多元线性回归

张伟平

zwp@ustc.edu.cn

Office: 东区管理科研楼 1006

Phone: 63600565

课件 http://staff.ustc.edu.cn/~zwp/

论坛 http://fisher.stat.ustc.edu.cn

简介

1.1 多重线性回归模型

- 回归分析是一类基于预测变量 (predictor variables)(a.k.a 解释变量, 自变量 (independent variables), 回归量 (regressors)) 来预测一个或多个响应变量 (response variable)(a.k.a 因变量 (dependent variable), 被解释变量 (explained variable), 回归应变量 (regressand)) 的统计方法
- 回归分析也可以用来评价解释变量对响应变量的作用, 常为解释变量的线性函数对响应变量的作用
- 解释变量可以为连续的或者离散的, 或者两者混合的
- 首先我们回顾一下用于一元响应变量的多重回归方法, 然后推广到响应变量是多维的.

Multiple Regression Analysis

- 假设解释变量为 $x_1, x_2, \ldots, x_{p-1}$, 这些变量认为是和响应变量 y 有关联
- 多重线性总体回归模型假设

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_{p-1} x_{p-1} + e$$

  - $\beta = (\beta_0, \beta_1, \ldots, \beta_{p-1})'$ 为 (固定的) 未知的参数向量
  - $x_1, \ldots, x_{p-1}$ 称为解释变量, 其可以为固定的 (设计的), 或者随机的.
  - e 称为随机误差项, 一般假设 $e \sim (0, \sigma^2)$, 且 $E(ex_i) = 0, i = 1, \ldots, p-1$.
- 当我们对总体进行随机抽样时候, 假设有 n 个个体, 每个个体有模型

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_{p-1} x_{i(p-1)} + e_i$$

表示成矩阵形式后有

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1(p-1)} \\ 1 & x_{21} & \cdots & x_{2(p-1)} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & \cdots & x_{n(p-1)} \end{pmatrix} \beta + \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}$$

$$\Longleftrightarrow Y_{n\times 1} = X_{n\times p}\beta + \epsilon$$

其中 $y_i, x_{i1}, \ldots, x_{i(p-1)}$ 表示对总体变量 $y, x_1, \ldots, x_{p-1}$ 的独立重复观测. 按照总体模型假设和抽样方式, 一般假设误差有下述性质:

- $Ee_i = 0$
- $Var(e_i) = \sigma^2$(常数)
- $Cov(e_i, e_j) = 0, i \neq j$

多重线性回归模型(Multiple linear regression)

$$Y_{n\times 1} = X_{n\times p}\beta + \epsilon$$

以及假设 $E\epsilon = 0, Var(\epsilon) = \sigma^2 I_n.$

例将下述单因素方差分析模型表示成回归模型的形式:

$$y_{ij} = \mu + \tau_i + e_{ij}, j = 1, \ldots, n_i, i = 1, 2, 3$$

此时有三个总体, 因此引入哑变量 (dummy variable) 来处理, 令 $x_{ij} = 1$, 如果 $i = j$; 否则为 0. 从而一元方差分析模型可以表示成回归分析的形式

$$y_{ij} = \mu + \tau_1 x_{i1} + \tau_2 x_{i2} + \tau_3 x_{i3} + e_{ij}$$

回归模型的推断

对多重回归模型, 一般关心的任务有:

- 参数及其函数的估计问题 (可估性, 最小二乘估计)
- 参数估计量的性质
- 模型诊断方面
  - 参数的检验问题 (正态性假设, 似然比检验)
  - 变量选择问题
  - 残差分析 (模型假设的检查, 数据清洁)
- 模型的预测功能

最小二乘估计

- 对 β 的估计方法之一是选择使得残差平方和达到最小:

$$\hat{\beta} = \arg\min_{\beta \in R^p} \|Y - X\beta\|^2 = \arg\min_{\beta \in R^p} \sum_{i=1}^{n} \left(y_i - \sum_{k=0}^{p-1} \beta_k x_{ik} \right)^2$$

其中 $x_{i0} \equiv 1$.

- 当 X 为满秩时候 $(p < n)$, 上述最小化残差平方和的 β 可以得出

$$\hat{\beta} = (X'X)^{-1} X'Y$$

  – 此时称 $\hat{\beta}$ 为 β 的最小二乘估计.

