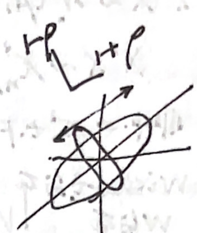
$X \sim \mathcal{N}(\mu, \Sigma)$ ← (3)种定义.

△ $\phi(t) = e^{it^T\mu - \frac{1}{2}t^T\Sigma t}$ ($\varphi(t) = e^{it\mu - \frac{1}{2}t^2\sigma^2}$)

△ ($X = AU + \mu$) $AA^T = \Sigma$ ($A = Q\sqrt{\Lambda} = \sqrt{\Sigma}$)



$U \overset{iid}{\sim} \mathcal{N}(0, \sigma^2 I)$.

($A\Sigma B^T = O$ 时 AX 与 BX id.

二次型 $X'AX$ 与 BX id. $\Leftrightarrow BA = 0$

$X'AX$ 与 $X'BX$ id. $\Leftrightarrow AB = BA = 0$.)

marginal $X_{r\times r} \sim \mathcal{N}(\mu_r, \Sigma_{r\times r})$

$f(x) = \frac{1}{(\sqrt{2\pi})^p\sqrt{|\Sigma|}} e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}$ ($U \to X$ 证)

(Σ 正定(非退化) 或 A, Σ 非奇异)

conditional

$X = \begin{pmatrix} X^1 \\ X^2 \end{pmatrix}_{p-r}^{r} \sim \mathcal{N}_p(\mu, \Sigma)$. ($\frac{\sigma_{ij\cdot r+1\cdots p}}{\sqrt{\sigma_{ii}\cdots\sigma_{jj}\cdots}}$ 偏相关系数)

(消元) $(X^1|X^2) \sim \mathcal{N}_r(\mu_{1|2}, \Sigma_{11|2})$ ↗ $\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ 条件协方差阵 $\Sigma_{11\cdot 2}$

($\Sigma_{12} = \Sigma_{21}^T$)

$\mu_{1|2} = \mu^1 + \Sigma_{12}\Sigma_{22}^{-1}(X^2 - \mu^2)$

X^1 对 X^2 回归 回归系数.

$E(X^1|X^2)$ 为对 X^1 最佳 (Varmin) 预测.

$\begin{pmatrix} I & \\ -\Sigma_{yx}\Sigma_{xx}^{-1} & I \end{pmatrix} \begin{pmatrix} \Sigma_{yx} & 0 \\ 0 & \sigma_{yy} \end{pmatrix} \begin{pmatrix} I & -\Sigma_{xx}^{-1}\Sigma_{xy} \\ & I \end{pmatrix}$ $\Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}$.

$Z = \begin{pmatrix} X \\ Y \end{pmatrix}_1^p \sim \mathcal{N}_{p+1}\left(\begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \sigma_{yy} \end{pmatrix}\right)$ $\mu_y - \Sigma_{yx}\Sigma_{xx}^{-1}\mu_x$.

$Y \leftarrow X$ 得 $\Sigma_{yx}\Sigma_{xx}^{-1}$ 回归.

$R = Corr(Y, \Sigma_{yx}\Sigma_{xx}^{-1}X) = \sqrt{\frac{\Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}}{\sigma_{yy}}}$. 全(复)相关系数 $\begin{pmatrix} \Sigma_{yx}\Sigma_{xx}^{-1} \\ I \end{pmatrix}$ $\begin{pmatrix} \Sigma_{yx} & \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy} \\ \Sigma_{yx} & \sigma_{yy} \end{pmatrix}$ $\begin{pmatrix} \Sigma_{yy}^{-1}\Sigma_{xy} \\ I \end{pmatrix}$

$Vec(\mu 1^T) = 1 \otimes \mu$

矩阵正态分布 $X \sim \mathcal{N}_{n\times p}(M, I_n \otimes \Sigma)$, $M = 1\mu^T$. $\Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}$.

$\begin{pmatrix} X_{11} \cdots X_{1p} \\ \vdots \\ X_{n1} \cdots X_{np} \end{pmatrix}_{n\times p}$ ($Vec(X^T) \sim \mathcal{N}(Vec(M^T), I_n \otimes \Sigma)$.) $\Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{yy}$.

$Z = AXB^T + D$, $Z \sim \mathcal{N}_{k\times q}(AMB^T, (AA^T) \otimes (B\Sigma B^T))$

estimation: $\bar{X} = \frac{1}{n}X^T 1_n$ $X_{(i)}$ 为第 i 个样品 ($i = 1, \cdots, n$). △ 取迹 (交换顺序, 算 E…) 的方法.

$S = \frac{1}{n-1}\sum_i (X_{(i)} - \bar{X})(X_{(i)} - \bar{X})^T = \frac{X^TX - n\bar{X}\bar{X}^T}{n-1} = \frac{X'(I_n - \frac{1}{n}1_n1_n')X}{n-1}$.

MLE $\hat{\mu} = \bar{X}$, $\hat{\Sigma} = \frac{1}{n}A$. A. (求导/不等式)

(Q) 中心极限定理 $A \overset{d}{\sim} \sum_{i=1}^{n-1} Z_iZ_i'$. $Z \overset{iid}{\sim} \mathcal{N}_p(0, \Sigma)$ —— $Cov(Z_i, Z_j) = E(X^T q_i q_j^T X) = \delta_{ij}\Sigma$.

$\bar{X} \sim \mathcal{N}_p(\mu, \frac{1}{n}\Sigma)$, $\bar{X}$ id. A.

$P(A > 0) = 1 \Leftrightarrow n > p$. ← $A = BB^T$, $B = (Z_1 \cdots Z_{n-1})_{p\times(n-1)}$.

无偏, 有效, 相合, 充分…

$I(\mu) = \int \frac{(\frac{\partial f}{\partial \mu})^2}{f} dx = \int f dx \cdot (\Sigma^{-1}(x-\mu))^2$

↓ $\Sigma^{-1}(x-\mu)(x-\mu)^T\Sigma^{-T}$

☆ 各分布形状熟悉: 峰, 尾…

$(n\Sigma^{-1})^{-1} = \frac{\Sigma}{n}$ ✓ $= \Sigma^{-1}$

△ $I(\Sigma)$?

⟶ $E((x-\mu)(x-\mu)^T(x-\mu)(x-\mu)^T)$

test: 1. $X \sim N(\mu, I_n)$. $\delta = \mu'\mu$.

$X'X \sim \chi^2_{n,\delta}$.

非中心 χ^2、t、F分布 → 第二类错误 一般

$X \sim N(0, \sigma^2 I)$, A sym, $r(A)=r$.

$X'AX \sim \sigma^2 \chi^2_r \iff A^2 = A$. (即 $Q\begin{pmatrix}1 & & \\ & 1 & \\ & & 0\end{pmatrix}Q^T$)

$X \sim N(\mu, \Sigma)$, $\Sigma > 0$.

$X'\Sigma^{-1}X \sim \chi^2_{p,\delta}$, $\delta = \mu'\Sigma^{-1}\mu$.

$(X'-\mu')A(X-\mu) \sim \chi^2_r \iff A\Sigma A = A$.

$(X-\mu)'A(X-\mu)$ 与 $(X-\mu)'B(X-\mu)$ id. $\iff A\Sigma B = 0$

2. Wishart.

W分布 $W = \underset{p\times n}{X'}\underset{n\times p}{X}$ $p=1$时 $W \sim \sigma^2\chi^2_n$

$W \sim W_p(n, \Sigma)$ $= \sum_{i=1}^n \underset{p\times p}{X_iX_i'}$ $(W = \sum_i X_i^2)$

$(X \overset{iid}{\sim} N_p(0, \Sigma))$

$X \sim N_{n\times p}$, $X'AX$ 与 $X'BX$ id. (A, B idempotent.) $\iff AB = 0$

$W_n(\sigma^2) = \sigma^2\chi^2_n$

非中心W: $X \sim N_p(\mu, \Sigma)$. $M = 1\mu^T$ $X_i \sim N_p(\mu_i, \Sigma)$. $\Delta = M'M$. $M = \begin{pmatrix}\mu_1' \\ \vdots \\ \mu_n'\end{pmatrix}$

$W_p(n, \Sigma, \Delta)$, $\Delta = M'M = n\mu\mu'$.

W与χ^2: X_i从一元变为p元.

$X \sim N_p(\mu, \Sigma)$. $A \sim W_p(n-1, \Sigma)$. $(= \sum_{i=1}^{n-1} Z_iZ_i')$ $W_{n-1}(\Sigma)$

$X \sim N_{n\times p}(M, I_n \otimes \Sigma)$. A sym, $r(A)=r$.

$X'AX \sim W_r(\Sigma, \Delta)$, $\Delta = M'AM \iff A^2 = A$.

$(\sqrt{\Lambda}Q'X = Y)$.

n可加. $CWC' \sim W_n(C\Sigma C')$. $E(W) = n\Sigma$.

对角元边际. $W_{jj} \sim \sigma_{jj}\chi^2_n$ $aW \sim W_n(a\Sigma)$

$v'Wv \sim \chi^2_n \cdot \underbrace{(v'\Sigma v)}_{\sigma_v^2}$

特征func:

$\phi_{\chi^2_n}(t) = (1-2it)^{-\frac{n}{2}}$.

$\phi_{W_n}(t) = E(e^{i\,tr(Wt)}) = |I - 2it\Sigma|^{-\frac{n}{2}}$

$f(W) = \dfrac{|W|^{\frac{n-p-1}{2}} e^{-tr(\frac{\Sigma^{-1}W}{2})}}{2^{\frac{np}{2}}|\Sigma|^{\frac{n}{2}}\Gamma_p(\frac{n}{2})}$.

3. Hotelling

T^2分布 $T^2 = nX'W^{-1}X$ (分出Σ^{-1}就即可).

与Σ无关 (对非退化变换不变)

$T^2 \sim T^2_n(p)$ $(X \sim N_p(0, \Sigma)$, $W \sim W_n(\Sigma)$, $\Sigma > 0$, $n > p)$. 否则 sin.

非中心T^2: $X \sim N_p(\mu, \Sigma)$. $T^2_n(p, \mu)$.

$X_i \sim N_p(\mu, \Sigma)$

$T^2 = (n-1)\cdot n\cdot(\bar{X}-\mu)'A^{-1}(\bar{X}-\mu) \sim T^2_{n-1}(p) = \frac{p}{n-p}(n-1)F_{p,n-p}$

4. Wilks.

Λ分布. 广义方差 $|\Sigma|$, $|\frac{A}{n-1}|$.

★ $\frac{n-p+1}{np}T^2_n(p) \sim F_{p,n-p+1}$.

$n \to \infty$时 $T^2_n(p) \sim \chi^2_p$

$A_1 \sim W_{n_1}(\Sigma)$, $A_2 \sim W_{n_2}(\Sigma)$

$\Lambda = \frac{|A_1|}{|A_1 + A_2|} \sim \Lambda_{n_1,n_2}(p) = \frac{|W_{n_1}(\Sigma)|}{|W_{n_1+n_2}(\Sigma)|}$

$p=1$, 一元 $t^2 \sim F_{1,n}$.

