

计量地理学导论

张 超 张长平 杨伟民 编译

高等教育出版社



计量地理学导论

张 超 张长平 杨伟民 编译

高等教育出版社

内 容 提 要

本书主要根据[英] Robert Hammond, Patrick McCullagh 合著“Quantitative Techniques in Geography: An Introduction”一书,参考有关文献编译。全书共分十章。1—3章介绍定量描述技术;4—9章介绍统计推断技术,包括统计假设检验;第10章介绍建立地理模式的技术。此书为导论性质,避免了过多的数学推导,数学基础只要求微积分和数理统计的基本知识。每章均配有习题,书后还附有习题答案。本书可供高等学校地理系师生阅读,也可供地理工作者参考。

计量地理学导论

张 超 张长平 杨伟民 编译

*

高等教育出版社出版

新华书店北京发行所发行

河北省香河县印刷厂印装

*

开本 850×1168 1/32 印张 12 字数 287,000

1983年1月第1版 1984年7月第1次印刷

印数 00,001—6,000

书号 12010·031 定价 1.85 元

序

社会主义建设的实践,要求地理科学对自然资源、自然环境和地域系统演变进行定量分析,应用数学方法和现代计算技术,寻求地理现象发生性质变化在数量方面的依据和尺度,从而对于地理环境的发展、变化提出预测及最优控制。计量地理学正是在地理学发展的这种需要和技术进步的基础上产生的。

所以,计量地理学是数学和地理学互相渗透、密切结合而形成的一门边缘学科。它应用数学方法研究地理学基础理论和解决地理学中的实际问题。电子计算技术是计量地理学研究的主要技术手段。

六十年代至七十年代有人称之为地学中的计量革命。在这个时期内,引进了新技术、新方法,注进了新的见解,并产生了新的理论;但是学科的发展有继承性,离不开原有基础。计量地理学也是如此。

地理学的内容丰富,有许多分支学科,各有不同的具体研究对象和任务。但总起来看,正如李春芬教授所指出的,当代地理学所强调研究的,大致可以归纳为下列三方面的主题:

第一,空间与过程的研究 这是关于地域分布与地域过程的研究,主要查明地表事物的分布位置和模式及其成因与变化,其特点是把分布与过程结合起来,强调动因。以往地理学虽然也关注分布位置和过程,但是对过程的研究是很薄弱的。常满足于静态的研究。过去关注位置,主要停留于描述;现在则着重于区位分析(location analysis)或区位因素的分析,即通过与其他有关因素的联系过程来进行分析,常用数学模型探索地表事物之间的规律

性的空间关系，并以此来为人类活动有关的地面设施提出适当的位置安排。也就是说，过去关注事物或现象的分布主要是为了反映现状，揭示地域差异；现在则强调分布规律和模式的形成，为生产布局服务。这里要指出，这一转向不是近期才见端倪。早在十九世纪初期，杜伦（J. H. von Thünen）已为农业区位分析和模式的制定作了开端。但这类工作的广泛开展，却是近期的事。对于空间过程或地域过程的研究，其中包括自然过程和人文过程，过去是很薄弱的。戴维斯曾就地貌发育的问题讲过这样一句话：地貌是构造、过程和时期的函数。后来，洛赛尔（R. J. Russell）又进一步提出地貌形成过程的研究，特别是现代地貌和过程的研究。七十年代初期，乔利（R. J. Chorley）强调自然地理学要研究空间过程——反应系统的重要性。最近我国地理学家吴传钧、沈玉昌、左大康等，也在有关论文中点出了这方面研究的重要性。总之，地理学强调空间与过程结合起来进行研究，是近期的一个重要发展，通过过程来研究分布位置与模式，并从过程的演化中来预测变化的趋势。所谓模式有多种多样，但有一个基本点是相同的，即：它是现实事物或真实世界的一种简化的模型，通过各种指标的相关来显示地域性质和空间关系。如前面提到的杜伦的农业土地利用模式，就是反映十九世纪初在资本主义国家内以城市为中心的一种简化的环带状模式，在今天看来，这类模式过于简单粗略，并停留在静止的平均状态，看不到动态变化。当代地理学在六十年代计量革命的基础上建立数学模式，要求更定量化、更精确地去研究空间过程，从而揭示地表事物的分布模式和空间关系，进一步为生产布局服务。由此可见，当代地理学在空间与过程的研究方面，是在地理学原有的基础上有了进一步的发展，其侧重点有了转移，向动态和理论与实践结合的方向前进了。

