

SHIYONG QIXIANG
TONGJI FANGFA

实用气象统计方法

—◎ 李湘阁 胡凝 等 编著

实用气象统计方法

李湘阁 胡 凝 等 编著



内容简介

《实用气象统计方法》的主要内容包括:气象数据及其预处理,气象领域常用的统计推断方法,气象科研与业务工作涉及的相关分析、类型分析、过程分析和场分析等一些最基本的定量分析方法。此外,还介绍了模糊数学、灰色系统论、信息论的一些分析方法,有益于读者扩大视野和知识面,增强创新能力。本书实用性强,以基本方法为主兼顾理论,各种方法均有实例,通俗易懂,便于自学。对于气象科技等文献中常见的统计方面的错误给予特别提示。各章节设置了习题,便于读者思考和练习巩固所学的知识。在附录中简要地给出概率论最基本的概念、定义和定理,并列出常用的统计表格,会给读者学习、应用带来更多的方便。

本书可作为大气科学类及其相关专业本科的教材。从统计方法应用的角度看,不仅对广大气象工作者,而且对其他学科或专业人员分析、处理有关数据、组建统计预测模型也有一定的参考价值。

图书在版编目(CIP)数据

实用气象统计方法/李湘阁等编著. —北京:
气象出版社, 2015. 9

ISBN 978-7-5029-6236-4

I. ①实… II. ①李… III. ①气象
资料—统计方法 IV. ①P468.0

中国版本图书馆 CIP 数据核字(2015)第 219607 号

Shiyong Qixiang Tongji Fangfa
实用气象统计方法

出版发行: 气象出版社

地 址: 北京市海淀区中关村南大街 46 号

总 编 室: 010-68407112

网 址: www.qxcbs.com

责任编辑: 马 可 黄红丽

封面设计: 博雅思企划

印 刷: 三河市鑫利来印装有限公司

开 本: 720 mm×960 mm 1/16

字 数: 466 千字

版 次: 2015 年 9 月第 1 版

定 价: 54.00 元

邮政编码: 100081

发 行 部: 010-68409198

E-mail: qxcbs@cma.gov.cn

终 审: 章澄昌

责任技编: 赵相宁

印 张: 24

印 次: 2015 年 9 月第 1 次印刷

前 言

《实用气象统计方法》是为大气科学及其相关专业本科教学编写的一本教材。主要内容包括了四个方面:统计学方法在气象领域中的任务和作用;气象数据的初步加工与处理;统计推断理论与方法在气象方面的应用;气象科研与业务工作涉及的定量分析方法,如相关分析、类型分析、过程分析和场分析等。此外,从实用出发,还引进了一些严格说来并非属于统计学范畴的模糊数学、灰色系统论、信息论等分析方法,目的是扩大读者的视野和知识面,增强创新能力以适应实际工作的需要。若从统计方法应用的角度看,这本教材不仅对广大气象工作者,而且对其他学科或专业人员分析、处理有关的数据也有一定的参考价值。

本教材是在多年教学实践和科研基础上编写的,以数理统计理论体系为框架,尽量涵盖气象领域应用到的一些统计学方法,同时关注统计理论与方法的新进展,适当引入了一些新的思路。然而,本书并不刻意追求介绍新方法,因为统计理论和方法在实际应用中还在不断地向前发展,可以说是没有止境的,一本教科书不可能包容一切。重要的是能让读者有较扎实的基础,以便未来能够更好地吸收、理解新的知识,学会应用。为此,编写过程中力图做到方法与其理论兼顾,立足于让读者掌握最基本的统计方法,同时了解其基本原理。对于气象科技等文献中,在统计方面常见的错误或毛病,以及学生们不易掌握、容易产生误解的概念与原理,在阐述时给予重点强调和特别提示,这也许对读者是有益的。教材中的文字论述部分,理论与实例相结合,尽量做到深入浅出、通俗易懂,以便于读者自学,克服学习过程中感到抽象难懂的困难。

