中 等 气 象 学 校 講 义

气　候　学

（气候資料整理方法）

气 象 专 业 用

北 京 气 象 专 科 学 校 編

1965年10月

目　　录

第一章 統計基本原理

一、概率基础知識

一、事件

在自然界中，一切事物之間都存在着普遍的联系和相互制約的关系，我們常常遇到这样的事情或現象，它們在一定的条件下，或者发生，或者不发生。

如果某一組条件实現时，某种現象一定发生，这种現象称为**必然事件**。例如，水在一个大气压和温度 100°C 时一定沸騰；当一团空气絕热上升时，其自身溫度一定下降；当溫度升高，风速加大，相对湿度較低时，水面蒸发必定增大，等等。所有这些現象都是必然事件。

如果某一組条件实現时，某种現象不会发生，则这种現象称为**不可能事件**。例如，冰在一个大气压，溫度低于 0°C 的条件下是不可能融化的；当有大量降水时，地面必然不会干燥；满天布满低云时，必然不是晴天，等等。所有这些現象都是不可能事件。

如果某一組条件实現后，某种現象可能发生，也可能不发生，则这种現象称**偶然事件**，或随机事件。例如，有 10 个灯泡，其中合格品 8 个，废品二个，如果在其中随便挑一个，可能挑到的有两种情況，即挑到合格品，或废品，因此这两件事都是偶然事件。又如，积雨云移近测站，可能有降水，也可能沒有降水，这些也是偶然事件。

上面所述事件发生的必然性和偶然性，是客观事物在相互联系和发展过程中，同时具备着的互相对立而又互相联系的两个方面。必然性是由事物的根本矛盾以及和它相联系的基本重要条件所决定的，有着它自身发展的規律。偶然性则是在事物发展过程中，受到許多个别条件的影响，从而使这种发展过程发生的不依人們主观意志为轉移的一种偏差。例如，下面所列是北京 1965 年 3 月份的最高气溫

日	1	2	3	4	5	6	7	8	9	10	11
最高溫度	7.0	10.4	4.9	4.2	4.1	7.7	11.4	12.4	6.7	8.8	19.2
日	12	13	14	15	16	17	18	19	20	21	22
最高溫度	15.0	14.2	14.4	11.1	13.6	14.4	16.7	13.4	14.0	15.9	16.2
日	23	24	25	26	27	28	29	30	31		
最高溫度	20.1	13.5	9.4	11.2	11.4	16.7	19.8	21.7	14.8		

从上述資料可以看出，上旬最高溫度的最低值为 4.1°C，最高值为 12.4°C；中旬最低值为 11.°C，最高值 19.2°C；下旬最低值 9.4°C，最高值为 21.7°C，全月溫度总的趋

势是升高的。这是北京溫度的必然性的表現，这种必然性是由于太阳高度逐漸增高，昼长时間不断增长等基本原因而形成的。但是就每天而論，还是有高有低的，溫度在逐日上升过程中又复降低或升高減緩的原因，可能是由于前一日阴天，該日下雨，或者是冷空气不时爆发等偶然性条件影响的結果。又如，北京春季多风沙，这是由于在春季太阳輻射日增，气溫不断升高，降水較少和环流基本情况等主要原因而形成的必然結果，但是每一年的春季所出現的风沙次数并不一定，这是由于上述各項因素不同的作用和不同的相互配合，以及一些其他因素影响而产生的一种偶然現象。因此，气象要素如溫度、湿度、降水、风向、风速和一些气象現象，它們所表現的数值和发生的次数等，在量的方面都是偶然事件，同样某一时期的气候情况的出現也是錯綜复杂的大气过程中的一种偶然事件。

事件发生的必然性和偶然性，决定于事件发生之前的条件，**如果离开条件只談某一事件会不会发生，是没有意义的。**例如，在大量的产品中有意地找出一件合格品，这时“有意地找出”是条件，如果产品中确有合格品，那么有意地找出一件合格品这一事件是必然会发生的。但若“随便抽出一件产品”，則所抽的一件可能是合格品，也可能是次品，这时“对产品随便抽取”是条件，如果把“抽出一件合格品”作为事件，則这一事件在条件实現后，可能发生，也可能不发生，具有偶然性，是一偶然事件。同时，偶然事件也是在一定条件具备情况下的必然結果，只是我們尙未弄清其錯綜复杂的原因（条件）而已。

