第 1 章
随机向量与多元正态分布

以往，我们只讨论一个随机变量的情况，与之相关的一元离散型随机分布（如二项分布、泊松分布、超几何分布等）和一元连续型随机分布（如一元正态分布、t 分布、χ^2 分布、F 分布等）在理论和实际应用中都有着重要的地位。

但在实际问题中，对于某些随机试验的结果需要同时用两个或两个以上的随机变量来描述。例如，为了研究某一地区学龄前儿童的发育情况，对这一地区的儿童进行抽查。对于每个儿童，都能观察到他的身高 X_1 和体重 X_2。因此，相对于某地区全部学龄前儿童这一样本空间，由它们构成的一个向量(X_1, X_2)，就形成二维随机向量。又如研究公司的运营情况时，要涉及公司的资金周转能力 X_1、偿债能力 X_2、获利能力 X_3 及竞争能力 X_4 等财务指标，这样向量(X_1, X_2, X_3, X_4)就形成一个四维随机向量。

引导案例

制造公司产品质量检测

在一家制造公司的质量控制部门，负责监督产品质量的工程师小李面临着一项重要任务。公司生产的产品涉及多个技术指标，包括尺寸、重量、硬度，这些指标共同构成了一个随机向量 $\boldsymbol{X}=(X_1, X_2, X_3)$，符合三元正态分布。小李通过质量控制分析得知，协方差矩阵$\boldsymbol{\Sigma}$如下：

$$\boldsymbol{\Sigma}=\begin{bmatrix} 4 & 1.5 & 0.8 \\ 1.5 & 9 & 2.5 \\ 0.8 & 2.5 & 5 \end{bmatrix}$$

现有两个产品 A 和 B 的技术指标数据为：产品 A，$\boldsymbol{X}_{\mathrm{A}}=(10, 20, 30)$；产品 B，$\boldsymbol{X}_{\mathrm{B}}=(12, 18, 28)$。小李知道，为了评估产品之间的差异程度，他需要使用统计距离来衡量随机向量之间的差异。最初，他选择了欧氏距离来计算产品之间的距离：

$$\begin{aligned} d_E(\mathrm{A},\mathrm{B}) &= \sqrt{(X_{\mathrm{A1}}-X_{\mathrm{B1}})^2+(X_{\mathrm{A2}}-X_{\mathrm{B2}})^2+(X_{\mathrm{A3}}-X_{\mathrm{B3}})^2} \\ &= \sqrt{(10-12)^2+(20-18)^2+(30-28)^2}=\sqrt{12}\approx 3.464 \end{aligned}$$

但很快他发现这种简单的距离度量没有考虑到变量之间的相关性。有时，产品的不同技术指标之间可能存在关联，而欧氏距离无法准确地反映这种关系。

本章将介绍随机向量、统计距离与马氏距离、多元正态分布的定义及其相关性质。

1.1 随机向量

1.1.1 随机向量基本定义

假定所讨论的是多个变量的总体，所研究的数据是同时观测 p 个指标(即变量)，进行了 n 次观测得到的(这里 n 实际是样本容量的概念)。可以将 p 个指标表示为 $X_1, X_2, \cdots, X_p$。常用向量

$$\boldsymbol{X} = (X_1, X_2, \cdots, X_p)^{\mathrm{T}}$$

表示对同一个个体观测的 p 个指标。这样经过 n 次观测，则可得到如表 1.1 所示的数据，称每个个体的 p 个指标为一个样品，而全体 n 个样品形成一个样本。

表 1.1 样本容量为 n 的一个多元随机样本

样品序号	变量(指标)			
	X_1	X_2	$\cdots$	X_p
1	x_{11}	x_{12}	$\cdots$	x_{1p}
2	x_{21}	x_{22}	$\cdots$	x_{2p}
⋮	⋮	⋮		⋮
n	x_{n1}	x_{n2}	$\cdots$	x_{np}

横看表 1.1，记

$$\boldsymbol{X}_{(i)} = (x_{i1}, x_{i2}, \cdots, x_{ip})^{\mathrm{T}}, i = 1, 2, \cdots, n$$

表示第 i 个样品的各个指标的观测值。

竖看表 1.1，第 j 列的元素：

$$\boldsymbol{X}_j = (x_{1j}, x_{2j}, \cdots, x_{nj})^{\mathrm{T}}, j = 1, 2, \cdots, p$$

表示对第 j 个变量 X_j 的 n 次观测值。

多元随机向量的样本数据可以用矩阵进行描述：

$$\boldsymbol{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} = (\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_p)$$

【例 1.1】 下面以一个实例来说明具有多个变量的样本数据。表 1.2 是蝴蝶花(IRIS)特征数据。

表 1.2 蝴蝶花(IRIS)特征数据

样品序号	变量			
	萼片长度 sepal length in cm	萼片宽度 sepal width in cm	花瓣长度 petal length in cm	花瓣宽度 width in cm
1	5.1	3.5	1.4	0.2
2	4.9	3	1.4	0.2
3	4.7	3.2	1.3	0.2

续表

样品序号	变量			
	萼片长度 sepal length in cm	萼片宽度 sepal width in cm	花瓣长度 petal length in cm	花瓣宽度 width in cm
4	4.6	3.1	1.5	0.2
5	5	3.6	1.4	0.2
6	5.4	3.9	1.7	0.4
7	4.6	3.4	1.4	0.3
8	5	3.4	1.5	0.2
9	4.4	2.9	1.4	0.2
10	4.9	3.1	1.5	0.1
11	5.4	3.7	1.5	0.2
12	4.8	3.4	1.6	0.2
13	4.8	3	1.4	0.1
14	4.3	3	1.1	0.1
15	5.8	4	1.2	0.2
16	5.7	4.4	1.5	0.4
17	5.4	3.9	1.3	0.4
18	5.1	3.5	1.4	0.3
19	5.7	3.8	1.7	0.3
20	5.1	3.8	1.5	0.3

在 SPSS 软件中，单击**文件→新建→数据表**，然后在“变量视图”表中定义变量名称、类型及其他属性，并在“数据视图”中填充表 1.2 中的数据，如图 1.1 所示。

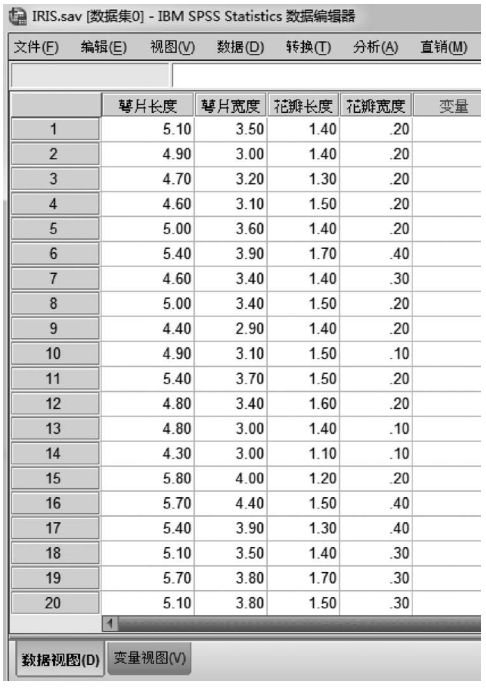


图 1.1　蝴蝶花(IRIS)多元随机样本的 SPSS 数据界面

1.1.2 随机向量分布函数

定义 1.1 设 $X_1,X_2,\cdots,X_p$ 为 p 个随机变量，由它们组成的向量 $\boldsymbol{X}=(X_1,X_2,\cdots,X_p)^{\mathrm{T}}$ 称为 p 维的随机向量。一维随机向量就是随机变量。

定义 1.2 设 $\boldsymbol{X}=(X_1,X_2,\cdots,X_p)^{\mathrm{T}}$ 为一 p 维的随机向量，对任意实数 $x_1,x_2,\cdots,x_p$，则 p 元函数 $F(\boldsymbol{X})=F(x_1,x_2,\cdots,x_p)=P\{X_1\leqslant x_1,X_2\leqslant x_2,\cdots,X_p\leqslant x_p\}$ 称为 p 维随机向量 $\boldsymbol{X}$ 的联合分布函数，或称为分布函数，式中 $\boldsymbol{X}=(x_1,x_2,\cdots,x_p)\in\boldsymbol{R}^p$，记作 $\boldsymbol{X}\sim F(\boldsymbol{X})$。

多元分布函数具有如下性质。

(1) $0\leqslant F(x_1,x_2,\cdots,x_p)\leqslant 1$。

(2) $F(x_1,x_2,\cdots,x_p)$ 是每个变量 x_i 的单调非降右连续函数。

(3) $F(-\infty,x_2,\cdots,x_p)=F(x_1,-\infty,\cdots,x_p)=\cdots=F(x_1,x_2,\cdots,-\infty)=0$。

(4) $F(+\infty,+\infty,\cdots,+\infty)=1$。

以二维随机变量 (X,Y) 为例。如果二维随机变量 (X,Y) 的所有可能取值是有限对或可列无穷多对，则称 (X,Y) 为离散型随机变量。设 (X,Y) 所有可能取的值为 (x_i,y_j)，$i,j=1,2,\cdots,p,\cdots$，并记 $P\{X=x_i,Y=y_j\}=p_{ij}$，$i,j=1,2,\cdots,p,\cdots$，则由概率的定义，我们可知：

$$\sum_{i=1}^{+\infty}\sum_{j=1}^{+\infty}p_{ij}=1,0\leqslant p_{ij}\leqslant 1$$

且分布函数

$$F(X,Y)=P\{X\leqslant x,Y\leqslant y\}=\sum_{x_i\leqslant x}\sum_{y_j\leqslant y}P\{X=x_i,Y=y_j\}$$

我们称 $P\{X=x_i,Y=y_j\}=p_{ij}$，$i,j=1,2,\cdots,p,\cdots$ 为二维随机变量 (X,Y) 的联合概率分布或联合分布律。

定义 1.3 设 p 维随机向量 $\boldsymbol{X}\sim F(\boldsymbol{X})=F(x_1,x_2,\cdots,x_p)$，若存在一个非负的函数 $f(x_1,x_2,\cdots,x_p)$，使得 $F(\boldsymbol{x})=\int_{-\infty}^{x_1}\int_{-\infty}^{x_2}\cdots\int_{-\infty}^{x_p}f(t_1,t_2,\cdots,t_p)\mathrm{d}t_1\mathrm{d}t_2\cdots\mathrm{d}t_p$ 对一切 $\boldsymbol{x}=(x_1,x_2,\cdots,x_p)\in\boldsymbol{R}^p$ 成立，则称 $f(x_1,x_2,\cdots,x_p)$ 为 $\boldsymbol{X}$ 的联合概率密度函数(或分布密度)，并称 $\boldsymbol{X}$ 为 p 维的连续型随机向量。

