谱分解：$A=T\Lambda T'=\sum_{i=1}^{p}\lambda_i t_i t_i'$

SVD 分解：$A=U\Lambda V'=\sum_{i=1}^{k}\sigma_i u_i v_i'$. 其中 $AA'u_i=\sigma_i^2u_i$, $A'Av_i=\sigma_i^2v_i$

Projection Matrix: $P_M=M(M'M)^{-1}M'$ is idempotent

对于幂等矩阵 A, $Rank(A)=tr(A)$, 特征值均为 0 或 1

Let $B=A_{22}-A_{21}A_{11}^{-1}A_{12}$; $D=A_{11}-A_{12}A_{22}^{-1}A_{21}$, then A^{-1}=

$$\begin{bmatrix} A_{11}^{-1}+A_{11}^{-1}A_{12}B^{-1}A_{21}A_{11}^{-1} & -A_{11}^{-1}A_{12}B^{-1} \\ -B^{-1}A_{21}A_{11}^{-1} & B^{-1}\end{bmatrix}$$

=

$$\begin{bmatrix} D^{-1} & -D^{-1}A_{12}A_{22}^{-1} \\ -A_{22}^{-1}A_{21}D^{-1} & A_{22}^{-1}+A_{22}^{-1}A_{21}D^{-1}A_{12}A_{22}^{-1}\end{bmatrix}$$

Moore-Penrose Inverse A^+.
$AA^+A=A$, $A^+AA^+=A^+$, $(AA^+)'=AA^+$, $(A^+A)'=A^+A$
For full column rank matrix A, $A^+=(A'A)^{-1}A'$
When $x = A^+b$, $Ax = b$ reaches the minimum MSE.

$\lambda^q|\lambda I_p-AB| = \lambda^p|\lambda I_q-BA|$, A^{p*q}, B^{q*p}.

$\max_{x\neq0}x'Ax/x'x=\lambda_1$, $\min_{x\neq0}x'Ax/x'x=\lambda_p$, x=对应特征向量
$\max_{x\neq0}x'Ax/x'Bx=\mu_1$, $\min_{x\neq0}x'Ax/x'Bx=\mu_p$, μ是 $B^{-1}A$ 对应特征值, B>0
$\max_{x'Bt_k=0,k=1\sim i-1,x\neq0}x'Ax/x'Bx=\mu_i$, x 取 t_i, $B^{-1}A$ 的特征向量时达到。

$(x'y)^2\le(x'Bx)(y'B^{-1}y)$, B>0 **扩展 Cauchy-Schwarz 不等式**

$Cov(Ax+b)=ACov(x)A'$
$Cov(Ax, By)=ACov(x,y)B'$
$Cov(\sum_{i=1}^nA_ix_i, \sum_{j=1}^mB_jy_j)=\sum_{i=1}^n\sum_{j=1}^mA_iCov(x_i, y_j)B_j'$

Sample Covariance $S=\sum_{i=1}^n(x_i-\bar{x})(x_i-\bar{x})'/(n-1)$

相关系数 Correlation Coefficients
$\rho(x, y)=Cov(x, y)/\sqrt{Var(x)Var(y)}$
相关系数矩阵 $\rho=V^{-1/2}\Sigma V^{-1/2}$, $V^{-1/2}=diag(1/\sqrt{\sigma_{11}}, ..., 1/\sqrt{\sigma_{pp}})$

Square Mahalanobis Distance
$d^2(x, y)=(x-y)'\Sigma^{-1}(x-y)$ 无单位

随机变量变换的分布
$J(\varphi\to y)=\partial\varphi/\partial y$
$g(y_1, ..., y_p)=f(\varphi_1(y_1, ..., y_p), ..., \varphi_p(y_1, ..., y_p))|J(\varphi\to y)|$

$x\sim N_p(\mu, \Sigma)\to y=Cx+b\sim N_p(C\mu+b, C\Sigma C')$
$x\sim N_p(\mu, \Sigma)\to (x-\mu)'\Sigma^{-1}(x-\mu)\sim\chi^2(p)$

III 多元正态分布
$f(x)=(2\pi)^{-p/2}|\Sigma|^{-1/2}\exp(-(x-\mu)'\Sigma^{-1}(x-\mu)/2)$

样本 $x_1, ..., x_n$ 的联合分布的似然函数
$L(\mu, \Sigma)=((2\pi)^p|\Sigma|)^{-n/2}\exp(-\sum_{i=1}^n(x_i-\mu)'\Sigma^{-1}(x_i-\mu)/2)$
$\mu_{MLE}=\bar{x}$, $\Sigma_{MLE}=\sum_{i=1}^n(x_i-\bar{x})(x_i-\bar{x})'=A/n$
A 被称为样本离差矩阵（Sample Dispersion Matrix）
样本协方差矩阵 $S=A/(n-1)$
样本相关系数
$r_{ij}=s_{ij}/\sqrt{s_{ii}s_{jj}}=\sum_{k=1}^n(x_{ki}-\bar{x}_i)(x_{kj}-\bar{x}_j)/\sqrt{\sum_{k=1}^n(x_{ki}-\bar{x}_i)^2\sum_{k=1}^n(x_{kj}-\bar{x}_j)^2}$

条件正态分布
$\mu_{1.2}=\mu_1+\Sigma_{12}\Sigma_{22}^{-1}(x_2-\mu_2)=E(x_1|x_2)$
复相关系数 $\rho_{y\cdot x}=\max_{l\neq0}\rho(y, l'x)=\sqrt{\sigma_{xy}'\Sigma_{xx}^{-1}\sigma_{xy}/\sigma_{yy}}=\sqrt{\rho_{xy}'R_{xx}^{-1}\rho_{xy}}$
偏相关系数矩阵 $\Sigma_{11.2}=\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}=[\sigma_{ij\cdot k+1\sim p}]$
p-k 阶偏相关系数 $\rho_{ij\cdot k+1\sim p}=\sigma_{ij\cdot k+1\sim p}/\sqrt{\sigma_{ii\cdot k+1\sim p}\sigma_{jj\cdot k+1\sim p}}$
1 阶偏相关系数 $\rho_{12\cdot3}=(\rho_{12}-\rho_{13}\rho_{23})/\sqrt{(1-\rho_{13})(1-\rho_{23})}$

对于实对称矩阵 A，有
$\partial(X'AX)/\partial X=2AX$; $\partial(X'AX)/\partial A=XX'$; $\partial\ln|A|/\partial A=A^{-1}$

$\bar{x}\sim N_p(\mu, \Sigma/n)$ 与 S 独立
$(n-1)S=\sum_{i=1}^{n-1}Z_iZ_i'\sim W_p(n-1, \Sigma)$

Wishart 分布
设 $X_i\sim N_p(0, \Sigma)$, i=1, ..., n 相互独立，记为 X_{n*p}.
则随机矩阵 $W_{p*p}=\sum_{i=1}^nX_iX_i'\sim W_p(n, \Sigma)$
是具有 n 个自由度的 Wishart 分布
$f(A)=|A|^{(n-p-1)/2}e^{-tr(\Sigma^{-1}A)/2}/(2^{np/2}\Gamma_p(n-2)|\Sigma|^{n/2})$
当 $p=1, \Sigma=\sigma^2$ 时，$W=\sum_{i=1}^nx_i^2\sim\sigma^2\chi^2(n)$
$W_i\sim W_p(n_i, \Sigma)\to\sum W_i\sim W_p(\sum n_i, \Sigma)$
C_{p*p} 可逆, $CWC'\sim W_p(n, C\Sigma C')$

X 典型相关分析 Canonical Correlation Analysis
$x=[x_1\ x_2 ... x_p]'$, $y=[y_1\ y_2 ... y_q]'$
$A=\Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1/2}$, $B=\Sigma_{22}^{-1/2}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1/2}$
$C=\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$, $D=\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$,
A、B、C、D 均共享非零特征值
$\rho_1^2\ge...\ge\rho_m^2>0$, $m=rank(\Sigma_{12})$