在客观世界中，大量的事件差不多都具有偶然性（也即随机性）。**偶然事件发生的可能性并不一样，有的发生的可能性大些，有的小些。**例如，北京某年 6 月的降水量大于 66.9 毫米（历年平均）这一事件是偶然事件，它发生的可能性就比北京連續两年 6 月降水量大于 66.9 毫米这一偶然事件发生的可能性要大些。

2．概率

每一偶然事件的发生既然都有它或大或小的可能性，那末，要在数量上来比較事件的可能性程度，必須对每一事件給予一个数字，使可能性越大的事件有越大的数値。将表示事件发生可能性大小的这个数，称为**事件的概率**，（**或称机率**）。

概率是客观存在的，只要客观条件确定，一件事物发生可能性的大小就已确定，不管人們知道或不知道，能計算或不能計算，都不能影响这一事物发生可能性的大小，也就是不能影响事件的概率。

对于概率的估計或計算只能是在事件发生之前，而不能在发生之后，当一件事情已經发生了，即当可能性已經变成了現实性，这时就不存在概率的問題了。例如，当一台发电机已經发生故障，这时来考察它发生故障的可能性大小，也就是，考察发生故障的概率，对于此时此刻的这台发电机就沒有意义了。

事件 A 发生的概率記作 $P(A)$，要比較各个事件发生的可能性程度，必需采取一个測度单位，一般取必然事件的概率作为这样的单位。一个必然事件发生的概率是 1，如果 A 是必然事件，那末

$$P(A)=1$$

一个不可能事件发生的概率是 0，如果 B 是不可能事件，那末

$$P(B)=0$$

其他一切偶然事件的概率都是在 0 与 1 之間的非負数。因此如果 C 是偶然事件，那末

$$0<P(C)<1$$

3. 概率的估算

根据不同的情况对偶然事件的概率进行估算，有以下两种方法：

利用事件发生机会的均等性进行估算 例如有 50 个零件，其中有 3 个不合格，現在要求任取一个是不合格零件这一事件的可能性的大小。这里 50 个零件都允許被抽取，也都有被取出的可能，現在只能取出一件，这时任何一件被取出的可能性都不能被証明比其他另一件会大些，因此，任何一个零件被取出的机会是均等的，或者是等可能的。

在上面求任取一个不合格零件的可能性大小的例子里，有机会可能被抽取到的总数是 50，称之为**可能机会数**。而我們希望抽取到的是不合格的零件，共中只有 3 件不合格，因此，必須取出这 3 件中的一件，才符合期望，也就是符合期望的数是 3，叫做**期望机会数**。我們把期望机会数与可能机会数的比$\frac{3}{50}$，用来表示任取一件遇到不合格产品的可能性的大小，这就是所要求的事件发生的概率。

一般說来，如果一定条件实現以后，所发生結果（包括所期望的和不期望的）的均等可能机会数为 n，而其中希望事件 A 发生的期望机会数为 m，則事件 A 发生的概率是$\frac{m}{n}$，即

$$P(A)=\frac{m}{n}(0\leqslant m\leqslant n)$$

例 1　設一个袋中有大小一致的 20 个白球和 30 个黑球，而且这 50 个球摸起來使人的感觉完全相同。問随手摸到一个白球的概率是多少？

解：設摸到一个白球的事件为 A，按照题意，50 个球任一球被摸到的机会是等同的，可能机会数是 50，而期望机会数是 20，因此，事件 A 发生的概率为

$$P(A)=\frac{m}{n}=\frac{20}{50}=0.4$$

利用統計方法进行概率估算 有些事件，特別是气象現象，其发生的机会并不是均等的，而其未来发生的可能机会数和期望机会数根本找不到。对于求这样事件的概率，

就必須根据大量統計的数据进行估算。

在讲利用統計数据对概率进行估算以前，先弄清楚頻率与概率的关系，現举一例說明。我們投擲一枚硬币，硬币落地正面（国徽一面）和背面（文字一面）向上的机会是等可能的，假如要求投擲硬币一次正面向上的概率，按照上面所述对概率估算的第一种方法，可能机会数为2，期望机会数（硬币正面向上）为1，因此投擲硬币一次正面向上（事件 A）的概率为

$$P(A)=\frac{1}{2}.$$

但是如果进行实际投擲，只擲一次，却只能得到一种結果，或是正面向上，或是背面向上。擲100次，也不一定正好是50次正面向上，50次背面向上。然而这并不等于硬币正面向上的可能性不是$\frac{1}{2}$，或概率不是$\frac{1}{2}$。以往有人作过实驗，发現擲的次数愈多，正面向上的頻率便愈接近$\frac{1}{2}$（見表1.1）