联合概率密度函数的性质如下。

(1) $f(x_1,x_2,\cdots,x_p)\geqslant 0$，$\forall(x_1,x_2,\cdots,x_p)\in\boldsymbol{R}^p$

(2) $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}f(t_1,t_2,\cdots,t_p)\mathrm{d}t_1\mathrm{d}t_2\cdots\mathrm{d}t_p=1$

(3) $f(x_1,x_2,\cdots,x_p)=\dfrac{\partial^p F(x_1,x_2,\cdots,x_p)}{\partial x_1\partial x_2\cdots\partial x_p}$

(4) 设 D 是 $\boldsymbol{R}^p$ 中的一个区域，点 $(x_1,x_2,\cdots,x_p)$ 落在 D 内的概率为

$$\underbrace{\int\cdots\int}_{D}f(x_1,x_2,\cdots,x_p)\mathrm{d}x_1\mathrm{d}x_2\cdots\mathrm{d}x_p$$

其中，性质(1)和(2)是一个 p 维变量的函数 $f(x_1,x_2,\cdots,x_p)$ 能作为 $\boldsymbol{R}^p$ 中某个随机向量的联合概率密度函数的充分必要条件。

【例 1.2】 若随机向量 (X_1,X_2,X_3) 有函数 $f(x_1,x_2,x_3)=x_1^2+6x_3^2+\frac{1}{3}x_1x_2$，其中 $0<x_1<1,0<x_2<2,0<x_3<\frac{1}{2}$，验证该函数能否成为三维随机向量 (X_1,X_2,X_3) 的分布密度函数。

解　因为 $f(x_1,x_2,x_3)\geqslant 0,\forall (x_1,x_2,x_3)\in\boldsymbol{R}^3$，而且

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}f(x_1,x_2,x_3)\mathrm{d}x_1\mathrm{d}x_2\mathrm{d}x_3=\int_0^{1/2}\int_0^2\int_0^1\left(x_1^2+6x_3^2+\frac{1}{3}x_1x_2\right)\mathrm{d}x_1\mathrm{d}x_2\mathrm{d}x_3$$

$$=\frac{x_1^3}{3}\Big|_0^1\times 2\times\frac{1}{2}+2x_3^3\Big|_0^{1/2}\times 2\times 1+\frac{x_1^2}{6}\Big|_0^1\times\frac{x_2^2}{2}\Big|_0^2\times\frac{1}{2}$$

$$=\frac{1}{3}+\frac{1}{2}+\frac{1}{6}=1$$

所以该函数满足上述分布密度函数性质中(1)和(2)，可以成为三维随机向量的分布密度函数。

1.1.3　多元变量的独立性

定义 1.4　设 $X_1,X_2,\cdots,X_p$ 为 p 个随机变量，如果对于任意的 $x_1,x_2,\cdots,x_p$，满足 $P\{X_1\leqslant x_1,X_2\leqslant x_2,\cdots,X_p\leqslant x_p\}=P\{X_1\leqslant x_1\}P\{X_2\leqslant x_2\}\cdots P\{X_p\leqslant x_p\}$，则称 $X_1,X_2,\cdots,X_p$ 是相互独立的。用以判定随机变量 $X_1,X_2,\cdots,X_p$ 是否独立的方法主要有以下两种。

(1) 如果 X_i 的分布函数为 $F_i(x_i)(i=1,2,\cdots,p)$，它们的联合分布函数为 $F(x_1,x_2,\cdots,x_p)$，则 $X_1,X_2,\cdots,X_p$ 相互独立的充分必要条件是对一切 $x_1,x_2,\cdots,x_p$，有

$$F(x_1,x_2,\cdots,x_p)=\prod_{i=1}^{p}F_i(x_i)$$

(2) 如果 X_i 的密度函数为 $f_i(x_i)(i=1,2,\cdots,p)$，它们的联合分布函数为 $f(x_1,x_2,\cdots,x_p)$，则 $X_1,X_2,\cdots,X_p$ 相互独立的充分必要条件是对一切 $x_1,x_2,\cdots,x_p$，有

$$f(x_1,x_2,\cdots,x_p)=\prod_{i=1}^{p}f_i(x_i)$$

1.1.4　随机向量的数字特征

1. 随机向量 $\boldsymbol{X}$ 的均值

设 $\boldsymbol{X}=(X_1,X_2,\cdots,X_p)^{\mathrm{T}}$ 有 p 个分量。若 $E(X_i)=u_i$ 存在，$i=1,2,\cdots,p$，定义随机向量 $\boldsymbol{X}$ 的均值为

$$E(\boldsymbol{X})=\begin{bmatrix}E(X_1)\\E(X_2)\\\vdots\\E(X_p)\end{bmatrix}=\begin{bmatrix}u_1\\u_2\\\vdots\\u_p\end{bmatrix}=\boldsymbol{u}$$

式中，$\boldsymbol{u}$ 是一个 p 维向量，称为均值向量。

对于任意的随机向量 $\boldsymbol{X}$、$\boldsymbol{Y}$ 及常数矩阵 $\boldsymbol{A}$、$\boldsymbol{C}$，有如下性质：

(1) $E(\boldsymbol{C})=\boldsymbol{C}$

(2) $E(\boldsymbol{X}+\boldsymbol{Y})=E(\boldsymbol{X})+E(\boldsymbol{Y})$

(3) $E(\boldsymbol{CX})=\boldsymbol{C}E(\boldsymbol{X})$

(4) $E(\boldsymbol{CXA})=\boldsymbol{C}E(\boldsymbol{X})\boldsymbol{A}$

2. 随机向量 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 的协方差阵

定义数 $\sigma_{ij}=E(X_i-EX_i)(Y_j-EY_j)$ 为 X_i 和 Y_j 的协方差，$i,j=1,2,\cdots,p$，通常记为 $\mathrm{cov}(X_i,Y_j)=\sigma_{ij}$。由数学期望的性质可以证明协方差有以下性质：

(1) $\mathrm{cov}(X_i,Y_j)=\mathrm{cov}(Y_j,X_i)$

(2) $\mathrm{cov}(X_i,X_i)=\sigma_{ii}=D(X_i)$

(3) $\mathrm{cov}(aX_i,bY_j)=ab\times\mathrm{cov}(X_i,Y_j)$

(4) $\mathrm{cov}(X_i+X_k,Y_j)=\mathrm{cov}(X_i,Y_j)+\mathrm{cov}(X_k,Y_j)$

设 $\boldsymbol{X}=(X_1,X_2,\cdots,X_p)^{\mathrm{T}}$ 和 $\boldsymbol{Y}=(Y_1,Y_2,\cdots,Y_n)^{\mathrm{T}}$ 分别为 p 维和 n 维随机向量，它们之间的协方差阵定义为一个 $p\times n$ 的矩阵，其元素是 $\mathrm{cov}(X_i,Y_j)$，即

$$\mathrm{cov}(\boldsymbol{X},\boldsymbol{Y})=(\mathrm{cov}(X_i,Y_j)),i=1,2,\cdots,p,j=1,2,\cdots,n$$

特别地：

(1) 若 $\mathrm{cov}(\boldsymbol{X},\boldsymbol{Y})=\boldsymbol{0}$，则 $\boldsymbol{X},\boldsymbol{Y}$ 是不相关的；

(2) $\boldsymbol{\Sigma}=\mathrm{cov}(\boldsymbol{X},\boldsymbol{X})=E(\boldsymbol{X}-E\boldsymbol{X})(\boldsymbol{X}-E\boldsymbol{X})^{\mathrm{T}}=D(\boldsymbol{X})$

$$=\begin{bmatrix} D(X_1) & \mathrm{cov}(X_1,X_2) & \cdots & \mathrm{cov}(X_1,X_p) \\ \mathrm{cov}(X_2,X_1) & D(X_2) & \cdots & \mathrm{cov}(X_2,X_p) \\ \vdots & \vdots & & \vdots \\ \mathrm{cov}(X_p,X_1) & \mathrm{cov}(X_p,X_2) & \cdots & D(X_p) \end{bmatrix}$$，可以看出其协方差阵是 $p\times p$ 的对称阵，同时总是非负定的。

对于任意的随机向量 $\boldsymbol{X}$、$\boldsymbol{Y}$ 及常数矩阵 $\boldsymbol{A}$、$\boldsymbol{C}$，协方差阵有如下性质：

(1) $D(\boldsymbol{C})=\boldsymbol{0}$

(2) $D(\boldsymbol{X}+\boldsymbol{C})=D(\boldsymbol{X})$

(3) $D(\boldsymbol{CX})=\boldsymbol{C}D(\boldsymbol{X})\boldsymbol{C}^{\mathrm{T}}=\boldsymbol{C\Sigma C}^{\mathrm{T}}$

(4) $\mathrm{cov}(\boldsymbol{AX},\boldsymbol{CY})=\boldsymbol{A}\,\mathrm{cov}(\boldsymbol{X},\boldsymbol{Y})\boldsymbol{C}^{\mathrm{T}}$

(5) 设 $\boldsymbol{X}=(X_1,X_2,\cdots,X_p)^{\mathrm{T}}$ 为 p 维随机向量，其期望和协方差存在，记 $\boldsymbol{u}=E(\boldsymbol{X})$，$\boldsymbol{\Sigma}=D(\boldsymbol{X})$，$\boldsymbol{C}$ 为 $p\times p$ 常数阵，则有

$$E(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{CX})=\mathrm{tr}(\boldsymbol{C\Sigma})+\boldsymbol{u}^{\mathrm{T}}\boldsymbol{Cu}$$

3. 随机向量 $\boldsymbol{X}$ 的相关阵

若 $i\neq j$，协方差 $\mathrm{cov}(X_i,X_j)=\sigma_{ij}=0$，则变量 X_i 和 X_j 不相关。

定义 X_i 和 X_j 的相关系数 $r_{ij}=\dfrac{\mathrm{cov}(X_i,X_j)}{\sqrt{D(X_i)}\sqrt{D(X_j)}}$，$i,j=1,2,\cdots,p$，显然，$r_{ij}=r_{ji}$，$r_{ii}=1$。

