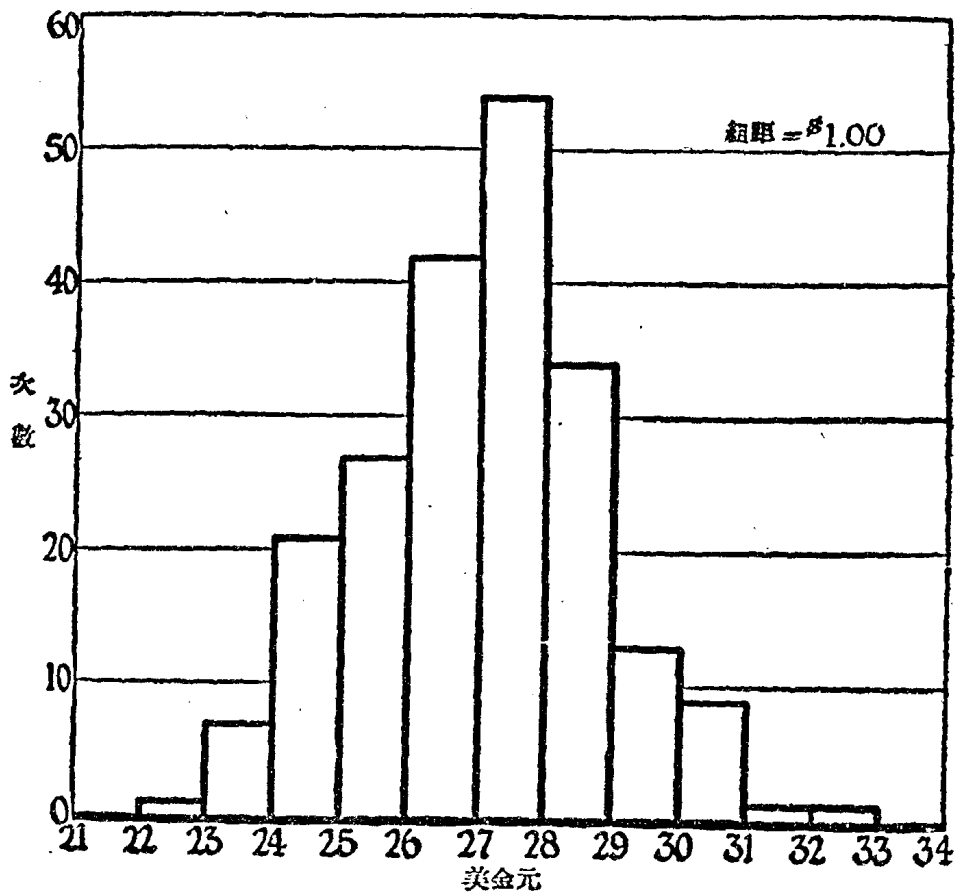
前節所述之次數分配，既將資料用簡明之形式表示，且可作進一步研究之準備，故在統計上實具有極大之功用。此種分配不僅可以表式顯示，且可用第二章內所述縱橫坐標制之圖解表示之。次數分配之各項特性，用圖解表示，最為明顯。



圖二十二. 直方圖：按某週工資數分組之工人 210 人之次數分配(組距 = \$2.00)

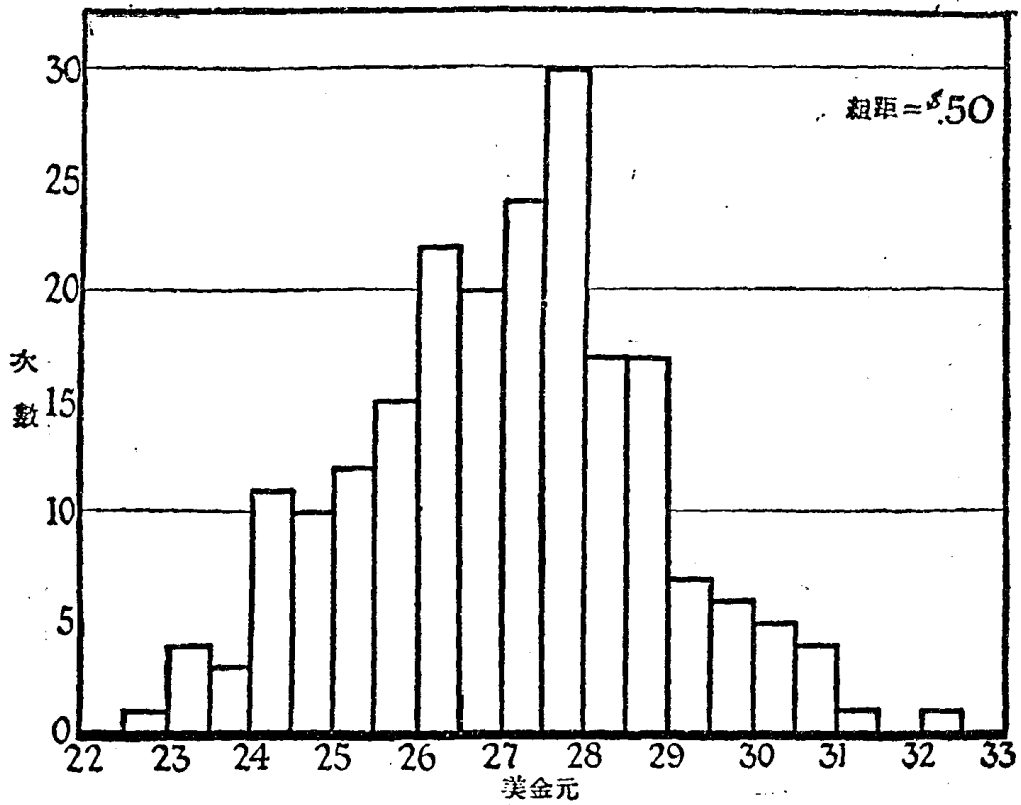
表六所列為工人 210 人某週所得之工資數，組距為美金二元；圖二十二即為該項資料之圖解。圖中組距沿 x 軸量之，次數沿 y 軸量之，兩軸各用相當之尺度表明；惟須注意者，橫坐標之尺度不自零點起，而自 \$20 起。為圖示便利起見，自 0 至 \$20 之部分未列入圖中，讀者對此兩變

量關係之印象不可因此發生錯誤。繪製此種直方圖 (column diagram or histogram) 時，應將每個組距上下限之兩點，用橫綫接連之。圖中各條形之面積與其所代表之次數成正比例，其總面積係代表各組次數之和，即 210。此種圖解頗能明白表示各工資組內工人分配之人數，吾人觀圖則對於次數分配之狀況，可獲極明確之印象。

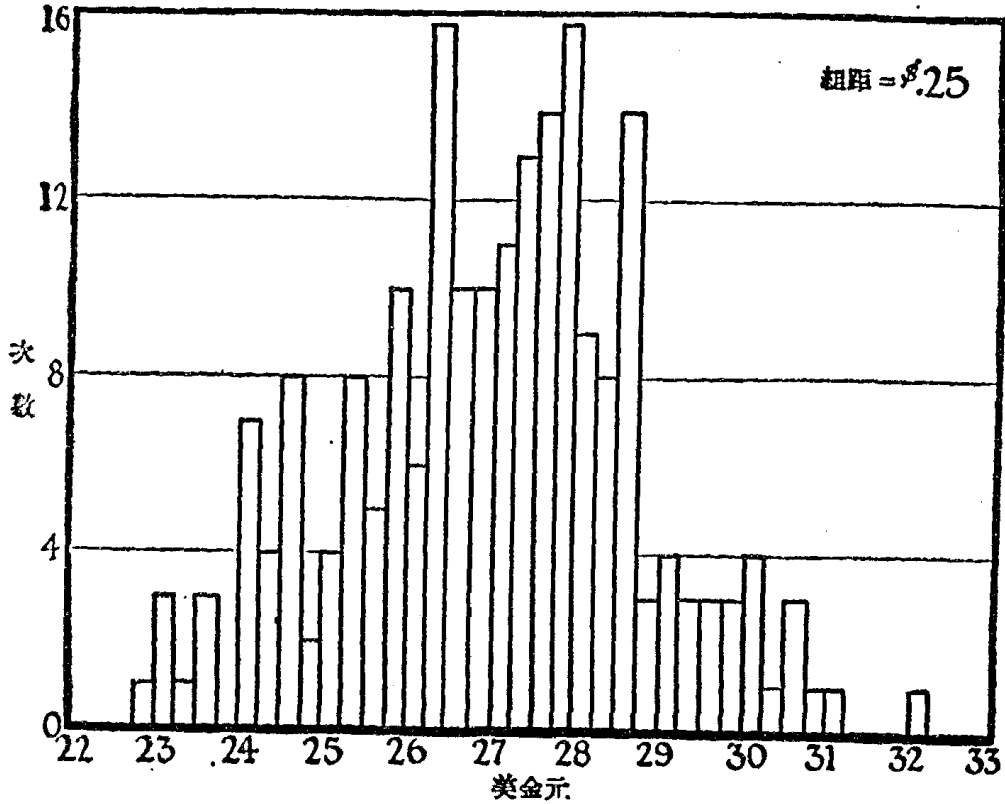


圖二十三. 直方圖：按某週工資數分組之工人210人之次數分配(組距 = \$1.00)

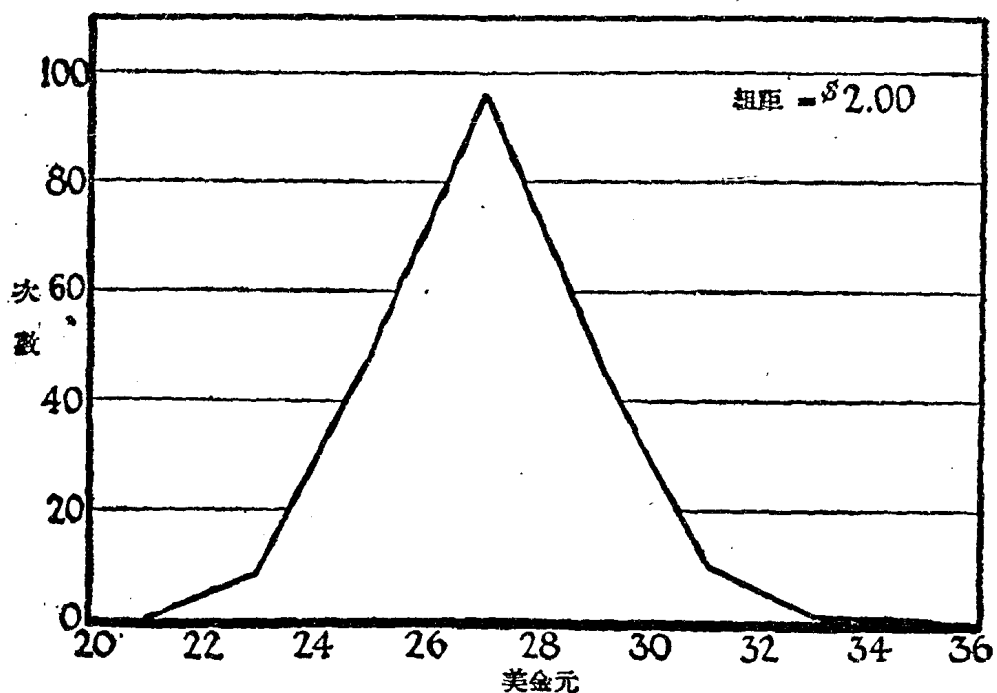
惟圖二十二之組距猶嫌過大，以致頗有與事實不符之處。表中將詳情刪路過多，足使吾人不易獲得項目分配之真確觀念。圖二十三係以美金一元為組距時所得工人分配狀況之直方圖。該圖組距較小，條形亦較狹，所得圖解較為整齊而對稱。圖二十四亦然，是圖係以美金五角為組距所得工人分配狀況之圖解。圖二十五所示之分配狀況，係以美



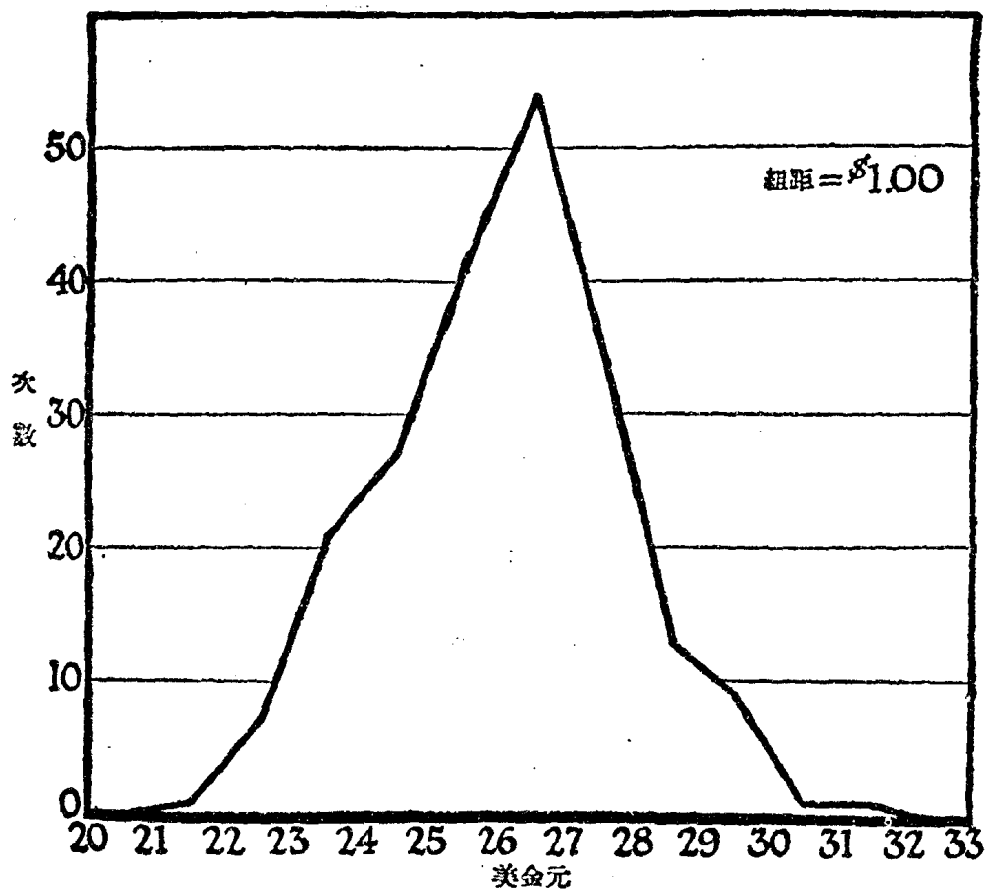
圖二十四. 直方圖:按某週工資數分組之工人210人之次數分配(組距=\$.50)



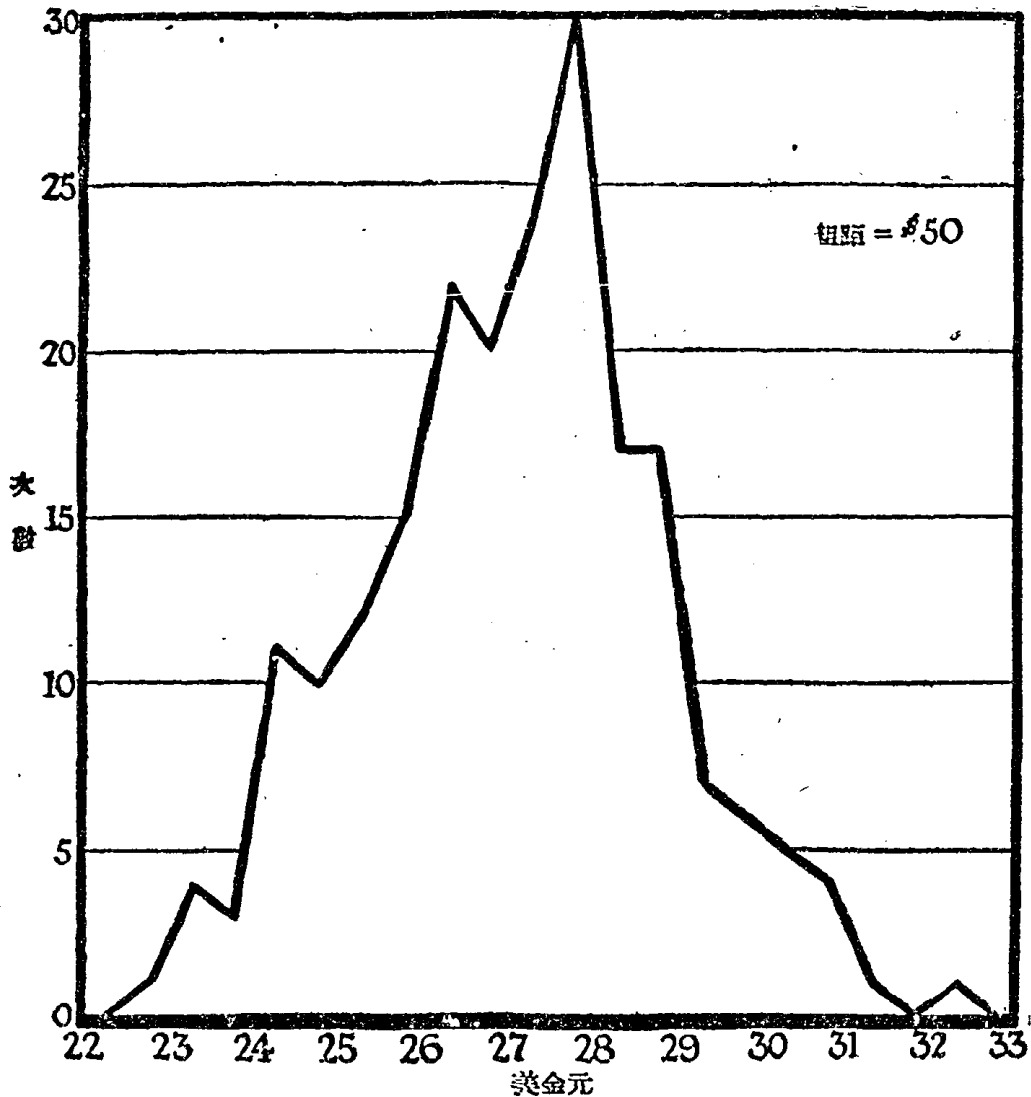
圖二十五. 直方圖:按某週工資數分組之工人210人之次數分配(組距=\$.25)



圖二十六. 次數多邊圖：按某週工資數分組之工人210人之次數分配(組距 = \$2.00)



圖二十七. 次數多邊圖：按某週工資數分組之工人210人之次數分配(組距 = \$1.00)



圖二十八. 次數多邊圖:按某週工資數分組之工人210人之次數分配(組距=\$.50)

金二角五分爲組距,惟該項組距用於此種資料失之過狹,故該圖之結構亦不規則(吾人應注意上列四圖之縱尺度並不相同,故欲比較各圖中各組之次數,不能逕以直方之高度作爲標準,而必須參考尺度上之數字)。

與圖二十二、二十三、二十四相對之次數多邊圖 (frequency polygon) 爲圖二十六、二十七及二十八。此種圖解係以每組中點沿橫軸量之,每組次數沿縱軸量之,然後將所定各點接連成一斷續綫而得者。圖中將最高組之上組及最低組之下組一併繪入,該兩組之次數均爲零,故此多邊圖之兩端,在該兩組之中點處,得與基綫相接。次數多邊圖中曲綫下之全部面積,表示次數之總數,但各組距內所包含之面積,並

不與各組之次數成正比例，因每組內次數分配並不勻整也。在各組中點處所有縱坐標之高度，則係代表各組之次數。

曲綫之修勻

變更組距之大小，又可得下述之結果。當組距逐漸縮小至某程度時，其所得直方圖與多邊圖亦漸見勻滑而整齊；逾此程度，則勻整之狀態漸見消失而呈現間斷之處，蓋在組距較大時所得各組次數有規則之變化，至此已為次數變化不規則之各組所破壞。在圖二十五中即可見次數變化不規則之一斑。此種不規則狀態，顯與次數分配之普通法則相悖；按普通法則，當工資數逐漸增加時，各內工人組數亦逐漸增加，至\$27.50附近而達最高點，由此點起，工人數復逐漸減少，至最後上限為\$32之一組時，工人數僅得一人。此210工人所擔任之工作相同，而其收入之大小係取決於各個工人工作之速率與技術之高下，故各工資組內次數之增減自應表現一有規則之狀態。設吾人所用工資資料非一週之數字，而為52週之數字，用此210人每人在全年內52週之平均工資，則雖用較小之組距，而所得次數分配必較前此為勻整，蓋各週之特殊變動，經此平均後業已消除也。又設吾人採用10,920工人(52×210)一週內之工資數，則所得結果必與用一年內之平均數字者相同。由是知吾人倘欲得整齊勻滑之狀態，不僅須縮小組距，且須添加項數，使少數之意外變動，得以消除；否則如僅將組距縮小，則所得圖形必將如圖二十五所示。雖然，增加項目常為事實所不許，吾人自不得不採用修勻曲綫之方法，以求組距極小、項目極多時次數分配之勻整狀態。採用此法吾人可得一勻滑而有規則之次數曲綫。

此種修勻後之次數曲綫(smooth frequency curve)，可以代表資料所應有之分配狀態。前節嘗謂次數多邊圖中，因每組內次數分配之不勻整，故各組距內所包含之面積並不與各組之次數成正比例；惟在修勻

表 十 一

1918 年美國個人收入之次數分配

(包括收入之在 \$4000 以下者)

收入組(註一)	人數(註二)
\$ 0 to \$ 100	62,809
100 to 200	103,704
200 to 300	209,087
300 to 400	489,963
400 to 500	961,991
500 to 600	1,549,974
600 to 700	2,154,474
700 to 800	2,668,466
800 to 900	3,013,034
900 to 1,000	3,144,722
1,000 to 1,100	3,074,351
1,100 to 1,200	2,850,526
1,200 to 1,300	2,535,285
1,300 to 1,400	2,205,728
1,400 to 1,500	1,832,230
1,500 to 1,600	1,512,649
1,600 to 1,700	1,234,397
1,700 to 1,800	999,996
1,800 to 1,900	811,236
1,900 to 2,000	663,789
2,000 to 2,100	549,787
2,100 to 2,200	463,222
2,200 to 2,300	395,115
2,300 to 2,400	340,141
2,400 to 2,500	295,490
2,500 to 2,600	258,650
2,600 to 2,700	227,731
2,700 to 2,800	201,488
2,800 to 2,900	178,901
2,900 to 3,000	154,499
3,000 to 3,100	142,802
3,100 to 3,200	128,217
3,200 to 3,300	115,583
3,300 to 3,400	104,504
3,400 to 3,500	94,803
3,500 to 3,600	86,405
3,600 to 3,700	79,023
3,700 to 3,800	72,562
3,800 to 3,900	66,900
3,900 to 4,000	61,894

(註一)各組之定義應釋為 '\$0 至不滿 \$100', '\$100 至不滿 \$200' 等是,故個人收入適為 \$100 者,應歸入第二組內。

(註二)據美國經濟研究所報告,謂所列人數,其末位數字未見全無差誤,此種數字所表示之確度,僅屬形式而已。

之次數曲綫圖中，此種不規則之狀態業已消除，故在橫尺度某兩點上所豎立之縱綫間之面積，應與該兩數值間所含之理想次數成正比例。此外又因該綫呈現勻滑之趨勢，故可用插補法 (interpolation) 以斷定表中未載之數值(註一)。

表十一所列資料 (註二) 爲一九一八年個人收入在美金四千元以下者之分配，可用以說明曲綫之修勻方法。

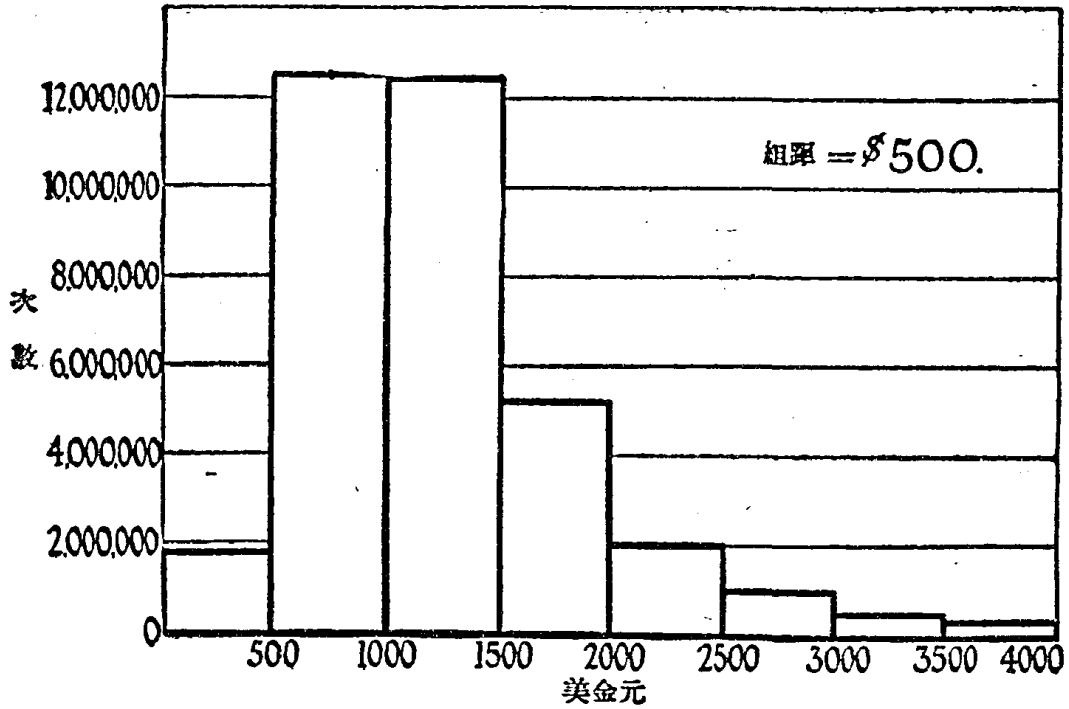
圖二十九、三十、三十一爲用 \$500, \$200 及 \$100 爲組距所得個人收入之直方圖。當組距逐漸縮小時，其直方圖亦漸見勻整，惟因原有資料之組距爲 \$100，故在本例中之組距，無法使其小於 \$100。吾人當前之問題，乃爲決定當組距縮小時此項資料所應有之分配狀態。如將直方圖中之斷續綫，易以勻滑曲綫，而同時使勻滑曲綫下包含之總面積與直方圖中原有之總面積相等，則此綫可以代表所求之分配形態。勻滑曲綫之繪法，係使曲綫通過直方圖上各點，同時使各條形上所切去之面積與其所圈入之面積大略相等，則曲綫下所包含之全面積，自必與直方圖中原有之全面積相等。繪勻滑次數曲綫時，綫下包含之全面積，須與原有面積相等，實屬必要之條件。至於各條形之面積，雖因有不規則變化之存在，每不能與此法則相合，然此仍不失爲繪圖時之一種普通原則 (用曲綫配合數字資料之精密方法，後文當再論述之。前節所述僅憑觀察以修勻曲綫之方法，其所得結果常與所求曲綫甚爲接近)。

圖三十二係將圖三十一之收入分配直方圖修勻後所得之圖形。該圖已將各組間不勻滑之輪廓加以修勻，故所得之圖成一依極微數量而增減之勻滑曲綫，可以代表收入分配之實際狀態，因其包含人數至數千

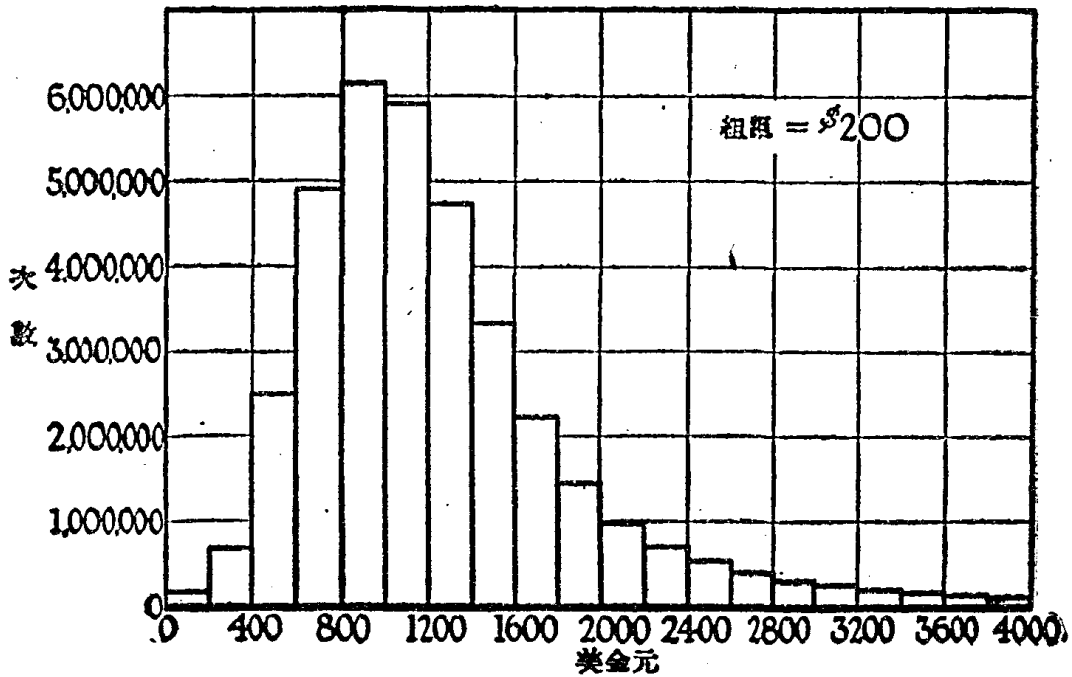
(註一)在統計實際工作上，每爲事實所限，所搜集之資料常有間斷之處，故各變量之數值往往不能連續。今用插補法可以估計兩已知數值間之各值，或估定曲綫上兩點間之各點。插補所得之數值以與已知各值相合者爲最準確。

(註二)錄自一九二一年 National Bureau of Economic Research 出版，紐約 Harcourt, Brace & Co. 發行之 "Income in the United States" 第一卷第 132—133 頁。

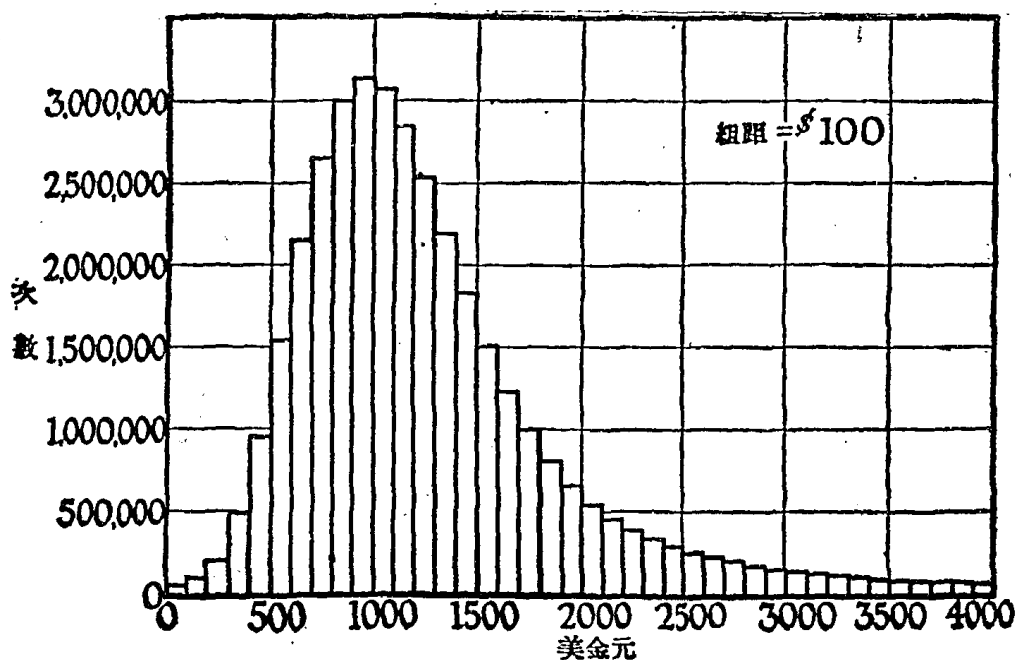
萬人時，其分配狀態固應如此也。此時吾人已得一可以代表基本分配之曲綫，其由分組而得之不規則變動及間斷之處均經消除矣。



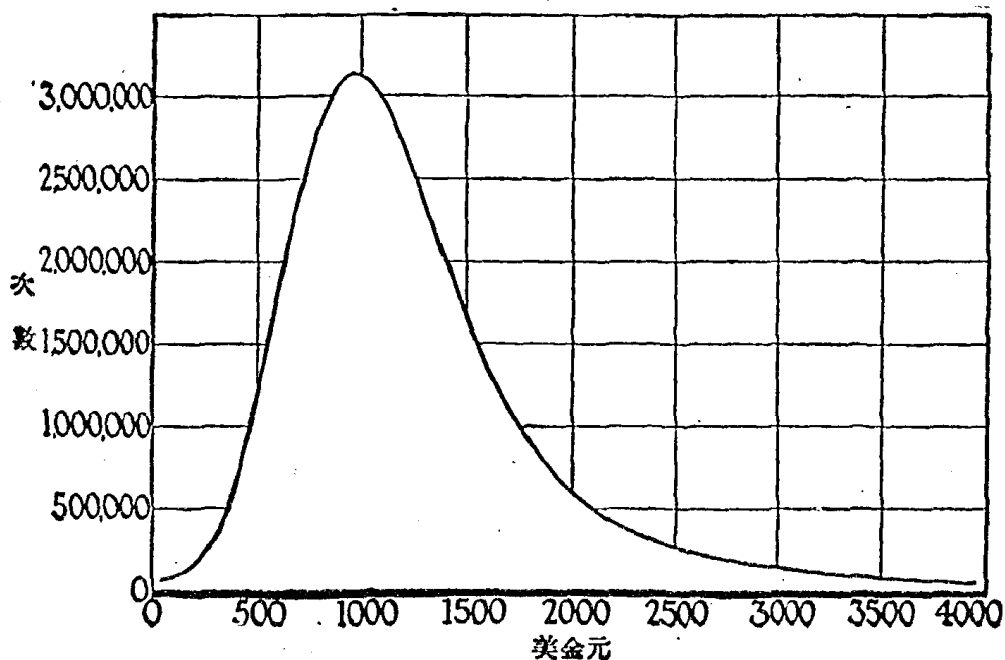
圖二十九. 直方圖: 1918年美國個人收入之次數分配
包括收入之在 \$4000 以下者(組距 = \$500)



圖三十. 直方圖: 1918年美國個人收入之次數分配
包括收入之在 \$4000 以下者(組距 = \$200)



圖三十一. 直方圖：1918年美國個人收入之次數分配
包括收入之在 \$1000 以下者(組距 = \$100)



圖三十二. 次數曲綫圖：1918年美國個人收入之次數分配
包括收入之在 \$4000 以下者(根據組距為\$100之直方圖繪成)

連續數列與間斷數列

曲綫修勻之是否合於邏輯，須視資料之性質而定。就資料之性質而言，次數數列可分為連續數列 (continuous series) 與間斷數列 (non-continuous or discrete series) 兩大類。數列中之自變量依無限小之差量而增減者，謂之連續數列；數列中之自變量依定量而增減者，謂之間斷數列。間斷數列之曲綫，不若連續數列之曲綫之勻滑，而呈凹凸之狀。

世間可供吾人研究之凡百現象，其性質實有連續與間斷之別。吾人所用以研究之樣本 (sample)，無論其所代表者為連續數列抑間斷數列，其自變量之數值，每呈間斷之狀，即在連續數列中，此種情形亦所不免，因測量器具之確度與吾人理解力均有限制之故。例如測量百人之高度所得之次數數列，其自變量 (高度) 必依定量而增減。在測量時，或以一英寸為最小單位，或以一英寸之八分之一為最小單位，但若將萬人或千萬人各按其高度排列，則鄰接兩人高度之差必為無限小。高度本為連續變量，惟吾人在樣本中所見者，常呈間斷之形狀耳。

至於利率或貼現率之變動，則與人體高度迥異。如將利率或貼現率之數字百項各按其大小排列，其數字之變動並不連續，與測量人體高度無異，但人體高度之測量，如包括人數極多，則其變動必呈連續狀態，貼現率則否，所用項目縱多至無限，其變動必仍呈間斷之狀，蓋貼現率乃以 $\frac{1}{2}\%$ 或 $\frac{1}{4}\%$ 而增減，而不以無限小之數量增減者也。此種數列謂之間斷數列。

當樣本之項目增至無限多時所應有之分配形態，亦可用曲綫修勻法以推求其大概。假定樣本中之不規則變動係屬偶然性質，而其基本變動實為連續不間斷之形狀，則可用修勻法將其曲綫修勻之，故曲綫修勻法實際僅能適用於連續數列。如根據樣本所得人羣高度之直方圖，即可依此法修勻，以求全體人羣高度分配之近似狀態，且可用作插補數值之

根據。此法如用於間斷數列，則殊不合邏輯。吾人設欲繪製一貼現率分配之勻滑曲綫，以求貼現率在 4.3675% 之一點上之理想次數，則無意義矣。雖然，在統計實務方面，每有以間斷數列視為連續數列，以助問題之解決者，此時曲綫修勻法亦可適用；惟在解釋與應用修勻曲綫時，此連續與間斷變動在邏輯上之區分，殊不可忽視耳(註一)。

統計資料之累積排列法

為欲達到某種目的，統計資料有須累積排列(cumulative arrangement)，而不宜採用各組次數分列之形式者。下列數字可說明此種排列法之優點。

此項數字為研究電桿木之耐用時期(註二)所得之結果。

表 十 二

按耐用時期長短分組之電桿木 248,707 根之次數分配

耐用時期 (年 數)	電桿木根數 (次 數)
0— 0.9	1,150
1— 1.9	4,221
2— 2.9	10,692
3— 3.9	13,966
4— 4.9	16,633
5— 5.9	18,211
6— 6.9	19,011
7— 7.9	19,260
8— 8.9	20,909
9— 9.9	19,879

(註一)讀者倘對於本節各點欲得更詳盡之論述，可參閱 G. H. Knibbs 氏著 The Theory and Justification of Curve Smoothing 一文，原文見 H. Secrist 氏著 “Readings and Problems in Statistical Methods” 第 278—282 頁，一九二〇年紐約 Macmillan 書館出版。

(註二)參看 Edwin Kurtz 氏著 Replacement Insurance，原文見一九二一年七月出版“administration” 第 41—69 頁。

表 十 二 (續)

耐用時期 (年 數)	電桿木根數 (次 數)
10—10.9	20,764
11—11.9	15,454
12—12.9	14,237
13—13.9	13,779
14—14.9	9,764
15—15.9	8,534
16—16.9	7,659
17—17.9	6,918
18—18.9	4,591
19—19.9	1,798
20—20.9	815
21—21.9	313
22—22.9	102
23—23.9	47

觀表可知電桿木在使用之第一年内即行損壞者計有 1,150 根,使用一年而未滿二年即行損壞者計有 4,221 根,餘類推;此為次數表之普通形式。在此例中如將次數排作累積形式,則其意義必更明瞭,可由下表見其一斑。

表 十 三

按耐用時期長短分組之電桿木 248,707 根之累積次數分配
(向上累積式)

耐用時期	電桿木根數 (次 數)
不滿1年者	1,150
不滿2年者	5,371
不滿3年者	16,063
不滿4年者	30,029
不滿5年者	46,662
不滿6年者	64,873
不滿7年者	83,884
不滿8年者	103,144

表 十 三 (續)

耐用時期	電桿木根數 (次 數)
不滿 9 年者	124,053
不滿 10 年者	43,932
不滿 11 年者	164,696
不滿 12 年者	180,150
不滿 13 年者	194,387
不滿 14 年者	208,166
不滿 15 年者	217,930
不滿 16 年者	226,464
不滿 17 年者	234,123
不滿 18 年者	241,041
不滿 19 年者	245,632
不滿 20 年者	247,430
不滿 21 年者	248,245
不滿 22 年者	248,558
不滿 23 年者	248,660
不滿 24 年者	248,707

次數數列之累積形式可分向上累積 (cumulated upward) 及向下累積 (cumulated downward) 兩種。上表係向上累積式，由此表吾人可推知耐用不滿某年數之電桿木根數；若將累積之方向顛倒，則欲推知耐用滿某年數以上之電桿木根數，常較便利。試仍用前項電桿木之數字以向下累積式排列之，可得下表。

表 十 四

按耐用時期長短分組之電桿木 248,707 根之累積次數分配

(向下累積式)

(1) 耐用時期	(2) 電桿木根數 (次數)	(3) 百分比
0年及以上	248,707	100.0
1年及以上	247,557	99.5
2年及以上	243,336	97.8
3年及以上	232,644	93.6

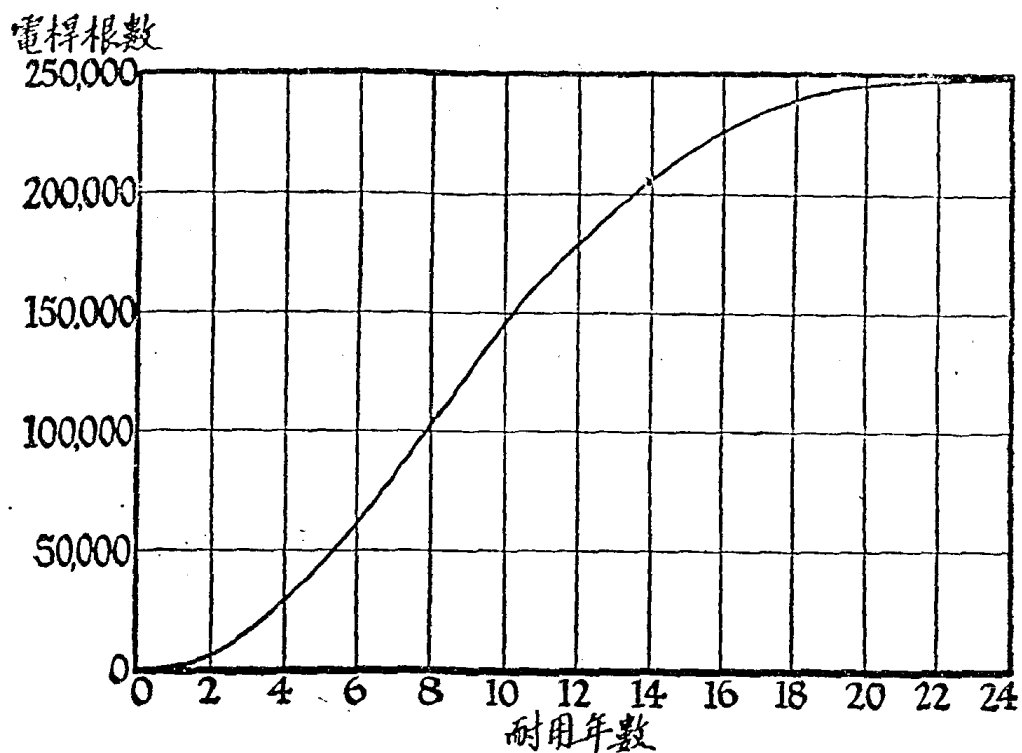
表 十 四 (續)

耐用時期	電桿木根數 (次 數)	百分比
4年及以上	218,678	88.0
5年及以上	202,045	81.2
6年及以上	183,834	73.8
7年及以上	164,823	66.3
8年及以上	145,563	58.5
9年及以上	124,654	50.1
10年及以上	104,775	42.1
11年及以上	84,011	33.8
12年及以上	68,557	27.6
13年及以上	54,320	21.8
14年及以上	40,541	16.3
15年及以上	30,777	12.4
16年及以上	22,243	8.9
17年及以上	14,584	5.9
18年及以上	7,666	3.1
19年及以上	3,075	1.2
20年及以上	1,277	0.5
21年及以上	462	0.2
22年及以上	149	0.06
23年及以上	47	0.02
24年及以上	0	0.00

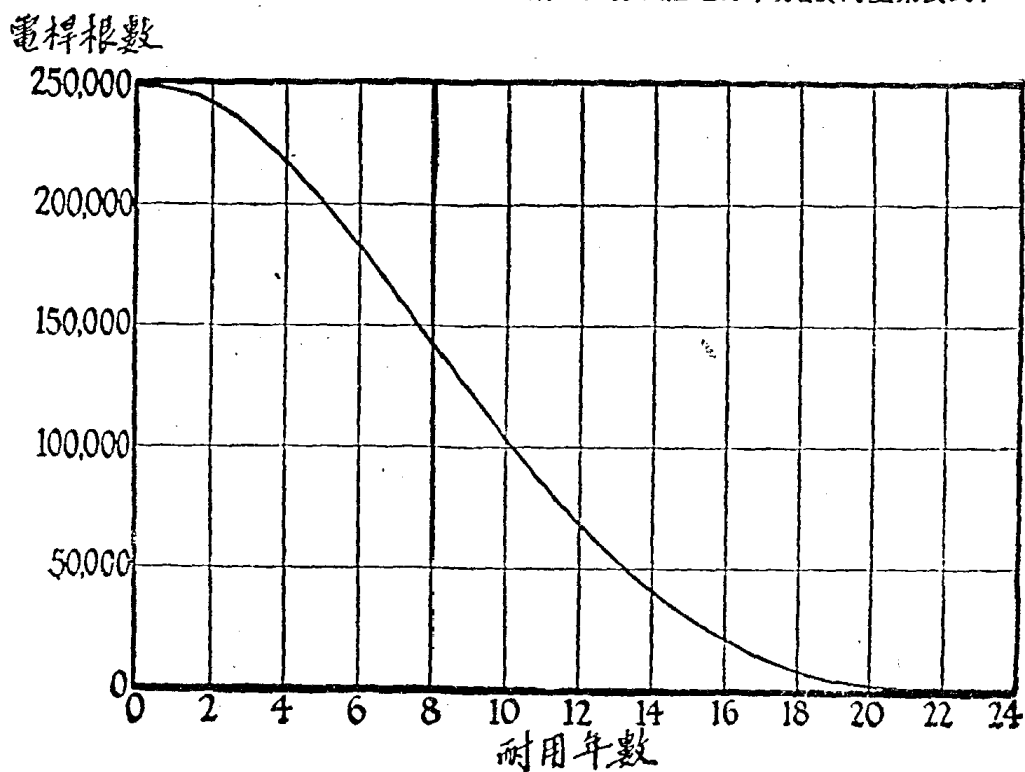
前項累積次數表在處理各種資料時，每有顯著之利便。吾人倘欲將折舊問題作科學上之研究時，常須為各種器物分別製就精密之生命表 (mortality table)，而此表以採取累積形式最為合宜。製此表時，每有將次數化為百分比之必要，如表十四第(3)欄所示。此種百分比之意義須視其所根據之原來數字而定。

累積次數曲綫

簡縮資料必須分組，而累積資料之功用，每因分組而受限制，倘吾人不用數學插補法，則所得之累積數字，僅限於表中所列各點，故吾人



圖三十三. 累積次數曲綫圖：按耐用時期長短分組之電桿木分配(向上累積式)



圖三十四 累積次數曲綫圖：按耐用時期長短分組之電桿木分配(向下累積式)

有求與前節所述之勻滑次數曲綫相似之累積曲綫之必要。將表十三內之各項數值繪於坐標圖上(耐用年數作為橫坐標,電桿木根數作為縱坐標),更依各點繪成一勻滑曲綫,即得如圖三十三所示之累積次數曲綫(ogive or cumulative frequency curve)。圖三十四係用表十四之資料所繪製者。

此種曲綫為表示次數數列最有效用之一法。此綫已將組距之界限大部消除,故組距與組數縱有變更,此曲綫之形態仍可大致不變。普通之次數曲綫,如其組距不同,每不能互相比較,而累積曲綫則不受此限制,即組距大小不等之資料,亦可繪為勻滑之累積曲綫而無錯誤。

累積曲綫對於數字之插補尤為適用。例如欲知耐用年數不滿十五年六個月之電桿木數,可在圖三十三中橫坐標等於 $15\frac{1}{2}$ 之處,讀其縱坐標之數值約為 222,000。如欲知耐用年數滿八年六個月以上之電桿木數,則可自圖三十四中依同法測定之,此數約為 135,000。

此種曲綫又可用以插補任何兩點間所包含之次數。此兩點之距離,並不以原表上所載之組距為限。例如欲知耐用年數在十年六個月以上十五年以下之電桿木數,則由表中或圖中可得耐用年數不滿十五年之電桿木數為 217,930 根;更用前述插補法,在圖中求得耐用年數不滿十年六個月之電桿木數為 154,000 根,由 217,930 中減去 154,000 得 63,930,即為十年六個月至十五年兩點間所包含之電桿木數。惟此數不過為一近似於實值之數,與其他由修勻曲綫插補所得之數字無異。

累積曲綫亦可逕由整列求之,而毋須經過編製次數表之手續。此種曲綫實質上僅可視為整列之圖解,為統計整理上最簡單之一種形式,而又為處理數量資料最有效力之一種方法。

累積曲綫與普通次數曲綫之關係

累積曲綫與普通次數曲綫乃係用同一資料所繪製兩種排列不同之

形態，而各有其特殊之優點。如吾人能瞭解此兩種曲綫結構上之關係，則對於兩種排列之特點更易明瞭。此種關係可以圖三十五表示之。

此圖係根據下列次數表繪製，表示按照每千呎木料所需勞工費用分組所得之美國鋸木廠分配(註)。

表 十 五

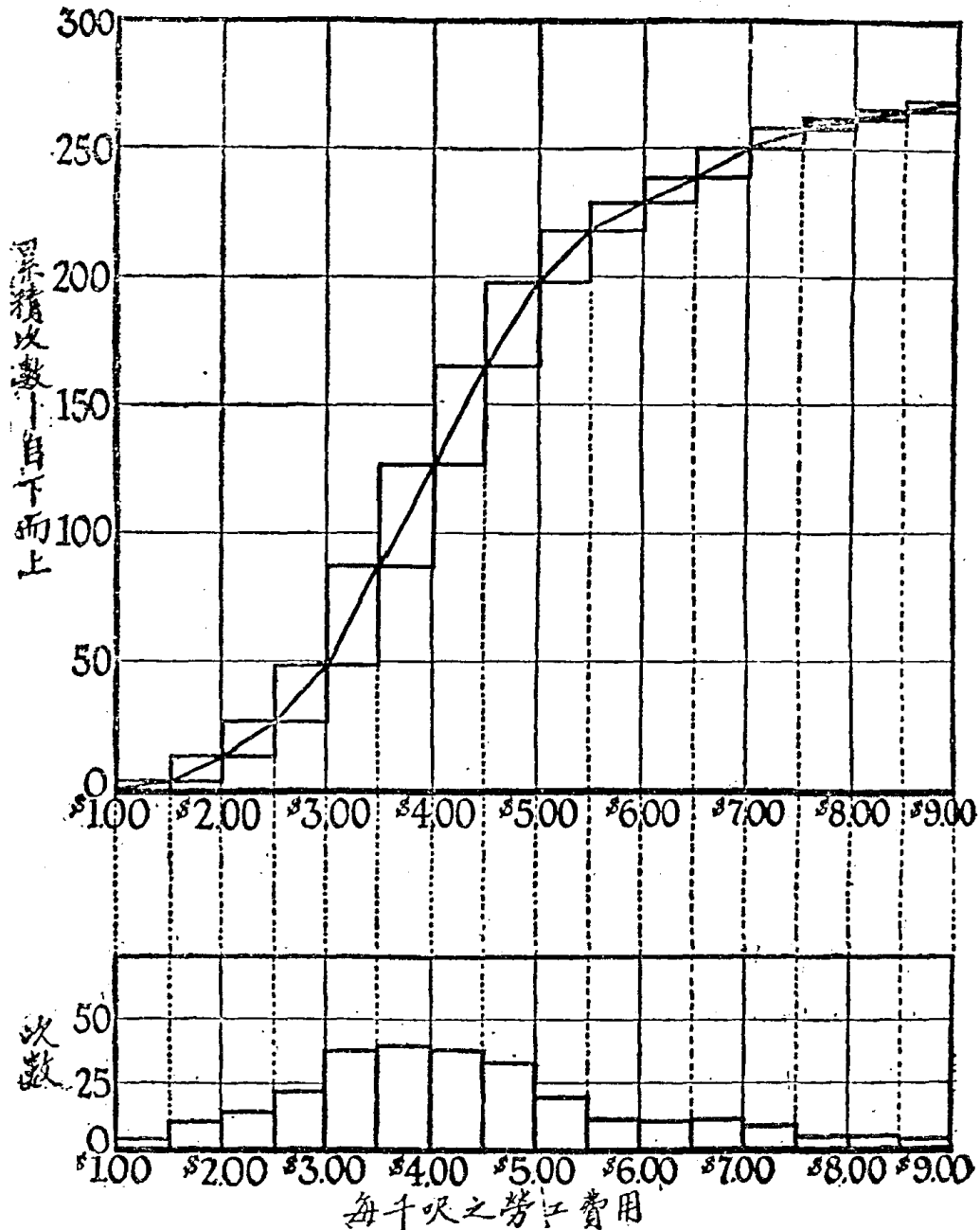
1921 年依勞工費用分組之美國鋸木廠 269 家之次數分配

每千呎之勞工費用 (包括一切雇員之薪工)	廠數 (次數)
\$1.00—1.49	3
1.50—1.99	10
2.00—2.49	14
2.50—2.99	22
3.00—3.49	38
3.50—3.99	40
4.00—4.49	38
4.50—4.99	33
5.00—5.49	20
5.50—5.99	11
6.00—6.49	10
6.50—6.99	11
7.00—7.49	8
7.50—7.99	4
8.00—8.49	4
8.50—8.99	3
	269

圖三十五上半部係表示累積曲綫之構造方法。圖中每一矩形之面積與各組所含次數成正比例，此與直方圖相同；惟就作法而言，此係累積形式，即每一矩形之底綫，各為其前列各組次數之累積數。例如第一矩形之 y 值(次數)為 3，係以 0 點作為底綫；第二矩形之 y 值為 10，則不

(註)摘錄一九二三年一月“Monthly Labor Review”第14頁所載 Ethelbert Stewart 氏著 Labor Efficiency and Productiveness in Sawmills 一文。\$9.00 以上之勞工費用，因其數字散漫，均未列入圖表中。

復以 0 點為底綫，而以 3 為底綫，餘類推。接連各矩形所成之曲綫，其斜度 (slope) 初以各組次數較小，頗為平坦，繼以各組次數漸大而漸陡峭，終以近上限處各組次數又漸減少而復轉平坦，此即所謂累積曲綫也。



圖三十五. 1920年按勞工費用分組之美國鋸木廠次數分配

說明累積曲綫與普通次數曲綫構造上之區別

假定各個 x 之數值不變，將代表各組次數之各個矩形，皆以零綫作為共同之基綫排列之，即得前文所述之直方圖，由此圖形即可繪成次數多邊圖或修勻次數曲綫。

參考書

- Bowley, A. L. "Elements of Statistics" (52—81).
- Chaddock, R. E. "Principles and Methods of Statistics" (Chaps. IV, V).
- Day, Edmund E. "Standardization of the Construction of Statistical Tables".
Quarterly Publications, American Statistical Association, March, 1920.
- Jones, D. C. "A First Course in Statistics" (5—21).
- Kelley, Truman L. "Statistical Method" (1—37).
- King, W. I. "Elements of Statistical Method" (83—107).
- Pearl, Raymond. "Medical Biometry and Statistics" (74—104).
- Rugg, H. O. "Statistical Methods Applied to Education" (57—94).
- Secrist, Horace. "Introduction to Statistical Methods" (116—157).
- Secrist, Horace. "Readings and Problems in Statistical Methods" (242—271).
- Yule, G. Udney, "An Introduction to the Theory of Statistics" (75—105).

第四章 次數分配之表述：平均數

數量資料之分組與次數分配之編製，為統計整理與分析之重要步驟。分組而後可表明資料之基本結構，揭示大量資料之劃一性質，惟此為統計分析之第一步，吾人尚須應用其他方法將資料全部之重要特性，為精密之測度與表述。次數分配之本身須簡縮之，使其特性可用三四有意義之數字表明之。

如每一次數分配含有特殊之性質，而各隨一獨具之法則，則研究與表述此種次數分配，誠屬困難，所幸事實並不如此。數量資料彼此性質縱不相同，然一經整理編為次數分配後，每表現共有之特性，而合於普通法則，故在某方面所獲得之經驗，即可用為他方面之南鍼。大量資料之變化具有劃一之性質，故由各種科學研究所得之數量資料，可用一普通方法整理之、分析之而比較之。

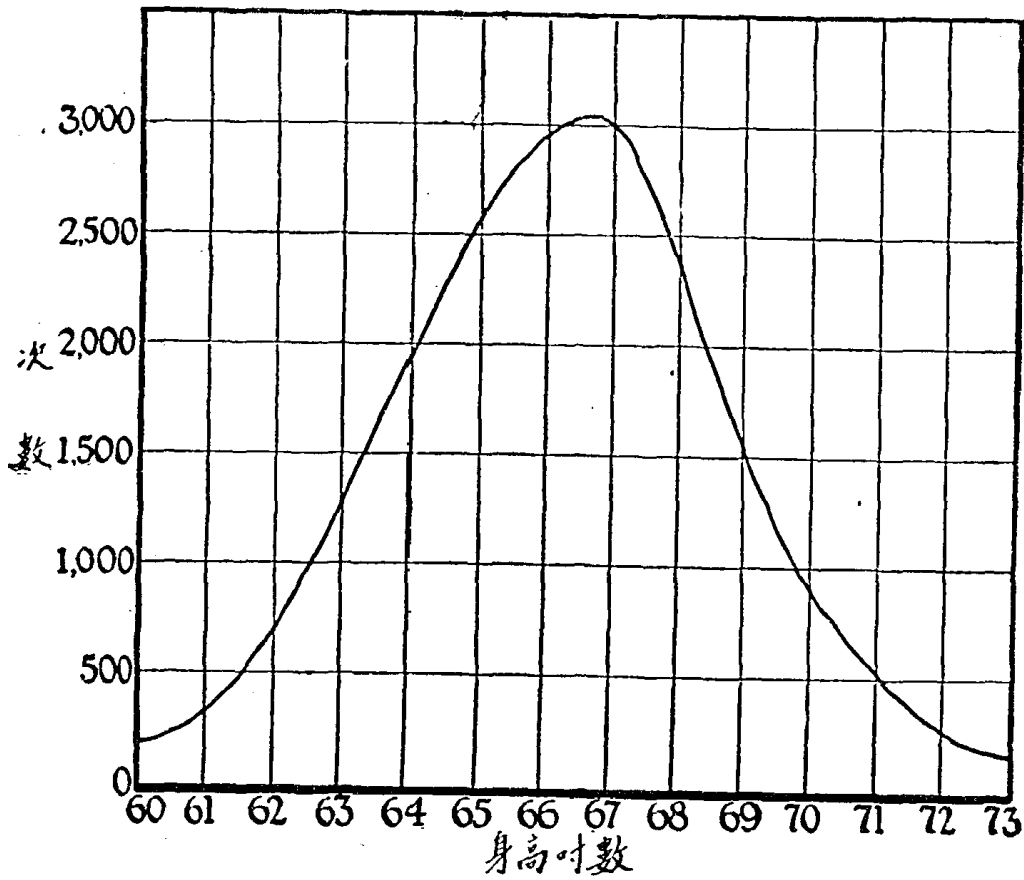
次數分配之比較

世間各種現象之數量排列，每遵循一普通法則，此可由比較各種資料之次數分配以發見之。吾人可觀察下列各種次數分配與次數曲綫之特性，並比較其分配狀態。

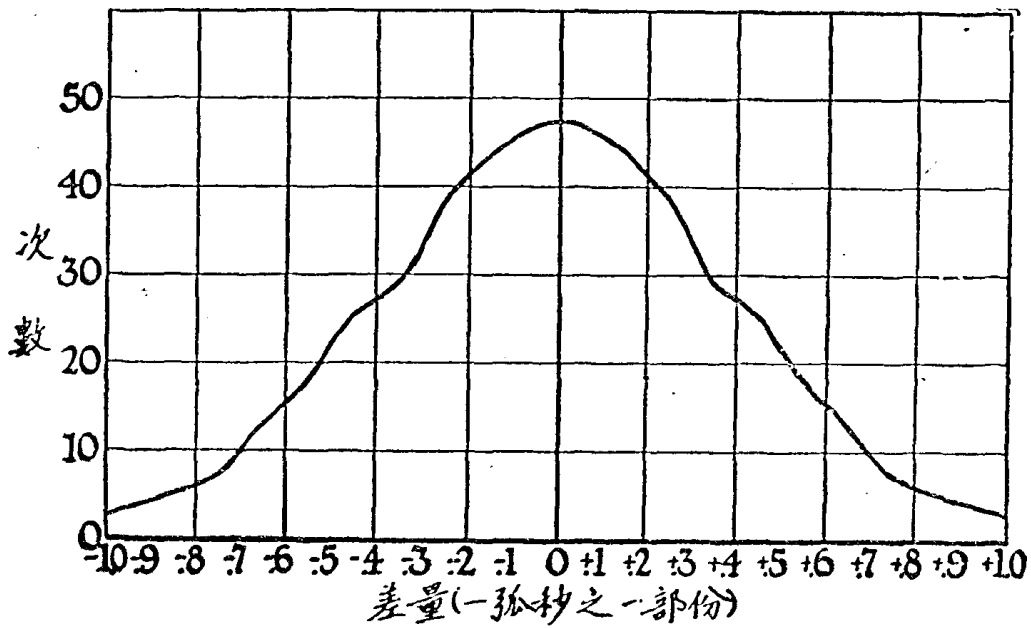
圖三十六係用下列某軍隊中兵士 18,780 人之身材高度分配所繪製之次數曲綫(註)。

身高吋數	兵士數	身高吋數	兵士數
60+	197	67+	3,017
61+	317	68+	2,287
62+	692	69+	1,599
63+	1,289	70+	878
64+	1,961	71+	520
65+	2,613	72+	262
66+	2,974	73+	174
		總計	18,780

(註)錄自 G. C. Whipple 氏著 "Vital Statistics" 第 377 頁。



圖三十六. 次數曲綫圖：按身材高度分組之兵士18,780人之次數分配



圖三十七. 次數曲綫圖：天文觀察差誤之分配

圖三十七係用卜拉德賽氏 (Bradley) 測量星球四百七十次所得觀察差誤 (註) 之次數曲綫。其原來數字如下：

表 十 七
某種天文測量之觀察差誤

差誤數量(一弧秒之一部分)	差誤次數
自0.0至0.1	91
自0.1至0.2	88
自0.2至0.3	78
自0.3至0.4	58
自0.4至0.5	51
自0.5至0.6	36
自0.6至0.7	26
自0.7至0.8	14
自0.8至0.9	10
自0.9至1.0	7
1.0以上	8
總計	470

前項資料因未將差誤分別正負，故不知其差誤為高於真實數值抑低於真實數值，若據以製圖，祇可得次數曲綫之半部。圖三十七所示為一完整之次數曲綫，乃係根據高於真實數值之差誤次數必與低於真實數值之差誤次數相等之一種假定，將其他半部補繪而成之曲綫。該項假定在此處頗可適用。

倘吾人移動砲口向某目標(即一點)精確瞄準，而後發射百次，則此百粒彈丸之碰撞點必分佈於該目標之四周。無論此砲之構造如何精密，瞄準之方法如何準確，彈丸之適中目標者必僅占一小部分，其他碰撞點將散佈於目標之四周，成一極有規則之形態。試繪一矩形圖，使包括彈丸碰撞點之全部，復將此矩形(或稱分佈區域 zone of dispersion)

(註)見 Mellor 氏著 "Higher Mathematics for Students of Chemistry and Physics" 第 514 頁。

分爲八個相等部分，則彈丸碰撞點在此八個部分間之分佈狀態，將如下圖所示(實際上之分佈或與下圖略有參差，惟普通必具此分佈狀態)：

2	7	16	25	25	16	7	2
---	---	----	----	----	----	---	---

圖三十八. 砲彈發射時彈丸百粒之分佈區域；表示彈丸之理想百分分配

此項普通法則可驗諸各種槍砲而不爽。槍砲之構造愈精密，則分佈區域之範圍亦愈狹隘，然在該區域內碰撞點分佈之狀態，在理論上應相類似。槍砲瞄準所用之法則即基此而成。

茲將實地射擊之結果與理想分配一比較之。下表係砲位與一靜止之目標相距二百碼處發射千彈所得之紀錄(註)。彈丸分佈區域係用平行綫等分爲十一部。

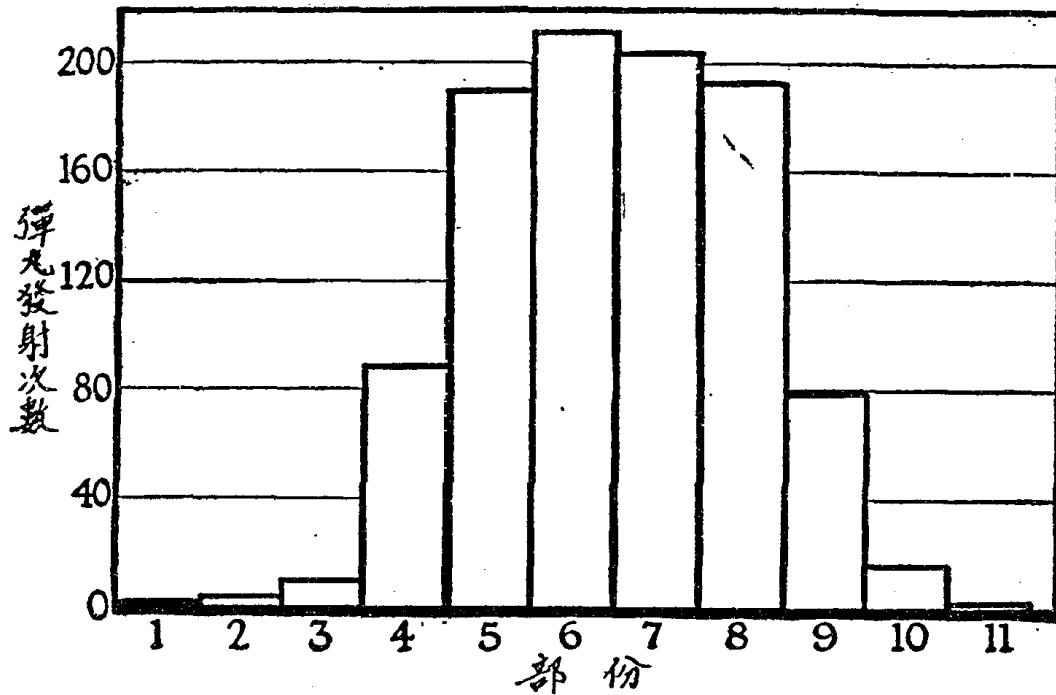
表 十 八

一砲發射彈丸千枚之次數分配	
部分	彈丸碰撞次數
1(頂)	1
2	4
3	10
4	89
5	190
6	212
7	204
8	193
9	79
10	16
11(底)	2
總計	1,000

(註)此項試驗見 "The Report of the Chief of Ordnance", 1878, Appendix S. Wiley 氏所著 "The Method of Least Squares" 第14頁中曾引用之。該書係 1897年紐約 Mansfield Merriman 公司出版。

此項彈丸分佈狀態如圖三十九所示。

彈丸分佈區域係分爲十一個部分，在說明上述之理想分配時，係分爲八個部分，故兩者不能直接比較，但其分配狀態，却與前述各例之普通形態相同。彈丸分配之較爲集中於目標下半部之原因，當係放置礮位時，未能完全顧及地心吸力所致。



圖三十九. 直方圖：一礮發射彈丸千枚之次數分配

茲更以擲錢爲例，假定錢幣落地時，正面與反面向上之次數分配，均係隨機遇而定。試以錢幣十枚同時拋擲百次，其每次正面向上之枚數如下表所示（在此項試驗中，每次拋擲正面向上之枚數，最多爲 10 枚，最少爲 0 枚）：

表 十 九

錢幣十枚同時拋擲百次所得之次數分配

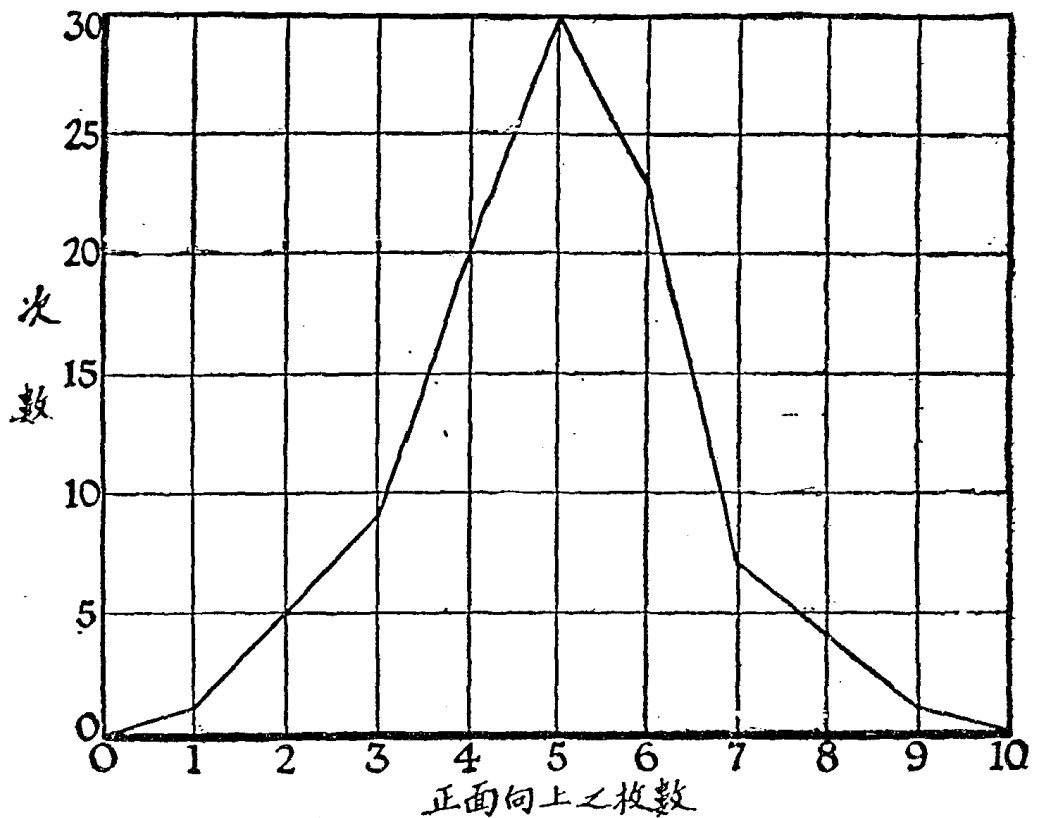
正面向上之枚數	次數
10	0
9	1
8	4
7	7

表 十 九 (續)

正面向上之枚數	次數
6	23
5	30
4	20
3	9
2	5
1	1
0	0

100

圖四十表示上項次數分配之狀態。



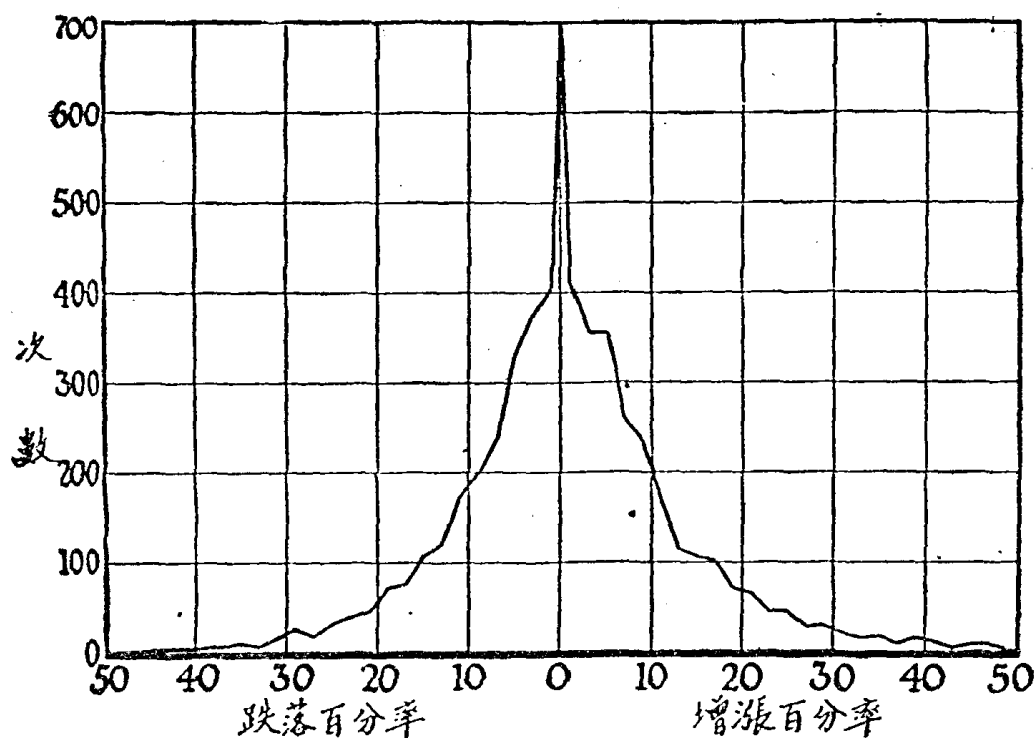
圖四十. 次數多邊圖：擲錢試驗中錢幣正面向上枚數之次數分配

經濟資料之分配

在上述四種絕不相同之現象中，吾人嘗發現數量資料排列之普通法則，惟此四項實例皆非代表經濟事實，現當研究經濟資料是否亦具有相似之特徵。參考第三章內所舉各例，而與上述四例比較之。該章內所

舉各例，爲工人某週工資之分配，電桿木之耐用年數，鋸木廠之勞工費用，及美國人民每年收入在四千元以下者之收入分配(此項收入分配，如將四千元以上之收入亦一併列入，則其曲綫將向右方極度延長，而成長尾狀)。茲更就經濟方面加舉數例於下。

圖四十一表示物價變動之分配狀態。圖中所用資料爲密哲爾氏(W. C. Mitchell) 研究所得某年 5,578 種物品之批發價格與前一年比較之結果(註)。例如一九一二年紐約高地米特令棉花 (middling upland cotton) 每磅之平均價爲美金 0.115 元，至一九一三年每磅平均價增至美金 0.128 元，計增 11.5%。在物價上漲之表中，此項即歸入“10—11.9%”之一組內，爲該組之一次數。全表包括 5,578 次之紀錄。圖四十一係用次數多邊圖以表示物價之變動者，該綫並未用修勻法將其修勻。



圖四十一 次數多邊圖：某年批發物價比前一年漲落百分率5540項之次數分配

(註)摘錄Bulletin 284, U. S. Bureau of Labor Statistics, Part I, "The Making and Using of Index Numbers" 第 18 頁。此項數字係表示自跌落 51% 起至上漲 51% 止之物價變動。此外物價跌落 55% 者計 1 項，上漲 52% 至 104% 者計 37 項，圖中均未繪入。

表二十係表示一八八二年至一九一三年間倫敦對紐約匯價（一鎊合美金元之數）之分配。該表係將三十二年內共 384 個月之匯兌率依數字大小所組成之次數分配（註）。

表 二 十

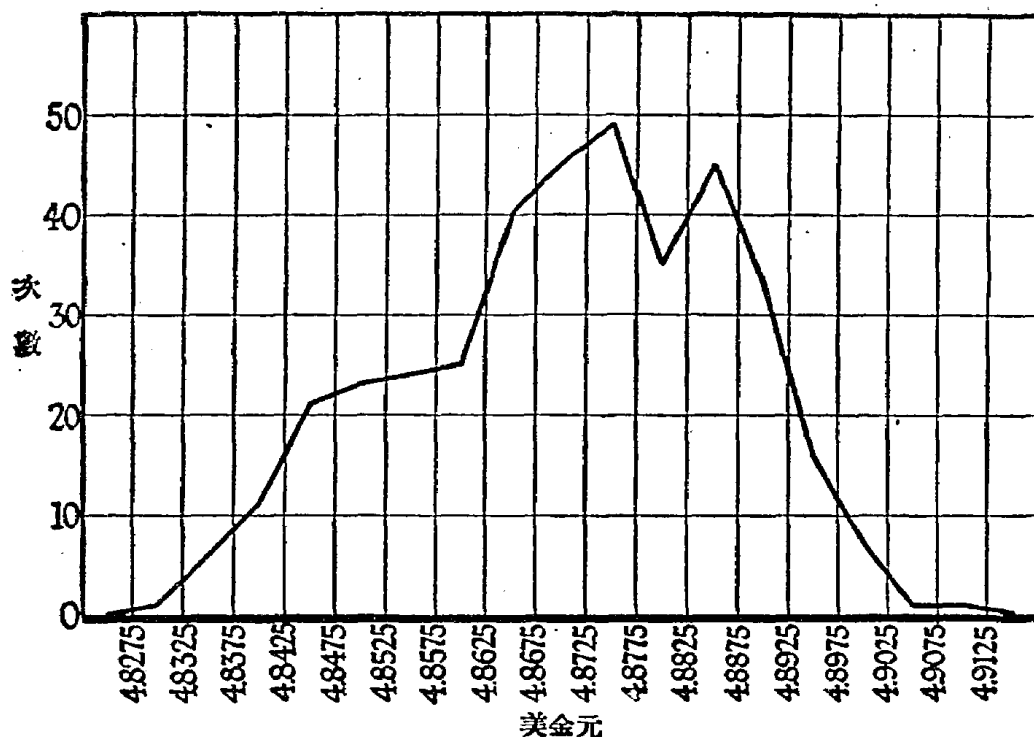
1882—1913 年倫敦對紐約每月匯價之次數分配

組 距	次 數 (發現左列匯價之月數)
\$1.8275—\$1.8324	1
4.8325— 4.8374	6
4.8375— 4.8424	11
4.8425— 4.8474	21
4.8475— 4.8524	23
4.8525— 4.8574	24
4.8575— 4.8624	25
4.8625— 4.8674	40
4.8675— 4.8724	45
4.8725— 4.8774	49
4.8775— 4.8824	35
4.8825— 4.8874	45
4.8875— 4.8924	33
4.8925— 4.8974	16
4.8975— 4.9024	8
4.9025— 4.9074	1
4.9075— 4.9124	1
	384

此項資料如圖四十二所示。

於此所應注意者，凡根據自然界資料繪製之曲綫，恆呈對稱 (symmetry) 與勻整之形態，而根據經濟資料繪製之次數曲綫及直方圖，則

(註)此項資料見 E. G. Peake 氏著 ‘An Academic Study of Some Money Market and Other Statistics’ 附錄 I, 1923 年倫敦 P. S. King 公司出版。原表註明 ‘此項數字係根據 ‘The Economist’ 每月初發表之匯價所計算之每月平均數。自 1886 年起係電匯價，以前則係短期匯價’。



圖四十二。次數多邊圖：倫敦對紐約匯價之次數分配(384個月之每月平均匯價)

未必皆然，有偏斜 (skew) 於右方或左方者，亦有因鄰接各組間次數增減之過驟，其曲綫呈不規則之形態者。惟兩種資料雖有此種區別，然不論其資料之屬於經濟、天文、人體測量、射擊或機遇方面者，其分配之形態恆有相似之處。此項次數分配之一般特徵當於下節論述之。

次數分配之一般特徵

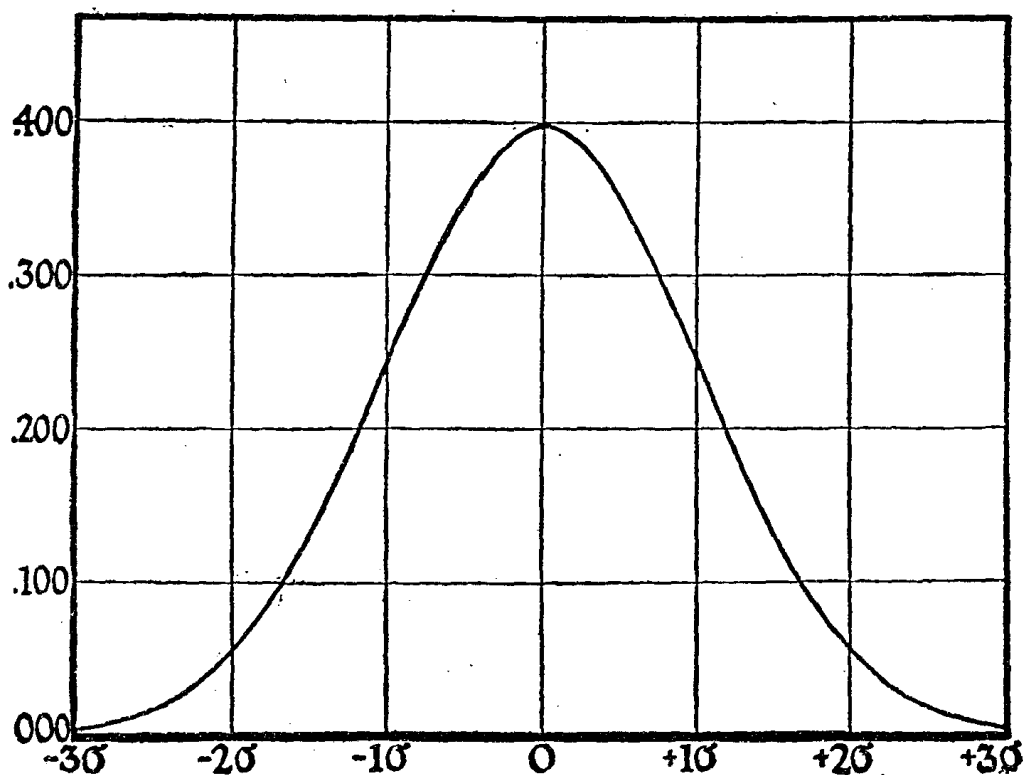
每種測度其數值必具有差異 (variation) 之特徵：人體之高矮不同，星球測量所得之結果不同，礮彈之發射即在同一情形之下，其碰撞點之地位不同，個人收入有多寡之分，匯兌市價有高低之別，凡此種種觀察數值，莫不變動於極高極低之兩限度間。

各種觀察之數值，其分配之形態，沿橫軸尺度自左而右量之，為由一極低之值，漸及於極高之值，而依次移動觀之，其在各點發現之次數 (即連續各組之次數)，漸次增多，而至一次數極多之點，由此乃逐漸減

少，而至次數極少之點，其增減之狀態頗有規律，而其次數每呈現集中之趨勢(central tendency)。集中趨勢者，次數密集於橫軸尺度上某數點之謂也。此集中趨勢又為各種次數分配所同具之第二特徵。

倘自最大集中點沿 x 尺度量其左右各組與此點之差異，則可發現差額較小各組所含之次數，必遠較差額較大各組者為多，且極端之差額頗不多見，而在集中點左右兩旁之分配，在自然科學與純粹機遇之各例中幾完全相等，而在經濟方面之各例中，亦近乎相等(惟此亦常有例外，個人收入之分配，即為其最著之一例，其最大集中點左右兩邊之分配，頗不相等也)。

圖四十三所示之曲線，謂之機率曲線(probability curve)，或稱差誤常態曲線(normal curve of error)。其特徵當俟後文詳論之，此處所以先行提示者，因其為次數分配之基本形式。前述各例中，其分配形態有與之極相類似者，亦有表現顯著之差別者。與此形式不同之分配，亦



圖四十三. 差誤常態曲線

所常見，惟就其基本形式而論，此種常態曲綫在統計學上極關重要。即與此形式極不相同之分配，亦不無相似之點，故仍可用一普通方法以表述該種次數分配。數量資料之分配必有不同之處，其彼此間之差異，及其與某種標準形式差異之處，固含有重要之意義，但不論其差異程度若何，彼此間終必有類似之處，故每一次數分配，決非一隔離之現象，而為全體之一部分，而可採取他種分配所用之方法以表述之、分析之。

既知此種普通之形式，又將如何表述一種次數分配或區別其與他種分配之差異乎？由前文所述各點，即可推知各種表述與區別之方法，請分論之。

表述次數分配之方法

各個觀察所得之數值，恆必分配於 x 尺度之各點上，已如前述。在此各點之中，吾人可擇一足以代表其全體之單獨數值以表述其次數分配。各點之次數既不相同，則選擇時理應選取包含次數最多之一數值，亦即尺度上次數最為集中之一點，作為該次數分配之代表。此數值即為測度分配之集中趨勢之一種量數 (measure of central tendency)。在個人收入次數分配之例中，應以包括人數最多之一個收入組之中點(在表十一之次數分配中，此中點為 \$950)作為該次數分配之代表數值。惟此最普通之數值，僅為測度次數集中趨勢所用各種量數中之一種。此各種量數概稱為平均數(average)。

此種單純之代表數值，雖為用甚廣，但關於次數分配之其他許多事實尚不能有所表示。其中以平均數兩旁次數分配之性質最關重要。次數之分配有密集於尺度上某數點者，有分散於各點者，平均數代表性之高下，須視其兩旁各數值之集中程度而定；故平均數之應用，必須藉一測量中心值兩旁分配之離散度量數(measure of variation)以補充之。

研究次數分配時，更應顧及分配之對稱程度(degree of symmetry)。

最大集中點之兩旁，其次數分配是否勻稱，或其次數曲綫是否如前述個人收入分配之例，有偏斜於一邊之形態，皆應考究明瞭。如其曲綫並不對稱，則當確知其不對稱(asymmetry)之程度，於是有測度分配之偏斜度量數(measure of skewness)。

最後，吾人可用差誤常態曲綫為標準，以測度次數曲綫之峻峭程度。如圖四十一所示物價變動之次數多邊圖，一經修勻，則此曲綫必較常態曲綫為峻峭，而此項目集中於中心數值之事實，頗具重大之意義。此種特徵謂之峯度(kurtosis)。測量峯度為表述次數分配之最後步驟。

前述各種量數既已求得，則統計分析工作可謂完成其大半。此項工作開始之時，係先將凌亂之資料一一分組，製成一次數表之形式；次將該表所呈現之重要事實，歸納為三個或四個有意義之量數。經此分析手續，則不僅可揭露一次數分配之特性，且可將該分配與性質相似之其他次數分配比較之。例如欲將數千萬未經整理之美國人民個人收入之數字與同樣之英國數字作比較，必不可能；但若由此兩類數字各求其平均數值，或最足代表之收入額，並將此中心值兩旁收入之分配狀態予以表述，則比較時可得一確切之標準。處理與分析大量資料時，不論其研究之目的若何，應藉表述次數分配之各種量數，將資料儘量簡縮、綜合而比較之。

下節當將表述次數分配之一種量數，即關於測度集中趨勢者，加以討論。此點討論以後，當進而論述測量離散度與偏斜度之方法。

平均數

吾人已知平均數——單獨之典型數值——所以能代表次數分配者，乃因大量數字有密集於中心值附近之趨勢，與各個數值散佈之狀態又多少具有規則與勻整形式之故，是以此項代表數值之有無意義，當全視次數之是否密集於尺度上之中心點而定。當平均數之值確為一典型數

值時，方可用此數以代表分配之全體。如次數分配中各數值差異頗大，且無集中之趨勢，則決無一單獨數值可代表次數分配之全體者。例如 3, 125, 1000 三個數字之算術平均數為 376，但此 376 之一數值絕不能用以代表其原有之三個數值。由是知次數分配必具備集中於中心數值之趨勢之基本條件，其平均數始有代表全體之價值。

吾人苟回憶次數分配之一般特徵，即可得一合理之平均數。前文中嘗提及在 x 尺度上有次數最為集中之一點，即包含次數最多之一數值，此值當可視為分配全體之典型數值。此典型數值謂之衆數 (mode)，衆數所在之一組，謂之衆數組 (modal group)。試繪一次數曲綫以代表某一次數分配，此衆數即為與最高縱坐標相對之 x 值 (註)。最高縱坐標乃係表示該衆數組之次數者，學者在決定衆數時，對於最高縱坐標之 x 值與最高縱坐標之值，每易混淆。衆數之值並非 y 軸上之距離，乃 x 軸上之距離也。至於縱坐標則用為測度各組之次數，而並不測度各次數所歸納之數值者。

除衆數外，吾人又可在 x 值之尺度上選擇左右兩邊各占全體項數之半之一點，作為次數分配之典型數值。此值當為大於全體項數之半而同時亦為小於全體項數之其他一半之數值，故謂之中位數 (median)。例如有人估計一九一八年美國人民個人收入之中位數之值為美金 1, 140 元，即謂享有收入者之 37,000,000 人之中，其半數之個人收入皆少於此數，其他半數之個人收入皆多於此數。如將次數分配繪成一次數曲綫，則在 x 軸上中位數值之一點上所豎立之縱綫，適劃分曲綫下全面積為兩個相等部分，此由於中位數本身之定義，同時亦由於次數曲綫下之面積，係代表該次數分配之全體項數故也。

算術平均數 (arithmetic mean) 為代表次數分配所用之第三種平均數，係由計算而得之一種平均數，故受次數分配中任何一項數值之影

(註)嚴格言之，衆數為配合於某一次數分配上之理想次數曲綫中、相對於最高縱坐標之 x 值。

響，此為與衆數、中位數不同之點，蓋後兩者均視各項目在次數表中之位置而定，而不受個別項目之數值之影響也。算術平均數實為次數分配之重心，如曲綫能繪成立體形，則此曲綫平衡支點處之 x 值，即為算術平均數。

平均數中尚有所謂幾何平均數 (geometric mean) 與倒數平均數 (harmonic mean) 兩種，其性質當俟後文討論之。

前述各種平均數之計算，在包括項數較多時，頗費手續，惟若運用適當方法，則可減省工作不少。茲為便於說明各種方法起見，引用下列各種符號：

M 算術平均數。

M_o 衆數。

Md 中位數。

m 個別觀察所得之數值；在次數分配中，此數值為各組中點之值。

f 次數分配中每組之項數。

N 一個數列或次數分配中之總項數。

Σ (sigma) 表示各項相加之符號，其意義為‘……之和’。

算術平均數之計算

應用前項符號，則算術平均數公式為

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

例如 2, 5, 6, 7 四個數值之算術平均數，等於此四數之和除以 4，即 $\frac{20}{4} = 5$ 是也。如各個數值係根據其實際數值，則算術平均數之計算，僅須費簡單之加法與除法手續。第三章內嘗表列工廠工人 210 人之某週工資數，如將此種數字一一相加，所得和數，再以 210 除之，則得每人每週平均工資為美金 26.982 元。惟如此計算，將 210 項數字一一相

加，工作已覺煩悶，假如吾人欲將美國 37,000,000 人個人收入之數字亦依此一一相加，則其計算工作幾屬不可能；故為便於實際工作起見，通常係根據次數分配表計算平均數，而不由未經分組之原有資料計算。茲以北達科多小麥產量為例，以說明平均數之計算手續。美國農業經濟局嘗記載一九一一年至一九二一年各年北達科多州五十三郡農田每英畝平均所產小麥之數量。茲將此項數字簡列於下：

表 二 十 一

北達科多州各郡小麥產量(註)算術平均數之計算

組 距 (每英畝蒲式耳數)	中 點 <i>m</i>	次 數 <i>f</i>	<i>fm</i>
0 — 1.9	1	3	3
2 — 3.9	3	26	78
4 — 5.9	5	78	390
6 — 7.9	7	101	749
8 — 9.9	9	113	1,017
10 — 11.9	11	65	715
12 — 13.9	13	40	520
14 — 15.9	15	22	330
16 — 17.9	17	45	765
18 — 19.9	19	41	779
20 — 21.9	21	21	441
22 — 23.9	23	8	184
		569	5,971

$$M = \frac{\Sigma(fm)}{N} = \frac{5,971}{569} = 10.49 \text{ 蒲式耳}$$

在討論數值分組時，嘗述及應需注意之各點，在本例中更可闡明其重要程度。組距大小之確定，應與每組內項目分配勻整之假定，不甚抵觸為主，此項原則嘗於前文論及之；蓋若每組內項目分配勻整，則其中點始可代表該組內之各項，但若項目分配欠勻，則該組中點即不能代表各項。算術平均數可根據次數分配計算者，係基於每組內項目分配

(註)該時期內之各郡平均產量見 Bulletin 165, Agricultural Experiment Station, North Dakota Agricultural College, 1922 及 "Cost of Production and Farm Organization on 126 Farms in North Dakota", 1921, 第120—121頁。

勻整之一個假設上，故以每組中點乘該組項數時，係假定其乘積略與該組內各個別數值相加之和數相等。算術平均數公式因之變為

$$M = \frac{\Sigma(fm)}{N}。$$

表二十一係舉示應用此式計算之手續。

依此手續計算所得之數值，亦稱加權算術平均數 (weighted arithmetic mean)。吾人計算時，實際上係求 m 欄內十二個數字之算術平均數，惟此平均數非十二個數字之簡單平均數，而係將此十二個組中點各按其組距內所包含之項數加權後計算而得。設有五人於此，其中兩人每年收入各 \$2,000，另三人各 \$3,000，則計算五人平均收入之手續，正與此相同。計算時並不僅將 \$2,000 及 \$3,000 相加而除以 2，而先將此 \$2,000 之數以權數 2 乘之，\$3,000 之數以權數 3 乘之，而後求其和數得 \$13,000，再以 5 除之而得。根據次數分配計算算術平均數之手續，雖具有此加權之形式，惟加權平均數 (weighted average) 之名詞尙有其較狹之意義，容後申論之，故此名詞通常不適用於按照次數分配計算之平均數。

計算算術平均數之簡捷法

通常算術平均數如由次數表計算，似較由未經分組之資料計算為易，但若包含項數過多，如照前法由次數表計算，手續仍頗煩瑣，惟此手續亦可設法化簡。

由算術平均數之計算法中可推知各數值對其算術平均數之正負離中差 (deviation) 之總和為 0，此點可以代數式證明之。今設 $m_1, m_2, m_3, \dots, m_n$ 代表一個數列中之各數值， M 代表其算術平均數， $d_1, d_2, d_3, \dots, d_n$ 代表各個數值對其算術平均數之離中差，則

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

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

今 m 之項數為 N , 故

$$(m_1 - M) + (m_2 - M) + (m_3 - M) + \dots + (m_n - M) = 0 \quad (3)$$

但 $m_1 - M = d_1, m_2 - M = d_2, \dots$ 故方程式 (3) 可寫作

$$\Sigma d = 0$$

吾人既知各數值對其算術平均數之正負離中差之和為 0, 則可先擇一假定算術平均數 (arbitrary origin or assumed mean), 再求各數值對此假定平均數之正負離中差之和, 更由此數值以斷定假定平均數與真實平均數 (true mean) 之差數。此項差數亦即各數值對其假定平均數之離中差之平均也。今設 M' 代表假定平均數, $c = M - M'$, 又設 $d'_1, d'_2, d'_3, \dots, d'_n$ 代表各數值對 M' 之離中差 (即 $d'_1 = m_1 - M', d'_2 = m_2 - M'$, 餘類推), 則

$$d'_1 = d_1 + c, d'_2 = d_2 + c, d'_3 = d_3 + c, \dots, d'_n = d_n + c$$

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

但 $\Sigma d = 0$

故 $\Sigma d' = Nc$

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

已知 M' 及 c 兩者之數值後, 則真實平均數亦可求得, 因 $M = M' + c$ 。其計算手續可用一簡例說明之。

表 二 十 二

算術平均數之計算(簡捷法)

(未經分組之資料)

m	f	d'	$M' = 20$
5	1	-15	
15	1	-5	$c = \frac{\Sigma d'}{N} = \frac{+25}{5} = 5$
25	1	+5	
35	1	+15	
45	1	+25	$M = M' + c = 20 + 5 = 25$
	5	+25	

如以各數值對其真實平均數之離中差作為標準，則依照假定平均數 20 以計算各數值之離中差時，每一離中差必含有一固定之差誤，此固定差誤即為真實平均數與假定平均數之差數。故離中差之總和中，實含有 N 個固定差誤，因每項計算時均含此差誤而共有 N 項也。離中差之和以 N 除之，即得此固定差誤之數值，而真實平均數亦可依此求得矣。

數量資料如依次數分配之形式排列，計算離中差時又以組距為單位，則平均數之計算手續更將便利不少。在校正差誤時，此真實平均數與假定平均數之差數仍可化為原來單位。茲仍以說明算術平均數普通計算法所用之小麥產量數字，說明此簡捷法之應用。

表二十三

1911—1921年北達科多州各郡小麥產量算術平均數之計算
(簡捷法)

組距 (每英畝蒲式耳數)	中點 m	次數 f	d' (以組距為單位)	fd'		M 之計算 $M' = 9$
				-	+	
0—1.9	1	3	-4	12		1.各數值對 M' 之離中差之和： +778
2—3.9	3	26	-3	78		-353
						+425
4—5.9	5	78	-2	156		2. c 之計算(以組距為單位)： $c = \frac{425}{569} = .7469$
6—7.9	7	107	-1	107		3. 化 c 為原來單位： 組距 = 2
8—9.9	9	113	0		65	c (按原來單位計算)
10—11.9	11	65	1		80	= .7469 × 2
12—13.9	13	40	2		66	= 1.4938
14—15.9	15	22	3		180	4. M 之決定： $M = M' + c$
16—17.9	17	45	4		205	$M = 9 + 1.4938$
18—19.9	19	41	5		126	$M = 10.4938$ 蒲式耳。
20—21.9	21	21	6		56	
22—23.9	23	8	7			
		569		-353	+778	

用簡捷法計算算術平均數之步驟，略如下述：

1. 將數字資料列成次數分配之形式。
2. 就靠近次數分配之中心處選定一組之中點作為假定平均數。
3. 在次數表中闢一欄，以便將每組各項對於假定平均數之離中差 (d') 列入，此離中差係以組距作為單位。在假定平均數所在組內各項之離中差為 0，其下一組內各項之離中差為 -1，其上一組內各項之離中差為 +1，餘類推。
4. 將每組之離中差乘以該組之次數，並注意其正負號，其乘積列入 fd' 欄內。
5. 就 fd' 欄內各項求其總和。
6. 將此總和除以次數之和 (N)，所得商數即為以組距作為單位之校正數 (c)。
7. 校正數 (c) 以組距乘之，所得乘積即為以原來單位表示之校正數。
8. 以此校正數與假定平均數 (M') 相加，其和數即為真實平均數 (M)。

中位數之決定：根據未經分組之資料

所謂中位數 (median) 者，在 x 數值各依其大小排列時，其地位最為居中之數值也。此值之地位係劃分項數為兩個相等部分，其中一部分之各數值大於中位數，另一部分小於中位數。此種中位數在許多次數分配中頗為適用。

當資料未經整理成為次數分配之形式時，中位數之決定至為簡易，祇須將資料依其數值之大小順序排列，然後由尺度之一端起依次點算項數，至劃分全部項數為兩個相等部分處之數值，即為中位數之值。茲以簡單數字說明中位數決定之方法。下列數字係七人之每年收入金額。

\$750 \$975 \$1128 \$1450 \$1475 \$1825 \$1950

前項數字之尺度，係自 \$750 起至 \$1950 止，依此尺度排列者共有

七項。在 \$1000 之數值上，其一邊有 2 項，另一邊有 5 項，故非中位數。此處中位數之值應為 \$1450，即七人中一人之收入金額。此數值之兩邊各有 3 項，假定將此項分割為二，則此點之兩邊各有 $3\frac{1}{2}$ 項。此可用圖四十四表示之。由該圖可見中位數之位置適將項數等分為二。

當項數為偶數時，則中位數之決定方法稍異，試用下表數字說明之。表內所列係一九二一年美國二十二州鋸木廠工人之平均工資率。

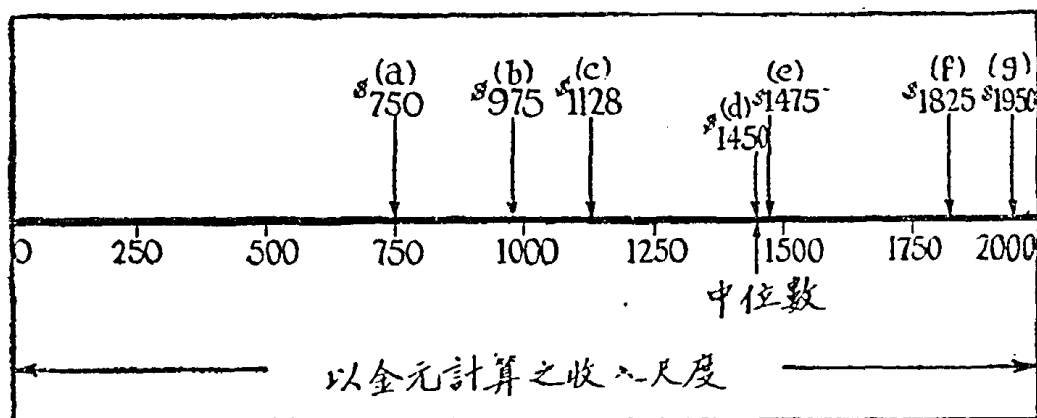
表 二 十 四

1921年美國22州鋸木廠工人之平均工資率(註)

州 別	每工人工作一小 時之平均工資率
弗 基 尼 阿	\$0.586
賓夕法尼亞621
泰 內 西627
北卡羅來那647
西弗基尼阿658
緬 因686
南卡羅來那721
威 斯 康 辛729
密 雪 千730
阿 拉 巴 瑪749
明 尼 蘇 達755
泰 克 薩 斯761
佐 治 亞779
阿 康 薩785
密 士 失 必798
路 易 斯 安 那824
佛 羅 里 達825
愛 達 荷833
加 利 福 尼 亞864
蒙 大 拿901
華 盛 頓	1.045
俄 勒 岡	1.106

(註)錄自 1923 年一月出版 "Monthly Labor Review", 第 9 頁。

本表之中位數當係兩邊各為十一州之數值，故大於 \$0.755\$ (即明尼蘇達之平均工資率) 與小於 \$0.761\$ (即泰克薩斯之平均工資率) 之間之任何一數值，均可適合於中位數之定義。在此情形下，中位數實無法確定，但通常係將此兩個數值之平均數作為中位數。故在本表中 22 項數字之中位數應為 \$0.758\$。



圖四十四 根據未經分組之資料決定中位數位置之圖解(七人之個人收入金額)

本例中中位數之值非為任何一州之平均工資率，此種情形在項數為偶數時所習見。又吾人所處理之資料如為間斷數列，則所得之中位數，不特為任何一項之數值所無，且為事實上所不致發現者。例如一九〇〇年至一九二二年紐約市場六十日至九十日期之商業票據貼現率，其中位數當為 4.865%，但此中位數純係假作之貼現率，而為實際市場上所絕無者。

中位數之決定：根據已經分組之資料

中位數之決定，在資料已分組而成次數分配之形式時，其手續亦大致相同，惟資料既已分組，即未能推知各個項目之實際數值，故決定中位數之手續較為複雜。茲就下列 276 個連續月份內紐約市場 60 日至 90 日期之商業票據每月平均貼現率，以說明決定中位數之方法。

表 二十五

紐約票據貼現率中位數之決定

(1900—1922年六十日至九十日期之雙記名商業票據)

組 距 (以百分率表示之貼現率)	中點 <i>m</i>	次數 <i>f</i>	
2.75—3.24	3.00	8	$\frac{N}{2} = \frac{276}{2} = 138$
3.25—3.74	3.50	31	
3.75—4.24	4.00	52	$Md = 4.75 + \left(\frac{9}{39} \times .50\right)$
4.25—4.74	4.50	38	
4.75—5.24	5.00	39	$= 4.75 + .115$
5.25—5.74	5.50	42	
5.75—6.24	6.00	35	$= 4.865$
6.25—6.74	6.50	13	
6.75—7.24	7.00	5	
7.25—7.74	7.50	4	
7.75—8.24	8.00	9	
		276	

在本例中決定中位數，即為決定其上下兩邊各占 138 項之一項數值。吾人試從尺度下端起逐漸向上移動以點算各組之次數，至第一組（即包含 2.75 至 3.25 各值之一組）上限時，則已點算 8 項，其餘 268 項尚未點算；至第二組上限時，已點算 39 項；至第四組上限時，已點算 129 項；在第五組上限處則已點至 168 項。因中位數兩邊之項數須各為 138 項，故中位數必在第五組上下兩限之間，然自該組下限 4.75 起移動至何處時，始為中位數之所在點乎？

吾人當可回憶為計算平均數起見，嘗假定每組內各項目有勻整之分配。在點算各組項數至第五組下限時，已點算 129 項，則在該組內 39 項中尚須點算 9 項以補足 138 項之數。吾人既已假定每組內項目分配勻整，則該組內第 9 項之數值，當為橫尺度上相對於該組組距之 $\frac{9}{39}$ 之一點。該組組距為 .50，故 .50 之 $\frac{9}{39}$ 等於 .115。吾人在尺度上自下端起逐漸移動已至 4.75 處，現仍須繼續向上移動 .115 之距離。此點在尺度

上之數值為4.865,即係使兩邊各占138項之劃分點,亦即中位數之值也。

此項中位數之計算手續,見上列次數表之右方。茲更將中位數之計算步驟簡述於下:

1. 將資料整理成一次數分配之形式。
2. 將總項數以2除之,即為中位數兩邊每邊所應有之項數。
3. 自尺度下端起,將各連續組之項數一一相加,至中位數所在組之下限為止。
4. 決定該組內所應補湊之項數。此項數與前節已求得之項數相加時須適等於 $\frac{N}{2}$ 項之數。
5. 將所應補湊之項數,除以中位數所在組之項數,此即補湊項數所占該組全距之分數。
6. 以所得分數與組距相乘。
7. 更將此乘積與中位數所在組之下限相加,即得中位數之值。

前述末三項步驟,即係簡單之插補法。

在點算項數時,亦可變其方向,自尺度上端起逐漸向下點算,惟此時前述第7項步驟,非加法而為減法,係從中位數所在組之上限,減去分數與該組組距相乘之積。

如 N 為奇數,則 $\frac{N}{2}$ 將為帶分數,至於中位數之決定手續,仍與前述相同。

四分位數與十分位數

除中位數以外,吾人又常須在 x 尺度上決定將次數分配之全體項數劃分為其他相等部分之界點。中位數係劃分全體項數為兩個相等部分,四分位數(quartile)、十分位數(decile)以及百分位數(percentile)之意義亦與此類似。所謂四分位數者,依其字義而言,當為尺度上劃分全體項數為四個相等部分之界點,十分位數為劃分全體項數為十個相

等部分之界點，百分位數為劃分全體項數為百個相等部分之界點。故第一四分位數為將尺度上全體項數分為 $\frac{1}{4}$ 與 $\frac{3}{4}$ 之比例之界點，第二四分位數與中位數之數值相同，第三十分位數係將尺度上全體項數分為 $\frac{3}{10}$ 與 $\frac{7}{10}$ 之比例之界點。以上所述，在點算項數時，皆自尺度之下端起。

例一. 第一四分位數(Q_1)之計算：貼現率(根據表二十五)

$$\frac{N}{4} = 69$$

$$Q_1 = 3.75 + \left(\frac{30}{52} \times .50\right) = 4.038$$

例二. 第四十分位數(D_4)之計算：貼現率(根據表二十五)

$$\frac{N}{10} = 27.6 \quad D_4 = 4.25 + \left(\frac{19.4}{38} \times .50\right)$$

$$\frac{4N}{10} = 110.4 \quad = 4.505$$

依圖決定中位數、四分位數、十分位數與百分位數之方法，當於後文論述之。

衆數之決定

衆數者，為一次數曲線上相對於最高縱坐標之 x 變量之數值也。衆數值之概念，極易悟解，因其為最普通之值。例如最普通之工資，最普通之收入，最普通之身材高度等。此值為次數最為密集之一點，故德人費希奈爾氏(Fechner)嘗稱此種平均數為 *dichtester wert*，英譯為 *thickest value*，意謂次數最為密集之數值也。據此定義，則衆數之特質已能確切表現，但有時真實之數值頗不易決定。通常從事統計工作時，欲求此衆數，祇能得其近似值，蓋在應用方面此近似值通常已够準確也(註)。

決定衆數近似值之方法，可就下列次數分配說明之。

(註)欲求更精確之衆數數值，其方法詳見第十五章末之附註。

表 二 十 六

五 釐 債 券 之 次 數 分 配

(此表係根據 1923 年十月三十一日紐約證券交易所

所開債息五釐鐵路及實業債券之行市編製)

價 格 組距	中 點 m	次 數 f
\$71.5 以下		30
71.5—73.49	72.5	2
73.5—75.49	74.5	5
75.5—77.49	76.5	5
77.5—79.49	78.5	5
79.5—81.49	80.5	12
81.5—83.49	82.5	13
83.5—85.49	84.5	15
85.5—87.49	86.5	11
87.5—89.49	88.5	17
89.5—91.49	90.5	17
91.5—93.49	92.5	33
93.5—95.49	94.5	45
95.5—97.49	96.5	51
97.5—99.49	98.5	59
99.5—101.49	100.5	23
101.5—103.49	102.5	9
103.5—105.49	104.5	1
		353

表內 71.5 以下之 30 項，其數值有極大之離散度，且該組祇有上限而無下限，故根據此表事實上無法計算其算術平均數。遇此情形，平均數以採用衆數爲宜。

在各組中，以 97.5—99.49 一組所包含之項數爲最多，該組似爲衆數組，該組之中點 98.5 似可作爲衆數之近似值；但若分組變更，則可得絕不相同之衆數值。試將原有債券市價，依照大小不同之數個組距分別製表，可得如下之結果（此處僅舉示中間各組之次數，已足證明衆數值

隨組距大小而有變更，似無須將全表各項一併列入)。

(a)		(b)		(c)	
組距 = 1		組距 = 2		組距 = 3	
組距	<i>f</i>	組距	<i>f</i>	組距	<i>f</i>
93.5—94.49	17	94.5—96.49	54	91.5—94.49	50
94.5—95.49	28	96.5—98.49	50	94.5—97.49	79
95.5—96.49	26	98.5—100.49	52	97.5—100.49	77
96.5—97.49	25				
97.5—98.49	25				
98.5—99.49	34				
99.5—100.49	18				
100.5—101.49	5				

以 1 為組距，得衆數值為 99，以 2 為組距，並使各組組限稍異於表二十六所列，得衆數值為 95.5，以 3 為組距，得衆數值為 96。如分組再予變更，則仍可得其他不同之衆數值。由此觀之，衆數似為一種不可捉摸之平均數，蓋根據同一資料，衆數值常因組距之大小與組限位置之變更而有差異。

惟此困難之發生，率由於事實上取樣不能過多所致。吾人果能將取樣之範圍為無限制之擴大，而將包含之項數為無限制之增加，則必有極大之次數密集於某一數值上，而真實之衆數乃可由此確定，蓋若包含之項數較多，則組距縮小時，衆數之近似值漸與真實衆數接近。組距過大，詳盡之事實每為所掩蔽；反之，組距縮小時，所表現之事實必較詳盡，而次數分配之真相乃可明白顯示；但事實上取樣不能過多，故組數增加時，各組間之次數每呈間斷與不規則之狀態，因之分配遂欠對稱與勻整，而最大集中點之實際地位，乃亦難於決定。前述根據債券市價所製各表，即為其一例。

如應用數學方法，則項數不為無限制之增加，亦可求得真實衆數之值。關於曲綫修勻之手續，前嘗述其梗概。修勻曲綫之一法，可依該曲綫所根據之次數分配資料，配以適當之理想次數曲綫。此曲綫在理論上係表示組距無限縮小與項數無限增加時所應有之次數分配。在此理想次數曲綫上所得最高縱坐標之 x 值，即為真實衆數。

在實際應用方面，僅求衆數之近似值，亦已適用，而可採用更簡單之各種方法以求此近似值。前節所述以次數最大一組之中點作為衆數，雖係一粗略之方法，但若分組時果能注意前述各項分組原則，則依此方法所得結果，大體尚不致發生重大差誤。

如次數分配尚屬勻整，可就衆數組內用插補法以推定其衆數值，由此所得之值較以該組中點作為衆數值者，為更與真實衆數相近似。茲仍以表二十六債券價格為例以說明之。按該表內衆數組兩邊之次數分配並不對稱，衆數組之中點為 98.5，較衆數組略小一組之中點為 96.5，包含價格 51 項，較衆數組略大一組之中點為 100.5，包含價格僅 23 項；而次於衆數組上下鄰接兩組之各組，其兩邊次數分配亦欠平衡，有偏向於小於衆數組之一邊之趨勢。以前計算算術平均數與中位數時，嘗假定每組內各項數值在該組上下兩限間分配勻整，但本例中衆數組內各數值之分配並不勻整，由衆數組以外各組之次數分配觀之，似可推定衆數組內各項之次數有偏向於該組下半部即 97.5 至 98.5 間之可能，故此衆數大抵在中點 98.5 以下，而未見適在中點。吾人試用加權法以確定衆數在該組內之位置，並假定此衆數被拉向尺度下端之力為 51（即係較衆數組略小一組之次數），被拉向尺度上端之力為 23（即係較衆數略大一組之次數）。此可用下列符號，以公式表示之。

l = 衆數組之下限。

f_1 = 較衆數組略小一組之次數。

f_2 = 較衆數組略大一組之次數。

i = 組距。

插補公式為

$$M_o = l + \frac{f_2}{f_2 + f_1} \times i$$

茲應用此公式以計算表二十六債券價格之衆數值，則得

$$M_o = 97.5 + \frac{23}{74} \times 2 = 97.5 + .62 = 98.12$$

如權數(以 f_1 及 f_2 代表者)根據大於衆數組與小於衆數組之鄰接各兩組或各三組之總次數,則所得衆數值更爲近似。在本例中如根據衆數組兩邊各三組之總次數計算,則得債券價格之衆數值爲 97.91。

衆數之位置,有時因次數並不如前例之密集於一點,而分散於各點,以致不易確定者。如表九所示以 \$.25 爲組距之工資分配,具有兩個明確之衆數點。此種分配謂之雙峯分配 (bi-modal distribution),因圖示時其次數曲線呈現兩個高峯也。如全部資料確屬同一性質,則此種欠勻整之分配,係由於所取資料不足與所採分組方法欠當所致。如所取樣本包含項數不多,而組距失之過小,則往往有得兩個衆數點之可能。遇此情形,吾人可變更組限與增加組距,使次數分配內祇存一個衆數組,而由此決定衆數之近似值。此與包含項數極多時求真實衆數之手續正相反。當項數爲無限多時,則組距可縮至無限小,以便求得真實衆數之值;然在包含項數無多時,則欲決定次數最爲密集之一點,必須將組距增大,以消除因所取樣本不足而發生次數分配不規則之狀態。

如變更組距與組限,次數分配仍具雙峯形態,此或由於資料性質不一,誤將兩個性質不同之分配,混爲一個所致。此種錯誤在研究生物測量中所習見,例如誤認兩種不同之動物爲一種,因之所得分配常現雙峯狀態,而表示其中實包含兩種不同之分配。如資料性質互異,則次數分配失其整個之意義,不僅在研究經濟統計時爲然,即研究其他統計時,亦莫不皆然。

由算術平均數及中位數決定衆數之方法

此外求衆數近似值之方法,亦有根據算術平均數、中位數及衆數三者間之關係以計算者。次數分配完全對稱時,算術平均數、中位數及衆數三者之值全相等;次數分配如不對稱,此三者之值各不相同;如次數