表 1.1　投擲硬币正面向上的頻率

投擲次数	擲出正面次数	頻　率
4040	2048	0.5080
12000	6019	0.5016
24000	12012	0.5005

我們在实际观測或实驗中，某現象在一定区間內出現的次数，称为**頻数**，而頻数与观測或实驗的总次数（总頻数）的比，称**相对頻数**，或**頻率**。

由上面所述投擲硬币的实驗可以看出，正面向上的頻率数值虽然不一定恰好等于概率$\left(\frac{1}{2}\right)$，但它是围繞在概率周围变动的，而且試驗次数愈多，頻率与概率差异很小这一事件发生的可能性愈大。不但投擲硬币的情况如此，其他任何偶然事件在大量試驗过程中，当試驗次数很大时，它的頻率总具有一定的**稳定性**，經常接近于一个**常数**，而且在这个常数的左右摆动，这个**常数**就是这个事件发生的**概率**。按照頻率和概率的这一关系，我們就有可能根据統計出来的頻率的数値来近似地推定概率的数値。

但应特別提出概率和頻率的关系虽然密切，而二者却是有着严格的区別的，概率是一个事件出現的“可能性”的度量，而頻率是在这个事件已經出現之后，在全体事件当中实际所占的比重，概率是存在于頻率之先的，我們只是利用頻率来推定現在还不能直接計算的概率而已。

在气象資料整理的实际工作中，概率往往无法知道，如果所統計的資料年代較长，則頻率比較稳定，因此**常把頻率值当做該現象出現的气候概率值**。

概率的計算問題，更多的情况是根据某些事件的已知概率来計算另一些較为复杂事件的概率。在概率計算中經常遇到的复杂事件是“事件和”。

事件和　表示在很多事件 $A_1, A_2, \cdots\cdots, A_n$ 中至少发生其中之一，就算該事件发生的

事件。例如，灾害性天气本身是一事件，但这一事件是由冰雹、霜冻、大风、暴雨、低溫、……等不同天气构成的，只要这些事件中的任一种天气发生，就等于灾害性天气发生了，当然若有两种或两种以上，譬如大风和暴雨，同时发生则也算灾害性天气发生了。灾害性天气就是冰雹、霜冻、大风、暴雨、低溫、……等的事件和。又如，雨包括大雨、中雨和小雨，只要出現大雨或中雨或小雨都算有雨发生，因此，雨是大雨、中雨和小雨的事件和。

事件和一般用$A_1+A_2+\cdots+A_n=\sum\limits_{i=1}^{n}A_i$表示，讀作“事件$A_1$或$A_2\cdots$或$A_n$”，可看作是一个具有$n$个事件性质中的任一个或一个以上的性质的事件。

互斥事件　在前面所举的雨是大雨、中雨和小雨的事件和，但大雨、中雨和小雨这三种天气現象并不能在同一地点的同一时間内同时发生。因此，在一定条件实現下，如果两件或两件以上事件絕对不能同时发生，则这些事件称为**互斥事件**。而冰雹、霜冻、大风、暴雨、低溫等有的是互斥事件，例如，霜冻和暴雨不能同时出現，所以它倆是互斥的，但大风和暴雨有时可能同时出現，因此大风和暴雨二者就不是互斥事件了，而称为**相容事件**。

如果A_1和A_2是互斥事件，则A_1或A_2出現的概率为A_1和A_2分别出現的概率的和，即

$$P(A_1+A_2)=P(A_1)+P(A_2) \tag{1.1}$$

推广到一般情形，则为

$$P(A_1+A_2+\cdots+A_n)=P(A_1)+P(A_2)+\cdots+P(A_n) \tag{1.2}$$

例2　有10匹布，其中一級品8匹，二級品及废品各一匹，如果把一級品和二級品算作合格品，間从中随便挑一匹而抽到的是合格品的概率？

解：設抽到一級品的事件为A_1，抽到二級品的事件为A_2，而抽到合格品的事件为A_1+A_2，且A_1和A_2为互斥事件，因此

$$P(A_1+A_2)=P(A_1)+P(A_2)=\frac{8}{10}+\frac{1}{10}=\frac{9}{10}=0.9.$$

例3　某地13年中，全年平均有雾45天，有沙暴0.7天，有雷暴47天，問全年里有雾或沙暴或雷暴出現的概率是多少？

解：設出現雾为事件A_1，出現沙暴为事件A_2，出現雷暴为事件A_3，按照題意：A_1，A_2，A_3为互斥事件，且$P(A_1)=\frac{45}{365}$，$P(A_2)=\frac{0.7}{365}$，$P(A_3)=\frac{47}{365}$，因此，