记$\beta_1, ..., \beta_m$ 为 B 的标准正交的特征向量，则
$\alpha_i=\Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1/2}\beta_i/\rho_i$ 是 A 的特征向量，
$a_i=\Sigma_{11}^{-1/2}\alpha_i$ 是 C 的特征向量
$b_i=\Sigma_{22}^{-1/2}\beta_i$ 是 D 的特征向量
当 $a=a_1, b=b_1$ 时，$u=a'x, v=b'y$ 使$\rho(u, v)=a'\Sigma_{12}b$ 达到最大
(限制：$Var(u)=a'\Sigma_{11}a=Var(v)=b'\Sigma_{22}b=1$)
证明：$\alpha_i=\Sigma_{11}^{1/2}a_i$, $\beta_i=\Sigma_{22}^{1/2}b_i$. 则约束为$\alpha_i'\alpha_i=1$, $\beta_i'\beta_i=1$
$(a'\Sigma_{12}b)^2=(\alpha'\Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1/2}\beta)^2$
$\le^{Cauchy}(\alpha'\alpha)[(\Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1/2}\beta)'(\Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1/2}\beta)]$
$=\beta'\Sigma_{22}^{-1/2}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1/2}\beta\le\rho_1^2$. ($\beta=\beta_1$ 时取等).
此时$\alpha=\alpha_1=\Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1/2}\beta_1/\rho_1$. 代回即证。

记 a_i, b_i 为第 i 对典型系数, i=1~m
记 $u_i=a_i'x$, $v_i=b_i'y$ 为第 i 对典型变量(Canonical Variables)
记ρ_i 为第 i 个典型相关系数(Canonical Correlation)
记 $A=[a_1 ... a_m]$, $B=[b_1 ... b_m]$, $u=A'x=[u_1 ... u_m]'$, $v=B'y=[v_1 ... v_m]'$
$\rho(u_i, v_i)=\rho_{u_i\cdot y}=\rho_{v_i\cdot x}=\rho_i$, $\rho(u_i, v_j)=0(i\neq j)$.
$Cov(u)=I$, $Cov(v)=I$, $Cov(u, v)=\Lambda=diag(\rho_1, ..., \rho_m)$
$Cov(x, u)=\Sigma_{11}A$, $Cov(x, v)=\Sigma_{12}B$, $Cov(y, u)=\Sigma_{21}A$, $Cov(y, v)=\Sigma_{22}B$
$\rho(x, u)=D_1^{-1}\Sigma_{11}A$, $\rho(x, v)=D_1^{-1}\Sigma_{12}B$, $D_1=diag(\sqrt{Var(x_1)}, ..., \sqrt{Var(x_p)})$
$\rho(y, u)=D_2^{-1}\Sigma_{21}A$, $\rho(y, v)=D_2^{-1}\Sigma_{22}B$, $D_2=diag(\sqrt{Var(y_1)}, ..., \sqrt{Var(y_q)})$

基于相关矩阵的典型相关计算
$a_i^*=D_1a_i$, $b_i^*=D_2b_i$. $\rho_i^*=\rho_i$. $u_i^*=a_i'D_1D_1^{-1}(x-\mu_1)=u_i-a_i'\mu_1$. $v_i^*=v_i-b_i'\mu_2$.

IV Multivariate Normal Population Statistical Inference
均值置信区间
$x_1, ..., x_n\sim N(\mu, \sigma^2)$.
σ^2 已知: $\mu\in[\bar{x}\pm u_{\alpha/2}\sigma/\sqrt{n}]$ u_q 为 N(0,1)的上 q 分位点
拒绝 $\mu=\mu_0$ if $P(|(\bar{x}-\mu_0)\sqrt{n}/\sigma|<u_{\alpha/2})\le\alpha$
σ^2 未知: $\mu\in[\bar{x}\pm t_{\alpha/2}(n-1)s/\sqrt{n}]$

$x_1, ..., x_{n_1}\sim N(\mu_1, \sigma_1^2)$. $y_1, ..., y_{n_2}\sim N(\mu_2, \sigma_2^2)$.
σ_1^2, σ_2^2 已知: $\mu_1-\mu_2\in[\bar{x}-\bar{y}\pm u_{\alpha/2}\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}]$
$\sigma_1^2=\sigma_2^2=\sigma^2$ 未知: $s_p^2=((n_1-1)s_1^2+(n_2-1)s_2^2)/(n_1+n_2-2)$
$\mu_1-\mu_2\in[\bar{x}-\bar{y}\pm t_{\alpha/2}(n_1+n_2-2)s_p\sqrt{1/n_1+1/n_2}]$

方差分析
$x_{i1}, ..., x_{in}$, i=1,...,k 从$\pi_i\sim N(\mu_i, \sigma^2)$中采样. H_0: $\mu_1=...=\mu_k$.
$F=SSTR(n-k)/(SSE(k-1))\sim_{H_0}F(k-1, n-k)$, $F<F_\alpha(k-1, n-k)$时拒绝 H_0

多元单总体均值推断
均值向量检验
$x_1, ..., x_n$ 是 $x\sim N_p(\mu, \Sigma)$总体的采样，n>p
检验：H_0: $\mu=\mu_0$ 对 H_1: $\mu\neq\mu_0$.
Σ已知：$T_0^2=n(\bar{x}-\mu_0)'\Sigma^{-1}(\bar{x}-\mu_0)\ge\chi^2(\alpha, p)$时拒绝 H_0
Σ未知：$T^2=n(\bar{x}-\mu_0)'S^{-1}(\bar{x}-\mu_0)\sim|_{H_0}T^2(p, n-1)$
证明：$(\bar{x}-\mu_0)\sqrt{n}\sim_{H_0}N(0, \Sigma)$, $(n-1)S\sim W_p(n-1, \Sigma)$. 由 T^2 的定义即得.
$(n-p)T^2/(p(n-1))\sim F(p, n-p)$
当 $T^2\ge T_\alpha^2=F(\alpha, p, n-p)p(n-1)/(n-p)$时拒绝 H_0
$x\sim N_p(0, \Sigma)$, $W\sim W_p(n, \Sigma)$,
$T^2=nx'W^{-1}x$ 是具有 n 个自由度的 Hotelling 分布 $T^2(p, n)$, 其与Σ无关。
证明：因为 $T^2=n(\Sigma^{-1/2}x)'(\Sigma^{-1/2}W\Sigma^{-1/2})^{-1}(\Sigma^{-1/2}x)$
$\Sigma^{-1/2}x\sim N_p(0, I)$, $\Sigma^{-1/2}W\Sigma^{-1/2}\sim W_p(n, I)$.

$(n-p+1)T^2(p, n)/(pn)\sim F(p, n-p+1)$

两个独立样本 $x_1, x_2, .., x_{n_1}\sim N_p(\mu_1, \Sigma)$; $y_1, y_2, .., y_{n_2}\sim N_p(\mu_2, \Sigma)$
H_0: $\mu_1=\mu_2$; H_1: $\mu_1\neq\mu_2$
均值无偏估计: $\bar{x}=\sum x_i/n_1$, $\bar{y}=\sum y_i/n_2$
联合方差无偏估计: $S_p=((n_1-1)S_1+(n_2-1)S_2)/(n_1+n_2-2)$
其中 $S_1=\sum_{i=1}^{n_1}(x_i-\bar{x})(x_i-\bar{x})'/(n_1-1)$, $S_2=\sum_{i=1}^{n_2}(y_i-\bar{y})(y_i-\bar{y})'/(n_2-2)$
Hotelling's T^2 检验：
$T^2=n_1n_2(\bar{x}-\bar{y})'S_p^{-1}(\bar{x}-\bar{y})/(n_1+n_2)$
$(n_1+n_2-p-1)T^2/(p(n_1+n_2-2))\sim F(p, n_1+n_2-p-1)$
证明：令 $d=(n_1+n_2-2)$
$T^2=d(1/n_1+1/n_2)^{-1/2}(\bar{x}-\bar{y})'[dS_p]^{-1}(1/n_1+1/n_2)^{-1/2}(\bar{x}-\bar{y})'$
注意到$(1/n_1+1/n_2)^{-1/2}(\bar{x}-\bar{y})'\sim N_p(0, \Sigma)$, $dS_p\sim W_p(d, \Sigma)$.
故 $T^2\sim T^2(p, d)$. 由 Hotelling's T^2 的性质即得。
$T^2\ge T_\alpha^2=F(\alpha, p, n_1+n_2-p-1)p(n_1+n_2-2)/(n_1+n_2-p-1)$拒绝 H_0.
$a'(\mu_1-\mu_2)$的 1-α同时置信区间:
$[a'(\bar{x}-\bar{y})\pm T_\alpha\sqrt{(n_1+n_2)/(n_1n_2)}\sqrt{a'S_pa}]$
k 很小时可将 T_α换成 $t(\alpha/2k, n_1+n_2-2)$

Pairing Test T^2 Statistic $(x_i, y_i)_{i=1, ..., p}$ 记 $d_i=x_i-y_i\sim N_p(\delta, \Sigma)$
H_0: $\delta=\mu_1-\mu_2=0$; H_1: $\delta\neq0$
$T^2=n\bar{d}'S_d^{-1}\bar{d}$. $S_d=\sum(d_i-\bar{d})(d_i-\bar{d})'/(n-1)$
$T^2\ge T_\alpha^2=F(\alpha, p, n-p)p(n-1)/(n-p)$时拒绝 H_0.