分配不對稱之程度不甚顯著，則三者之間恆有一定之關係：在次數曲綫圖中，衆數與算術平均數之相距最遠，而中位數與算術平均數之距離，約等於算術平均數與衆數距離之三分之一。但若次數分配極不對稱，則三者之間無此關係。在次數分配之不對稱程度不甚顯著時，三者之中如已知任何兩種平均數之值，即可由三者之關係以求得第三種平均數之值。惟此法在實際應用上僅可適用於衆數值之決定，因其餘兩種平均數之值，可採用其他更精確之計算方法也。即在求衆數值時，亦必在其他更精確之計算方法不能應用時，始可依此法計算。

下列公式係根據算術平均數、中位數及衆數三者之關係而得。

$$M_o = Mean - 3(Mean - Md)$$

茲將表十二電桿木之資料代入公式，得下列結果：

$$M_o = 9.33 - 3(9.33 - 9.015) = 8.385$$

此值略小於衆數組之中點 8.5，而同時亦小於按照加權法在衆數組內插補所得 8.49 之數值（係用衆數組左右鄰接各四組之次數作為權數者）。

吾人有須注意者，上列各個衆數值孰為準確，殊無標準可言。前述關於決定衆數之各種方法，亦僅能求其近似值，此點在解釋或應用其結果時，切宜注意及之。

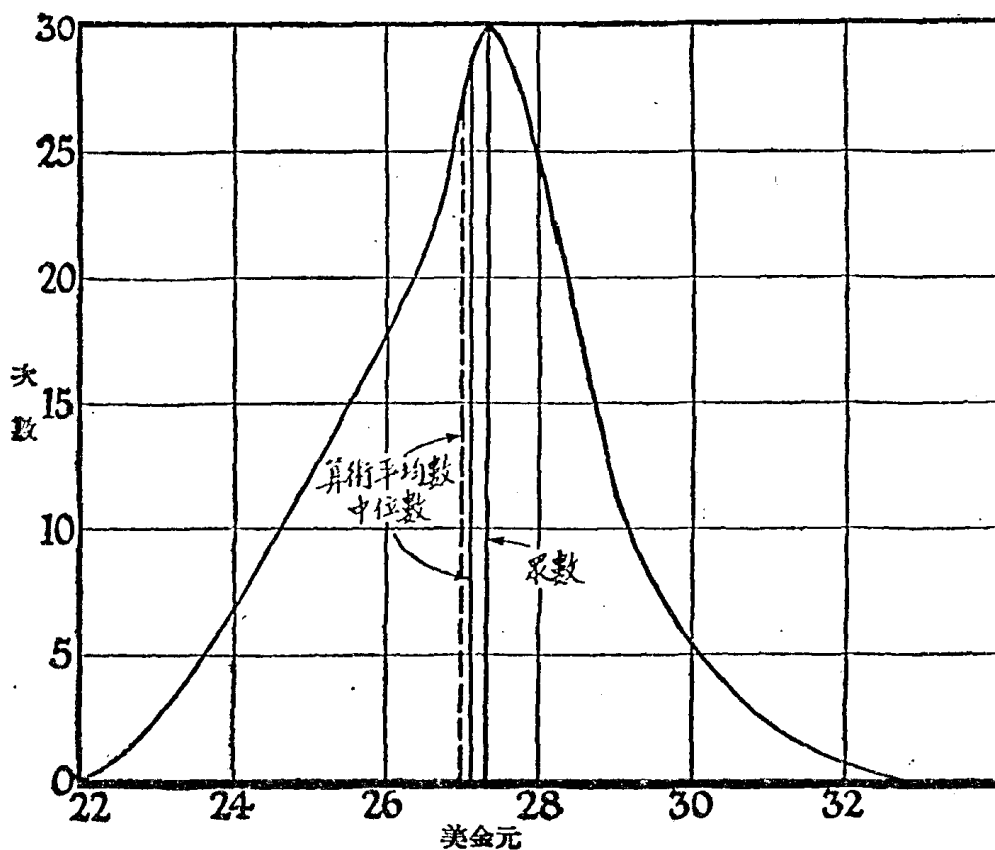
按圖決定衆數中位數四分位數與十分位數之方法

吾人如將按圖確定前述各種量數之方法，略加討論，則對於次數曲綫與累積次數曲綫可得更確切之瞭解。

在普通次數曲綫圖中，衆數之值頗易確定，蓋依衆數之定義，其值當為 x 尺度上相對於曲綫最高縱坐標之一點。在次數多邊圖中，亦可得粗略之衆數值，此值即為次數最多一組之中點。再如欲得更迫近於真實衆數之值，則須用觀察法或數學方法，在修勻之曲綫圖中求之。圖四

十五係說明按圖決定衆數之方法。圖中曲綫（係用觀察法修勻者）係根據表八之工資資料繪成者。在橫尺度上與曲綫最高縱坐標相對之值為 \$27.50，此僅係衆數之近似值，可與按照加權法求得之值 \$27.69 及由算術平均數與中位數計算所得之值 \$27.3470 比較之。

中位數與算術平均數之位置，已在該曲綫上記明。前嘗論及次數分配不對稱（偏斜）程度不甚顯著時，算術平均數、中位數及衆數三者間恆有一定之關係：中位數係在算術平均數與衆數之間，其與算術平均數之距離，約等於算術平均數與衆數兩者間距離之三分之一。證諸本例，此種關係確見存在，故在此修勻之曲綫圖中，可求得衆數之近似值，惟因原有資料欠整齊，故用觀察法以修勻曲綫，自不免稍覺武斷耳。



圖四十五. 按某週工資分組之工人次數分配之修勻曲綫
表示算術平均數、中位數與衆數三者關係之圖解

圖四十六係根據同一資料，按照表二十七之分組所繪成之累積次

數曲綫。此累積次數曲綫在某距離內之峭度 (steepness), 係隨該距離內之項數多寡而定。該曲綫其始緩漸上升, 繼而上升較速, 最後在曲綫之上端上升又轉緩漸, 故其衆數之值, 當為橫尺度上相對於峭度最高之一點, 在此點上次數增加最速, 亦即次數分配中項數最為集中之一點也。在修勻次數曲綫中, 此衆數值可就曲綫上斜度最高一點 (即轉向點 point of inflection) 之 x 值決定之, 而在本例累積次數曲綫中, 亦可用此法以推定其衆數之值約為 \$27.50。

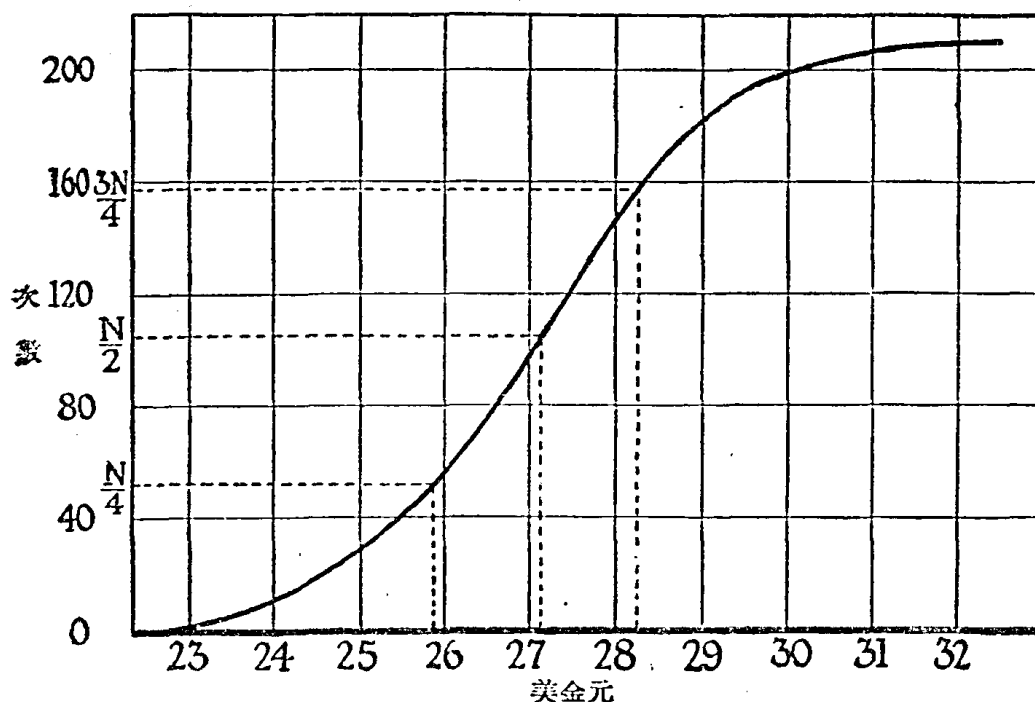
表 二 十 七

某製造廠工人之累積次數分配
(按照某週工資數分組)

某週工資數	工人數 (次數)
\$22.50以下	0
23.00以下	1
23.50以下	5
24.00以下	8
24.50以下	19
25.00以下	29
25.50以下	41
26.00以下	56
26.50以下	78
27.00以下	98
27.50以下	122
28.00以下	162
28.50以下	169
29.00以下	186
29.50以下	193
30.00以下	199
30.50以下	204
31.00以下	208
31.50以下	209
32.00以下	209
32.50以下	210

至於中位數、四分位數及十分位數之值, 亦可在累積次數曲綫圖中

得。此種曲綫一經修勻，則應用插補法可得良好之結果。又若圖之尺度够大時，用插補法且可得精確之數字。在縱尺度上（即表示累積次數之尺度）確定自底綫起約等於 $\frac{N}{2}$ 距離之一點，由此點起繪一水平綫，與累積次數曲綫相交，此交點之橫坐標，即為中位數之數值。在決定此值時，可由交點繪一縱綫，垂直於 x 軸上量其橫坐標即得。圖四十六說明此法之應用。用此法求得中位數之數值為 \$27.125，如選用插補法則得 \$27.1458。至於四分位數亦可依照此法求得，其縱尺度應劃作四等分，更由此分點各繪水平綫交於累積次數曲綫可矣。



圖四十六：根據某週工資之累積次數曲綫決定中位數及四分位數之圖解

關於平均數之選擇，在某種情形之下，尤其在求百分率 (rates) 或比率 (ratios) 之平均，而非實際數量之平均時，前述各種平均數皆不適用，惟幾何平均數及倒數平均數則較為適宜，故有研究之必要。

幾何平均數

幾何平均數係 N 項數字相乘之積之 N 方根，其值以公式表示之如

下：

$$M_g = \sqrt[n]{a_1 \cdot a_2 \cdot a_3 \cdots a_n}$$

例如2,4,8三個數字之幾何平均數為

$$\begin{aligned} M_g &= \sqrt[3]{2 \times 4 \times 8} \\ &= \sqrt[3]{64} \\ &= 4 \end{aligned}$$

依上式計算，可知一個數列中如有一項數值為零時，其幾何平均數之數值亦為零。

計算幾何平均數如用對數，可節省手續不少。其算式如下：

$$\text{Log } M_g = \frac{\log a_1 + \log a_2 + \log a_3 + \cdots + \log a_n}{N}$$

幾何平均數之對數，等於各數量之對數之算術平均數。

在計算幾何平均數時，如欲將各數量分別加權，則各個權數即為各數量之指數。今設權數之和為 N ， $a_1, a_2, a_3, \cdots, a_n$ 各數量之權數各為 $w_1, w_2, w_3, \cdots, w_n$ ，則加權幾何平均數之公式為

$$M_g = \sqrt[n]{a_1^{w_1} \cdot a_2^{w_2} \cdot a_3^{w_3} \cdot \cdots \cdot a_n^{w_n}}$$

按此式無異將每項數量依其權數重複若干次，正與計算加權算術平均數時情形相似。如用對數計算，則加權幾何平均數之公式當為

$$\text{Log } M_g = \frac{w_1 \log a_1 + w_2 \log a_2 + w_3 \log a_3 \cdots + w_n \log a_n}{N}$$

茲就下表數字以說明幾何平均數之計算方法。此表係紐約證券交易所所開一一五種紅利優先股票價格之分配。表內價格均係一九二二年八月二十六日之收盤價

表二十八
優先股票價格幾何平均數之計算

組距	中點 <i>m</i>	次數 <i>f</i>	組距中點之對數 <i>log m</i>	組距中點之對數×次數 <i>f log m</i>
\$35—\$44.9	\$ 40	1	1.602060	1.602060
45— 54.9	50	6	1.698970	10.193820
55— 64.9	60	8	1.778151	14.225208
65— 74.9	70	5	1.845098	9.225490
75— 84.9	80	14	1.903090	26.643260
85— 94.9	90	22	1.954243	42.993316
95—104.9	100	27	2.000000	54.000000
105—114.9	110	18	2.041393	36.745074
115—124.9	120	14	2.079181	29.108534
		115		224.736792

$$\text{Log } M_g = \frac{224.736792}{115}$$

$$\text{Log } M_g = 1.954233$$

$$M_g = \$ 90.00$$

幾何平均數之特質

幾何平均數之特質若何，可就該平均數與各數量間之關係推求而得。

在一個數列中，其算術平均數恆可替代各數量，而使各數量之原有總和不變。例如2, 4, 8三個數字之和為14，其算術平均數為 $4\frac{2}{3}$ ，如將原有三個數字各以此平均數替代之，則其總和仍為14，此為算術平均數之特質。至於幾何平均數之特質則為在一個數列中，其幾何平均數恆可替代各數量，而使各數量之原有乘積不變。例如2, 4, 8三個數字之乘積為64，其幾何平均數為4，如將原有三個數字各以此平均數替代之，則其乘積仍為64。

復次，算術平均數之特質，為大於平均數各項離中差之和，適等於小於平均數各項離中差之和(不計其正負號)；換言之，即大於平均數各項對於平均數差額之和，與小於平均數各項對於平均數差額之和，彼此相等也。幾何平均數之特質，則為大於幾何平均數與小於幾何平均數兩

邊各項對於該平均數所得比率之乘積彼此相等；換言之，即幾何平均數對於較小各項所得比率之乘積，等於較大各項對於幾何平均數所得比率之乘積。例如3, 6, 8, 9之幾何平均數為6，可得下列等式：

$$\frac{6}{3} \times \frac{6}{6} = \frac{8}{6} \times \frac{9}{6}$$

上述第二點，尤為幾何平均數最主要之特質。此係求平均比率之一種方法，其在經濟統計方面之主要用途，係供編製物價指數之用，因指數所注重者為物價變動之百分率。物價自50漲至100，其重要程度正與100漲至200相等，惟此種相等非算術平均數所能表示。如用算術平均數計算，則前者所增加之絕對數為50，而後者所增達100，是不啻予後者以倍重之權數。吾人習常所舉之例為兩項物價之變動，其一自100漲至1000，計漲十倍，其另一項自100跌至10，僅及原價之十分之一。1000與10之算術平均數為505，其幾何平均數為 $\sqrt{1000 \times 10}$ ，即100。由此幾何平均數可知一漲一跌程度相等，故相抵銷，而算術平均數乃高至505，顯屬謬誤，其不適於求平均比率之用，當可想見。此點當於第六章物價指數內再為詳述。

前章內所述用對數尺度製圖對於某種目的所具有之優點，實與幾何平均數之用途有關。幾何平均數亦有稱為對數平均數者，則以幾何平均數之對數，即係各數量對數之算術平均數之故。在計算變動百分率之平均時，所注重者為比率而非絕對差數，故幾何平均數最為適用。

在計算本金按照複利率累積之本利和時，尤非採用幾何平均數不可。今設 P_0 為本金， P_n 為若干時期後之本利和， r 為利率， n 為若干時期內之年數，則 n 年後按照每年複利一次， P_0 累積所得之本利和，當如下列方程式所示：

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

自上式化為

$$r = \sqrt[n]{\frac{P_n}{P_0}} - 1$$

例如本金1000元按複利計算，十二年後之本利和為1600元，較本金增60%，如按算術平均數計算，則利率為5%，但此非本利累積之百分率，其實際利率應為：

$$\begin{aligned} r &= \sqrt[12]{\frac{1600}{1000}} - 1 \\ &= \sqrt[12]{1.60} - 1 \\ &= 1.04 - 1 \\ &= .04 \text{ 即 } 4\% \end{aligned}$$

此外凡欲求增減百分率之平均時，當發生同樣問題，如用算術平均數必得錯誤之結果。

幾何平均數表示集中趨勢之討論

或問何種次數分配，其集中趨勢應以幾何平均數代表之乎？此一問題如加以討論，頗饒興趣。大抵當絕對數值繪於算術尺度上，其次數分配頗具對稱狀態時，則算術平均數恆優於幾何平均數；但若絕對數值繪於算術尺度上，其曲綫並不對稱，而繪於幾何尺度上可減除其不對稱狀態，而成一對稱曲綫時，則幾何平均數優於算術平均數。在後一種分配中，呈對稱形態者，為集中趨勢兩旁之相對離中差 (relative deviation)，即以比率表示之離中差，而非絕對離中差 (absolute deviation)。此種次數分配如欲確切表示其集中趨勢，應採用各數量對數之算術平均數（此值即為絕對數量之幾何平均數之對數，前已述及）以代表之。由各對數所得之曲綫，當在幾何平均數之對數之兩邊成對稱狀態。例如各項物價漲落百分率之對數所繪成之次數曲綫，必在其幾何平均數之對數之兩邊呈對稱狀態。但此種漲落百分率，其真數之分配狀態並不對稱，大於算術平均數各項之全距，必遠較小於算術平均數各項之全距為

大(註一),此因物價上漲時,可漲至1,000%或1,000%以上,而在物價跌落時,則不能跌去100%以上也。關於此點,在第六章物價指數內當有更詳盡之討論(註二)。

此種根據各個數值之對數而不根據其真數繪製之次數曲綫,對於個人收入分配之資料,尤為適用。此項資料如按真數繪製,則以收入分配之全距過大,每使曲綫各部份之特性不能在圖中表現,但若繪於雙對數圖中(即 x 及 y 皆用對數繪製者),即能避免此種困難,而可揭示全部分配之真象及各部分間相互之關係,同時並可表明為真數尺度所不能顯示之某種要點。此種雙對數圖似可使曲綫中大於衆數之部分,化為平滑而成直綫,故倡用此雙對數圖以表述個人收入資料之巴拉圖氏(Vilfredo Pareto)嘗賴以發明關於收入分配之巴拉圖法則(Pareto's Law)。當美國經濟研究所精密研究美國人民收入分配之結果後,對於根據巴拉圖法則所推演而得之各種結論雖置懷疑,而對於雙對數尺度之用於收入分配之資料,固仍承認其具有價值也。關於此種次數曲綫,在美國經濟研究所所編美國人民收入("Income in the United States")報告書中嘗作具有興趣之論述。

(註一)參看圖四十九及圖五十。

(註二)美國 C. M. Walsh 氏所著 "The Problem of Estimation" 第35頁對於平均數之選擇,嘗定有下列標準:

- (a) 某數列各項中其最大值或最小值之限度不可估量或不能確定時,以採用算術平均數為宜。
- (b) 某數列各項中之最小值可確定為0或大於0,而其最大值之限度不可估量或不能確定時,以採用幾何平均數為宜。物價變動之百分率即其一例,故 Walsh 氏認為編製物價指數,以採用幾何平均數最為適當。
- (c) 在實際應用方面,或由於資料之性質關係,最大值與最小值之限度皆存在,而不能應用前項標準時,則必須先行研究資料之實際離散度。如其衆數與算術平均數較為相近,則應採用算術平均數,如其衆數與幾何平均數較為相近,則應採用幾何平均數。

倒數平均數

倒數平均數 (harmonic mean) 為平均數中用途較為狹隘者，惟在某種資料中如採用倒數平均數，則可避免差誤之發生。在計算時間速率之平均時，非用倒數平均數不可，而對於他種特殊資料，亦常有應用之必要。下述之例係說明此種平均數之應用方法：

例如某汽車行駛4英里之距離，其速率為每小時20英里，嗣又續駛4英里，其速率為每小時30英里，欲求該汽車之平均速率。按上述兩項速率之算術平均數為每小時25英里，但照此計算係屬錯誤。該汽車係先按每小時20英里之速率行駛12分鐘，嗣按每小時30英里之速率續駛8分鐘，是兩次共計行駛8英里之距離，費時共20分鐘，故其平均速率當為每小時24英里。此實即20及30之加權平均數，前者之權數為12，後者權數為8也。但若按照兩個速率計算其倒數平均數，可得相同之結果。若干數值之倒數平均數，即為各數值倒數之算術平均數之倒數。今設 r_1, \dots, r_n 代表所欲平均之各個速率，則求倒數平均數 (H) 之公式為

$$\frac{1}{H} = \frac{\frac{1}{r_1} + \frac{1}{r_2} + \frac{1}{r_3} + \dots + \frac{1}{r_n}}{N}$$

用上項數字代入公式，得

$$\begin{aligned} \frac{1}{H} &= \frac{\frac{1}{20} + \frac{1}{30}}{2} \\ &= \frac{5}{120} \\ H &= 24 \end{aligned}$$

如欲求 a, b 兩數之倒數平均數，則由公式 $H = \frac{2ab}{a+b}$ 計算，手續更為簡便。如有 a, b, c 三個數字，則可用公式 $H = \frac{3abc}{ab+ac+bc}$ 。此兩公式與前述第一公式相等，但若數列中項數較多，則計算倒數平均數

時，如用製就之倒數表，手續可便利不少(註)。

倒數平均數在分析經濟資料方面之用途，可以物價為例以說明之。例如物品論價有按“每元若干件”者，欲求此種價格之平均數，應用算術平均數每得錯誤之結果。今設有高下不同之物價三項，一為“每元四件”，一為“每元五件”，一為“每元二十件”，問平均每元售若干件(即平均價)? 根據上述數字(4, 5, 20)求得算術平均數為 $9\frac{2}{3}$ ，驟視之似為平均每元所售件數，即每件平均價為10.34分；但按諸實際，原有之三項價格，亦即每件25分、每件20分及每件5分之謂，其算術平均數為 $16\frac{2}{3}$ 分，如易為與原價同一形式之價格，此平均數即為每元6件。如欲予每個物價以相等之權數，則後一平均數係準確之答案。蓋按照“每元售若干件”之價格所求得之算術平均數，乃係加權平均數，對於每元所售件數較多之物價，不啻即予以較大之權數。上項準確之結果，亦可直接按照原有物價用倒數平均數求之，蓋4, 5, 20三個數字之倒數平均數為6，即平均每元所售之件數也。

各種平均數之相互關係

如根據一個數列同時求得各種平均數，則各平均數間恆有各種關係之存在。

1. 在完全對稱之次數分配圖中，算術平均數、中位數與衆數三者同在一點。
2. 在偏態不甚顯著之次數分配圖中，中位數之位置處於算術平均數與衆數之間，中位數與算術平均數之距離約等於算術平均數與衆數距離之三分之一。故此種次數分配，其三種平均數間之關係約如下式所示：

(註)一九一九年紐約 Spar and Chamberlain 公司出版之 “Barlow's Tables of Squares, Cubes, Square Roots, Reciprocals” 一書頗切實用。

$$M_o = M - 3(M - M_d)$$

3. 任何數列之算術平均數，必大於其幾何平均數。
4. 任何數列之幾何平均數，必大於其倒數平均數；惟若該數列中所有數量各各相等，則算術平均數、幾何平均數與倒數平均數三者亦必相等。
5. 任何兩數列之幾何平均數，等於倒數平均數與算術平均數兩者之幾何平均數。例如2與8之倒數平均數為 $3\frac{1}{5}$ ，算術平均數為5，幾何平均數為4，而 $3\frac{1}{5}$ 與5之幾何平均數亦為4也。但數列包括項數在兩項以上者，不在此列。
6. 統計資料之離散度(dispersion)如合乎算術定律，則衆數及中位數兩者之值，與算術平均數相差較近，而與幾何平均數相差較遠；但若其離散度合乎幾何定律，則衆數及中位數兩者之值，與幾何平均數相差較近，而與算術平均數相差較遠。

各種主要平均數之特質

算術平均數：

1. 算術平均數之值係受數列中每個數量之影響，故遇有極端數量存在時，頗嫌其所受影響過鉅，而不適於某種用途。
2. 算術平均數易於計算，且在任何數列中其值皆可確定。
3. 算術平均數係由計算而得，故可用代數方法處理之。

中位數：

1. 中位數之值不受極端數量之影響。
2. 統計事項之不能以數量表示者，仍可求得其中位數。
3. 凡統計資料不齊備，而能確知其各項之總項數與其大概之位置以及次數分配中間各項數量之大小時，其中位數亦可決定。
4. 中位數非若算術平均數、幾何平均數與倒數平均數三者之可用代數方法處理者。

衆數：

1. 衆數之值不受極端數量之影響。
2. 衆數之近似值易於確定，但欲決定真實衆數，則計算手續頗繁。
3. 次數分配如包含項數無多，集中趨勢又不顯著，則其衆數無意義。
4. 衆數之位置居於次數分配之密集點，故允為分配中之典型數值。
5. 衆數不能用代數方法處理。

幾何平均數：

1. 幾何平均數所受極端數量之影響，較算術平均數為小。
2. 在任何數列中，幾何平均數之值皆可確定。
3. 幾何平均數對於等比之變動視為同等重要，故欲求數量變動百分率之平均，或比率之平均者，非用幾何平均數不可，而用以求物價變動之比率，尤為適當。
4. 幾何平均數可用代數方法處理。

倒數平均數：

1. 倒數平均數適用於平均時間速率與類似性質之事物。在經濟統計方面，亦有用以計算物價指數者。
2. 倒數平均數因計算費時與意義不通俗，故在普通統計分析方面每棄而不用。
3. 倒數平均數可用代數方法處理。

上述各點係欲說明每種平均數各有其特殊用途。為適應某種目的或在某種情形下，一種平均數常較優於其他各種平均數，吾人應先明瞭每種平均數之特質及其缺點，始可知所取捨。惟欲將次數分配作詳盡之表述，常須同時斷定二三種主要平均數及其他各種統計量數之數值。算術平均數或為最切實用之一種平均數，因其計算簡便，且可適用代數方法處理，其意義既確切，又通俗，故極合於統計上之應用，惟其用途並不普遍，應在適當情形之下始可採用之耳。幾何平均數之優點漸為一般所

瞭解，故在統計方面之應用亦日廣。平均數之選擇對於統計分析頗關重要，學者不可不注意也。

參 考 書

- Bowley, A. L. "Elements of Statistics"(82—109, 138—139).
- Chaddock, R. E. "Principles and Methods of Statistics"(Chaps. VI, VII, VIII).
- Jones, D. C. "A First Course in Statistics"(22—41).
- Kelley, Truman L. "Statistical Method"(44—68).
- King, W. I. "Elements of Statistical Method"(121—140).
- Pearl, Raymond. "Medical Biometry and Statistics"(264—273).
- Rugg, H. O. "Statistical Methods Applied to Education"(97—148).
- Secrist, Horace. "Introduction to Statistical Methods"(234—292).
- Walsh, C. M. "The Problem of Estimation"(1—67)
- Yule, G. U. "An Introduction to the Theory of Statistics"(106—132).
- Zizek, Franz. "Statistical Averages"(82—109).

第五章 次數分配之表述：離散度 與偏斜度之量數

在前數章內吾人首先論及大量資料之簡縮方法，使化爲簡單形式，俾全部資料之特性可以易於確定；次又論及表述資料特性之方法。當資料已組成次數分配之形式，則前項目的業已達到；迨已求得一單純數值即平均數以代表次數分配之集中趨勢後，則後項目的亦已達到一部分。但任何一種平均數不能單獨表述次數分配之全部情形，如欲測度次數分配之主要特性，或與另一次數分配相比較，尚須另求三種數值而後可。其第一種爲測度次數分配中各項與其中心值之相離程度者，即離散度（“scatter”，variation or dispersion）之量數是也；第二種爲測度中心值兩邊之分配狀態爲平衡或不平衡者，即測度次數分配對稱程度（degree of symmetry）之量數是也；第三種爲測度在衆數值上次數之密集程度者，即測度次數分配之峯度（kurtosis）之量數是也。本章所欲論述者，爲測度離散度與偏斜度之各種方法，至於測量峯度之方法，容後討論。

離散度之性質與意義

數量資料之呈現離散現象，已在前文提及，此種現象對於統計分析上之關係，亦已約略述及。吾人實際所蒐集之各種數量資料，無論其屬於社會、生物或經濟方面，皆具有離散之特質，即各個數量大小不同是也。此各個數量不同之特質實與各個數量類似之特質同一重要。在生物方面，變動之現象爲進化過程之一要素。今若吾人欲測驗某種族生理方面之特質，例如體高，不能以測量各人之體高爲已足，而又須測度彼等之平均離中差。又如一國人民之平均收入額固屬重要，但社會各階級間收入差異之情形尤關重要。又如物價之變動，每足阻害經濟制度之正常運行，致物價制度下各組織份子因其所受變動之影響不同，而有備受困

苦者，亦有不勞而獲者，此非由於物價水準之一般的變動，而實由於各個物價變動程度不一所致也。

平均數之本身無甚意義，吾人必須同時求得次數分配之離散度以補充之。如次數分配之離散度甚大，且無顯著之集中趨勢，則其平均數毫無價值；離散度愈小，平均數愈有意義。故欲表述某一次數分配，或與另一次數分配比較，除測度集中趨勢外，又須測度其離散度以補充之。

絕對離散度之量數

離散度量數有用原來資料所用之數量單位表示者，亦有用抽象數字而與原來單位絕無關係之百分率表示者。用原來單位所測度者為絕對離散度(absolute variability)；用抽象數字表示者為相對離散度(relative variability)。在比較兩種次數分配之離散度時，後者較前者更為適用。茲先討論絕對離散度之量數。

全距

測量離散度之粗略方法，係以全距(range)為標準。全距者，即次數分配中最小一項數值與最大一項數值之絕對差數也。表三十所列係一八八二年至一九一三年倫敦對紐約之每月匯價。表內所包含最小一項之原來匯價為\$4.83，最大一項為\$4.908，故該次數分配之全距為\$4.908 - \$4.83，即\$.078。在尺度上此\$.078之距離內包含次數分配中之一切項數。如未知該兩項之原來數字，則其全距亦可自次數表內估計而得，即以分配中最小一組之下限與最大一組之上限相減之差數作為全距。依此所得表三十之全距為\$.085。

由是知全距之大小，全視兩極端之數值而定。如遇一項數值極大或極小，則全距之數值必大受影響。全距之數值因其太不穩定，且不能代表各項之實際分配狀態，故統計方面不常採用。全距恆有用以測度股票

市價之變動者，但是否合於此用，不無疑問耳。

平均差

測量平均數兩邊各項離散度比較準確之量數，尚有所謂平均差（mean deviation）者。平均差之計算，係先求各項對於平均數之差數（即離中差），而後求其平均差數。其計算方法可就下列簡例說明之：

表 二 十 九
平均差之計算

m	f	d	
3	1	6	$M = 9$
6	1	3	
9	1	0	$M.D. = \frac{18}{5} = 3.6$
12	1	3	
15	1	6	

前項次數分配之平均數（在本例中算術平均數與中位數之值相等）為9。先將各項之離中差，不問其符號為正為負，一一相加，再將所得和數除以項數，即得該分配之平均差。其計算手續可用下式表示之：

$$M.D. = \frac{\sum d}{N}$$

如用文字說明，則一個數列之平均差即係該數列內各項對於平均數（或為算術平均數，或為中位數）之離中差之算術平均數也。在各項離中差相加及平均時，其符號之為正為負，均所不問。至於計算各項離中差所根據之平均數，或為算術平均數，或為中位數，實際上無大出入，惟理論上似以後者為宜。則因平均差之由中位數計算者，其數值為最小故也。

表三十係說明統計資料已組成次數分配表時，計算平均差之方法。吾人猶憶以前根據次數分配表計算算術平均數時，嘗假定每組內各項數值密集於其組中點，換言之，即每組內各項之數值，皆可視為等於各

該組之中點；今於平均差之計算，亦作同樣之假定。

表 三 十

1882—1913年倫敦對紐約匯價平均差之計算

(本表數字係根據每月月初之匯價)

組 距	組中點	次數	對於假定 平均數之 離中差		
	<i>m</i>	<i>f</i>	<i>d'</i>	<i>fd'</i>	
\$4.8275—\$4.8324	\$4.830	1	\$.040	\$.040	假定平均數 = \$4.8700
4.8325—4.8374	4.835	6	.035	.210	中位數 = 4.8721
4.8375—4.8424	4.840	11	.030	.330	差 數 = \$.0021
4.8425—4.8474	4.845	21	.025	.525	各項離中差須補足\$.0021
4.8475—4.8524	4.850	23	.020	.460	者196項
4.8525—4.8574	4.855	24	.015	.360	各項離中差須減去\$.0021
4.8575—4.8624	4.860	25	.010	.250	者188項
4.8625—4.8674	4.865	40	.005	.200	淨計：離中差須補足\$.0021
4.8675—4.8724	4.870	45	者8項
4.8725—4.8774	4.875	49	.005	.245	各項對於假定平均數之離
4.8775—4.8824	4.880	35	.010	.350	中差之和 = \$5.020
4.8825—4.8874	4.885	45	.015	.675	校正數(8×\$.0021) =
4.8875—4.8924	4.890	33	.020	.660	\$.0168
4.8925—4.8974	4.895	16	.025	.400	各項對於中位數之離中差
4.8975—4.9024	4.900	8	.030	.240	= \$5.020 + .0168
4.9025—4.9074	4.905	1	.035	.035	= \$5.0368
4.9075—4.9124	4.910	1	.040	.040	M. D. = $\frac{\$5.0368}{384}$
		384		\$5.020	= \$.01312

由此表計算平均差，亦可採取表二十九之計算方法，先求每組中點對於中位數之差數，乘以各該組之次數，然後一一相加，再以項數除之。例如本表匯價之中位數為 \$4.8721，第一組內 1 項對於中位數之差

數為\$.0421；第二組內共6項，其每項對於中位數之差數為 \$.0371，餘類推。惟照此一一計算其真實離中差，因有小數，手續頗為繁瑣，為求簡捷起見，可從假定中位數(或假定算術平均數)以計算離中差，再將各項離中差相加，而後再按校正數校正之。此校正數即為計算離中差時不根據真實平均數，而根據假定平均數所發生之差誤也。表三十即係說明此種簡捷計算法。

吾人如根據算術平均數計算平均差，亦可用此簡捷法。用簡捷法計算時，如以組距為單位，則平均差之計算手續更可簡化，與前章內計算算術平均數之簡捷法相似。

在表三十內計算各項離中差時，係根據假定平均數 \$4.870 而並不根據真實中位數 \$4.8721 計算者。各項對此假定平均數之差數，不問其符號之為正為負，一一相加，其總和為 \$5.020，然此總和究與根據中位數計算之各項離中差之和相差若干乎？在以組中點作為假定平均數之一組內共包含45項，每項之 d' (即對於假定平均數之離中差)皆為0，但實際該組內每項對於中位數之離中差應為 \$4.8721 - \$4.870，即\$.0021，故該組內每項由假定平均數所得之離中差須各補足 \$.0021 之數而後可。在計算小於該組之各組內每項之離中差時，亦發生同樣之差誤；總計發生此項差誤者共196項，每項均應補足 .0021 之數。在大於假定平均數所在組之各組內每項之離中差則發生異方向之差誤。例如該組之上一組共包含49項，每項對於假定平均數之離中差為 \$.005，而每項對於中位數之離中差應為 \$.0029，故應由各項離中差減去 \$.0021；總計發生此項差誤者共188項。盈虧相抵，淨計尚須補足 \$.0021 者8項，故各項離中差之總和 \$5.020 尚須補足 $8 \times $.0021，即$.0168 之數，由此求得各項對於中位數之離中差之和為 $5.0368，其平均差為 $.01312。$

此計算平均差之簡捷法可化為下列公式

$$M.D. = \frac{\Sigma(fd') + (N_s - N_c)c}{N}$$

公式中 N_1 = 對於假定平均數之離中差比較對於中位數 (或算術平均數) 之離中差為小之各項項數。

N_2 = 對於假定平均數之離中差比較對於中位數 (或算術平均數) 之離中差為大之各項項數。

c = 假定平均數與中位數 (或算術平均數) 之差數。

在應用此項公式時，假定平均數與中位數 (或算術平均數) 須在同一組內(註)。

茲更列計算平均差之程序於下：

1. 求中位數之數值。
2. 以中位數所在組之中點作為假定平均數，求每組各項對於假定平均數之離中差，更將此差數乘以各該組之次數，而後求乘積之總和，其乘積之符號為正為負，可不置論。
3. 點算對於假定平均數之離中差比較對於中位數之離中差為大之各項項數，及對於假定平均數之離中差比較對於中位數之離中差為小之各項項數。將此兩類項數之差乘中位數與假定平均數之差數，即得各項對於假定平均數離中差之和與各項對於中位數離中差之和之差額。應用上項差額以校正各項對於假定平均數所得離中差之和。
4. 將此已校正之離中差之和除以總項數，即得根據中位數計算之平均差。

(根據算術平均數計算平均差，手續與此相同)。

(註)在前例計算時，係假定中位數所在組內之各項數值密集於該組之中點者，但若假定該組各項在其組距內分配勻整，而將計算方法依此略予變更，則所得結果之確度當可稍高(關於此種計算法之說明，見 H.L. Rietz 氏所編“Handbook of Mathematical Statistics”第29—31頁)。

標準差

平均差之計算，將正負符號概置不問，在學理上不免牽強。標準差 (standard deviation) 之計算，則無此弊病，且在數學意義上亦較準確。統計學上通常以 σ 作為標準差之符號，即希臘字 sigma 也。

標準差之計算方法，係先將根據算術平均數求得之各項離中差一一自乘後相加，除以項數，再開平方根即得。故標準差者實即各項離中差自乘後之算術平均數之平方根也。標準差亦稱均方根差 (root-mean-square deviation)，為由計算方法而得之名詞。標準差之數值，以從算術平均數計算者為最小，故求各項離中差恆以算術平均數作為中心值。今以簡單之例說明標準差之計算方法於下：

表 三 十 一

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	
12	1	+3	9	$= \sqrt{18}$
15	1	+6	36	$\sigma = 4.24$
	5		90	

統計事項如未組成次數表，則計算標準差之公式為

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

統計事項如已組成次數表，則標準差之計算手續稍覺繁瑣，為求簡捷起見，各項離中差之計算恆以假定平均數為中心值。由已組成次數分配表之統計事項求標準差之普通公式為

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

上式中 f 表示每組次數， d 表示每組中點依算術平均數計算之離中差， N 表示總項數。由上式得

$$\sigma^2 = \frac{\sum fd^2}{N}$$

如以 d' 表示依假定平均數計算之每項離中差， s 表示依假定平均數求得之標準差，則得

$$s^2 = \frac{\sum f(d')^2}{N}$$

以算術平均數為中心值所求得之標準差 (σ)，恆較以尺度上其他任何一點為中心值所求得之標準差為小，故 s^2 恆大於 σ^2 。今設以 c 表示真實平均數與假定平均數之差，可得下列公式(註)：

$$\sigma^2 = s^2 - c^2$$

由是知標準差之數值甚易求得，祇須計算 s^2 及 c^2 即可。表三十二係說明標準差之計算手續。表內所列係 384 個月倫敦對巴黎之匯價。

(註)因	$\sigma^2 = \frac{\sum d^2}{N},$	但	$\sum d = 0,$
	$s^2 = \frac{\sum (d')^2}{N},$	故	$\sum (d')^2 = \sum d^2 + Nc^2,$
	$d' = d + c,$		$\frac{\sum (d')^2}{N} = \frac{\sum d^2}{N} + c^2,$
	$(d')^2 = d^2 + 2cd + c^2,$		$s^2 = \sigma^2 + c^2,$
	$\sum (d')^2 = \sum d^2 + 2c\sum d + Nc^2,$		$\sigma^2 = s^2 - c^2.$

參看 Yule 氏著 "Introduction to the Theory of Statistics" 第 134 頁。

表 三 十 二

1882—1913年倫敦對巴黎匯價標準差之計算

(表內數字根據 The Economist 所載每月月初之匯價)

(1) 組 距 (法郎)	(2) 組中點 (法郎) m	(3) 次數 f	(4) 對於假定 平均數之 離中差 d'	(5) fd'	(6) $f(d')^2$	(7) $(d'+1)^2$	(8) $f(d'+1)^2$
25.07—25.089	25.08	1	-8	- 8	64	49	49
25.09—25.109	25.10	4	-7	- 28	196	36	144
25.11—25.129	25.12	14	-6	- 84	504	25	350
25.13—25.149	25.14	20	-5	-100	500	16	320
25.15—25.169	25.16	45	-4	-180	720	9	405
25.17—25.189	25.18	60	-3	-180	540	4	240
25.19—25.209	25.20	40	-2	- 80	160	1	40
25.21—25.229	25.22	43	-1	- 43	43	0	
25.23—25.249	25.24	42	0			1	42
25.25—25.269	25.26	32	1	32	32	4	128
25.27—25.289	25.28	26	2	52	104	9	234
25.29—25.309	25.30	21	3	63	189	16	336
25.31—25.329	25.32	20	4	80	320	25	500
25.33—25.349	25.34	4	5	20	100	36	144
25.35—25.369	25.36	6	6	36	216	49	294
25.37—25.389	25.38	2	7	14	96	64	128
25.39—25.409	25.40	2	8	16	128	81	162
25.41—25.429	25.42	2	9	18	162	100	200
		384		-372	4,076		3,716

 $N=384$

組距 = .02法郎

$$c(\text{以組距爲單位}) = \frac{-372}{384} = -.969$$

$$c^2(\text{以組距爲單位}) = .9390$$

$$s^2(\text{以組距爲單位}) = \frac{\sum f(d')^2}{N} = \frac{4,076}{384} = 10.6146$$

$$\sigma^2(\text{以組距爲單位}) = s^2 - c^2 = 10.6146 - .9390 = 9.6756$$

$$\sigma(\text{以組距爲單位}) = 3.11$$

$$\sigma(\text{原來單位}) = 3.11 \times .02 = .0622$$

於此所當注意者，上表中逐步計算，皆以組距爲單位，故最後須將求得之結果化爲原來單位。真實平均數與假定平均數之差(c)之計算，係將各項離中差之和除以總項數求得。如欲求算術平均數之值，可先將 c 化爲原來單位，再將此數值依代數方式與假定平均數相加，惟此項手續並不需，因標準差之計算，可不必求真實平均數之值也。計算標準差時，在計算方面有無錯誤，可採用表三十二第(7)(8)兩欄所示之校勘法(查里歐校勘法 the Charlier check)(註)。在前例中求標準差時，各項離中差均係按照假定平均數計算，今若改按假定平均數所在組之下一組之中點計算，則所得各項離中差之值等於 $d' + 1$ ，其平方後之數值當如第(7)欄所示。再將第(7)欄數字乘以各組次數得第(8)欄數字，其和爲3,716。此和數與計算標準差時求得之各項數字有固定之關係，因

$$\begin{aligned} \Sigma f(d' + 1)^2 &= \Sigma f[(d')^2 + 2d' + 1] \\ &= \Sigma f(d')^2 + 2\Sigma fd' + \Sigma f \end{aligned}$$

$$\text{即} \quad \Sigma f(d' + 1)^2 = \Sigma f(d')^2 + 2\Sigma fd' + N$$

上式如以表三十二內所求得之各數值代入，可校核標準差之計算有無錯誤。其式如下：

$$\begin{aligned} 3,716 &= 4,076 + 2(-372) + 384 \\ &= 3,716 \end{aligned}$$

統計資料如已分組成爲次數分配，則標準差之計算手續，可簡述如下：

1. 選定靠近次數分配中心之一組之中點，作爲假定平均數。

(註)參看 C. V. L. Charlier 氏編“Vorlesungen Über Die Grundzüge Der Mathematischen Statistik”第19頁。

2. 就上下各組，以組距為單位，分別計算各組中點對於假定平均數之離中差；次將各組離中差乘以各該組之次數。
3. 將各項離中差與次數相乘之積一一相加而除以 N ，即得以組距為單位之 c ，再求 c^2 。
4. 將各組離中差自乘，再乘以各該組之次數。
5. 以 N 除前項乘積之和，即得以組距為單位之 s^2 。
6. 依照公式 $\sigma^2 = s^2 - c^2$ 求 σ^2 ，再將所得之值開平方，即為以組距為單位之 σ 。
7. 依此求得 σ 之數值，以組距乘之，即得以原來單位表示之 σ 。

標準差之各種特質及其與他種離散度量數之關係，當於下節論述之(註)。

四分位差

在討論平均數之一章內，嘗述及四分位數與十分位數之確定方法。四分位數者，係當次數分配內各數量依次排列於價值尺度上時，將全部項數截分為四等分之分點也；十分位數者，為將全部項數截分為十等分之分點也。故四分位數與十分位數一經確定，次數分配之離散度已不啻明白表示，惟此種平均數雖亦可表示次數分配之離散狀態，但欲為概括之敘述，或供比較之用，則較遜於平均差與標準差。蓋平均差或標準差祇有一個量數，其意義較為扼要，而四分位數與十分位數恆包括數個彼此相關之數值，其意義頗難領悟也。惟用一個單純量數表示離散度之方法，亦可自四分位數蛻化而得，即所謂四分位差(quartile deviation)是也。四分位差計算手續之簡便與意義之易於明瞭，均為他法所不可及。

在尺度上第一四分位數與第三四分位數之間，包含全體項數之半。

(註)標準差有時須加校正(舍巴德氏校正法 Sheppard's correction)，其校正法詳見

第十五章。

凡次數分配愈集中，離散度愈小，則此兩者之距離愈短，故欲根據此兩個四分位數之關係，以測度次數分配之離散度，亦可得相當之準確性。四分位差實即在尺度上第一四分位數與第三四分位數間之距離之半 (semi-interquartile range)。令 $Q.D.$ 表示四分位差， Q_1 表示第一四分位數， Q_3 表示第三四分位數，則以公式表示之四分位差如下：

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

如將尺度上第一四分位數與第三四分位數間之中點以 K 表示之，則次數分配全體項數之半當在 $K \pm Q.D.$ 之距離內。茲仍以前述倫敦對巴黎匯價之數字為例，說明於下：

$$\begin{aligned} Q_3 &= 25.262 \\ Q_1 &= 25.174 \\ Q.D. &= \frac{25.262 - 25.174}{2} \\ &= .044 \\ K &= 25.174 + .044 \\ &= 25.218 \end{aligned}$$

由是知該次數分配全體項數之半當在 $25.218 \pm .044$ 之距離內。此點在表述次數分配之意義上，與匯價平均數值 (算術平均數、中位數或衆數) 同等重要。次數分配如完全對稱，則 K 之數值與中位數之數值合而為一 (即中位數適在尺度上第一四分位數與第三四分位數間之中點)。今匯價之分配不甚對稱，故中位數之值為 25.214 ，而 K 之數值則為 25.218 ，兩者乃微有出入。

機 誤

吾人如將關於天文或其他形體的測量之結果予以考究，則可知對

於同一事物，各人測量所得之數值互異，惟此種不同數值之分配狀態常循一定法則，如繪為曲綫當與差誤常態曲綫相似。於此吾人遂亟需有一種測定差誤離散度之量數，作為測量結果可靠程度之指標。測量結果無論為各人所測量者，或為一人測量多次者，如數值上差異甚大，則此項測量結果即不能認為可靠；反之，如差異不鉅，則測量結果之可靠程度亦必較高。通常用以測定此離散度之量數，謂之機誤(probable error)。機誤為一數值，即當各個觀察值之差誤超過此數值者適占全體項數之半也。各個觀察所得之數值，以接近其算術平均數之機會為最多，故機誤之測度恆以算術平均數為原點。按機誤一名詞之由來，係因每次觀察所得數值與其算術平均數之程度，超過機誤之機率適為 $\frac{1}{2}$ 也，故當觀察所得數值已組成次數分配時，在其算術平均數左右兩邊各取等於機誤之距離，在此距離內應包含全體項數之半。

此種離散度之量數原為測定觀察差誤之用，但亦有供作與機誤原有意義不同之其他用途者，此時機誤可視為機差(probable deviation)，以免誤解。此機差當為算術平均數兩邊所應取之距離，使各項離中差之大於此距離者，適為全體項數之半也。

機誤用作測定離散度之量數，必須在次數分配合乎差誤常態法則時，始有明確之意義，如次數分配呈現偏斜狀態即不可用。在常態次數分配中，四分位差之數值與機誤相等，惟在表述偏斜之次數分配時，則前者之意義反較後者更為直接，此時以採用四分位差為宜。關於用機誤測定統計結果可靠程度之方法，容後申論之。

在常態次數分配中，機誤與標準差兩者之數值恆有一定之關係，故前者之數值可從後者求得。以公式表示之為 $P.E. = 0.6745\sigma$ 。

各種離散度量數之相互關係

茲為便於瞭解前述各種離散度量數之意義起見，再將其相互之關

係作簡略之說明並加以比較於下：

1. 全距者，乃尺度上一定距離，全體項數盡在此距離之內者也。
2. 四分位差者，為尺度上另一距離。就第一四分位數與第三四分位數間之中點，在其左右兩邊各取等於四分位差之距離，則在此距離間當包含全體項數之半。
3. 在完全對稱或不甚偏斜之次數分配中，依照算術平均數計算之平均差，約等於標準差 $\frac{4}{5}$ 。以算術平均數為中心，取平均差 $7\frac{1}{2}$ 倍之距離，約可包含全體項數99%。
4. 在完全對稱或不甚偏斜之次數分配中，若從算術平均數左右兩邊各取相等於標準差之距離，則在此距離內約包含全體項數三分之二（在常態次數分配中，則此距離內當包含全體項數 68.26%）。若從算術平均數左右兩邊各取兩倍於標準差之距離，則此距離內約包含全體項數 95%（在常態次數分配中，當包含全體項數95.46%）。若從算術平均數左右兩邊各取三倍於標準差之距離，則此距離內約包含全體項數 99%（在常態次數分配中，當包含全體項數 99.73%）。故吾人求標準差時，可以算術平均數為中心，試取六倍於標準差之距離，以觀此距離內是否約包含全體項數99%，則對於標準差之計算是否有誤，即可知其大略矣。試參看圖四十三，則對於常態次數分配之標準差之意義，當更明瞭。
5. 在常態次數分配中，機誤等於 0.6745σ 。以算術平均數為中心，在兩倍於機誤之距離內應包含全體項數50%。以算術平均數為中心，在八倍於機誤之距離內，應包含全體項數99%。

各種離散度量數之特質

全距

1. 全距之計算，手續簡單，而其意義亦易瞭解。如吾人僅欲測定離散度之大概，可採用之。
2. 全距之數值係依兩極端數值之大小而定，故極不穩定，常因一項數量之去留而受重大影響。
3. 全距對於兩極端數值間之次數分配狀態，全無提示。

四分位差

1. 四分位差之計算，手續簡單，意義亦易於瞭解，為測定離散度之一種粗略量數。就準確程度而言，四分位差較勝於全距。
2. 四分位差並不依照任何平均數計算各項數量之離中差。
3. 四分位差之數值不受第一四分位數與第三四分位數間各項數量分配狀態之影響，亦不受該距離以外各項數量分配狀態之影響。如兩個次數分配絕不相似，而四分位數偶然相同，則其四分位差亦必相等。四分位差既不受各項數量離中差之影響，故不能視為測定離散度之精確量數。
4. 四分位差不能用代數方法處理。

平均差

1. 平均差受每項數量之影響，且因係從各項單獨數量對其中位數（或算術平均數）之離中差平均而得，故其意義頗為確切。
2. 平均差所受極端離中差之影響，比標準差為小。
3. 依數學理論而言，平均差不及標準差之合理與合用。

標準差

1. 標準差受每項數量之影響。
2. 標準差之計算，將各項數量之離中差一一自乘，以消除其正負

號，而後相加，在代數學上絕無抵觸。

3. 標準差在數學上具有確切之意義，用代數方法處理絕對適當。
4. 標準差所受取樣差誤之影響，大致較測量離散度之其他量數為小。
5. 分析差誤常態曲綫，常以標準差為單位，頗足增廣標準差之用途。

機誤

1. 機誤之用於常態次數分配，具有確切之意義，而在偏斜之次數分配則否，故不可用以表述後一種次數分配。
2. 在適用機誤之次數分配，機誤極有用途，其最要者即可用作測量數量可靠程度之指標。
3. 在常態次數分配中，機誤與標準差間有一定之關係，故機誤之數值不難確定。

前述測量離散度之各種量數，各有其特殊用途，惟大體言之，以標準差為最優，而欲使離散度之測量結果十分準確，尤不能不採用標準差。至於機誤實際上祇係標準差數值之分數，惟其用途範圍較為狹隘。

相對離散度之測度

在前節內已將絕對離散度之測度方法加以討論。用上述各種方法所求得離散度之數值，皆係用原來單位表示者。例如倫敦對巴黎匯價之標準差係以法郎為單位，生鐵產量之標準差係以噸為單位等是。如吾人之目的僅為表述一單純之次數分配，則測量離散度自以採用原來單位較為相宜，但如欲比較兩個不同之次數分配，即不免發生困難。在兩個次數分配之原來單位彼此互異時，困難之發生固屬顯見，即原來單位彼此相同，此種困難仍所難免。例如測量犬馬體重之離散度，雖同以磅為單位，但當馬體重量之標準差大於犬體重量之標準差時，吾人不能遽謂前者之離散度必大於後者。測量離散度所得之絕對數值，必須與計算各

項離中差所根據之平均數比較，始有意義。如與平均數脫離，意義即不明瞭，故在比較時，離散度之數值須化成比率之形式。最簡便之手續係將該數值除以計算各項離中差所根據之平均數，而得該值與平均數之百分率。此值遂成抽象數字，而為測量次數分配之相對離散度之量數，乃可用以與其他次數分配之相對離散度之量數彼此比較矣。

離散度係數

在測定相對離散度之各種量數中，應用最為普遍者，當推披爾遜氏 (Pearson) 所創之離散度係數 (coefficient of variation)。此離散度係數係以 V 表示，實即以算術平均數除標準差所得之百分率也。其公式如下：

$$V = \frac{\sigma}{M} \times 100$$

以前述倫敦對巴黎匯價一例中求得之結果代入上式，得

$$\begin{aligned} V &= \frac{.0622}{25.2206} \times 100 \\ &= .25\% \end{aligned}$$

同時期倫敦對紐約匯價之離散度可從表三十之資料求得。其離散度係數為 .33%，由是知倫敦對紐約匯價之變動，顯較同時期倫敦對巴黎匯價為劇烈。

至於其他測量離散度之絕對量數，亦可依照計算此係數之方法，以其離中差所根據之平均數除之，以求離散度之指數，惟披爾遜氏之離散度係數，應用最為普遍耳。

偏斜度之量數

關於表述次數分配之集中趨勢與測定次數分配中心值兩邊之離散度之方法，已如上述。吾人尚需要一種量數以測定次數分配之偏斜度

(skewness or asymmetry), 蓋對於次數分配中心值兩邊各數量之分配之為對稱, 抑欠勻整而呈偏態, 亦理應研究之。此項量數既已求得, 則吾人可有三種簡單之量數——平均數、離散度量數及偏斜度量數——以簡述次數分配之特質。測定偏斜度之量數可有數種, 茲分述於下。

次數分配如係完全對稱, 則其算術平均數、中位數及衆數均在一點。當次數分配漸趨偏斜, 此三者之數值亦漸分離, 而以算術平均數與衆數之差異為最大, 此項差數即可用作測度偏斜程度之量數。此測定偏斜度之量數亦可仿照求相對離散度之方法, 化為抽象形式之係數, 俾可與另一次數分配之偏斜度係數互為比較。披爾遜氏主張以次數分配之標準差除算術平均數與衆數之絕對差數作為偏斜度量數。其公式如下:

$$sk(\text{skewness}) = \frac{M - Mo}{\sigma}$$

在對稱之次數分配中, 算術平均數與衆數合一, 故此時偏斜度量數之數值等於0。偏斜度量數之數值或為正數, 或為負數, 胥視尺度上此兩種平均數位置之關係而定。

如次數分配不甚偏斜, 則用下式測定偏斜度尤為便利。

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

用此式求得之結果, 與用前述第一式所求得者相近似, 此因次數分配不甚偏斜時, 中位數之位置介乎算術平均數與衆數之間, 而中位數與算術平均數之距離, 約等於算術平均數與衆數距離之三分之一也。

衆數之值不易確定, 故披爾遜氏公式常不適用, 而須採用計算較易之其他方法。鮑萊氏 (Bowley) 主張用第一四分位數、第三四分位數與中位數三者關係以為計算之方法。蓋如次數分配對稱, 此兩個四分位數與中位數之距離應各相等; 如次數分配不對稱, 則兩個距離不相等。今以 q_3 表示第三四分位數與中位數之差額, q_1 表示中位數與第一四分位數之差額, 吾人即可用下式測定偏斜度。

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

依上式求得之數值，當在0與±1之間，因次數分配如完全對稱，則 $q_2 = q_1$ ，即偏斜度量數之數值為0；如次數分配極不對稱，以致中位數與第一四分位數或與第三四分位數合而為一，則 q_2 或 q_1 等於0，偏斜度量數之數值當為+1或-1。鮑萊氏嘗解釋謂計算結果等於.1時，即次數分配略帶偏斜之表示，等於.3時，即次數分配呈現顯著偏斜之表示。

用此法求得偏斜度量數之數值不能與用披爾遜氏公式所求得者相比較。

峯 度

關於表述次數曲綫第四種特質之方法，前嘗提及，此乃以差誤常態曲綫為標準，測定某一曲綫峻峭程度之方法。在統計學上此種次數分配之特質謂之峯度 (kurtosis)。其測度方法當在第十五章內論述之。

參 考 書

- Bowley, A. L. "Elements of Statistics" (110—117).
 Chaddock, R. E. "Principles and Methods of Statistics" (Chap. IX).
 Davies, G. R. "Introduction to Economic Statistics" (29—46).
 Jones, D. C. "A First Course in Statistics" (42—51).
 Kelley, Truman L. "Statistical Method" (70—82).
 King, W. I. "Elements of Statistical Method" (141—158).
 Pearl, Raymond. "Medical Biometry and Statistics" (272—278).
 Rietz, H. L. (editor). "Handbook of Mathematical Statistics" (27—33).
 Rugg, H. O. "Statistical Methods Applied to Education" (149—179).
 Secrist, Horace. "An Introduction to Statistical Methods" (377—424).
 West, Carl J. "Introduction to Mathematical Statistics" (45—58).
 Yule, G. U. "Introduction to the Theory of Statistics" (133—156).

第六章 物價指數

指數之性質

指數 (index number) 之一名詞常用以指統計分析上數種相類之方法而言。指數之用於物價變動之研究者為最廣，惟其他各種用途亦應加討論，使讀者得明瞭其重要之性質。時間數列所化成之百分比，亦常被稱為指數，為指數中形式最簡單之一種。下表所列美國棉花消費量之指數，即為其一例。

表 三 十 三

1913—1923年美國國產棉花之消費量

(1913年消費量 = 100)

年份	棉花消費量 (單位：袋)	棉花消費量 百分比
1913	5,583,468	100.0
1914	5,448,760	97.6
1915	6,008,984	107.5
1916	6,620,415	118.6
1917	6,815,811	122.0
1918	6,176,547	110.5
1919	5,919,520	106.0
1920	5,843,200	104.5
1921	5,406,721	96.8
1922	6,087,520	109.0
1923	6,513,696	116.7

物品價格亦可採用同法以百分比表示，而以某日或某時期之價格作為基數。

表 三 十 四

1913年及1919—1923年美國米乃波立斯頭號北方春麥之平均市價

(1913年平均價 = 100)

年份	每蒲式耳平均價	比價
1913	\$0.8735	100
1919	2.5660	294
1920	2.5581	293
1921	1.4660	168
1922	1.3450	154
1923	1.1810	136

時間數列中各項，如以一固定基數為 100，以百分比表示之，不但可使各期之數值便於比較，而對於該數列之變動趨勢，亦較原有形式更易明瞭。在比較數個數列之趨勢時，用百分比之形式亦較為便利。

指數之名詞雖可用於此種百分比，但吾人應保留此名詞，以用於代表由數個數列綜合所得之數字較為適當。所綜合之各個數列，或關於物價，或關於物品之生產量，消費量，工資，貿易額，以及隨時間變化之任何事物。編製此項特種指數常發生極複雜之問題，惟其最後目的為欲求一單純與簡要之數字，以表示各個數列所受各種勢力之淨結果，則各指數皆同。

試根據美國煤與石油之產量，編製一種簡單指數，以表示產量之增減變化。在編製此種指數時，須將烟煤、白煤與石油之產量綜合計算。表三十五係一九一〇年至一九二三年烟煤、白煤、石油三個數列之產量及其百分比。

吾人設欲根據此三個數列編製一種粗略之燃料生產指數，則因物品數量單位之不同，不能逕將原有之產量數字相加，但若採用百分比，則可免除此種困難。將每年之三個百分比相加，而求其簡單平均數，即得所求之指數。依此求得之各年指數見表三十六：

表 三 十 五

1910—1923年美國烟煤白煤及石油產量

(1910年產量=100)

年 份	烟煤產量 百萬公噸	百分比	白煤產量 百萬公噸	百分比	石 油 產 量 百萬蒲式耳	百分比
1910	417.1	100	84.49	100	209.6	110
1911	405.9	97	90.46	107	220.4	105
1912	450.1	108	84.36	100	222.9	106
1913	478.4	115	91.52	108	248.4	118
1914	422.7	101	90.82	107	265.8	127
1915	442.6	106	89.00	105	281.1	134
1916	502.5	120	87.58	103	300.8	143
1917	551.8	132	99.61	118	335.3	160
1918	579.4	139	98.83	117	355.9	170
1919	465.9	112	88.09	104	378.4	181
1920	568.7	136	89.60	106	442.9	211
1921	415.9	100	90.44	107	472.2	225
1922	404.5	97	52.90	63	557.5	266
1923	545.3	131	95.20	113	725.7	346

表 三 十 六

1910—1923年美國煤與石油之生產指數

(1910年產量=100)

年份	指數	年份	指數
1910.....	100	1917.....	137
1911.....	103	1918.....	142
1912.....	105	1919.....	132
1913.....	114	1920.....	151
1914.....	112	1921.....	144
1915.....	115	1922.....	142
1916.....	122	1923.....	196

計算前項指數係將每年三個百分比相加，以三除之而得，是各個百分比所用之權數相等。此種根據權數相等之百分比所求得之指數，謂之未加權指數(unweighted index)，但此名詞易滋誤解，蓋實際權數固已採用，惟在本例中各個百分比之權數相等而已。此種指數因係根據相等之權數編製，故其指數不能準確反映三個數列之綜合趨勢，蓋若各個數列予以相等之權數，即謂各個數列之重要程度相等，此點實與事實相

左。下列數字係一九二一年烟煤、白煤、石油批發市場之成交價值，以示三種數列之比重程度(註)。

礦 產	1921年批發市場之成交價值
烟煤	\$1,948,000,000
白煤	731,000,000
石油	712,000,000

上列數字表示三種物品在批發市場之重要程度約為 3,1,1 之比。吾人可將此項比重資料作為權數，加於各該數列，以計算每年指數。茲將根據前述三個數列計算一九一〇年及一九二三年指數之手續列舉於下，以說明其計算方法。

表 三 十 七
煤與石油生產量加權指數之計算

礦 產	1910年生產量之百分比	權 數	權數×百分比	1923年生產量之百分比	權 數	權數×百分比
烟 煤	100	3	300	131	3	393
白 煤	100	1	100	113	1	113
石 油	100	1	100	346	1	346
		5	500		5	852

$$1910\text{年燃料生產指數} = 500 \div 5 = 100$$

$$1923\text{年燃料生產指數} = 852 \div 5 = 170$$

按照上法求得十四年間之每年指數如下：

表 三 十 八
1910—1923年美國煤與石油生產量之加權指數

年份	指數	年份	指數
1910.....	100	1917.....	135
1911.....	101	1918.....	141
1912.....	106	1919.....	124
1913.....	114	1920.....	145
1914.....	107	1921.....	126
1915.....	111	1922.....	124
1916.....	121	1923.....	170

(註) 此項數字係美國勞工統計局所編製。

表三十六及表三十八所列兩種指數，試加比較，則有顯著之差異。後一種指數因其所用權數較為合理，故所得結果自較準確，尤足以代表三個數列所受各種勢力之綜合結果。

此外尚有一種指數，其算法係將每期各項數值一一相加，得其綜合數值，作為全體數值之代表，而不求其平均數者。惟欲計算此種指數，各數值之單位必先一致而後可。此種指數常用以測度物價水準之變動，係將某時期若干物品之綜合價格，與另一時期各該物品之綜合價格比較而得。下列數字即為其一例。

表 三 十 九

1913—1923年美國勃蘭特街(Bradstreet)雜誌社之批發物價指數

年份	指數	年份	指數
1913.....	\$9.21	1919.....	\$18.66
1914.....	8.90	1920.....	18.81
1915.....	9.85	1921.....	11.37
1916.....	11.82	1922.....	12.12
1917.....	15.64	1923.....	13.40
1918.....	18.71		

上列每年綜合價格係各該年九十六種物品之平均批發市價之總和，惟在價格相加之先，各項價格均曾折合為每磅價格，俾各項價格間有一比較之標準。此種指數亦可化為百分比，以任何一年作為基期，將其他各年之綜合價格折算為基年(base year)綜合價格之百分率。

由前例可見指數之形式繁簡不一，自簡單之百分比以至各項百分比之平均數，絕對數或百分比之綜合數，概可稱為指數。顧指數之形式雖異，但其目的為測量某時期內之數值變動，或表示時間數列內各項數值之變動則一也。除前述各例外，指數尚有更廣泛之意義。例如吾人欲測驗售貨員之能力，可就決定此能力之各因素一一考查，評定其績分，再求其平均數，即得售貨員能力之指數。又如欲測驗企業組織下各部分之效率，亦可編製一種指數。惟是任何指數之編製，應將各個不同之因素化為可供比較之數值，而後根據此各個數值求一單純數字，以代表全

體之數值。故指數可用以比較一時期內某項事物之變動，或同時期內各項事物之變化。可知指數之各種形式(除第一例含義較狹僅指簡單百分比之數字以外)，實即統計上之一種平均數，故指數之計算及其用途，遂不免受平均數各項原則及缺點之拘束。

本章所欲論述者，祇為應用於時間數列之指數編製方法。指數之應用於各種時間數列時所涉及之原則及慣例並不一致，故須將各種數列分別討論。請先討論批發物價指數。

物價之變動

吾人如將物價之變動詳加研討，則錯綜紛紜，漫無準則，幾無顯著之趨勢可尋。試任取物價數項列表於下，以示物價變動凌亂紛歧之一斑。

表 四 十
批 發 物 價 (註)

物 品	單 位	批發價格 1923年10月	批發價格 1923年11月
磚,普通建築用,各產地之平均價	1000	\$14.752	\$14.746
生鐵,原料	噸(毛重)	23.500	20.875
水泥, Portland, 各廠平均價	桶	1.893	1.842
胡麻子油,未製者,紐約	加侖	.943	.910
銅條,匹茲堡,培塞麥	噸(毛重)	40.000	40.000
錫版,匹茲堡	百磅	5.500	5.500
電化紫銅,提錄廠出品	磅	.126	.128
鉛,紐約	磅	.069	.069
鋅,紐約	磅	.067	.067
白煤,紐約,徹斯納特	噸(毛重)	11.471	11.478
烟煤,支加哥	噸(淨重)	4.600	4.525
生石油,油井售價,本薛文尼	桶	2.500	2.388
汽發油,馬達用,紐約	加侖	.185	.170
棉花,米特令,新奧雷安斯	磅	.292	.349
小麥,二號赤色冬麥,支加哥	蒲式耳	1.097	1.061
糖,顆粒形,紐約	磅	.090	.087

(註)表內物價係美國勞工統計局所編。

表內十六種物品中，其一九二三年十一月與十月價格之比較，無變動者四種，漲者三種，跌者九種。其中某數種物品價格漲跌無多，而某數種物品之價格變動則甚劇烈。此一小部分物價之變動，已足象徵市場上全部物品價格變動之情況。各種物品價格之變動趨勢不一，有上漲者，有下跌者，亦有絕無變動者。在成千累萬種物品之中，每種物品各因在國內市場或國外市場所受影響不同，其價格變動因而各異，惟各種物品之價格，互有關聯，一物價格之漲跌，足以影響他物之價格，而同時亦受他物價格變動之影響。凡此皆係指每種物品本身所受之勢力而言，此外尚有範圍較廣之勢力，足以影響整個物價制度下全體物品之價格者。經濟統計學者之職責，即在從各時期紛亂之物價變動中求其規律，從許多繁瑣之變動中窺測經濟變動之大勢。

吾人所欲研究之物價變動，其所受勢力種類繁多，惟就大體言之，可歸納為兩大類。其一為關於各個物品本身之生產消費狀況者。各個物品生產消費狀況之變更，常可直接影響於其價格。例如新市場之發現，產地之擴展，生產技術之改進，風尚之變更，需求之轉移，以及季節發生之變化等，皆足使物價隨時變動。平時此種勢力最為明顯，商人與消費者均能見及。此種勢力亦能影響物價之水準，但不能使整個物價制度呈一致上漲或下跌之現象。

至於可使物價普遍漲落之勢力，則屬於另一類，其範圍較為廣泛。例如一般生產技術之進步，足以增進勞工之生產能力，其結果直接增加物品之供給數量，間接影響於其價格。又如貨幣制度之變更，黃金供給數量之增減，皆足以影響貨幣之流通數量，而可使物價直接受其影響。依此類推，則凡銀行與信用制度之改革以及商業習慣之變更，足以影響信用票據之使用範圍及貨幣與信用之流通速率者，對於物價亦有同樣之作用。凡此種種勢力，皆可影響一般物價，惟其所及之範圍，不若直接影響個別物品之各種勢力之確切耳。

普通批發物價指數之目的

前節所述各種勢力恆同時存在，而發生聯合作用，物價種種變動即爲此聯合作用之淨結果，故吾人殊無法將各種勢力一一隔絕，而分別估量其輕重。研究此種種變動，可從各種不同之觀點着手。吾人可研究物價制度內各部分所發生之變動，以決定其變動之性質及程度。此種研究對於物價變動之狀況及各個物價之相互關係，可獲明確之認識，惟吾人當前之問題則爲決定物價所受一切勢力之淨結果。各個物價之變動是否互相抵銷，使上漲之物價與下跌之物價適相平衡，以致物價之水準不生變動乎？抑在一時期內各個物價之漲勢與跌勢緩急不同，致使物價之水準爲之增高或降落乎？此種趨勢是否存在？如其存在，究爲何種趨勢？又將如何測度之乎？前所論述之各種統計方法可用以解答本問題乎？

此項研究之第一步，應先解答上述最後之一問題。前嘗論述簡縮數字資料之方法，但此種方法僅能在適合某種情形時始可應用。就平均數而言，統計資料必須性質一致，而呈現顯著之集中趨勢時，平均數始有意義。平均數之選擇又必須視統計資料之分配狀態而定。如對於統計原始資料之分配，事前絕未研究，則所選取之平均數或他種統計量數必欠恰當。吾人欲解答此問題，應先決定物價資料之性質，並研究該項資料整理後次數分配之形態。

設吾人研究物價變動之目的，爲普通決定兩個時期內批發物價水準之變動。決定物價水準之變動，實即測度兩個時期內貨幣價值之變動也。研究之資料爲兩時期內若干項物品之價格，每品之兩項價格可測度其在此兩時期內因受外界種種勢力而發生之變動。如將許多物價彙集一處，則此項集合資料可代表各種勢力之相互作用。此各種勢力中，有特殊者，有普通者；特殊勢力僅影響一二物品之價格，普通之勢力則可影響多種物品或全體物品之價格。吾人所欲決定者爲此各種勢力所引

起物價變動之淨結果，故必須用一種量數以測度足使各個物價漲落之一切勢力之混合結果。此量數即為批發物價指數。

此時吾人所用資料之單位，為每一個物價之變動。前所論述之統計方法，能否應用於此項資料之整理與分析，須視各項單位之動態而定。下述一例係舉示此項資料分組後之次數分配。

表 四 十 一

1914年物品346種之比價分配(註)

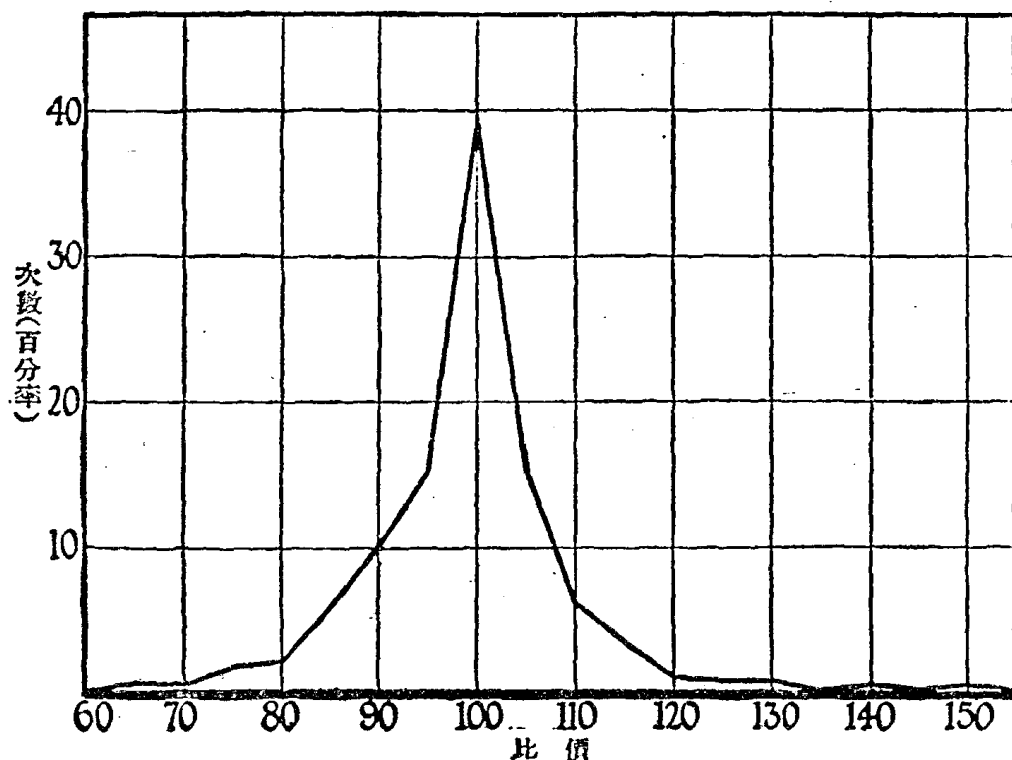
(1913年平均價 = 100)

比 價	中 點 <i>m</i>	項 數 <i>f</i>	占總項數 之百分率
62.5— 67.4	65	1	.3
67.5— 72.4	70	1	.3
72.5— 77.4	75	5	1.5
77.5— 82.4	80	7	2.0
82.5— 87.4	85	20	5.6
87.5— 92.4	90	35	10.0
92.5— 97.4	95	51	14.5
97.5—102.4	100	134	39.0
102.5—107.4	105	50	14.5
107.5—112.4	110	21	6.0
112.5—117.4	115	12	3.5
117.5—122.4	120	3	1.0
122.5—127.4	125	2	.6
127.5—132.4	130	2	.6
132.5—137.4	135		
137.5—142.4	140	1	.3
142.5—147.4	145		
147.5—152.4	150	1	.3
		346	100.0

(註)表內所包含之346種物品，係美國勞工統計局編製批發物價指數所用者。其原來價格及各品之百分比見該局叢書第269種“Wholesale Prices, 1890—1919”。

物價比率之次數分配

每項物價之變動可用一比率表示，即某一時期一物價格與另一時期同物價格之比率也。此比率可依本章前此所舉指數之各例，化為百分比形式，藉便比較。今以表四十內各對價格中之一對為例以說明之。如生鐵一九二三年十一月份價格與一九二三年十月份價格之比為 $\$20.875 : \23.500 ，化為百分比為 $88.8 : 100$ 。上列次數表(表四十一)係根據一九一四年三百四十六種物品之價格，各以一九一三年各該物品之價格作為基數計算而得之百分比編製而成。此三百四十六項百分比數字之分配如表四十一。



圖四十七. 次數多邊圖:1914年物品346種之比價分配(1913年平均價=100)

表示此項分配之次數多邊圖如圖四十七。圖中次數係按照各組項數占總項數之百分率繪製，俾可與其他性質相似之次數分配比較。此次數分配與前此所示標準形式頗有相似之處：其一，為顯著之集中趨勢。本例中物價變動頗有穩定之趨勢，各項物價比較其基年價格漲落在 2.5%

以內者，達總項數之 39%。其二，在此分配中大於衆數各項之全距離比較小於衆數各項之全距爲大，而此點又頗關緊要，但中心值兩旁之分配，大體呈對稱狀態。吾人暫不研究此次數分配應採用何種平均數以表示其集中趨勢，然平均數之可適用於此種次數分配，則屬顯見也。

上述之例係比較前一年物價與後一年物價之變動。在此兩時期間物價之水準幾無變動。密哲爾氏 (W. C. Mitchell) 嘗舉示更爲詳備之一例，係將一八九〇年至一九一三年間 5578 種物品價格之逐年變動，採用與上例相同之分組法，製爲次數分配表。表中大於衆數各項之全距所超過小於衆數各項之全距之情勢似更顯著，其原因由於在此二十三年中，各年物價上漲者居多所致。該次數分配如圖四十一所示。

在研究物價逐年變動時，物價之情性最爲明顯。吾人應再研究較長時期內物價變動之狀況，藉以明瞭是否具有與逐年變動相同之分配狀態。試舉兩例研究之：一爲十年間之物價變動，其期末之物價水準與期首約略相等者，一爲五年間之物價變動，其逐年物價之增漲頗速者。表四十二係一九〇〇年物品 222 項之價格，以一八九〇年爲基期之變動百分率。在此時期內，一八九〇年至一八九六年間物價之水準下跌，而一八九六年至一九〇〇年間物價之水準上漲，跌後回漲，故一九〇〇年之物價水準較一八九〇年僅低 1%。

以此項數字資料繪爲次數多邊圖如圖四十八。圖中次數係按照各組項數占總項數之百分率繪製（圖四十七與圖四十八之縱尺度大小不同，比較時宜注意）。

圖四十七與圖四十八所表示之兩種次數分配迥不相同。圖四十八中之全距顯較圖四十七中之全距爲大，此由於包含時期較長，比價分散之所致。再則圖四十八中之集中趨勢雖仍明顯，但密集於衆數組之次數之百分率遠較前一種爲低。兩種分配繪於算術尺度上，皆呈對稱狀態（吾人應注意兩種分配中用算術平均數所測度前後兩時期之物價水準

幾全相等),惟前一種分配之集中趨勢遠較後一種為明顯,而各個比價對於其算術平均數之離中差亦較小。此種分配頗與極準確之物體測量或由極準確之礮發出彈丸所得之分配相似。後一種曲綫則與不甚準確之物體測量或由陳舊不準確之礮發出彈丸所得之分配相似,其衆數值發現之次數不多,而各個比價之離中差亦較大。由是知計算物價變動包

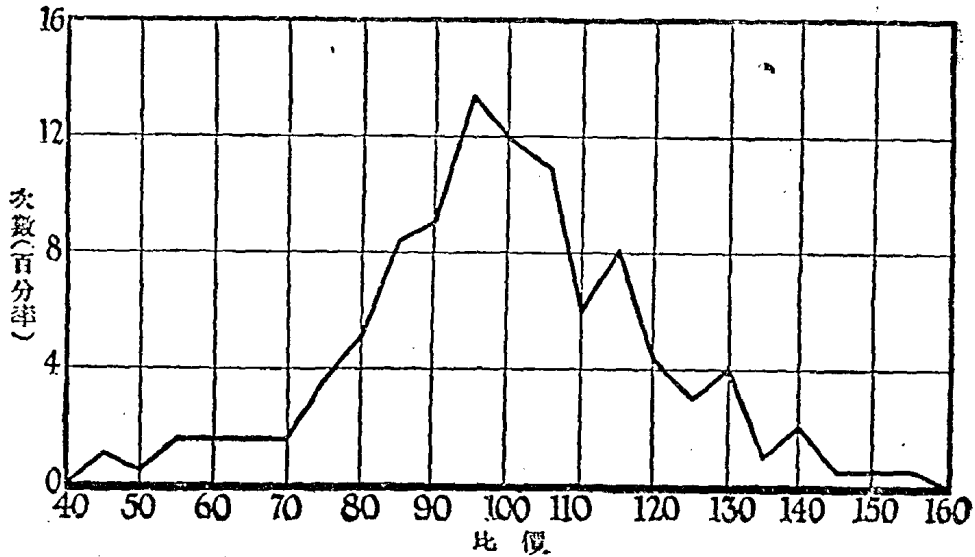
表 四 十 二

1900年物品222種之比價分配

(1890年平均價=100)

比 價	中 點 <i>m</i>	項 數 <i>f</i>	占總項數 之百分率
42.5— 47.4	45	2	1.0
47.5— 52.4	50	1	.5
52.5— 57.4	55	3	1.5
57.5— 62.4	60	3	1.5
62.5— 67.4	65	3	1.5
67.5— 72.4	70	3	1.5
72.5— 77.4	75	8	3.5
77.5— 82.4	80	12	5.0
82.5— 87.4	85	19	8.5
87.5— 92.4	90	20	9.0
92.5— 97.4	95	30	13.5
97.5—102.4	100	27	12.0
102.5—107.4	105	24	11.0
107.5—112.4	110	14	6.0
112.5—117.4	115	18	8.0
117.5—122.4	120	10	4.5
122.5—127.4	125	7	3.0
127.5—132.4	130	9	4.0
132.5—137.4	135	2	1.0
137.5—142.4	140	4	2.0
142.5—147.4	145	1	.5
147.5—152.4	150	1	.5
152.5—157.4	155	1	.2
		222	100.0

含時期愈長，則後一種分配所呈現之散漫趨勢愈顯著，其最高縱坐標之數值漸減小，分配之全距漸增大，而所得曲綫之形態愈趨平坦，曲綫之展佈亦愈廣泛，以致任何平均數之代表性爲之減弱。蓋若次數分配不集中，則依此求得之任何一種平均數僅爲一抽象數字而無切實意義也。



圖四十八. 次數多邊圖：1900年物品222種之比價分配(1890年平均價=100)

吾人於此可得一實驗之結論，即物價變動可用統計方法測度，又物價所包含之時期如不過長，其變動可用平均數表示。至於吾人應以若干年認爲最長時期，在此時期內之物價變動方可測度，事實上殊不易決定。物價指數如欲求其準確而具有意義，則須根據短時期內之物價編製，尤以根據物價逐年變動之指數最爲準確。如物價指數之編製目的僅在表示物價之大體趨勢，則包含時期自須稍長，惟編製與引用此種指數者須明瞭其缺點之所在耳(註)。

以上所舉兩例皆係選擇前後兩時期之物價用算術平均指數所測度之水準大致相等者，此係特殊之例，故吾人尚須研究物價水準變動較烈之兩個時期內各項價格變動之分配。下表所列係一九一八年 1437 種物

(註)此項結論大半根據 W.C.Mitchell 氏研究之結果。參看美國勞工統計局叢書(物價叢書)第284種 "The Making and Using of Index Numbers".

品比價之分配，以一九一三年七月至一九一四年六月之平均物價作為基數者(註)。

表 四 十 三

1918年物品1437種之比價分配

(1913年七月至1914年六月之平均價=100)

比 價	中點 <i>m</i>	項數 <i>f</i>	占總項數 之百分率
36	36	1	*
49	49	1	*
50—69	60	4	.3
70—89	80	17	1.2
90—109	100	61	4.3
110—129	120	64	4.5
130—149	140	130	9.0
150—169	160	212	14.7
170—189	180	219	15.2
190—209	200	164	11.4
210—229	220	135	9.4
230—249	240	104	7.2
250—269	260	76	5.3
270—289	280	64	3.8
290—309	300	42	3.0
310—329	320	30	2.1
330—349	340	31	2.1
350—369	360	16	1.1
370—389	380	13	.9
390—409	400	7	.5
410—429	420	7	.5
430—449	440	8	.6
450—469	460	4	.3
470—489	480	4	.3
490—509	500	4	.3
510—529	520	5	.4

(註)此項物品比價之分配係美國戰時工業局物價組(Price Section of the War Industries Board)所編，轉載於美國勞工統計局叢書第284種卷一第70頁。

表 四 十 三 (續)

比 價	中 點 '''	項 數	占總項數 之百分率
530—549	540	3	.3
550—569	560	4	.3
587	587	1	*
627	627	1	*
727	727	1	*
730	730	1	*
743	740	1	*
761	761	1	*
784	784	1	*
826	826	1	*
848	848	1	*
900	900	1	*
1165	1165	1	*
1356	1356	1	*
1585	1585	1	*
1764	1764	1	*
2049	2049	1	*
2863	2863	1	*
3009	3009	1	*

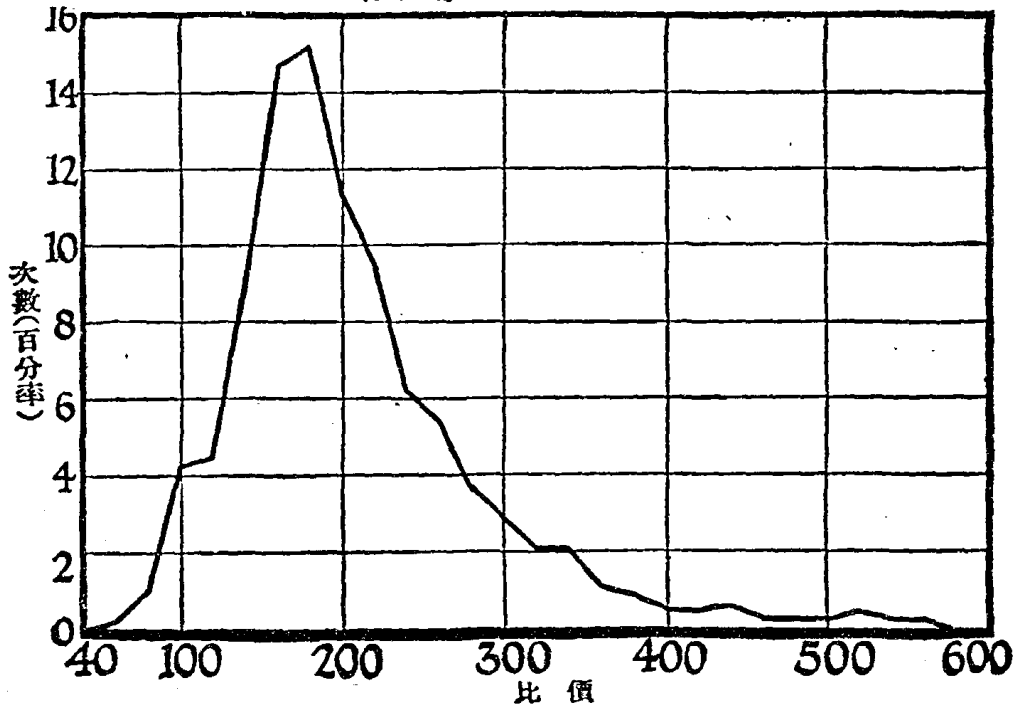
1437

*百分率在.1%以下。

上項次數分配見圖四十九(本圖所定尺度之大小與前兩圖不同)。

該次數分配細加研究,可獲與前述兩例相同之結論,即該分配之集中趨勢甚為明顯,故可用平均數以代表分配之全體。且因其衆數組為包含中點180之一組,故知項數密集之原因,並不由於物價之惰性,而由於足以影響整個物價制度之各種勢力之存在。惟此分配與前述兩種次數分配猶有顯著之差別,即分配之偏斜傾向,在前述第一種分配中已略呈端倪,而在此分配中則更較明顯。圖中繪於算術尺度上之曲綫,頗呈不對稱之狀態,項數最密集之點係迫近於尺度之下端,而曲綫向右延長,漸見尖細,越出本圖尺度範圍以外而成長尾狀。此種情形可參看上表而益明瞭。例如以100作為基數,其最小之比價為36,在尺度上小於基數64點;其最大之比價為3009,大於基數2909點。在一般物價水準劇漲之時,

此項比價分配偏斜之狀態亦所常有（按戰時工業局指數一九一八年之物價比較基期物價高 94%）。此種情形雖可解釋衆數組位置偏低之原因，但未說明此分配之性質。在說明此次數曲綫之形態時，吾人應解答關於編製指數之基本問題，即吾人應採用何種平均數，求一單純數字以表示物價變動百分率之集中趨勢乎？



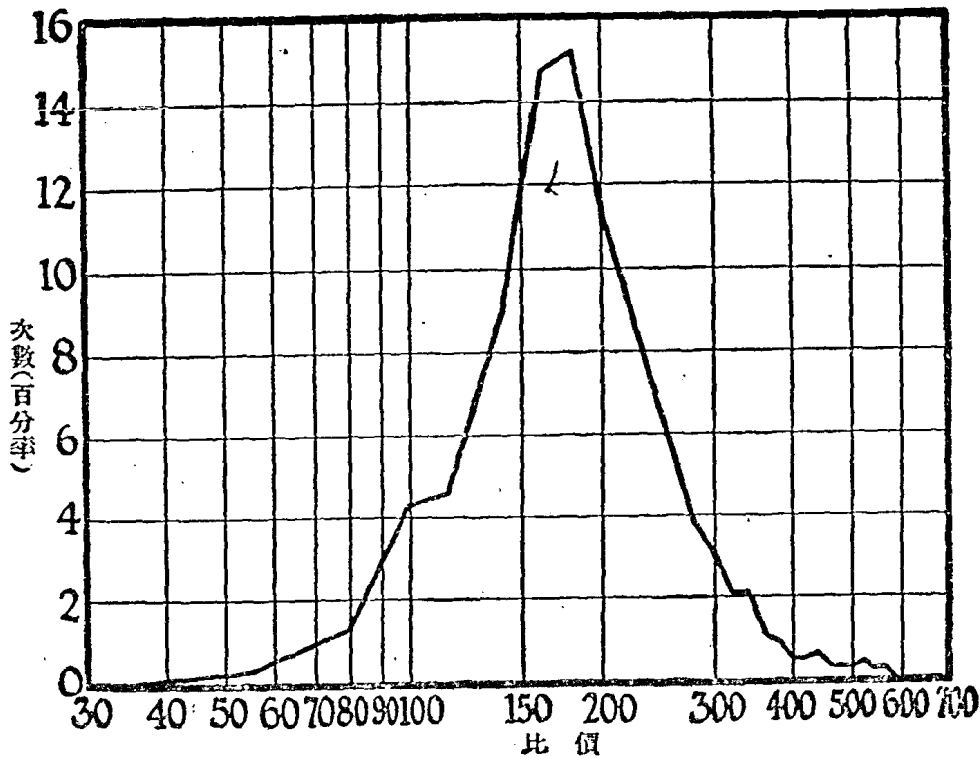
圖四十九. 次數多邊圖：1918年物品1437種之比價分配
(1913年七月至1914年六月之平均價 = 100)

物價變動率之平均問題

以百分率表示之物價增高率，其高度並無限制。物價增高100%，500%，1000%，或1000%以上，事實上均有可能。美國戰時工業局研究戰時物價嘗發現最大之增高率為 acetiphenetidin 一物之4981%。但物價跌落百分率最多不過100%，此時物價格蓋已跌落至0矣。圖中曲綫之呈現偏斜狀態者，即由於此。如比價項數極多，則製表後繪於算術尺度上之次數曲綫，即呈現此種特性，而當物價上漲時，此種特性尤為明顯。

次數分配之呈現偏斜狀態與平均數之選擇有關，此點在討論平均

數時已有論及。如所欲平均之各項數值為比率而非絕對數，以採用幾何平均數為宜。試根據前述之次數分配計算各種平均數，互為比較，可證此說之不謬。表中比價1437項之算術平均數為217，眾數為180，故算術平均數與眾數之距離相差頗遠。任何平均數如其與眾數相差之距離過遠，則此平均數顯已失去其代表性。根據同一資料求得之幾何平均數為194，此數值與眾數及中位數（191）均極接近。其他姑勿置論，即以此各種平均數之關係而言，此種分配自以採用幾何平均數而捨棄算術平均數為宜也。



圖五十. 次數多邊圖：繪於對數尺度上之1918年物品1437種之比價分配
(1913年七月至1914年六月之平均價 = 100)

吾人在論述平均數時，嘗論及凡次數分配繪於對數尺度上之曲綫比較繪於算術尺度上之曲綫更與常態曲綫相近時，即表示該分配之變動比較合乎幾何法則，而不合於算術法則。圖五十係繪於對數尺度上，表示以一九一三年七月至一九一四年六月為基期所得一九一八年1437

項比價之分配。前在算術尺度上繪成曲綫所呈現顯著之不對稱狀態，今在對數尺度上已不復存在，而其曲綫與常態曲綫極為相似。

物價比率變動之合乎幾何法則，雖可視為平均此種比率應採用幾何平均數之強有力之理由，但不能遽謂除幾何平均數以外，其餘各種平均數即不宜供作編製物價指數之用。就大體言之，幾何平均數固不失為一種合理之平均數，惟優良之指數儘有採用他種平均數者，此點容後再行討論。

茲將前數頁內所述及之理論，再作簡略之說明。在確定指數編製方法以前，吾人首應研究統計資料之性質，以及此種資料所組成之次數分配之狀態，俾便決定普通統計方法之是否適用。此處之統計資料為以比率表示之各個物價變動，彙合許多物價比率即可得一次數分配。吾人已知此種次數分配頗與合乎差誤常態法則之各種統計資料之分配相似，而常含有顯著之集中趨勢，故可用平均數以代表之；惟若時期愈長，則集中趨勢亦漸欠明顯，離散之程度愈見分明，而平均數之代表性乃愈見減弱。吾人又知物價變動之分配常呈偏斜狀態，而在物價上漲時為尤甚。此種偏態之存在，乃由於物價比率有一定之下限而無一定之上限所致。在前此討論平均數時，吾人已知物價比率之分佈合乎幾何法則，故欲平均此項變動率應採用幾何平均數。惟此項結論雖無可非議，但在某種情形之下，編製指數常有採用他種平均數較為相宜者。

編製指數之應採用何種平均數之問題，以及聯帶發生之加權方法以何者為最優之問題，皆為討論指數編製法之中心。假定批發物價指數之編製目的係為測度一般物價之平均變動率，則應採用何種平均數，上節已論及之，惟因加權問題乃引起新因素，故須將加權問題併入討論。

編製指數採用加權方法之主要理由，前嘗述及。指數不論其為測度生產數量之增減，貿易額之消長，或物價之漲落，其所根據之各項數量及比率，重要程度並不相等。例如編製生產指數，煤在生產方面之地位

比鉛爲重要；編製批發物價指數時，小麥價格比較胡蘆子價格更爲重要。欲測定指數中所包含各品之比重程度，方法不一而足，大抵權數之選擇，須視指數編製之目的及搜集權數資料有無困難而定。下節將舉示權數之實例數種，然後根據此種實例再作詳盡之討論。

編製指數之各種方法

現時編製批發物價指數所用方法不一，其所由不同之原因頗多。其一，編製指數究以何種方法爲最優，在理論上迄無定論。其二，從事統計工作之各機關，其財力人力均不相同，故實際上所用方法不能一致。其三，編製指數之目的不同，故所用方法亦不能相同。

編製指數方法既不相同，結果勢難一致。吾人可就同一資料，應用各種方法，將其結果互相比較，則對其不同之處可易於明瞭。表四十四所示一九一〇年至一九二三年每年十二月一日十二種主要農產品之農場價格，爲說明各種方法時所用者。

符號之意義

說明各種指數計算法所用之符號，具有下述之意義：

p_0' ：基期（“0”）第一種物品之價格。

q_0' ：基期同一物品之數量。

p_1' ：計算時期（“1”）同一物品之價格。

q_1' ：計算時期同一物品之數量。

p_0'' ：基期第二種物品之價格。

q_0'' ：基期第二種物品之數量。

p_1'' ：計算時期第二種物品之價格。

q_1'' ：計算時期第二種物品之數量。

$\frac{p_1'}{p_0'}$ ：比價（計算時期第一種物品之價格與基期同一物品價格之百

表 四 十 四

1910—1923年每年十二月一日主要農產品十二種之農場價格(註一)

農產品	單 位	1910	1911	1912	1913	1914	1915	1916	1917	1918	1919	1920	1921	1922	1923
玉蜀黍	蒲式耳	.480	.618	.487	.691	.644	.575	.889	.1279	\$1.365	\$1.347	.677	\$0.423	.657	.727
棉花	磅	.141	.088	.119	.122	.068	.113	.196	.277	.276	.356	.140	.162	.238	.310
乾草	公噸	12.14	14.29	11.79	12.43	11.12	10.63	11.22	17.09	20.13	19.55	16.72	12.11	12.59	14.07
小麥	蒲式耳	.883	.874	.760	.799	.986	.919	1.603	2.008	2.042	2.151	1.443	.926	1.009	.923
燕麥	蒲式耳	.314	.450	.319	.392	.438	.361	.524	.666	.709	.715	.472	.302	.394	.415
白番芋	蒲式耳	.557	.799	.505	.687	.487	.617	1.461	1.228	1.193	1.606	1.164	1.101	.582	.823
糖(註二)	磅	.0393	.0494	.0405	.0354	.0392	.0520	.0569	.0672	.0728	.0728	.0577	.0370	.0570	.0747
大麥	蒲式耳	.578	.869	.505	.537	.543	.516	.881	1.137	.917	1.210	.707	.419	.525	.540
菸草	磅	.093	.094	.108	.128	.098	.091	.147	.240	.280	.390	.211	.199	.231	.203
胡麻子	蒲式耳	2.317	1.821	1.147	1.199	1.260	1.740	2.486	2.966	3.401	4.383	1.766	1.451	2.114	2.108
黑麥	蒲式耳	.715	.832	.663	.634	.865	.834	1.221	1.660	1.516	1.345	1.278	.697	.692	.647
米	蒲式耳	.678	.797	.935	.858	.924	.906	.889	1.896	1.918	2.668	1.189	.952	.934	1.103

(註一)表內1910—1920年之物價係 Warren M. Persons 氏所搜集,參看“Review of Economic Statistics”初卷第三號第

105頁中“Fisher's Formula for Index Numbers”. 1921—1923年之物價係根據“Weather, Crops and Markets”.

(註二)各年糖價係每年十二月份粗製糖之批發價格,該品之農場價格無從搜集.