另外,教材中设置了部分习题供学生思考和练习,以便加深和巩固所学的知识。同时,在附录中简要地给出了一些概率论最基本的概念、定义和定理,提供了常用的一些表格,试图给读者学习、应用带来更多的方便。

本书主要由李湘阁负责策划、设计、编写以及最终修改定稿,胡凝参与了大部分工作和图表制作,黄红丽执笔了第2,4章,杨沈斌提供了小波分析的实例。本书作为讲义试用,得到滨江学院申双和、耿焕同教授很大的帮助;书稿承蒙缪启龙、门可佩教授审阅并提出宝贵意见;有关专家参与了编写大纲及初稿的审定;顾显跃同志帮助排版设计。在此均表示衷心感谢。对于本书的出版,还要特别感谢张永宏教授和南京信息工程大学教材基金的鼎力支持。

由于本人学识水平有限,书中不当之处在所难免,恳请读者指正。

编著者

2015年6月

目 录

前 言

第 1 章 绪 论	(1)
1.1 气象统计方法的重要性	(2)
1.2 气象统计方法的理论基础和基本内容	(3)
1.3 气象统计方法的基本任务	(4)
习题	(5)
第 2 章 气象数据的初步整理与加工	(6)
2.1 气象数据及其来源	(6)
2.2 气象数据的审查	(7)
2.3 气象数据序列订正	(11)
2.4 气象数据的特征量统计	(29)
2.5 气象数据的初步加工	(43)
习题	(47)
第 3 章 气象数据的统计推断	(49)
3.1 气象数据特征量的点估计和区间估计	(49)
3.2 假设检验	(53)
习题	(84)
第 4 章 气象科学试验结果的方差分析	(88)
4.1 一元方差分析的原理及其应用	(88)
4.2 多元方差分析的原理及其应用	(93)
4.3 正交试验的方差分析	(103)
4.4 使用方差分析方法应注意的问题	(104)
习题	(107)
第 5 章 气象领域中的相关分析方法	(109)
5.1 相关系数	(109)
5.2 回归分析	(114)
5.3 灰色系统分析	(169)
习题	(181)

第 6 章 气象领域中的类型分析	(183)
6.1 判别分析	(183)
6.2 聚类分析	(207)
习题	(231)
第 7 章 气象领域中的过程分析	(232)
7.1 随机过程的基本概念	(232)
7.2 气象要素的趋势变化分析	(238)
7.3 气象要素的阶段变化分析	(239)
7.4 气象要素的周期变化分析	(241)
7.5 气象要素的平稳时间序列分析	(262)
7.6 天气气候等转折性变化的马尔可夫链分析	(269)
习题	(274)
第 8 章 气象要素场分析	(276)
8.1 气象要素场概述	(276)
8.2 气象要素场的自然正交函数展开	(277)
8.3 气象要素场的切比雪夫多项式展开	(287)
习题	(296)
第 9 章 气象信息分析	(297)
9.1 熵及其在气象领域中的应用	(297)
9.2 信息及其在气象领域中的应用	(304)
习题	(312)
参考文献	(313)
附录 A 概率论基础知识备忘录	(314)
A1 随机事件及其概率	(314)
A2 随机变量及其概率分布	(317)
A3 随机变量的数字特征	(322)
附录 B 常用统计表	(325)
附表 B.1 二项分布表	(325)
附表 B.2 正态分布的密度函数表	(329)
附表 B.3 正态分布表	(330)
附表 B.4 正态分布双侧置信限表	(333)
附表 B.5 泊松分布表	(333)
附表 B.6 χ^2 分布置信限表	(341)

附表 B. 7	t 分布双侧置信限表	(342)
附表 B. 8	F 分布置信限表	(344)
附表 B. 9	检验相关系数 $\rho=0$ 的临界值表	(351)
附表 B. 10	检验复相关系数 R 的临界值表	(352)
附表 B. 11	符号检验表	(353)
附表 B. 12	秩和检验表	(355)
附表 B. 13	游程检验表	(355)
附表 B. 14	正交多项式表	(359)
附表 B. 15	“野点子”临界值表	(366)
附表 B. 16	$-P\log_2 P$ 表	(367)