轮廓分析 Contour Analysis
单总体均值成分结构关系检验
$x\sim N_p(\mu, \Sigma)$, $\Sigma>0$, $x_1, ..., x_n$ 是样本
H_0: $\mu_1=...=\mu_p$, H_1: 存在$\mu_i\neq\mu_j$.
令 C=[[1 -1 0 ...][1 0 -1 0 ...] ... [1 ... -1]](对比矩阵)
H_0 变为 $C\mu=0$. $T^2=n\bar{x}'C'(CSC')^{-1}C\bar{x}\sim_{H_0}T^2(p-1, n-1)$
$T^2\ge(p-1)(n-1)F(\alpha, p-1, n-p+1)/(n-p+1)$时拒绝 H_0.

双总体均值成分结构关系检验
$x_1, x_2, .., x_{n_1}\sim N_p(\mu_1, \Sigma)$; $y_1, y_2, .., y_{n_2}\sim N_p(\mu_2, \Sigma)$. $n_1+n_2-2\ge p$.
H_0: $C(\mu_1-\mu_2)=\varphi$, H_1: $C(\mu_1-\mu_2)\neq\varphi$.
$T^2=n_1n_2(C(\bar{x}-\bar{y})-\varphi)'(CS_pC')^{-1}(C(\bar{x}-\bar{y})-\varphi)\sim T^2(k, n_1+n_2-2)$.

多总体均值比较检验
$x_{i1}, ..., x_{in_i}$ 属于$\pi_i\sim N_p(\mu_k, \Sigma)$, i=1, ..., k.
H_0: $\mu_1=...=\mu_k$. H_1: 存在$\mu_i\neq\mu_j$.
$T=SST=\sum_{i=1}^k\sum_{j=1}^{n_i}(x_{ij}-\bar{x})(x_{ij}-\bar{x})'=SSE+SSTR$ 自由度 n-1
$E=SSE=\sum_{i=1}^k\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)(x_{ij}-\bar{x}_i)'$ 组内矩阵 自由度 n-k
$H=SSTR=\sum_{i=1}^kn_i(\bar{x}_i-\bar{x})(\bar{x}_i-\bar{x})'$ 组间矩阵 自由度 k-1
$\Lambda=|E|/|E+H|\sim\Lambda(p, k-1, n-k)$ Wilks' Λ分布
$\Lambda<\Lambda_{1-\alpha}(p, k-1, n-k)$时拒绝 H_0.

$(r-p+1)(1-\sqrt{\Lambda(p,2,r)})/(p\sqrt{\Lambda(p,2,r)})\sim F(2p, 2(r-p+1))$.
大于 $F(\alpha, 2p, 2(r-p+1))$时拒绝 H_0.

协方差矩阵相等性检验
设 k 个多元正态总体$\pi_1, \pi_2, ..., \pi_k$ 服从 $N_p(\mu_i, \Sigma_i)$.
从 k 个总体中各自独立地抽取一个样本，设来自总体 i 的样本是
$x_{i1}, x_{i2}, ..., x_{in_i}$.
H_0: $\Sigma_1=...=\Sigma_k$ vs. H_1: $\exists i\neq j, \Sigma_i\neq\Sigma_j$.
$S_i=\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)(x_{ij}-\bar{x}_i)'/(n_i-1)$, $S_p=\sum_{i=1}^k(n_i-1)S_i/(n-k)=E/(n-k)$
修正似然比统计量：$\lambda=\prod_{i=1}^k|S_i|^{(n_i-1)/2}/|S_p|^{(n-k)/2}$
证明：$\lambda=\max_{\mu_i,\Sigma\ge0}L(\mu_1, ..., \mu_k, \Sigma)/\max_{\mu_i,\Sigma_i\ge0}L(\mu_1, \Sigma_1, ..., \mu_k, \Sigma_k)$
分母$=\prod_{i=1}^kL_i(\bar{x}_i, S_i^{MLE})=\prod_{i=1}^k(2\pi)^{-n_ip/2}|S_i^{MLE}|^{-n_i/2}e^{-n_ip/2}$
$=(2\pi e)^{-np/2}\prod_{i=1}^k|S_i^{MLE}|^{-n_i/2}$.
分子$=(2\pi e)^{-np/2}|A/n|^{-n/2}$, $A=\sum_{i=1}^kA_i=\sum_{i=1}^kn_iS_i^{MLE}$.
故$\lambda=|A/n|^{-n/2}/\prod_{i=1}^k|A_i/n_i|^{-n_i/2}$.
根据无偏性，用 n_i-1 代替 n_i，用 n-k 代替 n，再取对数，即得
Box-M 统计量: $M=-2\ln\lambda^*=(n-k)\ln|S_p|-\sum_{i=1}^k(n_i-1)\ln|S_i|$
$c=(\sum_{i=1}^k1/(n_i-1)-1/(n-k))(2p^2+3p+1)/(6(p+1)(k-1))$
$u=(1-c)M\sim_{H_0,asymp}\chi^2((k+1)p(p+1)/2)$
$u>\chi^2(\alpha, (k+1)p(p+1)/2)$时拒绝 H_0.

V 判别分析 Discrimination Analysis
平方马氏距离 $d^2(x, \pi)=(x-\mu_1)'\Sigma(x-\mu_1)$

组间距离判别
① $\Sigma_1=\Sigma_2=\Sigma$
判据：$d^2(x, \pi_1)\le d^2(x, \pi_2)$时，认为 $x\in\pi_1$，否则认为 $x\in\pi_2$
令$\bar{\mu}=(\mu_1+\mu_2)/2$, $a=\Sigma^{-1}(\mu_1-\mu_2)$,
$d^2(x, \pi_1)-d^2(x, \pi_2)=-2a'(x-\bar{\mu})$
令 $W(x)=a'(x-\bar{\mu})$，则判据变为 $W(x)\ge0$ 时为π_1
$W(x)$称为线性判别函数
误判概率
$p(2|1)=P(W(x)<0|x\in\pi_1)$, $p(1|2)=P(W(x)\ge0|x\in\pi_2)$
$\Delta=\sqrt{(\mu_1-\mu_2)'\Sigma^{-1}(\mu_1-\mu_2)}$ is Mahalanobis distance
$W(x)\sim N(\Delta^2/2, \Delta^2)$, $p(2|1)=p(1|2)=\Phi(-\Delta/2)$
样本估计
π_1 中抽取 $x_{11}, ..., x_{1n_1}$, π_2 中抽取 $x_{22}, ..., x_{2n_2}$. $n_1+n_2-2\ge p$.
$\hat{\Sigma}=S_p=(A_1+A_2)/(n_1+n_2-2)$是无偏估计
$A_i=\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)(x_{ij}-\bar{x}_i)'$, $\hat{\Delta}=\sqrt{(\bar{x}_1-\bar{x}_2)'\Sigma^{-1}(\bar{x}_1-\bar{x}_2)}$
$\bar{x}=(\bar{x}_1+\bar{x}_2)/2$, $\hat{a}=S_p^{-1}(\bar{x}_1-\bar{x}_2)$, $\hat{W}(x)=\hat{a}'(x-\bar{\mu})\ge0$ 时为π_1.
π_1、π_2 是正态组时，$\hat{p}(2|1)=\hat{p}(1|2)=\Phi(-\hat{\Delta}/2)$
误判概率的非参数估计（各组非正态时）
$\hat{p}(2|1)=n(2|1)/n_1$; $\hat{p}(1|2)=n(1|2)/n_2$.

② $\Sigma_1\neq\Sigma_2$:
判据：$d^2(x, \pi_1)\le d^2(x, \pi_2)$时，认为 $x\in\pi_1$，否则认为 $x\in\pi_2$
$W(x)=d^2(x,\pi_1)-d^2(x, \pi_2)=(x-\mu_1)'\Sigma_1^{-1}(x-\mu_1)-(x-\mu_2)'\Sigma_2^{-1}(x-\mu_2)$
则判据等价于 $W(x)\le0$

多组判别分析
当 $d^2(x, \pi_l)=\min_id^2(x, \pi_i)$时，将 x 分到π_l 中。
当各组方差相等时，简化为 $l=\arg\max_i((\Sigma^{-1}\mu_i)'x-\mu_i'\Sigma^{-1}\mu_i/2)=\arg\max_i(I_i'x+c_i)$
无偏估计$\hat{\Sigma}=S_p=\sum_{i=1}^k(n_i-1)S_i/(n-k)$
其中，$S_i=\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)(x_{ij}-\bar{x}_i)'/(n_i-1)$

Box-M 检验可用于检验 H_0: $\Sigma_1=...=\Sigma_k$ vs. H_1: $\exists i\neq j, \Sigma_i\neq\Sigma_j$.

