生物统计之理论与实际

赵仁容
余松烈 合著

生物统计之理论与实践

[illegible]
[illegible]

大學教本

生物統計學理論與實際

趙仁鎔
余松烈 著

新農叢書

新農企業股份有限公司出版

生物統計之理論與實際

著作者　靖仁鎔　余松烈

發行委員會　鄭曼倩　余松烈　佴霖生
林子琦　高順濤

發行所　上海虎丘路14號41A室
新農企業股份有限公司

印刷者　上海徽寧路717弄12號
新農企業股份有限公司
印刷部

定價

中華民國三十七年九月再版

自　序

有關生物統計之中西名著,可能在坊間獲得者已屬不少,惟欲于其中選一合于吾國農學院學生程度而可用作教本者,實屬不易.或嫌過深,專注于理論之探討,非一般學子之數學程度所能瞭解,或嫌過淺,僅用實際示例介紹分析方法,讀後僅能依樣摸倣而不知所以然,較理想之生物統計教本,應理論與實際並重,介紹任一統計方法時,宜先闡明其基本原理,使讀者理解之,乃用極淺近之數學方法,引述計算公式來源,然後再以示例演算解釋之,不僅可作讀者實際應用時之模範,並可引導讀者作進一步之研究.

作者等鑒于此實際需要,並為達上述之目的計,乃于教學之餘,合力草成本書,並以其全部或一部材料,先後施教于協和大學,福建省立農學院,南通學院及復旦大學等校,本書雖經幾次試教與修正,仍覺遠遜于預定目標,錯誤之處,定所難免,尚望學者有以正之.

本書材料取自各學者之研究試驗結果者甚多,除分別于引述之處註明其來源外,特再于此對諸學者表示敬意.又作者等于撰作及試教本書時,蒙諸同仁同學之協助與合作,亦一併于此敬表謝意.

趙仁鎔　余松烈

三十六年四月一日識于上海

生物統計之理論與實際

目　　錄

第一章

生物統計簡史及重要述語簡釋

1.1 生物統計、統計、與統計方法 應用數學邏輯以解釋生物界數量的資料(Quantitative Data),稱曰生物統計(Biometry).在許多場合,生物統計常被視與統計(Statistics)同意義,由于統計乃生物統計之前驅,藉統計分析(Statistical Analysis)之基本原理及技術之發展,生物統計乃得奠定基礎.吾人若專指統計方法(Statistical Method),為以數學概念為基礎之各種推論(Reasoning)的表率,以生物統計為應用此種推論于生物界問題,則可對此二名詞有具體之瞭解與辨別.

統計與統計方法之意義為何?名統計學者 G. U. Yule 氏(1)嘗給予概括之定義,氏認為統計乃受多重原因影響至堪注意程度之數量資料.至特別適用于解釋受一多重原因影響數量資料之方法,則稱曰統計方法.

名統計學者 R. A. Fisher 氏(2)之見解則與 Yule 氏有異,氏認為"統計"一詞乃為理論與實用數學推論(Mathematical Reasoning)之一旁枝,用于處理資料之歸納分析(Inductive Analysis)而已.此項定義實源于在統計分析中,由研究某事物全體(集團**)之一部分(樣本**)以估計全體之特性.

1.2 生物統計簡史 統計為生物統計之前驅,故言生物統計歷史,探本索源,先明統計之歷史.

統計(德文 Statistik)一詞,為德學者 Achenwall 氏由中世紀拉丁文"Status"一字轉變而來,原意"國家"(Political State).其後為英國學者引用,且迅即推廣其意義,兼指一國之自然資源,產品,居民及兵力等數字.至十九世紀,統計一詞之意義,始漸由物質名詞轉變而意為一種使變數**資料概括成簡單明瞭之形式的一種方法.此種改變,一方面由于自 1800 年以來關于數學上的或然學說發展之故.

Bernoulli 氏或為想像應用或然原理于社會及經濟重要性之第一個數學家,法國數學家 La Place 氏及天文學家 Gauss 氏對或然學說亦有極大之貢獻,Gruss 氏更確立應用或然學說于真確科學之基礎.

及至十九世紀,比利時數學家兼天文學家 Quetelet (1796-1874) 將生物界之或

* 括弧內號碼係指參考文獻書號.