第1章 绪 论

统计学是研究各种数据的收集、分析和推断方法的科学。它涉及社会科学和自然科学诸多领域,应用十分广泛。统计学的实质是,通过对研究对象及其相关因素的大量观测、试验结果进行数学分析,寻求事物发展变化规律,以便解释事物发展变化的原因并做出相应的未来变化的预测和预报。气象统计方法是统计学方法在大气科学领域中的具体应用。它研究的是与气象有关的数据的收集、分析和推断。主要用来定量研究和描述大气科学研究对象本身的变化及其与相关因素之间的关系。

对于气象工作者来说,无论从事试验、观测、科学研究,还是从事各种日常的业务服务工作,总是要和与气象有关的数据打交道。因此,气象统计方法是不可缺少的重要工具。

应用统计学理论和方法,分析、解决自然科学和社会科学问题的事例枚不胜数。就自然科学而言,早在19世纪,比利时的A. Quetelet就把统计学理论和方法引入生物科学研究中。英国的F. Galton应用统计学方法分析了生物变异现象,并在1888年提出了相关系数的概念,他曾试图用相关分析方法探讨遗传定律,虽未成功但所用方法却奠定了生物统计学的基础。19世纪70年代,曾有人利用统计学方法分析糖类作物与雨量的关系,建立了预测糖类作物产量的模式。20世纪初,印度也曾进行过许多天气气候与作物产量关系的统计研究。英国的R. A. Fisher则对统计学做出了重大贡献,1924年他在研究雨量对小麦产量影响时提出了积分回归的概念和方法,至今仍很有意义。1925年,他的《实验设计》一书,阐述了实验设计理论和方差分析原理,确立了统计推断的基本方法。在我国,统计学方法用于气象学科也是比较早的,最早涉足的是关于气象观测资料的整编、各种气象要素指标的统计以及天气气候特征、气象灾害规律等方面的统计分析。20世纪60年代,中国科学院大气物理研究所的有关人员到气象台站参加预报实践,总结经验并推广统计学方法,使整个气象系统形成了学习使用气象统计方法进行科学研究、开展气象预报的热潮。由此,通过各种途径提高了气象工作人员的素质与水平,进一步为气象统计方法的广泛深入应用打下了厚实的基础。与此同时,我国的气候统计学也有了长足地发展。著名的气候学家、气候统计学权威么枕生先生,曾经在《现代气候统计诊断与预测技术》(魏凤英,1999)一书的序言中提到:我国的气候统计自20世纪60年代以来,一直站在世界气候学术研究的前沿,气候统计方面的研究成果,不管是数量还是内容也都位于国际学术界的前列。目前,随着数理统计理论和方法的发展、计算机技术的突飞猛进以及电

子计算机的普及,现代气象统计方法早已突破了原来对气象观测资料统计指标简单计算和分析的局限,产生了飞跃。这种飞跃不仅体现在统计技术内容的深度和广度上,而且体现在研究问题的新思维和新观点上。应用统计方法研究和处理大气科学有关问题的范围更广阔了,统计方法已渗透到监测、诊断及预测等各个方面,更涉及大气科学领域中气候、应用气象等分支学科的许多问题。

1.1 气象统计方法的重要性

众所周知,天气气候变化是相当复杂的,其中各种气象要素不仅本身在不停地变化,而且相互之间既有影响,又有制约。另外,天气气候变化不仅与气象要素有关,还与大量非气象要素(如海温、太阳活动、地理位置、地形地貌、植被、土壤、人类活动等)有关。大气科学中的有关分支学科,如卫星气象学、雷达气象学、应用气象学(包括农业气象、城市气象、生物气象、军事气象、建筑气象、医疗气象)等,也涉及许许多多非气象要素,它们的变化又是十分复杂的。因此,我们面对其中任何一种复杂的研究对象,要想科学定量地分析、描述它,就必须建立一定的数学表达式,或称为数学模式(模型)。