$T^2 =$

$\frac{n-p+1}{p}X'W^{-1}X = \dfrac{\frac{X'\Sigma^{-1}X}{p}}{\frac{X'W^{-1}X}{X'\Sigma^{-1}X}\Big/(n-p+1)} \sim F_{p,n-p+1}$.

$X'\Sigma^{-1}X$ — $\chi^2_p(\delta=0)$ ✓ (X, 2)

$\frac{X'\Sigma^{-1}X}{X'W^{-1}X}$ — χ^2_{n-p+1} (Δ?)

proof:

$\delta \to \Sigma$.

$T^2 = (\bar{X}-\mu)'\sqrt{n}\Sigma^{-\frac{1}{2}}\cdot\sqrt{n}\Sigma^{-\frac{1}{2}}(\bar{X}-\mu)$

$= Y'Y \sim \chi^2_p$.

$\Lambda_{n_1,n_2}(1) = \beta(\frac{n_1}{2}, \frac{n_2}{2})$

$\Lambda_{n,1}(p) = \frac{1}{1+\frac{1}{n}T^2_n(p)}$ $(T^2_n(p) = n\cdot\frac{1-\Lambda_{n,1}(p)}{\Lambda_{n,1}(p)})$

$(|W| = |W_1 + X_{n+1}X_{n+1}'| = \begin{vmatrix} W_1 & -X \\ X' & 1\end{vmatrix}$

$= |W_1|(1 + X_{n+1}'W_1^{-1}X_{n+1}))$

$n_1 \to \infty$时 $-r\ln\Lambda \sim \chi^2_{pn_2}$

$r = n_1 - \frac{1}{2}(p - n_2 + 1) \approx n_1$ — 方差分析里可与λ对应. $\sim\chi^2$

(F分布亦可: $\frac{\chi^2_p}{(\frac{\chi^2_n}{n}) \to 1}$)

似然里 似然比λ ~ χ^2

$\Lambda = B_1 \cdots B_p$ $(\prod_{i=1}^p B_i)$. $B_i \sim \beta(\frac{n_1-p+i}{2}, \frac{n_2}{2})$

$(n_2 < p)$ $\Lambda_{n_1,n_2}(p) = \Lambda_{p,n_1+n_2-p}(n_2)$

Δ

1. Weistein-Aronszajn Iden.

$\det(I_m + AB) = \det(I_n + BA)$

推论: $|A + c\vec{u}\vec{u}'|$ $= |A|(1 + c\vec{u}'A^{-1}\vec{u})$

$n_2 = 1/2$时 Λ可化为F.

$p=1$时 Λ即βeta.

2. λ, $A > 0$. then. $x'Ax \le \lambda \iff xx' \le \lambda A^{-1}$.

均等阔 LRT.

one-sample test:

Σ_0 ✓ $\sqrt{n}(\bar{X}-\mu) \sim N_p(0, \Sigma_0)$.

H_0时 $T_0^2 = n(\bar{X}-\mu_0)'\Sigma_0^{-1}(\bar{X}-\mu_0) \sim \chi^2_p$. $R=\{T_0^2 > \chi^2_p(\alpha)\}$ / P值

(β: $T_0^2 \sim \chi^2_p(\delta)$, $\delta = n(\mu_1-\mu_0)'\Sigma_0^{-1}(\mu_1-\mu_0)$.
$\beta = P\{T_0^2 \le \chi^2_p(\alpha)\}$ | H_1)

Σ ✗ $A \sim W_{n-1}(\Sigma)$.

H_0时 $T^2 = n(n-1)(\bar{X}-\mu_0)'A^{-1}(\bar{X}-\mu_0) \sim T^2_{n-1}(p)$.

$F = \frac{n-p}{(n-1)p}T^2 \sim F_{p,n-p}$.

LR 检: $H_0: \theta\in\Theta_0$ $H_1: \theta\notin\Theta_0$ $\Theta_0 \subset \Theta$

$$\lambda = \frac{\max_{\theta\in\Theta_0} L(X;\theta)}{\max_{\theta\in\Theta} L(X;\theta)}$$

$P(\lambda(X_1\cdots X_n) < \lambda_\alpha) = \alpha$.
λ分布? 近似: $n\to\infty$时 $-2\ln\lambda \sim \chi^2_f$ 其中 $f = \dim\Theta - \dim\Theta_0$.

N-P lemma.
($\to$ UMVUE)

验证 LRB: $\max_{\mu,\Sigma>0} L(\mu,\Sigma) = (2\pi)^{-\frac{np}{2}} e^{-\frac{np}{2}} \left|\frac{1}{n}A\right|^{-\frac{n}{2}}$ 则 $\lambda = \frac{|A_0|^{-\frac{n}{2}}}{|A|^{-\frac{n}{2}}} = \left(\frac{|A|}{|A_0|}\right)^{\frac{n}{2}} = \left(\frac{1}{1+\frac{1}{n-1}T^2}\right)^{\frac{n}{2}}$.

$A_0 = A + n(\bar{X}-\mu_0)(\bar{X}-\mu_0)'$.
$|A_0| = |A|(1+n(\bar{X}-\mu_0)'A^{-1}(\bar{X}-\mu_0))$.

$R = \{T^2 > T^2(\alpha)\}$
$= \{F_{p,n-p} > F_{p,n-p}(\alpha)\}$

$\sim \frac{T^2_{n-1}(p)}{n-1}$ 且: $-2\ln\lambda \approx -n\ln\left(1-\frac{\chi^2_p}{n-1}\right) \approx \chi^2_p$ ✓. QED.

confidence interval:

$T^2 = n(\bar{X}-\mu)'S^{-1}(\bar{X}-\mu) \le \frac{(n-1)p}{n-p}F_{p,n-p}(\alpha)$ $\frac{n-p}{p}\frac{T^2}{n-1}$

μ位于$\bar{X}$中心的椭球(p维)内.
(置信域)

μ的线性组合 $Z = a'X \sim N(a'\mu, a'\Sigma a)$

$t = \frac{\sqrt{n}(a'\bar{X}-a'\mu)}{\sqrt{a'Sa}}$. $a'\mu \in a'\bar{X} \pm t_{n-1}(\frac{\alpha}{2})\frac{\sqrt{a'Sa}}{\sqrt{n}}$ (单)置信区间.

a取e_i得μ_i —— 联立! ✗ $<1-\alpha$.

推广柯西ineq. $S>0$, Scheffe:

$(a'b)^2 \le (a'Sa)(b'S^{-1}b)$. $\longrightarrow$ $t^2 = n\frac{(a'(\bar{X}-\mu))^2}{a'Sa} \le t^2_{\frac{\alpha}{2}}$ 换为 $\max_a t^2 = n(\bar{X}-\mu)'S^{-1}(\bar{X}-\mu) = T^2 \sim$ (F).

(联立)置信区间. $\forall a$, $a'\mu \in a'\bar{X} \pm \sqrt{\frac{(n-1)p}{n(n-p)}F_{p,n-p}(\alpha)a'Sa}$ 此时$a = c\cdot S^{-1}(\bar{X}-\mu)$ ✓ T^2区间

$\mu_i \in \bar{X}_i \pm \sqrt{\frac{(n-1)p}{n-p}F(\alpha)}\cdot\sqrt{\frac{S_{ii}}{n}}$.

multi-sample test:

(等Σ)双正态 $T^2 = \frac{nm}{n+m}(\bar{X}-\bar{Y})'\left(\frac{A_X+A_Y}{n+m-2}\right)^{-1}(\bar{X}-\bar{Y})$ $\overset{H_0}{\sim} T^2_{n+m-2}(p) = \frac{(n+m-2)p}{n+m-p-1}F_{p,n+m-p-1}$.

$T(\alpha) \leftrightarrow t(\frac{\alpha}{2})$ ✓

△回归直线的范围:
(单) $T^2 = (\hat\beta-\beta)'(X'X)^{-1}\hat\sigma^{-2}(\hat\beta-\beta)$ $n-k-1, k+1$ β_i需小t改大T.

$m(x) \in \hat{Y}(x) \pm \hat\sigma\sqrt{\frac{1}{n}+\frac{(x-\bar{x})^2}{S_x}}$.

$\left(\max \frac{(x'(\hat\beta-\beta))^2}{\hat\sigma^2 x'(X'X)^{-1}x}\right.$ $T(\alpha)$

有 $\alpha'\beta \in \alpha'\hat\beta \pm T(\alpha)\cdot\hat\sigma\sqrt{\alpha'(X'X)^{-1}\alpha}$)

多总体均值检验:

(ANOVA) 多元阵 $T = A + B$?证明: (Σ一致下) 组内 组间

B id. A 且 $W_p(k-1,\Sigma)$ (Wishart可加性)

$\Lambda = \frac{|A|}{|T|} \sim \Lambda_{n-k,k-1}(p)$ $\sum_i A_i \sim W_{n-k}(p,\Sigma)$

($\downarrow$LRE: $\frac{\max\cdots(\mu)}{\max\cdots(\mu_i)}$ ← $|T|$ 即 $R = \left(\frac{|A|}{|T|}\right)^{\frac{n}{2}}$. QED.)

$A = \sum_i A_i$, $|A|$

(k=2时与T^2检验一致)

LR渐近 χ^2 proof:　1. $H_0: \theta=\theta_0$, $H_1: \theta\neq\theta_0$

GLLR loglike ← $\Lambda(x) = -2\ln\lambda = 2(l(x;\hat\theta) - l(x;\theta_0))$

Taylor $l(\theta_0) = l(\hat\theta) + (\hat\theta-\theta_0)l'(\hat\theta) + \frac{1}{2}(\hat\theta-\theta_0)^2 l''(\hat\theta) + \cdots$

$\Lambda(x) = -(\hat\theta-\theta_0)^2 l''(\hat\theta) = +(\hat\theta-\theta_0)^2 I(\theta) \cdot \frac{-l''(\hat\theta)}{I(\theta)}$

MLE is 渐相合: $(\hat\theta-\theta_0)\sqrt{I(\theta)} \xrightarrow{P} N(0,1)$.

LLN (大数定律): $\frac{-l''(\theta)}{I(\theta)} \xrightarrow{P} 1$.

(Slutsky Th.) $\Lambda(x) \xrightarrow{P} \chi^2_1$

θ 多元时有 p 个 χ^2_1 相加. ← ? $(\hat\theta_1-\theta_{10})(\hat\theta_2-\theta_{20})$ 独立?

Wilks' Th.

asymptotic dis. of the likelihood ratio statistic:

Ω_0 (i) θ_0 (ii) $A(\theta-a)=0$ (iii) $g_i(\theta)=0$, $i=1,\dots,p$.

$\theta=(\theta_1,\dots,\theta_k)$ ↓d.　$p\times k$ mt. ↓d　↓d.

Then $2\log\frac{L_n(\hat\theta_n)}{L_n(\hat\theta_{n,0})} \xrightarrow{d} \chi^2_k$　χ^2_p　χ^2_p.

$p = \dim\Omega - \dim\Omega_0$.

(if there are some $\theta_1\cdots\theta_r$ remains to be estimated → the same. Ω_0↑ dim.)

χ^2 (拟合优度)检验:　(Q方法见(中))　$H_0: \pi_i$, $\dim(H_0)=0$.

