

环境统计学

吴聿明 编著

中国环境科学出版社

环 境 统 计 学

吴丰明 编著

中国环境科学出版社

1991

内 容 简 介

本书主要介绍如何在环境科学领域中以数理统计方法收集、整理、分析受随机性因素影响的数据，并对所研究、分析的问题予以判断、预测。内容包括：一、环境统计调查及数据整理；二、环境统计的概率基础；三、环境观测值的误差分析和计算；四、环境参数的统计推断；五、观测值的方差分析；六、环境科学中常用的统计分析、预测方法；七、抽样技术；八、多元统计介绍。书末有附表，可供查用。

本书可供从事环境科学工作者及有关科技人员，高等院校师生参考。

环 境 统 计 学

吴聿明 编著

责任编辑 夏伟松

中国环境科学出版社出版

北京崇文区北岗子街8号

北京市通县永乐印刷厂印刷

新华书店总店科技发行所发行 各地新华书店经售

1991年10月 第 一 版 开本 850×1168 1/32

1991年10月 第一次印刷 印张 18

印数 1—2 000 字数 482千字

ISBN 7-80010-890-2/X·472

定价：9.80元

前 言

环境问题已成为世界性问题。环境统计学是通过环境数据的科学分析,对某些环境问题进行评价、判断、预测,并为环境科研、设计、生产实践提供依据的有效手段。例如,观测数据异常值的剔除,污染物分布的统计检验、推断,环境中诸因素间和人为活动对环境过程的影响及它们相互间的相关分析,环境区划,污染源成因分析,多指标体系的环境质量评价,环境质量的预测和控制等等,环境统计学均能发挥它的应有作用。

环境统计学是个部门统计学,它是数理统计原理在环境科学中的应用。运用数理统计原理,结合环境科学理论分析、解释环境问题,正确进行环境问题的试验、统计并处理试验结果从而对环境问题给出客观结论和予以正确处理。书中给出的部分实例是选自其它作者的报告,在此深表谢意。

环境统计学和环境科学均是综合性很强的学科,限于作者水平,错误与缺点难免,希望读者、专家批评指正。

本书承蒙朱德威教授审阅,在此谨表谢忱!

编 者

目 录

第一章 环境统计调查及数据整理	(1)
第一节 环境统计指标体系及统计调查.....	(1)
第二节 环境统计的资料性质.....	(8)
第三节 展示资料的表格法.....	(10)
第四节 展示资料的图示法.....	(16)
第五节 样本特征数.....	(23)
第二章 环境统计的概率基础	(32)
第一节 概率概念.....	(32)
第二节 几个基本公式.....	(39)
第三节 随机变量及其分布函数.....	(48)
第四节 典型分布函数介绍.....	(53)
第五节 多维随机变量及其分布介绍.....	(71)
第六节 随机变量的数字特征.....	(76)
第七节 大数定律和中心极限定理.....	(100)
第三章 环境观测值的误差分析和计算	(104)
第一节 环境统计中误差.....	(104)
第二节 观测值的整理和计算.....	(109)
第三节 误差的统计分布和传播.....	(114)
第四节 离群值的取舍.....	(124)
第四章 环境参数的统计推断	(137)
第一节 点(参数)估计.....	(138)
第二节 区间估计.....	(145)
第三节 假设检验原理及两种类型错误.....	(166)
第四节 单个总体参数的假设检验.....	(170)
第五节 两个正态总体参数的假设检验.....	(178)