贝叶斯判别
最大后验概率法
假设 k 个组的概率密度为 $f_i(x)$，样本来自各组的先验概率为 p_i.
$p_1+...+p_k=1$. 则后验概率 $P(\pi_i|x)=p_if_i(x)/\sum_{j=1}^kp_jf_j(x)$.
x 归为后验概率最大的组。

k 个组均为正态组时，
令 $g_i=\Sigma_1=...=\Sigma_k=\Sigma$? 0 : $\ln|\Sigma_i|$,
$h_i=p_1=...=p_k=1/k$? 0 : $-2\ln p_i$.
称 $D^2(x, \pi_i)=d^2(x, \pi_i)+g_i+h_i$ 为 x 到π_i 的广义平方距离
$P(\pi_i|x)=\exp(-D^2(x, \pi_i)/2)/\sum_{j=1}^k\exp(-D^2(x, \pi_j)/2)$

等价表达：令 $I_i=\Sigma^{-1}\mu_i=S_p^{-1}\bar{x}_i$, $c_i=-\mu_i'\Sigma^{-1}\mu_i/2=-\bar{x}_i'S_p^{-1}\bar{x}_i/2$,
$P(\pi_i|x)=\exp(I_i'x+c_i+\ln p_i)/\sum_{j=1}^k\exp(I_j'x+c_j+\ln p_j)$

最小期望误判代价（ECM）法
$P(2|1)=\int_{R_2}f_1(x)dx$, $P(1|2)=\int_{R_1}f_2(x)dx$
$ECM=c(2|1)P(2|1)p_1+c(1|2)P(1|2)p_2$. $c(|)$为误判代价.
判别规则：$f_1(x)/f_2(x)\ge c(1|2)p_2/c(2|1)p_1$? π_1 : π_2.
先验概率难以给出时通常取等。

两个正态组
$\Sigma_1=\Sigma_2=\Sigma$时，$\bar{\mu}=(\mu_1+\mu_2)/2$, $a=\Sigma^{-1}(\mu_1-\mu_2)$
$a(x-\bar{\mu})\ge\ln(c(1|2)p_2/c(2|1)p_1)$? π_1 : π_2.
$\Sigma_1\neq\Sigma_2$ 时，
$d^2(x, \pi_1)-d^2(x, \pi_2)\le2\ln(c(2|1)p_1|\Sigma_2|^{1/2}/c(1|2)p_2|\Sigma_1|^{1/2})$? π_1:π_2.

多组
$x\in\pi_l$ if $l=\arg\min_i\sum_{j\neq i}^k(p_jf_j(x)c(i|j))$
误判代价相同时，$ECM=1-\sum_{i=1}^kp_iP(i|i)$

Fisher 判别/典型判别
用 p 维向量 x 的 r<p 个线性组合 $a_1'x, ..., a_r'x$ 降维。
$E=\sum_{i=1}^k(n_i-1)S_i$. $H=\sum_{i=1}^kn_i(\bar{x}_i-\bar{x})(\bar{x}_i-\bar{x})'$
$\Delta(a)=SSTR/SSE=a'Ha/a'Ea$ 度量各组分离程度，应达到最大

$F=SSTR(n-k)/SSE(k-1)\sim F(k-1,n-k)$. F 越大，组间差异越大。

$E^{-1}H$ 的特征值为$\lambda_1\ge...\ge\lambda_s>0$, $s=rank(H)\le\min(k-1, p)$.
其对应的特征向量为 $t_1, ..., t_s$.
$y_1=t_1'x$, $y_2=t_2'x$,...为第 i 个 Fisher 线性判别函数
特性：各判别函数均具有单位方差，且彼此不相关，判别方向 t_i 不一定正交
$\Delta(t_i)=\lambda_i$，记前 r 个特征向量的累计贡献率为$\sum_{i=1}^r\lambda_i/\sum_{i=1}^s\lambda_i$.
判别规则：$l=\arg\min_i\sum_{j=1}^r(y_j-y_{ij})^2=\arg\min_i\sum_{j=1}^r(t_j'(x-\bar{x}_i))^2$.
组数为 2 时，有唯一非零特征值
$\lambda=(\bar{x}_1-\bar{x}_2)'E^{-1}(\bar{x}_1-\bar{x}_2)n_1n_2/(n_1+n_2)>0$
取 $y=(\bar{x}_1-\bar{x}_2)S_p^{-1}x$, $S_p=E/(n_1+n_2-2)$.
此时判别规则为$|y-y_1|\le|y-y_2|$?π_1:π_2.
等价于 $y\ge(\bar{y}_1+\bar{y}_2)$? π_1 : π_2. $y_i=t_i'\bar{x}$.
对于 2 组的判别，Fisher 判别等价于协方差矩阵相等的距离判别，对两个正态组也等价于协方差相等且先验概率和误判代价均相同的贝叶斯判别。

VI 聚类分析 Cluster Analysis
Q Type：样本聚类；R Type：变量聚类
相似性度量：Distance、Similarity Coefficient
距离 Distance：$d_{ij}\ge0$; $d_{ij}=0\leftrightarrow i=j$; $d_{ij}=d_{ji}$; $d_{ij}\le d_{ik}+d_{kj}$.
Minkowski Distance: $d(x, y)=[\sum_{i=1}^p|x_i-y_i|^q]^{1/q}$.
标准化处理：$x_i^*=(x_i-\bar{x}_i)/s_{ii}$.
Lance Distance: $d(x, y)=\sum_{i=1}^p|x_i-y_i|/(x_i+y_i)$.适用高度偏斜、异常的数据
Mahalanobis Distance: $d(x, y)=\sqrt{(x-y)'S^{-1}(x-y)}$.
斜交空间距离: $d(x, y)=[\sum_{i=1}^p\sum_{j=1}^p(x_i-y_i)(x_j-y_j)r_{ij}/p^2]^{1/2}$. r_{ij} 是相关系数。
基于非负定相关系数矩阵构造：$d_{ij}=\sqrt{2(1-c_{ij})}$

相似系数：$c_{ij}=\pm1\leftrightarrow x_i=ax_j+b, a\neq0$; $|c_{ij}|\le1$; $c_{ij}=c_{ji}$
夹角余弦：$c_{ij}=x_ix_j/(||x_i||_2\cdot||x_j||_2)$
相关系数：$c_{ij}=r_{ij}=\sum_{k=1}^n(x_{ki}-\bar{x}_i)(x_{kj}-\bar{x}_j)/\sqrt{[\sum_{k=1}^n(x_{ki}-\bar{x}_i)^2](\sum_{k=1}^n(x_{kj}-\bar{x}_j)^2)}$
基于距离的定义：$c_{ij}=1/(1+d_{ij})$.

系统/层次聚类法 Hierarchical Clustering Method

Agglomerative 聚集系统法：初始所有样本自成一类，每次迭代进行一次合并，下面记合并类为 M。

单连接法：选最小的 $D_{KL}=\min_{i\in GK,\,j\in GL}d_{ij}$., $D_{MJ}=\min(D_{KJ}, D_{LJ})$.

　　有链接倾向，不适合对分离得很差的群体进行聚类。

完全连接法：选最小的 $D_{KL}=\max_{i\in GK,\,j\in GL}d_{ij}$. $D_{MJ}=\max(D_{KJ}, D_{LJ})$.

　　容易被异常值扭曲

类平均法：$D_{KL}=\sum_{i\in GK,\,j\in GL}d_{ij}/n_Kn_L$. $D_{MJ}=n_KD_{KJ}/n_M+n_LD_{LJ}/n_M$.

类平均平方距离法：$D_{KL}^2=\sum_{i\in GK,\,j\in GL}d_{ij}^2/n_Kn_L$. $D_{MJ}^2=n_KD_{KJ}^2/n_M+n_LD_{LJ}^2/n_M$.

重心法：$D_{KL}^2=(\mathbf{x}_K^{-}-\mathbf{x}_L^{-})'(\mathbf{x}_K^{-}-\mathbf{x}_L^{-})$.

$D_{MJ}^2=n_KD_{KJ}^2/n_M+n_LD_{LJ}^2/n_M-n_Kn_LD_{KL}^2/n_M^2$. 较为稳健，但其他方面一般不如类平均法或离差平方和法好

中间距离法：$\mathbf{m}_M=(\mathbf{x}_K^{-}+\mathbf{x}_L^{-})/2$ 代替重心 $\mathbf{x}_M^{-}$. $D_{MJ}^2=D_{KJ}^2/2+D_{LJ}^2/2-D_{KL}^2/4$.

