第十讲 多元线性模型

2024.4.7

数据矩阵：横向排，竖向看

似然比检验

数理统计的似然理论指出，基于似然的极大似然估计是渐近最优的，似然比检验LRT以及渐近等价的Score、Wald检验也具有渐近最优性，原假设下它们都分布收敛于卡方。

Wilks定理. 假设$\mathbf{x}_i \in R^p, i=1,\dots,n,$的似然函数为$L(\boldsymbol{\theta}), \boldsymbol{\theta}\in\Theta,$
原假设$H_0:\boldsymbol{\theta}\in\Theta_0, \nu=\dim(\Theta), \nu_0=\dim(\Theta_0),$似然比检验统计量

$$\Lambda=\frac{\max\limits_{\boldsymbol{\theta}\in\Theta_0} L(\boldsymbol{\theta})}{\max\limits_{\boldsymbol{\theta}\in\Theta} L(\boldsymbol{\theta})}=\frac{L(\hat{\boldsymbol{\theta}}_0)}{L(\hat{\boldsymbol{\theta}})}, \quad Wilks'\,\Lambda^*=\Lambda^{2/n}$$

其中$\hat{\boldsymbol{\theta}}$为$\boldsymbol{\theta}$的极大似然估计，$\hat{\boldsymbol{\theta}}_0$为原假设下的$\boldsymbol{\theta}$的极大似然估计。
记$\nu=\dim(\Theta), \nu_0=\dim(\Theta_0)$，则$H_0$成立时，在正则(regular)条件下

$$-2\log\Lambda=-n\log\Lambda^*\xrightarrow{d}\chi^2_{\nu-\nu_0}, n\to\infty.$$

Bartlett校正方法修正n为nc_n，$c_n\approx 1$.

多正态总体的均值相同性检验：MANOVA

多正态问题

假设g个方差矩阵相同总体单均值可能不同的正态总体:

$\mathbf{x}_{11}, \ldots, \mathbf{x}_{1n_1}\ iid \sim N_p(\boldsymbol{\mu}_1, \Sigma)$, 样本均值和样本方差$\bar{\mathbf{x}}_1, S_1$

…

$\mathbf{x}_{g1}, \ldots, \mathbf{x}_{gn_g}\ iid \sim N_p(\boldsymbol{\mu}_g, \Sigma)$, 样本均值和样本方差$\bar{\mathbf{x}}_g, S_g$

考虑零假设H_0: $\boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_g$

总样本量：$n = n_1 + \cdots + n_g$

组内平均：$\bar{\mathbf{x}}_k = \frac{1}{n_k}\sum_{i=1}^{n_k} \mathbf{x}_{ki}$

组内方差: $S_k = \frac{1}{(n_k-1)}\sum_{i=1}^{n_k}(\mathbf{x}_{ki} - \bar{\mathbf{x}})(\mathbf{x}_{ki} - \bar{\mathbf{x}})^\top$

总平均：$\bar{\mathbf{x}} = \frac{1}{n}\sum_{k=1}^{g}\sum_{i=1}^{n_k} \mathbf{x}_{ki} = \frac{1}{n}\sum_{i=1}^{n_k} n_k \bar{\mathbf{x}}_k$

总方差：$S = \frac{1}{(n-1)}\sum_{k=1}^{g}\sum_{i=1}^{n_k}(\mathbf{x}_{ki} - \bar{\mathbf{x}})(\mathbf{x}_{ki} - \bar{\mathbf{x}})^\top = \frac{1}{(n-1)}\sum_{k=1}^{g}(n_k - 1)S_k$

总“平方和”：$T = (n-1)S = \sum_{k=1}^{g}\sum_{i=1}^{n_k}(\mathbf{x}_{ki} - \bar{\mathbf{x}})(\mathbf{x}_{ki} - \bar{\mathbf{x}})^\top$.

平方和分解

分解总“平方和”：

$$
\begin{aligned}
T &= (n-1)S = \sum_{k=1}^{g}\sum_{i=1}^{n_k}(\mathbf{x}_{ki}-\bar{\mathbf{x}})(\mathbf{x}_{ki}-\bar{\mathbf{x}})^\top \\
&= \sum_{k=1}^{g}\sum_{i=1}^{n_k}(\bar{\mathbf{x}}_k-\bar{\mathbf{x}}+\mathbf{x}_{ki}-\bar{\mathbf{x}}_k)(\bar{\mathbf{x}}_k-\bar{\mathbf{x}}+\mathbf{x}_{ki}-\bar{\mathbf{x}}_k)^\top \\
&= \sum_{k=1}^{g}\sum_{i=1}^{n_k}(\bar{\mathbf{x}}_k-\bar{\mathbf{x}})(\bar{\mathbf{x}}_k-\bar{\mathbf{x}})^\top+\sum_{k=1}^{g}\sum_{i=1}^{n_k}(\mathbf{x}_{ki}-\bar{\mathbf{x}}_k)(\mathbf{x}_{ki}-\bar{\mathbf{x}}_k)^\top \\
&\triangleq B+W
\end{aligned}
$$

其中

- B 为组间“平方和”（Between-group），代表各组之间的差异；
- W为组内“平方和”（Within-group），代表组内的变差。

H_0 的检验统计量基于 B相对于W的大小。

- 当 $p=1$时，$F=\frac{B/g-1}{W/(n-g)}$是一元anova的F检验统计量。
- 当 $p>1$时，B,W都是$p\times p$矩阵，此时检验统计量有多种构建方法，但都与矩阵BW^{-1}的某个数字特征，比如行列式、trace有关，比如Wilks' lambda
$$\Lambda^* = |W|/|W+B|$$

