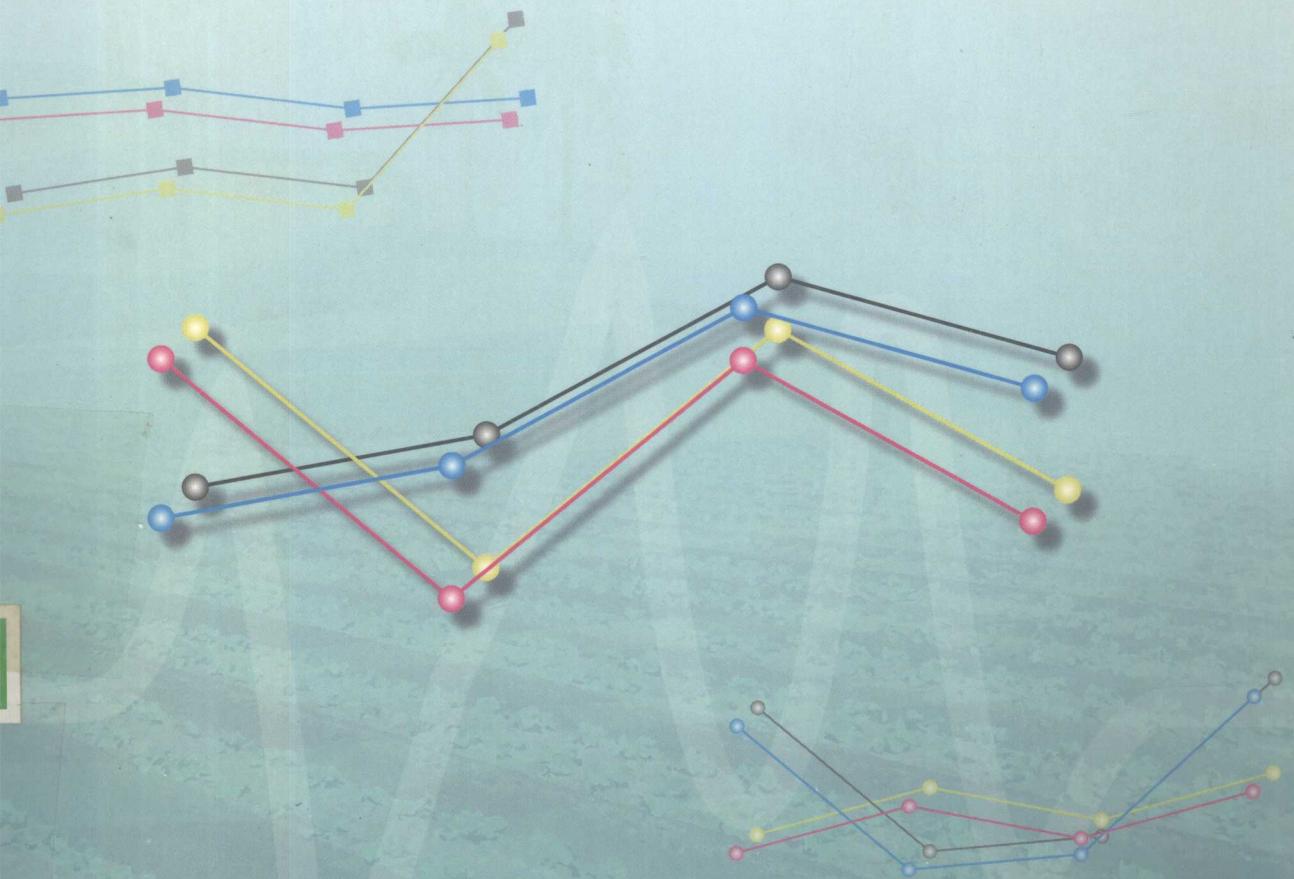


生物統計學

附 題解及練習

李英周 審校

杜榮騫 編著



藝軒圖書出版社

生物統計學

附

題解及練習

李英周 審校

杜榮騫 編著

藝軒圖書出版社

生物統計學附題解及練習／杜榮騫編著. --第一版.
--臺北縣新店市：藝軒，2006[民 95]
面： 公分。
參考書目：面
含索引

ISBN 957-616-866-X (平裝)

1. 生物學-統計-問題集 2. 數理生物學-問題集
3. SAS(電腦程式)

