

大學叢書

統計學大綱

上 冊

金國寶著

商務印書館發行

大學叢書
統計學大綱

上 冊
金 國 寶 著

(一九五〇年十二月修訂)

原國立上海商學院叢書
商務印書館發行

一九五〇年修正版序言

余早歲有統計學大綱一書之編。自開始編輯至交付手民，閱時不下十載，自印行以來，轉瞬亦已十六七年。此二三十年之中，世界各國各種科學靡不有長足之進步，而統計一門尤有一日千里之勢。獨我國以喪亂之餘，民生凋敝，學術衰退。而余個人自服務銀行以來，書籍久束高閣，學業尤為荒落，故此書迄未修訂，殊慚報焉。去歲秋，應復旦大學之聘，教授統計，乃得重溫舊書，並取海外新出之本稍稍瀏覽。同時為搜集教材故，乃乘機取統計學大綱而修訂之，其中謬誤者改正之，闕漏者補充之。原書二十章，減去二章，新添三章，結果反增一章，合計二十一章。除新增各章外，其原有各章中亦頗多增損之處。例如用累積頻數法計算標準差、偏態與繫聯係數等，公式過煩，故予刪除。第一章加入計數之方法，第十九章常態曲線加入曲線公式之證明，第八章指數亦加入若干公式，第十八章複繫聯以原書過簡，故重行寫過。書中例題亦頗多改用本國統計數字。惟以書局趕應秋銷，時間尙促，不克詳細訂正。疵謬之處仍所不免，尙祈海內方家予以指教，則幸甚矣。

金國寶 一九五〇年十月

引 言

(一)此書編輯實開始於十年之前，時余授課於上海中國公學、復旦大學、國立暨南大學、商科大學及政治大學等校。如第四、第五、第六、第七(一部)、第十、第十一、第十二及第十五等章均於此時編成。民國十七年余奉前大學院院長蔡子民先生命往歐美各國考察統計事業，凡一年。除搜集各種資料外，並就此書續編幾章，如第一、第十四、第十六及第十七等四章，均於此時續成者也。十八年返國以後，余供職前南京財政局及上海交通銀行，俗務紛集，日不暇給，此書遂又擱置筐篋中者三年。去歲應國立上海商學院之聘，授指數編製法一科，加以蔡正雅褚鳳儀二先生之敦促，因將舊稿重加整理，又續成若干章。並經褚君詳為校閱，勉得付印。其中疵謬之處自知不免。倘承博雅君子不吝指教，則幸甚矣。

(二)余前有統計新論及物價指數淺說二書之輯。但本書譯名與前二書稍有不同，其中尤重要者：frequency 譯為頻數，correlation 譯為繫聯。『頻數』一詞遠勝於舊譯『次數』；『繫聯』一詞似亦較勝於舊譯『相關』。此二名詞均胡明復先生所擬定，民國十五年科學名詞審查會所採用。明復先生人皆知為數學家，不知其於統計學亦有研究與貢獻，故特表而出之，以資紀念。

(三)統計符號各家所用頗不一致。例如算術平均數有以 A 表示者亦有以 M 表示者；即在同一書內亦不一致，有時以 M 代表算術平均數，有時以 M 代表中位數。甚至一個符號代表二三種以上之意義。初學之士最易混淆。本書有鑒於此，力矯其弊，在可能範圍內，務使一個符號祇代表一種意義。

(四)本書凡二十章，如用為學校教本，可分一年半讀完。如其祇讀

一年者，除前列十三章外，須加讀第二十章一章。如其用爲經濟統計或商業統計教本，則可選讀第八、第九、第十一、第十二、第十三、第十四及第十九等七章，半年讀完。

(五)本書附錄甲爲各公式之數理研究，學生數學程度不高者可以不讀。如已讀過高等代數微積分者，則可同時研習之。

(六)本書除余歷在各大學教授外，並承蔡正雅先生在暨南大學，褚鳳儀先生在上海法學院用作教本，前後講授已不下六七次，易稿亦不下二三次。此次付印，又承褚君詳細校閱增補，尤爲感佩，特贅數語，以誌不忘。