分比)。

$\frac{q_1'}{q_0'}$: 比量。

P_0 : 基期之物價水準。

P_1 : 計算時期之物價水準。

簡單物價指數

費暄氏 (Irving Fisher) 分析各種指數計算法時(註), 嘗分別為六種基本方法: 即綜合法 the aggregative (或稱實價綜合法 price aggregative)、算術平均法、倒數平均法、幾何平均法、中位數法及衆數法。其中衆數法事實上絕少採用, 可勿具論, 其餘五種方法之特性, 將以最簡單之方式說明之, 然後再研討比較繁複之方式。

實價綜合法

簡單綜合法指數之編製, 係將某時期之各項物價一一相加, 依此求得各時期物價之總和, 互為比較, 即可測定一般物價之變動。用前述符號表示之, 得

$$\frac{P_1}{P_0} = \frac{\sum p_1}{\sum p_0}$$

茲根據表四十四之數字, 計算簡單綜合法指數如表四十五。表內第(2)欄為實價之綜合數, 為便於比較起見, 將此項數字一律化為百分比, 而以一九一〇年之綜合價格作為基數, 如第(3)欄。

用此法計算物價指數所得之結果, 將與根據同一資料用他法計算所得之結果互為比較。實價綜合所得之指數既非不加權指數, 又非權數相等之指數, 因每一物品影響於指數之勢力大小, 全視每單位之價格而定, 而各品單位之大小, 乃隨市場習慣為轉移, 此為實價綜合法之主要

(註)見“The Making of Index Numbers”, 一九二二年 Houghton Mifflin Co. 出版。

缺點。在本指數中，乾草係以噸計價，故其權數較其餘十一品之總權數猶大，胡麻子之重要程度次之。用實價相加所求得之指數，因加權方法全不合理，故不能表示該項農產物價之變動。

表 四 十 五
農產品物價指數

(1)	(2)	(3)
年 份	指 數 (實價綜合數)	指數,百分比 (1910=100)
1910	\$18.9653	100
1911	21.5814	114
1912	17.3785	92
1913	18.5124	98
1914	17.4722	92
1915	17.3540	91
1916	21.5739	113
1917	30.5142	161
1918	33.8198	178
1919	35.7938	189
1920	25.8247	136
1921	18.7790	99
1922	20.0230	106
1923	21.9437	116

爲欲避免因物品買賣單位不同所發生權數輕重不等之弊病起見；於是有將各品價格折算爲同一單位之辦法，例如乾草、米、玉蜀黍、棉花以及其他物品一律折合爲每磅之價格，然後將此項價格一一相加，以求指數者，勃蘭特斯脫指數(Bradstreet's index)即係採用此法計算。實則此法不過將前述不合理之加權方法，易以同樣不合理之另一種加權方法，且此法原欲使各品之權數相等，而事實上固未嘗予各品以相等之權數。例如一九一〇年乾草每磅值\$.00607，棉花每磅\$.141，米每磅\$.015，

故在各品每磅價格之綜合數內，棉花之權數九倍於米，二十三倍於乾草矣。

比價之算術平均數

編製指數所用之另一方法，係將每品之價格，各以該品基期價格作為 100 化成百分比，然後根據各個百分比，採取通常所用之平均數方法，求其平均，即得所求之指數。下表係就兩年之資料，以一九一〇年作為基期，說明此項計算手續之初步。

表 四 十 六
編製指數所用比價之計算

(1) 物 品	(2) 單 位	(3) 1910年價格	(4) 比 價	(5) 1911年價格	(6) 比 價
玉蜀黍.....	蒲式耳	\$.480	100	\$.618	128.8
棉花.....	磅	.141	100	.088	62.4
藍草.....	公 噸	12.14	100	14.29	117.7
小麥.....	蒲式耳	.883	100	.874	99.0
燕麥.....	蒲式耳	.344	100	.450	130.9
白番芋.....	蒲式耳	.557	100	.799	143.5
糖.....	磅	.0393	100	.0494	125.6
大麥.....	蒲式耳	.578	100	.869	150.2
荻草.....	磅	.093	100	.094	101.1
胡蘆子.....	蒲式耳	2.317	100	1.821	78.7
黑麥.....	蒲式耳	.715	100	.832	116.2
米.....	蒲式耳	.678	100	.797	117.5
			1200		1371.6

由上表數字可計算兩個年份比價之算術平均數。每一比價之公式為 $\frac{p_1}{p_0}$ ，如其項數為 N ，則求計算時期之指數所用之公式為

$$\frac{\sum \left(\frac{p_1}{p_0} \right)}{N}$$

茲以表內數字為例，計算指數如下：

$$\text{指數}(1910年) = \frac{1200}{12} = 100$$

$$\text{指數}(1911\text{年}) = \frac{1371.6}{12} = 114.3$$

依此計算一九一〇年至一九二三年之各年指數見表四十九第(3)欄。

比數之簡單平均數含有加權之意味

前節所述指數計算法，通常稱為比價之“未加權”指數，但按諸實際仍含有加權之意味，與前此所舉實價綜合法之兩例相似。此處各品用作權數之數量，即為該品在基年得價\$100所售出之數量。在前例中用作權數之各品數量如下：

玉蜀黍.....	208.3	蒲式耳
棉 花.....	710.0	磅
乾 草.....	8.24	公噸
小 麥.....	113.3	蒲式耳
燕 麥.....	291.0	蒲式耳
白番芋.....	180.0	蒲式耳
糖.....	2650.0	磅
大 麥.....	173.2	蒲式耳
菸 草.....	1076.0	磅
胡蘆子.....	43.2	蒲式耳
黑 麥.....	140.0	蒲式耳
米.....	147.7	蒲式耳

是以比價簡單平均之計算，實不啻依照上列各數量，決定在十一年間每年各項物品所可售得之總價值。按照一九一〇年之價格，上列數量之各項物品，均可售得\$100，故其總價值為\$1200；按照一九一一年之價格，上列數量之物品共可售得\$1371.60。此兩年之總價值各以12除之，得表四十九第(3)欄內之指數：一九一〇年指數為100，一九一一年為114 (114.3)等等。由是知“比價之未加權指數”實際即係實價之加權綜合數，所謂各品權數相等者，僅可謂一九一〇年基期內用作權數之數量，其價值各為\$100而已(註)。

(註)此項簡單平均比價之特性，為 F. R. Macaulay 氏所發現，見一九一五年十二月“American Economic Review”第928頁 Macaulay 氏著作中。

比價之中位數

計算每年平均比價，亦有捨算術平均數而用中位數者。其計算法係先將表四十六第(6)欄內之比價，依大小次序排列，得下列之分配：

62.4	117.7
78.7	125.6
99.0	128.8
101.1	130.9
116.2	143.5
117.5	150.2

上列數字分配內，最小之比價為 62.4，最大之比價為 150.2，中位數為 117.6。此中位數之值即一九一一年之指數也。依比價之中位數求得各年之指數見表四十九第(4)欄。

比價之幾何平均數

各年比價亦可用幾何平均數求其平均，以與前述各例所得之結果相比較。如以 $\frac{p_1'}{p_0'}$ 表示一個單純比價，則求比價 N 項之幾何平均數之公式為

$$Mg = \sqrt[N]{\frac{p_1'}{p_0'} \times \frac{p_1''}{p_0''} \times \frac{p_1'''}{p_0'''} \times \dots}$$

幾何平均數通常係用對數計算，故上式可變為

$$\text{Log } Mg = \frac{\log\left(\frac{p_1'}{p_0'}\right) + \log\left(\frac{p_1''}{p_0''}\right) + \log\left(\frac{p_1'''}{p_0'''}\right) + \dots}{N}$$

茲說明一九一〇年及一九一一年比價之幾何平均數之計算方法如下。表內各品之比價係根據表四十六重行列入者。

表 四 十 七
比價幾何平均數之計算

(1) 物 品	(2) 1910年比價	(3) 第(2)欄數字之對數	(4) 1911年比價	(5) 第(4)欄數字之對數
玉蜀黍.....	100	2.0	128.8	2.10992
棉花.....	100	2.0	62.4	1.79518
乾 草.....	100	2.0	117.7	2.07078
小 麥.....	100	2.0	99.0	1.99564
燕 麥.....	100	2.0	130.9	2.11694
白香芋.....	100	2.0	143.5	2.15685
糖.....	100	2.0	125.6	2.09899
大 麥.....	100	2.0	150.2	2.17667
燕 草.....	100	2.0	101.1	2.00475
胡麻子.....	100	2.0	78.7	1.89597
黑 麥.....	100	2.0	116.2	2.06521
米.....	100	2.0	117.5	2.07004
		24.0		24.55694

$$\text{Log } Mg(1910\text{年}) = \frac{24}{12} = 2$$

$$Mg = \text{anti-log } 2 = 100$$

$$\text{Log } Mg(1911\text{年}) = \frac{24.55694}{12} = 2.04641$$

$$Mg = \text{anti-log } 2.04641 = 111.3$$

此 111.3 之數值即一九一一年之指數也。依此求得之各年指數，見表四十九第(5)欄。

比價之倒數平均數

倒數平均數之特質，已在第四章內有所論述。吾人猶憶倒數平均數之倒數，即係各數值倒數之算術平均數。本例各數值即為以 $\frac{p_1'}{p_0'}$ 表示之比價，每一比價之倒數為 $\frac{p_0'}{p_1'}$ ，故求比價N項之倒數平均數之公式為

$$\frac{1}{H} = \frac{\frac{p_0'}{p_1'} + \frac{p_0''}{p_1''} + \frac{p_0'''}{p_1'''} + \dots}{N}$$

或

$$H = \frac{N}{\sum \left(\frac{p_0}{p_1} \right)}$$

其計算方法如下表所示：

表 四 十 八
比價倒數平均數之計算

(1) 物 品	(2) 1910年比價	(3) 第(2)欄數字之倒數	(4) 1911年比價	(5) 第(4)欄數字之倒數
玉 蜀 棉 花 乾 草 小 燕 燕 麥 白 番 糖 芋 大 麥 菸 草 胡 子 黑 麥 米	100	.01	128.8	.007763975
	100	.01	62.4	.016025640
	100	.01	117.7	.008496177
	100	.01	99.0	.010101010
	100	.01	130.9	.007639419
	100	.01	143.5	.006968641
	100	.01	125.6	.007961783
	100	.01	150.2	.006657790
	100	.01	101.1	.009891197
	100	.01	78.7	.012706480
	100	.01	116.2	.008605852
	100	.01	117.5	.008510638
			.12	

$$H(1910年) = \frac{12}{.12} = 100$$

$$H(1911年) = \frac{12}{.111328602} = 107.8$$

依此求得之各年指數，見表四十九第(6)欄。

表 四 十 九
1910—1923年農產物價指數
(1910年=100)

(1) 年 份	(2) 實價綜合數 (以百分比表示)	(3) 比價之 算術平均數	(4) 比價之 中 位 數	(5) 比價之 幾何平均數	(6) 比價之 倒數平均數
1910	100	100	100	100	100
1911	114	114	118	111	108
1912	92	95	93	92	90
1913	98	104	98	100	97
1914	92	101	102	97	91
1915	91	104	104	102	101
1916	113	156	152	151	147
1917	161	208	208	204	198
1918	178	215	209	210	205
1919	189	252	226	241	231
1920	136	151	143	145	139
1921	99	115	99	107	101
1922	106	129	114	124	119
1923	116	142	134	135	129

上述五種指數之計算，未嘗採用合理之加權制度，故該五種指數皆稱為“未加權”指數。實則此名詞易滋誤解，蓋第一種根據實價綜合法求

得之指數，實爲加權甚重之指數，惟權數不合理耳；其後四種指數，實際亦係加權指數，每品所用權數爲一九一〇年\$100所可購得該品之數量。茲將此五種指數併列一表(表四十九)，互爲比較。各年指數皆取整數，整數後一位之小數，四捨五入。

該表內各指數繪爲曲綫，如圖五十一。

各種簡單指數之比較

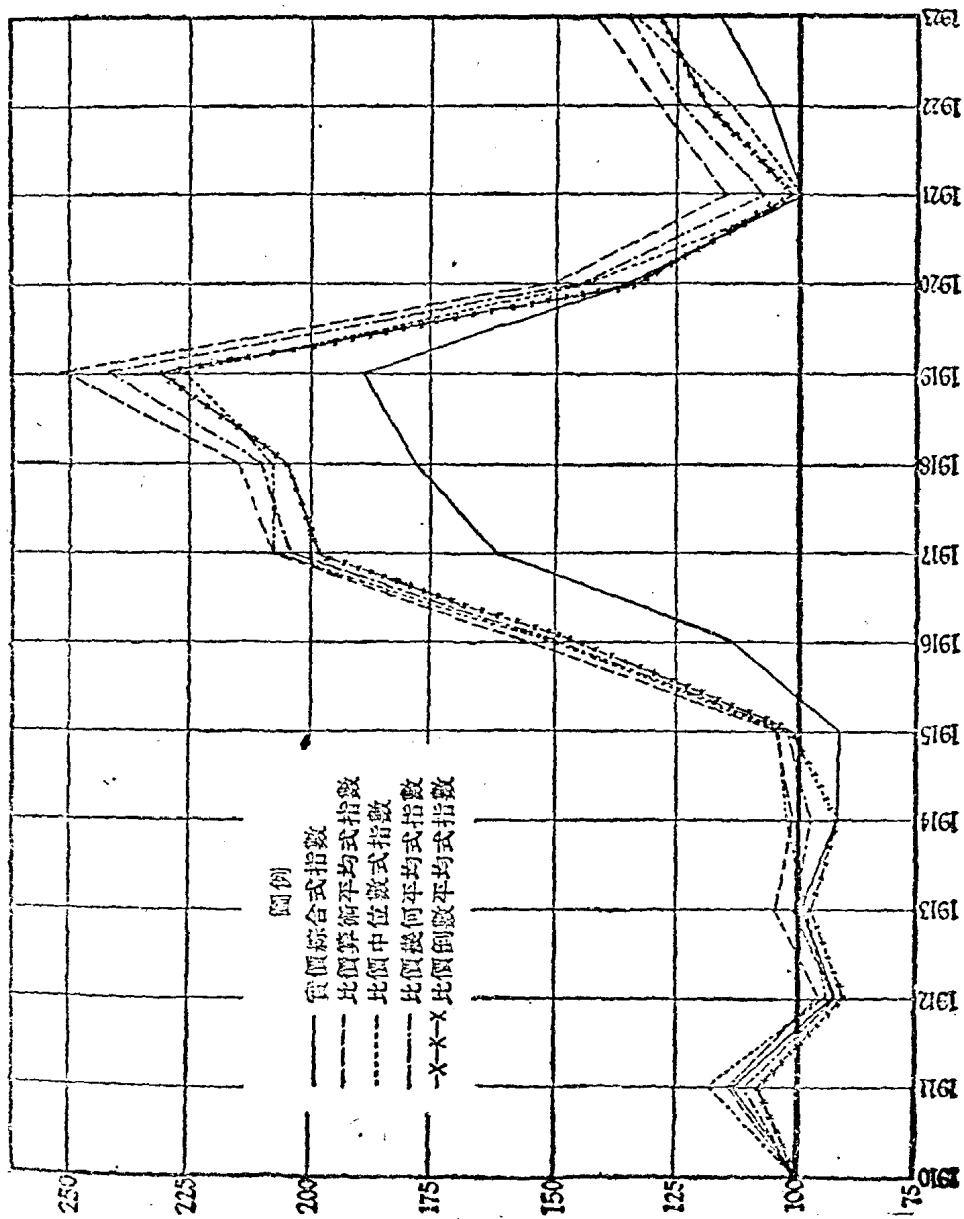
用比價平均法求得之四種指數，彼此極爲相似，惟與實價綜合式指數相比，則其相似程度較淺。實價綜合法之缺點，前嘗論及，故用此法測度物價變動殊不可靠。至於其餘四種指數中，算術平均數、幾何平均數及倒數平均數之指數，由於該平均數本身之性質，恆有一定之關係，即除基年以外，幾何平均指數恆較算術平均指數爲小，而倒數平均指數又恆較幾何平均指數爲小，且各項比價愈分散，此差數亦愈增大。中位數指數則因每期所平均者祇有十二項數字，故其數值不穩定，與其他三種平均指數之關係，遂難於確定。

吾人對於此種數值各不相同之指數，將何所取捨乎？在各種“未加權”指數中實無一足稱完備者，因其所採用之權數，並不能表示指數內所包含各品之相對的重要程度也。茲姑將權數問題暫勿置論，吾人亦可設法測驗物價變動所用各種測度方法之適當否乎？

時間顛倒測驗法

費暄氏(Irving Fisher)嘗用所謂時間顛倒測驗法(time reversal test)以測驗指數計算方法之優劣。此法實即測驗應用每種方法，計算前後兩時期之指數，經互換基期後所得之結果是否合理耳。例如糖每磅價格自一九一〇年之\$.04漲至一九一一年之\$.08，故一九一一年之價格，應爲一九一〇年之200%，一九一〇年之價格，應爲一九一一年之50%。一數應爲另一數之倒數，而其乘積(2.00×.50)應等於一。據此原則，如

用某法計算指數，而得某年一般物價之水準為其前一年水準之200%，則時間顛倒後，第一年之物價水準應為第二年水準之50%。故如根據任何兩年之物價，採用同一種指數計算法，將基期互換所求得之兩個指數，應互為倒數，而其乘積應等於一，否則此指數計算法必有偏誤之處。



圖五十一. 1910—1923年農產物價五種簡單指數之比較(1910年=100)

此項測驗可仍就一九一〇年及一九一一年之物價資料，以求前述各種指數計算法之是否合於上項測驗。以一九一〇年為基期，求得指數如下：

年 份	實價綜合數 (以百分比表示)	比價之 算術平均數	比價之 中 位 數	比價之 幾何平均數	比價之 倒數平均數
1910	100	100	100	100	100
1911	113.79414	114.3	117.6308	111.3	107.8

以一九一一年為基期，求得指數如下：

年 份	實價綜合數 (以百分比表示)	比價之 算術平均數	比價之 中 位 數	比價之 幾何平均數	比價之 倒數平均數
1910	87.87799	92.76	85.0117	89.85	87.47
1911	100	100	100	100	100

如將上列第一表內一九一一年指數，各乘以第二表內一九一〇年指數，得下列數值(相乘時指數須化為比率，並非即以百分比相乘)：

實價綜合數	比價之 算術平均數	比價之中位數	比價之 幾何平均數	比價之 倒數平均數
1.00	1.0602	1.00	1.00	.9429

上列各種指數計算法中，合乎時間顛倒測驗者計有三種。算術平均數及倒數平均數皆不能適合時間顛倒測驗。算術平均數有偏高之弊，其偏高之程度，以一九一〇年及一九一一年之差誤相混合後，在百分之六以上。倒數平均數偏誤之程度亦如之，惟其方向相反。此兩種平均數所有之偏誤，如不設法校正，則根據此項平均數之各種計算法，皆不宜供作編製指數之用。

指數之加權問題

在前節內已將五種簡單物價指數略加論述，惟指數常因權數問題之歸入，變化複雜，其計算法之種類乃大為增加。本書所欲論述者，則為比較重要之數種。

吾人欲求測度物價變動之準確指數，必須採用合理之權數，此種權數必須確能表示指數內各項物品之相對的重要程度。如權數問題未加注意，即難免發生偶然的加權或權數不合理之弊。

茲仍以前述各例所用之物價為例，以說明加權方法及指數所受權數變更之影響。編製此項農產物價指數所用權數，或為農產物之生產數量，或為其生產價值，全視所選指數方式而定。各項物品一九一〇年至一九二三年之生產數量如表五十所示(註)：

實價加權綜合法

由實價綜合法所得指數，其不合理之處，前已論述。惟在實價相加之先，如各以適當之權數乘之，則可免不合理之結果。如所用權數為基年(即“0”年)之生產數量，則實價加權綜合公式為

$$\frac{\sum p_1 q_0}{\sum p_0 q_0}$$

此法為美國勞工統計局所採用者，惟該局所用權數非基年之數量，而為另一年之數量而已。加權綜合式指數之計算手續，將在表五十一中說明之。

將表中第(5)及第(8)欄內之數值分別相加，以求其總價值。以兩年中任何一年作為基年，而以另一年之總價值化為其百分比，即得百分比之指數。例如以一九一〇年之總價值作為基數，則一九一一年之指數為111.5。依此求得其他各年之指數見表五十四第(2)欄。

加權綜合價值之計算，亦有不採用基期之權數，而用後一時期之權數者，則屬於另一種之加權綜合式指數。例如吾人可用 q_1 （第一計算年“1”之數量）為權數，以比較第一計算年“1”與基年“0”之物價，但亦可用

(註)一九一〇年至一九一九年之生產數量係根據台氏(E. E. Day)所著“An Index of the Physical Volume of Production”第8頁，一九二一年出版(係根據一九二〇年九月“Review of Economic Statistics”所載原文翻印者)。一九二〇年至一九二三年之生產數量，係根據“Weather, Crops and Markets”。

表 五 十
1910—1923年農產品十二種之每年產量

年 份	玉蜀黍 (百萬蒲 式耳)	棉花 (百萬包) (註一)	乾 草 (百萬 公噸)	小 麥 (百萬蒲 式耳)	燕 麥 (百萬蒲 式耳)	白 燕 麥 (百萬蒲 式耳)	糖 (百萬磅) (註二)	大 麥 (百萬蒲 式耳)	菸 草 (百萬磅)	胡 蔬 子 (百萬蒲 式耳)	黑 麥 (百萬蒲 式耳)	米 (百萬蒲 式耳)
1910	2886	11.61	69.38	635.1	1186	249.0	1618	173.8	1108	12.72	34.90	24.51
1911	2531	15.69	54.92	621.3	922	292.7	4296	160.2	905	19.37	33.12	22.91
1912	3125	13.70	72.69	730.3	1418	420.6	3764	223.8	963	28.07	35.66	25.05
1913	2447	14.16	64.12	763.4	1122	331.5	3941	178.2	854	17.85	41.38	25.74
1914	2673	16.14	70.07	891.0	1141	409.9	4134	195.0	1035	13.75	42.78	23.65
1915	2995	11.19	85.92	1025.8	1519	359.7	4230	228.8	1062	14.03	54.05	28.95
1916	2567	11.45	91.19	636.3	1252	287.0	4671	182.3	1153	14.30	48.86	40.86
1917	3065	11.30	83.31	636.7	1593	442.1	3950	211.8	1249	9.16	62.93	37.74
1918	2503	12.04	76.66	921.4	1538	411.9	4220	256.2	1439	13.37	91.04	38.61
1919	2917	11.42	91.33	941.0	1248	357.9	3634	165.7	1390	8.92	88.48	41.06
1920	3209	13.44	87.85	833.0	1497	403.3	4065	189.3	1582	10.77	60.5	52.07
1921	3069	7.95	82.38	814.9	1078	361.7	154.9	1070	8.03	61.7	37.61
1922	2906	9.76	95.88	867.6	1216	453.4	182.1	1247	10.37	103.4	41.40
1923	3054	10.08	98.90	785.7	1230	412.4	198.2	1475	17.43	63.0	33.25

(註一)每包毛重500磅。

(註二)糖之產量數字,係各該收穫年美國大陸所產甜菜糖及蔗糖之數量,再加上該收穫年七月一日起之會計年度內自英國殖民地輸入之數量之總額。計算1921年至1923年指數時,係根據估計之數量。

q_2 (第二計算年“2”之數量) 爲權數, 以比較第二計算年“2”與基年“0”之物價。今以代數式表示之, 則求第一計算年“1”指數之公式爲

$$\frac{\sum p_1 q_1}{\sum p_0 q_1}$$

用此式計算指數之手續, 與前例完全相同, 惟所用權數則須每年更換一次。依此求得之各年指數見表五十四第(3)欄。

表 五 十 一
實價加權綜合式指數之計算

(1) 物品	(2) 單位	(3) 1910年 價格	(4) 權 數 (1910年 生產數量, 單位百萬)	(5) 價格×權數 ($p_0 q_0$)	(6) 1911年 價格	(7) 權 數 (1910年 生產數量, 單位百萬)	(8) 價格×權數 ($p_1 q_0$)
玉蜀黍	蒲式耳	\$.480	2,886	\$1,385,280,000	\$.618	2,886	\$1,783,518,000
棉 花	磅	.141	5,805	818,505,000	.088	5,805	510,840,000
乾 草	公 噸	12.14	69.35	842,273,200	11.29	69.38	991,440,200
小 麥	蒲式耳	.883	635.1	560,793,300	.874	635.1	555,077,400
燕 麥	蒲式耳	.344	1,186	407,984,000	.450	1,186	533,700,000
白番芋	蒲式耳	.557	349.0	194,393,000	.799	349.0	278,851,000
糖	磅	.0393	3,618	142,187,400	.0494	3,618	178,729,200
大 麥	蒲式耳	.578	173.8	100,456,400	.869	173.8	151,032,200
菸 草	磅	.093	1,103	102,579,000	.094	1,103	103,682,000
胡 蘿 子	蒲式耳	2.317	12.72	29,472,240	1.821	12.72	23,163,120
黑 麥	蒲式耳	.715	34.90	24,953,500	.832	34.90	29,036,800
麥	蒲式耳	.678	24.51	16,617,780	.797	24.51	19,534,470
				\$4,625,494,820			\$5,158,634,390

前述兩例中之權數皆係數量, 因價格乘數量後得價格之總數; 但若加權於比價時, 即不能用數量作權數。此種抽象比價必須用價值作爲權數, 則所得乘積方可比較, 因價值有一共同之貨幣單位, 而數量單位則各品互異故也。用作權數之價值, 可用各種方式求得之。

費喧氏(註)嘗列舉計算價值之四種方式如下, 其中第二第三種係屬混合式。

(註)見 “The Making of Index Numbers” 第54頁。

- I. 每品之權數 = 基年價格 × 基年數量 (p_0q_0)
 II. 每品之權數 = 基年價格 × 計算年數量 (p_0q_1)
 III. 每品之權數 = 計算年價格 × 基年數量 (p_1q_0)
 IV. 每品之權數 = 計算年價格 × 計算年數量 (p_1q_1)

上列各種價值權數皆含有所謂權數之偏誤 (weight bias) (以數量作為權數, 即無偏誤發生), 正與某數種平均數之含有偏誤相似。第一第二兩種權數 (係按基年價格求得之價值) 常偏低, 第三第四兩種權數 (係按計算年價格求得之價值) 常偏高, 此點既可以數理證明之 (註), 亦可以數字證實之。

茲為便於說明起見, 下述各例皆以一九一〇年基年各品生產量之價值作為權數。此項價值見表五十二第(3)欄。各品權數均以百萬為單位, 所以便於計算也。

比價之加權算術平均數

此種指數之計算, 係將每品之比價各與其權數相乘, 而後以此乘積

(註) 用第三種權數求得之指數, 必較用第一種權數求得之指數為高。每品之比價如以第三種權數乘之, 得

$$\frac{p_1}{p_0} \times p_1q_0$$

如以第一種權數乘之, 得

$$\frac{p_1}{p_0} \times p_0q_0$$

倘 p_1 大於 p_0 (即比價大於 100), 則第三種權數 (p_1q_0) 必大於第一種權數 (p_0q_0), 故此比價在 100 以上之各品, 如用第三種權數, 其所得重量必較用第一種為大; 倘 p_1 小於 p_0 , 則第三種權數 (p_1q_0) 必小於第一種權數 (p_0q_0), 故此比價在 100 以下之各品, 如用第三種權數, 其所得重量必較用第一種為小。今以價格高漲之物品所得重量過大, 價格跌落之物品所得重量過小, 其結果按照第三種權數求得之指數, 恆較按照第一種權數求得之指數為小。第四種權數與第二種權數比較, 所得情形亦同。至於第一種與第四種權數之間, 並無一定之關係, 但就大體而論, 用第四種權數求得之指數, 當較用第一種權數求得者為大, 蓋基年之權數恆偏低, 而計算年之權數恆偏高也 (關於權數偏誤問題, 費噠氏嘗作詳盡之討論, 見 "The Making of Index Numbers" 第五章第 384—387 頁)。

相加，除以權數之和。其計算手續如下表所示：

表五十二 比價加權算術平均數之計算

(1) 物 品	(2) 1910年 比 價	(3) 權 數	(4) 比價×權數	(5) 1911年 比 價	(6) 權 數	(7) 比價×權數
玉蜀黍.....	100	\$1,385	\$138,500	128.8	\$1,385	\$178,388.0
棉花.....	100	819	81,900	62.4	819	51,105.6
乾 草.....	100	842	84,200	117.7	842	99,103.4
小 麥.....	100	561	56,100	99.0	561	55,539.0
燕 麥.....	100	408	40,800	130.0	408	53,407.2
白番芋.....	100	194	19,400	143.5	194	27,839.0
糖.....	100	142	14,200	125.6	142	17,835.2
大 麥.....	100	100	10,000	150.2	100	15,020.0
菸 草.....	100	103	10,300	101.1	103	10,413.3
胡麻子.....	100	29	2,900	78.7	29	2,282.3
黑 麥.....	100	25	2,500	116.2	25	2,905.0
米.....	100	17	1,700	117.5	17	1,997.5
		\$4,625	\$462,500		\$4,625	\$515,835.5

(表內權數係1910年基年各品生產數量之價值，以百萬元為單位。)

$$\text{加權算術平均數(1910年)} = \frac{\$462,500}{\$4,625} = 100.$$

$$\text{加權算術平均數(1911年)} = \frac{\$515,835.5}{\$4,625} = 111.5.$$

於此所應注意者，此處一九一一年之指數，與根據表五十一計算所得之指數完全相同。按後者係用實價加權綜合法計算者，其權數為基年之生產數量。比價算術平均式之指數，如用基年價值作為權數，則恆與根據此種綜合價值所化成百分比之指數相等(註)。

(註)此點可以代數式證明之。今設各品之基年價值為 p_0q_0 ，其第二年之比價為 $\frac{p_1}{p_0}$ ，則此種比價之加權算術平均數等於

$$\frac{\frac{p_1'}{p_0'} \times p_0'q_0' + \frac{p_1''}{p_0''} \times p_0''q_0'' + \frac{p_1'''}{p_0'''} \times p_0'''q_0''' + \dots}{p_0'q_0' + p_0''q_0'' + p_0'''q_0''' + \dots}$$

由上式化為

$$\frac{\sum p_1q_0}{\sum p_0q_0}$$

此式即前述實價加權綜合法之公式。

比價倒數平均數之公式，如其權數為第二年之數量，可化為

$$\frac{\sum p_1q_1}{\sum p_0q_1}$$

此式亦即以第二年數量作為權數之實價加權綜合法之公式也。

比價之加權幾何平均數

加權幾何平均數之計算手續，與未加權之幾何平均數相同，惟前者每一比價之對數，須與權數相乘，再以此項已加權之對數之和，除以權數之和，其結果即為所求指數之對數(註)。其計算法在下表內說明之。

表 五 十 三

1911年比價加權幾何平均數之計算

(1910年=100)

物 品	1911年 比 價	比價之對數	權數 (191 年 生產價值)	比價之對數×權數
玉蜀黍.....	128.8	2.10992	1,385	2922.23920
棉 花.....	62.4	1.79518	819	1470.25242
乾 草.....	117.7	2.07078	842	1743.59676
小 麥.....	99.0	1.99564	516	1119.55404
燕 麥.....	130.9	2.11694	408	863.71152
白 苜 蓿.....	143.5	2.15685	194	418.42890
糖.....	125.6	2.09899	142	298.05658
大 麥.....	150.2	2.17667	100	217.66700
蔡 草.....	101.1	2.00475	103	206.48925
胡 蘆 子.....	78.7	1.89597	29	54.98313
黑 麥.....	116.2	2.06521	25	51.63025
米.....	117.5	2.07004	17	35.19068
			4,625	9401.79973

$$\text{Log } Mg = \frac{9401.79973}{4625} = 2.032822$$

$$Mg = 107.9$$

如表所示，以一九一〇年為基期之一九一一年指數為 107.9。依此求得其他各年之指數見表五十四第(5)欄，與其他各種加權指數並列一表。

然則此三種加權指數，吾人將用何法區別其優劣乎？吾人可先用前次測驗五種簡單指數所用之時間顛倒測驗法，作同樣之測驗。該三種加權指數皆不合於時間顛倒測驗，其中加權幾何平均式指數亦非例外。

(註)加權幾何平均公式見第114頁。

按簡單幾何平均式指數雖合於此項測驗，但一經加權，指數即含有偏誤。以時間顛倒測驗而言，三種加權指數中無一能適應者。吾人可更以費暄氏所發現之第二種基本測驗法，即所謂因子顛倒測驗法 (factor reversal test) 測驗之。

因子顛倒測驗法

每一物品任何一年之總價值：本係該品是年之生產數量與每單位之價格相乘之積；以代數式表示之為 $p_0'q_0'$ 。某年總價值與前一年總價值之比為 $\frac{p_1'q_1'}{p_0'q_0'}$ 。今若某年之價格及數量各增一倍，則其價格之百分比為200，數量之百分比亦為200，其價值之百分比應為400，故第二年之總價值應為第一年總價值之四倍。價值之百分比為價格與數量兩個百分比相乘之積，此項關係在個別物品單獨觀察時，頗為明顯。

如吾人根據多種物品編製某年與後一年間價格變動之指數，及某年與後一年間數量變動之指數，則此兩種指數之乘積，應等於第二年總價值對於第一年總價值之比率。如此項乘積並不與價值之比率相等，則可斷定價格及數量兩種指數中，必有一種含有差誤，或兩者皆含有差誤。

今吾人以前述第一種加權綜合式指數(公式為 $\frac{\sum p_1 q_0}{\sum p_0 q_0}$) 試驗其是否合於時間顛倒測驗。依照此項公式，可將 p 及 q 之地位互易，以求物量指數。其公式為

$$\frac{\sum q_1 p_0}{\sum q_0 p_0}$$

上式之分子分母所含之物價因子，數值相同，因吾人現欲測度者，為數量變動之結果也。以前述十二種農產品之數字代入公式，得

$$1911\text{年之物量指數}(1910\text{年}=100) = \frac{\$1,446,264,630}{\$1,625,494,820} = .96125$$

以一九一〇年爲基期之一九一一年生產數量指數化爲百分率之形式得96.125,其加權綜合式之物價指數爲111.5,兩者之乘積爲

$$.96125 \times 1.115 = 1.0718$$

上式之意義,謂若物價增高 11.5%,同時物量減低 3.875%,則物值應增 7.18%也。

吾人如直接由兩年之價值以求一九一一年價值之比率,得

$$\frac{\sum p_1 q_1}{\sum p_0 q_0} = \frac{\$4,748,718,320}{\$4,625,494,820} = 1.02664$$

兩項結果相差 $4\frac{1}{2}\%$,可知加權綜合公式不合於因子顛倒測驗,故此式不能認爲完善。

吾人再以第二種加權綜合式指數試驗其是否合於此種試驗。按以一九一〇年爲基期之一九一一年指數如下:

$$\text{物價指數} = \frac{\sum p_1 q_1}{\sum p_0 q_1} = 106.8$$

$$\text{物量指數} = \frac{\sum q_1 p_1}{\sum q_0 p_1} = 92.05$$

$$\text{乘積} = .9205 \times 1.068 = .9831$$

(在物價及物量指數兩者相乘時,係按照比率計算,並不按照百分比計算)。

依此求得之結果與物值比率相比較,仍相差4%以上,惟其差異之方向與前相反耳。

至於加權幾何平均式指數,亦不合於因子顛倒測驗。按幾何平均式與綜合式之簡單指數本無偏誤,而因權數之採用,以致指數含有差誤,惟欲使指數確切表現事實,則權數又不能捨棄不用。事實上無論簡單指數或加權指數,無一能同時合於此兩項基本測驗者。費暄教授嘗試驗公式四十六種,其中合於時間顛倒測驗者祇有四種(即簡單之幾何平均式、中位數式、衆數式及綜合式),而無一能合於因子顛倒測驗者。

理想公式

吾人爲修正公式之偏誤，於是有配合法之採用。其法係將差誤方向相異之公式，求其幾何平均數，以消除其偏高偏低之差誤。費暄教授嘗用此法，將所有計算公式一一加以配合，以試驗公式配合後，能否合於兩種基本測驗。其結果乃發現配合之公式十三種，皆能合於該兩項基本測驗，並由費暄氏選定其中準確可靠與計算簡便之一種，作爲理想公式(“ideal” formula)。用此公式求得之指數，實即前述兩種綜合式指數之幾何平均數。其公式(註)爲

$$\sqrt{\frac{\sum p_1 q_0}{\sum p_0 q_0} \times \frac{\sum p_1 q_1}{\sum p_0 q_1}}$$

茲爲舉例說明起見，可將前已求得之兩個單純指數代入公式，以求理想指數。其一九一一年之指數爲

$$\begin{aligned} \text{理想指數} &= \sqrt{111.5 \times 106.8} \\ &= 109.14 \end{aligned}$$

此理想指數能合於時間顛倒及因子顛倒兩種測驗。茲先證明其合於第一種測驗如下：

$$\begin{aligned} 1911\text{年物價指數}(1910\text{年}=100) &= 109.14 \\ 1910\text{年物價指數}(1911\text{年}=100) &= 91.63 \\ 1.0914 \times .9163 &= 1.00 \end{aligned}$$

次證明其合於因子顛倒測驗。以一九一〇年爲基期，按照理想公式求得之一九一一年指數爲

$$\begin{aligned} \text{物價指數} &= \sqrt{\frac{\sum p_1 q_0}{\sum p_0 q_0} \times \frac{\sum p_1 q_1}{\sum p_0 q_1}} = 109.14 \\ \text{物量指數} &= \sqrt{\frac{\sum q_1 p_0}{\sum q_0 p_0} \times \frac{\sum q_1 p_1}{\sum q_0 p_1}} = 94.07 \end{aligned}$$

(註)此式亦爲Bowley, Pigou, Walsh, Young 諸氏所各自發現者。見“The Making of Index Numbers” 第十五章240—242頁。

$$\text{物價指數} = \frac{\sum p_1 q_1}{\sum p_0 q_0} = 102.664$$

物價指數及物量指數之乘積 = $1.0914 \times .9407 = 1.02668$

茲將一九一〇年至一九二三年之理想指數、理想公式所由配合之兩種加權綜合式指數、以及用基年價值作為權數之幾何平均式指數，併列下表。此四種指數繪成曲綫，如圖五十二。

表 五 十 四
1910—1923年農產物價加權指數之比較(註)
(1910年=100)

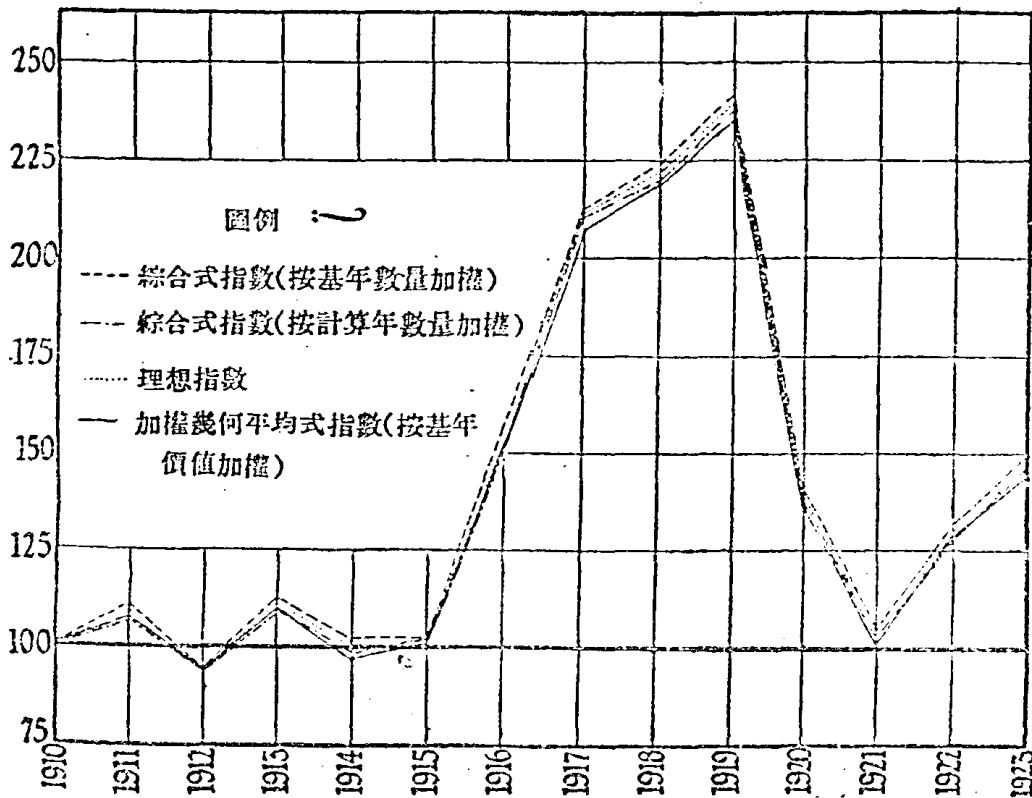
(1) 年份	(2) 綜合式指數 (按基年數量加權)	(3) 綜合式指數 (按計算年數量加權)	(4) 理想指數 (第(2)(3)兩指數 之幾何平均數)	(5) 加權幾何平均式指數 (按基年價值加權)
	$\frac{\sum p_1 q_0}{\sum p_0 q_0}$	$\frac{\sum p_1 q_1}{\sum p_0 q_1}$		
1910	100.0	100.0	100.0	100.0
1911	111.5	106.8	109.1	107.9
1912	94.4	93.7	94.0	94.0
1913	112.4	109.4	110.9	109.9
1914	102.8	98.9	100.8	97.0
1915	102.3	102.3	102.3	101.0
1916	156.8	151.3	154.0	151.2
1917	212.5	210.9	211.7	207.2
1918	224.2	220.3	222.2	219.3
1919	241.3	238.2	239.7	235.1
1920	139.9	137.9	138.9	137.2
1921	104.3	103.6	104.0	101.6
1922	131.6	127.9	129.7	128.9
1923	149.3	145.2	147.2	144.1

此四種加權指數互為比較，其相差程度比各該式之簡單指數相互間之差度，已大見減小。各年指數中雖尙不免略有出入之處，惟前此簡單指數中所呈現奇特之差異，確已不復存在矣。

上述測度一九一〇年與後此各年物價平均變動之四種指數中，自以理想指數為最優。於此所應注意者，此項指數係為測度兩個指定時

(註) 1910—1919年之兩種加權綜合式指數係 W. M. Persons 氏所編製者 (見 “Fisher's Formula for Index Numbers”，原文載 “Review of Economic Statistics” 初卷第三號第107頁)。

期間之變動，不能與中間各期指數直接比較。例如一九二三年之指數，係按照一九一〇年及一九二三年之價格及數量之關係求得，其間含有兩重權數，而權數又各年不同。如欲以一九二三年與一九二二年比較，則須編一新指數，而以一九二三年及一九二二年之價格及數量為根據，其他各年之價格及數量不與焉。如根據上表所列該兩年之理想指數，直接比較，則因採用加權制度，每易發生差誤。



圖五十二. 1910—1923年農產物價四種加權指數之比較(1910年=100)

幾何平均式指數，如其權數固定不變，則每年指數不特可與基年指數直接比較，且可與其他任何一年之指數直接比較，此為幾何平均式指數之優點。其基期可由指數直接轉換，所得結果與根據原有物價更換基期重行計算者完全相同。理想指數如以同法轉換基期，雖不致發生重大

差誤，然不能得極準確之結果(註)。

理想公式之未能普遍採用，尙有其他重要原因在，蓋若採用此式，則用作權數之數量資料須每年或每月搜集一次，其間不無困難之處，而計算手續又頗費時間也。惟指數果欲力求準確，則計算手續之困難，要所不免。費暄氏以爲指數如不採用理想公式，亦可採用下式以替代之，而計算手續之簡捷則遠過之。其式如下：

$$\frac{\Sigma(q_0 + q_1)p_1}{\Sigma(q_0 + q_1)p_0}$$

此式由形式簡單、計算便捷、結果準確及循環差誤微渺之四點觀之，費暄氏認爲最優良之公式，亦爲愛奇渥斯氏(Edgeworth)及馬莎爾氏(Marshall)所最倡導者。採用此式所得結果，大抵與根據理想公式求得者，相差不及 $\frac{1}{4}\%$ 。下表係採用一九一〇年及一九一一年之資料，以說明其計算方法。

表五十五
按聯合數量加權之綜合指數之計算

(1) 物 品	(2) 單 位	(3) 1910年 價 格	(4) 1910年數量+ 1911年數量 (單位：百萬)	(5) 1910年價格× 數量之總和 (3)×(4)	(6) 1911年 價 格	(7) 1911年價格× 數量之總和 (6)×(4)
玉蜀黍.....	蒲式耳	.480	5,417	\$2,600,160,000	\$.618	\$3,347,706,100
棉 花.....	磅	.141	13,650	1,924,650,000	.088	1,201,200,000
乾 草.....	公噸	12.14	124.30	1,509,002,000	14.29	1,776,247,000
小 麥.....	蒲式耳	.883	1,256.4	1,109,401,200	.874	1,098,093,600
燕 麥.....	蒲式耳	.314	2,108	725,152,000	.450	948,600,000
白番芋.....	蒲式耳	.557	641.7	357,426,900	.799	512,718,300
糖.....	磅	.6393	7,914	311,020,200	.0494	390,915,600
大 麥.....	蒲式耳	.578	334	193,052,000	.869	290,246,000
菸 草.....	磅	.093	2,008	186,744,000	.094	188,752,000
胡蘆子.....	蒲式耳	2.317	32.09	74,352,530	1.821	58,435,890
黑 麥.....	蒲式耳	.715	68.02	48,634,300	.832	56,592,640
米.....	蒲式耳	.678	47.44	32,164,320	.797	37,809,680
				\$9,071,759,450		\$9,907,352,710

(註)如指數注重於逐年之比較，則編製理想指數可採用連環基期制(chain system)，先以每年之物價作爲基價而求次年指數，得各年之環比指數(link index number)，然後任擇一年爲基期，將各年環比指數連續相乘，即得以該年爲基年之連環指數(chain index number)。Warren M. Persons 氏嘗發現連環指數包含累積性之差誤，如連環之時間過長，常有發生重大差誤之虞。

$$\frac{\Sigma(q_0 + q_1)p_1}{\Sigma(q_0 + q_1)p_0} = \frac{\$9,907,352,710}{\$9,071,759,450}$$

$$= 109.2 \text{ (以1910年爲基期之1911年指數)}$$

(此指數係以百分比表示者。)

此項公式所需數字資料，與理想公式相同，惟其中數量資料頗不易搜集，非在舉行普查時期，無從求得，故未值普查各年，不得不採用固定權數，此時或以採用加權綜合公式

$$\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0}$$

最爲相宜。加權幾何平均公式雖有種種優點，但受權數偏誤之影響。如權數資料不能蒐集，甚或不能約計，不得不採用簡單公式時，則比價之簡單幾何平均數及簡單中位數遠較其他各法爲優，而兩者之中，尤以幾何平均數更切實用。

各種指數之可靠性

物價指數皆係根據物價之樣本所編製，而求得之指數，則常視爲全體物價之代表，故吾人需要一種方法以判定各種指數之可靠性，並以試驗根據各種樣本求得之指數，是否有穩定性。就吾人經驗而言，根據各種樣本編製指數，每得不同之結果，然則究以何種指數在取樣上之差誤爲最小乎(註一)?

凱萊氏(Truman L. Kelley)(註二)嘗試算各種指數之機誤，並依機誤之大小分級評定各種指數之優劣。加權幾何平均式及加權中位數式兩種指數最爲可靠，所受取樣差誤之影響爲最小，故列居首位。費暄氏之理想指數位次較低，但仍較比價之加權算術平均式及倒數式之位次爲高。至於簡單未加權之算術平均式指數，則列居末位。

(註一)取樣問題與統計量數之可靠性有關之處，後文有更詳盡之討論。

(註二)見 Kelley 氏著 "Statistical Method" 第334—346頁。

凱萊氏認加權幾何平均指數爲最可靠，最有伸縮，而最優良之物價指數。此外用選擇之固定權數(未必定與交易量或消費量之數字完全相等者)之綜合式指數

$$\frac{\Sigma(p_1 w)}{\Sigma(p_0 w)}$$

因其具備優良指數之各要素，亦認爲最優之公式，其位次與加權幾何平均式相埒，而較理想公式爲高。此式中所用權數，不定爲實際數量，故解決加權問題，較有伸縮之餘地。

關於編製物價指數之其他問題

前文已將關於編製物價指數所用平均方法等技術問題，加以討論。在各種方法中，有不宜採用者，亦有適合於某種目的者。應用指數者務須透切瞭解該指數所採用之編製方法，俾可確知該指數之作用及其可靠程度，而不致誤解。

編製指數者所遇之問題，猶不限於前述數種，應用指數者所應注意者，亦非僅此數種。其與平均方法及加權問題同等重要者，厥爲關於選取樣本之實際問題。吾人如欲準確測度物價水準之變動，必須決定計算期內貨幣(包括信用)之流通數量與市場上一切物品實際交易數量之比率而後可，故欲測度兩時期間一般物價之變動，亦必須詳悉該兩時期內之該兩種要素。惟此項資料自不易蒐集，故物價指數之編製，不得不採用抽樣法。編製物價指數之最要問題，乃爲決定指數內所應包含之物品項數及其性質。

指數包含物品之項數問題

指數編製方法常隨目的而異，前已言之，故指數內所應包含之物品多寡及其種類，亦隨編製之目的爲轉移。指數如爲測度一般物價水準之

變動，則指數內應包含之物品項數問題，解答甚易，即樣本愈大，指數之代表性可以愈高，因樣本較大，其次數多邊圖必較樣本小者為更能合於全體物價變動之理想曲綫也。美國勞工統計局指數係根據物價404項編製，勃蘭特斯脫指數係根據物價96項編製。後者雖有其特殊之優點，然就測度一般物價變動之標準而言，兩者比較，當以勞工統計局指數更為可信，顧包含物品項數較少之指數，未必皆無價值而可委棄者。密哲爾氏嘗將物價項數多寡懸殊之數種指數，詳加比較。此項研究，足使吾人對於物價制度及指數之性質，可獲明確之認識。密氏試驗結果，發現各種指數變動之相似，頗為意料所不及。根據少數物價編製之指數，所表示物價變動之趨勢，與根據數百項物價編製之指數大致相同；惟若細加觀察，則物價之漲落變動，不無些微差異，而此種差異常足使吾人對於計算時期之物價變動不能確定。遇此情形，則包含品目較多之指數，如物品選擇適當，確能代表物價制度內之各原素時，應為最足代表一般物價之變動者也。

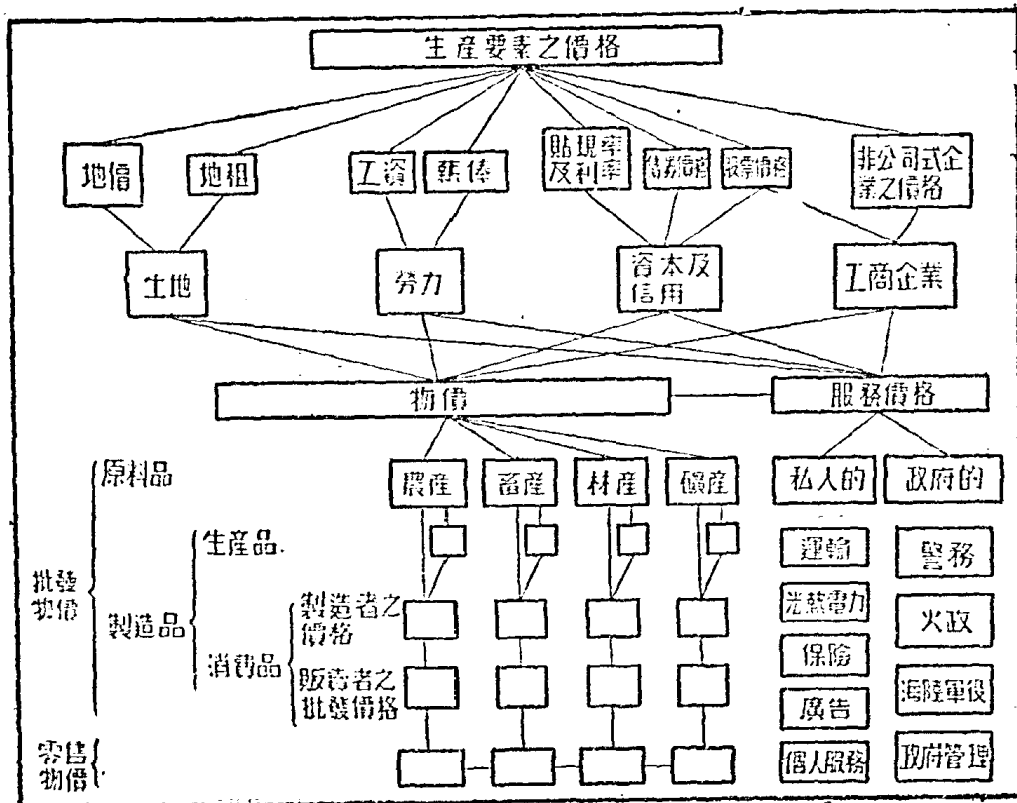
雖然，指數亦有為某種目的，物品項數不妨從少者。例如吾人欲編製一種感應靈敏(“sensitive”)之指數，以預測一般物價變動之趨勢，而非用為測度一般物價水準變動之精確標準者，則選取品目以少為宜。美國紐約聯邦準備銀行所編指數，即為其一例。該項指數最初僅包含基本物品(原料品)12項，現時則係根據22品之價格編製。潘遜氏所編之批發物價指數，包含物品僅10項，其性質與之相似。此種指數即係商情循環之物價指數，其編製目的係為測度商情由盛而衰之過程中之趨勢者。欲使指數感應靈敏，應選取價格變動劇烈之少數物品，包含物品之項數，切不可多，惟此種指數之用途，較為狹隘耳。蓋就一般而論，包含物品項數較多之指數，其感應性雖不免遲鈍，然此乃物價制度所固有之特性，在指數中應忠實表現者也。

討論指數內所應包含物品之項數問題，不能不與物品之性質問題

同時討論之。指數代表性之高下，雖視物價項數之多寡而定，但各項物品之性質若何，尤關重要。在選取物價時所應注意者，為各類物價之動態，各有其特殊之異點。此各類物價，其相互之關係、其動態、其與經濟制度運行之關係、及其與商情盛衰變動之關係等，皆為經濟學家及企業家所極端重視之事實。

物價制度之原素

吾人如欲將物價制度內各原素間之複雜關係，用圖示法表明之，自屬不可能。圖五十三僅係表示此項關係之大概，圖中所列之各原素，亦祇係比較重要者。其重要綱目為各生產要素之價格、批發物價、零售物價及私人因營利而服務與政府服務之價格。私人服務之價格，其一部份係歸納於批發物價項下，另一部份因係為消費者所服之勞役，故歸納於



圖五十三. 物價制度內各原素間關係之圖解

零售物價項下。在每一項內，各包含價格相聯之重要類目。在生產要素價格之第一門內所列地租及地價、俸薪及工資、貼現率及利率、債券及股票價格，在物價高漲或跌落時期，各有其特殊之動態。各種物品之價格，其動態互殊，公私團體服務之價格，其變動亦各有其特殊之點。

物價制度內各單位及各類目相互之關係，欲用圖解以關係綫一一表明之，殊不可能，即屬可能，則每一單位及每一類目皆可繪製無數關係綫，與其他單位及類目相連結，以各單位間莫不均有關聯也。此種關聯或為直接而又重要者，或頗微渺而在事實上難以根究者。

物價制度之用平面圖表示，事實上亦欠恰當，上圖僅能表明一轉瞬間之狀態。物價前期與後期必有關聯，現時與將來亦必有關聯。例如今日之製造品價格，必受有以往之原料品及勞力費用等等之影響也。

凡能準確測度貨幣購買力或一般物價水準變動之指數，應包含物價制度中大小各類目內分別選出之代表物品。現時所發表之指數，尙未有能照此辦理者，每見一種批發物價指數，編製者自信其為一般物價變動之測度標準，實則未必盡然也。

吾人現時所欲討論者，為批發物價指數之編製方法。前述種種，原為解釋此項指數之真實範圍及其意義，並以表示其所測度者，僅係物價制度內最重要之一種原素而已。其次尙須考慮者，則為此種原素內所包含之各各不同而又互相關聯之各類物價，以研究其與編製指數之關係。

批發物價之價格分類

批發物價指數既係根據物價樣本編製，則所選樣本必須確能代表全體物價，必須包含物價制度內各項原素中之典型價格。劃分原素之方法，必須根據各類物品價格變動之特性而定。依此原則分類，其最顯見者，為各類之工業品。如紡織品與鋼鐵之價格，皮革與化學品之價格，係受不同之勢力所支配。各種工業所受商情盛衰之影響，時期既先後，程

度亦有深淺，故編製批發物價指數務必在各種重要工業分類中一一選取物品以代表之。倘指數中某類物品特占優勢，則指數即失去若干之代表性。例如勃蘭特斯脫指數所加於棉品、生熟牛皮、燻腌肉等品之權數，各大於其實際貿易上原有之重要程度，因之非但指數之用途有所限制，且足減損其批發物價指數之代表價值。

各種工業品價格變動不同之範圍，可就一九一三年以來農產品及家用品之批發價格指數加以比較，即可得其大略。

欲使指數具有代表性，則所取樣本不僅須就各類工業中選取適當項數之代表品，即原料與製造品價格之變動亦各不同，故兩類中均須選取物品以代表之。大抵原料之價格，對於商情盛衰之感覺更為敏銳，其變動先於製造品，變動之程度亦較劇烈。此其原因有二：其一，原料之銷售於市場，或供消費，或供製造之用，在商情由衰而盛之時期，競爭之製造商人鑒於消費者之需要增加(或為預測需要之增加)，致使遂不惜重價競購原料，故需要增加之壓力首及於原料市場，而因製造商之競購，遂使原料價格在他種物價受有影響以前，先已騰漲(惟此非不變之定律，在一九一九年商情轉旺期間之物價變動，即為其例外)；反之，商情之恐慌，每始於製造品之積存過剩，故當商情衰落初現端倪之時，原料需要首先激減。簡單純粹之商業勢力，在原料市場之作用，恆較製造品市場為自由，故吾人可由原料價格之變動，以預測其他物價變動之趨勢焉。

製造品價格較為穩定之第二原因，由於製造品價格內所含變動比較穩定之各種成本費用如雜項開支及工資等，恆占較大之成分所致。工資、利率及地租之變動，常較物品價格遲鈍而和緩。物價之包含此種原素者，其價格必較為穩定，故當物品之漸由原料以達製造品之階段，其價格內所包含此種穩定之原素漸多，故其變動亦漸趨穩定(註)。各種製

(註)參看密哲爾氏著“The Making and Using of Index Numbers”一文中所舉之

例，見美國勞工統計局刊物第 284 種第 44—45 頁。

造品價格變動之所以分歧者，即由於此。

在上述各類中，每類又包含小類若干，其價格變動亦互不相同。原料品類中之農產、畜產、林產及礦產之價格變動，常呈顯著之差異。農產品價格之變動，不僅受商情盛衰之影響，且隨氣候及收穫豐歉為轉移，其變動程度雖未見和緩，然其反映商情之功用，較礦產品為低(註一)。畜產及林產之價格變動，反映商情之程度，似介乎農產及礦產之間。在選取原料品作為編製指數之物價樣本時，亦應使此四小類之物品有適當之比例(註二)。

復次，各種製造品之價格變動亦不盡一致，故各種製造品未見能合成一類而不可再分為若干小類者。製造品之中有可供再生產或再加工之用者，其價格變動受製造商出價高下之影響，頗與原料品相似，因之變動亦較劇烈；製造品如僅供消費者之需求，則其價格受商情之影響較淺，變動自較穩定。此與前所述及之理論相吻合，即凡物價內所包含穩定之原素，如工資及雜項開支之成分愈大時，則物價變動愈平穩是也。由此觀之，批發物價指數所根據之物價樣本，應包含製造者之物品、消費者之物品、及製造各階段所需之物品在內，自不待言(註三)。

物價分類雖尚有其他方法，惟依吾人觀察，前述分類最稱重要。欲使批發物價指數具有代表性，應在所有各類中選取物價，而各類之權數須與各該類中全部物品在貿易上所佔地位之輕重成正比例。

(註一) 惟吾人不據以確定農業生產量及農產價格兩者與一般商情毫無關聯。吾人可有充分之理由，足以徵信農業及商業狀況具有直接之因果關係，惟此關係發生之初，常呈現異方向之變動而已。

(註二) 參看 “The Making of Index Numbers” 第47頁。

(註三) 見前書第45及49頁。

美國之批發物價指數(註)

美國勞工統計局指數

美國現編之批發物價指數中，當以勞工統計局 (Bureau of Labor Statistics) 指數之聲譽為最著。此項指數始編於一九〇二年。在一九一三年以前之指數，為比價之未加權平均數，每項比價之基數，為一八九〇年至一八九九年之十年平均價格；迨一九一四年，據該局出版之報告所載指數，其編製方法頗多變更，指數之基期亦同時更換。現時此項指數任何一期之數字(月指數或年指數)，皆係實價之加權綜合數，惟為便於比較起見，此綜合數均已化為一九一三年基期之百分比。

此項指數現時根據物價 404 項計算(每種物品或採用數項物價，其物品等級及調查價格之城邑，詳見該局之報告中。例如棉花一品，係用紐奧雷安斯州米特令及紐約州高地米特令 [Middling New Orleans 及 Middling Upland, New York] 兩項價格)。每品之價格，各與其固定權數即一九一九年之交易數量相乘。

茲以棉花一品為例，說明其計算方法於下：

物 品	1919年平均價 (每 磅)	1919年交易數量	1919年平均價 × 1919年交易數量
棉花，紐奧雷安斯州米特令	\$.319	3,806,921,000磅	\$1,214,407,799
棉花，紐約州高地米特令	.325	1,903,461,000磅	618,624,825

全部 404 項物價皆依上法計算後，將末一欄內各品價值一一相加，即得計算時期(在本例中為一九一九年)之指數。在發表指數時，此綜合

(註)各指數之詳細說明及其比較，具載於美國勞工統計局刊物第 264 種 "Index Numbers of Wholesale Prices in the United States and Foreign Countries"。現時該局出版之各期 "Wholesale Prices"，對於該局批發物價指數之編製內容，均有詳盡之說明。

價值各化爲百分比，而以一九一三年之綜合價值作爲100。該項指數因係採用加權綜合法計算，故可轉換爲任何一年或一月之基期，僅須將求得之百分比，化爲以該年或該月爲基期之百分比可矣。

由此觀之，該項指數實係根據一批貨物之批發價格計算者。各期貨品並無變更，惟因各品價格之變動，總價值乃隨之增減，故指數實係測度個別物價變動所給予總價值之影響之淨結果。全部物品分爲九類，各有其綜合價值，故除包含全部物品之總指數外，尚有各分類指數。物品之分類如下(註)：

農產品	建築材料
食 物	化學品及藥材
衣 着	家用品
燃料及燈光	雜 類
金屬及其製品	

聯邦準備局指數

美國聯邦準備局 (The Federal Reserve Board) 編有美國批發物價指數兩種。其一係根據勞工統計局所調查之物價編製者，惟爲研究物價制度內各原素起見，將物品重行分類。該指數所採用之權數及計算方法，與勞工統計局指數相同。該局編製分類指數之類名如下：

原料品
農 產
畜 產
林 產
礦 產
生產品
消費品

此種分類之用途，最初係由密哲爾氏所闡明，其意義在前數頁內已有解釋。各類指數最初在一九一三年十月聯邦準備局月刊 (“Federal Reserve Bulletin”)內發表，該刊並載有各類之品目表。

(註) 各類所包含之品名表，見美國勞工統計局出版最近一期之“Wholesale Prices”。

聯邦準備局爲欲使國際物價便於比較起見，另編有世界主要商業國之批發物價指數。此項指數之編製，乃鑒於各國所編物價指數，其編製方法既不相同，而所包含之物品種類又復各殊，礙難比較之故。現時該局按期發表之國際物價指數，計有美、加、英、法、日五國。

此項指數之計算方法，與勞工統計局指數所採用者相同，係將每品之價格與其權數相乘，得每品之價值，而後將各品之價值一一相加，得綜合價值。此綜合價值係以百分比表示，而以一九一三年之綜合價值作爲100者。此項指數所以異於勞工統計局指數者，一爲所包含物品項數之多寡不同，一爲所採分類標準與權數之不同。

每一國際物價指數所包含之物價自90項至100項不等，按物品計，約70種。物品分類係採用兩種標準：其第一種標準，前已述及，即係將全部物品分爲原料品、生產品及消費品；其第二種標準，則將全部物品另分爲國產品及輸入品。此外又另編一輸出品指數，該指數內所包含各品，係由上述各類中選取之重要輸出品。前述五國之國際物價指數，皆按此六類發表分類指數及總指數；此外又發表按照各國物價折合爲金物價後計算之金物價指數。

編製前項分類指數所用權數，爲一九一三年之國內生產、輸入或輸出數量。求總指數時，係將國內產品之綜合價值與輸入品之綜合價值兩項相加，而後計算百分比。

此項指數已編製者，爲一九一三年之指數及自一九一九年一月至最近止之月指數（法國指數係自一九二〇年一月起）（註）。

戰時工業局指數

美國所有之批發物價指數中最稱詳備者，當推戰時工業局物價股

（註）關於此項指數之詳細說明，見一九二四年聯邦準備局出版“Prices in the United States and Abroad”。

(the Price Section of War Industries Board)爲研究歐戰期內物價所編之指數。此項研究所包含之時期，係自一九一三年起迄一九一八年止。研究結果詳見戰時工業局出版歐戰期內之物價（“History of Prices During the War”）一書內，其節略具載於該局出版物價叢書第一種。

各年物價計有1474品，皆由該局職員親自搜集，惟其中供編製指數之用者計1366品。物品按性質分類者計有五〇類，按工業分類者計有七類。計算指數除總指數外，並分別計算此五〇類、七大類及許多小類之指數。

此項指數皆係以百分比表示之實價加權綜合數，與勞工統計局指數相似。權數資料係根據一九一七年全年美國生產及輸入之總數量。每類各品之某月價格，先與其權數一一相乘，而後相加，即得該類是月份之綜合價值。當合併各小類之綜合價值以求七大類指數及總指數時，每類各按其重要程度，分別加權，其方法與加權於各個物品時所用者相同。

綜合價值之折合爲百分比，係選取一九一三年七月至一九一四年六月之十二個月作爲基期，因其適當歐戰發生之前一年也，惟此基期可任意轉換而不損及指數之準確性。

除前述各政府機關所編之指數外，尚有私家編製之批發物價指數數種，其最著者爲勃蘭特斯脫及滕氏所編之指數。

勃蘭特斯脫批發物價指數

勃蘭特斯脫 (Bradstreet) 指數係以實價綜合數發表者。此項指數之編製，係將商業上重要物品96種之價格，一律折合爲“每磅”之價格計算，並不採用加權制。其月指數係根據每月一日之價格計算，非若勞工統計局指數之根據每月平均價計算者。又因價格無須折算爲百分比，故無基期，惟引用指數者可根據發表之綜合價格，換算爲以任何一月或任何一年爲基期之百分比。此項指數所包含之時期，係自一八九二年起以

迄於今。

前嘗論及任何指數，無論其為有意的或無意的，均含有加權之意味。勃蘭特斯脫指數雖屬簡單指數，實則對於某數種物品加權特重。據一八九七年七月十日出版之勃蘭特斯脫物價表中，焦煤之每磅價格為\$0.0007，火酒每磅價為\$0.34，羊毛每磅價為\$0.50。在各品之每磅價格相加時，顯見無意中所加於後兩種物品之權數，遠較前一種物品為重。潘遜氏 (Warren M. Persons) 嘗就一九二一年九月一日勃蘭特斯脫指數(註)製成下表，以示各類物品之權數分配。

表五十六

1921年九月一日勃蘭特斯脫指數之權數分配

物品分類	權數(百分分配)
紡織品及其原料：	
棉織品.....	16.7
羊毛.....	5.6
其他.....	3.3
食物：	
牛乳、雞蛋、奶油、乳酪.....	8.6
燻豬肉及牛肉、豬油.....	6.9
其他食物.....	12.4
生熟皮.....	13.3
建築材料.....	1.5
煤及焦煤：	
煤.....	*
焦煤.....	*
金屬品：	
鋼鐵.....	.7
其他金屬品.....	4.3
化學品及藥品.....	9.6
畜產.....	3.6
油.....	4.0
水果.....	3.2
麵包原料.....	1.0
船用品.....	.9
雜項.....	4.2

*不滿 $\frac{1}{10}$ %。

(註)此項指數見“Review of Economic Statistics”初卷第三號第365—366頁。

上述權數之分配，足使棉織品、生熟皮、化學品及藥品、燻醃豬牛肉等，在指數中占有重大勢力，因之指數有偏重於各該項物品之特質；且此項權數並不固定，常因價格變動而有變更。例如一九二〇年二月一日紡織品之權數占總權數之 34.46%，而在十八個月以後，因紡織品價格之跌落遠較一般物價為劇烈，故至一九二一年九月一日紡織品之權數遂減為總權數之 25.58% 矣。

勃蘭特斯脫指數雖含有不合理之加權意義，亦自有其優點，倘為某種用途如供測度商情變化之用者，此為批發物價指數中最適當之一種。

滕氏批發物價指數

滕氏公司 (R. G. Dun & Co.) 之商務代理所發表之批發物價指數，按月具載於滕氏月刊 (“Dun's Review”)。該項指數以實價綜合數發表而無基期，此與勃蘭特斯脫指數相同，惟指數之編製方法，權數之分配，以及包含之時期，則兩種指數迥不相同。

滕氏指數實際祇係在一年內個人所需數項大宗物品之數量，以美金表示之費用總額而已。求月指數時，係先將指數內各品之價格，分別與其每人每年之平均消費量相乘。例如一九二一年一月一日之指數為 \$198.600，此即指數內包含之各品在一年內個人所需數量，按照一九二一年一月一日之價格求得之費用總額也。指數中之價格為每月一日之價格。指數所用之物品約 300 種，詳細品名表及其實際項數均無發表，惟知所包含者皆為生活必需品，奢侈品則概在擯棄之列。權數資料雖係根據每人平均消費量，但指數非為測度生活費用之變動，而為測度一般批發物價水準變動之用。主要物品如生鐵、煤、建築材料等非供個人直接消費之品亦皆採入。此項指數起編於一九一〇年，而計算指數則追溯至一八六〇年。

就理論上言之，滕氏指數所用每項物價按照每人平均消費量加權

之方法，固屬妥善，但此法在實際編製時究竟如何運用，則不無疑問之處，蓋美國現時關於物品之消費額，除少數物品外，尚無適當可靠之資料可供採用也。且就發表之數字觀之，食物之權數幾及總權數之半，按諸事實，似失之過大。滕氏指數之每呈現特殊之變動，而常與其他一般批發物價指數之趨勢未能一致者，亦即由於食物之權數過大之所致。

費暄氏每週批發物價指數

費暄氏(Irving Fisher)嘗於一九二三年一月間編製一種批發物價指數，每週發表一次。物價指數之按週編製而比較詳備者，當以此為嚆矢。

此項指數係以百分比數字發表，而以一九一三年為基期，惟此種百分比係由實價加權綜合數求得。其權數採用固定制，係根據戶口財產普查局(The Census Bureau)所紀載之一九一九年物品成交數量。其計算公式係前述加權綜合式之第一種，即

$$\frac{\sum p_1 q_0}{\sum p_0 q_0}$$

惟此處所用權數為一九一九年之數量，而非基期之數量，其辦法與勞工統計局所採用者相同。費暄氏認為就現有可供採用之資料而論，依此法計算指數所得之結果，最能與“理想”指數相接近。

此項指數所用加權方法，頗能別具心裁。指數係根據滕氏月刊所載物品 200 種之價格編製，其最重要之各品，係取一九一九年之實際數量為權數，惟在確定類權數時稍予變更而已；至於其他次要各品之權數，則為求計算上便利起見，採用下列各項數字：1000, 100, 10, 1, .1, 即以其中與每項物品實際數量之絕對數（或幾何比例）最為相近之一數作為該品之權數。費暄氏採取此項辦法，因鑒於次要各品採用整齊之權數，在指數上所發生可能之差誤，常不及百分之一，無足輕重之故。

費暄氏編製一九二三年每週指數時所用之基期綜合價值，實乃一九二二年十一月之價值，惟此項指數已與勞工統計局指數相接合。其接合方法如下：例如一九二二年十一月之勞工統計局指數為156，計算一九二三年每週指數時，在求得每週綜合價值與基期（一九二二年十一月）綜合價值之比率後，尚須乘以156，以換算為一九一三年之基期。現時該項指數之基期綜合價值已改為一九二三年十一月之價值，惟仍以同樣之接合手續，將指數折算為一九一三年基期之百分比。

每週之指數例在次星期一之日報內發表，故此項指數不特為現有一般批發物價指數中最切實用之一種，復以發表迅速，尤為增色不少。

潘遜氏測度商情循環之物價指數

前述各種指數皆以測度一般物價水準之變動為目的，惟實際僅以表示批發物價為限。此外尚有為特種目的編製之指數，此種特殊目的之指數之編製，頗有日見增多之傾向。在已編之指數中，潘遜氏 (Warren M. Persons) 與考伊爾氏 (Eunice S. Coyle) 合編之“商情循環之物價指數” (commodity price index of business cycles) 為最值注意(註)。

此項指數編製之目的，係為測度一般商情由興旺而入衰落與由衰落而入興旺時期內之商情變動者也；換言之，此項指數係由選配各種物品之批發物價而成之商情循環之指數也。所謂商情循環，實質上原係一種物價現象。商情之變化雖尚有其他重要因素，然此種種因素必先影響於物價而後有所感覺，而商人、受雇者以及消費者所感受商情循環之影響，亦必有見於物價之變動而後發現。各項物價固皆受商情變化之影響，但以證券價格與批發物價所受影響最為直接而亦最劇烈。潘遜教授乃編製一種批發物價指數以反映及測度與商情循環有直接關係之物價變動。因其編製之目的與普通目的 (“general purpose”) 之指數迥不相

(註)參看 “Review of Economic Statistics” 初卷第三號第353—369頁。

同，故所用編製方法亦隨之而異。

在決定此項指數之組織時，潘遜氏嘗將按照各工業分類之物價指數及若干物品之個別價格加以試驗，藉為選取各項物價之標準。凡其變動與一九〇三年至一九一四年間之物價循環變動及一般商情相吻合者，皆在選取之列。最初試驗之結果，發現勞工統計局及勃蘭特期脫所編物價指數及各種非物價數列之循環變動，在上述時期內，大致相似，其波動綫之頂點，皆在一九〇七、一九一〇及一九一二年。嗣又將各指數所共同呈現之波浪式變動作為標準，用以試驗各種物價數列是否與此波動趨勢相符，而將其不相符者擯棄之。

在試驗工業分類之物價指數時，係先按照各分類編製指數，而後用上法試驗。其結果發現礦產、鋼鐵、皮革、化學品及紡織品等類之物價指數，皆呈現上項之循環變動，且其趨勢與商情循環亦屬相似。至於農產、食品製造業以及石、粘土、玻璃製造業各類之物價指數，並無上項循環變動，其趨勢與一般物價及一般商情之變動，亦不相侔。

其次用同一方法，為個別物價變動之試驗。凡各種物價並未呈現循環變動，或其變動無伸縮性者，因其不合指數之編製目的，皆所不取。依此法試驗後，發現在未呈現循環變動之分類指數中，亦含有若干物價，其個別之變動呈現上述循環變動者；又發現個別物品之價格變動，比較分類指數更為劇烈而敏銳，於是決定商情循環之物價指數，按照個別物價，而不按分類指數編製。

最後決定選入指數之物品僅十種，為棉子油、焦煤、鋅塊、生鐵、鐵條、燻醃豬肉、生牛皮、印花布、粗布、絨綫。此數項物品之價格變動甚為劇烈，且於商情變動之感應性亦極敏銳，其所代表之各工業，亦皆佔重要之地位。

此十種物品自一八九〇年以後之價格資料，皆已蒐集，而編成商情循環之指數。所有每月及每年之價格，俱以百分比表示，而以一八九〇

至一八九九年爲基期（基期現已改爲一九一九年）。其指數爲此項比價之未加權幾何平均數。

其他批發物價指數

紐約聯邦準備銀行現編有包含基本物品20種之批發物價指數，其性質略與潘遜氏指數相似，惟其編製目的則未明言爲測度商情循環之變動。實際上此種指數係一種測度基本物品市價變動而感應敏銳之指數。此項指數刊載於紐約聯邦準備銀行出版之月刊“Monthly Review”，係以百分比表示，而以一九一三年爲基期者。指數起編於一九一三年，以迄於今。

安納利士脫（“The Annalist”）之批發食物價格指數，係按週發表者。此項指數所根據者，爲主要食物25種之價格。其計算方法係先將此項價格折算爲百分比，而以一八九〇至一八九九年之平均價作爲基價，而後求其未加權之算術平均數。此項指數之編製，僅爲測度家用食物帳所應包含各種食物之批發價格之變動而已。

此外紐約奇勃孫公司（Thomas Gibson）亦編有批發食物價格指數一種，在該公司出版之每週商情報告內發表。該項指數係根據主要食物22種之價格編製，至於所採何種加權制度，初未明言，僅謂該指數係採用滕氏指數之方法而已。該項指數爲比價之加權平均數，計算每項比價所用之基價，爲一八九〇至一八九九年之平均價，惟所得結果尚須與一常數相乘，使此項指數與滕氏指數相銜接，蓋其目的欲以接續滕氏指數也。

各指數編製方法之差異

吾人如欲將前述各種批發物價指數作長時期之比較（註），其價格

（註）密哲爾氏在其所著“The Making and Using of Index Numbers”一文中，曾將三種主要之美國批發物價指數作長時期之比較。原文載於美國勞工統計局叢書第284種卷一第94—112頁。

變動大體雖極相似，但細加審察，要不無變動互異之處。例如一種指數所表示物價水準之變動，本月較上月增高，而同期之另一種指數並無變動，甚有表示跌落者。此種矛盾情形，在價格變動劇烈時，如一九一六年至一九二一年之時期內，尤所常見。其變動不同之原因，在前此討論指數編製方法時，嘗有述及。任何矛盾之變動，必不外乎下述之原因：

- a. 物品項數多寡之不同。
- b. 採入指數之物品在性質上之不同。
- c. 物價蒐集方法之不同（或為合同價，或為市價；或採用每月一日之價格，或為每月之平均價）。
- d. 權數之不同。
- e. 平均方法之不同。

概略：批發物價指數

茲將各種指數之主要特質及其用途，就其相異各點，概括論述之。

1. 美國現編各種物價指數中，其編製最稱完善者，首推勞工統計局所編之一般批發物價指數。該指數所測度者，非物價變動之平均比率，而為購置一定數量之物品所需費用之增減。
2. 在一九一三年至一九一八年間所編之指數中，以戰時工業局所編之一般批發物價指數最為優良。
3. 滕氏指數測度批發物價之變動，尚屬完美，惜於食物製品加權過重，指數所包含之各品及其實際權數又未予發表，致指數之價值不免減損。
4. 勃蘭特斯脫指數於測度批發物價之變動，亦頗優良，惟於原料品過於重視，頗足減損指數之代表價值。其所含權數輕重之不合理，亦為一大弱點，然因其偏重於原料品之故，以之測驗商情盛

衰，反較勞工統計局及滕氏指數為勝任愉快。

5. 潘遜氏指數非為測度一般物價之變動者。其編製之初，即擬以測驗商情循環為目的，故測度此種現象自頗切當。
6. 紐約聯邦準備銀行所編20種物品之指數，對於測度基本原料之價格變動，為其特長。
7. 聯邦準備局為比較國際物價所編之指數，可使各國之物價變動易於比較，故頗切實用。其所編各國之指數對於測度各該國之一般物價變動，亦頗精確。
8. 費暄氏之每週批發物價指數，理解準確，編製精審，為用甚著。該項指數發表迅速，尤足增進其效用。
9. 安納利斯脫所編之批發食物價格指數，用途較狹。其編製方法因係採用比價之未加權平均數，故所得結果在價值方面每致發生懷疑，且其所採基期為一八九〇年至一九九九年，亦失之過遠。

其他物價指數

應用指數以測度物價之變動，並不以批發物價為限，亦有用以測度他種物價之變動者。此處略將其他指數之性質及其編製方法說明之。

零售物價指數

美國勞工統計局按期發表零售食物價格指數一種。其編製方法與該局用以計算批發物價指數者，大致相同，所稍異者乃由於採用之資料性質不同之故。

編製指數所用食物43種(註)之實際零售價格，例於每月十五日向零售商店直接調查。調查區域包括美國主要城市51處，間有各處習慣不盡相同者，在可能範圍以內，務使各處價格能相互比較。計算指數時，係先

(註)此項指數自一九二一年一月起，係根據食物43種編製，以前係根據食物22項。

求每品價格之總數，再以報告該品價格之商店數除此總數，即得每品之全國平均價格；次將各品平均價，按照該局家計調查所得一九一八年之每家平均消費數量，分別加權，以求實價加權綜合數，其法與編製批發物價指數所用者相同；最後將零售物價之綜合數，以一九一三年為基期，換算為百分比。此項百分比之指數在勞工月刊“Monthly Labor Review”內按期發表。

同時此51城市亦分別編製零售食物價格指數，其編製手續與全國指數相同。全國之平均物價及各城市之物價，均在零售物價刊物內按期發表。此種辦法頗足增進該局工作之實際價值。

編製指數以測度批發物價變動之種種困難，前嘗一再論及，而編製零售物價指數，其困難則猶有過之，蓋除編製批發物價指數所發生之一切理論問題仍須一一解決外，在實際方面，如蒐集適當之權數資料，調查準確之物價數字，以及判別價格之能否相互比較等，在在發生困難。復以物品之未能標準化，及商業習慣與各地風尚之不同，尤足使各地價格不易比較。因此種種原因，現時編行之零售物價指數，尚無一能受人信仰如優良之批發物價指數者。

生活費指數

若謂編製零售物價指數不易，則編製生活費指數之困難，當更倍蓰，蓋欲將家計調查所得每家消費之各項目，如食物價格、房租、零售衣着價格、燃料及燈光費用、零售傢具價格以及其他雜項物品之價格，一一求其平均數，以供編製此項價格指數之用，每發生統計上之種種困難，而須設法解決之；且除關於平均數及加權方法發生學理上之問題外，他如關於蒐集準確詳備之數字，亦必發生更困難之實際問題。在現時所編生活費指數尚未能確切表現其可靠程度以前，應用此項指數者，務宜特別注意。

現時已編之生活費指數有二：一爲勞工統計局所編，一爲屬於雇主公會之全國實業會 (National Conference Industrial Board) 所編，前者在勞工月刊 (“Monthly Labor Review”) 發表，後者在該會出版之刊物內發表。上項指數內所有家庭消費之主要項目，俱按照家用賬簿所示之各品比重程度，分別加權，其綜合價值亦皆以一九一三年爲基期，折算爲比價後發表者(全國實業會指數係以一九一四年七月爲基期)。

農產品之價格指數及其購買力指數

美國農業部編有指數一組，頗切實用。其中一種係爲測度主要農產品農場價格之按月變動者，另一種係爲測度畜產品農場價格之變動者，兩者復合編爲第三種指數，以測度一切農產品之平均價格變動。上述三種指數俱以百分比表示，而以一九一三年爲基期。

該部爲欲使農產品價格指數變動之意義更爲明顯起見，另計算第四種指數，以表示農產品購買力之變動。以貨幣表示之農產物價指數，因貨幣之購買力常在變動之中，其本身無甚意義，故吾人需要一種測度農民所購各種物品價格變動之指數，以供計算農產品購買力指數之用。農業部編製此種指數，係將勞工統計局之批發物價指數，剔除農產品及食物後，改編而成。茲以一九二三年一月份指數爲例，以說明農產品購買力指數之計算方法。按是月農產品價格指數爲 116 (一九一三年 = 100)，剔除農產品及食物後之批發物價指數爲 170。其意謂一九二三年一月農民於出售其產品時所獲得之代價，較一九一三年增 16%，同時一般批發物價較一九一三年增 70%，故農民之購買力，以一九一三年爲基期，應爲 $\frac{116}{170} \times 100$ ，即 68.2% 也。

茲將本節所述各種指數之一九一三年各月數字，列表於下：

表 五 十 七
物 價 指 數

1913年各月農產品價格及其購買力之指數
(一九一三年=100)

月 份	農 場 價 格			躉售物價*	農 產 品 之 購 買 力 §
	農 產 品 (每月十五日)	畜 產 品 (每月十五日)	農 產 品 及 畜 產 品 合 計		
一 月.....	126	106	116	170	68
二 月.....	130	107	118	172	69
三 月.....	134	106	120	175	69
四 月.....	139	107	123	176	70
五 月.....	140	105	123	172	71
六 月.....	139	100	120	168	71
七 月.....	136	102	119	165	72
八 月.....	136	102	119	163	73
九 月.....	138	109	123	164	75
十 月.....	139	103	121	161	75
十一 月.....	137	97	117	160	73
十二 月.....	137	94	116	158	73

*農產品及食物除外。

§以購買農產品及食物以外之各種物品表示者。

編製上項購買力指數時，嘗有一重要之假設，即謂農民所購物品之價格變動，可以校正後之勞工統計局批發物價指數測度之是也。此項假設不無商榷之餘地，蓋一則農民購買物品，通常並不躉批購入；再則該指數所包含者，亦不盡為農民所購買之物品也。雖然，在此尚無比較適當之指數可資採用以前，原係權宜之計，惟其中所含可能之差誤，在引用此項指數時應加注意耳。

前項指數自一九一三年起以迄於今，皆按月編製，在美國農業部出版之農產品及其市場（“Crops and Markets”）內發表。

貨幣工資及真實工資之指數

近年工業糾紛，風起雲湧，工資問題遂為一般人士所密切注意，此雖由於時代關係，然此問題實為一永久性之重要問題。工資指數已由政府機關、雇主公會、工會以及私人編製者，為數頗夥。各指數之優劣，此處不擬評論，本文所欲陳述者，為關於編製及解釋此種指數方面之重要數點而已。

編製貨幣工資指數所涉及之問題，其最要者不過為平均物價之問題，惟此處之物價，係指勞力而非商品耳。平均時所發生之困難，在於紀錄勞力價格方法之不同。吾人計算工資指數時，應按照每小時工資率計算乎？抑按照每週應得之工資總數（即依普通每小時工資率，核算全週應得之工資數）計算乎？抑按照每週實得工資數計算乎？如按每小時工資率計算，則每週實際工作之時間、失業、半失業、加工以及額外獎金等之變動，均未能顧及。如按每週應得工資數計算，則每週工作小時數之增減，雖可反映於其指數，但仍有其他缺點。如按每週實得工資數計算，倘能包括忙閒季節，自屬最切實用，惜此種資料殊不易獲得，因在事實上諸多窒礙也。紐約州勞工局（The New York State Department of Labor），嘗按照每週實得工資數編有此種指數，其工資資料乃得自該州重要各工廠之報告。此項指數係以貨幣單位表示之平均工資數，惟可任擇一基期，換算為百分比表示之指數。

除工資資料外，尚有關於加權及平均方法之問題，此處可無須討論。其更重要者，則有“代表性”（“representativeness”）之一問題，此與編製批發物價指數時所發生選取物品之種類問題相似。如吾人僅取熟練工人之工資，求其平均工資，則其結果不能用作代表普通工資率之指數；如以工頭及薪資甚厚之技術工人之工資，選入普通工資指數，則須各按其應有之權數，分別加權，且須明白解釋其結果。故遇此種情形，當仿照批發物價指數編製分類指數之辦法，分編各級工人之分類指數，並將分類指數用適當權數加權後，彙編全體工人工資之加權平均指數。吾人如將工資分類研究，編成分類指數，則對工資變動之性質，必較各類彙合編成之總指數易於明瞭，此點正與研究物價變動時之情形相似。

編製工資指數因有上述種種不同之點，故現時已編之各種工資指數，其所用編製之方法迥不相同。除其中嘗將編製方法及所用資料之性質作詳盡之說明，可供審擇者外，吾人對於各該種指數所表示工資變動

之經過情形，尙未可貿然置信也。

貨幣工資指數所表示者，必在生活費用無變動時，始有意義。蓋當物價與生活費用增高時，工資指數所表示工資之增高，或屬虛妄；反之，當物價與生活費用跌落時，工資指數所表示工資之跌落，亦不正確，故吾人如欲測度工人實際享受之豐吝，應顧及生活費用之增減變動，將貨幣工資指數折算爲真實工資指數，始有準確之意義。

茲就下表說明真實工資指數之編製手續。表內所列之工資指數，係根據紐約州工人475,000人之平均每週工資額計算，而以一九一四年十二月爲基期者。生活費指數係美國勞工統計局所編包括全國主要城市三十二處之指數，亦以一九一四年爲100(註)。

表 五 十 八
1914—1923年紐約州真實工資指數
(1914年十二月=100)

月 份	紐約工人每週 工資之百分比	生活費指數	真實工資指數 (即貨幣工資之購買力)
1914年十二月	100	100	100
1915年十二月	107	102	105
1916年十二月	123	115	107
1917年十二月	140	138	101
1918年十二月	185	169	109
1919年十二月	209	193	108
1920年十二月	225	195	115
1921年十二月	198	169	117
1922年十二月	210	165	127
1923年十二月	222	168	132

求真實工資指數時，係將工資指數除以同期之生活費指數即得。如僅以工資指數測度工人之收入狀況，所得結果之不正確，自屬顯見。校正以後之工資指數所表示真實工資數雖仍有增加，但遠不若貨幣工資數增加之劇烈。雖然，上例中各項數字，至多亦僅能近乎準確，蓋用以計算真實工資之兩種指數，在編製方面均有發生重大差誤之可能，故由此

(註)本例中紐約州之工資指數，如能以紐約一州之生活費指數校正之，自更適當，惟依此計算所得之結果，相差恐不致過大。

求得之真實工資指數，亦僅能視為測度紐約工人真實工資變動之一種粗略指數而已。

參 考 書

- Bowley, A. L. "Elements of Statistics" (196—213).
- Chaddock, R. E. "Principles and Methods of Statistics" (Chap. X).
- Fisher, Irving. "Revision of the Weekly Index Number". Journal of the American Statistical Association, Sept., 1924.
- Fisher, Irving. "The Making of Index Numbers".
- Flux, A. W. "The Measurement of Price Changes". Journal of the Royal Statistical Society, March, 1921(167—215).
- Kelley, Truman L. "Statistical Method" (331—347).
- Mitchell, W. C. "The Making and Using of Index Numbers". Part I, Bulletin 284, U. S. Bureau of Labor Statistics, Oct., 1921.
- Persons, Warren M. and Coyle, Eunice. "A Commodity Price Index of Business Cycles". Review of Economic Statistics, Prel. Vol. III(353—369).
- Persons, Warren M. "Fisher's Formula for Index Numbers". Review of Economic Statistics, Prel. Vol. III(103—113).
- Walsh, C. M. "The Measurement of General Exchange Value".
- Walsh, C. M. "The Problem of Estimation".
- Young, Allyn A. "The Measurement of Changes in the General Price Level". Quarterly Journal of Economics, Aug., 1921(557—573).
- Young, Allyn A. "Index Numbers" (in Rietz, H. L., "Handbook of Mathematical Statistics", 181—194).

第七章 時間數列之分析：長期趨勢之測度

以上各章均係討論次數數列以及關於整理及表述此項數列之各問題，吾人現時所欲討論者，爲因時間而變動之資料以及分析此項變動之方法。在經濟統計中，時間數列極爲重要，蓋各種經濟及商業資料，如銀行票據交換額、鋼鐵產量、銷貨數量等等大都含有時間性之變動也。時間數列在其他統計方面，則不甚重要，故分析時間數列之方法發展亦較緩，迨近數年統計方法之應用於經濟界方面漸見普遍後，始漸發展耳。

企業家在日常內部管理方面，以及經濟學家分析一般經濟狀況之變動時，每涉及時間數列之問題。例如銷貨、進貨、損益等內部管理方面之問題，股票價格、利率、企業倒閉等一般經濟問題，莫不與時間變動有關。在分析此種數列時，吾人應斷定其變動之性質及程度，並須將週期變動(*periodic fluctuations*)及意外變動(*accidental fluctuations*)之因素，予以隔離(*isolate*)，以便研究。在企業方面，公司經理所欲知者，爲銷貨數量之增減，其變動之時間與原因，以及銷貨數量與生產量之比較；在經濟方面，經濟學家所欲知者，爲各種物價變動之趨勢，以及一般物價水準升降之詳細情形。倘欲決定商業政策或預測未來之經濟情勢，必先研究已往之變動與趨勢，比較互有關聯之各種時間數列之變動以爲根據；他如商業循環之研究，亦必應用此種分析方法而後可。吾人現時所欲討論者，卽爲分析時間數列之方法也。

時間數列之初步整理

時間數列之初步整理，常較次數數列爲簡易，蓋後者須將原來資料編爲次數分配，而前者不論其爲原始資料或次級資料，往往可就原有形式着手分析，而無須另加編制也。惟有數點應加注意者，茲分述之。

資料所載數字之日期，其意義若何，應辨別清楚，並分別註明之。例如月數字之資料，或為每月一日之數字（如勃蘭特斯脫之批發物價指數），或為每月之平均數（如勞工統計局之物價指數），或為每月之總數（如棉花之消費數量）。此種數字又或為逐月之累積數，表示某年內至各該月終之總數，如煤產量之資料往往以累積數表示之。又數字如為每月或每年之平均數，則須查明求得此項平均數之方法。

時間數列中各時期之資料，其性質務必相同，庶可比較。如資料性質不一致(not homogeneous)，則分析所得之結果必生錯誤而不適用。吾人日常所見發表之統計資料，其性質每不一致，其例不勝枚舉。如政府機關或商業團體所發表商品之生產量或消費量，因係彙集各機關各社團之報告編製而成，故其日期、單位及性質往往不盡一致。如物價數列常因物品之單位或品質稍有參差，或調查地點之不同，以致各時期之物價難作比較。如戶口調查每因分類方法之有變更，各期調查所得之數字亦有礙難比較之慨。又如曩時進出口貿易之數字，其編製方法前後各期頗不一致，故比較時發生錯誤，常所難免。又如銷貨員之銷貨數量，或因派往地點有變更而受重大影響。又如美國鋼鐵公司所發表用以表示其債務性質之所謂‘未辦噸量’之數字，其意義亦時有變更：如在一九二二年七月此項噸量係指實際定貨約期交貨之數量，而在一年以前則僅指載於定貨合同，並未限期交貨，而隨時可以取消定貨之總數量。以上所舉為時間數列中所常見之錯誤，故在分析數列之前，必先辨別資料是否準確可靠，性質是否一致，以定取捨焉。

時間數列之圖示

研究時間數列之初步工作，須將資料作圖，俾便觀察，並作進一步之分析，蓋時間數列之特性及其趨勢，每可由圖而得其梗概。數列之資料有可繪於普通算術尺度之圖上者，有可繪於半對數或雙對數尺度之

圖上者。用對數尺度作圖之意義及其功用，前已詳言之矣。尺度之選擇常視資料之性質及研究之目的而定：倘研究之目的在於表示銷貨數額、物價或生鐵產量等數量變動之絕對值，或比較各數列間數量相差之絕對值，則以用普通算術尺度為宜；如研究之目的在於表示數量變動之百分率，或比較相對變動之差異，則以用半對數尺度為宜矣。大體言之，如吾人熟諳半對數尺度之圖解，自以採用該種圖形較為適當。經濟資料之屬於時間數列者，如以時間按照算術尺度，以另一變量按照對數尺度製圖，尤可得明確之意義，而欲比較兩種時間數列之關係，亦以此種半對數圖最為清晰。

時間數列之研究工作，有時在製成統計圖以後已告完成，蓋所有變動之大勢以及季節變動與其他週期性之變動，或各期變動及其趨勢之比較，皆可由圖中測知其大概。惟吾人須知由圖觀察，並不能獲精確之印象，即所得之比較亦僅屬試驗性質；雖然，圖中所示之變動趨勢及各種關係，已較原來數字易於領會，倘吾人僅欲粗知其大概，則此種圖解雖未能予吾人以確切之印象，亦已盡其能事。但若吾人認為獲得此項印象尚未滿足，則非作更精確之測度與更詳密之分析不可矣。以下所述，即為此種測度及分析之方法。

時間數列所受各種勢力之影響

分析時間數列之普通目的，厥為隔離 (isolate) 數列中所受一種或多種勢力之影響。吾人倘欲研究一數列已往之變動情形以預測其未來之變動，或與另一種或多種數列互為比較，則須將數列中所受各種勢力之影響予以隔離，以便研究。惟事實上欲隔離各個勢力必難絕對準確，有時即相當準確之程度亦不易得，但若吾人所用之數列包含長時期之資料，則其所受各種勢力之影響，並非不可測度而得一相當準確之結果者。現時經濟資料之分析，對於此種測度方法尚未能充分利用之。

然則時間數列係受何種勢力之影響乎？按各種勢力之存在與否每隨數列之性質而定，某種勢力或為某種數列所特有，惟大致可分為數種，下文當分別討論之。

長期趨勢

在經濟統計中，各種時間數列大抵含有顯著之趨勢 (trend)。此種趨勢或有一定之方向，或雖方向無定，而增減之比率大致相同者，亦有因新因素之隨時發生，其變動方向與比率時有表現驟然之變更者。例如物品之產量，商店之銷貨額，在一長時期內，其數字之變動，常有逐漸向上之傾向。此外，人口之變動，礦物產量之消長，以及汽車登記輛數之增減等等，皆其例也。在某種經濟現象中，此種傾向亦有屬於負方向，而呈現逐漸向下之傾向者，如前半世紀中美國利率之變動即為其一例。故吾人所謂長期趨勢 (secular trend) (即長時期內之傾向) 實包含逐漸向上或逐漸向下之變動傾向也。

分析時間數列所得某一時間之長期趨勢值 (trend value) 即作為該時期之常態數值 (normal value)。常態數值者，乃謂一數列在剔除長期趨勢以外所受各種意外及複雜勢力之影響後，而在正常狀態下，呈逐漸向上或逐漸向下之傾向時所應得之數值也。常態數值之觀念實為分析經濟資料之基本要點，因求得各期之常態數值後，即可以此項數值作為基數，以斷定其他各種勢力之影響。故斯尼登氏 (Carl Snyder) 嘗有，吾人處置經濟資料，欲求準確而迅捷，除此法以外，實無其他方法可資採用’之語，亦具見此項概念重要之一斑矣。

於此應注意所謂長期趨勢者，乃統計數列所表現平滑而有規則之長期變動也，故若絕對數值或比率之增減變動，倏忽無定，則無長期趨勢之可言。一種數列有因新因素之發生，或舊因素之消滅，亦可釀成偶然的意外變動者，但若將數列所包含之時間分為數階段而分別求其變

動之趨勢，則與長期趨勢之定義所謂在長期內逐漸變動之趨勢相抵觸；故一數列變動之因素，如常有極端變更，則該數列可認為無長期趨勢之存在。

以上所論，非謂任何時間數列皆有一向上或向下之顯著趨勢。例如氣壓之變化，雖經過一相當時間，而常沿一固定之水平綫升降，絕無向上或向下之顯著趨勢者也。

週期變動

吾人觀察時間數列之圖解時，雖可由曲綫之向上或向下傾向，而發現其長期趨勢，但常因其他種種變動之混入，不能僅憑觀察以作精確之斷定。此種種變動或有規則，或無規則，或劇烈，或和緩，或簡單，或複雜，故時間數列內各時期之數值，常為彙合長期趨勢及其他種種勢力所得之淨結果。此種種勢力之存在，足使各時期之數值不循長期趨勢綫而變動，以致該數列所表現之長期趨勢為之晦而不彰，如無此項勢力以擾亂長期趨勢，則該數列必可呈現緩漸與正常之變動。

此種種勢力大致可分為數種。經濟統計之資料，如採用月數字，大抵含有季節變動 (seasonal variations)。例如商品之消費量及生產量、利率、銀行票據交換額、鐵路運貨數量以及其他各種資料，莫不含有週期性之季節變動。此種變動週期性顯著，而週期之長短亦固定不變，常以十二月為一週。此外週期性不甚顯著，而其變動極有規則者，則為循環變動 (cyclical fluctuations)，係因受經濟與商業狀況循環變動之影響而發生者。例如物價、工資、工業生產量、股票交易額以及其他屬於各個企業組織活動方面之各種統計數列，莫不受商業興旺與商業衰落循環變動之影響。循環變動週期之長短或不盡同，但其起伏變動常整齊有序，可供研究。

意外變動

時間數列除受上述各種勢力而表現較有規則之變動外，復受另一種勢力而含有不規則與意外之變動 (random fluctuations)。如三藩市之地震、戰爭、水災、火災以及其他各種意外事件，皆足引起不規則而比較和緩之變動。此種變動雖與季節變動及循環變動有別，但其足使數列中一變量之數值離開其常態值，則一也。由是知各時期之實際數值與常態值相差之數，即為所受此種種擾亂因素彙合而成之淨結果。

時間數列之分析，即在可能範圍內，隔離上述各種勢力之影響也。吾人之問題，或在研究其中單獨一種因素，或在分別研究各種因素。如吾人所用之資料為每年數字，則季節變動之因素，自不存在。以下討論分析方法時所採用之資料均係年數字，故季節問題可暫置勿論。

長期趨勢之測度

下表所列之數字(註)可作為研究上述各種問題之資料。表中所列數值，為求計算手續便利起見，故以十萬萬美金元為單位。

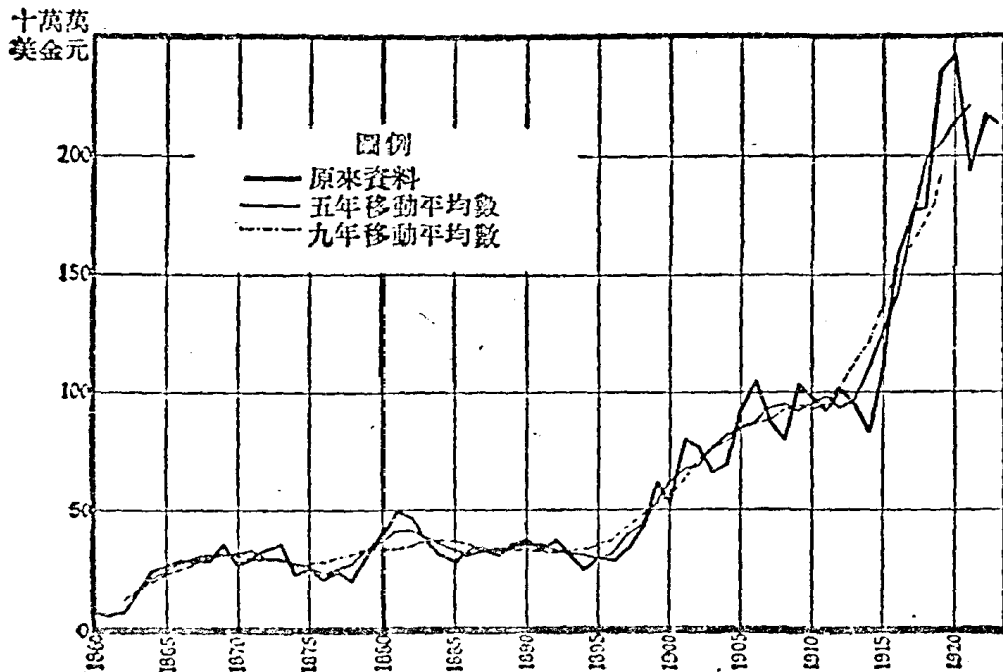
表五十九
1860—1923年紐約市各銀行票據交換額
(單位：十萬萬美金元)

1860.....	7.2	1876.....	\$ 21.6	1892.....	\$ 36.7	1908.....	\$ 79.3
1861.....	5.9	1877.....	23.3	1893.....	31.2	1909.....	103.6
1862.....	6.9	1878.....	19.9	1894.....	21.4	1910.....	97.3
1863.....	14.9	1879.....	29.2	1895.....	29.9	1911.....	92.4
1864.....	24.1	1880.....	38.6	1896.....	28.8	1912.....	100.7
1865.....	26.0	1881.....	49.4	1897.....	33.4	1913.....	94.6
1866.....	28.7	1882.....	46.9	1898.....	42.0	1914.....	83.0
1867.....	28.7	1883.....	37.4	1899.....	60.8	1915.....	110.6
1868.....	28.5	1884.....	31.0	1900.....	52.7	1916.....	159.6
1869.....	37.4	1885.....	28.2	1901.....	79.4	1917.....	177.4
1870.....	27.8	1886.....	33.7	1902.....	76.3	1918.....	178.6
1871.....	29.3	1887.....	33.4	1903.....	66.0	1919.....	235.8
1872.....	33.8	1888.....	31.1	1904.....	68.6	1920.....	243.2
1873.....	35.5	1889.....	35.9	1905.....	93.8	1921.....	194.4
1874.....	22.9	1890.....	37.4	1906.....	101.7	1922.....	217.9
1875.....	25.1	1891.....	33.7	1907.....	87.2	1923.....	214.0

(註)此項數列應先平減價值(deflate)，而後分析。其手續下文將有說明。

表中任何一年之數字，例如一九一二年之交換額 \$100.7 (單位十萬萬美金元)，係彙合長期趨勢、週期變動及意外變動等種種勢力而成之淨結果，此點前已言之矣。吾人當前之問題，即為測定該數列所受第一種勢力即長期趨勢之影響。

圖五十四所示，即一八六〇年至一九二三年紐約市各銀行票據交換額之圖解。此圖表現一顯著之長期趨勢，同時又顯示脫離長期趨勢之其他各種較有規則之變動。測度長期趨勢之方法有數種：一曰移動平均數法 (moving averages)，其法係根據各期數值求移動平均數，以消除



圖五十四. 1860—1923年紐約市各銀行票據交換額及其移動平均數

長期趨勢以外其他各種勢力之影響，而使求得之各期數值表示逐漸增高或逐漸減低之長期變動；二曰配合曲綫法，此法係假定時間與另一變量有函數關係之存在，而以另一變量作為時間之函數，應用數學方法以求與資料最相配稱之曲綫，即以此綫代表其長期趨勢；三曰隨手畫法，其法係憑觀察之經驗，就圖中表示原來資料之綫，詳察其變動之趨勢，隨手畫一平滑之趨勢綫。三者之中，後者最無標準，而含有試驗與揣測之性質。更有以一種統計數列作為另一種數列之基綫或趨勢綫之方法

者，惟此法必須在後種數列之資料性質一致時，始可採用之。

移動平均數法

用移動平均數以決定長期趨勢之方法，係在全時期內，求若干年（或若干月或若干週）之移動平均數，作為中間一年之趨勢值（trend value）或常態值（normal value）。下表所列為一九〇〇年至一九二三年紐約市各銀行票據交換額之三年、五年、七年及九年移動平均數。

表 六 十

1900—1923年紐約市各銀行票據交換額之三年、五年、七年及九年之移動平均數
(單位：十萬萬美金元)

年 份	原來資料	三年移動平均數	五年移動平均數	七年移動平均數	九年移動平均數
1900	\$ 52.7				
1901	79.4	\$ 69.5			
1902	76.3	73.9	\$ 68.6		
1903	66.0	70.3	76.8	\$ 77.4	
1904	68.6	76.1	81.9	82.3	\$ 78.7
1905	93.8	89.0	84.1	82.3	84.3
1906	104.7	95.2	86.7	86.2	86.3
1907	87.2	90.4	93.7	90.6	88.1
1908	79.3	90.0	94.4	94.0	92.0
1909	103.6	93.4	92.0	95.0	94.8
1910	97.3	97.8	94.7	93.6	93.6
1911	92.4	96.8	97.7	93.0	94.3
1912	100.7	95.9	93.6	97.5	102.3
1913	94.6	92.8	96.3	105.5	113.2
1914	83.0	96.1	109.7	116.9	121.6
1915	110.6	117.7	125.0	129.2	137.0
1916	159.6	149.2	141.8	148.5	153.7
1917	177.4	171.9	172.4	169.7	164.1
1918	178.6	197.3	198.9	185.7	177.8
1919	235.8	219.2	205.9	201.0	192.4
1920	242.2	224.5	214.0	208.8	
1921	194.4	218.5	221.1		
1922	217.9	208.8			
1923	214.0				

表內一九〇四年之三年移動平均數為自一九〇三年至一九〇五年三年間之平均數，其五年之移動平均數為自一九〇二至一九〇六年五年間之平均數，其餘各年之移動平均數亦用同法計算。每一平均數均集

中於時間之中點，而以平均數代表該點之常態數值。移動平均數之年數如為奇數，則集中手續 (centering process) 較易，即以移動平均數作為中間一年之常態值；惟所定年數不必定為奇數，如年數為偶數時，則於求得各年移動平均數後，再求其兩年移動平均數，而使此項平均數集中於一年之中間。表中五年與九年之移動平均數及原來資料均繪於圖五十四內。

移動平均數之特質

設有一數列，其所包含之長期趨勢成一水平綫，或斜度不變而成一傾斜之直綫，而其所包含之循環變動，其週期之長短 (length of cycle) 有定，每一週期內起伏變動之幅度 (amplitude) 亦有一定時，則用循環時期之年數或其倍數以求移動平均數，可得一直綫，恰與原來之長期趨勢綫相合。在同一情形之下，如計算移動平均數之年數，較週期之年數或多或少時，則移動平均數所得之綫並非直綫，而為表現另一種循環變動之曲綫。此新循環週期之長短與原來週期相同，惟其變動之幅度較小，且該綫所示循環之最高點及其最低點之位置亦未必與原來循環相吻合。大體言之，移動平均數所用之時期愈長，則由此所得新循環之幅度亦愈小(註)。

表六十一所列，為任意假設之數字，用以引證前述各點者。表內(1)行內每五個數字成一週期之資料，其長期趨勢為一水平綫。

由表觀之，(2)(3)兩行內所列之移動平均數均成一水平綫，而已將循環變動完全消除；但當平均數之項數與循環週期之長短或其倍數不相等時，則循環變動仍不能消除，如(4)(5)兩行所示。

如一數列之長期趨勢綫為一傾斜之直綫，而其循環變動有一固定之週期，及一固定之幅度，則用移動平均數所得之結果正與上例相同。

(註)幅度之減小並無一定規則，惟循環變動則仍存在。

表六十一
說明移動平均數之應用

(1) 循環資料	(2) 五項移動平均數	(3) 十項移動平均數 (集中數字)	(4) 三項移動平均數	(5) 八項移動平均數 (集中數字)
2				
6			$5\frac{1}{3}$	
8	$6\frac{1}{5}$		8	
10	$6\frac{1}{5}$		$7\frac{2}{3}$	
5	$6\frac{1}{5}$		$5\frac{2}{3}$	$6\frac{3}{8}$
2	$6\frac{1}{5}$	$6\frac{1}{5}$	$4\frac{1}{3}$	$6\frac{1}{16}$
6	$6\frac{1}{5}$	$6\frac{1}{5}$	$5\frac{1}{3}$	$6\frac{3}{8}$
8	$6\frac{1}{5}$	$6\frac{1}{5}$	8	$5\frac{6}{8}$
10	$6\frac{1}{5}$	$6\frac{1}{5}$	$7\frac{2}{3}$	$5\frac{11}{16}$
5	$6\frac{1}{5}$	$6\frac{1}{5}$	$5\frac{2}{3}$	$6\frac{3}{8}$
2	$6\frac{1}{5}$	$6\frac{1}{5}$	$4\frac{1}{3}$	$6\frac{1}{16}$
6	$6\frac{1}{5}$	$6\frac{1}{5}$	$5\frac{1}{3}$	$6\frac{3}{8}$
8	$6\frac{1}{5}$	$6\frac{1}{5}$	8	$5\frac{6}{8}$
10	$6\frac{1}{5}$	$6\frac{1}{5}$	$7\frac{2}{3}$	$5\frac{11}{16}$
5	$6\frac{1}{5}$	6	$5\frac{2}{3}$	$6\frac{3}{8}$
2	$6\frac{1}{5}$		$4\frac{1}{3}$	$6\frac{1}{16}$
6	$6\frac{1}{5}$		$5\frac{1}{3}$	
8	$6\frac{1}{5}$		8	
10			$7\frac{2}{3}$	
5				

[(3)(5)兩行所列平均數均係集中數字，即在求得十項或八項之移動平均數後，再求兩項移動平均數集中而得者。]

表六十二所列(1)行內之數字，係由上表內之資料以一定增量(constant increment) + 3 遞加而得者。此與將上項循環變動之數字加於斜度為 +3 之直線上時所得之結果相同。

表六十二內數字如採用與循環週期相等或與其倍數相等之項數求移動平均數，則所得之長期趨勢綫，為一不含循環變動，斜度為 +3 之直綫；但若所用項數與週期或其倍數不相等時，則循環變動依然存在，週期與前相同，惟其幅度較小。此可於(4)(5)兩行之數字見之。

上述兩例所假設之數字，係較為單純而屬於理想之資料，蓋謂循環變動之週期及幅度皆有一定也。倘此簡單情形並不存在，則準確可信之

移動平均數即難揣測，而其解釋亦較困難。當循環變動之週期無定時，欲決定移動平均數之年數，殊為困難，惟就大體而論，此時似以選取與各週期平均年數相等或較多之年數，更為相宜；否則如所選年數較少，所得移動平均數仍必含有顯著之循環變動，故選較長之時期者，所以消除循環變動之勢力也。

表 六 十 二

移動平均數應用於直綫趨勢之數列

(1) 循環資料	(2) 五項移動平均數	(3) 十項移動平均數 (集中數字)	(4) 三項移動平均數	(5) 八項移動平均數 (集中數字)
2				
9			$8\frac{1}{3}$	
14	$12\frac{1}{5}$		14	
19	$15\frac{1}{5}$		$16\frac{2}{3}$	
17	$18\frac{1}{5}$		$17\frac{2}{3}$	$18\frac{8}{8}$
17	$21\frac{1}{5}$	$21\frac{1}{5}$	$19\frac{1}{3}$	$21\frac{13}{6}$
24	$24\frac{1}{5}$	$24\frac{1}{5}$	$23\frac{1}{3}$	$24\frac{6}{8}$
29	$27\frac{1}{5}$	$27\frac{1}{5}$	29	$26\frac{6}{8}$
34	$30\frac{1}{5}$	$30\frac{1}{5}$	$31\frac{2}{3}$	$29\frac{11}{6}$
32	$33\frac{1}{5}$	$33\frac{1}{5}$	$32\frac{2}{3}$	$33\frac{3}{8}$
32	$36\frac{1}{5}$	$36\frac{1}{5}$	$34\frac{1}{3}$	$36\frac{13}{6}$
39	$39\frac{1}{5}$	$39\frac{1}{5}$	$38\frac{1}{3}$	$39\frac{3}{8}$
44	$42\frac{1}{5}$	$42\frac{1}{5}$	44	$41\frac{6}{8}$
49	$45\frac{1}{5}$	$45\frac{1}{5}$	$46\frac{2}{3}$	$44\frac{11}{6}$
47	$48\frac{1}{5}$	$48\frac{1}{5}$	$47\frac{2}{3}$	$48\frac{3}{8}$
47	$51\frac{1}{5}$		$49\frac{1}{3}$	$51\frac{13}{6}$
54	$54\frac{1}{5}$		$53\frac{1}{3}$	
59	$57\frac{1}{5}$		59	
64			$61\frac{2}{3}$	
62				

[(3)(5)兩行所列平均數均係集中數字，即在求得十項或八項之移動平均數後，再求兩項移動平均數集中而得者]。

又設循環變動之週期有定，而各週期內變動之幅度不等時，如採用與週期相等之年數計算移動平均數，則所得之長期趨勢綫，仍含有較小之循環變動。此種次級循環 (secondary cycles) 變動之幅度，在移動平均數之年數與週期相等或為其倍數時為最小。又若循環變動之週期無

定，而各週期內變動之幅度又不相同時，則移動平均數之年數，以選取與各週期之平均年數或其倍數相等者為宜，因由此所得之移動平均數用以表示長期趨勢，最為恰當也。

長期趨勢之形態，原不盡為直綫，如為曲綫，則問題更見複雜。倘長期趨勢綫呈向上內凹狀 (concave upward)，則由移動平均法求得之數值必大於實際趨勢值；倘長期趨勢綫呈向上外凸狀 (convex upward)，則由移動平均法求得之數值必小於實際趨勢值。

表六十三
移動平均數應用於非直綫趨勢之數列(遞加增加率)

(1)	(2)	(3)	(4)	(5)	(6)
x	x^2	循環資料	(2)+(3)	五項移動平均數	實際趨勢值 ($x^2+6.2$)
0	0	2	2		
1	1	6	7		
2	4	8	12	12.2	10.2
3	9	10	19	17.2	15.2
4	16	5	21	24.2	22.2
5	25	2	27	33.2	31.2
6	36	6	42	44.2	42.2
7	49	8	57	57.2	55.2
8	64	10	74	72.2	70.2
9	81	5	86	89.2	87.2
10	100	2	102	108.3	106.2
11	121	6	127	129.2	127.2
12	144	8	152	152.2	150.2
13	169	10	179	177.2	175.2
14	196	5	201	204.2	202.2
15	225	2	227	233.2	231.2
16	256	6	262	264.2	262.2
17	289	8	297	297.2	295.2
18	324	10	334		
19	361	5	366		

此種情形可舉例說明之。表六十三所載之數字，其趨勢綫呈向上內凹狀，此綫係以週期有定幅度相同之循環變動數字，加於向上內凹狀之長期趨勢綫而成。其長期趨勢係依遞加之增加率增加 (increasing at a constantly increasing rate)，故成向上內凹狀。如求得之移動平均數已將循環變動之影響完全消除，則所得之各項數值應為循環資料五

項平均數($6\frac{1}{5}$)與函數關係 $y=x^2$ 內 y 數值兩項之和。

然由表六十三觀之,移動平均數之數值,常大於實際趨勢值。凡用移動平均數以求此種數列之長期趨勢綫,皆不免發生此項錯誤。

表 六 十 四
移動平均數應用於非直綫趨勢之數列(遞減增加率)

(1)	(2)	(3)	(4)	(5)	(6)
x	\sqrt{x}	循環資料	(2)+(3)	五項移動平均數	實際趨勢值 ($\sqrt{x+6.2}$)
0	0	2	2.00		
1	1.00	6	7.00		
2	1.41	8	9.41	7.428	7.61
3	1.73	10	11.73	7.876	7.93
4	2.00	5	7.00	8.166	8.20
5	2.24	2	4.24	8.414	8.44
6	2.45	6	8.45	8.634	8.65
7	2.65	8	10.65	8.834	8.85
8	2.83	10	12.83	9.018	9.03
9	3.00	5	8.00	9.192	9.20
10	3.16	2	5.16	9.354	9.36
11	3.32	6	9.32	9.510	9.52
12	3.46	8	11.46	9.658	9.66
13	3.61	10	13.61	9.800	9.81
14	3.74	5	8.74	9.936	9.94
15	3.87	2	5.87	10.068	10.07
16	4.00	6	10.00	10.194	10.20
17	4.12	8	12.12	10.318	10.32
18	4.24	10	14.24		
19	4.36	5	9.36		

表六十四所載之數字,其趨勢綫呈向上外凸狀,此綫係以同一循環變動之數字,加於向上外凸狀之長期趨勢綫而成。其長期趨勢係依遞減之增加率增加 (increasing at a constantly decreasing rate), 故成向上外凸狀。此時如採用優良方法以求長期趨勢,而能完全消除循環變動之影響,則所得之各項數值應為循環資料五項平均數($6\frac{1}{5}$)與函數關係 $y=\sqrt{x}$ 內 y 數值兩項之和。

但由表六十四觀之,移動平均數之數值常小於實際趨勢值。此兩項數值之差異,在 x 數值較小時,尤為顯著,蓋 x 數值小,則受遞減增加率之影響較大也。

吾人於此可得如下之結論，即在趨勢綫為直綫時，移動平均數之年數至少應等於週期之年數，而在數列之變動無甚規則時，尤以採用週期年數之倍數為宜，蓋移動平均數之時期愈長，則平均數之數值亦愈穩定；但若趨勢綫之形式非直綫，而呈向上內凹狀或向上外凸狀，則移動平均數之時期愈長，發生之差誤亦愈大，故用移動平均數法以測度此種數列之長期趨勢時，所取時期以短為佳，而應等於各個週期之平均年數。

以上所論，均係根據假設之資料，討論移動平均數應用之方法，然實際上一個數列之變動常受種種複雜情形之影響，故欲決定移動平均數之年數，並不如前述之簡單。如一數列循環變動之週期與其幅度均不固定時，移動平均數之時期自以稍長者為佳，但一般數列之趨勢往往為非直綫形，則平均數之時期又以短者為宜。在實際應用上所當考慮者，移動平均數之數值絕不能包括全時期，平均數之時期愈短，則遺漏之年數亦愈少。吾人必先參酌資料之性質，並注意前述各點，然後始可擇定平均數之時期。

移動平均數法之應用，並不限於測度長期趨勢，而亦可用以消除不規則變動之影響，此時移動平均數之時期以選擇小於循環週期之年數者為宜。

吾人現時仍以紐約市各銀行票據交換額為例，以討論移動平均數之年數問題。由圖五十四內觀之，兩種移動平均數所繪成之曲綫頗多相異之處。五年移動平均數之綫與原來資料之綫最為接近；九年移動平均數綫雖最平滑，但與原來資料相差亦最鉅，尤以一八九三年至一八九八年及一九一一年至一九一五年兩時期為尤甚，因在此兩時期內票據交換額變動最為劇烈也。除此兩時期外，似以九年移動平均數最能代表其長期趨勢。

在確定移動平均數之年數時，吾人可參酌該時期內商情變遷之狀況以為準繩。紐約市各銀行票據交換額為一感應敏銳之一般商情指標，

對於投機交易及實業狀況之變遷，均具有特別靈敏之感應性，故在該時期內，無論大小之商情循環變動，一一反映於該數列。在一八七〇年至一九二〇年之以往之半世紀中，吾人已知所經歷商業循環變動之週期數，故可用此實例以為根據，以斷定以上所舉兩種移動平均數中，究以何種最能代表該時期內銀行票據交換額之長期趨勢，而用以測度循環變動之程度。惟依此研究，乃就已知之結論，作步驟相反之推考，要為平時工作所罕見耳。

假定以商情之由衰落以迄恢復常態之前一年為每一週期之起點，則在該全時期內，可得以下之各週期(註)：

1870—1878年	1904—1908年
1879—1885年	1908—1911年
1886—1897年	1911—1914年
1898—1904年	1915— 年

吾人現時可計算數列內各年數值對於其同一年份移動平均數之差，以試求循環變動之週期數。由三年移動平均數求得循環變動之週期數，自嫌過多，蓋該項變動實乃種種意外而又細微之變動，不能視為循環變動。由數列各項對於其七年及九年移動平均數之差，可得下列循環變動之各週期：

由七年移動平均數所得循環變動之各週期	由九年移動平均數所得循環變動之各週期
1871—1878年	1871—1878年
1879—1884年	1879—1887年
1885—1888年	1888—1897年
1888—1897年	1898—1900年
1898—1900年	1900—1903年
1900—1903年	1901—1907年
1904—1907年	1908—1914年
1908—1911年	1915— 年
1911—1914年	
1915—1918年	

(註)此項循環變動之週期，係根據 W. F. Ogburn 氏暨 Dorothy Thomas 氏所編之指數而得。參看一九二二年九月出版 “Quarterly Publications of the American Statistical Association” 第324—340頁。

上列兩種移動平均數所得循環變動之週期，可與前此所列之週期相比較，其中不同之處，一部分自由於所用資料迥異之故，惟有若干不同之點頗堪注意。例如由七年移動平均數所得之週期數較其餘兩種各多三週期。如七年移動平均數內一八八五年至一八八八年及一九一五年至一九一八年之兩週期均為其餘兩種所無，後者在此兩期內僅見微小之波動。一八九八年至一九〇〇年之週期，雖同見於兩種移動平均數，而為用作比較根據之指標所無。一九一一年至一九一四年之週期雖同見於指標，而不見於九年移動平均數。歐戰期間票據交換額劇增，足使九年移動平均數受其影響，以致最後一週期因平均作用而為之消滅矣。

由上推知，欲由票據交換額之資料，求其長期趨勢，則移動平均數之時期，不可少於七年。九年移動平均數較為合用，惟在期末交換額之增加率劇變時則為例外。七年之時期似覺過短，因其移動平均數過於偏重微小變動也。

就一般而論，移動平均數有伸縮性 (flexibility)，此為其優點。用數學方法配合曲綫以代表長期趨勢，常須將全時期分作數階段，而各配一曲綫。此在情形發生變化，或數列之增加率或減少率劇變時，用數學方法配合長期趨勢綫，則不能不將時期分為數段。此時倘用移動平均數，則因其具有伸縮性，而能迎合變遷之情況，不令趨勢猝然中斷，常較計算費時之數學趨勢綫尤為適用，此即移動平均數法之優點也。

除前述之簡單移動平均數外，又有所謂加權移動平均數者。加權方法，乃在計算每一移動平均數時，將各年數值分別加權，其數值愈在中間者，其所加權數亦愈大。此種加權移動平均數稱為遞進平均數 (progressive mean)，係以兩項展開式中各項之係數用作移動平均數各項之權數。如上例採用五項移動平均數時，其由第一項至第五項之權數，應依次為 1, 4, 6, 4, 1。遞進平均數雖可用於某種數列之分析，但欲用以消除時間數列中之週期變動或不規則變動，似不適宜耳。

配合數學曲綫法

時間數列之資料中，頗有可用數學曲綫以代表其長期趨勢者。倘數列中各數值之增加(或減少)爲一固定之增量(increments)(或減量decrements)，則用一直綫表示其長期趨勢，頗爲恰當；倘各數值之增加爲一固定之百分率，如依照複利法則，本利和之金額係按固定之利率而增加，則可用一數學曲綫以表示其長期趨勢。經濟統計之資料，頗有遵循一定法則而增減者，故此種資料之分析與解釋以及預測未來之趨勢，常可求得一數學方程式以表示該項法則，作爲根據。任何資料雖常有不依一定法則，而在長期趨勢綫之上下變動者，但在此種變動之中，仍有規則存在，故方程式之價值，並不因此而減損也。

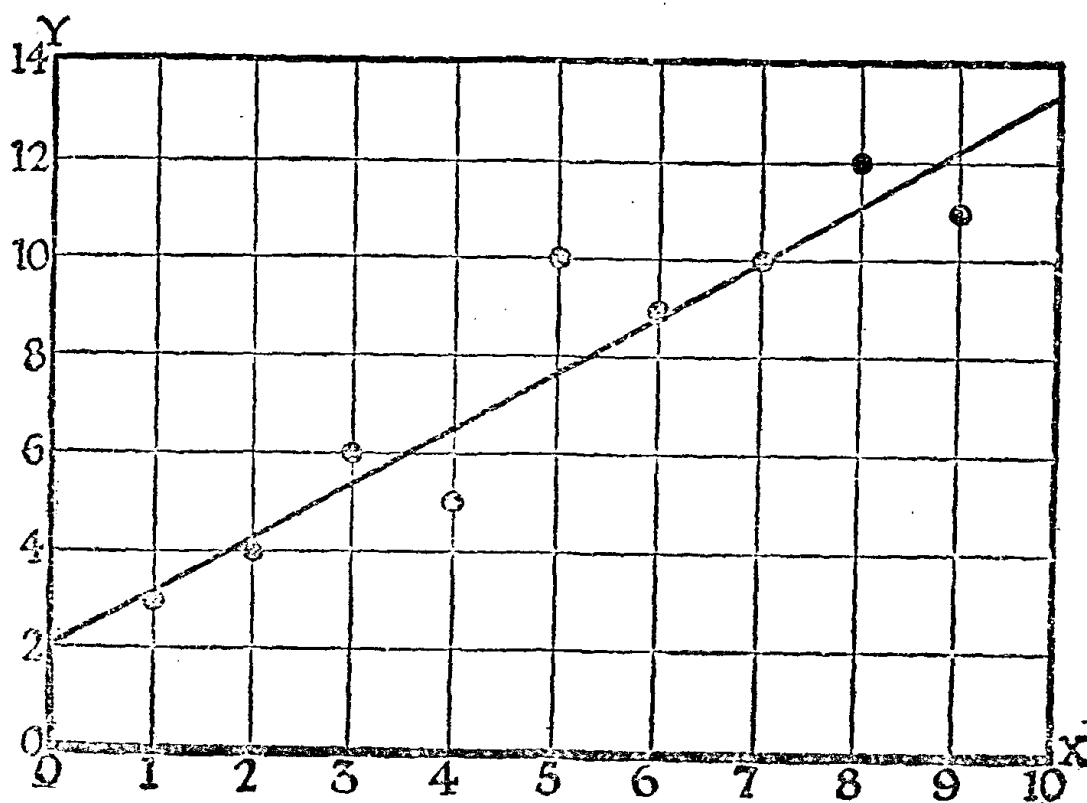
數學曲綫及移動平均數雖皆用以代表長期趨勢，但有一根本不同之點，應予注意。移動平均數並不假定資料之必遵循一定法則，而全以實際資料爲根據，如長期趨勢有所變更，則移動平均數亦必隨此新趨勢而變動，故此平均數能適應情況之變遷，而伸縮自如，其目的僅在表示數列變動之大體趨向。數學曲綫配合之目的，雖亦僅以表示其大體趨勢爲限，但其意義稍有不同。此法係假定數列之各種變動，在資料之表面上，無論其爲意外變動或他種變動，均遵循一定之法則。其所假定者固屬不能證實之經驗法則(empirical law)，但經數學方程式表示後，則成一固定劃一之趨勢綫。故欲使此種假定確實而有效，則在此法則所存在之時期內，一切經濟狀況務必相當穩定，而無劇烈變化足以影響研究之時間數列而後可。設吾人已求得一黃金產量趨勢綫之方程式，而適在此時採礦方法根本變更，則黃金產量之趨勢亦必立受鉅大之影響，已得之方程式即不能適用，蓋在採礦方法變更前後之兩時期內，因資料性質並不一致，自不當用同一之曲綫方程式，以表示兩時期內之長期趨勢也。

數學曲綫之形式甚多，故在實際工作時，欲測定長期趨勢，首須擇

定一相當之曲綫形式。此雖屬初步工作，而實亦最難者，因選擇時須憑個人之判斷力，絕無一定標準可循。關於此點，當俟各種主要曲綫之特性及其配合方法討論後，再行論述。茲先就第二章內所載各種曲綫或其類似之形式中擇定一種，以研究其配合方法。

直綫之配合：最小平方法

吾人將資料作圖後，如見圖中各點之趨勢能用直綫表示者，則配合之工作僅在決定 $y = a + bx$ 方程式中之 a, b 兩常數。所求 a, b 之數值，應以由此求得之直綫最能適合資料之趨勢者為限。茲可舉一簡單之實例，以說明求 a, b 之方法。圖五十五中共有九點，即 1, 3; 2, 4; 3, 6; 4, 5; 5, 10; 6, 9; 7, 10; 8, 12; 9, 11 是也。試就圖中各點配合一直綫。



圖五十五. 在九點上配合直綫之圖解

此處若用觀察法亦可決定 a, b 之近似值。如手執一綫之兩端，在各點左右移動，使此綫愈能適合各點之趨勢愈佳，即可由此求得一最適當

之位置，然後決定此綫之斜度及其 y 截部，則該綫方程式亦可確定矣。惟此法漫無標準，直綫之位置隨人而異，因之所得方程式亦無一定，而在理論上，圖中各點僅有一綫最為適合。此綫常數之值，可用最小平方法求之。