精确分布

原假设成立时（各组均值相同），B,W服从独立的Wishart分布，

$$U=(W+B)^{-1/2}W(W+B)^{-1/2}$$

服从多元beta分布 (Hsu 1939, Olkin & Rubin 1963)。

拓展到多个独立的Wishart矩阵，会得到多元Dirichlet分布。

若$W\sim W_p(m_1,\Sigma),B\sim W_p(m_2,\Sigma)$,独立，$U=(W+B)^{-1/2}W(W+B)^{-1/2}$

服从多元Beta分布$B_p(m_1/2,m_2/2)$

$$f(U)=\frac{\Gamma_p(m_1/2+m_2/2)}{\Gamma_p(m_1/2)\Gamma_p(m_2/2)}|U|^{(m_1-p-1)/2}|I_p-U|^{(m_2-p-1)/2}.$$

U是方阵，U的行列式的分布称为Wilks' lambda分布：

$$\Lambda^*\stackrel{\Delta}{=}|U|=|W|/|W+B|\sim\Lambda_p(m_1,m_2)\stackrel{d}{=}\prod_{i=1}^{m_2}\mathrm{beta}\left(\frac{m_1-p+i}{2},\frac{p}{2}\right),$$

该分布非常复杂。下面从似然比检验出发求出其近似逼近，即*Wilks*检验。

渐近分布

对于MANOVA问题（g个正态总体均值检验问题）：

$$\begin{cases} \mathbf{x}_{11},\dots,\mathbf{x}_{1n_1} \quad iid \sim N_p(\boldsymbol{\mu}_1,\Sigma), \text{ 样本均值和方差：} \bar{\mathbf{x}}_1, S_1 \\ \dots \\ \mathbf{x}_{g1},\dots,\mathbf{x}_{gn_g} \quad iid \sim N_p(\boldsymbol{\mu}_g,\Sigma), \text{ 样本均值和方差：} \bar{\mathbf{x}}_g, S_g \end{cases}$$

$$\Theta = \{\boldsymbol{\mu}_1,\dots,\boldsymbol{\mu}_g,\Sigma\}, \Theta_0 = \{\boldsymbol{\mu}=\boldsymbol{\mu}_1=\dots=\boldsymbol{\mu}_g,\Sigma\}.$$

- 原假设成立时，极大似然估计为：$\hat{\boldsymbol{\mu}}_0 = \bar{\mathbf{x}},\ \hat{\Sigma}_0 = (B+W)/n,$
- 没有限制时，极大似然估计：$\hat{\boldsymbol{\mu}}_k = \bar{\mathbf{x}}_k, k=1,\dots,g, \hat{\Sigma} = W/n,$

$$\Lambda = \frac{\max_{\boldsymbol{\theta}\in\Theta_0} L(\boldsymbol{\theta})}{\max_{\boldsymbol{\theta}\in\Theta} L(\boldsymbol{\theta})} = \frac{L(\hat{\boldsymbol{\mu}}_0,\hat{\Sigma}_0)}{L(\hat{\boldsymbol{\mu}}_1,\dots,\hat{\boldsymbol{\mu}}_g,\hat{\Sigma})} = \frac{|\hat{\Sigma}|^{n/2}}{|\hat{\Sigma}_0|^{n/2}} = \frac{|W|^{n/2}}{|B+W|^{n/2}}$$

$$\Lambda^* = \frac{|W|}{|B+W|}$$

由Wilks定理，$-2\log(\Lambda) = -n\log(\Lambda^*) = -n\log\left(\frac{|W|}{|B+W|}\right) \xrightarrow{d} \chi^2_{(g-1)p}$。

One-way MANOVA: Wilks'检验

令Wilks检验统计量(Wilks' Lambda）

$$\Lambda^* = \frac{\det(W)}{\det(W+B)}$$

则 H_0成立时 $T = -n\log\Lambda^* \xrightarrow{d} \chi^2_{(g-1)p}$， $n \to \infty$

当$g > 2$且$p > 1$时： 当$T \geq \chi^2_{p(g-1)}(\alpha)$ 时，在α水平下否定原假设。

Bartlett修正：$T_2 = -(n-1-\frac{p+g}{2})\log\Lambda^* \xrightarrow{d} \chi^2_{(g-1)p}$,

注意： 当$g = 1,2$或$p = 1$时，没必要应用上述近似检验。

$g = 2$时, $-2\log\Lambda = n\log(1+T^2/(n_1+n_2-2))$，其中

$$T^2 = \frac{n_1 n_2}{n}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^\top S_p^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)\text{服从}cF\text{分布(精确分布)}$$

$p = 1$时， B/W服从cF分布(精确分布).

$g=2$时，两样本Hotelling T2检验与Wilks检验1-1对应，单正态总体情形也是如此：

例1. 设$\mathbf{x}_1,\ldots,\mathbf{x}_n\ \ iid \sim N_p(\boldsymbol{\mu},\Sigma)$, $H_0:\boldsymbol{\mu}=\boldsymbol{\mu}_0$（$\boldsymbol{\mu}_0$已知),则Wilks$\Lambda^*$等价于$T^2$:

$$\Lambda^*=\frac{1}{1+T^2/(n-1)},\quad T^2=n(\bar{\mathbf{x}}-\boldsymbol{\mu}_0)^{\mathsf{T}}S^{-1}(\bar{\mathbf{x}}-\boldsymbol{\mu}_0)$$

证：$\hat{\Sigma}=\frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i-\bar{\mathbf{x}})(\mathbf{x}_i-\bar{\mathbf{x}})^{\mathsf{T}}$,原假设下，$\hat{\Sigma}_0=\frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i-\boldsymbol{\mu}_0)(\mathbf{x}_i-\boldsymbol{\mu}_0)^{\mathsf{T}}$,故