  – 此时响应变量的拟合值为 $\hat{Y} = X\hat{\beta} = X(X'X)^{-1}X'Y = HY$, H 称为帽子 (Hat) 矩阵.

  – 残差为 $\hat{\epsilon} = Y - \hat{Y} = (I - H)Y$

– 残差满足 $\hat{\epsilon}'X = 0, \hat{\epsilon}'\hat{Y} = 0.$

平方和分解

由于 $\hat{\epsilon}'\hat{Y} = 0$, 因此总的响应变量平方和 $Y'Y$ 可以分解为

$$Y'Y = (\hat{Y} + Y - \hat{Y})'(\hat{Y} + Y - \hat{Y}) = \hat{Y}'\hat{Y} + \hat{\epsilon}'\hat{\epsilon}$$

由于 X 的第一列为 $\mathbf{1}$, 因此 $X'\hat{\epsilon} = 0$ 表明 $\mathbf{1}'\hat{\epsilon} = 0 \Longrightarrow \mathbf{1}'Y = \mathbf{1}'\hat{Y}$. 从而两边同时减去 $\mathbf{1}'Y$ 和 $\mathbf{1}'\hat{Y}$ 得到

$$Y'Y - \mathbf{1}'Y = \hat{Y}'\hat{Y} - \mathbf{1}'\hat{Y} + \hat{\epsilon}'\hat{\epsilon}$$

$$\Longleftrightarrow \sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(\hat{y}_j - \bar{y})^2 + \sum_{i=1}^{n}\hat{\epsilon}_i^2$$

$$SST = SS_{reg} + SS_e$$

总平方和 = 回归平方和 + 残差平方和

总波动性 = 回归能解释的波动性 + 误差的波动性

由此分解, 模型拟合程度的一个度量标准为

$$R^2 = 1 - \frac{SS_e}{SST} = \frac{SS_{reg}}{SST}$$

称为判定系数 (coefficient of determination). R 即为总体多重相关系数的估计 (参看第四讲条件分布部分).

最小二乘估计的性质

- $E\hat{\beta} = \beta, cov(\hat{\beta}) = \sigma^2(X'X)^{-1}$
- $cov(\hat{\beta}, \hat{\epsilon}) = 0$
- $c'\beta$ 的最佳线性估计估计为 $c'\hat{\beta}$.(Gauss-Markov 定理)

1.2　多元线性回归模型

当响应变量为多元时候, 不妨设 m 个响应变量, $Y_1, \ldots, Y_m$, 解释变量为 $x_1, \ldots, x_{p-1}$, 考虑解释变量与响应变量之间的关系, 假设有如下总体回归模型:

$$
\begin{aligned}
Y_1 &= \beta_{01} + \beta_{11}x_1 + \cdots + \beta_{(p-1)1}x_{p-1} + e_1 \\
Y_2 &= \beta_{02} + \beta_{12}x_1 + \cdots + \beta_{(p-1)2}x_{p-1} + e_2 \\
&\vdots \\
Y_m &= \beta_{0m} + \beta_{1m}x_1 + \cdots + \beta_{(p-1)m}x_{p-1} + e_m
\end{aligned}
$$

也就是假设每个指标 Y_i 和解释变量之间存在线性关系. 误差项 $\mathbf{e} = [e_1, e_2, \ldots, e_m]'$ 满足假设

$$E\mathbf{e} = 0, \qquad Cov(\mathbf{e}) = \Sigma = (\sigma_{ij})$$

当对总体中 n 个个体观测时候, 记第 j 次观测样本的解释变量为 $x_{j1}, x_{j2}, \ldots, x_{j(p-1)}$, 而响应变量记为 $\mathbf{y}_j = [y_{j1}, y_{j2}, \ldots, y_{jm}]'$, $j = 1, \ldots, n$. 使用矩阵表达, 则

$$\mathbf{Y}_{n\times m} = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1m} \\ y_{21} & y_{22} & \cdots & y_{2m} \\ \vdots & & & \\ y_{n1} & y_{n2} & \cdots & y_{nm} \end{bmatrix} = [\mathbf{y}_{(1)}, \mathbf{y}_{(2)}, \cdots, \mathbf{y}_{(m)}]$$

$$X_{n\times p} = \begin{bmatrix} x_{10} & x_{11} & \cdots & x_{1(p-1)} \\ x_{20} & x_{21} & \cdots & x_{2(p-1)} \\ \vdots & & & \\ x_{n0} & x_{n1} & \cdots & x_{n(p-1)} \end{bmatrix}$$

其中 $x_{i0} \equiv 1, i = 1, \ldots, n$.