金國寶

二十三年四月二十五日

上 冊 目 次

引 言

第一章 緒論	1
第一節 統計學之定義及統計之應用	1
第二節 統計之誤用	2
第三節 統計之法則	5
第四節 統計方法之程序	7
第五節 計數之方法	7
第二章 統計資料之搜集與整理	12
第一節 統計資料之搜集方法	12
第二節 次級資料之編製	14
第三節 原始資料之搜集與整理	16
第三章 統計表	20
第一節 統計表之功用	20
第二節 統計事項之特性及其相互之關係	20
第三節 統計事項之分類	21
第四節 總表與摘要表	22
第五節 統計表之形式及製表規律	23
第六節 統計數列	27
第七節 頻數表	28
第四章 統計圖	36
第一節 統計圖之功用及製圖之原則	36

第二節	統計圖之分類	36
第三節	條形圖	37
第四節	統計地圖、面積圖及體積圖	45
第五節	線圖	48
第六節	作圖規則	56
第五章	平均數	59
第一節	平均數之意義與種類	59
第二節	算術平均數	60
第三節	中位數	70
第四節	四分位數、十分位數及百分位數	72
第五節	衆數	76
第六節	幾何平均數	80
第七節	倒數平均數	83
第八節	各種平均數之比較	87
第六章	離中趨勢	93
第一節	離中趨勢之意義及其測定之方法	93
第二節	全距	94
第三節	四分位差	94
第四節	平均差	96
第五節	標準差	99
第六節	均方差	107
第七節	各種離中差之關係	111
第七章	偏態峯態與動差	114
第一節	偏態之意義及其形式	114
第二節	測定偏態之方法	119

第三節 峯態	125
第四節 動差	126
第五節 計算動差之簡法	128
第六節 薛伯氏校正法	131
第八章 指數	134
第一節 指數之意義與種類	134
第二節 物價指數編製之方法	134
第三節 指數公式之測驗	149
第九章 直線繫聯	157
第一節 直線繫聯之意義	157
第二節 繫聯直線之測定	158
第三節 標準誤	161
第四節 繫聯係數	163
第五節 響應直線與響應係數	176
第十章 長期趨勢	184
第一節 長期趨勢之意義及其測定	184
第二節 直線趨勢之測定	185
第三節 曲線趨勢之測定	197
第四節 計算趨勢曲線之簡便法	201
第十一章 季節變動	205
第一節 季節變動之性質及其效用	205
第二節 季節變動存在之確定	206

第三節 季節指數之計算	210
第十二章 循環變動	230
第一節 循環變動之意義及其起因	230
第二節 循環變動之測定	231
第十三章 時間數列之繫聯	239
第一節 時間數列繫聯之特性	239
第二節 循環變動繫聯之測定	240
第三節 短期變動之繫聯	245
第十四章 曲線配合	248
第一節 概說	248
第二節 定點法	250
第三節 平均法	253
第四節 最小平方法	254
第五節 動差法	255
第六節 直線原點之移轉	256
第七節 指數曲線	259
第八節 乘方函數	265
第九節 拋物線	270
第十節 雙曲線等	273
第十一節 康撥茨曲線	277
第十二節 天文曲線	282
第十三節 適合度	284

統計學大綱

第一章 緒論

第一節 統計學之定義及統計之應用

統計學者，用計數或估量以數字表示社會或自然現象之動態或靜態並分析其數字間關係之學也。此定義須稍加以解釋。

統計學計量而不較質。欲比較人之智愚，在統計學內必須先有可以表示此智愚之數量方可以言比較，故數字不能與統計分離。統計學上大半數字均由計數而來，故雷翁氏以計數之學作為統計學之定義。此定義雖覺太狹，但計數為統計學之主要職務要無可疑。惟統計學上之數字未必均由計數而來，有時不得不用估量方法以求其近似之數值，故定義中計數與估量並列。

古代統計學研究之對象為國家，故有以研究國家之學為統計學之定義者。其後研究之範圍漸次推廣，研究之對象亦漸由國家而推及於社會與自然現象；此種現象或同時同地，或同時異地，或同地異時，故社會與自然現象之動態與靜態均在統計學研究範圍之內。

統計學亦有作為研究平均數之科學者。此定義亦覺太狹。統計學不特用數字表示社會與自然現象之動態與靜態，且用種種分析方法以推求其數字間之關係；此種關係不僅是平均數一種。故謂統計學之任務在根據大量觀察而闡明其數字間之平均關係則可。若謂統計學為研究平