$$\begin{aligned}P(A_1+A_2+A_3)&=P(A_1)+P(A_2)+P(A_3)\\&=\frac{45}{365}+\frac{0.7}{365}+\frac{47}{365}\\&=0.254\end{aligned}$$

在概率計算中經常遇到的另一复杂事件是"事件积"。

事件积 表示由事件 A 与 B 同时实現所組成的事件。例如，"旣刮风又下雨"，刮风是一事件，下雨是另一事件，而"旣刮风又下雨"表示刮风这一事件和下雨这一事件同时发生，因此，"旣刮风又下雨"是刮风和下雨这两事件的**事件积**。

事件积，一般用 AB 表示，讀作"事件 A 兼事件 B"，可看作是一个兼具两个或两个以上事件的性质的事件。

条件概率 一事件 B 在另一事件 A 发生的条件下而发生的概率，称为**事件 B 在条件 A 下的条件概率**，記作 $P(B/A)$。例如，假定袋中有三个白球二个黑球，从袋中继續摸出二球，第一次摸出的球不放回去。那末"第一次摸出的是白球"和"第二次摸出的是黑球"都是偶然事件，把前者叫做事件 A，后者叫做事件 B。不难看出 B 的概率是和 A 的发生与否密切相关的。如果 A 发生了，也就是，如果第一次摸出的是白球，这时袋中只剩下二个白球和二个黑球，因此 B（第二次摸出黑球）在 A 实现条件下的条件概率 $P(B/A)=\frac{2}{4}=\frac{1}{2}$。但是，如果 A 不发生（A 不发生記作 $\bar{A}$），就是說，如果第一次摸出的是黑球，則袋中剩下三个白球一个黑球，因此，B 在 A 不发生($\bar{A}$)条件下的条件概率，$P(B/\bar{A})=\frac{1}{4}$。又如，北京某年 7 月共观测风 124 次，其中有偏南风 100 次，而在这124 次观测中有 50 次下雨，其中在刮偏南风时下雨的有 40 次。設刮偏南风为事件 A，下雨为事件B,，則在偏南风条件下下雨的条件概率 $P(B/A)=\frac{40}{100}=\frac{2}{5}$。

两事件 A、B 同时出現的概率等于一个事件 A 出現的概率乘以另一事件 B 在第一事件发生的条件下的条件概率，即

$$P(AB)=P(A)\cdot P(B/A) \tag{1.3}$$

为了說明这一关系式的成立，仍然用上面摸球的例子。第一次摸出一个球（不管是白球或者是黑球）有五种均等的可能机会，第二次摸出一个球有四种均等的可能机会，因此連續摸两次，共有 5×4 种均等的可能机会。其中属于"第一次摸出的是白球，而第二次摸出的是黑球"这一事件的期望机会共有 3×2 种。这是因为第一次摸球时有三个白球都可能摸到，而第二次摸球时，由于第一次摸去的是白球，黑球沒有被摸去，仍有二个黑球可以摸到。所以，事件 A（第一次抽出的是白球）和事件 B（第二次抽出的是黑球）都发生（也可看作是同时发生）的概率

$$P(AB)=\frac{3\times2}{5\times4}=\frac{3}{5}\times\frac{2}{5}$$

上式右边的$\frac{3}{5}$是事件 A 发生的概率 $P(A)$，而$\frac{2}{5}$則是在 A 发生的条件下 B 的条件概率（見前面条件概率中摸球的例子），即 $P(B/A)$。因此

$$P(AB)=P(A)P(B/A)$$

独立事件 如果在若干事件中，某一事件的出現幷不影响其他事件的出現，則这些

事件叫做**不相关事件**，或**独立事件**。两件独立事件可以同时出现，同时不出现，或这一件出现而那一件不出現。在条件概率中，若事件B的出现与条件A无关即A、B都是独立事件，则