** 集團與樣本二名詞之意義見 1.3 節.變數一名詞之意義見 1.7 節.

然現象分為有系統之次數分配,渠將生物變異與社會研究成為機誤定律,即現在著名之常態曲線是也.同時 Francis Galton (1822-1911) 氏研究人種特性與遺傳,對資料之紀錄與分析極為詳細對英國生物數學上定一重要之基礎,氏對遺傳學方面之最大貢獻為相關之發現.

1890 年 Karl Pearson 對統計方法應用于生物方面,定下一重要之基礎,氏對生物統計之理論與實際均有極大之貢獻,1901年氏之第一本生物統計學(Biometrika)問世,至今已達數十版之多,此書為生物統計之至寶,Pearson教授之研究結晶盡在于此矣.

近三十年來英人 R. A. Fisher 對變量分析之發現,為生物統計劃時代之進步.氏對生物統計之貢獻至大,如對小樣本差異比較與 χ^2 比較等,均獨步一格,為近世統計學家之泰斗.最近十餘年來, F. Y. Yates 氏對田間試驗又有獨特之心得,先後發明混雜試驗與擬複因子,使田間試驗結果益為準確.餘如 Wishart, Snedecor, Goulden, Immer, Yule 諸氏,均為近代之生物統計學者,貢獻甚多.

生物統計近二十年來經 Fisher 氏變量分析之發現,進步極為迅速,尤其近十年來對田間試驗之研究方法突飛猛晉,有一日千里之勢,對科學之貢獻未可限量也.

1.3 集團與樣本 在統計學上,吾人常見集團(Population)與樣本(Sample)二名詞,前者乃指某事物之全體,包含屬于該事物之所有個體,後者乃言由集團中所選取之若干個體,為該集團之一部.

集團依其性質不同,有有限性(Finite)集團與無限性(Infinite)集團之分,及實有的(Real)集團與假設的(Hypothetical)集團之別.有限性集團之各個分子一一可數,大小一定,無限性集團則否,其所組成之分子無法可數,統計上有關抽樣機誤之各種公式幾皆假設源自無窮性集團,惟此等公式常應用于有限性之集團,蓋有限性集團在實際上亦包含甚大而可視為無窮大也.實有的集團為實際存在,假設的集團則否,實際上本無此物,僅存在于吾人之想像中而已.

1.4 集團之研究與隨機樣本 任何科學研究,其目的不外乎明瞭其研究對象之本質或特性,設吾人欲研究福建青蛙之體長情形,則吾人研究之對象即為福建青蛙集團,同樣若欲研究江蘇全省米價之變動情形,全國高中學生之國文程度,則其研究對象亦必為江蘇全省米價數字集團,與全國高中學生國文測驗之成績集團.

吾人之研究對象既為各該集團,而所欲明瞭者亦為各該集團之特性,是則吾人應將各該集團之個體或數字逐個研究,而冀得一結論.但實際上吾人無法能將福建青蛙集團之各個青蛙身體長度逐一度量之,無法普取江蘇全省各地各時期之米價

數字亦無法將全國所有高中學生予以同樣之國文測驗或雖可能,而所費之人力物力,必屬可觀,而不合乎經濟要求.

集團既無法或不易獲得,則該集團之特性,惟由其所屬一部之樣本研究之,而由研究樣本所得之結果以估計該集團之特性.由樣本之特性估計集團之特性時,當需以該樣本確能代表其所屬之集團或所來自之集團為前提換言之,即該樣本需合乎下列之要求:

(1) 樣本本質不偏不倚.

(2) 組成樣本之各個分子必需彼此獨立.

(3) 樣本所來自之區域間無顯著之基本差異.

(4) 組成樣本各項之條件必需相同.

如何獲得合乎上述要求之樣本實一難事,幸有或然律(Law of Probability)*之發現,乃得有所依據.或然律之意,乃謂在一群數量之中,隨機(At Random)擇取適當之若干數,而合法計算之,則所得結果,與由全體所得者,相差極近.所謂隨機,即不摻雜主觀于其內,不偏不倚,全憑機會.此種隨機方法所得之樣本,稱之曰隨機樣本(Random Sample),或簡稱樣本.由隨機方法所選得之樣本,方可代表集團,用以研究該集團之特性.

1.5 常軌數與估計常軌數 于此吾人又需介紹統計學上常用之另二述語,即常軌數(Parameter)與估計常軌數(Statistic)是也.所謂常軌數乃指根據整個集團所測得之數值,是為該集團之真正數值,固定而不變.至估計常軌數乃為根據樣本所測得之數值,而用以估計集團之常軌數者.估計常軌數因所用樣本之不同而有所變更.

