气象统计分析与预报方法

黄嘉佑 编著

气象出版社

气象统计分析与预报方法

黄嘉佑　　编著

气象出版社

内 容 简 介

本书主要介绍气象学中有关天气统计分析与预报方面的基本理论及计算方法，系统地阐述了目前国内外常用的有关方法，如多元分析中的回归分析、主分量分析、因子分析，判别分析、聚类分析及时间序列分析中的自回归滑动平均模型、谱分析及马尔可夫概型分析等。本书着重讲授这些方法的基本原理、计算步骤以及它们在天气分析及动力预报中的应用。

本书经国家气象局高等学校气象类教材编审领导小组审查，确认为作为大学本科通用教材。此外也可作为大专院校有关专业教学参考书，对气象业务人员也有参考价值。

气象统计分析与预报方法

黄嘉佑 编著

责任编辑 黄丽荣

*

气象出版社 出版

（北京西郊白石桥路46号）

北京昌平环球科技印刷厂印刷

新华书店北京发行所发行　全国各地新华书店经售

*

开本：850×1168 1/32　印张：12.5　字数：319 千字

1990年4月第一版　1990年4月第一次印刷

印数：1—2500　定价：2.60 元

ISBN 7-5029-0316-X/P·0176（课）

前　言

近三十年来，由于电子计算机的广泛使用，使得大量资料有可能得到迅速而有效的处理，近代统计方法亦随之迅速发展，这些方法也相继引入到气象学中，使用它们可以及时处理和分析来自地面、高空、甚至由卫星、雷达等先进大气探测工具所得到的大量资料，使得气象中的天气分析和预报也得到很快的发展。例如在传统天气学基础上发展起来的天气气候学、动力气候学，以及在传统动力学及数值预报基础上发展起来的统计动力预报，已成为现代气象分析与预报的新内容。事实上，大气运动不仅具有随机性的特征，也具有确定性的运动规律，只有把这两者有机地结合进行分析研究，才能使大气运动得到全面深刻的认识。因此，天气-统计以及动力-统计的结合是近年来气象分析预报发展的必然趋势。

在气象分析与预报的业务中，多元分析以及时间序列分析等统计方法使用十分广泛，其中一些基本方法已成为气象预报的基本方法。本书主要介绍涉及天气统计分析和预报方面的近代统计方法。这些方法除了在这个领域上广泛使用外，还在大气探测、大气湍流、大气污染、人工控制天气、雷达气象、卫星气象、农业气象、以及水文气象等方面都有着广泛的应用。

本书在内容上大致分如下四个方面：

第一部分介绍气象资料的整理，基本统计量的求法及其在气象中的应用（第一章）。

第二部分内容从第二章到第七章，主要介绍目前常用于气象要素场的分析和预报上的多元分析方法。

第三部分介绍时间序列分析、谱分析及马尔可夫概型分析（第

八章至第十章），其中涉及如何分析气象要素随时间变化的规律性，如持续性、周期性及状态转移可能性大小等规律以及如何利用它们作气象要素未来时刻的预报。

第四部分即最后一章介绍预报模式的评价及综合预报方法。

本书是为大学及专科气象专业学生编写的教材，在学习本书之前最好具备有一定的概率论和数理统计学及线性代数等方面必要的知识，自修学习的读者亦可参阅有关方面的书籍作必要的准备。

本书是在北京大学地球物理系“气象统计预报试用教材”基础上编写的。由于编者实践经验不足，理论水平有限，资料和方法收集也不够齐全，错误和不当之处在所难免，希望读者批评指正。

编著者

1989年1月于北京大学

目　　录

第一章　气象资料的整理

用统计方法作气象要素的分析和预报是依据大量的气象观测资料来进行的。从概率论或统计学的观点来看，某个气象要素可看成为一个变量(或随机变量)，它的全体在概率论中称为总体，而把收集到的该要素的资料称为样本。利用统计学方法对样本进行分析来估计和推测总体的规律性就是本书主要介绍的内容。气象中单个或多个要素可看成为统计学中单个或多个变量。本章将介绍对它们的资料(样本)进行初步整理的方法。

§1.1　气象资料的表示

1. 单个变量

我们要研究的对象是气象要素，比如气温、降水量、气压，它们可以是月平均值、年平均值、也可以是日平均值，这要看我们所要研究的气象问题而定。对于长期预报，经常分析的是气象要素的月资料，对于短期预报则常使用日资料，要作出预报就需要先研究它们随时间变化的规律性，比如我们抽取1951—1980年的1月份月气温平均值进行研究，这段资料就是我们研究的样本。