360.13022

95003308

本書經高等教育出版社授權藝軒圖書出版社出版發行繁體字版

◎本書任何部份之文字或圖片，如未獲得本社書面同意，
不得以任何方式抄襲、節錄及翻印

新聞局出版事業登記證局版台業字第1687號

生物統計學附題解及練習

審校者：李英周
編著者：杜榮騫
發行所：藝軒圖書出版社
發行人：彭賽蓮
總公司：台北縣新店市寶高路7巷1號5樓
電話：(02)2918-2288
傳真：(02)2917-2266
網址：www.yihsient.com.tw
E-mail:yihsient@ms17.hinet.net
總經銷：藝軒圖書文具有限公司
台北市羅斯福路三段316巷3號
(台大校門對面·捷運新店線公館站)
電話：(02)2367-6824
傳真：(02)2365-0346
郵政劃撥：01062928
台大醫學院展售處
台北市仁愛路台大醫學院聯教館醫工室B1
電話：(02)2397-5070
台中門市
台中市北區五常街178號
(健行路445號宏總加州大樓)
電話：(04)2206-8119
傳真：(04)2206-8120
大夫書局
高雄市三民區十全一路107號
(高雄醫學大學正對面)
電話：(07)311-8228
本公司常年法律顧問／魏千峰、邱錦添律師

二〇〇六年四月第一版

ISBN 957-616-866-X

本書如有缺頁、破損或裝訂錯誤，請寄回本公司更換。

讀者訂購諮詢專線：(02) 2367-0122

審校序

本書第一部份從數據的收集與整理開始，透過數據類型與分布介紹生物統計學的基本觀念、樣本特徵值與機率分布，進而介紹從族群到樣本的研究中，幾種常見的機率分布函數及抽樣分布函數。第二部份介紹從樣本如何去推論族群，透過對族群提出一個假設，再藉由樣本資料去推論這個假設是否可以被接受，以及通過樣本統計值去推估族群參數。第三部份介紹變方分析，迴歸分析與相關分析。本書在第 12 章也新增實驗設計單元，審校者在該章節前撰寫本章加註語，希望能使讀者更容易了解實驗設計學的精神與運用。第四部份是對第一版中“SAS 軟體基本操作”和相關的 SAS 程序，做了調整和補充，連同每一章的習題詳解及各章大量復習題附於書末。本書的全書設計是為供讀者理解並建立生物統計學的基礎概念，進而誘發對生命科學的學習興趣。

李英周

國立台灣大學漁業科學研究所副教授

第二版前言

本書自 1999 年出版以來，深受廣大讀者歡迎，被多所高校採用作為教材。讀者的喜愛便是對編者的激勵，為了能更好地為生物統計學教學服務。經多方徵求讀者意見後，在本書第二版的內容編排上做了比較大的調整。

第二版增加了第十二章“實驗設計”。統計分析與實驗設計是密不可分的，只知道統計分析方法，而不知道如何設計符合統計學要求的實驗，這樣的知識是不全面的。為了提高學生獨立分析問題、獨立設計實驗和獨立處理實驗結果的能力，我們認為增加實驗設計是十分必要的。

對第一版中“SAS 軟體基本操作”和相關的 SAS 程序，做了調整和補充，連同每一章的習題詳解及各章的大量復習題附在書末供讀者理解和鞏固所學知識以及學習如何用 SAS 軟體處理數據。

生物統計學的內容很廣泛，根據對本科生的要求和學時數的安排並徵求了多方意見，確定了本書所選內容。講完全書大約需要 50~60 學時。如果學時安排較少，可以適當減少兩因子和多因子變方分析以及複迴歸內容。我們建議，儘量保持生物統計學基本原理和統計假設檢驗內容的完整性，在此基礎上，學生通過自學便能很快掌握更多的統計學知識。

希望廣大讀者在使用本書過程中，對所發現的問題及不足之處能不吝賜教，並希望提出進一步的修改意見，使本書在生物統計學教學中發揮更大作用。對此，編者將致以深切謝意。編者的電子信箱地址為：
our gene@eyou.com。

