数理統計講义

北京農業大學

农业物理气象系气象教研組

1961.12.

目　　录

第一章 緒　論

§1—1 統計的概念

（一）統計一語的意义

統計一語系由拉丁文中Status导来，它原来的含义是"情况"。最初出現这个名詞是在十八世紀初期，那时，它所表达的意义十分狹窄，只是說明一些国家的政治情况，直到該世紀末叶以后，才开始流傳起来。而到現今，可以說已发展到相当丰富和十分实用的地步，并且已以一門完善的单元出現于科學的領域中。

目前統計一語已有了广泛的含义，通常把說明社会及自然現象的全部數字叫做統計，这些數字可以反映出社会中或自然界中各种大量現象的发展水平和发展过程，以及这些現象的質量內容變化等等。

統計的內容可分为：資料的搜集，所得資料的統計研究，統計观察方法的擬定和統計資料的分析，而后者实为數理統計學的內容。

（二）总体和样本

总体是客观存在的，是在同基础上結合个別单位的集团，也就是指有一个共同的联系結合起来的許多因素（观測单位）的群。而样本是总体中由若干单位組成的小群。因此总体可以分割成很多样本。

总体必須要包括所研究現象的全部情况，例如北京年降水量总体，必須包括北京自古迄今再延展至未来极其长远的岁月中所有的年降水量，而現在所掌握的有限資料，不过是总体中的一个很小部分，也就是一个样本。

总体的概念，我們还可以理解得更广一些，这需要决定于我們所研究的任务与內容。例如我們要研究黃河流域的年雨量，則全流域各地区的年雨量就构成一个总体。由于研究的任务不同，故上例中所包括的地区大小、时間长短都不同。不过，以后我們所要談論的总体，都是以时間过程来表示的，也就是研究气象現象在时序上的變化。

气象現象的总体有着与別的統計現象不同的特点，即总体是无限的。例如工业成品的总体，毕竟可以知道它有多少个或者知悉一个十分近似的數目，而气象序列究竟有多长呢？除非考古到地球开始的那一天和推測到地球存在的最后一天，而現所能获得的資料，都是无限总体中的一个很小样本。这样說来，气象序列的总体項數似乎是无法捉摸的，幸而在概率論和數理統計學中給出了估計总体的定理与方法，才使得有可能在样本數足够大时去推測总体的近似情况。因此，总体与样本的概念有着严格的区别，两者不能混淆。总体只有通过样本去認識它，而不能直接算得，尤其不能把样本認作是总体，我們才可以根据样本的观測資

料应用數理統計中的方法对总体作出充分精确的估計，以滿足实际分析的需要。

§1—2 統計观察

（一）統計研究的三个阶段

所有完整的統計研究都必須包含三个工作階段：統計观察，統計归納和統計分析。

統計观察是任何統計研究的开始階段，是科學地有組織地采集和登記所需要的材料，把組織成总体的各个单位連同它們的所有特征进行有計划的登記，此时要对大量現象和过程作出綜合說明。

統計归納是把由統計观察中所得到的大量的原始資料列入綜合說明全部总体的归納表內，或繪列于統計图中。

統計分析是把整理好的材料，进行科學的加工和分析，得出結果。而且还必須对各种現象作出总結性的描述，以及从質和量上作出多方面的分析。随着分析就要計算各种簡单的或者复杂的統計特征值，本書是着重介紹这个方面的知識。

統計研究的三个階段是依次遞連着的，相互之間都有密切的关系，实际工作时，不应忽視其中任何一个階段。

（二）統計观察的重要意义

統計观察是統計工作的第一个階段，它的基本任务就是要从所研究总体的各个单位中获得眞实而可靠的原始資料，它的基本方法是大量观察法。

大家知道，一切的統計研究都是在各項原始資料的基础上建立起来的，只有取得能反映現实情况的資料，才能对所研究現象的实际研究发生作用。就統計資料的构成內容来說，每一个单位，每一个數据，都是构成資料整体的一員，唯有对全部个体的調查記录都做到正确无誤，才能保証統計整体不发生偏差。因此，統計观察是整个統計工作的基础，只有先把基础打好，才能保証整个統計工作的順利开展。