第二，生态研究 所谓生态研究，对地理学来说，主要指人地

关系的研究。这是地理学的一个老课题。但过去研究人地关系，偏重于地对人的影响，甚至予以绝对化，以致环境决定论占了统治地位。现在不仅注意了两者的相互影响，更重要的是倾向于强调人对地的影响。因为科学技术的进步，人类利用环境的范围扩大了，程度加深了，人对地的干扰和影响愈来愈大，甚至使环境质量恶化。如最近发表的资料，我国西北干旱与半干旱地区，由于过去土地利用不合理，沙漠化的面积达到 17 万平方公里（包括各历史时期和近半个多世纪形成的在内），受到严重威胁的达到 16 万平方公里，占全国土地面积的 3.4%。再由于人口的日益增长，人类对物质的需求不断增加，人类利用自然的范围也必然会相应地不断扩大，利用的强度也不断提高。在这种情况下，人地关系的研究又重新被提了出来，并把着重点转移到人对地的影响。这说明在人地关系这一重大课题上起了质的变化，大大向前发展了。人地关系的研究，就是要讨论人与环境变数的相互关系，进而解释彼此间的连锁性。这表明地理学者不仅讨论地区间的空间差异，更重要的是在单一而有界限的地理区内，进行各种现象关系的分析研究。

第三，**区域研究** 整个地球表面的地理环境千差万别，不同的区域各有不同的特点；另一方面，区域之间也有一些相似的特征。这种差异性和相似性，是区域研究的根据，也是地理学的基本点之一。但是过去对区域研究只满足于描述和解释区域特点和区域差异（区内的和区际的），主要是为了解决这方面认识上的问题。现在，已开始重视它在实践上的意义。还要指出，过去研究区域强调区域个性，因而把注意力主要局限于形态一致的区域（formal or homogeneous regions）；现在则日益趋向功能区域（functional regions）。在实践中为了某一特定目的或问题，往往不是一个区（指形态一致区）所能单独解决，常涉及几个区域，为此，单因素

(或要素)的空间分布就成为必要的基础资料和科学依据,它可跨越几个形态一致的区域。如对流域之间调水的问题,首先就要查明有关流域内的水资源分布和需水情况,把水资源盈余和匮乏的地区结合起来进行通盘考虑。这类区域就是所谓功能区,区内各部分存在功能上的联系,也称之为空间组织区域(spatially organized regions)。以城市为中心的结节区(nodal regions)也属于这类区域。所谓区域空间组织,用一句通俗的话来讲,就是如何重新安排河山。在实践中这类研究多应用于区域规划,包括城市规划、流域规划等方面。在理论上主要从事区际关系,如区际交流等的研究。

针对上述三方面的主题研究,现在强调三种分析,即空间分析(spatial analysis)、生态分析(ecological analysis)和地域综合体分析(regional complex analysis)。上述三种分析,均须利用计量分析的方法,方能使当代地理学所强调的主题研究更臻于科学化、合理化。

地理事物,普遍地受到随机因素的影响,因此研究随机现象的概率论与数理统计,是目前计量地理学中主要的数学工具。地理系统是多级、多要素系统,探讨其类型结构、类型组合、空间关系、要素之间的相关等,经常应用多元统计分析的方法,如多元线性回归,逐步回归,聚类分析,判别分析等等。

计量地理学的研究在国内尚属发展阶段,但已为我国地理学工作者普遍重视。计量地理学已列为大学地理系的必修或选修课程,只是计量地理学的中文专著尚不多见。我们在为华东师范大学地理系地理和地貌专业讲授计量地理课的同时,以R. 哈蒙得(R. Hammond)和P. 麦卡勒(P. McCullagh)合著的《地理学定量技术导论》(Quantitative Techniques in Geography: an Introduction, 牛津大学出版社出版,1978)为基础,参考其他有