$\hat p_i$ 即 MLE. $=\frac{Y_i}{n}$

$\Lambda = 2n\sum_i \hat p_i \cdot \ln\frac{\hat p_i}{\pi_i} \xrightarrow{Taylor} 2n\cdot\frac{\sum(\frac{Y_i}{n}-\pi_i)^2}{2\pi_i} = \sum_i \frac{(Y_i-n\pi_i)^2}{n\pi_i}$

($n\to\infty$ 时 $\hat p_i \xrightarrow{P} \pi_i$) MLE.

与 Wilks Th. 一致.

$(\lambda\in[\,])$

H_0, φ: N-P引理, 两类错误, 一致最优

△ 似然比检验(LRT)与传统参数检验(NP理论)的一致性?

$-\frac{1}{2}(\bar x-\mu)'\Sigma_0^{-1}(\bar x-\mu)$　平移 χ^2_p.

$(H_0: \mu=0, \hat\mu=\bar X)$

$\sum \ln R(x) = \frac{1}{2}tr\Sigma_0^{-1}(XX'-A) = -\frac{1}{2}\sum_{i=1}^{n}(X_i-\mu)'\Sigma^{-1}(X_i-\mu)$

$L(\mu,\Sigma) = \frac{1}{(2\pi)^{\frac{np}{2}}}\cdot\frac{1}{|\Sigma|^{\frac{n}{2}}} e^{\cdots}$

cov test:

$\ln L(\mu,\Sigma) = -\frac{np}{2}\ln 2\pi - \frac{n}{2}\ln|\Sigma| - \frac{1}{2}tr(\Sigma^{-1}\sum_i(X_i-\mu)(X_i-\mu)')$

$= A + \sum_i(\bar X-\mu)(\bar X-\mu)'$

1. $H_0: \Sigma = I_p$ 球性检验.　$\lambda = \dots$

$L(\hat\mu,\hat\Sigma) = L(\bar X, \frac{1}{n}A) = (\frac{n}{2\pi e})^{\frac{np}{2}}|A|^{-\frac{n}{2}}$

2. $H_0: \Sigma_0$　$Y = \Sigma_0^{-\frac{1}{2}}X$ 变为1,

$A_Y = \Sigma_0^{-\frac{1}{2}}A\Sigma_0^{-\frac{1}{2}}$　$\stackrel{d}{\sim}\chi^2_{(\frac{p(p+1)}{2})}$

3. $H_0 = \sigma^2\Sigma_0$

σ^2 待估. $\hat\sigma^2 = \frac{1}{np}tr(\Sigma_0^{-1}A)$　$\stackrel{d}{\sim}\chi^2_{(\frac{p(p+1)}{2}-1)}$.

多总体方差检验　$\sim\chi^2_{\frac{p(p+1)(k-1)}{2}}$

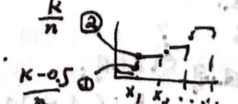
$\frac{\max L(\mu^1\cdots\mu^k,\Sigma)}{\max L(\mu^1\cdots\mu^k,\Sigma^1,\dots,\Sigma^k)}$

independence test:　$H_0: \begin{pmatrix}\Sigma_{11} & 0 & 0 & 0\\ 0 & \ddots & & \vdots\\ 0 & & \ddots & 0\\ 0 & \cdots & 0 & \Sigma_{kk}\end{pmatrix}$, $\lambda = \frac{\prod|A_{ii}|^{-\frac{n}{2}}}{|A|^{-\frac{n}{2}}}$

normality test: χ^2 test

↳(非参数检验)　K-S test　$D = \sup_x|F_n(x) - F_0(x)|$.

edf: $X_1 \le \dots \le X_n$



$F_n(x) = \frac{k}{n}$　$(X_k \le x \le X_{k+1})$

$F_n(x) = 0$ $(x\le X_1)$ / 1 $(x\ge X_n)$

拟合优度 { χ^2, KS }　偏峰 (JB test)

计数stat. 符号　Wilks. D test

秩stat. Wilcoxon　Q-Q / P-P ← χ^2图

符号秩stat. 秩和　36原则

Cochran Q　(A^2. W^2 统计量)

游程 …

P→1元　生成多 (X. (λ大的)

(等概率椭圆)

U T 检验 F

相关性检验等 pre:

方差分析 (正态性、方差齐性.)

Bayes. →

(极限~) 大样本情况 (制造总体)

参数检验: 基于对总体分布的假设, 对参数进行统计推断.

① 分布不定、非正态 ②(等级)分类数据

对某性质(假设)检验.

[分布特性]

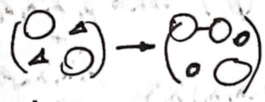
① 正态等分布

② 样本量化且独立

③ 均值方差存在、方差相等…

挑选: 分布; 样本大小; 等方差; 偏度 (已/未知); 待检量

△ 回归↔条件~

$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ 

X_1 对 X_2（X_1 被 X_2 回归掉）

（后剩一点多余）

$R^2 \sim F$?

复相关系数 $\begin{pmatrix} X \\ Y \end{pmatrix}$ $r^2(Y,\hat{Y}) = \frac{\Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}}{\sigma_{yy}^2}$ ⟵ $R^2 = \frac{SSR}{SST}$

偏相关系数（$X_2 \to X_1$ 后剩余协方差）$r_{ij\cdot(\dots)}$.

F 或开根 t.

$F = \frac{MSR}{MSE}$

△偏回归显著性（F）检验：$r(\dots)^2$，$H_0: \beta_1 = \dots = \beta_r = 0$

（偏 R^2 = $\frac{\dots}{k-r}$），$r=1$ 时等价：

$SSR_{\dots} = R_2 - R_1 = \hat{\beta}'(X^TX)\hat{\beta}$

$\frac{SSR}{SST} \sim \chi^2_r$

新 xxx

原 $\hat{\beta}_i \sum (X_i-\bar{X}_i)Y_X \neq \frac{\hat{\beta}_i^2}{\sigma_{ii}}$

$V(\hat\beta) = (X^TX)^{-1}\sigma^2$ 元素

$\hat{\beta}_i^2$ · 消化新元 ✓（把 i 的作用回归给其他，剩 error）

$a_{11} - v'A_{2-m}^{-1}v = \frac{1}{a_{11}^{-1}}$

$a_{11}^{-1} = \frac{|A_{2-m}|}{|A|}$

最优回归方程（变量选择）：逐步回归法（STEPWISE）

全子集法（$2^m - 1$）

（准则：R^2、R^2_{adj}、C_p…）

MIN/MAXR

增量法

$\begin{pmatrix} X \\ Y \end{pmatrix}$ $\begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \sigma_{yy} \end{pmatrix}$

$\hat{R}^2$ 累积 ~~中心化~~

X. $Y - \hat{Y}$

$\begin{pmatrix} \Sigma_{xx} & 0 \\ 0 & \sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy} \end{pmatrix}$

否则有 $\mu_{1.2}$.

（$\mu_x = 0, \mu_y = \bar{y}$）由 $\mu_{1.2} = \bar{y}$.

$\begin{pmatrix} I & \\ -\Sigma_{xy}\Sigma_{xx}^{-1} & I \end{pmatrix}$

准则：

1. $\sum_i \hat{\varepsilon}_i^2$（$= SSE$）↓ / R（$= \sqrt{\frac{SSR}{SST}}$）↑

2. $\hat{\sigma}$（$= \sqrt{\frac{\sum \hat{\varepsilon}_i^2}{n-m-1}}$）↓

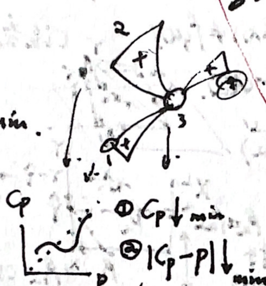
3. m ↓

pre: 固定 k, SSE↓ 时子集 $A(k)$.

① $S^2_{(k)} = s^2_{(A)}$（$\hat{\sigma}^2$）$= \frac{SSE(k)}{n-k-1}$ ↓min.

② 选模型。 $p = k+1$.

全（回归）模型 $C_p = \frac{SSE(k)}{\hat{\sigma}^2} + 2p - n$.



① C_p ↓min

② $|C_p - p|$ ↓min

选√且 n 大时 $E(C_p) = p$

③ $\hat{R}^2 = 1 - (1-R^2)\frac{n-1}{n-k-1}$

$= 1 - \frac{MSE(\hat{\sigma}^2)}{MST}$

Overfitting.

④ $J_p = \hat{\sigma}^2(n+k+1)$

$S_p = \frac{\hat{\sigma}^2}{n-k-2}$

Akaike info criterion

Schwartz Bayesian

⑤ AIC、SBC、BIC. ↓min

reg 中 $L = (2\pi)^{-\frac{n}{2}} \cdot \left(\frac{SSE}{n}\right)^{-\frac{n}{2}} \cdot e^{-\frac{n}{2}}$ ⟵ 熵、信息量

（似然 func）

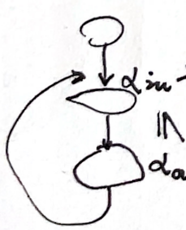
PRESS.

$AIC = 2k - 2\ln L = 2k + n\ln\frac{SSE}{n}$.

k 为参数个数（算 β_0 的）. ↓ MLE 的 $\hat{\sigma}^2$

? SBC：$2 \to \ln n$. $\ln n\, k + n\ln\frac{SSE}{n}$

BIC（还有其他解释（即回归））



α_{in} 量难以/不等 min 剔

$\Delta R^2_{SS\,i}$ { max λ

α_{out}.

$R^2_{(k)} = \frac{SSR(k)}{SST}$. $R^2_{k-r} = \frac{SSR_{k-r}}{SST}$

$R^2_k - R^2_{k-r} =$

$\hat{Y} = \Sigma_{yx}\Sigma_{xx}^{-1}X$ $\begin{pmatrix} \hat{Y} \\ Y \end{pmatrix}$

$\begin{pmatrix} \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy} & \\ & \sigma_{yy} \end{pmatrix}$ SSR. SST.

$\frac{SSR}{SST}$ $\hat{Y} =$

$(y - \bar{y})$ $\hat{Y} - \bar{y}$ Y

K-S test:

$D_n = \sup_x |F_n(x) - F_0(x)|$

（$D = \max(|F_n(X_i) - F_0(X_i)|,$
$|F_n(X_i) - F_0(X_{i-1})|)$

$n \to \infty$, $\sqrt{n}D_n \longrightarrow \sup_t |B(F(t))|$ Kolmogorov 分布.

Brownian bridge

$B(t) = W(t) - \frac{t}{T}W(T)$ $t \in [0,T]$

$= (W(t) \mid W(T) = 0)$

$W(0) = 0$

$W(t) \sim N(0, \sigma^2 t)$ Wiener process

$f_{W(t)}(x) = \frac{1}{\sqrt{2\pi t}}e^{-\frac{x^2}{2t}}$

△证明：残差的消元法求偏的正确性 $P(\sup_t |B(t)| < x) = \frac{\sqrt{2\pi}}{x}\sum_{k}^{\infty} e^{-\frac{(2k-1)^2\pi^2}{8x^2}}$

$\beta_1 = \begin{pmatrix} (X_1'X_1)^{-1}X_1'Y \\ 0 \end{pmatrix}$. $\beta_{12} = \left(\begin{pmatrix} X_1' \\ X_2' \end{pmatrix}(X_1\ X_2)\right)^{-1}\begin{pmatrix} X_1' \\ X_2' \end{pmatrix}Y$.

