QIXIANG TONGJI FENXI YU YUBAO FANGFA

气象统计分析与预报方法

◎ 黄嘉佑 编著

（第四版）

气象出版社
China Meteorological Press

气象统计分析与预报方法

（第四版）

黄嘉佑　编著

气象出版社
China Meteorological Press

内容提要

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。

本书经中国气象局高等学校大气科学类教材编审领导小组审查，并经教育部高等学校教学指导委员会确认，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。

图书在版编目(CIP)数据

气象统计分析与预报方法 / 黄嘉佑编著. -- 4版
. -- 北京 : 气象出版社, 2016.5
ISBN 978-7-5029-6346-0

Ⅰ. ①气… Ⅱ. ①黄… Ⅲ. ①气象资料-统计分析②气象预报 Ⅳ. ①P468.0②P457

中国版本图书馆 CIP 数据核字(2016)第 098493 号

QIXIANG TONGJI FENXI YU YUBAO FANGFA

气象统计分析与预报方法

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　　**邮政编码**：100081
电　　话：010-68407112(总编室)　010-68409198(发行部)
网　　址：http://www.qxcbs.com　　**E-mail**：qxcbs@cma.gov.cn
责任编辑：黄红丽　　**终　　审**：邵俊年
责任校对：王丽梅　　**责任技编**：赵相宁
封面设计：博雅思企划
印　　刷：三河市百盛印装有限公司
开　　本：720 mm×960 mm　1/16　　**印　　张**：19.5
字　　数：398 千字
版　　次：2016 年 5 月第 4 版　　**印　　次**：2016 年 5 月第 5 次印刷
定　　价：48.00 元

再版前言

目前我们处在信息时代，一切信息都可以用数字型的数据来表现，因此，这些数据形成巨量资料(big data)，或称大数据、海量资料，是人们了解社会和自然的基础。在大气科学中，也存在大量的资料数据，也是人们了解大气变化的基础。气象数据是描述和记录气象现象的性质和变化的符号，是信息形式化的表示。在自然科学和社会科学中广泛使用气象数据(定性的或定量的、直接的或间接获得的描述、观测或实验气象数据)。我们可以通过气象数据分析来研究大气变化规律，甚至研究自然、人类社会与大气的关系等。对气象数据进行分析的方法，常常被用来研究大气中大气系统之间变化关系和演变的规律，并建立相应的数学模型进行大气未来状态的预测。

气象统计分析与预报方法，作为气象学中三大分析与预报方法之一，是高等学校大气科学类专业普遍开设的课程。本书是在作者多年来为本科生讲授的同名课程讲义的基础上编写而成的，着重介绍大气科学中统计分析与预测方法的基础理论，也涉猎一些新技术和新方法，除具有较强的理论性外，还有相当丰富的应用实例。本书第一版曾经中国气象局高等学校大气科学类教材编审领导小组确认为大学本科生通用教材，1999 年 10 月本教材再次经教育部高等学校大气科学教学指导委员会确认，作为大学本科通用教材，并获 1996 年全国第三届大气科学类优秀教材一等奖。本书出版后受到读者的欢迎和好评，至今已经出版了三版，加上此教材也已使用多年，为了适应学科的发展和教学上的需要，应广大读者的要求，对本教材的内容进行了必要的增删，使之更臻完善。

本书所介绍的气象数据分析和预测方法是气象数据分析的最常用的方法。实际上，许多气象数据分析的新方法和新理论正在发展，众多的内容几乎无法用一本书加以概括和讲述，也不可能在一个学期内讲授完毕。因此，在第四版中着重选择了最基本的统计量和分析方法做了一些补充，

作为学生今后进一步学习的基础。同时为方便读者使用该书介绍的方法，还补充了一些基础方法的计算机源程序供参考使用，各章末的参考文献也尽量引用近年的文献，以方便读者参考。

黄嘉佑

2015 年 12 月于北京大学

目　录

第 1 章　大气变量

用统计方法做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（或气候要素）及其变化（数据）可看成为一个变量（或随机变量），称为大气变量。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数据称为样本。利用统计学方法对大气变量的样本进行分析其总体规律性，并估计和推断大气变量总体的未来状态就是本书主要介绍的内容。

1.1　大气变量的表示

1.1.1　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。对于长期预报或短期气候预测，经常分析的是气象要素的月或年资料，对于短期预报则常使用日资料，要做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。

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，习惯把它称为时间序列，并记为 x_t，其中 $t=1,2,\cdots,n$。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。此类变量称为连续型变量。

至于一些气象要素，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，一般用“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，经

常被用来做短期天气预报中的定性预报。对地区温度的气候状态描述，一般分为特冷、冷、正常、暖和特暖五个级别状态来描述，常用数字 1、2、3、4、5 描述。

1.1.2 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，既有空间变化，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

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

对第 i 个变量、第 j 个时刻的观测值，可表示为

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

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

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

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。

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

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。例如取 12 月、1 月、2 月平均气温的 1951—1980 年资料(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
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	−.03	−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

续表

年份	x_1	x_2	x_3	x_{d1}	x_{d2}	x_{d3}	x_{z1}	x_{z2}	x_{z3}
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

1.2.1 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

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

对 p 个要素(变量),可以分别求出它们的平均值 $\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_p$。由 p 个变量的平均值可以构成 $p\times 1$ 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