组建数学模式,大体上又分为三种类型,它们分别由三种观点决定。一种观点认为,研究对象,如大气运动规律等是不确定的、随机的,它受许多不确定因素的影响。要完全从理论出发,准确写出全部运动方程是不容易的。退一步说,即使有了一组方程,要给出所有的初始条件并求解,也是十分困难的。但只要有足够多的实验观测资料,利用概率统计原理和方法建立统计学模式(又称为经验模式),透过现象看本质,是可以找到规律并进行预测预报的。另一种观点认为,大气运动规律等研究对象的变化是确定的、非随机的。它是一些变量的函数。依据热力学、流体力学等原理,总可以建立相应的数学模式(即通常所说的数值模拟模式、动力学模式或理论模式),对其运动规律进行细致的数学描写。只要在给出初始条件的情况下输入相关数据,通过电子计算机运算得出合理的结果,即可做出预测预报。某些持这种观点的人认为,建立在概率论基础上的统计学模式缺乏物理内容和应有的理论依据,多少有些盲目性。同时,经验统计模式也根本无法预报出历史资料中未出现过的情况。显然,这两种观点是相互对立、互斥的,都有正确的一面,但又不尽完善。第三种观点是上述两种观点的结合,即在组建模式过程中充分考虑到研究对象形成的机制和理论,同时又把初始状态和未来可能状态看作是随机的。这种模式被称作动力—统计模式,或半经验—半理论模式。

不管怎样,三种定量分析的数学模式都缺少不了统计学方法。即使是第二种纯理论的数值模拟模式,其中某些参数也往往要通过实际观测资料用统计学方法来确

定。因为要想模式中的所有变量参数都从理论上获得,是不现实的,至少在当前或未来一段时间内,受科学技术发展进程等方面的制约,是不现实的。综上所述,气象统计方法的重要性是不言而喻的,它在我们进行试验观测、科学研究和业务服务中的地位是不可动摇的。学习、掌握它是有实际意义的。

当然,气象统计方法本身并非完美无缺,它还存在着一些缺欠,使用时有一定的局限性。例如,①它往往要求有足够大的样本容量,即要求有足够多的在相同条件下取得的观测数据。只有这样,统计得到的一些规律或关系才比较稳定可靠。然而,事实上所需要的观测数据常常是有限的。②有些气象统计方法,要求观测数据具有某种统计特性,如必须服从正态分布、时间序列必须平稳化,必须具有各态历经性等,而这些要求有时是难以满足的。③各种气象统计方法所得到的结果都有一定的随机误差,而此误差无论如何也是无法避免的。④各种气象统计方法所得到的结果,都受到建立统计模式时所用观测资料取值范围的限制。换句话说,统计学方法无法预报出历史上未出现过的现象或数值。⑤经统计得到的某种关系,说到底只是大量观测数据所体现出来的一种表面上的联系,而对其内在的因果关系无法做出适当的解释。有时所得到的某种关系可能就是一种谬误。如利用周期分析方法得到的数学模式,若不考虑其他因素变化的影响,仅按时间进行无限期的推延预报,实际上就是荒唐可笑的。又如有的相关分析模型得到的相关关系可能是一种假象,或根本就不存在,或与理论相矛盾。这些问题在我们应用气象统计方法时必须十分注意。

特别强调 气象统计方法重要且不可少,但它也有欠缺和局限性。

1.2 气象统计方法的理论基础和基本内容

气象统计的各种具体方法,都是以概率论与数理统计的有关理论作为基础的。实质上都是把大气科学中的某种研究对象当作随机事件或随机变量看待的。各种方法之间的区别仅在于,根据不同的目的要求进行不同的数学处理。气象统计方法的基本内容可概括为统计分析和统计推断两大部分。统计分析是最重要的内容,主要指对大气科学研究对象的概率、概率分布及其特征量进行统计,或进行相关分析(包括相关性质、相关程度及各种相关关系数学表达式的建立等)、类型分析、随机过程和随机场分析以及试验设计等。统计推断主要是指,对由样本资料得到的关于总体的概率、概率分布及某些特征量和相应的定量关系进行估计、判断,看它们是否能真正代表总体,包括估计和假设检验两个方面。