杜榮騫

前　　言

生物統計學是現代生物學研究不可缺少的工具。不論是傳統學科還是現代分子生物學，時時刻刻都在與數字打交道。為了揭示生物體內在規律或生物與環境之間的關係，都離不開因素分析，特別是多元分析。生物統計學不僅在傳統生物學、醫學和農學中被廣泛應用，而且在新興的分子生物學研究中也發揮著重要作用。例如，繪製連鎖圖，特別是繪製人類基因連鎖圖時，製圖函數的獲得，Lod Score 的計算以及 DNA 序列同源性分析等都是建立在統計學基礎上的。沒有良好的統計學基礎，這些工作只能知其然，而不能知其所以然，對於工作的深入開展是很不利的。因此，生物統計學已經成為每一位生物科學工作者必備的基礎。

這本教材是在 1985 年版本的基礎上，廣泛徵求各方面意見重新編寫的。為配合生物學的迅速發展，在內容和編排上做了適當調整，刪除了一些不常用的內容，增加了一些必要的基礎，如變方分析中均方期望的推演等。近十幾年來電腦在我國的迅速普及，出現了大量的統計軟件。許多過去望而卻步的繁重計算工作，現在已變得輕而易舉。利用統計軟體代替繁重的手工計算，是生物統計學發展的必然趨勢。SAS 系統是國際上公認的統計軟體，它的包容量大，伸縮性強，在全球範圍內被各行各業廣泛採用，因此，本書編進了介紹 SAS 軟體應用的章節，以滿足讀者的需要。書內的例題和習題除一部分是編者自己的工作外，很多是從書後所列參考資料中引用的，在這裏對原著者深表謝意。為了使例題更具代表性，對其中有些數據做了適當調整，因此，書中例題和習題中的數據只供學習和鞏固統計學知識使用，沒有真正的科學意義，請廣大讀者切勿引用。

本書在編寫過程中得到了各方面大力支持，四川大學劉天倫先生在內容編排上提出過寶貴建議，本校數學系沈世鎰先生，計算機系塗奉生先生曾鼎力相助，生命科學學院王穎老師在資料整理和謄寫上做了大量工作，在這裏對以上各位先生表示真摯謝意。

在這裏需要特別提出的是，美國 SAS 軟體研究所上海辦事處為本書的編寫提供了 SAS 軟體和多方支援，為促成本書起了很大作用。編者在



生物統計學
附題解及練習

viii

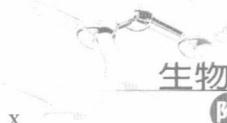
這裏對上海辦事處的關心和支持表示衷心感謝。

編者在編寫時雖已盡心竭力，但錯誤及不當之處仍在所難免，敬希讀者不吝指出，本人將不勝感謝。

編 者

目 次

第一章 統計數據的收集與整理	1
『1.1 族群與樣本	4
『1.2 數據類型及頻數（率）分布	4
『1.3 樣本的幾個特徵值	12
『習 題	27
第二章 機率和機率分布	31
『2.1 機率的基本概念	31
『2.2 機率分佈	39
『2.3 族群特徵值	44
『習 題	47
第三章 幾種常見的機率分布律	53
『3.1 二項分佈	53
『3.2 卜瓦松分佈	61
『3.3 另外幾種離散型概率分布	64
『3.4 常態分布	66
『3.5 另外幾種連續型機率分布	73
『3.6 中央極限定理	76
『習 題	79
第四章 抽樣分佈	83
『4.1 從一個常態族群中抽取的樣本統計量的分布	83
『4.2 從兩個常態族群中抽取的樣本統計量的分布	89
『習 題	93
第五章 統計推論	95
『5.1 單個樣本的統計假設檢驗	95
『5.2 兩個樣本的差異顯著性檢驗	110
『習 題	126
第六章 參數估計	129
『6.1 點估計	129
『6.2 區間估計	131
『習 題	141



第七章 適合度檢驗	143
『7.1 適合度檢驗的一般原理	143
『7.2 適合度檢驗	145
『7.3 獨立性檢驗	150
『7.4 χ^2 的可加性	156
『習 題	158
第八章 單因子變方分析	161
『8.1 變方分析的基本原理	161
『8.2 固定效應模型	165
『8.3 隨機效應模型	171
『8.4 多重比較	174
『8.5 變方分析應具備的條件	178
『習 題	180
第九章 兩因子及多因子變方分析	185
『9.1 兩因子變方分析中的一些基本概念	185
『9.2 固定模型	189
『9.3 隨機模型	200
『9.4 混合模型	205
『9.5 兩個以上因子的變方分析	208
『9.6 缺失數據的估計	213
『9.7 轉換	216
『習 題	217
第十章 簡單迴歸及簡單相關分析	223
『10.1 迴歸與相關的基本概念	223
『10.2 簡單迴歸方程	225
『10.3 簡單迴歸的檢驗	231
『10.4 簡單非線性迴歸	250
『10.5 相關	265
『習 題	275
第十一章 複迴歸及複相關分析	279
『11.1 複迴歸方程	279
『11.2 複相關分析	305