最小平方法之定理，此處不必詳細討論，可簡述於下：設有一數量經測量數次後，得若干不同之觀察值 (observation value)，而欲求此量最可能之數值 (the most probable value)。吾人可證明該數值應為剩餘數 (residual) 平方之和為最小時之值 (剩餘數係指某一數值與各觀察值之差)，在理論上即為各觀察值之算術平均數也。如有多人測量一固定之距離，則各人所量得之結果未必一致，但最可能之數值，應為各人所量尺度之算術平均數。計算算術平均數之方法包含下列步驟，述之於此，以便解釋。今所求者為上項距離最可能之數值，其形式應為

$$M = (\text{一常數})$$

如測量所得 M 之值共有三項如下：

$$M = 5672 \text{ 呎}$$

$$M = 5671 \text{ 呎}$$

$$M = 5676 \text{ 呎}$$

各項相加，得

$$\frac{3M = 17019 \text{ 呎}}{3}$$

此式中之未知數僅有一個 M ，故其值可逕由上式求得

$$M = 5673 \text{ 呎}$$

此值即為與各觀察值相差平方之和為最小時之值。

當測度兩變量間之關係時，其問題與上相似。吾人此時之目的，在於根據圖中各點之位置，求得一直綫方程式，以表示此種關係，惟由此所得之直綫不一而足，因之在其各個方程式中之常數，亦不相同，換言之，即各點並不在同一直綫之上也。各點既不在同一直綫上，然則以何綫為最合於各點分佈之情勢，亦即各方程式中之常數以何者為最可能之值乎？此其答案與測量一單獨數量時正復相同。直綫方程式之常數應

使該直綫繪於圖中時，其與各點相距平方之和為最小數。假定由觀察所得之每對變量均係表示兩變量真實關係之近似程度，吾人欲由此確定一最可能之相關程度，而以直綫表示之，則該綫正為用最小平方法求得之綫，蓋該綫與各點相距平方之和較其他任何直綫為最小也(註)。

吾人於此有 x 與 y 之九對數值，將各對數值代入直綫之普遍方程式 $y = a + bx$ 中，可得下列九個觀察方程式(observation equations)：

$$3 = a + 1b$$

$$4 = a + 2b$$

$$6 = a + 3b$$

$$5 = a + 4b$$

$$10 = a + 5b$$

$$9 = a + 6b$$

$$10 = a + 7b$$

$$12 = a + 8b$$

$$11 = a + 9b$$

上列九式中，任何兩式均可應用聯立方程式之解法，得 a, b 之數值，但每對方程式所得 a, b 之數值，不能適合其他各個方程式，故須聯合此九個觀察方程式以求兩個常態方程式(normal equations)，更由此兩式以求 a, b 之最可能值，代入直綫之普遍方程式中，即得所求之直綫。以每個觀察方程式中 a 之係數乘該式中各項，由此求得之九個方程式一一相加，即得第一個常態方程式。此處各個觀察方程式中之係數均為1，相乘後各式不變，故即可用此原來方程式相加。次以每個觀察方程式中 b 之係數乘該式中各項，由此求得之九個方程式一一相加，即得第二個常態方程式。此處各個觀察方程式中，第一式 b 之係數為1，故以1乘該式中各項；第二式 b 之係數為2，故以2乘該式中各項；餘類推。茲將本例中求常態方程式之計算手續舉示於下：

(註)關於最小平方法之詳細討論及其計算方面校核之方法，可參考附錄A。

表 六 十 五

由觀察方程式求常態方程式之方法

$3 = a + 1b$	$3 = 1a + 1b$
$4 = a + 2b$	$8 = 2a + 4b$
$6 = a + 3b$	$18 = 3a + 9b$
$5 = a + 4b$	$20 = 4a + 16b$
$10 = a + 5b$	$50 = 5a + 25b$
$9 = a + 6b$	$54 = 6a + 36b$
$10 = a + 7b$	$70 = 7a + 49b$
$12 = a + 8b$	$96 = 8a + 64b$
$11 = a + 9b$	$99 = 9a + 81b$
$70 = 9a + 45b$	$418 = 45a + 285b$

由此求得之兩個常態方程式爲

$$70 = 9a + 45b$$

$$418 = 45a + 285b$$

至此僅須求兩式中 a, b 之數值。以5乘第一式各項，由第二式中減去之，即可消除 a 而得 b 之數值爲 $\frac{68}{60}$ 即 1.133 ；再將此值代入兩式中任何一式，即得 a 之數值爲2.111，故本例中所求之直綫方程式爲

$$y = 2.111 + 1.133x$$

實際計算時，無須將所乘之方程式一一紀錄而後相加之，僅須將 x 及 y 之相當數值代入下列兩個典型方程式 (type equations) 即可(註)。

$$\Sigma(y) = na + b\Sigma(x)$$

$$\Sigma(xy) = a\Sigma(x) + b\Sigma(x^2)$$

式中所用各符號之意義，解釋於下：

$\Sigma(y)$: y 各數值之和。

$\Sigma(x)$: x 各數值之和。

$\Sigma(xy)$: x 與 y 之每對數值相乘之積之和。

(註)常態方程式作成之法，見附錄A。

$\Sigma(x^2)$: x 各數值平方之和。

n : x 與 y 數值相配之對數,亦即圖上之點數。

應用下列表式,則計算手續可較省便。

表六十六
配合直綫所需之各數值

x	y	xy	x^2	
1	3	3	1	$n=9$
2	4	8	4	$\Sigma(x)=45$
3	6	18	9	$\Sigma(y)=70$
4	5	20	16	$\Sigma(x^2)=285$
5	10	50	25	$\Sigma(xy)=418$
6	9	54	36	
7	10	70	49	
8	12	96	64	
9	11	99	81	
45	70	418	285	

將表內所列五個數值代入前述典型方程式中,即得所求之兩個常態方程式。此兩式與前所求得者完全相同。

直綫方程式既已求得,即可將 x 各數值代入式中,而分別求 y 各數值,再將所得各值與 y 之各觀察值(即原來數值)一一比較,而求其差。下表所示,即求差數所用之行列也。

表六十七
一變量之觀察值與其計算值之比較(註)

	y (觀察值)	y_c (計算值)	d	d^2	xd
1	3	$3.2\frac{4}{9}$	$-.2\frac{4}{9}$.0597	$-.2\frac{4}{9}$
2	4	$4.3\frac{7}{9}$	$-.3\frac{7}{9}$.1427	$-.7\frac{5}{9}$
3	6	$5.5\frac{1}{9}$	$+.4\frac{8}{9}$.2390	$+1.4\frac{6}{9}$
4	5	$6.6\frac{4}{9}$	$-1.6\frac{4}{9}$	2.7041	$-6.5\frac{7}{9}$
5	10	$7.7\frac{7}{9}$	$+2.2\frac{2}{9}$	4.9381	$+11.1\frac{1}{9}$
6	9	$8.9\frac{1}{9}$	$+.0\frac{8}{9}$.0079	$+.5\frac{3}{9}$
7	10	$10.0\frac{4}{9}$	$-.0\frac{4}{9}$.0020	$-.3\frac{1}{9}$
8	12	$11.1\frac{7}{9}$	$+.8\frac{2}{9}$.6760	$+6.5\frac{7}{9}$
9	11	$12.3\frac{1}{9}$	$-1.3\frac{1}{9}$	1.7190	-11.8
合計			0.0	10.4885	0.0

(註) d 及 xd 行中小數點後仍用帶分數者,欲使各項差數之和適等於0,無稍差異也。

由上表可測知圖五十五中各點與直綫縱距離之和爲0，以 x 各數值乘各項差數而後相加，所得之和亦爲0，故計算之準確與否，可據此校核之。各項差數平方之和 (10.4885) 爲最小之數值。如將直綫方程式中 a 或 b 之數值予以變更，則所得直綫與各點縱距離之平方之和，必大於 10.4885。

直綫之配合：特例

在任何情形下，吾人均可應用聯立方程式之解法，由兩個常態方程式求得 a, b 之最可能數值，惟在特殊情形下，計算手續尙可簡略。此種情形爲經濟資料中所常見。設 x 各數值爲連續數字，如時間數列中之連續年份然，則在圖中可用 x 之中數爲原點。若 x 之項數爲奇數，則此原點卽爲中間一項。此時 $\Sigma(x)$ 之值爲0，而兩個常態方程式變爲

$$\Sigma(y) = na$$

$$\Sigma(xy) = b \Sigma(x^2)$$

例如一數列包含之時期，係自一九〇〇年起至一九二〇年爲止，吾人可取一九一〇年爲原點，如是則一九〇九年 x 之數值爲 -1 ，一九一一年 x 之數值爲 $+1$ ，餘類推，依此求 a, b 之數值，手續上自可便利不少。如數列之年數爲偶數，亦可應用此項簡便方法，將中間兩年之中點作爲原點，而以半年爲單位，計算各年 x 之數值，則 $\Sigma(x)$ 亦必等於0。

又如 x 各數值爲由0點起之連續正數時， $\Sigma(x)$ 及 $\Sigma(x^2)$ 之數值亦易斷定。自1起至第 n 項止之連續正數之和，可用公式 $\frac{n(n+1)}{2}$ 求之。例如自1至10各數之和爲 $\frac{10(10+1)}{2} = 55$ ，卽常態方程式中 $\Sigma(x)$ 之值也。自1起至第 n 項止之連續正數平方之和，可用公式 $\frac{2n^3 + 3n^2 + n}{6}$ 求之。例如自1至5各數平方之和爲 $\frac{250 + 75 + 5}{6} = 55$ ，卽常態方程式中 $\Sigma(x^2)$ 之值也。故常態方程式亦可改爲

$$\Sigma(y) = na + b\left(\frac{n(n+1)}{2}\right)$$

$$\Sigma(xy) = a\left(\frac{n(n+1)}{2}\right) + b\left(\frac{2n^3 + 3n^2 + n}{6}\right)$$

用此式求 a, b 之數值，往往較原來方程式為簡易。凡時間數列均可應用此式，以第一年 x 之數值編列為 1，將以後各年 x 之數值順次編列連續數字，可節省計算手續。

遞升冪級數式曲綫之配合

以上所論，均限於直綫之長期趨勢，但時間數列之長期趨勢亦常有不宜用直綫表示，或雖用直綫表示，而須將時間分作數段，分別配合直綫者。此種分段辦法，如因全時間內情勢驟變，前後兩期資料之性質不一致時，未嘗不可引用；但如全時期內情況大致相同，資料性質亦尚一致時，則此種分段辦法，實與長期趨勢之意義相背。數列之不宜以直綫配合者，吾人常可配合遞升冪級數式之曲綫，以表示其長期趨勢。配合此種曲綫之計算手續，可簡述於下。

遞升冪級數方程式之普遍形式為 $y = a + bx + cx^2 + dx^3 \dots$ 。此種方程式之曲綫雖未必為拋物綫式之曲綫，但通常稱為拋物綫。式中 x 之最高冪數為二次冪時，稱為二次拋物綫；為三次冪時，稱為三次拋物綫；餘類推。普通應用時，式中之冪數以不過二次或三次為宜。如 x 之最高冪數為二次冪時，則式中含有三個未知數，而須用三個常態方程式以求未知數之數值矣。

求常態方程式之方法與配合直綫時所用者相同，即以每個觀察方程式中第一個未知數之係數乘該式中各項，由此求得之各個方程式一一相加，即得第一個常態方程式；以每個觀察方程式中第二個及第三個未知數之係數分別乘各該式中各項，由此求得兩組方程式各別相加，即得第二個及第三個常態方程式。由此三個常態方程式即可求得 a, b, c 三

個未知數之值。此三值為 a, b, c 三個常數最可能之數值。三個常態方程式之普遍形式如下：

$$\Sigma(y) = na + b\Sigma(x) + c\Sigma(x^2)$$

$$\Sigma(xy) = a\Sigma(x) + b\Sigma(x^2) + c\Sigma(x^3)$$

$$\Sigma(x^2y) = a\Sigma(x^2) + b\Sigma(x^3) + c\Sigma(x^4)$$

茲以 1, 2; 2, 6; 3, 7; 4, 8; 5, 10; 6, 11; 7, 11; 8, 10; 9, 9 九點為例，說明配合二次拋物綫之計算手續。吾人無論在配合曲綫，或從事其他繁重之計算工作時，均須有條不紊，按步進行。使每一步驟與前後各步互相關聯，而在可用校核法時，又應儘量採用，蓋計算時縱極端審慎，錯誤或仍所難免也。計算之步驟應排列成表，使每步之計算，及其所得之結果，俱可詳列於表內。

本例所用之資料及其計算步驟，列如下表。

表 六 十 八

配合二次拋物綫所需各項數值之計算

x	y	xy	x^2	x^2y	
1	2	2	1	2	$n=9$
2	6	12	4	24	$\Sigma(x)=45$
3	7	21	9	63	$\Sigma(x^2)=285$
4	8	32	16	128	$\Sigma(x^3)=2,025$
5	10	50	25	250	$\Sigma(x^4)=15,333$
6	11	66	36	396	$\Sigma(y)=74$
7	11	77	49	539	$\Sigma(xy)=421$
8	10	80	64	640	$\Sigma(x^2y)=2,771$
9	9	81	81	729	
45	74	421	285	2,771	

本例中 x 各數值為自 1 起之連續整數，故 $\Sigma(x)$, $\Sigma(x^2)$, $\Sigma(x^3)$ 及 $\Sigma(x^4)$ 之數值又可自預先製就之表中求得之(註)。

(註)參考 Pearson 氏所編“Tables for Statisticians and Biometricians”第 28 表，一九一四年劍橋大學出版。

將表內各數值代入前述三式中，可得下列常態方程式：

$$74 = 9a + 45b + 285c$$

$$421 = 45a + 285b + 2,025c$$

$$2,771 = 285a + 2,025b + 15,333c$$

由此聯立方程式可得 a, b, c 三個常數之值如下：

$$a = -.929$$

$$b = +3.523$$

$$c = -.267$$

所求之二次拋物綫方程式爲

$$y = -.929 + 3.523x - .267x^2$$

此綫及原來之九點均繪於圖五十六內。

倘 x 各數值爲連續數字，如本例中所用者，亦可將 x 之中位數作爲原點，以節省計算工作，因此時 $\Sigma(x)$ 及 $\Sigma(x^3)$ 俱等於 0，而常態方程式可變爲：

$$\Sigma(y) = na + c\Sigma(x^2)$$

$$\Sigma(xy) = b\Sigma(x^2)$$

$$\Sigma(x^2y) = a\Sigma(x^2) + c\Sigma(x^4)$$

在配合 $y = a + bx + cx^2 + dx^3$ 形式之三次拋物綫時，吾人須斷定式中四個常數之值，而需用四個常態方程式。此四式之形式如下：

$$\Sigma(y) = na + b\Sigma(x) + c\Sigma(x^2) + d\Sigma(x^3)$$

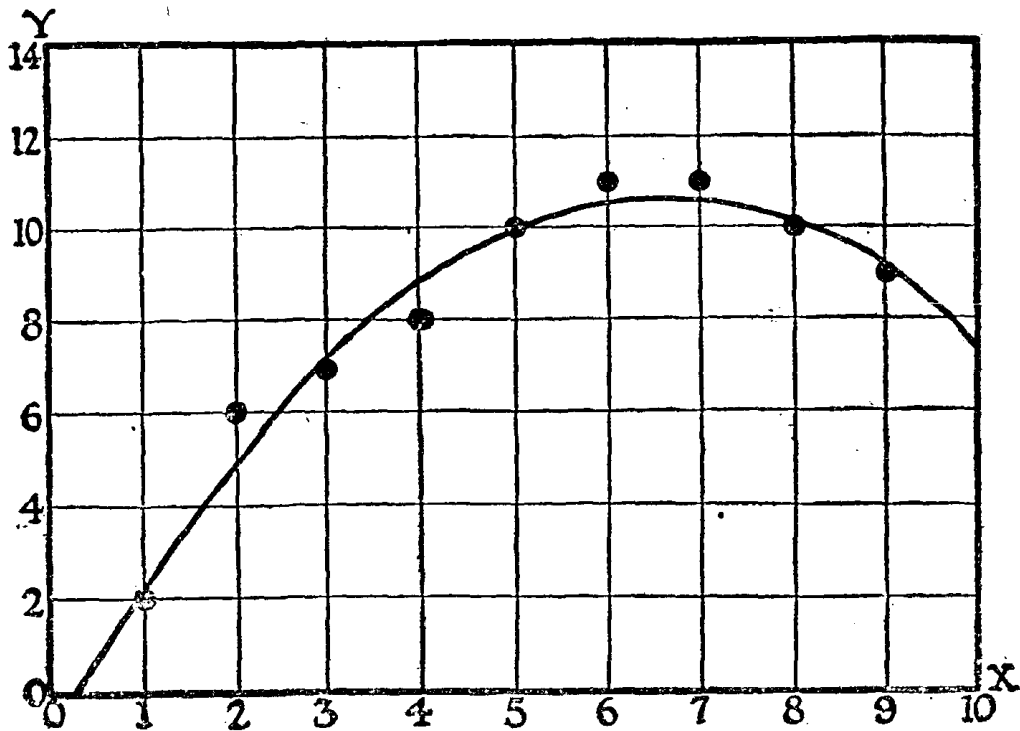
$$\Sigma(xy) = a\Sigma(x) + b\Sigma(x^2) + c\Sigma(x^3) + d\Sigma(x^4)$$

$$\Sigma(x^2y) = a\Sigma(x^2) + b\Sigma(x^3) + c\Sigma(x^4) + d\Sigma(x^5)$$

$$\Sigma(x^3y) = a\Sigma(x^3) + b\Sigma(x^4) + c\Sigma(x^5) + d\Sigma(x^6)$$

欲求四個或四個以上之常數數值，不僅所需計算工作頗爲繁重，且此種曲綫在理論上是否值得採用以表示數列之長期趨勢，亦成問題，蓋若方程式中常數增多，雖可使資料之轉變在曲綫上一一表現，然此種曲

綫斷非表示長期趨勢之綫也(註)。在經濟資料中,果有一真實趨勢之存在,其趨勢綫無論其為直綫或為曲綫,應為一簡單勻滑之綫,而實際數值在所配曲綫之兩旁雖不免有微小之起伏變動,但極端之起伏變動應所罕見也。如綫之兩旁遇有極端之變動,是必由於數列之本質有所變更,決不能適用一綫以代表其長期趨勢,而須將時期截為兩段或數段,



圖五十六. 在九點上配合二次拋物綫之圖解

(註)按 Steinmetz 氏之意見,在用上述遞升器級數式曲綫以代表經驗曲綫時,須合下列條件,方可採用。

1. 如 a, b, c (連續係數)之數值逐項減低,且減低之速度甚大,以致在觀察範圍內, x 高次器之數迅速變小,而成爲式中之次要項目者。
2. 如連續係數係依隨一定法則,而自成一收斂級數 (convergent series), 如指數之級數,或三角函數之級數等等者。
3. 如各係數中,僅有數個係數之數值較大,其餘均甚小,但欲偏重於數個較大之係數,不必計及較小之係數者(參看 Steinmetz 氏著 "Engineering Mathematics" 第 214—215 頁)。

分別配以趨勢綫。史丹曼茲(Steinmetz)氏嘗謂“在觀察範圍之內，自然界情形不變，始可用一單獨方程式以代表經驗曲綫(empirical curve)”，此言雖指自然科學方面之資料而言，然亦可適用於經濟資料也。

測定商業倒閉家數長期趨勢之實例

前數節所述為用簡單之資料以配合直綫或拋物綫之手續，在討論其他曲綫以前，請舉實例以說明之。茲就一八九七年至一九二一年間美國商業倒閉家數之資料，分別配合直綫、二次拋物綫及三次拋物綫。吾人用同一資料配合三種長期趨勢綫後，可比較其結果，而注意其不同之點，以供研究。

為計算便利起見，可用中間一年，即一九〇九年，作為原點。常態方程式中所需各數值可由表六十九得之。表內 x 值代表時間， y 值代表商業倒閉之家數。

表六十九
1897—1921年美國商業倒閉家數
配合長期趨勢綫所需各項數值之計算

(1) 年份	(2) x	(3) y (商業倒閉家數)	(4) xy	(5) x^2y	(6) x^3y
1897	-12	13,083	-156,996	1,883,952	-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	488,677	-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

表 六 十 九 (續)

年 份	x	y	xy	x^2y	x^3y
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	686,260	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
合 計	0	301,958	+188,084	15,516,228	+9,881,156

$$n = 25$$

$$\Sigma(x) = 0$$

$$\Sigma(y) = 301,958$$

$$\Sigma(xy) = 188,084$$

$$\Sigma(x^2) = 1,300$$

$$\Sigma(x^2y) = 15,516,228$$

$$\Sigma(x^3) = 0$$

$$\Sigma(x^3y) = 9,881,156$$

$$\Sigma(x^4) = 121,420$$

$$\Sigma(x^5) = 0$$

$$\Sigma(x^6) = 13,471,900$$

配合直綫之演算

x 之原點既在時期之中點,則求 a, b 兩常數所用之直綫常態方程式爲

$$\Sigma(y) = na$$

$$\Sigma(xy) = b \Sigma(x^2)$$

將表六十九中求得之數值代入上式,則

$$301,958 = 25a$$

$$188,084 = 1,300b$$

由此兩式可得 a, b 之數值爲

$$a = 12,078$$

$$b = 144.7$$

故直綫方程式爲

$$y = 12,078 + 144.7x$$

配合二次拋物綫之演算

配合二次拋物綫時所用之常態方程式有三。其形式如下

$$\Sigma(y) = na + c \Sigma(x^2)$$

$$\Sigma(xy) = b \Sigma(x^2)$$

$$\Sigma(x^2y) = a \Sigma(x^2) + c \Sigma(x^4)$$

將相當之數值代入式中，則

$$301,958 = 25a + 1,300c$$

$$188,084 = 1,300b$$

$$15,516,228 = 1,300a + 121,420c$$

解式中之常數，可得

$$a = +12,258$$

$$b = +144.7$$

$$c = -3.45$$

故二次拋物綫之方程式爲：

$$y = 12,258 + 144.7x - 3.45x^2$$

配合三次拋物綫之演算

配合三次拋物綫時，須用四個常態方程式以求四個常數之數值。其形式如下：

$$\Sigma(y) = na + c \Sigma(x^2)$$

$$\Sigma(xy) = b \Sigma(x^2) + d \Sigma(x^4)$$

$$\Sigma(x^2y) = a \Sigma(x^2) + c \Sigma(x^4)$$

$$\Sigma(x^3y) = b \Sigma(x^4) + d \Sigma(x^6)$$

由此所得之四個聯立方程式爲

$$301,958 = 25a + 1,300c$$

$$188,084 = 1,300b + 121,420d$$

$$15,516,228 = 1,300a + 121,420c$$

$$9,881,156 = 121,420b + 13,471,900d$$

解四式中常數之數值,可得

$$a = +12,258$$

$$b = +482$$

$$c = -3.45$$

$$d = -3.61$$

故三次拋物綫之方程式爲

$$y = 12,258 + 482x - 3.45x^2 - 3.61x^3$$

表 七 十

1897—1921年美國商業倒閉家數

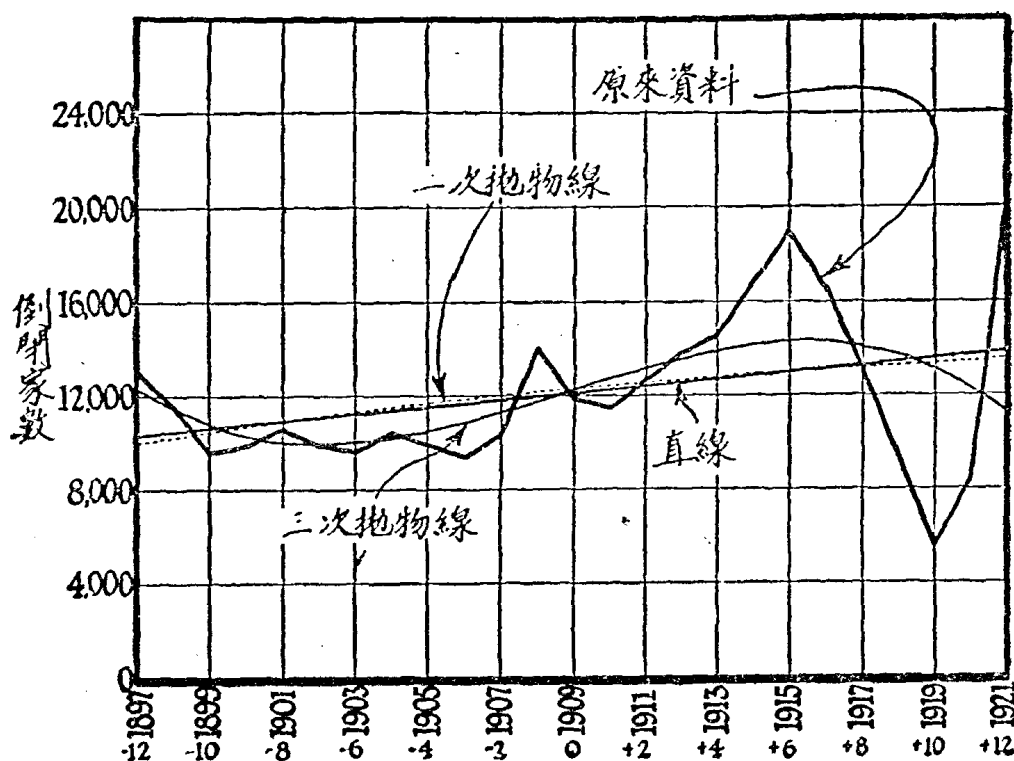
實際數值、由三種長期趨勢綫求得之常態數值、及實際數值對常態數值之百分差

年份	實際數值	直綫之 常態數值	二次拋物綫 之常態數值	三次拋物綫 之常態數值	實際數值對各趨勢綫常態數值之百分差		
					直 綫	二次拋物綫	三次拋物綫
1897	13,083	10,342	10,025	12,215	+26.5%	+30.5%	+7.0%
1898	11,615	10,486	10,249	11,344	+10.8	+13.2	+2.3
1899	9,642	10,631	10,466	10,703	-9.3	-8.0	-10.0
1900	9,912	10,776	10,676	10,272	-8.0	-7.0	-3.8
1901	10,648	10,920	10,880	10,030	-2.5	-2.0	+6.5
1902	9,973	11,065	11,076	9,953	-9.9	-10.0	+0.2
1903	9,775	11,210	11,266	10,022	-12.8	-13.2	-2.5
1904	10,417	11,354	11,448	10,213	-8.3	-9.0	+2.1
1905	9,967	11,499	11,624	10,506	-13.3	-14.3	-5.1
1906	9,385	11,644	11,793	10,878	-19.4	-20.4	-13.7
1907	10,274	11,789	11,955	11,313	-12.9	-14.1	-9.2
1908	14,066	11,933	12,110	11,776	+17.8	+16.0	+19.4
1909	11,872	12,078	12,258	12,258	-1.7	-3.2	-3.2
1910	11,588	12,223	12,400	12,733	-5.2	-6.6	-9.1
1911	12,679	12,367	12,533	13,179	+2.5	+1.0	-3.8
1912	13,832	12,512	12,660	13,575	+10.5	+9.2	+1.9
1913	14,553	12,657	12,781	13,900	+15.0	+13.9	+4.6
1914	16,780	12,801	12,895	14,131	+31.0	+30.0	+18.6
1915	19,035	12,946	13,002	14,246	+47.0	+46.2	+33.5
1916	16,498	13,091	13,102	14,225	+26.0	+25.9	+15.9
1917	13,073	13,236	13,194	14,045	-1.2	-0.9	-7.0
1918	9,331	13,380	13,281	13,685	-30.2	-29.8	-31.8
1919	5,515	13,525	13,330	13,123	-59.1	-58.7	-57.8
1920	8,463	13,670	13,432	12,338	-38.0	-37.0	-31.4
1921	19,982	13,814	13,498	11,307	+44.6	+48.0	+76.8

商業倒閉家數之原來資料及由前述三個方程式所得之綫，均繪於圖五十七中。茲將 y 之實際數值（商業倒閉家數）、及由各方程式計算所得之常態數值、以及實際數值對常態數值之百分差，列於表七十內。

各種長期趨勢綫之比較

由以上討論及圖五十七之圖解，可見用各種曲綫代表長期趨勢所得結果頗有差異。在本例中，直綫與二次拋物綫極為接近，但三次拋物綫所表現之長期趨勢則與兩者迥不相同。吾人選取此項資料，即欲極度顯示此種不同之點。各年之常態數值，係由趨勢綫上各該年之縱坐標測度而得，故若所用趨勢綫不同，則視作常態之標準自亦不同，而由實際數值對常態數值之差所表示之循環變動，亦必隨所選曲綫之種類而異。此點可由表七十所列實際數值對常態數值之百分差見之。



圖五十七. 1897—1921年美國商業倒閉家數及其三種長期趨勢綫

實際數值對常態數值之百分差，由直綫求得與由二次拋物綫求得者極為相似，但由三次拋物綫求得者，即顯有不同之處。表內百分差有正負之別。由三次拋物綫求得之差與由其他兩種趨勢綫求得之差彼此比較，正負號不同者共有四年，其餘各年正負號雖同，而百分數頗有相差甚鉅者。吾人常有根據已往資料所配合之趨勢綫延長至資料時期以外，以推測未來之可能趨勢者；若就此點而論，則三次拋物綫延長後所得結果，必不合理。趨勢綫之延長，無論其為直綫或曲綫，雖皆含有揣測之意味，然延長直綫及二次拋物綫，則所得結果皆比較可靠。

關於選擇趨勢綫之原則，下文當更有詳盡之討論，此處可就數點作簡要之說明。如在長時期內，資料之實際數值始終高於趨勢綫或低於趨勢綫時，則此綫之配合恐不適當。在商業倒閉家數之一例中，直綫及二次拋物綫均不能緊隨實際資料之綫，而相距甚遠；尤以一八九九年至一九〇七年之九年間，實際數值始終在兩綫之下。三次拋物綫與實際資料頗為接近，故就原有資料範圍以內，三者相較，似以此綫之配合最為貼切。

對數應用於曲綫之配合

前述各綫之形式俱屬簡單而甚合用者，惟分析時間數列如用半對數式之直綫或曲綫，則尤為適用。用半對數或比例尺度作圖之優點，前已論之詳矣，此種圖形最優之點，在能確切表示相對的或比例的變動。經濟資料之分析，常注重於相對的或比例的變動，故選擇趨勢綫時，此點應考慮及之。

在測度經濟資料比例變動之長期趨勢時，吾人仍可應用上述各種直綫或曲綫之普遍方程式，所稍異者，式中 y 一項須用 $\log y$ 代替之，故直綫之方程式應為 $\log y = a + bx$ ，遞升冪級數曲綫之方程式應為 $\log y = a + bx + cx^2 + dx^3 + \dots$ 。作圖時，如繪於算術尺度之圖上，可以 x 之真數

與 y 之對數繪製；如繪於半對數尺度之圖上，則縱軸用對數尺度，橫軸用算術尺度，而 x 及 y 俱以真數繪製。兩者比較，當以後者較為簡便。

茲將配合 $\log y = a + bx + cx^2$ 式之曲綫時所需計算之步驟說明之。設吾人欲決定一九〇八年至一九二二年美國石油產量之趨勢綫，則須將相當數值代入常態方程式，以求 a, b, c 各常數。其所需數值列表於下：

表七十一

1908—1922年美國石油生產量

配合趨勢綫時所需各數值之計算

年份	x	y 產量 (百萬桶)	$\log y$	$x \cdot \log y$	$x^2 \cdot \log y$
1903	-7	178.5	2.25164	-15.76148	110.33036
1909	-6	183.2	2.26269	-13.57614	81.45684
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
			36.99528	9.41834	694.23282

$n = 15$

$\Sigma(x) = 0$

$\Sigma(\log y) = 36.99528$

$\Sigma(x^2) = 280$

$\Sigma(x \cdot \log y) = 9.41834$

$\Sigma(x^3) = 0$

$\Sigma(x^2 \cdot \log y) = 694.23282$

$\Sigma(x^4) = 9352$

此處所欲求解之三個常態方程式之形式如下：

$$\begin{aligned}\Sigma(\log y) &= na + b\Sigma x + c\Sigma x^2 \\ \Sigma(x \cdot \log y) &= a\Sigma x + b\Sigma x^2 + c\Sigma x^3 \\ \Sigma(x^2 \cdot \log y) &= a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4\end{aligned}$$

將求得之數值代入上式，則

$$\begin{aligned}36.99528 &= 15a + 280c \\ 9.41834 &= 280b \\ 694.23282 &= 280a + 9352c\end{aligned}$$

解式中之常數，得

$$\begin{aligned}a &= +2.450508 \\ b &= +.033637 \\ c &= +.0008488\end{aligned}$$

故所求趨勢綫之方程式，爲

$$\log y = 2.450508 + .033637x + .0008488x^2,$$

此式之原點在一九一五年。

表七十二

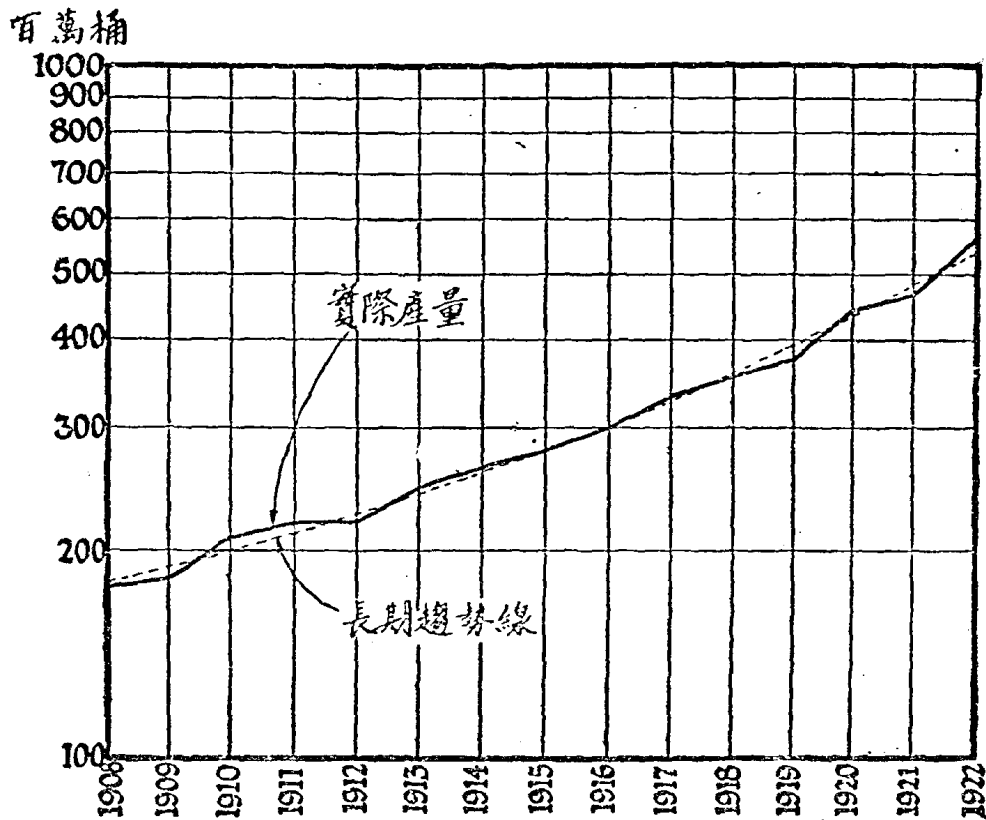
美國石油產量之長期趨勢及其實際值與趨勢值之比較

(趨勢綫係配合於產量之對數上者)

年份	x	y(實際值)產量 (單位:百萬桶)	趨勢值 之對數	y(計算值)趨勢值 (單位:百萬桶)	實際值對趨勢值 之百分比
1908	-7	178.5	2.256639	180.6	98.8
1909	-6	183.2	2.279242	190.2	96.3
1910	-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	102.0
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

將代表某年之 x 值代入式中，即可求得該年趨勢值或常態值之對數，更由此對數以求其真數。各年常態值及實際值對趨勢值之百分比，見表七十二。

各年實際產量之數字及其趨勢綫見圖五十八。由前項方程式所得之綫，極為配稱，頗能表示石油產量之長期趨勢，可由圖中窺見一斑。



圖五十八. 1908—1922年美國石油產量及用對數配合之長期趨勢綫

試以此綫與根據產量之真數配合所得相類之曲綫(二次拋物綫)互為比較，即可立見其優點。該拋物綫常數之數值用最小平方法求得後，代入該綫之普遍方程式中，即得下式：

$$y = 279.387 + 24.262x + 1.650x^2$$

式中原點亦在一九一五年。石油之實際產量、及由上式求得之趨勢值、以及兩者之百分比，均列入下表內。

表 七 十 三

美國石油產量之長期趨勢及其實際值與趨勢值之比較

(趨勢綫係配合於產量之真數上者)

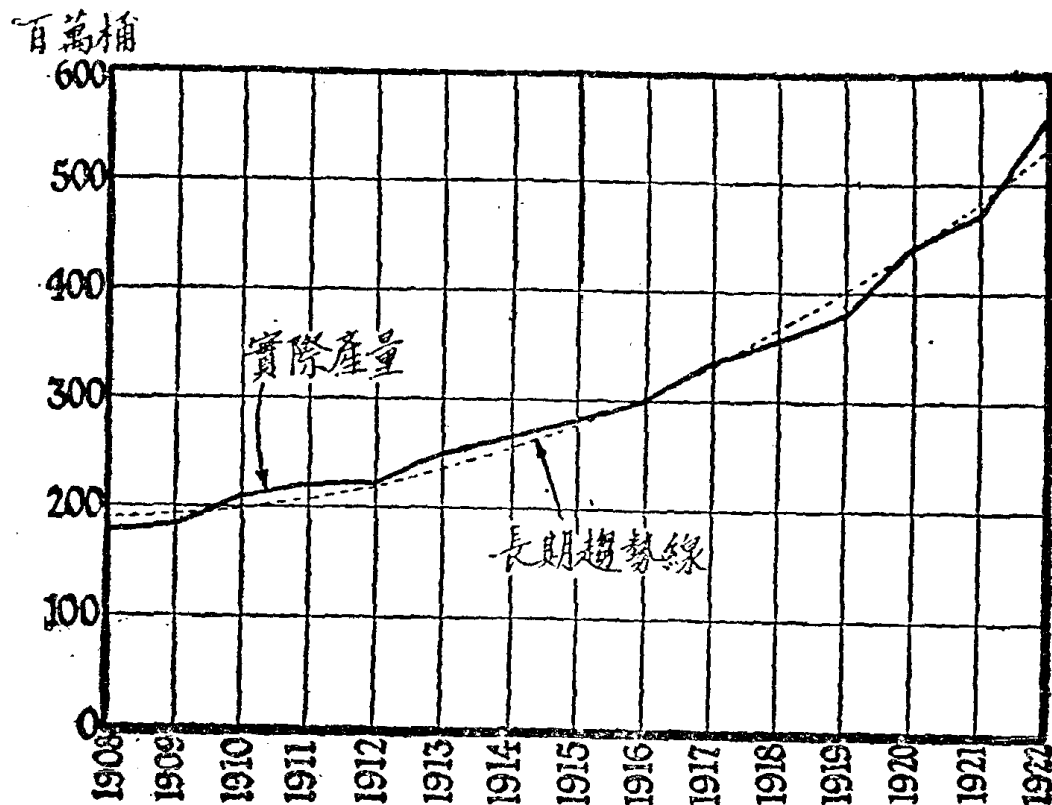
年份	n	x (實際值)產量 (單位:百萬桶)	y (計算值)趨勢值 (單位:百萬桶)	實際值對趨勢值 之百分比
1908	-7	178.5	190.403	93.7
1909	-6	183.2	193.215	94.8
1910	-5	209.6	199.327	105.2
1911	-4	220.4	208.739	105.6
1912	-3	222.9	221.451	100.7
1913	-2	248.4	237.463	104.6
1914	-1	265.8	256.775	103.5
1915	0	281.1	279.387	100.6
1916	1	300.8	305.299	98.5
1917	2	335.3	334.511	100.2
1918	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

該趨勢綫及石油之實際產量,見圖五十九。

此處用以表示石油產量長期趨勢之兩種方程式中,均含有三個常數,故其結果可互相比較。由圖中觀之,配合於對數上之曲綫似較配合於真數上者為優,但實際須用一準確之量數以測度之,方可斷定。吾人可用均方根差作為測度之標準。均方根差之計算與前此用於次數分配時之算法相同,所稍異者,各項差數係實際值對趨勢值之差,而非實際值對算術平均數之差。

配合於真數之趨勢綫所求得之均方根差(即標準差)為 12.46,配合於對數之趨勢綫所求得之均方根差為 9.44 (兩者之單位均為百萬桶)。

於此可見對數趨勢綫兩旁散佈之程度遠較真數趨勢綫為小，故在本例中，應用對數曲綫以表示長期趨勢較為適當(註)。



圖五十九. 1908—1922年美國石油產量及用真數配合之長期趨勢綫

前述兩類曲綫在經濟統計方面，最為適用。時間數列之長期趨勢大都可用遞升冪級數之曲綫表示之。此項曲綫或用資料之真數配合，或用資料之對數配合(其區別前者係用 y 值之真數配合，後者則用 y 之對數配

(註)上列兩標準差之數值，係由單位百萬桶之實際差數求得者。如用百分差計算，則對數趨勢綫之優點當更顯著。由百分差求得真數趨勢綫之標準差為 4.01，對數趨勢綫之標準差為 2.69。

於此所應注意者，在比較兩綫之標準差時，該兩綫之方程式中所含常數之數目必須相同，因所含常數或多或少與標準差數值之大小有關。若所含常數甚多，而與所給點數相同，則由此所配合之趨勢綫可通過各點，此時標準差之值將等於 0，但此綫已失去趨勢綫之意義，斷不可用以測度長期趨勢也。

合；至於代表時間之 x 值，則均用其真數）。此兩類曲綫均含有伸縮性，爲在曲綫中應用甚廣者。此外尚有數種曲綫，在時間數列方面雖爲用較狹，然有時所得結果亦甚佳，請分述於下。

普通拋物綫式之曲綫($y = ax^b$)不適用於經濟統計之時間數列，因吾人不應將時間變量依幾何級數處理之也。此種曲綫在雙對數尺度之圖中，可變爲直綫，前嘗言之。此綫如能準確表示某種數列之長期趨勢，亦未始不可採用耳。

此種曲綫可用對數配合，使成直綫形式，其配合手續亦頗簡易。方程式

$$y = ax^b$$

應用對數時，即變爲

$$\log y = \log a + b \log x$$

配合此綫所用之兩個常態方程式如下：

$$\Sigma(\log y) = n \log a + b \Sigma(\log x)$$

$$\Sigma(\log x \cdot \log y) = \log a \Sigma(\log x) + b \Sigma(\log x)^2$$

根據實際資料計算上兩式中所需之各項數值，代入式中，即可求得 $\log a$ 及 b 之數值，其手續與配合普通直綫時相同(註)。

簡單之指數曲綫 $y = ab^x$ ，用於分析時間數列，頗爲普遍。其對數方程式爲

$$\log y = \log a + (\log b)x$$

配合此種曲綫所用之兩個常態方程式爲

$$\Sigma(\log y) = n \log a + \log b \Sigma(x)$$

$$\Sigma(x \cdot \log y) = \log a \Sigma(x) + \log b \Sigma(x^2)$$

(註)Raymond Pearl 氏所著“Medical Biometry and Statistics”一書附錄內(一

九二三年菲勒特爾菲亞 Saunders 出版)載有一表，將真數1起至100止之對數之和

一一列入，頗切實用。

此綫與以前所述半對數曲綫之直綫形式相類似。配合曲綫時，先解常態方程式中兩個常數之對數，更由此對數求真數，即可將前項對數方程式還原為指數曲綫之形式。

統計學家亦有用康氏曲綫(Gompertz curve)以解釋經濟統計之資料者。此綫之方程式原為保險學家計算保險率時所發現。其形式為

$$y = ab^{c^x}$$

提倡應用此式以分析經濟統計者，謂人口之增加常遵循一滋長之普通法則，而此法則在工業生產方面亦當存在，則以工業生產量為人口滋長之函數故也(註一)。

布爾氏(Raymond Pearl)暨李特氏(Lowell J. Reed)曾用類似之滋長曲綫以預測人口之蕃殖(註二)，嗣經發現該綫亦可用以表示某種經濟資料之長期趨勢。其方程式如下：

$$y = d + \frac{b}{e^{-ax} + c}$$

配合曲綫之簡捷法：平均數法

曲綫之配合，欲求十分準確，自應採用最小平方法；但若配合之目的，僅欲求其近似值，則有時可採用他種方法。

吾人若用平均數法配合直綫，可將一數列分為時期長短相等或近乎相等之兩部分，由兩部分所含 x 及 y 之數值，可求得如下式之兩個形式相同之方程式：

(註一)關於此綫之特性，在 Raymond B. Prescott 氏所著 "Law of Growth in Forecasting Demand" 一文內曾有敘述。原文見一九二二年十二月出版 "Journal of the American Statistical Association" 第471—479頁。

(註二)參看一九二三年 Committee of Plan of New York and Its Environs 發表 Raymond Pearl 暨 Lowell J. Reed 兩氏所著 "Predicted Growth of Population of New York and Its Environs"。

$$\Sigma(y) = na + b\Sigma(x)$$

解此兩個聯立方程式，即可得常數 a, b 之數值。此法與求每部分內以 x 及 y 數列之兩個算術平均數作為縱橫坐標之一點，而後將兩點用直綫接連之意義相同。通常用平均數法求得之曲綫，與用最小平方法求得者，並不完全相同。

茲就表六十九所列商業倒閉家數之資料，以說明此法之應用。該數列共包含十五年，故分為兩部分時，一部分可包含十三年，另一部分包含十二年。由此兩部分所得之方程式為

$$140,629 = 13a - 78b$$

$$161,329 = 12a + 78b$$

解兩式，可得

$$a = 12,078$$

$$b = 210$$

故用此簡捷法求得長期趨勢綫之方程式為

$$y = 12,078 + 210x$$

而由最小平方法所得之方程式為

$$y = 12,078 + 144.7x$$

後者自屬配合最為恰當之直綫。在本例中，該兩綫相差甚遠，自在吾人意料之中，蓋原來資料本與直綫形之趨勢綫相差甚多也。

選點法

配合趨勢綫時，如已擇定曲綫之形式，則常數之近似值可用簡捷法求得。其法係就實際資料繪成之曲綫上，選定與常數個數相等之數點，將各點之縱橫坐標分別代入數目相等之各個曲綫方程式中，即可解出常數之數值。在商業倒閉家數之一例中，如欲配合三次拋物綫，則應選定四點。吾人可觀察實際資料變動之大體趨勢，先繪一相當之經驗曲

綫，以助選擇各點之用。假如吾人選定下列各點，作為計算之根據（參看表六十九及圖五十七關於商業倒閉家數之資料）：

x	y
-10	10,000
-5	10,000
+5	14,000
+10	13,000

將 x 及 y 各數值代入三次拋物綫之方程式中，即得下列四式：

$$10,000 = a - 10b + 100c - 1,000d$$

$$10,000 = a - 5b + 25c - 125d$$

$$14,000 = a + 5b + 25c + 125d$$

$$13,000 = a + 10b + 100c + 1,000d$$

解此四式，可得四個常數之數值為

$$a = +12,166$$

$$b = +483$$

$$c = -6.7$$

$$d = -3.3$$

故所求之方程式為

$$y = 12,166 + 483x - 6.7x^2 - 3.3x^3$$

此式可與用最小平方法求得之方程式比較之。

普通經濟資料用簡捷法計算，常有發生重大差誤之可能，因此項資料之實際數量與其趨勢綫常不甚接近所致。經濟資料雖可由觀察而得，但常含有極大之差誤。在自然現象方面，如資料遵循一定法則而變動，而觀察之差誤又甚小時，則用平均數法或選點法所得之結果，當不致發生重大差誤；但在經濟現象方面，實際資料與任何形式之趨勢綫常相差甚遠，故以不用簡捷法為是。

長期趨勢綫之選擇

經濟資料可配合種種曲綫以表示其長期趨勢，已如前述，但每種曲綫決不盡能適用於各種經濟資料，於是發生曲綫之選擇問題。由各種曲綫求得之常態值並不相同，故曲綫之選擇與常態值之大小有關。在各種常態值中，究以何者最爲適合乎？關於此點，前曾有所指示，但未涉及一般原則，惟嚴格言之，曲綫之選擇原無一般原則可循，蓋吾人欲確定曲綫配合之是否適當，實無絕對可靠之測驗方法也。某種經濟資料應選擇何種曲綫以表示其長期趨勢，常憑個人之判斷而定，而在個人判斷之中，經驗實占重要地位。惟有數種方法可作吾人選擇曲綫之助，茲分述之。

1. 選擇曲綫之初步手續爲作圖。將原來資料繪成綫圖後，常可由觀察選定一適當之長期趨勢綫。此項綫圖可繪成下列四種不同之混合形式，其中前兩種用於分析經濟資料，頗占重要地位。

- a. x 值及 y 值均用真數（即將數字繪於普通算術尺度之紙上者）。
- b. x 值用真數， y 值用對數（即繪於半對數尺度之紙上，而以 x 值用真數尺度， y 值用對數尺度繪製者）。
- c. y 值用真數， x 值用對數（即繪於半對數尺度之紙上，而以 x 值用對數尺度繪製者）。
- d. x 值及 y 值均用對數（即繪於雙對數尺度之紙上者）。

觀察四種圖形時，如發現其中一種呈直綫趨勢，則應依該圖所用之尺度，以求出該綫之方程式（參看第二章）。如四圖之中，無一呈現直綫趨勢，則該項圖解亦可作選擇他種簡單曲綫之助。在用圖解以選擇趨勢綫時，吾人必須熟悉各種簡單方程式所代表之曲綫形式（註）。

（註）在下列各書中，可得各種曲綫之圖形：

T. R. Running氏著：“Empirical Formulas”，1917年紐約 Wiley 公司出版。

C. P. Steinmetz氏著：“Engineering Mathematics”，1917年紐約 McGraw-Hill 公司出版。

Joseph Lipka 氏著：“Graphical and Mechanical Computation”，1918年紐約 Wiley 公司出版。

2. 曲綫之選擇，可由研究 x, y 兩變量之關係而定。在極簡單情形之下， x 及 y 可有下列數種關係(註)：

a. 如將 x 之數值排列成一算術級數，而相對於各 x 值之 y 數值成一幾何級數時，則兩者之關係可由指數曲綫表示之。其方程式為

$$y = ab^x$$

b. 如將 x 之數值排列成一幾何級數，而相對於各 x 值之 y 數值亦成一幾何級數時，則兩者之關係可用簡單拋物綫或雙曲綫表示之。其方程式為

$$y = ax^b$$

c. 如將 x 之數值排列成一算術級數，而相對於 x 值之 y 數值，其第一階差 (first differences) 為常數時，則兩者之關係可用直綫表示之。其方程式為

$$y = a + bx$$

當 x 之數值排列成一算術級數時，其 y 數值之各連續項之差數謂之第一階差，而以 Δy 之符號代之；其第一階差各連續項之差數謂之第二階差，而以 $\Delta^2 y$ 之符號代之；其他高次階差之求法及符號，依此類推。下表係說明階差之求法。

x	y	Δy	$\Delta^2 y$	$\Delta^3 y$
1	11			
2	40	29		
3	101	61	32	
4	206	105	44	12
5	367	161	56	12
6	596	229	68	12
7	905	309	80	12
8	1306	401	92	12
9	1811	505	104	12
10	2432	621	116	12

(註)關於兩變量之關係，吾人前有論及，謂算術級數內各項數值依一固定之相對增量而增加，幾何級數內各項數值係依一固定之百分率而增加。

d. 如將 x 之數值排列成一算術級數，而相對於各 x 值之 y 數值，其第 n 階差為常數時，則兩者之關係可用 x 最高冪數為 n 之遞升冪級數之方程式表示之。其式如下：

$$y = a + bx + cx^2 + dx^3 + \dots + qx^n$$

例如上表內 y 之第三階差為常數， x 與 y 之關係應以下式表示之。

$$y = a + bx + cx^2 + dx^3$$

以上所述及其他類似之測驗方法，均可適用於各種資料，以作選擇趨勢綫之助，在本章末所列倫寧(Running)及立蒲克(Lipka)兩氏所著各書內，論述頗詳。惟於此所當注意者，上述各種測驗方法如用於經濟資料以選擇趨勢綫，鮮有能完全合於任何一種測驗者，蓋如欲完全合於測驗，則所選曲綫非通過原來資料之各點不可矣。故在測驗資料時，但求與某種測驗方法大致相合，即可以該種測驗所代表之曲綫選充趨勢綫。

3. 如將原來資料詳加研究後，仍不能確定所應選取之趨勢綫形式，則可同時配合各種曲綫，比較其結果，而後擇定一最適合之綫，如在商業倒閉家數及石油產量之兩例中選擇趨勢綫時之所為。吾人所欲比較之數種曲綫，如其方程式中所含常數之數目相同，則欲就資料範圍以內，斷定配合最適當之綫，可比較原來資料與各趨勢綫所得均方根差之大小而得，此為比較準確之測驗方法也。

各綫之均方根差可由下列關係求之。

$$\Sigma(d^2) = \Sigma(y^2) - a\Sigma(y) - b\Sigma(xy) - c\Sigma(x^2y) - \dots$$

式中之 $\Sigma(d^2)$ 為各項 y 值與其常態值相差之平方之和(此方程式之來源，詳見附錄A，惟所用者為普遍形式)。但若吾人所欲比較之數種曲綫其方程式中所含常數之數目不同時，則欲斷定配合最適當之綫，不可用此法測驗，而僅能用觀察法加以比較。此時則以個人之判斷力，為選擇曲綫之唯一根據矣。

於此更應注意者，吾人所配之曲綫，就資料範圍以內，配合極貼

切，未必即為最優之綫，蓋若方程式中常數之數目與資料之點數相同，則由該方程式所得之綫可通過各點，然此綫未必能代表其長期趨勢。長期趨勢者，謂一數列於紛亂錯雜之變動中，在長時期內所表現之一種逐漸向上或逐漸向下而較有規則之變動也。故就長期趨勢之意義而論，曲綫之形式應求簡單；但此點非謂吾人可用不合原來資料之簡單曲綫以表示複雜之長期趨勢，亦當注意耳。

4. 在選擇曲綫之前，尚有一先決問題應加討論者，即趨勢綫之應用在時間上之限度是也。如曲綫之應用，僅限於資料時期以內者（如作內插補 interpolation 之用），則曲綫與資料之是否適當，即可作為選擇曲綫之唯一標準。如曲綫應用之範圍須伸展至資料時期以外，係為預測之用，而作為決定未來時期內常態值之根據者，則選擇曲綫時，除曲綫與資料務求相配外，更須注意趨勢綫伸展至資料時期以外，是否合理，而不與以往情形相矛盾。此項要點在商業倒閉數之一例中，曾論及之。

曲綫之伸展或外插補 (extrapolation) 原係一種揣測，僅在趨勢綫與資料極相配稱，而所有影響於該項資料之過去狀況，在未來時期內仍將存在時，始屬可行。如未來之狀況有所變更，或有新因素發生，則不可將曲綫伸展至資料時期以外。惟在處理經濟統計時，經濟狀況之變更，常不易立時發現，而在經過相當時期以後方可知之，由伸展趨勢綫所得之預測常含有差誤者，即由於此。雖然，在統計實際工作方面，吾人亦嘗伸展趨勢綫以為預測之根據者，蓋吾人以為未來之變化常有遵循過去途徑之可能也。由是知預測之時期愈遠，則包含之差誤亦愈大，故當新資料源源而來時，趨勢綫亦當隨時校正之。

趨勢綫作伸展之用者，則以選取含有常數較少之簡單曲綫為宜。三次或四次拋物綫對於資料或極相配，但不合於伸展之用，故潘林氏 (Perrin) 嘗有適合於內插補之曲綫未必即合於外插補之言也。

大體言之，用 y 值之對數所配合之簡單曲綫，伸展後預測所得之結

果，似較用真數所配合者更為可靠。關於此點，加斯頓氏 (Karl G. Karsten) 曾作一頗有興趣之討論(註)。氏以為依一固定變動率 (rate of change) 而變動之現象，常較依一固定變動量 (amount of change) 而變動之現象，為能保持其以往之趨勢。測量變動率之綫自以半對數式之曲綫為最佳。

5. 配合曲綫之目的，如為測度商情循環變動時求循環變差之用，則另有一種測驗方法，可作選擇曲綫之標準。其法係以各種曲綫所求得之循環變動與表示實際商情循環變動之其他各種數列比較觀之，如由某綫測度所得之循環變動與實際商情循環變動最相符合，則該綫可認為最適宜之綫。吾人在配合商業倒閉數之趨勢綫時，曾用此法測驗之。在該例中，由直綫及二次拋物綫測度所得之循環變動，雖能反映全時期內顯著之商情循環，但未能顯示細微之商情循環。例如一八九九年至一九〇七年為商業興旺時期，該時期內商業倒閉數之實際數值皆在該兩綫之下，可見由兩綫所求得之循環變動，確能反映商情循環；但由三次拋物綫觀之，則自一九〇一年至一九〇四年間商業倒閉數之實際數值略高於該綫，蓋該數年為在一八九九年至一九〇七年商業興旺之全期內所包含細微循環變動中商業衰落之一部，而未能反映於直綫及二次拋物綫者也。由此法測驗之結果，可見三次拋物綫不僅能顯示商情變動之大勢，且能測度細微之循環變動，似為最能適合實際資料之綫，但此趨勢綫配合之是否適當，仍須視配綫之目的是否為測度商情循環變動之用而定耳。

6. 吾人配合一單獨之任何趨勢綫，常覺不能完全適合於全時期內之資料，此因全時期內經濟情況或有變更，足以影響其長期趨勢之所致。例如自南北大戰至一八九六年間批發物價變動之趨勢逐漸向下，可用一直綫表示之；自一八九六年至歐戰之初期，物價趨勢轉變而逐漸

(註)參看 "Charts and Graphs" 第423—425頁。

上升，則須用二次拋物綫表示之。類此情形之見於他種經濟數列者，其例蓋不勝枚舉。如遇此種情形，雖可將全時期截為數段，而分別配以適當之趨勢綫，但若分段過多，即難免有極不合理之處，蓋長期趨勢原為長時期內一種逐漸變動之趨向，今為配合數種適當之趨勢綫起見，而將數列分作數段，即不合於長期趨勢之基本觀念矣。故除因經濟情況確有變更，而有分段之必要時以外，在任何情形之下，應選擇一綫以表示全時期內之長期趨勢。

用關聯之統計數列代表長期趨勢之方法

當一數列所包含之時期過短，或資料不足，或因經濟情況有所變更，因之以前資料所表示之趨勢顯有轉變時，吾人自不能用配合數學曲綫之方法，以求該數列之長期趨勢。歐戰以後之物價數列，即其一例。歐戰以前物價之長期趨勢，在戰後顯有變更，但因戰後經過之時期太短，即不能配合長期趨勢綫，故其未來之物價趨勢亦屬無從推測。遇此困難情形，潘遜氏(Warren M. Persons)嘗倡用一新穎方法，即採用一數列，作為測度另一關聯數列之循環變動之基數(base)(註一)。

潘遜氏所用者為兩種物價指數之數列，一為勃蘭特斯脫批發物價指數，一為潘遜氏本人所編用以測度商情循環之物價指數。該兩種指數之編製方法，已詳前文(註二)。前一種係根據物品九十六種之價格編製，後一種則係根據物品十種之價格編製者。兩種指數之變動大致相同，惟潘遜氏指數之變動較為劇烈。氏之研究目的僅為測度其指數之循環變動，故用勃蘭特斯脫指數作為基數，由此基數計算所得之差數，即可表示循環變動；是無異以勃蘭特斯脫指數用作普通之趨勢綫也。

(註一)參看 Harvard Economic Service 出版“Review of Economic Statistics”，

1923年四月號第73—74頁及1923年七月號第102—104頁。

(註二)參看第六章。

潘遜氏又曾用同法以求利率之循環變動。氏所用循環變動之資料，為六十日至九十日期及四月至六月期兩種商業票據之貼現率，所用基數為十種鐵路債券之利息。該三種數列之變動，大致相同，惟商業票據貼現率對於商情之感應性更為靈敏，而鐵路債券利息之變動則較平穩，故用後者作為前兩者之基數。惟計算差數時，略需調整工作，俾鐵路債券利息綫之交叉點（即數列所繪成之曲綫與基綫相交之點）得與商業票據貼現率綫之交叉點相合。調整之法可在本章末所列參考書中得之。

分析時間數列所需平減價值之手續

經濟資料之數值多有用元、鎊、法郎等貨幣單位表示者，此種數列每因物價之漲落而失却其比較性。例如一九一三年美國二十七州所訂建築工程合同之總值為858,000,000美金元，一九二二年總值為3,344,000,000美金元（註）。驟視之，一九二二年之建築數量四倍於一九一三年，其實不然，蓋建築合同之價值，不僅視建築之實際數量而定，且視建築材料之價格及建築工資而定，而自一九一三年至一九二二年間材料價格與工資皆增高甚鉅也。吾人倘欲根據合同價值以測度建築數量之變動，則不可不用建築材料之價格及工資校正之，此即平減價值(deflation)之謂也。

平減價值時，最重要之問題為選擇一相當之平減指數(deflating index)。在討論物價指數一章內，吾人亦嘗用生活費指數平減貨幣工資(money wages)，以求真實工資(real wages)。在前例中，吾人可用建築材料批發價格之指數以為平減建築合同價值之用。其平減手續如下：

(1) 年 份	(2) 建築合同之價值 (美金百萬元)	(3) 建築材料 物價指數	(4) 建築合同之平減價值 (美金百萬元)
1913	858	100	858
1922	3344	168	1990

(註)此項數字為 F. W. Dodge & Co. 所編。

第四行內之平減價值，係由第三行內之指數除第二行內之合同價值，乘以100而得。經此平減手續後，一九二二年之實際價值即已移至一九一三年基期價格之水平綫上。由平減價值所示一九二二年建築之真實數量，僅較一九一三年高出一倍餘。

在平減合同價值時，如能將工資之變動一併予以校正，則所得之結果，必更準確。美國電話電報公司曾用建築材料價格及工資之混合指數，以平減建築合同之價值。茲將各年數字列表於下：

表七十四
建築合同之實際價值及平減價值

	建築合同之實際價值 (美金百萬元)	美國電話電報公司 建築成本指數 (1914=100)	建築合同之平減價值 (美金百萬元)
1914			
一月	51.1	100	51.1
二月	39.1	101	38.7
三月	58.9	102	57.7
四月	79.7	101	78.9
五月	72.0	100	72.0
六月	81.8	100	81.8
七月	72.0	100	72.0
八月	77.3	100	77.3
九月	47.1	100	47.1
十月	53.4	99	53.9
十一月	45.5	99	46.0
十二月	42.3	98	43.2
1915			
一月	43.3	99	43.7
二月	48.8	99	49.3
三月	75.6	99	76.4
四月	76.5	100	76.5
五月	77.1	102	75.6
六月	92.2	104	88.7
七月	94.7	104	91.1
八月	90.4	102	88.6
九月	82.0	103	79.6
十月	88.6	104	85.2
十一月	88.0	107	82.2
十二月	82.9	112	74.0
1916			
一月	62.8	118	53.2
二月	66.3	121	54.8
三月	94.5	124	76.2
四月	100.9	127	79.4
五月	131.4	127	103.5
六月	140.7	128	109.9
七月	114.4	125	91.5
八月	127.0	122	104.1
九月	132.2	123	107.5
十月	151.4	125	121.1
十一月	122.4	128	95.6
十二月	112.9	136	83.0

表 七 十 四 (續)

	建築合同之實際價值 (美金百萬元)	美國電話電報公司 建築成本指數 (1914=100)	建築合同之平減價值 (美金百萬元)
1917			
一月	90.8	138	65.8
二月	95.2	141	67.5
三月	132.7	144	92.2
四月	148.5	148	100.3
五月	157.6	155	101.7
六月	206.5	160	129.1
七月	159.2	171	93.1
八月	165.6	170	97.4
九月	122.5	166	73.8
十月	154.5	154	100.3
十一月	94.3	155	60.8
十二月	90.8	155	58.6
1918			
一月	161.6	156	103.6
二月	137.3	161	85.3
三月	115.3	162	71.2
四月	128.9	164	78.6
五月	120.4	167	72.1
六月	248.2	171	145.1
七月	153.0	176	86.9
八月	146.4	180	81.3
九月	124.5	180	69.2
十月	166.1	183	90.8
十一月	130.3	183	71.2
十二月	57.3	186	30.8
1919			
一月	54.1	186	29.1
二月	98.7	186	53.1
三月	121.8	186	65.5
四月	188.8	178	106.1
五月	234.7	178	131.9
六月	285.4	180	158.6
七月	317.7	183	173.6
八月	295.1	194	152.1
九月	228.7	199	114.9
十月	307.5	201	153.0
十一月	220.6	207	106.6
十二月	226.7	221	102.6
1920			
一月	226.1	227	99.6
二月	205.3	240	85.5
三月	302.2	248	121.9
四月	304.9	254	120.0
五月	264.9	264	100.3
六月	260.1	266	97.8
七月	204.5	265	77.2
八月	202.7	267	75.9
九月	182.2	261	69.8
十月	178.6	256	69.8
十一月	132.9	250	53.2
十二月	100.1	236	42.4

表七十四 (續)

	建築合同之實際價值 (美金百萬元)	美國電話電報公司 建築成本指數 (1914=100)	建築合同之平減價值 (美金百萬元)
1921			
一月	111.6	231	48.3
二月	102.4	223	45.9
三月	163.8	221	74.1
四月	221.3	212	104.4
五月	240.9	206	116.9
六月	226.2	201	112.5
七月	212.2	197	107.7
八月	220.4	194	113.6
九月	244.9	188	130.3
十月	222.4	186	119.6
十一月	190.8	187	102.0
十二月	198.3	186	106.6
1922			
一月	166.3	185	89.9
二月	172.7	185	93.4
三月	293.4	187	156.9
四月	353.0	186	189.8
五月	361.8	188	192.4
六月	342.4	195	175.6
七月	350.1	199	175.9
八月	322.0	197	163.5
九月	271.5	201	135.1
十月	253.1	201	125.9
十一月	243.4	197	123.6
十二月	214.2	196	109.3
1923			
一月	218.7	202	108.3
二月	229.9	209	110.0
三月	338.2	212	159.5
四月	362.9	217	167.2
五月	373.9	217	172.3
六月	323.6	221	146.4
七月	274.2	219	125.2
八月	253.1	219	115.6
九月	253.5	217	116.8
十月	319.9	216	148.1
十一月	289.3	211	137.1
十二月	267.9	215	124.6

實際價值及平減後之價值，均繪於圖六十內。

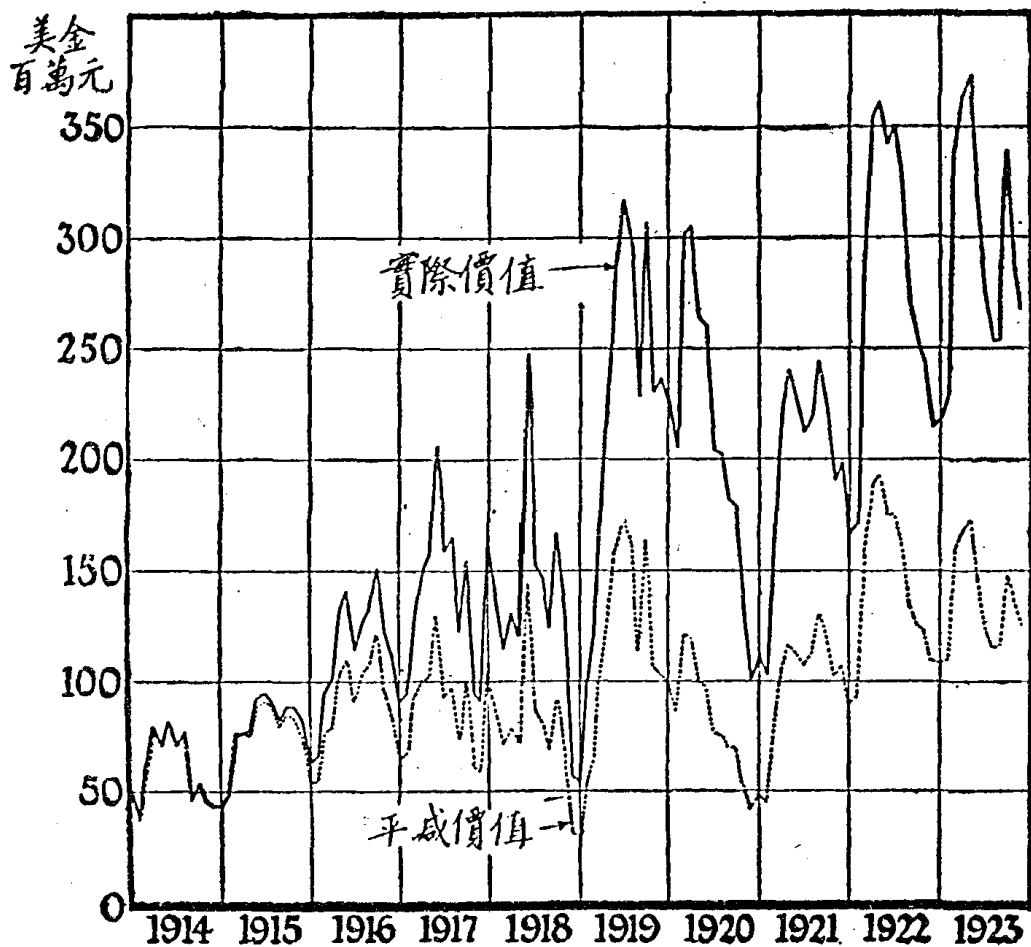
價值數列往往受物價變動之影響，故在分析之先，應將此項因素校正之，惟各種價值數列，性質不同，一種指數決不能適合於各種數列，以爲平減之用。美國勞工統計局所編之批發物價指數，常被用以平減美元表示之經濟資料，但頗有應用極不適當者。例如用批發物價指數以作平減貨幣工資之用，即不免有悖理之處。一數列所用之平減指數，應爲該

數列受直接影響之一種物價所編成之指數。

紐約聯邦準備銀行統計部所用平減紐約以外各城市銀行票據交換額之方法，頗饒興趣。該行所用之平減指數，係由下列各種指數依所列權數加權混合編製而成者。

房租指數	1
批發物價指數	2
工資指數	$3\frac{1}{2}$
生活費指數	$3\frac{1}{2}$

依此編製之混合指數，確能消除銀行票據交換數額所受物價及工資變動之影響。該項平減後之票據交換數額，常用作美國貿易數量消長之指標焉。



圖六十. 1914—1923年各月建築合同之實際價值與平減價值之比較

平減價值常為研究價值數列之初步手續，價值數列經平減後，乃可應用本章及以後各章內所述種種方法着手分析。

參 考 書

BOWLEY, A. L. *Elements of Statistics* (132—169).

DAVIES, G. R. *Introduction to Economic Statistics* (100—130).

KARSTEN, KARL. *Charts and Graphs* (477—481).

LIPKA, JOSEPH. *Graphical and Mechanical Computation*.

MOORE, H. L. *Forecasting the Yield and the Price of Cotton* (28—37).

PEARL, RAYMOND. *Medical Biometry and Statistics* (332—341).

PERSONS, W. M. *Indices of Business Conditions*. Review of Economic Statistics. Prel. Vol. I, 1919 (1—107).

RUNNING, T. R. *Empirical Formulas*.

STEINMETZ, C. P. *Engineering Mathematics* (209—255).

(References to books dealing with the method of least squares are given at the end of Appendix A.)

第八章 時間數列之分析：季節變動 及循環變動之測度

在分析時間數列時，測度長期趨勢僅屬其中問題之一。時間數列又每受季節性及循環性週期變動之影響，而此種變動對於商業之關係常較長期趨勢尤為密切。吾人當前之問題即為研究此種週期變動。下表所載之時間數列含有顯著之季節變動之特性，用以說明測度季節變動之方法，最為適宜。

表 七 十 五

1910—1921年美國雞蛋產主所得之平均價格(註)

(表內各月價格係每月一日每打雞蛋值美金分之

數，各年平均價則係根據每月一日之價格計算)

	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

(註)此項資料係美國農業部所編製。

每月項目平均法

測度季節變動最簡單之方法，係將每月項目依照月份分別相加而求其算術平均數。上表數字包含十二年，故每月均含有十二項目。表七十六第(2)欄內數字即每月之平均數也。

表七十六

根據實際價格之算術平均數計算季節指數

(1)	(2)	(3)	(4)
月份	各月蛋價之平均數 (每打值美金分)	校正長期趨勢後之平 均數(每打值美金分)	季節指數
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
平均	29.357	100.0

第(2)欄內所列之簡單算術平均數，已足表現蛋價季節變動之程度。自一月至四月價格慘落，旋即逐漸升騰，七月稍見回跌，至十二月而達最高價。此項平均價格，如未受長期趨勢之影響，即可化為百分數，以測度季節變動。但自一九一〇年至一九二〇年間蛋價顯有增漲之趨勢，故知各月之平均數不僅受季節之影響，且受長期趨勢之影響。如僅就長期趨勢而論，則各年二月份之價格例應高於一月份，三月份之價格應高於二月份，其他各月份依此類推，故須將各月份平均數所受長期趨勢之影響先予校正，然後可用以測度季節變動。

自一九一〇年至一九二一年之十二年間，每年蛋價之平均數若以直綫配合之，可求得該時期內每打蛋價每年平均增加數為2.32分。每月

平均增加數為2.32分之十二分之一，即0.193分。各月份之平均價格所受長期趨勢之影響，應恰為此數。如僅有長期趨勢單獨存在，則二月份之平均價格應較一月份高出0.193分，三月份之平均價格應較二月份高出此數，其他各月份亦如之。惟此長期趨勢之因素不難消除，吾人可任擇一月為基月，而將其他各月數字校正之。如以正月為基月，可由二月之平均價格減去0.193分，由三月之平均價格減去0.386分，由四月減去0.579分，其他各月依此類推。如以十二月為基月，則須將十一月之平均價格加上0.193分，十月之平均價格加上0.386分，餘類推。但此處宜以一年中間之月份作為基月。如以六月為基月，吾人可由七月之平均價格減去0.193分，八月減去0.386分，至十二月止皆依此推算；同時將五月之平均價格加上0.193分，四月加上0.386分，至一月止皆依此推算。表七十六第(3)欄所列即業已校正之各月平均價格也。

上項數字係表示十二年間之季節變動，故可用作季節指數，但為便於應用起見，當以十二月之平均價格(29.357)作為基數，將各月平均價格化為百分數。表七十六第(4)欄所列即係以百分數表示之季節指數也。此項指數對於分析時間數列之用途，將於後文說明之。

環比中位數法

測度季節變動之另一方法，係潘遜氏 (Warren M. Persons) 所倡。美國曾有數種重要統計採用此法，其較著者為潘遜氏所編之商情循環指數。此法第一步為計算環比(link relative)，所謂環比者，乃將每月數字以其前一月數字為基數計算而得之百分比也。例如一九一〇年一月之數字應以一九〇九年十二月為基數化為百分比；一九一〇年二月之數字應以一九一〇年一月為基數化為百分比；其他各月依此類推。下表所列係根據表七十五內數字求得之環比。

表七十七
各月蛋價之環比

月 份	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	99.2	113.4
八 月	96.7	109.2	104.2	101.2	103.4	101.2	105.1	105.3	112.1	106.8	109.0	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

環比求得後，吾人尚須斷定各年同一月份價格與前一月份價格之平均關係，如一月之與十二月，二月之與一月等是。此種關係，在每月項目平均法中，係先求每月價格之平均數，再化爲百分比而得之。如用潘遜氏之方法則須求每月之環比中位數 (median link relatives)。其法係將一月份之十二個環比，依數字大小排列，而測定其中位數，是爲一月份之環比中位數；其他各月亦依此計算。表七十八第(2)欄所列即各月之環比中位數也。

表七十八
用環比中位數法計算季節指數

(1) 月 份	(2) 各月蛋價之 環比中位數	(3) 鎖 比	(4) 校正後之鎖比	(5) 季 節 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	71.69	100.0

此項環比中位數，雖可測度每月變動，惟因基期逐月變更，不易比較，故更須擇一固定基期，將此項環比化爲百分數。基月既定，即當進而計算各月之鎖比 (chain relatives)。如以一月爲基期，則一月之鎖比爲 100，二月爲 90.1，與其環比中位數相同，因二月之環比本以一月爲基期也。三月之環比爲 80.5，係以二月爲基期；故以一月爲基期之鎖比應爲 80.5×90.1 ，即 72.5。其他各月亦可依此以各月之環比中位數乘前一月之鎖比而得。表七十八第(3)欄所列即各月之鎖比也。

雞蛋之價格若不受長期趨勢之影響，則十二月之鎖比與一月之環比中位數相乘之積，應爲 100，但此兩者一爲 1.014，一爲 103.4，其相乘之積爲 104.8，此因各月之環比中位數俱受長期趨勢之影響，故有此差誤。且此差誤爲累積差誤 (cumulative error) 之結果，各月環比中位數化爲鎖比時，均須與其前一月之鎖比相乘，至十二月之環比化爲鎖比時，相乘次數乃至十二次之多，故其差誤係累積的。此與按照固定利率，計算複利時本利和金額之累積情形正復相同。此處之差誤 (即因長期趨勢而發生之每月增加率) 與複利中之利率無異，而可用同樣公式求得之。以本金若干元 (P_0) 按固定利率 (r) 複利計算，則求 n 年後本利和 (P_n) 之公式如下：

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

故若已知本利和之金額，求利率之公式應爲：

$$r = \sqrt[n]{\frac{P_n}{P_0}} - 1$$

吾人如將本例求得之乘積代入上式，得：

$$104.8 = 100(1+r)^{12}$$

$$r = \sqrt[12]{\frac{104.8}{100}} - 1$$

$$= 1.004 - 1$$

$$= .004$$

由是推知因長期趨勢而發生之每月增加率為 .004。吾人如欲校正此長期趨勢，可將二月之鎖比以 1.004 即 $(1+r)$ 除之，三月以 1.008 即 $(1+r)^2$ 除之，餘類推，其十二月之鎖比應以 1.044 即 $(1+r)^{11}$ 除之。如此則各月鎖比所受長期趨勢之影響業已銷除，而可用以測度季節變動矣。表七十八第(4)欄所列即校正後之各月鎖比也。

惟季節變動之指數宜以全年各月之平均數為基數，不應以一月份數字為基數，蓋以平均數為基數時，所得各月百分數之和應為 1200，各月數字之解釋，似較簡易也。上表內各月數字，以其平均數(71.69)為基數，用百分數表示之，即得表七十八第(5)欄所列之指數。

移動平均數法

測度季節變動之第三法，乃利用移動平均數測度之方法。其法係先求十二個月之移動平均數以銷除季節變動之影響，然後將原有數列除以各月相當之移動平均數，以顯示其季節變動。季節變動以十二個月為一週，週期有定，故用移動平均數在此尚為適宜。惟因季節變動之範圍(即價格漲落之幅度)年有不同，由移動平均數所得之綫，仍不能完全剔除季節之影響，故須於求得各月實際項目與移動平均數相除所得之百分比後，更須平均之。由此所得之平均數即可用作計算季節變動指數之根據。

惟移動平均數集中之日期應與各月實際項目之日期一致，此則又須作第二次平均之手續。例如上表所列雞蛋價格俱係每月一日之價格，而一九一〇年十二個月之平均數，則集中於六月十五日，一九一〇年二月至一九一一年一月十二個月之平均數，集中於七月十五日，故欲求七月一日之平均數，又須將此兩項平均之。經過此兩重平均之後，所得各數方可與原來數字相比。下表所列即經過兩次平均後所得每月一日之移動平均數。

表 七 十 九

1910—1921年蛋價之移動平均數

(十二個月移動平均數,並已以兩個月移動平均數校正為每月一日雞蛋每打值美金分之平均價格)

月 份	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	20.79	22.91	22.06	22.86	31.59	37.38	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	19.46	21.94	20.99	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	21.63	21.88	22.64	21.84	25.60	35.10	40.31	44.84	47.26
九 月	21.63	20.20	21.16	22.32	22.55	21.74	26.51	35.93	39.96	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	22.02	29.20	36.68	40.77	46.67	41.81

由是可將各月之原來價格除以各月相當之移動平均數而計算其百分比,則得下表。

表 八 十

雞蛋之實際價格與其十二個月移動平均數之百分比

月 份	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	79.1	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

任何月份之百分比各年頗有差異,其所表示與該月份移動平均數之關係各年並不一致。如一月份之原來價格雖皆高於其移動平均數,但其百分比介乎125.4與151之間,而並不固定。故各月份之十一個百分比尚須求其平均數,方可求出指數。此處所用平均數或為算術平均數,或為中位數,皆無不可。由此兩種平均數所得指數,見下表第(2)(3)兩欄。惟此種指數未經修正,不便比較,故須將十二個月之指數相加,以其平均數

爲基數，分別求各月指數。第(4)(5)兩欄所列即係此兩種修正後之指數也。

表 八 十 一

用移動平均數法計算蛋價季節指數

(1)	(2)	(3)	(4)	(5)
月 份	算術平均數 (未修正者)	中位數 (未修正者)	算術平均數 (已修正者)	中位數 (已修正者)
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	77.3	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	89.7	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

實際數值對長期趨勢值之比率平均法

關於季節指數之計算，近又發現一優良之方法(註)。此法係先將資料配合一適當之長期趨勢綫，或爲曲綫，或爲直綫，次將每月實際數值除以長期趨勢綫所得同月之常態數值，而分別求其百分比，最後以每月之各個百分比平均之。此法與移動平均數法相似，惟各月之百分比係與長期趨勢之常態數值(normal value)相比而得之耳。在斷定各月之平均數值時，可應用複式次數表(multiple frequency table)。此表可助吾

(註)此法亦爲 Helen D. Falkner 女士(著有“The Measurement of Seasonal Variation”一文，具載於“Journal of the American Statistical Association”一九二四年六月號167—179頁)及 Lincoln W. Hall 氏(著有“Seasonal Variation as a Relative of Secular Trend”一文，具載於“Journal of the American Statistical Association”一九二四年六月號156—166頁)所各自發現者。

人觀察季節變動之存在與否及其變動之情況，俾可擇定一種適當之平均數，作為各月之代表。茲仍以前項蛋價為例，說明此法之應用。

表七十五所列一九一〇年至一九二一年每年之平均蛋價，可以一直線的長期趨勢綫配合之。此綫之方程式為：

$$y = 29.45 + 2.32x$$

此式之原點在一九一六年一月一日。每月之常態數值可依前章所述各法計算之。常態數值求得後，即將原來價格除以各月之常態數值，而求其百分比。本例中蛋價資料共有十二年，故此項百分比每月共有十二個。一月份之十二個百分比介乎103.5與194.8之間，五月份之十二個

	一月	二月	三月	四月	五月	六月	七月	八月	九月	十月	十一月	十二月
170 以上	/	/										
165 至 169.9	//											
160 至 164.9												//
155 至 159.9	/											/
150 至 154.9												/
145 至 149.9	//	/										
140 至 144.9		//	/								//	/
135 至 139.9	/										/	
130 至 134.9		/									/	//
125 至 129.9										/		/
120 至 124.9	/	//								/		/
115 至 119.9	///		///							//	///	/
110 至 114.9		//		/	/	/			/	/	/	
105 至 109.9		/					/		//			/
100 至 104.9	/	/	/			/		//	/	/	//	/
95 至 99.9			/	/	/		/	//	/	//		
90 至 94.9		/		/	//	//	/				//	
85 至 89.9			//	//	/		//	/	//	//		
80 至 84.9			/	/	/	/		//	/			
75 至 79.9			/	/	/	//	/		/	//		
70 至 74.9			//	/		/	//	/	//			
65 至 69.9				/	//	/	/	//	/			
60 至 64.9					/	/	/	/				
55 至 59.9				//	/	/	/	/				
50 至 54.9							/					
50 以下				/	/	/						

圖六十一. 各月蛋價對常態數值百分比之次數分配
(根據1910—1921年之資料編製)

百分比則介乎48.2與113.7之間，餘類推。圖六十一所示之複式次數表，係由每月百分比之次數分配編製而成。

觀此表可見蛋價含有顯著之季節變動。冬季各月份之百分比常高於春夏兩季，故可斷定季節變動之存在，而可以指數表示之。

此表又可助吾人選擇一種適當之平均數以測度此季節變動。倘各月次數之分配無顯著之集中趨勢，則用中位數似覺不妥，因一二項目之增減，足使中位數之數值受重大影響；算術平均數因易受極端項目之影響，亦所不取；若將兩法混合採用而取中間數項之算術平均數，似可兼兩者之長而補其短。倘參考複式次數表尚不能抉擇一適宜之平均數，則可由每月之次數分配，編製數種指數，比較其結果，而擇其最適宜者。吾人可由前例中各月百分比之次數分配，每月求得六種平均數。下表所列‘未修正指數’即此六種不同之平均數也。茲更將每種平均數之全年十二個月份數字相加而平均之，以此平均數為基數，分別計算各月之百分比，即得下表內之已修正指數（此項平均數係根據各月實際數值對其長期趨勢值之百分比直接計算，非由圖六十一之次數分配求得者）。

表八十二

未修正及已修正之蛋價季節指數

(根據實際數值對長期趨勢值之百分比計算)

月 份	未修正指數						已修正指數					
	(根據中間項數2,4,6,8,10,12項計算者)						(根據中間項數2,4,6,8,10,12項計算者)					
	2	4	6	8	10	12	2	4	6	8	10	12
一 月	140.5	138.2	138.0	139.2	139.8	141.4	142.1	139.4	138.4	139.1	139.6	140.5
二 月	120.8	121.6	123.4	123.8	123.6	125.7	122.2	122.6	123.8	123.7	123.5	124.9
三 月	92.8	95.0	96.9	97.1	96.7	98.3	93.9	95.8	97.2	97.1	96.6	97.7
四 月	80.0	79.0	79.1	78.1	78.3	78.9	80.9	79.7	79.3	78.1	78.2	78.4
五 月	79.8	79.3	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	78.6	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	77.5	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	90.1	90.2	90.0	89.3	89.9	90.0	90.1	90.1	89.5
十 月	100.7	102.2	102.4	102.4	102.3	102.4	101.8	103.1	102.7	102.4	102.2	101.8
十一 月	116.6	115.7	116.0	117.5	117.3	117.3	117.9	116.7	116.3	117.5	117.2	116.6
十二 月	133.0	134.0	134.7	135.6	135.5	135.2	134.6	135.1	135.1	135.6	135.4	134.4

上表中‘未修正指數’項下第一欄之數字，為每月百分比次數分配之中間兩項之算術平均數；第二欄之數字為中間四項之算術平均數，餘類推。

在本例中蛋價之季節變動甚為顯著，且其變動亦頗有規律，故根據項數不等之中間各項所求得之各種指數，俱大同小異。此處殆以中間四項平均所得之指數最為適用。大體言之，中間三四項或五項之平均數，每較中位數或全體項目之平均數為穩妥。複式次數表中，每月次數分配之集中程度愈高，則求指數時所用中間項目之項數可以愈少。

茲將根據上述四種測度季節變動之方法所求得蛋價之季節指數，併列於下，以資比較。

表 八 十 三

已修正之蛋價季節指數

月 份	移動平均數法 (中位數)	每月項目平均法	環比中位數法	實際數值對長期趨勢 值之比率平均法 (中間四項之平均)
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	100.0	100.0	100.0	100.0

各種季節指數之比較

此四種指數所表現蛋價之季節變動，頗屬相似。其中參差較多者為

二三兩月份之指數，而以三月份最高與最低指數之相差為尤鉅，竟達百分之四以上。至於究應採用何種方法最為適宜，並無絕對標準可言，惟有數點可申述之。每月項目平均法之計算手續雖頗簡便，但若所用資料包含極端變動，或在資料時期內數值變動甚烈，則不宜採用之。環比中位數法計算頗費時間，如資料勻整，變動有序，則此法未見較優於其他各法(註)。就中以移動平均數法及實際數值對長期趨勢值之比率平均法最切實用，而後者計算手續簡單，且求百分比之平均數，亦比求實際數值之平均數為合理焉。

各月常態數值之計算法

在討論校正時間數列之長期趨勢及季節影響以前，吾人可先將關於計算常態數值時，由年數字化為月數字之方法，略加說明。此處所涉及者祇算術方面簡單之計算手續而已。

茲選取下列數字，以說明其計算方法。

年份	某種物品之常態產量
1910	186
1911	330
1912	474
1913	618

(註)潘遜氏嘗列舉環比中位數法之優點如下(其第一第二兩點為移動平均數法及實際數值對長期趨勢值之比率平均法所同具)：

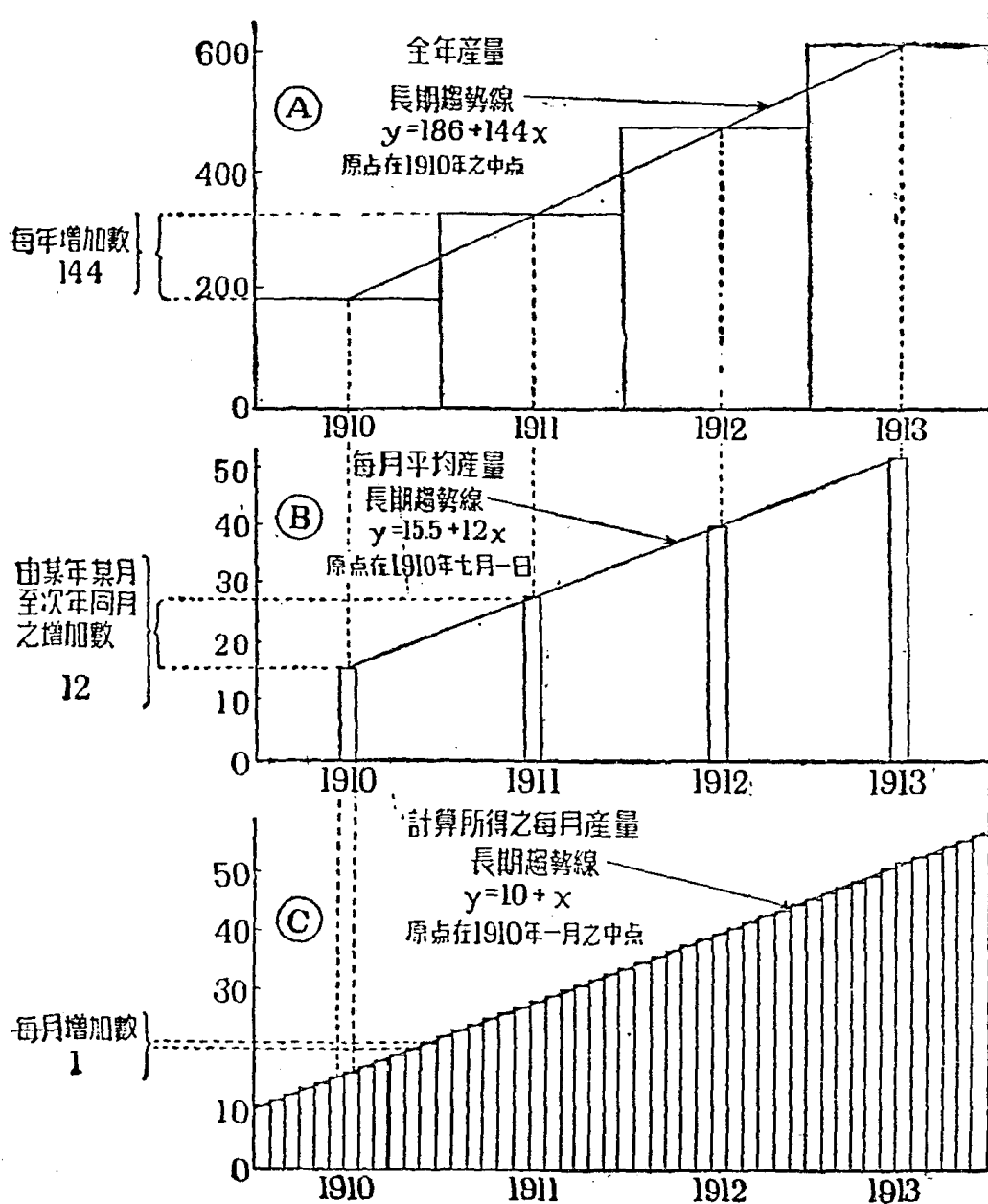
1. 環比之次數分配，可助吾人斷定季節變動之有無規律及其變動之程度。
2. 用中位數(或中間項目之平均數)可減弱非季節性之強烈影響。
3. 應用此法以測度季節變動，可不受統計數列性質上之嚴格限制，例如某時期內包括五十城市之銀行票據交換額(bank clearings)之環比可與另一時期內包括一百城市之銀行票據交換額之環比湊合計算而無妨礙。

(摘錄“Journal of the American Statistical Association”一九二三六月號第717頁)

圖六十二A即係根據此項資料繪製者。

前項數字為四個連續年份之常態產量，其長期趨勢綫可用下列方程式表示之。

$$y = 186 + 144x$$



圖六十二. 說明每月常態數值計算法之圖解
(在A,B兩圖中,x之單位為年 在C圖中,x之單位為月。)

此式之原點係在一九一〇年。每年之數字既為全年之產量，故其集中點應在每年之中點（即七月一日）。長期趨勢綫即為經過各該點而繪成者。此綫之斜度為144，即此時期內每年產量之增加數也。

由年數字化為月數字之手續，可分兩步驟。第一步係先求以按月平均數表示每年產量之方程式，將前式中186及144兩項各以12除之即得。此方程式為：

$$y = 15.5 + 12x$$

該式之原點仍在一九一〇年之中點。以圖表示之如圖六十二 B。圖中每一狹長方格表示每年六月十五日至七月十五日之產量而集中於七月一日。長期趨勢綫之斜度為12，此數表示某年某月較上年同月所增加之產量，故以一九一〇年七月一日為中點之一個月內產量為15.5；以一九一一年七月一日為中點之一個月內產量為27.5；餘類推。於此所當注意者，在此式中 x 之單位仍為一年。

第二步為計算每月之常態數值或長期趨勢值。自一九一〇年七月一日至一九一一年七月一日常態產量之增加數既為12，則一九一〇年七月一日至同年八月一日之增加數應為1。吾人試將上式橫坐標之單位由一年改為一月，即得

$$y = 15.5 + x$$

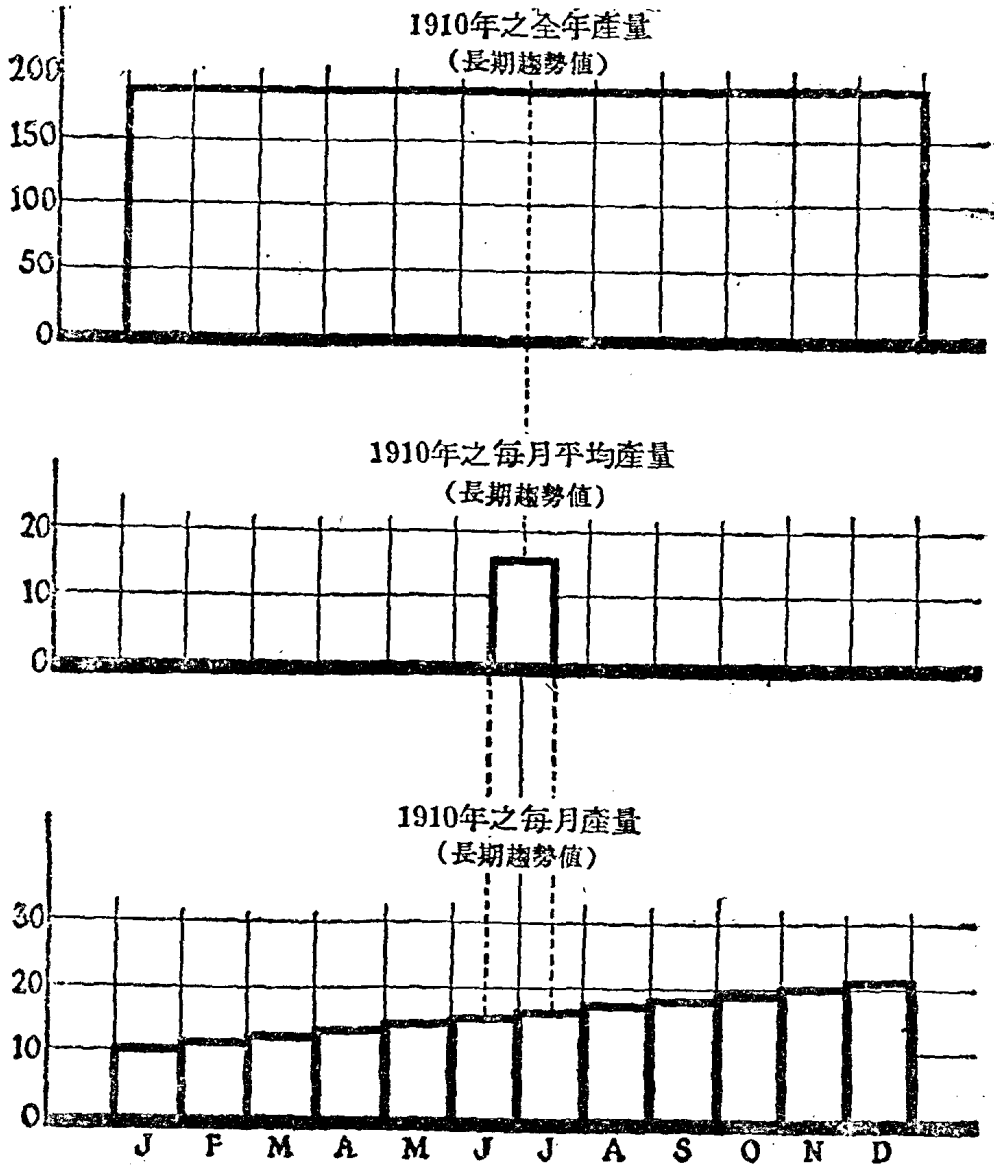
該式之原點仍在一九一〇年七月一日，但此原點在此不甚適用，因每月之產量通常應集中於月之中點也。今一九一〇年七月一日之縱坐標為15.5，而每月之增加數為1，則一九一〇年一月十五日之縱坐標應為10。故將原點移至一九一〇年一月十五日以後，其公式當為：

$$y = 10 + x$$

此即計算每月常態產量之公式也。圖六十二 C 係表示一九一〇年至一九一三年間每月常態產量之關係。

圖六十三係表示一九一〇年全年之常態產量、每月之平均常態產

量與每月之常態產量三者相互關係之詳細圖解(註)。該圖所最注重者，在於顯示每月平均產量改算為每月常態產量時，因日期上有半個月之參差，應予校正之一點。



圖六十三. 全年常態產量、每月平均常態產量與每月常態產量之關係

上例中因長期趨勢之存在所得之每月增加數，等於全年增加數除

(註)圖六十二及六十三係 D. H. Davenport 氏所設計繪製者。

以144。各種產量資料，其每年產量如係十二個月產量之總和，則求每月增加數之方法與上例同。物價數列(price series)或他種數列，其年數字常為十二個月之平均數，故每月之增加數適為全年增加數之十二分之一。如此則前述由年數字化為月數字之兩步驟中，僅用第二步驟足矣。

循環變動之測度

吾人既知測度長期趨勢及季節變動之方法，更須進而求校正長期趨勢及季節變動之方法，以為測度循環變動之用。在各種經濟研究中，循環變動恆居最重要之地位，而測度此種變動遂為分析時間數列問題之中心。吾人可用一九〇一年至一九二三年美國烟煤產量之數字，加以分析，以說明其計算手續。

先將每年產量之數字配合一適當之長期趨勢綫。此綫之方程式似以採取下式最為相宜。

$$\log y = a + bx + cx^2$$

上式中 a, b, c 三個常數之值可用最小平方法(method of least squares)求之。此項常數求得後，則表示長期趨勢綫之方程式為：

$$\log y = 2.646052 + .014226x - .0010246x^2$$

由此式即可求得每年之常態產量。茲將此項產量及每年之實際產量列入表八十四，以資比較。

烟煤每年產量所表現之循環變動，可由表中測定之，惟吾人當前之問題乃為測度其每月產量所受循環變動之影響。

測度此項循環變動之步驟，第一步應先斷定長期趨勢綫上各月份所有之縱坐標，此項縱坐標即各月之常態產量也。惟前此長期趨勢綫係配合於年數字上者，故又須將此綫改為以月為單位之綫。

該項數列中每期之數字均應視為集中於該期之中點。例如一九一三年全年之產量應視為集中於一九一三年七月一日；一九一三年一月

表 八 十 四

1901—1923年美國烟煤實際產量與常態產量之比較

(1) 年份	(2) 實際產量 (淨重百萬噸)	(3) 長期趨勢值 之對數	(4) 常態產量 (淨重百萬噸)	(5) 實際產量對於常 態產量之百分比
1901	225.83	2.365589	232.05	97.3
1902	260.22	2.401332	251.96	103.2
1903	282.75	2.435025	272.29	103.8
1904	278.66	2.466670	292.87	95.1
1905	315.06	2.496265	313.52	100.5
1906	342.87	2.523810	334.05	102.6
1907	394.76	2.549307	354.25	111.4
1908	332.57	2.572754	373.90	88.9
1909	379.74	2.594153	392.78	96.7
1910	417.11	2.613502	410.68	101.6
1911	405.91	2.630801	427.36	95.0
1912	450.10	2.646052	442.64	101.7
1913	478.44	2.659253	456.30	104.9
1914	422.70	2.670406	468.17	90.3
1915	442.62	2.679509	478.09	92.6
1916	502.52	2.686562	485.92	103.4
1917	551.79	2.691567	491.55	112.3
1918	579.38	2.694522	494.91	117.1
1919	465.86	2.695429	495.94	93.9
1920	568.67	2.694286	494.64	115.0
1921	415.92	2.691093	491.01	84.8
1922	404.51	2.685852	485.12	83.4
1923	545.30	2.678561	477.05	114.3

之產量應視為集中於一月十五日。

一九一三年之長期趨勢值 456.30 百萬噸，應視為集中於一九一三年七月一日之該年常態數值。此值以12除之，得38.025百萬噸，即集中於一九一三年七月一日之該月平均常態產量也。但此項產量既不能與是年六月之實際產量相比，又不能與是年七月之實際產量相比，因實際

產量皆集中於每月之十五日，其間各相差半月也。吾人須用下法以校正此半月之差異。其法以12除一九一四年之常態數值得39.014百萬噸，即為集中於一九一四年七月一日之該月常態產量。此數與38.025之差為.989，亦即一九一三年七月一日之常態產量與一九一四年七月一日之常態產量相距一年間之差量，因長期趨勢之存在而發生者也。故再以12除.989得.0824，此即一九一三年七月一日至一九一四年七月一日之一年間逐月之差量也。一九一三年七月一日至同年八月一日一個月之差量既為.0824，則一九一三年七月一日至同年七月十五日半個月之差量應為此數之半，即.0412。今集中於一九一三年七月一日之該月常態產量既為38.025百萬噸，則以38.025與.0412相加，得38.0662百萬噸，即為集中於一九一三年七月十五日之該月常態產量。至於集中於八月十五日之常態產量應為 $38.0662 + .0824$ ，即38.1486；九月之常態產量應為 $38.1486 + .0824$ ，即38.2310。由一九一三年七月起至一九一四年六月止各月之常態產量，皆可用此差量加減計算。

長期趨勢如係一直綫，則在全時期內每月之常態數值與次月常態數值之差為一常數，故如既定某月之常態數值，其他各月之常態數值即可據此求得之。在本例中實際產量之逐年增加率不等(註)，故常態產量之每年增加數亦隨之而異。表八十五所列，係一九一二年至一九二二年間各年七月之常態數值(集中於七月十五日)及各年間每月之增加數。

計算各月之常態數值時，應注意各年增加率之變動。一九一三年七月至一九一四年六月間每月常態數值之計算方法已如上述，其他各年間每月之常態數值亦可依此計算。

(註)嚴格言之，實際產量之逐月增加率亦不等。今將每十二個月期間之趨勢作為直綫，用插補法求該期內每月之常態數值，自不免含有差誤，但此差誤極微，而計算手續則較用繁複公式計算者，簡便多矣。

表 八 十 五

1912—1922年烟煤常態產量增加率之變動

時 期	常態產量 (淨重百萬噸)	由每年七月起十二個月內每月常 態產量之增加數(淨重百萬噸)
1912年七月	36.9341	.0948
1913年七月	38.0662	.0824
1914年七月	39.0490	.0690
1915年七月	39.8680	.0543
1916年七月	40.5120	.0391
1917年七月	40.9742	.0233
1918年七月	41.2461	.0071
1919年七月	41.3238	-.0090
1920年七月	41.2074	-.0252
1921年七月	40.8970	-.0409
1922年七月	40.3987	-.0560

求得各月常態數值後，即可進而計算季節指數。其計算方法可就上述各法中任擇一種。此處所用者為移動平均數法，所根據之資料為一九一三年至一九二一年間烟煤之每月產量數字(註)。其季節指數如表八十六所示。

測度烟煤產量循環變動之第一步，係將每月之實際數值化為各該月常態數值之百分比。表八十六第(4)欄所列，即係一九一三年及一九一四年兩年間每月之百分比，以說明計算手續者也。此項百分比中仍含有季節變動，故須以第(5)欄內之季節指數校正之。其法係求第(4)欄內各項百分比與第(5)欄內相對項之百分比之差，如前者小於後者，差數為負，反是則為正。第(6)欄內之差數，即由第(4)欄內之百分比減去第(5)欄內之季節指數而得。此項數字即可用為測度烟煤產量循環變動之量數，因其所受長期趨勢及季節變動之影響皆已消除也。至於此項數字內所受意外及‘不規則’變動 (accidental and irregular movements) 之影響雖未剔除，惟此意外變動在長時期內常相沖銷(此種假說對於上例中之煤產

(註)此項烟煤產量之季節指數，係紐約聯邦準備銀行統計科所編。

表八十六

1913—1914年烟煤產量循環指數之計算

年	(1) 月	(2) 實際產量 (淨重百萬噸)	(3) 常態產量 (淨重百萬噸)	(4) 實際產量對常態 產量之百分比 (2)÷(3)	(5) 季節指數	(6) 校正季節變動後 實際產量與常態 產量之百分差 (4)-(5)
1913	一月	42.274	37.503	112.7	107	+5.7
	二月	37.057	37.598	98.6	92	+6.6
	三月	37.536	37.693	99.6	102	-2.4
	四月	34.169	37.788	90.4	83	+7.4
	五月	37.205	37.882	98.2	92	+6.2
	六月	37.405	37.977	98.5	95	+3.5
	七月	38.858	38.066	102.1	97	+5.1
	八月	41.590	38.149	109.0	106	+3.0
	九月	41.424	38.231	108.4	106	+2.4
	十月	46.164	38.313	120.5	113	+7.5
	十一月	43.233	38.396	112.6	104	+8.6
	十二月	41.519	38.478	107.9	103	+4.9
1914	一月	40.191	38.561	104.2	107	-2.8
	二月	35.472	38.643	91.9	92	-.1
	三月	45.455	38.725	117.4	102	+15.4
	四月	23.609	38.808	60.8	83	-22.2
	五月	28.551	38.890	73.4	92	-18.6
	六月	31.412	38.973	80.6	95	-14.4
	七月	34.305	39.049	87.9	97	-9.1
	八月	37.751	39.118	96.5	106	-9.5
	九月	39.019	39.187	99.6	106	-6.4
	十月	37.685	39.256	96.0	113	-17.0
	十一月	33.392	39.325	84.9	104	-19.1
	十二月	35.862	39.394	91.0	103	-12.0

量未必適用),故第(6)欄內之數字即可用為測度循環變動之指數。

上表僅舉示兩年間循環指數之計算方法,其餘各年之指數亦可用此法計算。其自一九一三年至一九二三年各月之實際產量及校正季節

變動後實際產量與常態產量之百分差，見表八十七。

表 八 十 七

1913—1923年間烟煤每月實際產量及校正後實際產量與
常態產量之百分差

	1913		1914		1915	
	實際產量 (淨重百萬噸)	校正季節變 動後實際產 量與常態產 量之百分差	實際產量 (淨重百萬噸)	校正季節變 動後實際產 量與常態產 量之百分差	實際產量 (淨重百萬噸)	校正季節變 動後實際產 量與常態產 量之百分差
一月	42.274	+5.7	40.191	-2.8	37.194	-12.7
二月	37.057	+6.6	35.472	-1.1	29.321	-17.8
三月	37.536	-2.4	45.455	+15.4	31.801	-21.7
四月	34.169	+7.4	23.609	-22.2	29.968	-7.5
五月	37.205	+6.2	28.551	-18.6	30.938	-14.1
六月	37.405	+3.5	31.412	-14.4	33.957	-9.7
七月	38.858	+5.1	34.305	-9.1	35.573	-7.8
八月	41.590	+3.0	37.751	-9.5	38.161	-10.4
九月	41.424	+2.4	39.019	-6.4	40.964	-3.5
十月	46.164	+7.5	37.685	-17.0	44.198	-2.6
十一月	43.233	+8.6	33.392	-19.1	44.737	+7.6
十二月	41.519	+4.9	35.862	-12.0	45.814	+11.1
	1916		1917		1918	
一月	46.593	+8.9	47.969	+10.7	42.227	-4.3
二月	45.187	+20.3	41.353	+9.4	43.777	+14.4
三月	43.829	+6.8	47.869	+15.3	48.113	+14.9
四月	33.628	+3	41.854	+19.4	46.041	+28.8
五月	38.804	+4.0	47.086	+23.1	50.443	+30.4
六月	37.742	-1.7	46.824	+19.4	51.138	+29.0
七月	38.113	-3.0	46.292	+16.0	54.971	+36.3
八月	42.696	-1.7	47.372	+9.5	55.114	+27.6
九月	42.098	-2.3	45.108	+4.0	51.183	+18.0
十月	44.807	-2.7	48.337	+4.8	52.300	+13.7
十一月	44.927	+6.5	47.690	+12.1	43.895	+2.3
十二月	44.098	+5.8	44.037	+4.2	40.184	-5.7
	1919		1920		1921	
一月	42.193	-4.8	49.748	+13.5	41.148	-6.8
二月	32.103	-14.3	41.055	+7.5	31.524	-15.2
三月	34.293	-19.0	47.850	+14.0	31.055	-26.3
四月	32.712	-3.8	38.764	+11.0	28.164	-14.3
五月	38.186	+4	39.841	+4.6	34.057	-8.8
六月	37.685	-3.8	46.095	+16.8	34.635	-10.4
七月	43.425	+8.1	45.988	+14.6	31.047	-21.1
八月	43.613	-4	49.974	+15.3	35.291	-19.6
九月	48.209	+10.7	50.241	+16.1	35.893	-18.1
十月	57.200	+25.5	53.278	+16.5	44.686	-3.4
十一月	19.006	-58.0	55.276	+30.5	36.805	-19.6
十二月	37.235	-12.8	53.257	+26.6	31.627	-25.3

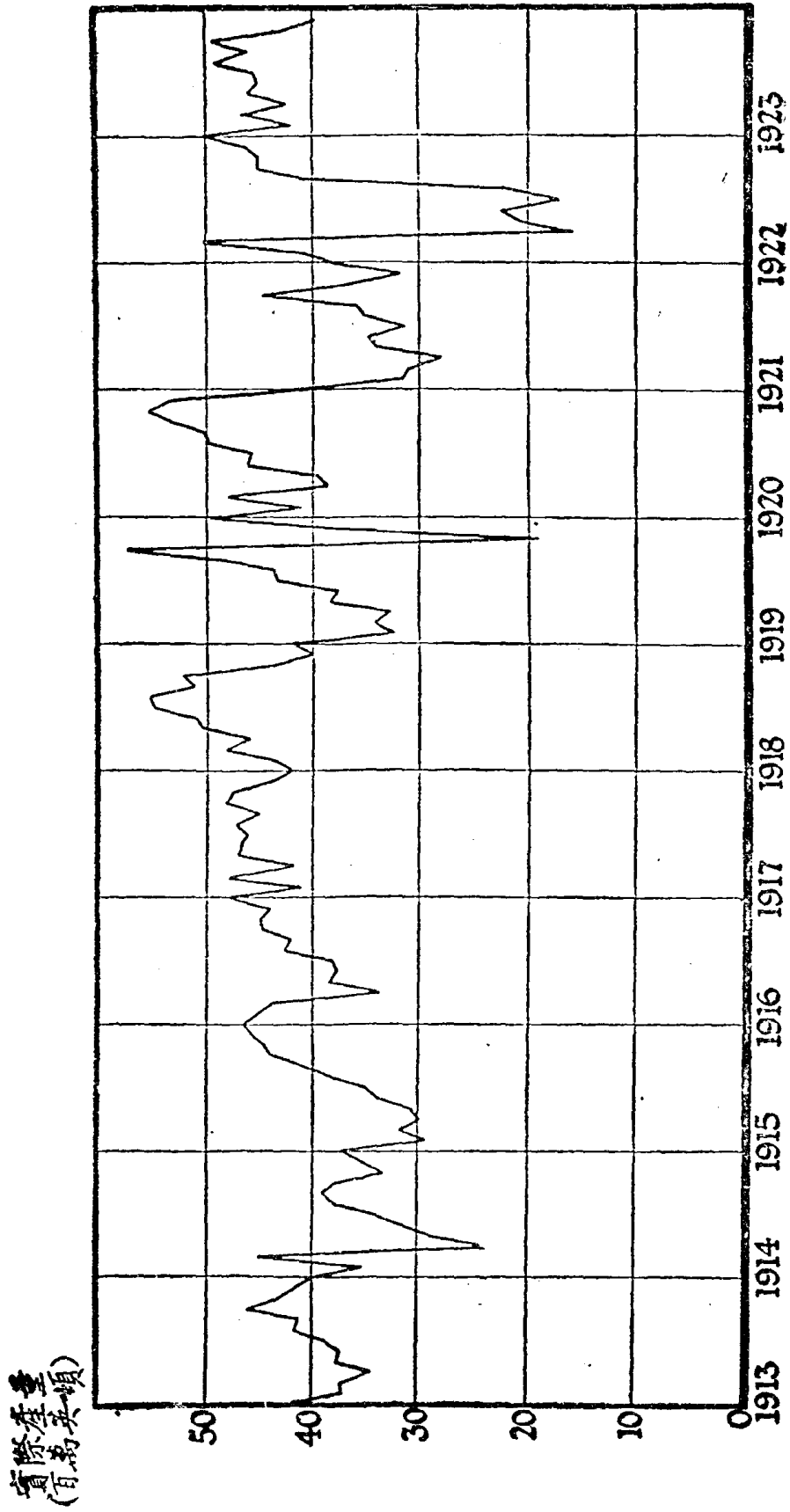
表八十七 (續)

	1922		1923			
	實際產量 (淨重百萬噸)	校正季節變動後實際產量與常態產量之百分差	實際產量 (淨重百萬噸)	校正季節變動後實際產量與常態產量之百分差	實際產量 (淨重百萬噸)	校正季節變動後實際產量與常態產量之百分差
一月	37.600	-14.5	50.123	+18.1		
二月	40.951	+8.8	42.160	+13.4		
三月	50.193	+21.7	46.807	+15.2		
四月	15.780	-44.1	42.564	+23.7		
五月	20.501	-41.4	46.076	+23.7		
六月	22.309	-39.8	45.490	+19.3		
七月	17.003	-54.9	45.644	+17.9		
八月	22.328	-50.7	48.864	+17.2		
九月	40.964	-4.3	46.216	+10.7		
十月	45.173	-.7	49.171	+11.3		
十一月	45.262	+8.7	42.946	+4.7		
十二月	46.450	+12.8	39.707	-2.3		

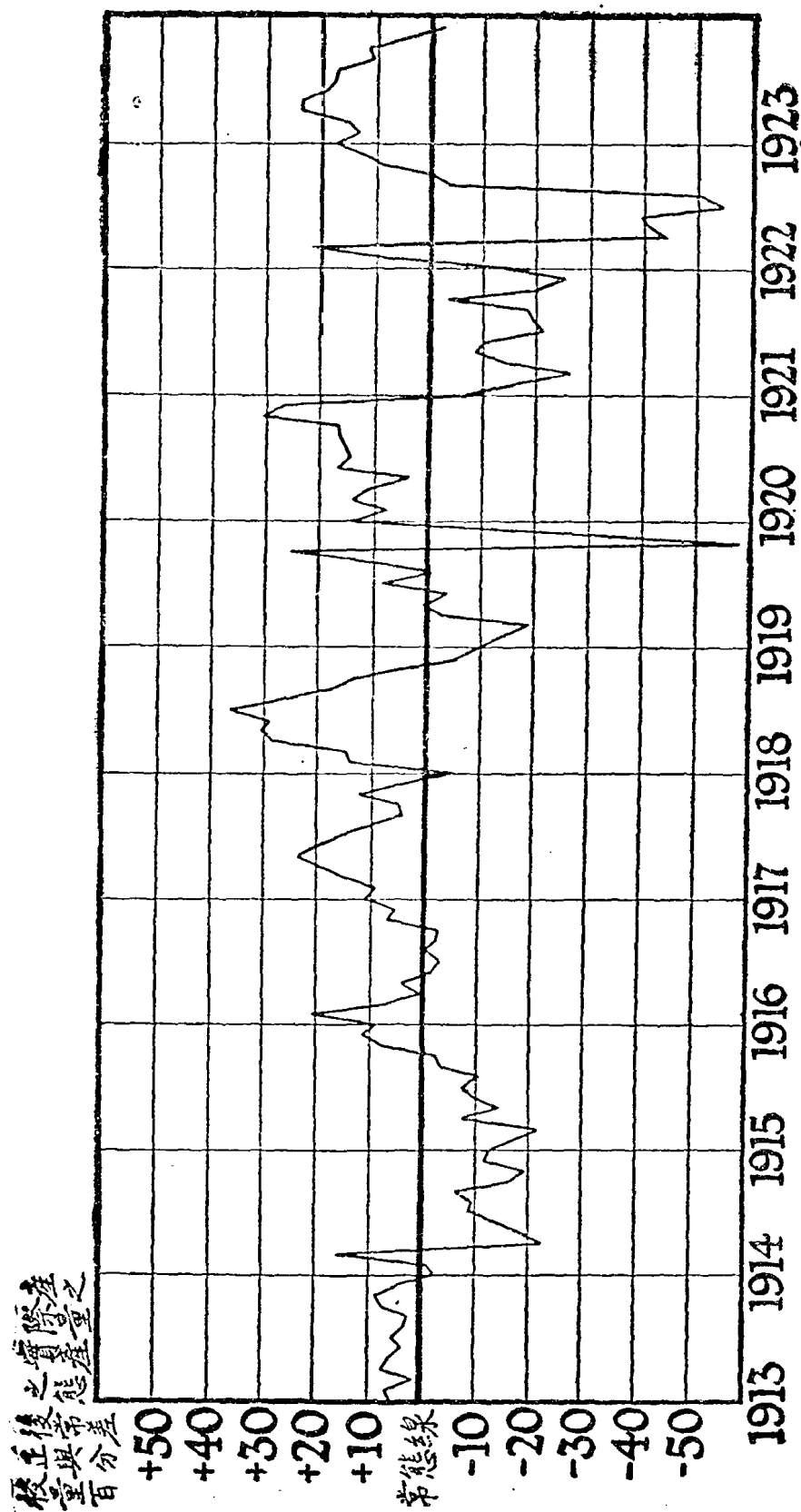
圖六十四係按照每月實際產量繪製者，圖六十五係按照循環指數(校正季節變動後實際產量與常態產量之百分差)繪製者。此處所得循環變動之指數所反映一般商情之循環變動，大致尚屬準確，惟烟煤產量所受意外及不規則變動之影響常較其他各種經濟數列所受者為大，此種影響在計算指數時尚未能剔除。如一九一九年十一月及一九二二年四月至八月因受礦工罷工之影響，故產量特低。吾人研究經濟數列(economic series)時，對於此種意外變動不可不加以注意焉。

又在比較兩種數列之循環變動時，首須將兩種數字化為可以比較之形式。數列之對於商情盛衰感應敏銳者，其循環變動必較劇烈，倘無一公分母以化為相對之數字，則比較時必感困難。此公分母可以標準差充之。在比較之前應將兩數列之每年或每月實際數值與常態數值之百分差各以其標準差除之，使化為標準差單位，則易於比較矣。

分析之手續至此乃告完成。吾人最初係將所欲分析之資料配合一長期趨勢之方程式，更由此式計算常態數值；次計算季節指數以測度季節變動；常態數值及季節指數分別求得後，即用以校正長期趨勢及季節變動，使循環變動與此兩種因素相隔離，以便單獨研究此種現象。至



圖六十四. 1913—1923年各月美國烟煤生產量



圖六十五. 1913—1923年美國煤產量數字內所含之循環變動及意外變動

於循環變動之指數內所含意外與不規則之變動雖可用數學方法剔除，使其表現粹純之循環現象，惟精密之數學方法應用於經濟資料時，宜極端審慎耳。此種方法因不在本書討論範圍之內，故不備述。

分析時間數列之要點

吾人分析時間數列時須注意下列各點。其中數點前已論及，於此更作概括之論述。

1. 在全時期內資料性質必須一致 (homogeneous)，是為基本條件之一。此不僅對於資料之外表、來源及其計價方法等應加注意，且對於其社會及經濟情況之背景亦當顧及。前述兩種因素之分析，對此條件皆有直接關係。
 - a. 求長期趨勢綫時，全時期內所有足以影響長期趨勢之各種情況必須固定不變，方可以一綫代表其長期趨勢。
 - b. 求季節指數時，在全時期內所表現之季節變動必須較有規律而後可。受意外變動影響之時期，不宜包含在內；如遇情況變更之時，應另編一新指數。故於編製季節指數之先，應將資料及有關之要素加以縝密之研究，以確定全時期內足以影響季節變動之情況曾否變更。
2. 求長期趨勢之方程式時，全時期應包含若干整個之循環週期。全時期之兩端應為變動平穩，近乎常態之年份。
3. 求長期趨勢綫及計算季節指數時，資料所包含之時期愈長愈佳。十年為最短時期，較此時期愈長愈為適宜。
4. 時間數列之以貨幣單位表示，而其意義實代表數量者，應先銷除貨幣購買力變動之因素，而後分析。如銀行票據交換額 (bank clearings)、出口貨值、建築執照所開列之房屋造價等，皆其著例也。校正購買力時應採用一種適宜之物價指數。

參 考 書

- CHADDOCK, R. E. *Principles and Methods of Statistics* (Chap. XIII).
- CRUM, W. L. *The Use of the Median in Determining Seasonal Variation*.
Journal of the American Statistical Association, March, 1923.
- FALKNER, HELEN D. *The Measurement of Seasonal Variation*. Journal of the
American Statistical Association, June, 1924.
- HALL, LINCOLN W. *Seasonal Variation as a Relative of Secular Trend*. Jour-
nal of the American Statistical Association, June, 1924.
- HART, W. L. *The Method of Monthly Means for Determination of a Seasonal
Variation*. Journal of the American Statistical Association, Sept., 1922.
- KING, W. I. *An Improved Method for Measuring the Seasonal Factor*. Journal
of the American Statistical Association, Sept., 1924. (In this article Dr.
King proposes a method by which year-to-year changes in the seasonal
factor may be measured.)
- MOORE, H. L. *Economic Cycles: Their Law and Cause*.
- MOORE, H. L. *Generating Economic Cycles*.
- PERSONS, W. M. *Correlation of Time Series*. (In *Handbook of Mathematical
Statistics*, edited by H. L. Rietz, 151—158.)
- PERSONS, W. M. *Indices of Business Conditions*. Review of Economic Sta-
tistics. Prei. Vol. I, 1919.