参数矩阵和随机测量误差记为

$$B_{p\times m} = \begin{bmatrix} \beta_{01} & \beta_{02} & \cdots & \beta_{0m} \\ \beta_{11} & \beta_{12} & \cdots & \beta_{1m} \\ \vdots & & & \\ \beta_{(p-1)1} & \beta_{(p-1)2} & \cdots & \beta_{(p-1)m} \end{bmatrix} = [\beta_{(1)}, \beta_{(2)}, \cdots, \beta_{(m)}]$$

$$\boldsymbol{\epsilon}_{n\times m} = \begin{bmatrix} \epsilon_{11} & \epsilon_{12} & \cdots & \epsilon_{1m} \\ \epsilon_{21} & \epsilon_{22} & \cdots & \epsilon_{2m} \\ \vdots & & & \\ \epsilon_{n1} & \epsilon_{n2} & \cdots & \epsilon_{nm} \end{bmatrix} = [\epsilon_{(1)}, \epsilon_{(2)}, \cdots, \epsilon_{(m)}] = \begin{pmatrix} \epsilon_1' \\ \epsilon_2' \\ \vdots \\ \epsilon_n' \end{pmatrix}$$

从而, **多元线性回归模型**(Multivariate linear model) 的矩阵表达:

$$\begin{aligned} \mathbf{Y}_{n\times m} &= X_{n\times p} B_{p\times m} + \boldsymbol{\epsilon}_{n\times m} \\ &= [X\beta_{(1)}, \cdots, X\beta_{(m)}] + [\epsilon_{(1)}, \epsilon_{(2)}, \cdots, \epsilon_{(m)}] \end{aligned}$$

其中 $E\epsilon_{(i)}=0, Cov(\epsilon_{(i)},\epsilon_{(j)})=\sigma_{ij}I_n, i,j=1,2,\ldots,m$. 虽然对第 k 个观测的测量误差 ϵ_k 有协方差矩阵 Σ, 但对不同个体的观测值不相关.

从上面的模型中可以看出对第 i 个响应 $\mathbf{y}_{(i)}$, 其服从线性回归模型

$$\mathbf{y}_{(i)}=X\beta_{(i)}+\epsilon_{(i)}$$

其中 $E\epsilon_{(i)}=0, Cov(\epsilon_{(i)})=\sigma_{ii}I_n, i=1,\ldots,m$.

- **对第 i 个响应变量来说, n 次观测之间不相关**
- **不同响应变量的观测之间存在相关**

基于第 i 个响应变量 $\mathbf{y}_{(i)}$ 的回归模型可得 $\beta_{(i)}$ 的最小二乘估计

$$\hat{\beta}_{(i)}=(X'X)^{-1}X'\mathbf{y}_{(i)}$$

将这些估计量放在一起组成矩阵, 我们有

$$\begin{aligned}\hat{B} &= [\hat{\beta}_{(1)}, \hat{\beta}_{(2)}, \ldots, \hat{\beta}_{(m)}] = (X'X)^{-1}X'[\mathbf{y}_{(1)}, \mathbf{y}_{(2)}, \cdots, \mathbf{y}_{(m)}] \\ &= (X'X)^{-1}X'\mathbf{Y}\end{aligned}$$

最小二乘估计

定理 1. 设 X 为满秩的, 则 $\hat{B}$ 为 B 的最小二乘估计.

证明. 利用拉直运算表达, 多元回归模型可以表示为

$$vec(\mathbf{Y}') = (X \otimes I_m)vec(B') + vec(\boldsymbol{\epsilon}')$$

其中 $Cov(\boldsymbol{\epsilon}') = I_n \otimes \Sigma$.