相关系数有如下重要性质。

(1) $|r_{ij}|\leqslant 1$。

(2) $|r_{ij}|=1$ 的充分必要条件是存在常数 a,b,使得 $P\{X_j=aX_i+b\}=1$,即线性相关。

(3) 当 X_i,X_j 相互独立时,$r_{ij}=0$。

(4) 如果 $r_{ij}>0$,则 X_i,X_j 趋于正相关;如果 $r_{ij}<0$,则 X_i,X_j 趋于负相关。

(5) 对于任意常数 a_i,a_j,b_i,b_j,设 $Y_i=a_iX_i+b_i,Y_j=a_jX_j+b_j$,那么 Y_i 和 Y_j 之间的相关系数 r_{yij} 满足

$$r_{yij}=\begin{cases} r_{ij} & a_ia_j>0 \\ -r_{ij} & a_ia_j<0 \\ 0 & a_ia_j=0 \end{cases}\text{(证明留给读者)}$$

在相关系数的基础上,随机向量 $\boldsymbol{X}$ 的相关阵定义为

$$\boldsymbol{R}=\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1p} \\ r_{21} & r_{22} & \cdots & r_{2p} \\ \vdots & \vdots & & \vdots \\ r_{p1} & r_{p2} & \cdots & r_{pp} \end{bmatrix}=(r_{ij})_{p\times p}$$

在数据处理时,为了避免由于指标的量纲不同给统计分析结果带来的影响,往往在使用某种统计分析方法之前,将每个指标“标准化”,即做如下类似以前一元统计分析中的 Z 转换:

$$X_i^*=\frac{X_i-E(X_i)}{\sqrt{D(X_i)}},\quad i=1,2,\cdots,p$$

这样得到,$\boldsymbol{X}^*=(X_1^*,X_2^*,\cdots,X_p^*)$,可知 $E(\boldsymbol{X}^*)=\boldsymbol{0}$,$D(\boldsymbol{X}^*)=\boldsymbol{R}$,即标准化数据的协方差阵正好是原指标的相关阵。

【例 1.3】 计算表 1.2 中蝴蝶花(IRIS)的协方差阵和相关阵

视频 1-1

解　在 SPSS 软件中,提取协方差阵和相关阵操作步骤如下。

(1) 依次单击**分析→度量→可靠性分析**打开可靠性分析对话框,如图 1.2 所示。将变量移入项目列表。

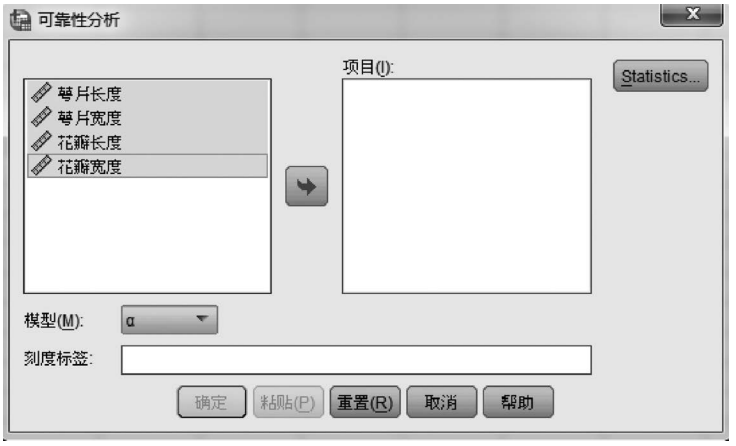


图 1.2　可靠性分析对话框

（2）单击 **Statistics**，弹出图 1.3 所示对话框，勾选项之间的**相关性**和**协方差**，然后单击**继续**和**确定**按钮，提交运算，输出结果，如图 1.4 所示。

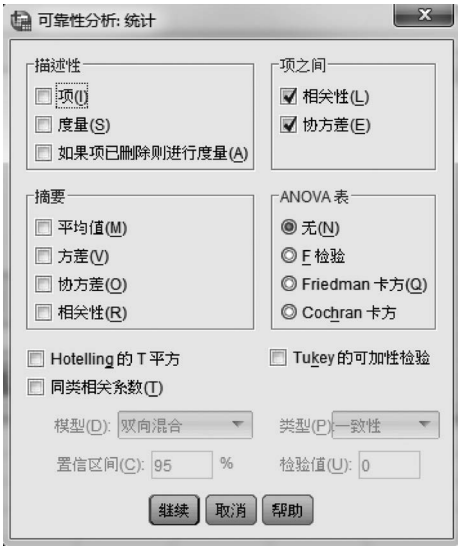


图 1.3 “可靠性分析：统计”对话框

项间相关性矩阵

	萼片长度	萼片宽度	花瓣长度	花瓣宽度
萼片长度	1.000	.876	.317	.562
萼片宽度	.876	1.000	.260	.753
花瓣长度	.317	.260	1.000	.369
花瓣宽度	.562	.753	.369	1.000

项间协方差矩阵

	萼片长度	萼片宽度	花瓣长度	花瓣宽度
萼片长度	.182	.152	.020	.022
萼片宽度	.152	.166	.015	.029
花瓣长度	.020	.015	.021	.005
花瓣宽度	.022	.029	.005	.009

图 1.4 相关矩阵与协方差矩阵结果

【例 1.4】 设随机向量 $\boldsymbol{X}=(X_1,X_2,X_3)=\begin{bmatrix}5 & 6 & 8\\ 7 & 9 & 4\\ 2 & 4 & 5\\ 3 & 4 & 6\end{bmatrix}$，计算协方差阵和相关阵。

解 已知变量个数 $p=3$，样品数 $n=4$，随机向量中的元素可用 x_{ij} 表示，$i=1,2,3,4$，$j=1,2,3$。首先计算每个变量下样品对应的平均值：

	x_{i1}	x_{i2}	x_{i3}
	5	6	8
	7	9	4
	2	4	5
	3	4	6
$\bar{x}_{\cdot j}$，$j=1,2,3$	4.25	5.75	5.75

然后计算每个变量值与其平均值的差值：

$x_{i1}-\bar{x}_{\cdot 1}$	$x_{i2}-\bar{x}_{\cdot 2}$	$x_{i3}-\bar{x}_{\cdot 3}$
0.75	0.25	2.25
2.75	3.25	−1.75
−2.25	−1.75	−0.75
−1.25	−1.75	0.25

根据协方差的定义计算

$$\begin{aligned}\operatorname{cov}(X_1,X_1)=D(X_1)&=\frac{1}{4-1}\times(0.75,2.75,-2.25,-1.25)\times\\&\quad(0.75,2.75,-2.25,-1.25)'\\&=4.916\ 67\end{aligned}$$

类似地，$\operatorname{cov}(X_1,X_2)=\frac{1}{4-1}\times(0.75,2.75,-2.25,-1.25)\times(0.25,3.25,-1.75,-1.75)'=5.083\ 33$，相应地可以得到协方差矩阵中的所有元素。

对于相关矩阵，譬如求 $r_{12}=\frac{\operatorname{cov}(X_1,X_2)}{\sqrt{D(X_1)}\sqrt{D(X_2)}}=\frac{5.083\ 33}{\sqrt{4.916\ 67}\sqrt{5.583\ 33}}=0.970\ 2$，具体的相关矩阵和协方差矩阵可见表 1.3 和表 1.4。

表 1.3　相 关 矩 阵

项间相关性矩阵			
	X_1	X_2	X_3
X_1	1.000	0.970	−0.154
X_2	0.970	1.000	−0.351
X_3	−0.154	−0.351	1.000

表 1.4　协方差矩阵

项间协方差矩阵			
	X_1	X_2	X_3
X_1	4.917	5.083	−0.583
X_2	5.083	5.583	−1.417
X_3	−0.583	−1.417	2.917

1.2　统 计 距 离

在多指标统计分析如判别分析和系统聚类分析时，距离的概念非常重要，通常用距离来度量不同样本间的相似性。距离的定义方法有多种，通常情况下，我们所说的距离是指欧氏距离，但在多元统计分析中，也常常使用马氏距离。本节主要介绍这两种距离。

1.2.1　欧氏距离

欧氏距离是最易于理解的一种距离计算方法，几何上它是空间中两个点之间的真实距离。

若点 $\boldsymbol{x}=(x_1,x_2,\cdots,x_p)^{\mathrm{T}}$ 和 $\boldsymbol{y}=(y_1,y_2,\cdots,y_p)^{\mathrm{T}}$ 是 p 维空间中任意两点，则这两点的欧氏距离定义为

$$\boldsymbol{d}(\boldsymbol{x},\boldsymbol{y})=\sqrt{(\boldsymbol{x}-\boldsymbol{y})^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{y})}$$
$$=\sqrt{(x_1-y_1)^2+(x_2-y_2)^2+\cdots+(x_p-y_p)^2}$$

点 $\boldsymbol{x}=(x_1,x_2,\cdots,x_p)^{\mathrm{T}}$ 到总体 $\boldsymbol{G}$ 的欧氏距离定义为

$$\boldsymbol{d}(\boldsymbol{x},\boldsymbol{G})=\sqrt{(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{\mu})}$$
$$=\sqrt{(x_1-\mu_1)^2+(x_2-\mu_2)^2+\cdots+(x_p-\mu_p)^2}$$

其中，$\boldsymbol{\mu}=(\mu_1,\mu_2,\cdots,\mu_p)^{\mathrm{T}}$ 是总体 $\boldsymbol{G}$ 的均值。

欧氏距离虽然简单、易于计算，但它也存在明显的缺点，主要包括以下几方面。

1. 没有考虑各分量量纲的差异

欧氏距离与向量各分量的量纲有关，(如果各分量)量纲不全相同，则上述公式计算的欧氏距离常常是无意义的。

2. 同等看待各分量对距离的贡献

在多指标统计分析时，采用欧氏距离的统计结果有时并不符合实际。这是因为各分量往往具有不同的波动程度，波动程度大的分量对欧氏距离起着决定性作用，而波动程度小的分量起到的作用则常常微乎其微。图 1.5 显示了某次收入和受教育年限调研的结果，很明显收入的波动程度远高于受教育年限，如果用欧氏距离计算样本点之间的距离，则会夸大收入的影响。

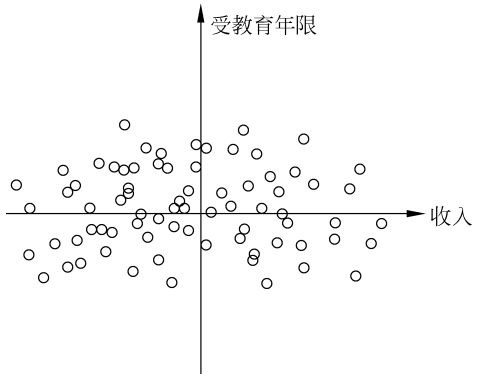


图 1.5　收入—受教育年限散点图

在实际统计分析时，为了消除量纲和不同波动程度的影响，通常对各分量进行标准化预处理，然后再计算欧氏距离。经过简单的推导就可以得到两点间标准化欧氏距离计算公式：

$$\boldsymbol{d}(\boldsymbol{x}^*,\boldsymbol{y}^*)=\sqrt{(\boldsymbol{x}^*-\boldsymbol{y}^*)^{\mathrm{T}}(\boldsymbol{x}^*-\boldsymbol{y}^*)}$$
$$=\sqrt{\frac{(x_1-y_1)^2}{S_1}+\frac{(x_2-y_2)^2}{S_2}+\cdots+\frac{(x_p-y_p)^2}{S_p}}$$