設由福建青蛙集團測得該青蛙集團平均長度為6.50Cm,則此6.50Cm即為該青蛙集團之真正平均長度(True Mean),而為該集團常軌數之一.同樣若根據該青蛙集團測得真正標準差(標準差之意義見3.7節)為2.15Cm,此數值則為該集團之另一常軌數.設上述二數非根據集團測得,而由代表該集團之某樣本測得,則其名稱應依次改稱為估計平均數及估計標準差,而屬于估計常軌數矣.

因集團之無法或不易獲得,自亦無法測得集團之常軌數,僅惟有代表該集團之樣本所測得之估計常軌數以估計而已.統計及生物統計所考慮者,大都為估計常軌數.

---

*因或然律之產生,乃得統計學上重要法則之一,即所謂現象齊一法則(Law of Statistical Regularity)是也.現象齊一法則乃云:凡完善之統計對于某集團之特性,不必浪費複雜手續,將集團中數字一一枚舉,而乃能示此集團相當精確之真相.

1.6 **樣本之種類與大小** 樣本因其所選取之方法及其代表集團之情形不同而有所各異如1.4節所述,樣本之由隨機方法選得者曰隨機樣本.所謂隨機,乃言集團之每一分子皆有被選取構成該樣本之機會.與隨機樣本相對者為有偏見的樣本(Biased Sample).偏見樣本不能用以代表集團.由偏見樣本所測得之估計常軌數與集團之常軌數有明顯之差別.

整個集團可因其某種性質之不同而可分為若干群.例如水稻集團,可因其秈粳,糯性質而分為三亞群.每亞群又可因其品種之不同再分為若干小組.此等集團之樣本乃為各小組樣本之總合.樣本之由各小組樣本總合者,或言樣本可分為若干小群,而每小群可視為整個集團中相當小組的樣本.此等樣本,稱曰層疊樣本(Stratified Sample).

若樣本之選取非由于隨機,或僅部分由于隨機,而係依據某種已知性質之比率分配者,此種樣本稱曰控制樣本.例如吾人研究某種性狀時,知此性狀與性畜之性別有關,則選取樣本時當顧及該種性畜之性別比率.由此所選得之樣本即屬于控制樣本.(Controlled Sample) 之列.與控制樣本性質相同者尚有所謂相等樣本(Equated or Matched Sample),乃言由同一集團選取二個或二個以上相同或相等之控制樣本.在進行試驗時為求試準確計,常採用相等樣本.當于以後述之.

樣本依其所包含個數之多寡,尚有大小之別.大小樣本之界限為何,迄今尚無明白之規定.學者意見紛歧,或云樣本具個數不滿100方為小樣本,或云不及30者方為小樣本,或云所謂小樣本乃指樣本含個數小于30者而言.惟大概以30為大小樣本之界限,較為普遍.

昔日認為統計方法僅能應用于大樣本,今日已知其謬誤,其實在小樣本更需用統計方法解釋.惟分析方法稍異.關于小樣本之試驗方法,乃為統計學權威Fisher氏之最大貢獻,容于以後詳述之.

1.7 **數量資料與變數** 生物統計所考慮者大都為數量(Quantity),例如研究人體之高度,則人體高度之樣本即為一群數字組成.此種由數字構成之樣本,可稱曰數量資料(Quantitative Data),或簡稱資料(Data).組成資料之每一數字,稱曰變數(Variable or Variate*).變數因其性質不同,可分為連續性變數(Continuous Variable)與非連續性變數(Discontinuous Variable)二大類.連續性變數乃為在變異範圍內,可取得任何個數之變數.例如人體之高度是,在此種變數之任何二變數間(

*Variable 與 Variate 無顯明之區別,或以前者為通稱,指一般高度,體重,或其他種類之變數.例如一群高度變數(Variable).後者為專稱,特別指明種類及單位,如某人高160cm.,此160cm.,可稱曰Variate.本書將此二名詞混用,未加區別.