◀ 11.3 逐步迴歸分析	311
◀ 習題	319
第十二章 實驗設計	321
◀ 12.1 實驗設計的基本原則	325
◀ 12.2 實驗計畫書的編制	328
◀ 12.3 簡單實驗設計	335
◀ 12.4 單因子實驗設計	341
◀ 12.5 兩因子實驗設計	359
◀ 12.6 正交設計	372
◀ 習題	380

題解及練習

第一章 SAS 軟體基本操作	357
◀ 1.1 SAS 系統的啓動	357
◀ 1.2 顯示管理系統	357
◀ 1.3 SAS 程式的編寫	365
◀ 1.4 數據步	366
◀ 1.5 過程步	369
◀ 1.6 執行程式	370
第二章 幾種常用的 SAS 程式	373
◀ 2.1 利用 SAS 軟體描述樣本資料	373
◀ 2.2 統計假設檢驗的 SAS 程式	382
◀ 2.3 χ^2 檢驗的 SAS 程式	384
◀ 2.4 單因子變方分析的 SAS 程式	388
◀ 2.5 常用實驗設計變方分析的 SAS 程式	392
◀ 2.6 相關與迴歸分析的 SAS 程式	409
◀ 2.7 複迴歸分析的 SAS 程式	417
第三章 習題答案	421
◀ 3.1 第一章習題答案	421
◀ 3.2 第二章習題答案	427
◀ 3.3 第三章習題答案	434
◀ 3.4 第五章習題答案	444



生物統計學

附題解及練習

xii

『3.5 第六章習題答案	449
『3.6 第七章習題答案	451
『3.7 第八章習題答案	455
『3.8 第九章習題答案	463
『3.9 第十章習題答案	479
『3.10 第十一章習題答案	501
『3.11 第十二章習題答案	503
第四章 復習鞏固題	513
『4.1 第一章復習鞏固題	513
答 案	515
『4.2 第二章復習鞏固題	518
答 案	522
『4.3 第三章復習鞏固題	523
答 案	526
『4.4 第四章復習鞏固題	528
答 案	529
『4.5 第五章復習鞏固題	530
答 案	536
『4.6 第六章復習鞏固題	541
答 案	541
『4.7 第七章復習鞏固題	542
答 案	545
『4.8 第八章復習鞏固題	550
答 案	553
『4.9 第九章復習鞏固題	554
答 案	556
『4.10 第十章復習鞏固題	558
答 案	559
附表	561
參考書目	593
索引	595

第一章 統計資料的收集與整理

1.1 族群與樣品

1.1.1 統計資料的不齊性

人類在生活、生產和科學研究中經常與資料打交道。在對特定的研究對象進行測量、記錄並分析所得資料之後，你會發現，即使從同一類對象中所得到的資料也不完全相同，有大有小、參差不齊。或者說，產生這些資料的個體間存在著廣泛變異。

造成生物體變異的原因有很多，概括起來可以分為遺傳因子、環境因子及發育噪音（development noise）。遺傳因子的影響是顯而易見的。就拿身高來說，子女身高直接受父母身高的影響，通常是父母高，子女也高；父母矮，子女也矮。環境因子表現在很多方面。仍以身高為例，包括：食量、蛋白質攝入量、營養成分平衡、維生素和微量元素的獲得量、鍛煉、勞動強度、睡眠時間、不良嗜好、修養、心理承受力等。我們會發現，即使在遺傳與環境因子都得到控制的情況下，個體間仍然存在變異。例如，小麥純系是經過多代自交得到的，遺傳上已經純化，個體間遺傳成分可以認為是均一的。將自交系的單株後代種植在生長條件都相同的環境中，例如，種植在人工氣候室中，使用電腦控制肥力、水分、光照、溫度、通風等，即使這樣，個體間仍存在變異。它們的株高、穗長、穗重、幹物重等還會有一定的波動。這種波動的產生是由發育噪音引起的。或者說是由於在個體發育過程中的某些隨機因子造成的。如果把影響生物變異的各種遺傳因子、形形色色的環境因子以及種種隨機因子自由組合起來，其組合數將是一個天文數字。不同個體的組合方式不同，由此造成了生物個體之間的廣泛變異。由此可見，變異性是自然界存在的客觀規律。