$\begin{pmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{pmatrix}^{-1}$

$\beta_{12}'\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\beta_{12}$ I $\beta_{22}^2(\Sigma_{22} - \Sigma_{12}\Sigma_{11}^{-1}\Sigma_{12})$

$\beta_1'(\ )\beta_1$

√（先后）

$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$: $\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$

$\begin{pmatrix} X_1 \\ X_2 - \Sigma_{12}\Sigma_{11}^{-1}X_1 \end{pmatrix}$: $\begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \end{pmatrix}$.

△ C_p Criterion：Mallow's C_p =

子模型 (X_1) 回归后剩

$$E(HY-X\beta)=(H-I)X\beta=(H-I)(X_1\beta_1+X_2\beta_2)$$

$$HY = X_1 b_1$$

$$=-(I-H)X_2\beta_2 = -X_2^{\perp}\beta_2$$

平方误差 $V(HY-X\beta)=V(HY)=tr H\cdot\sigma^2$

总SSE $=p\sigma^2+\|X_2^{\perp}\beta_2\|^2=m.$

$p=k+1$ ↑ $X_1(x_1,\dots,x_k)$

准则应让子模型的SSE尽量小（较好的子）（与正交化（准确说 (X_1) 单独带有1即 β_0，偏回归 (X_2) 时无需带1.）

（注意 $n\to\infty$ 时且 β_2 真为0，SSE仍有 $\to p\sigma^2$ 的方差。）（大数→E）

（X_1 为一元时）等同，但 (H) 投影是更广的理解）（hat mat：用它 (X) 去估计什么 (Y) 就乘什么 (HY) 得到回归估计值）

用估计 $\hat m = p\hat\sigma^2+\|X_2^{\perp}\hat\beta_2\|^2$ ？$E(\hat m)=m$ 吗？

$\hat\sigma^2\to\sigma^2$ ✓. $\hat\beta_2\to\beta_2$ ✓ △ $E(x)=\mu$ 时通常 $E f(x)\neq f(\mu)$!

$\hat\beta_2^2\to\beta_2^2$ ✗. $E(X'AX)=\mu'A\mu+tr(A\Sigma)$.

考察 $E(\|X_2^{\perp}\hat\beta_2\|^2)=\beta_2'X_2^{\perp\prime}X_2^{\perp}\beta_2+tr(X_2^{\perp}V(\hat\beta_2)X_2^{\perp\prime})$

$$\hat\beta\sim N(\beta,\sigma^2(X'X)^{-1})\quad=\|X_2^{\perp}\beta_2\|^2+tr(I_{m-k})\sigma^2$$

$$\begin{pmatrix}X_1'X_1 & X_1'X_2\\ X_2'X_1 & X_2'X_2\end{pmatrix}^{-1}=\begin{pmatrix}\cdot\cdot & \cdot\cdot\\ \cdot\cdot & (X_2'(I-H)X_2)^{-1}\end{pmatrix}$$

X_2 $(x_1,\dots,x_k,x_{k+1},\dots,x_m)$ （对于 $X_2^{\perp}=(I-H)X_2$）

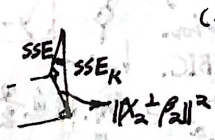
故无偏估计

$$\hat m=\|X_2^{\perp}\hat\beta_2\|^2+2p\hat\sigma^2-(m+1)\hat\sigma^2$$

↓

子模型 SSE_k − 全模型 SSE_m

$(n-(m+1))\hat\sigma^2$



$$\hat m=SSE_k+2p\hat\sigma^2-n\hat\sigma^2$$

（标化）$C_p=\frac{\hat m}{\hat\sigma^2}=\frac{SSE_k}{\hat\sigma^2}+2p-n$ ⟶ 子模型尽量 min 的预测误差.

C_p min 或 $|C_p-p|$ min.

（子基本正确时 $C_p\approx p$）

（子模型有偏，只要 $(X_2)\beta_2$ 不真的是0）

△ 证明：残差消元法. ⟶ 需分块矩阵逆公式

① 对 (X_1,X_2) 回归 $\begin{pmatrix}\hat\beta_1\\ \hat\beta_2\end{pmatrix}$（全） $SSR=\hat\beta_1'X_1'X_1\hat\beta_1+\hat\beta_1'X_1'X_2\hat\beta_2+\hat\beta_2'X_2'X_1\hat\beta_1+\hat\beta_2'X_2'X_2\hat\beta_2$

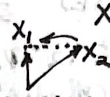
② 对 (X_1) 回归 $\hat b_1=(X_1'X_1)^{-1}X_1'Y$

$SSR=\hat b_1'X_1'X_1\hat b_1$

③ (X_2) 对 (X_1) 回归的残差，

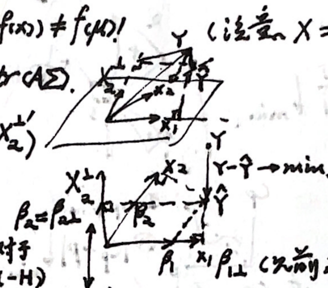
（使用消元法）$SSR=\hat\beta_2'(X_2'X_2-X_2'X_1(X_1'X_1)^{-1}X_1'X_2)\hat\beta_2$

$X_2^{\perp}=X_2-X_1(X_1'X_1)^{-1}X_1'X_2=(I-H)X_2.$



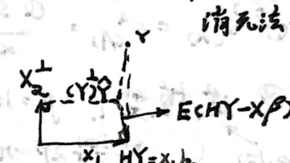
SSR：① = ② + ③

（注意 $X=(1, X_1, X_2)$. 模長的 乘是是 $X'X$）消元法为左乘.



（先前正交化(X_1与2)解相同.）

（检 β_2 即检 X_2 即 $X_2^{\perp}$ 方向是否为0.）



（BIC 基于 Bayes 框架）

△AIC：K-L 距离度量预测误差：

真 $y\sim g(y)$ 候选模型 $f(y;\theta)$.

$L(\theta)=\prod f(y_i;\theta)$，$L_{max}\to\hat\theta$.

$\hat f$ 与 g 距离 $K(g,\hat f)=\int g\log\left(\frac{g}{\hat f}\right)\geq 0$

$K\downarrow_{min}$：$\max\left(\int g\log f(y;\hat\theta)dy\right)$

$k(\hat\theta)$ $E(\log\hat f)$

注意 $E\left(\frac{1}{n}\sum\log f(y_i;\theta)\right)=k(\theta)$

? $E\left(\frac{1}{n}\sum\log f(y_i;\hat\theta)\right)=E(k(\hat\theta))+\frac{k}{n}+o(1)$ ↑ 偏差.

用 $\bar K-\frac{k}{n}\to k(\hat\theta)$

$$AIC=-2n\cdot\left(\bar K-\frac{k}{n}\right)=2k-2\log L(\hat\theta)\ \underset{min.}{\longrightarrow}\ k(\hat\theta)_{max}\text{ 即 } K\text{-}L_{min}.$$

多多回归：$Y = X\beta + E^{n\times p}$ $\sim \left(\begin{array}{c|c} & \end{array}\right) \beta_{(i)}$, β_i

$n\,(y_1,\cdots,y_p) = X\,(\beta_1,\cdots,\beta_p) + (\varepsilon_1,\cdots,\varepsilon_p)$

$\vec{Y} = (I_p \otimes X)\vec{\beta} + \vec{E}$ $\quad \varepsilon_{(i)} = (\varepsilon_{i1},\cdots,\varepsilon_{ip}) \overset{iid}{\sim} N_p(0,\Sigma)$.

$SSE = \sum_{ij} \varepsilon_{ij}^2 = \vec{E}'\vec{E} = (\vec{Y}-\hat{X}\vec{\beta})'(\vec{Y}-\hat{X}\vec{\beta})_{min}\ (\partial\vec{\beta}=0)\ (tr R_{min})$ estimation

残差平方和 投影(正规)方程(组) $\hat{X}'\hat{X}\vec{\beta} = \hat{X}'\vec{Y}$. and the nature of the estimator.

多残差阵 $R = (Y-\hat{Y})'(Y-\hat{Y}) = Y'(I-H)Y$. 於 $X'X(\beta_1\cdots\beta_p) = X'(Y_1,\cdots,Y_p)$

$\hat{\Sigma} = \dfrac{R}{n-k-1}$. $\begin{pmatrix} Y_1'(I-H)Y_1 & \cdots & \\ \vdots & & \vdots \\ & \cdots & Y_p'(I-H)Y_p \end{pmatrix}$ → 亦即 $\Sigma_{YY} - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}$. 同前所述.

(元K的$\hat{\beta}_K$的第i个分量Y) $E(Y_i'(I-H)Y_i) = (n-m-1)\sigma^2$ 　若多对一则 $R = R_{ii}$, (复)$R = \sqrt{1-\dfrac{R_{ii}}{(Y'Y)_{ii}}}$. (SSE)

$Cov(\hat{\beta}_{ik}',\hat{\beta}_{jl}) = \sigma_{kl}^2\cdot(X'X)^{-1}_{ij}$. $Y_i \sim N(X\beta_i,\sigma^2 I_n)$ (Q旋转(Λ)、定理、取tr)

则 $Cov(\hat{\beta}_i,\hat{\beta}_j) = \sigma_{ij}^2 (X'X)^{-1}$ — △技巧：$e_i'(X'X)^{-1}X'Ee_K$

$Cov(\hat{\beta}_{(i)}',\hat{\beta}_{(j)}') = (X'X)^{-1}_{ij}\Sigma$ $\ (X'X)^{-1}_{ij}$ $i\in(ij)$ 　用单位向量表示出元.

有 $\hat{\beta} \sim N_{mat.}$, $R \sim W_{n-K-1}(\Sigma)$, $\hat{\beta}$与R id. (⊥).