三統計观察的方法

（1）經常性观察和一次性观察

經常性观察就是随着事实的产生而随时进行登記，例如：每日定时气象的观测。一次性观察就是根据需要临时举行的观察，例如旱涝的調查、冰雹或大风灾害損失調查等等。

（2）全面观察和非全面观察

全面观察是在調查过程中要把被調查总体的所有单位都包括进去，但不是无条件地包括一切，只是在具体任务的范圍內进行。例如人口普查等。

非全面观察就是对被調查总体中的一部分单位进行观察。非全面观察应用很广，它只需要較少的人力与物力，且获得資料迅速而及时。非全面观察可分为抽样观察法、重点法和专題論述法。抽样观察是抽取总体的一部分进行观察，再把这种抽样样本的观察結果推論于总体。重点法是对总体中占絕大比重的那些单位进行观察，而把比重小的部分予以省略。专題論述法是深入研究一个或少數几个典型的对象，雖則这种研究結果不能据以推論至总体，但可更深入、更詳細地研究大規模調查中所不能說明的問題。气象分析中以抽样观察法为最常用，我們将在第六章中单独論述。

（3）資料来源和取得資料的方式

一、台站报表：是按統計表报舖形式，从气象機关中系統地取得必要的材料，如气象月报、年报等。

二、訪問农民：有些調查工作，不能只靠台站报表，而需要組織大批人力下乡訪問农民，因为农民是最有經驗和对当地情况是最熟悉的。

§1—3 統計归納

(一) 統計归納与分組的意义

在經过統計观察取得了大量的原始材料以后，繼之而来的任务是将这些材料进行汇总并加以系統化。汇总的方法首先就是按照一定的标志，把登記材料分門别类地分成若干部分，按照某一标志划分为各种不同的組，这种分組的方法就叫做統計分組法。

分組的目的是把同一現象、同一类型加以合并，使它們与其他現象、其他类型区别开来。它的作用就是按不同的标志，把資料的特征反映出来，以便进一步运用各种統計方法进行加工，来研究它們的規律性和相互关系。

由于資料的分組是正确地反映出社会或自然現象特征的基础，因此对現象的各种形态必須进行具体研究，以及正确地和客观地定出各种不同的組。如果分組分得不科學不正确，不管已取得的資料是如何广泛、正确和全面，我們仍不能根据这些材料得出客观的結論，所以統計归納階段也具有相当重大的意义。

气象計算中的分組比較簡单，只需把現象的不同类型、不同性質、不同地区（或地点）及不同作用条件区分开来，也就是說分組时必須把上述非同位相的資料分开，而把同位相的資料归納在一起。例如：降水量与徑流量是属于不同类型，平流霜与輻射霜的性質是不同的；人类活动影响前后的气象資料，由于产生的条件不同，因此都不能把它們归納在一起。

为了計算方便，同位相資料也要按數值的不同分成若干組。

(二) 統計表

原始材料經过統計归納后，其結果需要用數字来說明，所以我們可将整理和分組的結果，列于縱横表格內，以便使統計特征一目了然，这种表式就称为統計表。

用統計表，可有很多优点：

（1）避免了文字上重复解釋的麻煩，故用統計表常較长篇大論为生动。

（2）錯誤及遺漏的地方容易发現，便于及时檢查。

（3）統計參數的計算和分析工作可以簡化，并利于數字間的比較。

（4）表中將資料按次序排列，有明显的邏輯系統，为作統計分析的依据，又其所表現的內容、規律和相互关系，易得明确的概念和容易記忆。

(三) 統計图

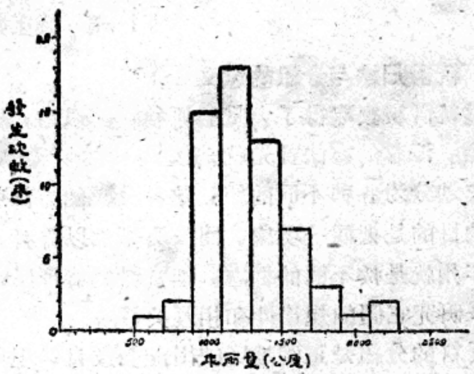
統計資料的數字用图形表現出来的，就叫做統計图。它可以明显地表达出現象的簡明关系，雖則极大极繁的序列，用統計表較已大为明晰，但对許多數字的整个情况，仍难完全領会，为此用統計图来表示，更为优越，且图中可以注明數字，以便一一对照，并有助于分析和比較，既通俗又适用于高深分析，实为統計表的一个有力助手。在頻率研究中，我們所用到的直方图和概率分布曲綫图等，都属于統計图。

（四）数字分組举例

取汉口站年雨量为例。自1882年至1954年共62年記录（其中缺測1884、1938至1946、1948計11年）中，使每隔200公厘划分为一組，如表（1，）所示，相应的直方图如图（1）。

表（1）汉口年雨量分組記数表

組段（公厘）	发生次數（年）
501—700	1
701—900	2
901—1,100	15
1,101—1,300	18
1,301—1,500	13
1,501—1,700	7
1,701—1,900	3
1,901—2,100	1
2,101—2,300	2
共　計	62