离差平方和（Ward）法：

$W_K=\sum_{i\in GK}(\mathbf{x}_i-\mathbf{x}_K^{-})'(\mathbf{x}_i-\mathbf{x}_K^{-})$, $D_{KL}^2=W_M-W_K-W_L=n_Kn_L(\mathbf{x}_K^{-}-\mathbf{x}_L^{-})'(\mathbf{x}_K^{-}-\mathbf{x}_L^{-})/n_M$

$D_{MJ}^2=(n_J+n_K)D_{KJ}^2/(n_J+n_M)+(n_J+n_L)D_{LJ}^2/(n_J+n_M)-n_JD_{KL}^2/(n_J+n_M)$.

　　两个大类容易分离；两个小类容易合并；对异常值较敏感

一般的系统聚类法

$D_{MJ}^2=\alpha_KD_{KJ}^2+\alpha_LD_{LJ}^2-\beta D_{KL}^2+\gamma|D_{KJ}^2-D_{LJ}^2|$.

聚类法的性质：

①单调性：

单连接法、完全连接法、类平均法、离差平方和法具有单调性

重心法、中间距离法不具有单调性

②空间浓缩与扩张　略

Divisive 分割系统法：所有样本归为 1 类，每次迭代进行一次分割

树形图 Dendrogram：并类的距离 - 类

确定类的个数

　给定距离阈值 T，所有类间距离均大于 T 时停止

　目测

统计量：　$W_i=\sum_{j\in Gi}(\mathbf{x}_j-\mathbf{x}_i^{-})'(\mathbf{x}_j-\mathbf{x}_i^{-})$是组内离差平方和

$R^2=1-P_k/W=1-\sum_{i=1}^{k}W_i/W$. $W=P_k+\sum_{i=1}^{k}n_i(\mathbf{x}_i^{-}-\mathbf{x}^{-})'(\mathbf{x}_i^{-}-\mathbf{x}^{-})$. R^2 越大越好。

半偏 $R^2=(W_M-W_K-W_L)/W$, 即一次迭代后和迭代前 R^2 之差。越大越好。

伪 $F=(W-P_k)(n-k)/(P_k(k-1))=(n-k)R^2/((k-1)(1-R^2))$. 越大越好。

伪 $t^2=(W_M-W_K-W_L)(n_K+n_L-2)/(W_K+W_L)$. 越大越好。

动态聚类法

　k 均值法：选择 k 个凝聚点……

VII 主成分分析 Principal Component Analysis

希望求得 $\mathbf{a}_1$, 使 $\mathrm{Var}(y_1)=\mathrm{Var}(\mathbf{a}_1'\mathbf{x})=\mathbf{a}_1'\Sigma\mathbf{a}_1$ 达最大。y_1 称为第一主成分

限制$||\mathbf{a}_1||=1$. 则 $\mathbf{a}_1=\mathbf{t}_1$,为协方差矩阵$\Sigma$最大特征值$\lambda_1$ 对应的特征向量.

考虑正交性，可继续定义 $\mathbf{a}_2=\mathbf{t}_2$, ...共 p 个主成分。

记 $T=[\mathbf{t}_1\ \mathbf{t}_2 \dots \mathbf{t}_p]$, $\Lambda=\mathrm{diag}(\lambda_1, \lambda_2, ..., \lambda_p)$. $\mathbf{y}=T'\mathbf{x}$, $\Sigma=T\Lambda T'$.

$V(\mathbf{y})=\Lambda\rightarrow$各主成分互不相关。　$\sum_{i=1}^{p}\lambda_i=\sum_{i=1}^{p}\sigma_{ii}=\mathrm{tr}(\Sigma)$.

$\mathbf{y}=T'\mathbf{x}\leftrightarrow y_k=t_{1k}x_1+...+t_{pk}x_p$.称 t_{ik} 第 k 主成分在第 i 原始变量 x_i 上的载荷,其反映了 x_i 对 y_k 的重要程度。$\sigma_{ii}=t_{i1}^2\lambda_1+t_{i2}^2\lambda_2+...+t_{ip}^2\lambda_p$.

$\lambda_1\geq\max_i\sigma_{ii}\geq\min_i\sigma_{ii}\geq\lambda_p$.

基于相关矩阵的主成分

各变量的单位不全相同时，或各变量的方差数值差异较大时，应基于标准化变量的协方差矩阵进行主成分分析。$x_i^*=(x_i-\mu_i)/\sqrt{\sigma_i^2}$。$\sum_{i=1}^{p}\lambda_i^*=p$.

累计贡献率：$\sum_{i=1}^{m}\lambda_i/\sum_{i=1}^{p}\lambda_i$

变量 x_i^*与主成分 y_k^*的相关系数　$\rho(x_i^*, y_k^*)=t_{ik}^*\sqrt{\lambda_k^*}$.

主成分对变量 x_i^*累计贡献率：$\rho^2_{i,\,1\sim m}=\sum_{k=1}^{m}\rho^2(x_i^*, y_k^*)$

$\sum_{k=1}^{p}\rho^2(x_i^*, y_k^*)=t_{ik}^{*2}\lambda_k^*=1$

样本主成分分析

将Σ替换为样本协方差矩阵 $S=\sum_{i=1}^{n}(\mathbf{x}_i-\mathbf{x}^{-})(\mathbf{x}_i-\mathbf{x}^{-})'/(n-1)=[s_{ij}]$

$R^{\wedge}=[r_{ij}]=[s_{ij}/\sqrt{(s_{ii}s_{jj})}]$

在实际应用时，通常减去均值进行中心化，此时有中心化的主成分

$y_i^{\wedge}=\mathbf{t}_i^{\wedge}{}'(\mathbf{x}-\mathbf{x}^{-})$. $\mathbf{x}=[x_1 \dots x_p]'$, $\mathbf{x}^{-}$是 $\mathbf{x}$ 的样本均值向量。

若将第 j 个观测值向量 $\mathbf{x}_j$ 代入 $\mathbf{x}$, 则有

$y_{ji}^{\wedge}=\mathbf{t}_i^{\wedge}{}'(\mathbf{x}_j-\mathbf{x})$称为观测值 $\mathbf{x}_j$ 的第 i 主成分得分。

所有观测值的平均主成分得分 $y_i^{\wedge -}=\mathbf{t}_i^{\wedge}{}'(\sum_{j=1}^{n}\mathbf{x}_j-n\mathbf{x}^{-})/n=0$.

基于样本相关矩阵的主成分分析

$\mathbf{x}^*=D^{\wedge-1}(\mathbf{x}-\mathbf{x}^{-})$, $D^{\wedge}=\mathrm{diag}(\sqrt{s_{11}}, ..., \sqrt{s_{pp}})$. $\mathbf{x}_j^*=D^{\wedge-1}(\mathbf{x}_j-\mathbf{x}^{-})$

则第 i 样本主成分 $y_i^{\wedge *}=\mathbf{t}_i^*{}'\mathbf{x}^*$.

观测值 $\mathbf{x}_j$ 的第 i 主成分得分 $y_{ji}^{\wedge}=\mathbf{t}_i^{\wedge}{}'\mathbf{x}_j^*$.

VIII 因子分析 Factor Analysis

正交因子模型

因子分析的一般模型可以表达为 $\mathbf{x}=\boldsymbol{\mu}+A\mathbf{f}+\boldsymbol{\varepsilon}$.　（★）

$\mathbf{f}=[f_1 \dots f_m]'$为公共因子向量；$\varepsilon=[\varepsilon_1 \dots \varepsilon_p]'$为特殊因子向量

$A_{p\times m}=[a_{ij}]$为因子载荷矩阵

假定 $E(\mathbf{f})=\mathbf{0}$; $E(\boldsymbol{\varepsilon})=\mathbf{0}$; $V(\mathbf{f})=I$; $V(\boldsymbol{\varepsilon})=D=\mathrm{diag}(\sigma_1^2, ..., \sigma_p^2)$. $\mathrm{Cov}(\mathbf{f}; \varepsilon)=0$.

$\Sigma=V(A\mathbf{f})+D=AV(\mathbf{f})A'+D=AA'+D$.

因子载荷矩阵

$\mathrm{Cov}(\mathbf{x}, \mathbf{f})=\mathrm{Cov}(A\mathbf{f}+\varepsilon, \mathbf{f})=AV(\mathbf{f})=A$.　$\mathrm{Cov}(x_i, f_j)=a_{ij}$.