易知，该公式计算的距离是一标量。如果将方差的倒数视为权重，则上述公式计算的距离可以看成一种加权欧氏距离。以图 1.5 为例，标准化的处理使波动程度大的收入变量相对压缩，而波动程度小的受教育年限变量相对扩张，标准化后的各变量波动程度一致。

3. 没有考虑变量间的相关性

欧氏距离没有考虑变量间的相关性，加权欧氏距离能够消除量纲和波动差异的影响，但不能消除变量之间相关性的影响，以致采用欧氏距离统计得到的结果有时并不理想。对此，印度统计学家 P. C. 马哈拉诺比斯(P. C. Mahalanobis)于 1936 提出“马氏距离”的距离度量方法。

1.2.2　马氏距离

以二元统计分析为例，设对 n 个样品观测两个指标 x_1 和 x_2 得到的散点图如图 1.6 所示。

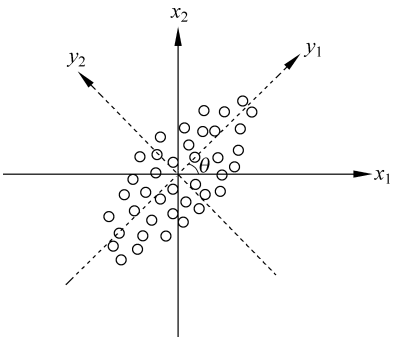


图 1.6　散点图

由图 1.6 可知，指标 x_1 和 x_2 存在着某种相关性。为了消除相关性的影响，在几何上可将坐标轴按逆时针旋转 θ 度，得到新的坐标轴，使得样本点在新的坐标系下的指标 y_1 和 y_2 互不相关。在此基础上，再通过计算加权欧氏距离来消除量纲和波动程度的影响，此距离即为马氏距离。

具体来讲，假设 $\boldsymbol{x}_1$ 和 $\boldsymbol{x}_2$ 是均值为$\boldsymbol{\mu}$、协方差阵为$\boldsymbol{\Sigma}$的总体 $\boldsymbol{G}$ 中的两个样本，则通过求协方差阵的特征值 $\lambda_1,\lambda_2,\cdots,\lambda_p$ 与对应的标准正交特征向量$\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2,\cdots,\boldsymbol{\gamma}_p$ 将协方差阵对角化，存在$\boldsymbol{\Lambda}=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_p)$，$\boldsymbol{P}=(\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2,\cdots,\boldsymbol{\gamma}_p)$，使得$\boldsymbol{\Sigma}=\boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^{\mathrm{T}}$。在此情形下，$\boldsymbol{y}=\boldsymbol{P}^{\mathrm{T}}\boldsymbol{x}=(y_1,y_2,\cdots,y_p)^{\mathrm{T}}$ 则表示点 $\boldsymbol{x}$ 在经正交旋转后的新坐标体系下对应的坐标。于是 $\boldsymbol{y}_1=\boldsymbol{P}^{\mathrm{T}}\boldsymbol{x}_1(y_{11},y_{12},\cdots,y_{1p})^{\mathrm{T}}$ 和 $\boldsymbol{y}_2=\boldsymbol{P}^{\mathrm{T}}\boldsymbol{x}_2=(y_{21},y_{22},\cdots,y_{2p})^{\mathrm{T}}$。由于

$$D(\boldsymbol{y})=D(\boldsymbol{P}^{\mathrm{T}}\boldsymbol{x})=\boldsymbol{P}^{\mathrm{T}}D(\boldsymbol{x})\boldsymbol{P}=\boldsymbol{\Lambda}$$

所以在新坐标体系下，分量 $y_1,y_2,\cdots,y_p$ 互不相关，消除了原始数据中各分量间相关性的影响。为了进一步消除分量 $y_1,y_2,\cdots,y_p$ 不同波动程度的影响，对各分量进行标准化处理，计算标准化欧氏距离得

$$\begin{aligned}d_p^2(\boldsymbol{x}_1,\boldsymbol{x}_2)&=\frac{(y_{11}-y_{22})^2}{\lambda_1}+\frac{(x_{12}-y_{22})^2}{\lambda_2}+\cdots+\frac{(x_{1p}-y_{2p})^2}{\lambda_p}\\&=(\boldsymbol{y}_1-\boldsymbol{y}_2)^{\mathrm{T}}\boldsymbol{\Lambda}^{-1}(\boldsymbol{y}_1-\boldsymbol{y}_2)\\&=(\boldsymbol{P}^{\mathrm{T}}\boldsymbol{x}_1-\boldsymbol{P}^{\mathrm{T}}\boldsymbol{x}_2)^{\mathrm{T}}\boldsymbol{\Lambda}^{-1}(\boldsymbol{P}^{\mathrm{T}}\boldsymbol{x}_1-\boldsymbol{P}^{\mathrm{T}}\boldsymbol{x}_2)\\&=(\boldsymbol{x}_1-\boldsymbol{x}_2)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_1-\boldsymbol{x}_2)\end{aligned}$$

显然，如果协方差矩阵为单位矩阵，马氏距离就简化为欧氏距离；如果协方差矩阵为对角阵，马氏距离则可称为加权欧氏距离。

基于上述讨论，我们可以得到马氏距离的定义：$\boldsymbol{x}_1$ 和 $\boldsymbol{x}_2$ 是均值为$\boldsymbol{\mu}$、协方差阵为$\boldsymbol{\Sigma}$ 的总体 $\boldsymbol{G}$ 中的两个维度为 p 的样本，则 $\boldsymbol{x}_1$ 和 $\boldsymbol{x}_2$ 两点的马氏距离为

$$d_p^2(\boldsymbol{x}_1,\boldsymbol{x}_2)=(\boldsymbol{x}_1-\boldsymbol{x}_2)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_1-\boldsymbol{x}_2)$$

点 $\boldsymbol{x}$ 到总体 $\boldsymbol{G}$ 的马氏距离为

$$d_p^2(\boldsymbol{x},\boldsymbol{G})=(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$$

设 D 表示一个点集，$d(x_1,x_2)$表示点 x_1 和 x_2 之间的距离，可以证明，马氏距离满足以下距离的基本公理。

(1) $d(x_1,x_2)\geqslant 0$，当且仅当 $x_1=x_2$ 时，$d(x_1,x_2)=0$，$\forall x_1,x_2\in D$。(非负性)

(2) $d(x_1,x_2)=d(x_2,x_1)$，$\forall x_1,x_2\in D$。(对称性)

(3) $d(x_1,x_2)\leqslant d(x_1,x_3)+d(x_3,x_2)$，$\forall x_1,x_2,x_3\in D$。(三角不等式)

【例 1.5】 体温测量与身高体重关系

医学研究人员正在研究体温与身高、体重之间的关系。为了收集数据，你随机选取了一组参与者，并测量了他们的体温、身高和体重。现在，有了每个参与者的三个变量数据：体温、身高和体重。假设研究以下四位参与者的体温、身高和体重数据：参与者 A：(体温＝36.5 ℃，身高＝170 cm，体重＝65 kg)；参与者 B：(体温＝36.8 ℃，身高＝175 cm，体重＝70 kg)；参与者 C：(体温＝37.0 ℃，身高＝165 cm，体重＝60 kg)；参与者 D：(体温＝36.6 ℃，身高＝172 cm，体重＝68 kg)。

(1) 欧氏距离计算。对于参与者 A 和 B，欧氏距离 $d(\mathrm{A},\mathrm{B})$可以通过以下公式计算：

$$\begin{aligned}d(\mathrm{A},\mathrm{B})&=\sqrt{(36.8-36.5)^2+(175-170)^2+(70-65)^2}\\&=\sqrt{(0.09+25+25)}\approx 7.08\end{aligned}$$

类似地，我们可以计算 A 和 C，A 和 D，B 和 C，B 和 D，C 和 D 之间的欧氏距离：$d(\mathrm{A},\mathrm{C})\approx 7.09$；$d(\mathrm{A},\mathrm{D})\approx 3.61$；$d(\mathrm{B},\mathrm{C})\approx 14.14$；$d(\mathrm{B},\mathrm{D})\approx 3.61$；$d(\mathrm{C},\mathrm{D})\approx 10.64$。

(2) 马氏距离计算。在计算马氏距离时，我们需要先计算数据的协方差矩阵 $\boldsymbol{S}$。在这个例子中，我们计算体温、身高和体重的样本协方差矩阵如下：

$$\boldsymbol{S}=\begin{bmatrix}0.010\,0 & 1.583\,3 & 0.708\,3\\ 1.583\,3 & 26.666\,7 & 11.333\,3\\ 0.708\,3 & 11.333\,3 & 6.666\,7\end{bmatrix}$$

然后，我们计算马氏距离 $d_p(\mathrm{A},\mathrm{B})$及其他组合的马氏距离：

$$d_p(\mathrm{A},\mathrm{B})=\sqrt{(\boldsymbol{X}_{\mathrm{A}}-\boldsymbol{X}_{\mathrm{B}})^{\mathrm{T}}\boldsymbol{S}^{-1}(\boldsymbol{X}_{\mathrm{A}}-\boldsymbol{X}_{\mathrm{B}})}\approx 1.00$$

其中，$\boldsymbol{X}_{\mathrm{A}}$，$\boldsymbol{X}_{\mathrm{B}}$ 表示两个参与者 A，B 的数据向量(体温、身高、体重组成的向量)。类似地，我们可以计算 $d_p(\mathrm{A},\mathrm{C})$，$d_p(\mathrm{A},\mathrm{D})$，$d_p(\mathrm{B},\mathrm{C})$，$d_p(\mathrm{B},\mathrm{D})$，$d_p(\mathrm{C},\mathrm{D})$的值：$d_p(\mathrm{A},\mathrm{C})\approx 1.41$；$d_p(\mathrm{A},\mathrm{D})\approx 0.35$；$d_p(\mathrm{B},\mathrm{C})\approx 1.73$；$d_p(\mathrm{B},\mathrm{D})\approx 1.63$；$d_p(\mathrm{C},\mathrm{D})\approx 1.34$。