图（1）汉口年雨量分組图

这个例子，我們称它为組距序列。因为年雨量的變化可用一个距离来表示，这个距离（本例中为200公厘）即称組距，但也可根据不同情况采用不同的組距值。通常我們称組段两端的數值叫組限，而把該組段中的最大值称上限，最小值称下限。

如果序列的項數趋于无穷而組距取无限小时，一般可以得出光滑的鈴形曲綫，不再象图（1）中呈跳跃式的直方图了。

§1—4　数理統計在气象及农业气象上的应用

（一）資料整理方面

随着国民經濟建設的飞跃发展，尤其在1958年我国工农业全面大跃进以来，对于气候資料工作的要求日益迫切和日益广泛。为了对某地或某地区气候有科學的認識，为了及时和有效的滿足各方面的需要，做好气候資料服务工作，就必須将日積月累的气象观測記录进行气候學的加工整理。气象观測記录是反映大气中某种物理性質的，也就是用量的變化来說明性質的變化，因此整个大气过程的发展和變化可以归納用一堆統計數字予以表达。观測記录进行加工整理正是为了把这一堆數字利用分組分类和各种統計指标来总結归納出大气过程发展的規律。因此可以說观測記录的加工整理的主要方法，也就是統計方法，即是指統計資料的搜集，整理研究的方法。我們可以利用整理的資料研究寻找气候形成規律和不同地区的气候特征，更重要的是能以各种方法整理出最适用的更能广泛服务于国民經濟建設的材料。

我們知道，农业与气候的关系最为密切，各种作物的整个生长发育过程及病虫害的发展与分布，都受气候要素變化的影响，与溫度高低雨水多少，蒸发量大小，日照时間长短，风力强弱以及結冰情况，凍土深度等气候条件有密切关系，在进行农业区划，农业預报，在考虑农业的增产措施都离不了气候資料。为了滿足农业上的需要，我們就必須适合于农业特殊

要求来統計一般的气候資料，再結合物候观测資料，設計出农业气象的指标，和提供农业气象预报和增产措施的参考材料。

气象學的研究是离不开观測紀录的。气象观测紀录是非常龐大的，要从龐大的材料中抽取其有用部分，换言之，要能“以簡馭繁”，就要借助于統計方法。目前状况是只运用了最初步的方法，如平均數，离均差，相对頻數等的計算；較复杂的統計方法就沒有用，致紀录中所包含的有价值部分大多沒有被利用。要从資料中正确地認識这些資料所包含的意义是不容易的；尤其我們所有的紀录只是大气中片段的材料，要从片段材料来推論整体更不容易。只有借助于統計方法，方能“以偏見全”，不致为片段的，表面的現象所迷惑。无論以簡馭繁，或由偏見全，都需要統計方法，如頻數理論，抽样理論等都是。

（二）預报方面

无論在短期預报，中长期預报，单站补充預报和农业气象預报方面，數理統計的相关法，周期性研究、概率法、最小二乘法和經驗公式等是应用最广的一些。例如：利用相关法来研究两个或两个以上變量間相互关系的紧密程度；农业气象预报、尤其是土壤水分預报，应用數理統計的最小二乘法建立一連串的預报經驗公式，当然其他預报方面也要用到它。

气象要素的變化，是有相当周期性的，要发現这些隐蔽着的周期，及判断所得周期的眞实性，就是統計學上的周期性研究。

概率法就是先找出影响某一气象要素的各控制因素，計算各控制因素不同配合时該要素出現的概率，再就当时各控制因素的情形来預报。苏联 Fedorov 在綜合气候學里建立了相当完善的天气型分法，我們还可以用这些天气型作基本因素，再应用概率原理来作較长期的天气預报。

（三）理論的証实方面

近来气象學进展很快，新的理論，新的學說出現很多。但有一缺点，許多理論的提出，往往只举若干例子，或与平均状态对照一下，很难判别它正确或不正确。那末怎样客观地、有效地来証明一种理論的“正确”或“不正确”呢？这只有利用統計學原理和方法才能解决。在这方面如适合性檢定，独立性檢定，均匀性檢定，差異性檢定等都可供我們应用。我們要使气象學获得眞正的进步，一定要能够有效地証实或否定所提出的理論。

（四）气象学的基本研究方面

大气中的騷动，可說是或大或小的渦动。渦动的特点是偶然性非常大，开始时很小一点差别，甚至小到不能覺察，最后的結果可以大相悬殊。因此要百分之百地从开始状态預計未来状态，在理論上是不可能的。我們要处理这种复杂多變的渦动运动，要在永久不稳定的体系中发現可預告的規律性，这些規律性一定是統計的，所用的方法必須是类似統計力學的方法。現在国际气象文献中，已有很多关于扩散的統計研究的文献。我們如能将統計力學的方法应用到大气騷动的研究中来，一定可以促使气象基本理論研究的提高。