$\hat{\beta}_i \sim N(\beta_i, \sigma_{ii}^2(X'X)^{-1})$. 　$R = Y'(I-H)Y$ 　△ 二次型独立 (A.B sym.)：(充要) $X'AX$ id. $X'BX \Leftrightarrow AB=0$ (BA,取') 　(证明：$A = Q\Lambda Q'$, $\Lambda = \begin{pmatrix}0 & \\ & \Delta\end{pmatrix}_{k-r}$, $BQ = (O\ \Delta_{k-r})\ldots$) $Y'Y$与$Y'QBQ'Y\ldots$

$\hat{\beta}_{(i)} \sim N(\beta_{(i)}, l_{ii}\Sigma)$ 　$\hat{\beta}'X'X\hat{\beta} = Y'HY$. 　hypo. test. (of significance)

$\beta_{(0)},\beta_{(1)}\cdots\beta_{(m_1)},\beta_{(m_1+1)}\cdots\beta_{(m_1+m_2=k)}$ 　编 $H_0: \beta_{(i)} = 0_p$ 　$\hat{\beta}_{(i)}/\sqrt{l_{ii}} \sim N(0,\Sigma)$,

$H_0: B_2 = 0$. 　(单) H_0√时 $T^2 = (n-k-1)\hat{\beta}_{(i)}'R^{-1}\hat{\beta}_{(i)}/l_{ii} \sim T^2_{n-k-1}(p)$

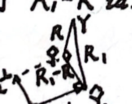
$R_1 - R = -Y'(H-H_1)Y = +Y(I-H_1)K(I-H_1)Y \sim W_{m_2}(\Sigma)$. 　即 $\dfrac{n-k-p}{p}\cdot\dfrac{\hat{\beta}_{(i)}'R^{-1}\hat{\beta}_{(i)}}{l_{ii}} \sim F_{p,n-k-p}$

(拆解H出H_1,X_2) $K = X_2(X_2'(I-H_1)X_2)^{-1}X_2'$. X_i的贡献 $V_i = \dfrac{\hat{\beta}_{(i)}'R^{-1}\hat{\beta}_{(i)}}{l_{ii}}$ 　($p=1$时 $V_i = \dfrac{\hat{\beta}_{(i)}^2}{l_{ii}\cdot SSE} = \dfrac{PSSR}{SSE}$) (∂SSR)

$H_1 = HH_1 = H_1H$ (理解) 　($n > m_1+m_2+1$) $\sim W_{m_2}(\Sigma)$ (秩m_2) → 亦即 $YH_{2^\perp}Y$ $(+Y'X_2^\perp(X_2^{\perp\prime}X_2^\perp)^{-1}X_2^{\perp\prime}Y)$

$(I-H)(H-H_1) = 0$ 则 $R_1 - R$ id. R. 现求LRS：

$L = (2\pi)^{-\frac{np}{2}}|\Sigma|^{-\frac{n}{2}}e^{-\frac{1}{2}tr\Sigma^{-1}(Y-X\beta)'(Y-X\beta)}$ 　$(+\hat{B}_2'X_2^{\perp\prime}X_2\hat{B}_2)$



$\max_{\beta,\Sigma} L$ 时 $\beta = \hat{\beta}$, $\Sigma = \frac{1}{n}R$.

$\lambda = \left(\dfrac{|R_1|}{|R|}\right)^{-\frac{n}{2}} = \left(\dfrac{|R|}{|R+(R_1-R)|}\right)^{\frac{n}{2}}$ 　H_0√时 $\Lambda \to 1$.

→ $\Lambda^{(p)}_{n-k-1,m_2}$

(单) $m_2 = 1$时 $\Lambda = \dfrac{1}{1+\frac{T^2}{n-k-1}}$, $(k = m_1+m_2)$ 　$P value = P(\Lambda \le \Lambda_{ob}) \gtrless \alpha$.

$\dfrac{n-k-p}{p}\cdot\dfrac{1-\Lambda}{\Lambda} = F_{p,n-k-p}$. (Wilks' Lambda)

多多下的 STEPREG：

变量u(外部)对p个Y贡献 $V_u = u'(I-H)u\cdot\hat{\beta}_{(u)}Q^{-1}\hat{\beta}_{(u)}'$ 　$1/l_{uu}$ 　原m个自变量 ↑ (未算β_0/Y) 　Q为残差R(阵) 　后(m+1)个 ↑

$u = \begin{pmatrix}u_1\\ \vdots \\ u_n\end{pmatrix}$ 　LR有 $1-V_u \to \Lambda$. ① $(T^2 = u'(I-H)u\hat{\beta}_{(u)}Q_u^{-1}\hat{\beta}_{(u)}' = \dfrac{u'(I-H)u\hat{\beta}_{(u)}Q^{-1}\hat{\beta}_{(u)}'}{1-u'(I-H)u\hat{\beta}_{(u)}Q^{-1}\hat{\beta}_{(u)}'}\cdot(n-m-2)$ $\cdot(n-m-2)$)

(此处$\beta_{(i)}$ — 上多$\beta_{(i)}$|) 　β_u的显著度/残差 或 的贡献量 (比Q) 　$(\Lambda = \dfrac{1}{1+\frac{T^2}{n-m-2}})(= \dfrac{1}{1+u'(I-H)u}$. Morrison. Formula. 　$\hat{\beta}_{(u)}Q_u^{-1}\hat{\beta}_{(u)}'$,可验证) 由 $Q_u = Q - u'(I-H)u\hat{\beta}_{(u)}'\hat{\beta}_{(u)}$.

$\dfrac{n-m-p-1}{p}\dfrac{V_i}{1-V_i} \to F$, $(n-m-2)\dfrac{V_i}{1-V_i} \to T^2$ ③ (①②两种推导路径，③为延伸)

$(R_1 - R = \hat{\beta}_{(u)}'u'(I-H)u\hat{\beta}_{(u)} \sim W_1(\Sigma_p)$ Λ 亦即 ② $\dfrac{|Q_u|}{|Q|} = \dfrac{|Q_u|}{|Q_u+W_1(\Sigma)|} \sim \Lambda \overset{\triangle}{=} \dfrac{|Q - \hat{\beta}_{(u)}'u'(I-H)u\hat{\beta}_{(u)}|}{|Q|} = 1 - V_i$.

$p=1$ 时 ΔSSR_i

$p\geq 1$ 时 V_i ($\frac{\Delta SSR_i}{SSE}$) 扩状向分

直接采取消元

初阵: $L^{(0)} = \begin{pmatrix} X'X & X'Y \\ Y'X & Y'Y \end{pmatrix}$ （m, p）

（消去变换）

↓

$L^{(r)}$ ⓪ DOUBLE STEPWISE 具体步骤？

定 α_X (in=out), α_Y.

① 标准化/中心化，~~Cov~~ Corr 阵 L

② 利用 $L^{(r)}$ ②: V_i / $\Lambda^{(p)}_{n-m-1,1}$; min/max V_i

③ ③ F 检验，剔/引

④ 继续或转入自/因.

⑤ 结果: "$Y_{...} = X_{...}\beta_{...} + E$.

2. $\Omega \to$ 全 3. $R = \sqrt{1-\Lambda}$, $R_{(Y\cdot X_{\text{更}})} = \sqrt{1-\text{…}}$（与 个[illegible]更）

4. test.

"最优"方程组与方程:

(P个) (某个 i) 双重筛选 (分组)

$\triangleright i$ 因 —— 因 —— 因

自 自 自

$\triangleright j$ · — — ·

Y_i △ $\Lambda = \frac{|R|}{|Y'Y|}$ (↓, 线性↑)

则 $\frac{\Lambda_i}{\Lambda} < 1$ ($1 - \frac{\Lambda_i}{\Lambda}$ ↑, Y_i 越显著)

△ $(1-\frac{\Lambda_2}{\Lambda_1})$ $\frac{\Lambda_1}{\Lambda_2} - 1$ 做F检验.

用绝对值公式把多出的一元$^{(i)}$拆出来, $\frac{\Lambda_1}{\Lambda_2} - 1$ 即 V_i (Y对X回归, 两模型等价. 考虑 Y_i 进入)

△ sweep. 消去变换/扫描算法:

⊙ STEPREG 重述:

?△ Wilks 定理

?△ $\frac{\Lambda_1}{\Lambda_2} \sim \Lambda$ 证明

判别分析(归类)：(dim m) $G_1,\cdots,G_k$. $F_i(x)$ ✓. ($i=1,\cdots,k$) $X_{new} \xrightarrow{?} F_i(G_i)$

overfitting

(训练集)

1. 马氏距离 $d^2=(X-\mu)'\Sigma^{-1}(X-\mu)$ 样本估计 μ_i,Σ_i.
(距离判别)

(各个G球性时为阿波罗尼乌斯球(判界)) $d_l^2(X)=\min_i d_i^2(X)$ 则 $X\in G_l$.

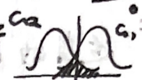
判别准则 | (判分).

- 两总体同方差假设：合并样本协方差阵 $S=\frac{1}{n_1+n_2-2}\sum_{i=1,2}(X-\bar{X})(X-\bar{X})'$

线性判别函数 $W(X)=d_2^2(X)-d_1^2(X)=2(X-\frac{\bar{X}_1+\bar{X}_2}{2})'S^{-1}(\bar{X}_1-\bar{X}_2)$ $G_2 \overset{W}{,} G_1$

2. Bayes 判别 先验 G_i-q_i (历史；样品 $\frac{n_i}{n}$；同一) LOOCV/K-fold CV 错判概率

(回判，留出，留一) ↑



$P(1|2)=P(2|1)=1-\phi(\frac{d_{12}}{2})$, $d_{12}^2=(\bar{X}_1-\bar{X}_2)'S^{-1}(\bar{X}_1-\bar{X}_2)$.

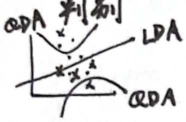
(多总体下不能用 DISCRIM).

esp. 后验概率最大. $P(j|i;D)=\int_{D_j}f_i(x)dx$

故 $\mu_1\neq\mu_2$ 显著时才有意义.

(q_i, $L(j|i)$ 一致时) $\max q_i f_i(x)$. $P(i|X)$. 错判损失 $L(j|i)$ < 经验赋值 $1-\delta_{ij}$.

二次判别



$\ln q_i-\frac{1}{2}\ln|\Sigma|-\frac{1}{2}d_i^2(x)$. 平均~

$L(D)=\sum_i q_i\sum_j P(j|i)L(j|i)$. (原)

即 $D_i^2=d_i^2+\ln|S_i|-2\ln q_i$ 最小. ($\leq\sum_i q_i r_i(D)$) $\in i$ 的判为 $j\neq i$ 的损失

作为判别 D_i. ↑

Bayes Discrim. Solution: $L(D^*)=\min_{\forall D} L(D)$

(Σ_i 相等时为线性判别) $D_i^*=\{X|\ h_i(x)<h_j(x),\ j\neq i\}$. 其中 $h_i(x)=\sum_j q_j f_j(x)L(i|j)$.

$G_2 \overset{W}{,} G_1$, $2\ln\frac{q_2 L(1|2)}{q_1 L(2|1)}$

3. Fisher 判别 (proof: $L(D)=\sum_i\int_{D_i}h_i(x)dx$) ∟ $\in j\neq i$ 的判为 i 的损失

(投影判别) $u(X)=l'X$. 当 $L(j|i)=1-\delta_{ij}$ (一致)时, $D_i^*:\{x|\ q_i f_i(x)>q_j f_j(x),\ j\neq i\}$. (原)

(齐方差)+距离

A 组内方和 $\frac{l'Bl}{l'Al}=\Delta(l)$ 判别效率 max.

B 组间方和 (一般令 $l'Al=1$) △ k=2 时投影与否等价于距离判别.

