畜禽育种中的线性模型

张沅　张勤　编著

北京农业大学出版社

前　　言

线性模型是数理统计学中创业较早、发展较快、理论系统且应用性极强的一个重要学科领域。在过去的几十年中，线性模型不仅在理论研究上不断深入和完善，而且随着计算机技术的飞速发展，自七十年代中期以来，线性模型在各学科和各行业中的应用有了突破性的进展，特别是在以 C.R.HENDERSON 为代表的一批动物遗传育种学家倡导下，线性模型理论在畜禽育种中得到了广泛的应用，从而不仅丰富了畜禽育种学科体系，同时也给畜牧生产带来很大的效益。近年来，欧美许多大学相继将线性模型列入畜牧专业或动物遗传育种专业的必修课或有关课程的基本内容。我国学者自 1982 年以来开始研究和应用畜禽育种中的线性模型，并取得了可喜的成果。但是，到目前为止，还没有一本系统论述畜禽育种中线性模型基本理论及其应用的专著。本书正是为了适应这一要求而作的。

本书的目的是力求较系统地论述在畜禽育种中所涉及到的线性模型的理论、方法及其应用。本书作者近年来在这个领域里做了一些教学和科研工作，对其发展现状比较了解。但由于不少新的发展目前尚未定型成熟，加之作者并非理论数学工作者，因此本书在论述线性模型理论的系统性和完整性上，是无法与纯数学性的同类论著相比的。但是，我们以自己的研究工作为基础，着重介绍了实用性十分强的线性模型方法及其理论基础，对从事畜牧或畜禽育种教学和科研工作者，是一本理论提高和实际应用的参考用书。本书还可作为动物遗传育种专业研究生的教科书或教学参考书。

全书共分十章。第一章是关于线性代数方面的预备知识。其中重点介绍了在一般线性代数教材中没有及讨论不充分，而阅读本书又十分必需的一些内容。第二章介绍了线性模型方面的基本知识。第三章讨论了线性模型的参数估计方法，重点讨论了在畜牧学科中应用最广的最小二乘估计。由第四章到第七章是本书的主体部分之一，专门讨论了目前在畜禽育种中应用最广泛的最佳线性无偏预测(BLUP)育种值的原理、方法、各种遗传评定模型以及在计算机上运算的技术问题。第八章到第十章是本书的另一主体部分，讨论了线性模型的方差组分和协方差组分估计的诸多方法。这对于畜禽遗传育种中的重要任务之一——准确地估计群体遗传参数，是十分有意义的。

本书的写作实际上开始于 1985 年。从那时起陆续收集文献资料，撰写了各种适用于不同层次讲习班的教材，在使用中不断补充完善。到 1989 年已基本形成本书目前的基本框架，并将它编入一本研究生用内部教材“动物育种原理和方法”中。在全国许多教学科研单位使用中，许多读者提出不少宝贵意见，并建议将线性模型部分独立出来，形成专著，公开发行。为此作者又作了进一步的修改和补充，尤其又将 1989 年以来的新进展以及我们的部分工作补充进来，力求能够反映该领域的最新面貌。作者的一个希望是使从事畜牧和畜禽育种工作的读者相信，线性模型是提高畜禽育种成效的重要数理统计方法，而不是高不可攀抽象理论。只要具有一定数学知识的读者，耐心地将本书读完，就会感到思路有一个明显的扩展。

本书的写作始终得到我们的导师吴仲贤教授的热情关怀和指导。盛志廉教授和吴常信教授对本书的写作也给予了极大的支持和关心，并提出了许多宝贵意见，这对提高本书的质量起了很大的作用。为此作者谨向三位老师表示衷心的感谢。

本书稿是由作者自行计算机汉字输入和排版的，这样一方面减少了校对的工作量和因漏校所带来的错误。但是另一方面由于我们排版的技术原因，也造成了书中有一些不规范的编排与表述。尤其是在书中反复出现的矩阵和向量等“量”的表达符号，按照“国标”均应以斜体字母，例如A，b，X来表示，但因受我们使用的计算机排版软件的限制，书中所有的量均用正黑体字母，**A**，**b**，**X** 表示。请读者原谅，尽管如此我们还是向为本书汉字输入和编排工作的陈遵送同志表示深切的谢意。