三、数理統計学的方法論

在现实世界里，我們还有一些尙未能掌握条件使其发生唯一結果的現象，所謂“偶然現象”。例如人們在显微鏡下可观察到的布朗运动；再如在气象观測中，人們必然会观察到偏大或偏小的非系統性誤差。这些變化的性質，和必然性變化完全不同，我們无法确切預料这

些變化的結果，我們只能够指出：这些事件可能有哪些結果。例如在布朗运动中，微粒可能往无穷多个方向移动，我們沒有方法确切断定个別微粒将要按照那一个方向繼續它的运动。属于这一类的事件，叫做随機事件。这些事件的變化，也就是哲學上所謂属于偶然性范疇的變化。

按照唯物辯証法的观点，偶然現象与必然現象是并存的，是可以相互滲透和相互轉變的。正因为这样，我們的世界并不是機械唯物論者所認为一切均归"宿命"的世界；也不是機会主义者所認为一切均属"偶然"的世界；而是一个有規律的，可以循着自然規律来加以控制、加以推动，加以改造的世界。偶然現象个别看来，好象是杂乱无章；但从集体看来，則有一定規律。概率和數理統計就是处理这类"集体性規律"的工具。它可以使偶然現象进入我們掌握之中。概率論在自然科學和生产方面雖然有了很多貢献，但是在另一方面，例如在政治經濟方面及其它社会科學方面，則不能濫用。所以我們应認清它的固有地位及应用范圍，不可过于誇張；尤其不可把概率論泛泛地看做全部統計學的基础，以免发生观点上的錯誤。还必須掌握辯証唯物观点，毛澤东思想，才能正确地应用數理統計，尤其是在社会科學方面的应用。离开了辯証唯物主义和毛澤东思想，很难得到正确的应用。資产階級的統計學者，采取伪妄的唯心主义观点，常用局部的資料，或脫离現象本質的資料歪曲事实的眞相，利用數理統計方法为掩护，导致虛假的結論，这一点必須特别注意。为了要掩盖階級本質，資产階級學者們常把資本主义社会制度所固有的許多現象解釋为自然的偶然現象，因而夸大了概率論与數理統計在所研究社会現象中的作用，这是不恰当的。

列宁的著作中，統計學始終是作为一門社会科學、具有深刻党性的科學，是和无产階級为反对資本主义、实現共产主义而进行的革命斗爭密切相联的科學。列宁无情地批判了資产階級統計的缺点，揭露了这种統計的袒护性的本質。列宁有力地捍卫了統計的科學基础，发展了馬克思主义的統計理論，将統計用作認識社会最有力工具之一。

总之，应用數理統計于生产和科學研究，必須掌握辯証唯物主义的基本观点，很好地學习毛澤东思想，才能完全客观地認識事物的眞面目，而不致迷失方向。

第二章 統計参数

§2—1 总 述

統計資料經过归納后，很多的实踐表明，中間的某一組出現次數最多，在这組以前和以后各組的出現次數逐漸減少，至两端的次數最少。如果画成出現次數的分配图形，就成为中間高两側下降的形状，若項數很多而組距很小，总可以繪成光滑的鈴形曲綫。但在一般情况下，也可以把出現次數的分配图形近似地用鈴形曲綫去配合，为了探討便利起見，很多學者設法求出了各种鈴形曲綫的方程式。

鈴形曲綫的形状是随着資料內各項數值的不同而異，例如，有的高而瘦，有的矮而胖，有的对称，有的偏斜等。为了能概括表示出这种特性，在方程式內参插了几个必要的参數。但都可以化为本章即将談的統計参數。

統計参數中有平均數，它是表示大量过程的水平，也就是对总体或样本中的各个标志值作概括的說明，把所研究現象的特征或代表性表現出来。另外还有表示資料分布的离散及偏斜程度的均方差、敧斜度及峭度等。所有这些，我們都可以通过已有資料适当地計算出来。

§2—2 平 均 数

（一）平均数的种类

平均數在統計學中常用的有五种，即

（1）算术平均數（均值）；$\bar{x}$；

（2）中位數（中值）；$\tilde{x}$；

（3）众數（众值）；$\hat{x}$；

（4）几何平均數；

（5）調和平均數。

（二）平均数的功用

（1）利用平均數可以說明在大量現象中的典型水平。例如某站的溫度，每时每刻都在變化着，我們要反映該站这一天的溫度就可以用平均數来代表。

（2）利用平均數可以按各个不同的變动标志达到在总体或样本間进行比較的目的。例如我們要比較两个地区的年降雨量，就可将两地的平均年降雨量来比較，以避免某一时間不正常的情况。