A 的行元素平方和 $h_i^2=\sum_{j=1}^{m}a_{ij}^2$. $\sigma_{ii}=h_i^2+\sigma_i^2$. 其反映了公共因子 $f_1, ..., f_m$ 对 x_i 的方差贡献，称为共性方差，σ_i^2 称为特殊方差。

若 $\mathbf{x}$ 已经经过标准化，则 $h_i^2+\sigma_i^2=1$.

A 的列元素平方和 $g_j^2=\sum_{i=1}^{m}a_{ij}^2$.反映了公共因子 f_j 对 $x_1, ..., x_p$ 的影响，可以视为 f_j 对 $x_1, ..., x_p$ 的总方差贡献。贡献率$=g_j^2/\sum_{i=1}^{p}\mathrm{Var}(x_i)$.

参数估计：使用样本均值 $\mathbf{x}^{-}$和样本协方差 S 时，需要估计 A 和 D。

①主成分法

$S=\lambda_1^{\wedge}\mathbf{t}_1^{\wedge}\mathbf{t}_1^{\wedge}{}'+...+\lambda_m^{\wedge}\mathbf{t}_m^{\wedge}\mathbf{t}_m^{\wedge}{}'+...+\lambda_p^{\wedge}\mathbf{t}_p^{\wedge}\mathbf{t}_p^{\wedge}{}'\approx\lambda_1^{\wedge}\mathbf{t}_1^{\wedge}\mathbf{t}_1^{\wedge}{}'+...+\lambda_m^{\wedge}\mathbf{t}_m^{\wedge}\mathbf{t}_m^{\wedge}{}'+D^{\wedge}$

$=A^{\wedge}A^{\wedge}{}'+D^{\wedge}$. 令$\sigma_i^{\wedge 2}=s_{ii}-\sum_{j=1}^{m}a_{ij}^2$.

其中，$A^{\wedge}=[\mathbf{t}_1^{\wedge}\sqrt{\lambda_1^{\wedge}} \dots \mathbf{t}_m^{\wedge}\sqrt{\lambda_m^{\wedge}}]$, $D^{\wedge}=\mathrm{diag}(\sigma_1^{\wedge 2}, ..., \sigma_p^{\wedge 2})$.「主成分解」

$A^{\wedge}$的第 j 列的元素平方和等于$\lambda_j^{\wedge}$.

残差矩阵 $S-(A^{\wedge}A^{\wedge}{}'+D^{\wedge})$的对角元为 0，其元素平方和$\leq\lambda_{m+1}^2+...+\lambda_p^2$.

证明：$||S-(A^{\wedge}A^{\wedge}{}'+D^{\wedge})||_F^2\leq||S-A^{\wedge}A^{\wedge}{}'||_F^2=||\sum_{i=m+1}^{p}\lambda_i^{\wedge}\mathbf{t}_i^{\wedge}\mathbf{t}_i^{\wedge}{}'||_F^2$

$=||[\mathbf{t}_{m+1} \dots \mathbf{t}_p]\Lambda[\mathbf{t}_{m+1} \dots \mathbf{t}_p]'||_F^2$

$=\mathrm{tr}([\mathbf{t}_{m+1} \dots \mathbf{t}_p]\Lambda[\mathbf{t}_{m+1} \dots \mathbf{t}_p]'[\mathbf{t}_{m+1} \dots \mathbf{t}_p]\Lambda[\mathbf{t}_{m+1} \dots \mathbf{t}_p]')$

$=\mathrm{tr}(\Lambda^2)=\lambda_{m+1}^2+...+\lambda_p^2$

也可以先将 $\mathbf{x}$ 做标准化，使用样本相关矩阵 R 代替 S。

②主因子法

假定 $\mathbf{x}$ 已经做过了标准化变换，且满足因子模型★。

相关矩阵 $R=AA'+D$,　　约相关矩阵 $R^*=R-D=AA'$.

若$\sigma_i^{\wedge 2}$ 是特殊方差σ_i^2 的一个合适的初始估计，则约相关矩阵可估计为 $R^{*\wedge}=R^{\wedge}-D^{\wedge}$. 设 $R^{*\wedge}$的前 m 个特征值是$\lambda_1^{*\wedge}\geq...\geq\lambda_m^{*\wedge}>0$.对应特征向量 $\mathbf{t}_1^{*\wedge}, ..., \mathbf{t}_m^{*\wedge}$. 则 A 的主因子解为 $A^{\wedge}=[\mathbf{t}_1^{*\wedge}\sqrt{\lambda_1^{*\wedge}} \dots \mathbf{t}_m^{*\wedge}\sqrt{\lambda_m^{*\wedge}}]$

此时可以重新估计特殊方差$\sigma_i^{\wedge 2}=1-h_i^{\wedge 2}=1-\sum_{j=1}^{m}a^{\wedge}{}_{ij}^2$, i=1, 2, ..., p.

也可以基于重新估计的特殊方差迭代求解。

特殊方差的初始估计：

I. $\sigma_i^{2\wedge}=1/r^{ii}$. r^{ii} 是 $R^{\wedge-1}$的第 i 个对角元。要求 $R^{\wedge}$满秩。

II. 取 $h_i^{2\wedge}=\max_{j\neq i}|r_{ij}|$, 此时$\sigma_i^{\wedge 2}=1-h_i^{\wedge 2}$.

III. 取 $h_i^{2\wedge}=1$. 如此得到的是一个主成分解。

③极大似然法

设公共因子 $\mathbf{f}\sim N_m(\mathbf{0}, I)$, 特殊因子$\boldsymbol{\varepsilon}\sim N_p(\mathbf{0}, D)$.且相互独立。则有 $\mathbf{x}\sim N_p(\boldsymbol{\mu}, \Sigma)$。可由样本 $\mathbf{x}_1, ..., \mathbf{x}_n$ 导出似然函数 $L(\boldsymbol{\mu}, \Sigma)$.

因为$\Sigma=AA'+D$, 所以似然函数可以表示为 $L(\boldsymbol{\mu}, A, D)$.

记参数的极大似然估计分别为$\boldsymbol{\mu}^{\wedge}, A^{\wedge}, D^{\wedge}$.

即 $L(\boldsymbol{\mu}^{\wedge}, A^{\wedge}, D^{\wedge})=\max L(\boldsymbol{\mu}, A, D)$.

$\boldsymbol{\mu}^{\wedge}=\mathbf{x}^{-}$. 记$\Sigma^{\wedge}=\sum_{i=1}^{n}(\mathbf{x}_i-\mathbf{x}^{-})(\mathbf{x}_i-\mathbf{x}^{-})'/n$, 则 $A^{\wedge}, D^{\wedge}$满足方程组

①$\Sigma^{\wedge}D^{\wedge-1}A^{\wedge}=A^{\wedge}(I_m+A^{\wedge}{}'D^{\wedge-1}A^{\wedge})$. ②$D^{\wedge}=\mathrm{diag}(\Sigma^{\wedge}-A^{\wedge}A^{\wedge}{}')$

在条件 $A'D^{-1}A$ 是对角矩阵的限制下，A 有唯一解。

因子旋转

希望 A 的元素都接近 0 或±1，区别度最大，使模型的因子易于解释

正交旋转

$\mathbf{f}^*=T'\mathbf{f}$, $A^*=AT$. 若记 $A=[\mathbf{a}_1 \dots \mathbf{a}_p]'$, $A^*=[\mathbf{a}_1^* \dots \mathbf{a}_p^*]'$,则 $\mathbf{a}_i^*=T'\mathbf{a}_i$.

正交旋转不改变共性方差，也不改变残差矩阵。

最大方差旋转法

记 $A^*=[a_{ij}^*]$, $d_{ij}=a_{ij}^*/h_i$, $d_j^{-}=\sum_{i=1}^{p}d_{ij}^2/p$.

A^*第 j 列的元素平方的相对方差定义为 $V_j=\sum_{i=1}^{p}(d_{ij}^2-d_j^{-})^2$.

希望选择 T，使 $V=V_1+...+V_m$ 达到最大

因子得分

希望给出 $\mathbf{x}_j$ 关于 m 个公共因子的得分。即对不可观测变量 $f_1, ..., f_m$ 做出估计。

加权最小二乘法（Bartlett）

希望得到一组估计 $f_1^{\wedge}, ..., f_m^{\wedge}$. 使得加权的「残差」平方和达到最小

即 $\min_{\mathbf{f}}\sum_{i=1}^{p}[x_i-(\mu_i+a_{i1}f_1^{\wedge}+...+a_{im}f_m^{\wedge})]^2/\sigma_i^2$.