$$P(B/A)=P(B) \tag{1.4}$$

式（1.4）是事件A与B相互独立的条件。

两件或两件以上独立事件同时出现的概率等于各独立事件单独出現的概率的积。即由（1.3）式

$$P(AB)=P(A)P(B/A)$$

因为A与B互为独立事件，则

$$P(B/A)=P(B)$$

所以

$$P(AB)=P(A)P(B) \tag{1.5}$$

推广到一般情形，如果有$A_1, A_2, \cdots\cdots$，A_n个互为独立的事件，则

$$P(A_1A_2A_3\cdots\cdots A_n)=P(A_1)P(A_2)\cdots\cdots P(A_n) \tag{1.6}$$

例 4　两射击手彼此独立地射击同一目标，設甲射中的概率为$P(A)=0.9$，乙射中的概率为$P(B)=0.8$，求目标同时被射中的概率。

解：因为甲、乙独立地射击，所以事件A与B互为独立事件，而二人同时射中的事件AB的概率为：

$$P(AB)=P(A)P(B)=0.9\times 0.8=0.72$$

例 5　某站每年7月降水量超过300毫米的概率为$\frac{1}{3}$，求該站連續三年7月降水量超过300毫米的概率。

解：設第一年7月降水量超过300毫米为事件A_1，第二年为事件A_2，第三年为事件A_3，假設每年7月的降水量互无影响，彼此是独立的，所以A_1，A_2，A_3是独立事件，而連續三年7月降水量超过300毫米这一事件，即$A_1A_2A_3$的概率为

$$P(A_1A_2A_3)=P(A_1)P(A_2)P(A_3)$$

$$=\frac{1}{3}\cdot\frac{1}{3}\cdot\frac{1}{3}=\frac{1}{27}.$$

4. 随机变量

在概率論中除"事件"以外，还有一个常用的基本概念是**随机变量**。

在一次試驗之后所得的数量結果可以取得各种数值，但究竟能取哪一个数值，事先并不知道，不过它所取的数值（或数值区間）都对应有一定的概率，这样的数量称为**随机变量**。例如，观测某一测站日、月或年的降水量；每日观测的最高、最低气溫等等，都是随机变量。如果試驗結果不是具体数值，而是某种属性，例如擲一次硬币出現阳面

和背面；天气有雨和无雨；冷和暖；干和湿……等等，可以把某种属性发生記为1，不发生記为0，这样就可把試驗的属性結果也可以用数值来表示，而每次試驗該种属性究竟是发生(1)，还是不发生(0)，事先并不知道，因此是随机变量。

如果一个随机变量所能取得的数值可以全部一个个的例举出来，而且每个数值都对应有确定的概率，则这样的随机变量称为**离散型的随机变量**。例如，某种天气現象6月分出現的日数，可以是0，1，2……，30；每年在我国登陆的台风个数，可以是0，1，2……；某地一年中干旱月数，可以是0，1，2，……，12。这些都是离散型随机变量。

如果随机变量可以取一个数值区間內任意实数值时，这个随机变量称为**連續型随机变量**。例如，某地某月的降水量可以是某一数值区間（譬如30—50毫米）的任意数，因此是連續型随机变量。連續型随机变量的可能值可以是无穷多的，因此分配到任何一个特定值的概率都是零。

不同性质的随机变量各有其不同的理論分布規律和表征它分布特点的数字特征值。概率論就是利用如事件及其概率，随机变量和它的分布規律、数字特征值的运算，提供了用理論方法从一些事件的概率来确定另一些事件的概率，或者从一些随机变量的分布律和数字特征，来确定另一些随机变量的分布律和数字特征等的可能性。这种間接方法可以大大节約在試驗方面的时間和經費。但是对于屬于偶然現象性质的研究仍然必須来自实践，来自試驗所得的数据，来自大量观察的結果。关于观測大量偶然現象所获得的許多数据的收集、整理和分析等等方法，是数理統計学的主要內容。下面讲的就是有关整理、分析大量观測数据的一些方法。

二、頻数分配

1. 总体和样本

总体　在我們用数理統計方法研究大量同类現象的数量特征和規律性时，所研究大量同类現象的全体，称为总体。总体是所有組成总体的各个个体的集合。例如，北京年降水量資料的总体，必須包括亘古以来直到今后无限长岁月中所有的年降水量，而每年的年降水量則是組成总体的个体。又如，中国年降水量資料的总体，就是指全中国每一个地方长年年平均降水量所組成的集合，而每一个地方的平均年降水量就是組成总体的个体。