关专著及论文,编译了这本《计量地理学导论》,介绍定量描述技术(第一章至第三章)、推断技术(第四章至第九章)、模型制作技术(第十章)。希望通过本书的论述,使读者了解现今地理学的发展趋势,并利用计量分析的方法,解决地理学研究中的实际问题。本书既可供地理工作者和大、中学地理教师参考,也可作为大学地理系各专业学习计量地理课程的参考。

本书编译过程中,得到李春芬教授、周淑贞教授及其他师友的热情关怀和鼓励;在确定本书选题的时候,得到人民教育出版社大学地理编辑室有关同志的积极支持,并提供最新版本的英文参考书,使编译工作得以顺利进行,在此一并表示衷心感谢。全书插图由刘永瑜同志绘制,谨表谢忱。

由于编译者的水平所限,书中错误及不足之处,希望读者批评指正。

编译者

1982年10月于上海

目 录

序

第一章 数据处理	1
§ 1. 频数分布	1
§ 2. 平均数和离散的测度	4
一、众数	4
二、中位数	5
三、百分位数	5
四、累积频率(或累积百分比)曲线	8
五、算术平均数(使用个体数据)	11
六、算术平均数(使用分组数据)	13
七、定向	14
八、标准差(个体数据)	19
九、标准差(分组频数数据)	23
十、变差系数	24
§ 3. 洛伦兹(Lorenz)曲线	25
§ 4. 吉布斯-马丁(Gibbs-Martin)多样化指数	29
§ 5. 三元分类图	33
§ 6. 威弗(Weaver)组合指数	35
§ 7. 资料的转换与组合	39
一、秩	41
二、比例	41
三、标准化	42
四、对数变换	43
五、加权	45
习题	46
第二章 空间分布的测度	48

§ 1. 空间分布的四种类型	49
§ 2. 点状分布	50
一、中心位置	50
二、离散程度和集中程度	53
§ 3. 线状分布—网络	61
一、运输网络	61
二、界线网络	69
§ 4. 离散区域分布的测度	72
一、空间洛伦兹曲线	72
二、差异性指数	78
三、洛伦兹曲线与差异性指数的局限性	80
§ 5. 连续区域分布的测度	82
习题	87

第三章 时间序列91

§ 1. 图象法	91
§ 2. 增长和下降	92
§ 3. 指数数字	93
§ 4. 对数比例尺	96
§ 5. 趋势和波动	101
§ 6. 用最小二乘法求趋势线	106
§ 7. 指数序列(或对数序列)趋势线	108
§ 8. 指数模式的计算	111
一、预测	111
二、预报	113
三、事后估计	115
习题	118

第四章 概率分布 120

§ 1. 概率	120
§ 2. 加法定理	121
§ 3. 乘法定理	122

§ 4. 概率分布·····	123
§ 5. 二项分布·····	124
§ 6. 概率和频率分布·····	127
§ 7. 正态概率分布·····	127
§ 8. 正态曲线的性质·····	130
§ 9. 正态分布的累积概率曲线·····	135
习题·····	138
第五章 样本和估计 ·····	141
§ 1. 总体和样本·····	141
§ 2. 抽样结构·····	142
§ 3. 抽样过程·····	144
一、随机抽样·····	144
二、系统抽样·····	147
三、分层抽样·····	149
四、分组抽样·····	149
五、抽样设计举例·····	150
§ 4. 替代法·····	151
§ 5. 总体参数和样本统计量·····	153
§ 6. 抽样分布和标准差·····	154
§ 7. 连续尺度的样本估计量·····	155
§ 8. 计数尺度的样本估计量·····	160
一、二项分布的标准差·····	160
二、二项分布标准差的百分比形式·····	162
§ 9. 抽样比修正·····	163
§ 10. 人口抽样调查的标准差·····	164
§ 11. 抽取样本量的估计·····	165
习题·····	166
第六章 假设检验 ·····	169
§ 1. 假设检验的基本思想·····	169
一、原假设(H_0)·····	170

二、假设检验	171
三、选择统计检验的临界值	177
§ 2. χ^2 (卡方) 检验	180
一、例 1	180
二、例 2	183
三、例 3	187
§ 3. 二项检验	190
一、研究比例差异的二项检验	190
二、研究数值差异的二项检验	193
§ 4. 比较样本值大小的检验	196
§ 5. t -检验	198
一、独立样本的 t -检验	198
二、成对样本的 t -检验	200
三、关于 t -检验的一些注意事项	201
§ 6. 曼-惠特尼 U 检验	207
§ 7. 成对样本的威尔科克森检验	216
§ 8. 方差分析	222
§ 9. 定向的瑞利(Rayleigh)检验	227
习题	228