第九章 物量指數

現代經濟學者以及具有遠識之企業家對於物量統計之缺乏，常引以為憾，蓋吾人除若干主要原料品有較準確之生產數量可考外，而於一般製造品幾全無可靠之資料足供參考。近今政府及私人機關有見及此，正在多方蒐集與編制之中，故不久必將有準確完備之資料以供社會之需要也。此項物量統計資料不僅為研究物價及商情循環者所不可少，且為研究商業制度之經濟學家及企業家所必需。

當吾人研究物品產量之範圍，漸由一二個別物品推及於一般物品時，即覺有彙合各種數字編製指數之必要。研究一般物品產量者，其所應付之問題正與研究一般物價變動者相同。研究之目的，如在於斷定產量之長期趨勢，或其循環及季節變動，則非彙集個別物品之產量合編意義賅要之一種指數不可。一九一九年美國戰時工業局（War Industries Board）物價股，在密哲爾氏（W. C. Mitchell）指導之下，嘗編製此種指數。該指數包含原料品九十種，起訖時期為自一九一三年至一九一八年止。近年來編製此種指數者，尤不在少數，且其範圍較廣，而所包含之時期亦較長。茲擇其重要者分別說明其編製方法於後。

戰時工業局之物量指數

戰時工業局物價股編製物量指數之動機，當時係為研究物價時參考之用。其編製方法與該局所編物價指數相似。該項物價指數係以每月或每年各品之價格，與其一九一七年之生產及進口之綜合數量相乘，逐品相加，求其總值而得之指數。在該指數中，各時期之物量不變，而物價則時有變更，蓋為測度物價之變動者也。編製物量指數時，情形適與此相反，而以每年各品之生產及進口之綜合數量，與其一九一七年之價格

相乘，而後逐品相加，以求每年之總值，化爲百分比，即成物量指數(註)。在該指數中，各時期物價不變，故指數所表示者，純爲物品生產及進口數量之變動。

該局嘗根據原料品九十種，分別計算三種綜合價值：其一係以每年之物量與一九一七年之價格相乘而得；其二係以每年之物價與一九一七年之物量相乘而得；其三係以每年之物量與每年之物價相乘而得。由第一種綜合價值即可計算物量指數，第二種計算物價指數，第三種計算價值指數(value index)。價值指數係以表示物量與物價兩種變動之淨結果者也。茲將此項綜合價值及指數分列於下：

表 八 十 八

原料品九十種之物量、物價及價值指數

1913—1918年*

年 份	綜合價值(美金百萬元)			指 數		
	每年物量× 1917年物價	每年物價× 1917年物量	每年物價× 每年物量	物 量	物 價	價 值
1913	\$30,375	\$19,973	\$17,390	100	100	100
1914	30,207	19,224	16,694	99	96	96
1915	32,482	19,699	18,455	107	99	106
1916	33,700	23,363	22,785	111	117	131
1917	34,748	34,748	34,748	114	174	200
1918	35,169	38,251	39,153	116	192	225

* 錄自戰時工業局出版物價叢書第一種第45頁所載“History of Prices During the War”。原表列有原料品分類指數，即植物產品、畜產品、林產品、礦產品及水產品，共五類。

由上表第一種指數所示，在一九一四年至一九一六年之間，物量增加甚鉅，一九一七年及一九一八年則所增無幾。物價指數所表示價格增漲之情形與他種物價指數相似。在第三種價值指數中，物量與物價俱爲

(註)各種物品除以物價爲權數外，又加用一種權數。該項權數係根據各該種原料於製成物品時所需經過製造步驟之繁簡而定，故礦產原料之權數大於農產原料，因後者大抵無須經過製造手續也。至於各品所用權數，該局並未發表。

變量，故其指數之漲落與數量及價值俱有關係，然常見誤用類似之數字，以證明戰時物量之增加者，實則戰時物品價值之增加，由於價格變動之原因為多，而所受數量增加之影響猶較小也。

前項指數所用計算方法，可以下式表明之。以一九一三年為基期，而求一九一四年物量指數之公式為

$$\frac{\sum q_{1914} \cdot p_{1917}}{\sum q_{1913} \cdot p_{1917}}$$

此式為加權綜合法之一種。費暄氏(Fisher)研究其“理想”指數時，曾將此法加以試驗，惟式中 p 及 q 之位置互易耳。理想指數之公式及通常所用之此種簡單的加權綜合公式，皆為適用於測度物量或物價之變動者。理想公式用於物量指數時，其形式如下：

$$\sqrt{\frac{\sum(q_1 p_0)}{\sum(q_0 p_0)} \times \frac{(\sum q_1 p_1)}{(\sum q_0 p_1)}}$$

上式中 q_0 及 p_0 係分別代表各項物品之基年數量及價格， q_1 及 p_1 代表各品計算時期之數量及價格者。此項指數之計算手續與計算理想物價指數相同，惟物價與物量須互易位置耳。

台氏之物量指數

台氏(Edmund E. Day)所編之物量指數共有兩種，均按期發表，物量指數中最合實用之指數也。此兩種指數所根據之原理不同，故其用途亦異。茲將其中一種所謂美國基本原料產量之“未修整指數”(Un-adjusted index)先加論述。

此項指數起編於一八九九年，以迄於今。其間指數所根據之資料嘗數度變更，茲所討論者，為自一九〇九年至一九二二年間之指數。

編製此項指數所用物品之產量資料，分為農產品、畜產品、林產品及礦產品四類，除總指數外，各編分類指數。此四類物品之產量所受外

來影響不盡相同，因之變動亦有差異，故有此分類也。

編製“未修整”物量指數之方法，與以前所論編製物價指數所用者幾全相同。先將計算時期之各品產量以一九一九年為基期化為百分比（農產品係以一九一九年及一九二〇年之平均產量為基數），然後分別加權，以求四類之加權幾何平均指數（惟林產品指數係根據一單獨數列，即全國每年斫伐木材之數量編製者）。所用權數係依各個物品之重要程度而定，即係按照各品在基年所得之總值以定權數之大小者也。

四類指數所包含物品之項數不同，其自一九〇九年至一九二二年間之各類項數如下：

農產品	16
畜產品	8
林產品	1
礦產品	13

（上列農產品類所包含之物品16項，為該時期內後數年之品數。）

上項分類指數求得後，復將四類物品合編一基本原料產量之總指數。總指數之編製，係由分類指數所求得之加權幾何平均數，各類指數加權時所用權數則係根據各該類物品基年之總價值。

此外台氏又嘗採用類似之方法，另編一九一四年至一九二二年製造品產量（volume of manufacture）之指數。惟編製此項指數，問題更較複雜，而於權數為尤甚。一九一四年至一九一八年間之各年指數為兩種指數之幾何平均數，該兩種指數係分別以一九一四年及一九一九年價值作為權數所求得者。其餘各年之指數，則均以一九一九年之價值作為權數而得。此項指數係根據製造品或製造業所用之原料品共三十一種編製而成。

茲將未修整之原料品各類產量指數及其總指數以及製造品產量指數列表於下：

表 八 十 九

1909—1922年美國物品產量之未修整指數(註)

(1919年=100)

年 份	原 料 品					製 造 品
	農產品	林產品	畜產品	礦產品	總指數	
1909	86.0	128.8	82.8	74.3	83.4
1910	88.5	115.8	75.0	81.7	82.6
1911	83.8	107.1	87.0	81.0	84.8
1912	99.6	113.3	84.3	86.7	91.7
1913	88.4	111.1	83.9	93.0	87.8
1914	99.5	108.1	80.3	87.6	90.1	75
1915	104.9	90.4	86.1	93.3	95.3	86
1916	91.6	100.7	93.6	104.8	94.6	102
1917	97.5	96.0	93.2	112.2	98.1	105
1918	98.1	84.9	103.2	112.7	102.0	102
1919	100.0	100.0	100.0	100.0	100.0	100
1920	109.2	86.4	92.5	114.2	103.0	104
1921	90.5	78.1	94.9	91.6	92.0	80
1922	100.7	100.7	102.1	97.5	100.7	109

(註)表內數字根據1922年七月及1923年七月出版之“Review of Economic Statistics”。

上列各種指數之變動加以比較，頗堪玩味。各種指數之變動趨勢莫不有其顯著之特性，而物品產量之逐年變動亦不相同。大抵農產品之產量所受氣候之影響最大，其他產品則與物價及商情之關係較為密切，在前此討論物價指數時亦已有所提示矣。

修整之物量指數

循環變動之研究，常為分析時間數列之主要工作，而尤為研究商情者所注視；此在研究物品產量時亦然，因物品之生產及貿易數量俱為表現商業循環之基本要素也。測度時間數列循環變動之方法，曾在前章內有所論述。吾人如欲研究一般商情之變遷，則應將有關各種數列之循環變動合編一指數，此種表示商情循環變動之過程中物品產量增減之指數，其功用之大，自不待言也。

編製循環變動之指數時，如採用年數字，則計算手續較為簡單，蓋

年數字內季節變動並不存在，僅須將各數列內所受長期趨勢之影響予以銷除可矣。此項指數編製方法有二，台氏在修整物量指數時，曾將兩法分別測驗之。

其第一法係先就個別物品產量之時間數列各配一適當之長期趨勢綫；次將各品每年之實際產量化為同年常態產量之百分比；最後將各品同一年份之百分比用加權算術平均數法求其平均百分比，即為是年之修整指數。計算平均百分比時，各品所用權數與計算未修整指數時所用者相同。此項指數成百分比之形式，惟此百分比係由常態產量求得，非由定基而得，此即物品產量循環變動之指數也。

第二法較第一法更為簡單。此法係根據前項未修整指數加以改正而得。未修整指數中各項物品產量所受長期趨勢之影響依然存在，故未修整指數所表現之長期趨勢，實乃彙合各項物品之長期趨勢而成。修整之法，即係根據未修整指數求長期趨勢綫，以銷除其長期趨勢之影響，如此則可免去在每個數列上各配一長期趨勢綫，而一一銷除其長期趨勢之繁瑣工作。此即由未修整指數直接求修整指數之方法也。

第二法之適當與否，可由兩法所得結果之比較而定。台氏曾採用兩法，而比較其結果，發現兩種指數幾完全相同，故台氏乃決用計算手續較為簡單之第二法以求修整指數。

表九十所列係根據表八十九之未修整指數，用第二法所求得農產品及礦產品之修整指數。

根據月數字編製之物量指數

上述各例中所用資料均為年數字，若用各品生產及貿易之每月數量以求指數，雖仍可採用前述各法，但所用方法必較繁複。現時此項按月編製之指數中最為詳備者，當推聯邦準備局 (Federal Reserve Board) 所編之指數。

表 九 十

1909—1922年美國物品產量之修整指數(註)

年份	農產品	礦產品
1909	98	97
1910	99	102
1911	93	98
1912	109	101
1913	95	104
1914	106	95
1915	110	98
1916	95	107
1917	100	111
1918	99	108
1919	99	93
1920	107	104
1921	88	81
1922	96	84

(註)此項指數根據1922年及七月及1923年七月出版之“Review of Economic Statistics.”

聯邦準備局之基本工業生產指數

聯邦準備局之指數起編於一九一三年一月，以迄於今，每月指數刊載於該局出版之聯邦準備月報(Federal Reserve Bulletin)。此項指數係根據統計數列二十二種編製而成，其每月資料亦逐期發表。茲將其中二十種數列錄入表九十一內，以供參考。求修正指數時，各數列均分別加權，權數係根據一九一九年工業調查所得之結果，並參酌各種原料在加工後所增價值之多寡及所僱工人數而定。

該局所編之指數，並未將長期趨勢予以消除。此項指數係以百分比表示，而以一九一九年之每月平均產量作為基數者，惟此項基數與由長期趨勢綫求得之常態數值不同，故其意義亦不相同。

計算平均數之前，係將各數列以一九一九年之每月平均產量為基

數分別化為百分比；但在化百分比時曾將一九一九年之每月平均產量用季節指數校正之。此項每月平均產量經此校正後，用作基數，則由此求得各數列之百分比即可不含季節變動之影響。此種校正方法頗值注意(註)。

該局所用之季節指數，係由每月實際數量對於十二個月之移動平均數之差所求得之中位數。此法已在研究季節變動時討論及之。各數列之季節變動因其對於該項產量指數頗關緊要，故將其季節指數列表於下，以供參考。

表 九 十 一
二十種統計數列之季節變動指數(註)

數列種類	一月	二月	三月	四月	五月	六月	七月	八月	九月	十月	十一月	十二月
生鐵產量.....	99	88	103	100	102	98	100	104	101	105	102	98
鋼條產量.....	99	89	102	100	102	99	100	104	101	107	102	95
棉花消費量.....	104	97	107	100	105	102	99	100	95	99	95	97
麵粉產量.....	112	89	89	82	84	79	77	106	113	124	124	121
糖消費量.....	69	93	128	125	121	128	127	126	99	71	60	53
宰豬數量.....	142	119	99	86	99	106	84	65	62	83	109	146
宰牛數量.....	102	81	88	84	86	92	96	99	111	128	122	111
宰犢數量.....	85	73	102	121	128	114	109	96	99	100	92	81
宰羊數量.....	104	89	91	80	83	90	100	110	120	117	110	106
木材研伐量.....	80	84	97	101	119	113	107	113	110	109	92	75
烟煤產量.....	107	92	102	83	92	95	97	106	106	113	104	103
白煤產量.....	97	84	100	96	105	105	100	102	99	110	102	100
紫銅產量.....	100	93	105	103	104	102	97	100	97	103	97	99
鞋底皮產量.....	102	90	100	103	108	109	98	101	97	105	92	97
印報紙產量.....	106	92	101	101	101	102	98	101	95	103	98	102
水泥產量.....	55	64	83	104	120	117	109	119	117	121	105	86
石油產量.....	98	90	102	100	102	101	105	103	99	103	98	99
雪茄烟產量.....	89	88	100	96	100	105	103	103	103	113	106	94
紙烟產量.....	95	89	99	91	99	106	114	108	104	110	99	86
製成菸草產量.....	94	94	107	100	102	102	100	107	106	108	95	85

(註)表內數字根據1922年十二月聯邦準備月報第1416頁。

表內各指數係根據一九一三年至一九一九年間之資料編製。此項指數頗可供研究各種工業分類之季節變動者之助。

季節變動之校正方法，係先將基年各月之數量用季節指數校正後，

(註)該局採用此法以校正季節變動，係 National Bureau of Economic Research 麥考萊氏(Frederick R. Macaulay)所設計者。

作為基數，而後計算各月之百分比。吾人可就聯邦準備月報所載棉花消費數量以說明此項校正方法。下表所列為一九一九年各月棉花之實際消費量及其季節變動指數，以及校正季節變動後一九一九年各月之平均消費量。

表 九 十 二
季節變動之校正法
(棉花消費量)

(1) 月 份	(2) 1919年棉花消費量 (單位:百包)	(3) 季節指數	(4) 校正後1919年各 月之平均數量
一月.....	5,669	104	5,130
二月.....	4,333	97	4,785
三月.....	4,335	107	5,278
四月.....	4,759	100	4,933
五月.....	4,879	105	5,179
六月.....	4,743	102	5,032
七月.....	5,103	99	4,884
八月.....	4,973	100	4,933
九月.....	4,911	95	4,686
十月.....	5,560	99	4,884
十一月.....	4,913	95	4,686
十二月.....	5,117	97	4,785
	<u>59,195</u>		<u>59,195</u>
平均.....	4,933		

一九一九年棉花之每月平均消費量為4,933(百包)。倘無季節變動之存在，吾人頗可假定該項消費量除所受長期趨勢及意外變動之影響皆不計及外即可認為該年每月之實際生產量，但由季節指數之所示，則一月份之消費量較全年平均消費量高百分之四，二月份之消費量較全年平均消費量低百分之三，其餘各月份亦高低不一。若棉花消費量在全年中僅受季節變動之影響，則以各月季節指數乘全年平均消費量4,933所得之積，即為校正季節變動後之各月平均數量，亦可認為一九一九年各月之實際生產量，故此項已經校正之各月平均數量，即可用作基數，而將其他各年每月之數量化為百分比之形式。

例如一九二一年一月之棉花實際消費量為3,665(百包)，欲化為一九一九年基期之百分比，可以校正後之一九一九年一月平均數量5,130

除之，再乘以 100，得 71.4，此即校正季節變動後一九二一年一月之百分比也。欲求二月之百分比，可用同法以校正後之一九一九年二月平均數量除之。此法之最大優點，即在基期數字經過一次校正後，其他各年之數字即無須加以校正。

求每月生產指數時，即可由上法所求得各數列每月之百分比彙合計算而得。其法係將每個百分比各以其權數乘之，而後一一相加，再以權數之和除之，即得所求之指數。該局在月報內發表物品之生產總指數時，且將校正後之百分比一併刊載，對於該項指數之功用，尤為增色不少。

除上項基本工業生產之普通指數外，聯邦準備局研究部另編各種企業活重之指數，包括農產變動之指數、礦物產量之指數、及工業生產之指數等。惟編製各項指數時，未將季節變動及長期趨勢加以校正耳。自一九一九年一月起，各月指數均有發表。

美國電話電報公司之一般商情指數

美國電話電報公司統計部所編一般商情之組合指數，其編製方法又與前者稍異。此項指數按月編製，其時期追溯至一八七七年，惟自編製迄今，指數所採用之資料業經變更多次。

就一九二三年年末之指數而言，係根據與一般商情最有關聯之數列十一種編製而成。每種數列均曾採用前兩章所述方法加以分析，即先配合一長期趨勢綫，而以實際數值化為常態數值之百分比，再將此項百分比內所含之季節變動校正之。惟在彙合十一個數列編為指數之前，尚須經過一次校正手續，此因各數列中循環變動之程度參差不齊，有變動極大者，故須在平均以前，將各數列之循環變動化為可以比較之單位，否則循環變動較大之數列所影響於平均數者必嫌過大。校正之法係將每一數列內各項對於常態數值之差以其標準差除之，使化為標準差之

表 九 十 三

1877—1923年美國電話電報公司所編一般商情之組合指數(註)

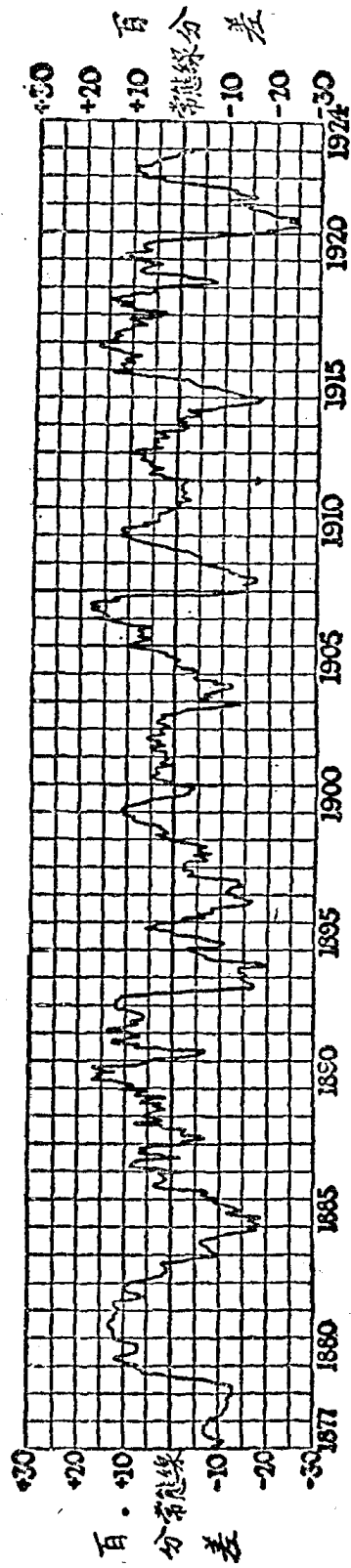
(表內數字爲對於常態值之百分差)

	1877	1878	1879	1880	1881	1882	1883	1884	1885	1886	1887	1888
一月	-9	-8	-13	+7	+12	+11	+9	-10	-16*	-10	-2	-2
二月	-9	-9	-13	+9	+12	+11	+7	-9	-19	-8	+3	-2
三月	-10	-10	-13	+11	+12	+12	+4	-8	-17	-6	+9	-7
四月	-11	-11	-13	+12	+12	+12	+2	-8	-16	-8	+8	-6
五月	-10	-11	-11	+10	+12	+10	+2	-7	-19	-7	+8	-2
六月	-8	-12	-9	+8	+13	+8	+1	-6	-14	+3	+6	-2
七月	-7	-12	-7	+7	+13	+5	+1	-6	-11	+4	-2	-3
八月	-7	-12	-3	+7	+13	+7	+1	-8	-15	+2	+2	+2
九月	-7	-12	+2	+7	+12	+9	+2	-10	-14	+2	+7	+2
十月	-7	-12	+6	+7	+12	+10	+2	-12	-12	+1	+5	+8
十一月	-7	-12	+6	+9	+12	+10	-2	-15	-11	+2	+7	+2
十二月	-8	-13	+7	+11	+11	+9	-6	-18	-9	+3	+1	+3
	1889	1890	1891	1892	1893	1894	1895	1896	1897	1898	1899	1900
一月	+6	+11	+5	+9*	+12	-14	-3	-1	-13	-3	+1	+10
二月	+5	+8	+1	+14	+12	-16	-9	+2	-11	-2	+1	+11
三月	+3	+10	-5	+11	+12	-14	-10	-7	-12	-3	+4	+10
四月	+1	+12	-6	+9	+11	-14	-11	-5	-12	-7	+1	+8
五月	+4	+17	-7	+7	+11	-14	-8	-8	-15	-7	+2	+6
六月	+2	+15	0	+9	+5	-20	-6	-8	-13	-4	+3	+6
七月	+7	+16	+8	+6	-2	-18	-1	-8	-14	-8	+3	+3
八月	+6	+12	+7	+6	-13	-10	0	-14	-10	-5	+6	0
九月	+1	+15	+13	+7	-17	-7	+1	-16	-3	-5	+8	-3
十月	+8	+17	+12	+8	-17	-7	+6	-17	-4	-7	+9	-3
十一月	+7	+8	+7	+10	-16	-6	+4	-16	-4	-4	+8	-4
十二月	+5	+5	+7	+12	-15	-4	+3	-13	-2	+2	+11	-4
	1901	1902	1903	1904	1905	1906	1907	1908	1909	1910	1911	1912
一月	-1	+2	+3*	-10	-5	+10	+17	-14	-5*	+12	0	-2
二月	0	+2	+1	-5	-5	+9	+14	-15	-5	+10	-1	+3
三月	+2	+2	+2	-8	-2	+7	+15	-15	-5	+12	0	+3
四月	+5	+5	+4	-6	-1	+5	+16	-14	-5	+11	-3	+6
五月	+4	+5	+3	-8	0	+8	+18	-17	-3	+6	-2	+5
六月	+3	+2	+2	-10	0	+7	+16	-17	-2	+7	-1	+3
七月	+4	+4	+3	-12	-1	+7	+18	-16	+1	+4	-2	+6
八月	+4	+3	-1	-12	0	+8	+16	-13	+2	+4	-1	+6
九月	0	+6	-2	-8	+2	+5	+12	-11	+5	+4	-1	+5
十月	+3	+5	-6	-9	+2	+10	+13	-11	+8	+4	-1	+8
十一月	+2	+2	-11	-4	+4	+11	-4	-9	+9	+2	-1	+7
十二月	+1	+3	-14	-5	+7	+13	-12	-8	+10	+1	-2	+7
	1913	1914	1915	1916	1917	1918	1919	1920	1921	1922	1923	
一月	+10*	0	-16	+11	+17	+4	+4*	+13	-19*	-16	+8	
二月	+8	-2	-13	+12	+13	+3	-3	+10	-18	-11	+8	
三月	+3	0	-12	+13	+14	+10	-8	+12	-22	-10	+10	
四月	+5	-1	-9	+9	+12	+10	-6	+9	-25	-16*	+9	
五月	+5	-5	-8	+12	+14	+13	-2	+6	-25	-14	+10	
六月	+4	-4	-4	+11	+12	+10	0	+8	-23	-11	+7	
七月	+5	-2	-3	+8	+10	+14	+7	+8	-25	-12	+6	
八月	+2	-8	-2	+12	+10	+15	+9	+7	-21	-10	+5	
九月	+3	-10	+2	+13	+7	+11	+8	+5	-19	-6	+1	
十月	+5	-13	+5	+15	+9	+10	+7	-1	-18	-2	+1*	
十一月	-1	-17	+9	+17	+11	+5	+4	-6	-15	+4	0	
十二月	-2	-18	+14	+16	+6	+6	+8	-11	-14	+4	-2	

*表示指數所包含之數列有變更之月份。

(註)此項指數刊載於此,係經美國電話電報公司之同意。

一般商情之組合指數



圖六十六. (美國電話電報公司統計部所編之一般商情指數)

單位即可矣。

茲將一九二三年十二月指數編製時所用之十一種數列及其權數列表於下：

數 列	權 數
生鐵產量	20
運輸之淨“噸哩”數	10
運貨車之運輸數量	10
鋼條產量	10
毛織機之活動量	10
棉花消費量	10
紙產量	10
木料產量	5
皮革產量	5
烟煤產量	5
電力產量	5
總 計	100

自一八七七年至一九二三年之指數見表九十三及圖六十六。

貿易數量指數

美國現時所編之各種物量指數中，包含數列最多者，當推紐約聯邦準備銀行情報部所編之貿易數量指數(註)。此項指數之性質及其範圍，可由下表內該指數所包含統計數列之類別見之。計算指數時，各數列均一一加權，其權數亦併列表內，以供參考。

(註)關於此項指數之編製內容，可參看一九二三年十二月出版“Journal of the American Statistical Association”第49—63頁 Carl Snyder 氏所著 A New Index of the Volume of Trade.

表 九 十 四

紐約聯邦準備銀行貿易數量指數所包含之統計數列

屬於生產方面者：	權數	
1. 消費品.....	8%	
2. 製造品.....	9	
3. 工廠僱用工人數.....	6	
4. 汽車及運貨車.....	2	
5. 建築.....	4	
	——	29%
屬於初步分配方面者：		
6. 商品運輸量.....	5	
7. 其他運輸量.....	2	
8. 批發交易量.....	8	
9. 出口貨.....	3	
10. 進口貨.....	2	
11. 通過巴拿馬運河之水運噸位.....	1	
12. 穀類出口.....	1	
	——	22%
屬於消費者之分配者：		
13. 百貨商店銷貨量.....	8	
14. 聯繫商店 (chain store) 銷貨量.....	3	
15. 聯繫雜貨店銷貨量.....	6	
16. 郵寄銷貨量 (mail order sales).....	3	
17. 新人壽保險.....	2	
18. 娛樂場收入.....	2	
19. 廣告費用.....	2	
	——	26%
屬於一般商業者：		
20. 紐約市郊各區之貨項.....	8	
21. 紐約市內之貨項.....	5	
22. 郵政收入.....	1	
23. 交通.....	1	
24. 電力產量.....	2	
	——	17%

表 九 十 四 (續)

屬於金融方面者：

25. 紐約證券交易所股票成交數.....	2
26. 新金融合作事業.....	2
27. 芝加哥穀類期貨交易.....	1
28. 紐約市及紐奧利安市 (New Orleans) 棉花期貨 交易.....	1

6%
100%

表內各數列中，尚有由數種數列組合而成者。如屬於生產方面消費品產量之數列，係由十五種數列組合而成，製造品產量之數列亦係根據十五種數列編成者。

在確定各數列之權數時，每遇種種特殊問題而須分別解決者，惟就一般而論，各數列之權數大抵根據原料經過加工後所增之價值、產品之價值或物品交易之價值而定。

由各數列合併編製單獨指數之方法，與前此所述者相同，即每一數列須各配一適當之長期趨勢綫以確定其常態數值。如數列中季節變動甚為顯著時，又須計算其季節指數以作校正季節變動之用。此處又涉及平減價值之問題，蓋所用各數列中，屬於物價範圍以內者不在少數也。在平減價值時，某數種數列雖有相當之平減指數可用，但其餘多種數列常須另行搜集數種數列合編一平減指數，方可適用。例如“紐約市郊各區之貸項”之數列，其所用平減指數係由以下各指數組合而成。

數 列	權數
批發物價指數	2
生活費指數	3
工人每週平均工資指數	3
股票價格指數	1

各數列經此種種校正手續後，乃可合編貿易指數。茲將一九一九年至一九二三年之貿易數量指數列表於下：

表 九 十 五

1919—1923年之貿易數量指數

(紐約聯邦準備銀行編)

	1919	1920	1921	1922	1923
一 月	97	109	90	93	109
二 月	97	104	92	96	110
三 月	95	108	91	102	113
四 月	101	105	92	99	109
五 月	106	104	91	103	110
六 月	108	102	92	103	107
七 月	110	101	91	97	100
八 月	107	99	95	99	102
九 月	107	96	94	103	100
十 月	109	95	92	103	106
十一月	105	94	92	104	105
十二月	105	93	93	105	100

以上所舉，自未能將現時所編各種物品生產及貿易數量之指數包括無遺(註)，蓋吾人之目的僅欲說明各種編製方法也。編製方法雖無一定標準，但就以上所論，可見物量指數之編製無需十分繁複之手續。各種困難問題，如平減價值、確定常態數值及測度季節變動等等，均可隨時發生，必須多加試驗始可完全解決。但現有之編製方法已足使吾人獲得測度企業活動之各種指數，此不僅可增進吾人對於經濟制度之智識，並可使吾人對於此項制度作更有效之管理焉。

(註)參看 1923 年四月出版“Review of Economic Statistics”內載 Warren M. Persons氏所編 the Index of Trade for the United States. 氏嘗將 1903 年以來美國現時所編各種貿易數量指數之每月數字彙合刊載。又參看同書所載工業及礦業之每月指數及“Federal Reserve Bulletin”按月發表之零售貿易指數。

參 考 書

- DAY, EDMUND E. *An Index of the Physical Volume of Production*. Review of Economic Statistics, Sept., 1920, Jan., 1921.
- DAY, EDMUND E. *The Volume of Production of Basic Materials in the United States*. Review of Economic Statistics, July, 1922.
- PERSONS, W. M. *An Index of Trade for the United States*. Review of Economic Statistics, April, 1923.
- SNYDER, CARL. *A New Index of the Volume of Trade*. Journal of the American Statistical Association, Dec., 1923.
- SNYDER, CARL. *A New Clearings Index of Business for Fifty Years*. Journal of the American Statistical Association, Sept., 1923.

第十章 關係之測度：直綫相關

在討論平均數、離散度及偏斜度時，吾人之目的在於研究一單獨次數分配之表述方法。此種次數分配僅含一變量，故可用前述各種量數（measure）以定其中心值，及表述中心值兩邊之分佈情形。在分析時間數列時，吾人所欲研究之問題則與此稍異。此時之變量因時間之變遷而發生變動，吾人之問題在於斷定此種變動所受各種勢力——長期趨勢、循環變動、季節變動及意外變動——之影響程度。前數章即在討論此種種勢力（意外變動除外）之測度方法。

分析時間數列之各種方法中，亦有可用以測度相關者。在研究時間數列時，吾人嘗用數學方程式以表示時間與變量之關係。經濟數列中之長期趨勢常為時間之函數（function），故如時間與變量間確有關係存在，則此種關係無論其完全與否，倘用一數學方程式以表述之，實有顯著之便利。測度相關時，所遇之問題類似於此，特範圍較廣耳。在時間數列中，既可用一數學方程式以表示時間與常態數值之關係，則他種變量間之關係亦可用一方程式以表示之否？如棉花產量與價格之關係，稻麥收穫與雨量之關係，工人工資與其工作效率之關係等亦可用數學方法以測度之否？倘此而可能，則經濟學家可獲得一有力之工具，使其方法達到一種相當精確之程度，如物理學家之工作然。

繳納個人所得稅人數與汽車登記數之關係

吾人試舉美國各州繳納個人所得稅之人數與汽車登記數為例，以研究兩種數列之關係。此項數字見表九十六第(2)(3)兩欄。

圖六十七係根據此項數字繪製者(註)。圖中每點之橫坐標表示納稅

(註)紐約、賓夕法尼亞、紐澤西、伊利那、馬薩諸塞五州皆未包括在內。因在商業繁盛區域，

其納稅人數與汽車登記數之關係每與普通區域不同，未可相提並論也。

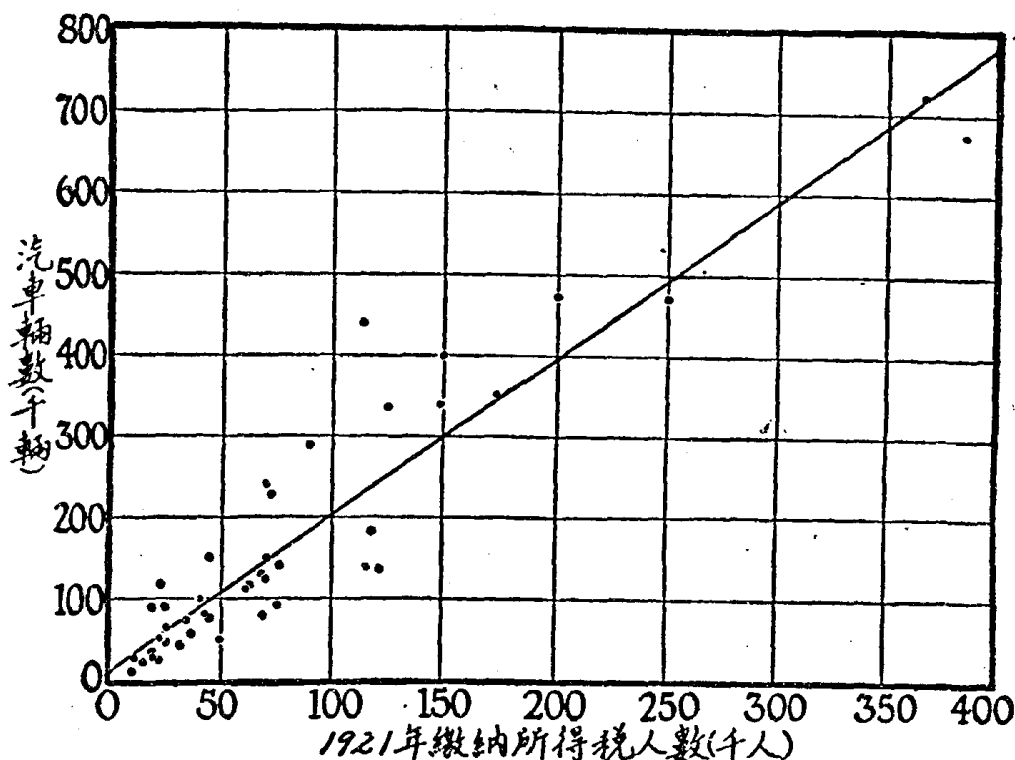
人數，縱坐標表示汽車登記數。此圖在統計學中稱為散佈圖 (scatter diagram)。由圖觀之，此兩種變量間確有關係存在。大抵各州中繳納個

表九十六

1921年美國四十三州繳納個人所得稅人數及汽車登記數

(1) 州 別	(2) 繳納所得稅 人 數 (單位:千) X	(3) 汽車登記數 (單位:千輛) Y	(4) XY	(5) X^2	(6) Y^2
阿拉巴瑪...	43	82	3,526	1,849	6,724
亞利桑那...	18	35	630	324	1,225
阿康薩...	34	67	2,278	1,156	4,489
加利福尼亞...	386	673	259,778	148,996	452,929
科羅拉多...	70	145	10,150	4,900	21,025
康涅狄格...	123	137	16,851	15,129	18,769
德拉瓦...	16	21	336	256	441
佛羅里達...	42	97	4,074	1,764	9,409
佐治亞...	68	131	8,908	4,624	17,161
愛達荷...	23	51	1,173	529	2,601
印第安納...	150	400	60,000	22,500	160,000
愛俄瓦...	111	460	51,060	12,321	211,600
堪薩斯...	89	291	25,899	7,921	84,681
肯塔基...	69	126	8,694	4,761	15,876
路易斯安那...	68	80	5,440	4,624	6,400
緬因...	44	77	3,388	1,936	5,929
馬利蘭...	113	140	15,820	12,769	19,600
密雪于...	250	477	119,250	62,500	227,529
明尼蘇達...	125	328	41,000	15,625	107,584
密士失必...	26	65	1,690	676	4,225
密蘇利...	173	346	59,858	29,929	119,716
蒙塔那...	37	58	2,146	1,369	3,364
內布拉斯加...	72	238	17,136	5,184	56,644
內華達...	10	10	100	100	100
紐罕布什爾...	32	42	1,344	1,024	1,764
新墨西哥...	12	24	288	144	576
北卡羅來納...	44	148	6,512	1,936	21,904
北達科他...	18	92	1,656	324	8,464
俄亥俄...	367	720	264,240	134,689	518,400
俄克拉荷馬...	69	221	15,249	4,761	48,841
俄勒岡...	63	118	7,434	3,969	13,924
羅得島...	48	54	2,592	2,304	2,916
南卡羅來納...	25	90	2,250	625	8,100
南達利他...	22	119	2,618	484	14,161
泰內西...	61	117	7,137	3,721	13,689
得克薩斯...	200	467	93,400	40,000	218,089
猶他...	26	47	1,222	676	2,209
威爾滿...	18	36	648	324	1,296
弗基尼阿...	76	141	10,716	5,776	19,881
華盛頓...	116	185	21,460	13,456	34,225
西弗基尼阿...	75	93	6,975	5,625	8,649
威斯康辛...	148	341	50,468	21,904	116,281
歪俄明...	22	26	572	484	676
	3,602	7,616	1,215,966	603,968	2,612,066

人所得稅之人數較多者，其汽車登記數亦較多，但此關係並不完全，間有納稅人數兩州相同，而其汽車登記數多寡懸殊者。例如一九二一年肯塔基州與俄克拉荷馬州納稅人數均為69,000人，但前者之汽車登記數僅12,600輛，而後者則多至221,000輛。如其關係為完全的，則X變量為某數值時，Y變量應有一定之數值。



圖六十七. 表示1921年各州繳納個人所得稅人數與汽車登記數兩者關係之散佈圖及其平均關係綫

上例中兩變量之關係雖不完全，却有存在，吾人之問題，即在求一適當之方程式以表示之。此項關係與長期趨勢相似，且可用直綫表示，故用最小平方法配合於散佈圖中各點所得之直綫，當可表示兩者之平均關係。此綫雖亦可用觀察法 (inspection) 配合，然以用最小平方法配合者較為精確。

用最小平方法時，吾人須求解下列方程式：

$$\Sigma(Y) = na + b\Sigma(X)$$

$$\Sigma(XY) = a\Sigma(X) + b\Sigma(X^2)$$

解此兩式所需之數值可由表九十六內得之。將 X, Y 數值之總和代入方程式，得

$$7,616 = 43a + 3,602b$$

$$1,215,966 = 3,602a + 603,968b$$

由此聯立方程式求得

$$a = 16.92$$

$$b = 1.91$$

故所求之直綫方程式爲

$$Y = 16.92 + 1.91X \text{ (註)}$$

此綫如圖六十七所示。

此際吾人已將表示兩變量關係之數學方程式求出。兩變量者，即各州繳納所得稅之人數及汽車登記數也。在方程式中前者爲自變量（即 X 變量），後者爲倚變量（即 Y 變量）。此方程式構成兩變量間之函數相關（functional relationship），其所表示者爲兩者之平均關係。吾人試研討此式之意義。倘兩者之關係爲完全的，則所有各點應均在表示此種關係之直綫上，吾人且可據此方程式由一變量之數值，求另一變量之數值，而毫無錯誤。但此式之直綫係配合於分散之各點，其與各點之距離常有相距甚遠者。在此情形之下，方程式雖已將兩者之關係確切斷定，但其所表示之關係，按諸實際情形，相差尙遠，故不能據此方程式由一變量之數值以求另一變量之數值而無錯誤。此與採用平均數時所遇之問題相同。吾人必先明瞭平均數兩邊各項之集中程度，方可斷定此平均數代表性之強弱而採用之。故表示變量相關之方程式，如不能確知其與實際情形相差之程度，則此式亦必毫無價值。於此吾人必須求一方法以

（註）在討論相關各章內， W, X, Y 等係用以表示各變量之原來數值， w, x, y 等表示各變量對其算術平均數之差。

測定各點在直綫兩邊之離散度。

在敘述次數分配時，吾人嘗謂標準差為通常測量離散度最優之量數，此量數當亦可用以確定表示平均關係之方程式之可靠程度。直綫之標準差不僅可表示其方程式之意義，且可測定由此式所得估計數之準確程度。

標準誤之計算

平均關係綫之標準差，因其可以測量估計數之確度，故稱估計標準誤 (standard error of estimate)，或簡稱標準誤 (standard error)。所謂標準差 (standard deviation) 通常僅指算術平均數兩旁各項之均方根差 (root-mean-square deviation)，係以 σ 表示者，故標準誤可用 S 表示，以資區別。

計算 S 時，應先求與各 X 數值相對之 Y 常態數值。將已知 X 之數值代入下式

$$Y = 16.92 + 1.91X$$

即可求得 Y 之常態數值。次求 Y 之實際數值對其常態數值之相差數。此項差數之均方根 (root-mean-square) 即為所求之 S 數值。其算法可就下表說明之。

表九十七

(1) 州 別	(2) 1921年汽車 登記數 (單位:千輛) (Y 之實際數值)	(3) 計算所得之 Y 值 (Y 之常態數值)	(4) d (2) - (3)	(5) d^2
阿拉巴瑪.....	82	99	-17	289
亞利桑那.....	35	51	-16	256
阿康薩.....	67	82	-15	225
加利福尼亞.....	673	756	-82	6,724
科羅拉多.....	145	151	-6	36
康涅狄格.....	137	262	-115	13,225
德拉瓦.....	21	48	-27	729

表 九 十 七 (續)

(1)	(2)	(3)	(4)	(5)
州 別	1921年汽車 登記數 (單位:千輛) (Y之實際數值)	計算所得之Y值 (Y之常態數值)	d (2) - (3)	d ²
佛羅里達.....	97	97	0
佐治亞.....	131	147	-16	256
愛達荷.....	51	61	-10	100
印第安納.....	400	304	+96	9,216
愛俄瓦.....	460	229	+231	53,361
堪薩斯.....	291	187	+104	10,816
肯塔基.....	126	149	-23	529
路易斯安那.....	80	147	-67	4,489
緬因.....	77	101	-24	576
馬利蘭.....	140	233	-93	8,649
密雪干.....	477	495	-18	324
明尼蘇達.....	328	256	+72	5,184
密士失必.....	65	67	-2	4
密蘇利.....	346	348	-2	4
蒙塔那.....	58	87	-29	841
內布拉斯加.....	238	155	+83	6,889
內華達.....	10	36	-26	676
紐罕布什爾.....	42	78	-36	1,296
新墨西哥.....	24	40	-16	256
北卡羅來納.....	148	101	+47	2,209
北達科他.....	92	61	+41	1,681
俄亥俄.....	720	719	+1	1
俄克拉荷馬.....	221	149	+72	5,184
俄勒岡.....	118	137	-19	361
羅得島.....	54	109	-55	3,025
南卡羅來納.....	90	65	+25	625
南達科他.....	119	59	+60	3,600
泰內西.....	117	134	-17	289
得克薩斯.....	467	399	+68	4,624
猶他.....	47	67	-20	400
威爾滿.....	36	51	-15	225
弗基尼阿.....	141	162	-21	441
華盛頓.....	185	239	-54	2,916
西弗基尼阿.....	93	160	-67	4,489
威斯康辛.....	341	300	+41	1,681
歪俄明.....	26	59	-33	1,089
				<hr/> 157,790

$$S_y = \sqrt{\frac{157,790}{43}}$$
$$= 6.06$$

依此求得之 S_y 值為60.6，此係 Y 變量之標準誤，故以 S_y 表示之。此標準誤之解釋與標準差完全相同。散佈於相關綫(line of relationship)兩旁之各項，如其分配近乎常態，則全體項目之68%應包含於 $\pm S$ (此處為60.6)之距離內，95%包含於 $\pm 2S$ (此處為121.2)之距離內，99.7%包含於 $\pm 3S$ (此處為181.8)之距離內。由表示 X, Y 兩數值之各點所配合之相關綫，其兩旁如無散佈之點，則 S 之數值為0，而 Y 之數值亦可由 X 之數值推算，而完全無誤。各點在相關綫兩旁之離散度愈低，則 S 之數值愈小，故 S 之數值可用以表示相關綫之意義及其確度。於此所應注意者，標準誤所用之單位，應與 Y 之原來數值所用者相同。

估計方法

現時吾人可研究估計結果之是否可靠。今設未知某州汽車登記數而須估計之。估計之方法有二。第一法係單就 Y 變量之數值作為估計之根據。例如美國四十三州所有汽車之總數為7,616,000輛，此數以43除之，求得平均數為177,116輛。倘吾人未知某州之汽車登記輛數而欲加估計，則此估計數當取各州登記數之算術平均數，因此數為最可能之數值也(各個觀察數值中，最可能之數值為其算術平均數)。然則此估計數值之準確程度若何？按原來次數分配之標準差，原係測量各項目對其平均數離散程度之量數，故亦可用以測量根據此平均數估計所得數值之確度。如原來之次數分配近乎常態，則該州汽車實際輛數與此平均數之差數無過於標準差數值之機率為68%。根據表九十六各州汽車登記數計算所得之標準差為171.4。由是知算術平均數可用為估計之根據，而標準差即所以表示此估計差誤之機率者也(此處原可以居民數或其他

因素作為估計之根據，姑從略)。

第二法可就該州繳納個人所得稅之人數估計之。吾人前嘗研究各州汽車登記數與繳納個人所得稅人數之平均關係，其結果得下列方程式：

$$Y = 16.92 + 1.91X \text{ (兩變量均以千為單位)}$$

如某州繳納所得稅之人數為200,000人，則由此式估計該州汽車登記為數399,000輛。此為根據平均關係之方程式估計所得最可能之數值。此種估計方法比較第一法之以平均數作為估計數者孰優孰劣乎？吾人於確知 X 與 Y 間之平均關係後，即可由已知之 X 數值以估計 Y 之數值乎？

此問題之答案可於 Y 之標準誤、及 Y 之標準誤與 Y 之標準差兩者之關係中求之。 Y 之標準誤(即各數值對平均關係綫之標準差)為60.6， Y 之標準差為171.4。由是知根據方程式所得之估計，確較以 Y 之平均數作為估計數者為準確。前者之估計與汽車實際輛數之差數無過60.6之機率為68%，此差數若用原來單位表示之，為60,600輛。但根據平均數所得之估計，其與實際輛數相差之機率同為68%時，則其相差之數可多至171,400輛(註)。故若已知兩變量之關係，此種關係縱不完全，吾人亦可使估計之差誤大為減少也。

相 關 係 數

吾人現時已得兩種量數以表示變量間之關係：一為相關之基本方程式，此式可以表示在一變量變動時另一變量隨之變動之平均程度；一

(註)在本例中，項數既少，次數分配又迥異於常態，故吾人不能由 S_y 及 σ_y 之數值以斷定準確之機率。且各項中，有六項或八項之數值皆甚大，餘則較小，數值較大之各項，足使此兩種量數受極大之影響。在解釋時必有見及此，方可採用 S_y 及 σ_y 作為離散度之量數焉。

為標準誤，此可測量相關綫兩旁散佈之程度。標準誤之單位為絕對數，以 Y 之原來單位為單位，此與標準差相同。由相關方程式求得估計數後，即可用標準誤以斷定此估計數在某種限度內之機率。

在測量離散度時，吾人嘗論及離散度有消除絕對單位，而用一抽象量數之必要，而在比較兩種次數分配時，此點尤為重要；故在測量或比較離散度時，宜用一離散度係數 (coefficient of variation) 表示，此處亦有同樣之需要。吾人需要一不用原來單位之抽象係數以測量兩變量間相關之程度。此種係數為披爾遜 (Carl Pearson) 氏所創始。

此量數之意義可解釋如下。前嘗論及由相關方程式所得之估計數，其有價值與否，可由比較 Y 之標準誤 (即測量相關綫兩旁散佈程度之量數) 與 Y 之標準差以決定之。如標準誤之數值與標準差之數值相等，則相關方程式毫無功用可言；但如標準誤小於標準差，則用此式以求估計數可比較準確；故此式之有無價值，全視此兩種量數間之關係而定。此兩種量數均用絕對單位，故若彼此相除，即可得一抽象之量數。其式如下：

$$\text{相關量數} = \frac{S_y}{\sigma_y}$$

如將上項比率用另一形式表示，則此量數當更切實用。此式為

$$\text{相關量數} = \sqrt{1 - \frac{S_y^2}{\sigma_y^2}}$$

此量數如用於直綫相關之方程式，即稱相關係數 (coefficient of correlation)，而可以符號 r 代表之。

吾人可將此式略加討論，以明 r 之意義。如相關綫之兩旁絕無散佈點，則 S_y 之數值為零，此時用上式求得 r 之數值為 1，意謂兩變量間有完全關係 (perfect relationship) 之存在也。

S_y 之數值最大時可與 σ_y 相等。此時相關方程式對於吾人毫無裨助，由公式求得 r 之數值為零，此即表示兩變量間絕無一定關係之存在，換

言之，相關綫爲一經過 Y 之算術平均數點之水平綫。此係表示當 X 數值增高時， Y 數值既無聯帶增高之傾向，而成正相關；而當 X 數值減低時， Y 數值亦無聯帶增高之傾向，而成負相關，其兩變量間之變動絕無關聯也。此時各點與相關綫之距離，與各點與其平均數之距離相等，而標準誤與標準差之數值亦相等。

由是知 r 之最小數值爲0，最大數值爲1。在實際工作上， r 之數值常在此兩限度之間，相關之程度愈高，則 r 之數值愈接近於1。 r 之數值愈大，則吾人對於相關方程式之可信程度亦愈高，因其所表示之關係與實際關係接近之機會愈多也。就上例汽車登記數與繳納個人所得稅人數求得之相關係數爲

$$r = \sqrt{1 - \frac{(60.6)^2}{(171.4)^2}} = .935$$

此數值表示兩變量間確有一定之密切關係(註)。

吾人如將相關方程式中常數 b 之正負號加於相關係數之前，則其意義更爲顯著。此符號在相關方程式中，原係表示直綫斜度之爲正爲負，用於相關係數，則可使吾人明瞭兩變量之相互關係爲正相關(direct relationship)，抑負相關(inverse relationship)。本例中當一變量之數值增高時，其另一變量之數值聯帶增高，故爲正相關，其係數應寫爲+.935。至於棉花產量與價格兩者之關係，則爲負相關，因一變量之數值增高時，其另一變量之數值聯帶減低也。

當吾人求得三種量數後，測度關係之工作即告完成。此三種量數者，一爲平均關係之方程式，在兩變量間如有一定法則之存在，則此式即所以表示此種法則者也；一爲標準誤，爲用絕對數測度相關綫兩旁之離散度者也；一爲相關係數，係用抽象量數以測度上項方程式所表示之

(註) r 之數值在此所以較高者，一部份乃受六個或八個極大項目之影響也。關於極端項目

對於相關係數之影響，下文當再詳論之。

平均關係在實際上是否確切之程度者也。

相關量數計算手續之詳述

上節僅就測度兩變量關係時所用之各種量數，加以申述，而未詳示其計算步驟。吾人現時當說明此項計算手續及各種簡捷計算法。

前例所示之計算步驟，係求三種量數時所應有者。此種步驟通常雖可採用，惟其計算手續頗為繁重，吾人倘用簡捷法以求 S_y ，則計算工作可減省不少。此法可先就前例之資料說明之。惟此際所討論者，僅以兩變量關係之可以直綫表示者為限。

計算之第一步，應先求一相關方程式。用最小平方法配合直綫之普遍方程式為：

$$Y = a + bX$$

第二步為計算 S_y^2 ，即標準誤之平方也。在上例中係先求各點距此直綫之差數，而後將此差數自乘，更將各項乘積相加而平均之即得。但此值亦可由下式求得(註)。

$$S_y^2 = \frac{\Sigma(Y^2) - a\Sigma(Y) - b\Sigma(XY)}{N}$$

(註)標準誤係由下列公式求得。

$$S_y^2 = \frac{\Sigma(d^2)}{N}$$

此式中之 d 代表每點與所配直綫之距離，亦即 Y 之實際數值與 Y 之計算數值之差數。

Y 之計算數值係由下式求得。

$$Y_c = a + bX \quad (\text{符號 } Y_c \text{ 代表 } Y \text{ 之計算數值})$$

如以 Y 代表實際數值，則每項差數為

$$d = Y_c - Y$$

$$\text{即} \quad d = a + bX - Y \quad (1)$$

各點之 d 均可以此種方程式表示，故點數有若干點，即有相似之方程式若干個。以 d 乘

上式內 a 與 b 為所配直綫方程式中之常數，餘為觀察所得之原來數值。將表九十六中之各項數值及第310頁上所求得 a, b 之數值代入上式(註)，可得

$$S_y^2 = \frac{2,612,066 - (16.92 \times 7,616) - (1.91238 \times 1,215,966)}{43}$$

$$= \frac{157,814}{43} = 3,670$$

$$S_y = 60.6$$

此後吾人即可用前述步驟，按照下式求 r 。

$$r = \sqrt{1 - \frac{S_y^2}{\sigma_y^2}}$$

但求此相關係數 r 時，亦可不必先求 S_y 之數值，因上項求 r 之公式可化爲

$$r^2 = \frac{a \Sigma(Y) + b \Sigma(XY) - Nc_y^2}{\Sigma(Y^2) - Nc_y^2}$$

(續上頁註)

每一方程式之各項而後相加，得

$$\Sigma(d^2) = a \Sigma(d) + b \Sigma(dX) - \Sigma(dY) \quad (2)$$

但因直綫係用最小平方法所求得，故

$$\Sigma(d) = 0$$

$$\Sigma(dX) = 0 \quad (\text{證明見附錄A})$$

故 $\Sigma(d^2) = -\Sigma(dY) \quad (3)$

吾人若以 Y 乘(1)式內各項，而後將各個方程式相加，則得

$$\Sigma(dY) = a \Sigma(Y) + b \Sigma(XY) - \Sigma(Y^2) \quad (4)$$

將(3)式內 $\Sigma(dY)$ 之數值代入，得

$$\Sigma(d^2) = \Sigma(Y^2) - a \Sigma(Y) - b \Sigma(XY) \quad (5)$$

由此式可得 S_y^2 之公式。

(註)此處 b 之小數點後之位數，較前所用者多三位。

上式中 c_y 為 Y 之算術平均數與計算時所用假定原點之差(註)。如此原點為原來 Y 尺度上之零點, 則 c_y 等於 Y 之算術平均數。在本例中用表九十六之資料, 可得

$$c_y = \frac{7,616}{43} = 177.116$$

其他數值均與求 S_y 所用者相同。將各項數值代入式中, 得

$$r^2 = \frac{1,105,337.98}{1,263,152.7} = .875$$

$$r = .935$$

實際上在用最小平方方法配合直綫後, 已將求 S 及 r 時所需之各數值求出, 故應用上式時, 所應計算之數值僅 $\Sigma(Y^2)$ 與 c_y 而已。

相關表之編製

表九十六所舉之例祇包含四十三項。在研究相關時, 倘遇項數過多, 則因計算手續過於繁複, 其勢不能保留各項之原來數值, 而須將資料歸併成組, 方可着手計算; 換言之, 即吾人須根據組成次數分配之資

(註)
$$r^2 = 1 - \frac{S_y^2}{\sigma_y^2}$$

之公式可寫作

$$r^2 = 1 - \frac{\Sigma(d^2)}{\Sigma(y^2)}$$

此處 y 為各點對於 Y 之算術平均數之離中差。但

$$\frac{\Sigma(y^2)}{N} = \frac{\Sigma(Y^2)}{N} - c_y^2$$

式中 Y 代表各點與假定原點(此處即為原來尺度上之零點)之離中差, c_y 代表此原點與 Y 之算術平均數之差。故

$$r^2 = 1 - \frac{\Sigma(d^2)}{\Sigma(Y^2) - Nc_y^2}$$

將第318頁附註中所得 $\Sigma(d^2)$ 之相等數值代入此式, 得

$$r^2 = 1 - \frac{\Sigma(Y^2) - a\Sigma(Y) - b\Sigma(XY)}{\Sigma(Y^2) - Nc_y^2}$$

將此式化簡, 得

$$r^2 = \frac{a\Sigma(Y) + b\Sigma(XY) - Nc_y^2}{\Sigma(Y^2) - Nc_y^2}$$

(correlation table)。其排列方式，均以便於計算相關為主。

茲以聯邦準備銀行對於標準式商業票據 (standard type of commercial paper) 貼現率與各商業銀行對於此種票據貼現率之相關資料為例，說明相關表之編製方法。因此種票據得由各會員銀行 (member bank) 向聯邦準備銀行請求重貼現，故兩種貼現率之間，在理論上應有相當關係之存在。此處所欲討論者，即為測度此種關係之問題。

吾人首應將原來資料表列之。此處所用之資料為美國十二個聯邦準備城市 (Federal Reserve cities) 一九二〇年七月一日至一九二三年十二月一日止四十二個月內每月之貼現率(註)。表列此種項目時，每一聯邦準備銀行之貼現率須與其同一城市之商業銀行貼現率相配。圖六十八係舉示此種表列方法。

項目既經表列，更須作一相關表，藉便計算。茲將此表列下：

表九十八

美國聯邦準備銀行貼現率與商業銀行貼現率之相關表

X—聯邦準備銀行貼現率(百分率)

組距	中點	f	d'	fd'	f(d') ²	3.75	4.25	4.75	5.25	5.75	6.25	6.75	總數
						4.24	4.74	5.24	5.74	6.24	6.74	7.24	
						4.00	4.50	5.00	5.50	6.00	6.50	7.00	
						24	227	48	20	123	20	42	504
						0	1	2	3	4	5	6	
						0	227	96	60	492	110	252	1227
						0	227	192	180	1968	510	1512	4579
										1	1	2	
7.75-8.24	8.00	4	8	32	256								
7.25-7.74	7.50	17	7	119	833					7	9	1	
6.75-7.24	7.00	117	6	702	4212			5	4	63	9	36	
6.25-6.74	6.50	47	5	235	1175		2	9	10	22	1	3	
5.75-6.24	6.00	156	4	624	2496	1	90	29	6	30			
5.25-5.74	5.50	126	3	378	1134	11	110	5					
4.75-5.24	5.00	34	2	68	136	10	24						
4.25-4.74	4.50	3	1	3	3	2	1						
總數		504		2161	10245								

(註見下頁)

在此表中, X 及 Y 變量皆取一假定原點, 各按組距單位計算之。此處亦可用 X' , Y' 以代表各項與假定原點之差 (此原點在原來尺度上為 4.4 之一點)。

相關量數之計算

配合相關直綫與計算 S 及 r 所需之資料, 俱列於相關表中。吾人當已明瞭 $\Sigma(X')$, $\Sigma(X')^2$, $\Sigma(Y')$ 及 $\Sigma(Y')^2$ 之計算方法。至於 $\Sigma(X'Y')$ 一項原可由相關表中計算而得, 但如將表內項目用直行式列如下表, 則初學者可易於瞭解。表中每一橫行代表相關表中每小方格內之數。

表 九 十 九

美國聯邦準備銀行之貼現率與商業銀行之貼現率

配合曲綫所需數值之計算			
(1)	(2)	(3)	(4)
X'	Y'	f	$f(X'Y')$
0	1	2	0
0	2	10	0
0	3	11	0
0	4	1	0
1	1	1	1
1	2	24	48
1	3	110	330
1	4	90	360
1	5	2	10
2	3	5	30
2	4	29	232
2	5	9	90
2	6	5	60
3	4	6	72
3	5	10	150

(上頁附註)聯邦準備銀行之貼現率係九十日期商業承兌匯票之貼現率, 商業銀行之貼現率則係三十日至六十日期(亦有三十日至九十日期者)商業票據之貼現率。普通三十日期之貼現率應視為該期中間一日之貼現率。

表九十九 (續)

(1) X'	(2) Y'	(3) f	(4) $f(X'Y')$
3	6	4	72
4	4	30	480
4	5	22	440
4	6	63	1,512
4	7	7	196
4	8	1	32
5	5	1	25
5	6	9	270
5	7	9	315
5	8	1	40
6	5	3	90
6	6	36	1,296
6	7	1	42
6	8	2	96
			6,289

配合直綫與計算標準誤及相關係數所需之各項數值爲

$$\begin{aligned}
 N &= 504 & \Sigma(X')^2 &= 4,579 \\
 \Sigma(X') &= 1,227 & \Sigma(X'Y') &= 6,289 \\
 \Sigma(Y') &= 2,161 & \Sigma(Y')^2 &= 10,245
 \end{aligned}$$

用上列數值求得最適合之直綫方程式爲

$$Y' = 2.7155 + .6458X'$$

將各項數值代入下式，

$$S_y^2 = \frac{\Sigma(Y')^2 - a\Sigma(Y') - b\Sigma(X'Y')}{N}$$

可得

$$S_y^2 = \frac{10,245 - (2.7155 \times 2,161) - (.6458 \times 6,289)}{504} = .6257$$

$$S_y = .791$$

求相關係數時，僅須將各項數值代入下式，

$$r^2 = \frac{a\Sigma(Y') + b\Sigma(X'Y') - Nc_y^2}{\Sigma(Y')^2 - Nc_y^2}$$

即得

$$\begin{aligned}
 r^2 &= \frac{(2.7155 \times 2,161) + (.6458 \times 6,289) - (504 \times 18.38437)}{10,245 - (504 \times 18.38437)} \\
 &= \frac{663.9}{979.27} \\
 &= .67795 \\
 r &= +.82
 \end{aligned}$$

此處逐步計算俱以原來尺度 (original scale) 上之4,4點作為原點，並以組距為單位者，其 r 之數值並不因此而受影響，惟直綫之方程式及標準誤之數值則須校正之。

用組距單位求得 S_y 之數值為.791， Y 變量之組距為.5%，故用原來單位表示 S_y 之數值應為

$$\begin{aligned}
 S_y &= .5\% \times .791 \\
 &= .40\%
 \end{aligned}$$

直綫方程式可用同樣方法校正之。 X, Y 之組距皆為.5%，故此式可寫作

$$2Y' = 2.7155 + .6458(2X')$$

$$\text{或} \quad Y' = 1.3577 + .6458X'$$

此即用原來單位表示之直綫方程式也。

如欲將此式原點移至固有原點，可以 $Y' = Y - 4$ 及 $X' = X - 4$ 代入上式，得

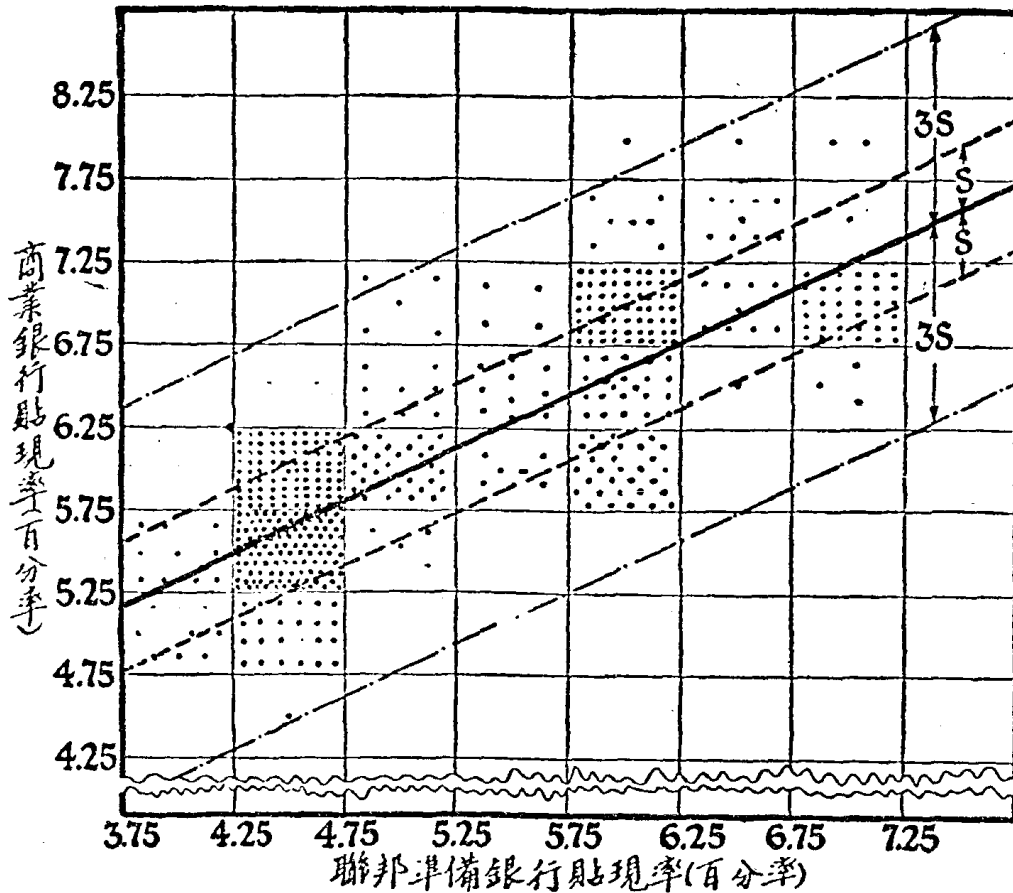
$$Y - 4 = 1.3577 + .6458(X - 4)$$

$$\text{或} \quad Y = 2.7745 + .6458X$$

此時測度聯邦準備銀行貼現率與商業銀行貼現率關係之三數值俱已求得。其中直綫方程式係用以表述兩變量間之平均關係，標準誤用以測驗由此式所得估計之可靠程度，相關係數則係用抽象數字以測度兩變量間之相關程度者也。

標準誤 S_y 之意義可用圖六十九解釋之。在此散佈圖 (scatter dia-

gram)上,先繪相關直綫,而後在此綫兩旁各繪平行綫,以記明所謂‘估計區域’(zones of estimate)。假定次數分配合乎常態,則在此綫兩邊各占1S之區域內(即2S之距離內),應含有全體點數之68%;在此綫兩旁各占3S之區域內(即6S之距離內),應含有全體點數之99.7%。S之數值愈小,則此估計區域愈狹,而根據此相關直綫所得之估計亦愈精確。



圖六十九. 美國聯邦準備銀行與商業銀行貼現率之散佈圖(附相關綫及估計區域)

計算相關係數之積差公式(product-moment formula)

在上例中相關係數俱由公式

$$r^2 = \frac{a \sum(Y) + b \sum(XY) - Nc_y^2}{\sum(Y^2) - Nc_y^2}$$

算得。此式乃直接由配合直綫所用之最小平方法而來。普通求r之公式

與此稍異，學者亦應知之。

當配合直綫時，倘用 X 及 Y 之算術平均數(M_x, M_y)點爲原點，則下列兩常態方程式(normal equations)

$$\Sigma(Y) = na + b \Sigma(X)$$

$$\Sigma(XY) = a \Sigma(X) + b \Sigma(X^2)$$

應簡化爲一式如下：

$$\Sigma(xy) = b \Sigma(x^2)$$

因 $\Sigma(x) = 0$, $\Sigma(y) = 0$ 也。

此處僅須求直綫之斜度 b ，而可由下式求得之：

$$b = \frac{\Sigma(xy)}{\Sigma(x^2)}$$

故公式

$$r^2 = \frac{a \Sigma(Y) + b \Sigma(XY) - Nc_y^2}{\Sigma(Y^2) - Nc_y^2}$$

在此處亦可化爲

$$r^2 = \frac{b \Sigma(xy)}{\Sigma(y^2)}$$

因用 Y 之平均數爲原點以計算離中差時， $c_y = 0$ 也。將前項 b 之相等數值代入上式，得

$$r^2 = \frac{\Sigma(xy) \cdot \Sigma(xy)}{\Sigma(y^2) \cdot \Sigma(x^2)}$$

但 $\Sigma(y^2) = N\sigma_y^2$, $\Sigma(x^2) = N\sigma_x^2$

故 $r^2 = \frac{\Sigma(xy) \cdot \Sigma(xy)}{N^2 \sigma_y^2 \sigma_x^2}$

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

在此式中， x, y 爲以 X, Y 之平均數點爲原點時所得之差。

r 之公式又可寫作

$$r = \frac{\bar{p}}{\sigma_x \sigma_y}$$

在此式中

$$\bar{p} = \frac{\Sigma(xy)}{N}$$

\bar{p} 實即 x, y 相配數值之均積數(mean product)也。

由上式計算相關係數之步驟與前稍異。吾人在計算算術平均數及標準差時，為求計算手續簡捷起見，嘗取一假定原點，以計算各數量之離中差，最後就所得結果，校正其因用假定原點所發生之差誤。計算均積數時，亦可採用簡捷法而由假定原點計算，最後校正其差誤。此假定原點即假定平均數點也。

如以 x', y' 表示由假定平均數點計算所得之離中差，以 \bar{p}' 表示此項離中差之均積數，則

$$\bar{p}' = \frac{\Sigma(x'y')}{N}$$

\bar{p}' 之計算並不困難，因離中差 x', y' 可由各組之中點計算之，而又可以組距作為單位也。 \bar{p}' 之數值求得後，即可由下式(註)

$$\bar{p} = \bar{p}' - c_x c_y$$

求真實之均積數 \bar{p} 。該式中 c_x 為 x 之真實平均數與其假定平均數之差， c_y 為 y 之真實平均數與其假定平均數之差。

(註)此式證明於下：

$x' = x$ 數列之各項對其假定平均數之離中差

$x = x$ 數列之各項對其真實平均數之離中差

$c_x = x$ 數列之真實平均數與其假定平均數之差

$y' = y$ 數列之各項對其假定平均數之離中差

$y = y$ 數列之各項對其真實平均數之離中差

$c_y = y$ 數列之真實平均數與其假定平均數之差

$x' = x + c_x$

$y' = y + c_y$

$x'y' = (x + c_x)(y + c_y) = xy + c_x y + c_y x + c_x c_y$

將 N 項之 $x'y'$ 相加，得

積差法之應用(根據未分組之資料計算)

應用積差法求相關係數之計算手續，可先就前述美國四十三州繳納個人所得稅人數(X)與汽車登記數(Y)之未分組資料說明之。計算時所需之數值，可由表九十六得之如下：

$$\begin{aligned} N &= 43 \\ \Sigma(X) &= 3,602 \\ \Sigma(Y) &= 7,616 \\ \Sigma(X^2) &= 603,968 \\ \Sigma(Y^2) &= 2,612,066 \\ \Sigma(XY) &= 1,215,966 \end{aligned}$$

均積數可由下式求之：

$$\hat{p} = \frac{\Sigma(xy)}{N} = \frac{\Sigma(x'y')}{N} - c_x c_y$$

如用X, Y原來尺度上之原點作為假定原點，則上式可化為

$$\hat{p} = \frac{\Sigma(XY)}{N} - c_x c_y$$

X, Y之標準差 σ_x, σ_y 可由下列兩式求得之：

$$\begin{aligned} \sigma_x &= \sqrt{\frac{\Sigma(X^2)}{N} - c_x^2} \\ \sigma_y &= \sqrt{\frac{\Sigma(Y^2)}{N} - c_y^2} \end{aligned}$$

(續上頁註)

$$\Sigma(x'y') = \Sigma(xy) + c_x \Sigma(y) + c_y \Sigma(x) + N c_x c_y$$

$$\text{但} \quad \Sigma(y) = 0, \Sigma(x) = 0$$

$$\text{故} \quad \Sigma(x'y') = \Sigma(xy) + N c_x c_y$$

$$\frac{\Sigma(x'y')}{N} = \frac{\Sigma(xy)}{N} + c_x c_y$$

$$\frac{\Sigma(xy)}{N} = \frac{\Sigma(x'y')}{N} - c_x c_y$$

$$\hat{p} = \hat{p}' - c_x c_y.$$

(當假定原點在原來尺度上之零點時,則 X, Y 兩符號與公式中之 x', y' 相同)。

前述各量數均可由表九十六內之數值計算而得。

$$c_x = \frac{3,602}{43} = 83.767 \qquad c_y = \frac{7,616}{43} = 177.116$$

$$c_x^2 = 7,016.910 \qquad c_y^2 = 31,370.077$$

$$p = \frac{1,215,966}{43} - 14,836.476$$

$$= 13,441.803$$

$$\sigma_x = \sqrt{\frac{603,968}{43} - 7,016.910} \qquad \sigma_y = \sqrt{\frac{2,612,066}{43} - 31,370.077}$$

$$= 83.84$$

$$= 171.41$$

由此求得相關係數為

$$r = \frac{p}{\sigma_x \sigma_y} = \frac{13,441.803}{83.84 \times 171.41} = +.935$$

表示 X 與 Y 兩變量平均關係之直綫方程式,亦可由計算相關係數所需之各數值求得之。當原點在 X, Y 之平均數點時,此方程式可寫作

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

將各數值代入上式,得

$$y = +.935 \frac{171.41}{83.84} x$$

$$= 1.91x$$

此式與以前用‘最小平方法’所求得之直綫方程式相似,所稍異者,在此式中僅缺少表示 y 截部(y -intercept)之常數一項耳。此因式中之原點在 X, Y 之平均數點,而由‘最小平方法’所求得之直綫必經過此平均數點

之故也(註)。

如用積差法求相關係數及迴歸方程式 (equation of regression) 時,標準誤 S_y 之數值可由前述求 r 之公式推演而得。此公式原為

$$r^2 = 1 - \frac{S_y^2}{\sigma_y^2}$$

可變為

$$S_y = \sigma_y \sqrt{1 - r^2}$$

如吾人已知 σ_y 及 r 之數值,即可求得 S_y 。在上例中

$$\begin{aligned} S_y &= 171.41 \sqrt{1 - (.935)^2} \\ &= 60.6 \end{aligned}$$

積差法之應用(根據已分組之資料計算)

如遇資料項數過多,而必先編成雙項次數表(double frequency table,或稱相關表 correlation table)時,吾人亦可用積差法以計算相關係數。其計算手續詳見表一〇〇內。

該表之格式與前述根據同一資料所製之相關表完全相同,惟所取假定原點不同耳。

表中 X 之假定平均數為5.5, Y 之假定平均數為6.5,由此兩原點所得之離中差均以組距為單位。相關表之每一小格內共有三項數字,用以計

(註) $y = r \frac{\sigma_y}{\sigma_x} x$ 之方程式與最小平方方法方程式相同之原因,極易證明。當直線經過 X, Y 之

平均數點時, $Y = a + bX$ 之方程式可變為 $y = bx$,

但 $b = \frac{\Sigma(xy)}{\Sigma(x^2)}$, 故此式可寫為 $y_0 = \frac{\Sigma(xy)}{\Sigma(x^2)} x$

此與 $y_c = r \frac{\sigma_y}{\sigma_x} x$ 相同,

因 $y_c = r \frac{\sigma_y}{\sigma_x} x$ 可寫作:

$$(1) \quad y_c = \frac{\Sigma(xy)}{N\sigma_y\sigma_x} \cdot \frac{\sigma_y}{\sigma_x} x \quad (3) \quad y_0 = \frac{\Sigma(xy)}{N \sqrt{\frac{\Sigma(x^2)}{N}} \cdot \sqrt{\frac{\Sigma(y^2)}{N}}} x$$

$$(2) \quad y_0 = \frac{\Sigma(xy)}{N\sigma_x \cdot \sigma_x} x \quad (4) \quad y_0 = \frac{\Sigma(xy)}{\Sigma(x^2)} x$$

(在上列各式中 y_c 皆係表示 y 之計算數值,以免與同式右邊真實數值 y 相混淆。)

表一〇〇
美國聯邦準備銀行貼現率與商業銀行貼現率之相關表
相關係數之計算

組距		X—聯邦準備銀行貼現率(百分率)					Y—商業銀行貼現率(百分率)					總數	$\Sigma(x'y')$
中點	f	d'	fd'	f(d') ²	3.75-4.24	4.25-4.74	4.75-5.24	5.25-5.74	5.75-6.24	6.25-6.74	6.75-7.24		
					4.0	4.5	5.0	5.5	6.0	6.5	7.0		
					24	227	48	20	123	20	42	504	
					-3	-2	-1	0	+1	+2	+3		
					-72	-454	-48	0	+123	+40	+126	-285	
					216	908	48	0	123	80	378	1753	
7.75-8.24	8.0	4	+3	+12	36				+3	+6	+9		+27
									1	1	2		
									(+3)	(+6)	(+9)		
7.25-7.74	7.5	17	+2	+34	68				+2	+4	+6		+56
									7	9	1		
									(+14)	(+36)	(+6)		
6.75-7.24	7.0	117	+1	+117	117		-1	0	+1	+2	+3		+184
							5	4	63	9	36		
							(-5)	(0)	(+63)	(+18)	(108)		
6.25-6.74	6.5	47	0	0	0	0	0	0	0	0	0		
						2	9	10	22	1	3		
						(0)	(0)	(0)	(0)	(0)	(0)		
5.75-6.24	6.0	156	-1	-156	156	+3	+2	+1	0	-1			+182
						1	90	29	6	30			
						(+3)	(+180)	(+29)	(0)	(-30)			
5.25-5.74	5.5	126	-2	-252	504	+6	+4	+2					+516
						11	110	5					
						(+66)	(+440)	(+10)					
4.75-5.24	5.0	34	-3	-102	306	+9	+6						+234
						10	24						
						(+90)	(+144)						
4.25-4.74	4.5	3	-4	-12	48	+12	+8						+32
						2	1						
						(+24)	(+8)						
總數		504		-369	1235								+1231

$A.M_x = 5.5$

$A.M_y = 6.5$

$p = \frac{\Sigma(x'y')}{N} - c_x c_y$

$c_x = \frac{-285}{504}$

$c_y = \frac{-359}{504}$

$= \frac{1231}{504} - .402$

$= -.565$

$= -.712$

$= 2.4425 - .402$

$c_x^2 = .319$

$c_y^2 = .507$

$= 2.0405$

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

$s_y^2 = \frac{1235}{504}$

$r = \frac{p}{\sigma_x \sigma_y}$

$= \frac{1753}{504}$

$= 2.450$

$r = \frac{2.0405}{1.777 \times 1.394}$

$= 3.478$

$\sigma_y^2 = 2.450 - .507$

$= \frac{2.0105}{2.477}$

$\sigma_x^2 = s^2 - c^2$
 $= 3.478 - .319$

$= 3.159$

$= 1.943$

$r = +.82$

$\sigma_x = \sqrt{3.159}$

$= 1.943$

$M_x = 5.218$

$= 1.777$

$\sigma_y = 1.394$

$M_y = 6.244$

註：表內各項計算時，皆以組距為單位。

算 $\Sigma(x'y')$ 者也。其中間一項數字爲此小格內之項數，其上一數爲 x' 與 y' 之乘積，其下一數則爲此兩數（即項數與 $x'y'$ ）之乘積也。例如 X 值在 5.75—6.25 組內（中點爲 6），同時 Y 值在 7.25—7.75 組內（中點爲 7.5）時，共有七對數字，故項數爲 7。每對數字之 x' （ X 之數值對其假定平均數之離中差）爲 +1（以組距爲單位）， y' （ Y 之數值對其假定平均數之離中差）爲 +2（以組距爲單位），故每對數字之 $x'y'$ 爲 +2，此爲該小格內上端之數字；但該小格內共有七對數字，故其 $x'y'$ 之和爲 +14，此爲該小格內下端附有括弧之數字。欲求 $x'y'$ 之總和〔即 $\Sigma(x'y')$ 〕，須將各小格內之下端數字相加，並須注意其正負號。計算時吾人可先將每一橫行內之 $x'y'$ 各別相加，而一一記其和數（sub-totals）於表之右方直行內，而後將各個和數相加，即得 $x'y'$ 之總和 $\Sigma(x'y')$ 。此處 $\Sigma(x'y')$ 之數值爲 +1231，並係以組距爲單位者。

c_x 與 c_y 兩個數值乃用以求 X, Y 之標準差者，其計算方法前已述及。此項數值如按組距單位計算，則爲

$$c_x = -.565 \qquad c_y = -.712$$

由此可得

$$\begin{aligned} \frac{\Sigma(xy)}{N} &= \frac{\Sigma(x'y')}{N} - c_x c_y \\ &= \frac{1231}{504} - .402 \\ &= 2.0405 \end{aligned}$$

此即以組距單位表示之均積數 p 之數值也。由此數值乃可進而計算 r 如下：

$$\begin{aligned} r &= \frac{\Sigma(xy)}{N\sigma_x\sigma_y} \\ &= \frac{p}{\sigma_x\sigma_y} \end{aligned}$$

$$= \frac{2.0405}{1.777 \times 1.394}$$

$$= +.82$$

上式中分子及分母(均積數及兩個標準差)皆以組距為單位,故計算 r 時亦可逕以組距為單位,不必將此項數字化為原來單位。惟須注意者,在計算其他各種量數時,亦皆以組距為單位耳。

由公式

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

求直綫方程式,以表示 x 與 y 之平均關係時, σ_y 及 σ_x 之數值,應以原來單位表示之(註)。此可以組距乘求得之數值而得。

$$\sigma_x(\text{原來單位}) = 1.777 \times .50 = .8885$$

$$\sigma_y(\text{原來單位}) = 1.394 \times .50 = .697$$

將此兩數值代入前式,可得

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

$$= .64x$$

迴歸綫

以上討論相關時,曾將通常所用之數種名詞故意略去,茲擇要解釋之。

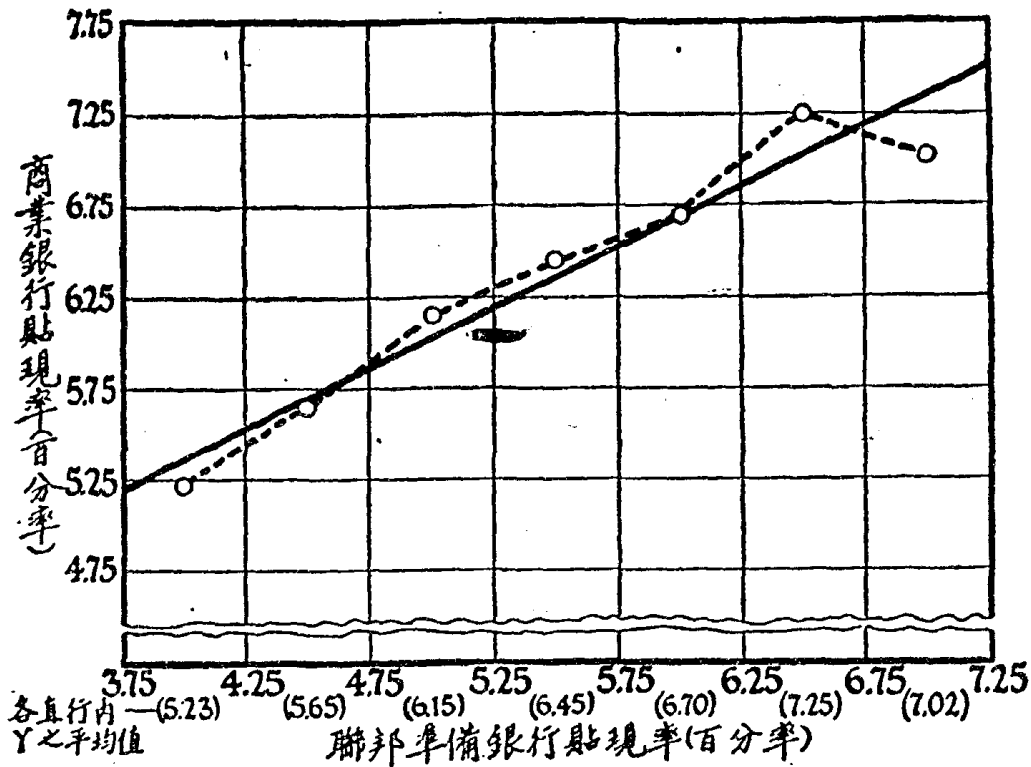
在前例中,吾人已知配合最為適宜之直綫方程式為

$$y = .64x$$

此式原點係在 X, Y 之平均數點。式中 y 為 x 之函數,即 x 為自變量(independent variable), y 為倚變量(dependent variable)也。此式所表示

(註)本例中 X 與 Y 之組距相同,由組距單位化為原來單位之手續原可省略,因式中分子分母之關係可不受其影響,但通常計算至此階段時,仍應將 σ_x 及 σ_y 之單位化為原來單位。

者為當 x (聯邦準備銀行貼現率)增減一單位時, y (商業銀行貼現率)所聯帶發生之平均變動。此綫之性質與長期趨勢綫極相類似,蓋長期趨勢綫為表示一變量因時間變遷而發生之平均變動,此綫則為表示兩變量間之平均關係者也。後者在統計學中謂之迴歸綫(lines of regression),其方程式稱為迴歸方程式(regression equation);表示該綫斜度之 $r \frac{\sigma_y}{\sigma_x}$ 數值稱為迴歸係數(coefficient of regression)。上列諸名詞最初為加爾頓氏(Galton)所採用。氏於研究父子高度之關係時,嘗發現兒子平均高度與全族平均高度之相差,比較其與父親平均高度之相差為小,且無論父親之高度為高於或低於全族之平均高度,其子之高度每有迴歸於全族平均高度之趨向,故加爾頓氏稱此兩變量平均關係之直綫為迴歸綫。惟現時普通用此名詞時,大都已失其原意矣。



圖七十. 商業銀行貼現率與聯邦準備銀行貼現率之關係(斷綫係接連各直行內Y之平均值而成,直綫係表示聯邦準備銀行貼現率增減一個單位時,商業銀行貼現率所聯帶發生之平均變動;此即Y倚X變動之迴歸綫)。

任何兩數列之相關皆可得兩條迴歸綫之方程式，其一為以 y 為倚變量 x 為自變量所得平均關係之綫，另一綫為以 x 為倚變量 y 為自變量所得平均關係之綫，兩綫之意義可用圖表示之。

試將圖七十及圖七十一加以比較即可明瞭該兩綫之意義。圖七十係由圖六十九之散佈圖直接求得者。圖內每一直行內之小圈係表示該行內 Y 之算術平均數。例如第一直行內共有24項，其各項之 X 值皆在3.75%至4.25%之間，但各項之 Y 值高下不同，其次數分配如下：

表一〇一

由一整列計算算術平均數之方法

組距	中點 (m)	次數 (f)	fm
5.75—6.24	6.0	1	6.0
5.25—5.74	5.5	11	60.5
4.75—5.24	5.0	10	50.0
4.25—4.74	4.5	2	9.0
		24	125.5

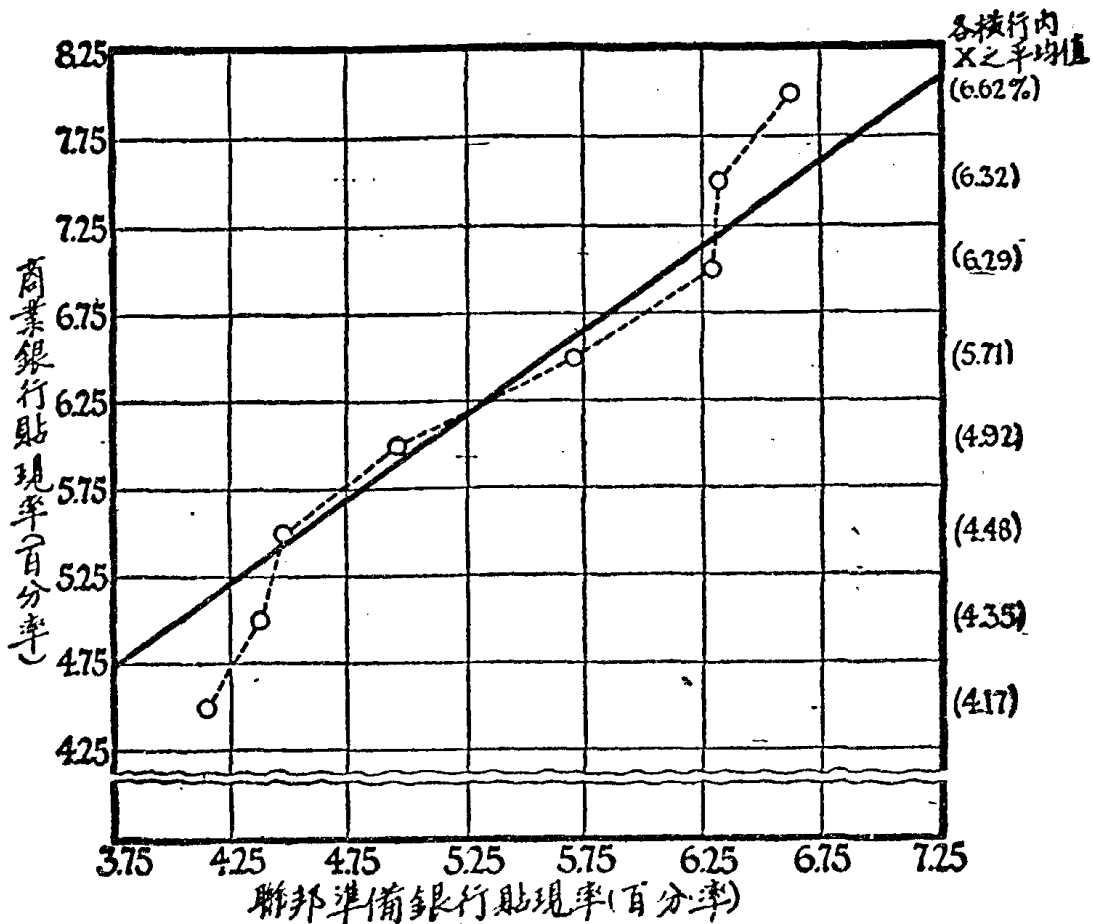
$$M = \frac{125.5}{24} = 5.23$$

由上表求得該直行內 Y 之算術平均數為5.23。其他各直行內 Y 之算術平均數可用同法求之。各直行之算術平均數與 Y 倚 X 而變動之迴歸綫(line of regression of Y on X)皆繪於圖七十內。

在此圖內 X (聯邦準備銀行貼現率)為自變量。當 X 值由4.0%漸增至4.5%, 5.0%, 5.5%及以上時，商業銀行之平均貼現率亦隨之增高。準備銀行之平均率為4%時，商業銀行之平均率為5.23%；準備銀行之平均率為4.5%時，商業銀行之平均率為5.65%；餘類推。圖中直綫(即迴歸綫，亦係表示平均關係之綫)之斜度可測度聯邦準備銀行貼現率增高一個單位時，商業銀行貼現率所聯帶發生之平均增加數。

但此兩變量之關係又可由另一方面觀之。例如已知某一商業銀行之貼現率，而欲推算與該率相配之準備銀行平均貼現率，或已知商業銀

行貼現率之變動，而欲推知準備銀行貼現率所聯帶發生之平均變動，則此時商業銀行貼現率應視為自變量，準備銀行貼現率應視為倚變量，此可用圖七十一表之。圖中接連斷續綫之小圈係表示橫行內各項之算術平均數。例如第一橫行內 X 共有三項，其平均數為4.17%，此即商業銀行貼現率為4.5%時，準備銀行之平均貼現率也。又由圖中可得商業銀行貼現率為5%時，準備銀行之平均貼現率為4.35%，餘類推。由是知配合於各點上之直綫，亦可表示兩種貼現率間之關係，惟其斜度所測度者，乃係商業銀行貼現率增減一個單位時，準備銀行貼現率之平均增減數也。



圖七十一。聯邦準備銀行貼現率與商業銀行貼現率之關係(斷續綫係接連各橫行內 X 之平均數值而成，直綫係表示商業銀行貼現率增減一個單位時，聯邦準備銀行貼現率所聯帶發生之平均變動；此即 X 倚 Y 而變動之迴歸綫)。

此綫為 X 倚 Y 而變動之迴歸綫。求此迴歸綫之普通公式為

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

將本例中之數值代入上式，可得

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

即 $x = 1.045y$

此式中之各因子與 Y 倚 X 而變動之迴歸綫公式中所用者相同(註)。倘 r 等於1，則此兩綫合而為一，又若兩標準差相同，則迴歸綫平分 X 軸與 Y 軸所成之直角。在任何一例中，倘將各點繪於以標準差為單位之圖上，則 $y = rx$ ，而迴歸綫之斜度適等於 r 。

迴歸係數通常用符號 b 表示之。在一簡單相關之例中，此種係數有二，所以代表兩迴歸綫之斜度者也。此兩係數為

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

(b 後二小字表示兩變量之關係，其前一字表示倚變量。)

兩式中均含係數 r ，故 r 之數值又可由兩迴歸係數求得之，因

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

故若吾人已知兩迴歸綫之斜度，即可決定 r 之數值。在本例中將兩項數

(註)公式 $x = r \frac{\sigma_x}{\sigma_y} y$

可化為 $x = \frac{\Sigma(xy)}{\Sigma(y^2)} y$

此為配合於圖七十一內各點之直綫方程式。各點離此直綫之橫距(horizontal deviations)，其各別自乘後相加所得之數值，比離其他任何直綫為小。

公式 $y = \frac{\Sigma(xy)}{\Sigma(x^2)} x$

為另一直綫方程式。各點離此直綫之縱距(vertical deviations)，其各別自乘後相加所得之數值，比離其他任何直綫為小。明乎此則於兩迴歸綫之異同，不難瞭解矣。

值代入，得

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

迴歸方程式之應用

前例中求得迴歸綫之兩個方程式為

$$y = .64x$$

$$x = 1.045y$$

此兩式係表示 X 各點之離中差與 Y 各點之離中差兩者間之關係。兩式之原點均在 X 及 Y 之平均數點，故應用兩式時，不可運用 X, Y 之原來數值，而必須採用其離中差。例如吾人欲斷定聯邦準備銀行之貼現率為6%時商業銀行之貼現率，可計算如下。

X （聯邦準備銀行貼現率）之平均數為5.218%，故6%之離中差為+.782，將此值代入前述第一式得

$$y = .64(+.782)$$

$$= +.500$$

故當 X 之離中差為+.782時， Y 之平均離中差為+.500。更將此數與商業銀行之平均貼現率6.144%相加，得6.644%，此即準備銀行貼現率為6%時，商業銀行之正常貼現率也。

上項計算頗為曲折，如將方程式之形式稍加變更，則計算較易。

令 $\bar{X} = X$ 之算術平均數

$\bar{Y} = Y$ 之算術平均數

則 $y = r \frac{\sigma_y}{\sigma_x} x$

之方程式可寫作

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

該式中 X, Y 代表此兩變量之原來數值，而非離中差。在坐標圖（coör-

dinate chart) 上言之,此即等於將原點由平均數點移至原來尺度上之零點也。

此式之功用可舉例以明之。將

$$y = .64x$$

之方程式變為

$$\begin{aligned} Y - 6.144 &= .64(X - 5.218) \\ &= .64X - 3.340 \\ Y &= .64X + 2.804 \end{aligned}$$

則式中之原點已由 X, Y 之平均數點移至原來尺度上之零點,故計算時吾人可運用 X, Y 之原來數值,而不必計算其離中差。例如欲求聯邦準備銀行貼現率為6%時之商業銀行貼現率,可逕將此數代入上式,即

$$\begin{aligned} Y &= (.64 \times 6.0) + 2.804 \\ &= 6.644 \end{aligned}$$

此與前此用原式所得之答數相同,但此式因可運用原來數值計算,故較原式為便。

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

亦可同樣變為

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

概略

本章對於測度兩變量間之關係曾陳述兩法。茲再將兩法之步驟,在此撮要敘述之。兩者俱以‘最小平方法’為根據,但第一法首先應用最小平方之原理配合直綫,故可以最小平方法名之。

I. 最小平方法

A. 根據未分組之資料計算:

1. 用最小平方法配合一直綫。將資料分直行排列，俾便計算 $\Sigma(X)$, $\Sigma(Y)$, $\Sigma(X^2)$, $\Sigma(Y^2)$, $\Sigma(XY)$ 之數值。由此求得之直綫方程式即表示兩變量間之平均關係。

$$2. \text{ 由 } S_y^2 = \frac{\Sigma(Y^2) - a\Sigma(Y) - b\Sigma(XY)}{N}$$

之公式計算標準誤 S_y 。 S_y 可以測度由相關方程式所得估計之可靠性 (reliability)。標準誤對於相關綫之意義正與標準差對於算術平均數之意義相同。

$$3. \text{ 由 } r^2 = \frac{a\Sigma(Y) + b\Sigma(XY) - Nc_y^2}{\Sigma(Y^2) - Nc_y^2}$$

求相關係數 r 之數值，並以迴歸方程式中 b 之正負符號作為 r 之正負符號。兩變量間之關係如可用直綫表示時，則 r 為一測度此種關係之抽象係數。

4. 如欲求 X 倚 Y 之迴歸方程式 (X 為倚變量)，可將相當數值代入下列兩常態方程式

$$\Sigma(X) = na + b\Sigma(Y)$$

$$\Sigma(XY) = a\Sigma(Y) + b\Sigma(Y^2)$$

由此兩式即可求得 a , b 之數值，而得如下形式之迴歸方程式

$$X = a + bY$$

S_y 之公式，經相當改變後，即可用以計算標準誤 S_x 。惟 r 之數值，無論以 X 為倚變量，或以 Y 為倚變量，其結果常相同。

B. 根據分組之資料計算：

1. 擇一相當組距，將各項按組列成一相關表。
2. 計算配合直綫時所需之各數值。計算時兩變量須各擇一假定原點，並用組距單位計算之。將各項分直行排

列，俾便計算

$$\Sigma(X'Y')$$

之數值。

3. 用標準誤之公式計算標準誤 S_y 。
4. 用相關係數之公式計算相關係數 r 。
5. 前項計算如以組距為單位，則直綫方程式及標準誤之數值均應化為原來單位。又方程式中如曾採用一假定原點，此時須加改正，使式中之各變量皆為對於真實原點之差。

II. 積差法

A. 根據未分組之資料計算：

1. 將 X, Y 相配之各項平行排列之，而後計算 $\Sigma(X), \Sigma(Y), \Sigma(X^2), \Sigma(Y^2), \Sigma(XY)$ 之數值。
2. 以 N 除(1)項中各數值。其首兩項 $\Sigma(X), \Sigma(Y)$ 被除後所得之商數，即等於 c_x, c_y ，因

$$\frac{\Sigma(X)}{N} = c_x \qquad \frac{\Sigma(Y)}{N} = c_y$$

故也。

3. 由下式計算均積數 p 。

$$p = \frac{\Sigma(XY)}{N} - c_x c_y$$

4. 由下列兩式，計算 X 及 Y 之標準差。

$$\sigma_x = \sqrt{\frac{\Sigma(X^2)}{N} - c_x^2}$$

$$\sigma_y = \sqrt{\frac{\Sigma(Y^2)}{N} - c_y^2}$$

5. 更由下式計算相關係數 r 。

$$r = \frac{p}{\sigma_x \sigma_y}$$

6. 將相當數值代入下兩式，以決定迴歸方程式。

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

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

(註：此兩式之原點均在平均數點)

7. 將 X, Y 之平均數代入下兩式，而使原點移至原來尺度之零點上。

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

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

8. 由下兩式計算標準誤 S_y 與 S_x 。

$$S_y = \sigma_y \sqrt{1 - r^2}$$

$$S_x = \sigma_x \sqrt{1 - r^2}$$

B. 根據分組之資料計算：

1. 編製一相關表，如I.B.中所述。
2. 兩變量各選一假定平均數，用組距為單位，以計算 X, Y 各項對於假定平均數之離中差。
3. 用組距單位計算 c_x, c_y 之數值。
4. 用組距單位計算 σ_x, σ_y 之數值。
5. 用組距單位計算相關表每小格中 $\Sigma(x'y')$ 之數值，並將此項數值一一相加，而求全表內 $\Sigma(x'y')$ 之總值。
6. 用組距單位由下式計算均積數 p 之數值。

$$p = \frac{\Sigma(x'y')}{N} - c_x c_y$$

7. 由下式求 r 之數值。

$$r = \frac{\phi}{\sigma_x \sigma_y}$$

8. 將 σ_x, σ_y 化爲原來單位。
9. 將相當數值代入下兩式，以決定兩迴歸方程式。

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

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

10. 由下列兩式，可將原點移至原來尺度上之零點。

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

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

11. 由下兩式計算標準誤 S_x 與 S_y 之數值。

$$S_y = \sigma_y \sqrt{1 - r^2}$$

$$S_x = \sigma_x \sqrt{1 - r^2}$$

應用前述方法求得各種量數後，更須繪製散佈圖，並在圖上加繪迴歸綫兩條。於此吾人對於兩變量間之關係以及所用計算方法之是否適當，皆可由圖而得確切之概念。

相關量數應用上之限制

上述各種相關量數(measures of relationship)在實際應用方面是否受有何種限制，換言之，即此種量數能否普遍適用於任何次數分配而無弊病，抑僅能適用於某種次數分配，在此不能不加以討論。

吾人前嘗論及次數分配合乎常態法則(normal law)時，標準差 σ 始有確切之意義。此時如已求得算術平均數及標準差之數值，即可推知在橫尺度上某距離內所包含之項數應占全體項數之百分率。如次數分配

與常態法則不甚相合時，標準差之用途仍存在，惟解釋時，其意義並不如前之確切耳。此點既已闡明，乃可進而解釋下式：

$$r^2 = 1 - \frac{S_y^2}{\sigma_y^2}$$

如倚變量各項在其平均數左右之分配，以及在其最小平方綫兩旁之分配，皆合乎常態法則時，則 S_y 與 σ_y 始有確切之意義，而由此兩種數值相互關係之比較所得之 r ，亦含有確切之意義。反之，該兩種分配之中，倘有一種不合乎常態法則，則此兩種數值之比較為不準確，而 r 之意義亦必因之減損。但即次數分配不合常態法則相當顯著時，標準差之功用仍可存在，惟其意義不及分配合乎常態時之確切。在同一情形之下，如次數分配即使不完全合乎常態之條件，吾人仍可計算標準誤及相關係數而利用之，惟解釋 S_y 及 r 之意義時，須先說明分配之狀態耳。吾人須知 S_y 及 r 之意義，僅在倚變量各項在其平均數左右，以及在其最小平方綫兩旁之分配為常態或近乎常態時，始為確切完全。

下述簡單之例，用以說明次數分配與常態法則極不相合時相關係數所受之影響。下表所列為一九一九年紐約州工業調查中所得一部分之資料。

表 一 〇 二

1919年紐約州十一城市工廠工人數及產品之價值

城 市	工人數 (千人) (X)	產品總值 (百萬美金元) (Y)
Batavia.....	2.2	9
Beacon.....	2.2	10
Corning.....	3.5	11
Geneva.....	2.5	10
Glens Falls.....	2.8	12
Ithaca.....	1.7	10
Middletown.....	2.2	10
Peekskill.....	2.1	11
Rensselaer.....	1.4	10
Tonawanda.....	1.8	16
New York City.....	638.8	5,261

如僅就表內所列前十城之資料計算，則得

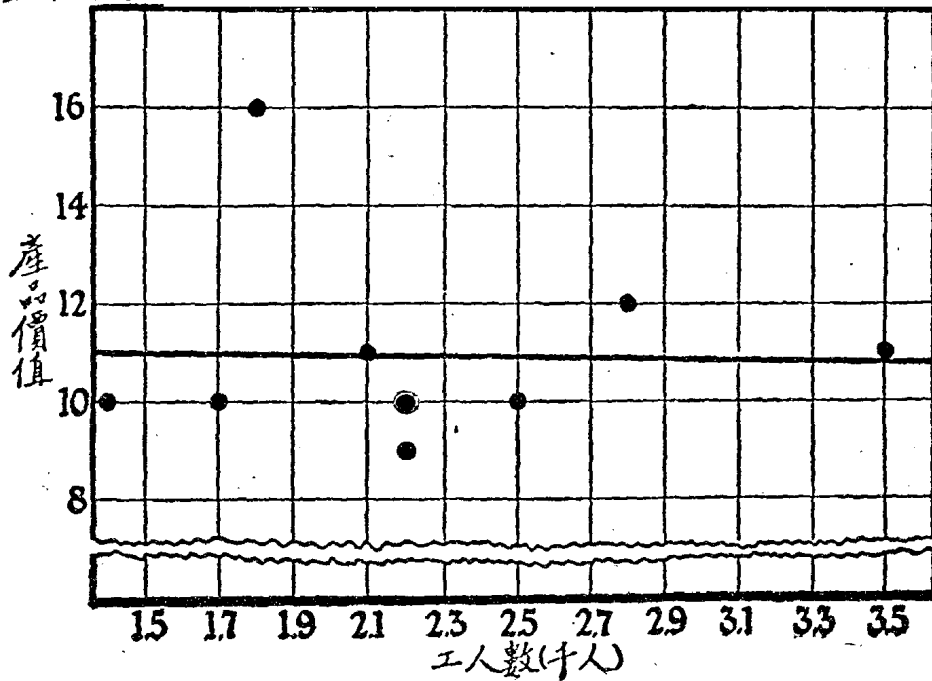
$$\sigma_y = 1.8682$$

$$S_y = 1.8669$$

$$r = -.034$$

此十點及其迴歸綫俱繪於圖七十二中（上所求得之相關係數不應視為具有何種意義，因選取此十城之目的，乃在舉示其一特殊之點也）。

美金百萬元



圖七十二 紐約州十城市工廠工人數與產品價值之關係

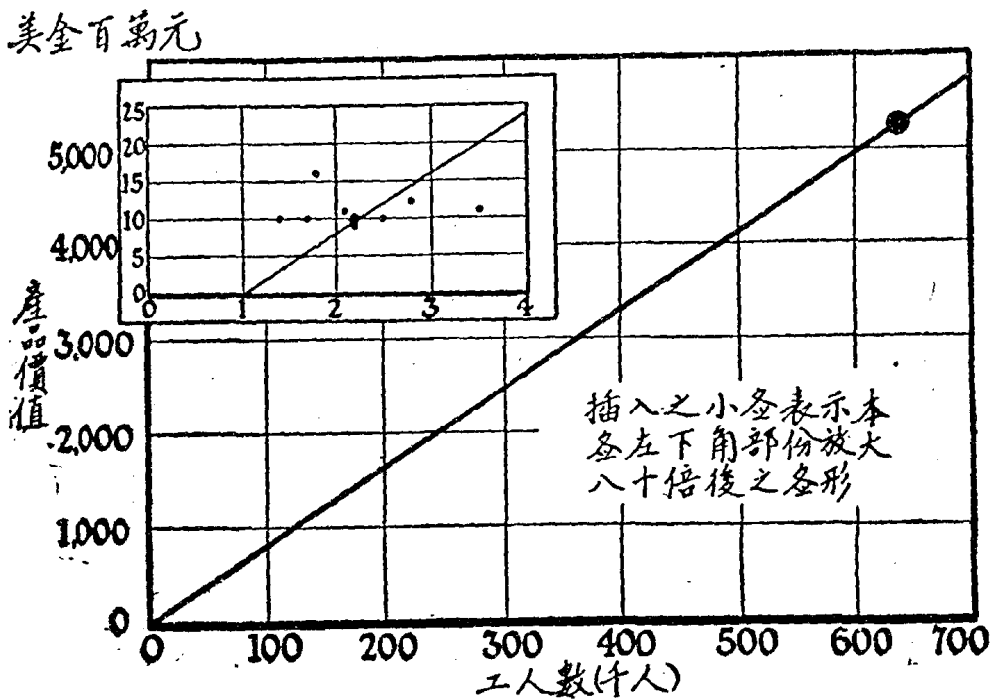
如更將紐約市併入計算，則由十一城市所得之各數值為

$$\sigma_y = 1509.3$$

$$S_y = 7.53$$

$$r = +.999988$$

此十一點及其迴歸綫見圖七十三。



圖七十三. 紐約州十一城市工廠工人數與產品價值之關係

上述各數值與根據十城市之資料所得者相差極大，其原因固極明顯，蓋當一極大城市與十小城市合併計算時， X 與 Y 之標準差數值俱因此增加甚鉅， Y （產品價值）之標準差遂由1.8682增至1509.3，而用以測度迴歸綫兩旁散佈狀態之 S_y 之數值雖亦同見增加，但遠不如 σ_y 之甚。根據十城計算 S_y 之數值為1.8669，根據十一城計算為7.53，此因採用最小平方法求直綫時，直綫之位置為一極端項目（紐約市之數值）所左右，而必須通過代表該項目之一點，或與該點極為接近，結果 S 之數值所受該極端項目之影響遂較 σ 為小。 r 之數值既由下式

$$r = \sqrt{1 - \frac{S_y^2}{\sigma_y^2}}$$

內 S_y 與 σ_y 之關係而得，故加入一極端項目，即有增加其數值之趨向。在上例中，一極端項目之加入，幾將 r 之數值由零而變為1，其所得結果自無意義。

上節所舉雖為極端之例，亦可證明次數分配不合常態時，所得各種

相關量數之意義必多牽強附會之處。雖然，在實際工作方面，各種相關量數之應用，自不能僅限於完全合乎常態之次數分配，但當吾人察覺次數分配具有足以影響相關量數之某種特質時，則解釋此種相關量數，不可不加以注意焉。

參 考 書

- BOWLEY, ARTHUR L. *Elements of Statistics* (350—397).
- BRUNT, DAVID. *The Combination of Observations* (148—170).
- CHADDOCK, R. E. *Principles and Methods of Statistics* (Chap. XII).
- ELDERTON, W. P. *Frequency Curves and Correlation* (106—124).
- GALTON, FRANCIS, *Correlations and their Measurement*. Proceedings of the Royal Society. Vol. XLV, 1888 (136—145).
- JONES, D. C. *A First Course in Statistics* (102—131).
- KELLEY, TRUMAN L. *Statistical Method* (151—195).
- MOORE, H. L. *Forecasting the Yield and the Price of Cotton* (12—51).
- PEARL, RAYMOND. *Medical Biometry and Statistics* (292—318).
- PEARSON, KARL. *Regression, Heredity and Panmixia*. Phil. Transactions, Royal Society. Series A. Vol. CLXXXVII, 1896(253—318).
- RIETZ, H. L. AND CRATHORNE, A. R. *Simple Correlation* (In Rietz, H. L. *Handbook of Mathematical Statistics*, 120—129).
- RUGG, H. O. *Statistical Methods Applied to Education* (233—307).
- WHITAKER, E. T. AND ROBINSON, G. *The Calculus of Observations* (317—336).
- YULE, G. U. *An Introduction to the Theory of Statistics* (151—209).