总体和个体并不是一成不变的，要看每一次的研究任务而定。如果任务是要研究北京1965年降水的分配情况，則1965年各月的月降水量就构成了一个总体，但要研究北京长年的降水情况，而1965年年降水量就只是北京无限长岁月中年降水量总体中的一个个体了。

总体的性质由其中各个体的性质而定，所以要了解总体的性质，就必須測定各个个

体的性质。也就是，为了表征总体的数量特征，須将总体内所有的个体資料加以测定綜合，得出一些描述总体特征和規律性的特征数字，即**总体的統計特征数**。

抽样 不过，有时所研究的現象总体中的个体数目很多，甚至是无限多，事实上不可能把总体中全部个体进行全面观察。或者即使能够全面观察，但因人力、物力和时間的浪費太大而无条件进行。例如，水文和气候現象的总体是无限长的，不可能进行全面观察。又如，我們要知道一片森林树木的平均高度，虽然是可以进行全面观察，但是在人力、物力和时間上很不經济。

另外，也有不少情形，总体中的个体数目并不很多，但是为了取得所观察对象的某些資料，就必須将观察对象破坏，在这种情况下，总体中个体数目虽然不多，仍不能对全部个体一一測定。例如，为了鑑定一批灯泡的质量而观察灯泡的耐用时数，就需将灯泡直到烧坏为止；为了鑑定棉紗质量而观察棉紗力，就需将棉紗进行試驗直至拉断，等等。在上述情况下，自然不能对所研究的对象进行全面观察。

由于以上种种原因，我們只能从总体中随机地抽取一部分个体，对这些个体进行观察，并根据所获得的观察資料推测全部研究对象的数量特征和規律性。这些被抽取出来的个体的集合体，称为**样本**。其中所包括的个体单位数，称**样本单位数**。例如，从一批灯泡的总体中抽出 50 个灯泡，即組成一个样本，其样本单位数即为 50。

抽样問題在气候資料統計中具有非常重要的意义。因为气象現象的总体是无限的，要得到气象数列的总体，必須从地球上有气象現象的那一天开始进行气象观测，一直到地球存在的最后一天为止，这显然是不可能的。而現在所有的气象資料，可以說仅仅是无限总体中的一个小小的样本，因而气象数列总体是无法获得的。但是，概率論和数理統計給出了，根据对样本单位所作的观察或实驗而得到的每个个体的具体数值，来推断或估計总体单位的一些分布規律和特征数字。

因此，总体与样本是两个有着严格区别的概念，总体只能通过样本去认識，而不易直接获得，但不能把样本当作总体。为了满足实际需要，我們可以根据样本的观测資料，应用数理統計原理和方法对总体作充分估計。

既然总体必須通过样本去认識，而且我們經常遇到和处理的是属于样本性质的气候資料，那么就必須了解样本具有一些什么数量特征，需要对样本資料进行計算，得出表示样本分布特征的統計特征数。下面就着重地討論有关这方面的問題。

2. 頻数分配的列表法和图示法

通常由实驗、观察或其他方法所收集到的一批資料中的各个数据，是非常分散的，好象沒有什么系統和次序，必須經过整理和归納，才能表現出这一批数据所遵循的規律。在整理和归納过程中，我們总是希望了解所研究的对象的某些特征，例如要了解某一地方气象要素：經常出現在什么范围内；这种要素的变化范围和差异程度如何，集中

程度又如何；这种要素在各个范围内的分布情况怎样等等。

为了达到上述目的，整理和归納的方法，常是将所收集到的数据用列表和制图方法来处理。有了統計表或图，就能大体上看出这批数据所代表的事物的性质。然后根据表或图，运用数学方法，可以計算出一些簡单的特征数，作为这批事物的性质的代表，因此，列表与制图是数理統計工作中很重要的步驟。

1）頻数分配表

将观测或收集到的資料数据按大小分成若干組，使全部記录各自分配到相应的組內，并找出各小組內包括的数据个数(**頻数**)，列成表格。这种表示頻数分布情况的表格，称为**頻数分配表**。