第七章 相关 231

§ 1. 相关系数 r	231
§ 2. 斯皮尔曼(Spearman)秩相关系数 r_s	236
§ 3. 肯德尔相关系数 (τ)	240
§ 4. 偏相关	245
§ 5. 双点列相关	247
§ 6. 相关矩阵	250
§ 7. 形态分析的相关约束	252
§ 8. 卡方列联系数	256
习题	257

第八章 回归分析 260

§ 1. 因变量和自变量	260
§ 2. 散点图、回归线和残差	261
§ 3. 内推、外推和预报	263
§ 4. 制作回归线的方法	265
一、等级尺度的数据	266
二、半平均法回归	267
三、最小二乘法回归	268
§ 5. 最小二乘回归线的置信限	273
§ 6. 非线性相关的回归	275
§ 7. 多元线性回归	281
一、回归方程的建立	281
二、回归方程的显著性检验	282
§ 8. 逐步回归分析	284
习题	286
第九章 空间分布检验	287
§ 1. 最近邻点指数	288
§ 2. 线状分布的最近邻点指数	296
§ 3. 样本的随机性检验	299
习题	304
第十章 定量技术模型	305
§ 1. 模拟模型——随机迈步方法	307
§ 2. 模拟模型——扩散方法	310
§ 3. 引力模型	320
§ 4. 坡的数学模型	327
§ 5. 计算机模拟模型	335
§ 6. 线性规划	339
附录	343
习题答案	363

第一章 数据处理

地理工作者经常会接触到各种类型的不同的地理数据，这些数据为我们认识地理事物的内在规律、研究它们之间的关系、预测地理事物的今后发展等，提供了丰富的材料和科学的依据。但是，这些资料通常是分散的，没有系统和次序，有些甚至不能直接地刻画出地理事物的特征。因此，要想从这样一些庞大的数据堆中找到有用的东西，得到可靠的结论，就必须很好地下一番工夫，对地理数据进行必要的整理。

例如，要对两个地区进行气候统计分析，就需要有气温和降水方面的资料，根据这些资料算出年降雨量的平均值、距平值（即逐年偏离平均年降雨量的数值），及年、月平均气温、月平均最高和最低气温等。这种把大量详细的气候资料，经过整理，计算成几个简单而有代表性的数据的方法，就是数据处理的一个方面。本章将详细介绍数据处理的各种方法。

§ 1. 频数分布

地理统计资料是由一种或几种地理要素的一系列数据组成。一般有两种不同的形式：第一种是对某一现象经过一系列观察而得到的连续标尺上测得的数据，例如，不同地区的小麦单产，某河流断面上洪峰的流量等；第二种是研究对象或某些指定数值重复出现的次数，也称频数，一般记为 f 。

假定要研究一个地区的商业经济状况，我们已有一组根据店员多少来确定商店规模的数据，其中所有 76 家商店的店员人数都

在 5—145 人之间。为了方便起见，把数据分成 15 组，使每个商店的店员人数都落在一个以 10 为组距单位的某一组（即 1—10，11—20，21—30……）中。

在本例中我们把变量定为商店的店员人数，记为 x ，而频数 f 就是落在每一组中的商店个数。

本例的观察频数分布如下：

分组	1—10	11—20	21—30	31—40	41—50	51—60	61—70	71—80	81—90
f	1	3	4	6	11	9	10	9	6

分组	91—100	101—110	111—120	121—130	131—140	141—150
f	8	3	2	2	1	1

图 1-1 为频数直方图，水平轴表示分组，垂直轴表示频数（落在每一组中的商店个数）。而每一个小矩形的高便代表一个频数，在本例中频数只能是整数。

变量可分为离散型和连续型二类。凡是可以一一列出的数据构成的变量，称为离散型变量，例如，店员人数就是离散型变量，象 3.75 个店员则是没有意义的。

连续型变量则产生另一类频数分布，气象站所记录的降雨量就是一个例子。不是每天都下雨，但可以倚着一个连续标尺来测量降雨量（读出数据的精度只受到仪器本身精度的限制），而视其为连续的。如，测得日降雨量是 3.75 mm。温度同降雨量一样也是一个连续型变量，它也可以在一个连续标尺上测量，但是有不同的地方，因为气温本来就是连续形成的（对于一个气象站可以一天没有降雨量，但不可能一天没有温度）。因此，可将连续型变量定义为：它在整个给定的范围内可以取到任意值，但它在空间或时间上可以是连续的，也可以是不连续的。