分解$\sum_{i=1}^{n}(\mathbf{x}_i-\boldsymbol{\mu}_0)(\mathbf{x}_i-\boldsymbol{\mu}_0)^{\mathsf{T}}=\sum_{i=1}^{n}(\mathbf{x}_i-\bar{\mathbf{x}})(\mathbf{x}_i-\bar{\mathbf{x}})^{\mathsf{T}}+n(\bar{\mathbf{x}}-\boldsymbol{\mu}_0)(\bar{\mathbf{x}}-\boldsymbol{\mu}_0)^{\mathsf{T}}$,

所以$$\Lambda^*=\frac{\det(\hat{\Sigma})}{\det(\hat{\Sigma}_0)}=\frac{\det\left(\sum_{i=1}^{n}(\mathbf{x}_i-\bar{\mathbf{x}})(\mathbf{x}_i-\bar{\mathbf{x}})^{\mathsf{T}}\right)}{\det\left(\sum_{i=1}^{n}(\mathbf{x}_i-\bar{\mathbf{x}})(\mathbf{x}_i-\bar{\mathbf{x}})^{\mathsf{T}}+n(\bar{\mathbf{x}}-\boldsymbol{\mu}_0)(\bar{\mathbf{x}}-\boldsymbol{\mu}_0)^{\mathsf{T}}\right)}$$

$$=\frac{\det(S)}{\det\left(S+n(\bar{\mathbf{x}}-\boldsymbol{\mu}_0)(\bar{\mathbf{x}}-\boldsymbol{\mu}_0)^{\mathsf{T}}/(n-1)\right)}=\frac{1}{\det\left(I_p+n(\bar{\mathbf{x}}-\boldsymbol{\mu}_0)(\bar{\mathbf{x}}-\boldsymbol{\mu}_0)^{\mathsf{T}}S^{-1}/(n-1)\right)}$$

$$=\frac{1}{1+n(\bar{\mathbf{x}}-\boldsymbol{\mu}_0)^{\mathsf{T}}S^{-1}(\bar{\mathbf{x}}-\boldsymbol{\mu}_0)/(n-1)}=\frac{1}{1+T^2/(n-1)}.$$

其它检验

因为B, W都是矩阵，Wilks Λ^*

$$\Lambda^* = \frac{|\mathrm{W}|}{|\mathrm{W}+\mathrm{B}|} = \frac{1}{|\mathrm{I}_p + \mathrm{BW}^{-1}|},$$

以行列式度量B相对于W是否接近0，其它检验方法还有：

- Lawley-Hotelling trace: $tr(\mathrm{BW}^{-1})$
- Phillai trace: $tr(\mathrm{B}(\mathrm{B}+\mathrm{W})^{-1})$
- Roy's largest root(最大特征根): $\lambda_{\max}(\mathrm{W}(\mathrm{B}+\mathrm{W})^{-1})$

例2 (例6.10， Johnson and Wichern p233). 威斯康星州卫生和社会服务部需要给养老院（nursing home）提供补贴, 补贴多少依据护理等级、护理成本或职工工资水平等。养老院分私营、非盈利经营和国企等三种所有权形式。我们希望了解养老院的运行成本是否与所有权形式有关，共统计了 $p = 4$种人力成本（护理、膳食、设备运行和维护、清洁维护）。各组数据的均值方法如下：

Group	Number of observations	Sample mean vectors
$\ell = 1$ (private)	$n_1 = 271$	
$\ell = 2$ (nonprofit)	$n_2 = 138$	$\bar{\mathbf{x}}_1 = \begin{bmatrix} 2.066 \\ .480 \\ .082 \\ .360 \end{bmatrix}$; $\bar{\mathbf{x}}_2 = \begin{bmatrix} 2.167 \\ .596 \\ .124 \\ .418 \end{bmatrix}$; $\bar{\mathbf{x}}_3 = \begin{bmatrix} 2.273 \\ .521 \\ .125 \\ .383 \end{bmatrix}$
$\ell = 3$ (government)	$n_3 = 107$	
	$\sum_{\ell=1}^{3} n_\ell = 516$	

Sample covariance matrices

$$\mathbf{S}_1 = \begin{bmatrix} .291 & & & \\ -.001 & .011 & & \\ .002 & .000 & .001 & \\ .010 & .003 & .000 & .010 \end{bmatrix}; \qquad \mathbf{S}_2 = \begin{bmatrix} .561 & & & \\ .011 & .025 & & \\ .001 & .004 & .005 & \\ .037 & .007 & .002 & .019 \end{bmatrix};$$

$$\mathbf{S}_3 = \begin{bmatrix} .261 & & & \\ .030 & .017 & & \\ .003 & -.000 & .004 & \\ .018 & .006 & .001 & .013 \end{bmatrix}$$