矩阵形式：$\min_{\mathbf{f}}(\mathbf{x}-\boldsymbol{\mu}-A\mathbf{f}^{\wedge})'D^{-1}(\mathbf{x}-\boldsymbol{\mu}-A\mathbf{f}^{\wedge})$　解得 $\mathbf{f}^{\wedge}=(A'D^{-1}A)^{-1}A'D^{-1}(\mathbf{x}-\boldsymbol{\mu})$

第 j 个样本的因子得分即为 $\mathbf{f}_j^{\wedge}=(A^{\wedge}{}'D^{\wedge-1}A^{\wedge})^{-1}A^{\wedge}{}'D^{\wedge-1}(\mathbf{x}_j-\mathbf{x}^{-})$.

回归法（Thompson）

假设$[\mathbf{f}'\ \boldsymbol{\varepsilon}']'$服从 m+p 元正态分布，则

$$\begin{bmatrix}\boldsymbol{f}\\ \boldsymbol{x}\end{bmatrix}=\begin{bmatrix}\boldsymbol{f}\\ \boldsymbol{\mu}+A\boldsymbol{f}+\boldsymbol{\varepsilon}\end{bmatrix}=\begin{bmatrix}0\\ \boldsymbol{\mu}\end{bmatrix}+\begin{bmatrix}I_m & 0\\ A & I_p\end{bmatrix}\begin{bmatrix}\boldsymbol{f}\\ \boldsymbol{\varepsilon}\end{bmatrix}$$

亦服从 m+p 元正态分布。

$$\mathrm{E}\begin{bmatrix}\boldsymbol{f}\\ \boldsymbol{x}\end{bmatrix}=\begin{bmatrix}0\\ \boldsymbol{\mu}\end{bmatrix},\mathrm{Var}\begin{bmatrix}\boldsymbol{f}\\ \boldsymbol{x}\end{bmatrix}=\begin{bmatrix}I_m & A'\\ A & AA'+D\end{bmatrix}=\begin{bmatrix}I_m & A'\\ A & \Sigma\end{bmatrix}$$

$\mathbf{f}^{\wedge}=E(\mathbf{f}\mid\mathbf{x})=A'(AA'+D)^{-1}(\mathbf{x}-\boldsymbol{\mu})=(I+A'D^{-1}A)^{-1}A'D^{-1}(\mathbf{x}-\boldsymbol{\mu})$

为减少误差，我们更倾向于用 $\mathbf{x}^{-}$, $A^{\wedge}$, S 代替$\boldsymbol{\mu}$, A, Σ计算得分。

第 j 个样本的因子得分即为 $\mathbf{f}_j^{\wedge}=A^{\wedge}{}'S^{-1}(\mathbf{x}_j-\mathbf{x}^{-})$.

若经过标准化，则有 $\mathbf{f}^{*\wedge}=A^{*\wedge}R^{\wedge-1}\mathbf{x}^*$.

无偏性：

加权最小二乘法因子是无偏的，因为

$E(\mathbf{f}_b^{\wedge}|\mathbf{f})=(A'D^{-1}A)^{-1}A'D^{-1}E(A\mathbf{f}+\boldsymbol{\varepsilon}|\mathbf{f})=(A'D^{-1}A)^{-1}A'D^{-1}A\mathbf{f}=\mathbf{f}$.

回归法因子是有偏的，因为

$E(\mathbf{f}_t^{\wedge}|\mathbf{f})=(I+A'D^{-1}A)^{-1}A'D^{-1}E(A\mathbf{f}+\boldsymbol{\varepsilon}|\mathbf{f})=\mathbf{f}-(I+A'D^{-1}A)^{-1}\mathbf{f}$.

有效性：

考虑平均预报误差矩阵 $E[(\mathbf{f}^{\wedge}-\mathbf{f})(\mathbf{f}^{\wedge}-\mathbf{f})']$，有

加权最小二乘法：

$\mathbf{f}_b^{\wedge}-\mathbf{f}=(A'D^{-1}A)^{-1}A'D^{-1}(A\mathbf{f}+\boldsymbol{\varepsilon})-\mathbf{f}=(A'D^{-1}A)^{-1}A'D^{-1}\boldsymbol{\varepsilon}$.

$E[(\mathbf{f}_b^{\wedge}-\mathbf{f})(\mathbf{f}_b^{\wedge}-\mathbf{f})']=(A'D^{-1}A)^{-1}A'D^{-1}E(\boldsymbol{\varepsilon\varepsilon}')D^{-1}A(A'D^{-1}A)^{-1}=(A'D^{-1}A)^{-1}$.

回归法：

$\mathbf{f}_t^{\wedge}-\mathbf{f}=(I+A'D^{-1}A)^{-1}A'D^{-1}(A\mathbf{f}+\boldsymbol{\varepsilon})-\mathbf{f}=(I+A'D^{-1}A)^{-1}A'D^{-1}\boldsymbol{\varepsilon}-(I+A'D^{-1}A)^{-1}\mathbf{f}$

$E[(\mathbf{f}_t^{\wedge}-\mathbf{f})(\mathbf{f}_t^{\wedge}-\mathbf{f})']=(I+A'D^{-1}A)^{-1}$.

考虑到

$(A'D^{-1}A)^{-1}-(I+A'D^{-1}A)^{-1}=(I+A'D^{-1}A)^{-1}(I+A'D^{-1}A-A'D^{-1}A)(A'D^{-1}A)^{-1}$

$=(I+A'D^{-1}A)^{-1}(A'D^{-1}A)^{-1}=(A'D^{-1}A+A'D^{-1}AA'D^{-1}A)^{-1}>0$　正定

因此回归法因子比加权最小二乘法因子具有更高的估计精度。

回归法的导出不依赖正态性假设，估计精度不受正态性影响。

辨析

①主成分分析涉及的只是一般的变量变换,它不能作为一个模型来描述,本质上几乎不需要任何假定;而因子分析需要构造一个因子模型,并伴有几个关键性的假定。

②主成分是原始变量的线性组合;而在因子分析中,原始变量是因子的线性组合,但因子却一般不能表示为原始变量的线性组合。

③在主成分分析中，强调的是用少数几个主成分解释总方差;而在因子分析中,强调的是用少数几个因子去描述协方差或相关关系。

④主成分的解是唯一的(除非含有相同的特征值或特征向量为相反符号);而因子的解可以有很多,表现得较为灵活(主要体现在因子旋转上),这种灵活性使得变量在降维之后更易得到解释,这是因子分析比主成分分析有更广泛应用的一个重要原因。

⑤主成分不会因其提取个数的改变而变化,但因子往往会随模型中因子个数的不同而变化。

下雨：

1:

$\mathbf{a}=S_p^{-1}(\mathbf{x}_1^{-}-\mathbf{x}_2^{-})$, $W=\mathbf{a}'(\mathbf{x}_0-\mathbf{x}^{-})$

$W(\mathbf{x}_0)=0.9499$; $W(\mathbf{x})=[0.078\ -0.178](\mathbf{x}-\mathbf{x}^{-})\geq0?\ \mathbf{x}\in\pi_1:\ \mathbf{x}\in\pi_2$.

2:　$p_1=0.4$, $p_2=0.6$

$\mathbf{I}_i=S_p^{-1}\mathbf{x}_i^{-}$, $c_i = -\mathbf{x}_i'S_p^{-1}\mathbf{x}_i/2$

$\mathbf{x}\in\mathrm{argmax}_{\pi i}\Pr(\pi_i|\mathbf{x})=\exp(\mathbf{I}'_i\mathbf{x}+c_i+\ln p_i)/\sum_{j=1}^{k}\exp(\mathbf{I}_j'\mathbf{x}+c_i+\ln p_j)$.

$\Pr(\pi_1|\mathbf{x})=0.6328 > \Pr(\pi_2|\mathbf{x})=0.3672$. 所以预测为下雨。

3:

$W(\mathbf{x})=[0.078\ -0.178](\mathbf{x}-\mathbf{x}^{-})\geq\ln((c(1|2)p_2)/(c(2|1)p_1)=-0.69?\ \mathbf{x}\in\pi_1:\ \mathbf{x}\in\pi_2$.

机场：

Old: $\mathbf{X}_{11}=[10\ 2]'$ $\mathbf{X}_{12}=[9\ 3]'$ $\mathbf{X}_{13}=[11\ 2]'$ $\mathbf{X}_1^{-}=[10\ 7/3]'$ $S_1=[[1\ -0.5][-0.5\ 1/3]]$

New: $\mathbf{X}_{21}=[8\ 3]'$ $\mathbf{X}_{22}=[9\ 4]'$ $\mathbf{X}_{23}=[7\ 2]'$ $\mathbf{X}_2^{-}=[8\ 3]'$ $S_2=[[1\ 1][1\ 1]]$.