由上述拉直向量化模型表示, 根据最小二乘方法, 若记

$$Q(vec(B')) = \|vec(Y') - (X \otimes I_m)vec(B')\|^2$$

则 $vec(B')$ 的最小二乘估计为

$$\begin{aligned}\widehat{vec(B')} &= [((X \otimes I_m))'(X \otimes I_m)]^{-1}(X \otimes I_m)'vec(\mathbf{Y}') \\ &= [(X'X)^{-1}X' \otimes I_m]vec(\mathbf{Y}')\end{aligned}$$

于是, 将估计重新表示成矩阵形式即证. □

定理 2. 设 $\hat{B}$ 为 B 的最小二乘估计, 则

1. $\hat{B}$ 为 B 的无偏估计.
2. $Cov(\hat{\beta}_{(i)}, \hat{\beta}_{(k)}) = \sigma_{ik}(X'X)^{-1}, i, k = 1, \ldots, p.$
3. Σ 的无偏估计为

$$\hat{\Sigma} = \frac{1}{n-p}\mathbf{Y}'P\mathbf{Y}.$$

其中 $P = I_n - X(X'X)^{-1}X' = I_n - H.$

证明. (1) 显然. 对 (2),

$$\begin{aligned}
Cov(\hat{\beta}_{(i)}, \hat{\beta}_{(k)}) &= (X'X)^{-1}X'Cov(\epsilon_{(i)}, \epsilon_{(k)})X(X'X)^{-1} \\
&= \sigma_{ik}^2(X'X)^{-1}.
\end{aligned}$$

下证 (3). 注意到 $PX = X'P = 0$, 从而

$$\mathbf{Y}'P\mathbf{Y} = (\mathbf{Y} - XB)'P(\mathbf{Y} - XB)$$

因此

$$\begin{aligned}
(E\mathbf{Y}'P\mathbf{Y})_{ij} &= (E[(\mathbf{Y} - XB)'P(\mathbf{Y} - XB)])_{ij} \\
&= E[(\mathbf{y}_{(i)} - X\beta_{(i)})'P(\mathbf{y}_{(j)} - X\beta_{(j)})] \\
&= tr\{PE[\epsilon_{(j)}\epsilon'_{(i)}]\} \\
&= tr\{P(\sigma_{ji}I_n)\} = (n-p)\sigma_{ij}.
\end{aligned}$$

从而得证. □

1.2.1 最小二乘估计的性质

- 使用 B 的最小二乘估计, 我们可以得到

$$\hat{\mathbf{Y}} = X\hat{B} = X(X'X)^{-1}X'\mathbf{Y}$$
$$\hat{\boldsymbol{\epsilon}} = \mathbf{Y} - \hat{\mathbf{Y}} = [I - X(X'X)^{-1}X']\mathbf{Y} = P\mathbf{Y}$$

- 使用平方和与交叉积分解有

$$\underbrace{\mathbf{Y}'\mathbf{Y}}_{totalSSCP} = \underbrace{\hat{\mathbf{Y}}'\hat{\mathbf{Y}}}_{RegSSCP} + \underbrace{\hat{\boldsymbol{\epsilon}}'\hat{\boldsymbol{\epsilon}}}_{ErrorSSCP}$$

- 估计的残差 $\hat{\epsilon}_{(i)}$ 满足

$$E\hat{\epsilon}_{(i)} = 0, \qquad E\hat{\epsilon}'_{(i)}\hat{\epsilon}_{(j)} = (n-p)\sigma_{ij}$$

从而

$$E\hat{\boldsymbol{\epsilon}} = 0, \quad E\hat{\boldsymbol{\epsilon}}'\hat{\boldsymbol{\epsilon}} = (n-p)\Sigma$$

- $\hat{B}$ 和 $\hat{\boldsymbol{\epsilon}}$ 不相关

若假定

$$\boldsymbol{\epsilon} \sim N_{n\times m}(0, I_n \otimes \Sigma), \Sigma_{m\times m} > 0$$

则

$$\mathbf{Y} \sim N_{n\times m}(XB, I_n \otimes \Sigma)$$

从而导出 B 和 Σ 的最大似然估计.