由于作者水平所限，书中肯定尚有许多不妥和错误之处，深望同行专家和广大读者不吝赐教。

作　者

1993 年 4 月于北京农业大学

目　　录

第一章　线性代数预备知识

矩阵是研究线性模型最基本的工具。根据本书的性质和任务，读者应具备线性代数的基础知识。本章的目的是对通常教材中没有论及或讨论不够充分，而在线性模型讨论又经常用到的一些知识，给予系统而扼要地叙述。为了保证知识的连贯性，在本章开始时还是梗概地介绍一下矩阵的基础知识。

1.1　纯量，矩阵与向量

只有大小的一个数值称为**纯量**(scalar)，也称为**标量**，**数量**或**无向量**。纯量除用数字表示以外，也可用经过定义的拉丁字母来表示，例如 m 代表质量，t 代表时间，b 代表任一数等。有时也将定义的缩写直接代表纯量，例如 SSE 代表误差平方和，用 F 代表 F 分布中的数。

由一定行数和一定列数的纯量，按一定顺序排列的表称为**矩阵**(matrix)，矩阵形式可以简化计算过程，由于其运算法则有异于普通代数，于是就形成了一个特殊的数学领域，即**矩阵代数**(matrix algebra)。矩阵一般均用大写字母表示，为了区别于其他的字母表示含义，对表示矩阵的字母或加下横线或用粗线体，本书采用后者。例如：**A**，**X′X**，**R**，**G** 等，任一矩阵可一般地表示为

$$\mathbf{A}=\{a_{ij}\}=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix}$$

矩阵中的每一个纯量称为矩阵的元素，通常用 a_{ij} 表示矩阵 **A** 的第 i 行和第 j 列的元素。矩阵的大小，即它的行数和列数，称为矩阵的维数或阶数，通常用 $\mathbf{A}_{n\times m}$ 来表示矩阵 **A** 有 n 行和 m 列。

仅有一列或一行的矩阵称为**向量**(vector)，前者称为**列向量**(column vector)，后者称为**行向量**(row vector)。向量用小写拉丁字母或希腊字母表示。同样要粗线体印刷，例如：**b**，**y**，β，**u** 等，为区别行向量和行向量，通常在字母的右上角加一撇表示行向量，如 **b′**，**y′**，β'等，不加撇则表示列向量。

1.2　一些特殊的矩阵

方阵(square matrix)　行数与列数相等的矩阵，$\mathbf{A}_{n\times n}$

对称阵(symmetric matrix)　元素间满足于$a_{ij}=a_{ji}$的方阵。

三角阵(triangular matrix)　分上三角阵，下三角阵两种，前者是当 j<i 时，$a_{ij}=0$ $(j<i)$，后者是当 i<j 时所有的元素为0，$a_{ij}=0$ $(i<j)$。

例如：

$$\text{上三角阵}\quad \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix} \qquad \text{下三角阵}\quad \begin{bmatrix} 1 & 0 & 0 \\ 2 & 3 & 0 \\ 4 & 5 & 6 \end{bmatrix}$$

对角阵(diagonal matrix)　除 i=j 时的元素(主对角线元素)外，其他元素均为 0 的方阵，即$a_{ij}=0(i\neq j)$。对角阵通常可用Diag$\{a_i\}$来表示，其中a_i为该阵的第 i 个对角线元素。

单位阵(identity matrix 或 unit matrix)　所有主对角线元素均为1($a_{ii}=1$)的对角阵。单位阵常用 **I** 表示。

零阵(null matrix)　所有元素均为 0 的矩阵。即$a_{ij}=0$

分块阵(block matrix)　用水平和垂直虚线将矩阵分为若干小块，此时的矩阵称为分块阵，其中的小块称为子阵 (sub-matrix)，例如：

$$\mathbf{A}=\begin{bmatrix} a_{11} & a_{12} & \vdots & a_{13} & a_{14} \\ a_{21} & a_{22} & \vdots & a_{23} & a_{24} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{31} & a_{32} & \vdots & a_{33} & a_{34} \\ a_{41} & a_{42} & \vdots & a_{43} & a_{44} \\ a_{51} & a_{52} & \vdots & a_{53} & a_{54} \end{bmatrix}=\begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}$$