均數之學，則不免令人誤解矣。

統計之應用甚廣，不勝枚舉，茲擇其最重要者分述如下：

(一)統計與公共衛生 人口之疾病死亡統計對於公共衛生尤有密切之關係。衛生當局之唯一參考即在統計；凡以後施政之方針及以前設施之成績無不取決於此。當疾病發生之際，即可由統計之報告而設法防止其蔓延；平日亦可注意社會之弱點而徐圖補救之辦法。且若將此等統計及其效用公告人民，尤不難得社會之合作與經濟之贊助，誠推廣公共衛生事業之第一急務也。

(二)統計與工業 現代工業範圍擴大，故其問題亦日趨複雜。例如浪費之減少，成本之分析以及生產品質之控制等等，均不能不乞靈於統計。故歐、美各國之各大工業機關，不論國營民營，每皆特設統計部以專司檢查分析之職責。

(三)統計與農業 農產品之收成往往與溫度、雨量有密切之關係，而土壤肥料對於農產品之收穫亦有甚大之影響。故一方面經濟學家與統計學家利用雨量、溫度等統計資料以預測各種農產品之收成，一方面農學家復利用統計技術，以測定各種肥料與土壤，對於各種農作物之影響，並藉以改進農業之技術。此皆統計應用於農業之實例也。

(四)統計與財政 財政以收支適合為原則；支出雖較能預定，然收入卻頗難預言。例如所得稅之多寡須視人民所得之數額而定；關稅之收入須視外國輸入品之種類與數額而定；然歐、美各國之財政專家每能根據歷年之統計而預測未來之收入，雖亦有時與實收數目相差甚遠，然適合者其常，而相差極大者僅例外事耳。

第二節 統計之誤用

統計之用固極神妙，然用之失當，其流弊所屆亦有不可勝言者。吾

人格物務須平心靜氣屏除成見，取懷疑之態度，戒獨斷之行爲，如是細心分析方能得事物之真相；否則，毫釐千里未有不陷於絕大之謬誤者。美國統計學家卻獨克氏對於統計學之誤用論列甚詳，據其所論可分謬誤爲四種：一曰，不同事物比較之謬誤；二曰，百分比之謬誤；三曰，原因脫漏之謬誤；四曰，偏見之謬誤；而統計自身之謬誤猶不與焉。卻氏對於每種謬誤各附以若干例證，茲節取其一二如下：

（一）不同事物比較之謬誤 統計之妙用端在比較。單獨一個數字毫無意義可言，必有兩處地方或兩個時期相互比較，意義始明；然事物之性質不相同者則無比較之可能。例如一八九九年美國陸軍部長論菲律賓之美國兵士死亡率事，即犯此病。當時外界對於多數兵士之死亡頗有責言，該部長乃出而置辯；大意謂兵士之死亡率不過萬分之一七二，與華盛頓波士頓一般人口之死亡率相差無幾，故兵士之死亡率不得謂爲過高云云。其實軍隊與一般人口，性質完全不同，安能相提並論？一般人口之中，老少齊全，而極老極少者之死亡率尤高出尋常，斷不可與中年人相比較；而兵士則既悉爲強壯之青年，且均經過體格檢查者，故此種比較實自欺欺人耳。

又如美國麻薩諸塞州勞工統計局嘗有母性比較之報告，亦犯同一之謬誤。其根據爲一九〇五年之人口調查；調查之時，先就州內現有之母親詢其所生子女之數目，再詢此等母親之母親所生子女之數目。於是據此以計算此兩代各代所生子女之平均數；其結果爲前代之母親每人平均生 6.47 人，而後一代祇 2.77 人，此後一代母親之生殖力似乎大減；其實不然，蓋在一九〇五年調查之時，此後一代之母親尙未至生育期告終之時，則其未來之子女正方興未艾，不能遽以已生之數爲定論也。

（二）百分比之謬誤 百分比之使用亦須格外注意，偶一不慎，即可令人發生謬誤之感想。例如美國約翰哈金斯大學初收女生之時，即發生

一有趣之新聞，謂該校女生百分之三十三又三分之一均與本校教員成眷屬云。不知底細者必以為該校教員均風流人物，然細加考察，則與教員結婚之女生僅一人而已。蓋當時女生共祇三人，一人即為其全體百分之三十三又三分之一。故數目甚小之時，不宜用百分比。如欲用百分比時，亦必須將實在數同時並列，方不致令人發生謬誤之印象也。