学习、掌握气象统计方法,除了要有比较扎实的概率统计知识外,还必须有一定的高等数学和线性代数知识作为基础。另外,还必须具有应用计算机的能力,会编程或会使用统计软件。

本书引进了一些模糊数学、灰色系统论、信息论等方面的内容与方法。虽然,严格地讲,这些并不属于统计学范畴,论述上也不够系统、不甚严谨,但为解决大气科学中常见的实际问题提供了新的思路,扩大了视野,可供读者学习参考。

1.3 气象统计方法的基本任务

1.3.1 初步加工整理有关数据

大气科学研究对象及其相关因素的出现大多都有偶然性,是否出现并不确定。所有气象要素与气象条件抑或相关的非气象要素和条件,每次的观测值时大时小,看起来似乎是杂乱无章的,但它们绝不是没有规律可循的。为了便于应用和进一步分析,需要对大量原始的观测数据进行初步的加工整理,以显示出某些特征。例如,数据可靠性的审查,可疑资料的处理,数据序列的订正与延长,数据中心化、标准化、归一化与平稳化等处理,特定的指标值(如稳定通过某界限温度的初始或终止日期、积温等)和重要的特征量(如平均值、累积值、标准差、变异系数等)的统计,这些都是气象统计方法的任务。

把经过初步加工整理的气象数据建成便于利用的数据库,可按用户的要求分门别类地提供服务,其效果比仅提供原始资料要好得多。同时,这种经过初加工的数据也为气象科学研究,组建各种气象模式带来方便。

1.3.2 定量分析气象变化规律

根据初步加工处理过的气象数据资料,利用各种统计手段,分析大气科学研究对象随时间、空间的变化规律,确定其相互间的定量关系,建立相应的经验(统计)模式是气象统计方法的主要任务之一。对于气象科学研究和开展业务服务来说,这是从定性分析向定量分析转变,提高工作质量所不可缺少的一步。

1.3.3 判断由有限数据统计得出某种规律或关系是否符合实际情况

由于气象观测及其相关的数据是有限的,也即样本容量是有限的,抽样存在着随机性,因而依据这些样本资料统计出的某些特征、关系或规律是否能代表总体的真实情况,需要进行判断。利用统计推断原理和方法,判明所得到的统计规律或关系的可靠程度;区别样本特征和总体特征之间的差异是本质的还是由偶然因素造成的;分析引起事物即研究对象变化的各种因素作用的大小等。这些由样本资料估计或推断总体的工作,也是气象统计方法的一项重要任务。

1.3.4 研究大气科学试验和取样的科学方法

为了最大限度地降低试验误差,在提取所需信息的过程中,充分减少试验的工作量和尽量缩短试验所需时间,根据问题的要求从统计学角度出发寻求最优的试验设计、取样方式和数学分析方法等,同样也是气象统计方法的任务。

习 题

- (1)什么是气象统计方法?
- (2)为什么要学习气象统计方法?
- (3)气象统计方法的基本内容有哪些?
- (4)气象统计方法有哪些缺点和局限性?

第2章 气象数据的初步整理与加工

2.1 气象数据及其来源

2.1.1 气象数据

大气科学的研究对象主要是天气气候本身的变化及其与有关方面的相互作用和影响。关于研究对象的所有资料可以统称为气象资料,其中包括各种气象要素(如气温、气压、湿度、风向风速、日照、太阳辐射、降水等)和天气气候现象及其灾害(如旱涝、冰雹、低温、霜冻、风暴、雷电等)方面的资料,以及影响天气气候变化或受天气气候变化影响的种种非气象要素方面的资料。前者作为气象资料是容易理解和接受的,而后者与大气科学研究对象有密切关系的各种非气象要素方面的资料(如海水温度、太阳活动、地理信息、受灾面积、灾害损失、卫星遥感和气象雷达探测数据以及农作物生长发育状况与产量等)实际上也可看作是广义气象资料的组成部分。