分块对角阵(block diagonal matrix)　主对角线上的子阵都为方阵，其余子阵都是零阵的分块阵，例：

$$\mathbf{A}=\begin{bmatrix} \mathbf{A}_{11} & 0 & \dots & 0 \\ 0 & \mathbf{A}_{22} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \mathbf{A}_{kk} \end{bmatrix}$$

1.3 矩阵的运算

加法(addition)　当矩阵 **A** 与 **B** 同阶时，$\mathbf{A}+\mathbf{B}=\{a_{ij}+b_{ij}\}$

乘法 (multiplication)　当矩阵 **A** 的行数与矩阵 **B** 的列数相等时，**A** 与 **B** 可乘，其乘积为$\mathbf{A}_{r\times c}\cdot\mathbf{B}_{c\times l}=\mathbf{C}_{r\times l}=\{c_{ij}\}$，其中$c_{ij}=\sum_{k}^{c}a_{ik}b_{kj}$。矩阵的乘积一般不适合交换律，即 $\mathbf{AB}\neq\mathbf{BA}$（当 **AB** 和 **BA** 均存在，即 **A** 为 r×c 阶，**B** 为 c×r 阶时）。但矩阵的乘法满足结合律，即$(\mathbf{AB})\mathbf{C}=\mathbf{A}(\mathbf{BC})$，矩阵的乘法对加法适合分配律，即$(\mathbf{A}+\mathbf{B})\mathbf{C}=\mathbf{AC}+\mathbf{BC}$，$\mathbf{C}(\mathbf{A}+\mathbf{B})=\mathbf{CA}+\mathbf{AB}$。

下面是应该记住的有关矩阵乘法的一些规律：

一矩阵被一行向量左乘，其乘积为一行向量，即

$$\mathbf{a}'_{1\times c}\mathbf{B}_{c\times r}=\mathbf{p}'_{1\times r}$$

一矩阵被一列向量右乘，其乘积为一列向量，即

$$\mathbf{A}_{r\times c}\mathbf{b}_{c\times 1}=\mathbf{p}_{r\times 1}$$

一行向量被一列向量右乘，其乘积为一纯量，即

$$\mathbf{a}'_{1\times c}\mathbf{b}_{c\times 1}=\mathrm{p}$$

一列向量被一行向量右乘，其乘积为一矩阵，即

$$\mathbf{b}_{c\times 1}\mathbf{a}'_{1\times r}=\mathbf{P}_{c\times r}$$

一矩阵与一纯量相乘，其乘积为该矩阵的每一个元素与该纯量相乘，即

$$\mathrm{a}\mathbf{B}=\mathbf{C}=\{\mathrm{ab}_{ij}\}$$

一矩阵被一零阵左乘或右乘，其乘积为一零阵：

$$\mathbf{0}_{c\times r}\mathbf{A}_{r\times s}=\mathbf{0}_{c\times s},\quad \mathbf{A}_{r\times s}\mathbf{0}_{s\times p}=\mathbf{0}_{r\times p}$$

一矩阵被一单位阵左乘或右乘，其乘积仍为原矩阵：

$$\mathbf{I}_p\mathbf{A}_{p\times q}=\mathbf{A}_{p\times q}\mathbf{I}_q=\mathbf{A}_{p\times q}$$

转置(transpose)　矩阵的行与列对调即为**转置**。通常用 **A**′来表示 **A** 的转置，即

$$(\mathbf{A}_{m\times n})'=\mathbf{A}_{n\times m}=\{a_{ji}\}$$

转置矩阵也可以用$\mathbf{A}^T$表示。矩阵的转置具有以下性质：

(1) 当 **A** 矩阵为对称方阵时，$\mathbf{A}'=\mathbf{A}$

(2) $(\mathbf{A}')'=\mathbf{A}$

(3) $(\mathbf{AB})'=\mathbf{B}'\mathbf{A}'$

(4) $(\mathbf{AB}'\mathbf{C})'=\mathbf{C}'\mathbf{BA}'$

(5) $(\mathbf{A}+\mathbf{B}+\mathbf{C})'=\mathbf{A}'+\mathbf{B}'+\mathbf{C}'$

方阵的行列式(determinant)　设有一 p 阶方阵 $\mathbf{A}=\{\mathrm{a}_{ij}\}$，则此方阵的行列式为：

$$|\mathbf{A}|=\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ a_{p1} & a_{p2} & \cdots & a_{pp} \end{vmatrix}$$