我们把单个气象要素记为 x，取它某一时间段的资料记录作为样本，样本中包含 n 个数据，记为

$$x_1, x_2, \cdots, x_n, \tag{1.1}$$

n 称为样本容量，每一个资料称为所抽取的一个样品。

如果取某要素月平均值的 n 年资料，那么这些数据就是一串随时间变化的序列，我们习惯把它称为时间序列，并记为

$$x_t \quad (t=1,2,\cdots,n),$$

这种表示法在时间序列分析中常用。

对于气温、气压及降水量等气象要素，观测值变化在正负无穷大之间，这种类型要素可看成为连续型随机变量。至于有一些气象要素，例如冰雹、晕、华等天气现象，气象资料中仅记录此现象“有”或“无”，这类无法用连续型变量表示的变量，一随用“1”或“0”二值数字化表征，这类变量可看成为离散型随机变量。至于云量，用数字1—10来分级表示的，也属于这一类型。当然，变量类型可以互相转化，例如对连续型变量如气温，规定一个临界值T_0。凡$T \geqslant T_0$记为“1”，$T < T_0$的记为“0”，那么这时的气温就处理成二值变量，这种做法在模式输出预报技术中经常被采用来作短期天气预报中的定性预报。

2. 多个变量

气象要素观测是三维空间的，有各种等压面上的要素资料，既有空间，又有时间变化。这时就可以把多个要素在某一段时间收集的资料看作为多个变量的样本，每个变量的样本可看成为一个向量。p个变量n次观测的样本可看成为n维空间中p个向量，每个向量可用行向量($1\times n$矩阵)表示

$$\begin{cases} \boldsymbol{x}_1 = (x_{11} \ x_{12} \ \cdots \ x_{1n}), \\ \boldsymbol{x}_2 = (x_{21} \ x_{22} \ \cdots \ x_{2n}), \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \boldsymbol{x}_p = (x_{p1} \ x_{p2} \ \cdots \ x_{pn}). \end{cases} \tag{1.2}$$

对第i个变量、第j个时刻的观测值，可表示为

$$x_{ij} \quad (i=1,\cdots,p; \quad j=1,\cdots,n).$$

有时为处理方便对多个要素的资料(样本)，可用一个矩阵来表示，记p个气象要素的$p\times n$个资料(设$p<n$)的样本为

$$\boldsymbol{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ x_{p1} & x_{p2} & \cdots & x_{pn} \end{pmatrix}, \tag{1.3}$$

把$\boldsymbol{X}$称为资料阵，它是把每个变量资料作为行的形式排列，称为

横资料阵。对 p 个要素 $p\times n$ 个资料的样本也可以把每一个变量资料以列的形式排列，写为如下矩阵形式称为竖资料阵。

$$X^* = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}, \tag{1.4}$$

上面(1.2)—(1.4)式仅是表示 p 个要素 $p\times n$ 个资料的样本表示方式。它们形式上不同，只是为在不同问题上处理方便而设的记号，本质上是一样的。为了不致使读者混淆，我们尽量采用(1.3)式的形式来表示 p 个要素 n 个资料的样本。例如我们取 12 月、1 月、2 月平均气温的 1951—1980 年资料(对 1 月、2 月是下一年资料)(见表1.1中第 2 至 4 列)。变量记为 x_1, x_2, x_3。对应的各变量的向量表示为

$$\boldsymbol{x}_1 = (1.0 \quad -5.3 \quad \cdots \quad -3.9)$$
$$\boldsymbol{x}_2 = (-2.7 \quad -5.9 \quad \cdots \quad -4.8)$$
$$\boldsymbol{x}_3 = (-4.3 \quad -3.5 \quad \cdots \quad -1.4).$$

由上面三个向量的分量亦可构成一个资料阵

$$\boldsymbol{X} = \begin{pmatrix} 1.0 & -5.3 & \cdots & -3.9 \\ -2.7 & -5.9 & \cdots & -4.8 \\ -4.3 & -3.5 & \cdots & -1.4 \end{pmatrix}.$$