由於個體間的變異，給我們處理資料帶來很多困難。例如，考察我國 18 歲男青年身高，若個體間沒有變異，我們隨便測量一個人就可以了。然而，由於個體間存在著變異，為了測得 18 歲男青年身高，從理論上講，應當把全國所有 18 歲男青年身高都測量一遍，用其平均值來代表身高數值。把所有 18 歲男青年身高都測量一遍是很難做到的。退一步講，雖然很難做到，但只要投入足夠的人力和財力，還是可以測量出這些資料的。如果要測量所有新生兒體重，則無論如何也拿不到全部資料。因為新生兒不斷出生，要想收集到所有新生兒體重，就要不斷測量，只要有新生兒出生，測量就不能停止。由此可見，測量全部對象既不現實也不可能。我們只能從全部研究對象中抽出一部分個體來，通過對這一部分個體的研究來推論全體的情況。這就出現了我們下面將要提出的兩個概念：族群與樣品。

1.1.2 族群與樣品

統計學研究的核心問題是如何通過樣品推測族群。因此，族群與樣品是生物統計學中的兩個最基本概念。

族群 (population) 是我們研究的全部對象。族群又分為無限族群 (infinite population) 和有限族群 (finite population)。例如，我們要研究在某種條件下生長的小麥的株高，因為無法估計出在這種條件下生長的小麥的數量，可以設想這一族群是無限的。或者研究新生兒體重，因為新生兒是無止境的，所以這一族群也可以設想是無限的。如果我們要調查一所學校今年新生的身高，這一族群則是有限的。生物統計學中所遇到的族群多數都是無限族群。構成族群的每個成員稱為個體 (individual)。

樣品 (sample) 是族群的一部分，樣品內包含的個體數目稱為樣品大小 (sample size)。

1.1.3 抽 樣

從族群中獲得樣品的過程稱為抽樣 (sampling)。抽樣的目的，是希望通過對樣品的研究，推測其族群。例如，希望由 100 株“三尺三”高粱的株高，推測在這種條件下生長的該品種的株高。這就要求樣品應能在最大程度上代表族群的情況。為此，在從族群中抽取樣品時，族群中的每一個個體被抽中的機會必須都一樣，不能帶有偏見。例如，在小麥育種工作中，我們常常希望得到矮稈品種。為了滿足個人願望，在抽樣時便多抽矮

稈的，這樣得到的樣品沒有代表性，屬於偏性抽樣，不能代表族群的情況。我們需要的樣品應該是一個族群的縮影。為了達到這個目的，就需要用隨機抽樣 (random sampling) 的方法獲得樣品。

隨機抽樣的方法很多，例如抽籤、拈鬮等。最好的方法是使用亂數字表（見附表1）進行抽樣。現舉例說明怎樣用亂數字表進行抽樣。假設需要從包含4,728個個體的族群中，抽出一個含量為20的樣品。因為個體族群4,728是一個四位數，所以族群中每一個個體的編號都應是四位數，即從0001號到4,728號。第一步，閉上眼睛用鉛筆在亂數字表上任意點上一點，假若點到奇數上，就用第一頁表；點到偶數上，就用第二頁表。第二步，在選定的那一頁上，再點一次，決定從哪個字開始。決定了起點以後，開始以四位元數位為一節連續讀下去，不用考慮數位間的間隙。可以正讀、倒讀、橫向讀、縱向讀，也可以沿對角線方向讀。選出小於等於4,728的數位，大於4,728的則捨棄，直到取滿20個數為止。這20個數所對應的個體，即為我們選中的樣品。