③利用平均數可以比較現象在不同时間內的平均水平。例如拿各年的平均溫度来比較，

可以发現它的升降趨勢，但在个别情况下，可能是看不出这种趨勢的，所以只有依据大量現象，平均數才能表現出事物的規律性。

（三）科学平均数方法論

平均數是在統計學中起着很重大的作用的，但資产階級學者一遇到标識的數量變化，不論在任何情况下，都把它平均化，而平均數的真正科學方法，首先是由列宁奠定的，主要的有下列三点：

（1）平均數只能应用于同位相的总体。馬克斯曾經說过："平均數永远是同一种类的許多不同个别數值的平均值"（資本論）。因为只有在同位相的总体中平均數才能反映出总体質的特点，才能成为总体的典型水平，如果总体的性質不同則平均數不能称为总体的典型水平，列宁称这种掩盖了总体質的差異的平均數为"虛构平均數"。

例如，某一河流的水位，虽然有了好多年的記录，但是由于上游开了沟渠，下游筑了拦河坝，或上游有了植林和水土保持工作，河流的情况有了變更，这样平均起来的水位，不但不能反映真实情况，反而易使人誤解，因为这些水位資料都是不同位相的，如把它們混在一起来求平均數，它們的性質差别就完全掩盖了。由于今后中小型农田水利工程大量兴建，地面徑流可以大大改變，因此我們在用統計法研究河流情况时需特別注意人类活动所引起的影响。

（2）組平均數对总平均數的补充作用。在同位相的总体中所求得的平均數，还是籠統的，使我們还不能去正确認辨現象的規律性，必須要用实际影响总体平均數的組平均數来补充說明。例如将月平均降水量来补充說明年平均降水量。显然月平均与年平均是有很大差别的，如只取年平均，則月平均的特征就完全掩沒了。

（3）利用分配系列补充总平均數。統計中的平均數虽則可以代表一般情况，但是不能表达該总体內大小两极端的情况和突出的情况，所以我們还要用分配系列来揭露它。例如年平均气溫是不能說明問題的，还必須知道最高的是几度，最低的是几度，各种不同溫度各占几天，当有了分配系列才能使我們更明晰更了解。

§2—3 算术平均数

算朮平均數在以后几章中我們叫它为均值，以符号$\bar{x}$来表示。因为它在統計學中占着相当重要的地位，且是平均數中用途最广的一个，在一般情况下，如不作特别說明，平均數总是指算朮平均數而言。

（一）簡單的算术平均数

我們如有一系列的變數 $x_1, x_2, \cdots\cdots x_n$，将它們的总和除以項數n，即得算朮平均數。

$$\bar{x}=\frac{x_1+x_2+\cdots\cdots+x_n}{n}=\frac{1}{n}\sum_{i=1}^{n}x_i \qquad (2、3、1)$$

式中Σ为总和的符号，讀为Сигма 或 Sigma，該符号上下附标的意思是 i 从 1 起一直累加到 n 止，如在不致混淆的情况下，可以省去附标的書写，即

$$\bar{x}=\frac{1}{n}\Sigma x_i.$$

上面公式是用同等比重来看待每个變數的，而只是把變數簡单地加起来用項數 n 去除，故称它为簡单的算朮平均數，其单位应与變數的单位一样。

（二）加权的算术平均数

如果變數 $x_1, x_2, \cdots\cdots, x_n$，各含有不同的比重，則比重占得大的，它們对平均數的影响也大，而比重占得小的，对平均數的影响也就小了。这样我們就不能平等地来看待这些變數，所以另用一种权衡其輕重的方法。

例如，x_1出現f_1次，x_2出現f_2次，……，x_n出現f_n次，則它們的平均數为：

$$x=\frac{x_1f_1+x_2f_2+\cdots\cdots x_nf_n}{\Sigma f_1}=\frac{1}{n}\Sigma x_1f_1 \qquad (2.3.2)$$

其中 n 为总的出現次數，即 $n=f_1+f_2+\cdots\cdots+f_n=\Sigma f_1$。这种算法是把各个不同的變量作不同等看待，因而需将各个變數x_i乘上它的出現次數f_i，这个f_i就給了變數 x_i 在全体中以一定的份量，这种份量就称为权數。所以在計算时如把各个變數乘上它自己的权數，再經过累加后除以权數的总和，这样我們就称它为加权的算朮平均數。

簡单的算朮平均數只是加权的算朮平均數中的一个特殊情況，就是各个變數的权數都是 1，或都相等。

在观測記录中，有时将某記录認为比其他記录精确些或有代表性的意义，統計时就把它看重，因此加权于它，如一記录精确程度比其他記录大二倍，把精确記录加权为 2，也就是把精确的記录計算二次，在气候資料統計中也可用来作为补充观測次數用，常将認为有代表性意义的記录加权，以使其平均數更接近我們的需要。