如35寸與36寸間），可再取得無窮個數之變數，蓋35寸與36寸之間，尚有無窮個分厘毫絲之差也。非連續性變數則為在某變異範圍內僅能取得一定個數之變數，例如福建永安大豆每荚簇（Raceme）之具荚數為1—8朵，吾人選取各荚簇而記其荚數，祇有1.2、3.4、5.6、7.8八種變數，大小一定，而不能另有1.2，2.5，3.7等數值之變數。

1.8 **錯誤與機誤**　錯誤與機誤二名詞，大相逕庭，讀者需切實瞭解錯誤乃由吾人疏忽所致之結果，如抄錯數目字，看錯磅碼數目，及其他在試驗過程中之各種作業過失，如田面積量錯，播種時忘記播下一行等是，優良之試驗工作者絕不應有此種錯誤。任何一試驗設發生錯誤，少則損失其精確度，甚則含有偏見（Bias），勢必前功盡棄，毫無價值，故進行試驗時必需謹慎從事，寧可多費人力物力，按步做去，設或不幸而發生錯誤，即需設法補救，或犧牲一切，重新做起。

機誤（Error）或稱誤差，大別為定機誤（Constant Error）與不定機誤（Accidental Error）二種。定機誤乃由吾人無意或因偏見所發生之差誤，例如以甚高或甚低之人，視察掛于一定高度之溫度計，其觀察結果因觀察角度之關係，必較原來為大或小，而生人為之偏見，嚴格言之，此種機誤亦應屬于錯誤之列，而應設法避免之。

統計及生物統計所生之機誤，係指不定機誤而言，所謂不定機誤，乃言錯誤之發生，由于機會所致，或知其原因，或不知其原因，但皆無法避免之，僅惟能使之盡量減少而已。例如在同一集團中隨機選取一樣本，其由第一次選得者與第二次選得者多少有若干不同，此種不同，僅由于選取時機會所致，即為機誤。又如隨機栽植同品種于二小區，二小區之產量有所不同，此種不同，除由于小區之土壤差異外，尚有其他不易控制之因子，如病蟲害程度並不一律，試驗區面積不能毫厘不差等是，此等差誤，乃由于機會，非人力所能控制使之完全避免，吾人僅惟能設法使之降低至最低限度而已。

1.9 **無效假說**　由上節所述，已知由同一集團選取二個樣本，因機誤關係，二樣本間可有若干差異。今有二事物或二樣本，已知其彼此間有若干差異，惟此種差異竟由于機誤，或由于二樣本本質之不同，無法辨別，欲辨別此二樣本之差異為何，乃需藉統計分析之助。在行此種情形之統計分析前，常先假設此二樣本所來自集團之真差為零，（意即來自同一集團）此種真差為零之假設，即為Fisher氏所創有名之無效假說（Null Hypothesis），由此假說，乃可進行統計分析，以判斷此假說之可靠性（Reliability）。若經統計分析結果，認為此假說能成立，即二樣本所來自集團之真差確為零，則來此二集團的樣本之差亦應為零，今此二樣本所以有若干差異者，由于機誤之故，非由于真的差異也。若無效假說不能成立，即示二樣本之差異顯著，非由于機誤，至差異之原因為何，尚不能知，而需另加研究者。

無效假說非僅限于真差為零，真差為零僅為無效假說之一種，例如測量二種變

數互相關係時，A種變數增大時，B種變數增大抑或減少抑或不生關係，在進行此種統計分析時，常先假設二種變數的真相關為零，（即彼此無關係）此種真相關為零之假設，亦為無效假說總之舉凡常軌數為零之假設，皆稱曰無效假說，如何用統計方法判斷無效假說之能否成立，詳述于以後有關各章。

1.10 統計之主要功用 1.4 及1.5已曾述及集團之特性可由樣本估計之，研究如何得到適當之樣本乃為統計工作者主要課題之一。既得適當之樣本，乃需進而研究如何用適當之統計方法以分析之，以獲得合理之結論，統計之主要功用亦即在此。重複言之，統計之功用在：

a. 使吾人獲得足能代表集團之樣本。

b. 使吾人能由樣本得合理可靠之結論。

主要參考文獻

(1) Yule: Theory of Statistics

(2) Fisher: Statistical Method for Research Works.

(3) Treloar: Outline of Biometrical Analysis. Chapter I.

(4) Lindquist: Statistical Analysis in Educational Research. Chapter I.

