

927982

高等学校教材

专题制图 数学模型

张克权 郭仁忠 编著

测绘出版社

高等学校教材

专题制图数学模型

张克权 郭仁忠 编著

测绘出版社

内 容 简 介

本书介绍了专题现象统计分析的各种数学模型的基本原理和解算方法,以及成果制图的技术方法。其中包括数学制图模型的基本概念,现象系统要素变量的特征与处理,现象系统要素的分级模型、相互关系模型、空间分布特征结构模型、发展动态分析和预测模型、内在关系分析和信息简化模型、组合特征的结构分析模型和类型划分的调整及区域界线的确定等各种模型的特点,以及复杂模型的结构类型。

本书主要作为地图制图专业专题制图数学模型课程的教材,也可作为有关研究方向硕士研究生教学参考,还可供测绘、地理及其它有关专业科技人员参考。

高 等 学 校 教 材
专题制图数学模型
张克权 郭仁忠 编著

测绘出版社出版
测绘出版社印刷厂印刷
新华书店总店科技发行所发行

开本 787×1092 1/16·印张 13·字数 292 千字
1991年6月第1版·1991年6月第1次印刷
印数 0,001—1,000 册·定价3.50元
ISBN 7-5030-0427-4/K·160

前 言

数学制图模型是运用数理统计（多变量统计）分析的方法建立数学模型和相应的制图模型，来研究分析多指标描述区域特征的内在分布规律和性质结构关系的理论和方法，这种定量分析的成果能为区域特征的研究和规划提供重要的参考依据。由于电子计算机的广泛应用，以及科研、生产实践对定量化研究的发展需要，数学制图模型的方法已被日益广泛地应用于自然和社会经济等许多领域。作为国民经济建设中最主要的研究手段之一的制图模型与数学模型的结合，是当前研究发展的方向。

数学制图模型研究的内容可分为三大部分：根据研究任务正确选择指标和数学模型；数学模型的建立和解算及制图模型的制作；对模型成果的专业解释。本书内容主要是根据武汉测绘科技大学地图制图专业“专题制图数学模型”课程的要求而编写的，重点在于对现象系统要素内在分布规律和性质结构关系等作统计分析时，能掌握各种数学模型的基本原理和解算方法，以及成果的制图等基本理论和技术方法；对于模型的选择和成果的专业解释问题，由于涉及有关专业的具体情况，书中不作详细讨论。作为一门课程的教材，不过多地重复数学的基本知识；同时，考虑到有关研究方向的硕士研究生教学需要和本科生专业知识的提高，也适当包括一些与数学模型有关的知识（用小字印刷）。本书中尽可能选用少量同样类型的原始数据作为详细介绍计算步骤的实例贯穿整个教材，使之便于对比分析；同时，由于指标数据收集困难，只得引用了部分国内外其它文献中已有的实例，以便说明模型的功能。

本书是在武汉测绘科技大学地图制图专业多年来使用教材的基础上，三次修改编写而成。全书由张克权、郭仁忠共同编写，最后由张克权整理；书中插图主要由高淑宁同志绘制。在编写过程中得到了许多同志的帮助，特别是王家耀教授和黄顺玉副教授对本书进行了初审，王家耀教授作了全面复审，提出了许多宝贵意见；童扬芬副教授也对原稿提出了不少修改意见。在此表示衷心感谢。

由于水平所限，书中难免还有不少错误和缺点，恳请批评指正。

编著者

1988年3月

目 录

第一章 概述.....	(1)
§ 1-1 现象系统要素的特点和分析方法.....	(1)
§ 1-2 数学制图模型的意义和功能.....	(2)
§ 1-3 数学制图模型的类型及其作用.....	(3)
第二章 现象系统要素变量的特征与处理.....	(6)
§ 2-1 现象系统要素的常用统计量和分布特征参数.....	(6)
§ 2-2 现象系统要素变量的归算处理.....	(8)
§ 2-3 变量数据的规格化方法.....	(10)
§ 2-4 几种主要相似性统计量.....	(12)
§ 2-5 数量化理论的基本原理.....	(15)
第三章 现象系统要素分级模型.....	(17)
§ 3-1 要素分级的一般要求和方法.....	(17)
§ 3-2 根据要素数据分布特征的分级模型.....	(18)
§ 3-3 根据要素分级划分的面积统计特征的分级模型.....	(21)
第四章 现象系统要素相互关系模型.....	(29)
§ 4-1 要素空间分布特征相互关系的相关模型.....	(29)
§ 4-2 要素空间分布特征相互关系的信息模型.....	(36)
§ 4-3 现象空间分布特征相互关系数学制图模型的建立.....	(51)
§ 4-4 现象内容结构特征相互关系数学制图模型的建立.....	(56)
第五章 现象系统要素空间分布特征结构模型.....	(63)
§ 5-1 点状物体空间分布模型.....	(63)
§ 5-2 现象结构的“重力”模型.....	(64)
§ 5-3 现象空间分布规律分析模型.....	(67)
第六章 现象系统要素发展的动态分析和预测模型.....	(76)
§ 6-1 现象要素发展动态的回归模型.....	(76)
§ 6-2 现象发展动态模型中的要素选择.....	(85)
§ 6-3 现象要素发展动态的时间序列模型.....	(94)
§ 6-4 现象空间分布变化的动态模型.....	(96)
§ 6-5 数量化理论在动态模型中的应用.....	(99)
第七章 现象系统要素内在关系分析和信息简化模型.....	(101)
§ 7-1 现象间影响因素的简单评价模型.....	(101)
§ 7-2 多变量因子分析的数学模型.....	(103)