第六节	非参数检验(一).....	(186)
第七节	非参数检验(二).....	(199)
第五章	观测值的方差分析.....	(234)
第一节	方差分析基础.....	(234)
第二节	单因素方差分析.....	(238)
第三节	两因素方差分析.....	(251)
第四节	三因素方差分析.....	(267)
第五节	方差分析中几个问题.....	(276)
第六章	环境科学中常用的统计分析、预测方法.....	(287)
第一节	一元线性回归.....	(287)
第二节	多元线性回归.....	(313)
第三节	逐步回归分析.....	(337)
第四节	时间序列预测法.....	(347)
第五节	马尔柯夫预测模型.....	(367)
第七章	抽样技术.....	(374)
第一节	抽样调查步骤、抽样类型和抽样误差的估计.....	(374)
第二节	简单随机抽样.....	(377)
第三节	机械抽样和多级抽样.....	(385)
第四节	分层抽样.....	(395)
第五节	某些环境问题抽样误差控制和样本容量的 确定.....	(408)
第八章	多元统计分析应用介绍.....	(413)
第一节	主成分分析.....	(413)
第二节	因子分析.....	(427)
第三节	聚类分析.....	(451)
第四节	判别分析.....	(473)
第五节	典型相关分析.....	(500)
第六节	数量化理论.....	(506)
附表1—19		(515)
参考文献		(566)

第一章 环境统计调查及数据整理

环境统计学也可称为环境统计方法，它是应用概率论和数理统计原理、方法对环境和实验资料从事调查、搜集、整理、分析，借以表现环境问题真相的学科，属于部门统计学。由于记录、观测的数量资料或属性资料是大量、复杂并且常是杂乱无章的。怎样根据统计学原理结合环境科学特点，科学、系统、有目的地去调查、搜集、整理、分析观测资料，并能将所获资料，所关心的环境问题以若干简单的数值、图表、公式表示其结果、性质、所研究对象的相互关系以及评价、预测环境状态，显然是环境科学工作者的重要任务。因此环境科学工作者或多或少地都应掌握环境统计方法。

第一节 环境统计指标体系及统计调查

一、环境统计指标

环境问题与人类活动的相互关系，相互影响的规律、原因和现象多是通过环境统计资料的研究、分析再以某种相互间数量关系变化来表达。由于环境科学涉及有天文、地理、数、理、化等诸学科并纵贯人类各种活动，所以环境统计包括的内容也是既多且广，如土地、能源、自然资源、环境污染情况及人类各类活动的记录、观测与统计，因此，环境科学是涉及自然科学和社会科学各领域的综合性学科。但在研究分析各种特定问题时，为反映问题本身特定性质，就必须从质的方面制定针对各类问题的指标。在一定指标体系中，从数量方面认识环境的状态及其发展趋势

势，根据环境指标实现程度进而提出管理、控制的措施。通常，环境统计指标是从数量方面反映环境总体现象。因而，人们首先要确定环境统计的指标及其有关概念。

环境统计指标是指反映客观存在的环境总体现象的数量特征概念和具体数值。它包括：指标名称，计量单位，^①计算方法，时空限制和指标数值。所以它既含有通常所指的统计指标概念又包括有统计指标数值，具体说明了总体数量大小、性质，反映了特定方面的环境问题的数量关系，起了对环境现象认识的指示器作用。当人们对环境问题作出各种决策或进行科学研究时，统计指标给出了基本根据，提供了用数量表示的事实。

环境统计指标可分为数量指标和质量指标。污染物种类、排放量等属于数量指标。它反映了环境受人为、自然或其它因素影响的规模及数量的大小。例如某化工厂每天排放 SO_2 、 NO_2 各为 10t、5t，这里排放物种类和数量即是数量指标。而废水处理率，污染物负荷量等则属于质量指标，它着重说明环境现象本身属性的量值。例如某造纸厂废水处理达地面三级水标准占废水排放量 50%，这里就是质量指标。当人们以环境要素来讨论统计指标，则有水体污染指标，如 BOD_5 、 COD 、氨氮等污染物含量的指标；土壤污染指标，如土壤中重金属含量，酸碱性，细菌含量等指标。大气污染指标则有大气中 SO_2 、 NO_2 ，固体微粒含量等指标。有时人们喜欢按社会要求划分来定指标，这时又常分为环境统计指标、农村环境统计指标等。按计量单位分，又可分为实物指标、价值指标等。在每种大的分类指标下都还有更小的分类指标——基本种类的统计分类指标。诸如污染治理统计指标，废水排放情况统计指标等。在具体制定环境统计指标时，依照问题性质，涉及范围来确定环境统计指标。即针对环境现象进行分析、判断、区分以确定特定环境问题的范围，以便于制定有关环境的策略和决策。例如，从 1985 年开始，经过两年半努力，全国首次工业污染源调查，涉及全国 29 个省、市、自治区（除台湾省外）的 2300 多个县，40 多个行业。具体制定调查内容、范围时，指标