下面以上海年降水量作例，說明頻数分配表的制作方法和步驟。

表 1.2 上海年降水量(毫米)**(自 1873 年至 1950 年)**

年(个位) 千百十位	0	1	2	3	4	5	6	7	8	9
187				974.8	1006.3	1588.1	769.8	1008.9	1206.8	1271.4
188	1101.9	1340.2	1331.0	1085.4	1184.4	1113.4	1204.9	1170.7	975.4	1462.3
189	747.1	1416.0	709.2	1147.5	935.0	1016.3	1031.6	1105.7	849.9	1233.4
190	1008.6	1063.8	1004.9	1086.2	1022.5	1330.9	1439.4	1236.5	1088.1	1288.7
191	1015.8	1217.5	1320.7	1078.1	1203.4	1480.0	1269.9	1049.2	1318.4	1191.8
192	1015.5	1507.1	1159.6	1021.3	986.1	794.7	1131.5	1170.6	1161.7	791.2
193	1143.8	1602.0	951.4	1002.5	859.4	870.6	912.0	1025.2	1264.2	1196.5
194	1140.7	1659.3	942.7	1123.3	910.2	1405.0	1208.7	1305.5	1167.2	1225.9
195	1402.5									

首先，从上表資料中找出最大值 $M=1659.3$ 毫米（1941 年），最小值 $m=709.2$ 毫米(1892年)，可知上海78年間的年降水量資料是在709.2毫米与 1659.3 毫米之間变动的。

其次，把最大值和最小值之間分成若干个小組，第一組要包括最小值，最后一組要包括最大值。在分組时，应先确定**組数**。組分得愈少，計算愈方便，但組分得很少时，每一組的組距就比較大，当一組之內变值相差較大时，若用其中的一个平均数为代表，来計算各种特征值，必将引起較大的計算誤差，所以分組不宜过少。反之，分組愈多时，計算誤差愈小，但計算将愈繁杂。因此将观測数据划分为多少个組，并无一成不变的規則。一方面要看有多少个观測数据，数据多要多分几組，数据少可少分几組；另一方面还要看計算要求的精确度有多高。按經驗，所分組数不宜超过 $5\lg N$ （N 为数据总項数），但也不能少于 $2.5\lg N$。例如，根据表 1.2 資料 $N=78$，$5\lg 78=9.4605$，$2.5\lg 78=4.73025$，也就是将上述資料分为 5—9 組較为适宜。

組数决定以后，就要决定**組距**和**組限**。組限是小組的界限，左边低値項称**低限**（或**下限**），右边高値項称**高限**（或**上限**）（見表 1.3）。組距是上限与下限之間的差値，表示小組內数値的活动范围。用上限与下限的平均数作为小組的代表数，称**組中値**，并用 x_i 表示。

然后計算所获資料在各組范围內出現的次数，也就是**頻数**，以 f_i 表之。当計算資料在各范围內出現的次数时，一般若資料数値恰好与該組的下限値相等，算作是在該組內出現的一次，但資料数値恰好等于上限，則算作是出現在下一組內的数。例如，在表 1.3 中如年雨量恰好为 600 毫米，此数就算作是出現在第一組內的数，而若年雨量为 800 毫米时，則应算是出現在第二組內的数了。

表 1.3 上海年降水量（毫米）頻数分配表（1873—1950）

組序	組限		組中値	頻数	累积頻数 F		頻率（%）	累积頻率 F'	
	下限	上限	x_i	f_i	自小至大	自大至小	f'	自小至大	自大至小
1	600	800	700	5	5	78	6.4	6.4	100.0
2	800	1000	900	11	16	73	14.1	20.5	93.6
3	1000	1200	1100	34	50	62	43.5	64.0	79.5
4	1200	1400	1300	18	68	28	23.1	87.1	36.0
5	1400	1600	1500	8	76	10	10.3	97.4	12.9
6	1600	1800	1700	2	78	2	2.6	100.0	2.6

将頻数自低値向高値（或自高値向低値）方向累加起来，就得到**累积頻数**。用 F 表示任一組的自低値向高値方向累加的累积頻数，表示小于該組上限的所有資料出現的总次数，反之，自高値向低値方向累加的累积頻数，表示大于該組下限的所有資料出現的总次数。例如，表 1.3 中第三組自低値至高値的累积頻数为 50，表明上海 78 年間不足 1200 毫米的年降水量，共出現过 50 次。而該組自高値至低値的累积頻数为 62，表明 78 年內年降水量超过 1000 毫米的共出現过 62 次。其他依此类推。