§ 7-3	因子模型的解算	(109)
§ 7-4	因子模型的算例及其简要解释	(113)
§ 7-5	对应分析的方法	(129)
§ 7-6	数量化理论在简化信息中的应用	(137)
第八章	现象系统要素组合特征的结构分析模型	(140)
§ 8-1	概述	(140)
§ 8-2	系统聚类法	(143)
§ 8-3	变量平均值逐步替代法 (贝利方法)	(148)
§ 8-4	树状图表法	(150)
§ 8-5	典型样品单元法	(151)
§ 8-6	模糊聚类法	(154)
§ 8-7	二次逐步聚类——区划聚类方法	(158)
§ 8-8	分类数目的估计和类型特征的描述	(160)
§ 8-9	应用举例及其简要解释	(163)
第九章	现象系统类型划分的调整和区域界线的确定	(167)
§ 9-1	概述	(167)
§ 9-2	费歇准则下的判别分析	(168)
§ 9-3	贝叶斯准则下的多组判别	(173)
§ 9-4	要素选择的逐步判别	(179)
第十章	数学制图模型的结构类型	(189)
§ 10-1	串状模型结构	(189)
§ 10-2	网状模型结构	(192)
§ 10-3	树状模型结构	(195)
主要参考文献		(202)

第一章 概 述

§1-1 现象系统要素的特点和分析方法

在我国经济建设过程中,制定规划、研究自然环境的条件和影响、了解人类活动的发展过程、预测未来的趋势等,都是十分重要的战略任务。为了做出符合客观实际的预测和规划,必须要有正确的研究分析方法和可靠的资料,才能保证研究成果的质量。随着科学技术的不断发展,一方面对研究成果的要求愈来愈高,另一方面也为研究工作的进行提供了科学的方法和条件。分析的方法,是把众多的、抽样统计的、离散的现象特征,采用一定的数学方法(即建立数学模型)来探求现象客观存在的规律性,分析现象在时间上的发展和空间上的分布,按此规律来研究现象,做出预测,为制定长远规划提供重要的参考依据。为了直观、明显地反映现象的规律性,可用地图或各种图表的形式(制图模型),来显示现象的研究过程和最后成果。随着统计数学、电子计算技术的不断发展,以及制图方法的改进,使之从研究问题的资料收集到结论成果的表达,逐步出现了各种数学模型与图象显示的结合,形成了研究现象的新手段——数学制图模型。

研究地理系统要素的过程,就是把各种方法获得的数据和资料,从离散的状态,通过各种途径探求出要素间相互作用的规律。科学技术的进步,生产力的发展,使人类对地理环境作用增强;同时,地理环境变化对人类活动也产生积极影响。50年代以来,是人类社会生产力发生深刻变革的年代,开始了地理科学研究方法的变革。60年代以来,开始把多元统计分析和随机过程引进研究领域,加强了系统分析的理论。

地理系统是地球表面的岩石圈、大气圈、水圈及生物圈与人类相互作用的物质能量和信息运动系统。地理系统的特点是:①它是由彼此有联系的诸多“部分”组成,各“部分”相互依赖构成统一体;②地理系统有一定的空间结构特征,是一个多级系统,并且是有序排列的,它可分为地质-地形、大气、土壤-生物、水、人类活动等亚系统;③系统状态有时间概念,是动态的;④不可能从数量上把握系统内全部要素的变化规律,系统状态及要素具有随机性质,发展过程中有一定的规律性,但人们还不能全部把握因素的影响,从个别事件探讨总体的规律性;⑤系统有实在的(物理的)系统和概念的(规划的、设计的)系统;⑥地理系统按性质不同可分成两大系统——地理环境系统和人类活动系统,需要研究两个基本要素相互作用的统一体。其中,人类活动系统是有意识的、有计划的主动系统。技术是中介,技术能加强人类对地理环境系统的干预能力,环境对人类也产生反馈。