$l_i=\lambda_i$'s eigenvector. $l=\frac{1}{d}A^{-1}(\bar{X}_1-\bar{X}_2)$ (A^{-1} 与 S^{-1} 仅差倍数)

准则 I: $\lambda_1(\rho(A^{-1}B))$ $\lambda_i=(A^{-1}B)$'s eigenvalue. $W(X)=u(X)-u(\frac{\bar{X}_1+\bar{X}_2}{2})$ ✓. (d 归一数 $=\sqrt{(\bar{X}_1-\bar{X}_2)'A^{-1}(\bar{X}_1-\bar{X}_2)}$)

∴ $u_1(X)=l_1'X$, $\bar{u}_1(X)=l_1'\bar{X}$, $\bar{u}_{1j}(X)=l_1'\bar{X}_j$ ($j=1,\cdots,k$)

准则 II. 取 $l\leq r$. $P=\frac{\lambda_1+\cdots+\lambda_l}{\lambda_1+\cdots+\lambda_r}<1$ (≥ 0.7)

再距离判别. 化为 l 元总体作距离判别.

$\frac{|u_1(X)-\bar{u}_{1j}|}{\hat{\sigma}_j}=\min(\cdots)$ 则 $X\in G_j$. (l 个不相关)

其中 $\hat{\sigma}_j^2=l_1'S_jl_1$ ($\frac{1}{n_j-1}l_1'A_jl_1$) 判别检验：

(若有多个=, 则序贯使用 λ_2, l_2 再判) 1. 判别效果 test: $H_0: \mu_i=\mu_j=\cdots$ 若✓则无意义.

逐步判别 STEPWISE △ STEPDISC 重述 (Σ 假设相等): $\Lambda=\frac{|A|}{|T|}\sim\Lambda_{n-k,k-1}^{(m)}$

2. 判别能力 test: (某变量 $H_0:\mu_{ih}=\mu_{jh}\cdots$ 均值是否有显著差异) (仅二元时亦即 (T^2) $d_{ij}^2\sim F$)

聚类分析：系统聚类法、动态聚类法(调优)、有序聚类法(分割), 模糊~. 图论~… 聚类预报.

$\frac{\Lambda_m}{\Lambda_{m-1}}=1$ 做 F 检验? $\frac{n-k-m+1}{m(n-k)}\cdot\frac{n_in_j}{n_i+n_j}$ "

R型聚类. Q型聚类. 未考虑相关性 ⁓ Λ? ex.标准化 $(\bar{X}_j-\bar{X}_i)'S^{-1}(\bar{X}_j-\bar{X}_i)$

△ 距离：Minkorski $d_{ij}(q)=(\sum_t|X_{it}-X_{jt}|^q)^{\frac{1}{q}}$ Lance $d_{ij}(L)=\frac{1}{m}\sum_{t=1}^{m}\frac{|X_{it}-X_{jt}|}{X_{it}+X_{jt}}$ 变换↓ ($S=\frac{A_1+\cdots+A_k}{n-k}$)

($\frac{X_{it}-X_{jt}}{S_t}$) $d_{ij}(1)$, $d_{ij}(2)$, $d_{ij}(\infty)$ Mahalanobis $d_{ij}(M)=(X_i-X_j)'S^{-1}(X_i-X_j)$ (去量纲；对(大)奇异值不敏感)

斜交空间距离 $d_{ij}=(\frac{1}{m^2}\sum_{k=1}^{m}\sum_{l=1}^{m}(X_{ik}-X_{jk})(X_{il}-X_{jl})r_{kl})^{\frac{1}{2}}$, r_{kl} 为 $Corr(X_k,X_l)$. (S 使用全部样品计算效果 X, 使用各类 S_kX)

相似系数：$\cos\alpha_{ij}=\frac{\sum X_{it}X_{jt}}{\sqrt{(\sum X_{it}^2)(\sum X_{jt}^2)}}$ $\rho_{ij}=\frac{\sum(X_{ti}-\bar{X}_i)(X_{tj}-\bar{X}_j)}{\sqrt{\sum_t(X_{ti}-\bar{X}_i)^2\cdot\sum(X_{tj}-\bar{X}_j)^2}}$ 定性变量的样品 $d_{ij}=\frac{m_2}{m_1+m_2}$ 无法实现

变量距离. 相似系数同上, 定性变量 χ_{ij}^2, $\sqrt{\frac{\chi_{ij}^2}{\chi_{ij}^2+n}}$ … 或 $=\sum_{i,j}(\delta_i-\delta_j)^2$ …

($d_{ij}=1-|\rho_{ij}|$ 或 $=1-\rho_{ij}^2$)

系统聚类法：

$G_1, \cdots, G_t$

$t=n$ 开始 $G_i=\{X_{(i)}\}$ ⟶ 计算类间距 D_{ij}，合并最小两类（repeat）⟶ $G_1(t=1)$ 画谱系图，分析并取类数

min；max；median；centroid；average；flexible average；MCQ；Ward ⟶ 类间平方和 $D_{pq}^2 = W_{p+q} - (W_p + W_q)$

对 D_{ab}^2 递推

（欧式 d^2 时 $\frac{n_p n_q}{n_p+n_q}(\bar{X}_p - \bar{X}_q)^2$ & centroid）

单调性：median ✗，centroid ✗ （空间）收缩/扩张性

min < centroid < average < ward < max

类的定义：≤ Threshold

1. $\forall i,j,\ d_{ij} \le T$
2. $\forall i,\ \frac{1}{n-1}\sum_j d_{ij} \le T$
3. $\frac{1}{n(n-1)}\sum_i\sum_j d_{ij} \le T,\ d_{ij} \le T'$
4. $\forall i, \exists j,\ d_{ij} \le T$

5. $\forall a,b$ 划分为两类，$D(G_a, G_b) \le T$

类的个数：

适当的 d_0；散布图；统计量：$R_k^2 = 1 - \frac{P_k}{T} = \frac{B_k}{T}$，$\Delta R^2 = R_{k+1}^2 - R_k^2$（某步）

（大致随 $k\uparrow$）$pseudoF = \frac{B}{P}\cdot\frac{n-k}{k-1}$ ✱F

$pseudo\ t^2 = \frac{B_{kL}^2}{\frac{W_k+W_L}{n_k+n_L-2}}$（某步，合并 k 和 L）

$(B_{kL}^2 = W_{k+L} - W_k - W_L)$

特征：$\bar{X}_G$，A_G，D_G（trA_G 或 $\max_{ij} d_{ij}$）✱t

动态聚类法：

初始分类 ⟵ 凝聚点

挑选：1. 人选 2. 重心（形心）3. 密度（画圆）4. 均值图 5. 随机 X

按批修改：

a. min condensation point → group

b. centroid as conden. point：$\min_i d_i^2$

⇒ $\sum_i W_i$ 组内离差 min（Ward 的系统聚类一致）

（证明：此时随便更改一个都将增大 W）

逐个修改（K-means）：

选 K，min，max

（first round）前 K 个作 conden. p.（或其他方式）

exam. $ij\in O$，$d_{ij} \ge Min$，否则合并取 means（centroid）

intro. $h\notin O$

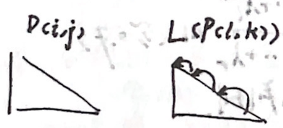
逐个再 exam.（迭代）：$d_{ih} \le Max$，归于最近点，取 means（centroid）；否则 > Max 新成一点。

无 Min Max 时：（按批）（取 K）→ 划入 min 中，重算 K（centroid）

有序聚类法：类直径（离差和）$D(i\sim j) = \sum_{i\sim j}(X-\bar{X})'(X-\bar{X})$

最优分割法（Fisher 算法）

$L(P(l,k))$ ⟵ 递推 $\min(L' + D)$



据 k 取用

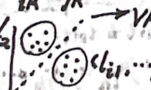
△缺点：
① K 把握
② 凸集收敛
③ 迭代：局部最优
④ 噪点敏感 outlier

2. PCA 分类：（标化，取 R）

变量距离 $\|X_i - X_j\|^2 = \sum_{k=1}^m (l_{ik} - l_{jk})^2$

因子载荷散布图 ⟶ VARCLUS

（此处 l 即 ρ）

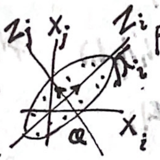


（样本 PCA 即 $\Sigma \to S = \frac{1}{n-1}X'X$ 或标准化后为 R）

样品距离 $\|X_{(i)} - X_{(j)}\|^2 = \sum_{k=1}^m (z_{ik} - z_{jk})^2$

3. 排序评估指数 Z_1（也叫主成分得分）（可加权 X_i）~散布图

4. normality test. $Z_1 \sim Z_m$ 无关，作一元正态检验



PCA：$Z = Q'X$

$L = Q\sqrt{\Lambda}$ $(\Sigma^{\frac{1}{2}})$

λ_i …… trD（$= tr\Sigma$）

（行/列元素平方和 $\sum_i / \sum_j l_{ij}^2$）

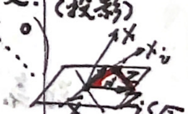
（标准化后 $\sigma_i^2 = 1$ 即 $trR = p$）

$\Sigma = R$，$L = \rho_{XZ}$

负荷 L：l_{ij} 为 X_i 分量（的离差）在主成分 j 方向上的长度（Q）（投影）

$\Sigma = LL'$，$\Lambda = L'L$

$D = \begin{pmatrix}\sigma_1^2 & 0 \\ 0 & \ddots\end{pmatrix}$



△ 秩了！

$Cov(X, Z) = Q\Lambda$

$\rho(X, Z) = \frac{L}{\sqrt{D}}$ $\left(\frac{Cov(X,Z)}{\sqrt{\Lambda}\sqrt{D}}\right)$

（PCR）with OLS：

$X_{n\times p} = Z_{n\times m} Q'_{m\times p} + E$，即 OLS 系数（残差 min）

（复）$R_i^2 = \frac{\sum_{k=1}^m l_{ik}^2}{\sigma_i^2}$ (X_i)

$SSR_i = (Q_i(Z'Z)Q_i')$，Λ_m

（相对于上方定义的 Q 这里要'）

解释难度（?逐步换回）

$Y \sim X$ β_X

$Y \sim Z$ $= Q\beta_Z$ ⟵ 有偏估计！

因子分析：因子模型 $X_i=\sum_{k=1}^{m}a_{ik}F_m+\varepsilon_i$

（Q型使用样品相似阵，寻找控制样品的因子）

{ 简化结构 / (变量/样品)分类

因子负荷（m 对 i 上）

特殊因子

公(共)因子（潜）

正交因子模型 $X=\mu+AF+\varepsilon$

A 为 loading mat. $\Sigma=AA'+D$. $\begin{pmatrix}F\\ \varepsilon\end{pmatrix}\sim N(0,\begin{pmatrix}I&0\\0&D\end{pmatrix})$

$\sigma_{ij}^2=$ [A] $\quad \sigma_{ii}^2=$ [ ] $+\sigma_{\varepsilon i}^2$. F 归一过.

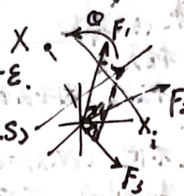
（A 列⊥）

X 标化则 $a_{ij}=\rho_{ij}$. $\sigma_{ij}^2=r_{x_ix_j}$. $\quad a_{ij}=Cov(X_i,F_j)$

A 不唯一：$V(X_i)=h_i^2+\varepsilon_i^2$.

$AQ\cdot Q'F$ 　公因子方差（共同度）　特因子方差（剩余方差）

因子得分：估计 $F_{(i)}$. $X=AF+\varepsilon$.