定理 3. 设 $\mathbf{Y} \sim N_{n\times m}(XB, I_n \otimes \Sigma)$, 其中 X 满秩, $B \in R^{p\times m}$ 和 $\Sigma > 0$, 则 B 和 Σ 的最大似然估计为

$$\hat{B}^* = (X'X)^{-1}X'\mathbf{Y}$$
$$\hat{\Sigma}^* = \frac{1}{n}\mathbf{Y}'P\mathbf{Y}$$

证明. 由矩阵多元正态分布的定义, 知对数似然函数

$$l(B,\Sigma) = -\frac{1}{2}nln|2\pi\Sigma| - \frac{1}{2}tr[(\mathbf{Y}-XB)\Sigma^{-1}(\mathbf{Y}-XB)']$$

从而

$$
\begin{aligned}
& tr[(\mathbf{Y} - XB)\Sigma^{-1}(\mathbf{Y} - XB)'] \\
& = tr[\Sigma^{-1}(\mathbf{Y} - X\hat{B})'(\mathbf{Y} - X\hat{B})] + tr[\Sigma^{-1}(\hat{B} - B)'X'X(\hat{B} - B)] \\
& \geq tr[\Sigma^{-1}(\mathbf{Y} - X\hat{B})'(\mathbf{Y} - X\hat{B})] \\
& = tr[\Sigma^{-1}\mathbf{Y}'P\mathbf{Y}] = tr[\Sigma^{-1}\hat{\Sigma}^*]
\end{aligned}
$$

等号成立当且仅当 $B = \hat{B}$, 于是

$$
\max_{\Sigma>0,B} l(B,\Sigma) = \max_{\Sigma>0} l(\hat{B},\Sigma) = \max_{\Sigma>0}\{-\frac{1}{2}nlog|\Sigma| - \frac{1}{2}ntr[\Sigma^{-1}\hat{\Sigma}^*]\}
$$

最后一步容易得到最大值在 $\Sigma = \hat{\Sigma}^*$ 处达到. 从而得证. □

注 B 的最大似然估计与最小二乘估计相同, 即 $\hat{B}^* = \hat{B}$; Σ 的最大似然估计不是无偏估计.

定理 4. 对多元线性回归模型, 若 $\boldsymbol{\epsilon} \sim N_{n\times m}(0, I_n \otimes \Sigma)$, 最小二乘估计 $\hat{B}$ 和无偏估计 $\hat{\Sigma}$ 由前给出, 则

(1) $\hat{B} \sim N_{p\times m}(B, (X'X)^{-1} \otimes \Sigma)$*;*

(2) $\hat{B}$ *和* $\hat{\Sigma}$ *相互独立;*

(3) $(n-p)\hat{\Sigma} \sim W_m(n-p, \Sigma)$*.*

证明. (1) 由假设知 $\mathbf{Y} \sim N_{n\times m}(XB, I_n \otimes \Sigma)$, 从而由矩阵正态分布的性质知 $\hat{B} = (X'X)^{-1}X'\mathbf{Y} \sim N_{p\times m}(B, (X'X)^{-1} \otimes \Sigma)$.

(2) 由多元正态的性质 (第四讲定理 9), 注意到

$$vec(\hat{B}) = [I \otimes (X'X)^{-1}X']vec(\mathbf{Y})$$
$$vec(P\mathbf{Y}) = [I \otimes P]vec(\mathbf{Y})$$

从而 $\hat{B}$ 和 $P\mathbf{Y}$ 相互独立 $\Leftrightarrow$ $vec(\hat{B})$ 和 $vec(P\mathbf{Y})$ 相互独立 $\Leftrightarrow$ $[I \otimes (X'X)^{-1}X'][I \otimes P]' = I \otimes (X'X)^{-1}X'P' = 0$.

(3) 因为 $\mathbf{Y} \sim N_{n\times m}(X\hat{B}, I \otimes \Sigma)$, 以及 $Rank(P) = n-p$, 所以由 Wishart 分布的性质 (第五讲 Cochran 定理) 知 $(n-1)\hat{\Sigma} = \mathbf{Y}'P\mathbf{Y} \sim W_m(n-p, \Sigma)$. □

定理 5 (Gauss-Markov 定理). 记 $\Theta = XB = [\theta_{(1)}, \ldots, \theta_{(m)}]$, 令 $\phi = \sum_{j=1}^m c_j' \theta_{(j)}$, 这里 $c_1, \ldots, c_m$ 是任意 m 个 n 维常数向量, 记 $\hat{\mathbf{Y}} = X\hat{B} = [\hat{\mathbf{y}}_{(1)}, \ldots, \hat{\mathbf{y}}_{(m)}]$, $\hat{\phi} = \sum_{j=1}^m c_j' \hat{\mathbf{y}}_{(j)}$, 则 $\hat{\phi}$ 为 ϕ 的最佳线性无偏估计.