就显出其重要性来。例如，调查之前应首先确定污染物种类，主要污染物指标，污染程度的排序等等。这些指标的确定使调查目的明确，为工业污染治理确定决策、策略及方式。

由于环境统计指标是以具体数字表达，因此结合调查方式还应制定出具体的计算、分析方法以保证环境统计的质量。这就是说，对各类指标取得应放在科学法则上，对每个污染物的具体指标计算都有其具体规定、换算。例如，废气排放中 SO_2 、 NO_2 或其它有害物质的计算，废水排放中汞、铬、镉等重金属及其它有毒物质指标的确定，它们的分析、计算方法等等在不同行业、不同排放物 and 不同排放方式就不一样。所有这些在有关环保手册或环境统计手册中均已给出，读者可参阅有关环保手册。

例1 某厂电镀车间每年用铬酸酐 8t，其中约 20% 的铬沉积在镀件上，约有 20% 的铬以铬酸雾的形式排入大气，约有 50% 的铬从废水中流失，其余的耗损在镀槽上，求每年从废水中排放多少六价铬？

解 铬酸酐分子式为 CrO_3 ，其中铬含量为

$$p = \frac{\text{Cr}}{\text{CrO}_3}$$

所以全年废水中排放出六价铬为

$$50\% \times 8 \times p = 4 \times \frac{\text{Cr}}{\text{CrO}_3} = 4 \times \frac{52}{52 + 16 \times 3} = 2.08\text{t}$$

例2 求在标准状态下（一大气压， 0°C ）40ppm 的氟化氢的 (mg/m^3) 浓度？

解 氟化氢分子量为 20，在常温下 1 摩尔体积为 22.4L，换算成 mg/m^3 的公式为

$$x = \frac{M \times C}{22.4}$$

这里 C 为污染物的 ppm 数值， M 为分子量，所以氟化氢浓度为

$$x = \frac{20 \times 40}{22.4} = 35.7 (\text{mg}/\text{m}^3)$$

二、环境统计指标体系

每个具体的环境统计指标只是反映复杂的环境现象或人为活动的某方面特征，能正确地对某地区环境状态或某个环境总体作出判断、预测，就应把相互有关的环境统计指标有机地结合起来。这些结合体习惯上称为环境统计指标体系。除单项指标外，各项环境统计指标多是通过一定的指标体系来起作用的。环境统计资料的调查、搜集、整理、分析是在一定的科学指标体系下完成的。

因环境内容的广泛性，所以环境统计指标体系可分成各种类型。在各种类型总体中又可分出子体系。某一类型总体是反映该类型环境统计现象的整体，而各子体系则反映了该类型的局部指标体系。它还可再分为更小体系以反映某类总体中各部分现象。例如：人类环境总体中有土壤环境统计指标体系、环境污染统计指标体系、能源环境统计指标体系、水环境统计指标体系等子体系。而在环境污染统计指标体系中，又有反映水污染与治理、噪声污染和治理、大气污染和治理等等子体系。象水污染和治理又可分为河流水污染和治理、湖泊水污染和治理等等更小的体系。由于环境统计是个刚发展不久的领域，许多指标体系还需要在人类对环境进一步认识中不断地扩大、充实、提高和完善。但总的来看，一个新的环境指标体系的建立，应遵循的原则是该体系的建立有利于环境污染的消除，自然生态系统恢复到良性的循环和人民生产、生活的环境是清洁、优美、符合人类健康生活要求，达到环境效益、经济效益和社会效益的统一，既有质量指标又有数量指标。把人类生产、生活活动的需要和可能结合起来，以达到环境保护的目的。