方阵的行列式为一纯量。

当$|\mathbf{A}|=0$，则 $\mathbf{A}$ 为奇异阵，反之则为非奇异阵。

矩阵的迹(trace)　一个方阵的迹为其对角线元素之和，表示为:

$$\operatorname{tr}(\mathbf{A})=\sum_i a_{ii}$$

迹和运算有以下性质:

(1)　$\operatorname{tr}(\mathbf{AB})=\operatorname{tr}(\mathbf{BA})$

(2)　$\operatorname{tr}(\mathbf{ABC})=\operatorname{tr}(\mathbf{BCA})=\operatorname{tr}(\mathbf{CAB})$

(3)　$\operatorname{tr}(\mathbf{aq}')=\operatorname{tr}(\mathbf{q}'\mathbf{a})=\mathbf{q}'\mathbf{a}$

范数(norm)　矩阵与其转置阵乘积的迹和的平方根为该矩阵的**范数**。

$$\|\mathbf{A}\|=[\operatorname{tr}(\mathbf{A}'\mathbf{A})]^{0.5}=\left[\left(\sum_i\sum_j a_{ij}^2\right)\right]^{0.5}$$

范数有以下性质:

(1)　$\|\mathbf{A}\|>0$，除非 $\mathbf{A}=0$

(2)　$\|k\mathbf{A}\|=|k|\|\mathbf{A}\|$($k$为一纯量)

(3)　$\|\mathbf{x}+\mathbf{y}\|\leqslant\|\mathbf{x}\|+\|\mathbf{y}\|$

逆矩阵(inverse matrix)　对于一方阵 $\mathbf{A}$，若存在另一矩阵 $\mathbf{B}$，使得 $\mathbf{BA}=\mathbf{I}$，则称 $\mathbf{B}$ 为 $\mathbf{A}$ 的逆矩阵，并通常将它表示为 $\mathbf{A}^{-1}$，即 $\mathbf{A}^{-1}\mathbf{A}=\mathbf{I}$。$\mathbf{A}^{-1}$ 存在的先决条件是(1)$\mathbf{A}$ 必须是一方阵，(2)$\mathbf{A}$ 的行列式$|\mathbf{A}|\neq 0$，即 $\mathbf{A}$ 为非奇异阵。此时我们称 $\mathbf{A}$ 可逆。$\mathbf{A}^{-1}$ 具有如下性质:

(1)　$\mathbf{A}^{-1}\mathbf{A}=\mathbf{A}\mathbf{A}^{-1}=\mathbf{I}$

(2)　$\mathbf{A}^{-1}$ 是唯一的，即 $\mathbf{A}$ 只能有一个逆矩阵。

(3)　$|\mathbf{A}^{-1}|=\dfrac{1}{|\mathbf{A}|}$

(4)　$(\mathbf{A}^{-1})^{-1}=\mathbf{A}$，因而 $\mathbf{A}^{-1}$ 也是非奇异阵。

(5)　$(\mathbf{A}')^{-1}=(\mathbf{A}^{-1})'$

(6)　如 $\mathbf{A}$ 为对称阵，则 $\mathbf{A}^{-1}$ 也是对称阵。

(7)　若 $\mathbf{A}$、$\mathbf{B}$ 两矩阵均可逆，则　$(\mathbf{AB})^{-1}=\mathbf{B}^{-1}\mathbf{A}^{-1}$

逆矩阵对于线性方程组的求解是十分有用的，例如对于方程组：$\mathbf{Ax}=\mathbf{r}$，若 $\mathbf{A}$ 为一方阵且可逆，则在等式两边同时左乘 $\mathbf{A}^{-1}$，可得 $\mathbf{x}=\mathbf{A}^{-1}\mathbf{r}$。如何对一可逆矩阵求逆矩阵，是线性代数中的一项重要内容，在此不能详细介绍，在以后的章节中将介绍一些特殊的简化方法。