第十一章 時間數列之相關

前章所論測度相關之方法係適用於次數數列之分析，而非應用於時間數列者。測度時間數列之相關，含有性質特殊之問題，故應另文討論之。

吾人已知時間數列受有各種勢力之影響，此種勢力曾分爲四種，即長期趨勢、季節變動、商業循環及意外變動是也。關於各種勢力隔離之方法，上文亦已討論及之。測度時間數列之相關時，吾人亦須就數列所受之各種勢力分別研究，否則所得結論每不準確而易於發生誤解。相關之問題在於求得足以表示變量間相關程度之精確量數，但每一時間數列包含數個變量，故在可能範圍內，須將各個變量之相關分別研究之。

兩時間數列，例如利率與債券價格間之關係，可按照下列項目分別研究之。

- a. 長期趨勢
- b. 循環變動
- c. 季節變動
- d. 由一時間單位至次一時間單位（如週與週間、月與月間、年與年間）中所發生之變動

此種關係可先用圖比較，由圖中之比較，吾人常可得一具體印象。兩數列長期趨勢之相似與否及其循環變動之有無關聯，常可由圖決定之。如欲得更精確之比較，則須應用相關係數，惟在應用時，對其性質及所得結果之意義，不可不明瞭耳。

長期趨勢之比較，不必應用相關係數，因兩數列之長期趨勢縱屬相同，兩者之間未必即有相倚之關係，由長期趨勢值求得之相關係數，亦必毫無意義。在比較長期趨勢時，原可不必應用相關係數，因尚有較爲簡易之方法也。

同理，計算兩時間數列之關係時（如兩數列均無顯明之長期趨勢，則爲例外），不可根據兩數列原始之絕對數值（original absolute values）。計算普通資料之 r 時，須先求各項對其算術平均數之離中差，次將兩數列中相配之離中差（paired deviations）一一相乘，然後相加而求其和。如各對相配之離中差其正負號相同時， r 之數值爲正；各不相同時， r 之數值爲負。但兩數列中如各有顯著向上或向下之長期趨勢，則用此法所求得之 r 即無甚意義。此可舉一例以明之：例如吾人欲測度一九〇〇年至一九二〇年汽車產量與火腿價格兩種數列之關係。此兩數列在此期內均有顯著向上之長期趨勢，故當計算數列之各年數字對其算術平均數之離中差時，其所得前數年相配之離中差均帶負號，後數年均帶正號，故由各相配離中差之乘積相加所得 r 之數值頗高，並爲正號，但此數值果有意義乎？此處 r 所測度者，應爲兩數列之長期趨勢間之相關，而事實上汽車產量與火腿價格固無真實之關係存在也。

測度兩數列季節變動之相關時，或可利用 r ，但效用如何亦屬疑問，且測度季節變動之關係，尙有較爲簡易之方法可用。

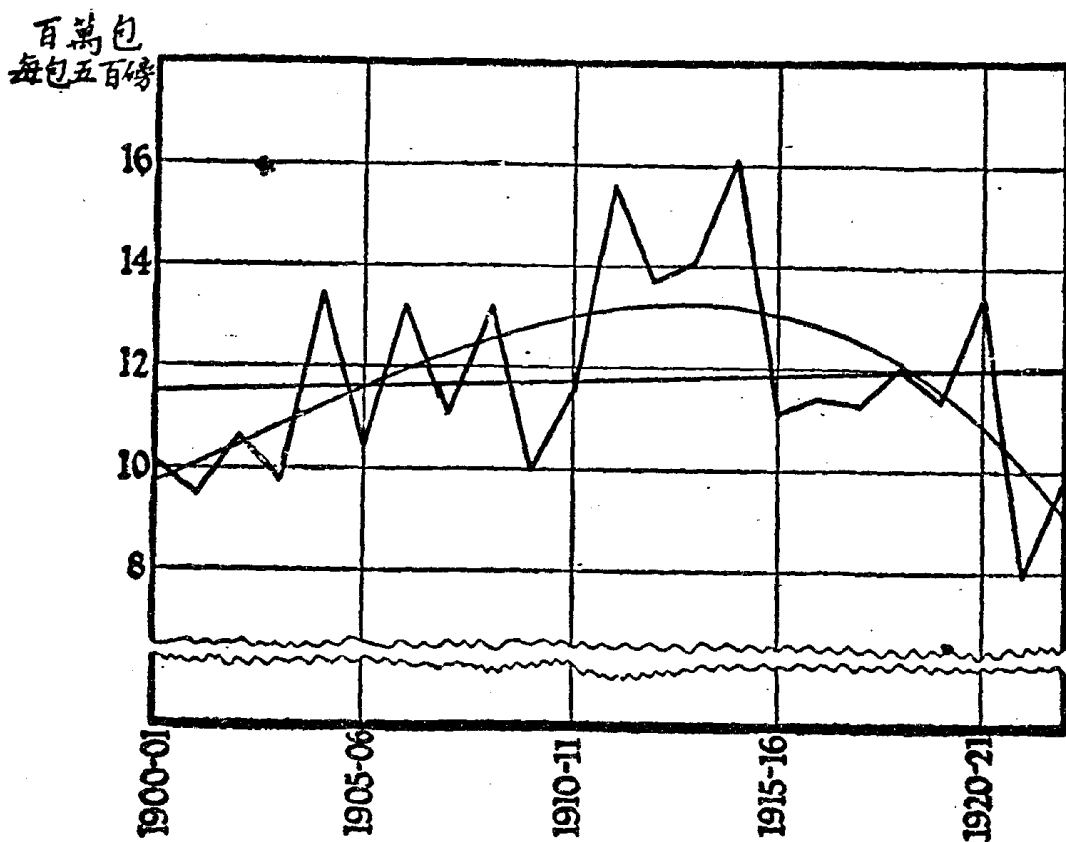
在實際應用方面，相關之方法，既不適用於測度長期趨勢之關係，亦不適用於測度季節變動之關係，其用途應限於比較兩數列之循環變動，以及月與月間或年與年間短期變動之關係。吾人比較此種關係時，須在可能範圍內，除去兩數列所受其他勢力之影響，故在計算相關以前，首須將原來資料予以校正，以消除循環變動與短期變動以外之各種勢力，否則對於相關係數之解釋必感困難矣。

循環變動相關之測度

在第七、第八兩章中，吾人曾討論測度及消除時間數列中長期趨勢及季節變動之方法。計算相關時，倘兩數列各項之差數，不根據算術平均數計算而根據長期趨勢值計算，即可避免因長期趨勢之存在而發生

之虛偽關係；且在測度循環變動之相關時，此種由各項與常態數值比較所得之差，亦為極重要之數值。至於銷除季節變動之問題，吾人倘用資料之年數字，即不復發生矣。

茲以棉花產量與棉花價格循環變動之資料為例，以說明時間數列相關之測度方法。此項資料之年份，以收穫年度 (crop year) 為準，自 1900—01 年度起至 1922—23 年度止。



圖七十四. 1900—01年度至1922—23年度美國棉花產量及其長期趨勢

在計算相關之前，棉花價格之數字尚須加以校正。棉花之原來價格為每年九月至次年五月之收穫年度內紐約‘高地米特令’ (Middling upland) 棉花現貨之平均批發價格。此項價格除因棉花市況之變遷而變動外，並受一般物價水準漲落之影響而隨之變動，故為銷除一般物價變動之因素起見，此種價格須先用物價指數校正之。此處所採用者為勃蘭

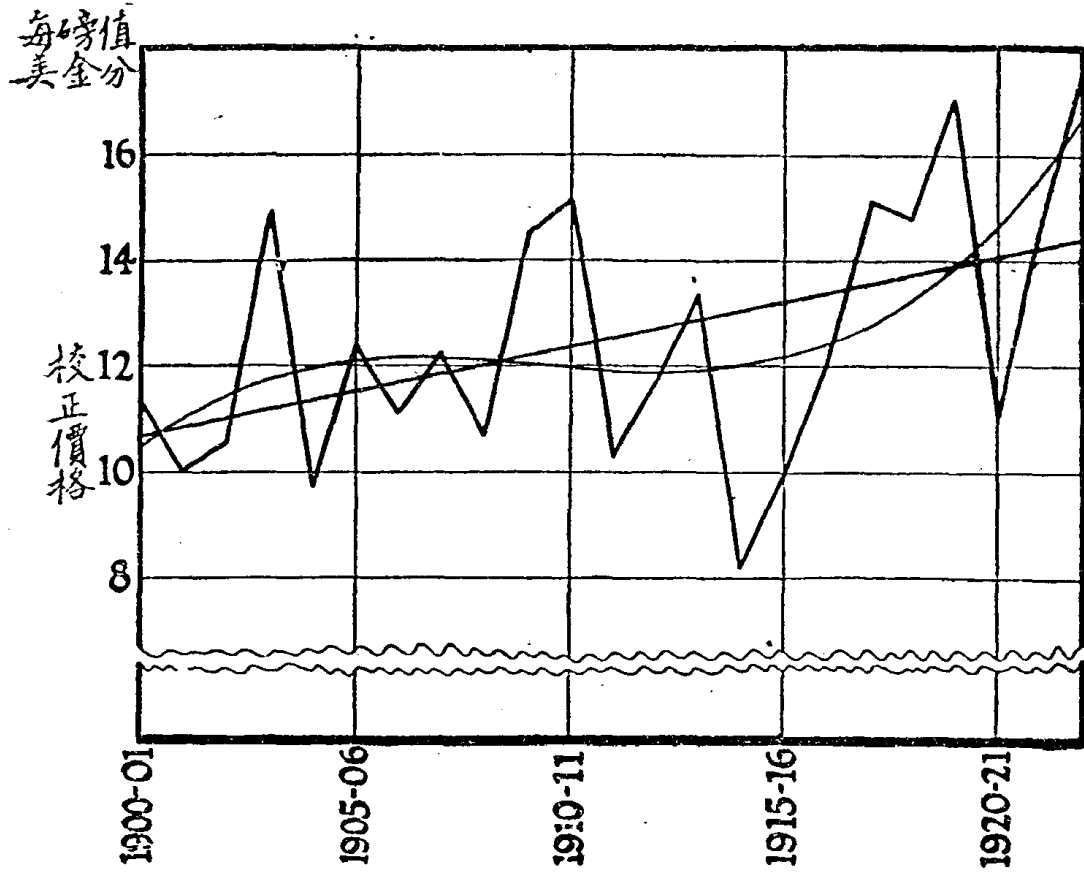
特爾斯脫指數 (Bradstreet price index) 每年九月至次年五月之平均, 並以一九一三年九月至一九一四年五月間之平均作為 100, 將其餘各年度之指數化為百分比之形式, 以便應用。茲將棉花產量及價格之原來數字以及校正後之價格列表於下:

表一〇三

1900—1923年間棉花之產量與價格

(1) 收穫年度	(2) 美國棉花產量 (麻絨不在內) (千萬袋)	(3) 棉花價格 紐約高地米特令 現貨九月至 五月之平均價 (每磅值美金分 之數)	(4) 勃爾特爾斯脫 物價指數 九月至五月之平均 (1913—14年度 = 100)	(5) 校正後之棉花價格 (每磅值美金 分之數)
1900—01	10,123	9.58	84.8	11.30
1901—02	9,510	8.64	86.2	10.02
1902—03	10,631	9.50	90.0	10.56
1903—04	9,851	13.20	88.6	14.90
1904—05	13,438	8.69	89.3	9.73
1905—06	10,575	11.40	92.3	12.35
1906—07	13,274	10.97	98.8	11.10
1907—08	11,107	11.41	93.2	12.24
1908—09	13,242	9.81	91.3	10.74
1909—10	10,005	14.62	100.6	14.53
1910—11	11,609	14.80	97.8	15.13
1911—12	15,693	10.34	100.0	10.34
1912—13	13,703	12.35	104.8	11.78
1913—14	14,156	13.40	100.0	13.40
1914—15	16,135	8.63	105.2	8.20
1915—16	11,192	12.04	121.2	9.93
1916—17	11,450	18.29	151.0	12.11
1917—18	11,302	29.96	197.9	15.14
1918—19	12,041	30.06	203.1	14.80
1919—20	11,421	38.63	226.3	17.07
1920—21	13,440	16.90	152.9	11.05
1921—22	7,954	18.67	127.2	14.68
1922—23	9,762	26.26	149.7	17.54

上表內棉花產量及其校正後之價格, 分繪於圖七十四及圖七十五內; 該兩數列之長期趨勢綫亦一併繪入。



圖七十五. 1900-01年度至1922-23年度紐約'高地米特令'

棉花價格及其長期趨勢綫

(收穫年度之平均價格爲用勃蘭特斯脫批發物價指數改正後之價格)

此時可計算兩數列各項對其長期趨勢綫之差，更進而計算兩組差數間之相關係數。如兩數列之長期趨勢均用三次拋物綫表示之，則其相關係數之計算如表一〇四所示(註)。

(註)兩數列所配合之長期趨勢綫，並不拘於同類之曲綫。

表一〇四
棉花產量與棉花價格相關係數之計算

(1) 年 度	(2) 棉花產量對其 長期趨勢拋物 綫之差 (千包) x	(3) 棉花價格對其 長期趨勢拋物 綫之差 (每磅值美金 分之數) y	(4) x^2	(5) y^2	(6) xy
1900-01	+371	+ .79	137,641	.6241	+293.09
1901-02	-603	-1.05	363,609	1.1025	+633.15
1902-03	+145	-.93	21,025	.8649	-134.85
1903-04	-1,012	+3.11	1,024,144	9.6721	-3,147.32
1904-05	+2,200	-2.26	4,840,000	5.1076	-4,972.00
1905-06	-1,027	+.24	1,054,729	.0576	-246.48
1906-07	+1,325	-1.06	1,755,625	1.1236	-1,404.50
1907-08	-1,164	+.08	1,354,896	.0064	-93.12
1908-09	+681	-1.38	463,761	1.9044	-939.78
1909-10	-2,806	+2.47	7,873,636	6.1009	-6,930.82
1910-11	-1,405	+3.14	1,974,025	9.8596	-4,411.70
1911-12	+2,530	-1.60	6,400,900	2.5600	-4,048.00
1912-13	+452	-.13	204,304	.0169	-58.76
1913-14	+887	+1.47	786,769	2.1609	+1,303.89
1914-15	+2,923	-3.81	8,543,929	14.5161	+11,136.63
1915-16	-1,878	-2.24	3,526,884	5.0176	+4,206.72
1916-17	-1,388	-.30	1,926,544	.0900	+416.40
1917-18	-1,206	+2.37	1,454,436	5.6169	-2,858.22
1918-19	-31	+1.56	961	2.4336	-48.36
1919-20	-102	+3.21	10,404	10.3041	-327.42
1920-21	+2,586	-3.59	6,687,396	12.8881	-9,283.74
1921-22	-2,104	-.91	4,426,816	.8281	+1,914.64
1922-23	+636	+.82	404,496	.6724	+521.52
總 計	0	0	55,236,930	93.5284	-40,752.29

$$\sigma_x = \sqrt{\frac{55,236,930}{23}} = 1,550$$

$$\sigma_y = \sqrt{\frac{93.5284}{23}} = 2.017$$

$$r = \frac{\sum(xy)}{N\sigma_x\sigma_y} = \frac{-40,752.29}{23 \times 1,550 \times 2.017}$$

$$= -.567$$

由表求得 r 之數值為 $-.567$ 。此值表示美國棉花產量對其長期趨勢綫之差與紐約棉花價格對其長期趨勢綫之差之兩者間在此期內確有相當之負相關存在。倘將該期內所包含之戰事年份剔除計算，則 r 之數值自必較高，惟剔除此種項目亦有妨礙耳。

由上表所得之各項數值，可求得兩數列相關綫之方程式。吾人已知迴歸方程式之形式為：

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

上式中之 y 及 x 俱代表各項對其長期趨勢拋物綫之差，將 r, σ_x, σ_y 之數值代入，可得：

$$y = - .567 \frac{2.017}{1,550} x$$

$$y = - .0007378x$$

此式之意義可解釋如下：平均而論，當棉花產量之差(x)高於其長期趨勢綫一單位時，則棉花價格之差(y)低於其長期趨勢綫.0007單位。此處產量之單位為千包，價格之單位為美金分，倘將 x 之單位變為百萬包，則方程式之解釋可較易，上式亦即變為：

$$y = - .7378x$$

由上式解釋之，棉花之產量較常態數值增百萬包時，棉花價格約比常態價格低四分之三美金分（此處均指校正價格而言），此即一九〇〇年至一九二三年間產量與價格之平均關係也；但因 r 之數值僅 $-.567$ ，故上述平均關係未見完全準確，倘 r 之數值為 -1 ，則此關係可以完全矣。

測度迴歸綫兩旁散佈程度之 S 數值，可用下式計算：

$$S_y = \sigma_y \sqrt{1 - r^2}$$

此處 S_y 為1.66美金分。 S_y 之意義，前已討論及之。

（吾人能否應用上式以預測未來價格，須視兩長期趨勢綫之延長是否合理而定）。

以上分析，在計算各項對其長期趨勢綫之差數時，俱用絕對單位，故由公式所得之結果亦以絕對單位為準，如棉花產量之用包，價格之用分是也，惟此種差數亦有用百分數表示者。在用百分數時，則 σ_x 及 σ_y

兩標準差及迴歸方程式亦必以百分數為單位。至於單位之選擇，則須視用途而定。

時間數列相關之測度，實際尚含有一不易確定之因素，而為前章討論次數數列之相關問題中所不存在者。在測度次數數列之關係時， σ_x ， σ_y 之計算，係由 x, y 兩數列之各項對其算術平均數之差求得；在測度時間數列之關係時，則係由兩數列之各項對其長期趨勢綫之差求得，而此長期趨勢綫為一不定因素，倘配合另一種形式不同之趨勢綫，則所求得 r 之數值亦必隨之而異。上例中所配合之長期趨勢綫俱為三次拋物綫，但吾人亦可配合兩直綫或其他曲綫以表示產量及價格之長期趨勢。倘該兩數列各配一直綫而計算各項對此直綫之差，以求相關係數，則 r 之數值為 -0.608 ， S_y 為1.72美金分（此項數值俱較前此用拋物綫所求得者為大，驟視之似有矛盾之處，實則 σ_y 之數值亦較前為大，而 r 之數值，乃由 S_y 與 σ_y 之關係而定者也）。

測度時間數列相關之困難

然則前所求得相關係數之兩數值，其所表示兩數列循環變動之關係，究以何者較為確切乎？吾人對此問題殊不易獲一肯定之答案，此為研究時間數列之相關時最困難之一點，蓋吾人如能斷定該兩種曲綫所表示各數列之長期趨勢孰為確切，則此問題本可迎刃而解，但對於長期趨勢綫之配合是否確當，亦無法測定耳。由圖上觀之，若僅就現有之棉花產量與價格之資料而論，則三次拋物綫所表示兩數列之長期趨勢似皆較直綫為確切，但又無法證明之。若以求得相關係數數值之大小以定長期趨勢綫之是否確當，則又不可，因相關係數數值之大小，往往由於兩長期趨勢綫發生同方向之偏誤所致（註）。

（註）參看一九二三年六月出版“Journal of the American Statistical Association”

第725頁 W. M. Persons 氏所著之 Correlation of Time Series 一文。

在計算兩數列各項對其長期趨勢綫之差，以求其循環變動之相關時，因此種不定因素之存在，足使吾人對於時間數列相關係數之可靠性發生懷疑，因相關係數數值之大小常隨所選長期趨勢綫之形式而異，故此處基本問題不在於相關係數之計算手續方面，而在於選取適當之長期趨勢綫。倘就吾人對於外界經歷之種種事實觀察之所得，俱覺以用某種長期趨勢綫表示各該數列之長期趨勢最為確切時，則由此綫所得之相關係數當可認為滿意，惟解釋及應用該項相關量數時，對於解決此種不定因素之主觀觀念應加以說明耳。如吾人研究之目的在於確定兩數列循環變動之函數關係，或根據此項函數關係以尋求一估量方程式(或迴歸方程式)，則對此點尤應注意。

相關係數及時間次序之測度

計算棉花產量與棉花價格相關之目的，在於測度棉花價格所受棉花產量增減之影響。吾人曾用方程式以表示兩數列各項對其長期趨勢綫之差數間之關係；在式中吾人又曾以棉花之價格作為其產量之函數，更進而斷定此項函數之關係。兩種數列之循環變動如確有函數關係之存在，則可藉此函數關係，由一數列之循環變動以斷定另一數列循環變動之程度，此原為研究時間數列相關時之主要問題也。

但在此處又發生一次要而性質不同之問題，例如兩數列均含有顯著之循環變動，惟其循環變動起伏之時期未必吻合，而有先後之別，乃有斷定時間先後之距離問題之發生，此“引前”(lead)或“落後”(lag)之問題又可藉相關係數以解答之。通常相關係數係用以測度函數之關係(functional relationship)，此時之相關係數則又可用以斷定時間先後之關係(temporal relationship)。

股票價格循環變動與一般商情循環變動之關係

茲以實業股票價格之循環變動與一般商情之循環變動為例，說明其時期先後關係之斷定方法。一般商情循環變動之資料係採用美國電話電報公司所編之組合指數，其自一八七七年至一九二三年間之每月數字在表九十三內已有刊載。茲將股票價格循環變動之數字列表於下：

表一〇五

1903—1923年實業股票價格之循環變動(註)

(下列數字為各月價格對常態數值之差，以標準差為單位)

	1903	1904	1905	1906	1907	1908	1909	1910	1911	1912	1913
一月	-.2	-2.0	-.1	+2.3	+1.6	-1.3	+.5	+1.0	-.1	-.4	-.2
二月	-.1	-2.1	+.2	+2.2	+1.4	-1.5	+.3	+.5	+.1	-.4	-.5
三月	-.3	-2.1	+.6	+1.9	+.6	-1.1	+.3	+.8	-.1	-.1	-.7
四月	-.5	-2.0	+.7	+1.7	+.6	-.8	+.5	+.5	-.1	+.3	-.6
五月	-.5	-2.1	+.2	+1.4	+.4	-.5	+.8	+.4	0	+.2	-.7
六月	-.9	-2.1	+.3	+1.5	+.2	-.5	+.9	+.1	+.1	+.2	-1.1
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	-.3	-.7	+.4	-.5
十月	-2.3	-.9	+.8	+1.7	-1.3	+.2	+1.4	0	-.7	+.3	-.8
十一月	-2.4	-.3	+1.1	+1.7	-1.9	+.5	+1.4	+.1	-.5	+.2	-.9
十二月	-2.1	-.1	+1.8	+1.7	-1.6	+.5	+1.3	-.2	-.4	0	-.9
	1914	1915	1916	1917	1918	1919	1920	1921	1922	1923	
一月	-.7	-2.42	+.42	+.61	-1.00	-.56	+1.16	-1.04	-.61	+.66	
二月	-.6	-2.47	+.23	+.15	-.65	-.53	+.40	-1.01	-.37	+.95	
三月	-.6	-2.30	+.31	+.41	-.80	-.13	+.78	-1.02	-.13	+1.14	
四月	-.8	-1.09	+.05	+.35	-.86	+.14	+.88	-.92	+.19	+.93	
五月	-.8	-1.72	+.08	+.32	-.61	+.77	+.17	-.85	+.34	+.53	
六月	-.7	-1.54	+.11	+.59	-.63	+1.22	+.17	-1.49	+.32	+.35	
七月	-1.1	-1.25	-.04	+.29	-.59	+1.56	+.11	-1.54	+.45	+.01	
八月	*	-.74	+.15	-.03	-.54	+1.01	-.27	-1.66	+.69	+.13	
九月	*	-.27	+.62	-.41	-.50	+1.38	-.11	-1.41	+.81	+.09	
十月	*	+.25	+.97	-.75	-.22	+1.86	-.31	-1.32	+.84	-.11	
十一月	*	+.40	+1.40	-1.32	-.35	+1.62	-.80	-.97	+.51	+.14	
十二月		-2.54	+.58	+.70	-1.41	-.50	+1.24	-1.28	+.64	+.37	

*證券交易所停市。

(註)此項數字為 W. M. Persons 氏分析之結果，錄自哈佛經濟研究委員會出版之“Review of Economic Statistics”。1913年一月至1914年七月之數字係根據十二種實業股票之平均價格計算；自1914年十二月以後之數字則係根據二十種實業股票之平均價格 (Dow-Jones 氏所編之指數) 計算。

兩數列之資料俱繪於圖七十六內。由圖中兩綫比較觀之，可知此兩種數列之循環變動確有關係存在，惟此種比較難得確切之結論。吾人此時之目的，在於斷定此兩種循環變動時期之先後是否吻合，如不吻合，則更須斷定一種循環變動比另一種循環變動平均引前或落後之時間。此種研究對於分析商情循環之重要，自可不言而喻。

現時所用以研究此問題者，為一九〇三年一月至一九一四年六月間之資料。歐戰期間之各年，因兩數列皆受特殊影響，故在此棄而不用。

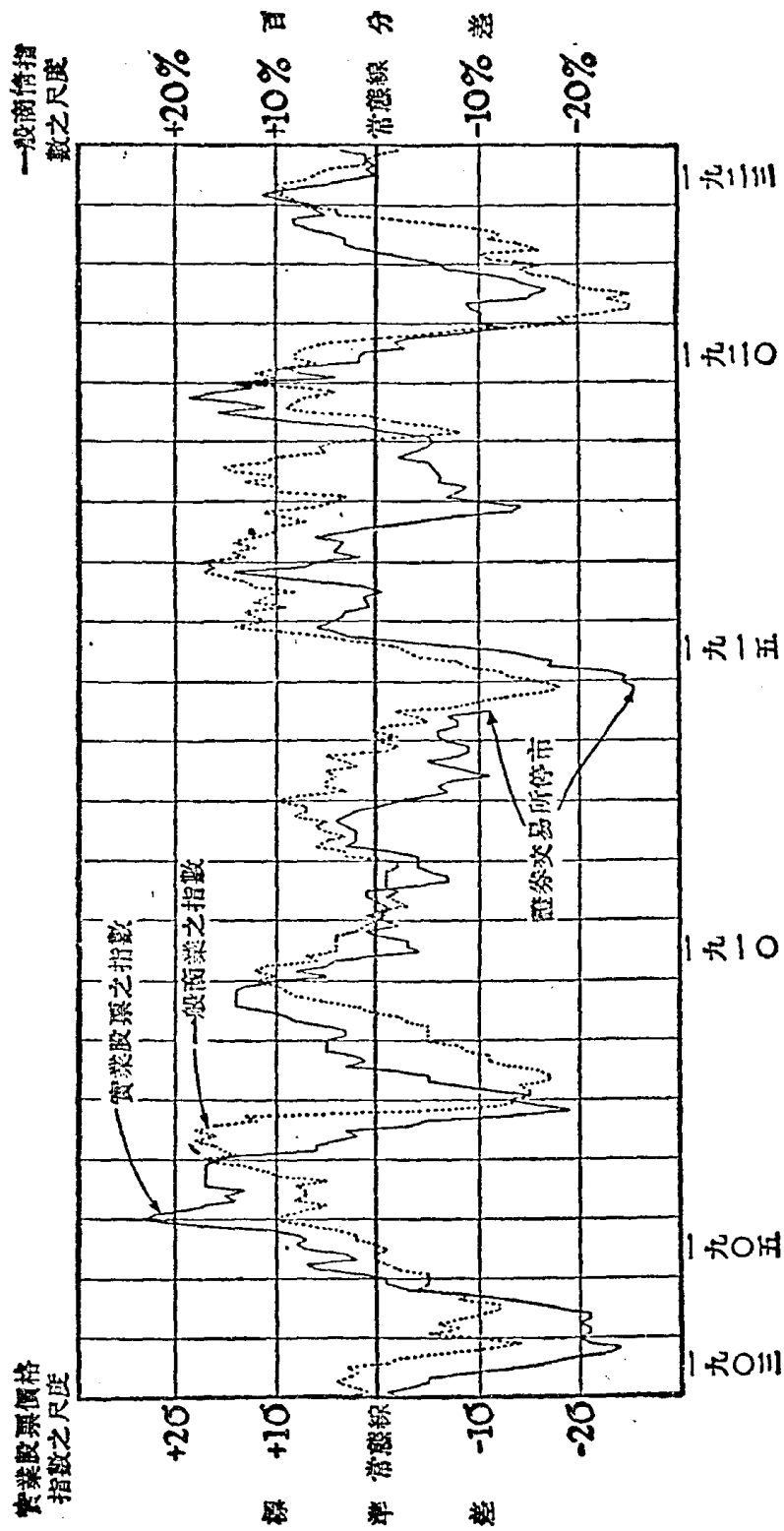
先將同時期之項目相配，以計算其相關係數，此數值為 $+ .55$ ；次將實業股票價格之前一月數字與一般商情指數之本月數字一一相配，以計算其相關係數，換言之，即以一九〇三年一月之股票數字與一九〇三年二月之商情數字相乘；以二月之股票數字與三月之商情數字相乘，依此類推，而分別計算相關係數中之 xy 也。自一九〇三年一月起至一九一四年六月止皆依此計算，惟此處祇有 137 個月之數值，而前此計算時則有 138 個月之數值，此因一九〇三年一月之商情數字與一九一四年六月之股票數字因無相配數字，不能列入計算所致。 c_x 及 c_y （因原點不在兩平均數點，故須應用此兩校正數）及兩標準差之數值亦均與前者稍異。由此求出之相關係數為 $+ .65$ 。如將兩數列各月之數值更用他種相配方法，以計算相關係數，其結果當如下表：

表 一 〇 六

實業股票價格與一般商情指數之相關係數

（根據1903年至1914年間之資料）

	相關係數
本月商情指數與同一月股票價格相配	$+ .55$
本月商情指數與前一月股票價格相配	$+ .65$
本月商情指數與前二月股票價格相配	$+ .70$
本月商情指數與前三月股票價格相配	$+ .73$
本月商情指數與前四月股票價格相配	$+ .76$
本月商情指數與前五月股票價格相配	$+ .76$

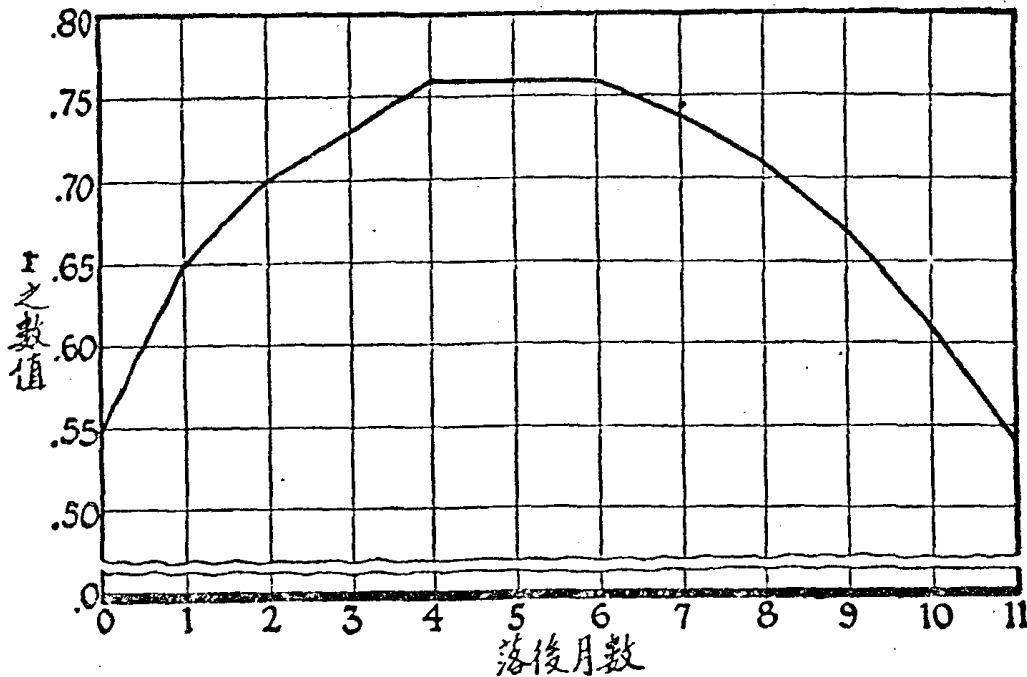


圖七十六. 1903—1923年實業股票價格變動與一般商情循環變動之比較

表 一 〇 六 (續)

本月商情指數與前六月股票價格相配	+ .76
本月商情指數與前七月股票價格相配	+ .74
本月商情指數與前八月股票價格相配	+ .71
本月商情指數與前九月股票價格相配	+ .67
本月商情指數與前十月股票價格相配	+ .61
本月商情指數與前十一月股票價格相配	+ .54

上表內各項係數均繪於圖七十七內。

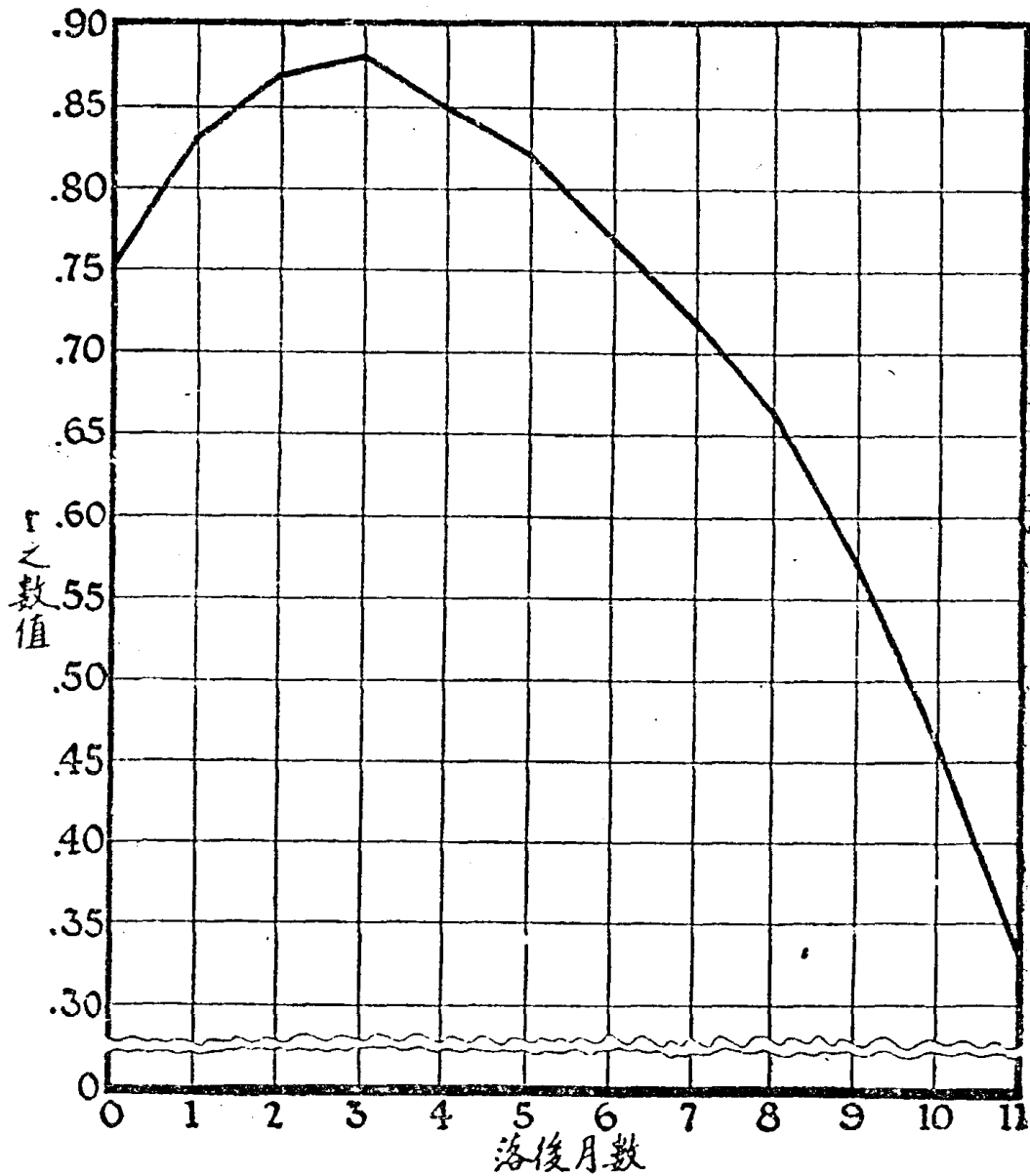


圖七十七. 1903—1914年實業股票價格指數與一般商情指數依各種月份相配法求得之相關係數

(在所有相配各對中,除同月份外,商情指數之變動均較股票價格指數為落後)

當商情指數與前四月、前五月、前六月之股票價格相配時,所得相關係數之數值最大,各為 +.76; 股票價格引前三月至引前七月所得之各個相關係數大致相同,故在此兩數列之循環變動間不易確定一時間先後之關係,惟就相關係數之數值觀之,股票價格之變動比一般商情指數引前之平均時間大約為五個月,但此亦絕非一確定不變之關係。

如根據歐戰以後即一九一九年一月至一九二三年十二月之資料，計算此兩數列之關係，則所得結果與前者稍異。茲將該時期內之相關係數列於表一〇七，並繪入圖七十八內。



圖七十八. 1919—1923年實業股票價格指數與一般商情指數依各種月份相配法求得之相關係數

(在所有相配各對中，除同月份外，商情指數之變動均較股票價格指數為落後)

此項資料所包含之時期過短，難得一肯定之結論。由表內相關係數

之數值觀之，歐戰以後一般商情與股票價格變動之關係，似較戰前更為密切，而商情變動追隨股票價格之變動亦較接近。當商情指數與前三月之股票價格相配時，所得相關係數之數值為最大。

表 一 〇 七

實業股票價格與一般商情指數之相關係數

(根據1919至1923年之資料)

	相關係數
本月商情指數與同一月股票價格相配	+ .75
本月商情指數與前一月股票價格相配	+ .83
本月商情指數與前二月股票價格相配	+ .87
本月商情指數與前三月股票價格相配	+ .88
本月商情指數與前四月股票價格相配	+ .85
本月商情指數與前五月股票價格相配	+ .82
本月商情指數與前六月股票價格相配	+ .77
本月商情指數與前七月股票價格相配	+ .72
本月商情指數與前八月股票價格相配	+ .66
本月商情指數與前九月股票價格相配	+ .57
本月商情指數與前十月股票價格相配	+ .45
本月商情指數與前十一月股票價格相配	+ .33

移動平均數之應用於計算時間數列循環變動之相關

以上討論循環變動之相關時，所有兩數列各項之差數，俱係根據數學方法所配合之長期趨勢綫計算而得。但長期趨勢綫原可應用移動平均數法測定，則表示循環變動之差數亦可根據此綫計算，而後依照上法以求相關係數；惟根據此綫求得之差數，其算術平均數並不等於零，而根據最小平方法所配之曲綫計算者，其算術平均數則等於零，故如用移動平均數法求長期趨勢綫，更由此綫計算各項差數以求相關係數時，對於此點應予校正。

用移動平均數法求長期趨勢綫，未見較優於用數學方法配合所得

之綫。根據前者計算所得之各項差數，亦非純係循環變動而無他種變動包含在內。且移動平均數之時期不同，其長期趨勢綫自必隨之而異，而決定時期之長短，亦無絕對標準之可言，是又加一不定之因素矣。雖然，應用移動平均數有時或可得有意義及有效用之相關係數，惟吾人須認明此項不定因素之存在，而在應用此法時亦須注意及之。

短期變動之相關

測度兩時間數列之相關，不可根據兩者之長期趨勢，或兩者之季節

表一〇八

1901年至1923年棉花產量與棉花價格相關係數之計算

(根據第一階差 first differences 計算)

(1) 收穫年度	(2) 各年產量與 前一年產量 之相差數 (單位百萬包) X	(3) 各年價格與前 一年價格之相 差數(校正價 格，每磅值美 金分之數) Y	(4) X ²	(5) Y ²	(6) XY
1901—02	- .613	- 1.28	.375769	1.6384	+ .78464
1902—03	+ 1.121	+ .54	1.256641	.2916	+ .60534
1903—04	- .780	+ 4.34	.608400	18.8356	- 3.38520
1904—05	+ 3.587	- 5.17	12.866569	26.7289	- 18.54479
1905—06	- 2.863	+ 2.62	8.196769	6.8644	- 7.50106
1906—07	+ 2.699	- 1.25	7.284601	1.5625	- 3.37375
1907—08	- 2.167	+ 1.14	4.695889	1.2996	- 2.47038
1908—09	+ 2.135	- 1.50	4.558225	2.2500	- 3.20250
1909—10	- 3.237	+ 3.79	10.478169	14.3641	- 12.26823
1910—11	+ 1.604	+ .60	2.572816	.3600	+ .96240
1911—12	+ 4.084	- 4.79	16.679056	22.9441	- 19.56236
1912—13	- 1.990	+ 1.44	3.960100	2.0736	- 2.86560
1913—14	+ .453	+ 1.62	.205209	2.6244	+ .73386
1914—15	+ 1.979	- 5.20	3.916441	27.0400	- 10.29080
1915—16	- 4.943	+ 1.73	24.433249	2.9929	- 8.55139
1916—17	+ .258	+ 2.18	.066564	4.7524	+ .56244
1917—18	- .148	+ 3.03	.021904	9.1809	- .44844
1918—19	+ .739	- .34	.546121	.1156	- .25126
1919—20	- .620	+ 2.27	.384400	5.1529	- 1.40740
1920—21	+ 2.019	- 6.02	4.076361	36.2404	- 12.15438
1921—22	- 5.486	+ 3.63	30.096196	13.1769	- 19.91418
1922—23	+ 1.808	+ 2.86	3.268864	8.1796	+ 5.17088
	- 22.847 + 22.486	- 25.55 + 31.79	140.548313	208.6688	- 117.37216
	- .361	+ 6.24			

表一〇八(續)

$$\begin{aligned}
 c_x &= \frac{-.361}{22} = -.0164 & \sigma_y &= \sqrt{9.40429} \\
 c_x^2 &= .00026896 & \sigma_y &= 3.067 \\
 c_y &= \frac{+6.24}{22} = +.284 & p &= \frac{\sum XY}{N} - c_x c_y \\
 c_y^2 &= .080656 & &= \frac{-117.37216}{22} - (.284 \times -.0164) \\
 \sigma_x &= \sqrt{\frac{140.548313}{22} - .00027} & &= -5.3304404 \\
 &= \sqrt{6.38829} & r &= \frac{p}{\sigma_x \sigma_y} \\
 \sigma_x &= 2.528 & r &= \frac{-5.3304404}{7.753376} \\
 \sigma_y &= \sqrt{\frac{208.6688}{22} - .08066} & &= -.687
 \end{aligned}$$

變動，以求相關係數，但數列中所受長期趨勢及季節變動之影響如能消除，則可用相關係數以測度循環變動之函數關係 (functional relations)，或循環變動時間之先後關係 (temporal relations)，前已言之矣。相關係數之應用於時間數列之研究，除測度上述兩種關係外，尚有其他用途。吾人可用此量數以測度兩數列短期變動之關係，如年與年間、月與月間、週與週間或日與日間之變動等是。此問題與前所討論者性質不同，故解釋其相關係數時不可不分別清楚。

比較短期變動之方法，先求每數列中各項與前一項之差數，然後將同時期兩數列之差數相配而求其相關係數，惟此差數可用絕對數值表示，亦可用百分數或比率表示。表一〇八所示係根據棉花產量與棉花價格逐年變動之絕對數值(第一階差)以求相關係數之計算手續。表中(2)(3)兩行內各項之原來數值見表一〇三。

r 之計算手續與前此所舉各例用假定原點計算者相同。此處所用假定原點即為零點，惟因各項之和不等於零，故所得各數值均須加以校正。由此求得 r 之數值為 $-.687$ 。

迴歸方程式及 S_y 之數值求得如下：

$$y = .8335x$$

$$S_y = 2.23 \text{ 美金分}$$

吾人如將前述關於棉花各例所得結果加以比較，則對於相關問題可更明瞭而感興趣。實際此項結果係分別根據三種不能嚴格比較之數字以計算相關而得，此三種數字者，乃棉花產量、棉花價格兩數列之各項與其三次拋物綫之差、與其直綫長期趨勢之差、以及各項逐年變動之絕對數值也。惟若吾人已知棉花產量而欲估計棉花價格，則可在此三種數字所求得之相關數值中任擇一種作為估量之根據。此三種相關之數值如下：

	r	S_y
棉花產量及價格循環變動之相關 (各項差數係用三次拋物綫求得者)	-.567	1.66美金分
棉花產量及價格循環變動之相關 (各項差數係用直綫長期趨勢綫求得者)	-.608	1.72美金分
棉花產量及價格逐年變動之相關	-.687	2.23美金分

r 之數值以第三種相關所得者為最大，其標準誤亦最大。此種表面上矛盾之事實及其原因，前已述及；吾人須注意者，逐年變動之標準差亦較其他兩種相關之標準差為大也。

吾人觀於三種相關之結果，必以為估量之差誤，以根據三次拋物綫求得者為最小，根據逐年變動求得者為最大，殊不知第一種相關實含有一內在之假定，即產量與價格之長期趨勢拋物綫均可適用於資料時期以外之假定是也。此種假定可發生極大之差誤，而差誤之機率又非其標準誤所能測度。直綫之長期趨勢綫亦有此弊，惟根據逐年變動測度相關時，則並無此種假定，但逐年變動之時期內如包含一非常年份，則由此發生之困難，又較其他兩種相關為大矣。

參 考 書

- MOORE, H. L. *Economic Cycles: Their Law and Cause.*
- MOORE, H. L. *Forecasting the Yield and the Price of Cotton.*
- MOORE, H. L. *Generating Economic Cycles.*
- PERSONS, W. M. *Correlation of Time Series.* Journal of the American Statistical Association, June, 1923. (This article is also published in Rietz, H. L., *Handbook of Mathematical Statistics* 150-165.)
- PERSONS, W. M. *Indices of Business Conditions.* Review of Economic Statistics, Prel. Vol. I. 1919.
- PERSONS, W. M. *The Variate Difference Correlation Method and Curve Fitting.* Quarterly Publications of the American Statistical Association, June, 1917.
- SNOW, E. C. *Trade Forecasting and Prices (with discussion).* Journal of the Royal Statistical Society, May, 1923 (332-398).
- STAMP, J. C. *The Effect of Trade Fluctuations Upon Profits (with discussion).* Journal of the Royal Statistical Society, July, 1918 (563-608).
- YULE, G. U. *On the Time Correlation Problem, with Especial Reference to the Variate Difference Correlation Method (with discussion).* Journal of the Royal Statistical Society, July, 1921 (497-527).

第十二章 關係之測度：非直綫相關

前兩章內所討論兩變量間之關係，皆可用直綫表示者。相關係數 r 為測度兩變量間直綫相關之程度，惟在 X 及 Y 數值相配所得各點之散佈確可用直綫代表時， r 始有明顯之意義。

在時間數列上配合長期趨勢綫時，吾人已知此綫未必定為直綫，而常須用一高次曲綫(a curve of higher degree) 表示之，研究相關時亦然。有時兩變量間之相關程度固甚高，但其關係或非直綫所能表示者，如強用直綫配合，則該綫兩旁各點之散佈區域必廣， r 之數值亦必減低，足以發生錯誤之印象；此時倘能配合一適當之曲綫，則該綫兩旁各點之散佈區域必可收狹，真實之關係於是可見。茲就下表內之數字以說明之。

表一〇九

香草(alfalfa)之收穫與灌溉之水量(註)

(1910年至1915年加利福尼亞州台維斯城調查之概略)

灌溉水量 之深度 (吋數)	每英畝收穫之噸數						平均
	1910	1911	1912	1913	1914	1915	
0	3.85	5.94	5.52	2.75	2.89	2.35	3.88
12	4.78	7.52	6.51	4.31	5.83	4.84	5.63
18	*	*	7.02	5.69	8.02	6.46	6.80
24	6.00	8.38	8.32	6.89	9.96	7.96	7.92
30	7.53	9.54	9.43	7.97	11.06	8.32	8.98
36	7.58	9.33	9.38	8.22	12.48	8.63	9.27
48	8.45	9.52	8.63	8.83	10.62	8.05	9.02
60	*	*	10.17	7.25	10.70	5.55	8.42

*無調查。

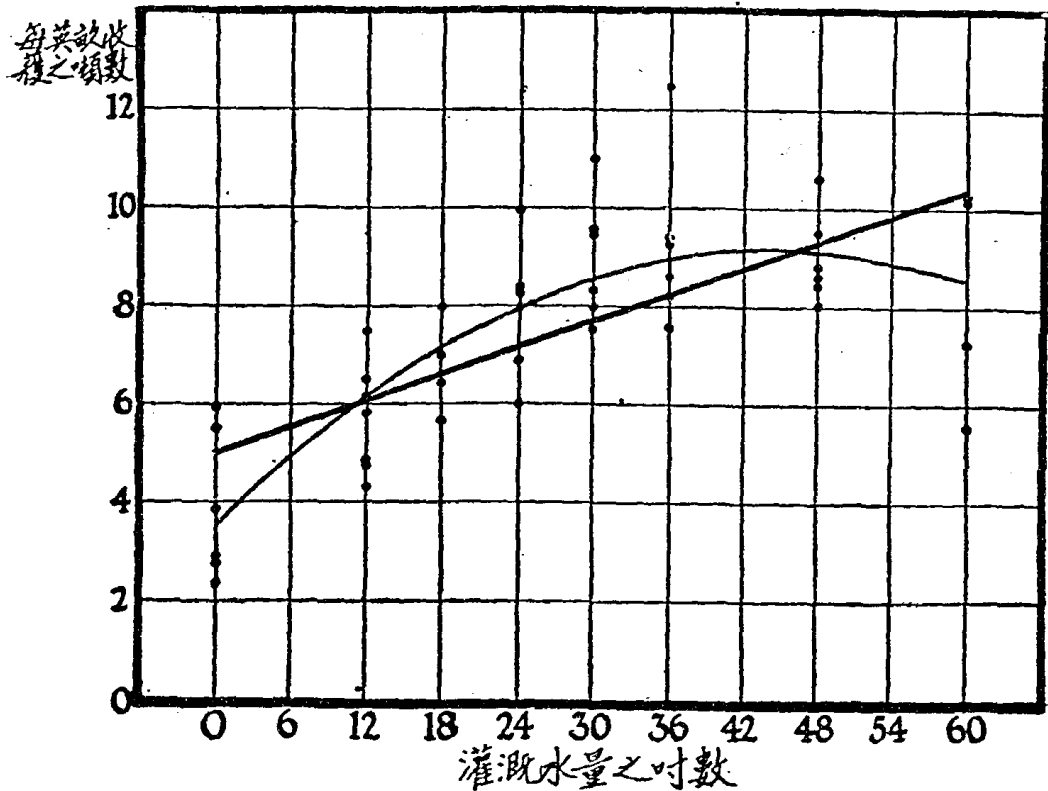
表內資料繪於圖七十九。

在圖中各點上，配合形式不同之兩綫。其一為直綫，方程式為

$$Y = 5.038 + .0886X$$

(註)此表錄自1917年五月加利福尼亞大學農業實驗所叢書第280種 S. H. Beckett 暨 R. D. Robertson 兩氏所著 "The Economical Irrigation of Alfalfa in the Sacramento Valley"。

式中之 Y 為每英畝收穫之噸數， X 為灌溉水量之吋數，此兩變量用直綫配合所得之相關程度可由 r 之數值表示之。 r 之數值在此為 $+ .68$ 。



圖七十九。表示香草 (alfalfa) 之收穫與灌溉水量兩者關係之散佈圖及其兩迴歸綫

觀上圖可知直綫並非一配合最適當之綫，故由此所得之 r 決不能準確測度 alfalfa 香草收穫量與灌溉水量之相關程度。

拋物綫相關

圖中另一綫為用最小平方法所配合之二次拋物綫，其方程式為

$$Y = 3.55 + .252X - .002816X^2$$

灌溉水量之增加對於 alfalfa 香草收穫量之關係，顯以用拋物綫所表示者較為準確，蓋水量漸增，收穫亦隨之增加，但有一定之限度，過此限度，水量增而收穫反減，此為研究此項相關中最重要之一點，而非直綫所能表示者也。

由是知此種關係應以拋物綫之方程式表示之。計算標準誤 S_y 時，可先求實際數值 Y 與計算數值 Y_c 之差；次用此項差數計算均方根差。其計算步驟如下表所示。表內常態收穫量(normal yield)係根據前項拋物綫方程式計算者。

表 一 一 〇

香草(alfalfa)之實際收穫量與常態收穫量之比較

(1) 灌溉水量 之深度	(2) 實際收穫量	(3) 常態收穫量 (根據拋物 綫方程式 計算者)	(4) 實際收穫量 與常態收穫 量之差 (2)-(3)	(5)
X	Y	Y_c	d	d^2
0	3.85	3.55	+ .30	.0900
0	5.94	3.55	+ 2.39	5.7121
0	5.52	3.55	+ 1.97	3.8809
0	2.75	3.55	- .80	.6400
0	2.89	3.55	- .66	.4356
0	2.35	3.55	- 1.20	1.4400
12	4.78	6.16	- 1.38	1.9044
12	7.52	6.16	+ 1.36	1.8496
12	6.51	6.16	+ .35	.1225
12	4.31	6.16	- 1.85	3.4225
12	5.83	6.16	- .33	.1089
12	4.84	6.16	- 1.32	1.7424
18	7.02	7.17	- .15	.0225
18	5.69	7.17	- 1.48	2.1904
18	8.02	7.17	+ .85	.7225
18	6.46	7.17	- .71	.5041
24	6.00	7.97	- 1.97	3.8809
24	8.38	7.97	+ .41	.1681
24	8.32	7.97	+ .35	.1225
24	6.89	7.97	- 1.08	1.1664
24	9.96	7.97	+ 1.99	3.9601
24	7.96	7.97	- .01	.0001
30	7.53	8.57	- 1.04	1.0816
30	9.54	8.57	+ .97	.9409

表 一 一 ○ (續)

(1)	(2)	(3)	(4)	(5)
30	9.43	8.57	+ .86	.7396
30	7.97	8.57	- .60	.3600
30	11.06	8.57	+2.49	6.2001
30	8.32	8.57	- .25	.0625
36	7.58	8.97	-1.39	1.9321
36	9.33	8.97	+ .36	.1296
36	9.38	8.97	+ .41	.1681
36	8.22	8.97	- .75	.5625
36	12.48	8.97	+3.51	12.3201
36	8.63	8.97	- .34	.1156
48	8.45	9.15	- .70	.4900
48	9.52	9.15	+ .37	.1369
48	8.63	9.15	- .52	.2704
48	8.83	9.15	- .32	.1024
48	10.62	9.15	+1.47	2.1609
48	8.05	9.15	-1.10	1.2100
60	10.17	8.53	+1.64	2.6896
60	7.25	8.53	-1.28	1.6384
60	10.70	8.53	+2.17	4.7089
60	5.55	8.53	-2.98	8.8804
			+24.22	80.9871
			-24.21	

將各項差數平方之和,代入下式

$$S_y = \sqrt{\frac{\sum d^2}{N}}$$

則得

$$S_y = \sqrt{\frac{80.9871}{44}} = 1.36$$

相關指數

吾人於此尚須用一測度關係疎密之抽象量數。在直綫相關中,測度相關程度之抽象量數即為由 S_y 及 σ_y 之關係所求得之相關係數 r ; 在非

直綫相關中，吾人亦可用同法以求得一同樣之抽象量數，此抽象量數謂之相關指數，用符號 ρ （讀如Rho）表示（註一），以取別於相關係數 r 。

計算相關指數之普遍公式（註二）為

$$\rho^2_{yx} = 1 - \frac{S_y^2}{\sigma_y^2}$$

S_y 之數值前已求出， σ_y 之數值用普通方法求得為 2.27，將此兩值代入 ρ 之公式，可得

$$\begin{aligned} \rho_{yx} &= \sqrt{1 - \frac{1.8406}{5.177}} \\ &= .80 \end{aligned}$$

此數值較用同資料求得之相關係數（ r 之數值為+.68）高出甚多，兩者之差乃因在此資料上配合拋物綫實遠較直綫為佳也。此處之相關顯為非直綫的，故不可用 r 以測度其相關程度。

相關指數之意義

相關指數 ρ 之意義及其數值之限度若何，亦為吾人所應瞭解者。 ρ 之數值係由各點在其所配曲綫兩旁之散佈程度與在 Y 數列之算術平均數兩旁之散佈程度比較而得。所配之綫如為直綫，則 ρ 與 r 相等， r 不過為 ρ 中之一種特殊形式。 ρ 之數值在0與1之間，如其為0，即謂兩變量間苟有

（註一）此符號 Spearman 氏曾用以代表等級相關中所求得之相關係數。在普通經濟分析中，此種用法無顯著之意義，故此處以代表相關指數，當不致發生誤解。

（註二）如 X 為倚變量，則此式應為

$$\rho^2_{xy} = 1 - \frac{S_x^2}{\sigma_x^2}$$

上式中 ρ 旁之兩字母，其第一字母常指倚變量，第二字母則指自變量。因 X 為倚變量時所得 ρ_{xy} 之數值未必與 Y 為倚變量時所得 ρ_{yx} 之數值相同，故 ρ 旁 x 及 y 兩字母之次序有區別之必要；至於直綫相關之係數，即不必如此區別，因不論以何者為倚變量， r 之數值常相等故也。

關係存在，此關係不可以所用方程式表示；如其為1，則謂此方程式所表示者為一完整之關係。 ρ 之數值之前不必附以正負符號，因高次曲綫所示兩變量之關係，一部份為正，另一部份為負，上述之 alfalfa 香草收穫量與灌溉水量，即其一例也。

計算相關指數 ρ 時，必須說明所用曲綫之形式，蓋離開曲綫，相關指數即無意義。 r 之意義較為明白，因所配之綫均為直綫故也；但在 ρ 則不然，如不說明曲綫之形式，則其意義必致混淆。相關指數亦可視為某種曲綫表示兩變量關係是否確當之標準。

任何資料如所配曲綫之方程式中所含常數之多寡與原來資料之點數相同，即可求得一通過各點之綫，此時 ρ 之數值必等於1，但此數值並無意義，蓋吾人採用數學方程式，倘令式中常數之數目與點數相等，常可發生虛謬之結論，而 ρ 之數值適足以表示此種虛謬也。故在計算相關指數時，必須注意配合曲綫之各種原則，蓋此種指數本身原無絕對意義，其意義常為相對的，須與所用曲綫聯帶解釋，始可明瞭。此點不僅對於 ρ 為然，對於其他各種相關量數亦然，但吾人每忽視之，而得虛謬之結論，是誠不可不注意也。

計算相關指數之簡捷法

以上求標準誤及相關指數時，逐步演算，手續繁重，原欲使學者明瞭此兩種量數之意義；實際計算時如利用求 r 公式中所得之關係，則計算手續可大為減省。用於下列遞升冪級數形式

$$Y = a + bX + cX^2 + dX^3 \dots$$

之各種拋物綫時，求 S_y 之公式可由直綫相關中所用公式伸引而得。其式如下：

$$S_y^2 = \frac{\Sigma(Y^2) - a\Sigma(Y) - b\Sigma(XY) - c\Sigma(X^2Y) - d\Sigma(X^3Y) - \dots}{N}$$

吾人又可伸引 r 之公式而求得用於上項曲綫時求 ρ 之普遍公式。其式如下(註)：

$$\rho^2_{yx} = \frac{a\Sigma(Y) + b\Sigma(XY) + c\Sigma(X^2Y) + d\Sigma(X^3Y) + \dots - Nc_y^2}{\Sigma(Y^2) - Nc_y^2}$$

當原點在 Y 之平均數點時，則 $\Sigma y = 0, c_y = 0$ ，上式可化爲

$$\rho^2_{yx} = \frac{b\Sigma(Xy) + c\Sigma(X^2y) + d\Sigma(X^3y) + \dots}{\Sigma(y^2)}$$

茲再將 S 與 ρ 兩公式之特點加以討論。用前式求此兩種量數時所需各數值爲曲綫方程式中常數之數值、配合曲綫時所用之各種數值、以及 $\Sigma(Y^2)$ 與 c_y^2 之數值。其中除 $\Sigma(Y^2)$ 及 c_y^2 須另行計算外，其餘各數值在配合曲綫時均已求得，故 S 及 ρ 之數值可視爲配合曲綫時之副產品，惟吾人如欲使兩變量間之關係完全表示，則除曲綫方程式以外，須求 S 及 ρ 之數值以補充其意義，蓋兩變量之關係如能用某種曲綫表示時，該曲綫方程式所以表示其平均關係， S 所以測度根據此式所得估量之可靠性，而 ρ 則用以測度相關程度之疎密者也。

茲仍以 alfalfa 香草收穫量之資料爲例，以說明前項公式之應用。下列各數值係根據表一〇九之資料計算而得及配合曲綫時所已求得者：

$$a = 3.5468$$

$$\Sigma(X^2Y) = 407,448.00$$

$$b = .2520$$

$$c_y^2 = 55.9504$$

$$c = -0.028162$$

$$\Sigma(Y) = 329.03$$

$$\Sigma(Y^2) = 2688.3129$$

$$\Sigma(XY) = 10,269.96$$

$$N = 44$$

將各數值代入下列二次拋物綫標準誤之公式中

$$S_y^2 = \frac{\Sigma(Y^2) - a\Sigma(Y) - b\Sigma(XY) - c\Sigma(X^2Y)}{N}$$

可得

(註)此式之來源見附錄A。

$$\begin{aligned}
 S^2_y &= \frac{2688.3129 - (3.5468 \times 329.03) - (.2520 \times 10,269.96)}{44} \\
 &\quad - \frac{-(-.0028162 \times 407,448)}{44} \\
 &= \frac{80.7345}{44} \\
 &= 1.8349 \\
 S_y &= 1.36
 \end{aligned}$$

此曲綫之相關指數可用下式計算：

$$r^2_{yx} = \frac{a \sum(Y) + b \sum(XY) + c \sum(X^2Y) - Nc_y^2}{\sum(Y)^2 - Nc_y^2}$$

將各數值代入，得

$$\begin{aligned}
 r^2_{yx} &= \frac{145.7608}{2688.3129 - (44 \times 55.9504)} \\
 r_{yx} &= .80
 \end{aligned}$$

上述求 S 及 P 之方法，應用頗廣，惟所用計算公式須合乎所配曲綫之方程式耳。此項方法下章內舉例尚多，在附錄A內並有詳細之說明。

相關率

表示相關程度之量數，除上述 r 及 P 兩種外，尚有一種，名曰相關率 (correlation ratio)，為披爾遜氏所創，吾人常以符號 η (讀如Eta) 表之。此量數可視為 P 之一種特例，惟其計算方法稍有不同耳。

吾人已知兩變量之相關，無論其為直綫或為曲綫之形式，其相關程度均可用下式求之。

$$\text{相關量數} = \sqrt{1 - \frac{S_y^2}{\sigma_y^2}}$$

當 S_y 係表示對於直綫之標準差時，則此相關量數即為相關係數 r ，而相關指數 P 則為此項量數之一般的表示。相關率與此亦同，其公式亦可用上式表之，惟式中 S_y 所代表者，為對於通過相關表內各直行中點

(即算術平均數點)之綫之標準差,故求此綫時,其方程式所含常數之多寡必與直行之行數相等。如各直行之中點均在一直綫上,則相關率之數值與相關係數相等;如各直行中點不在一直綫上,則相關率之數值大於相關係數。

由是知相關率之概念並不含有新穎之原理,在迴歸綫為非直綫,而兩變量之關係可用一通過各直行中點之曲綫表示時,吾人可用相關率以測量此種相關程度之疎密。此項關係如其完全,而該綫兩旁絕無散佈點者,則 η 之數值為1;如其兩變量間並無關係存在,而各點在該綫兩旁之散佈程度恰與 Y 數列平均數兩旁之散佈程度相等者,則 η 之數值為0。

通常所用計算相關率之公式與上式稍異,因對於通過各直行中點之綫之標準差,在此不用 S_y 而以 σ_{ay} 表示,其意義與 S_y 相同,惟 σ_{ay} 常指相關表而言耳。計算 η 之公式如下:

$$\eta_{yx} = \sqrt{1 - \frac{\sigma_{ay}^2}{\sigma_y^2}}$$

當 η 寫為 η_{yx} 時,係表示 Y 倚 X (Y 為倚變量)之迴歸,寫為 η_{xy} 時,則表示 X 倚 Y (X 為倚變量)之迴歸。後者之數值常隨通過各橫行中點之曲綫兩旁各點之散佈程度而定。 η_{yx} 與 η_{xy} 之數值,除直綫之迴歸外,各不相等,此為與 r 不同之點,蓋後者無論在 Y 為倚變量或 X 為倚變量時,其數值常相等也。

相關率之計算

下表係根據美國某次農業試驗用淡氣作肥料時;每英畝所用淡氣磅數與每英畝小麥收穫蒲式耳數(bushels)所製成之相關表(註)。表內

(註)此表係根據達文博氏(E. Davenport)之農業實驗報告,原文見達氏所著“Comparative Agriculture”論文內,刊載於Bailey氏著“Cyclopedia of American Agriculture”。表內數字係由該報告中任意摘錄者,惟採用此項資料研究所得之結論,與達氏實驗結果相似。

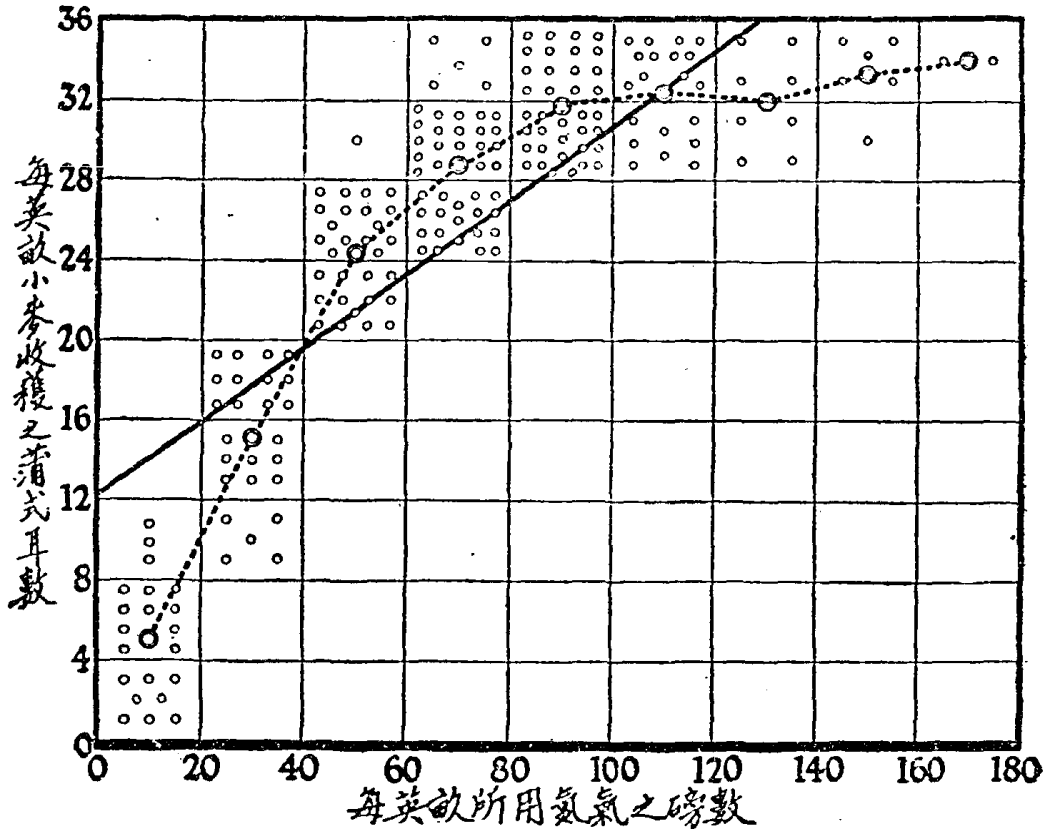
各點均繪於圖八十內。

表 一 一 一

每英畝小麥收穫量與所用淡氣磅數之相關表

		X—每英畝所用淡氣之磅數										各橫行之算術平均數
		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	合計	
Y—每英畝小麥收穫量之蒲式耳數	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	30.0
	8-11.9	3	5								8	22.50
	4-7.9	10									10	10.0
	0-3.9	8									8	10.0
	合計	21	25	30	44	37	20	8	6	2	193	
	各直行之算術平均數	5.05	15.12	24.4	28.73	31.73	32.4	32.0	33.33	34.0		

應用上式計算 η_{yx} 時，吾人須先求 r_y 及 σ_{ay} 之數值。 σ_y 之算法前已言之，無庸贅述；至於 σ_{ay} 則為對於通過各直行中點（即算術平均數點）一綫之均方根差。計算 σ_{ay} 可採用計算 S_y 時所用之第一法，先求各點對於迴歸綫之差，次求其平方之和而除以點數，再開平方即得。此表示兩變量關係之綫既通過各直行之中點，則各中點即可用以替代迴歸方程式所求得之常態數值，故計算 σ_{ay} 時可根據各點對於各直行中點之差，一一平方後相加而除以點數，再開平方即得，如計算標準差然。茲就表一一一第一直行之數字以說明每行之計算手續。該直行內共有21項，其X值均在0與20之間。此21項之Y算術平均數為5.05，各項離中差均根據此平均數計算。



圖八十. 小麥收穫量與氮氣用量之散佈圖及其迴歸直綫與通過各直行之中點綫

表 一 一 二

表 一 一 一 第一直行內各項離中差平方之計算

組 距 (每英畝小麥收 穫之蒲式耳數)	<i>m</i>	<i>f</i>	各項對於直行 平均數(5.05) 之離中差 <i>d</i>	<i>d</i> ²	<i>fd</i> ²
8—11.9	10	3	4.95	24.5025	73.5075
4—7.9	6	10	.95	.9025	9.0250
0—3.9	2	8	-3.05	9.3025	74.4200
合 計					156.9525

其他直行內各項離中差之平方,均可依此計算。對於各直行算術平均數之標準差 σ_{ay} 之數值求得為2.420, σ_y 之數值為9.188。

將各數值代入下式

$$\eta^2_{yx} = 1 - \frac{\sigma_{ay}^2}{\sigma_y^2}$$

$$\begin{aligned}\eta^2_{yx} &= 1 - \frac{(2.42)^2}{(9.188)^2} \\ &= 1 - .0694 \\ &= .9306 \\ \eta_{yx} &= .965\end{aligned}$$

此即測量通過各直行中點綫兩旁散佈程度之相關率之數值也。其意義論述於下。

前項計算 η 之手續過繁，可另求簡捷之法。令 σ_{my} 代表各直行算術平均數對於 Y 總平均數之標準差。計算此值時，各直行平均數對於總平均數之離中差應以各直行之次數乘之。吾人可證明(註)

$$\sigma_{ay}^2 = \sigma_y^2 - \sigma_{my}^2$$

(註)以下採用猶爾(Yule)氏之證明。

已知一數列之算術平均數為 M ，分此數列為兩部，令 M_1 為第一部之算術平均數， M_2 為第二部之算術平均數， N 為總項數，則 N 等於 $N_1 + N_2$ ，即兩部項數之和也。今欲知 σ ， σ_1 及 σ_2 三者之關係。設

$$M_1 - M = c_1$$

又設 S_1 為第一部各項對於 M 之標準差，則

$$S_1^2 = \sigma_1^2 + c_1^2$$

同理

$$S_2^2 = \sigma_2^2 + c_2^2$$

但 $N_1 S_1^2$ 等於第一部各項對 M 離中差平方之和， $N_2 S_2^2$ 等於第二部各項對 M 離中差平方之和，故

$$\sigma^2 = \frac{N_1 S_1^2 + N_2 S_2^2}{N}$$

$$N\sigma^2 = N_1 S_1^2 + N_2 S_2^2 \quad (1)$$

但

$$S_1^2 = \sigma_1^2 + c_1^2, \quad S_2^2 = \sigma_2^2 + c_2^2$$

故

$$N\sigma^2 = N_1(\sigma_1^2 + c_1^2) + N_2(\sigma_2^2 + c_2^2) \quad (2)$$

在本例中，設 Y 之算術平均數為 M_y ，分此數列為數部(每一直行內各項作為一部)，其

將下式內 σ_{ay}^2 一項以 $\sigma_y^2 - \sigma_{my}^2$ 代入

$$\eta^2_{yx} = 1 - \frac{\sigma_{ay}^2}{\sigma_y^2}$$

得

$$\eta^2_{yx} = 1 - \left(\frac{\sigma_y^2 - \sigma_{my}^2}{\sigma_y^2} \right)$$

$$= \frac{\sigma_{my}^2}{\sigma_y^2}$$

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

因計算 σ_{my} 之手續遠較 σ_{ay} 為簡易，故通常皆用上式計算 η 。茲就表一一一之資料以說明其計算手續。

(續上頁註)

算術平均數各為 m_{y1}, m_{y2} 等。又設 S_{ay} 代表任何一直行內各項對於該行算術平均數之標準差，則各直行之標準差為 S_{ay1}, S_{ay2} 等，各直行之算術平均數與 Y 之總平均數之差為 $M_y - m_{y1}, M_y - m_{y2}$ 等。將各數代入(2)式內，得

$$N\sigma_y^2 = n_1[S_{ay1}^2 + (M_y - m_{y1})^2] + n_2[S_{ay2}^2 + (M_y - m_{y2})^2] + \dots \quad (3)$$

$$N\sigma_y^2 = \sum n[S_{ay}^2 + (M_y - m_y)^2] \quad (4)$$

但 $N\sigma_{ay}^2 = \sum (n \cdot S_{ay}^2)$

因在每行內 $S_{ay}^2 = \frac{\sum d^2}{n}$

d 為每行內各項對於該行算術平均數之離中差，故

$$\sigma_{ay}^2 = \frac{\sum d^2}{N} = \frac{\sum (n \cdot S_{ay}^2)}{N}$$

將此數代入(4)式，得

$$N\sigma_y^2 = N\sigma_{ay}^2 + \sum n(M_y - m_y)^2 \quad (5)$$

由各行算術平均數對於 Y 總平均數之標準差之定義，得

$$\sigma_{my}^2 = \frac{\sum n(M_y - m_y)^2}{N}$$

故由(5)式得

$$\sigma_y^2 = \sigma_{ay}^2 + \sigma_{my}^2 \quad (6)$$

表 一 一 三
相 關 率 之 計 算

X各組中點 (磅)	Y各行之 算術平均數 (蒲式耳) m_y	對於 Y 總平均 數(25.605) 之離中差 d	離中差之 平方 d^2	次數 f	fd^2
10	5.05	-19.955	398.202	21	8,362.242
30	15.12	-9.885	97.713	25	2,442.825
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
合計				193	15,162.769

$$\sigma_{my} = \sqrt{\frac{15,162.769}{193}}$$

$$= 8.864$$

將各數值代入下式

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

$$\eta_{yx} = \frac{8.864}{9.188}$$

$$= .965$$

計算相關率之步驟,可概括於下:

1. 將各項目製成相關表。
2. 求各直行(即X行)內Y數值之算術平均數。
3. 計算Y數列全體之算術平均數(即Y之總平均數)。
4. 就各直行內Y數值之算術平均數而計算其對於Y數列總平均數之離中差,求其平方,乘以各行次數,而求其總和。

5. 求得之總和以總項數除之，再開平方，即得 σ_{my} 之數值。

6. 計算 σ_y 之數值。

7. 以 σ_y 除 σ_{my} ，即得 η_{yx} 。

X 倚 Y 之相關率，亦可依此將相當數值代入下式計算即得。

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

式中 σ_{mx} 代表各橫行內 X 數值之算術平均數對於 X 數列總平均數之標準差。 X 倚 Y 之相關率之數值，係隨通過各橫行中點綫兩旁之散佈程度而定。其值常與 Y 倚 X 之相關率不同。在本例中 η_{xy} 之數值為.824。相關綫之形式愈近於直綫，則此兩相關率之數值亦愈相近。

η 之數值不能大於1，此與 r 相同。如等於1，即謂通過各直行（或橫行）中點綫兩旁全無散佈是也。由公式

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

觀之，如 σ_{my} 等於0，則 η_{yx} 亦等於0。在各直行算術平均數之值與 Y 之總平均數相等時， σ_{my} 即等於0，換言之，即 X 變量之值或增或減時， Y 變量之值並不發生聯帶變動，故相關表中各直行之次數分配與 Y 之總次數分配相似，此時兩變量間自無關係之可言也。

此外應加注意者，相關率之數值決無負數。兩變量間之關係為正為負，或正負兼有，可觀相關表而得之。

以相關係數與相關率相較，則前者有一顯著之優點。吾人如已知相關係數及兩標準差之數值，可立得一迴歸綫之方程式； η 則不然，在求得 η 數值後，尚須計算通過各行中點一綫之方程式，以表示兩變量間之關係焉。

相關率之校正

相關率 η 之應用，以資料豐富而能排列成相關表者為限。如項數無

多，而在製成相關表時每行祇有一項，則 σ_{my} 與 σ_y 之數值相等，而 η 之數值當然為 1。故若項數甚少而分組甚多，則由此求得之相關率將無意義可言。

披爾遜氏(Karl Pearson)為欲校正相關率因分組過多而發生之錯誤起見，曾創一校正公式。以 k 代表行數，則 η 之數值可用下式校正之。

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

此式應用於前例時，則

$$\begin{aligned} \text{校正後之 } \eta^2 &= \frac{(.965)^2 - \frac{(9-3)}{193}}{1 - \frac{(9-3)}{193}} \\ &= .929 \end{aligned}$$

$$\text{校正後之 } \eta = .964$$

此數值與未校正前相差甚微，但若項數(N)甚少，或行數(k)甚多，則 η 之數值常因校正而大見減小。

相關率與相關係數之關係

如兩變量間之關係確為一直綫，則通過各直行中點之綫自必與計算 r 時所根據之迴歸綫相合，而 η 與 r 之數值亦必相等；如兩變量間之關係由直綫而漸變為非直綫，則 η 與 r 之數值亦漸有差異，而前者常高於後者，此由於通過各行中點綫兩旁之散佈程度常較各該中點上所配直綫兩旁之散佈程度為小故也。相關綫兩旁之散佈程度愈小，則相關量數之值愈大，故在前例中 r 之數值為 $+ .68$ ，由二次拋物綫所得相關指數之值為 $.80$ ，用同樣資料求得相關率之數值為 $.83$ 。又自表一一一之資料求得 η_{yx} 之數值為 $.964$ ， r 之數值為 $+ .793$ ，兩者相差甚鉅者，其原因可由圖八十觀察而得。圖中 Y 倚 X 之迴歸綫應為曲綫而非直綫，結果迴歸直綫兩旁之散佈程度乃遠較通過各中點綫兩旁之散佈程度為高，故兩變量間

之關係是否為直綫抑為曲綫，可以 r 與 η 之關係作為測驗之標準。如迴歸正為一直綫，則兩數值相等；如迴歸離直綫之形式愈遠，則兩數值相差亦愈大。此項測驗之公式為

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

惟應用上式時，如迴歸確係直綫， η 與 r 之數值未見能絕對相等，此純由於機遇(chance)所致。但若 r 與 η 相差甚鉅，而 ζ (讀如 Zeta)之數值甚大時，即謂兩變量間之關係不應以直綫表示，而 r 亦不應作為測量相關程度之標準矣。

前例中 η 之數值為.964， r 為.793，故 ζ 之數值為.300，此值已甚大，而足以表示該迴歸實為非直綫之形式。關於斷定 ζ 數值大小之意義，後文(註)當再論述之。

參 考 書

- BOWLEY, A. L. *Elements of Statistics* (365—367).
- KELLEY, TRUMAN L. *Statistical Method* (238—245).
- PERAL, RAYMOND. *Medical Biometry and Statistics* (311—318).
- PEARSON, KARL. *Mathematical Contributions to the Theory of Evolution (XIV)*.
On the General Theory of Skew Correlation and Non-Linear Regression.
 Draper's Company Research Memoirs, Biometric Series II. 1906.
Notes on the History of Correlation. *Biometrika*, Vol. 13, 1920 (25—45).
On a Correction Needful in the Case of the Correlation Ratio. *Biometrika*,
 Vol. 8, 1911 (254—256).
On the Correction Necessary for the Correlation Ratio. *Biometrika*, Vol.
 14, 1923 (412—417).
- RIETZ, H. L. (editor) *Handbook of Mathematical Statistics* (129—131).
- YULE, G. U. *An Introduction to the Theory of Statistics* (204—207).

第十三章 關係之測度與估量之問題

數量方法之應用於經濟及商業方面，常注重於估量之問題。吾人嘗研究迴歸方程式、標準誤以及相關係數，因其能助吾人解決各種實際問題，如斷定最可能之產量、最可能之價格以及最可能之商業變動等；但估量之應用並不限於預測將來之變動，即如吾人欲由許多觀察數值中，確定一最可能之數值，或用數學方程式以表示變量間之關係，亦莫不含有估量之意味。統計技術之價值大抵在於估量方法之實際運用。

以前各章之討論，亦均以此問題為中心，而對於由一變量之數值以估量另一變量之數值之方法，亦嘗有所闡明。此時可先將應用估量方法時所作之種種假定概括論述之。

估量所含之假定

在觀察所得之各數值中，最可能之數值為其算術平均數，此點前嘗論及。假如平均數兩旁之分配為常態時，則根據此平均數所得估量之機誤可用標準差測度之。又假如迴歸綫兩旁之分配為常態時，則根據此迴歸綫所得估量之機誤可用標準誤測度之。迴歸方程式之意義及其功用則可由一變量之標準誤及其標準差之關係而定。

由上兩項數值之關係又可得一抽象之相關量數，即相關係數或相關指數是也。但此種係數或指數之意義及其確度常為常態分配之假定所限。如迴歸綫兩旁之分配及平均數兩旁之分配均為常態或近乎常態時，吾人方可用此兩種量數以測度相關之程度而無誤；如分配不合常態，則此兩種量數之意義亦必因之減損矣。

惟以上所討論者，均係就算術數值(即絕對數值，或真數數值)而言，所謂根據平均數以求估量者，此平均數常指算術平均數；所謂假定平均數兩旁及迴歸綫兩旁之分配合乎常態者，其各項差量亦常指絕對量 標

準差及標準誤所用之單位亦均為算術單位，而以絕對數值表示者也。但經濟資料之分配不僅限於算術的形式，估量及測度估量差誤所用之單位亦不必限於算術單位。吾人當前之問題即在於斷定前所討論之各種方法是否適用於假定範圍以外之他種分配。此問題可用一實例解答之。

表 一 一 四

美國燕麥之產量及價格

年 份	燕 麥 產 量 (單位:百 萬蒲式耳)	產量之直綫 長期趨勢值 (註一)	實際產量對 長期趨勢值 之比率	芝加哥燕麥 價格 (每蒲式耳 值美金分)	價格之直綫 長期趨勢值 (註二)	實際價格對 長期趨勢值 之比率
1881	416	448	.929	47	36.0	1.30
1882	488	471	1.036	37	35.3	1.05
1883	571	494	1.156	31	34.6	.90
1884	583	517	1.128	29	34.0	.85
1885	629	540	1.165	28	33.2	.84
1886	624	563	1.108	25	32.5	.77
1887	659	586	1.124	30	31.2	.96
1888	701	609	1.151	24	30.5	.79
1889	751	632	1.188	24	29.8	.81
1890	523	655	.798	43	29.0	1.48
1891	738	678	1.088	31	28.3	1.10
1892	661	701	.943	30	27.5	1.09
1893	639	724	.882	31	26.8	1.16
1894	662	747	.886	28	26.1	1.07
1895	824	770	1.070	19	25.3	.75
1896	780	793	.983	18	23.6	.76
1897	791	816	.969	24	25.0	.96
1898	843	839	1.005	25	26.4	.95
1899	926	862	1.074	23	27.8	.83
1900	914	885	1.033	25	29.2	.86
1901	778	908	.857	42	30.6	1.37
1902	1053	931	1.131	33	32.0	1.03
1903	869	954	.911	38	33.4	1.14
1904	1009	977	1.033	30	34.8	.86
1905	1090	1000	1.090	31	36.2	.86
1906	1036	1023	1.013	39	37.6	1.04
1907	805	1046	.770	51	39.0	1.31
1908	851	1069	.796	52	40.4	1.29
1909	1068	1092	.978	43	41.8	1.03
1910	1186	1115	1.064	35	43.2	.81
1911	922	1138	.810	51	44.6	1.14
1912	1418	1161	1.221	37	46.0	.80
1913	1122	1184	.948	41	47.4	.87

(註一)配合此長期趨勢綫所根據之原來資料，包含時期較此為長。

(註二)此直綫長期趨勢綫係採用 H. B. Killough 氏所配合者。配綫時係將全時期分為兩部份，以1881年至1895年為一部，1896年至1913年為一部，而各配以直綫。

估量問題之實例

表一一四所列爲一八八一年至一九一三年美國燕麥之產量及價格，與該兩數列所配合之長期趨勢綫，以及各項實際數值對其長期趨勢值之比率。

吾人現欲測度此兩比率變量間之關係。此項關係似可用雙曲綫方程式 $Y = aX^b$ 表示之。惟用最小平方法配合此綫，須將該式化爲對數之形式：

$$\log Y = \log a + b \log X$$

配綫時所用之兩常態方程式如下：

$$I \quad \Sigma(\log Y) = N \log a + b \Sigma(\log X)$$

$$II \quad \Sigma(\log X \cdot \log Y) = \log a \Sigma(\log X) + b \Sigma(\log^2 X)$$

解此兩式所需各數值可由表一一五得之如下(註)：

$$N = 33$$

$$\Sigma(\log Y) = -.32849$$

$$\Sigma(\log X \log Y) = -.1143005$$

$$\Sigma(\log X) = .037535$$

$$\Sigma(\log^2 X) = .096423$$

將各數值代入上兩式，得

$$-.32849 = 33 \log a + .037535b$$

$$-.1143005 = .037535 \log a + .096423b$$

更將此兩式解之，得

$$\log a = -.00861$$

$$b = -1.18206$$

故吾人所求之方程式爲

(註)表一一四及一一五內之資料係經美國農業經濟局 (Bureau of Agricultural Economics) 凱洛氏(H. B. Killough)之允許而採用者，深爲感謝。氏爲研究燕麥價格漲落之因素，曾搜集此項數字，本表即係由此項數字中摘錄者。

$$\log Y = (9.99139 - 10) - 1.18206 \log X$$

或

$$Y = .9804X^{-1.18206}$$

此為表示燕麥產量與價格兩者平均關係之方程式（兩數列所用數值各為實際數值對其長期趨勢值之比率）。此綫如圖八二所示。

表 一 一 五

配合燕麥產量與價格之相關綫時所需各項數值之計算

第一例

(1) 年份	(2) 實際價格對常態價格之比率 Y	(3) 實際產量對常態產量之比率 X	(4) $\log Y$	(5) $\log X$	(6) $\log^2 Y$	(7) $\log^2 X$	(8) $\log Y \cdot \log X$
1881	1.30	.929	.1139434	.9680157-1	.01298310	.001022995	-.0036444
1882	1.05	1.036	.0211893	.0153598	.00044899	.000235923	.0003255
1883	.90	1.156	.9542425-1	.0629578	.00209375	.003963685	-.0728808
1884	.85	1.128	.9294189-1	.0523091	.00498169	.002736242	-.0336920
1885	.84	1.165	.9212793-1	.0663259	.00573362	.004399125	-.0050222
1886	.77	1.108	.8864907-1	.0445398	.01288436	.001983794	-.0050557
1887	.96	1.124	.9822712-1	.0507663	.00031431	.002577217	-.0009000
1888	.79	1.151	.8976271-1	.0610753	.01048021	.003730192	-.0062524
1889	.81	1.188	.9084850-1	.0748164	.00837500	.005597494	-.0068468
1890	1.48	.798	.1702617	.9020029-1	.02898905	.009603432	-.0166852
1891	1.10	1.088	.0413927	.0366289	.00171336	.001341676	.0015162
1892	1.69	.943	.0374265	.9745117-1	.00140074	.000649653	-.0009539
1893	1.16	.882	.0644580	.9454686-1	.00415483	.002973674	-.0035150
1894	1.07	.886	.0293838	.9474337-1	.00086311	.002763216	-.0015446
1895	.75	1.070	.8750613-1	.0253838	.01560968	.000863408	-.0036712
1896	.76	.983	.8808136-1	.9925535-1	.01420540	.000055450	.0008875
1897	.96	.969	.9822712-1	.9863238-1	.00031431	.000187038	.0002425
1898	.95	1.005	.9777236-1	.0021661	.00049624	.000004692	-.0000483
1899	.83	1.074	.9190781-1	.0310043	.00654835	.000961267	-.0025089
1900	.86	1.033	.9344985-1	.0141003	.00429045	.000198818	-.0009236
1901	1.37	.857	.1367206	.9329808-1	.01725316	.004491573	-.0091629
1902	1.03	1.131	.0128372	.0534626	.00016479	.002858250	.0006863
1903	1.14	.911	.0569049	.9595184-1	.00323817	.001638760	-.0023036
1904	.86	1.033	.9344985-1	.0141003	.00429045	.000198818	-.0009236
1905	.86	1.090	.9344985-1	.0374265	.00429045	.001400743	-.0024515
1906	1.04	1.013	.0170333	.0056094	.00029013	.000031465	.0000955
1907	1.31	.770	.1172713	.8864907-1	.01375256	.012824361	-.0133113
1908	1.25	.796	.1105897	.9009131-1	.01223008	.009818214	-.0109580
1909	1.03	.978	.0128372	.9903389-1	.00016479	.000093337	-.0001240
1910	.81	1.064	.9084850-1	.0269416	.00837500	.000725850	-.0024656
1911	1.14	.810	.0569049	.9084850-1	.00323817	.008374812	-.0052076
1912	.86	1.221	.9030900-1	.0867157	.00939155	.007519613	-.0084036
1913	.87	.948	.9395193-1	.9768083-1	.00365792	.000537855	.0014027
合計	32.83	33.338	17.6715068-18	14.0375350-14	.21721807	.096422642	-.1194567 +.0051562 -.1143005

對數表示之估量標準誤

然則前項方程式之可靠性及由該式所得估量之可靠程度若何？欲解答此項問題，吾人須求標準誤 S 。前項相關綫既用對數配合，則計算標準誤亦可用對數。用對數求該綫標準誤之公式，可用前述直綫及拋物綫求 S_y 之公式稍加改變即得。其式如下：

$$S^2_{\log y} = \frac{\Sigma(\log^2 Y) - \log a \Sigma(\log Y) - b \Sigma(\log X \cdot \log Y)}{N}$$

將相當數值代入上式，得

$$\begin{aligned} S^2_{\log y} &= \frac{.21721807 - (-.00861 \times -.32849) - (-1.18206 \times -.1143005)}{33} \\ &= \frac{.07927928}{33} \end{aligned}$$

$$S^2_{\log y} = .0024024$$

$$S_{\log y} = .04901$$

以對數表示之該綫標準誤為 .04901。此標準誤除係用對數表示者外，其意義與用真數表示之標準誤相同，可解釋如下：假定相關綫兩旁各項對數之分配為常態時，則由此綫估量所得之對數，對於實際數值之對數差誤無過 .04901 者，其機率為 68%；差誤無過 .09802 者，其機率為 95%；差誤無過 .14703 者，其機率為 99.7%。

估量標準誤之解釋；估量區域

以上所述在真數上之意義若何？簡言之，上例所用者為比率而非絕對數值。凡兩數對數之相差即為由該兩數所得比率之對數，故 S 之絕對數值須視所用數值之大小而定。如吾人欲將 S 化為絕對數值，則必先知估量數 Y 後可。其法以一已知 X 之數值，代入平均關係之方程式，即可求得與其相配之 Y 值之估量數。如該方程式為對數形式，則此估量數亦

爲對數，以 $S_{\log y}$ 與估量數之對數相加，而求其真數，取此真數作距離，沿相關綫之上方繪一綫，是爲離相關綫 $1S$ 距離所成區域之上限；更由估量數之對數內減去 $S_{\log y}$ ，而求其真數，取此真數作距離，沿相關綫之下方繪一綫，是爲離相關綫 $1S$ 距離所成區域之下限。估量數 Y 之數值在此區域內之機率爲 68%。離相關綫 $2S$ 及 $3S$ 距離所成區域之上下限，可用同法求之。

由此對數綫所得之估量區域，與前此所述由簡單直綫所得之估量區域大不相同。在直綫上下兩方以 $1S$ 爲距離所得之估量區域，如用絕對數量其寬度，各段常相同，而相關綫適居其中。對數之估量區域則否，倘用真數量其寬度，各段皆不相同，且相關綫上下兩方之距離亦不相等，但相關綫上方之距離與下方距離之比率則各段相同；換言之，亦即估量值之對數減去 $1S$ 之對數後所得反對數之值對估量值之比率，等於估量值對估量值之對數加上 $1S$ 之對數後所得反對數之值之比率也。如將曲綫繪入對數尺度之圖內，則相關綫上下兩方 $1S$ 距離之估量區域成完全對稱之形式，與前此所論簡單之例中所得者相同。吾人倘對此常存有比率之觀念，而又慣用對數尺度者，當不難索解也。

比率表示之估量標準誤

相關綫上下兩方距離之比率既各段相同，則吾人可將估量標準誤化爲比率之形式。在本例中，得

$$S_r = \text{anti-log } S_{\log y} = \text{anti-log } .04901 = 1.12$$

此處 S_r 爲用比率表示之估量標準誤。 $S_{\log y}$ 既爲正數，故此比率大於 1。此爲較大數對較小數之比率，其意義可解釋如下：在 68% 之機率中，倘實際數值大於估量數值，則所大之數無過 12%；倘小於估量數值，則估量數值所大於實際數值之數無過 12%。此種形式殊不便解釋，因其比率係較大數對較小數之比也。吾人倘將此比率化爲估量數值之百分比，則

解釋較為便利。此可將 $S_{\log y}$ 之對數數值作為負對數而求其真數，則 $-.04901 = 9.95099 - 10$ ，其真數為 .8933。此處之比率業已化為較小數對較大數之比率。若將兩項比率混合採用，則 S_r 之意義即易於解釋矣。其式可寫為

$$S_r = .89 \text{ 至 } 1.12$$

此式可解釋如下：相關綫兩旁之分配為常態時，則實際數值對估量數值之百分比，不小於 89% 而又不大於 112% 者，其機率為 68%。此式內之大小兩數值均以估量數值為百分，故其意義較為簡括而確定，較絕對數值之形式尤合實用(註)。

求 $2S$ 與 $3S$ 之數值時，吾人不可逕將前項百分比用 2 或 3 乘之，而須以 2 或 3 乘 $S_{\log y}$ ，然後求其真數；又為便利應用起見，須將正負兩項之對數各求其真數，如上例然。其計算手續尚屬簡單，舉示於下：

$$2S_{\log y} = .09802$$

此對數如為正數，則其真數為 1.25；如為負數，其真數為 .80。

$$3S_{\log y} = .14703$$

此對數為正數時，其真數為 1.40；為負數時，其真數為 .71。茲再將標準誤併列於下：

$$S_r = .89 \text{ 至 } 1.12$$

$$2S_r = .80 \text{ 至 } 1.25$$

$$3S_r = .71 \text{ 至 } 1.40$$

各 S_r 之大小兩數值均為實際值對估量值百分比之上下兩限，此百分比在 S_r 所示之兩限以內者，其機率為 68%；在 $2S_r$ 所示之兩限以內者，其機率為 95%，在 $3S_r$ 所示之兩限以內者，其機率為 99.7%。惟凡此所述均

(註)達文博氏(D. H. Davenport)於1922年在其未刊佈之論文中即曾指出測度估量數可靠性之量數，應以百分形式表示之，而在數種研究中亦曾有採用此項形式者，惟彼時尚無計算此項量數之具體方法，故為用不彰。

以實際數值之對數在相關綫兩旁之分配合乎常態為要件耳。

估量標準誤之應用

茲仍就前述燕麥之資料，以說明 $S_{log y}$ 之應用。例如已知某年燕麥之實際產量為高於常態產量50%（即實際產量對常態產量之比率為1.50），而欲求是年最可能之價格比率，並斷定此估量之準確程度。

此處所用估量之方程式為

$$\log Y = (9.99139 - 10) - 1.18206 \log X$$

將 .176091 之數值（即 1.50 之對數）代入式中之 X ，得 $\log Y$ 之數值為 $9.78324 - 10$ ，故知 Y 之真數為 .607。此值亦為一比率，即謂實際產量為其常態產量（係由所配之長期趨勢綫求得者）之150%時，則最可能之實際價格對其常態價格（亦係由所配之長期趨勢綫求得者）之百分比應為 60.7%。

由此式所得 Y 之數值，僅係一種估量，欲知此估量之準確程度，須求其標準誤。應用前所求得之標準誤數值，則60.7之89%為54，60.7之112%為68。由此可斷定燕麥實際產量為其常態產量之150%時，其實際價格對其常態價格之百分比，不少於54%而無過68%者，其機率為68%（註一）。 $2S_r$ 及 $3S_r$ 之上下限亦可依上法求之。

對數表示之相關指數

於此吾人又可求第三種量數，即抽象之相關指數是也（註二）。倘方

（註一）本例中所配之長期趨勢綫是否適當，應加考慮，蓋據此以決定估量之可靠程度，其所得結論是否準確與曲綫配合之適度有聯帶關係也。此問題在另節內當有詳盡之討論。

（註二）此處之關係用對數表示雖為直綫形式，但其相關量數不用符號 r ，而用 P 代表之，蓋此量數既用對數表示，則解釋時即不能與普通之相關係數完全相同也。

程式之形式爲

$$\log Y = \log a + b \log X$$

則 ρ 之公式應爲

$$\rho^2_{\log y \log x} = \frac{\log a \sum(\log Y) + b \sum(\log X \cdot \log Y) - Nc^2_{\log y}}{\sum(\log^2 Y) - Nc^2_{\log y}}$$

式中 $c_{\log y}$ 爲 Y 值之對數之算術平均數與原點 (即對數尺度上之零點) 之差。將相當數值代入上式, 得

$$\begin{aligned} \rho^2_{\log y \log x} &= \frac{(-.00861 \times -.32849) + (-1.18206 \times -.1143005) - (33 \times .00009909)}{.21721807 - 33 \times .00009909} \\ &= \frac{.13466882}{.2139481} \\ &= .629445 \end{aligned}$$

$$\rho_{\log y \log x} = .793$$

由此求得相關指數之值爲 .793。此相關指數既係根據對數計算, 則其意義應如何解釋乎?

此數如用下式

$$\rho^2_{\log y \log x} = 1 - \frac{S^2_{\log y}}{\sigma^2_{\log y}}$$

表示, 則其意義更爲明瞭。在此

$$S_{\log y} = .04901$$

$$\sigma_{\log y} = .08052$$

將此兩值求其平方, 代入上式, 得

$$\rho^2_{\log y \log x} = 1 - \frac{.002402417}{.00648328}$$

即

$$\rho_{\log y \log x} = .793$$

然則此值所測度者爲何? 吾人已知兩變量間之關係如能以函數表

示，則 r 及相關指數 ρ 皆為測度此相關程度之抽象量數。 ρ 之價值常由迴歸綫兩旁散佈程度與 Y 之算術平均數兩旁散佈程度比較而得。如用迴歸方程式作為估量根據時所得之散佈程度，較用 Y 之算術平均數作為估量根據時所得之散佈程度大為減少時，則此方程式可認為有表示此關係之價值。故 ρ 之價值全視 S_y 與 τ_y 之關係而定。

前章內測量散佈程度俱用絕對差數， ρ 之數值亦視兩種散佈程度之絕對量數之關係而定。現時吾人所欲測度者則為對數或比率之散佈程度，即各項差數不用真數而用對數計算者也。

測量散佈程度既係根據對數或比率，則解釋相關指數 ρ 時，即不可不認明此項事實。 ρ 之數值固仍視 S^2 及 σ^2 之關係而定，但此時該兩種量數均須用對數計算；換言之，即 ρ 之數值須視迴歸綫兩旁比率之散佈程度與 Y 之幾何平均數兩旁比率之散佈程度間之關係而定（此處 Y 之平均數為幾何平均數，因 Y 之對數之算術平均數即其幾何平均數之對數也）。

此處所有各種量數在比率方面之功用，與前所討論之 S 及 ρ （如係直綫則為 r ）在真數方面之功用完全相同，其求法與前所討論之 S 及 ρ 亦復相同，惟所根據之相關方程式中，其倚變量為 $\log Y$ （或為 $\log X$ ）。計算各種量數所用之公式亦與用真數時相同，惟須將 $\log Y$ 代替 Y 耳。如用對數圖紙代替真數圖紙，則各項手續均相同矣。

吾人應加注意者， X 數值無論用真數或對數表示，如 Y 數值用對數，則所得各量數之值亦為對數或比率形式。例如上例所配曲綫之形式為

$$\log Y = \log a + b \log X$$

即普通之拋物綫或雙曲綫之對數形式。式中 X 係用對數表示， S 及 ρ 之數值固為對數形式；即所配曲綫為指數曲綫 $Y = a(b^X)$ 之下列對數形式時

$$\log Y = \log a + X \log b$$

X 雖用真數表示，而 S 及 ρ 之數值亦仍為對數形式也。

前兩種曲綫，其對數方程式俱為直綫形式，但 S 及 ρ 之用途並不受其限制；不論所用者為對數或真數，函數之關係為直綫或非直綫而 S 及 ρ 仍可用以測量相關之程度。

茲將代表各種量數所用符號之意義說明之。倘兩變量之關係用真數表示時，吾人常用 S_y , σ_y 及 ρ 以代表各量數，前兩者為用絕對數值測量散佈程度之量數，後者為用真數表示相關程度之指數。如 Y 數值係用對數表示者，則須用 $S_{\log y}$ 及 $\sigma_{\log y}$ 以取別於 S_y 及 σ_y 。前者為測量迴歸綫兩旁對數散佈程度之量數，後者為測量 Y 對數值之算術平均數兩旁散佈程度之量數。如 $S_{\log y}$ 化為比率形式，則可用 S_r 代表之。此時相關指數 ρ 因意義稍異，故應寫作 $\rho_{\log y \log x}$ 或 $\rho_{\log y x}$ 以資區別。

倒數之應用於相關之測度

燕麥產量與價格之關係，吾人可用另一種曲綫表示之。此綫頗含新意義，其所表示之相關頗為新穎，因之各種量數之解釋亦與前不同。此綫之形式為

$$Y = \frac{1}{a + bX}$$

此式又可在分母內伸展一項而成為

$$Y = \frac{1}{a + bX + cX^2}$$

此為一種雙曲綫之方程式，在某種研究方面，嘗用以代表數種物品之需要綫(demand curve)者。

前項方程式亦可寫作

$$\frac{1}{Y} = a + bX$$

此式亦為直綫方程式之一種，可表示 Y 值之倒數與 X 之原來數值間之關係。配合該綫所用之常態方程式為：

$$I \quad \Sigma\left(\frac{1}{Y}\right) = Na + b\Sigma(X)$$

$$II \quad \Sigma\left(\frac{X}{Y}\right) = a\Sigma(X) + b\Sigma(X)^2$$

式中所需之各數值，其計算方法如表一一六所示。

將各數值代入上兩式，可得

表 一 一 六

配合燕麥產與價格之相關綫時所需各項數值之計算

第 二 例

(1) 年 份	(2) 價格之 比 率 Y	(3) 產量之 比 率 X	(4) $\frac{1}{Y}$	(5) $\frac{X}{Y}$	(6) $\left(\frac{1}{Y}\right)^2$	(7) X^2
1881	1.30	.929	.7692308	.7146154	.59171602	.863041
1882	1.05	1.036	.9523810	.9866667	.90702957	1.073296
1883	.90	1.156	1.1111111	1.2844444	1.23456788	1.336336
1884	.85	1.128	1.1764706	1.3270588	1.38408307	1.272384
1885	.84	1.165	1.1904762	1.3869048	1.41723358	1.357226
1886	.77	1.108	1.2987013	1.4389610	1.68662507	1.227664
1887	.96	1.124	1.0416667	1.1708334	1.08506951	1.263376
1888	.79	1.151	1.2655228	1.4569620	1.60230736	1.324801
1889	.81	1.188	1.2345679	1.4666667	1.52415790	1.411344
1890	1.48	.798	.6756757	.5391892	.45653765	.636804
1891	1.10	1.088	.9090909	.9890909	.82644626	1.183744
1892	1.09	.943	.9174312	.8651376	.84168001	.889249
1893	1.16	.882	.8620690	.7603449	.74316296	.777924
1894	1.07	.886	.9345791	.8280373	.87343865	.784996
1895	.75	1.070	1.3333333	1.4266666	1.77777769	1.144900
1896	.76	.983	1.3157895	1.2934211	1.73130201	.966289
1897	.96	.969	1.0416667	1.0093750	1.08506951	.938961
1898	.95	1.005	1.0526316	1.0578948	1.10808329	1.010025
1899	.83	1.074	1.2048193	1.2939759	1.45158955	1.153476
1900	.86	1.033	1.1627907	1.2011628	1.35208221	1.067089
1901	1.37	.857	.7299270	.6255480	.53279343	.734449
1902	1.03	1.131	.9708738	1.0980583	.94259594	1.279161
1903	1.14	.911	.8771930	.7991228	.76946756	.829921
1904	.86	1.033	1.1627907	1.2011628	1.35208221	1.067089
1905	.86	1.090	1.1627907	1.2674419	1.35208221	1.188100
1906	1.04	1.013	.9615385	.9740385	.92455629	1.026169
1907	1.31	.770	.7633588	.5877863	.58271666	.592900
1908	1.29	.796	.7751938	.6170543	.60092543	.633616
1909	1.03	.978	.9708738	.9495146	.94259594	.956484
1910	.81	1.064	1.2345679	1.3135802	1.52415790	1.132096
1911	1.14	.810	.8771930	.7105263	.76946756	.656100
1912	.80	1.221	1.2500000	1.5262500	1.56250000	1.490841
1913	.87	.948	1.1494258	1.0896552	1.32117852	.898704
	32.83	33.338	34.3360320	35.2571485	36.85702940	34.168554

$$34.3360320 = 33a + 33.338b$$

$$35.2571485 = 33.338a + 34.168554b$$

解此兩式，得

$$a = -.1357$$

$$b = 1.1643$$

故所求之方程式為

$$\frac{1}{Y} = -.1357 + 1.1643X$$

倒數表示之估量標準誤及相關指數

欲確定上項方程式之可靠程度，必須求其標準誤及相關指數，所用公式可依前法推求之。用 $\frac{1}{Y}$ 代表每項實際數值之倒數，則由相關綫求得每項之差應為

$$d = a + bX - \frac{1}{Y}$$

以 d 乘上式後，各式相加，得

$$\Sigma(d^2) = a\Sigma(d) + b\Sigma(dX) - \Sigma\left(\frac{d}{Y}\right)$$

因

$$\Sigma(d) = 0, \quad \Sigma(dX) = 0$$

故

$$\Sigma(d^2) = -\Sigma\left(\frac{d}{Y}\right)$$

更以 $\frac{1}{Y}$ 乘每項差數 d 之方程式，而後各式相加，得

$$\Sigma\left(\frac{d}{Y}\right) = a\Sigma\left(\frac{1}{Y}\right) + b\Sigma\left(\frac{X}{Y}\right) - \Sigma\left(\frac{1}{Y^2}\right)$$

將上式所求得 $\Sigma\left(\frac{d}{Y}\right)$ 之等值代入前式，得

$$\Sigma(d^2) = \Sigma\left(\frac{1}{Y^2}\right) - a\Sigma\left(\frac{1}{Y}\right) - b\Sigma\left(\frac{X}{Y}\right)$$

故 $S_{\frac{1}{y}}^2$ 之公式爲

$$S_{\frac{1}{y}}^2 = \frac{\Sigma\left(\frac{1}{Y}\right)^2 - a\Sigma\left(\frac{1}{Y}\right) - b\Sigma\left(\frac{X}{Y}\right)}{N}$$

將上式之數值代入下列相關指數之普遍公式

$$r^2 = 1 - \frac{S_{\frac{1}{y}}^2}{\sigma_{\frac{1}{y}}^2}$$

更將此式化簡，得

$$r^2_{\frac{1}{y}x} = \frac{a\Sigma\left(\frac{1}{Y}\right) + b\Sigma\left(\frac{X}{Y}\right) - Nc_{\frac{1}{y}}^2}{\Sigma\left(\frac{1}{Y}\right)^2 - Nc_{\frac{1}{y}}^2}$$

將相當數值代入 $S_{\frac{1}{y}}$ 及 $r_{\frac{1}{y}x}$ 之公式內，可得

$$S_{\frac{1}{y}} = .1191$$

$$r_{\frac{1}{y}x} = .766$$

Y 數值之倒數之標準差爲

$$\sigma_{\frac{1}{y}} = .1851$$

(以上各種量數旁俱以小字 $\frac{1}{y}$ 註明，所以取別於用真數或對數計算之量數也)。

倒數表示之估量標準誤之意義

以上求得各量數之值，其意義若何，應加討論。本章內以前所引用之方程式，均爲助吾人由一已知之 X 值以估量 Y 值之用。標準誤 $S_{\frac{1}{y}}$ 之功用在於測度此估量之可靠程度， $r_{\frac{1}{y}x}$ 之功用則在於測度兩變量之相關程度。惟此處各量數俱以倒數表示者，故由方程式所估量者爲 Y 值之倒數，標準誤所測度者爲迴歸綫兩旁倒數之散佈程度，而 r 之數值須視其他兩量數($S_{\frac{1}{y}}$ 與 $\sigma_{\frac{1}{y}}$)之關係而定，此兩量數亦係以倒數表示者也。

茲更舉一實例以說明各種量數之意義。如吾人已知某年燕麥之實

際產量爲其常態產量之150%，而欲求是年最可能之價格，可將該 X 值1.50代入方程式

$$\frac{1}{Y} = -.1357 + 1.1643X$$

由該式可得

$$\frac{1}{Y} = 1.6108$$

或

$$Y = .621$$

故由此估量所得該年燕麥之價格爲其常態價格之62%。如欲測度此估量之可靠程度，則得

$$S_{\frac{1}{Y}} = .1191$$

此值係對倒數而言，故應適用於以倒數表示之估量數值如下：

$$1.6108 + .1191 = 1.7299$$

$$1.6108 - .1191 = 1.4917$$

將上兩式所得之兩個倒數一律化爲真數，得.578與.670。故最可能之價格爲其常態價格之62.1%，而當迴歸綫兩旁倒數之分配近乎常態時，則實際價格在常態價格之57.8%至67.0%範圍內之機率爲68%。 $2S$ 及 $3S$ 之上下限可用同法將兩倍於.1191與三倍於.1191之數，在以倒數表示之估量數值內分別加減，所得之倒數再化爲真數即得。此處 Y 之估量區域，其寬度亦各段不同，正與用對數時無異，故用 $S_{\frac{1}{Y}}$ 測度估量之可靠程度時，須先求倒數表示之估量數值，次以 $S_{\frac{1}{Y}}$ 之值分別加減，而後可斷定估量區域之上下限。

由真數、對數及倒數求相關量數之比較

解釋 ρ 時，其理論與前相同。相關指數之數值本視迴歸綫兩旁之散佈程度與倚變量平均數兩旁之散佈程度比較而定。在用真數時，迴歸綫

兩旁之散佈程度係與倚變量算術平均數兩旁之散佈程度相比較，而此兩種散佈程度俱係用真數測度者（即 S_y 與 σ_y 之比）。在用對數時，迴歸綫兩旁之散佈程度則係與倚變量對數之算術平均數兩旁之散佈程度相比較，而此兩種散佈程度俱係用對數測度者。對數之差亦可用比率解釋之。由迴歸綫所求得對數之差，為實際數值對計算數值之比率；由對數之算術平均數所求得之差，為實際數值對幾何平均數之比率。 $\rho_{\log y}$ 之數值係由此兩種差數之關係而定，亦即 $S_{\log y}$ 與 $\sigma_{\log y}$ 之比也。

如配綫時所用數字為倚變量之倒數，則該綫兩旁之散佈程度亦須用倒數測度之。倚變量之標準差亦然，即 $\sigma_{\frac{1}{y}}$ 須由各實際數值之倒數對倒數之算術平均數之差計算而得。惟倒數之算術平均數即係倒數平均數之倒數。故簡言之，相關指數 $\rho_{\frac{1}{y}}$ 之數值須視迴歸綫兩旁散佈程度與倚變量倒數平均數兩旁散佈程度之關係（即 $S_{\frac{1}{y}}$ 與 $\sigma_{\frac{1}{y}}$ 之比）而定，而兩種散佈程度俱係用倒數測度者。

至是吾人知兩變量之關係，可用三大類曲綫表示之。此三大類曲綫為：

1. 用倚變量之真數所配合之綫，其方程式屬於下列形式：

$$Y = f(X)$$

2. 用倚變量之對數所配合之綫，其方程式屬於下列形式：

$$\log Y = f(X)$$

3. 用倚變量之倒數所配合之綫，其方程式屬於下列形式：

$$\frac{1}{Y} = f(X)$$

以上三種方程式中任何一種，其形式可為直綫，亦可為非直綫。就方程式之意義而論， X 之函數不拘何種，均可應用（前此所述計算 S 與 ρ 之方法，在應用上尚有若干限制，前已論及之矣）。

第一類曲綫之估量標準誤係用倚變量之原來單位求得者，故其意

義簡而易明。其相關指數 ρ 爲一絕對數值，所以測度倚變量之絕對散佈程度 (absolute variability) 之由迴歸綫各項差數計算而得者，較由算術平均數計算而得者，所可減少之程度者也。

第二類曲綫之估量標準誤係用對數求得者，解釋此量數時以用比率爲便。其相關指數 $\rho_{\log y}$ 所以測度倚變量之對數或比率散佈程度之由迴歸綫各項差數(或比率)計算而得者，較由幾何平均數計算而得者，所可減少之程度者也。

第三類曲綫之估量標準誤係用倒數求得者，故此量數亦爲倒數，其相關指數 $\rho_{\frac{1}{y}}$ 所以測度倚變量之倒數散佈程度之由迴歸綫各項差數計算而得者，較由倒數平均數計算而得者，所可減少之程度者也。

選擇相關量數之原則

由上所論，可知選擇曲綫以表示兩變量關係之前，必先根據各項基本原則，以確定何種平均數最能表示其集中趨勢；換言之，亦即斷定某數列用真數、對數及倒數所表示之三種分配之中，究以何者最與常態分配相近也。此點與曲綫之選擇及 S 與 P 之應用，皆有關係。

倘吾人注重於絕對數值，而倚變量之分配，繪於算術尺度之圖上，又近乎常態形式時，可用普通真數曲綫以表示兩變量之關係。但若吾人處理某種數列，注重於變動率，而不注重變動之絕對量，而數列之分配又合乎幾何法則時，則算術平均數及其他用真數計算之量數顯不適用。此時用對數曲綫似較真數曲綫爲優，而根據比率測量估度之可靠程度及相關程度亦較根據絕對值計算者更爲相宜。

倒數平均數之應用，不及算術平均數及幾何平均數之廣。在相關量數方面，應用此平均數之原則，此處可簡述之。大體言之，倒數平均數之弱點正與算術平均數相同，惟其差誤之方向彼此相反。在普通情形之下，幾何平均數常較前兩者爲優。但倒數平均數用於某種現象特爲相

宜，此時各項目均須以倒數表示，前述用倒數計算相關量數，即其例也。

計算相關量數用倒數平均數者，係假定倒數之分配合乎常態也；故用真數時，其分配必呈偏斜，而高於平均數之散佈範圍必遠較低於平均數者為廣。迴歸綫採用

$$\frac{1}{Y} = a + bX$$

形式時對於 Y 與 X 之關係亦嘗作同樣之假定，即 X 值以一絕對量增加時， Y 值隨之減少若干；而當 X 值以同一絕對量減少時， Y 值即隨之增加，但所增之量大於以前所減之量是也。在物品產量與價格兩變量間，以價格為倚變量時，含有類此之關係者，其例不少概見。例如產量以一絕對量增加時，價格隨之跌落若干；而當產量以同一絕對量減少時，則價格必隨之增漲，而所漲必遠過於前所跌落之數。在平均此種物品之價格時，以其倒數平均數代表全部價格，亦較其他各種平均數更有代表性(註)，故此時欲求相關量數，則相關綫自應根據倒數配合，而測度估量之確度，亦當根據倒數計算矣。

(註)“販買番芋者常不能瞭解經濟情形決定價格之原理。例如秋初所定價格過高，則番芋之消費必減，而農人及販賣者必不能將所有存貨悉按此高價脫售。番芋既不能由本年儲藏至翌年，則勢非減低售價至相當程度，使存貨在年內售盡不可。故在秋初應釐定一相當價格，使在全季之中，價格不致參差過甚，即欲取價貯存費用，亦僅能緩漸提高其價格，庶全季內之消費得維持比較穩定之狀態。否則季初售價特高，消費必銳減，迨季末非廉價脫售，即不能將存貨售罄，而必遭受虧損。

同理，如在秋初所定價格過低，則一般消費必劇增，漸致供不應求，而在季末貯存之少量存貨，其售價亦必特高。

季初價格過高或過低，皆非所宜，然則吾人將如何參酌其供求狀況，由此過高或過低之價格，以計算一平均價，作為釐定價格之標準，以調劑消費之情形乎？吾人知季初定價較低，以後必將過分提高以補償之，故用全季各月價格之算術平均數必嫌過高，非季初所應釐定之平均價，而須用各月價格之倒數平均數，始為相宜。”

上文錄自 University of Minnesota Agricultural Experiment Station 出版之 Technical Bulletin 第10號第8—10頁所載 Holbrook Working 氏著 Factors Determining the Price of Potatoes in St. Paul and Minneapolis.