通过这个例子，我们可以看到欧氏距离衡量了样本在多个变量上的差异，而马氏距离考虑了变量之间的相关性，可以更准确地衡量样本之间的差异。在这个例子中，可以观察到参与者 C 和 D 在欧氏距离上的差异较大，但在马氏距离上的差异较小，这反映了他们在体温、身高和体重之间存在一定的相关性。

【例 1.6】 引导案例解答

马氏距离不仅考虑了变量之间的相关性，还考虑了随机向量 $\boldsymbol{X}$ 的协方差矩阵。产品 A 和 B 之间的马氏距离计算公式为：$d_p(\mathrm{A},\mathrm{B})=\sqrt{(\boldsymbol{X}_{\mathrm{A}}-\boldsymbol{X}_{\mathrm{B}})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_{\mathrm{A}}-\boldsymbol{X}_{\mathrm{B}})}$，其中，T 表示矩阵转置，$\boldsymbol{\Sigma}^{-1}$ 表示协方差矩阵的逆矩阵。通过代入数值计算，可得到：

$$(\boldsymbol{X}_{\mathrm{A}}-\boldsymbol{X}_{\mathrm{B}})=\begin{bmatrix}-2\\ 2\\ 2\end{bmatrix},\quad (\boldsymbol{X}_{\mathrm{A}}-\boldsymbol{X}_{\mathrm{B}})^{\mathrm{T}}=[-2\quad 2\quad 2]$$

$$\boldsymbol{\Sigma}^{-1} \approx \begin{bmatrix} 0.277\,8 & -0.062\,5 & -0.083\,3 \\ -0.062\,5 & 0.125 & -0.083\,3 \\ -0.083\,3 & -0.083\,3 & 0.25 \end{bmatrix}$$

将这些结果代入马氏距离的公式可得到 $d_p(\mathrm{A},\mathrm{B}) \approx \sqrt{0.555\,6+0.25-0.333\,2} \approx \sqrt{0.472\,4} \approx 0.687\,4$。

马氏距离相对于欧氏距离在描述和度量多维数据之间的差异时更为准确和可靠。考虑到不同技术指标之间可能存在的相关性，小李使用马氏距离有助于更精确地评估产品之间的差异程度，提高质量控制的效果。在整个质量控制过程中，随机向量、统计距离与马氏距离、多元正态分布紧密相连，为小李提供了强大的分析工具。借助这些工具，他成功地提高了产品质量的控制准确性和效率，为公司的发展贡献了自己的一份力量。

1.3　多元正态分布

多元正态分布是一元正态分布及二元正态分布的自然推广，内容更为丰富。迄今为止，多元分析的主要理论都是建立在多元正态总体基础上的。

1.3.1　多元正态分布的定义

我们已经知道一元正态分布的密度函数为

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x-u)^2}{2\sigma^2}},\quad \sigma>0$$

该函数可以改写成

$$f(x)=(2\pi)^{-1/2}\sigma^{-1}\exp\left[-\frac{1}{2}(x-u)'(\sigma^2)^{-1}(x-u)\right]$$

定义 1.5　如果 p 维随机向量 $\boldsymbol{X}=(X_1,X_2,\cdots X_p)^{\mathrm{T}}$ 的联合概率密度函数为

$$f(x_1,x_2,\cdots,x_p)=\frac{1}{(2\pi)^{p/2}\,|\boldsymbol{\Sigma}|^{1/2}}\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{u})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{u})\right\},$$

$$\boldsymbol{\Sigma}>0$$

则称 $\boldsymbol{X}=(X_1,X_2,\cdots X_p)^{\mathrm{T}}$ 服从 p 元正态分布，也称 $\boldsymbol{X}$ 为 p 元正态变量。记为：$\boldsymbol{X}\sim N_p(\boldsymbol{u},\boldsymbol{\Sigma})$，其中 $|\boldsymbol{\Sigma}|$ 为协方差阵 $\boldsymbol{\Sigma}$ 的行列式。

当 $p=2$ 时，可以得到二元正态分布的概率密度函数。

设 $\boldsymbol{X}=(X_1,X_2)^{\mathrm{T}}$ 服从二元正态分布，则 $\boldsymbol{\Sigma}=\begin{bmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{21} & \sigma_{22}\end{bmatrix}=\begin{bmatrix}\sigma_1^2 & \sigma_1\sigma_2 r\\ \sigma_2\sigma_1 r & \sigma_2^2\end{bmatrix}$，$r\neq\pm1$，

其中 σ_1^2 和 σ_2^2 分别是 X_1 与 X_2 的方差，r 是 X_1 与 X_2 的相关系数。这样，

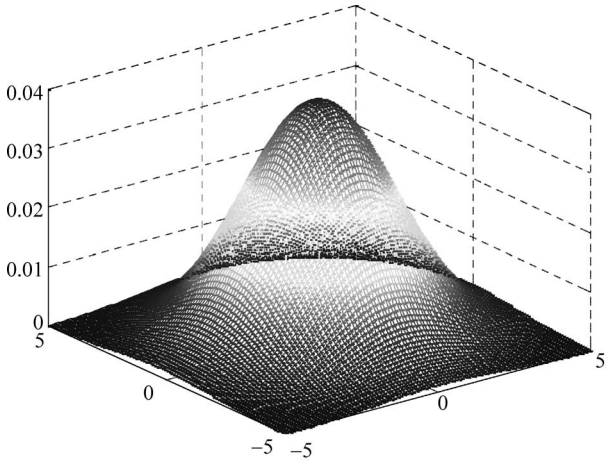
$$|\boldsymbol{\Sigma}|=\sigma_1^2\sigma_2^2(1-r^2)$$

$$\boldsymbol{\Sigma}^{-1}=\frac{1}{\sigma_2^2\sigma_1^2(1-r^2)}\begin{bmatrix}\sigma_2^2 & -\sigma_1\sigma_2 r\\ -\sigma_2\sigma_1 r & \sigma_1^2\end{bmatrix}$$

故，X_1 与 X_2 的概率密度函数为

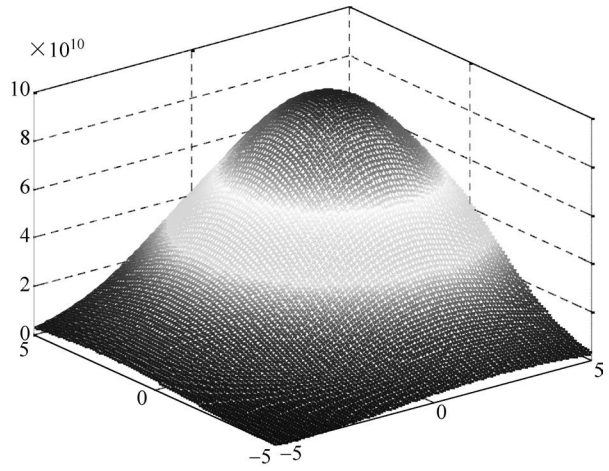
$$f(x_1,x_2)=\frac{1}{2\pi\sigma_1\sigma_2(1-r^2)^{1/2}}\exp\left\{-\frac{1}{2(1-r^2)}\left[\frac{(x_1-u_1)^2}{\sigma_1^2}-2r\frac{(x_1-u_1)(x_2-u_2)}{\sigma_1\sigma_2}+\frac{(x_2-u_2)^2}{\sigma_2^2}\right]\right\}$$

二维正态分布的图形如图 1.7 和图 1.8 所示。



$u_1=0.1,u_2=0.2,\sigma_1=\sigma_2=2,r=0$

图 1.7　二维正态分布(1)



$u_1=u_2=2,\sigma_1=\sigma_2=3,r=0.2$

图 1.8　二维正态分布(2)

当 $\boldsymbol{X}\sim N_p(\boldsymbol{u},\boldsymbol{\Sigma})$，$\boldsymbol{\Sigma}>0$ 时，$\boldsymbol{X}$ 的密度等高面是一族椭球或椭圆。$(\boldsymbol{x}-\boldsymbol{u})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{u})=\alpha^2$，随着 α 的大小不同，得到不同的椭球。

【例 1.7】　设随机向量 $\boldsymbol{X}=(X_1,X_2)^{\mathrm{T}}$ 服从二元正态分布 $f(x_1,x_2)=\dfrac{1}{2\pi\sigma_1\sigma_2(1-r^2)^{1/2}}$

$\exp\left\{-\frac{1}{2(1-r^2)}\left[\frac{(x_1-u_1)^2}{\sigma_1^2}-2r\frac{(x_1-u_1)(x_2-u_2)}{\sigma_1\sigma_2}+\frac{(x_2-u_2)^2}{\sigma_2^2}\right]\right\}$，证明 X_1 与 X_2 的相关系数就是概率密度函数中的 r。

证明　可以求得 X_1、X_2 的边缘概率密度函数分别是

$$f(x_1)=\frac{1}{\sqrt{2\pi}\sigma_1}\exp\left\{-\left(\frac{x_1-u_1}{\sqrt{2}\sigma_1}\right)^2\right\},\quad -\infty<x_1<+\infty$$

$$f(x_2)=\frac{1}{\sqrt{2\pi}\sigma_2}\exp\left\{-\left(\frac{x_2-u_2}{\sqrt{2}\sigma_2}\right)^2\right\},\quad -\infty<x_2<+\infty$$

故 $E(X_1)=u_1$，$E(X_2)=u_2$，$D(X_1)=\sigma_1^2$，$D(X_2)=\sigma_2^2$。

而

$$\begin{aligned}\operatorname{cov}(X_1,X_2)&=E(X_1-EX_1)(X_2-EX_2)\\&=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x_1-u_1)(x_2-u_2)f(x_1,x_2)\mathrm{d}x_1\mathrm{d}x_2\\&=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\frac{(x_1-u_1)(x_2-u_2)}{2\pi\sigma_1\sigma_2(1-r^2)^{1/2}}\exp\left\{-\frac{1}{2(1-r^2)}\left[\frac{(x_1-u_1)^2}{\sigma_1^2}-\right.\right.\\&\quad\left.\left.2r\frac{(x_1-u_1)(x_2-u_2)}{\sigma_1\sigma_2}+\frac{(x_2-u_2)^2}{\sigma_2^2}\right]\right\}\mathrm{d}x_1\mathrm{d}x_2\\&=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\frac{\sigma_1\sigma_2\mu\nu}{2\pi(1-r^2)^{1/2}}\exp\left\{-\frac{1}{2(1-r^2)}(\mu^2-2r\mu\nu+\nu^2)\right\}\mathrm{d}\mu\mathrm{d}\nu\end{aligned}$$