正交矩阵 (orthogonal matrix)　如矩阵 $\mathbf{A}$ 满足 $\mathbf{AA}'=\mathbf{I}=\mathbf{A}'\mathbf{A}$，则称 $\mathbf{A}$ 为**正交矩阵**。根据此定义可知 $\mathbf{A}$ 也一定是一方阵，且有$\mathbf{A}^{-1}=\mathbf{A}'$。若有 2 个向量 $\mathbf{u}$ 和 $\mathbf{v}$，如 $\mathbf{u}'\mathbf{u}=1=\mathbf{v}'\mathbf{v}$，$\mathbf{u}'\mathbf{v}=0$，则称这 2 个向量为正交向量。

幂等矩阵 (idempotent matrix)　如矩阵 $\mathbf{A}$ 满足 $\mathbf{A}^2=\mathbf{A}$，则称 $\mathbf{A}$ 为**幂等矩阵**。如 $\mathbf{A}$ 为幂等矩阵，则 $\mathbf{A}$ 必为方阵，且$\mathbf{A}^r=\mathbf{A}$（r 为任意正整数），若 $\mathbf{K}$ 是幂等矩阵，则 $(\mathbf{I}-\mathbf{K})^2=\mathbf{I}-\mathbf{K}$，即 $\mathbf{I}-\mathbf{K}$ 也是幂等矩阵，但 $\mathbf{K}-\mathbf{I}$ 不是幂等矩阵。

1.4　二次型 (quadratic form)

由变量x_1，x_2，…，x_p 组成的下列函数

$$\begin{aligned} q = & a_{11}x_1^2 + a_{12}x_1x_2 + \cdots + a_{1p}x_1x_p \\ & + a_{21}x_2x_1 + a_{22}x_2^2 + \cdots + a_{2p}x_2x_p \\ & + \cdots\cdots\cdots \\ & + a_{p1}x_px_1 + a_{p2}x_px_2 + \cdots + a_{pp}x_p^2 \\ = & \sum_i\sum_j a_{ij}x_ix_j = \sum_i a_{ii}x_i^2 + \sum\sum_{i\neq j} a_{ii}x_ix_j \end{aligned}$$

称为**二次型函数**。若令

$$\mathbf{A}=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ a_{p1} & a_{p2} & \cdots & a_{pp} \end{bmatrix},\quad \mathbf{x}=\begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_p \end{bmatrix}$$

则上式可表示为

$$q=\mathbf{x}'\mathbf{A}\mathbf{x}$$

矩阵 $\mathbf{A}$ 称为**二次型的方阵**。对于任意一特定的二次型，其方阵并不是唯一的，例如

$$\mathbf{x}'\mathbf{A}\mathbf{x}=\mathbf{x}'\begin{bmatrix} 1 & 2 & 3 \\ 4 & 7 & 6 \\ 2 & -2 & 5 \end{bmatrix}\mathbf{x} \quad 与 \quad \mathbf{x}'\mathbf{B}\mathbf{x}=\mathbf{x}'\begin{bmatrix} 1 & 1 & 1 \\ 5 & 7 & 0 \\ 4 & 4 & 5 \end{bmatrix}\mathbf{x}$$

这两个二次型是完全相等的，尽管它们的方阵 **A** 和 **B** 并不相等。事实上一个二次型可以有多个不同的方阵，但它们都有二个共同的特点，一是它们都有相同的主对角线元素，二是它们的每对对称的非对角线元素之和 $a_{ij}+a_{ji}$ 相等，如上列中的 2+4＝1+5，3+2＝1+4。根据这个特点，我们可将任一二次型写为

$$\mathbf{x}'\mathbf{A}\mathbf{x}=\mathbf{x}'\left[\frac{1}{2}(\mathbf{A}+\mathbf{A}')\right]\mathbf{x}$$

矩阵$\frac{1}{2}(\mathbf{A}+\mathbf{A}')$是一个对称阵，并且它对于一个特定的二次型是唯一的。由于对称阵有很多有用的特性，因而在讨论二次型时，我们总将其方阵当作对称阵来考虑。

若对于任意 $\mathbf{x}\neq 0$，均有 $q=\mathbf{x}'\mathbf{A}\mathbf{x}>0$，则称此二次型是**正定的**，若均有 $q<0$，则称之为**负定的**，若均有 $q\geqslant 0$，则称之为**半正定的**。其中的对称方阵 **A** 也相应地被称为**正定阵**，**负定阵和半正定阵**。正定阵和半正定阵统称**非负定阵**。任意正定阵均存在逆矩阵。