設 $x_1, x_2, x_3\cdots\cdots x_n$ 为某要素記录值

$f_1, f_2, f_3, \cdots\cdots f_n$为所加的权，$f_1$为对$x_1$的权，$f_2$为对$x_2$的权，$x_n$的权为$f_n$，在$x_1\cdots\cdots x_n$权不相等情况下写成表示式为：

$$x=\frac{f_1x_1+f_2x_2+f_3x_3\cdots\cdots f_nx_n}{f_1+f_2+f_3\cdots\cdots f_n} \text{ 或 } x=\frac{\sum\limits_{i=1}^{n} f_ix_i}{\sum\limits_{i=1}^{n} f_i}\cdots\cdots (2)$$

x 即为不同數量的平均數，

例如气温观測一日三次，为T_7，T_{13}，T_{19}，在一日观測中認为接近于一天温度的平均是 7 时观測，經过研究 T_7 多計算一次（加权）、才比較符合真正的（24 次观測的）日平均，則在統計日平均气温时多統計 7 时观測值一次，也就是說 7 时观測值的$f_7=2$，其他T_{13}，T_{19}的权为1，$f_{13}=f_{19}=1$。

代入②式：

$$\text{則 } \overline{T}=\frac{2T_7+T_{13}+T_{19}}{4}$$

（T代表温度，$\overline{T}$ 为平均气温）

最近我們用青島等八站气温記录以 $x=\frac{T_7\times 2+T_{13}+T_{19}}{4}$ 与 $x=\frac{T_7+T_{13}+T_{19}\times 2}{4}$ 加权

两种不同时間的加权平均做了比較，而以加权 7 时观测的更能接近各地的眞正日平均气温，如下表：

站名	一月			七日		
	24次观測	加权 7 时观測平均値与24次平均値之差	加权19时的观測平均值与24次平均值之差	24次观測	同前	同前
青島	−1.7	+0.3	−0.4	22.6	−0.1	−0.2
福州	10.7	0.2	−0.4	29.1	−0.3	−0.5
海拉尔	−27.0	−0.6	0.1	19.0	−0.6	−1.6
佳木斯	−18.8	0.5	−0.5	21.7	−0.1	−1.2
成都	5.7	0.4	0.3	25.8	0.0	+0.6
酒泉	−10.7	0.7	+0.6	23.7	−0.6	−1.5
烏魯木齐	−16.8	0.1	−0.1	23.1	−0.6	−1.8
汉口	1.7	0.3	−0.5	28.6	−0.3	−0.5

（三）算术平均数的数学性質

在实际工作中，往往由于數字很大計算麻烦，必須設法簡化，下面是几个算朮平均數的主要數學性質：

①算朮平均數与总項數的乘積，永远等于各个變數的总和

$$n\bar{x}=\Sigma x_i \qquad (2.3.3)$$

由这个數學性質可以說明算朮平均數是各个變量的共同代表。

②如果每一个變量中加上或减去任一常數 c，則新的算朮平均數中也增减加或少了常數 c。即

$$\frac{1}{n}\Sigma(x_i\pm c)=\bar{x}\pm c \qquad (2.3.4)$$

（証）：$\frac{1}{n}\Sigma(x_i\pm c)=\frac{1}{n}\Sigma x_i\pm\frac{1}{n}\Sigma c=\bar{x}\pm c$.

③如果每一个變量用任一常數 c 乘或除之，則所得新算朮平均數等于原算朮平均數乘以或除以此一常數 c，即

$$\frac{1}{n}\Sigma cx_i=c\bar{x} \qquad (2.3.5)$$

$$\frac{1}{n}\Sigma\frac{x_i}{c}=\frac{1}{c}\bar{x} \qquad (2.3.6)$$

由于常數可以提出至总和符号外，故証明从略。

④所有變量与算朮平均數的离差的和恒等于零，即

$$\Sigma(x_i-\bar{x})=0 \qquad (2.3.7)$$

（証）$\Sigma(x_i-\bar{x})=\Sigma x_i-\Sigma\bar{x}=n\bar{x}-n\bar{x}=0$.