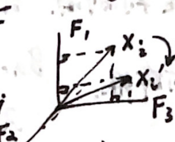
Bartlet —— 加权最小二乘（WLS）

$\min\varepsilon' D^{-1}\varepsilon$. $\hat F=(A'D^{-1}A)^{-1}A'D^{-1}X$.

$(D=V(\varepsilon))$

（或 L_{max}. $L=-\frac{1}{2}\varepsilon' D^{-1}\varepsilon-\frac{1}{2}\ln|2\pi D|$）

$\hat F_{(i)}$ 与主成分得分差一倍数 $=\frac{z_{ij}}{\sqrt{\lambda_j}}$

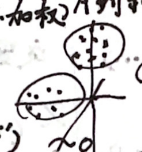


$(\|X_iQ\|=\|X_i\|$ 即 $\hat\sigma_i^2$ 不变）

Thompson —— 回归（多多）

（主成分解常不加权）A 列和改变，但 Σ 列和$^{(2)}=\Sigma\lambda$ (Q_1Q_2) 不变。

$F=BX$
$k\times n\ \ k\times p\ \ p\times n$ （因子得分函数）



（旋过的轴由 QF 尽可能穿过几个点群）

（可能是斜交旋转）

由载荷阵估计 B：$A=E(XF')=\Sigma B'$

$B=A'\Sigma^{-1}$

$\hat F=A'\Sigma^{-1}X$. ← Bayes 思想：后验分布的均值（先验：$\begin{pmatrix}X\\F\end{pmatrix}\sim N(0,\begin{pmatrix}\Sigma&A\\A'&I\end{pmatrix})$）

（常 Σ 代 S 或 R） $E(F-A'\Sigma^{-1}X)=0$. ✓

主成分解：$S\approx\sum_{i=1}^{m}\lambda_i q_i q_i'+D$

其误差平方和 $\sum_i\sum_j\varepsilon_{ij}^2\le\lambda_{m+1}^2+\cdots+\lambda_p^2$.

主因子解：有初始 $\hat\sigma_{\varepsilon i}^2$.

极大似然解：$\max_{A,D}L(\bar X,AA'+D)\Rightarrow\hat A,\hat D$.

$\hat\sigma_i^2=\sum_{t=1}^{m}a_{it}^2$, $\hat\varepsilon_i^2=1-\sum_{t=1}^{m}a_{it}^2$, $\sum_{i=1}^{p}a_{it}^2=\lambda_t$

① $\bar X$、S、R ② $R=Q\Lambda Q'$ ③ 定 m. $A=(Q\sqrt{\Lambda})_m$ ④ $\hat\sigma^2$, $\hat\sigma_\varepsilon^2$.（如 a_i 了）

解释因子（可先旋转）

△ 比较：$F_B=(A'D^{-1}A)^{-1}A'D^{-1}X$. ① $F_B=(I+(A'D^{-1}A)^{-1})F_T$ 常近似相等.（需要一点矩阵等式）

$F_T=A'(AA'+D)^{-1}X$. ② $E(F_B)=F$. $E(F_T)=A'\Sigma^{-1}AF$ 有偏.

Q型：矩阵大× 标准化难×（给定 F，$X\sim N(AF,D)$）（准确应加 |F） ③ $E((F_B-F)(F_B-F)')=(A'D^{-1}A)^{-1}$.
（依缺）

$E((F_T-F)(F_T-F)')=(I+A'D^{-1}A)^{-1}$, 误差更小.

相似系数(cos)阵求特征向量（及旋转）

可从代表性（≈1）系数看出样本聚类等.（或 F_i 聚类）（各因子）

对应分析：列联表的一类加权主成分分析.

（R-Q 因子分析） 克服变量和样品间量级的差异. ⟶

data transformation:

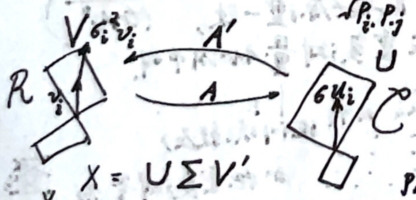
中心化、标准化、极差标准化/规格化.

卡方（对应、列联）变换

函数变换. Box-Cox 变换 $\hat\lambda$ 用 MLE 算. $y'=x\beta+\underline{e}$

$X\to Z\quad z_{ij}=\frac{p_{ij}-p_{i\cdot}p_{\cdot j}}{\sqrt{p_{i\cdot}p_{\cdot j}}}=\frac{x_{ij}-\frac{x_{i\cdot}x_{\cdot j}}{T}}{\sqrt{x_{i\cdot}x_{\cdot j}}}$ (p_{ij}) 对应阵.

$y'=\begin{cases}\frac{y^\lambda-1}{\lambda} & (\lambda\ne0)\\ \ln y & (\lambda=0)\end{cases}$



加权平方距离（卡方距离）$Q=\sum_i p_{i\cdot}\sum_j\frac{1}{p_{\cdot j}}(\frac{p_{ij}}{p_{i\cdot}}-p_{\cdot j})^2$

$X=U\Sigma V'$
$m\times m\ \ m\times p\ \ p\times p$
（r 个非 0，取）

$p/m\,\lceil^{r}\ \mathrm{L}$

总惯量（行点/样本至重心 c 的 weighted 距离平方和）$=\frac{\chi^2}{T}=\sum_i\sum_j z_{ij}^2=\sum_i\lambda_i$

$X_{(i)}\left|\begin{matrix}1\sim p\\ \vdots\\ m\end{matrix}\right.$

$S_R=X'X$　$S=LL'$

$S_Q=XX'$　$\Lambda=\Sigma^2=L'L$

Q 型 （principal coord.）

行轮廓（形象）（坐标）$G_r=G_Q=D_r^{-\frac{1}{2}}U\Sigma$ （除以 $\frac{1}{\sqrt{p_i}}$）

列轮廓 $G_c=G_R=D_c^{-\frac{1}{2}}V\Sigma$

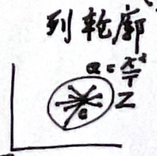
R 型

$L_R=V\Sigma'$（取 r 时 Σ sym.）

$L_Q=U\Sigma$

$\chi^2=T\,trS_R=T\,trS_Q$

$(T\,trZZ')=T\sum_i\lambda_i$.



△? coordinates and graphical reprenstation ↓ see in Wiki.

卡方分解（$Q=\sum_i\lambda_i$）（惯量）

一个 λ ⟶ 解释了一部分 inertia

⟶ 一个因子

△ Z: $u\sqrt{\lambda}v'+\cdots$

R $(v\sqrt{\lambda})$　$v\lambda v'+\cdots$

Q $(u\sqrt{\lambda})$　$u\lambda u'+\cdots$

（Z 及其 $\sum\sum z_{ij}^2$ 拆解）

sum(sum(Z)) = λ.

（χ^2 乘以 $\sqrt{T}$. $\sum\frac{(f_o-f_e)^2}{f_e}$）

$\sim_{df=(m-1)(p-1)}$.

▲ PCA、LDA、FDA 关系重述：

$\max \frac{x^TAx}{x^TBx} = \max \frac{x^T(B^{-\frac12}AB^{-\frac12})x'}{x'^Tx'}$ 也称(广义)瑞利商 Rq.

取 ρ(最大特征值)，方向为 (x') $B^{-\frac12}AB^{-\frac12}$ 特征方向；

或用求导(Lagrange λ) 得到. (x) $B^{-1}A$ 特征方向.)

$u(x)=w'x+b$ (超平面) ≥0

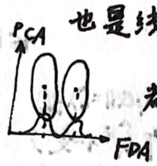
LDA：假设各类样本 Cov 相同(且满秩). S^{-1} (马氏距离).

FDA：逐方向(往大类间距方向)判别 esp. m=1 维时方差不同. $(W(x)=d_1^2-d_2^2\geq 0$ 也是线性判别的. 可开方 $d_1-d_2\geq 0)$

投影 (或降为k个) 可以认为是 LDA.

$u(x)=l'X$ 注意 PCA — A+B 特征方向

FDA — $A^{-1}B$ 特征方向

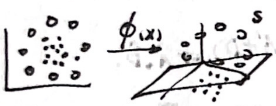


都可降维.

QDA：Cov 不同，只能写 d^2. Bayes：扣 $\ln q_i$, $\ln|S_i|$, $L(j|i)$

(先验概率 + 损失 $h_j=\sum_i\cdots$)

KFDA：kernel '核方法' ~



$K(\vec{x},\vec{y})=\phi(\vec{x})\cdot\phi(\vec{y})$ 核函数

$X=(x_1,\cdots,x_n)$ n 样本

非线性 — 线性可分，再 FDA. $\underset{m\times n}{\phi(X)}=(\phi(x_1),\cdots,\phi(x_n))$ m 变量.

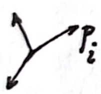
KPCA 同理.

$S=\frac{1}{n-1}\phi(X)\phi(X)'$ $Sp=\lambda p$

不重要.

kernel trick: $\phi(X)\phi(X)'p=\lambda p$. (SVD 角度：$\phi(X)$ 的奇异向量)

$(\sum_i\phi(X_i)\phi(X_i)'p=\lambda p)$



$p=\frac{1}{\lambda}\sum_i\phi(X_i)[\phi(X_i)'p]$. 令 $\alpha=\phi(X)'p\cdot\frac1\lambda$

$p=\phi(X)\alpha$. (α_i) (系数向量)

$\alpha=\phi(X)'p\cdot\frac{1}{\lambda}$. $\phi(X)'\phi(X)\alpha=\lambda\alpha$.

(λ≠0 的) $K=\phi(X)'\phi(X)$, $k_{ij}=\phi(X_i)\cdot\phi(X_j)$.

$(p'p=1\rightarrow \alpha'K\alpha=\lambda\alpha'\alpha=1$. (半)正定核.

$p=\phi(X)(\alpha\frac{1}{\sqrt\lambda})$ 即可归一.)

polynomial — $(x'y+c)^d$ $-\gamma\|x-y\|^2$

Gaussian/RBF — $\exp(\frac{-\|x-y\|^2}{2\sigma^2})$

sigmoid — $\tanh(ax'y+r)$

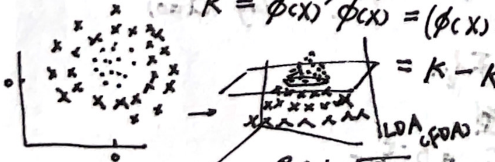
故主成分 $Z_i=\phi(X)'p_i=K\alpha_i=\lambda_i\alpha_i$. FDA 类似.

$u(x)=p'\phi(x)$

centralize: $\tilde\phi(X_i)=\phi(X_i)-\frac1n\phi(X)1$ $=\sum\alpha_iK(X_i,X)$.

$\tilde K=\tilde{\phi(X)}'\tilde{\phi(X)}=(\phi(X)-\phi(X)\frac{1}{n})'(\phi(X)-\phi(X)\frac{1}{n})$

$=K-K1_n-1_nK+1_nK1_n$.



▲ 球形(性)检验 $(H_0:\sigma^2 I)$

sphericity test：

Pearson r t 分布

Bartlet 球性验验 $H_0: I$ (对相关阵)

2 分类 (独立性) 相关性 检验 (因子分析前：变量是否独立的提供信息?)