1.5 矩阵的导数(differentiation)

有式:

$$c=3x_1+5x_2+9x_3$$

$$=(3\quad 5\quad 9)\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=\mathbf{b}'\mathbf{x}=\mathbf{x}'\mathbf{b}$$

则 $$\frac{\partial c}{\partial x_1}=3,\ \frac{\partial c}{\partial x_2}=5,\ \frac{\partial c}{\partial x_3}=9$$

或 $$\frac{\partial c}{\partial \mathbf{x}}=\begin{bmatrix}\frac{\partial c}{\partial x_1}\\\frac{\partial c}{\partial x_2}\\\frac{\partial c}{\partial x_3}\end{bmatrix}=\begin{bmatrix}3\\5\\9\end{bmatrix}=\mathbf{b}$$

对于一个行向量 $\mathbf{y}'=\mathbf{x}'\mathbf{A}$，若将 **A** 表示为

$$\mathbf{A}=(\mathbf{a}_1\quad \mathbf{a}_2\quad \cdots\quad \mathbf{a}_n)$$

其中 $\mathbf{a}_i$(i＝1，2，⋯，n)为 **A** 中的第 i 列，则 $\mathbf{y}'$可表示为

$$\mathbf{y}'=(y_1,\ y_2,\ \cdots,\ y_n)$$

$$= \mathbf{x}' \begin{pmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{pmatrix}$$
$$= \begin{pmatrix} \mathbf{x}'\mathbf{a}_1 & \mathbf{x}'\mathbf{a}_2 & \cdots & \mathbf{x}'\mathbf{a}_n \end{pmatrix}$$

其中 **y** 的第 i 个元素为$y_i = \mathbf{x}'\mathbf{a}_i$，$\mathbf{a}_i$ 为 **A** 的第 i 列。

$$\frac{\partial \mathbf{y}'}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial y_1}{\partial \mathbf{x}} & \frac{\partial y_2}{\partial \mathbf{x}} & \cdots & \frac{\partial y_n}{\partial \mathbf{x}} \end{bmatrix}$$
$$= \begin{bmatrix} \frac{\partial(\mathbf{x}'\mathbf{a}_1)}{\partial \mathbf{x}} & \frac{\partial(\mathbf{x}'\mathbf{a}_2)}{\partial \mathbf{x}} & \cdots & \frac{\partial(\mathbf{x}'\mathbf{a}_n)}{\partial \mathbf{x}} \end{bmatrix}$$
$$= (\mathbf{a}_1 \quad \mathbf{a}_2 \quad \cdots \quad \mathbf{a}_n) \quad = \mathbf{A}$$

因而 $\qquad \partial(\mathbf{x}'\mathbf{A}) / \partial\mathbf{x} = \mathbf{A}$

对于一列向量 $\mathbf{y} = \mathbf{A}\mathbf{x}$，有:

$$\begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1p}x_p \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2p}x_p \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{np}x_p \end{bmatrix}$$

$$\frac{\partial y_i}{\partial \mathbf{x}} = \begin{bmatrix} a_{i1} \\ a_{i2} \\ \cdots \\ a_{ip} \end{bmatrix}$$

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \left(\frac{\partial y_1}{\partial \mathbf{x}} \quad \frac{\partial y_2}{\partial \mathbf{x}} \quad \cdots \quad \frac{\partial y_n}{\partial \mathbf{x}}\right) = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \cdots & \cdots & \cdots & \cdots \\ a_{1p} & a_{2p} & \cdots & a_{np} \end{bmatrix}$$