H_0: $\boldsymbol{\mu}_1=\boldsymbol{\mu}_2$ vs. H_1: $\boldsymbol{\mu}_1\neq\boldsymbol{\mu}_2$.

$S_p=(S_1+S_2)/2=[[1\ 1/4][1/4\ 2/3]]$.

$T^2=n_1n_2(\mathbf{x}_1^{-}-\mathbf{x}_2^{-})S_p^{-1}(\mathbf{x}_1^{-}-\mathbf{x}_2^{-})/(n_1+n_2)$

$(n_1+n_2-p-1)T^2/((p(n_1+n_2-2))\sim F(p, n_1+n_2-p-1)=F(2,3)$.

WLSE:

$\mathbf{Y}_{n\times1}=Z_{n\times(r+1)}\boldsymbol{\beta}_{(r+1)\times1}+\boldsymbol{\varepsilon}$. $\boldsymbol{\varepsilon}\sim N_n(\mathbf{0}, \sigma^2V)$. $V=KK'>0$. 考虑 $\mathbf{Y}^*=K^{-1}\mathbf{Y}$.

$Z^*=K^{-1}Z$, $\boldsymbol{\varepsilon}^*=K^{-1}\boldsymbol{\varepsilon}$. 则重新表述模型为 $\mathbf{Y}^*=Z^*\boldsymbol{\beta}+\boldsymbol{\varepsilon}^*$. $\boldsymbol{\varepsilon}^*\sim N_n(\mathbf{0}, \sigma^2I)$.

于是$\boldsymbol{\beta}^*$的最小二乘估计是$(Z^*{}'Z^*)^{-1}Z^*{}'\mathbf{Y}^*=(Z'K^{-1}K^{-1}Z)^{-1}Z'K^{-1}K^{-1}\mathbf{Y}=(Z'V^{-1}Z)^{-1}Z'V^{-1}\mathbf{Y}$.

考虑到$(\mathbf{Y}^*-Z^*\boldsymbol{\beta}_W^{\wedge})'(\mathbf{Y}^*-Z^*\boldsymbol{\beta}_W^{\wedge})/(n-r-1)$是$\sigma^2$ 的无偏估计，故

$(\mathbf{Y}^*-Z^*\boldsymbol{\beta}_W^{\wedge})'(\mathbf{Y}^*-Z^*\boldsymbol{\beta}_W^{\wedge})/(n-r-1)=(\mathbf{Y}-Z\boldsymbol{\beta}_W^{\wedge})'V^{-1}(\mathbf{Y}-Z\boldsymbol{\beta}_W^{\wedge})/(n-r-1)$也对$\sigma^2$ 无偏

$$|A_{p\times p}-\lambda I|=\begin{vmatrix}1-\lambda & \rho & \cdots & \rho\\ \rho & 1-\lambda & \ddots & \vdots\\ \vdots & \ddots & 1-\lambda & \rho\\ \rho & \cdots & \rho & 1-\lambda\end{vmatrix}=\begin{vmatrix}1+(p-1)\rho-\lambda & \rho & \cdots & \rho\\ 1+(p-1)\rho & 1-\lambda & \ddots & \vdots\\ \vdots & \ddots & \ddots & \rho\\ 1+(p-1)\rho & \cdots & \rho & 1-\lambda\end{vmatrix}$$

$$=\begin{vmatrix}1+(p-1)\rho-\lambda & \rho & \cdots & \rho\\ 0 & 1-\rho-\lambda & \ddots & \vdots\\ \vdots & \ddots & \ddots & \rho\\ 0 & \cdots & 0 & 1-\rho-\lambda\end{vmatrix}$$

$$\boldsymbol{t}_1=\frac{1}{\sqrt{p}}\begin{bmatrix}1\\1\\ \vdots\\1\end{bmatrix},\boldsymbol{t}_2=\frac{1}{\sqrt{2\times1}}\begin{bmatrix}1\\-1\\ \vdots\\0\end{bmatrix},\boldsymbol{t}_3=\frac{1}{\sqrt{3\times2}}\begin{bmatrix}1\\1\\-2\\ \vdots\end{bmatrix},\dots,\boldsymbol{t}_p=\frac{1}{\sqrt{p\times(p-1)}}\begin{bmatrix}1\\1\\ \vdots\\-(p-1)\end{bmatrix}$$

$$\mu=\begin{bmatrix}10\\4\\7\end{bmatrix},\ \Sigma=\begin{bmatrix}9 & -3 & -3\\ -3 & 5 & 1\\ -3 & 1 & 5\end{bmatrix},\ \rho_{1.23}=\sqrt{\frac{\begin{bmatrix}-3 & -3\end{bmatrix}\begin{bmatrix}5 & 1\\ 1 & 5\end{bmatrix}^{-1}\begin{bmatrix}-3\\-3\end{bmatrix}}{9}}=\frac{\sqrt{3}}{3}$$

$$\Sigma_{12.3}=\begin{bmatrix}9 & -3\\ -3 & 5\end{bmatrix}-\begin{bmatrix}-3\\ 1\end{bmatrix}[5]^{-1}\begin{bmatrix}-3 & 1\end{bmatrix}=\begin{bmatrix}36/5 & -12/5\\ -12/5 & 24/5\end{bmatrix}$$

$$\rho_{12.3}=\frac{-12/5}{\sqrt{24\times36}/5}=-\frac{\sqrt{6}}{6}$$

深圳经济：$\lambda_1^{\wedge}=2.25=9/4$, $\lambda_2^{\wedge}=1.96=49/25$, $\lambda_3^{\wedge}=0.16=4/25$

$\mathbf{e}_1^{\wedge}=[1/2\ 1/2\ 1/\sqrt{2}]'$ $\mathbf{e}_2^{\wedge}=[1/2\ 1/2\ -1/\sqrt{2}]'$ $\mathbf{e}_3^{\wedge}=[1/\sqrt{2}\ -1/\sqrt{2}\ 0]'$

1. 重建样本协方差矩阵

$S=[\mathbf{e}_1^{\wedge}\ \mathbf{e}_2^{\wedge}\ \mathbf{e}_3^{\wedge}]\mathrm{diag}[\lambda_1^{\wedge}\ \lambda_2^{\wedge}\ \lambda_3^{\wedge}][\mathbf{e}_1^{\wedge}\ \mathbf{e}_2^{\wedge}\ \mathbf{e}_3^{\wedge}]'=\sum_{i=1}^{3}\lambda_i^{\wedge}\mathbf{e}_i^{\wedge}\mathbf{e}_i^{\wedge}{}'$.

=[[453/400　389/400　29√2/400]
　[389/400　453/400　29√2/400]
　[29√2/400　29√2/400　421/200]

=[[1.1325 0.9725 0.1025] [0.9725 1.1325 0.1025] [0.0125 0.1025 2.105]]

2. 用主成分法求载荷矩阵、共性方差、特殊方差、残差矩阵、被第一因子解释的方差比例(m=1)

$A^{\wedge}=[\mathbf{e}_1^{\wedge}\sqrt{\lambda_1^{\wedge}}]=[3/4\ 3/4\ 3\sqrt{2}/4]'$, $h_1^2=h_2^2=9/16$, $h_3^2=9/8$.

A^A^'=[[9/16　9/16　9√2/16]
　[9/16　9/16　9√2/16]
　[9√2/16　9√2/16　9/8]]

特殊方差$\sigma_1^2=S_{11}-h_1^2=57/100$, $\sigma_2^2=57/100$, $\sigma_3^2=49/50$.

S-A^A^'-D=[[0　164/400　-196√2/400]
　[164/400　0　-196√2/400]
　[-196√2/400 -196√2/400　0　]]

方差比例=2.25/(2.25+1.96+0.16)=0.5149

3. (m=2)

$A^{\wedge}=[\mathbf{e}_1^{\wedge}\sqrt{\lambda_1^{\wedge}}\ \mathbf{e}_2^{\wedge}\sqrt{\lambda_2^{\wedge}}]=[[3/4\ 3/4\ 3\sqrt{2}/4]\ [7/10\ 7/10\ -7\sqrt{2}/10]]'$

$h_1^2=h_2^2=421/400$, $h_3^2=421/200$

A^A^'=[[421/400　421/400　29√2/400]
　[421/400　421/400　29√2/400]
　[29√2/400　29√2/400　421/200]]

特殊方差$\sigma_1^2=2/25$, $\sigma_2^2=2/25$, $\sigma_3^2=0$.

S-A^A^'-D=[[0　-2/25　0]
　[-2/25　0　0]
　[0　0　0]]

方差比例=(2.25+1.96)/(2.25+1.96+0.16)=0.9634.