表1.1 北京气温资料表*

年份	x_1	x_2	x_3	x_{d1}	x_{d2}	x_{d3}	x_{z1}	x_{z2}	x_{z3}
1951	1.0	−2.7	−4.3	3.7	1.8	−2.1	2.10	1.69	−1.14
1952	−5.3	−5.9	−3.5	−2.6	−1.4	−1.3	−1.49	−1.24	−0.71
1953	−2.0	−3.4	−0.8	0.7	1.1	1.4	0.39	1.05	0.76
1954	−5.7	−4.7	−1.1	−3.0	−0.2	1.1	−1.72	−0.14	0.59
1955	−0.9	−3.8	−3.1	1.8	0.7	−0.9	1.01	0.68	−0.49
1956	−5.7	−5.3	−5.9	−3.0	−0.8	−3.7	−1.72	−0.69	−2.01
1957	−2.1	−5.0	−1.6	0.6	−0.5	0.6	0.33	−0.42	0.32

续表

年　份	x_1	x_2	x_3	x_{d1}	x_{d2}	x_{d3}	x_{z1}	x_{z2}	x_{z3}
1958	0.6	−4.3	0.2	3.3	0.2	2.4	1.87	0.23	1.30
1959	−1.7	−5.7	2.0	1.0	−1.2	4.2	0.56	−1.06	2.27
1960	−3.6	−3.6	1.3	−0.9	0.9	3.5	−0.52	0.87	1.89
1961	−3.0	−3.1	−0.8	−0.3	1.4	1.4	−0.18	1.33	0.76
1962	0.1	−3.9	−1.1	2.8	0.6	1.1	1.58	0.59	0.59
1963	−2.6	−3.0	−5.2	0.1	1.5	−3.0	0.05	1.42	−1.63
1964	−1.4	−4.9	−1.7	1.3	−0.4	0.5	0.73	−0.32	0.27
1965	−3.9	−5.7	−2.5	−1.2	−1.2	−0.3	−0.70	−1.06	−0.17
1966	−4.7	−4.8	−3.3	−2.0	−0.3	−1.1	−1.15	−0.23	−0.60
1967	−6.0	−5.6	−4.9	−3.3	−1.1	−2.7	−1.89	−0.97	−1.47
1968	−1.7	−6.4	−5.1	1.0	−1.9	−2.9	0.56	−1.70	−1.58
1969	−3.4	−5.6	−2.0	−0.7	−1.1	0.2	−0.41	−0.97	0.10
1970	−3.1	−4.2	−2.9	−0.4	0.3	−0.7	−0.24	0.32	−0.38
1971	−3.8	−4.9	−3.9	−1.1	−0.4	−1.7	−0.64	−0.32	−0.93
1972	−2.0	−4.1	−2.4	0.7	0.4	−0.2	0.39	0.41	−0.11
1973	−1.7	−4.2	−2.0	1.0	0.3	0.2	0.56	0.32	0.10
1974	−3.6	−3.3	−2.0	−0.9	1.2	0.2	−0.52	1.14	0.10
1975	−2.7	−3.7	0.1	0.0	0.8	2.3	−0.01	0.78	1.24
1976	−2.4	−7.6	−2.2	0.3	−3.1	0.0	0.16	−2.80	0.00
1977	−0.9	−3.5	−2.3	1.8	1.0	−0.1	1.01	0.96	−0.06
1978	−2.7	−4.2	−0.5	0.0	0.3	1.7	−0.01	0.32	0.92
1979	−1.6	−4.5	−2.9	1.1	0.0	−0.7	0.62	0.04	−0.38
1980	−3.9	−4.8	−1.4	−1.2	−0.3	0.8	−0.70	−0.23	0.43
平　均	−2.7	−4.5	−2.2						
标准差	1.75	1.09	1.84						

* 表中第2至4列分别表示北京各月月平均气温（℃）的数值，5至7列表示各月对应的距平值，8至10列表示各月对应的标准化值。

§1.2　基本统计量

1. 平均值

我们要用一些统计量来表征某一要素样本中资料分布的特

点．平均值是一个常用的重要统计量．气象上的月平均气温、年平均气温及某气象要素多年平均值等就是这种统计量．平均值可作为要素总体数学期望的一个估计．

平均值是描述资料数字平均状况的量．对形如（1.1）式的单要素（变量）资料，其平均值为

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i \text{ .} \tag{1.5}$$

对 p 个要素（变量），可以分别求出它们的平均值 $\bar{x}_1$，$\bar{x}_2$，…，$\bar{x}_p$．由 p 个变量的平均值可以构成 $p\times1$ 的矩阵

$$\bar{\boldsymbol{x}}=\begin{pmatrix}\bar{x}_1\\ \bar{x}_2\\ \vdots\\ \bar{x}_p\end{pmatrix}.$$