因而 $\qquad \partial(\mathbf{A}\mathbf{x}) / \partial\mathbf{x} = \mathbf{A}'$

对于一二次型 $y = \mathbf{x}'\mathbf{A}\mathbf{x}$，则有

$$\frac{\partial y}{\partial \mathbf{x}} = \frac{\partial}{\partial \mathbf{x}}(\mathbf{x}'\mathbf{A}\mathbf{x}) = \frac{\partial}{\partial \mathbf{x}}\left(\sum_i \sum_j a_{ij} x_i x_j\right)$$

$$
= \begin{bmatrix} \frac{\partial}{\partial x_1} \\ \frac{\partial}{\partial x_2} \\ \cdots \\ \frac{\partial}{\partial x_n} \end{bmatrix} (\sum_i \sum_j a_{ij} x_i x_j)
$$

$$
= \begin{bmatrix} \sum_j a_{1j} x_j + \sum_i a_{i1} x_i \\ \sum_j a_{2j} x_j + \sum_i a_{i2} x_i \\ \cdots\cdots\cdots\cdots \\ \sum_j a_{nj} x_j + \sum_i a_{in} x_i \end{bmatrix}
$$

$$
= \mathbf{Ax} + \mathbf{A'x}
$$

上述推导也可根据微分学中分部求导的法则求得。若令 $\mathbf{P} = \mathbf{Ax}$，$\mathbf{Q} = \mathbf{x'A}$，则

$$
\begin{aligned}
\frac{\partial}{\partial \mathbf{x}}(\mathbf{x'Ax}) &= \frac{\partial}{\partial \mathbf{x}}(\mathbf{x'P}) + \frac{\partial}{\partial \mathbf{x}}(\mathbf{Qx}) \\
&= \mathbf{P} + \mathbf{Q'} \\
&= \mathbf{Ax} + \mathbf{A'x}
\end{aligned}
$$

当 $\mathbf{A}$ 为对称阵时，有

$$
\frac{\partial}{\partial \mathbf{x}}(\mathbf{x'Ax}) = \mathbf{Ax} + \mathbf{Ax} = 2\mathbf{Ax}
$$

1.6 方差协方差矩阵（Variance-covariance matrix）

设 x_1，x_2，…，x_n 是 n 个随机变数，它们的数学期望分别是 μ_1，μ_2，…，μ_n，方差为 σ_1^2，σ_2^2，…，σ_n^2。协方差为 σ_{12}，σ_{13}，…，σ_{1n}，σ_{23}，…，σ_{2n}，…，$\sigma_{n-1,n}$。

如果这些随机变量以向量 $\mathbf{x}$ 表示，它们的期望用向量 μ 表示。

$$\mathbf{x}=\begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{bmatrix} \quad \mu=\begin{bmatrix} \mu_1 \\ \mu_2 \\ \cdots \\ \mu_n \end{bmatrix}$$

则有 $\mathrm{E}(\mathbf{x})=\mu$。它们的方差和协方差可以用矩阵的形式写为：

$$\mathrm{Var}(\mathbf{x})=\mathbf{G}=\begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{1n} & \sigma_{2n} & \cdots & \sigma_{n2} \end{bmatrix}$$

这是一个对称阵，即 $\mathbf{G}'=\mathbf{G}$，它的主对角线上的第 i 个元素σ_i^2是x_i的方差。在第 i 行与第 j 列上的非主对角线元素$\sigma_{ij}(i\neq j)$是x_i和x_j的协方差。这个矩阵称为随机变量x_1，x_2，…，x_n的**方差协方差矩阵**，或简称为**协方差矩阵**。

由方差和协方差的基本定义，即：

$$\sigma_i^2=\mathrm{E}(x_i-\mu_i)^2$$
$$\sigma_{ij}=\mathrm{E}(x_i-\mu_i)(x_j-\mu_j) \qquad (i\neq j)$$

很容易看出

$$\mathrm{Var}(\mathbf{x})=\mathrm{E}\begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2 \\ \cdots\cdots \\ x_n-\mu_n \end{bmatrix}(x_1-\mu_1 \quad x_2-\mu_2 \quad \cdots \quad x_n-\mu_n)$$
$$=\mathrm{E}(\mathbf{x}-\mu)(\mathbf{x}-\mu)'$$