又如甲城人口十萬而外國移民居百分之二十，乙城人口五十萬而移民居百分之三十，丙城人口百萬而移民居百分之四十，今若將此三城合而計之，試問其移民成分幾何？對此問題往往即將此三城之百分比相加而以三除之為答，則其答數為百分之三十：

$$\frac{20+30+40}{3}\% = 30\%$$

然應得之百分比當為 35.625 而非 30 也；因準確之計算法當以三城人口之總數除其移民之總數。甲城之移民有二萬，乙城之移民有十五萬，丙城之移民有四十萬，故三城移民之總數共有五十七萬，而其人口之總數則有一百六十萬，列成算式即得：

$$\frac{570,000}{1,600,000} = 35.625\%$$

(三) 原因脫漏之謬誤 有時事實之原因甚多，若獨取其一，盡置其他於不顧，亦常發生不確之結論。如美國某大學調查學生吸煙程度以斷定吸煙為學業不及格之原因，即其一例。茲將學生分為吸煙極多者，吸煙不多者，完全不吸者三類，而其中不及格之人數如下：

調查學生數 全年平均分數 不及格之百分比			
完全不吸者	111	85.2	3.2
吸煙不多者	35	73.3	14.1
吸煙極多者	18	59.7	24.1

根據此項調查結果，遂認吸煙為不及格之原因則未免失當；蓋學生之好

吸煙者其人往往視他種活動較重於學業，同時或爲體育家，或爲極貧舒適之人，其所以不及格者由於其不重學業所致，而吸煙不過爲其不重學業之一種間接表示。科學家之探討必須盡窺事物之全豹，不當以部分的理由作全體之解釋也。

(四)偏見之謬誤 統計學家當有超然中正之態，切不可先有偏見存乎胸中，乃覓統計以實其說；自欺欺人，莫此爲甚。例如前年美國曾有反對種痘同盟會之職員投函於紐約晚報（一九一四年五月四日），引用英國之統計如下：

天花死亡總數（1905—1910年）	199
種痘死亡總數（1905—1910年）	99
五歲以下天花死亡總數（1905—1910年）	26
五歲以下種痘死亡總數（1905—1910年）	98

大意謂就全體而論，種痘死者僅佔天花死者之半，就五歲以下而論，種痘死者尚多出天花死者之上幾有四倍之多，故強迫種痘之舉殊可不必云云。所引統計固屬確實，但其議論殊與事實相反。天花死亡的減少即由於種痘。六年之中因種痘致死者祇有九十九人。假使不實行強迫種痘，則當時以天花死者必非少數，以之與九十九人相比，孰多孰少，不言可知矣。

然而以上種種猶非就統計本身言也。統計本身亦難免謬誤，或由於調查之疏忽，或由於計算之錯誤，故於他人所披露之統計，吾人須詳加分析，不可輕信。差以毫釐，謬以千里，不可不慎之又慎也。

第三節 統計之法則

抽樣爲近世所發見最有價值之調查方法。所謂抽樣即自一大羣極複雜之事項中抽取一小部分作爲調查之標準，由此所得之結果即可用

以代表全部。例如欲調查上海工人所得之平均工資，吾人不必遍查全部工人所得之工資再求其平均數，吾人祇須抽查其中可以代表全體的一小部工人所得之工資而求其平均數。由是而得之平均工資雖未必與全體工人所得之平均工資完全一致；然相差甚微，實際上可以略而不計，故即以之作爲全體工人所得平均工資之代表亦無不可。又設有雞蛋十萬枚而欲求其平均重量，吾人不必將此十萬枚雞蛋一一秤其重量再求其平均數。吾人祇須任取（當然不能故意選擇最大或最小之雞蛋）其中一千枚雞蛋秤其重量而求其平均數。由是而得之平均重量雖未必與十萬雞蛋之平均重量完全一致；然相差無幾，實際上已可用爲全部雞蛋平均重量之代表。此種調查方法係根據統計常態之法則。所謂統計常態即謂由一大羣中任意選擇之一小部平均差不多可以保持全部之特性。

由一大羣中任意抽出之一部既能代表全體，則由此一大羣中抽出之其他一部自當與第一部相似。若第一部中有幾項具有異常之特性，則在第二部中吾人亦可預期發見具此異常特性之幾項，其項數亦與前無甚差異；此即所謂小數永存之法則，蓋由統計常態之法則脫胎而來也。統計學家蒲蘭謂各種職業之專家白專醫特種難症之耳科醫生以至販賣古董之商人，靡不賴此「小數永存之法則」而生。