气象资料通常以三种形式存在:一是文字记载,如对某时期气象条件和状况的描述、关于气象灾害的调查记录或报告等;二是数量记载,如各种气象要素的观测数值等;三是图像资料,如遥感和雷达图像等。所谓的气象数据,一般是指以数量表示的气象资料。上述以文字或图像形式存在的气象资料,只有在经过一定处理转化为数据后才能为气象统计方法所利用。

特别注意 气象统计方法面对的气象数据是广义的,是与研究对象有关的那些数据。

2.1.2 气象数据的来源

气象数据主要来源于几个方面:一是来自按统一的观测规范进行的正规气象观测,包括气象台站的地面观测、高空探测、气象雷达探测以及农业气象观测等,这些观测结果是气象数据的主要构成部分;二是来自短期气象调查、考察和为某种目的而设的自动气象站观测以及海洋船只航行中的气象观测等;三是来自特定的气象科学实验观测及相关的科学研究,如小气候观测和农业气象的平行观测等;四是来自民航、水文、海洋、农业科研等部门的一些气象观测,如风向风速、云、能见度、天气现象、降水、蒸发、径流、海温等;五是来自航空与人造卫星遥感观测;六是来自史书、史料等文

献的记载,如关于旱涝、冰雹、霜冻、风雪和病虫害等灾害状况的历史记载。

气象数据来源的多样性使得气象数据丰富多彩,既给利用气象统计方法进行分析研究带来方便,同时也增加了处理气象数据的难度。

2.2 气象数据的审查

2.2.1 审查的目的和意义

在研究实际的气象问题时,面对来源各异的气象数据,都有一个去粗取精、去伪存真的问题。为了检查、提高所用气象数据的质量,在使用之前需要进行严格、科学的审查。审查的主要内容包括以下四个方面。

(1) 准确性

准确性审查,即对现有的气象数据能否真实地反映客观实际情况进行审查。对于来自国家气象台站网正规的气象观测、农业气象观测数据,应了解有没有可能因观测仪器的故障或观测人员的失职失误而产生了误差。对于非正规的气象观测,如特定的科学试验、考察项目以及有关部门观测到的气象数据,要特别注意其所使用仪器的精度,观测记录的方法是否正确,以便评定其准确的程度。对于调查所得的间接的数据或历史文献中记录的气象数据,应设法考证或验证其可靠性。对于通过遥测遥感或其他技术手段测得的,经过换算而成的气象数据,也要从原理上考查其准确性、合理性。总之,在使用这些气象数据进行分析之前,首先要对它们的准确性有所了解。若准确性存在问题,那么所有的一切工作都会是无意义的。

(2) 均一性

均一性审查,即审查现有的气象数据其观测条件是否一致,气象数据序列的变化是否真正反映了气象要素或条件本身的变化。对气象数据中出现的个别异常值,或某一时段的数值明显偏高或偏低现象,均应查清原因。若属于观测地点、时间、方法改变,观测仪器和人员变动或其他技术等方面的原因造成数据的不均一性,则要进行适当的订正处理。

(3) 代表性

代表性审查,即针对所研究的气象问题和对象,审查所用的气象数据在观测的地理位置、观测的时间等方面有无代表性,能否用来解决面临的问题。如研究农业气象或山地气候问题,不仅要求气象数据具有准确性和均一性,同时还要有代表性。若气象数据多来自海拔高度较低的山谷、河滩地上的观测场,那么它就缺乏代表性,很难用来分析海拔高度较高地方的具体气象问题。

(4) 比较性

当研究气象要素或气象条件在地域上或时间上的差异时,就要求气象数据具有比较性。如要求气象数据的样本容量相同,观测的年代和时期一致,使用的单位、量纲及统计方法一样等。

一般来自气象台站网观测和整编的气象数据,都是已经过审查的,可以比较放心地使用。而来自其他方面的气象数据,使用前是需要认真对待的。

2.2.2 审查的依据和方法

气象数据的审查应以大气科学及其相关学科的基本原理和气象观测规范等为依据。着重考虑测站(点)位置和周围环境的变迁;观测时间、次数、方法及仪器的变化;统计方法的变更等是否影响到气象数据的准确性、均一性、代表性和比较性。审查的方法主要有以下四种。