如果数学期望为零，即 $\mu=\mathbf{0}$，则上式变为

$$\mathrm{Var}(\mathbf{x})=\mathrm{E}(\mathbf{x}\mathbf{x}')$$

对 $\mathbf{x}$ 作线性变换 $\mathbf{y}=\mathbf{T}\mathbf{x}$，我们也很容易求出 $\mathbf{y}$ 的期望向量和方差–协方差矩阵。

$$\mathrm{E}(\mathbf{y})=\mathrm{E}(\mathbf{T}\mathbf{x})=\mathbf{T}\mathrm{E}(\mathbf{x})=\mathbf{T}\mu$$
$$\mathrm{Var}(\mathbf{y})=\mathrm{E}[\mathbf{y}-\mathrm{E}(\mathbf{y})][(\mathbf{y}-\mathrm{E}(\mathbf{y})]'$$
$$=\mathrm{E}(\mathbf{T}\mathbf{x}-\mathbf{T}\mu)(\mathbf{T}\mathbf{x}-\mathbf{T}\mu)'$$
$$=\mathrm{E}[\mathbf{T}(\mathbf{x}-\mu)][(\mathbf{T}(\mathbf{x}-\mu)]'$$

$$
\begin{aligned}
&= \mathbf{T}\mathrm{E}(\mathbf{x}-\mu)(\mathbf{x}-\mu)'\mathbf{T}' \\
&= \mathbf{T}\mathrm{Var}(\mathbf{x})\mathbf{T}' \\
&= \mathbf{TGT}'
\end{aligned}
$$

如果我们有 $\mathbf{y}=\mathbf{t}'\mathbf{x}$，则

$$\mathrm{Var}(\mathbf{y}) = \mathbf{t}'\mathbf{G}\mathbf{t}$$

因为 $\mathrm{Var}(\mathbf{y})$是一个方差，根据定义，它不可能为负，因而对于任意 $\mathbf{t}\neq\mathbf{0}$，总有 $\mathbf{t}'\mathbf{G}\mathbf{t}\geqslant 0$，所以 $\mathbf{G}$ 是一个非负定对称矩阵。这对任何方差-协方差矩阵都成立，这是方差—协方差矩阵的一个基本特征，在绝大多数情况下，方差—协方差矩阵是正定的。

1.7 特征值与特征向量

对于一个 n 阶方阵 $\mathbf{A}$ 和一 n 维非零向量 $\mathbf{u}$，若存在一个数 λ，使得

$$\mathbf{A}\mathbf{u}=\lambda\mathbf{u}$$

则称 λ 为矩阵 $\mathbf{A}$ 的**特征值**（eigen value）或**特征根**（characteristic root），$\mathbf{u}$ 为 $\mathbf{A}$ 的特征值 λ 所对应的**特征向量**（eigen vector）。

若将上式重写为

$$(\mathbf{A}-\mathbf{I}\lambda)\ \mathbf{u}=\mathbf{0},$$

这是一个线性齐次方程组，当（$\mathbf{A}-\mathbf{I}\lambda$）为非奇异阵时，它的唯一解为 $\mathbf{u}=0$，而当（$\mathbf{A}-\mathbf{I}\lambda$）为奇异阵时，它可以有多个非零解，因而对于矩阵 $\mathbf{A}$，其特征根和特征向量存在的充要条件是（$\mathbf{A}-\mathbf{I}\lambda$）为奇异阵，即

$$|\mathbf{A}-\mathbf{I}\lambda|=0$$

上式称为 $\mathbf{A}$ 的**特征方程**（characteristic equation），它是 λ 的 n 阶多项式方程，有 n 个根，记为λ_1，λ_2，…，λ_n，其中某些根可能为零，它们都是矩阵 $\mathbf{A}$ 的特征值，其中最大的一个特征值称为第一特征值。而对于每一个特征值 λ_i，都可有一相应的特征向量 $\mathbf{u}_i$，（注意：$\mathbf{u}_i$ 不是唯一的），使得

$$\mathbf{A}\ \mathbf{u}_i=\lambda_i\ \mathbf{u}_i$$

若令$\mathbf{D}=\mathrm{Diag}\{\lambda_1\ \ \lambda_2\ \ \cdots\ \ \lambda_n\}$ $\mathbf{U}=(\mathbf{u}_1\ \ \mathbf{u}_2\ \ \cdots\ \ \mathbf{u}_n)$ （$\mathbf{U}$ 不是唯一的），则有

$$\mathbf{AU}=\mathbf{UD}$$

特征值和特征向量具有如下性质：

(1) n 阶方阵具有 n 个特征值。

(2) 一方阵的所有特征值之和等于它的迹：

$$\mathrm{tr}(\mathbf{A})=\sum_i^n a_{ii}=\sum\lambda_i$$