由統計常態之法則脫胎而來尚有大量惰性之法則。所謂大量惰性乃謂在外界原因不變之情況下若觀察之範圍擴大甚廣則每年之統計常得相似之數量。例如火災之損失就一城而言歷年之損失或相差甚多；然就全國或全世界而言，若房屋之建築或防火之設備未有改進，則每年火災之損失常能保持一定之數量；蓋各處每年火災損失數量之變動，其方向不同，有較去年增加者，亦有較去年減少者，甲乙等地增加之量適與丙丁等地減少之量約略相抵，故其結果變動甚微。

第四節 統計方法之程序

統計方法之程序，可分為四大步驟：

- (一) 搜集資料
- (二) 整理資料
- (三) 發表資料
- (四) 分析資料

請舉例以明之。今設欲清查某地之人口，自當首先確定其清查之範圍、調查表之問題及格式、施行方法、以及其舉行日期等等，此皆屬於其初步「搜集資料」之工作。迨資料既已齊集，則當進而點明其人口之總數，男女各若干？已婚與未婚者各若干？識字與不識字者又各若干？其年齡之分配如何？其職業之分配又如何？類此之工作皆屬於其第二步「整理資料」之範圍。今既得其統計各事項之數目，即得酌量情形製成圖表以公布之，此即其第三步「發表資料」之工作。大半統計機關之工作至此為止。完成此三項工作亦可謂為已盡「計數」之能事；然統計學之效用卻不但示吾人以各種事項之確切數目，尤當由其所示之數目間發現一定之規律；例如由各國人口年齡統計之比較而發現其分配曲線大致有一定之形式，再若由其已婚者年齡之研究而得計算其夫婦間年齡之繫聯係數，諸如此類之工作概屬於其第四步「分析資料」之範圍。規模較大之統計機關亦甚努力於此種工作。本書第二章述統計資料之搜集與整理，第三、第四兩章述圖表之繪製，第五章至第十九章則分別詳論統計資料之分析。

第五節 計數之方法

統計學上所用之數值均由觀察而來。觀察值之準確性恃乎量度者

之技術與審慎以及所用測量儀器之種類。例如有木一株，由一百人量之，雖不致有一百個不同之結果，然取得五六個不同之結果往往有之。在測量之人均自以爲最準確，然此木之真實長度祇有一個，斷不能謂五六個不同之結果均正確，即使中間一人確能量得真實之長度，然誰能知之，而誰能斷言之，故一切觀察值均祇是幾近值而已。

故準確一詞亦是相對的，並非絕對的。絕對的準確爲事實上不可能之事。初學之士以爲 9.7 之意義與 9.70 相同，並與 9.700 亦相同。不知其意義大不相同。照習慣上之解說，9.7 者謂其真值在 9.65 與 9.75 之間。如有一繩量得長度爲 18 寸者，其意義謂在 17.5 寸與 18.5 寸之間。吾人謂月球離地球之距離爲 240,000 英里，考其意謂其實在距離在 235,000 英里與 245,000 英里之間。

如有一物，量得 15 寸者，其意蓋謂此物之量度準確至最近之一寸 (correct to the nearest inch)。有一物量得 9.7 公分者，則謂此數準確至最近一公分之十分之一。此種數字往往書作 9.7 ± 0.05 ，此處 0.05 則謂可能有 .05 之差也。同理，9.70 則謂在 9.695 與 9.705 之間，可書作 9.70 ± 0.005 ，故相對的誤差爲 $.005/9.70 = .0005 = .05\%$ 。

無論變數爲連續與不連續，一個數目之真值總是在測量之最後一位以下半單位至以上半單位之間。

由此觀之，9.7 與 9.70 之意義不同，蓋 9.7 之有效數字爲兩位，而 9.70 之有效數字爲三位。下列諸數之有效數字均五位：108.05, 184.10, 0.0086391, 400.00。但月球之距離 240,000 英里，則有效數字祇二位。

照通常意義解說地球與日之距離爲 93,000,000 英里，似謂其真確之距離在 92,999,999.5 與 93,000,000.5 英里之間。其實不然，此處之有效數字祇 9 與 3 兩個數字，故實在距離在 92,500,000 與 93,500,000 英里之間。爲免此項誤會起見，吾人可書作 9.3×10^7 。故零字最易引起