地理系统的研究方法可分为解析法与组合法。解析是把复杂的、高级的地理系统分解、简化成低一级的系统,深入探讨要素之间数量关系,查明系统间、要素间的结构形式和功能特征。地理系统往往是多元线性系统(非线性可处理成线性),组合就是按要素间的相互关系,以及各级系统联结的规律,从简单到复杂、低级到高级的合成过程。解析是简化、

分解大的系统,建立简化的数学模型;组合是联结、合成复杂的系统。组合的过程也是地理系统的模拟过程,是舍去实际系统的许多次要因素和次要部分,建立起便于应用的地理系统,属于概念系统范畴。从建立简单的两要素之间的数量关系,利用多元方程组来模拟比较复杂的地理系统,都是组合。地理系统是多级复杂系统,各级系统目标彼此有关联,用图论概念,可以联结目标“树”。对地理系统研究的要求是:以所需的精度对分析的成果定位,并作出必要的结论和预测。

§ 1-2 数学制图模型的意义和功能

20多年来,在自然、经济等诸多领域中,迅速发展了采用数学模拟的研究方法,也促进了地理学和制图学研究中的数量化,即数学模型化的研究。利用数学表达式描述客观事物或其属性,称为数学模型。建立数学模型的过程,称为系统的数学模拟。建立模型,单独求解,进行解释,寻求解决方案,是数学模型的优点。然而,一般的数学模型并不能十分完善地解决地理系统研究的任务。因为地理系统各部门主要研究某种自然和人文现象的结构、发展和职能的区域分布特点,因而将必然给采用数学抽象的研究方法带来某些困难,使数学模型的研究与解决实际问题脱节。为了提高数学模型的实用意义,除保持其优越性之外,还应注意具体的区域分布特征。

如果在数学模型中导入区域分布参数,这就使得模型的算法复杂化。因此,常用综合性的模型,既有数学模型,同时也配合采用其它的模型,使之简单明确地表达现象和过程的区域分布特征。

地图制图模型具有具体的区域分布特征,并具有许多对地学研究极为重要的性质。地图既是有关自然和人文现象,以及它们综合性、系统性的信息来源,又是研究自然环境的时间、空间规律性的可靠手段。

地图在科学研究中的作用,很大程度上取决于地图上反映的科学信息的丰富程度和可靠程度,以及这些信息与所进行研究任务的适应程度。为了描述地理信息,要在地图上完善地表示一种或几种地理现象或地理过程的指标,即按观测或量测的资料来描述基本特征。虽然许多地图只局限于表示最简单的质量特征,但是随着对分析研究要求的提高和地图在国民经济建设中作用的加强,对地图所表达信息的精度和完备程度的要求也有了明显地提高。由表示单要素的解析特征向综合反映多要素组合成果发展,并且逐步深入地研究现象间相互联系,编制显示分析成果的结论地图。在编制这些地图时,必须按一定的目的方向,选择与所研究任务相适应的数学模型来处理地理系统的数据资料,并编制相应的各种地图。为了编制反映客观现实的最优地图,使这些地图能解决具体问题,往往导致试验性地选用数学模型,寻求结论地图的最优方案,这就形成了数学制图模型的新方向。综上所述,数学制图模型,就是指在分析研究区域的自然和人文现象时,把数学和制图的模型进行系统的有机结合,这种相结合的数学制图模型与其它研究方法也很容易配合。

目前已经应用的各种数学制图模型,虽然它们的结构有某些差别,但都可以由独立的数学模型和由此模型编制的地图组成每个基本环节,基本环节的数目与所研究的任务和编

制地图的困难程度有关。当建立简单的地理现象或过程的模型时（即解决不复杂的地理研究任务），常是单一的基本环节，即使这种情况，也需设计多种方案，编制具有不同数量分划表的结论地图，以寻求解决研究任务的较优方案，较精确地反映模型化的现象和过程。如果任务复杂，需揭示某种现象和过程的内部性质和关系时，则采用不同的数学模型和地图，组成具有一组基本环节的复杂的模型结构；每个数学模型和每幅地图，在数学制图模型化中起着各自的作用，它们具有特定的职能和专门的性质。但是，在每个基本环节中，地图是主导因素。