用真數、對數及倒對表示各種量數之比較

由真數、對數及倒數求相關量數，其不同之處，可就同一資料配合該三類曲綫，比較其結果而可明瞭。前嘗就燕麥產量及價格之資料，配合第二第三兩類曲綫(即用對數及倒數配合者)，並說明其計算手續(表一一四)；茲就同一資料，配合一直綫(即用真數配合者)，並計算其相關量數。由該三類曲綫計算所得之結果，併列於下：

表 一 一 七

1881年至1913年燕麥產量與價格之關係

配合三類曲綫所得結果之比較

(在各例中價格均視為倚變量)

	方 程 式	估量標準誤	相關指數
A	$Y = 2.24 - 1.236X$	$S_y = .12$	$r = .783$
B	$\frac{1}{Y} = -.1357 + 1.1643X$	$S_{\frac{1}{y}} = .1191$	$\rho_{\frac{1}{y}} = .766$
C	$\text{Log } Y = -.00861 - 1.18206 \log X$	$S_{\log y} = .04901$	$\rho_{\log y} = .793$

上列三種標準誤尚不能根據其數值之大小，彼此比較，因其單位不同，僅第一種所用者為原來之單位也(實際產量對常態產量之比率)。下表所列為根據每一方程式，由五種產量估量所得之五種最可能之價格(此價格係以實際價格對常態價格之比率表示者)。表內每一估量值右方所列各數值，係表示由標準誤所定之上下限。表內並將各估量值加減 S 、 $2S$ 及 $3S$ 後所得之值一併列入，所以表示實際值在估量值兩旁可能之散佈程度者也。實際值對每一相關綫可能之差離，可由估量值與其上下限之實際差數測度之，此項差數列入 Δ 行內。表內各數值俱已化為原來單位(實際價格對常態價格之比率)，故可彼此比較。

表一八
根據燕麥產量與價格之三類相關綫方程式所得
估量價格及估量標準誤之比較

(1) X之數 值(實際 對產量 比率)	(2) 由數方 程式所 得Y之 估量值 (A)	(3) 根據真 數所得 估量值 之上下 限	(4) Δ	(5) 由倒數 方程式 所得Y 之估量 數值 (B)	(6) 根據倒 數所得 估量值 之上下 限	(7) Δ	(8) 由對數 方程式 所得Y 之估量 數值 (C)	(9) 根據對 數所得 估量值 之上下 限	(10) Δ
.5	1.622	+3S=1.982 +2S=1.862 +S=1.742 -S=1.502 -2S=1.382 -3S=1.262	+ .36 + .24 + .12 - .12 - .24 - .36	2.240	+3S=11.223 +2S= 4.803 +S= 3.055 -S= 1.768 -2S= 1.461 -3S= 1.244	+8.983 +2.563 + .815 - .472 - .779 - .996	2.224	+3S=3.114 +2S=2.780 +S=2.491 -S=1.979 -2S=1.779 -3S=1.679	+ .890 + .556 + .267 - .245 - .445 - .645
.8	1.251	+3S=1.611 +2S=1.491 +S=1.371 -S=1.131 -2S=1.011 -3S= .891	+ .36 + .24 + .12 - .12 - .24 - .36	1.257	+3S=2.281 +2S=1.794 +S=1.478 -S=1.093 -2S= .967 -3S= .867	+1.024 + .537 + .221 - .164 - .290 - .390	1.276	+3S=1.786 +2S=1.595 +S=1.429 -S=1.136 -2S=1.021 -3S= .906	+ .510 + .319 + .163 - .140 - .255 - .370
1.0	1.004	+3S=1.364 +2S=1.244 +S=1.124 -S= .884 -2S= .764 -3S= .644	+ .36 + .24 + .12 - .12 - .24 - .36	.972	+3S=1.490 +2S=1.265 +S=1.100 -S= .871 -2S= .789 -3S= .722	+ .518 + .293 + .128 - .101 - .183 - .250	.980	+3S=1.372 +2S=1.225 +S=1.098 -S= .872 -2S= .784 -3S= .696	+ .392 + .245 + .118 - .108 - .196 - .284
1.2	.757	+3S=1.117 +2S= .997 +S= .877 -S= .637 -2S= .517 -3S= .397	+ .36 + .24 + .12 - .12 - .24 - .36	.793	+3S=1.106 +2S= .977 +S= .875 -S= .724 -2S= .667 -3S= .618	+ .313 + .184 + .082 - .069 - .126 - .175	.790	+3S=1.106 +2S= .987 +S= .885 -S= .703 -2S= .632 -3S= .561	+ .316 + .197 + .095 - .087 - .158 - .229
1.5	.386	+3S= .746 +2S= .626 +S= .506 -S= .266 -2S= .146 -3S= .026	+ .36 + .24 + .12 - .12 - .24 - .36	.621	+3S= .798 +2S= .728 +S= .670 -S= .578 -2S= .541 -3S= .508	+ .177 + .107 + .049 - .043 - .080 - .113	.607	+3S= .852 +2S= .761 +S= .680 -S= .542 -2S= .484 -3S= .433	+ .245 + .154 + .073 - .065 - .123 - .174

估量區域及其意義

細觀上表，可明瞭由此三類方程式所得估量之性質。三類方程式不同之要點，不在於估量之數值，而在於所以測度此項估量可靠程度之標準誤。換言之，其不同之點乃在於吾人對各該曲綫兩旁散佈狀態所作之

種種假定也。

由真數所配之綫求得之量數 S_y , 不論估量數值或大或小, 其所示估量差誤之絕對距離, 在該綫各段均屬相同, 蓋係假定該綫兩旁真數之散佈合乎常態者也。每一估量數值為該值加 $1S$ (或 S 之倍數) 及該值減 $1S$ 後所得兩數值之算術平均數。凡此種種均可由圖八十一見之。圖內將原來各點與其相關直綫 (即用真數所配之綫)、及以直綫為中心寬度為 $2S, 4S, 6S$ 之估量區域, 一併繪入。

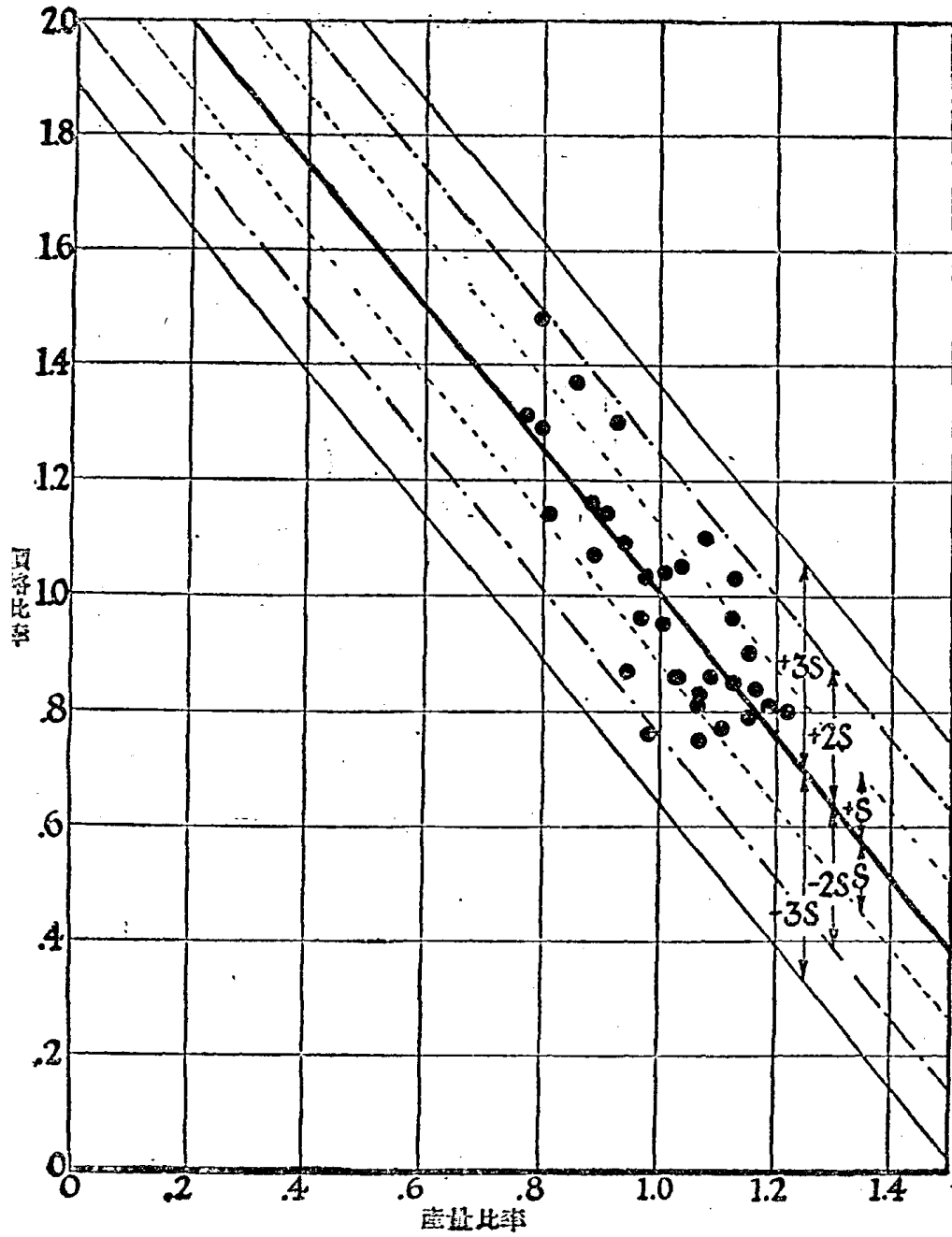
量數 $S_{\log y}$ 所示估量差誤之相對或百分距離, 不論估量數值或大或小, 在相關綫各段亦皆相同, 但此估量差誤之絕對距離則隨估量數值而有不同。估量數值大則絕對距離亦大, 估量數值小則絕對距離亦小。此處係假定相關綫兩旁對數之散佈合乎常態者也。每一估量數值為該值加 $S_{\log y}$ (或 $S_{\log y}$ 之倍數) 及該值減 $S_{\log y}$ 後所得兩數值之幾何平均數。凡此種種均可由圖八十二見之。圖內將原來各點與方程式

$$Y = .9804X^{-1.18206}$$

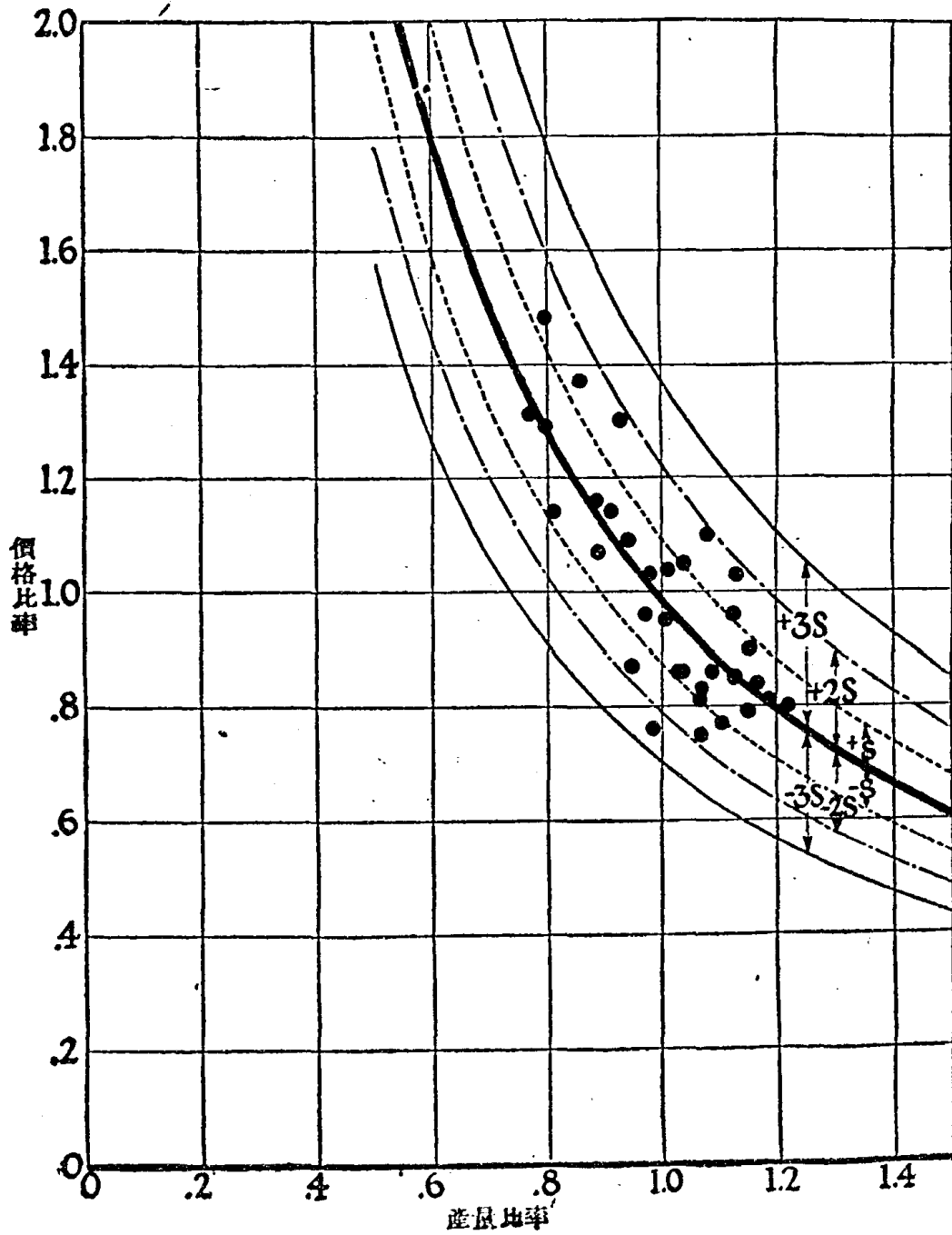
之相關綫、及以該綫作中心(幾何的)寬度為 $2S_r, 4S_r, 6S_r$ 之估量區域, 一併繪入。以圖八十一及圖八十二相較, 則根據真數散佈合乎常態與對數散佈合乎常態之兩種假定所得估量不同之點, 更易明瞭矣。

圖八十三係用雙對數尺度所繪圖八十二內之各點及其相關綫與估量區域。在該圖內相關綫已變為直綫, 而估量區域亦變為對稱, 故其寬度各段相同。以同一資料繪於對數尺度上圖形之改觀, 足使吾人明瞭以對數數值為估量之根據時所作假定之簡單一斑矣。

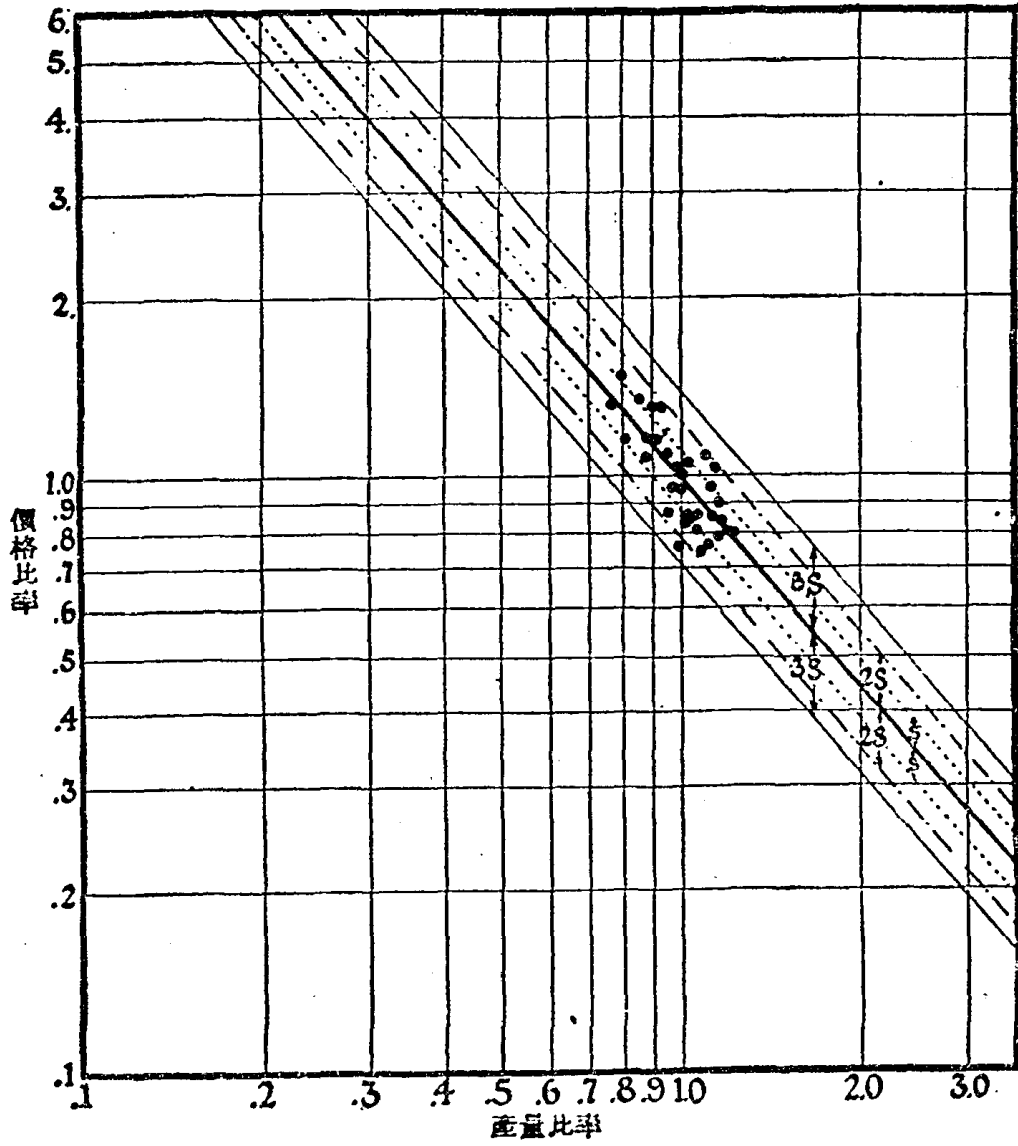
應用量數 $S_{\frac{1}{y}}$ 時, 吾人又須假定相關綫兩旁之散佈程度隨價格之高低而異, 價格愈高, 散佈愈大。當估量之價格低時, 其差誤之距離必甚小, 估量之價格高時, 其差誤之距離必甚大, 蓋係假定相關綫兩旁倒數之散佈合乎常態者也。每一估量數值常為該值加 $S_{\frac{1}{y}}$ (或 $S_{\frac{1}{y}}$ 之倍數) 及該值減 $S_{\frac{1}{y}}$ 後所得兩數值之倒數平均數。



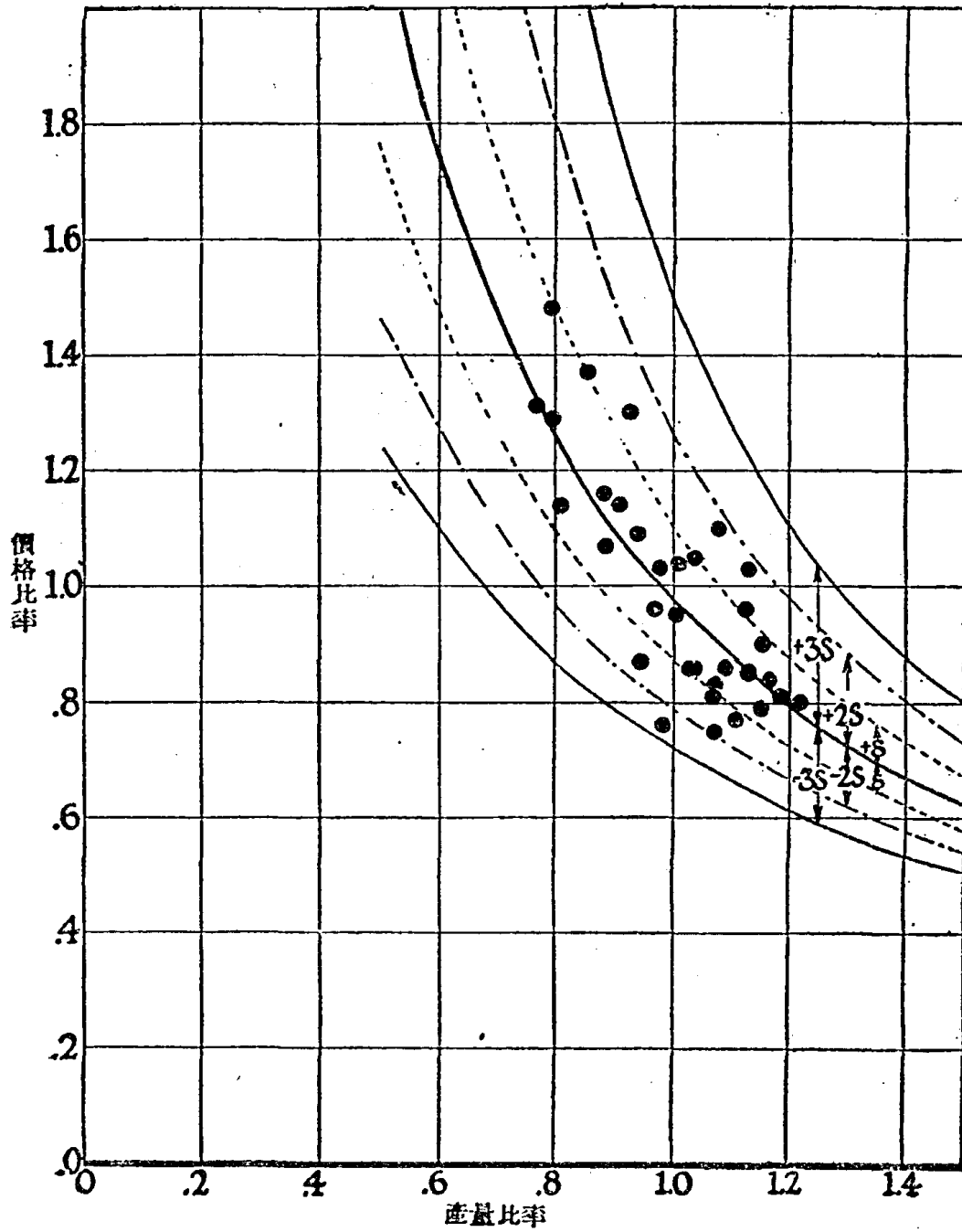
圖八十一. 表示燕麥產量與價格兩者關係之真數迴歸綫及真數估量區域



圖八十二 表示燕麥產量與價格兩者關係之對數迴歸綫及對數估量區域



圖八十三. 表示燕麥產量與價格兩者關係之對數迴歸綫及對數估量區域
(繪於雙對數尺度之紙上者)



圖八十四. 表示燕麥產量與價格兩者關係之倒數迴歸綫及倒數估量區域

圖八十四所示為原來各點與 $\frac{1}{Y} = -.1357 + 1.1643X$ 之相關綫，及以該綫作中心（倒數的）寬度為 $2S_{\frac{1}{Y}}$, $4S_{\frac{1}{Y}}$, 及 $6S_{\frac{1}{Y}}$ 之估量區域。此圖與前兩圖之差異甚為顯著，尤以估量區域為甚。假定相關綫兩旁之倒數分配合乎常態，則在最寬之估量區域內（其寬度為 $6S$ ）應包含全體點數 99.7%，最狹之估量區域內（其寬度為 $2S$ ）應包含全體點數 68%。倘不用各項數值之真數而用其倒數繪於圖上，則此極不對稱之分配亦可化為對稱形式，如對數數值之繪於對數圖上者然。

無論估量數值之或大或小，對數標準誤 $S_{\log y}$ 之值必介乎真數標準誤 S_y 與倒數標準誤 $S_{\frac{1}{y}}$ 之間。後兩者雖各有其特殊功用，而適合於特殊情形，但在採用此種方法以分析經濟資料時，根據對數計算之相關量數，似較根據真數及倒數計算者更合於一般用途，蓋比率之變動常較絕對數之變動更為重要，故估量時自以根據方程式

$$\log Y = f(X)$$

較為合理。採用此方程式時，估量之可靠程度應以 $S_{\log y}$ 測度之，相關指數則以 $\rho_{\log y}$ 測度之。 $\rho_{\log y}$ 之數值，係視相關綫兩旁比率之散佈程度與幾何平均數兩旁比率之散佈程度兩者比較而定（註）。

（註）關於此點之理論，C. M. Walsh 氏在其所著“*The Problem of Estimation*”（一九二一年倫敦 King 書局出版）一書內，曾有指示，頗可適用於本題。此外 Galileo 氏亦贊同採用幾何平均數以計算估量值之平均。氏嘗引 Walsh 氏原著一段（第12頁）如下：“是以差誤應就估量數值與實際數值之比率測度之，而不應根據絕對數量測度之。吾人測度差誤決不可用磅、呎或貨幣單位等絕對數量表示，而須以估量所得之磅、呎或貨幣數對實際數量之磅、呎或貨幣數之比率表示。”此種論據與本書前文所述吾人應採用對數函數作為估量根據及應用對數以計算估量標準誤之說，不謀而合，而其持論則更為有力矣。

參 考 書

KILLOUGH, H. B. *A Statistical Analysis of Oat Prices*. A manuscript prepared for the U. S. Bureau of Agricultural Economics.

MOORE, H. L. *Elasticity of Demand and Flexibility of Prices*. Journal of the American Statistical Association, March, 1922. *Empirical Laws of Demand and Supply and the Flexibility of Prices*. Political Science Quarterly, Dec., 1919.

WALSH, C. M. *The Problem of Estimation* (1—67).

Examples of curves employed to describe relationships of the type discussed in this chapter will be found in

SCHULTZ, HENRY. *The Statistical Measurement of the Elasticity of Demand for Beef*. Journal of Farm Economics, July, 1924 (254—278).

WORKING, HOLBROOK. *Factors Determining the Price of Potatoes in St. Paul and Minneapolis*. Technical Bulletin 10, University of Minnesota Agricultural Experiment Station.

第十四章 關係之測度：複相關及淨相關

以前各章討論相關問題僅限於兩個變量，一為倚變量，一為自變量。在所舉各例中，嘗發現兩變量相關之程度甚高者。但就一般而論，各種經濟現象關係複雜，變量之變動不僅受另一變量之影響，而常受許多其他變量之共同影響。前此測度兩變量之相關，即測度一倚變量與一自變量之相關，而忽視此倚變量與其他自變量之相關者，乃假定一個單獨自變量為此倚變量變動之主要因素也(註)。如前述香草一例中，吾人係假定香草之收穫量僅受灌溉水量多寡之影響，實則雨量之多寡，溫度之高低，皆與收穫量有關。在經濟現象中，任何變量變動之因素不一而足，吾人分析此種現象，欲求詳盡，自非應用各種方法，以同時研究兩種以上變量之關係不可。此時吾人所需者，為測度一變量所受各種因素共同影響之各種工具，而此工具可由以前所述各種方法中推求而得。

表一一九所列，為一八九〇年至一九二二年堪薩斯州每英畝玉蜀黍之收穫及每年六、七、八三個月之平均溫度。收穫量之數字可配以直綫趨勢綫，其實際收穫量對常態收穫量之百分比，見第(4)行。

玉蜀黍收穫量與溫度之關係；初步分析

玉蜀黍之收穫常受生長時期內溫度高低之影響。現時研究之目的，在於決定玉蜀黍收穫量與六、七、八三個月溫度之關係，以便根據此種關係，由已知之溫度以預測收穫量之多寡；但溫度影響於收穫量之程度，各月不同，故吾人可先確定各月溫度與收穫量之關係。

每英畝收穫量與六月溫度之關係，可用下列形式之方程式表示之。

$$X_1 = a + b_{12}X_2$$

每英畝收穫量與七月溫度之關係，可用下列形式之方程式表示之。

(註)此非謂相關係數係測度兩變量之因果關係者。關於此點可參看第十六章。

表 一 一 九

1890年至1922年堪薩斯州玉蜀黍之收穫量及平均溫度(註一)

(1) 年 份	(2) 每英畝實際 收穫量 (蒲式耳)	(3) 每英畝常態 收穫量 (蒲式耳) (註二)	(4) 實際收穫量 對常態收穫 量之百分比 X_1	(5) 六月之平 均溫度 X_2	(6) 七月之平 均溫度 X_3	(7) 八月之平 均溫度 X_4
1890	15.6	22.4	69.6	77.6	83.1	76.1
1891	20.7	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	79.5	75.7
1911	14.5	19.2	75.5	80.5	78.6	76.4
1912	23.0	19.1	120.4	69.3	79.9	77.4
1913	3.2	18.9	16.9	74.2	82.1	84.2
1914	18.6	18.8	98.4	78.2	79.9	78.2
1915	31.0	18.6	166.7	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
1918	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

(註一)表內玉蜀黍產量數字轉載美國農業部(U. S. Department of Agriculture)出版

之叢刊第515種及各年年鑑,溫度係根據美國氣象研究局(U. S. Weather Bureau)

之報告。

(註二)長期趨勢綫之方程式為 $Y = 19.967 - .1502X$,其原點在1906年。

$$X_1 = a + b_{1,2} X_2$$

(上兩式中 X_1 代表實際收穫量對常態收穫量之百分比， X_2, X_3 等代表華氏寒暑表度數之絕對數值)。此處不用 Y 及 X 作為變量之符號，而以 X_1 代表倚變量， X_2, X_3 代表自變量。第一式中代表直綫斜度之常數(迴歸係數)為 $b_{1,2}$ 。 b 旁1及2兩小字所以表示其所指之變量，其第一字係指倚變量(即本例中之 X_1)，第二字指自變量(即本例中之 X_2)。方程式所含變量在兩個以上時，表示迴歸綫斜度之常數即不止一個，故應有此區別也。至於此項常數之意義，與以前方程式中僅含兩變量而無須加註小字之常數完全相同。

將上列表示每英畝收穫量(實際收穫量對常態收穫量之百分比)與六月溫度兩者平均關係之直綫方程式所用之常態方程式解之，即可求得常數之數值。以常數之數值代入直綫方程式中，可得

$$X_1 = 522.31 - 5.743X_2$$

$S_{1,2}$ 之數值可用下式求之

$$S_{1,2}^2 = \frac{\Sigma(X_1^2) - a \Sigma(X_1) - b_{1,2} \Sigma(X_1 X_2)}{N}$$

(上式內 S 旁及後列公式內 r 旁之兩小字，其意義與 b 旁所註者相同。)

將已知各數值代入上式，可解得

$$S_{1,2}^2 = 913.178$$

故

$$S_{1,2} = 30.22$$

標準誤 S 為測度以相關方程式為根據所得估量可靠程度之量數，其義意前已言之詳矣。欲判別所得相關方程式之有無意義，須比較 $S_{1,2}$ 與 σ_1 (X_1 之標準差)兩數值之大小而定。 σ_1 可視為測度以 X_1 變量算術平均數為根據所得估量之可靠程度之量數。此處 σ_1 之數值為

$$\sigma_1 = 34.477$$

該值大於 $S_{1,2}$ 之數值，可見根據方程式所得之估量實較根據算術平均

數所得者為可靠。兩者之關係可用抽象數字以相關係數 r 表示之。 r 之數值可由下式求得：

$$r_{12} = \frac{a \Sigma(X_1) + b_{12} \Sigma(X_1 X_2) - N c_1^2}{\Sigma(X_1^2) - N c_1^2}$$

將相當數值代入上式以求 r ，並於 r 數值之前加註與 b_{12} 相同之正負號，得

$$r_{12} = -.4814$$

此值表示堪薩斯州每英畝玉蜀黍收穫量與六月溫度為一負相關，但相關之程度不高。吾人可改用七月之溫度作為估量之根據，以測驗估量之準確程度能否增高。

解收穫量與七月溫度兩者關係之直綫常態方程式時所需之各數值，可由表一一九得之。常數之數值求得後，即得下列方程式

$$X_1 = 827.64 - 9.302X_2$$

標準誤之數值為

$$S_{13} = 24.73$$

相關係數之數值為

$$r_{13} = -.6968$$

此處相關程度較高，故用七月溫度作為估量之根據，所得結果必較用六月溫度者為可靠。

用同法計算收穫量與八月溫度之各種相關量數，得

$$X_1 = 517.86 - 6.098X_4$$

$$S_{14} = 29.98$$

$$r_{14} = -.4937$$

由是知八月之溫度亦可影響於玉蜀黍之收穫量，溫度低則收穫增加。其相關程度雖不及七月溫度之密切，但相關之存在要無疑義。吾人現時所需者，即為彙合三個月之溫度，而同時根據此三個月溫度以估

量收穫量之方法。惟吾人斷不可由此三個月之溫度以求其平均溫度，因七月溫度顯較其他兩月為重要也。茲所應用之方法，其原理亦尚簡單，請論述於下。

由三個自變量估量玉蜀黍收穫量之方法

此處所用估量方程式(即迴歸方程式)為含有一個倚變量(玉蜀黍收穫量)及三個自變量之方程式。其形式如下：

$$X_1 = a + b_{1.2.34}X_2 + b_{1.3.24}X_3 + b_{1.4.23}X_4$$

吾人如能求得上式中四個常數之數值，則將已知 X_2, X_3, X_4 之數值代入式中，即可求得 X_1 之估量數，其手續與前此用兩個變量時相同。常數之數值可用最小平方法求之。

上列方程式因應用各種符號關係，外表似頗複雜，實際甚為簡單，惟此項符號須加以解釋耳。在研究 X_1 及 X_2 兩變量間之關係時， $b_{1.2}$ 之符號為 X_1 倚 X_2 而變動之迴歸係數(即 X_1 為倚變量時，兩變量相關綫之斜度)，此點前曾述及。 $b_{1.2.34}$ 之符號則係代表 X_1 倚 X_2 而變動之淨迴歸係數(coefficient of net regression)。 b 旁所註各小字，在點之右方3及4兩字表示除 X_1 及 X_2 外，曾將 X_3 及 X_4 兩變量加入研究，惟在求 $b_{1.2.34}$ 之數值時，已將 X_3 及 X_4 之影響消除，故此數值可視為用 X_2, X_3, X_4 三個自變量以估量 X_1 之數值時 X_2 變量所應得之權數。 $b_{1.2.34}$ 之數值與 $b_{1.2}$ 不同，因後者乃僅用 X_2 以估量 X_1 時 X_2 之權數也。同理常數 $b_{1.3.24}$ 為 X_1 倚 X_3 而變動之淨迴歸係數，係用 X_2, X_3, X_4 三個自變量以估量 X_1 時 X_3 所應得之權數。每個係數係代表一單獨而又簡單之常數，其右旁所註小字係表明該常數之正確意義者也。點之左方兩小字稱為主要註字(primary subscripts)，點之右方兩小字稱為次要註字(secondary subscripts)

常態方程式之形式及其解法

現時初步分析工作(註一)在於求得常態方程式，以便解出估量方程式內所含各常數之數值。依照普通手續(註二)可得下列四個常態方程式：

$$I \quad \Sigma(X_1) = Na + b_{12.34} \Sigma(X_2) + b_{13.24} \Sigma(X_3) + b_{14.23} \Sigma(X_4)$$

$$II \quad \Sigma(X_1 X_2) = a \Sigma(X_2) + b_{12.34} \Sigma(X_2^2) + b_{12.24} \Sigma(X_2 X_3) \\ + b_{14.23} \Sigma(X_2 X_4)$$

$$III \quad \Sigma(X_1 X_3) = a \Sigma(X_3) + b_{12.34} \Sigma(X_2 X_3) + b_{13.24} \Sigma(X_3^2) \\ + b_{14.23} \Sigma(X_3 X_4)$$

$$IV \quad \Sigma(X_1 X_4) = a \Sigma(X_4) + b_{12.34} \Sigma(X_2 X_4) + b_{13.24} \Sigma(X_3 X_4) \\ + b_{14.23} \Sigma(X_4^2)$$

將已知各數值代入上列聯立方程式，即可求得四個常數之數值。惟吾人可將上列四式減為三式，以節省計算手續，蓋若每一變量之各項不用絕對值，而用各項對其算術平均數之差，則估量方程式中之常數 a 即可消除矣。

今設以 A_1, A_2, A_3 等代表各變量之算術平均數， x_1, x_2, x_3 等代表各項對其算術平均數之差，則 X_1, X_2, X_3 等絕對數值可用其相等數值 $x_1 + A_1, x_2 + A_2, x_3 + A_3$ 等代之。將各數值代入四個常態方程式，並用代數方法簡約之，則第一式即可消除，其餘三式化為：

$$\frac{\Sigma(x_1 x_2)}{N} = \frac{\Sigma(x_2^2)}{N} b_{12.34} + \frac{\Sigma(x_2 x_3)}{N} b_{13.24} + \frac{\Sigma(x_2 x_4)}{N} b_{14.23}$$

(註一)此處研究複相關問題所取之初步分析，係依照 H. R. Tolley 暨 M. J. B. Ezekiel 兩氏在其所著 "A Method of Handling Multiple Correlation Problems" 一文中之方法。原文具載於 1923 年十二月出版之 "Journal of the American Statistical Association" 第 993—1003 頁。兩氏闡明此項簡單而有效之方法，對於統計學者貢獻頗大。關於常態方程式之逐步化簡手續，此處不具述，讀者可參看兩氏之原文。

(註二)此項手續在附錄 A 內有討論，可參看。

$$\frac{\Sigma(x_1x_2)}{N} = \frac{\Sigma(x_2x_3)}{N} b_{12.34} + \frac{\Sigma(x_3^2)}{N} b_{13.24} + \frac{\Sigma(x_3x_4)}{N} b_{14.23}$$

$$\frac{\Sigma(x_1x_3)}{N} = \frac{\Sigma(x_2x_3)}{N} b_{12.34} + \frac{\Sigma(x_3x_4)}{N} b_{13.24} + \frac{\Sigma(x_4^2)}{N} b_{14.23}$$

上列各式內每一變量之各項均指各數值對其算術平均數之差，故 $\frac{\Sigma(x_1x_2)}{N}$ 即為變量 x_1 及 x_2 之均積數 (mean product)， $\frac{\Sigma(x_3^2)}{N}$ 為 σ_3^2 ，其餘各項依此類推。以 p_{12}, p_{13} 等代表各均積數，同時將各標準差平方之符號一併代入上列三式，則得常態方程式如下：

$$p_{12} = \sigma_2^2 b_{12.34} + p_{23} b_{13.24} + p_{24} b_{14.23}$$

$$p_{13} = p_{23} b_{12.34} + \sigma_3^2 b_{13.24} + p_{34} b_{14.23}$$

$$p_{14} = p_{24} b_{12.34} + p_{34} b_{13.24} + \sigma_4^2 b_{14.23}$$

此為求常數所用各種形式之常態方程式中，計算最為便利者。由表一一九所列之資料，可得下列各數值：

$$\Sigma(X_1) = 3298.1 \qquad \Sigma(X_1^2) = 368846.67$$

$$\Sigma(X_2) = 2426.9 \qquad \Sigma(X_2^2) = 178755.75$$

$$\Sigma(X_3) = 2581.5 \qquad \Sigma(X_3^2) = 202163.79$$

$$\Sigma(X_4) = 2553.8 \qquad \Sigma(X_4^2) = 197890.32$$

$$\Sigma(X_1X_2) = 240967.22$$

$$\Sigma(X_1X_3) = 255954.11$$

$$\Sigma(X_1X_4) = 253664.85$$

$$\Sigma(X_2X_3) = 189941.83$$

$$\Sigma(X_2X_4) = 187909.38$$

$$\Sigma(X_3X_4) = 199845.00$$

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

$$= 99.9424$$

$$c_1^2 = 9988.4833$$

$$c_2 = 73.5424$$

$$c_2^2 = 5408.4846$$

$$c_3 = 78.2273$$

$$c_3^2 = 6119.5105$$

$$c_4 = 77.3879$$

$$c_4^2 = 5988.8871$$

由上列各數值，即可進而計算解常態方程式時所需之各數值如下：

$$\begin{aligned} \sigma_1^2 &= \frac{\Sigma(X_1^2)}{N} - c_1^2 \\ &= \frac{368,846.67}{33} - 9988.4833 = 1188.688 \end{aligned}$$

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

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

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

$$\rho_{12} = \frac{\Sigma(X_1 X_2)}{N} - c_1 c_2$$

$$\rho_{12} = \frac{240,967.22}{33} - (99.9424 \times 73.5424) = -47.967$$

$$\rho_{13} = \frac{255,954.11}{33} - (99.9424 \times 78.2273) = -62.039$$

$$\rho_{14} = \frac{253,664.85}{33} - (99.9424 \times 77.3879) = -47.519$$

$$\rho_{23} = \frac{189,941.83}{33} - (73.5424 \times 78.2273) = 2.790$$

$$\rho_{24} = \frac{187,909.38}{33} - (73.5424 \times 77.3879) = 2.932$$

$$\rho_{34} = \frac{199,845.00}{33} - (78.2273 \times 77.3879) = 2.063$$

將以上各數值代入常態方程式得

$$-47.967 = 8.3564b_{12.34} + 2.790b_{13.24} + 2.932b_{14.23}$$

$$-62.039 = 2.790b_{12.34} + 6.6645b_{13.24} + 2.063b_{14.23}$$

$$-47.519 = 2.932b_{12.34} + 2.063b_{13.24} + 7.7893b_{14.23}$$

解此三個聯立方程式(註一)，可得各常數之值如下：

$$b_{1.2.34} = -2.095 \quad b_{1.3.24} = -7.394 \quad b_{1.4.23} = -3.354$$

故本例所求之方程式為

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

此為 x_1 倚 x_2, x_3, x_4 之淨迴歸方程式。將已知三個自變量(六月溫度、七月溫度、八月溫度)之數值代入式中，即可斷定倚變量(每英畝玉蜀黍收穫量)最可能之數值。惟在此方程式中各變量皆以原來數值對其算術平均數之離中差表示者，為實用上便利起見，此式有改以原來數值表示之必要；換言之，即吾人應將原點由平均數點移至原來尺度上之零點是也。但原點經此移動後，則前已消除之常數 a 又須列入方程式中矣。

a 之數值可由下式求之(註二)：

$$A_1 = a + A_2 b_{1.2.34} + A_3 b_{1.3.24} + A_4 b_{1.4.23}$$

式中之 A_1, A_2, A_3, A_4 為各變量之算術平均數。以相當數值代入上式，可得

$$\begin{aligned} 99.9424 &= a + (73.5424 \times -2.094816) \\ &+ (78.2273 \times -7.39374) + (77.3879 \times -3.353796) \quad (\text{註三}) \end{aligned}$$

(註一)解此聯立方程式可用任何方法，但若聯立方程式共有三個或三個以上，似以採用 Doolittle 氏之方法最為便利。此法在附錄A內有詳細之說明。

(註二)此式係由第416頁所載第一常態方程式

$$\Sigma(X_1) = Na + b_{1.2.34}\Sigma(X_2) + b_{1.3.24}\Sigma(X_3) + b_{1.4.23}\Sigma(X_4)$$

求得者。將 X_1, X_2 等絕對數值以其相等數值 $x_1 + A_1, x_2 + A_2$ 等代入，則得

$$\begin{aligned} \Sigma(x_1) + NA_1 &= Na + b_{1.2.34}[\Sigma(x_2) + NA_2] + b_{1.3.24}[\Sigma(x_3) + NA_3] \\ &+ b_{1.4.23}[\Sigma(x_4) + NA_4] \end{aligned}$$

因式中 $\Sigma(x_1) = 0, \Sigma(x_2) = 0, \Sigma(x_3) = 0, \Sigma(x_4) = 0$ ，故此四項完全消除，再以 N 除之，即得此式。

(註三)此處假定原點係在原來尺度上之零點，故 $A_1 = c_1, A_2 = c_2$ ，餘類推。式中 $b_{1.2.34}, b_{1.3.24}$ 等所用小數之位數較迴歸方程式中為多，欲使所求 a 之數值更為準確也。

解上式,得

$$a = 1091.94$$

故以原來數值表示之迴歸方程式為

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

估量標準誤之計算

由上式所得之估量果較前此用一個自變量之方程式所得之估量為可靠乎?欲解答此問題,須計算此式之標準誤。此處標準誤可用 $S_{1.234}$ 代表之, S 旁所註小字係表示一個單獨之倚變量 (X_1) 與三個自變量。 $S_{1.234}$ 之數值可由下式求之(註)。

$$S^2_{1.234} = \sigma_1^2 - b_{12.34}p_{12} - b_{13.24}p_{13} - b_{14.23}p_{14}$$

(註)此式證明如下:如已知方程式之形式為

$$x_1 = b_{12.34}x_2 + b_{13.24}x_3 + b_{14.23}x_4$$

(式中各變量均為各項對其平均數之離中差)則每項之實際離中差與根據此式計算所得離中差之差數,可用下列差數方程式表之。

$$d = b_{12.34}x_2 + b_{13.24}x_3 + b_{14.23}x_4 - x_1 \dots \dots \dots (1)$$

以 d 乘上式內各項,而後各式相加,得

$$\Sigma(d^2) = b_{12.34}\Sigma(dx_2) + b_{13.24}\Sigma(dx_3) + b_{14.23}\Sigma(dx_4) - \Sigma(dx_1)$$

但該迴歸綫係用最小平方法所配合,故

$$\Sigma(dx_2) = 0$$

$$\Sigma(dx_3) = 0$$

$$\Sigma(dx_4) = 0$$

故

$$\Sigma(d^2) = -\Sigma(dx_1) \dots \dots \dots (2)$$

以 x_1 乘差數方程式(1)內各項,而後各式相加,得

$$\Sigma(dx_1) = b_{12.34}\Sigma(x_1x_2) + b_{13.24}\Sigma(x_1x_3) + b_{14.23}(\Sigma x_1x_4) - \Sigma(x_1^2)$$

將(2)式內 $\Sigma(dx_1)$ 之等值代入上式,可得

$$\Sigma(d^2) = \Sigma(x_1^2) - b_{12.34}\Sigma(x_1x_2) - b_{13.24}\Sigma(x_1x_3) - b_{14.23}\Sigma(x_1x_4)$$

將相當數值代入上式，得

$$S^2_{1.234} = 1188.688 - 100.482 - 458.700 - 159.369$$

$$= 470.137$$

$$S_{1.234} = 21.68$$

此量數之意義與以前所用之標準誤完全相同。根據 X_1 之平均數求得之估量，其可靠程度係用 σ_1 測度之。 σ_1 之數值在此為34.477。以收穫量作為六、七、八三個月溫度之函數，而根據其淨迴歸方程式所得之估量，其可靠程度係用 $S_{1.234}$ 測度之。 $S_{1.234}$ 之數值在此為21.68。由是知根據方程式求得之估量確較根據 X_1 之平均數所得者為可靠。吾人在此雖未能將所有影響玉蜀黍收穫量之一切因素完全包括在內，但已將三種因素所影響於收穫量之程度確切表示之矣。

複相關係數

茲進而討論複相關 (multiple correlation) 之相關係數。吾人已知此量數之數值須視 S 與 σ 之關係而定。在複相關中其數值可由下式求之：

$$R^2_{1.234} = 1 - \frac{S^2_{1.234}}{\sigma_1^2}$$

此處所討論者為一個倚變量與數個自變量間之關係，故此量數稱為複

(續上頁附註)

$$S^2_{1.234} = \frac{\Sigma(d_2)}{N} = \frac{\Sigma(x_1^2)}{N} - b_{12.34} \frac{\Sigma(x_1x_2)}{N} - b_{13.24} \frac{\Sigma(x_1x_3)}{N} - b_{14.23} \frac{\Sigma(x_1x_4)}{N}$$

但變量之數值為各項對其平均數之離中差，故

$$S^2_{1.234} = \sigma_1^2 - b_{12.34}p_{12} - b_{13.24}p_{13} - b_{14.23}p_{14}$$

標準誤亦可用原來尺度上零點為原點之相關綫普遍方程式求之。本例中因含有一個倚變量與三個自變量，故普遍方程式為

$$\Sigma(X_1^2) - a\Sigma(X_1) - b_{12.34} \dots n\Sigma(X_1X_2)$$

$$S^2_{1.234 \dots n} = \frac{-b_{13.24} \dots n\Sigma(X_1X_3) - b_{14.23} \dots n\Sigma(X_1X_4) - \dots}{N}$$

此式之證明見附錄A。

相關係數，而以符號 R 代表之。 R 旁所註小字，在點之左方者表示倚變量，在點之右方者表示自變量。將 $S^2_{1.234}$ 之等值代入此式，得

$$R^2_{1.234} = 1 - \frac{\sigma_1^2 - b_{12.34}p_{12} - b_{13.24}p_{13} - b_{14.23}p_{14}}{\sigma_1^2}$$

此式可化爲(註)

$$R^2_{1.234} = \frac{b_{12.34}p_{12} + b_{13.24}p_{13} + b_{14.23}p_{14}}{\sigma_1^2}$$

將相當數值代入上式，可得

$$R^2_{1.234} = \frac{100.482 + 458.700 + 159.369}{1188.688}$$

$$R_{1.234} = .778$$

複相關係數者，乃測度一個倚變量同時與數個自變量間相關程度之指數也。當各個自變量聯合變動時，倚變量必隨之變動，其相關程度即可用此複相關係數測度之。吾人倘將此數個自變量視爲一個獨立數列(independent series)，則複相關係數之意義更易明瞭。如此則複相關係數即爲測度一個倚變量與一個獨立數列間相關程度之量數，其意義與兩變量間之相關係數無異。在複相關中，此獨立數列雖係由數個自變量所構成，但相關係數之基本意義並不因此而有所變更。 R 之前無正號與負號之別。在本例中各個自變量與玉蜀黍收穫量之關係皆爲負相關，故 R 之前亦可加一負號，惟各個自變量與倚變量間之關係有爲正相關，有爲負相關，故 R 之前常不加正負號。至於每一自變量與倚變量相關之爲正爲負，可由淨迴歸方程式中各自變量之常數所附之符號定之。

(註)複相關係數亦可由以原來尺度上零點爲原點之普遍公式求之。其式爲：

$$R^2_{1.234\dots n} = \frac{a\Sigma(X_1) + b_{12.34\dots n}\Sigma(X_1X_2) + b_{13.24\dots n}\Sigma(X_1X_3) + b_{14.23\dots n}\Sigma(X_1X_4) + \dots - Nc_1^2}{\Sigma(X_1)^2 - Nc_1^2}$$

複相關量數之比較

吾人已知同時根據數個因素以估量玉蜀黍之收穫，則所得估量愈為可靠，而於收穫量增減之原因亦可獲得確切之認識，此點可由比較前所求得各量數之數值而益信。

表 一 二 〇

堪薩斯州玉蜀黍收穫量 各種相關量數之比較

估量之根據	測度估量可靠程度之量數	相關係數
X_1 之算術平均數	$\sigma_1 = 34.477$	
$X_1 = 522.31 - 5.743X_2$	$S_{12} = 30.22$	$r_{12} = -.4814$
$X_1 = 827.64 - 9.302X_3$	$S_{13} = 24.73$	$r_{13} = -.6968$
$X_1 = 517.86 - 6.098X_4$	$S_{14} = 29.98$	$r_{14} = -.4937$
$X_1 = 1091.94 - 2.095X_2 - 7.394X_3 - 3.354X_4$	$S_{1,234} = 21.68$	$R_{1,234} = .778$

計算相關量數時，倘將玉蜀黍生長時期內之雨量等其他因素一併列入，則 S 之數值尚可減小，而 R 之數值尚可增大。其計算方法可由前述方法伸引而得，惟因素每增加一項，則聯立方程式亦須增多一式耳。

計算複相關方法與直綫相關之關係

以上討論複相關問題時，尚有一要點未加提示，即該計算複相關之方法必在每對變量間之關係為一直綫關係時方可適用是也。在用四個變量時，可有六種配法而成六對變量（即可求得六種均積數），倘六對變量中任何一對之關係與直綫相差甚遠，則由此所得之相關量數，其意義自較薄弱。如仍用前述估量方程式雖無何種錯誤，但用此式作為估量之根據，不及表示真實關係之方程式為優耳。此時由 S 及 R 之數值亦可推知根據該式所得之估量為不可靠矣（註）。

（註）1924年美國統計學社月刊（“Journal of the American Statistical Association”）第19卷第148號內載有 M. J. B. Ezekiel 氏所著 “A Method of Handling Curvilinear Correlation for any Number of Variables”。此文專論各變量間之關係為非直綫時處理複相關之方式。

計算複相關方法之應用

吾人可舉一實例以說明估量方程式之應用。堪薩斯州一九二二年六月之平均溫度為華氏75.2度，七月為77度，八月為80.1度；如欲求是年每英畝玉蜀黍最可能之收穫量，則可將各月溫度代入下式

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

內之 X_2, X_3, X_4 ，即得

$$X_1 = 1091.94 - (2.095 \times 75.2) - (7.394 \times 77.0) \\ - (3.354 \times 80.1)$$

$$X_1 = 96.4$$

此值為實際收穫量對常態收穫量之百分比。玉蜀黍收穫量長期趨勢綫之方程式為

$$Y = 19.967 - .1502X$$

(式中之 Y 代表收穫量， X 代表時間，其原點在一九〇六年)。由此式可得一九二二年之常態收穫量為每英畝17.6蒲式耳。估量之收穫量為常態收穫量之96.4%，即每英畝17蒲式耳也。

然則實際收穫量對此估量收穫量之差異限度若何？此處求得 $S_{1,2,3,4}$ 之數值為21.68；此即謂實際收穫量對常態收穫量之百分比在 $96.4 + 21.68$ (即118.1)及 $96.4 - 21.68$ (即74.7)之限數以內者，其機率為68%。此兩限數皆指實際收穫量對常態收穫量之百分比而言，但吾人既知是年常態收穫量為17.6蒲式耳，則此兩限數亦可化為蒲式耳數，而得20.8及13.1，故知每英畝實際收穫量在20.8蒲式耳及13.2蒲式耳之限數以內者，其機率為68%。表內一九二二年每英畝之實際收穫量為19.3蒲式耳。

本例所估量者為表內所已包含之一年，吾人倘欲估量未來年份之收穫量，亦可用同一估量方法求之，惟此時因須將長期趨勢綫引長之故，又多一不定因素矣。

淨相關之意義

在上文內吾人曾同時用六、七、八三月之溫度作為三個自變數以斷定堪薩斯州玉蜀黍收穫量所受溫度影響之程度。彼時之目的在於測度該三個月之溫度對於玉蜀黍收穫量所發生之混合作用。吾人現可進而研究統計分析上極占重要地位之淨相關(partial or net correlation)問題。所謂淨相關者，乃在其他各因素保持常態時，一倚變量與一自變量間之純淨關係也。如在上例中，吾人可令六月、八月之溫度以及足以影響玉蜀黍收穫量之其他因素保持常態，而研究收穫量所受七月溫度之影響，此即玉蜀黍收穫量與七月溫度之淨相關也。

吾人倘能發明一種方法，使一因素與其他各因素相隔離，而成為單獨研究之對象，則大可增進經濟學家及一般社會學家分析各種社會經濟現象之能力。研究社會經濟學者乃可應用此法以消除各種無關係之因素，而致其全力於分析一單獨之重要因素。化學家在斷定一種原素所受另一種原素之影響時，亦往往將該種原素所受其他因素之勢力予以消除，以便分析，而此項分析之成就若何，又須視其所用隔離方法精密之程度而定。

在經濟分析上自不易將一變量所受各種因素之影響予以消除，而單獨研究一種因素之作用。經濟現象所由構成之直接及間接原因錯綜複雜，故經濟學家決不能如化學家之便利，而可將複雜之經濟現象化為僅含兩個變量之單純問題。雖然，在可能之限度內，統計學家亦嘗效法科學家令某種因素保持常態，而專研究另一因素變動之形態及其影響。此種方法實為研究社會科學者最有力之工具。

茲仍就堪薩斯州玉蜀黍收穫量之例，以說明淨相關之計算方法。此時之目的在於斷定玉蜀黍收穫量與各月平均溫度之淨相關。

淨相關與單相關之區別

淨相關與通常所測度兩變量間之關係，性質不同，應有區別之必要。吾人前已求得玉蜀黍收穫量與七月溫度之平均關係，其方程式為

$$X_1 = 827.64 - 9.302X_2$$

其標準誤及相關係數為

$$S_{1.2} = 24.73$$

$$r_{1.2} = -.697$$

上述各種量數僅就七月溫度測度收穫量與溫度之單純關係，而絕未顧及其他各種因素，聽其參雜其間，此與化學家研究一種原素對另一種原素所發生之反應時，聽任其試驗管中含有雜質而不設法清除之情形相似。大體言之，經濟學者原不易發現問題中所含之‘雜質’而予以消除，但應認明此項事實之存在，並須將分析所得之結果就此點率直說明之。

淨相關之測度方法

在斷定玉蜀黍收穫量與七月溫度之淨相關時，吾人所欲測度者，為假令其他因素保持常態時，此兩者所應有之純淨關係也；故吾人不能忽視其他因素之勢力，而應同時顧及之，惟欲求之量數須表示收穫量所受七月溫度單獨因素之純淨影響耳。

吾人所用之資料倘包含時期極長，則有一法可供試用。將全部資料內選取六月、八月平均溫度相同之年份，假定所選為三十年，在該時期內六月之平均溫度均為74度，八月之平均溫度均為78度，而收穫量及七月溫度皆有變動，則六、八兩月溫度既無變化，根據此項一部份資料所求得之結果，自可表示收穫量與七月溫度之純淨關係。但此法以能選取其他因素固定不變之資料為要件，實際上難於採用，蓋吾人所用之資料，包含時期往往有限，各種因素亦常有變動，故不易選取相當項目以

供需要。惟除此法以外，尚有其他方法足供採擇，請分述於下。

由迴歸係數與相關係數之關係測度淨相關之方法

在討論此法以前，請先就單相關 (simple correlation) 問題，作簡略之敘述。表示兩變量關係之方程式可用兩種形式表示：其一以 y 倚為變量，其二以 x 為倚變量。以兩變量之算術平均數作為原點，則此兩種方程式之形式為

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

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

兩式內之 $r \frac{\sigma_y}{\sigma_x}$ 及 $r \frac{\sigma_x}{\sigma_y}$ 稱為迴歸係數，所以測度兩迴歸綫之斜度者也。相關係數 r 即等於此兩迴歸係數之平方根，可用下式表示之：

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

上式內各數值改以通常所用之符號代之，得

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

式中 b_{yx} 為 y 倚 x (y 為倚變量) 之迴歸係數， b_{xy} 為 x 倚 y (x 為倚變量) 之迴歸係數。如 y 及 x 之符號改以 x_1 及 x_2 代之，則

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

於此所應注意者，無論 x_1 或 x_2 為倚變量， r 之數值常相等，即 $r_{12} = r_{21}$ 是也。

吾人於此可討論淨相關問題。此處所欲求者為兩變量間之淨相關係數，亦即其他因素保持常態時表示兩者關係之量數也。前在討論複相關時，吾人嘗求得數個淨迴歸係數。表示一個倚變量與數個自變量間關係之迴歸方程式為

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

上式內 $b_{12.34}$, $b_{13.24}$, $b_{14.23}$ 均為淨迴歸係數，其所表示者為同時根據三個自變量以估量倚變量之數值時各自變量所應得之權數。例如 $b_{13.24}$ 為同時根據 X_2, X_3, X_4 三個變量以估量 X_1 時 X_3 所應得之權數。此三個係數之值可用聯立方程式之簡單解法求得之。

惟用此法求淨相關係數時，尚須應用其他數值。相關係數 r_{yx} 為 b_{yx} 及 b_{xy} 之平方根。相關係數 r_{13} 為 b_{13} 及 b_{31} 之平方根。同理，當 X_2 (六月溫度) 及 X_4 (八月溫度) 保持常態時，測度 X_1 (玉蜀黍收穫量) 與 X_3 (七月溫度) 兩者相關程度之淨相關係數可由下式求之：

$$r_{13.24} = \sqrt{b_{13.24} \cdot b_{31.24}}$$

式中 $b_{13.24}$ 為 X_1 倚 X_3 之淨迴歸係數， $b_{31.24}$ 為 X_3 倚 X_1 之淨迴歸係數。

此處所用符號之意義，在逐步討論時曾已提及。旁註小字表示各數值所指之特殊關係。點之左方所註兩小字 (主要註字) 表示此量數所測度之兩變量，點之右方所註小字 (次要註字) 表示測度該兩變量之關係時保持常態之其他變量。在本例中保持常態之變量有二，但此數可為一，亦可為任何他數。故求 X_1, X_3 兩變量間淨相關係數之普遍公式為

$$r_{13.2456 \dots n} = \sqrt{b_{13.2456 \dots n} \cdot b_{31.2456 \dots n}}$$

在兩個主要註字中，前一字表示倚變量，後一字表示自變量。用前式求 $r_{13.24}$ 時，式中迴歸係數 $b_{13.24}$ 之數值前已求得，故僅須計算迴歸係數 $b_{31.24}$ 之數值。後者為方程式

$$X_3 = a + b_{31.24} X_1 + b_{32.14} X_2 + b_{34.21} X_4$$

中常數之一。此式為以七月溫度為倚變量，六月、八月溫度及玉蜀黍收穫量為自變量所得之相關方程式。將已知 X_1, X_2, X_4 之數值代入此式，即可估量最可能之七月溫度 (此式實為表示相聯，而非表示相倚之方程式)。

常態方程式之解法

解上式中常數之方法，與前相同。先求四個常態方程式，再將該四式化爲下列三式：

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

$$p_{23} = p_{12} b_{31.24} + \sigma_2^2 b_{32.14} + p_{24} b_{34.21}$$

$$p_{34} = p_{14} b_{31.24} + p_{24} b_{32.14} + \sigma_4^2 b_{34.21}$$

解此三式所需之各數值，除 σ_1^2 以外，所有各均積數及各標準差在解前三個聯立方程式時均曾採用。將各數值代入，得

$$-62.039 = 1188.688b_{31.24} - 47.967b_{32.14} - 47.519b_{34.21}$$

$$2.790 = -47.967b_{31.24} + 8.3564b_{32.14} + 2.932b_{34.21}$$

$$2.063 = -47.519b_{31.24} + 2.932b_{32.14} + 7.7893b_{34.21}$$

解此三式可得下列數值：

$$b_{31.24} = -.0531$$

$$b_{32.14} = +.0574$$

$$b_{34.21} = -.0807$$

淨相關係數之計算

前例中曾求得

$$b_{13.24} = -7.394$$

將各數值代入下式：

$$r_{13.24} = \sqrt{b_{13.24} \times b_{31.24}}$$

可得

$$\begin{aligned} r_{13.24} &= \sqrt{-7.394 \times -.0531} \\ &= -.6266 \end{aligned}$$

(r 之正負號與迴歸係數之符號相同，此與僅含兩變量之單相關情形無異。又因兩迴歸係數之正負號常相一致，一爲正時，一亦爲正，一爲負時，一亦爲負，故若兩迴歸係數皆爲正時， r 亦爲正，皆爲負時， r 亦爲

負)。

此即堪薩斯州玉蜀黍收穫量與七月溫度之淨相關係數，所以測度六月及八月溫度保持常態時收穫量與七月溫度間之純淨關係者也。

計算淨相關係數之又一法

計算淨相關係數除前法外，又可用另一法。此法所用公式較有條理，為其優點。

如計算相關係數時所包含之變量僅有兩個，則此兩變量間之單相關係數稱為零次係數 (coefficient of zero order)，可用符號 r_{12} , r_{24} 等代表之。如包含變量共有三個，而令其中一變量保持常態，則所求得其他兩變量之淨相關係數稱為一次係數 (coefficient of the first order)，可用 $r_{12.3}$, $r_{24.3}$ 等代表之。同理，保持常態之自變量為二個、三個、四個或 n 個，而測度一倚變量與另一自變量之關係時，吾人亦可求二次、三次、四次或 n 次係數。

每一淨相關係數皆可由較低一次之淨相關係數之值推求而得，故一次係數可由零次係數求之。其式如下：

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

二次係數之公式為

$$r_{12.34} = \frac{r_{12.3} - r_{14.3} \cdot r_{24.3}}{(1 - r_{14.3}^2)^{\frac{1}{2}} (1 - r_{24.3}^2)^{\frac{1}{2}}}$$

n 次淨相關係數之普遍公式為

$$r_{12.345 \dots n} = \frac{r_{12.345 \dots (n-1)} - r_{1n.345 \dots (n-1)} \cdot r_{2n.345 \dots (n-1)}}{(1 - r_{1n.345 \dots (n-1)}^2)^{\frac{1}{2}} (1 - r_{2n.345 \dots (n-1)}^2)^{\frac{1}{2}}}$$

如是則吾人可由零次係數起逐漸推算至高次係數，惟應用此式，計算費時，幸有表可用(註)，故計算工作可減至最低度。茲就前例所用之資

(註) 麥納氏 (J. R. Miner) 著有 "Tables of $\sqrt{1-r^2}$ and $1-r^2$ for use in Partial Correlation and in Trigonometry"，一九二二年 Johns Hopkins Press (Baltimore, Md.) 出版。

料說明此法之應用。

吾人現時所欲求者為 $r_{12.34}$, $r_{13.24}$ 及 $r_{14.23}$ 三個二次係數，此為玉蜀黍收穫量與六、七、八各月之淨相關係數。計算 $r_{12.34}$ 之公式前已舉示，至於 $r_{13.24}$ 及 $r_{14.23}$ 之數值可用下式計算：

$$r_{13.24} = \frac{r_{13.2} - r_{14.2} \cdot r_{34.2}}{(1 - r_{14.2}^2)^{\frac{1}{2}} (1 - r_{34.2}^2)^{\frac{1}{2}}}$$

$$r_{14.23} = \frac{r_{14.2} - r_{13.2} \cdot r_{43.2}}{(1 - r_{13.2}^2)^{\frac{1}{2}} (1 - r_{43.2}^2)^{\frac{1}{2}}}$$

但此三個二次係數之數值亦可應用排列稍異之一次係數公式求之，其式如下：

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

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

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

故若同時用此兩種公式以計算二次係數，可作校核計算錯誤之用。

一次係數之計算

計算二次係數時，須先求式中所需一次係數之數值。由淨相關係數之普遍公式，可得形式如下之一次係數之公式：

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

吾人所需一次係數之數值不止一項，故計算手續須有條不紊。如將各項數字編成表一一之排列方式，則計算上較為省事，而應用相關數表時亦較為便利。

計算一次係數之手續頗為簡單。其公式內之分子包含三個零次係數，故求每個一次係數時須先計算三個零次係數之數值。各個零次係數應排列成表，並須按照其在一次係數公式內位置之先後順次排列，如表

一二一。由第一零次係數減去第二及第三零次係數之乘積 即得該式之分子。此項乘積見表內第四行。公式內之分母為屬於 $\sqrt{1-r^2}$ 形式之兩項相乘之積，而由三個零次係數所成一組中之第二、第三零次係數求得者也。表內行列既定，乃可作有系統之計算。

表 一 二 一

一次淨相關係數之計算

堪薩斯州玉蜀黍收穫量與溫度之關係

r 零次		$(1-r^2)^{\frac{1}{2}}$	分子中之乘積	分子全部	分母	r 一次	
註字	係數					註字	係數
12	-.4814	.7173 .9275	-.2604	-.2210	.6653	12.3	-.3322
13	-.6968						
23	+.3737						
14	-.4937	.7173 .9582	-.1994	-.2943	.6873	14.3	-.4282
13	-.6968						
43	+.2862						
21	+.3633	.9275 .9582	+.1070	+.2563	.8887	21.3	+.2884
23	+.3737						
43	+.2862						
13	-.6968	.8765 .9275	-.1799	-.5169	.8130	13.2	-.6358
12	-.4814						
32	+.3737						
14	-.4937	.8765 .9317	-.1749	-.3188	.8166	14.2	-.3904
12	-.4814						
42	+.3633						
34	+.2862	.9275 .9317	+.1358	+.1504	.8642	34.2	+.1740
32	+.3737						
42	+.3633						
12	-.4814	.8696 .9317	-.1794	-.3020	.8102	12.4	-.3727
14	-.4937						
24	+.3633						
13	-.6968	.8696 .9582	-.1413	-.5555	.8333	13.4	-.6666
11	-.4937						
34	+.2862						
23	+.3737	.9317 .9582	+.1640	+.2697	.8928	23.4	+.3021
24	+.3633						
34	+.2862						
11	-.4937	.7173 .9582	-.1994	-.2943	.6873	11.3	-.4282
13	-.6968						
43	+.2862						
12	-.4814	.7173 .9275	-.2604	-.2210	.6653	12.3	-.3322
13	-.6968						
23	+.3737						
42	+.3633	.9582 .9275	+.1070	+.2563	.8887	42.3	+.2884
43	+.2862						
23	+.3737						

係數 $r_{23.4}$ 與 $r_{32.4}$ 之數值相同, $r_{34.2}$ 與 $r_{43.2}$ 亦彼此相同, 餘類推, 故計算時可不必重複求相同之數值。

二次係數之計算

吾人求得一次係數後, 即可用類似方法以求三個二次係數。其計算手續如下表所示。表內用一次係數之兩種配合方法同時求得二次係數之數值, 以為校核計算錯誤之用。

表 一 二 二

二次淨相關係數之計算

堪薩斯州玉蜀黍收穫量與溫度之關係

r 一次		$(1-r^2)^{\frac{1}{2}}$	分子中之乘積	分子全部	分母	r 二次			
註字	係數					註字	係數		
12.3	-.3322	.9037 .9575	-.1235	-.2087	.8653	12.34	-.2412		
14.3	-.4282								
24.3	+.2884								
13.2	-.6358	.9206 .9847	-.0679	-.5679	.9065	13.24	-.6265		
14.2	-.3904								
34.2	+.1740								
14.2	-.3904	.7719 .9847	-.1106	-.2798	.7601	14.23	-.3681		
13.2	-.6358								
43.2	+.1740								
12.4	-.3727	.7454 .9533	-.2014	-.1713	.7106	12.34	-.2411		
13.4	-.6666								
23.4	+.3021								
13.4	-.6666	.9280 .9533	-.1126	-.5540	.8847	13.24	-.6262		
12.4	-.3727								
32.4	+.3021								
11.3	-.4282	.9432 .9575	-.0958	-.3324	.9031	11.23	-.3681		
12.3	-.3322								
42.3	+.2884								

$r_{13.24}$ 之數值與前用兩淨迴歸係數之數值所求得者相同。

淨相關係數之意義前已言之, 姑不贅述。茲將計算所得結果併列於下, 以便比較, 且以顯示上項分析足以啓迪吾人智識之一斑焉。

$$r_{12} = -.4814$$

$$r_{12.34} = -.2412$$

$$r_{13} = -.6968$$

$$r_{13.24} = -.6265$$

$$r_{14} = -.4937$$

$$r_{14.23} = -.3681$$

由上列數字比較觀之，可知淨相關係數所表示六月溫度對於玉蜀黍收穫量之影響顯較單相關係數所表示者為小。此因六月溫度與七八兩月溫度間有一正相關之關係，故單相關係數所表示者遂較真實之影響為大。七月及八月之相關係數亦然，其淨相關係數皆小於單相關係數。但就各月溫度對於收穫量之影響而言，則七月溫度實較其他兩月為重要。

淨相關係數之所謂淨，其意義自僅對計算係數時所已顧及而令其保持常態之自變量而言，實則足以影響於玉蜀黍之收穫量者，除溫度以外，尚有六月、七月、八月之雨量等其他原因，且雨量與溫度又彼此相關，故如將雨量列入，則所得淨相關係數之數值又必不同矣。

測量離散度之量數

吾人既得淨相關係數，可進而求另一重要之量數。此量數所以測度各變量保持常態時一變量之離散度者也。在上例中，如六月、七月、八月之溫度皆保持常態，欲問玉蜀黍收穫量之離散度如何？換言之 即吾人倘將玉蜀黍收穫量由溫度變化所發生之變動完全剔除，則尚有何種變動存在是也。測量此離散度之量數可用符號 $\sigma_{1.234\dots n}$ 代表之。 $\sigma_{1.234\dots n}$ 稱為 n 次之標準差。

求此量數可用下列普遍公式：

$$\sigma^2_{1.23\dots n} = \sigma_1^2(1-r^2_{12})(1-r^2_{13.2})(1-r^2_{14.23})\dots(1-r^2_{1n.23\dots n-1})$$

將此式應用於玉蜀黍收穫量之資料，得

$$\sigma^2_{1.234} = 1188.688[1-(-.4814)^2][1-(-.6358)^2][1-(-.3681)^2]$$

$$\sigma^2_{1.234} = 470.3364$$

$$\sigma_{1.234} = 21.69$$

此值與前所求得 $S_{1.234}$ 之數值相同(其差數為.01)，由是知 X_2, X_3, X_4 三個變量保持常態時， X_1 之標準差實即測度迴歸綫兩旁散佈程度

之量數；而在根據 X_2, X_3, X_4 三個變量以估量時，此量數亦即估量之標準誤也。此中原因頗易明瞭。當 X_1, X_2, X_3 保持常態時， X_1 之變動乃因實際數值與估量數值之不同而起，而 S 即為測度此不同程度之量數也。吾人倘能瞭解此兩種量數相同之處，則其意義可益明瞭矣。

$\sigma_{1.234}$ 既與 $S_{1.234}$ 相同，則複相關係數 $R_{1.234}$ 可依下式計算：

$$R^2_{1.23\dots n} = 1 - \frac{\sigma^2_{1.23\dots n}}{\sigma_1^2}$$

如用 $\sigma^2_{1.234\dots n}$ 之公式，則 $R_{1.234\dots n}$ 又可由下式求之：

$$1 - R^2_{1.23\dots n} = (1 - r^2_{12})(1 - r^2_{13.2})(1 - r^2_{14.23}) \dots (1 - r^2_{1n.23\dots(n-1)})$$

但上述計算 $R_{1.234\dots n}$ 所用之前一式係直接根據最小平方法求得，意義簡明，一般而論，較為適宜。

測度相關方法應用上之限制

以上所述複相關及淨相關量數之各種計算方法，僅能適用於兩變量關係之可用直綫表示者。在用四個變量時，可有六種配法而成六對變量，此六對變量中每對之關係均須成直綫，方可應用前述各法以研究其複相關或淨相關。若用真數求得之迴歸綫並非直綫，則用對數或倒數或可求得直綫之關係。又若 X_1 之對數與其他各變量之真數間之關係成直綫時，則可應用形式如下之估量方程式：

$$\log X_1 = a + b_{1.234}X_2 + b_{13.24}X_3 + b_{14.23}X_4$$

惟此時所求得 S 及 R 之數值表示比率而非絕對值，如第十三章所舉之例。

此外更須注意者，求複相關或淨相關係數時，如包含變量頗多，則所用資料務必豐富，包含項目務必衆多。倘資料不足而包含變量又多，則相關係數即無意義，所得數值縱或甚高，不足置信。本章所論複相關

及淨相關之各種方法，在實際應用上雖受上述種種限制，但用於經濟分析，實為極有效之工具。

參 考 書

- BOWLEY, ARTHUR L. *Elements of Statistics* (398—408).
- EDGEWORTH, F. Y. *On Correlated Averages*. *Phil. Mag.* 5th series, Vol. XXXIV, 1892 (194).
- EZEKIEL, M. J. B. *A Method of Handling Curvilinear Correlation for any Number of Variables*. *Journal of the American Statistical Association*, Vol. XIX, N. S. 148, 1924.
- HAAS, G. C. *Sale Prices as a Basis for Farm Land Appraisal*. Technical Bulletin No. 9. University of Minnesota Agricultural Experiment Station.
- KELLEY, TRUMAN L. *Statistical Method* (279—310).
- KELLEY, TRUMAN L. *Partial and Multiple Correlation* (in Rietz, H. L. *Handbook of Mathematical Statistics*, 139—149).
- MINER, J. R. *Tables of $\sqrt{1-r^2}$ and $1-r^2$ for use in Partial Correlation and Trigonometry*.
- PEARL, RAYMOND. *Medical Biometry and Statistics* (319—331).
- PEARSON, KARL. *Regression, Heredity and Panmixia*. *Phil. Transactions Royal Society*, Series A. Vol. CLXXXVII, 1896 (253—318).
- SMITH, BRADFORD B. *The Use of Punched Card Tabulating Equipment in Multiple Correlation Problems*. (Prepared for the use of statisticians of the Bureau of Agricultural Economics, U. S. Dept. of Agriculture.)
- TOLLEY, H. R. and EZEKIEL, M. J. B. *A Method of Handling Multiple Correlation Problems*. *Journal of the American Statistical Association*, Dec. 1923.
- YULE, G. U. *An Introduction to the Theory of Statistics* (229—253)
(The notation usually employed in the correlation of several variables was developed by Yule. It is explained in this reference.)

第十五章 機率及差誤常態曲綫

前在討論次數分配時，嘗論及各種現象之次數分配具有類似之形態；同時又論及一種基本形態之次數分配，而可用對稱而成鐘形之曲綫表示之。此綫在統計學上稱爲常態曲綫(normal curve)或差誤常態曲綫(normal curve of error)。曩昔頗有以此綫可代表各種次數分配之基本法則者，但以現代目光觀之，此種論斷未免錯誤，蓋常態曲綫不過爲表示次數分配之各種曲綫中之一種，惟所占地位最爲重要，故統計學者對於此種曲綫之特質有瞭解之必要。

機率之簡單定理

欲使學者明瞭常態曲綫之意義起見，先將機率之單純定理在此作簡略之討論。機率原理頗爲繁複，詳細解釋非本書範圍所能及。以下所述僅爲介紹此項原理之性質，並引用簡單數字以說明機率原理與差誤常態法則之關係而已。

茲先將以下各例所用之標準符號說明之。設一事件(event)之發生共有 n 種可能情形，而其中 a 種情形認爲成就(successful)， b 種情形認爲不成就(unsuccessful)，則成就之機率 p 爲

$$p = \frac{a}{n}$$

而不成就之機率 q 爲

$$q = \frac{b}{n}$$

但因成就之情形與不成就之情形相加，等於事件發生可能情形之總數，故

$$a + b = n$$

以 n 除各項，得

$$\frac{a}{n} + \frac{b}{n} = 1$$

故

$$p + q = 1$$

亦即必然 (certainty) 之意也。

由是知機率可用比率表示，比率中之分子代表成就（或不成就）之情形，其分母代表可能情形。

簡單機率之舉例

試取一錢而擲之，如錢幣正面向上認為成就，則成就之機率為

$$p = \frac{1}{2}$$

失敗之機率為

$$q = \frac{1}{2}$$

試取一骰而擲之，如擲得 6 點者認為成就，則

$$p = \frac{1}{6}$$

$$q = \frac{5}{6}$$

又若由撲克牌 52 張中抽取一張，則抽得黑桃 ace 之機率為 $\frac{1}{52}$ ，不能抽得之機率為 $\frac{51}{52}$ 。

機率之相加

設由撲克牌 52 張中抽取一張，則抽得黑桃 ace 或黑桃 2 之機會若何？在本例中，無論抽得黑桃 ace 或黑桃 2 皆認為成就，故成就之機率應為兩個機率之總和，即

$$p = \frac{1}{52} + \frac{1}{52} = \frac{1}{26}$$

由全副撲克牌中抽取一雞心或一黑桃之機率應為

$$p = \frac{13}{52} + \frac{13}{52} = \frac{1}{2}$$

機率之相乘

兩種事件若其中一種事件之實現不影響於其他一種之實現者，謂之獨立事件。例如取一骰擲之，其第一次所得結果應毫不影響於第二次之結果。一複合事件 (compound event, 即兩種各自獨立而依次發生之事件) 之機率應為各獨立事件之機率相乘之積。例如第一次擲出 1 點與第二次擲出 2 點之機率為

$$p = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$$

吾人計算機率時，每須兼用加法與乘法。例如同時取兩骰擲之，欲求兩骰共得 5 點之機率。令一骰為 a ，一骰為 b ，則擲得 5 點不出下列四種配法：

a 骰	b 骰
1	4
2	3
3	2
4	1

a 骰擲得 1 點之機率為 $\frac{1}{6}$ ， b 骰擲得 4 點之機率亦為 $\frac{1}{6}$ ，故此兩者同時擲得之機率為 $\frac{1}{36}$ ，此為第一種配法。其他三種配法之機率亦各為 $\frac{1}{36}$ 。

四種之中，任何一種皆可認為成就，故擲得 5 點之機率為

$$p = \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} = \frac{1}{9}$$

在此例中，吾人對於兩骰共得 5 點之機率，已得一肯定之答案。若

將此問題稍加改變，而求兩骰同擲至少可得 5 點之機率。此時兩骰共得點數，或為 5 點，或多於 5 點，俱可認為成就。吾人可依照上例求各種成就方式之機率。此項機率簡列於下：

$$\text{兩骰擲得 12 點之機率} = \frac{1}{36}$$

$$\text{兩骰擲得 11 點之機率} = \frac{2}{36}$$

$$\text{兩骰擲得 10 點之機率} = \frac{3}{36}$$

$$\text{兩骰擲得 9 點之機率} = \frac{4}{36}$$

$$\text{兩骰擲得 8 點之機率} = \frac{5}{36}$$

$$\text{兩骰擲得 7 點之機率} = \frac{6}{36}$$

$$\text{兩骰擲得 6 點之機率} = \frac{5}{36}$$

$$\text{兩骰擲得 5 點之機率} = \frac{4}{36}$$

$$\text{上項機率之總和} = \frac{30}{36}$$

故擲得 5 點或 5 點以上之機率為 $\frac{30}{36}$ 或 $\frac{5}{6}$ 。

兩項展開式與機率之測度

上述事件可用一普遍方式表示之。茲用一簡例，即可說明此普遍方式之由來。

例如取錢幣兩枚同時擲之，則其可能之結果可有下列四種：

a	b	a	b	a	b	a	b
面	面	背	面	背	面	背	背

(此兩幣以 a 及 b 代表) 兩幣面向上、一面一背及兩幣背向上之機率各為

$\frac{1}{4}$, $\frac{1}{2}$ 及 $\frac{1}{4}$ 。又若取錢幣三枚同時擲之 (此三幣以 a, b, c 代表), 則有

下列八種可能之結果:

$abc \quad abc \quad abc \quad abc \quad abc \quad abc \quad abc \quad abc$

面面面 面面背 面背背 面背面 背面面 背面背 背背面 背背背

三面、二面、一面、三背之機率各為 $\frac{1}{8}, \frac{3}{8}, \frac{3}{8}, \frac{1}{8}$

但上列各種結果之機率, 其實不必如此詳舉, 亦可求得。吾人曾用 p 及 q 分別代表成就及失敗之機率。如有兩個獨立事件, 而其 p, q 之機率相同 則其複合機率 (compound probabilities) 可以下列展開式表示之:

$$(p+q)^2$$

本例中 p (面向上之機率) $= q = \frac{1}{2}$, 故各種結果之機率如下式所示:

$$\left(\frac{1}{2} + \frac{1}{2}\right)^2 = \frac{1}{4} + \frac{1}{2} + \frac{1}{4}$$

此與第一例所得之結果相同。如有三個獨立事件, 則由下式

$$\left(\frac{1}{2} + \frac{1}{2}\right)^3 = \frac{1}{8} + \frac{3}{8} + \frac{3}{8} + \frac{1}{8}$$

所得之機率又與第二例所得之結果相同。

如吾人所欲知者: 非各項單獨之機率, 而為若干次試驗中各種結果可能之實現次數, 則可依下式計算之:

$$N(p+q)^n$$

式中 N 代表試驗之次數, n 代表獨立事件之數。例如獨立事件有二, 共試驗 200 次, 則各種結果之次數可依下式求之:

$$200(p+q)^2 = 200(p^2 + 2pq + q^2)$$

如 $p = q = \frac{1}{2}$, 則得:

$$200\left(\frac{1}{4}\right) + 200\left(\frac{1}{2}\right) + 200\left(\frac{1}{4}\right) = 50 + 100 + 50$$

上式係表示二事成就、一事成就、及兩事無一成就之可能次數。

又如獨立事件有三，試驗 N 次，則各項之次數可就下列展開式得之：

$$N(p+q)^3$$

倘 N 為200，則上式為：

$$200(p^3 + 3p^2q + 3pq^2 + q^3)$$

倘 p 等於 $\frac{1}{2}$ ，則得：

$$200\left(\frac{1}{8}\right) + 200\left(\frac{3}{8}\right) + 200\left(\frac{3}{8}\right) + 200\left(\frac{1}{8}\right)$$

上式內各項依次為三事成就、二事成就、一事成就、及三事無一成就之可能次數。各項次數之和應等於試驗之總次數，因各項可能之結果均包括於展開式內也。

故若吾人預知成就及失敗之機率(註)，即可斷定各種成就或失敗之可能次數，此在統計學理之發展上極關重要者也。

實在次數與理論次數之比較

吾人如將擲錢擲骰試驗中所得各項之實在次數與由兩項展開式求得之理論次數加以比較，則可明瞭若干重要之點。試取十二骰同時擲若干次，而以得4,5,6點者為成就，1,2,3點者為失敗（例如某次擲骰之結果為3,1,5,1,2,4,4,6,3,2,3,5，則成就之數為5，而紀錄之）。維爾頓

(註)機率通常可分為二種：一曰預知機率 (priori probabilities)，如上例中所舉者；一曰經驗機率 (empirical probabilities)，乃由觀察或經驗而得之機率也。試舉一經驗機率之例。根據 American Experience Table of Mortality 所載年齡三十五歲之81,822人中，過十年後尚能生存者為74,173人，故年齡三十五歲之人能繼續生存十年之機率為 $\frac{74173}{81822}$ 。第十六章係討論此種由實驗而得之量數之應用。

氏(W. F. R. Weldon)(註)曾作此項擲骰試驗,共擲4096次,並將每次結果依上法一一紀錄之。其所得結果見表一二三第(2)行;其次數分配見圖八十五。由該次數分配求得算術平均數為6.139,標準差為1.712。

表 一 二 三

擲骰試驗中實在次數與理論次數之比較

(1)	(2)	(3)
得4,5,6各點之骰子數	實在次數	理論次數
0	0	1
1	7	12
2	60	66
3	198	220
4	430	495
5	731	792
6	948	924
7	847	792
8	636	495
9	257	220
10	71	66
11	11	12
12	0	1
	4096	4096

茲將此項試驗所得實在次數與理論上所應得之次數比較之。每次擲12骰,故獨立事件為12。共擲次數為4096。擲4,5,6點者皆認為成就,故 $p = q = \frac{1}{2}$ 。

兩項展開式之各項為

$$(a + b)^n = a^n + na^{n-1}b + \frac{n(n-1)}{1 \cdot 2}a^{n-2}b^2 + \frac{n(n-1)(n-2)}{1 \cdot 2 \cdot 3}a^{n-3}b^3 + \dots$$

(註)猶爾氏(G. U. Yule)在其所著“An Introduction to the Theory of Statistics”

(第五版)第258頁暨愛奇渥斯氏(F. Y. Edgeworth)在 Encycl. Brit. (第十一

版)第十二卷第394頁曾分別引載維爾頓氏此項實驗之結果。

在本例中可依上式將

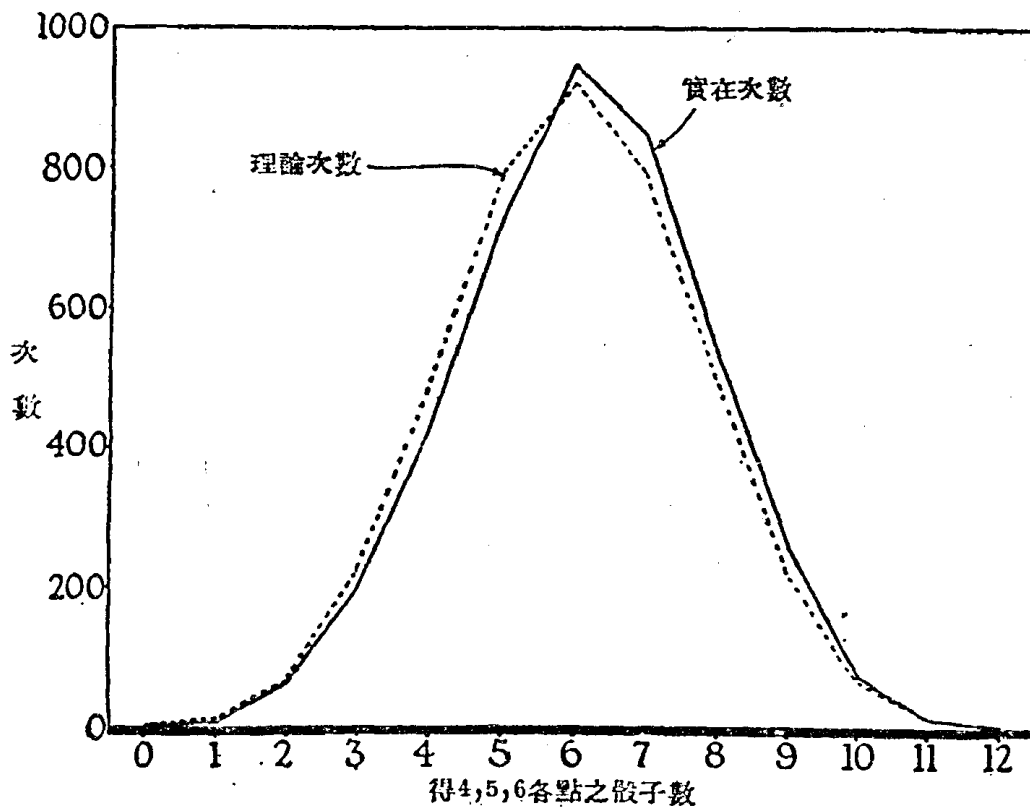
$$4096\left(\frac{1}{2} + \frac{1}{2}\right)^{12}$$

展開,得

$$4096\left(\frac{1}{4096} + \frac{12}{4096} + \frac{66}{4096} + \frac{220}{4096} + \frac{495}{4096} + \frac{792}{4096} + \frac{924}{4096} + \frac{792}{4096} + \frac{495}{4096} + \frac{220}{4096} + \frac{66}{4096} + \frac{12}{4096} + \frac{1}{4096}\right)$$

將上式括弧內各項乘4096,可得十二骰同時擲4096次之各種成就方式之理論次數。此項次數分配見上表第(3)行。

理論次數之分配所繪曲綫見圖八十五。



圖八十五. 擲骰試驗中實在次數分配與理論次數分配之比較

吾人以此理論次數分配作為標準,而與實在分配比較,可得種種利益。吾人比較理論次數與實在次數所得之差異,可斷定所受實驗時由取樣而發生偶然變化(或由於骰子之構造而發生之偏誤)之影響程度;並

可斷定實驗次數(如擲骰次數)無限增加後所應有之真實分配,蓋由實驗所得之實在分配,其意義常為實驗次數所限,非若理論分配之可應用於各種機率相同之例也。是以吾人倘知某種現象之理論分配,即可用作概括論斷 (generalization) 之基礎,而僅憑實驗之結果則不可能,此為應用理論分配最顯著之利益也。

在上例中,其成敗之機率既已預知,故可計算其理論次數分配之算術平均數與標準差。如已知獨立事件之總數及其成就之機率,則平均成就數之普遍公式為:

$$M = np$$

以本例之數值代入,則

$$M = 12 \times \frac{1}{2} = 6$$

由實在次數之分配計算所得之算術平均數為6.139。

標準差之普遍公式為:

$$\sigma = \sqrt{npq}$$

在本例中

$$\begin{aligned} \sigma &= \sqrt{12 \times \frac{1}{2} \times \frac{1}{2}} = \sqrt{3} \\ &= 1.732 \end{aligned}$$

由實在次數分配計算所得之標準差為1.712。

差誤常態曲綫

茲就圖八十五代表擲骰試驗中理論次數分配之曲綫加以討論。此綫為一完全對稱之十二邊多邊形,其邊數(底邊除外)與獨立事件之總數相等。如獨立事件數為六,則為六邊形,獨立事件數為二十,則為二十邊形,餘類推。 n 之數逐漸增加,則多邊形邊數逐漸增多,而代表 $(p+q)^n$ 兩項展開式之圖形漸變為一平滑之曲綫。 n 之數增至無限大時,可得一

完全平滑之綫。此即圖八十七所作之差誤常態曲綫也。

差誤常態曲綫亦稱高斯氏曲綫 (Gaussian curve)。由此種曲綫特質上研究之所得，對於現代統計技術頗多貢獻。此綫之理論次數雖可用兩項展開式求之，但計算繁重，不若用此曲綫之積分表計算較為簡便。積分表之應用方法，可舉例說明於後。

此曲綫方程式可用種種形式表示，其最普通者為

$$y = y_0 e^{-\frac{x^2}{2\sigma^2}}$$

式中 y_0 為一常數，代表最高之縱坐標， e 亦為一常數(即那浦爾氏對數之基數)，其值為2.7182818， σ 代表標準差(註一)。最高縱坐標又可用下式求之：

$$y_0 = \frac{N}{\sigma\sqrt{2\pi}}$$

故常態曲綫之方程式亦可寫作

$$y = \frac{N}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}$$

此曲綫及其機率積分表應用頗廣，姑不詳論，茲僅就以下所舉簡單之例，以說明其用法。

應用於經濟資料之舉例

美國電話電報公司統計部曾研究勃發羅城 (Buffalo) 裝用四部綫 (four-party line) 電話之住戶995家每年通話之次數，並將每家通話次數分組列表(註二)。其結果以及各種計算詳見下表。

(註一)高斯氏差誤常態方程式之演繹法可在討論最小平方法之各種標準著作中見之，其書名見附錄A末頁。

(註二)摘載美國電話電報公司統計部主任所主編之統計公報項內統計方法叢書第一種 "Introduction to Frequency Curves and Averages".