各組的頻数占总頻数（总項数）的百分比称为該組的**頻率（相对頻数）**，用 f' 表示。頻率表示該組資料在全部資料中所占的**比重**。把頻率由低値向高値（或由高値向低値）方向累加，可以得出**累积頻率**（F'），其意义与累积頻数一样，只是累积頻数是一絕对数，而累积頻率是一相对数，表示低于或高于某一数値的資料出現次数的总和占全部資料的成数。

按照上述步驟制出如表 1.3 所示的頻数分配表，可分析得出如下結果：

(1) 上海年雨量变化在 709.2—1659.3 毫米之間。

(2) 上海最常出現的年雨量在 1000—1400 毫米之間，这种年雨量出現的机会約占

总数的66.6%。

頻数分配表不但能說明观測数列的分布規律，而且也是計算观測数列各种特征值的重要計算工具。

2）頻数(率)分配图

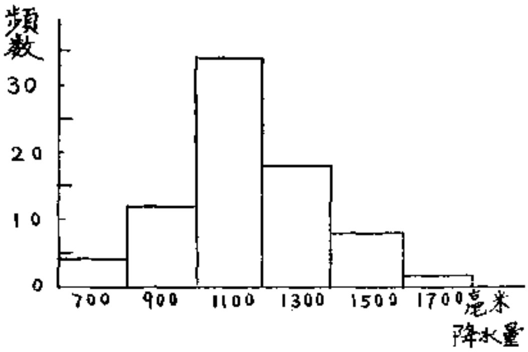

图 1.1 上海年降水量頻数直方图

与頻数（率）分配表相对应的图，称**頻数（率）分配图**。图的横坐标表示組限(或組中值)，纵坐标表示頻数（或頻率），图形可用直方块、折綫或曲綫表示，通常多用曲綫。

图形如果以直方块表示，則每小块高表示其所对应的頻数（或頻率）(图1.1)。

如果图形以折綫表示，則应将各組的頻数（或頻率）与相对应的組中值，作为纵、横坐标值，分别点在座标图上，然后以折綫連接各点，即得**頻数(或頻率)多边形**图（1.2)。若用曲綫表示，将上述各点以曲綫連接，即会得出**頻数(或頻率)曲綫**（图1.3)。

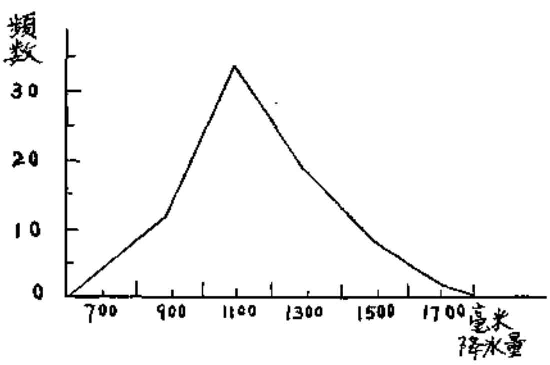

图 1.2 上海年降水量頻数多边形

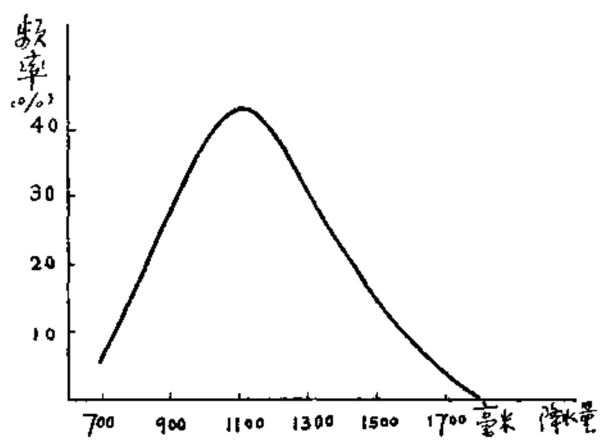

图 1.3 上海年降水量頻率曲綫

累积頻数（或累积頻率）图的横坐标表示組限，纵坐标表示累积頻数（或累积頻率），其图形也可用直方块、折綫或曲綫表示（图1.4，1.5)。

用表格和图形，可以簡单、直观地把一批观測数据的特征表現出来，但这仅能給人一个粗略的了解，而且如前面所述，数理統計学的基本任务是要根据样本的統計性质来推断或判定总体的統計性质。因为样本和总体数据的統計性质以及它們之間的內在联系往往隐蔽在現象背后，单单通过人們的直观认識是不够的，必須利用一定的数理統計方