在建立数学制图模型过程中，地图完成不同的职能，大致可分为三种类型。

1. 地图作为建立模型的资料来源和确定研究任务的依据。这种地图反映了研究现象所需的重要性质，根据这些地图确定研究的途径，检核原始资料数据的详细性和可靠性，以及作为原始信息的基本来源。因此，在地图上最有价值的是：区域组群和综合体的空间分布，即定位信息（坐标）；以数量量度反映的抽象分类；时间、空间信息的集合；获得集合和离散特征的可能性，以及它们的相互转换；地图上解析性和组合性特征信息的结合。

2. 模型化过程中的工作地图能明显地反映数学计算成果的空间分布概念，并作出全面的评价。并且很容易反映出所采用数学模型的优缺点，确定模型化的前阶段成就和明确后阶段的任务；从研究的目的与所采用的数学模型相比较，评价原始信息，获得模型本身或其信息保障所产生的错误，确定是否需要返工计算或改换其它模型。

3. 模型化最后阶段的成果地图是反映模型化的最后成果，解决所定的研究任务。地图本身作为数学模型的图象显示，必须对模型进行全面评价，检查其可靠性、现势性等，明确实际中应用的可能性。

利用地图对模型成果进行地理解释，根据模型化过程中处理的信息来认识区域规律性，以及反映某些地理现象和过程的结构、发展及职能特征等，是十分方便的。并且，在标定地理现象的性质方面（包括各种地理带），地图也具有特殊的功能。

在数学制图模型化中，数学模型的主要职能是有目的的加工处理原始信息。由于所研究的任务不同，加工处理的复杂性也各不相同，为了适应制图自动化技术的需要，需对原始数据资料进行简单的处理，并对编制组合性地图进行根本性的转换处理。

采用的数学模型与地图的内容、类型和编制方法有一定的关系，在电子计算机上加工信息的容量和质量，影响到地图的完备性和详细性，并在很大程度上能确定其表示方法的特点。

数学模型和制图模型的结合，是一种分析研究现象的有效方法，必须要研究各种模型的相互关系，每种模型的特殊职能，以及对数学制图模型建立过程的统一管理；同时，正确采用数学和制图模型，应最大程度地利用各种模型的优点，并使建立的模型精确、经济和目的明确。

§1-3 数学制图模型的类型及其作用

合理选择数学制图模型是比较困难的，这是由于所研究的地理现象和过程具有复杂性，

因而要求模型成果与所研究任务严格适应。通常, 数学制图模型带有试验性质, 这是由于一方面是对模型化过程的方法论研究不足, 另一方面是与所选用的原始数据资料是否有代表性有关。这就使得数学模型方法较难统一标准化, 不易编制标准的框图和通用的程序。应该在各阶段内由地理学者和制图学者来检核, 针对具体情况拟订模型化的独特框图, 对数学模型的种类和专题地图的方案进行试验。由于数学制图模型具有多种方案, 因而可避免产生片面的或错误的结论和评价, 以及资料不足等缺点。此外, 多方案性能使复杂的数学模型改变为简化模型, 并且能够精确地反映地理现象的具体性质, 满足研究任务的要求。

建立数学制图模型最常用的方法, 就是对一系列数学模型和地图进行试验, 试验可能有两种情况: 一种是使用同一的信息基础, 并用同一原则编制结论地图 (统一的分划表及其编制方法), 如果采用几种不同的数学模型方法编制几种不同成果的地图, 可以显示出不同数学模型结果的差异, 评价几种数学模型的效果; 另一种是根据一种数学模型成果编制一系列地图, 其差别在于不同详细性的分类、分级表, 分析这些地图之后, 可以知道哪种分类的详细性能满足具体的地理研究任务。

在数学制图模型化中, 当采用一系列数学模型时, 常进行数学模型的转化, 在一系列模型的配合中, 必须有一定的相互联系组合。这种数学模型有目的的组合系统与地图配合, 能解决许多十分复杂的任务。

根据所研究任务的复杂性、原始资料信息的多样性以及各种数学模型的复杂结构, 可将组合系统大致分为三类。

1. 串状结构。它是由原始数据资料建立一个模型系列, 由多个数学模型和制图模型组成, 其前个模型成果是后个模型的资料, 逐步形成最终模型的结果。

2. 网状结构。由原始信息平列建立一系列数学模型, 进行比较后, 最后形成一种结论性的制图模型。

3. 树状结构。由原始数据建立同一分析目的的几种数学模型及其相应的几种制图模型, 形成一个地图系列。

就基本的数学模型而言, 常用的有下列几种:

1. 现象要素分级模型。现象研究中常需要对其数量特征进行分级, 以便有利于研究现象空间分布的趋势和差异, 针对不同的目的要求和数据分布本身的特征, 采用不同的数学模型并编制相应的地图。