三、环境统计调查

环境统计调查可分为普查、抽样调查、重点调查和典型调查四种。

1. 普查 指为完成某种调查任务需要而专门组织的一次性调查。它是在一定时点或期间内进行，以求得比较全面的材料。例如自然资源普查，水文地质资源普查等。截至1988年初，我国水文地质普查已达700km²，这是在几年内完成的。象人口普查，时间就相对短得多，并且是在同一时点进行。因普查工作涉及面广，规模大，情况复杂，所以在动手普查前应做好充分准备，制定详细方案和统一组织规划，普查内容应重点突出。现在计算机处理普查已很普遍。因此在普查项目的编制上应结合整理分组的要求便于资料的汇总，资料的整理和计算机的数据处理。除了特殊情况，普查应选择标准时点同时进行，力争在短期内完成。

2. 抽样调查 在调查时从所研究的总体中抽取部分进行调查，并从其结果去推断总体的某个参数特征的非全面抽查方法。抽样调查可以在比较短时间内以比较少的人力、物力获得资料，从而提高了资料质量。对于那些因财力、技术等原因尚无法做到或没有必要进行普查的某些环境问题，抽样调查是经常采用的方法。它既可以由局部推算总体对全面资料进行检查和修正，还可对总体假设进行检验。

抽样调查一般是采取随机抽样调查。在环境统计中可分为单项指标抽样调查，多项指标抽样调查或全部指标抽样调查等。此外抽样调查样本点的设置，采用的技术等等有专门的研究。抽样调查在社会各行业中相当普及，现代抽样理论和抽样技术已得到很大发展。实际上，本书许多内容都是涉及到对样本资料的数据处理、统计推断、分析等。对于抽样技术后面章节将作一些简单介绍。

3. 典型调查 指在调查对象中选择具有典型意义的单位进

行调查。在我国现有条件下，许多环境问题常采取典型调查。解剖麻雀，了解全体。特别是对一些环境污染问题的认识常是从典型调查下手以便了解问题所在，说明环境污染状况，发展趋势，选择应采取的策略、措施等。自然，这种调查所获资料中误差较难确定，用它去推断总体其结论仅是近似值，并且有可能是误差比较大的近似值。

4. 重点调查 在调查总体中选择一部分重点单位进行抽查。它们在所研究的对象中占有很大比重或起了重大作用，掌握了这些重点单位也就掌握了基本情况。在工业污染的调查中，常是选取特定厂家大户，了解这些厂家排污，对环境影响情况，一些基本情况也就掌握了。

环境统计调查是环保工作的基础。通过对污染源有害物质的治理、排放、排污企业的能源、资源利用情况及工艺生产过程的调查，然后进行统计分析、研究和评价而提出污染治理，企业的技术改造及综合利用等规划、决策，从而对环境做出一定的评价和制订环境标准，制定环境管理规划策略。因而环境统计调查是环境科学研究和环境策略制定的重要基础。至于调查形式、内容、规模，依调查问题的内容、目的及人力、财力、可能性而采用上述四种形式中某一种或若干种相结合。由于调查任务不同，调查内容差异较大。例如环境背景调查和污染源调查就有很大不同。见表1.1.1，表1.1.2。

表 1.1.1 区域环境背景特征调查内容

项 目	内 容
地 理	调查区域地理位置、地形、地貌、地质构造、地层、矿物资源等
水文气象	温度变化、降水、蒸发、日照、能见度、云量、大气含尘量、风的情况；在丰水、枯水、平水年时河流水位、流量、流速、泥沙、水化学情况；河床断面，坡降，水温变化，湖泊水来源、去向变化情况等等

续表

项 目	内 容
土 壤	土壤理化特性、类型、土壤剖面构型等
生 物	水生、陆生生物种类,种群数量,分布状况和相互关系,生态特征等
社会经济 结构概况	村镇、城市分布及功能,人口,名胜古迹,工、农、林、牧、渔业结构、布局、产品种类,各行业人口情况,在区域中经济作用等