由上页数据计算得到总平均$\bar{\mathbf{x}}$，组间和组内“平方和”B,W:

$$\bar{\mathbf{x}} = \frac{n_1\bar{\mathbf{x}}_1 + n_2\bar{\mathbf{x}}_2 + n_3\bar{\mathbf{x}}_3}{n_1 + n_2 + n_3} = \begin{bmatrix} 2.136 \\ .519 \\ .102 \\ .380 \end{bmatrix}$$

$$\mathbf{B} = \sum_{\ell=1}^{3} n_\ell(\bar{\mathbf{x}}_\ell - \bar{\mathbf{x}})(\bar{\mathbf{x}}_\ell - \bar{\mathbf{x}})' = \begin{bmatrix} 3.475 & & & \\ 1.111 & 1.225 & & \\ .821 & .453 & .235 & \\ .584 & .610 & .230 & .304 \end{bmatrix}$$

$$\mathbf{W} = (n_1 - 1)\mathbf{S}_1 + (n_2 - 1)\mathbf{S}_2 + (n_3 - 1)\mathbf{S}_3$$
$$= \begin{bmatrix} 182.962 & & & \\ 4.408 & 8.200 & & \\ 1.695 & .633 & 1.484 & \\ 9.581 & 2.428 & .394 & 6.538 \end{bmatrix}$$

$$\Lambda^* = \frac{\det(\mathrm{W})}{\det(\mathrm{W}+\mathrm{B})} = 0.7714, \quad -n\log\Lambda^* = 132.8 > \chi^2_{p(g-1)}(\alpha) = \chi^2_8(0.01) = 20.09,$$

拒绝原假设。$pvalue = P(\chi^2_{p(g-1)} \geq 132.8) = 0.001$

总结：正态均值检验

g个总体：$\mathbf{x}_{k1}, \dots, \mathbf{x}_{kn_k} \sim N_p(\boldsymbol{\mu}_k, \Sigma), k = 1, \dots, g;$
零假设：$H_0: \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_g$

	精确检验($g \le 2$)		近似检验($g > 2$)	
	单总体 ($g = 1$)	两总体 ($g = 2$)	多总体 ($g > 2$)	$g = \infty$
统计量	$T^2 = n\bar{\mathbf{x}}^\top S^{-1}\bar{\mathbf{x}}$	$T^2 = \frac{n_1 n_2}{n_1 + n_2}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^\top S^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$	$-n\log\frac{\lvert W\rvert}{\lvert W + B\rvert}$	
零分布	$\frac{(n-1)p}{n-p} F_{p,n-p}$	$\frac{(n-2)p}{n-p-1} F_{p,n-p-1}$	$\chi^2_{(g-1)p}$	
多元 ($p > 1$)	Hotelling T^2检验（F检验）	Hotelling T^2检验（F检验）	MANOVA（卡方检验）	多元线性模型（卡方检验）
一元 ($p = 1$)	t检验	t检验	ANOVA（F检验）	一元线性模型（F检验）

以上所有检验都只与协方差矩阵有关

多元线性回归模型

- 多元线性模型将多元响应变量与自变量以线性形式联系起来.
- 多元线性模型的标准解法是最小二乘法。
- MANOVA（包括Hotelling T^2检验)是多元线性模型的特殊情形。
- 通常多元线性模型可以转化为每个响应与自变量的一元线性回归问题。

多元线性回归模型的一个难点在于理解模型公式的含义

例3. 同一研究对象有3个相关的响应(response) $\mathbf{y}=\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$，比如血压、脉搏、心率，假设这些响应与自变量$x$满足一元线性模型：

$$y_k = a_k + b_k x + \varepsilon_k,\ \varepsilon_k \sim (0, \sigma_k^2),\quad k=1,2,3$$

3个模型合在一起即是多元线性回归模型：

$$\mathbf{y}=\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}=\begin{pmatrix} a_1+b_1x+\varepsilon_1 \\ a_2+b_2x+\varepsilon_2 \\ a_3+b_3x+\varepsilon_3 \end{pmatrix}=\begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \\ a_3 & b_3 \end{pmatrix}\begin{pmatrix} 1 \\ x \end{pmatrix}+\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix}$$

$$\triangleq B^{\mathsf{T}}\mathbf{x}+\boldsymbol{\varepsilon},\quad B=\begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{pmatrix}$$

其中$\boldsymbol{\varepsilon}=(\varepsilon_1,\varepsilon_2,\varepsilon_3)^{\mathsf{T}}$, 假设 $\mathrm{var}(\boldsymbol{\varepsilon})=\Sigma$ (Σ非对角, 三个响应相关)。

例4. 假设x是二值变量，$x=0$或1，代表两组，假设$q\times 1$
响应$\mathbf{y}$在两组分别服从多元正态

$$\mathbf{y}|_{x=0} \sim N_q(\boldsymbol{\mu}_0, \Sigma), \quad \mathbf{y}|_{x=1} \sim N_q(\boldsymbol{\mu}_1, \Sigma),$$

$$\Leftrightarrow \mathbf{y} = \boldsymbol{\mu}_0 + (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)x + \boldsymbol{\varepsilon},\ \boldsymbol{\varepsilon} \sim N_q(\mathbf{0}, \Sigma), \qquad \mathbf{a} = \boldsymbol{\mu}_0, \quad \mathbf{b} = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_0$$

$$\Leftrightarrow \mathbf{y} = \mathbf{a} + \mathbf{b}x + \boldsymbol{\varepsilon} = (\mathbf{a}, \mathbf{b})\begin{pmatrix}1\\x\end{pmatrix} + \boldsymbol{\varepsilon} \triangleq B^{\top}\mathbf{x} + \boldsymbol{\varepsilon}, \qquad B^{\top} = (\mathbf{a}, \mathbf{b}), \mathbf{x} = \begin{pmatrix}1\\x\end{pmatrix}$$

一般地，x未必二值，误差未必正态，

$$\mathbf{y}_{q\times 1} = B^{\top}\mathbf{x}_{2\times 1} + \boldsymbol{\varepsilon}_{q\times 1} = \mathbf{a}_{q\times 1} + \mathbf{b}_{q\times 1}x + \boldsymbol{\varepsilon}_{q\times 1}, \quad \boldsymbol{\varepsilon} \sim (\mathbf{0}, \Sigma)$$