2. 现象系统要素相互关系模型。地理现象研究中广泛采用研究统计相关关系的相关模型。应用相关模型编制显示地区和部门结构特征的地图, 确定地理现象和过程的时间、结构及区域参数的相似程度。同时, 相关模型作为其它数学制图模型的起始环节, 是建立其它数学模型和地图的基础。

3. 现象系统要素空间分布特征的结构模型。它是研究点状物体分布的离散程度及均匀性, 现象空间分布方面的潜力场, 以及建立显示自然和人文现象与过程在空间分布的规律和特点的统计面地图。

4. 现象系统要素发展的动态分析与预测模型。回归模型在数学制图模型化中用来分析时间和空间分布系列, 以及显示现象空间分布的动态变化。此外, 回归模型与相关模型、

因素及成分分析模型相配合，用于编制各种组合性地图（类型图、评价图和预测图等）。

5. 现象系统要素内在关系分析和信息简化模型。在相关分析的基础上建立的因子分析模型，可以在相应地图上的多维信息中，区分和反映出最为实质的部分，它们说明了最重要的地理规律性，在对要素原始信息的目视分析或制图时，是很难确定这些规律的，但是若按每个单独主因子或其某种组合编制不同类型的结构地图时，则十分方便。同时，因子分析模型和制图的成果，可根据组合因子的数据作为资料来编制各种类型地图的分类模型以及其它数学模型的基础。

6. 现象系统要素组合特征系统结构分析模型。聚类分析(群分析)模型是在数学制图模型化的结论阶段来探讨的。采用这种数学方法，就已转向编制类型图、评价图和其它组合型地图，产生了自动化分类的广泛可能性。

7. 现象系统类型划分的调整模型。判别分析模型，用于检核改进自动化分类成果的合理性，或在原有分类样品单元基础上，补充局部新样品单元的现象分类方法。

上述各种数学模型除了具有各自独立的功能和意义外，相互之间还可以有机结合，互相补充，综合分析，以便更深刻、更全面、更准确地揭示所研究现象的相互关系、内部规律、空间变化、发展趋势等。

第一章 复习思考题

1. 举例解释地理系统研究的解析和组合方法。
2. 什么是数学制图模型，数学模型和制图模型各有何功能。
3. 基本的数学模型有哪几种，各有何作用。

第二章 现象系统要素变量的特征与处理

§ 2-1 现象系统要素的常用统计量和分布特征参数

对地理数据变量统计加工, 计算要素在时间或空间变化的统计特征数和反映这种变化规律的统计量, 才能反映地理系统的特性, 特别是空间统计的概念。

地理数据变量可有定量数据和定性数据两大类。定量数据取值是实数, 是一种连续的量, 可分为间隔尺度数据和比率尺度数据。间隔尺度数据是实际标量衡量的要素数据 (例如生产总值、总量等); 比率尺度数据是以一个基准量作为衡量标准测度的数据, 称为某类指数 (如百分含量、某种比值等)。定性数据不是连续的量, 甚至不表示数量关系, 定性数据可分为三种: ①有序尺度数据。它不用连续的量测度; 而是要素规模的量级 (分级), 或者说是表示次序关系的数据, 它不表示具体的数量, 而是给出一个等级或次序, 例如人口密度的分级等。②二元数据。它表示样本是否具有某种性质, 例如研究土壤是否具有某种属性, 可以用二元数据 0 (没有某种属性) 或 1 (具有某种属性) 表示。③名义尺度数据。它是用以表示要素性质上差异 (分类) 的类型的数据, 例如土壤分类等。对名义尺度数据可以采用以下三种处理方法: 一是按照评等评分的方法近似地把这类数据作为定量数据处理; 二是按某种标志给予等级的编号处理 (如计算等级相关系数); 三是当作定性数据来处理, 用一个向量来表示, 它的每个分量分别对应一个等级, 每个分量都是 0、1 型数据。

一、常用统计量

(一) 频数与频率

表示现象分布的基本特征有频数与频数分布直方图、组距、组数; 相对频数 (即频率) 和相对频数分布表、相对频数分布直方图、相对频数分布曲线。这些可以用来分析工矿企业规模 (人数、设备能力、产值、利润)、城市规模 (人口、用地)、农作物产量、线路运输量、进行区域对比等。

