向量，记作 vec($\boldsymbol{X}$)，即

$$\operatorname{vec}(\boldsymbol{X})=\begin{pmatrix}\boldsymbol{x}_1\\ \boldsymbol{x}_2\\ \vdots\\ \boldsymbol{x}_q\end{pmatrix} \tag{3.5.3}$$

称“vec”为**拉直运算**。当 $\boldsymbol{X}$ 是 p 阶对称矩阵时，因 $x_{ij}=x_{ji}$，故只需取其下三角部分组成一个缩减了的长向量，记作 vech($\boldsymbol{X}$)，即

$$\operatorname{vech}(\boldsymbol{X})=(x_{11},\cdots,x_{p1},x_{22},\cdots,x_{p2},\cdots,\ x_{p-1,p-1},x_{p,p-1},x_{pp})' \tag{3.5.4}$$

随机矩阵 $\boldsymbol{X}$ 的分布是指 vec($\boldsymbol{X}$)或(当 $\boldsymbol{X}'=\boldsymbol{X}$ 时)vech($\boldsymbol{X}$)的分布，拉直运算将矩阵分布问题转化为了向量分布的问题，这就方便了对问题的研究。

设随机向量 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n$ 独立同分布于 $N_p(\boldsymbol{0},\ \boldsymbol{\Sigma})$，$\boldsymbol{\Sigma}>0$，$n\geqslant p$，则 p 阶矩阵 $\boldsymbol{W}=\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i'$ 的分布称为自由度为 n 的(p 阶)**威沙特**(Wishart)**分布**，记作 $W_p(n,\ \boldsymbol{\Sigma})$。当 $p=1$，$\Sigma=\sigma^2=1$ 时，显然有 $W=\sum_{i=1}^{n}x_i^2\sim\chi^2(n)$，即有

$$W_1(n,1)=\chi^2(n) \tag{3.5.5}$$

因此，威沙特分布是卡方分布在多元场合下的一种推广。

从上述定义出发容易推得威沙特分布具有如下**性质**：

(1) 设 $\boldsymbol{W}_i\sim W_p(n_i,\ \boldsymbol{\Sigma})$，$i=1,2,\cdots,k$，且相互独立，则

$$\boldsymbol{W}_1+\boldsymbol{W}_2+\cdots+\boldsymbol{W}_k\sim W_p(n_1+n_2+\cdots+n_k,\ \boldsymbol{\Sigma}) \tag{3.5.6}$$

(2) 设 $\boldsymbol{W}\sim W_p(n,\ \boldsymbol{\Sigma})$，$\boldsymbol{C}$ 为 $q\times p$ 常数矩阵，则

$$\boldsymbol{CWC}'\sim W_q(n,\ \boldsymbol{C\Sigma C}') \tag{3.5.7}$$

设 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n$ 是取自 $N_p(\boldsymbol{\mu},\ \boldsymbol{\Sigma})$，$\boldsymbol{\Sigma}>0$ 的一个样本，$n>p$，则可以证明，$\bar{\boldsymbol{x}}$ 和 $\boldsymbol{S}$ 相互独立，且有

$$(n-1)\boldsymbol{S}\sim W_p(n-1,\boldsymbol{\Sigma}) \tag{3.5.8}$$

* §3.6　二次型分布

设 $\boldsymbol{x}=(x_1,x_2,\cdots,x_p)'\sim N_p(\boldsymbol{\mu},\boldsymbol{I})$，即 $x_1,x_2,\cdots,x_p$ 相互独立，且 $x_i\sim N(\mu_i,1)$，$i=1,2,\cdots,p$，则 $\boldsymbol{x}'\boldsymbol{x}=x_1^2+x_2^2+\cdots+x_p^2$ 的分布称为**非中心的卡方分布**，记作 $\boldsymbol{x}'\boldsymbol{x}\sim\chi^2(p,\lambda)$，其中 $\lambda=\boldsymbol{\mu}'\boldsymbol{\mu}=\mu_1^2+\mu_2^2+\cdots+\mu_p^2$ 称为**非中心参数**。当 $\lambda=0$ 时，$\chi^2(p,\lambda)$ 就化为 $\chi^2(p)$。下面我们列举出二次型分布的一些**性质**。

(1) 设 $\boldsymbol{x}\sim N_p(\boldsymbol{0},\boldsymbol{I})$，$\boldsymbol{A}'=\boldsymbol{A}$，则 $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}\sim\chi^2(r)$，当且仅当 $\operatorname{rank}(\boldsymbol{A})=r$ 且 $\boldsymbol{A}^2=\boldsymbol{A}$。