問題及習題

(1)有限性集團，無限性集團，實有的集團與假設的集團之意義各若何？

(2)集團之特性，為何常由該集團之樣本研究之，由樣本研究所得之特性可用以代表該集團之特性，其理由安在？

(3)常軌數與估計常軌數之區別何在？

(4)何謂機誤，試詳述之。

(5)何謂無效假說，試舉例詳述之。

第二章

次數分配

2.1 資料之整理 研究問題之初步手續為調查或觀察有關該問題樣本之個體情形,乃進而整理分析,冀能由此資料獲得該問題之基本概念,再解釋問題之本身.設樣本個體數甚少,則其資料自無須整理,即可分析而能明其意義.但當樣本含個體數甚多之時,則由此所得之資料,若不加以適當之整理與分類,僅是一堆數字而已,其所表示之性質為何,無法明瞭.例如協和大學農藝試驗場212個邵武水稻品種的產量記錄(表2.1),在未加整理前,吾人觀之,味同嚼臘,蓋吾人一時無法明瞭此項記錄之意義也.若能于短時間內測知此212個邵武水稻品種中之最高產量與最低

表2.1 邵武協和大學農藝試驗場212個水稻品種之產量(市斤)(民國30-31年)

127.5	329	401	322.5	272.5	363	298	353.5
271.5	222.5	341	359	335	291	241	313.5
319	343.5	241.5	387.5	297	312.5	244	266.5
336.5	305	265	251.5	165.5	265.5	275.5	339
365	327	371.5	335	255.5	371.5	299	232.5
189	329.5	352	268	359.5	310.5	224.5	292.5
291.5	241	368.5	241	277.5	333	226	266.5
335	321.5	276	344	167	248	264.5	375
229.6	285.5	243	355	257	338.5	276	311.5
366	266.5	285.5	293.5	346.5	331	371.5	273.5
203.5	329.5	292.5	268.5	296.5	355.5	226	287
312	260.5	330	210	235	266	285.5	375.5
353.5	303.5	362.5	332.5	277.5	297.5	296	291
277	264.5	304.5	273.5	237.5	358.5	365	273.5
368	332.5	447	269.5	257.5	344	262	326.5
101.5	329.5	310	214.5	324	286	305	375.5
332.5	280	283.5	312.5	338.5	277.5	336	270.5
345.5	283.5	324	275	252	341.5	328	292.5
296.5	244	403	327.5	273	322.5	281.5	349.5
368.5	380	290	193	305.5	307	324	377
107.5	300.5	303.5	291.5	328	278	346.5	253.5
354.5	320	345.5	294.5	270	320	329	378
324	262.5	384	337.5	291.5	303.5	302.5	255
337	244.5	270	161	286	325	345.5	
279	285	279.5	263	297	243	326	
302.5	346.5	281.5	335.5	261.5	295	365	
321.5	356.5	307.5	372	342.5	334	287	

產量各為若干,大多數品種之產量為若干而無誤乎?由此可見整理資料之重要矣.在統計學上,有統計處理一詞,即指用數學之原理與方法,以整理原始資料而言.

2.2 整列 統計處理之第一步,即將原始數字由小而大,按其大小次序排列成一依次表(Array),此種排列方法名曰整列(Array). 例如表2.1 中水稻品種之最低產量為101.5,最高為447,可依其大小排列成一依次表(表2.2),由表2.2 吾人可不加細察即能看出邵武212個水稻品種之最低產量為101.5市斤,最高產量為447市斤,及品種最低與最高產量之差為345.5 市斤,此345.5 數字稱謂全距(Range),即資料內最大數與最小數差異之謂也.

依次表之作法甚易,即將每一數字各寫于每一小卡片上,然後根據數字之大小依次重疊卡片,再將卡片上數字依次錄入表內即成.

表2.2 邵武協和大學農藝試驗場212個水稻品種產量之依次表 (市斤)(民國30-31年)

101.5	244.5	270.5	286	303.5	326	338.5	365
107.5	248	271.5	286	303.5	326.5	339	365
127.5	251.5	272.5	287	304.5	327	341	365
161	252	273	287	305	327.5	341.5	366
165.5	253.5	273.5	290	305	328	342.5	368
167.	255	273.5	291	305.5	328	343.5	368.5
189	255.5	273.5	291	307	329	344	368.5
193	257	275	291.5	307.5	329	344	371.5
203.5	257.5	275.5	291.5	310	329.5	345.5	371.5
210	260.5	276	291.5	310.5	329.5	345.5	371.5
214.5	261.5	276	292.5	311.5	329.5	345.5	372
222.5	262	277	292.5	312	330	346.5	375
224.5	262.5	277.5	293.5	312.5	331	346.5	375.5
226	263	277.5	294.5	312.5	332.5	346.5	375.5
226	264.5	277.5	295	313.5	332.5	349.5	377
229.6	264.5	278	296	319	332.5	352	378
232.5	265	279	296.5	320	333	353.5	380
235	265.5	279.5	296.5	320	334	353.5	384
237.5	266	280	297	321.5	335	354.5	387.5
241	266.5	281.5	297	321.5	335	355	392.5
241	266.5	281.5	297.5	322.5	335	355.5	401
241	266.5	283.5	298	322.5	335.5	356.5	403
241.5	268	283.5	299	324	336	358.5	447
243	268.5	285	300.5	324	336.5	359	
243	269.5	285.5	302.5	324	337	359.5	
244	270	285.5	302.5	324	337.5	362.5	
244	270	285.5	303.5	325	338.5	363	