称为简单线性回归模型($B^{\top}: q\times 2$)。

多个自变量情形：$\mathbf{x} = (x_1, x_2, \ldots, x_p)^{\top}$ 维数为p, $x_1 = 1$

$$\mathbf{y}_{q\times 1} = B^{\top}\mathbf{x}_{p\times 1} + \boldsymbol{\varepsilon}_{q\times 1} = \boldsymbol{\beta}_1 x_1 + \cdots + \boldsymbol{\beta}_p x_p + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim (\mathbf{0}, \Sigma)$$

称为多重（自变量）多元（响应）线性回归模型($B^{\top}: q\times p$)。

Multiple　　Multivariate

多元线性
回归模型
(总体版本)

假设响应为$q\times 1$向量$\mathbf{y}$，自变量$\mathbf{x}=(x_1,x_2,\dots,x_p)^\top, x_1=1$

$$\mathbf{y}_{q\times 1}=B^\top \mathbf{x}_{p\times 1}+\boldsymbol{\varepsilon}_{q\times 1}=\boldsymbol{\beta}_1 x_1+\cdots+\boldsymbol{\beta}_p x_p+\boldsymbol{\varepsilon},\ \boldsymbol{\varepsilon}\sim(\mathbf{0},\Sigma),\boldsymbol{\varepsilon}\perp\!\!\!\perp\mathbf{x}$$

其中$B=(\beta_{ij})=(\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_p)^\top=(\boldsymbol{\beta}_{(1)},\dots,\boldsymbol{\beta}_{(q)})$是$p\times q$回归系数矩阵，$\boldsymbol{\beta}_{ij}$为$x_i$对$y_j$的效应。当$q=1$时，即一元的线性回归模型。

上述模型满足Gauss-Markov假设：
- 回归函数线性：$E(\mathbf{y}|\mathbf{x})=B^\top\mathbf{x}$
- 方差齐性：$\mathrm{var}(\mathbf{y}|\mathbf{x})=\Sigma$，与$\mathbf{x}$无关
- 外生性：$\boldsymbol{\varepsilon}=\mathbf{y}-B^\top\mathbf{x}\perp\!\!\!\perp\mathbf{x}$

各个响应关于$\mathbf{x}$都服从一元线性回归模型：

$$\mathbf{y}=\begin{pmatrix}y_1\\ \vdots\\ y_q\end{pmatrix}=\begin{pmatrix}\boldsymbol{\beta}_{(1)}^\top\mathbf{x}\\ \vdots\\ \boldsymbol{\beta}_{(q)}^\top\mathbf{x}\end{pmatrix}+\begin{pmatrix}\varepsilon_1\\ \vdots\\ \varepsilon_q\end{pmatrix}\Leftrightarrow y_j=\boldsymbol{\beta}_{(j)}^\top\mathbf{x}+\varepsilon_j, j=1,\dots,q.$$

$\boldsymbol{\beta}_{(j)}$为第j个响应的所有回归系数

多元线性回归模型(样本版本)

数据$\mathbf{y}_i, \mathbf{x}_i, i = 1, \dots, n$满足多元线性回归模型：

$$\mathbf{y}_i = B^\top \mathbf{x}_i + \boldsymbol{\varepsilon}_i, \boldsymbol{\varepsilon}_i, i = 1, \dots, n\ iid \sim (\mathbf{0}, \Sigma)$$

逐行排列： $Y_{n\times q} = \begin{pmatrix} \mathbf{y}_1^\top \\ \vdots \\ \mathbf{y}_n^\top \end{pmatrix}, X_{n\times p} = \begin{pmatrix} \mathbf{x}_1^\top \\ \vdots \\ \mathbf{x}_n^\top \end{pmatrix}, \mathcal{E} = \begin{pmatrix} \boldsymbol{\varepsilon}_1^\top \\ \vdots \\ \boldsymbol{\varepsilon}_n^\top \end{pmatrix}$

则我们有多元线性回归模型

$$Y_{n\times q} = X_{n\times p} B_{p\times q} + \mathcal{E}_{n\times q}, \mathcal{E}\text{的各行iid}\sim(\mathbf{0}, \Sigma), \Sigma = \left(\sigma_{ij}\right)$$

q元模型: q个一元模型

$Y = XB + \mathcal{E}$：横向排（行内相关），竖向看（列内球对称）

记$Y, B, \mathcal{E}$的第k列为$\mathbf{y}_{(k)}$，$\boldsymbol{\beta}_{(k)}$，$\boldsymbol{\varepsilon}_{(k)}$，即

$$Y = \left(\mathbf{y}_{(1)}, \dots, \mathbf{y}_{(q)}\right),\ B = \left(\boldsymbol{\beta}_{(1)}, \dots, \boldsymbol{\beta}_{(q)}\right), \mathcal{E} = \left(\boldsymbol{\varepsilon}_{(1)}, \dots, \boldsymbol{\varepsilon}_{(q)}\right),$$

第k个响应的一元模型：

$$\mathbf{y}_{(k)} = X\boldsymbol{\beta}_{(k)} + \boldsymbol{\varepsilon}_{(k)},\ \ \boldsymbol{\varepsilon}_{(k)} \sim N_n(\mathbf{0}, \sigma_{kk} I_n),\ \ k = 1, \dots, q$$

为了拟合q元模型，我们通常对每一个响应拟合一元模型。

最小二乘法

Frobenius模

对任何矩阵$A, B \in R^{n\times q}$, 定义内积

$$\langle A, B\rangle = tr(A^\top B) = \sum_{i,j} a_{ij} b_{ij}$$

矩阵A的Frobenius/欧氏模

$$\|A\| = \sqrt{\langle A, A\rangle} = \sqrt{tr(A^\top A)} = \sqrt{\sum_{i,j} a_{ij}^2}$$