算朮平均數的这个性質告訴我們，變量向两方面（正的与負的）的偏差是可以相互抵銷的。

的。

⑤各變量与算朮平均數的离差的平方和为最小，即

$$\Sigma(x_i-\bar{x})^2=\text{最小} \tag{2.3.8}$$

（証）：設A为不等于$\bar{x}$的任意常數，則

$$\begin{aligned}\Sigma(x_i-\bar{x})^2&=\Sigma[(x_i-A)-(\bar{x}-A)]^2\\&=\Sigma(x_i-A)^2-2\Sigma(x_i-A)(\bar{x}-A)+\Sigma(\bar{x}-A)^2\\&=\Sigma(x_i-A)^2-\Sigma(\bar{x}-A)^2\end{aligned}$$

由于$\Sigma(\bar{x}-A)^2>0$，故

$$\Sigma(x_i-\bar{x})^2<\Sigma(x_i-A)^2$$

也就是說，如果用一任意數$A(\neq\bar{x})$来計算离差平方，是一定大于由$\bar{x}$算得的离差平方的，因此式(2.3.8)成立。

⑥所有权數都乘以或除以任意數，則所得的新算朮平均數与原来的算朮平均數相等。

这个數學性質很容易了解，因为算朮平均數的权數只是权衡在系列中所占的比重，若将所有权數乘以或除以任意數，只不过把权數加大或縮小若干倍，而幷未改變它們之間的比例关系，所以算朮平均數仍然不会改變。

（四）算术平均數的优缺点

①优点：

一、意义淺近而具体，不含抽象的數學性質，使人容易領会。

二、有严密的計算公式，計算手續簡单而迅速，不会因計算者的不同而得出不同的結果。

三、因为把全部观察值（變量）都包括在內，只要量數均属同位相就能代表全体分配的特征，也即富有代表性的意义。

四、只要知道變量的总數和項數，而不須知道各變量的詳細數目，即能求得算朮平均數。

五、容易用代數的方法来处理（如上述的几种數學性質），且便于从理論上来研究和証明。

六、能与所研究現象的物理性質联系起来，使其更具体化。

②缺点：

一、算朮平均數的數目，往往不是具体事实。如平均每家人口为3.84人，实际上人口不可能有小數的。

二、易受极端項的影响。

三、算朮平均數既从全体量數中算出，如任何量數有差誤，則它即受其影响，极端項有差誤，影响更大。

四、依品質分类或分等級的系列，不能求算朮平均數，如顏色的深淺程度，成積甲乙丙丁等。

五、系列两端的量數不定者，如用“以上”或“以下”表示的不能求得算朮平均數。

③結語

在实际气象計算时，缺点中的第二条第三条是可以碰到的，但若当項數增加时，这些影响就会減小，且极端項是不經常出现的，所以問題不大。这样說来，算朮平均數的优点多于缺

点，故能成为平均數中最优良而用途最广的一个。其他的几种平均數，只能适用于某些特殊情况，以作补充算术平均數的不足。

§2—4 中位数

（一）中位数的意义

中位數也称中值，以符号me来表示，中位數是一个在n个观察值（變量）中比它大的數和比它小的數各占一半时的數值，下面是两种情况的計算法。

①如變量都占有同等的权，設x_1，x_2，……，x_n为一系列已按大小次序排列的n个變量。

当n为奇數时，中位數就是它們最中間的一項，即

$$me=x_{\frac{1}{2}(n+1)}. \tag{2.4.1}$$

例如有一系列數：1、3、6、9、13，中位數就是第三項的6。

当n为偶數时，中位數就是它們最中間两項的算术平均數，即

$$me=\frac{1}{2}\left(x_{\frac{1}{2}n}+x_{\frac{1}{2}n+1}\right) \tag{2.4.2}$$

例如有一系列：30、26、23、20、17、13，中位數就是第三項和第四項的算术平均數21.5。

中位數是意义极不清楚和极不肯定的一种平均數的形式，假如中位數不能为小數时，則中位在事实上就不存在。如上例中，21.5用四舍五入的方法，得出22，但实际上22已不是它們的中位數了。又如有另一系列18、16、13、13、9，中位數应为13，但大于13的有两項，小于13的只有一項，这样13就不能分系列为两个相等的部分。

②如變量各含有不同比重时，計算中位數可用例来闡述：

表 2

組序	變量	頻數	累積頻數			
1	600—500	3	3	由大至小累積	25	由小至大累積
2	500—400	5	8		22	
3	400—300	7	15		17	
4	300—200	6	21		10	
5	200—100	4	25		4	

表（2）中的中位數应該在25次中的最中間一次，即第13次，但这里找不到第13次，只有第8次和第15次或第10次和第17次的數，不过我們知道中位數一定在300—400的一組里，一般是用直綫补插法来推求，可用下面公式：

$$me=A-\frac{\frac{1}{2}n-n_a}{n_{me}}\cdot h \tag{2.4.3}$$

及

$$me=B+\frac{\frac{1}{2}n-n_b}{n_{me}}\cdot h. \tag{2.4.4}$$

式中 A 是中位數所在組的上限，B 是中位數所在組的下限，n_m 是中位數的出現次數，n_a 是大于中位數組上限的累積次數，n_b 是小于中位數組下限的累積次數，n 是总的出現次數。