证明. 首先我们来证明 $\hat{\phi}$ 的线性无偏性. 无偏性显然, 另外注意到 $\hat{\mathbf{Y}} = H\mathbf{Y} = [H\mathbf{y}_{(1)}, \ldots, H\mathbf{y}_{(m)}]$, 其中 $H = X(X'X)^- X'$. 因此

$$\hat{\phi} = \sum_{j=1}^m c_j' H \mathbf{y}_{(j)}$$

即 $\hat{\phi}$ 为 $\mathbf{Y}$ 的列向量的线性函数.

下面我们讨论 $\hat{\phi}$ 在所有线性无偏估计里方差最小. 假设 $\phi^* = \sum_{j=1}^m d_j' \mathbf{y}_{(j)}$ 为 ϕ 的任意线性无偏估计, 则

$$E\phi^* = \sum_{j=1}^m d_j' \theta_{(j)} = \phi = \sum_{j=1}^m c_j' \theta_{(j)}$$

因此

$$\sum_{j=1}^{m}(d_j - c_j)'\theta_{(j)} = 0, \quad 对任意\ \theta_{(j)} \in \mathcal{L}(X)$$

从而

$$(d_j - c_j)'\theta_{(j)} = 0, \quad 对任意\ \theta_{(j)} \in \mathcal{L}(X)$$

这等价于

$$H(d_j - c_j) = 0 \Leftrightarrow Hd_j = Hc_j, j = 1, \ldots, m$$

于是由 $cov(\mathbf{y}_{(j)}, \mathbf{y}_{(k)}) = \sigma_{jk} I_n$ 有

$$\begin{aligned} Var(\phi^*) - Var(\hat{\phi}) &= \sum_{j=1}^{m}\sum_{k=1}^{m} d_j'(I_n - H)d_k\sigma_{jk} \\ &= tr[D'(I_n - H)D\Sigma] \geq 0 \end{aligned}$$

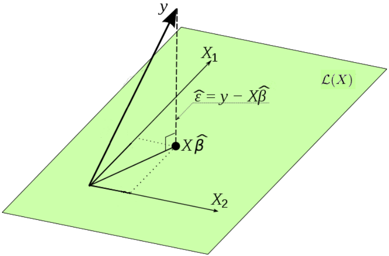
其中 $D = [d_1, \ldots, d_m]$, 等号成立当且仅当 $D(I_n - H) = 0$, 即 $d_j = Hd_j = Hc_j, j = 1, \ldots, m$. 这时, $\hat{\phi}^* = \hat{\phi}$. □

1.2.2 最小二乘估计的几何解释

- 对一元多重线性回归模型

$$y_{n\times 1} = X_{n\times p}\beta_{p\times 1} + e_{n\times 1}$$

我们知道, y 向 X 的列向量所张成的线性子空间 $\mathcal{L}(X)$ 的投影为 $\hat{y} = X\hat{\beta} = Hy$. 而 $y - Hy = (I - H)y = \hat{e}$ 为残差, 其中 $H = X(X'X)^{-1}X$. 如下图所示



- 由前面最小二乘估计的构造知，将 $\mathbf{Y}$ 的每一列 $\mathbf{y}_{(i)}$ 向空间 $\mathcal{L}(X)$ 上投影，得到

$$\hat{\mathbf{Y}} = X\hat{B}, \qquad \mathbf{Y} - \hat{\mathbf{Y}} = (I_n - P)\mathbf{Y} = \hat{\boldsymbol{\epsilon}}$$

1.2.3 有线性约束的线性模型

考虑带有线性约束条件的多元回归模型:

$$\mathbf{Y} = XB + \boldsymbol{\epsilon}, Rank(X) = p$$
$$AB = C, A_{s\times p}, C_{s\times m}, Rank(A) = s < m$$
$$\boldsymbol{\epsilon}_{n\times m} \sim (0, I_n \otimes \Sigma)$$

为在此约束条件下求 B 的最小二乘估计,

- 记 $\hat{B}_H$ 为此线性约束条件下 B 的最小二乘估计.
- $\hat{B} = (X'X)^{-1}X'\mathbf{Y}$ 为没有线性约束时候 B 的最小二乘估计.