最小二乘

多元线性模型$Y = XB + \mathcal{E}$，$\mathcal{E}$的各行iid$\sim(\mathbf{0}, \Sigma)$,$\mathcal{E} \perp\!\!\!\perp X$. 极小化误差平方和

$$\|\mathcal{E}\|^2 = \|Y - XB\|^2 = tr\left((Y - XB)^\top (Y - XB)\right)$$

得到的最优解称为为LS估计。

命题1. 假设X列满秩，最小二乘估计 $\hat{B} = (X^\top X)^{-1} X^\top Y$。

定义. 拟合值：$X\hat{B}$，残差：$Y - X\hat{B}$。Σ的LS估计定义为

$$\hat{\Sigma} = \frac{1}{n-p}(Y - X\hat{B})^\top \left(Y - X\hat{B}\right) = \frac{1}{n-p} Y^\top (I_n - P_X) Y.$$

求导
方法

证明1：目标函数

$$f(B) = tr\left((Y - XB)^\top (Y - XB)\right)$$
$$= tr(Y^\top Y) - 2tr(Y^\top XB) + tr(B^\top X^\top XB)$$

令 $\frac{\partial f(B)}{\partial B} = 2X^\top XB - 2X^\top Y = 0$，得正则方程

$$X^\top XB = X^\top Y$$

$\Rightarrow$ LS估计 $\hat{B} = (X^\top X)^{-1}X^\top Y$。

$$\frac{\partial tr(Y^\top XB)}{\partial B} = X^\top Y$$
$$\frac{\partial tr(B^\top X^\top XB)}{\partial B} = 2X^\top X\, B$$

投影
方法

证明2：多元线性模型的最小二乘问题：

$$\min_{B \in R^{p\times q}} \|Y - XB\|^2 = \min_{U = XB\, \in C(X)} \|Y - U\|^2$$

最优的$U = XB$为Y在X的列向量张成的空间 $C(X)$上的投影：

$$\hat{Y} = P_X Y = X(X^\top X)^{-1}X^\top Y = X\hat{B},$$

X的系数即LS估计 $\hat{B} = (X^\top X)^{-1}X^\top Y$ 。

注：因为矩阵Y的投影$P_X Y$等于其各列分别投影

$$\hat{Y}=P_X Y=\left(P_X\mathbf{y}_{(1)},\dots,P_X\mathbf{y}_{(q)}\right)=\left(\hat{\mathbf{y}}_{(1)},\dots,\hat{\mathbf{y}}_{(q)}\right)$$

所以多元线性模型的最小二乘等价于q个响应分别对X回归：

$$\mathbf{y}_{(j)}\sim X\ \Rightarrow\hat{\mathbf{y}}_{(j)}=X\widehat{\boldsymbol{\beta}}_{(j)},\qquad\widehat{\boldsymbol{\beta}}_{(j)}=(X^\top X)^{-1}X^\top\mathbf{y}_{(j)}$$

所有回归系数的LS估计

$$\hat{B}=\left(\widehat{\boldsymbol{\beta}}_{(1)},\dots,\widehat{\boldsymbol{\beta}}_{(q)}\right)=(X^\top X)^{-1}X^\top\left(\mathbf{y}_{(1)},\dots,\mathbf{y}_{(q)}\right)=(X^\top X)^{-1}X^\top Y$$

拟合q个一元模型

证明3： $Y=\left(\mathbf{y}_{(1)},\dots,\mathbf{y}_{(q)}\right)$，$B=\left(\boldsymbol{\beta}_{(1)},\dots,\boldsymbol{\beta}_{(q)}\right)$， $\mathcal{E}=\left(\boldsymbol{\varepsilon}_{(1)},\dots,\boldsymbol{\varepsilon}_{(q)}\right)$，对于第$k$个响应，有一元线性模型：

$$\mathbf{y}_{(k)}=X\boldsymbol{\beta}_{(k)}+\boldsymbol{\varepsilon}_{(k)},\ \ \boldsymbol{\varepsilon}_{(k)}\sim N_n(\mathbf{0},\sigma_{kk}I_n),\ \ k=1,\dots,q$$

$$\|E\|^2=\|Y-XB\|^2=\sum_{k=1}^q\left\|\boldsymbol{\varepsilon}_{(k)}\right\|^2=\sum_{k=1}^q\left\|\mathbf{y}_{(k)}-X\boldsymbol{\beta}_{(k)}\right\|^2,$$

$$\min\|E\|^2\Leftrightarrow\min\left\|\mathbf{y}_{(k)}-X\boldsymbol{\beta}_{(k)}\right\|^2,k=1,\dots,q$$

$$\Rightarrow\hat{\mathbf{y}}_{(1)}=X(X^\top X)^{-1}X^\top\mathbf{y}_{(k)},\ \ \widehat{\boldsymbol{\beta}}_{(k)}=(X^\top X)^{-1}X^\top\mathbf{y}_{(k)}$$

$$\Rightarrow\hat{B}=\left(\widehat{\boldsymbol{\beta}}_{(1)},\dots,\widehat{\boldsymbol{\beta}}_{(q)}\right)=(X^\top X)^{-1}X^\top\left(\mathbf{y}_{(1)},\dots,\mathbf{y}_{(q)}\right)=(X^\top X)^{-1}X^\top Y。$$