$\left(\text{令 }\mu=\frac{x_1-u_1}{\sigma_1},\nu=\frac{x_2-u_2}{\sigma_2}\right)$

$$\begin{aligned}&=\frac{\sigma_1\sigma_2}{2\pi(1-r^2)^{1/2}}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\mu\nu\exp\left\{-\frac{1}{2(1-r^2)}[(\mu-r\nu)^2+(1-r^2)\nu^2]\right\}\mathrm{d}\mu\mathrm{d}\nu\\&=\frac{\sigma_1\sigma_2}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\left\{\nu\exp\left(-\frac{\nu^2}{2}\right)\int_{-\infty}^{+\infty}\frac{\mu}{\sqrt{2\pi}\sqrt{1-r^2}}\exp\left[-\frac{1}{2(1-r^2)}(\mu-\nu r)^2\right]\mathrm{d}\mu\right\}\mathrm{d}\nu\end{aligned}$$

其中，积分 $\int_{-\infty}^{+\infty}\frac{\mu}{\sqrt{2\pi}\sqrt{1-r^2}}\exp\left[-\frac{1}{2(1-r^2)}(\mu-\nu r)^2\right]\mathrm{d}\mu$ 恰好是服从正态分布 $N(r\nu,1-r^2)$ 的随机变量的数学期望 $r\nu$，于是，

$$\begin{aligned}\operatorname{cov}(X_1,X_2)&=\frac{\sigma_1\sigma_2 r}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\nu^2\exp\left\{-\frac{\nu^2}{2}\right\}\mathrm{d}\nu\\&=\frac{\sigma_1\sigma_2 r}{\sqrt{2\pi}}\left[-\nu\exp\left\{-\frac{\nu^2}{2}\right\}\right]\Bigg|_{-\infty}^{+\infty}+\int_{-\infty}^{+\infty}\exp\left\{-\frac{\nu^2}{2}\right\}\mathrm{d}\nu=\sigma_1\sigma_2 r\end{aligned}$$

所以有相关系数

$$r=\frac{\operatorname{cov}(X_1,X_2)}{\sqrt{D(X_1)}\sqrt{D(X_2)}}=\frac{\sigma_1\sigma_2 r}{\sigma_1\sigma_2}=r$$

定理 1.1　设 $\boldsymbol{X}\sim N_p(\boldsymbol{u},\boldsymbol{\Sigma})$，则 $E(\boldsymbol{X})=\boldsymbol{u}$，$D(\boldsymbol{X})=\boldsymbol{\Sigma}$。

定理 1.1 将正态分布的参数 $\boldsymbol{u}$ 和 $\boldsymbol{\Sigma}$ 赋予了明确的统计意义。

1.3.2 多元正态分布的性质

(1) 如果 p 维随机向量 $\boldsymbol{X}\sim N_p(\boldsymbol{u},\boldsymbol{\Sigma})$，随机向量 $\boldsymbol{Y}=\boldsymbol{BX}+\boldsymbol{C}$，其中 $\boldsymbol{B}$ 是一个 $n\times p$ 的常数矩阵，$\boldsymbol{C}$ 是 n 维常向量，则 $\boldsymbol{Y}$ 服从 $N_n(\boldsymbol{Bu}+\boldsymbol{C},\boldsymbol{B\Sigma B}^{\mathrm{T}})$，即 $\boldsymbol{Y}$ 服从 n 元正态分布，其均值向量为 $\boldsymbol{Bu}+\boldsymbol{C}$，协方差阵为 $\boldsymbol{B\Sigma B}^{\mathrm{T}}$。由此可见，正态随机向量的线性组合仍然是正态向量。

(2) 多元正态分布随机向量 $\boldsymbol{X}\sim N_p(\boldsymbol{u},\boldsymbol{\Sigma})$，则 $\boldsymbol{X}=(X_1,X_2,\cdots,X_p)^{\mathrm{T}}$ 的任何一个分量子集的分布(称为 $\boldsymbol{X}$ 的边缘分布)仍然服从正态分布；但是，反过来，如果一个随机向量的任何边缘分布均为正态，并不能推导出该随机向量服从多元正态分布。

例如，设 $\boldsymbol{X}=(X_1,X_2)^{\mathrm{T}}$ 有联合分布密度

$$f(x_1,x_2)=\frac{1}{2\pi}\mathrm{e}^{-\frac{1}{2}(x_1^2+x_2^2)}\left[1+x_1x_2\mathrm{e}^{-\frac{1}{2}(x_1^2+x_2^2)}\right]$$

容易验证，$X_1\sim N(0,1)$，$X_2\sim N(0,1)$，但 $\boldsymbol{X}=(X_1,X_2)^{\mathrm{T}}$ 不是正态分布。

(3) 如果 p 元随机向量 $\boldsymbol{X}=(X_1,X_2,\cdots,X_p)^{\mathrm{T}}$ 与 $\boldsymbol{Y}=(Y_1,Y_2,\cdots,Y_p)^{\mathrm{T}}$ 的联合分布是正态分布 $N_p\left(\begin{pmatrix}\boldsymbol{u}_X\\ \boldsymbol{u}_Y\end{pmatrix},\begin{pmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{pmatrix}\right)$，则 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 相互独立的充分必要条件是 $\boldsymbol{\Sigma}_{12}=\boldsymbol{\Sigma}_{21}=0$。

(4) 设 n 个 p 维随机向量是 $\boldsymbol{X}^{(1)},\boldsymbol{X}^{(2)},\cdots,\boldsymbol{X}^{(n)}$ 相互独立的，且每个随机向量都服从正态分布，即 $\boldsymbol{X}^{(i)}\sim N_p(\boldsymbol{u}^{(i)},\boldsymbol{\Sigma})(i=1,2,\cdots,n)$，如果 n 阶方阵 $\boldsymbol{P}=(p_{ij})_{n\times n}$ 是正交阵，则 p 维随机向量 $\boldsymbol{Y}^{(i)}=\sum_{j=1}^{n}p_{ij}\boldsymbol{X}^{(j)}$ 仍服从正态分布 $N_p\left(\sum_{j=1}^{n}p_{ij}\boldsymbol{u}^{(j)},\boldsymbol{\Sigma}\right)(i=1,2,\cdots,n)$，且 $\boldsymbol{Y}^{(i)}$，$i=1,2,\cdots,n$ 也是相互独立的。

1.3.3 条件分布和独立性

设 $\boldsymbol{X}\sim N_p(\boldsymbol{u},\boldsymbol{\Sigma})$，$p\geqslant 2$，将 $\boldsymbol{X}$、$\boldsymbol{u}$ 和 $\boldsymbol{\Sigma}$ 剖分如下：

$$\boldsymbol{X}=\begin{bmatrix}\boldsymbol{X}^{(1)}\\ \boldsymbol{X}^{(2)}\end{bmatrix},\quad \boldsymbol{u}=\begin{bmatrix}\boldsymbol{u}^{(1)}\\ \boldsymbol{u}^{(2)}\end{bmatrix},\quad \boldsymbol{\Sigma}=\begin{bmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{bmatrix}$$

其中，$\boldsymbol{X}^{(1)}$、$\boldsymbol{u}^{(1)}$ 为 q 维向量，$\boldsymbol{\Sigma}_{11}$ 为 $q\times q$ 阵，我们希望求出当 $\boldsymbol{X}^{(2)}$ 给定时 $\boldsymbol{X}^{(1)}$ 的条件分布，即 $(\boldsymbol{X}^{(1)}\mid\boldsymbol{X}^{(2)})$ 的分布。下面的定理不仅指出了正态分布的条件分布与边缘分布仍是正态分布，而且给出了条件分布的均值和方差阵的计算公式。

定理 1.2 设 $\boldsymbol{X}\sim N_p(\boldsymbol{u},\boldsymbol{\Sigma})$，$\boldsymbol{\Sigma}>0$，则

$(\boldsymbol{X}^{(1)}\mid\boldsymbol{X}^{(2)})\sim N_q(\boldsymbol{u}_{1\cdot 2},\boldsymbol{\Sigma}_{11\cdot 2})$，其中

$$\boldsymbol{u}_{1\cdot 2}=\boldsymbol{u}^{(1)}+\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{X}^{(2)}-\boldsymbol{u}^{(2)})$$

$$\boldsymbol{\Sigma}_{11\cdot 2}=\boldsymbol{\Sigma}_{11}-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$$

证明：先将证明思路介绍一下。我们知道，条件密度等于联合密度除以给定条件的边缘分布密度。联合密度及给定条件的边缘分布密度都是可以求得的。然而，问题在于形式上两个密度函数除不尽，不能简单写成整式，因此无法直接证明两个密度函数之商是正态分

布密度。为了使两个密度函数能除尽，我们需要将联合密度表示为给定条件的边缘分布密度与另一个函数的乘积。

显然，另一个函数即商是条件分布密度函数，也就是我们需要找到这个函数使联合密度表示成两个密度函数的乘积。接着，根据独立性的性质，这两个密度函数所对应的随机向量应该相互独立。因为其中一个向量为给定条件的随机向量，所以另一个随机向量应该与给定条件独立。由一元正态分布的性质可知，对于正态分布而言，独立与不相关是等价的。因此，我们只需要确保另一个随机向量与给定条件的随机向量不相关，即协方差阵为零矩阵，或等价地将这两个随机向量合成为一个向量的协方差阵是对角块阵。

进而可以将问题简化为如何将协方差阵对角化。由于协方差阵是对角阵，所以一定可以通过相似对角化将其表示为对角矩阵，然后再利用线性变换矩阵进行线性变换。通过这样的步骤，我们可以得到两个不相关的随机向量，其中一个与给定条件的随机向量独立，从而证明了所求的条件分布是正态分布。

由$\boldsymbol{\Sigma}>0$，根据正定阵的性质，$\boldsymbol{\Sigma}_{11}>0$，$\boldsymbol{\Sigma}_{22}>0$，令

$$\boldsymbol{Z}=\begin{bmatrix}\boldsymbol{Z}^{(1)}\\ \boldsymbol{Z}^{(2)}\end{bmatrix}=\begin{bmatrix}\boldsymbol{I}_q & -\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\\ \boldsymbol{0} & \boldsymbol{I}_{p-q}\end{bmatrix}\begin{bmatrix}\boldsymbol{X}^{(1)}\\ \boldsymbol{X}^{(2)}\end{bmatrix}=\boldsymbol{B}\boldsymbol{X}$$

$$=\begin{bmatrix}\boldsymbol{X}^{(1)}-\boldsymbol{X}^{(2)}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\\ \boldsymbol{X}^{(2)}\end{bmatrix}$$