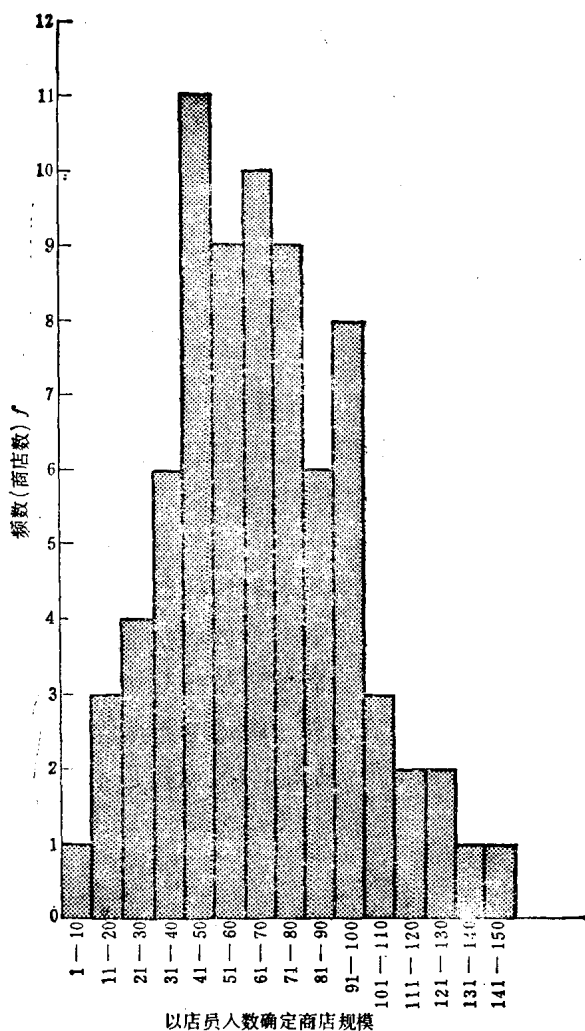


图 1-1 商店规模的频数分布图

§ 2. 平均数和离散的测度

频数直方图, 实际上就是地理资料规律性的直观反映, 但我们看到作直方图的过程颇为麻烦, 有时可以用某些统计特征数来反映地理资料的各种情况。

平均数是数据处理中最常用的一个指标, 在地理文献中, 可以经常看到“平均”产值、“平均”个人收入、“平均”雨量、“平均”流量等等。在地理学研究中经常以这种或那种形式用到它, 因此必须严格地加以定义。在数学中, “平均”一词是与“算术平均”同义的。本书中除了个别特别指出者外, 也都作为“算术平均”来理解。下面首先介绍两种不太严格的平均数的概念——众数和中位数。

一、众数

在地理数据中有最大频率(频数)的那个数, 即具有最大可能性出现的数值, 称为众数。因此众数所在的那一组就是频数最多的那个区间。图 1-2 表示商店的频数分布直方图, 它是按商品的类型来分组的。众数所在的那一组是电器商品商店的那一组。而在前述商店规模的例子中(见图 1-1), 众数是 41—50, 因为, 该处

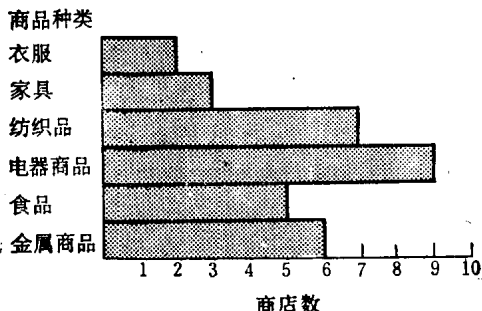


图 1-2 商店的分布(按商品的类型分组)