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

式中 $\boldsymbol{1}$ 为 n 个元素为 1 组成的列向量。

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做一次平均值的计算，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

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

式中 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

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

距平是气象上常用的量，它也就是通常所说的异常，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：$x_{di}=x_i-\bar{x}$；对应的变量称为距平变量为：$x_d=x-\bar{x}$。

大气状态的异常是大气状态分析的重点，因为平均态随时间很少变化，而异常是随时间有剧烈变化的。大气异常状态是针对气候常态而言的状态。度量大气变量的异常状态，通常使用距平来表示，某时刻的变量距平表示为

$$x_{di}=x_i-\bar{x}\quad(i=1,2,\cdots,n) \tag{1.7}$$

对大气变量做距平处理后，得到的新变量称为距平变量。记为

$$x_d=x-\bar{x}$$

距平变量的数据向量表示为

$$\boldsymbol{x}_d=\boldsymbol{x}-\bar{x}\mathbf{1}$$

式中 $\boldsymbol{x}$ 为原始数据在 n 维空间中的向量，$\bar{x}$ 为数据平均值，$\mathbf{1}$ 为 n 维空间中的所有元素为 1 的向量。变量的距平数据向量，可以看成是原变量数据向量在 n 维空间中的平移。

距平变量的数据随时间变化，可以反映变量异常状态随时间变化的情况。对大气变量所有异常状态的平均情况的分析，可以了解大气的变化幅度。

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

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

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 的距平变量序列。

由 p 个距平变量向量构成的距平资料阵为

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

式中 $\boldsymbol{x}_{dk}$ 为第 k 个变量距平值组成的行向量($1 \times n$ 矩阵),$k = 1, 2, \cdots, p$。

1.2.2 标准差与方差

描述样本中资料与平均值差异的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

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

更常用的是标准差的平方,称之为方差。记为

$$s_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{1.9}$$

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

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

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

例如，12 月气温的方差为

$$s_1^2 = \frac{1}{n}\boldsymbol{x}_{d1}\boldsymbol{x}'_{d1} = \frac{1}{30}(3.7 \quad -2.6 \quad 0.7 \quad \cdots \quad -1.2)\begin{pmatrix} 3.7 \\ -2.6 \\ 0.7 \\ \vdots \\ -1.2 \end{pmatrix} = 3.078$$

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：$s_2^2=1.190$，$s_2=1.09$；

2 月气温的方差及标准差为：$s_3^2=3.402$，$s_3=1.84$。

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

$$s_{n+1}^2 = \left(\frac{n}{n+1}\right)s_n^2 + \frac{n}{(n+1)^2}(x_{n+1} - \bar{x}_n)^2 \tag{1.10}$$

式中 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 3 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）。

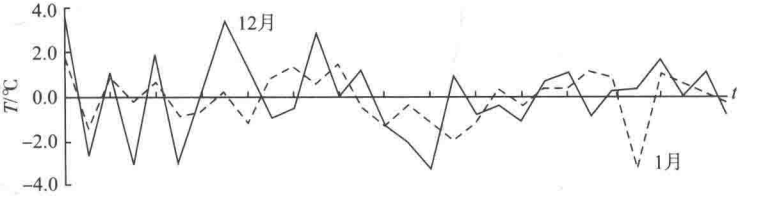

图 1.1　北京 12 月气温与 1 月气温距平变化曲线比较

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（1.9）与（1.8）式做估计。但有的用下面两式估计：

$$(s_2^*)^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2,\quad s_2^*=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

在 n 较大时，$(s_2^*)^2$ 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，但在显著性检验中亦常用无偏估计量 s_2^*。

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。表示为

$$x_{zi}=\frac{x_{di}}{s}=\frac{x_i-\bar{x}}{s}\quad(i=1,2,\cdots,n)\tag{1.11}$$

或表示为

$$x_{zi}=\frac{x_{di}}{s}$$

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

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

标准化变量具有如下性质：

(1)x_z 的平均值为 0，因为

$$\bar{x}_z=\frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{s}\right)=\frac{1}{ns}\sum_{i=1}^{n}x_{di}=0$$

(2)x_z 的方差为 1，因为

$$s_{xz}^2=\frac{1}{n}\sum_{i=1}^{n}(x_{zi}-\bar{x}_z)^2=\frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{s}\right)^2=\frac{1}{s^2}\cdot\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2=\frac{s^2}{s^2}=1$$

假定样本容量为 n，由 p 个要素的标准化行向量构成横标准化资料阵，记为 $\boldsymbol{X}_{z(p\times n)}$，即

$$\boldsymbol{X}_z=\begin{pmatrix}x_{z11}&x_{z12}&\cdots&x_{z1n}\\x_{z21}&x_{z22}&\cdots&x_{z2n}\\\vdots&\vdots&&\vdots\\x_{zp1}&x_{zp2}&\cdots&x_{zpn}\end{pmatrix}$$

由 p 个标准化变量向量构成标准化资料阵也可表示为：

$$\boldsymbol{X}_z=\begin{pmatrix}\boldsymbol{x}_{z1}\\\boldsymbol{x}_{z2}\\\vdots\\\boldsymbol{x}_{zp}\end{pmatrix}$$