(二) 平均值、数学期望、几何平均数、中位数和众数

这些统计量均表示现象分布的集中趋势。

1. 平均数: 表示要素在时间或空间分布上的集中位置 (如职工人均收入、平均高度、平均单产等)。

2. 数学期望: 常用加权平均值表示, 权就是频率 (P_i), 则 $M\zeta = \sum P_i x_i$ 。

3. 几何平均数: 是 n 个变量相乘开 n 次方的值, 常用于计算人口平均增长率等。

4. 中位数: 是按数值大小排列的中间数, 偶数列则为中间两个数的平均值。

5. 众数：大样本随机变量中如有某一变量数据出现的次数最多，这变量数据就称为众数。

(三) 极差、离差、方差与标准差、变异系数

这些是为了查明数据的离散程度，反映其内部差异的统计量。

1. 极差：是最大值与最小值的差值。

2. 离差与离差平方和：各数值 x_i 与其平均值 $\bar{x}$ 的离散程度称为离差 (d_i)。

$$d_i = x_i - \bar{x} \quad \Sigma d_i = 0$$

离差绝对值求和的平均值称为平均离差 (M_d)。

$$M_d = \frac{\Sigma |x_i - \bar{x}|}{n} \quad (i = 1, 2, \dots, n)$$

离差平方和为

$$\Sigma d_i^2 = \Sigma (x_i - \bar{x})^2$$

两个变量的叉积和为

$$L_{xy} = \Sigma (x_i - \bar{x})(y_i - \bar{y})$$

这些都是对平均值的离散程度的重要统计量。

3. 方差与标准差：方差是用离差平方和除以样本容量得出的， σ^2 为总体方差，它是反映一组数据与平均值的离散程度的重要指标，方差的开平方即为标准差(均方差)。当用样本标准差对总体标准差进行估计时，则采用无偏估计值，即

$$s = \sqrt{\frac{\Sigma (x_i - \bar{x})^2}{n - 1}}$$

在实际计算中，当在大样本的情况下，为方便起见， $n - 1$ 可用 n 代替。

两个变量间的协方差为

$$S_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

4. 变异系数：也称变差系数。用以衡量要素在时间与空间上的相对变化(波动)的程度，以 C_v 表示， $C_v = \frac{s}{\bar{x}} \times 100\%$ 。在空间分析时可作出变异系数 C_v 的等值线图，来表示各区域某种要素的相对稳定状况。如出生率、增长率、城乡人口比相对波动程度、农作物稳定程度等。

(四) 分组数据平均值与标准差的计算

当收集到的是已分组的数据时，知道了组间距、分点数值和频数，需要计算近似平均值和标准差。

1. 分组数据的近似平均值计算，常采用假定中项离差的方法。假定中项应符合两个条件，即出现的频数最大和可能最接近于平均值的数值。与各组中项的离差记为 0， ± 1 ， ± 2 ， $\dots$ 。用假定中项离差与频数，近似计算平均值

$$M_x = x_0 + \Delta \left(\frac{\Sigma d_k f_k}{\Sigma f_k} \right) \quad (k = 1, 2, \dots, n \text{ 组}) \quad (2-1)$$

式中: x_0 是假定中项的中点值; Δ 为组距; d_k 为第 k 组的假定中项离差; f_k 为第 k 组的频数。

2. 分组数据的标准差计算:

$$s = \Delta \sqrt{\frac{\sum d_k^2 f_k}{\sum f_k} - \left(\frac{\sum d_k f_k}{\sum f_k} \right)^2} \quad (2-2)$$

二、分布特征参数

时间和空间分布特征可用概率函数来描述, 一阶导数是概率密度(密度函数), 其特征参数称为分布特征参数, 常用偏态(偏度)系数和峰态(峰度)系数来衡量。

(一) 偏态系数

偏态系数表示要素分布的不对称性。统计偏态系数为。

$$C_s = \frac{\mu_3}{\mu_2 \sqrt{\mu_2}} = \frac{\mu_3}{s^3} \quad (2-3)$$

式中: μ_2 、 μ_3 是第二、第三阶中心距; s 为标准差。当 $C_s < 0$ 时, 要素以小于平均值的波动占优势; $C_s = 0$ 时, 说明要素的时间或空间统计在均值两侧均匀分布; $C_s > 0$ 时, 是高于均值的波动占优势。

(二) 峰态系数

峰态系数表示密度函数图形顶峰的凸平度, 是要素的时间、空间分布在均值附近的集中程度。统计峰态系数为

$$C_e = \frac{\mu_4}{\mu_2^2} - 3 = \frac{\mu_4}{s^4} - 3 \quad (2-4)$$

式中: μ_4 是第 4 阶中心距; 3 表示当峰态系数大于 3 时, 图形凸度过大。

§ 2-2 现象系统要素变量的归算处理

在建立地理系统要素的数学模型过程中, 主要是根据各种不同的变量数据状况, 通过数据的归算处理, 采用多元统计分析的数学方法, 建立各种相应的数学模型。此外, 近年来同时发展了多边形叠加(相同统计单元或不同统计单元)的方法, 来编制综合型或组合型地图。其叠加方法可有常规制图分析的方法和用电子计算机处理的方法。其中, 采用多元统计分析方法研究地区的特征, 将是本教材的基本内容; 多边形叠加方法只在本节的最后作概略介绍。