表 一 二 四
按每年通話次數分組之電話用戶995家之次數分配

計算次數分配動差之舉例

(1) 通話次數 之組距 (註)	(2) 中 點	(3) 次 數	(4) 對於假定 原點之 離中差 (以組距 為單位)	(5)	(6)	(7)	(8)
	<i>m</i>	<i>f</i>	<i>x'</i>	<i>fx'</i>	<i>f(x')²</i>	<i>f(x')³</i>	<i>f(x')⁴</i>
0- 50	25	0	-10	0	0	0	0
50- 100	75	1	-9	-9	81	-729	6561
100- 150	125	9	-8	-72	576	-4608	36864
150- 200	175	19	-7	-133	931	-6517	45619
200- 250	225	38	-6	-228	1368	-8208	49248
250- 300	275	50	-5	-250	1250	-6250	31250
300- 350	325	95	-4	-380	1520	-6080	24320
350- 400	375	85	-3	-255	765	-2295	6885
400- 450	425	116	-2	-230	460	-920	1840
450- 500	475	132	-1	-132	132	-132	132
500- 550	525	144	0	0	0	0	0
550- 600	575	116	1	116	116	116	116
600- 650	625	79	2	158	316	632	1264
650- 700	675	54	3	162	486	1458	4374
700- 750	725	31	4	124	496	1984	7936
750- 800	775	11	5	55	275	1375	6875
800- 850	825	5	6	30	180	1080	6480
850- 900	875	6	7	42	294	2058	14406
900- 950	925	2	8	16	128	1024	8192
950-1000	975	1	9	9	81	729	6561
1000-1050	1025	1	10	10	100	1000	10000
1050-1100	1075	1	11	11	121	1331	14641
		995		-956	9676	-22952	283564

(註)本表所用分組法凡與組限相同之數值均歸入較小之一組，例如50之數值係歸入以50為上限之一組內。

次數分配之動差

以上討論時，尚有數種名詞及符號未經提示，此際可略加說明。下列四種符號之意義為：

$$v_1 = \frac{\sum f(x')}{N} = \text{次數分配對於假定原點之第一動差}$$

$$v_2 = \frac{\sum f(x')^2}{N} = \text{次數分配對於假定原點之第二動差}$$

$$v_3 = \frac{\Sigma f(x')^3}{N} = \text{次數分配對於假定原點之第三動差}$$

$$v_4 = \frac{\Sigma f(x')^4}{N} = \text{次數分配對於假定原點之第四動差}$$

動差 (moment) 爲力學上習見之名詞, 所以測度旋轉趨勢之力量者也。此力量之大小隨用力點與原點距離之遠近而異。動差在統計學上之意義與在力學上之意義相似, 各組次數可視爲力學上之力, 每組之中點與原點之距離爲計算動差最重要之因素。次數分配對於任何原點之動差可用下法計算。其法以每組之次數乘 x 軸上該組中點對原點距離之冪方, 而後將各組之乘積相加, 再以次數之和除之。如所求者爲第一動差, 則用各組中點對原點距離之一次方乘其次數; 如所求者爲第四動差, 則用此距離之四次方乘之, 餘類推。符號 v 旁所註小字, 所以表示動差之級次者也。

次數分配之動差雖可根據任何原點求之, 但以根據算術平均數點所求得者, 在統計學上最爲重要。以 π 代表此種動差, 則可得下列各種關係:

$$\text{次數分配對算術平均數之第一動差} = \pi_1 = 0$$

$$\text{次數分配對算術平均數之第二動差} = \pi_2 = v_2 - v_1^2$$

$$\text{次數分配對算術平均數之第三動差} = \pi_3 = v_3 - 3v_1v_2 + 2v_1^3$$

$$\text{次數分配對算術平均數之第四動差} = \pi_4 = v_4 - 4v_1v_3 + 6v_1^2v_2 - 3v_1^4$$

根據分組資料計算此種動差時, 常含有一種假定, 卽每組內各項數值皆集中於該組之中點是也。惟由此假定計算所得之結果往往發生常誤 (constant error), 尤以第二動差與第四動差爲然, 其由分組資料所求得之數值與由不分組資料所求得者常有差異。

舍巴德氏 (W. F. Sheppard) (註) 爲校正此項偏誤起見, 嘗求得一校正公式。次數分配如具有下列兩條件, 卽可應用舍巴德氏之校正公式。

(註) 參看 "Proceedings of the London Mathematical Society" 第29卷第353—380頁。

1. 次數分配之變量, 須爲一連續變量。
2. 次數曲綫須含有“高接”性 (“high contact”), 亦即次數曲綫之兩端須向左右延長而緩漸下降, 呈平坦之形態者。

符號 μ 係用以代表次數分配對算術平均數之校正動差者。應用舍巴德氏之校正法可得動差之公式如下:

$$\mu_1 = 0$$

$$\mu_2 = \pi_2 - \frac{1}{12}$$

$$\mu_3 = \pi_3$$

$$\mu_4 = \pi_4 - \frac{1}{2} \pi_2 + \frac{7}{240}$$

(應用該校正法時, $\frac{1}{12}$ 及 $\frac{7}{240}$ 通常可用小數 .083333 及 .029167 代之。) 惟應用上項公式時, 計算各項對算術平均數之離中差, 須以組距作爲單位耳。

吾人應注意者, 所謂標準差實即根據算術平均數計算之第二動差之平方根也。其未校正之標準差數值係由下式求得:

$$\sigma = \sqrt{\pi_2}$$

如應用舍巴德氏之校正法, 則求校正標準差之公式爲:

$$\sigma = \sqrt{\mu_2}$$

茲仍就前例所用按每年通話次數分組之電話用戶次數分配說明動差之計算於下。此項次數分配之變量雖非連續變量, 但因其單位(1)與全距之比極小, 可視爲連續變量, 且其曲綫之兩端尙具有高接性, 故計算動差時曾採用舍巴德氏之校正法。

$$v_1 = \frac{-956}{995} = -.960804$$

$$v_2 = \frac{9676}{995} = 9.724623$$

$$v_3 = \frac{-22952}{995} = -23.067337$$

$$v_4 = \frac{283564}{995} = 284.988945$$

$$\pi_1 = 0$$

$$\pi_2 = v_2 - v_1^2 = 9.724623 - .923144 = 8.801479$$

$$\pi_3 = v_3 - 3v_1v_2 + 2v_1^3 = -23.067337 + 28.030370$$

$$-1.773922 = 3.189111$$

$$\pi_4 = v_4 - 4v_1v_3 + 6v_1^2v_2 - 3v_1^4$$

$$= 284.988945 - 88.652760 + 53.863384 - 2.556586$$

$$= 247.642983$$

$$\mu_1 = 0$$

$$\mu_2 = \pi_2 - \frac{1}{12} = 8.801479 - .083333 = 8.718146$$

$$\mu_3 = \pi_3 = 3.189111$$

$$\mu_4 = \pi_4 - \frac{1}{2}\pi_2 + \frac{7}{240} = 247.642983 - 4.400739 + .029167$$

$$= 243.271411$$

斷定曲綫形態之標準

以上各項數值既已求得，吾人可進而討論應用常態曲綫之方法。一種次數分配之是否可用常態曲綫表示，可根據各種標準斷定之。此項標準又可由該次數分配計算所得校正動差之數值得之於下：

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{10.170429}{662.632015} = .01534853$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{243.271411}{76.006070} = 3.200683$$

$$\kappa_2 = \frac{\beta_1(\beta_2 + 3)^2}{4(4\beta_2 - 3\beta_1)(2\beta_2 - 3\beta_1 - 6)}$$

$$= \frac{.01534853 \times 38.448470}{4(12.756686)(.355320)} = \frac{.5901275}{18.130823}$$

$$\kappa_2 = .032548$$

根據常態曲綫求得之此項標準之數值爲

$$\beta_1 = 0$$

$$\beta_2 = 3$$

$$\kappa_2 = 0$$

故吾人可暫時斷定該次數分配可用常態曲綫表示之(註一)。

常態曲綫之配合;縱坐標表之應用

配合常態曲綫時,因應用縱坐標表,手續可減省不少。此表所表示者爲常態曲綫各部分之縱坐標對最高縱坐標 y_0 之比率。下列簡表所以說明配合之手續,惟在實際應用上,此表尙嫌其不詳備(註二)。

表 一 二 五
常態曲綫之縱坐標
係以對最高縱坐標之比率表示者

x/σ	y/y_0	x/σ	y/y_0
0.0	1.00000	2.5	.04394
0.1	.99501	2.6	.03405
0.2	.98020	2.7	.02612
0.3	.95600	2.8	.01984
0.4	.92312	2.9	.01492
0.5	.88250	3.0	.01111
0.6	.83527	3.1	.00819
0.7	.78270	3.2	.00598
0.8	.72616	3.3	.00432
0.9	.66689	3.4	.00309
1.0	.60663	3.5	.00219
1.1	.54607	3.6	.00153
1.2	.48676	3.7	.00106
1.3	.42956	3.8	.00073
1.4	.37631	3.9	.00050
1.5	.32465	4.0	.00034
1.6	.27804	4.1	.00022
1.7	.23575	4.2	.00015
1.8	.19790	4.3	.00010
1.9	.16448	4.4	.00006
2.0	.13534	4.5	.00004
2.1	.11025	4.6	.00003
2.2	.08892	4.7	.00002
2.3	.07100	4.8	.00001
2.4	.05614	4.9	.00001
		5.0	.00000

(註一)(註二)見下頁。

表內 x/σ 行下所列為 x 軸上各數值離平均數點之距離而以標準差為單位者。相對於各距離之縱坐標，則列於 y/y_0 行下，係以各縱坐標對最高縱坐標之比率表示者。例如離平均數點 1σ 處之縱坐標等於 $.60653y_0$ ， 2.8σ 處之縱坐標等於 $.01984y_0$ 。故欲就一種次數分配配以常態曲綫，須先將該分配中各項對其平均數之距離用 σ 之單位表示；次斷定其最高縱坐標之數值；最後應用縱坐標表以計算各縱坐標之數值。最高縱坐標之數值可由下式求之。

$$y_0 = \frac{N}{\sigma\sqrt{2\pi}}$$

將 $\sqrt{2\pi}$ 之數值代入，則公式為

$$y_0 = \frac{N}{2.506628\sigma}$$

在本例中

$$\begin{aligned} N &= 995 \\ \sigma &= \sqrt{\mu_2} \\ &= \sqrt{8.718146} \\ &= 2.953 \end{aligned}$$

故

$$\begin{aligned} y_0 &= \frac{995}{2.506628 \times 2.953} \\ &= 134.42 \end{aligned}$$

此項分配之算術平均數為476.96。

其他計算之結果見表一二六。

(上頁註一)此項次數分配如用披爾遜氏第四類形式之曲綫配合之，固較適當，惟用常態曲綫配合亦尚相稱，且其配合手續尤較簡易也。

(上頁註二)在披爾遜氏所著之“Tables for Statisticians and Biometricians”可得一詳盡之常態曲綫縱坐標表。該表所列縱坐標之數字，係以橫坐標為單位者。

在應用 y_0 之公式及計算表一二六內之各項數值時，標準差須以組距為單位，蓋實在次數既以組表示，則求得之計算次數亦應按組表示，以便比較也。

表 一 二 六

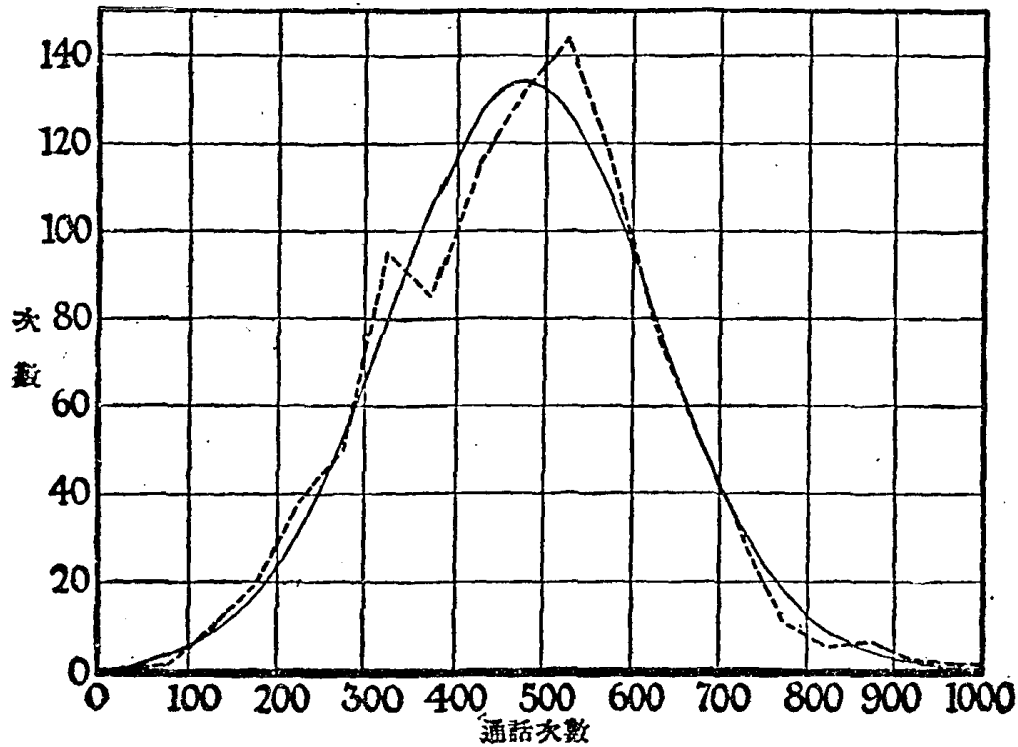
配合於電話用戶次數分配之常態曲綫時縱坐標之計算

通話次數 m	對於算術平均數 之離中差 (以組距為單位) x	$\frac{x}{\sigma}$	$\frac{y}{y_0}$	計算次數 (縱坐標) y_c
25	-9.0392	-3.06	.00926	1.24
75	-8.0392	-2.72	.02474	3.33
125	-7.0392	-2.38	.05888	7.91
175	-6.0392	-2.05	.12230	16.44
225	-5.0392	-1.71	.23176	31.15
275	-4.0392	-1.37	.39123	52.59
325	-3.0392	-1.03	.58534	79.08
375	-2.0392	-.69	.78817	105.95
425	-1.0392	-.35	.94055	126.43
475	-.0392	-.01	.99995	134.41
525	.9608	.32	.94856	127.71
575	1.9608	.66	.80429	108.11
625	2.9608	1.00	.60653	81.53
675	3.9608	1.34	.40747	54.77
725	4.9608	1.68	.24385	32.78
775	5.9608	2.02	.13000	17.47
825	6.9608	2.36	.06174	8.30
875	7.9608	2.70	.02612	3.51
925	8.9608	3.03	.01015	1.36
975	9.9608	3.37	.00342	.46
1025	10.9608	3.71	.00103	.14
1075	11.9608	4.05	.00027	.04
				<hr/> 994.71

曲綫配合適度之測驗

按通話次數分組之電話用戶實在次數分配及其所配之常態曲綫見圖八十六。觀圖可知常態曲綫配合於此項資料尚屬適當，惟有數點兩者相差較大耳。常態曲綫之不能與各點完全適合者，其原因不外兩端：其一，或由於抽樣之偶然變化 (chance fluctuations)，蓋此項偶然變化為任何樣本所難免。按通話次數分組之電話用戶次數分配或本有一基本法則之存在，而此法則又或與差誤常態法則完全符合，祇以所取樣本偶

有不規則之狀態，致與差誤常態曲綫有矛盾之處，是即抽樣之偶然變化也。如將抽樣所包含之項數逐漸增多，則此不規則狀態亦可逐漸減少。其二，實在次數之分配與常態曲綫之差異，或由於電話用戶之次數分配本與差誤常態法則不合，故根本不能用常態曲綫表示所致。



圖八十六。按通話次數分組之電話用戶次數分配及其所配之常態曲綫

上述兩種解釋究以何者為是，可由兩點觀察之。吾人可就實在次數與計算次數相差最大之各組加以研究，以斷定此種差異是否僅由於樣本之偶然變化。吾人又可就曲綫之全部加以研究，以斷定實在分配與差誤常態法則是否相符，如屬相符，即可斷定實在次數與計算次數之差異是否由於抽樣之偶然變化。此兩種觀察方法皆可採用之。

惟應用此兩種觀察方法時，必先有更精密之理論次數。常態曲綫之縱坐標所表示之理論次數並不十分精確。此曲綫與闊度為一組距單位所成之直條，倘其上端之曲綫部分成直綫時，則直條中點之縱坐標始可準確表示該條內所包含之面積。此在常態曲綫則不然，故吾人必須測度

曲綫下面積以斷定理論次數。

理論次數之斷定

次數分配之次數總和可用該次數曲綫下之全面積表示之，故若吾人能知曲綫下各部分之面積對於全面積之比，則各部分之次數即不難求得。常態曲綫下各部分面積對全面積之比可用機率積分表求之。表一二七所示爲一簡單之機率積分表。如計算時欲求準確，當採用更精密之表(註)。

常態曲綫在其最高縱坐標之兩邊完全對稱，故下表內各數值可適用於平均數點之左右兩邊。

表 一 二 七
常態曲綫下之面積

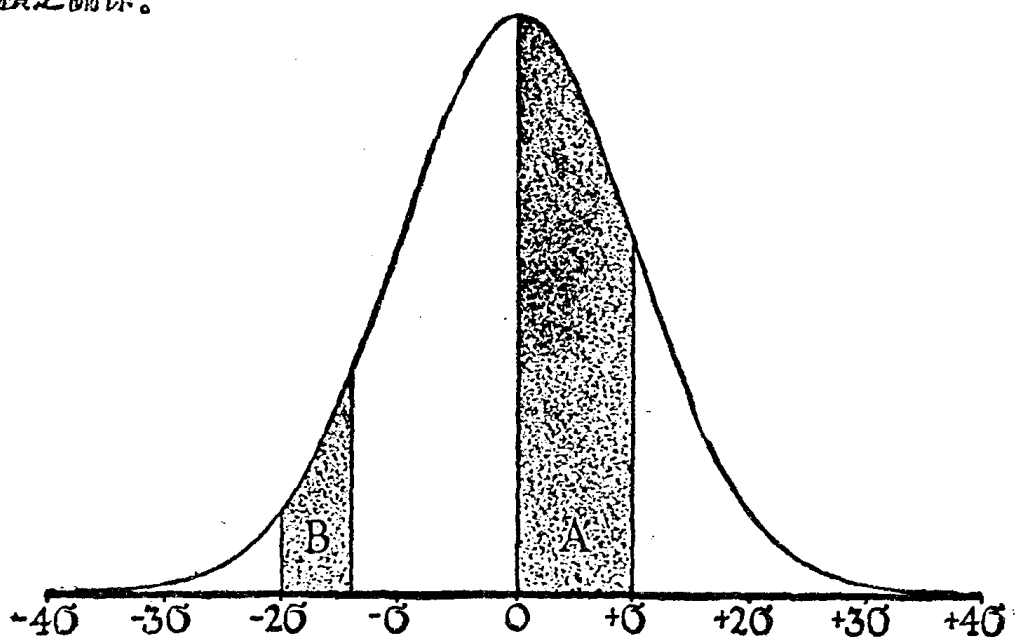
(表內所載爲各縱坐標與 y_0 所包含之面積對於全面積之比)

x/σ	a	x/σ	a
0.0	.00000	2.0	.47725
0.1	.03983	2.1	.48214
0.2	.07926	2.2	.48610
0.3	.11791	2.3	.48928
0.4	.15542	2.4	.49180
0.5	.19146	2.5	.49379
0.6	.22575	2.6	.49534
0.7	.25804	2.7	.49653
0.8	.28814	2.8	.49744
0.9	.31594	2.9	.49813
1.0	.34134	3.0	.49865
1.1	.36433	3.1	.49903
1.2	.38493	3.2	.49931
1.3	.40320	3.3	.49952
1.4	.41924	3.4	.49966
1.5	.43319	3.5	.49977
1.6	.44520	3.6	.49984
1.7	.45543	3.7	.49989
1.8	.46407	3.8	.49993
1.9	.47128	3.9	.49995
		4.0	.49997

(註)舍巴德氏(W. F. Sheppard)所製之機率積分表，頗爲詳備，應用甚廣。披爾遜氏之“Tables for Statisticians and Biometricians”內亦列有此項機率積分表，其形式極合實用。

應用該表時，各縱坐標對平均數點之距離須先化為標準差之單位，則任何兩縱坐標間所含面積對全面積之比，即可由該表求得。例如欲求最高縱坐標與平均數點相距 $+\sigma$ 處所豎之縱坐標間之面積對全面積之比，可查表中 $\frac{x}{\sigma} = 1$ 得 .34134，此即該距離內所含面積對全面積之比。此比率如以百分比表示，則為 34.134%。

圖八十七表示該距離內之面積（圖中黑暗部分之 A）對曲綫下全面積之關係。



圖八十七. 測度差誤常態曲綫下面積之圖解

設由離平均數點 -1.4σ 及 -2σ 之兩點各豎縱坐標與常態曲綫相交，則此兩縱坐標間所包含之次數對次數總和之比究為幾何？由表中可得 y 。至 -1.4σ 處所豎縱坐標間之面積為全面積之 41.92%； y 。至 -2σ 處所豎縱坐標間之面積為 47.73%。兩者相減得 5.81%，亦即 -1.4σ 與 -2σ 間所含面積對全面積之百分比也。此百分比如以次數之總和乘之，即可求得該部分之理論次數。圖八十七中黑暗部分之 B 即代表該部分之面積者也。

由上舉例當可明瞭由面積表求理論次數之方法。吾人僅須查明最

高縱坐標與由各組限處所豎縱坐標間之面積對全面積之比，而後用簡單之減法，求各組所包含之面積，即可由此求得各組之理論次數。下表係就電話用戶之次數分配，以說明計算理論次數之方法。

表 一 二 八
由面積表計算理論次數

(1) 組 限	(2) 對於算術平均 數之離中差 $\frac{x}{\sigma}$	(3) y_0 與在 $\frac{x}{\sigma}$ 處各 縱坐標間所包 含之面積對全 面積之比	(4) y_0 與 $\frac{x}{\sigma}$ 處各 縱坐標間所 包含之次數	(5) 各組之理論次數	
0	-3.23	.4993810	496.88		
50	-2.89	.4980738	495.58	0-50	1.92(註)
100	-2.55	.4946139	492.14	50-100	3.44
150	-2.22	.4867906	484.36	100-150	7.78
200	-1.88	.4699460	467.60	150-200	16.76
250	-1.54	.4382198	436.03	200-250	31.57
300	-1.20	.3849303	383.01	250-300	53.02
350	-.86	.3051055	303.58	300-350	79.43
400	-.52	.1984682	197.48	350-400	106.10
450	-.18	.0714237	71.07	400-450	126.41
500	+.16	.0635595	63.24	450-500	134.31
550	+.495	.1896931	188.74	500-550	125.50
600	+.83	.2967306	295.25	550-600	106.51
650	+1.17	.3789995	377.10	600-650	81.86
700	+1.51	.4344783	432.31	650-700	55.21
750	+1.85	.4678432	465.50	700-750	33.19
800	+2.19	.4857379	483.31	750-800	17.81
850	+2.53	.4942969	491.83	800-850	8.52
900	+2.87	.4979476	495.46	850-900	3.63
950	+3.20	.4993129	496.82	900-950	1.36
1000	+3.54	.4997999	497.30	950-1000	.48
1050	+3.88	.4999478	497.45	1000-1050	.15
1100	+4.22	.4999878	497.49	1050以上	.05
					<u>995.00</u>

(註)按此項理論分配在 -3.23σ 以下尚有次數.62,但在本例為無意義,故將該次數併入0-50一組內。

抽樣標準誤之應用於測驗曲綫配合之適度

吾人既已求得較為準確之各組理論次數，乃可據此推測常態曲綫配合於電話用戶之次數分配是否適當。茲先就實在次數與理論次數相差最大之各點加以研究，以斷定此項差異之意義。本書前嘗論及，倘吾人已知成敗之機率，則標準差可由下式求之。

$$\sigma = \sqrt{npq}$$

式中 p 代表成就之機率， q 代表失敗之機率。在處理次數分配時，可用 $\frac{f}{N}$ (f 代表 x 尺度上某一點之理論次數， N 代表次數之總和)以代替 p ，則 q 等於 $\frac{N-f}{N}$ 。將兩者代入上項標準差之普通公式中，則

$$\begin{aligned}\sigma_s &= \sqrt{N \times \frac{f}{N} \times \left(\frac{N-f}{N}\right)} \\ &= \sqrt{\frac{f(N-f)}{N}}\end{aligned}$$

此量數稱為抽樣標準誤(standard error of sampling)。

茲將上例中實在次數與理論次數相差最大之各組列表於下：

(f_0 代表實在次數， f 代表理論次數)

m	f_0	f	$f_0 - f$
325	95	78.63	+16.37
375	85	106.90	-21.90
525	144	125.50	+18.50
775	11	17.81	-6.81

上表內第一組之抽樣標準誤為

$$\sigma_s = \sqrt{\frac{78.63(995-78.63)}{995}} = 8.51$$

其他三組之抽樣標準誤亦可依公式求得。茲將各組之 σ_s 數值列下：

中點為325一組之 $\sigma_s = 8.51$

中點為375一組之 $\sigma_s = 9.73$

中點爲525一組之 $\sigma_s = 10.47$

中點爲775一組之 $\sigma_s = 4.18$

電話用戶之次數分配如確依差誤常態之法則，則前述實在次數與理論次數之差異是否可能？此問題現時當可作確切之解答矣。按中點爲325之一組，其實在次數與理論次數之差數爲+16.37，此組之 σ_s 爲8.51，故差數合 σ_s 之1.9倍。如該次數分配確係合乎差誤常態法則，則此差數在實際上是否可能乎？

此問題可用機率積分表解答之(表一二七)。該表內 $\frac{x}{\sigma} = 1.9$ 之數值爲.47128，此即謂最高縱坐標與離平均數點 1.9σ 處所豎縱坐標之間，約包含曲綫下全面積之47%。在離平均數點 $\pm 1.9\sigma$ 處所豎之兩縱坐標間約包含全面積之94%，故在機率6%中，一數值與平均數之相差亦可大於 1.9σ 。此差異之數，或由於抽樣之偶然變化，並不能證明常態曲綫配合於電話用戶之次數分配爲不當也。

其他各組之差數亦可作同樣之測驗。在各組中以中點爲375一組之差數爲最大。此差數爲-21.90，合 $2.2\sigma_s$ 。查表可知在機率1.5%中，一數值與平均數之負差可等於或大於 2.2σ 。若將正負差數併計，則在機率3%中，實在次數與理論次數之相差可等於或大於上述差數。故就本例中之最大差數而論，此差數在常態分配條件之下亦屬可能，而不能遽以證明常態曲綫之不能用以表示電話用戶之次數分配也。

測定曲綫配合適度之 χ^2 測驗法

上法僅就實在次數與理論次數相差較大之各組加以研究，吾人若將次數分配內所有各組及其差數一併研究，則所得結果可更確切。披爾遜氏(Karl Pearson)曾創 χ^2 測驗法(Chi-square test)，應用次數分配之全部資料以測驗曲綫配合之適度。吾人既知實在次數，又得理論次數，則 χ^2 之計算甚易。茲舉例說明於下：

表 一 二 九

測驗曲綫配合之適度時 χ^2 之計算

差誤常態曲綫配合於電話用戶之次數分配

(1) 組限	(2) 實在次數 f_i	(3) 理論次數 f	(4) $(f_o - f)$	(5) $\frac{(f_o - f)^2}{f}$
150 以下	10	13.14	-3.14	.75
150-200	19	16.76	+2.24	.30
200-250	38	31.57	+6.43	1.31
250-300	50	53.02	-3.02	.17
300-350	95	78.63	+16.37	3.41
350-400	85	106.90	-21.90	4.49
400-450	115	126.41	-11.41	1.03
450-500	132	134.31	-2.31	.04
500-550	144	125.50	+18.50	2.73
550-600	116	106.51	+9.49	.85
600-650	79	81.85	-2.85	.10
650-700	54	55.21	-1.21	.03
700-750	31	33.19	-2.19	.14
750-800	11	17.81	-6.81	2.60
800 以上	16	14.19	+1.81	.23
	995	995.00	15 組	$\chi^2 = 18.18$

吾人應注意者，製此表時已將次數分配下端三組及上端六組各併成一組，蓋欲避免曲綫兩端不適當之微小差異過於顯著也。

吾人可用 χ^2 之數值作為測定曲綫配合適度之量數，惟其解釋須用詳細計算表(註)。茲摘錄其一部以說明計算之步驟於下：

已知 χ^2 及 n' 之數值求 P 之數值

χ^2	$n' = 14$	$n' = 15$	$n' = 16$
16	.249129	.313374	.382051
17	.199304	.256178	.318864
18	.157520	.206781	.262666
19	.123104	.164949	.213734
20	.095210	.130141	.171932

(註) W. Palin Elderton 氏曾作此表。可於 Karl Pearson 氏所編之 “Tables for Statisticians and Biometricians” 第26—30頁得之。

表內 χ^2 代表 $\Sigma\left(\frac{(f_o-f)^2}{f}\right)$, n' 代表組數。在本例中

$$\chi^2 = 18.18$$

$$n' = 15$$

當 $n' = 15, \chi^2 = 18$ 時, P 之數值為 .206781; 當 $n' = 15, \chi^2 = 19$ 時, $P = .164949$ 。用插補法可求得 $n' = 15, \chi^2 = 18.18$ 時, $P = .199$ 。

P 為測度機率之量數。如實在次數分配確合乎差誤常態法則, 則由抽樣可得一種相等於此或更劣於此之配合的機率。就本例而言, $P = .199$, 即謂假定實在次數合乎常態時, 由抽樣可得一種相等於此或更劣於此之配合的機率為 $\frac{199}{1000}$, 或 $\frac{1}{5}$ 弱。此項機率並不過小, 故吾人可斷定常態曲綫配合於電話用戶之次數分配尙屬適當, 頗可用以表示該項次數分配。

吾人既已斷定某種分配可用常態曲綫配合, 則凡關於常態曲綫所有之特徵, 皆可適用於此種分配之研究。在原來次數表中, 吾人所知者, 僅以表中所載各組之次數為限, 今於配合常態曲綫後可進而求 x 尺度上任何兩點間所應包含之次數。吾人並可計算某一數值在尺度上任何一點, 或在某點之上, 或在其下之機率。且吾人既已明瞭抽樣之代表性, 即可依此作概括之論斷, 此絕非單獨研究一特殊分配時所可能者也(註)。

關於表述次數分配之註釋

吾人倘利用本章所述各種測度之標準, 則表述次數分配可更詳確, 並可補以前所用各種量數之不足。惟此問題不在本書討論範圍之內, 此處僅能就前未論及之各種量數之性質略加申述耳。

(註)本書前嘗論及常態曲綫僅為次數曲綫之一種, 惟最為重要耳。Karl Pearson 氏曾推演各種次數曲綫標準形態之方程式, 並詳示其特性。至於其他基本形式之次數曲綫, 在本章末所列 Arne Fisher 氏著述之兩書中, 均有討論。各種次數曲綫之詳細研究, 因不在本書範圍之內, 姑不備述。

β_2 之數值係用以測量次數曲綫峯度之一種量數。倘曲綫所表示者為差誤常態分配，而得 $\beta_2 = 3$ ，則該曲綫稱為常態峯的 (mesokurtic)。倘 $\beta_2 < 3$ ，則曲綫稱為平闊峯的 (platykurtic)，其峯度較常態曲綫為扁平。倘 $\beta_2 > 3$ ，則曲綫稱為高狹峯的 (leptokurtic)，其峯度較常態曲綫為峻峭。

次數曲綫偏斜度之量數亦可由 β_1, β_2 各量數之數值計算而得，而其結果則較以前所用公式更為準確。此項計算公式為披爾遜氏 (Karl Pearson) 所創，而用以測量曲綫之偏斜度者也(註)。其式如下：

$$\chi = \frac{\sqrt{\beta_1}(\beta_2 + 3)}{2(5\beta_2 - 6\beta_1 - 9)}$$

將電話用戶之次數分配中所求得 β_1, β_2 之數值代入上式 得

$$\chi = -.05558$$

(如算術平均數大於中位數， χ 之數值為正數；小於中位數，則為負數。本例中算術平均數為 476.96，中位數為 482.39，故其偏斜度之數值為負數)。

再算術平均數與衆數之距離(d)可由下式決定之：

$$d = \chi \times \sigma$$

就電話用戶次數分配之例， σ 之數值為 147.65 (係用原來單位表示者)。將各數值代入上式，得

$$d = -.05558 \times 147.65 = -8.21$$

但

$$M_o = M - d$$

故

$$M_o = 476.96 + 8.21 = 485.17$$

(註) χ 為測量偏斜度之量數， χ^2 則為測驗曲綫配合適度之量數，其意義絕不相同，學者應鑑別清楚。

此數值較用第四章內求衆數之任何方法所得者，爲更接近於真實衆數之數值。

參 考 書

- BOWLEY, ARTHUR L. *Elements of Statistics* (259—286).
- BRUNT, DAVID. *The Combination of Observations* (11—28).
- CARVER, H. C. *Frequency Curves* (In Rietz, H. L., *Handbook of Mathematical Statistics*, 92—119).
- ELDERTON, W. P. *Frequency Curves and Correlation*.
- FISHER, ARNE. *An Elementary Treatise on Frequency Curves. The Mathematical Theory of Probabilities*.
- KELLEY, TRUMAN L. *Statistical Method* (94—108). (The Kelley-Wood Table of the Normal Probability Integral is given as an appendix.)
- PEARL, RAYMOND. *Medical Biometry and Statistics* (220—263).
- PEARSON, KARL. *Tables for Statisticians and Biometricians*. (The Introduction to these tables will be found particularly useful.)
- SHEPPARD, W. F. *On the Calculation of the Most Probable Values of Frequency Constants for Data Arranged According to Equi-distant Divisions of a Scale*. Proceedings of the London Mathematical Society, Vol. XXIX, 1898.
- The Calculation of the Moments of a Frequency Distribution*. Biometrika, Vol. V.
- YULE, G. U. *An Introduction to the Theory of Statistics* (291—316).

第十六章 統計歸納及抽樣問題

以前各章專論統計分析時所用之各種工具問題，所舉各例亦均注重於技術方面，所以解釋此項工具應用於各種特殊問題之方法者也。吾人現時應高瞻遠矚，以考量關於應用此項工具之一般問題。此中所應注意者爲此項工具在經濟及商業之研究方面所占之地位，以及應用此項工具時所含之假設及其所受之限制。凡此種種問題皆甚重要，惟以本書篇幅所限，未能詳加討論耳。

統計表述及統計歸納

在討論前項問題以前，吾人應明瞭統計表述 (statistical description) 與統計歸納 (statistical induction) 之區別。吾人應用統計方法可將大量之數字資料作簡略之表述。成千累萬之單獨事項可用分組方法化簡而成次數分配。此種分配之特質又可用四種量數——集中趨勢、離散度、偏斜度及峯度之量數——表述之。以有限之量數表述萬千繁複之單獨事項，其間所獲之裨益自非淺鮮。此項工具之應用可使吾人有限之理解力得領會大量事項之要義。統計方法又可用以表述變量間之關係。吾人可就研究之資料，應用數學方法配以相當曲綫，而求得該綫之方程式，更由此方程式以斷定一個或數個相關變量變動時，另一個相關變量平均變動之程度。此項方程式之意義又可用測度曲綫兩旁散佈程度之量數及測度一個倚變量與一個或數個自變量間相關程度之抽象量數以補充之。

惟用上述各種量數求得之結果，其計算所根據者，僅以實際搜集供作研究之資料爲限，故此種量數不過爲表述該項資料之特質或其變量間之關係之工具。在該項資料範圍以內，此種量數固可準確表述其各種特質，但若吾人欲將求得之結果伸引至資料範圍以外，認爲可以適用於

未搜集之資料，而作為概括論斷之根據，是則又成一新問題矣。

由特殊事例之研究，歸納而得概括論斷之方法，在論理學上謂之歸納法 (induction)，此與由概括之論斷以推斷特殊事例之演繹法 (deduction) 相反。所謂統計之歸納 (statistical induction) 或統計之推論 (statistical inference) 者，係假定由特殊事例研究所得之結果，可適用於較原來資料包含更廣之事例，而歸納為概括論斷之方法也。吾人在統計實際工作方面，無不採用此法，惟對於前項假定及其所受之限制，未見常能充分瞭解耳。此項假定及限制可簡略討論之。

根據統計分析之結果作概括論斷之問題

吾人現時所欲討論者，為統計歸納是否成立之問題。此問題可論述於下：設由一大羣衆 (population) 中抽取一部分資料作為樣本 (sample) 而已求得其一種統計量數——如平均數、次數比率 (frequency ratio) 或相關係數 (此處所用“population”一字係指具有某種共同特質之現象或一羣事物而言)。如由此同一羣衆中連續抽取數個樣本而分別計算其同屬一種之統計量數，吾人可否假定由此各個樣本中所得統計量數之數值與前所求得者完全相同？如不相同，則其差異之近似限度亦可根據一特殊事例所得之結果以斷定之否？此為統計學上最重要之問題，故披爾遜氏稱之為實用統計上之基本問題。倘吾人對於各個樣本求得之結果，不能斷定其差異之限度，則由有限之事例所得之結果，自不能作為概括論斷之根據。是以研究時非將羣衆所由組成之一切現象及事物全部包括在內，則所得結論必無價值，然吾人研究時欲將經濟現象之全部網羅無遺，亦為事實所不可能。例如物價指數、工資指數、生活費指數、表示物品產量與物價關係之方程式、溫度與玉蜀黍收穫量之相關係數等各種統計，其勢皆不得不應用抽樣法，就抽取之樣本加以研究也。

自然現象劃一性之假定

根據連續抽取之樣本所求得之統計量數，如能適合兩種假定，則此種量數之數值即可假定其相當穩定而不含顯著之差異。其第一種假定爲自然現象之劃一性，即謂在自然界中，就一種現象而論，其個別變化必有一定之限度。在處理數量資料時，此種自然現象之劃一性表現於大數穩定之事實，例如人類之生殖率及死亡率，每能保持其常態而無大變化。自然現象非絕對錯綜紛亂者，在自然界之各種演變中，莫不有整齊有序與穩定之原理可尋，此項原理在大量數字資料中尤爲顯著。吾人在根據一種量數如批發物價指數以作概括論斷時，實際已含有如下之假定：在一大羣衆中，其特質及關係常有劃一之傾向，故由抽樣所得之結果可適用於整個之羣衆。根據此劃一性之假定，即可推論由羣衆中連續抽取之樣本計算而得之統計量數，其變化必有一定之限度，而此限度吾人可預測其大概。

惟此種劃一性之假定不能全用純粹之統計方法證實之，故在統計歸納中常含有一種不能證實之成分 (a priori element)。統計結論之本身未見皆能成立，吾人自不可妄加信賴，而須憑理解力及判斷力以衡量之。設由樣本之研究而得香蕉進口數量與自殺率之高度正相關係數，吾人斷不能遽以此係數作爲證明該兩變量間因果關係之有力證據，蓋吾人絕無任何理由可假定在該現象之全域內亦有此種關係之劃一性之存在也。

樣本代表性之重要

第二種假定實已包含於第一種假定之內。此項假定即謂由羣衆中抽取之樣本，具有代表羣衆全體之資格是也。統計歸納之有無價值，全視吾人能否求得一具有代表性之樣本以爲斷。

然則吾人將如何求一具有代表性之樣本乎？由一大羣衆中抽樣時，其最要條件爲隨手 (random) 抽取，不可存有取捨之偏見，務使全體羣衆中之每一項目皆有被抽爲樣本之均等機會。據凱恩氏 (J. M. Keynes) 之意見，謂此項抽樣之普通要件，應解釋如下：在對某種問題作概括之論斷時，則由該問題有關之羣衆中抽作樣本之項目應爲隨手抽取之項目。抽樣時雖謂隨手抽取，亦宜極端審慎，蓋在羣衆之中每有若干項目最易抽取，其被抽爲樣本之機會較大，此與上述條件不能適合。故若抽樣時稍不留意，每發生無意識之偏誤，以致所抽之樣本常有將全體現象所含之重要原素於無意中脫漏之虞。

簡單抽樣之條件

關於簡單抽樣之要件，除以上所舉者外，猶爾氏 (G. U. Yule) 亦嘗有所指陳(註)。此項要件在推演關於抽樣之公式時，本爲默認之事實。茲分述於下：

1. 樣本內所包含之各事件應完全獨立。例如一種物品之價格變動足以影響另一種物品之價格變動時，則此兩種物品之價格變動並非獨立；倘將兩者同時包含於一個樣本內，則與簡單抽樣之條件不合。

2. 倘吾人之樣本係由各地方抽取而得，則各地方之本質須無根本不同之點；在觀察之全時期內，各種情形亦須無重大之變化。

3. 凡各種情形之足以影響觀察事件之特質者，則抽樣時不僅各個樣本所處之情形務須相同，即每個樣本中各個項目所處之情形亦必求其相同。

吾人於抽樣時倘能注意上述三要件，則由同一羣衆中所取各個樣本計算而得之統计量數，其數值之差異限度即可加以預測。此其意義亦

(註)參看 G. U. Yule 氏著 "Introduction to the Theory of Statistics" 第五版第

259—261頁。

即吾人由樣本之研究所得之統計量數，倘應用於整個羣衆，作為概括論斷之根據，則此種量數雖不能謂絕對穩定，但可約略推知該項論斷所可發生錯誤之限度。吾人果能遵守此數要件，則統計之歸納始為合理而有價值。

可靠程度量數之意義

測量穩定程度之量數(measures of stability)，其意義若何，應有透切瞭解之必要。當吾人已知統計量數之數值而作概括之論斷時，實即將一個樣本所得之平均數、標準差或相關係數等統計量數應用於全部也。此時吾人對於根據同一羣衆中所取各樣本計算而得之結果，其差異之限度須加斷定，而測量此種限度之量數，即可用作測量前此所得結果可靠程度之量數(measures of reliability)。

此項量數可由連續採取許多不同之樣本(如擲骰試驗之例，吾人可擲4096次)實驗而得。吾人設欲測驗某種工人每週工資平均數之可靠程度，則可先行抽取包含250人工資數之樣本而計算其平均數。為測驗此平均數之可靠程度起見，另行抽取樣本499個，每個樣本各包含250人之工資數，而一一計算其平均數。此500個平均數之數值未必相同，但若編為次數表，則該次數分配必與常態分配極為接近。由此分配可計算500個平均數之平均數及其標準差，而此標準差即可用以測量由各個樣本所得每週工資平均數之差異程度。

但吾人欲抽取400或500個樣本以確定一統計量數之可靠程度，每為事實上所不可能。倘取樣時能注意簡單抽樣之要件，則可靠程度之量數可應用公式直接求得。

在前章內嘗將每次同時擲骰十二粒，以每骰擲得4,5,6點為成就，共擲4096次之試驗結果編成次數分配，並求得標準差之數值為1.712。但因吾人已知每次擲骰之骰子數及其成敗之機率，故其理論標準差可

由下式求之：

$$\sigma = \sqrt{npq}$$

其所得之數值為1.732。

如由樣本求得之統計量數，其數值不穩定之原因純由於抽樣之變化 (fluctuations of sampling)，則吾人亦可應用上述求標準差之同樣方法，依照公式以計算統計結果可靠程度之量數。茲將測量可靠程度所用之各種量數或統計量數標準誤之公式分列於下。此項公式之來源可參看本章末所列之參考書，茲不具論。

主要統計量數之標準誤

由一個樣本所求得之算術平均數，其可靠程度須視其標準差之數值與樣本內所包含之項目數而定。其關係可用下式表之：

$$\sigma_M = \frac{\sigma}{\sqrt{N}}$$

(符號旁未加註小字之 σ 代表樣本之標準差) 此式係用標準誤表示抽樣之算術平均數之可靠程度，惟通常係用機誤表示之。在常態分配中機誤等於 .6745 σ ，故

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

上式之意義可舉例說明之。設吾人欲求某城工人100,000人之某週平均工資數。如將全體工人之某週工資數列入計算殆不可能，故非採用抽樣法不可。今設由全體中抽取具有代表性之900人之工資資料作為樣本，而求得其某週平均工資數為27.50元。然則此平均數之準確程度若何？此數值與根據其他樣本所得之平均數相差之限度又若何？假定該樣本之標準差為2.00元，則代入公式得

$$P.E._{Mean} = .6745 \frac{\$2.00}{\sqrt{900}} = \$.045$$

故吾人在發表抽樣研究之結果時，此某週平均工資數應寫作 \$27.50 \pm .045\$。此其意義謂如在全體中另行抽取項目相同之樣本，則由此樣本所得之平均工資在 $27.50 + .045$ 元與 $27.50 - .045$ 元間之機率為二分之一。

其他統計量數之準確程度亦可用同法求之。以下所列為最重要統計量數之標準誤之公式(各種量數之機誤可以 .6745 乘其標準誤得之)。

$$\sigma_{\bar{M}_d} = 1.25331 \frac{\sigma}{\sqrt{N}} \quad P.E._{\bar{M}_d} = .84535 \frac{\sigma}{\sqrt{N}}$$

$$\sigma_{Q_1} = 1.36263 \frac{\sigma}{\sqrt{N}} \quad P.E._{Q_1} = .91908 \frac{\sigma}{\sqrt{N}}$$

(Q_3 之標準誤及機誤數值與 Q_1 相同)

$$\sigma_{\sigma} = \frac{\sigma}{\sqrt{2N}}$$

根據抽樣所得之次數分配為常態形式時，標準差之標準誤始可應用上項公式。至於由其他形式之次數分配求 σ 之可靠程度，須用下式決定之。

$$\sigma_{\sigma} = \sqrt{\frac{\mu_4 - \mu_2^2}{4\mu_2 \cdot N}}$$

在常態分配中，相關係數之可靠程度可用下式測度之。

$$\sigma_r = \frac{1-r^2}{\sqrt{N}}$$

複相關係數及淨相關係數之標準誤，亦可用上式求之。如用此式求相關率之標準誤，亦可得其近似數值。

倘次數分配合乎常態，則迴歸係數之標準誤可用下式求之。

$$\sigma_{b_{12}} = \frac{\sigma_1 \sqrt{1-r^2}}{\sigma_2 \sqrt{N}}$$

此式可適用於各種次數(order)之迴歸係數。

關於相關綫之是否屬於直綫形式，前曾採用下式測驗，

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

但吾人欲決定 η^2 與 r^2 之差,是否僅由於抽樣之變化,抑由於相關綫之本為非直綫形式,則須求 ζ 之標準誤。勃雷克門氏(Blakeman)所創 ζ 之標準誤公式為

$$\sigma_{\zeta} = 2\sqrt{\frac{\zeta}{N}}\sqrt{(1-\eta^2)^2 - (1-r^2)^2 + 1}$$

此量數之應用方法,可就前此所舉小麥收穫量之例以說明之。在小麥收穫量與淡氣用量之相關問題中,曾求得

$$r = +.793$$

$$\eta = .964$$

故

$$\zeta = \eta^2 - r^2 = .300$$

將各數值代入下式

$$\sigma_{\zeta} = 2\sqrt{\frac{\zeta}{N}}\sqrt{(1-\eta^2)^2 - (1-r^2)^2 + 1}$$

得

$$\sigma_{\zeta} = .074$$

ζ 之數值為.300,約等於 ζ 之標準誤之4.05倍,由此可斷定 η^2 與 r^2 之相差,並不由於抽樣之變化,而由於兩變量間關係本為非直綫形之故也。

曲綫合乎常態時,用以測量偏斜度之 X 之標準誤可由下式求之。

$$\sigma_x = \sqrt{\frac{3}{2N}}$$

應用前章內電話用戶次數分配之資料,可求得

$$\sigma_x = \sqrt{\frac{3}{1990}} = .03883$$

該次數分配中, X 之數值為 $-.05558$,約等於 X 之標準誤之1.43倍,故知 X 之數值或由於抽樣之變化而發生,而不能證明真實分配之不合乎常態。

曲綫合乎常態時,測度算術平均數與衆數距離 d 之標準誤,可用下式求之。

$$\sigma_d = \sqrt{\frac{3}{2N}} \sigma$$

就電話用戶次數分配之資料，將各數值代入上式，得

$$\begin{aligned} \sigma_d &= \sqrt{\frac{3}{1990}} 147.65 \\ &= 5.733 \end{aligned}$$

前曾求得 d 之數值為 8.21，尚不及其標準誤之一倍半，故吾人可斷定算術平均數與衆數之距離乃由於取樣之變化而來。如 d 之數值三倍於其標準誤，則此距離不可謂僅由於抽樣之變化而發生，而或由於次數分配不合乎常態之所致。

在實用統計上，常有比較兩個平均數之重要問題發生。例如抽取兩個樣本分別計算其平均數，而欲斷定此兩個平均差異之原因，是否由於抽樣之變化，抑由於兩個樣本所由抽取之兩大羣衆本質上不同之所致。此時吾人須用兩個平均數相差數之標準誤公式決定之。其式如下：

$$\sigma_D = \sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}$$

式中 σ_1 及 σ_2 代表兩個樣本之標準差， N_1 及 N_2 代表兩個樣本之項數。由上式求得 σ_D 之數值應與兩個平均數之相差數比較。如差數大於 σ_D 之數值三倍以上，則此差數恐非由於抽樣之變化，而或由於兩個樣本所由抽取之兩大羣衆在本質上不同之故。

抽樣標準差之公式常用於比較實在次數與理論次數間之差異。其式如下：

$$\sigma_s = \sqrt{\frac{f(N-f)}{N}}$$

此式之功用在前章內曾有說明。

以上所述可靠程度之各種量數多以標準誤表示，而不用機誤表示。吾人既已求得標準誤，則機誤之數值甚易求得，以 .6745 乘標準誤即得，惟此項可靠程度之量數，通常以用標準誤表示較為便利，因普通所用較

爲詳備之機率積分表，其橫坐標之距離大都用標準差單位表示，且任何標準誤之機率及其意義，亦可由此項機率積分表決定也。

可靠程度之量數在應用上之限制

在應用上述各種可靠程度之量數時，吾人應明瞭其所受之種種限制。其第一種限制爲樣本內所包含之項數必須衆多，如是則可靠程度之量數始有意義。倘項數 N 少於15，前述各標準誤之公式即不可用，大抵樣本內包含之項數愈多，則標準誤之價值愈高。求相關係數時，包含之項數不得少於25。蓋在解釋各標準誤時，吾人嘗假定由連續抽取之各個樣本中求得之統計量數，其次數分配常與差誤常態法則相合，當樣本中所包含之項數衆多，則雖原來資料分配之形式不合乎常態，此項假定仍近乎真實，但若項數過少，則此假定即難成立，而標準誤機率之意義亦必因之消失矣。

可靠程度量數在應用上之第二種限制尤爲重要。吾人係假定標準誤僅能用以測度由簡單抽樣所發生之差誤，惟依此假定可使吾人對於任何統計歸納發生懷疑，其故有二：吾人就所取樣本加以研究時，對於簡單抽樣之條件曾否完全履行，不易斷定，而在處理經濟資料時，此種條件尤難絕對適合，一也。根據前述公式求得之標準誤，除抽樣之變化外，絕不能指明足使連續抽樣所得結果發生差異之其他原因，實際此項差異之發生或由於某種偏誤之存在，或由於抽樣時未能遵守隨手抽取之原則，或由於其他任何繼續存在之差誤，凡此種種原因足以減損標準誤之功用，二也。由此可見吾人應用此種量數時，對於上述各點切宜注意，以免誤用，庶幾統計分析之結果可不致發生曲解。

測量機誤之量數應用於經濟資料，既受上述種種限制，妨礙孔多，故欲斷定抽樣之可靠程度，應以採取統計實驗方法爲主，而以計算該項量數爲輔。在實驗時，吾人不妨多取樣本，或將一個樣本分爲有意義之

若干附屬部分而研究之，俾能斷定統計量數之可靠程度與可否用作概括論斷之根據，是則遠較貿然採用標準誤或機誤之數學公式為愈矣。

參 考 書

- BOWLEY, ARTHUR L. *Elements of Statistics* (312—342).
- BROAD, C. D. *On the Relation between Induction and Probability*. *Mind*. N. S. Vol. 27, 1918, and Vol. 29, 1920.
- EDITORIAL. *On the Probable Errors of Frequency Constants*. *Biometrika*. Vol. 2 (273—281).
- ELDERTON, W. P. *Frequency Curves and Correlation* (131—138).
- FISHER, ARNE. *The Mathematical Theory of Probabilities*.
- KELLEY, TRUMAN L. *Statistical Method* (94—108).
- KEYNES, J. M. *A Treatise on Probability*.
- MILLS, FREDERICK C. *On Measurement in Economics* (in Tugwell, R. G. ed. *The Trend of Economics*, 37—70).
- PEARL, RAYMOND. *Medical Biometry and Statistics* (209—219).
- PEARSON, KARL. *The Fundamental Problem of Practical Statistics*. *Biometrika*. Vol. 13.
- RIETZ, H. L. *Random Sampling* (in Rietz, H. L. ed. *Handbook of Mathematical Statistics*, 71—81).
- WHITAKER, E. T. and ROBINSON, G. *The Calculus of Observations* (164—208).
- YULE, G. U. *An Introduction to the Theory of Statistics* (254—290, 335—356).

附 錄 A

最小平方法在統計上之應用

應用最小平方法以測量一個單獨未知數量時，祇須根據觀察所得之各個數值中求其最可能之數值 (the most probable value)，此最可能之數值為該值與各觀察值相差數 (或稱剩餘數 residuals) 平方之和為最小時之數值，即各觀察值之算術平均數也。

但若吾人所欲測量者非一個單獨未知數量，而為數個未知數量之函數 (functions)，則此問題與前稍異，蓋在前者每次觀察所得為一個單獨數值，而在後者則每次觀察所得為一個觀察方程式 (observation equation)，而為各變量觀察所得之數值彙合而成者。此數個未知數量，乃表示各變量函數關係所包含之各常數也。吾人當前之問題，因未知各個常數之真實數值，而欲求其最可能之數值。

在測量數個未知數量之函數時，此最可能之數值亦為剩餘數平方之和為最小時之數值，惟此剩餘數並非對於一個單獨數值 (算術平均數) 所求得之差數，而為對於表示其最可能之函數關係之曲綫所求得之距離；換言之，此剩餘數亦即倚變量之實際數值與計算數值 (computed value) 之差也。

常態方程式之求法

設以 Y 代表倚變量之實際數值， Y_0 代表其計算數值， v 代表剩餘數，即 Y 與 Y_0 之差數， W_1, W_2, W_3, W_4 代表各個自變量 (或一個自變量所包含之各個不同函數)，則

$$Y_0 = f(W_1, W_2, W_3, W_4)$$

$$v = Y_0 - Y$$

$$= f(W_1, W_2, W_3, W_4) - Y$$

$$\Sigma(v^2) = \Sigma [f(W_1, W_2, W_3, W_4) - Y]^2$$

如在一特例中, Y_0 之形式爲

$$Y_0 = aW_1 + bW_2 + cW_3 + dW_4$$

則

$$\Sigma(v^2) = \Sigma [(aW_1 + bW_2 + cW_3 + dW_4) - Y]^2$$

吾人現時即欲斷定表示上項函數關係之 a, b, c, d 各個常數之最可能數值(式中 W_1, W_2, W_3, W_4 均爲已知數, 因在求得各個觀察方程式後, 各自變量之數值即已確定也。該項符號通常係指一個單獨變量之各個不同函數, 但此非重要之點)。假定觀察差誤之分配合乎差誤常態法則, 則上式內 a, b, c, d 之最可能數值當爲 $\Sigma(v^2)$ 爲最小時之數值, 亦即

$$\Sigma [(aW_1 + bW_2 + cW_3 + dW_4) - Y]^2 = \text{最小值} \quad (A)$$

時之數值也。解常數數值所用之常態方程式可由下法得之: 依次求 a, b, c, d 各未知數之偏微分 (partial derivatives), 而各令其等於零。即先令 b, c, d 不變而求 a 之偏微分, 次令 a, c, d 不變而求 b 之偏微分, 次用同法分別求 c, d 之偏微分。求方程式 (A) 中 a 之偏微分而令其等於零, 得

$$\frac{\partial}{\partial a} \Sigma [(aW_1 + bW_2 + cW_3 + dW_4) - Y]^2 = 0$$

或

$$I \quad \Sigma W_1 [(aW_1 + bW_2 + cW_3 + dW_4) - Y] = 0$$

求方程式 (A) 中 b 之偏微分而令其等於零, 得

$$\frac{\partial}{\partial b} \Sigma [(aW_1 + bW_2 + cW_3 + dW_4) - Y]^2 = 0$$

或

$$II \quad \Sigma W_2 [(aW_1 + bW_2 + cW_3 + dW_4) - Y] = 0$$

求方程式 (A) 中 c 之偏微分而令其等於零, 得

$$\frac{\partial}{\partial c} \Sigma [(aW_1 + bW_2 + cW_3 + dW_4) - Y]^2 = 0$$

或

$$III \quad \Sigma W_3 [(aW_1 + bW_2 + cW_3 + dW_4) - Y] = 0$$

求方程式(A)中 d 之偏微分而令其等於零,得

$$\frac{\partial}{\partial d} \Sigma [(aW_1 + bW_2 + cW_3 + dW_4) - Y]^2 = 0$$

或

$$IV \quad \Sigma W_4 [(aW_1 + bW_2 + cW_3 + dW_4) - Y] = 0$$

將依此求得之四個常態方程式 (即標以 I, II, III, IV 者) 聯立解之, 即可求得 a, b, c, d 各常數之最可能數值。

常態方程式之組成

當各個觀察方程式皆為一次方程式 (即 a, b, c 等未知數均為一次方) 時, 常態方程式可由下法求得:

1. 將表示假定關係之方程式錄出, 更將每次觀察所得各變量之數值同時代入該方程式內, 即得每個觀察方程式。
2. 以每個觀察方程式內第一未知數之係數乘該式內各項, 再將由此所得各式一一相加, 即得第一常態方程式。
3. 以每個觀察方程式內第二未知數之係數乘該式內各項, 再將由此所得各式一一相加, 即得第二常態方程式。

依照上法分別求第三、第四、第五等常態方程式, 至所得之方程式數與未知數之個數相等為止。

實際上配合曲綫時求常態方程式之手續尚可化簡, 而不必將各個觀察方程式一一錄出, 此點前嘗言之。以下為求常態方程式之普通程序。

1. 將所欲配合之曲綫方程式錄出。為便於解釋起見, 吾人可用普遍之形式如下:

$$Y = aW_1 + bW_2 + cW_3 + dW_4 + \dots \quad (1)$$

該式內 Y 為倚變量, a, b, c, d, \dots 等為常數(即現時之未知數), $W_1, W_2, W_3, W_4, \dots$ 等為各未知數之係數。此種係數係代表各自變量。令該式為(1)式。

2. 以第一未知數之係數(即 W_1)乘(1)式內各項, 而將總和之符號 Σ 置於各變量之前, 即得第一常態方程式(I)。
3. 以第二未知數之係數(即 W_2)乘(1)式內各項, 而將總和之符號置於各變量之前, 即得第二常態方程式(II)。
4. 以第三未知數之係數(即 W_3)乘(1)式內各項, 而將總和之符號置於各變量之前, 即得第三常態方程式(III)。
5. 以第四未知數之係數(即 W_4)乘(1)式內各項, 而將總和之符號置於各變量之前, 即得第四常態方程式(IV)。

用同法分別求第五第六等常態方程式, 至所得之方程式數與未知數之個數相等為止(註)。

常態方程式之標準形式

依照前述手續所求得適用於下列曲綫方程式

$$Y = aW_1 + bW_2 + cW_3 + dW_4 + \dots$$

之各個常態方程式之普遍式如下:

$$I \quad \Sigma(W_1 Y)$$

$$= a \Sigma(W_1^2) + b \Sigma(W_1 W_2) + c \Sigma(W_1 W_3) + d \Sigma(W_1 W_4) + \dots$$

$$II \quad \Sigma(W_2 Y)$$

$$= a \Sigma(W_1 W_2) + b \Sigma(W_2^2) + c \Sigma(W_2 W_3) + d \Sigma(W_2 W_4) + \dots$$

$$III \quad \Sigma(W_3 Y)$$

(註)此處所述求常態方程式之程序係錄自 Raymond Pearl 氏所著“Medical Biometry and Statistics”第341頁。

$$= a\Sigma(W_1W_3) + b\Sigma(W_2W_3) + c\Sigma(W_3^2) + d\Sigma(W_3W_4) + \dots$$

$$IV \quad \Sigma(W_4Y)$$

$$= a\Sigma(W_1W_4) + b\Sigma(W_2W_4) + c\Sigma(W_3W_4) + d\Sigma(W_4^2) + \dots$$

將 W_1, W_2, W_3, W_4 等所代表之函數代入上列四式內，則任何曲綫之可用最小平方法配合者均可用該項常態方程式解之。如配合時所用曲綫之方程式爲

$$Y = a + bX + cX^2 + dX^3$$

則吾人可將下列各值

$$W_1 = 1$$

$$W_2 = X$$

$$W_3 = X^2$$

$$W_4 = X^3$$

代入前述之常態方程式內。 $\Sigma(W_1Y)$ 即變爲 $\Sigma(Y)$ ； $\Sigma(W_1^2)$ 變爲 $\Sigma(1^2)$ ，此值等於觀察之總項數 N 。第一常態方程式亦即變爲：

$$\Sigma(Y) = Na + b\Sigma(X) + c\Sigma(X^2) + d\Sigma(X^3)$$

其他各個常態方程式亦須同樣加以變更。

本例中之各係數均係一個單獨自變量 X 之各個不同函數，但在採用最小平方法時此非必要條件， W_1, W_2, W_3 等係數亦可代表各個自變量，如在研究複相關時所用者。

最小平方法在應用上亦有相當限制，不可不加以注意。在直接應用此法時，曲綫方程式內所包含之 a, b, c 等各常數均須爲直綫形式，亦即謂各個觀察方程式對其所含之 a, b, c 等未知數而言，均須爲直綫形也（此與所配曲綫本身之形式無關，該曲綫可爲直綫，亦可爲非直綫）。例如曲綫之方程式爲 $y = abc^x$ 時，即不可直接應用最小平方法以求其常數之數值。但若觀察方程式爲非直綫形，而用對數後可變爲直綫形者，則變成直綫形式後，仍可應用最小平方法。

估量標準誤公式之求法

本書前嘗論及估量標準誤之求法原係最小平方法之副產物。茲詳述該標準誤公式之求法於下。

求下式

$$\Sigma[(aW_1 + bW_2 + cW_3 + dW_4) - Y]^2 = \text{最小值}$$

中 a 之偏微分而令其等於零，可得

$$\Sigma W_1[(aW_1 + bW_2 + cW_3 + dW_4) - Y] = 0$$

但

$$aW_1 + bW_2 + cW_3 + dW_4 - Y = v$$

故

$$\Sigma(vW_1) = 0$$

此為配合曲綫時必要之條件。

求同式中 b 之偏微分而令其等於零，則得

$$\Sigma W_2[(aW_1 + bW_2 + cW_3 + dW_4) - Y] = 0$$

將 v 代入，則

$$\Sigma(vW_2) = 0$$

依同法求 c 及 d 之偏微分而得

$$\Sigma v(W_3) = 0$$

$$\Sigma v(W_4) = 0$$

概括言之，欲用最小平方法以求各未知數之最可能數值，則各未知數之係數 W_1, W_2, W_3, W_4 與 v 間須具有下列之關係，始可應用。

$$\Sigma(vW_1) = 0$$

$$\Sigma(vW_2) = 0$$

$$\Sigma(vW_3) = 0$$

$$\Sigma(vW_4) = 0$$

如上項關係確已存在，則 $\Sigma(v^2)$ 及估量標準誤之數值甚易求得。假定吾人已用最小平方法求得下式

$$Y_c = aW_1 + bW_2 + cW_3 + dW_4$$

內常數之數值，則每個剩餘數 v (即 Y 與 Y_c 之差數) 可用下式表示之：

$$v = aW_1 + bW_2 + cW_3 + dW_4 - Y \quad (1)$$

以 v 乘上式內各項，而後各式相加，得

$$\Sigma(v^2) = a\Sigma(vW_1) + b\Sigma(vW_2) + c\Sigma(vW_3) + d\Sigma(vW_4) - \Sigma(Yv) \quad (2)$$

但

$$\Sigma(vW_1) = 0$$

$$\Sigma(vW_2) = 0$$

$$\Sigma(vW_3) = 0$$

$$\Sigma(vW_4) = 0$$

故

$$\Sigma(v^2) = -\Sigma(Yv) \quad (3)$$

以 Y 乘(1)式內各項，而後各式相加，得

$$\Sigma(Yv) = a\Sigma(W_1Y) + b\Sigma(W_2Y) + c\Sigma(W_3Y) + d\Sigma(W_4Y) - \Sigma(Y^2) \quad (4)$$

將(4)式內 $\Sigma(Yv)$ 之值代入(3)式，得

$$\Sigma(v^2) = \Sigma(Y)^2 - a\Sigma(W_1Y) - b\Sigma(W_2Y) - c\Sigma(W_3Y) - d\Sigma(W_4Y) - \dots \quad (5)$$

由上式求 $\Sigma(v^2)$ 之數值時，可無須求各個剩餘數之數值。凡所配曲綫之方程式屬於下列形式

$$Y = aW_1 + bW_2 + cW_3 + dW_4$$

或用對數、倒數或其他方法可使曲綫方程式化爲此種形式時，均可應用(5)式以求 $\Sigma(v^2)$ 。

在實際應用時，可將代表曲綫方程式中各未知數之實際係數之函數代替 W_1, W_2, W_3, W_4 等。例如所配曲綫之方程式爲

$$Y = a + bX + cX^2 + dX^3$$

時，則

$$\begin{aligned} W_1 &= 1 \\ W_2 &= X \\ W_3 &= X^2 \\ W_4 &= X^3 \end{aligned}$$

將以上各值代入(5)式，得

$$\Sigma(v^2) = \Sigma(Y^2) - a\Sigma(Y) - b\Sigma(XY) - c\Sigma(X^2Y) - d\Sigma(X^3Y) \quad (6)$$

標準誤 S_y 之公式爲

$$S_y = \frac{\Sigma(d^2)}{N} \text{ (註)}$$

上式中 d 爲各項對所配曲綫之差，故此差數 d 實即剩餘數 v 之別名。

當 W_1, W_2, W_3, W_4 爲各個自變量時， Y 之標準誤之普遍公式爲

$$S_y^2 = \frac{\Sigma Y^2 - a\Sigma(W_1Y) - b\Sigma(W_2Y) - c\Sigma(W_3Y) - d\Sigma(W_4Y)}{N} \quad (7)$$

式中 W_1, W_2, W_3, W_4 等亦可用代表曲綫方程式中未知數之實際係數之函數代替之，如前例相同。

相關指數公式之求法

測量兩變量間相關程度所用之相關指數 ρ (*Rho*)之公式爲

$$\rho^2_{yx} = 1 - \frac{S_y^2}{\sigma_y^2} \quad (8)$$

此處變量有二，一爲自變量 X ，一爲倚變量 Y 。如變量不止兩個，而爲個倚變量 Y 及數個自變量 W_1, W_2, W_3, W_4 ，則 ρ 之公式應改爲

(註)吾人之目的既爲測量曲綫兩旁之散佈程度，故當用公式 $\frac{\Sigma(d^2)}{N}$ 而不用 $\frac{\Sigma(d^2)}{N - N_c}$

(式中 N 爲觀察之項數， N_c 爲所配曲綫方程式中所含常數之個數。)，依照最小平方之理論，吾人如欲求觀察值或觀察方程式之均方差誤(mean square error)，應採第二公式。

$$r^2_{y \cdot w_1 w_2 w_3 w_4} = 1 - \frac{S_y^2}{\sigma_y^2} \quad (9)$$

如所用符號有變更時，則 r 旁所註小字應同樣變更。上式亦可寫為

$$r^2_{y \cdot w_1 w_2 w_3 w_4} = 1 - \frac{\Sigma(d^2)}{\Sigma(y^2)}$$

式中 y 為各項對 Y 之算術平均數之離中差。但

$$\frac{\Sigma(y^2)}{N} = \frac{\Sigma(Y^2)}{N} - c_y^2$$

式中 Y 代表 Y 變量之原來數值， c_y 代表原來原點與 Y 之算術平均數之差（符號 c_y, c_x 不可與 c 相混，後者為所配曲綫方程式中之一個常數），

故 r 之公式又可寫為

$$r^2_{y \cdot w_1 w_2 w_3 w_4} = 1 - \frac{\Sigma(d^2)}{\Sigma(Y^2) - Nc_y^2} \quad (10)$$

吾人前已求得用最小平方法配合曲綫時求 $\Sigma(v^2)$ 之公式，而此 $\Sigma(v^2)$ 之數值等於 $\Sigma(d^2)$ 。茲舉一普通含有數個自變量之例，將 $\Sigma(d^2)$ 之等值代入上式，則

$$r^2_{y \cdot w_1 w_2 w_3 w_4 \dots} = 1 - \frac{\Sigma(Y^2) - a\Sigma(W_1 Y) - b\Sigma(W_2 Y) - c\Sigma(W_3 Y) - d\Sigma(W_4 Y) - \dots}{\Sigma(Y^2) - Nc_y^2}$$

將上式化簡，即得相關指數之普遍公式如下：

$$r^2_{y \cdot w_1 w_2 w_3 w_4 \dots} = \frac{a\Sigma(W_1 Y) + b\Sigma(W_2 Y) + c\Sigma(W_3 Y) + d\Sigma(W_4 Y) + \dots - Nc_y^2}{\Sigma(Y^2) - Nc_y^2} \quad (11)$$

此式應用於一特殊之例時，可將代表原來曲綫方程式中各未知數係數之函數代替式中之 W_1, W_2, W_3, W_4 等。如各係數均為一個單獨自變量之函數，則相關指數之符號可用 r_{yx} 表示之。

特殊之例

前在討論複相關時，吾人嘗用 X_1, X_2, X_3, X_4 等代表各變量，至

於此項變量之爲自變量抑倚變量皆所不問。又嘗用 R 以代表相關量數，並在該符號旁加註亞刺伯小字，此在第十四章內已有敘明。

如兩變量間之關係可用直綫表示時，則用符號 r 以代替 ρ ， r 即爲普通之相關係數。 r 之普遍公式爲

$$r^2 = \frac{a \Sigma(Y) + b \Sigma(XY) - Nc_y^2}{\Sigma(Y^2) - Nc_y^2}$$

用於兩種特例時， r 之公式尚可化簡。如以 X 之算術平均數作爲原點，則

$$a = c_y = \frac{\Sigma Y}{N}$$

$$a^2 = c_y^2 = \frac{a \Sigma Y}{N}$$

$$Nc_y^2 = a \Sigma Y$$

而 r 之公式可化簡爲

$$r^2 = \frac{b \Sigma(xY)}{\Sigma(Y^2) - Nc_y^2}$$

如以 Y 之算術平均數作爲原點（不必同時以 X 之算術平均數作爲原點），則

$$\Sigma(y) = 0, c_y = 0$$

而相關係數之公式又可化爲：

$$r^2 = \frac{b \Sigma(Xy)}{\Sigma(y^2)}$$

在後一種特例中， ρ 之公式亦可化簡，而將 $a \Sigma(y)$ 及 Nc_y^2 兩項消去。

組成常態方程式之校核法

在組成常態方程式及解常態方程式時，每易發生錯誤，故在可能範圍以內，應採用校核法。吾人倘欲校核常態方程式組成之有無錯誤，其法簡易，僅須在每一觀察方程式之後，加列 s 一項，將該式內已知數之和填入，即可作校核之用。下列各觀察方程式係將直綫配合於(1, 3); (2, 4);

(3, 6); (4, 5); (5, 10); (6, 9); (7, 10); (8, 12); (9, 11)各點時所得，並將 s 之數值附列於各式之後，以說明校核法之應用。

	s
$3 = a + 1b$	5
$4 = a + 2b$	7
$6 = a + 3b$	10
$5 = a + 4b$	10
$10 = a + 5b$	16
$9 = a + 6b$	16
$10 = a + 7b$	18
$12 = a + 8b$	21
$11 = a + 9b$	21

(各式內 a 之係數皆為 1，在求 s 時此數與其他已知數之數值一併計入。)

倘所配曲綫之方程式屬於下列形式

$$Y = aW_1 + bW_2 + cW_3 + dW_4$$

則在每個觀察方程式中， s 與式中各數間當有下列之關係：

$$Y + W_1 + W_2 + W_3 + W_4 = s$$

在常態方程式中， s 與其他各數間又有下列之關係：

$$\Sigma(W_1 Y) + \Sigma(W_1^2) + \Sigma(W_1 W_2) + \Sigma(W_1 W_3) + \Sigma(W_1 W_4) = \Sigma(W_1 s)$$

$$\Sigma(W_2 Y) + \Sigma(W_1 W_2) + \Sigma(W_2^2) + \Sigma(W_2 W_3) + \Sigma(W_2 W_4) = \Sigma(W_2 s)$$

$$\Sigma(W_3 Y) + \Sigma(W_1 W_3) + \Sigma(W_2 W_3) + \Sigma(W_3^2) + \Sigma(W_3 W_4) = \Sigma(W_3 s)$$

$$\Sigma(W_4 Y) + \Sigma(W_1 W_4) + \Sigma(W_2 W_4) + \Sigma(W_3 W_4) + \Sigma(W_4^2) = \Sigma(W_4 s)$$

此種形式可適用於任何特殊問題。在每一例中， s 公式之組成與常態方程式相同。

應用 s 校核法時，須在計算表上加添數行，雖似費時 但較通常逐步校核之工作猶為省事。下表所示為配合二次拋物綫時校核法之應用。此綫之方程式為

$$Y = a + bX + cX^2$$

係配合於 (1, 2); (2, 6); (3, 7); (4, 8); (5, 10); (6, 11); (7, 11); (8, 10); (9, 9) 各

點之上者。

表 A

組成常態方程式時校核法之應用

Y	X	X ²	XY	X ² Y	s	Xs	X ² s
2	1	1	2	2	5	5	5
6	2	4	12	24	13	26	52
7	3	9	21	63	20	60	180
8	4	16	32	128	29	116	464
10	5	25	50	250	41	205	1,025
11	6	36	66	396	54	324	1,944
11	7	49	77	539	68	476	3,332
10	8	64	80	640	83	664	5,312
9	9	81	81	729	100	900	8,100
74	45	285	421	2,771	413	2,776	20,414

(X³, X⁴兩行表內未列入,因 $\Sigma(X^3)$ 及 $\Sigma(X^4)$ 之數值均可由製就之表中查得也。)

s行內之各數值均由各個觀察方程式求得。例如由第一觀察方程式

$$2 = 1a + 1b + 1c$$

可求得s之數值為5(即2與三個常數之係數之和)。求s之數值時,可將Y, X, X²各行內之數值與常數a之係數1相加而得。

將上表每行內數值相加,即可校核常態方程式之有無錯誤如下:

$$\Sigma(Y) + N + \Sigma(X) + \Sigma(X^2) = \Sigma(s)$$

$$74 + 9 + 45 + 285 = 413$$

$$\Sigma(XY) + \Sigma(X) + \Sigma(X^2) + \Sigma(X^3) = \Sigma(Xs)$$

$$421 + 45 + 285 + 2,025 = 2,776$$

$$\Sigma(X^2Y) + \Sigma(X^2) + \Sigma(X^3) + \Sigma(X^4) = \Sigma(X^2s)$$

$$2,771 + 285 + 2,025 + 15,333 = 20,414$$

在解常態方程式時,亦可應用s校核法。此點將在另節討論之。

他種校核法

前節討論時,嘗提及他種校核法。吾人已知所配曲綫中各常數之係

數倘用 W_1, W_2, W_3, W_4 代表, 則

$$\Sigma(vW_1) = 0$$

$$\Sigma(vW_2) = 0$$

$$\Sigma(vW_3) = 0$$

$$\Sigma(vW_4) = 0$$

如所配曲綫之形式爲

$$Y = a + bX + cX^2 + dX^3$$

則

$$\Sigma(v) = 0$$

$$\Sigma(vX) = 0$$

$$\Sigma(vX^2) = 0$$

$$\Sigma(vX^3) = 0$$

故計算上有無錯誤, 可用上列各式校核之。

吾人又可用兩法計算估量標準誤之數值, 以試驗所得兩個數值之是否相同, 俾作計算上校核之用。第一法係計算各個剩餘數之數值, 即實際數值與計算數值之差, 以求標準誤 S 之數值。第二法係直接應用標準誤之普遍公式, 求 S 之數值。上節用以說明校核常態方程式之方法時, 根據實例配合二次拋物綫所得之方程式爲:

$$Y = -.92860 + 3.52316X - .267316X^2$$

由實際數值與用上式求得之計算數值之差, 可得

$$S_y = .4941$$

由下列標準誤之公式

$$S_y^2 = \frac{\Sigma(Y^2) - a\Sigma(Y) - b\Sigma(XY) - c\Sigma(X^2Y)}{N}$$

求得

$$S_y = .4947$$

此亦爲校核計算上準確程度之一法也。

常態方程式之解法

通常在經濟統計方面，解常態方程式並無若何困難。如式中僅含二個未知數，則常態方程式亦祇有二個，可用簡單之代數方法解之。惟若未知數共有三個，欲解三個常態方程式，雖亦可應用代數方法，但仍以採取較有條理之步驟為是；若方程式共有三個以上，則此項步驟更有採用之必要。高斯氏(Gauss)暨杜立特氏(Doolittle)對於用最小平方法所得聯立方程式之解法，均曾規劃有條理之步驟。杜氏之法尤合於一般之應用，茲說明於下：

常態方程式中各未知數之係數，在其主要對角綫兩邊成對稱形式。求方程式

$$Y = aW_1 + bW_2 + cW_3 + dW_4$$

中常數之最可能數值時，其四個常態方程式為：

$$a\Sigma(W_1^2) + b\Sigma(W_1W_2) + c\Sigma(W_1W_3) + d\Sigma(W_1W_4) - \Sigma(W_1Y) = 0$$

$$a\Sigma(W_1W_2) + b\Sigma(W_2^2) + c\Sigma(W_2W_3) + d\Sigma(W_2W_4) - \Sigma(W_2Y) = 0$$

$$a\Sigma(W_1W_3) + b\Sigma(W_2W_3) + c\Sigma(W_3^2) + d\Sigma(W_3W_4) - \Sigma(W_3Y) = 0$$

$$a\Sigma(W_1W_4) + b\Sigma(W_2W_4) + c\Sigma(W_3W_4) + d\Sigma(W_4^2) - \Sigma(W_4Y) = 0$$

該四式中，如將Y各項除外，即可顯示對角綫兩旁之排列成對稱形式。如在主要對角綫上任取一項，則在其上方之係數與在其左方之係數完全相同。例如取 $c\Sigma(W_3^2)$ 一項，在其上方之係數為 $\Sigma(W_2W_3)$ 及 $\Sigma(W_1W_3)$ ，而在其左方之係數亦為 $\Sigma(W_2W_3)$ 及 $\Sigma(W_1W_3)$ 。故在對角綫左方之各項可一併略去，而將常態方程式寫為

$$a\Sigma(W_1^2) + b\Sigma(W_1W_2) + c\Sigma(W_1W_3) + d\Sigma(W_1W_4) - \Sigma(W_1Y) = 0$$

$$+ b\Sigma(W_2^2) + c\Sigma(W_2W_3) + d\Sigma(W_2W_4) - \Sigma(W_2Y) = 0$$

$$+ c\Sigma(W_3^2) + d\Sigma(W_3W_4) - \Sigma(W_3Y) = 0$$

$$+ d\Sigma(W_4^2) - \Sigma(W_4Y) = 0$$

杜立特氏之解法

吾人可用杜立特氏之方法以解第418頁上之常態方程式。此三個常態方程式可簡書如下：

$$\begin{aligned}
 8.3564b_{12.34} + 2.790b_{13.24} + 2.932b_{14.23} + 47.967 &= 0 \\
 + 6.6645b_{13.24} + 2.063b_{14.23} + 62.039 &= 0 \\
 + 7.7893b_{14.23} + 47.519 &= 0
 \end{aligned}$$

解該三式中之常數時，為便於指引起見，令代A表 $b_{12.34}$ ，B代表 $b_{13.24}$ ，C代表 $b_{14.23}$ 。茲將計算步驟及其校核方法列表舉示於下：

表 B

用杜立特氏法解常態方程式

綫	(1) 倒數	(2) A	(3) B	(4) C	(5)	(6) s
I		8.3564	2.790	2.932	47.967	62.0454
II			6.6645	2.063	62.039	73.5565
III				7.7893	47.519	60.3033
1		8.3564	2.790	2.932	47.967	62.0454
2	-.11966876	-1.000000	-.333876	-.350869	-5.740151	-7.424896 check
3			6.6645	2.063	62.039	73.5565
4			-.931514	-.978924	-16.015030	-20.715470
5			5.732986	1.084076	46.023970	52.841030 check
6	-.17442917		-1.000000	-.189094	-8.027923	-9.217017 check
7				7.7893	47.519	60.3033
8				-1.028748	-16.830133	-21.769807
9				-.204992	-8.702857	-9.991922
10				6.555560	21.986010	28.541571 check
11	-.15254227			-1.000000	-3.353796	-4.353796 check

回 解 法

C	B	A
-3.353796	-8.027923	-5.740151
<u>-3.353796</u>	<u>+ .634183</u>	<u>+ 2.468592</u>
	-7.393740	+ 1.176743
		<u>- 2.094816</u>

$A = b_{12.34} = -2.094816$

$B = b_{13.24} = -7.393740$

$C = b_{14.23} = -3.353796$

校核：

方程式 I：

$$8.3564b_{1,2,3,4} + 2.790b_{1,3,2,4} + 2.932b_{1,4,2,3} = -47.967$$

將以上所求得之各數值代入式中，得

$$\begin{aligned} 8.3564(-2.094816) + 2.790(-7.393740) + 2.932(-3.353796) \\ = -47.966985 \end{aligned}$$

說明——將 A, B, C 各未知數之係數列於 A, B, C 之各直行內，每個常態方程式中之末一項已知數列入第(5)直行內(此項已知數之正負號應與常態方程式全式等於零時該已知數應得之符號相同)。第(6)直行 s 係作校核之用。該 s 行內與 I, II, III 各橫行相對之三個數值為各常態方程式中已知數之代數之和。在求此和數時，所有前經略去之對角綫左方之各係數仍應併入計算。

茲將採用杜立特氏法解常態方程式之步驟簡述於下：

1. 將常態方程式 I 書於橫行(1)內。
2. 將橫行(1)直行(2)內數值之倒數書於橫行(2)直行(1)內，惟須用其相反之正負號(此即 A 之係數之倒數)。以此倒數乘橫行(1)內各項，並將此各乘積分別列入橫行(2)之各該直行內(橫行(2)直行(2)、(3)、(4)、(5)內各數值代數之和，應與直行(6)內之數值相等)。經此手續後則未知數 A 之數值即可用 B 及 C 表示矣(直行(2)橫行(2)內 -1 之數值係為便於校核之用，故列入表內。橫行(6)及(11)內 -1 之數值用途亦同)。茲可劃一粗綫於橫行(2)之下，橫貫全表。
3. 將常態方程式 II 書於橫行(3)內。
4. 以橫行(2)內 B 之係數(即 -0.333876)乘橫行(1)直行(3)、(4)、(5)、(6)內之數值，並將此各乘積分別列入橫行(4)之各該直行內。
5. 將橫行(3)及(4)內之數值相加，並將其和數分別列入橫行(5)之各該直行內(橫行(5)直行(3)、(4)、(5)內各數值代數之和應與直行(6)

內之數值相等。)

6. 將直行(3)橫行(5)內數值之倒數書於直行(1)橫行(6)內，惟須用其相反之符號。以此倒數乘橫行(5)內之各數值，並將此各乘積分別列入橫行(6)之各該直行內(橫行(6)直行(3)、(4)、(5)內各數值之和應與直行(6)內之數值相等)。經此手續後則未知數 B 之數值即可用 C 表示矣。茲又可劃一粗綫於橫行(6)之下，橫貫全表。
7. 將常態方程式 III 書於橫行(7)內。
8. 以橫行(2)內 C 之係數(即 -0.350869)乘橫行(1)直行(4)、(5)、(6)內之數值，並將此各乘積分別列入橫行(8)之各該直行內。
9. 以橫行(6)內 C 之係數(即 -0.189094)乘橫行(5)直行(4)、(5)、(6)內之數值，並將此各乘積分別列入橫行(9)之各該直行內。
10. 將橫行(7)、(8)、(9)內之數值相加，並將其和數分別列入橫行(10)之各該直行內(橫行(10)直行(4)、(5)內各數值代數之和應與直行(6)內之數值相等)。
11. 將直行(4)橫行(10)內數值之倒數書於直行(1)橫行(11)內，惟須用其相反之符號。以此倒數乘橫行(10)內之各數值，並將此各乘積分別列入橫行(11)之各該直行內(橫行(11)直行(4)、(5)內各數值代數之和應與直行(6)內之數值相等)。經此手續後即可在橫行(11)直行(5)內求得 C 之數值。茲又可劃一粗綫於橫行(11)之下，橫貫全表。

如未知數尚有 D, E 等，則上述之最後一項手續可使 C 之數值用 D 及 E 之函數表示之，而須繼續求 D, E 之數值。其手續係將第四個常態方程式書於橫行(12)內，以橫行(2)、(6)、(11)內 D 之係數依次乘橫行(1)、(5)、(10)內之數值，而將此各乘積分別列入橫行(13)、(14)、(15)內。橫行(12)、(13)、(14)、(15)內各數值之和應書於橫行(16)之各該直行內，並用 s 行以校核各數值之正誤。以橫行(16)內 D 之係數之倒數(須用其相反符號)乘之，即

可求得用 E 表示之 D 之數值。 E 之數值可用同法求之。

各項計算之校核可由表得之。逐步計算之結果經校核後，即可減少錯誤可能性至最低之限度。

A, B, C 之數值可用回解法，由橫行(11)得

$$C = -3.353796$$

由橫行(6)得

$$B = -.189094C - 8.027923$$

由橫行(2)得

$$A = -.333876B -.350869C - 5.740151$$

(直行(6)內各數值列入表內，係作校核之用。橫行(2)、(6)、(11)內 -1.000000 之數值亦併列表內，所以便於校核也。)

回解法之各項計算，俱見上表內。

將各該數值代入一常態方程式中，以作最後之校核。校核方程式I之手續，詳見表B之下。

參 考 書

- ADAMS, OSCAR S. *Application of the Theory of Least Squares to the Adjustment of Triangulation*. Special Publication No. 28, U. S. Coast and Geodetic Survey, 1916.
- BRUNT, DAVID. *The Combination of Observations*.
- HUNTINGTON E. V. *Curve Fitting by the Method of Least Squares and the Method of Moments* (in Rietz, H. L. ed. *Handbook of Mathematical Statistics*, 62—70).
- MERRIMAN, MANSFIELD. *The Method of Least Squares*.
- SMITH, BRADFORD B. *The Use of Punched Card Tabulating Equipment in Multiple Correlation Problems*. Washington, Bureau of Agricultural Economics, 1923.
- WELD, L. D. *Theory of Errors and Least Squares*.
- WHITAKER, E. T. and ROBINSON, G. *The Calculus of Observations* (209—259).
- WRIGHT and HAYFORD. *Adjustment of Observations*.

附 錄 B

統計符號彙解

以下所列為本書所用各種符號中之最重要者。一種符號偶或表示不同之事物，但其準確意義，可就其所用處所確定之。

1. 變量與常數之普通符號：

x ：一變量。

y ：一變量。

凡在英文字母尾端之各字母常可用以代表一變量。各個變量亦可用同一字母加註不同之小字，如 X_1, X_2, X_3 ，或 W_1, W_2, W_3 等表示之〔代表變量所用之大寫字母與小寫字母（如測量相關時所用者）常含有不同之意義〕。

a ：一常數（即在討論方程式時，其值固定不變之一數量）。英文字母之前數個，通常均可用以代表常數。

2. 分析及表述次數分配所用之符號：

m ：一個單獨觀察所得之數值；亦可用以表示一組中點之數值（ a_1, a_2, a_3 等符號亦有用以代表一數列中之各個觀察數值者）。

f ：一組內觀察數值之項數，亦即一組之次數。

i ：組距。

l ：一組之下限。

N ：一數列或一個次數分配中項目之總數。

d ：一觀察數值對平均數（通常為算術平均數）之離中差。倘 d 旁附有小字，如 d_x 或 d_y ，則此符號表示該小字所代表之變量對其算術平均數之離中差。符號 d 亦有用以代表算術平均數與衆數兩數值之差者。

d' ：一觀察數值對假定原點或假定平均數之離中差。

c : 假定原點或假定平均數對真實平均數之差(如用後文所述之符號表示,則 $c = M - M'$)。

Σ (Sigma): 各項相加之符號。例如 Σd 係表示各項離中差之總和。

w_1, w_2, w_3 : 一數列平均時,各項所用之權數(此項符號不可與代表各變量之 W_1, W_2, W_3 相混)。

y_0 : 次數曲綫之最高縱坐標。

平均數四分位數等之符號:

M : 算術平均數。

Md : 中位數。

Mo : 衆數。

M_g : 幾何平均數。

H : 倒數平均數。

M' : 假定算術平均數之數值。

Q_1 : 第一四分位數,即下四分位數。

Q_2 : 第二四分位數,即中位數。

Q_3 : 第三四分位數,即上四分位數。

K : 在第一與第三四分位數間中點之數值。

D_3 : 第三十分位數。

離散度及偏斜度量數之符號:

$M.D.$: 平均差。

σ : 標準差;亦即各項對其算術平均數之均方根差。

s : 各項對其算術平均數以外任何原點之均方根差。

$P.E.$: 機誤。

$Q.D.$: 四分位差。

q_1 : 中位數與第一四分位數之差($Md - Q_1$)。

q_2 : 第三四分位數與中位數之差($Q_3 - Md$)。

V : 離散度係數。

sk : 偏斜度之量數。

$\chi(Chi)$: 根據 β_1 及 β_2 之數值求得之偏斜度量數。

3. 關於指數之符號:

p_0' : 第一種物品之基期價格。

q_0' : 第一種物品之基期數量。

p_1' : 第一種物品之計算時期價格。

q_1' : 第一種物品之計算時期數量。

p_0'' : 第二種物品之基期價格。

q_0'' : 第二種物品之基期數量。

p_1'' : 第二種物品之計算時期價格。

q_1'' : 第二種物品之計算時期數量。

$\frac{p_1'}{p_0'}$: 一價比(即第一種物品計算時期之價格與其基期價格之比)。

$\frac{q_1'}{q_0'}$: 一量比

P_0 : 基期之物價水準。

P_1 : 計算時期之物價水準。

4. 用於相關量數之符號:

X : 一變量之一個觀察數值。

Y : 一變量之一個觀察數值(各變量之觀察數值亦可用 X_1, X_2, X_3 , 或 W_1, W_2, W_3 等符號代表之。)

\bar{X} : X 變量各個觀察數值之算術平均數。其他變量之算術平均數可用類似之符號表示之(在討論複相關之一例中曾用 A_1, A_2, A_3, \dots 等符號以代表 X_1, X_2, X_3, \dots 各變量之算術平均數。符號 M_x, M_y 亦常用以代表各變量之算術平均數。)

x : 以觀察數值對其算術平均數之離中差所表示一變量之數值。 y 及

- x_1, x_2, x_3, \dots 等符號亦可同樣用以代表各變量之數值。至其原來數值則用 Y, X_1, X_2, X_3, \dots 等符號代之。
- x' : 以觀察數值對其假定原點之離中差所表示一變量之數值。符號 y' 意義相似。
- Y_c : 根據平均關係方程式所求得一變量之計算或估量數值; y_c 之符號可用以代表根據算術平均數計算之離中差所表示之計算或估量數值。
- p : 以根據算術平均數計算之離中差所表示兩變量之均積數, 即 $p = \frac{\Sigma(xy)}{N}$ 。如 p 旁加註小字, 例如 p_{12} , 則此兩小字表示其所指之 x_1 及 x_2 兩變量。
- p' : 以根據假定算術平均數計算之離中差所表示兩變量之均積數, 即 $p' = \frac{\Sigma(x'y')}{N}$ 。
- r : 披爾遜氏之相關係數。 r 旁加註小字時, 則此項小字係表示相關係數所指之各變量。例如 r_{yx} 表示其變量為 y 及 x , r_{12} 表示其變量為 x_1 及 x_2 。
- ρ (Rho): 相關之普通指數(此符號不可與斯比亞門氏 (Spearman) 之由級差平方而定之相關係數相混)。 ρ 旁應加註小字以表示該量數所根據之各變量, 如 $\rho_{yx}, \rho_{xy}, \rho_{\log yx}, \rho_{\log y \log x}, \rho_{\frac{1}{y}x}$ 等是。每一符號中, 其第一小字係指倚變量。
- d : 一個觀察數值與所配曲綫上數值之差, 亦即一變量之觀察數值與其相對之計算數值之差。
- v : 一剩餘數; 與上述 d 之意義完全相同。
- S : 各項對所配曲綫之均方根差, 亦即估量之標準誤。此量數旁應加註小字以表示其所指之變量, 如 $S_y, S_x, S_{\log y}$ (用對數表示之估量標準誤), S_r (用比率表示之估量標準誤), $S_{\frac{1}{y}}$ (用倒數表示之

估量標準誤)。

η (Eta): 相關率。 η 旁應加註小字如 η_{yz}, η_{xy} 以表示其所指之變量。小字中之第一字母係指倚變量。

σ_{ay} : 各項對於通過相關表中各直行中點一綫之均方根差;即 y 之各行列與其各中點之標準差。符號 σ_{ax} 之意義亦同,惟其所指者為相關表中之各橫行或 x 之行列。

σ_{my} : 相關表中各直行之中點對於 y 之平均數之標準差,各直行之中點應以各該行所含次數分別加權。符號 σ_{mx} 之意義亦同,惟指相關表中之各橫行而言。

ζ (Zeta): 試驗迴歸直綫性之量數($\zeta = \eta^2 - r^2$)。

κ : 計算相關率時所用行列之數。

b : 迴歸係數;亦即迴歸綫之斜度。 b 旁如註有小字,則此項小字表示其所指之變量,如 b_{yz}, b_{12} (即指 x_1, x_2 各變量而言)。小字之第一字係指倚變量, b_{yz} 為 y 倚 x 之迴歸係數, b_{xy} 為 x 倚 y 之迴歸係數。

$R_{1.234}$: 複相關中一倚變量 x_1 與各自變量 x_2, x_3 及 x_4 之複相關係數。小字之次序可變更,但主要註字常指倚變量。

$r_{1.234}$: 自變量 x_3, x_4 保持常態時,變量 x_1 與變量 x_2 之淨相關係數。各變量之組合不同時, r 旁所註小字之排列次序亦隨之而異,惟兩個主要註字常指此淨相關係數所測度之兩變量。

$b_{1.234}$: 估量方程式中之自變量 x_3 及 x_4 保持常態時,倚變量 x_1 與自變量 x_2 間之淨迴歸係數;亦即彙合 x_2, x_3, x_4 各自變量以估量 x_1 時 x_2 所應得之權數。各變量之組合不同時, b 旁所註小字之排列次序亦隨之而異。

$S_{1.234}$: 由表示一倚變量 x_1 與自變量 x_2, x_3, x_4 間關係之綫所求得之均方根差;亦即在此情形下 x_1 之估量標準誤。

$\sigma_{1.234}$: 四次標準差;與 $S_{1.234}$ 完全相同。

(上節末所列各量數,其所註小字之字數應與研究時所包含之變量數相同。爲求說明時簡便起見,每一量數均係假定僅包括四個變量)。

5. 用於測度差誤之符號:

σ_M : 算術平均數之標準誤。

$\sigma\sigma$: 標準差之標準誤。同樣,符號 σ 旁如註有一小字,則此小字係代表該標準誤所指之量數。

$P.E._M$: 算術平均數之機誤 ($P.E._M = .67449\sigma_M$)。同樣,符號 $P.E.$ 旁如註有一小字,則此小字係代表該機誤所指之量數。

6. 其他符號:

p : 某事件成就之機率。

q : 某事件失敗之機率。

s : 一觀察方程式中已知數量之總和; s 可作組成常態方程式及解常態方程式時校核之用。

Δ (Delta): 代表兩數量之差或一變量之兩連續數值之差之普通符號。符號旁可加註小字如 Δ_v ,以表示其所指之變量。第一階差用 Δ^1 代之,第二階差數用 Δ^2 代之,餘類推。

v_1, v_2, v_3 ,等: 在假定原點兩旁次數分配之動差。

π_1, π_2, π_3 ,等: 在算術平均數兩旁次數分配之未校正動差。

μ_1, μ_2, μ_3 ,等: 用舍巴德氏校正法所得算術平均數兩旁次數分配之動差(如無採用舍巴德氏校正法之必要,則算術平均數兩旁未校正之動差亦可用此種符號表之)。

$$\beta_1: \frac{\mu_3^2}{\mu_2^3}$$

$$\beta_2: \frac{\mu_4}{\mu_2^2}$$

κ_2 : 根據 β_1 及 β_2 之數值以確定曲綫峯度之標準。

χ_2 : 試驗曲綫配合適度之量數。

參 考 書 目 錄

- ADAMS, OSCAR S. *Application of the Theory of Least Squares to the Adjustment of Triangulation.* Special Publication No. 28. U. S. Coast and Geodetic Survey. 1915.
- BARLOW. *Tables of Squares, Cubes, Square Roots, Cube Roots, Reciprocals.* Spon and Chamberlain. New York. 1919.
- BECKETT, S. H., and ROBERTSON, R. D. *The Economical Irrigation of Alfalfa in the Sacramento Valley.* Bulletin 280, Agricultural Experiment Station. University of California. May, 1917.
- BOWLEY, A. L. *Elements of Statistics.* P. S. King and Son, London. 1920.
- BRINTON, W. C. *Graphic Methods for Presenting Facts.* The Engineering Magazine Co., New York. 1914.
- BROAD, C. D. *On the Relation between Induction and Probability.* Mind N. S. Vol. 27, 1918, Vol. 29, 1920.
- BRUNT, DAVID. *The Combination of Observations.* Cambridge University Press. 1917.
- CHADDOCK, R. E. *Principles and Methods of Statistics.* Houghton Mifflin, Boston. (In preparation.)
- CLARK, WALLACE. *The Gantt Chart.* Ronald Press, New York. 1922.
- CRUM, W. S. *The Use of the Median in Determining Seasonal Variation.* Journal of the American Statistical Association. March, 1923.
- DAVENPORT, E. *Comparative Agriculture.* In Bailey's Cyclopedia of American Agriculture.
- DAVIES, G. R. *Introduction to Economic Statistics.* Century, New York. 1922.
- DAY, EDMUND E. *An Index of the Physical Volume of Production.* Review of Economic Statistics. Sept., 1920, Jan., 1921. *Standardization of the Construction of Statistical Tables.* Quarterly Publications of the American Statistical Association. March, 1920. *The Volume of Production of Basic Raw Materials in the United States.* Review of Economic Statistics. July, 1922.
- EDGEWORTH, F. Y. *On Correlated Averages.* Phil. Mag., 5th series. Vol. 34. 1892
- EDITORIAL. *On the Probable Errors of Frequency Constants.* Biometrika, Vol. 2 (273-281).
- ELDERTON, W. PALIN. *Frequency Curves and Correlation.* Layton, London. 1908.
- EZEKIEL, M. J. B. *A Method of Handling Curvilinear Correlation for Any Number of Variables,* Journal of the American Statistical Association, Dec., 1924.
- FAUTKNER, HELEN D. *The Measurement of Seasonal Variation.* Journal of the

- American Statistical Association, June, 1924.
- FIELD, J. H. *Some Advantages of the Logarithmic Scale in Statistical Diagrams.* Journal of Political Economy. Oct., 1917.
- FISHER ARNE. *An Elementary Treatise on Frequency Curves.* Macmillan, New York. 1922. *The Mathematical Theory of Probabilities.* Macmillan, New York. 1922.
- FISHER, IRVING. *A Weekly Index Number of Wholesale Prices.* Journal of the American Statistical Association. Sept., 1923. *The Making of Index Numbers.* Houghton Mifflin, Boston. 1922. *The "Ratio" Chart.* Quarterly Publications of the American Statistical Association. June, 1917. *Revision of the Weekly Index Number.* Journal of the American Statistical Association, Sept., 1924.
- FLUX, A. W. *The Measurement of Price Changes.* Journal of the Royal Statistical Society. March, 1921.
- GALTON, FRANCIS. *Correlations and Their Measurement.* Proceedings of the Royal Society. Vol. 45, 1888.
- GRIFFIN, F. L. *Introduction to Mathematical Analysis.* Houghton Mifflin, Boston. 1922.
- HAAS, G. C. *Sale Prices as a Basis for Farm Land Appraisal.* Technical Bulletin, No. 9. University of Minnesota Agricultural Experiment Station. Nov., 1922.
- HALL, LINCOLN W. *Seasonal Variation as a Relative of Secular Trend.* Journal of the American Statistical Association, June, 1924.
- HART, W. L. *The Method of Monthly Means for Determination of a Seasonal Variation.* Journal of the American Statistical Association. Sept., 1922.
- HASKELL, A. C. *Graphic Charts in Business.* Codex Book Co., New York. 1922.
How to Make and Use Graphic Charts. Codex Book Co., New York. 1919.
- JONES, D. C. *A First Course in Statistics.* Bell, London. 1921
- KARSTEN, KARL. *Charts and Graphs.* Prentice Hall, New York. 1923.
- KELLEY, TRUMAN L. *Statistical Method.* Macmillan, New York. 1923.
- KEYNES, J. M. *A Treatise on Probability.* Macmillan, New York. 1921.
- KILLOUGH, H. B. *A Statistical Analysis of Oat Prices.* Bureau of Agricultural Economics.
- KING, W. I. *Elements of Statistical Method.* Macmillan, New York. 1912. *An Improved Method for Measuring the Seasonal Factor.* Journal of the American Statistical Association, Sept., 1924.
- KNIBBS, G. H. *The Theory and Justification of Curve Smoothing.* In H. Secrist, *Readings and Problems in Statistical Methods.* Macmillan, New York. 1920.
- KURTZ, EDWIN. *Replacement Insurance.* Administration, July, 1921.

- LIPKA JOSEPH. *Graphical and Mechanical Computation*. Wiley, New York. 1918.
- MELLOR, J. W. *Higher Mathematics for Students of Chemistry and Physics*. Longmans, London. 1922.
- MERRIMAN, MANSFIELD. *The Method of Least Squares*. Wiley, New York. 1897.
- MILLS, FREDERICK C. *On Measurement in Economics* (in Tugwell, R. G. ed., *The Trend of Economics*. Knopf, New York. 1924).
- MINER, J. R. *Tables of $\sqrt{1-r^2}$ and $1-r^2$ for use in Partial Correlation and Trigonometry*. Johns Hopkins Press, Baltimore. 1922.
- MITCHELL, W. C. *History of Prices During the War*. Price Bulletin, No. 1, War Industries Board. 1919. *The Making and Using of Index Numbers*. Part I. Bulletin No. 284. U. S. Bureau of Labor Statistics. Oct., 1921.
- MOORE, H. L. *Economic Cycles: Their Law and Cause*. Macmillan, New York. 1914. *Elasticity of Demand and Flexibility of Prices*. Journal of the American Statistical Association. March, 1922. *Empirical Laws of Demand and Supply and the Flexibility of Prices*. Political Science Quarterly. Dec., 1919. *Forecasting the Yield and the Price of Cotton*. Macmillan, New York. 1917. *Generating Economic Cycles*. Macmillan, New York. 1923.
- NATIONAL BUREAU OF ECONOMIC RESEARCH. *Income in the United States*. (Edited by W. C. Mitchell.) Harcourt Brace and Co., New York. 1921.
- NORTH DAKOTA AGRICULTURAL COLLEGE. *Cost of Production and Farm Organization on 126 Farms in North Dakota*. Bulletin No. 165, Agricultural Experiment Station. 1922.
- OSBURN, W. F., and THOMAS, DOROTHY. *Influence of the Business Cycle on Certain Social Conditions*. Quarterly Publications of the American Statistical Association. Sept., 1922.
- PEAKE, E. G. *An Academic Study of Some Money Market and Other Statistics*. P. S. King, London. 1923.
- PRAEL, R., and REED, L. J. *Predicted Growth of Population of New York and its Environs*. Committee of Plan of New York. 1923.
- PEARL, RAYMOND. *Medical Biometry and Statistics*. Saunders, Philadelphia. 1923.
- PEARSON, KARL. *Mathematical Contributions to the Theory of Evolution*. *On the General Theory of Skew Correlation and Non-Linear Regression*. Draper's Company Research Memoirs. Cambridge University Press. 1905. *Notes on the History of Correlation*. Biometrika, Vol. 13. *On a Correction Needful in the Case of the Correlation Ratio*. Biometrika, Vol. 8. *On the Correction Necessary for the Correlation Ratio*. Biometrika, Vol. 14. *Regression Heredity and Panmixia*. Phil. transactions, Royal Society, Series A, Vol. 187. 1896. *Tables*

- for Statisticians and Biometricians.* Cambridge University Press. 1914. *The Fundamental Problem of Practical Statistics.* Biometrika, Vol. 13.
- PERSONS, WARREN M., and COYLE, EUNICE. *A Commodity Price Index of Business Cycles.* Review of Economic Statistics. Prel. Vol. 3.
- PERSONS, WARREN M. *An Index of Trade for the United States.* Review of Economic Statistics. April, 1923. *Correlation of Time Series.* Journal of the American Statistical Association. June, 1923. *Fisher's Formula for Index Numbers.* Review of Economic Statistics. Prel. Vol. 3. *Indices of Business Conditions.* Review of Economic Statistics. Prel. Vol. 1. 1919. *The Variate Difference Correlation Method and Curve Fitting.* Quarterly Publications of the American Statistical Association. June, 1917.
- PRESCOTT, RAYMOND. *Law of Growth in Forecasting Demand.* Journal of the American Statistical Association. Dec., 1922.
- RIETZ, H. L. ed. *Handbook of Mathematical Statistics.* Houghton Mifflin, Boston. 1924.
- RIETZ, H. L., and CRATHORNE, A. R. *College Algebra.* Holt, New York. 1917.
- RUGG, H. O. *Statistical Methods Applied to Education.* Houghton Mifflin, Boston. 1917.
- RUNNING, T. R. *Empirical Formulas.* Wiley, New York. 1917.
- SCHULTZE, ARTHUR. *Graphic Algebra.* Macmillan, New York. 1918.
- SCHULTZ, HENRY. *The Statistical Measurement of the Elasticity of Demand for Beef.* Journal of Farm Economics. July, 1924.
- SECRET, HORACE. *Introduction to Statistical Methods.* Macmillan, New York. 1917. *Readings and Problems in Statistical Methods.* Macmillan, New York. 1920.
- SHEPPARD, W. F. *On the Calculation of the Most Probable Values of Frequency Constants for Data Arranged According to Equi-distant Divisions of a Scale.* Proceedings of the London Mathematical Society, Vol. 29. 1898. *The Calculation of the Moments of a Frequency Distribution.* Biometrika, Vol. 5.
- SMITH, BRADFORD B. *The Use of Punched Card Tabulating Equipment in Multiple Correlation Problems.* (Prepared for the use of statisticians of the Bureau of Agricultural Economics., U. S. Dept. of Agriculture.) 1923.
- SNOW, E. C. *Trade Forecasting and Prices.* Journal of the Royal Statistical Society. May, 1923.
- SNYDER, CARL. *A New Index of the Volume of Trade.* Journal of the American Statistical Association. Dec., 1923.
- STAMP, J. C. *The Effect of Trade Fluctuations upon Profits.* Journal of the Royal

- Statistical Society. July, 1918.
- STEIMMETZ, C. P. *Engineering Mathematics*. McGraw Hill, New York. 1917.
- STEWART, ETHELBERT. *Labor Efficiency and Productiveness in Sawmills*. Monthly Labor Review. Jan., 1923.
- TOLLEY, H. R., and EZEKIEL, M. J. B. *A Method of Handling Multiple Correlation Problems*. Journal of the American Statistical Association. Dec., 1923.
- WALSH, C. M. *The Measurement of General Exchange Value*. Macmillan, New York. 1901. *The Problem of Estimation*. P. S. King and Son, London. 1921.
- WELD, L. D. *Theory of Errors and Least Squares*. Macmillan, New York. 1916.
- WEST, CARL J. *Introduction to Mathematical Statistics*. Adams, Columbus. 1918.
- WHIPPLE, G. C. *Vital Statistics*. Wiley, New York. 1919.
- WHITAKER, E. T., and ROBINSON, G. *The Calculus of Observations*. Blackie and Son, London. 1924.
- WHITEHEAD, A. N. *An Introduction to Mathematics*. Holt, New York. 1911.
- WORKING, HOLBROOK. *Factors Determining the Price of Potatoes in St. Paul and Minneapolis*. Technical Bulletin, No. 10. University of Minnesota Experiment Station. Oct., 1922.
- WRIGHT, T. W., and HAYFORD, J. F. *Adjustment of Observations*. Van Nostrand, New York. 1906.
- YOUNG, ALLYN A. *The Measurement of Changes in the General Price Level*. Quarterly Journal of Economics. Aug., 1921.
- YULE, G. UDNEY. *An Introduction to the Theory of Statistics*. Griffin, London. 1919. *On the Time Correlation Problem, with Especial Reference to the Variate Difference Correlation Method*. Journal of the Royal Statistical Society. July, 1921.
- ZIZEK, FRANZ. *Statistical Averages*. Holt, New York. 1913.

統計譯名

(中國統計學社審定)

A

Abscissa: 橫坐標
Absence: 缺
Absolute dispersion: 絕對離中差
Absolute error: 絕對差誤
Absolute number: 絕對數
Absolute value: 絕對值
Accuracy: 確度
Actual frequency: 實際次數; 實際頻數
Adding machine: 加數機
Additive constant: 常加數
Adjustment: 修整
Adjustment of chain relatives: 鎖比之修整
Aggregate: 綜合
Aggregate-value method of weighting: 總值加權法
Aggregative index number: 綜合指數
Amplitude: 廣度
Anti-logarithm: 逆對數
Approximation: 近似數
Arbitrary or Assumed average: 假定平均數
Arbitrary origin: 假定原點
Area bar-chart: 面積條圖
Area graph: 面積圖
Arithmetic mean: 算術平均數
Arrangement: 排列
Array: 整列
Ascending order: 遞升次序
Association: 相聯
Assumed median: 假定中位數
Asymmetrical distribution: 不對稱分配
Asymmetry: 不對稱
Asymptote: 漸近線
Attenuation: 減微

Attribute: 品質
Attributive classification: 品質分類
Average deviation or Mean deviation: 平均差
Average error: 平均差誤
Average index number: 平均指數
Average or Mean: 平均數
Axis: 軸
Axis of abscissa: 橫軸
Axis of ordinate: 縱軸

B

Balance chart: 平衡圖
Band curve chart: 帶形曲線圖
Bar chart: 條圖
Base: 基
Base line: 基線
Base number: 基數
Base period: 基期
Base year or month: 基年或基月
Bead map: 標珠地圖
Bell-shaped curve: 鐘形曲線
Bell-shaped symmetrical curve: 鐘形對稱曲線
Bias: 偏性
Biased error: 偏誤
Bimodal: 雙峯的
Bimodality: 雙峯
Binomial distribution: 二項分配
Binomial expansion: 二項展開式
Binomial theorem: 二項式定理
Biometry: 生物測量學
Birth rate: 出生率
Bi-serial coefficient of correlation: 二數列相關係數
Bi-serial ratio of correlation: 二數列相關比率
Blank form: 空白表式

Block diagram: 方柱圖
 Box: 格
 Broadened base system: 擴張基期法
 Business barometer: 商情測變表
 Business cycle: 商情循環
 Business forecasting: 商情預測

C

Calculating chart: 計算圖
 Calculating machine: 計算機
 Caption: 縱標目
 Cartogram: 統計地圖
 Categorical series: 類別數列
 Census: 普查
 Central ordinate: 中點縱坐標
 Central tendency: 集中趨勢
 Centroid: 中心
 Chain relative: 鎖比
 Chance: 機遇
 Chance error: (見 Probable error)
 Characteristics: 表徵
 Chart: 圖
 Chart field: 圖域
 Charting: 製圖
 Chronological order: 時間次序
 Circle graph: 圓形圖
 Circular number: 環數
 Circular test: 循環試驗
 Class: 組; 類
 Class frequency: 組次數; 組頻數
 Classification: 分類; 分組
 Two-fold-classification or dichotomy: 兩組分類
 Classification chart: 分類圖
 Class interval: 組距
 Class limit: 組限
 Class or group index number: 分類指數
 Coefficient: 係數
 Coefficient of alienation: 離異係數
 Coefficient of association: 相聯係數
 Coefficient of contingency: 相依係數

Coefficient of correlation: 相關係數
 Coefficient of first order: 一次係數
 Coefficient of multiple correlation: 複相關係數
 Coefficient of partial correlation: 淨相關係數
 Coefficient of regression: 迴歸係數
 Coefficient of skewness: 偏態係數
 Collection of data: 資料之搜集
 Colored map: 彩色地圖
 Column: 縱行
 Column diagram: 直方圖
 Combination: 組合
 Comparability: 公比性
 Compensating error: 償補差誤
 Compensation curve: 償補曲線
 Compensatory movement: 補償變動
 Component bar-chart: 區分條圖
 Composite bar-chart: 組合條圖
 Composite curve chart: 組合曲線圖
 Compound event: 複合事件
 Concentric circle diagram: 同心多圓圖
 Consensus method: 同意法
 Consistency: 相容
 Constancy of great numbers: 大數恆性
 Constant: 常數
 Constant error: 常誤
 Constriction: 相關減弱
 Contingency: 相依
 Contingency table: 相依表
 Continuous data: 連續資料
 Continuous series: 連續數列
 Continuous variable: 連續變量
 Contra-harmonic mean: 反倒數平均數
 Contrary class: 反組
 Coördinate axes: 坐標軸
 Coördinate: 坐標
 Correction: 校正
 Correlated measures: 相關數量
 Correlation or Co-relation: 相關; 繫聯
 Correlation ratio: 相關率
 Correlation surface: 相關面

Correlation table: 相關表
 Cost index number: 成本指數
 Cost of living index number: 生活費指數
 Cross check: 相互校對
 Crossing formula: 交叉公式
 Crude mode: 概約衆數
 Cubic chart: 立體圖
 Cumulative chart: 累積圖
 Cumulative curve or ogive: 累積曲線
 Cumulative curve chart: 累積曲線圖
 Cumulative error: 累積差誤
 Cumulative table: 累積表
 Curve fitting: 曲線配合
 Curve chart: 曲線圖
 Curve plotting: 曲線繪製
 Curvilinearity: 曲線性
 Cycle: 循環
 Cyclical movement: 循環變動

D

Data: 資料
 Death rate: 死亡率
 Decile: 十分位數
 Decrement: 減量
 Deflating index: 減縮指數
 Demography: 人口統計學
 Dependent variable: 倚變量
 Derivative table: 導來表
 Descending order: 遞降次序
 Deviation: 離中差
 Diagonal method: 對角線法
 Diagonal scale: 對角線尺度
 Diagram: 圖
 Diagrammatic presentation: 圖示
 Dichotomous classification: 兩分法
 Dichtester wert or Thickest value: 最密值
 Difference: 差; 差數 (如 First, Second or Third difference: 第一, 第二, 或第三差)
 Difference correlation: 差相關

Direct correlation: 正相關
 Discontinuous variable: 間斷變量
 Discrete data: 間斷資料
 Discrete series: 間斷數列
 Dispersion or variation: 離散度
 Dissociation: 不相聯
 Distribution: 分配
 Divergence curve chart: 分歧曲線圖
 Dotted map: 加點地圖
 Double-entry table: 雙列表
 Double lines: 雙線
 Double-logarithmic: 雙對數
 Double or Dual tabular form: 二次表式
 Downward bias: 向下偏

E

Empirical curve: 經驗曲線
 Empirical formula: 經驗公式
 Empirical mode: 經驗衆數
 Enumeration: 挨查
 Enumerator: 調查員
 Estimates from correspondents: 通信估計
 Exponential average: 指數平均數
 Exponential curve: 指數曲線
 Extensive sampling: 擴大抽樣
 Extrapolation: 外推法
 Extreme: 極端

F

Factor reversal test: 因子還元試驗
 Fair average: 適當平均
 Field of investigation: 調查範圍
 First quartile: 第一四分位數
 Fitting a second degree parabola: 二次拋物線配合
 Fitting a straight line: 直線配合
 Fitting of curves: 曲線配合
 Fixed base: 固定基期
 Fixed weighting: 固定加權
 Fluctuation: 變動

Forecasting: 預測
 Forecasting sequence: 預測順序
 Foreign trade index number: 國際貿易指數
 Formula bias: 公式之偏誤
 Fourfold correlation: 四重相關
 Free-hand method: 徒手畫法
 Frequency: 次數; 頻數
 Frequency curve: 次數曲線
 Frequency distribution: 次數分配
 Frequency histogram: 次數直方圖
 Frequency polygon: 次數多邊圖
 Frequency series: 次數數列
 Frequency surface: 次數面
 Frequency table: 次數表
 Functional relationship: 函數相關

G

Gap: 斷缺
 Gaussian curve: 蓋氏曲線
 Geneological chart: 系譜圖
 Geographical order: 地理次序
 Geographic classification: 地理分類
 Geographic series: 地理數列
 Geometric mean: 幾何平均數
 Given period: 計算期
 Given price: 計算價
 Given year: 計算年
 Goodness of fit: 配合之適度
 Graduation: 分度
 Graphic calculation: 圖算
 Graphic interpolation: 作圖插補法
 Graphic method: 圖示法
 Graphic presentation: 圖示
 Graphic solution: 圖解
 Graph: 圖
 Great numbers: 大數
 Grouped data: 分類資料
 Grouped or class index number: 類指數
 Guessed mean: (見 Arbitrary mean)
 Guide line: 導線

H

Hand-sorting method: 手工分類法
 Harmonic mean: 倒數平均數
 Heading: 標目
 Heterogeneous: 異質
 High correlation: 高度相關
 Histogram: 直方圖
 Historical series: 歷史數列
 Historigram: 歷史圖
 Homogeneous: 同質
 Horizontal bar-chart: 橫條圖
 Horizontal scale: 橫尺度
 Hundred-per-cent bar-chart: 百分比條圖
 Hybrid system: 混合法
 Hyperbola: 雙曲線
 Hypernormal: 超常態

I

Ideal frequency distribution: 理想次數分配
 Increment: 增量
 Independent event: 單獨事件
 Independent variable: 自變量
 Index number: 指數
 Index of correlation: 相關指數
 Inertia of large numbers: 大數惰性
 Informant: 答問人; 填報者
 Inspectional mode: (見 Crude mode)
 Inter-census year: 兩普查間年份
 Interpolation: 內推法; 插補法
 Interval: 距度
 Inverse correlation: 負相關
 Irrelevancy: 無關
 Item: 項目

J

J-shaped curve: J形曲線

K

Key: 圖例; 表例
 Kurtosis: 峯度

L

Lag: 落後
 Lagging correlation: 落後相關
 Lagging curve: 落後曲線
 Large numbers: 大數
 Law of causality: 原因律
 Law of large numbers: 大數律
 Law of possibility: 或然律
 Law of regularity: 常性律
 Lead: 引前
 Least square: 最小平方
 Legend: (參看 Key)
 Leptokurtic: 高狹峯的
 "Less-than" cumulation: 以下累積
 Life table: 生命表
 Linear correlation: 直線相關
 Linearity: 直線性
 Line chart: 線圖
 Line of means: 平均數線
 Line of regression: 迴歸線
 Link-relative: 環比
 Logarithmic chart: 對數圖
 Logarithmic curve chart: 對數曲線圖
 Logarithmic equation: 對數方程式
 Logarithmic scale: 對數尺度
 Long-time change or fluctuation: 長期變動
 Low correlation: 低度相關
 Lower limit: 低限
 Lower quartile: 下四分位數

M

Magnitude: 大小度
 Major heading: 大標目
 Manifold classification: 多組分類
 Marriage rate: 結婚率
 Mass: 大量
 Mass observation: 大量觀察
 Material: 材料
 Maximum: 最大
 Maximum ordinate: 最高縱坐標

Mean: (參看 Average)
 Mean deviation: 平均差
 Mean difference: 均互差
 Mean square contingency coefficient: 均方相依係數
 Mean-square-deviation: 均方差
 Mean-square error: 均方差誤
 Measure of characteristics: 表徵數量
 Measure of differences: 差量數
 Median: 中位數
 Median-deviation: 中位差
 Median-link-relative: 環比中位數
 Mesokurtic: 常態峯的
 Mid-point: 中點
 Mid-score: 中分數
 Mid-value: 中值
 Mid-value of class: 組距中值
 Mid x/y ratio correlation: x/y中比相關
 Minimum: 極小
 Modality: 峯性
 Mode: 衆數; 範數
 Modulus: 模差
 Moment: 動差
 "More-than" cumulation: 以上累積
 Mortality: 死亡率
 Moving average: 移動平均數
 Multimodality: 多峯性
 Multiple association: 複相關
 Multiple contingency: 複相依
 Multiple correlation: 複相關
 Mutilated distribution: 偏破分配
 Mutually exclusive event: 互拒事件

N

Natural scale: 自然尺度
 Natural unit: 自然單位
 Negative correlation: 負相關
 Negative skewness: 負偏態
 Nomograph: 圖算法
 Non-comparable data: 無公比性資料
 Non-linear or Curvi-linear regression: 非直線或曲綫迴歸

- Non-linear or Non-rectilinear correlation: 非直線相關
 Normal correlation: 常態相關
 Normal correlation surface: 常態相關面
 Normal curve: 常態曲線
 Normal curve of error: 差誤常態曲線
 Normal curve of frequency or Normal frequency curve: 常態次數曲線
 Normal distribution: 常態分配
 Normal equation: 標準方程式
 Normal frequency distribution or Normal distribution: 常態次數分配或常態分配
 Normal histogram: 常態直方圖
 Normal law of error: 差誤常態律
 Normal probability curve: 常態機率曲線
 Normal surface: 常態面
 Notation: 記法
-
- Observation: 觀察
 Obtained mean: 實得平均數
 Order distribution: 次序分配
 Ordinate: 縱坐標
 Organization chart: 組織圖; 系統圖
 Origin: 原點
 Original table: 原始表
 Oscillation: 顫動
 Overlapping: 交疊
 Overlapping of frequency curve: 次數曲線之交疊
- P
- Part correlation: 部分相關
 Partial association: 淨關聯
 Partial contingency: 淨相依
 Partial correlation: 淨相關
 Partial regression: 淨迴歸
 Partial regression coefficient: 淨迴歸係數
 Percentile: 百分位數
- Percentile curve: 百分曲線
 Percentile rank: 百分等級
 Perfect correlation: 整相關
 Periodicity: 週期
 Periodogram analysis: 週期循環分析
 Permanence of small numbers: 小數恆性
 Permutation: 錯列
 Personal inquiry: 親身訪查法
 Pictogram: 像形圖
 Pin chart: 標針圖
 Pin map: 標針地圖
 Platykurtic: 平闊峯的
 Plot: 繪製
 Point measure: 點量數
 Point of inflection: 轉向點
 Polygon: 多邊圖
 Positive correlation: 正相關
 Positive skewness: 正偏態
 Possible accuracy: 可能確度
 Possible error: 可能差誤
 Precision: 精度
 Prediction: 預料
 Predominant value: 最衆值(即 Mode)
 Presence: 存
 Price level: 物價平準
 Price relative: 價比
 Primary or Original data: 原始資料
 Primary record: 原始紀錄
 Primary source: 原始來源
 Primary table: 原始表
 Probability: 機率
 Probability curve: 機率曲線
 Probability integral: 機率積分
 Probability surface: 機率面
 Probable error: 機誤
 Product moment correlation: 積差相關
 Profile curve: 側面曲線
 Progress chart: 進行圖
 Prophecy: 預占
 Prophecy formula: 預占公式
 Punching machine: 打孔機

Q

Quadrants: 象限
 Quadruple tabular form: 四次表式
 Qualitative observation: 質的觀察
 Quality: 質
 Quantitative classification: 數量分類
 Quantitative criteria: 數量標準
 Quantitative data: 數量資料
 Quantitative observation: 量的觀察
 Quantity: 量
 Quartile: 四分位數
 Quartile deviation: 四分位差
 Question blank or Questionnaire: 查問表

R

Random: 隨機
 Random or Accidental or Irregular fluctuation: 隨機或意外或不規則變動
 Random sampling: 隨機抽樣
 Range: 全距
 Rank: 等級
 Rank correlation: 等級相關
 Rank difference: 等級差
 Rank difference square correlation: 等級差方相關
 Rank distribution: 等級分配
 Rank method: 等級法
 Rate: 率
 Rate-of-change curve chart: 變率曲線圖
 Ratio chart: 比率圖
 Ratio method: 比率法
 Ratio scale: 比率尺度
 Reciprocal: 倒數
 Rectangular distribution: 長方分配
 Rectilinear correlation: (見 Linear correlation)
 Rectilinearity: 直線性
 Reference point: 參照點
 Registration method: 登記法

Regression: 迴歸
 Regression coefficient: 迴歸係數
 Regression equation: 迴歸方程式
 Regression line: 迴歸線
 Relative: 比
 Relative dispersion: 相對離散度
 Relative error: 相對差誤
 Relative number: 比數
 Relative price: 比價
 Relative quantity: 比量
 Relative value: 比值
 Reliability: 可靠性
 Representative data: 代表資料
 Representative sampling: 代表抽樣
 Residual movement: 剩餘變動
 Reversal test: 還元試驗
 Rolling average: (見 Moving average)
 Root-mean-square-deviation: 均方根差
 Row: 橫行

S

Sample: 樣本
 Sampling: 抽樣
 Sampling bias: 抽樣偏誤
 Sampling method: 抽查法
 Scale: 尺度
 Scale designation: 尺度標記
 Scale number: 尺度數目
 Scale point: 尺度點
 Scale unit: 尺度單位
 Scatter diagram: 散佈圖
 Schedule: 調查表格
 Seasonal variation or fluctuation: 季節變動
 Secondary data: 次級資料
 Secondary source: 次級來源
 Secondary table: 次級表
 Sectioned bar-diagram: 分段條形圖
 Sector diagram: 扇形圖
 Secular trend: 長期趨勢
 Self-correlation: 自身相關
 Semi-average: 半平均數

- Semi-interquartile range: 四分位差
 (即 Quartile deviation)
 Semi-logarithmic: 半對數
 Series: 數列
 Shaded or Cross-hatched map: 陰影
 線或交叉線地圖
 Shifting base: 變動基期
 Short-cut method: 簡捷法
 Short-time fluctuation: 短期變動
 Significant difference: 重要差
 Simple arithmetic mean: 簡單算術平
 均數
 Simple bar-chart: 單式條圖
 Simple correlation: 單相關
 Simple curve chart: 單式曲線圖
 Single or Simple tabular form: 簡單
 表式
 Size: 大小
 Skew: 偏斜
 Skewed histogram: 偏態直方圖
 Skew line: 偏斜線
 Skewness: 偏斜度
 Slope: 斜度
 Smoothed curve: 修勻曲綫
 Sort: 分類
 Sorting machine: 分類機
 Source of data: 材料來源
 Spot map: (見 Dotted map)
 Spread: (見 Deviation)
 Spurious correlation: 不純相關
 Staircase curve: 階形曲綫
 Standard deviation: 標準差
 Standard error: 標準誤
 Standard of accuracy: 確度標準
 Statistical method: 統計方法
 Statistical unit: 統計單位
 Statistics: 統計或統計學
 Statistics of attributes: 質的統計
 Statistics of variables: 量的統計
 Step: 級
 Straight line correlation: (見 Linear
 correlation)
 Stub: 橫標目
 Subclass: 小組
 Subclassification: 小分類
 Sub-group: 小組
 Sub-heading: 小標目
 Sub-normal: 亞常態
 Subordinate class or Subsidiary class:
 附類
 Subscript: 下標
 Sub-total: 小計
 Superimpose: 重疊
 Superscript: 上標
 Surface chart: 平面圖
 Symbol: 符號
 Symmetrical curve: 對稱曲綫
 Symmetrical distribution: 對稱分配
 Symmetry: 對稱
- T**
- Table: 表
 Tabular form: 表式
 Tabular order: 表次
 Tabulating card: 列表卡片
 Tabulating method: 表列法
 Tabulation: 表列
 Tail (of a distribution): 尾端 (在一分
 配表中)
 Temporal classification: 時間分類
 Ten percentile: 十分位數
 Theoretical frequency: 理論次數
 Theoretical mode: 理論衆數
 Third quartile: 第三四分位數
 Time reversal test: 時間還元試驗
 Time series: 時間數列
 Title: 標題
 Total: 總數
 Total association: 全相關
 Total correlation: 全相關
 Total regression: 全迴歸
 Transcription: 轉錄
 Transmutation: 蛻變
 Trend: 趨向
 Triple tabular form: 三次表式

True mean: 眞實平均數
 True median: 眞實中位數
 True mode: 眞實衆數
 Type: 型
 Type bias: 型偏
 Typical data: 代表資料
 Typical form: 代表式

U

Ultimate class: 末組
 Unclassified: 未分類
 Uncorrelated: 不相關
 Ungrouped data: 未歸類資料
 Uniformity: 劃一
 Unimodal: 單峯的
 Unimodality: 單峯性
 Unit: 單位
 Universality: 普遍性
 Unlike signed correlation: 異號相關
 Unreliability: 不可靠性
 Unsmoothed curve: 未修勻曲線
 Unweighted mean: 未加權的平均數
 Upper limit: 高限
 Upper quartile: 上四分位數
 Upward bias: 向上偏
 U-shaped distribution: U形分配

V

Value: 值
 Variability: (見 Dispersion)
 Variable: 變量

Variable error: 變量差誤
 Variable weighting: 變量加權
 Variate: 變數
 Vertical bar-chart: 縱條圖
 Vertical scale: 縱尺度
 Vital index: 生命指數
 Vital statistics: 生命統計
 Volume graph: 容積圖

W

Weight: 權數
 Weight bias: 偏權
 Weighted mean or average: 加權平均數
 Weighting method: 加權法

X

"X"-axis: X軸

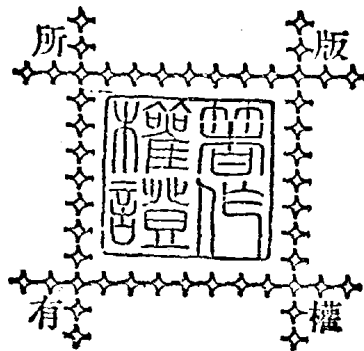
Y

"Y"-axis: Y軸

Z

Zee chart: Z形曲線圖
 Zero correlation: 零相關
 Zero line: 零線
 Zero point: 線點
 Zero score: 零分
 Zigzag form: 鋸形
 Zone curve: 距限曲線

一九四一年四月發行
一九四九年十月四版



原著者

F. C. Mills

譯者

李黃孝貞
陸宗蔚

發行人

李虞杰
中華書局股份有限公司代表

印刷者

上海澳門路八九號
中華書局永寧印刷廠

發行處

各埠中華書局

大學
書統
計方法 (全一册)

◎

基價並精裝本三十一元五角

二五元五角

(郵運匯費另加)

(111709)(海)



(12709)