- 可以使用 Lagrange 乘子法和最小二乘方法进行直接求解. 但习惯上常常将约束条件解出来, 代入原线性模型中, 这样将带约束的回归模型转换为不带约束的回归模型.

显然约束条件方程 $AB = C$ 的通解可以表示为

$$B = A^{-}C + (I - A^{-}A)\mathbf{z}, \quad \mathbf{z}\text{为任意向量}$$

其中 A^{-} 为任意固定的广义逆, 特别, 取 $A^{-} = (X'X)^{-1}A'(A(X'X)^{-1}A')^{-1}$. 代入到回归模型中有

$$\mathbf{Y} - XA^{-}C = X(I - A^{-}A)\mathbf{z} + \boldsymbol{\epsilon}$$

因此 $\mathbf{z}$ 的最小二乘估计为

$$\hat{\mathbf{z}} = [(I - A^{-}A)'X'X(I - A^{-}A)]^{-1}(I - A^{-}A)'X'(\mathbf{Y} - XA^{-}C)$$

从而 B 的最小二乘估计为

$$\begin{aligned}\hat{B}_H &= A^- C + (I - A^- A)[(I - A^- A)' X' X (I - A^- A)]^{-1} \\ &\quad \cdot (I - A^- A)' X' (\mathbf{Y} - X A^- C) \\ &= \hat{B} - (X'X)^{-1} A' (A (X'X)^{-1} A')^{-1} (A\hat{B} - C)\end{aligned}$$

此时, Σ 的无偏估计为

$$\hat{\Sigma}_H = \frac{1}{n - p + s} (\mathbf{Y} - X\hat{B}_H)' (\mathbf{Y} - X\hat{B}_H)$$

证明. 证明作为补充作业. $\square$

1.2.4 预测

- 给定一组协变量值 $\mathbf{x}_0' = [1, x_{01}, \ldots, x_{0(p-1)}]$, 则 $\mathbf{x}_0' B$ 的预测值为 $\hat{Y}_0' = \mathbf{x}_0' \hat{B}$, 其中 $\hat{B}$ 为 B 的最小二乘估计.

- $\hat{Y}_0'$ 为 $\mathbf{x}_0' B$ 的无偏估计, 即 $E\hat{Y}_0' = \mathbf{x}_0' E\hat{B} = \mathbf{x}_0' B$.

- 估计误差 $\mathbf{x}_0'\hat{B} - \mathbf{x}_0' B$ 第 i 个和第 k 个分量的协方差为

$$E[\mathbf{x}_0'(\hat{\beta}_{(i)} - \beta_{(i)})(\hat{\beta}_{(i)} - \beta_{(i)})'\mathbf{x}_0] = \sigma_{ik}\mathbf{x}_0'(X'X)^{-1}\mathbf{x}_0.$$

- 预报误差 $Y_0' - \hat{Y}_0'$ 的第 i 个和第 k 个分量的协方差为

$$\begin{aligned}
&E(Y_{0i} - \mathbf{x}_0'\hat{\beta}_{(i)})(\hat{Y}_{0k} - \mathbf{x}_0'\hat{\beta}_{(k)}) \\
&\quad = E(\epsilon_{0i} - \mathbf{x}_0'(\hat{\beta}_{(i)} - \beta_{(i)}))(\epsilon_{0k} - \mathbf{x}_0'(\hat{\beta}_{(k)} - \beta_{(k)})) \\
&\quad = E(\epsilon_{0i}\epsilon_{0k} + \mathbf{x}_0' E(\hat{\beta}_{(i)} - \beta_{(i)})(\hat{\beta}_{(k)} - \beta_{(k)})'\mathbf{x}_0 \\
&= \sigma_{ik}(1 + \mathbf{x}_0'(X'X)^{-1}\mathbf{x}_0)
\end{aligned}$$