命题2. 假设X列满秩， $\Sigma=(\sigma_{ij})>0$, 则
(1) $E(\hat{B}|X)=B, E(\hat{\Sigma}|X)=\Sigma$,
(2) $\mathrm{var}(\hat{B}|X)=\Sigma\otimes(X^\top X)^{-1}$
$\Leftrightarrow \mathrm{var}(\hat{\boldsymbol{\beta}}_{(i)}|X)=\sigma_{ii}(X^\top X)^{-1}, cov(\hat{\boldsymbol{\beta}}_{(i)}, \hat{\boldsymbol{\beta}}_{(j)}|X)=\sigma_{ij}(X^\top X)^{-1}$,
其中 $B=(\boldsymbol{\beta}_{(1)},\dots,\boldsymbol{\beta}_{(q)})$， $\hat{B}=(\hat{\boldsymbol{\beta}}_{(1)},\dots,\hat{\boldsymbol{\beta}}_{(q)})$

证明: (1) $\mathcal{E}=(\boldsymbol{\varepsilon}_{(1)},\dots,\boldsymbol{\varepsilon}_{(q)})$, $\Sigma=(\sigma_{ij})$, 则

$$(n-p)\hat{\Sigma}=(Y-X\hat{B})^\top(Y-X\hat{B})=Y^\top(I_n-P_X)Y$$
$$=\mathcal{E}^\top(I_n-P_X)\mathcal{E}=\left(\boldsymbol{\varepsilon}_{(i)}^\top(I_n-P_X)\boldsymbol{\varepsilon}_{(j)}\right)$$

因为 $var(\boldsymbol{\varepsilon}_{(i)})=\sigma_{ii}I_n, cov(\boldsymbol{\varepsilon}_{(i)},\boldsymbol{\varepsilon}_{(j)})=\sigma_{ij}I_n$,

$$E\left(\boldsymbol{\varepsilon}_{(i)}^\top(I_n-P_X)\boldsymbol{\varepsilon}_{(j)}\right)=tr\left((I_n-P_X)E\left(\boldsymbol{\varepsilon}_{(j)}\boldsymbol{\varepsilon}_{(i)}^\top\right)\right)=\sigma_{ij}(n-p)$$

所以$E((n-p)\hat{\Sigma})=E(\mathcal{E}^\top(I_n-P_X)\mathcal{E})=(n-p)\Sigma$.

(2) $\mathrm{var}(\hat{\boldsymbol{\beta}}_{(i)}|X)=\mathrm{var}((X^\top X)^{-1}X^\top\mathbf{y}_{(i)}|X)$

$$=\mathrm{var}\left((X^\top X)^{-1}X^\top\boldsymbol{\varepsilon}_{(i)}|X\right)=\sigma_{ii}\,(X^\top X)^{-1}\text{。}$$

最小二乘估计

一元回归中误差方差σ^2不影响最小二乘，多元回归也是如此：$\hat{B}$极小化误差平方和$\|\mathcal{E}\|^2$, 它也使得 $\left\|\mathcal{E}\Sigma^{-1/2}\right\|^2$达到极小。

命题3. LS估计 $\hat{B}=(X^\top X)^{-1}X^\top Y$使得标准化误差平方和

$$\left\|\mathcal{E}\Sigma^{-1/2}\right\|^2=tr(\Sigma^{-1}\mathcal{E}^\top\mathcal{E})=tr(\Sigma^{-1}(Y-XB)^\top(Y-XB))$$

达到极小。

证明：

$$\begin{aligned}
&tr(\Sigma^{-1}(Y-XB)^\top(Y-XB))\\
&\quad=tr\left(\Sigma^{-1}(Y-X\hat{B}+X\hat{B}-XB)^\top(Y-X\hat{B}+X\hat{B}-XB)\right)\\
&\quad=tr\left(\Sigma^{-1}(Y-X\hat{B})^\top(Y-X\hat{B})\right)+tr\left(\Sigma^{-1}(X\hat{B}-XB)^\top(X\hat{B}-XB)\right)\\
&\quad\geq tr\left(\Sigma^{-1}(Y-X\hat{B})^\top(Y-X\hat{B})\right)
\end{aligned}$$

其中因为$X^\top\left(Y-X\hat{B}\right)=0$，交叉项

$$\begin{aligned}
tr\left(\Sigma^{-1}(X\hat{B}-XB)^\top(Y-X\hat{B})\right)&=tr\left(\Sigma^{-1}(X\hat{B}-XB)^\top(Y-X\hat{B})\right)\\
&=tr\left(\Sigma^{-1}\hat{B}^\top X^\top(Y-X\hat{B})\right)=0.
\end{aligned}$$

正态模型

命题4. 多元线性模型$Y=XB+\mathcal{E}$, $\text{vec}(\mathcal{E}^\top)\sim N_{nq}(0, I_n \otimes \Sigma)$，即$\mathcal{E}$的各行$\sim N_q(0,\Sigma)$, 则$B,\Sigma$的极大似然估计

$$\hat{B}=(X^\top X)^{-1}X^\top Y,\quad \tilde{\Sigma}=\frac{1}{n}(Y-X\hat{B})^\top\left(Y-X\hat{B}\right)$$

且$n\tilde{\Sigma}=(n-p)\hat{\Sigma}=(Y-X\hat{B})^\top\left(Y-X\hat{B}\right)\sim W_q(n-p,\Sigma)$。

证明： $\mathbf{y}_i|\mathbf{x}_i\sim N_q(B^\top\mathbf{x}_i,\Sigma)$，似然函数

$$L(B,\Sigma)=p(\mathbf{y}_1,\dots,\mathbf{y}_n)$$

$$=\frac{C}{|\Sigma|^{n/2}}\prod_{i=1}^n \exp(-(\mathbf{y}_i-B^\top\mathbf{x}_i)^\top\Sigma^{-1}(\mathbf{y}_i-B^\top\mathbf{x}_i)/2)$$

$$=\frac{C}{|\Sigma|^{n/2}}\exp\left(-\frac{1}{2}\,tr((Y-XB)\Sigma^{-1}(Y-XB)^\top)\right)$$