这样的矩阵也可看成是 p 维空间中的一个向量（列向量），故也称 $\bar{\boldsymbol{x}}$ 为平均向量．如果用（1.3）形式的横资料阵，平均向量还可以表示为

$$\bar{\boldsymbol{x}}=\frac{1}{n}\boldsymbol{X}\mathbf{1},$$

式中$\mathbf{1}$为 n 个元素为 1 组成的向量．

例如对12月、1月、2月北京气温，其平均向量可利用表1.1的资料组成的矩阵算得

$$\bar{\boldsymbol{x}}=\frac{1}{30}\begin{pmatrix}1.0 & -5.3 & -2.0 & \cdots & -3.9\\ -2.7 & -5.9 & -3.4 & \cdots & -4.8\\ -4.3 & -3.5 & -0.8 & \cdots & -1.4\end{pmatrix}\begin{pmatrix}1\\1\\ \vdots\\1\end{pmatrix}=\begin{pmatrix}-2.7\\-4.5\\-2.2\end{pmatrix}.$$

在气象统计中，要素的资料是逐年增加的，如果每增加一个就从头又作一次平均值的计算，那么将会增加很多不必要的计算。实际上如果我们记某一变量 n 个资料的平均值为 $\bar{x}_n$，增加一个资料时样本平均值为$\bar{x}_{n+1}$，则$\bar{x}_{n+1}$可按下面公式直接计算得到

$$\bar{x}_{n+1}=\left(\frac{n}{n+1}\right)\bar{x}_n+\left(\frac{1}{n+1}\right)x_{n+1}, \tag{1.6}$$

其中 x_{n+1} 为增加一个资料时变量的实测值. 因为据平均值定义有

$$\bar{x}_{n+1}=\frac{1}{n+1}\sum_{i=1}^{n+1}x_i=\frac{1}{n+1}\left(\sum_{i=1}^{n}x_i+x_{n+1}\right)$$

$$=\frac{1}{n+1}(n\bar{x}_n+x_{n+1})=\left(\frac{n}{n+1}\right)\bar{x}_n+\left(\frac{1}{n+1}\right)x_{n+1}.$$

距平是气象上常用的量，它也即通常所说的异常，即对平均值的正常情况的偏差。资料中某一个数值与平均值之差就是距平，例如第 i 点资料的距平为

$$x_{di}=x_i-\bar{x}.$$

单变量样本(序列)中每个样品资料点的距平值组成的序列称为距平序列

$$x_1-\bar{x},\ x_2-\bar{x},\cdots,x_n-\bar{x}.$$

某一变量距平序列也可以用距平向量记为

$$\boldsymbol{x}_d=(x_1-\bar{x}\quad x_2-\bar{x}\quad\cdots\quad x_n-\bar{x})$$

$$=(x_{d1}\quad x_{d2}\quad\cdots\quad x_{dn}).$$

由 p 个要素(变量)的距平值的资料用横的次序排列组成的资料阵称为横距平资料阵，记为

$$\boldsymbol{X}_d=\begin{pmatrix}x_{d11} & x_{d12} & \cdots & x_{d1n}\\ x_{d21} & x_{d22} & \cdots & x_{d2n}\\ \cdots & \cdots & \cdots & \cdots\\ x_{dp1} & x_{dp2} & \cdots & x_{dpn}\end{pmatrix}.$$

气象上经常用距平值代替原样本中资料数值作为研究对象，因为在气象要素的研究中，它们受年变化周期影响很大，各月的平均值不一样。例如12月、1月、2月平均值就各不相同。为使之能在同一水平下进行比较，常使用距平值。用距平值作为变量的资料值，使得各变量的平均值为0，可以带来研究上的方便，

也便于计算．有时直接以它作为预报值，可以给人们一个偏高或偏低的直观了解．例如对12月、1月、2月北京气温距平序列（见表1.1中5—7列），可进行相互比较它们偏离平均值的变化状况。

对于使用距平值的变量，其平均值为0的证明如下：

$$\bar{x}_d=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})=\frac{1}{n}\sum_{i=1}^{n}x_i-\frac{1}{n}\sum_{i=1}^{n}\bar{x}$$

$$=\bar{x}-\frac{1}{n}(n\bar{x})=0.$$

因而任何变量序列，经过距平化处理，总可以化为平均值为0的变量序列。

由p个距平变量向量构成的距平资料阵为

$$\boldsymbol{X}_d=\begin{pmatrix}\boldsymbol{x}_{d1}\\ \boldsymbol{x}_{d2}\\ \vdots\\ \boldsymbol{x}_{dp}\end{pmatrix}.$$