表 1.1.2 工业污染源调查内容

项 目	内 容
企业环境和 基本情况	地理位置、地形、环境功能区(如文化区、工业区、商业区等)环境状况、企业名称、规模、经济类型、主要产品、利润,职工人数、资产、动工年份、厂区绿化面积、厂区布局等
能源、水源、原材料、生产工艺、 生产发展情况	能源类型、产地、成份、消耗量,水源类型、供水方式、供水量、处理措施,原材料种类、产地成份、消耗量,工艺流程、技术指标、企业发展方向,布局、规模、经济指标等
排污和治理	污染物种类、排放量、排放去向、浓度、生产方式、治污方法、工艺、成本、效益、规划治理的项目方法,工艺、资金、预期效益等
污染危害	危害对象、程度、时间、原因,处理情况和企业内外群众反映等

第二节 环境统计资料的性质

一、数量属性和质量属性

在环境科学的研究和应用领域中,有许多监测、试验、调查所获得的资料需经统计方法处理后才能给出所研究、分析对象的环境问题信息。由于测试和调查的手段、工具、对象不同,所获得的资料有很大差异性。如有的是数据,有的则是文字描述或是实物、影象等。通常,这些资料按其性质、特性可分为两大类:

1. 数量属性 数量属性指调查、测试的客体物具有可以度量的性质。例如,环境污染范围(面积),被污染对象的种类(数量)等。在对客体物某些方面测试、调查的数据中,有的类型个体间数据存在很小差别。当总体足够大时,随着度量精度提高这一差别可达到人为可测的任意精度。习惯上人们即认为由这类客体物个性度量得的数据集合存在有连续性;反之,不具有这类特性的客体物个体测量出来数据集合认为具有间断性或不连续性。例如,环境中活的机体总是以自然数计算,尽管总体无限大,两类群的机体差数永远不会小于1的。

2. 质量属性 在环境研究中,有不少测量调查对象的一些属性虽然可以观察和记录,却无合适单位度量。例如污染物颜色、形状、不同气味比较等。这类特性一般看作客体物的质量属性。许多质量属性尽管是无法度量的量,但有时人们却非常需要赋以度量值。为了研究分析的方便,人们往往给以某种人为的度量特性。例如对不同的颜色赋予不同的数值,取红为-1,白为0,黑为1等。经过这种赋予质量属性数值的程序后,就可以用统计方法处理所研究、分析的环境问题,以便人们进一步弄清环境问题所在。这就是数量化方法中的一种。

现在,在环境科学中,新的概念、参数不断出现。例如环境

质量、心理影响、景观损害、污染的社会代价、资源资本及各种质量评价等。因为文明技术发展的限制,这些量虽然属极需较精确度量,目前却又无法做到,现在凭人为主观加以计算或专家主观的评分综合来定案,这无疑会在实践中出现失误。完善这方面度量理论,对环境科学研究、应用是很重要的。其原因是因环境问题中所获得信息,在一定程度上仅用数量信息表达出来,才能进行有效的定量研究分析,才能利用某些数学工具做统计处理,预测预报一些环境问题趋势、规律性等。

二、总体和样本

上面我们已经提过总体一词,在环境统计中,人们把研究对象具有某种可能结果的集合称为总体。更通俗些说总体指含有许多共同性质的个别物体的群体,它也称统计总体,也叫做母体。这里的个体即是构成总体的基本单元。例如,当研究人畜水银中毒情况时,被研究的人畜集合就是总体,被研究的个人、畜即为个体。总体分为有限总体和无限总体。在实际问题中大多认为总体是无限或近似于无限。在讨论问题时如没有特别说明,我们总是把总体当做无限总体来讨论。我们知道,环境统计因各个体自身的条件,所处环境差异,测量、观察工具、人的不同,和其它各种偶然因素,使各类度量值和性质的观察差异较大,各种数量资料、质量属性也均有一定的误差。当我们从总体中随机抽取一个个体来研究其特性时,个体的特性值往往无法预先予以确定,它依赖于被抽取的个体。所以该特性值具有很大的随机性或偶然性。在环境统计中,人们把这种依赖于抽样个体的特征数称为随机变数。在多数场合,尽管人们对总体中各个体的某种特性很感兴趣,但往往无法一一对个体进行度量、观察或监测。这时,人们只是以总体中抽取部分个体进行测试、观察,抽取这部分即叫做样本或子样。样本所含个体数目称为样本容量。而从总体中抽取样本或子样就叫做抽样。抽取的程序和方法叫做抽样技术或采样技术,