为了采用各种多元统计分析方法建立数学模型, 来分析研究制图地区的特点, 必须对各种地理数据进行归算预处理。各种原始数据, 在数据来源、数据类型和数据采集的地区分布特征等方面具有不同的特点, 现分别说明之。

一、地理数据的来源

数据资料的来源主要有以下四种:

1. 人文和自然要素的统计数据。在国家和地方各级的统计机构中, 进行有关方面的统计工作。如人口、人口增长; 自然资源; 工业、农业及交通; 商业和贸易; 科学文化以及国民收入等。

2. 遥感数据。遥感技术为专题制图提供了丰富的信息资料, 例如使用仪器通过象片获取数据, 即在象片上建立均匀密集的方格网, 读取网格点的要素数据, 称为数字模型。借助电子计算机可以全面地分析地表形态和地貌类型, 如计算平均高程、坡向、地面起伏、切割密度等多种指标, 并进行地貌类型划分。

3. 地图量测数据。利用手工或借助设备从地图上可获得很多数据。如各级河流及湖泊的数量、长度、面积, 各类土地的面积, 各种密度数据等。还可以用类似于在象片上布设均匀方格网的方法, 量测所需要素的数据。

4. 观测和实测数据。如气象、水文台站定期观测记录的数据, 野外勘测数据, 地球物理方面(磁场、重力等)的勘测数据, 环境污染监测数据等。

二、地理数据的类型

前已述及, 地理数据(变量)按其类型可分为两大类, 即定量变量和定性变量。定性变量各评价等级本身也是根据多种因素确定的。例如确定对某种作物的有利程度, 就考虑到土壤的肥力、酸碱度、质地、有机质含量等一系列指标。在许多情况下, 往往根据有关专家的分析确定, 其中不少是定性分析, 带有一定的主观性。较为理想的是定性和定量分析相结合的方法, 能够比较客观的确定评价等级。

在多个变量中, 由于它们在所研究的问题当中所处的地位不同, 常分为两种情况: 某些变量被视为变化的原因, 称之为自变量(或说明变量); 另一种变量被视为变化的结果, 称之为因变量(或基准变量)。

定量变量与定性变量之间是可以转化的。如果将数轴划分为互不相交的若干个区间, 当一些变量取值于同一区间时, 认为是同一等级, 这样便将此定量变量转化为定性变量, 相应的数据也转化为定性数据。反之, 对于定性变量及其数据, 设法按某一合理的原则, 实现向定量方面的转化, 并以得到的定量数据为基础进行预测或分类等研究, 这就是数量化理论的内容和目的。由此可见, 数量化理论使得我们不仅可以利用定量变量, 而且还可以利用定性变量, 更全面地研究并发现事物间的联系和规律性。

三、地理数据的区域分布特点

众多的地理数据都是针对一定的点、线、面(区域)单元, 有些是统计数据, 有些是量测数据。根据图形量测的数据有:

1. 根据地图上布设的规则网格量测其交点或中心点的数据值, 例如在等值线地图上量测数据。

2. 根据不同现象分布的相同对应区域获得数据或等级, 例如同样政区的人口密度和医疗保障。

3. 根据不同现象分布的各自轮廓范围, 量测各类或各级及其交叉重合的面积, 例如土

壤和植被各类型间的重合面积。

地图上对现象的分布采用不同的方法表示,常用的量测方法可归纳为:把符号法或线状符号法表示的现象归算为单位面积的数值或密度;把点数法表示的现象按规则网格归算为现象绝对值或密度,并可转换为伪等值线;把质底法或范围法表示的现象可量测各类轮廓面积或归算为单位面积的占有比重;等值线表示的现象可量测得出任何点的数据,其分层可作为分区使用;分区统计图表的数据也可转换为分级值计算。

面状的区域单元比较复杂。有的数据的区域单元是一致的,例如不少统计资料都是针对政区单元(省、县、乡等),完全可以用前述的方法进行数据处理。但是,在综合分析一个制图地区时,更多的情况是各种地理数据对于不同的区域单元,往往既有人文要素统计数据的政区单元,又有自然要素(例如地貌、土壤、气候等)各种类型的不同区域界线,即使用综合自然景观来分析的区域界线,也与政区单元(或几何格网)不一致。这个问题目前尚未得到很好的解决,但在实际分析中已做了一些试验。大体可以归纳为下述方法步骤:

- 1.首先在几种不同的区域单元中选择需要分析的区域作为基准单元,如果没有严格的基准单元时,可根据具体情况假定某种归算的区域为基准单元(较多变量使用政区单元或规则网格)。对每个基准单元所包含其它区域单元的某类变量的不同数据或级别,计算各自的权系数(往往是面积比重值),即在某个基准单元内每种类别所占面积与基准单元总面积之比,该基准单元内各类别权值之和为1。

- 2.如果每种类别的指标可以用定量数据反映,或者其中对分析目的起本质影响的某些定量指标,经多元统计方法处理成一种或少数数量指标变量后,用权系数为比重归算到基准单元总的变量数据。

- 3.如果是定性指标或等级数量指标,则需对要素各类别予以评级或评分,按前述的权系数为比重归算为基准单元的平均等级或评分,再参与数据处理。

- 4.以各类别的权值表示基准单元在定性变量中诸类目上的取值情况,这就是数量化方法的推广,称之为“广义数量化理论”。

四、多边形叠加地图的方法

上面所述的是把不同区域单元的多种变量,采用加权平均归算的方法进行预处理后,再建立相应的统计分析数学模型。70年代以来,逐步发展了采用常规分析或电子计算机直接处理几种不同区域单元(相当于不一致的几种多边形结构)叠置后形成各种小多边形图形的方法,并通过输出设备将图形输出。对各种要素变量不同方式叠置的等级或类别,需要用专业分析的模式方法确定。本教材中不作详细讨论。

§ 2-3 变量数据的规格化方法

每个样本单元有多种变量的原始数据,各种变量的量纲和数量大小是很不一致的,变化的幅度也不一样。假如直接用原始数据进行计算,就会突出绝对值大的变量,而压低绝对值小的变量的作用。为了给每种变量以同一量度,在进行建立模型的统计计算前,往往

需要对原始数据进行规格化变换。

一、标准差标准化（普通标准化）

如有 n 个单元，每个单元有 m 个数据，每个变量可记为 x_{ij} ； $i = 1, 2, \dots, n$ ； $j = 1, 2, \dots, m$ 。

标准化后的变量为 x'_{ij} ，则

$$x'_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \quad \text{或} \quad x'_{ij} = \frac{x_{ij}}{\sqrt{\frac{1}{n} \sum_{i=1}^n x_{ij}^2}} \quad (2-5)$$

式中： $\bar{x}_j$ 为第 j 个变量的平均数； s_j 为第 j 个变量的标准差。

经过 $x'_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}$ 变换后，每种变量平均值为 0，方差为 1。

二、极差标准化

普通标准化要计算标准差，为了避免麻烦，并把变量变换到 0 和 1 的范围之内，也可采用极差变换。选出每个变量的最大值 $x_{\max}$ 和最小值 $x_{\min}$ 后，按下列三种公式之一予以变换（令标准化后的变量为 y_{1ij} 、 y_{2ij} 和 y_{3ij} ）。

$$y_{1ij} = \frac{x_{ij} - \bar{x}_j}{x_{j\max} - x_{j\min}} \quad (2-6)$$

极差标准化后的新变量的值在 ± 1 之内变化。

$$y_{2ij} = \frac{x_{ij} - x_{j\min}}{x_{j\max} - x_{j\min}} \quad (2-7)$$

极差标准化后的新变量的值在 0 ~ 1 之间变化，也称为“正规化”。

$$y_{3ij} = \frac{|x_{ij} - \bar{x}_j|}{|x_{j\max/\min} - \bar{x}_j|} \quad (2-8)$$

极差标准化后的新变量的值在 0 ~ 1 之间变化。 $x_{j\max/\min}$ 是变量与 $\bar{x}_j$ 差异最极端的最大或最小值，如果 $|x_{j\max} - \bar{x}_j| < |x_{j\min} - \bar{x}_j|$ ，则用 $x_{j\min}$ ，反之用 $x_{j\max}$ 。同时， $\bar{x}_j$ 也可以用 $x_{j\min}$ 或 $x_{j\max}$ 代替，但有时可能没有用算术平均值稳定。

由于定性观测中可把某一变量的出现记作 1，缺失记为 0，因此用后两种公式变换就便于把定量数据和定性数据混合在一起使用。

二、加权极差标准化

上面所介绍的数据规格化方法，都是把每种指标变量在各样品单元中的平均值作为比较的假设标准，同时假定各指标变量对分析组合评价的影响是同等的，这在一般情况下，尤其是当采用比较同类的指标来分析地区的组合分类时适用。如果各种评价的指标对样品单元组合分类的影响很不一致，就必须考虑对各种指标给予不同的权（比重）值 P_j ；当有些指标具有某种标准评价值（例如影响农作物的适宜降水量）时，就用标准评价值 $\bar{x}_j$ 代替