2.3 次數分配 由依次表觀察資料性質較之于未整理時已易于瞭解,但其所能給予吾人對問題基本概念之認識尚嫌過淺,乃需作更進一步之處理,而有次數分配(Frequency Distribution)方法之應用.所謂次數分配,即以數字之大小而

分類將某一定限度內之各數字歸為一組,將另一限度內之數字歸為另一組,資料之意義乃能明白顯露,使吾人藉此對事物之本質可作精確有用之計算,比較與結論.由此所作成之表,曰次數分配表(Frequency Distribution Table).

2.4 組與組距 作次數分配表時,必先將資料之全距分成若干組(Class).某資料應分成若干組,需視其組距(Class Interval)之大小而定.所謂組距者,乃指相鄰二組間之距離.各組各有其確定之上下限.如表2:3 A,0為第一組之下限(Lower Limit),10為第二組之下限,20為第三組之下限.第一組之上限(Upper Limit),應為相近10之數,即9.99,第二組之上限應為相近20之19.99.是各組上下限之中點,或該組之下限與其次組下限之中點,稱為該組之中價(Mid-point).中價可用作代表該組內所有各數之平均數,由于吾人假設組內各數值均密集于各該組之中價,及任何組距內各數值均勻分配于各該組距內是也.

2.5 組距之寫法 組為構成次數分配表之主要部份,其界限需力求明顯,俾讀者觀之,無需解釋即能一目瞭然.組距之寫法甚多,茲摘述一二佳者如表2.3,以供參考.

表2.3 組距之詳細寫法

A			B		
組距	中價	次數	組距	中價	次數
0—9.99	5	0	0		
10—19.99	15	0	10	5	0
20—29.99	25	4	20	15	0
30—39.99	35	8	30	25	4
40—49.99	45	18	40	35	8
50—59.99	55	60	50	45	18
60—69.99	65	17	60	55	60
70—79.99	75	9	70	65	17
80—89.99	85	6	80	75	9
90—99.99	95	2	90	85	6
			100	95	2

表2.3 A之寫法,為目前各統計教科書所採用者,以其各組界限明顯,一目瞭然故也.例如組距之為30-39.99者,其意為在此組中各數其大小為自30至39.99,決不到40,一至40即屬下一組.惟寫法較為麻煩,是為其缺點.

表2.3 B之寫法,為吾國統計學者艾偉氏所提倡,氏倡此種寫法之主要論點(3)為"一個數目在一組距內代表其終點,決不能在另一組內代表其起點或終點;一個組距有一個數目,其數目為其起點,至另一數目則為另一組距之起點".因此吾人觀察表2.3 B即可知0為第一組起點,10為第二組起點,20為第三組起點,……組與組間之組距為10,第一組之中價為(10+0)/2=5.第二組之中價為(20+10)/2=15.資料中

凡自0起而不達到10(小于10,即10除外.)之各數悉屬于第一組內,自10(包含10)而不達20(小于20,即20除外.)各數均屬于第二組,餘類推.此種直綫組距寫法,簡明優美.有採用之價值.

2.6 組距之選擇 組之決定,或組距之選擇,因變數(Varable)種類之不同而有異.所謂變數者,乃資料中各種單位數字之謂也.在質量變數(Qualitative Varable)中,可將不同質之各級分別為各組,如以人類之髮色為例,紅,淡紅,淡棕,暗棕,深黑等,各作為單獨之組.