$KMO=\frac{\sum_{i\neq j}r_{ij}^2}{\sum_{i\neq j}r_{ij}^2+\sum_{i\neq j}r_{ij\cdot}^2}$ → 简单相关系数；→ 偏~ $r_{ij\cdot(12\cdots)}$

KMO↑1 越有相关，适合做. KMO 应 >0.5 (0.7)

Bartlet: 似前 Cov test.

$\lambda=\frac{|\Sigma_0^{-1}A|^{\frac n2}}{\left(\frac{tr(\Sigma_0^{-1}A)}{p}\right)^{\frac{np}{2}}}$ ($\hat\sigma^2$ 由 $\frac{\partial\ln L(\bar X,\sigma^2\Sigma_0)}{\partial\sigma^2}=0$)

变量为相关阵. $\Sigma_0=I$.

$\chi^2(\frac{p(p-1)}{2}-0)\sim-2\ln\lambda=-n\ln|A|$. ↑ 即相关阵.

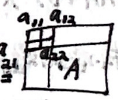
(详细版公式带 p…)

(关于求 KMO 方法 ← Matlab 代码).

$AIS=\frac{1}{D_{A^{-1}}}\cdot R^{-1}\cdot\frac{1}{D_{R^{-1}}}$?

$AIR=\frac{1}{\sqrt{D_{AIS}}}\cdot AIS\cdot\frac{1}{\sqrt{D_{AIS}}}$ $(=\frac{1}{\sqrt{D_{R^{-1}}}}\cdot R^{-1}\cdot\frac{1}{\sqrt{D_{R^{-1}}}})$

∂r_{ij} 阵. ★ 简单~与偏~关系.



证明：R^{-1}

$\frac{a_{12}-v_1'A^{-1}v_2}{\sqrt{a_{11}-v_1'A^{-1}v_1\quad a_{22}-v_2'A^{-1}v_2}}=\frac{\frac{|A|(\ )}{|\Sigma|}}{\sqrt{\frac{|A|(\ )}{|\Sigma|}\cdot\frac{|A|(\ )}{|\Sigma|}}}$

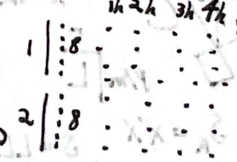
(可与先前 $\frac{\hat\rho_i^2}{\sigma_{ii}}$ 作一对比.) $\frac{R^{-1}_{12}}{\sqrt{R^{-1}_{11}\cdot R^{-1}_{22}}}$

(只剩一个对其间 y 的)

(Manchly sph. test $W=\lambda^{\frac2n}=\frac{p^p|A|}{(trA)^p}$)

与 Bartlet 原理一致，常在重复测量 ANOVA 中用.

two-factor: (mixed anova)



★ 把重复测量视作另一变量('组内效')的关键是球形检验.

(16 个样本，看 4 个变量相关性) 应无关 (每次认抽 8 个样本)

$Y=\bar Y+A_i+B_j+AB_{ij}+e$

64−1 = 63, 1, 3, 3, 56

(正态、方差齐、球)

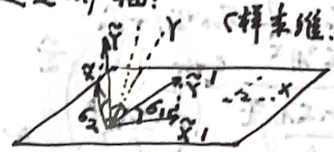
然后再检验 B 间、A 间效应. AB 交互. (二阶效应)

▲ 如? A: df — 1, e — 14

df?

CCA. RDA. PLS:　　△? RDA 待定：是提取 $\hat{Y}$ 还是 $\tilde{Y}$ 的主轴还是 α,β 轴？

典型相关分析. 典型冗余分析. 偏最小二乘法　　（样本维：可以做到 $\tilde{X},\tilde{Y}$ 的上）



$\tilde{X}=\alpha'X$　（设计型
$\tilde{Y}=\beta'Y$　$\tilde{X}=X\alpha,\ \tilde{Y}=Y\beta$）
（归一，经奇异案上 proj.）

$\max\frac{\alpha'\Sigma_{12}\beta}{\sqrt{\alpha'\Sigma_{11}\alpha}\sqrt{\beta'\Sigma_{22}\beta}}=\max\frac{\alpha'T\beta}{\sqrt{\alpha'\alpha}\sqrt{\beta'\beta}}.$

$\alpha=\Sigma_{11}^{-\frac{1}{2}'}U,\ \beta=\Sigma_{22}^{-\frac{1}{2}'}V$

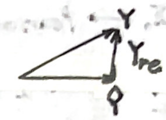
$T=\Sigma_{11}^{-\frac{1}{2}}\Sigma_{12}\Sigma_{22}^{-\frac{1}{2}'}=U\Sigma V'$　　$\Sigma_{\hat{Y}}=\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$. $\beta'\Sigma_{\hat{Y}}\beta=\Sigma^2$

（$\Sigma^{\frac{1}{2}'}=Q\sqrt{\Lambda},\ \Sigma^{\frac{1}{2}}=\sqrt{\Lambda}Q$ 注意）（有些地方说对 $\hat{Y}=HY$ 作主轴（β）也对）（$\tilde{Y}$ 长 σ_1）

△解决 X（Y）病态（collnearity），
　剥离共线性（挑出前 k 对）.

△相当于回归和 PCA.　　（$Y\beta=\hat{Y}\beta+Y_{re}\beta$）

此处 $\Sigma_{11}^{-\frac{1}{2}'}U$ 的载荷阵：同 F 一样归一；



但作了旋转 U（转后不能写 $L'L=\Lambda$ 了；
（共同的是，载荷要正规为 $L'\Sigma_{11}^{-1}L=I$）
　阵都是 Cov（O，O）.）且 U 考虑了 Y 的信息.
　　　　　　　　↑ 归一

predictor mt.　— scores mt.　— loading mt.

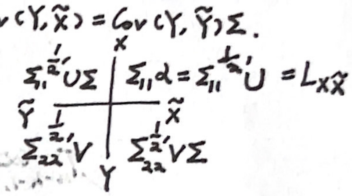
$X=F_XL_{X\tilde{X}}'+E_X$

$Y=F_YL_{Y\tilde{Y}}'+E_F$ — error mt.
response mt.

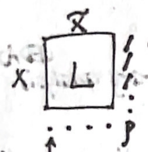
k principle components.

$\alpha'\Sigma_{11}\alpha=I$. $\beta'\Sigma_{22}\beta=I$.
（$L_{\tilde{X}X}\Sigma_{11}^{-1}L_{X\tilde{X}}=I$）
仍有 $L_{X\tilde{X}}L_{\tilde{X}X}'=\Sigma_{11}$（中间夹 I）
$L_{X\tilde{Y}}\Sigma^{-1}L_{\tilde{X}Y}=\Sigma_{12}$. 依此类推.

$Cov(X,\tilde{Y})=Cov(X,\tilde{X})\Sigma$
$Cov(Y,\tilde{X})=Cov(Y,\tilde{Y})\Sigma$.



Σ 即为 $\tilde{X}$. $\tilde{Y}$ 回归系数.



$L'L=U'\Lambda_{11}U$.
$tr(L'L)=tr(LL')=tr(\Sigma_{11})=p$.

$\tilde{X}_i$ 解释占比（$\frac{\cdot}{p}$）

△ Bartlet Σ_0 ⟷ id. test ? : 有区别.

$H_0: \sigma^2\Sigma_0$, $L(\bar{X}, \sigma^2\Sigma_0) \leftarrow \hat{\sigma}^2$

$$\Lambda = \frac{|\Sigma_0^{-1}A|^{\frac{n}{2}}}{\left(\frac{tr(\Sigma_0^{-1}A)}{p}\right)^{\frac{np}{2}}} \overset{Bartlet}{=} \left(\frac{|A|}{\left(\frac{tr(A)}{p}\right)^p}\right)^{\frac{n}{2}}$$

id. test $H_0: \Sigma_{ij} = 0$, $\forall i \neq j$

$\prod_i L_i(\bar{X}_i, \Sigma_{ii})$

$$* \quad \Lambda = \left(\frac{|A|}{\prod_i |A_{ii}|}\right)^{\frac{n}{2}}$$

Bartlet 只估一个σ^2,

假设好了各元的方差比例关系；

id. test 估每个元各自的方差. 知 $\Lambda_{Bart} < \Lambda_{id.}$

$\chi^2_{Bart} > \chi^2_{id.}$

(更易拒绝?)

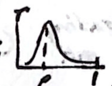
△ Bartlet Σ_0 ⟷ Pearson $t = r\sqrt{\frac{n-2}{1-r^2}}$? : ←

$-n \ln|R| \sim \chi^2\left(\frac{p(p-1)}{2}\right)$

$r\sqrt{\frac{n-2}{1-r^2}} \sim t(n-2)$

$(X^2 \sim Ga(\frac{1}{2}, \frac{1}{2}) \xrightarrow{CLT.} \bar{\chi}^2_n \sim N(1, \frac{2}{n}))$ $(n \to \infty)$

$(n-2)$ $\sqrt{\frac{SSR(MSR)}{MSE}}$ (esp. 仅当二元正态时) △ 相关系数分布



(表达式复杂)

$\left(\frac{\hat{\beta}_1^2}{\sigma_{\beta_1}^2}\right.$ 一样，一元时 $\left.\sigma_{\beta_1} = \frac{\hat{\sigma}}{X_1^T X_1}\right)$

$\rho = 0$ 时化为 $f(r) = \frac{(1-r^2)^{\frac{n-4}{2}}}{B(\frac{1}{2}, \frac{n-2}{2})}$ 与左侧等同.

(可化为t分布)

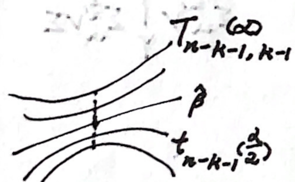
Fisher 变换：

$z = arcth^{-1} r = \frac{1}{2}\ln\frac{1+r}{1-r}$.

$z \approx N\left(\frac{1}{2}\ln\frac{1+\rho}{1-\rho}, \frac{1}{\sqrt{n-3}}\right)$

$\lim_{n \to \infty, r, \rho \to 0}$.

(逆 $r = \frac{e^{2z}-1}{e^{2z}+1}$)



NB ...

neutral network

SVM — algorithm

trees

classification

regression > unsupervised

dustering — supervised

△ 一般回归直线的联合置信区间?: 见置信椭球处(大T区间)

△ Skills for data analysis:

1. mathmatics
2. statistics
3. data visualization
4. programming
5. software engineering
6. machine learning
7. soft skills.

△ common process of data ana.:

collection

Data

selection, integration

Target Data

pre-processing:
- cleaning
- curation (handle missing data)
- transformation (ex. stdize)
- splitting

Processed Data

exploratory analysis (EDA)
- imbalance
- feature selection, extraction

building ML model:
- learning algorithm ($\xrightarrow{\text{Data mining}}$ patterns)
- parameter optimization
- cross-validation

Trained Model

prediction (test set).
evaluation (↺ modify)
interpretation

Deployment

△ overfitting:

1. more data
2. punish complexity (regulization)
3. less parameter (simpler model)
4. dedimension < feature selection / feature extraction

{ filter r. χ^2. $I(x,y)$. Var.
wrapper (screen)
embedded eval func → subset
ML scores.