從一有限族群中抽樣，可分為歸還取樣 (sampling with replacement) 和不歸還取樣 (sampling without replacement)。所謂歸還取樣是指：從族群中抽出一個個體，記下它的特徵後，放回族群中，再做第二次抽樣。這種抽樣方式可能會重複抽中某一個體。不歸還取樣是指：從族群中抽出個體後，不再放回。在上述的例子中，若保留重複的隨機數位，則為歸還取樣；若捨棄重複的數位，則為不歸還取樣。對於無限族群來說，歸還取樣和不歸還取樣，實際上沒有區別。

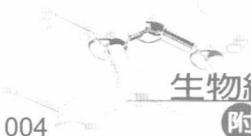
樣品的含量越大越有代表性。但是，太大的樣品研究起來是很困難的。因此，樣品的大小必須合適。

1.2 資料類型及頻數（率）分佈

1.2.1 連續型資料和離散型資料

統計學的最基本工作是收集資料。把原始資料收集上來之後，首先要對資料進行整理並分析這些資料的特性和變化規律。生物統計學中經常遇到的資料有兩種類型，一種是連續型資料，另一種是離散型資料。

與某種標準做比較所得到的資料稱為連續型資料 (continuous data)，



又稱為度量資料 (measurement data)。例如，長度、時間、重量、OD 值、血壓值等。這類資料通常是非整數。雖然有時記載的是整數，如身高的釐米數，但是當提高精確度後，總會出現小數。對連續型資料進行分析的方法，通常稱為變數的方法 (method of variable)。

由記錄不同類別個體的數目所得到的資料，稱為離散型資料 (discrete data)，又稱為計數資料 (count data)。例如，某一類別動物的頭數，具有某一特徵的種子粒數，血液中不同類型的細胞數目等。所有這些資料全都是整數，而且不能再細分，也不能進一步提高它們的精確度。對離散型資料進行分析的方法，通常稱為屬性的方法 (method of attribute)。

在判斷資料的類型之後，就要進一步研究資料的變化規律。描述資料變化規律的最簡單方法是將這些資料列成頻度表 (frequency table) 或繪成頻數圖 (frequency graph)，根據頻數分佈進行研究。

1.2.2 頻數(率)表和頻數(率)圖的編繪

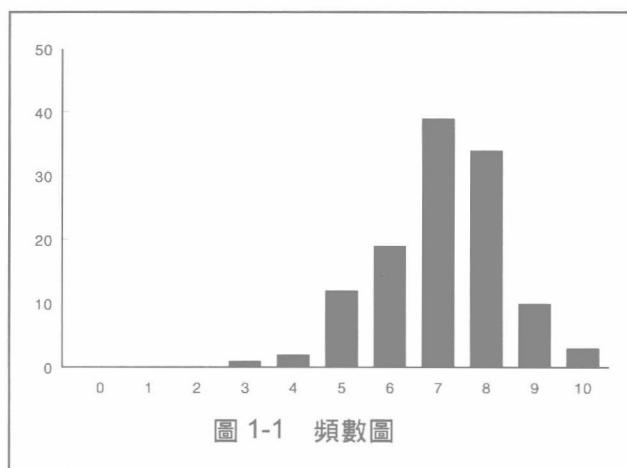
離散型資料及連續型資料的頻度表和頻數圖的編繪方法略有不同，下面各舉一例說明。先看離散型資料頻數(率)表和頻數(率)圖的編繪方法。

例 1.1 調查每天出生的 10 名新生兒中，體重超過 3 kg 的人數，共調查 120 d。每天的 10 名新生兒中，體重超過 3 kg 的人數，可能有 11 種情況：1 名也沒有，有 1 名，有 2 名，…，10 名都是，如表 1-1 的第一列所示。這一列稱為組值 (class value)。表 1-1 的第 2 列所記載的是調查結果。

如第 1 天調查的結果，有 6 名超過 3 kg 的，則在組值為 6 的一行做

►► 表 1-1 每 10 名新生兒中體重超過 3 kg 的人數的頻數(率)表

組值 (體重超過 3 kg 的人數)	頻數計算	頻 數	頻 率
0		0	0.000
1		0	0.000
2		0	0.000
3	—	1	0.008
4	丁	2	0.017
5	正正丁	12	0.100
6	正正正正	19	0.158
7	正正正正正正正	39	0.325
8	正正正正正正正	34	0.283
9	正正	10	0.083
10	下	3	0.025
總 計		120	0.999