(1) 核对核算法

通过核对和核算找出记录、抄录或初步整理计算中的错误。如某次气温观测记录为 20.1°C ,而在计算日平均气温时误认为 20.7°C ;某地 4 月份冰雹灾害实际出现 16 次,统计抄录时却记为 10 次等。这类问题比较简单,经核对和核算很容易发现。

(2) 变化规律分析法

根据大气科学不同的研究对象本身的变化规律,判断气象数据是否合理。例如,某日 14 时气温比当日最低气温还低,就可能出错了;若一个月中,降水日数与非降水日数之和大于 31 天,肯定也有错误;又如,某次农业气象观测中,水稻植株高度为 12 cm,而 10 天前的观测记录为 14 cm,出现倒退现象显然是有问题的。

(3) 相关分析法

根据大气科学不同的研究对象相互之间的关系,审查气象数据是否正确合理。如研究干旱程度总是离不开土壤湿度的,而土壤湿度又受降水量、灌溉量、蒸发量及风速等要素的制约和影响。若某一时段降水量和灌溉量没有明显增加,蒸发量和风速也没有多少变化,土壤湿度却增加了不少,就可能存在问题;又如,在农田小气候观测中,小麦植株间离地面 25 cm 处风速是 1.6 m/s ,而冠层 1.5 m 处的风速却为 0.9 m/s ,这肯定是有问题的;再如,南京与附近合肥、扬州、镇江三站的日最低气温(或最高气温)之间是有一定关联的,若南京的记录比周围偏高或偏低很多,也就值得怀疑了。

(4) 时空分布规律分析法

任何大气科学的研究对象,在时间和空间的分布上都是有一定规律的。根据这些规律可以审查其观测数据的均一性。通常有两种形式:时间比审和地区比审。

时间比审指的是,审查同一地点某气象要素在不同时期的变化是否稳定。若存在显著差异,则说明其均一性遭到了破坏。例如,某地春季日平均气温稳定通过

10℃的初日,1969 年以前均在 3 月 10 日之前,而 1970 年之后均出现在 3 月 15 日以后。如表 2.1 所示,显然前后存在着不均一的特点。究其原因是 1970 年气象站迁址,地理环境前后大不相同所致。

表 2.1 某地日平均气温稳定通过 10℃的初日

年份	1964	1965	1966	1967	1968	1969	1970	1971
月-日	3-7	3-6	3-8	3-9	3-2	3-8	3-15	3-16
年	1972	1973	1974	1975	1976	1977	1978	1979
月-日	3-16	4-1	3-17	3-20	3-22	3-17	3-19	3-15

地区比审指的是,由于相邻两测站同一气象要素同一时刻观测值的差值或比值是相对稳定的,它们的变化要比该气象要素本身的变化小得多。因此,若发现此气象要素两相邻测站之间差值或比值有较大变化时,可以认定其均一性出现了问题。例如,沈阳与抚顺的年平均气温差值,1935 年以前为负值,以后则为正值(如表 2.2 所示)。经审查结果发现,抚顺年平均气温 1935 年以前比沈阳偏高,是由于计算过程中其日平均气温以每日 10 时 1 次观测值代替造成的。实际上抚顺所在的纬度比沈阳高,一般来说,年平均气温应比沈阳要稍低一些才对。同样道理,对于降水量等那些变化较大、不连续的气象要素可用相邻两站对应的比值进行均一性审查。

表 2.2 沈阳与抚顺年平均气温(℃)资料

年份	1929	1930	1931	1932	1933	1934
沈阳	7.5	7.7	7.0	7.8	6.7	6.9
抚顺	9.1	10.1	9.4	10.0	9.1	9.2
差值	-1.6	-2.4	-2.4	-2.2	-2.4	-2.3
年份	1935	1936	1937	1938	1939	1940
沈阳	8.5	6.5	7.9	8.1	8.8	8.0
抚顺	10.5	5.8	7.1	7.4	8.5	7.6
差值	-2.0	0.7	0.8	0.7	0.3	0.4