环境统计学主要内容是研究、分析从环境总体中抽取样本后怎样推断总体情况。一般来说，样本中各个体是相互独立的，而且样本应具有代表性。每个个体的度量值习惯上称为观测值。对所收集、记录的观测值、历史数据必须进行分析，以便弄清总体情况。对观测值作统计分析时包括两部分：统计叙述和统计推断。统计叙述可以说是统计资料显示。由于一般搜集来的数据比较分散，难于直接分析或与其它数据比较，统计叙述就是把所得观测数据进行整理、归纳，用若干统计量（样本特征数）和必要的图表把环境资料的主要特征显示出来以便于对环境状况加以分析、比较。统计推断是对样本资料依数理统计原理进行分析、解释并进而推断所研究的总体问题，它是本书叙述的重点。

第三节 展示资料的表格法

绝大多数环境科学的报告、论文、文献中都需用表、图来展示经过分析、研究、整理而得的各种调查、测试的信息资料。这些图表使人们尽快明了问题所在及研究分析的内容和结果。它还可使人们对资料做进一步分析、比较和利用。因此，我们首先须了解环境统计中常用的图、表结构和种类。这一节我们先讨论统计表的构成和常见的类型。

一、统计表的构成

所谓统计表系指经过调查、测试而得的资料直接或是经过分析、整理后以表格形式表示出来。好的统计表，使人一目了然，充分表达主题，即使事先不清楚内容的人一眼看过也能正确理解表的内容，统计表结构包括表头、表体及附注三部分。前二部分为各统计表所必须，附注部分有的统计表可以没有。

1. 表头 每个统计表上方称表头，它包括有标题。表不应无标题或是“词不达意”的标题。标题也不能使人费解甚至产生误解。

标题应简明扼要说明内容。在标题下有的还注明表中资料来源、地点、度量单位。它也属表头范围。如统计表较多还在统计表表头写明标号。

2. 表体 表体是统计表主要部分,是表的主体。它由纵横线分割成不同区块。标目,每一项目的名称,必需的单位及数字等填写其中。如有的标目可用符号说明问题,可直接用符号表示。习惯上统计主语即主要标目总是放在表中最左一列或最上一行,统计谓语即分项指标名称放在最上一行或最左一列。表中应写明分项指标,有的还应在项目栏中注明单位。但表中数字通常不注明单位。至于合计之类栏目放在最右列或最下行。表中填写的观测值或分析计算所得数据资料属同一性质的精确度,应相同。上下位数据资料属同一性质的精确度也应相同,且上下位数应对齐。若数值太大或太小可用指数表示,没有资料的应用破折号“——”或省略号“……”表示。为便于阅读、理解,若无必要,统计表内容不应太庞杂。太多的内容可分别制成几个表。

3. 附注 当标题或表中某些项目须做必要说明时则应在表底给予必要的附注说明。

二、几种常见的环境统计表

1. 原始资料展示表 这种表一般是先编定栏目,把在环境研究领域实际观测值依实际情况记录。最常见的是依日期、地点顺序填表(见表1.3.1)。

2. 依次表 有的原始资料统计表数据多时,显得杂乱(见表1.3.2),无法使人很快看出数据分散程度,差异的大小。为此,常有人把原始数值按大小顺序排列。排列时数值可以由大到小,也可以由小到大。这样处理后的统计表称依次表。人们从表上数字的集中和分散程度,对某些环境统计参数就有大概的了解。依次表1.3.3的数据是由表1.3.2整理而得。

由表1.3.3立即可看出监测值大小变化和集中、分散程度,由