個記號，一般使用“正”字或“++”號表示。全部調查完畢，累加各行結果，填入頻數一欄。或者將各行的結果除以總數而得出頻率。把頻數或頻率按超過 3 kg 的人數的順序排列起來，便得到了頻數分佈 (frequency distribution) 或頻率分布 (percentage distribution)。頻度表可以比較清楚地描述出資料變化規律。為了更直觀地描述資料變化規律，還可以繪成頻數圖表示（圖 1-1）。圖 1-1 的橫軸表示每 10 名新生兒中，體重超過 3 kg 的人數，縱軸表示每一組的頻數。若將縱軸改為頻率的話，則得到頻率圖。頻率圖與頻數圖的圖形完全一樣。

下面這個例子介紹了連續型資料頻數（率）表和頻數（率）圖的編繪方法。

例 1.2 表 1-2 列出了某農場在做“三尺三”高粱提純時所調查的 100 個資料。表中所列出的資料雖然是連續型的，但看上去好象是離散型資料。產生這種誤解的原因，是由於株高的單位是“cm”，在這種精確度下出現了許多高度相同的植株，當進一步提高精確度後，便很難再找到兩個高度相同的植株了。從表 1-2 的原始資料中，除可以找出最大值是 170 cm，最小值是 141 cm 以及估計出它們的平均高度大約在 150~160 cm 之外，很難再看出什麼規律來。但是，當我們將表 1-2 中的資料列成頻度表之後，便可以比較清楚地看出這些資料的變化規律。高粱的株高是連續型資料，不是一個個孤立的值。因此，不能像例 1.1 那樣製表。連續型資料頻度表的製作過程如下：首先將資料分組，一般來說，100 個資料可以分成 8~10 組。根據範圍 $R = \max x - \min x = 170 - 140 = 30$ ，分為 10 組比

►►► 表 1-2 “三尺三” 株高測量結果

155	153	159	155	150	159	157	159	151	152
159	158	153	153	144	156	150	157	160	150
150	150	160	156	160	155	160	151	157	155
159	161	156	141	156	145	156	153	158	161
157	149	153	153	155	162	154	152	162	155
161	159	161	156	162	151	152	154	157	162
158	155	153	151	157	156	153	147	158	155
148	163	156	163	154	158	152	163	158	154
164	155	156	158	164	148	164	154	157	165
158	166	154	154	157	167	157	159	170	158

較合適，每一組的間距剛好是 3 cm。用比較簡單的組間距分組，編制頻度表比較方便。將分好的組填入表 1-3 的第 1 列。表 1-3 的第 1 列稱為組限（class limit），組限是根據原始記錄中的數值確定的。本實驗是以 cm 為單位統計數值的，所以第一組的上限“143” cm 的實際值，可能在大於等於 142.5 cm，小於 143.5 cm 範圍內。同樣，第一組的下限“141” cm 的實際值，可能在大於等於 140.5 cm，小於 141.5 cm 範圍內。因此，這一組的全部實際可能值是在 140.5~143.5 cm 範圍內。140.5~143.5 這個範圍稱為組界（class boundary）。對於其他各組，同樣可以定出相應的組界。

中間值（midvalue）是每一組的兩個組限的平均值，但是也有例外。例如，習慣上通常以“歲”為計算年齡的單位。假若有一組的組限是 20~29 歲，上限 29 歲包括這個人可能剛剛 29 歲，也可能即將進入 30 歲。所以這一組既包括 20 歲的，也包括 29 歲的，共有 10 個年齡級。因此，中間值應是 $(20 + 30) \div 2 = 25$ ，而不是 $(20 + 29) \div 2 = 24.5$ 。類似

►►► 表 1-3 “三尺三” 株高頻數（率）表

組限/cm	組界/cm	中 值	頻數計算	頻 數	頻 率
141~143	140.5~143.5	142	—	1	0.01
144~146	143.5~146.5	145	丁	2	0.02
147~149	146.5~149.5	148	正	4	0.04
150~152	149.5~152.5	151	正正	13	0.13
153~155	152.5~155.5	154	正正正正	23	0.23
156~158	155.5~158.5	157	正正正正正	28	0.28
159~161	158.5~161.5	160	正正正	15	0.15
162~164	161.5~164.5	163	正正	10	0.10
165~167	164.5~167.5	166	下	3	0.03
168~170	167.5~170.5	169	—	1	0.01
總 計				100	1.00