此外,也可以利用要素的空间分布图、时间变化曲线图、相关点聚图等工具进行比审,以发现那些与正常变化规律不同的数据资料。

(5)“野点子”的识别

从统计学角度看,气象数据序列就是一个随机样本。例如,上海的历年年降水量、南京历年 7 月平均气温、某地 1949 年以来历年大豆平均单产等,都可看作随机样本。它们有时会出现个别数据明显偏大或偏小的情况,若样本中某个过大或过小的数据(个体)是因为某种无法控制的因素引起的,则这个数据(个体)就称为“野点子”。

一般在相关点聚图中比较容易发现,因为它往往远离群体。很有可能这个“野点子”和其他数据根本就不属于同一个总体。这种过大或过小的数据混在样本中,将破坏样本反映总体性质的客观性,尤其是当样本容量较小时更是如此。

那么,如何识别“野点子”呢?不难理解样本中的最大值或最小值是首先最可能成为“野点子”的。若它们不是“野点子”,则其他的数据(个体)肯定也不会是“野点子”。关于识别“野点子”的原理和思路,涉及统计假设检验,我们将在第3章予以讨论。这里只介绍识别“野点子”的具体方法和步骤。待学习过假设检验之后,就比较容易理解识别“野点子”为什么要这样做。

第一步 把样本容量为 n 的观测数据由小到大顺序排列,如

$$x_1, x_2, \dots, x_{n-2}, x_{n-1}, x_n$$

式中 x_1 和 x_n 分别为样本中的最小值和最大值。例如,某地历年7月份降水量(mm)观测结果分别为:52.1, 62.3, 40.5, 204.3, 71.2, 65.0, 23.1, 34.6, 45.2, 31.2, 15.3, 36.7, 和 45.3。将它们从小到大排列为:15.3, 23.1, 31.2, 34.6, 36.7, 40.5, 45.2, 45.3, 52.1, 62.3, 65.0, 71.2, 204.3。其中最小值 x_1 为 15.3, 最大值 x_{13} 为 204.3。

第二步 假设 x_1 和 x_{13} 并非“野点子”。

第三步 根据样本容量 n 计算有关的统计量 r_{ij} , 见表 2.3。

表 2.3 检验“野点子”的统计量

n 的范围	检验 x_1	检验 x_n
$3 \leq n \leq 7$	$r_{10} = \frac{x_2 - x_1}{x_n - x_1}$	$r_{10} = \frac{x_n - x_{n-1}}{x_n - x_1}$
$8 \leq n \leq 10$	$r_{11} = \frac{x_2 - x_1}{x_{n-1} - x_1}$	$r_{11} = \frac{x_n - x_{n-1}}{x_n - x_2}$
$11 \leq n \leq 13$	$r_{21} = \frac{x_3 - x_1}{x_{n-1} - x_1}$	$r_{21} = \frac{x_n - x_{n-2}}{x_n - x_2}$
$14 \leq n \leq 25$	$r_{22} = \frac{x_3 - x_1}{x_{n-2} - x_1}$	$r_{22} = \frac{x_n - x_{n-2}}{x_n - x_3}$

在本例中 $n=13$, 因此应分别计算检验最小值 $x_1=15.3$ 和最大值 $x_{13}=204.3$ 的两个 r_{21} , 即

$$r_{21} = \frac{x_3 - x_1}{x_{n-1} - x_1} = \frac{31.2 - 15.3}{71.2 - 15.3} = 0.2844$$

$$r_{21} = \frac{x_n - x_{n-2}}{x_n - x_2} = \frac{204.3 - 65.0}{204.3 - 23.1} = 0.7688$$

第四步 查附表 B. 15, 由已知的 n 和给定的信度 α (0.05 或 0.01) 可得统计量 r_{ij} 的临界值 $r_{ij, \alpha}$ 。在本例中, 由 $n=13$ 和 $\alpha=0.05$ 查得检验 r_{21} 的临界值 $r_{21, 0.05}$