则根据 1.3.2 节中的性质(1)，$\boldsymbol{Z}$ 服从正态分布，且，

$$E(\boldsymbol{Z})=\boldsymbol{B}E(\boldsymbol{X})=\begin{bmatrix}\boldsymbol{u}^{(1)}-\boldsymbol{u}^{(2)}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\\ \boldsymbol{u}^{(2)}\end{bmatrix}$$

$$D(\boldsymbol{Z})=\boldsymbol{B}D(\boldsymbol{X})\boldsymbol{B}^{\mathrm{T}}=\begin{bmatrix}\boldsymbol{\Sigma}_{11}-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{\Sigma}_{22}\end{bmatrix}=\begin{bmatrix}\boldsymbol{\Sigma}_{11\cdot 2} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{\Sigma}_{22}\end{bmatrix}$$

根据正定性质$\boldsymbol{\Sigma}_{11\cdot 2}>0$，由前文分析可知，$\boldsymbol{Z}$ 的分布密度应该是 $\boldsymbol{X}^{(2)}$ 与 $\boldsymbol{X}^{(1)}-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{X}^{(2)}$ 两个正态分布密度之积，即

$N_q(\boldsymbol{u}^{(1)}-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{u}^{(2)},\boldsymbol{\Sigma}_{11\cdot 2})N_{p-q}(\boldsymbol{u}^{(2)},\boldsymbol{\Sigma}_{22})$，这样 $\boldsymbol{Z}$ 的密度函数可表示为

$$\begin{aligned}f(\boldsymbol{Z})&=(2\pi)^{-q/2}(2\pi)^{-\frac{p-q}{2}}\mid\boldsymbol{\Sigma}_{11\cdot 2}\mid^{-\frac{1}{2}}\mid\boldsymbol{\Sigma}_{22}\mid^{\frac{1}{2}}\\&=\exp\left\{-\frac{1}{2}(\boldsymbol{Z}^{(1)}-\boldsymbol{\mu}^{(1)}+\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\mu}^{(2)})^{\mathrm{T}}\boldsymbol{\Sigma}_{11\cdot 2}^{-1}(\boldsymbol{Z}^{(1)}-\boldsymbol{\mu}^{(1)}+\right.\\&\left.\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\mu}^{(2)})\right\}\exp\left\{-\frac{1}{2}(\boldsymbol{Z}^{(2)}-\boldsymbol{\mu}^{(2)})^{\mathrm{T}}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{Z}^{(2)}-\boldsymbol{\mu}^{(2)})\right\}\\&=(2\pi)^{-\frac{q}{2}}\mid\boldsymbol{\Sigma}_{11\cdot 2}\mid^{-1/2}\exp\left\{-\frac{1}{2}(\boldsymbol{Z}^{(1)}-\boldsymbol{\mu}^{(1)}+\boldsymbol{\Sigma}_{12}\Sigma_{22}^{-1}\boldsymbol{\mu}^{(2)})^{\mathrm{T}}\right.\\&\left.\boldsymbol{\Sigma}_{11\cdot 2}^{-1}(\boldsymbol{Z}^{(1)}-\boldsymbol{\mu}^{(1)}+\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\mu}^{(2)})(2\pi)^{-\frac{p-q}{2}}\mid\boldsymbol{\Sigma}_{22}\mid^{-\frac{1}{2}}\right\}\times\\&\exp\left\{-\frac{1}{2}(\boldsymbol{Z}^{(2)}-\boldsymbol{\mu}^{(2)})^{\mathrm{T}}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{Z}^{(2)}-\boldsymbol{\mu}^{(2)})\right\}\end{aligned}$$

当然，这样的 $\boldsymbol{Z}$ 函数与给定条件 $\boldsymbol{X}^{(2)}$ 的密度仍不好相除，因此需再做逆变换回到 $\boldsymbol{X}$，逆

变换的公式为

$$\begin{cases}\boldsymbol{X}^{(1)}=\boldsymbol{Z}^{(1)}+\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{Z}^{(2)}\\ \boldsymbol{X}^{(2)}=\boldsymbol{Z}^{(2)}\end{cases}$$

变换当然要考虑雅可比行列式，但它是三角阵，对角线元素全为 1，故 $J=1$，这时再考虑联合密度与给定条件的密度函数之商。根据 $\boldsymbol{X}^{(2)}=\boldsymbol{Z}^{(2)}$，后一个因子即 $\boldsymbol{X}^{(2)}$ 的分布密度即给定条件的分布密度，因而可以约去，此时就只剩下

$$(2\pi)^{-\frac{q}{2}}\mid\boldsymbol{\Sigma}_{11\cdot 2}\mid^{-1/2}\exp\left\{-\frac{1}{2}(\boldsymbol{X}^{(1)}-\boldsymbol{\mu}^{(1)}+\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\mu}^{(2)}-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{X}^{(2)})^{\mathrm{T}}\times\right.$$

$$\left.\boldsymbol{\Sigma}_{11\cdot 2}^{-1}(\boldsymbol{X}^{(1)}-\boldsymbol{\mu}^{(1)}+\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\mu}^{(2)}-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{X}^{(2)})\right\}$$

$$=(2\pi)^{-\frac{q}{2}}\mid\boldsymbol{\Sigma}_{11\cdot 2}\mid^{-1/2}\exp\left\{-\frac{1}{2}(\boldsymbol{X}^{(1)}-\boldsymbol{\mu}_{1\cdot 2})^{\mathrm{T}}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{X}^{(1)}-\boldsymbol{\mu}_{1\cdot 2})\right\}$$

这正是正态分布密度函数的标准形式，均值为$\boldsymbol{\mu}_{1\cdot 2}$，方差为$\boldsymbol{\Sigma}_{11\cdot 2}$。

该定理告诉我们，$\boldsymbol{X}^{(1)}$ 的分布与 $\boldsymbol{X}^{(1)}\mid\boldsymbol{X}^{(2)}$ 的分布均为正态分布，它们的协方差阵分别为$\boldsymbol{\Sigma}_{11}$ 与$\boldsymbol{\Sigma}_{11\cdot 2}=\boldsymbol{\Sigma}_{11}-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$，由于$\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\geqslant 0$，故$\boldsymbol{\Sigma}_{11}\geqslant\boldsymbol{\Sigma}_{11\cdot 2}$，等号成立当且仅当$\boldsymbol{\Sigma}_{12}=0$。协方差阵是用来描述指标之间相关关系及散布程度的，$\boldsymbol{\Sigma}_{11}\geqslant\boldsymbol{\Sigma}_{11\cdot 2}$ 说明了在已知 $\boldsymbol{X}^{(2)}$ 的条件下，$\boldsymbol{X}^{(1)}$ 散布的程度比不知道 $\boldsymbol{X}^{(2)}$ 的情况下缩小了，只有当$\boldsymbol{\Sigma}_{12}=0$时，两者才相同。同时，还可以证明，当$\boldsymbol{\Sigma}_{12}=0$ 时等价于 $\boldsymbol{X}^{(1)}$ 和 $\boldsymbol{X}^{(2)}$ 相互独立，这时，即使给出 $\boldsymbol{X}^{(2)}$，对 $\boldsymbol{X}^{(1)}$ 的分布也是没有影响的。

定理 1.3 设 $\boldsymbol{X}\sim N_p(\boldsymbol{u},\boldsymbol{\Sigma})$，$\boldsymbol{\Sigma}>0$，将 $\boldsymbol{X}$，$\boldsymbol{u}$，$\boldsymbol{\Sigma}$ 剖分如下：

$$\boldsymbol{X}=\begin{bmatrix}X^{(1)}\\X^{(2)}\\X^{(3)}\end{bmatrix}\begin{matrix}r\\s\\t\end{matrix},\quad \boldsymbol{u}=\begin{bmatrix}\boldsymbol{u}^{(1)}\\\boldsymbol{u}^{(2)}\\\boldsymbol{u}^{(3)}\end{bmatrix}\begin{matrix}r\\s\\t\end{matrix},\quad \boldsymbol{\Sigma}=\begin{bmatrix}\boldsymbol{\Sigma}_{11}&\boldsymbol{\Sigma}_{12}&\boldsymbol{\Sigma}_{13}\\\boldsymbol{\Sigma}_{21}&\boldsymbol{\Sigma}_{22}&\boldsymbol{\Sigma}_{23}\\\boldsymbol{\Sigma}_{31}&\boldsymbol{\Sigma}_{32}&\boldsymbol{\Sigma}_{33}\end{bmatrix}\begin{matrix}r\\s\\t\end{matrix},\text{则}$$

$X^{(1)}$ 有如下的条件均值和条件协方差阵的递推公式：

$$E(X^{(1)}\mid X^{(2)},X^{(3)})=\boldsymbol{u}_{1\cdot 3}+\boldsymbol{\Sigma}_{12\cdot 3}\boldsymbol{\Sigma}_{22\cdot 3}^{-1}(X^{(2)}-\boldsymbol{u}_{2\cdot 3})$$

$$D(X^{(1)}\mid X^{(2)},X^{(3)})=\boldsymbol{\Sigma}_{11\cdot 3}-\boldsymbol{\Sigma}_{12\cdot 3}\boldsymbol{\Sigma}_{22\cdot 3}^{-1}\boldsymbol{\Sigma}_{21\cdot 3}$$

其中，$\boldsymbol{\Sigma}_{ij\cdot k}=\boldsymbol{\Sigma}_{ij}-\boldsymbol{\Sigma}_{ik}\boldsymbol{\Sigma}_{kk}^{-1}\boldsymbol{\Sigma}_{kj}$，$i,j,k=1,2,3$，$\boldsymbol{u}_{i\cdot 3}=E(X^{(i)}\mid X^{(3)})$，$i=1,2$。证明可参见方开泰编著的《实用多元统计分析》(上海：华东师范大学出版社，1989)。

定理 1.2 和定理 1.3 在 20 世纪 70 年代中期为国家标准部门制定服装标准时有成功的应用。

【例 1.8】 在制定服装标准时需抽样进行人体测量，现从某年龄段女子测量取出部分结果如下。

X_1：身高，X_2：胸围，X_3：腰围，X_4：上体长，X_5：臀围。已知它们遵从 $N_5(\boldsymbol{u},\boldsymbol{\Sigma})$，其中，$\boldsymbol{u}=\begin{bmatrix}154.98\\83.39\\70.26\\61.32\\91.52\end{bmatrix}$，

$$\boldsymbol{\Sigma}=\begin{bmatrix}29.660 & 6.514 & 1.847 & 9.358 & 10.336\\ 6.514 & 30.530 & 25.536 & 3.540 & 19.532\\ 1.847 & 25.536 & 39.859 & 2.227 & 20.703\\ 9.358 & 3.540 & 2.227 & 7.033 & 5.213\\ 10.336 & 19.532 & 20.703 & 5.213 & 27.363\end{bmatrix}$$