公式（2、4、3）是用于累積頻數由大變量累積至小變量的，上例中：

$$me=400-\frac{\frac{1}{2}\times 25-8}{7}\times 100=400-64.3=335.7.$$

公式（2、4、4）是用于累積頻數由小變量累積至大變量的，上例中：

$$me=300+\frac{\frac{1}{2}\times 25-10}{7}\times 100=300+35.7=335.7$$

不論用式（2、4、3）或（2、4、4），两者所算得的結果，应該完全一样。

（二）中位數的优缺点

①算朮平均數有一缺点，即受某些极端項的影响很大，而中位數則不然，它只要位置决定后，两边變量无論怎样變动，都不会影响中位數，故为避免受极端項影响时，可采用中位數，以补助算朮平均數的不足。

在記录較短时，众數算朮平均數沒有意义时，可用中位數，如干燥地区，雨量很少，變化性很大，求平均雨量意义較小，就用中位數較有意义。又观測能見度时也常取各方向能見度的中位數。中位數在云高統計上也是常用的。

②确定方法簡單，計算容易。

③极端量不确定者，如只知系列中的中間各項數值及其在总項中的大小次序，这样我們因不知全体變量，就无法来推求算朮平均數，但中位數是可以求得的。例如某整編資料中有溫度在 10°C以上的有若干次、在5.0°C以下的有若干次，两端的限度都不明确，而中位數是可以計算出来的。

④如現象无法用數字大小来表达的品質資料或分等級的系列，則可用中位數来表示。

⑤中位數有时为抽象數目，与实际不符。

⑥难用代數方法处理，如已知變量的总數及項數，不能求中位數。上述前四条为优点，后两条为缺点，由于中位數缺少优良的數學特性，不易用代數方法处理，因此使它在应用上大受限制。

（三）四分位数

将一系列的變量數值，按大小次序排列，如选擇三个点恰能分該系列为四个相等部分，則称此三点为第一四分位數、第二四分位數和第三四分位數。中位數也就是第二四分位數。因为四分位數的位置决定較易，計算方法与計算中位數相似，且在实际上应用很少，故此处我們不作詳細討論。普通在統計學上是用它来說明機誤的意义，它是一种分割數，不是平均數。

§2—5 众 数

（一）众數的意义

众數也叫做众值，用符 mo 号来表示，它是在一系列观察值中頻數（出現次數）最多的一个值，也就是頻數分配曲綫（機率密度曲綫）上頂点的横坐标。

众數的求法，有下列四种：

①把頻數最多的一組的組中值取出来作为众數。但若一旦将分組的方法即組距及組的上下限改變，它也随之改變，这样就成为非常活动而不可捉模的數了。

②皮尔逊在多次观察中，发現在微偏态分布的情形中，算朮平均數、中位數、众數的关系如下：

$$\left.\begin{array}{l}\bar{x}-me=\frac{1}{3}(\bar{x}-mo)\\ mo=\bar{x}-3(\bar{x}-me)\end{array}\right\} \quad (2.5.1)$$

或

③如观察項數相当多时，用下面两公式能求得比較准确的結果：

密而思公式 $$mo=a+\frac{f_{+1}}{f_{-1}+f_{+1}}h \quad (2.5.2)$$

鮑 莱公式 $$mo=a+\frac{f_0-f_{-1}}{2f_0-f_{-1}-f_{+1}}h \quad (2.5.3)$$

式中 a 为众數組的下限，f_{-1} 为众數組下一組（組值比众數組較小的邻組）的頻數，f_0 为众數組的頻數，f_{+1} 为众數組上一組（組值比众數組較大的邻組）的頻數，h 为众數組的組距。

众數也是表示趨于中常（平均）情况的一种量數，也可以說是表示平均状态的一种方法，它是一組序列中出现次數最多的數值，用它可以描写气候要素的大多數情况，如是分組記录，則先找众數組，众數組中点就是众數，仍以上海为例：

表 3

組距(°C)	中点（組距的中間值——組值）	頻數
1.1—2.0	1.6	4
2.1—3.0	2.6	2
3.1—4.0	3.6	9
4.1—5.0	4.6	12
5.1—6.0	5.6	9
6.1—7.0	6.6	11
7.1—8.0	7.6	6
8.1—9.0	8.6	8
9.1—10.0	9.6	6
10.1—11.0	10.6	5
11.1—12.0	11.6	1
总 計		73

由表看出4.1—5.0°C为众數組，4.6°C为众數，为更精确起見，也常可用直綫內插法所求的二个公式来計算：