其中$\boldsymbol{x}_{dk}$为第k个变量距平值组成的向量($1\times n$矩阵)。

2. **标准差与方差**

描述样本中资料与平均值差异平均状况的统计量就是标准差，它衡量资料围绕平均值的平均变化幅度．平常说：“内陆台站气温日变化较沿海地区要大”．这个日变化大小的比较就是用它们的标准差来比较的．

某气象要素(变量)x(含n个资料的样本)的标准差计算公式为

$$s_x=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}. \qquad (1.7)$$

更常用的是标准差的平方，称之为方差．记为

$$s_x^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2. \qquad (1.8)$$

如果对多个变量，其中第 k 个变量的资料写成行向量的形式时，其方差可表示为

$$s_k^2 = \frac{1}{n}\boldsymbol{x}_{dk}\boldsymbol{x}'_{dk},$$

其中 $\boldsymbol{x}_{dk} = (x_{dk1}\ x_{dk2}\ \cdots\ x_{dkn})$. $\boldsymbol{x}'_{dk}$ 为 $\boldsymbol{x}_{dk}$ 的转置。

例如，12月气温的方差为

$$s^2 = \frac{1}{n}\boldsymbol{x}_{d1}\boldsymbol{x}'_{d1}$$

$$= \frac{1}{30}(3.7\ \ -2.6\ \ 0.7\ \ \cdots\ \ -1.2)\begin{pmatrix} 3.7 \\ -2.6 \\ 0.7 \\ \vdots \\ -1.2 \end{pmatrix} = 3.078.$$

12月气温的标准差为

$$s_1 = \sqrt{3.078} = 1.75.$$

同样算出 1 月气温的方差及标准差为

$$s_2^2 = 1.190, \quad s_2 = 1.09.$$

2 月气温的方差及标准差为

$$s_3^2 = 3.402, \quad s_3 = 1.84.$$

在统计工作中，当资料增加一个，计算样本标准差较计算平均值更麻烦. 如果记 n 个资料时的样本方差为 s_n^2，$n+1$ 个资料时的方差为 s_{n+1}^2，则计算某一变量增加一个样品时的方差可按下式直接算出

$$s_{n+1}^2 = \left(\frac{n}{n+1}\right)s_n^2 + \frac{n}{(n+1)^2}(x_{n+1} - \bar{x}_n)^2. \qquad (1.9)$$

其中 x_{n+1} 为增加一个资料时的样品值，$\bar{x}_n$ 为某一变量 n个样品时的平均值(读者可自行证明).

从三个月气温的标准差值比较(见表1.1)可见，12月份比 1 月份大，这反映12月气温随时间变化幅度比 1 月大(见图1.1).

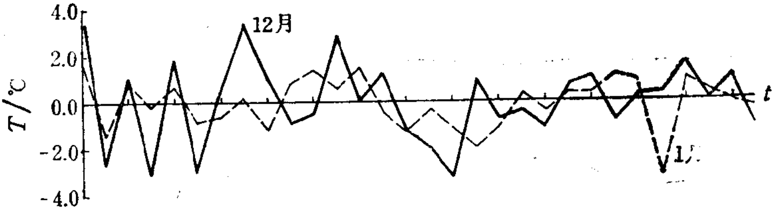

图1.1　北京12月气温与1月气温距平变化曲线比较

在估计要素总体方差与标准差中，可使用样本的方差与标准差，即用(1.8)与(1.7)式作估计。但有的用下面两式估计。

$$(s_x^*)^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2,$$

$$s_x^*=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2}.$$

在 n 较大时，$(s_x^*)^2$ 与 s_x^2 差别不大，气象中常用 s_x^2 作为总体的方差估计量，但在显著性检验中无偏估计量亦常用。

在气象要素中，各个要素的单位不一样，平均值及标准差亦有所不同。为使它们能在同一水平上进行比较，常使用标准化的方法，使它们变成同一水平的无单位的变量，这种变量就称为标准化变量。对单要素(变量)样本容量为 n 的资料，标准化变量的时间序列为

$$\frac{x_1-\bar{x}}{s},\ \frac{x_2-\bar{x}}{s},\ \cdots,\ \frac{x_n-\bar{x}}{s}.$$

标准化行向量记为

$$\boldsymbol{x}_z=(x_{z1}\ \ x_{z2}\ \cdots\ x_{zn}),$$

其中　$x_{zi}=\dfrac{x_i-\bar{x}}{s}\quad(i=1,2,\cdots,n).$

标准化变量具有如下性质：

（1）x_z 的平均值为0，因为