由命题3， $\hat{B}$极小化$tr((Y-XB)\Sigma^{-1}(Y-XB)^\top)$，代入$\hat{B}$

$$\log L\left(\hat{B},\Sigma\right)=-\frac{n}{2}\log(|\Sigma|)-\frac{1}{2}tr\left(\Sigma^{-1}\left(Y-X\hat{B}\right)^\top\left(Y-X\hat{B}\right)\right)$$

$$=-\frac{n}{2}\log(|\Sigma|)-\frac{n}{2}tr\left(\Sigma^{-1}\tilde{\Sigma}\right)=\frac{n}{2}\{\log(|\Omega|)-tr\left(\Omega\tilde{\Sigma}\right)\},\quad \Omega=\Sigma^{-1}$$

完全同第9讲命题2， Σ的最优解为$\tilde{\Sigma}$。易证$n\tilde{\Sigma}\sim W_q(n-p,\Sigma)$。

Wilks 检验

- 多元回归模型的假设检验问题很多可转化为一元模型的检验问题（比如检验特定自变量是否对某个响应有显著影响）。
- 可利用命题4构造精确检验，但于复杂且精确分布不是通常的 F ，t，χ^2，一般我们使用近似检验方法，即似然比检验。

Wilks检验：考虑回归系数B的某个零假设H_0，似然函数：

$$L(B,\Sigma) \quad = \frac{C}{|\Sigma|^{n/2}} \exp\left(-\frac{1}{2}\, tr((Y-XB)\Sigma^{-1}(Y-XB)^{\top})\right)$$

我们已知

$$\max L(B,\Sigma) = L\left(\hat{B},\tilde{\Sigma}\right) = C|\tilde{\Sigma}|^{-n/2}$$

假设H_0下B的极大似然估计为$\hat{B}_0$, 完全同命题4的证明过程可得

$$\log L\left(\hat{B}_0,\Sigma\right) = \frac{n}{2}\{\log(|\Omega|) - tr\left(\Omega\tilde{\Sigma}_0\right)\}，\ \tilde{\Sigma}_0 = \frac{1}{n}(Y-X\hat{B}_0)^{\top}\left(Y-X\hat{B}_0\right)$$

从而

$$\max_{H_0} L(B,\Sigma) = L\left(\hat{B}_0,\tilde{\Sigma}_0\right) = C|\tilde{\Sigma}_0|^{-n/2},$$

$$\Rightarrow \Lambda = \frac{\max_{H_0} L(B,\Sigma)}{\max L(B,\Sigma)} = \frac{|\tilde{\Sigma}|^{n/2}}{|\tilde{\Sigma}_0|^{n/2}}，\ \text{Wilks 统计量}\Lambda^* = \Lambda^{2/n} = \frac{|\tilde{\Sigma}|}{|\tilde{\Sigma}_0|}$$

$$-2\log(\Lambda) = -n\log(\Lambda^*) = -n\log\left(\frac{|\tilde{\Sigma}|}{|\tilde{\Sigma}_0|}\right) \to \chi_k^2, pq - \dim(\Theta).$$

例如，Johnson&Wichern教材中的Result 7.11给出了部分回归系数的似然比检验：

$$Y = XB + \mathcal{E} = (X_1, X_2)\begin{pmatrix} B_1 \\ B_2 \end{pmatrix} + +\mathcal{E} = X_1 B_1 + X_2 B_2 + \mathcal{E},$$

$$\mathrm{vec}(\mathcal{E}^{\top}) \sim N_{nq}(0, I_n \otimes \Sigma)$$

$H_0: B_2 = 0$, 即$B = \begin{pmatrix} B_1 \\ 0 \end{pmatrix}$, Wilks lambda:

$$\Lambda^{*} = \frac{|\tilde{\Sigma}|}{|\tilde{\Sigma}_0|}$$

其中$\tilde{\Sigma}$，$\tilde{\Sigma}_0$分别是全模型和零模型下的误差方差矩阵的LS估计。

当$-n\log\left(\frac{|\tilde{\Sigma}|}{|\tilde{\Sigma}_0|}\right) > \chi^2_{\mathrm{df}}(\alpha)$时否定$H_0$， $df = kq = B_2$元素个数(B_2行数k)。

Bartlett校正：

$$-\left(n - p - \frac{q-k+1}{2}\right)\log\left(\frac{|\tilde{\Sigma}|}{|\tilde{\Sigma}_0|}\right) > \chi^2_{\mathrm{df}}(\alpha)\text{时否定}H_0$$

例3（续）. 假设三个响应y1,y2,y3和一个自变量x数据如下($n = 5$)，假设线性模型

x	y1	y2	y3
1	1	-1	0
2	4	-1	1
3	3	2	1
4	8	3	2
5	9	2	4

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \\ a_3 & b_3 \end{pmatrix} \begin{pmatrix} 1 \\ x \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix}$$

```
# LS估计
> fit1= lm( cbind(y1,y2,y3) ~ x)
Coefficients:
              y1    y2    y3
(Intercept) -1.0  -2.0  -1.1
x            2.0   1.0   0.9
```

回归系数LS估计$\hat{B}$

```
> summary(fit1)
⇔ summary(y1~x);
  summary(y2~x);
  summary(y3~x)
```

```
#假设检验
> fit0= lm( cbind(y1,y2,y3) ~ 1 ) #null，  自变量x不影响y1-y3
> anova(fit1, fit0） #检验H0: b1 = b2 = b3 = 0
```

$H_0: b_1 = b_2 = b_3 = 0$