(2) 设 $\boldsymbol{x}\sim N_p(\boldsymbol{0},\boldsymbol{I})$，$\boldsymbol{A}_i'=\boldsymbol{A}_i$，$i=1,2$，则 $\boldsymbol{x}'\boldsymbol{A}_1\boldsymbol{x}$ 和 $\boldsymbol{x}'\boldsymbol{A}_2\boldsymbol{x}$ 相互独立，当且仅当 $\boldsymbol{A}_1\boldsymbol{A}_2=\boldsymbol{0}$。

(3) 设 $\boldsymbol{x}\sim N_p(\boldsymbol{\mu},\boldsymbol{I})$，$\boldsymbol{A}_i'=\boldsymbol{A}_i$，$q_i=\boldsymbol{x}'\boldsymbol{A}_i\boldsymbol{x}$，$i=1,2,\cdots,k$，其中 $\sum_{i=1}^{k}\boldsymbol{A}_i=\boldsymbol{I}$，则如下三个命题是**相互等价**的：

(i) $q_1,q_2,\cdots,q_k$ 相互独立，且均服从非中心的卡方分布；

(ii) $\sum_{i=1}^{k}\operatorname{rank}(\boldsymbol{A}_i)=p$；

(iii) $\boldsymbol{A}_i^2=\boldsymbol{A}_i$， $i=1,2,\cdots,k$，且 $\boldsymbol{A}_i\boldsymbol{A}_j=\boldsymbol{0}$， $1\leqslant i\neq j\leqslant k$。

小 结

*1.多元正态分布有非退化的和退化的两种，前者存在概率密度，后者不存在概率密度，是一个奇异型的分布。

*2.利用§3.2 中的性质(1)和(2)可以给出多元正态分布的另两个等价定义，即用特征函数定义和用 $\boldsymbol{x}$ 分量的一切线性组合均为正态变量的性质定义。

3.多元正态概率密度的等高面($p=2$ 时为等高线)是一个(超)椭球曲面($p=2$ 时为一条椭圆曲线)。

4.多元正态变量的线性变换、边缘分布和条件分布仍是(多元)正态的。

5.维数相同的独立多元正态变量的线性组合仍为多元正态变量。

6.对多元正态变量来说，其子向量间的不相关性与独立性是等价的。

7.极大似然估计是一种参数方法。若总体的分布类型已知，则常采用该方法。它的优点在于能够较充分地利用总体分布类型的信息，并常使获得的估计量有较好的性质。

8.评价估计量的好坏有这样四个常用准则：无偏性、有效性、一致性和充分性。对于多元正态分布，$(\bar{\boldsymbol{x}},\boldsymbol{S})$ 是$(\boldsymbol{\mu},\boldsymbol{\Sigma})$的一致最优无偏估计、一致估计和充分估计量。

9.复相关系数度量了一个随机变量 y 和一组随机变量 $x_1,x_2,\cdots,x_p$ 之间相关关系的强弱，它也是 y 和用 $x_1,x_2,\cdots,x_p$ 对 y 的最优线性预测的相关系数。复相关可看成是(将在第十章介绍的)典型相关的一个特例。

10.偏相关系数度量了剔除给定变量的(线性)影响(或控制给定变量)之后变量间的相关性。对于多元正态变量，偏相关系数与条件相关系数是相同的，此时的条件相关系数与已知条件变量的取值无关，这是多元正态分布的一个特点。

11.偏相关系数为零，并不意味着相关系数为零，反之亦然；偏相关系数和相关系数未必同号；偏相关系数和相关系数之间孰大孰小没有必然的结论。

12.在本章的统计推断中，通常要求样本容量 n 要大于总体的维数 p，否则样本协方差矩阵 $\boldsymbol{S}$ 将是退化的。对于多元正态总体，$n>p\Leftrightarrow\boldsymbol{S}>0$ (以概率 1 成立)。

13.对于多元正态总体，$\bar{\boldsymbol{x}}$ 和 $\boldsymbol{S}$ 相互独立，$\bar{\boldsymbol{x}}$ 和$(n-1)\boldsymbol{S}$ 的抽样分布分别为多元正态分布和威沙特分布。对于非多元正态总体，根据多元中心极限定理，在大样本情形下，$\bar{\boldsymbol{x}}$ 的抽样分布可用正态来近似。这些结论都可看作是一元情形向多元情形的直接推广。

附录 3-1 SAS 的应用

以下我们利用例 3.4.2 中的数据来介绍 SAS 9.3 的一些应用，包括使用编程和菜单两种不同途径来获得相关分析的一些结果，以及对数据所作的两种图形表示。建议未接触过 SAS 菜单系统的读者参阅参考文献[18]中的附录 5－1。

一、用编程作相关分析

SAS 程序：

```
proc corr data=sasuser.examp342 cov;
   var x1-x7;
```