非連續性變數之組距,常以其自然單位為組距之大小,如以大豆每莢荅之荅數為例,可以一朵荅為其組距,即其各組依次為一朵荅,二朵荅,三朵荅,四朵荅……等.設此種非連續性變數具較大之變異範圍,如播穗之具穀實粒數,少者僅有數十粒,多者可達數百粒,則不能以自然單位每粒作為組距,而需合併數個單位,至應合併若干單位,即組距之大小應為幾粒,則可參考以下所述連續性變數之組距決定法而斟酌變通之.

變數為連續性時,組距大小之選擇,取決于

1. 資料數字變異之全距,
2. 資料所含之個數.(總次數)
3. 所得之組數是否便于計算.
4. 由次數分配表計算估計常軌數是否精確.

組數越多,則由次數表計算所得之估計常軌數越形精確,但計算更麻煩.反之若組數減少,計算固較簡易,但所得之估計常軌數較不準確矣.為欲避免此二項之矛盾,乃有組距不超過標準差(Standard Deviation)四分之一之適用通則.惟在製表前常不知標準差之數值,幸Tippelt氏曾發表全距變異與標準差關係之詳表,藉變異之全距以估計標準差,此項關係經Snedecor氏彙成簡表.表2.4係錄自Goulden氏修整Snedecor氏表完全為二位數者(1).

表2.4 資料或樣本(大小自20至1000)之全距與標準差之比率 (Goulden氏)

樣本內所含變數個數	全距/標準差	樣本內所含變數個數	全距/標準差
20	3.7	200	5.5
30	4.1	300	5.8
50	4.5	400	5.9
75	4.8	500	6.1
100	5.0	700	6.3
150	5.3	1000	6.5

現設有含300個變數之資料,其變異之全距為365−155=210,由上表查得,

該含300個變數資料之 $\frac{\text{全距}}{\text{標準差}}=5.8$

則標準差 $=\frac{\text{全距}}{5.8}$

為欲使組距約等于標準差之四分之一,故

$$組距=\frac{1}{4}S.D.=\frac{1}{4}\times\frac{全距}{5.8}=\frac{1}{4}\times\frac{210}{5.8}=\frac{210}{4\times5.8}=9.05$$

由上算式算得之組距應為9.05,但為計算便利計,此.05之小數可畧去,無甚大碍.又組距為偶數時,較奇數為便利,蓋組距如為偶數,則其中價數值之數字位數毋須較原有組距數值之位數多加一位也,應用時亦可畧加變通.

至第一組之下限應為何數,應小于資料中最小一數字若干?需酌視資料情形而加變通.通常皆以不增加計算麻煩,及次數分配二端各組內之中價能代表其所屬組內之各變數之均數為原則.

用上法所成之次數表以計算任何估計常軌數,可有足夠之精確.但如次數表祇用以圖示時(詳2.9),則由此方法所得之組距常嫌過小,組中有僅含極少之次數,甚至完全不含,在此種場合,可增大組距為標準差之三分之一乃至二分之一.

2.7 次數分配表之製作 其方法可歸納為下列各項:

1. 決定組距.

2. 將最小一組寫在頂端,依次寫好組距與各組中價.

3. 由原始資料或依次表過入各組數值範圍內所發現變數之次數.

每發現一個,用"/"符號表示之,至四個"/"後,在"////"中加一橫劃成"𝍸",即為記數五個之符號.或同依次表作法,先將每數分別寫在每一張小卡片上,屬于同組內之各卡片彙集一起,其張數即為該組之次數,而可直接填入表內各組.後法較前法為優,因能校對故也.

茲用表2.2 邵武協和大學農藝試驗場212個水稻品種產量資料作例,說明次數分配表之作法.

× × × × ×

【例2.1】 邵武協和大學農藝試驗場212個水稻品種產量次數分配表之作法,如次頁表2.5所示.本資料共包含邵武水稻品種212個,已知全距為345.5,由表2.4查得含200個變數樣本之全距與標準差之比率為(全距/標準差)=5.5.(因212與200接近)故其組距應為標準差四分之一,=345.5/5.5×4=15.7,謀計算方便計,定組距為16,乃排列組距如表2.5第一直行,與各組中價如第二直行.小于116(116除外)至100各數(包含100)悉歸諸第一組.同樣,凡小于132(132除外)至116各數(包含116)悉歸諸第二組.每組發現在其範圍內之一數字時,即于第三直行該組之同列處作"/"符號,至四個"/"號後,在"////"中加橫劃成"𝍸",即為計數五次之符號.餘類推.至第一組下限之所以定為100,乃由于排列組距及以後計算較為方便故也.