解

若取 $\boldsymbol{X}^{(1)}=(X_1,X_2,X_3)^{\mathrm{T}}$，$\boldsymbol{X}^{(2)}=(X_4)$，$\boldsymbol{X}^{(3)}=(X_5)$，则由定理 1.2 得

$$E\left[\begin{array}{c}X_1\\X_2\\X_3\\X_4\end{array}\middle|X_5\right]=\begin{bmatrix}154.98\\83.39\\70.26\\61.32\end{bmatrix}+\begin{bmatrix}10.34\\19.53\\20.70\\5.21\end{bmatrix}(27.36)^{-1}(X_5-91.52)$$

$$=\begin{bmatrix}154.98+0.38(X_5-91.52)\\83.39+0.71(X_5-91.52)\\70.26+0.76(X_5-91.52)\\61.32+0.19(X_5-91.52)\end{bmatrix}$$

其中第一个常数向量来自 $\boldsymbol{u}$ 的上半部分，$(10.336,\cdots)^{\mathrm{T}}$ 来自 $\boldsymbol{\Sigma}$ 的右上角，27.363 是 $\boldsymbol{\Sigma}$ 的右下角，即 $\boldsymbol{\Sigma}_{22}$，91.52 是 $\boldsymbol{u}^{(2)}$，

$$D\left[\begin{array}{c}X_1\\X_2\\X_3\\X_4\end{array}\middle|X_5\right]=\begin{bmatrix}29.66 & 6.51 & 1.85 & 9.36\\ 6.51 & 30.53 & 25.54 & 3.54\\ 1.85 & 25.54 & 39.86 & 2.23\\ 9.36 & 3.54 & 2.23 & 7.03\end{bmatrix}-\begin{bmatrix}10.34\\19.53\\20.70\\5.21\end{bmatrix}(27.36)^{-1}(10.34,19.53,20.70,5.21)$$

$$=\begin{bmatrix}25.76 & -0.86 & -5.97 & 7.39\\ -0.86 & 16.59 & 10.76 & -0.18\\ -5.97 & 10.76 & 24.19 & -1.72\\ 7.39 & -0.18 & -1.72 & 6.04\end{bmatrix}$$

其中第一个矩阵是 $\boldsymbol{\Sigma}$ 的左上角。

再利用定理 1.3 得到

$$D\left[\begin{array}{c}X_1\\X_2\\X_3\end{array}\middle|\begin{array}{c}X_4\\X_5\end{array}\right]=\begin{bmatrix}25.76 & -0.86 & -5.97\\ -0.86 & 16.59 & 10.76\\ -5.97 & 10.76 & 24.19\end{bmatrix}-\begin{bmatrix}7.39\\-0.18\\-1.72\end{bmatrix}(6.04)^{-1}(7.39,-0.18,-1.72)$$

$$=\begin{bmatrix}16.72 & -0.64 & -3.87\\ -0.64 & 16.58 & 10.71\\ -3.87 & 10.71 & 23.71\end{bmatrix}$$

其中三阶阵是 $D\left[\begin{array}{c}X_1\\X_2\\X_3\\X_4\end{array}\middle|X_5\right]$ 的左上角，6.04 是右下角，(7.39，−0.18，−1.72)是左下角。

我们可以看到，

$$\mathrm{var}(X_1 \mid X_4, X_5) = 16.72 < 29.66 = \mathrm{var}(X_1)$$
$$\mathrm{var}(X_2 \mid X_4, X_5) = 16.58 < 30.53 = \mathrm{var}(X_2)$$
$$\mathrm{var}(X_3 \mid X_4, X_5) = 23.71 < 39.86 = \mathrm{var}(X_3)$$

这说明，若已知一个人的上体长和臀围，则身高、胸围和腰围的条件方差比原来的方差大大缩小。

定义 1.6 当 $\boldsymbol{X}^{(2)}$ 给定时，X_i 和 X_j 的偏相关系数为

$$r_{ij\cdot q+1,\cdots,p} = \frac{\sigma_{ij\cdot q+1,\cdots,p}}{(\sigma_{ii\cdot q+1,\cdots,p}\sigma_{jj\cdot q+1,\cdots,p})^{\frac{1}{2}}}$$

在上面制定服装标准的例子中，给定 X_4 和 X_5 的偏相关系数为

$$r_{12\cdot 45} = \frac{-0.643}{\sqrt{16.717 \times 16.582}} = -0.038\,6$$
$$r_{13\cdot 45} = \frac{-3.873}{\sqrt{16.717 \times 23.707}} = -0.195$$
$$r_{23\cdot 45} = \frac{10.707}{\sqrt{16.582 \times 23.707}} = 0.540$$

定理 1.4 设 $\boldsymbol{X} \sim N_p(\boldsymbol{u}, \boldsymbol{\Sigma})$，$\boldsymbol{\Sigma} > 0$，将 $\boldsymbol{X}, \boldsymbol{u}, \boldsymbol{\Sigma}$ 剖分如下：

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}^{(1)} \\ \vdots \\ \boldsymbol{X}^{(k)} \end{bmatrix}, \quad \boldsymbol{u} = \begin{bmatrix} \boldsymbol{u}^{(1)} \\ \vdots \\ \boldsymbol{u}^{(k)} \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \cdots & \boldsymbol{\Sigma}_{1k} \\ \vdots & & \vdots \\ \boldsymbol{\Sigma}_{k1} & \cdots & \boldsymbol{\Sigma}_{kk} \end{bmatrix}$$

其中：$\boldsymbol{X}^{(j)} : S_j \times 1$，$\boldsymbol{u}^{(j)} : S_j \times 1$，$\boldsymbol{\Sigma}_{jj} : S_j \times S_j$，$j = 1, \cdots, k$，则 $\boldsymbol{X}^{(1)}, \cdots, \boldsymbol{X}^{(k)}$ 相互独立，当且仅当 $\boldsymbol{\Sigma}_{ij} = 0$，对一切 $i \neq j$，即 $\boldsymbol{\Sigma}$ 为对角块阵。

证明：必要性　如果 $\boldsymbol{X}^{(1)}, \cdots, \boldsymbol{X}^{(k)}$ 相互独立，则 $\forall\, i \neq j$，$\boldsymbol{X}^{(1)}, \cdots, \boldsymbol{X}^{(k)}$ 之间独立，一定不相关，所以 $\boldsymbol{\Sigma}_{ij} = \mathrm{cov}(\boldsymbol{X}^{(i)}, \boldsymbol{X}^{(j)}) = \boldsymbol{0}$。

充分性　$\forall\, i \neq j$，$\boldsymbol{\Sigma}_{ij} = 0$，$\boldsymbol{\Sigma}$ 为对角块阵，所以

$$f(\boldsymbol{X}^{(1)}, \cdots, \boldsymbol{X}^{(k)}) = (2\pi)^{-\frac{1}{2}\sum_{j=1}^{k} S_j} \mid \mathrm{diag}(\boldsymbol{\Sigma}_{11}, \cdots, \boldsymbol{\Sigma}_{kk}) \mid^{-\frac{1}{2}} \times$$
$$\exp\left\{-\frac{1}{2}[(\boldsymbol{X}^{(1)} - \boldsymbol{u}^{(1)})^{\mathrm{T}}, \cdots, (\boldsymbol{X}^{(k)} - \boldsymbol{u}^{(k)})^{\mathrm{T}}] \times \right.$$
$$\left. \mathrm{diag}(\boldsymbol{\Sigma}_{11}^{-1}, \cdots, \boldsymbol{\Sigma}_{kk}^{-1})[(\boldsymbol{X}^{(1)} - \boldsymbol{u}^{(1)}), \cdots, (\boldsymbol{X}^{(k)} - \boldsymbol{u}^{(k)})]\right\} =$$
$$\prod_{j=1}^{k} (2\pi)^{-\frac{S_j}{2}} \mid \boldsymbol{\Sigma}_{jj} \mid^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\boldsymbol{X}^{(j)} - \boldsymbol{u}^{(j)})^{\mathrm{T}} \boldsymbol{\Sigma}_{jj}^{-1} (\boldsymbol{X}^{(j)} - \boldsymbol{u}^{(j)})\right\} =$$
$$\prod_{j=1}^{k} f(\boldsymbol{X}^{(j)})$$

所以，$\boldsymbol{X}^{(1)}, \cdots, \boldsymbol{X}^{(k)}$ 相互独立。

值得注意的是，对于多元正态分布而言，“$\boldsymbol{X}^{(1)}$ 和 $\boldsymbol{X}^{(2)}$ 不相关”等价于“$\boldsymbol{X}^{(1)}$ 和 $\boldsymbol{X}^{(2)}$ 相互独立”。

1.4　JMP 软件操作

视频 1-2

1. 多元样本数据输入

在 JMP 软件中，依次单击**文件→新建→数据表**，然后通过增加列的方式建立多个变量并填充数据完成样本输入。

2. 协方差矩阵和相关系数矩阵提取

在 JMP 中计算相关系数和协方差矩阵步骤如下。

(1) 依次单击**分析→多元方法→多元**，如图 1.9 所示。

图 1.9　操作流程示意图

(2) 在弹出的图 1.10 的对话框中，将各列放入 **Y**，单击**确定**按钮，可得到相关系数矩阵和散点图矩阵。

图 1.10　多元对话框

(3) 单击**多元**旁的按钮,在弹出的菜单中勾选**协方差矩阵**,即可得到协方差矩阵。

习题

1. 已知一二维正态总体 $\boldsymbol{G}$ 的分布为 $N_2\left(\begin{pmatrix}0\\0\end{pmatrix}\begin{pmatrix}1 & 0.9\\0.9 & 1\end{pmatrix}\right)$,求点 $\boldsymbol{A}=\begin{bmatrix}1\\1\end{bmatrix}$ 和 $\boldsymbol{B}=\begin{bmatrix}1\\-1\end{bmatrix}$ 到均值 $\boldsymbol{\mu}=\begin{bmatrix}0\\0\end{bmatrix}$ 的距离。

2. 设随机向量(X,Y)的两个分量相互独立,且均服从标准正态分布 $N(0,1)$。

(1) 分别写出随机变量 $X+Y$ 与 $X-Y$ 的分布密度。

(2) 试问:$X+Y$ 与 $X-Y$ 是否独立?并说明理由。