第七章　线性回归模型

§7.1　引　言

线性回归模型是现代统计学中应用最为广泛的模型之一，它也是其它统计模型研究或应用的基础。这主要有下列几个原因：

1. 在实际问题中，变量之间的关系常具有线性或近似线性的依赖关系。

2. 在现实世界中，虽然许多变量间的关系是非线性的，但经过适当的变换，将会成为线性关系。

3. 线性关系是变量之间最简单的关系，容易处理，理论和方法比较完善，这些为实际应用提供了有效算法。

本节将通过实例说明线性统计模型的背景和分类。

一、一元线性回归模型

变量之间的关系大致可分为确定性关系和非确定性关系两大类，数理统计是处理非确定性变量统计规律性的学科。线性回归模型是非确定性(具有随机性)变量之间关系的最基本的模型之一，如人的体重(Y) 与身高(X) 之间有一定的相依关系：当 X 大时，Y 也倾向于大，但 X 不能严格决定 Y。小麦产量(Y) 与小麦品种(X_1)、施肥量(X_2) 和浇水量(X_3) 有一定的关系，但还不能严格利用数学函数关系表达它们之间的关系。

以上例子中，通常称 Y 为因变量或响应变量，称 X 为自变量。Y 的值有两部分组成：一部分是能够由 X 决定的部分，它是 X 的函数，记为 $f(X)$；另一部分是由其它众多未加考虑的因素产生的影响，称为随机误差，故有：

$$Y = f(X) + e, \tag{7.1.1}$$

这里 e 作为随机误差，假定 $E(e) = 0$。特别，当 $f(X)$ 是线性函数时，$f(X) = \beta_0 + \beta_1 X$，则有

$$Y = \beta_0 + \beta_1 X + e. \tag{7.1.2}$$

(7.1.2) 式称为线性回归模型或线性回归方程，其中 β_0 和 β_1 未知，常数项 β_0 是回归直线 $y = \beta_0 + \beta_1 X$ 的截距，β_1 是斜率。

设有一组样本$(x_i, y_i), i = 1, 2\cdots n$，将上述模型用样本表示为

$$y_i = \beta_0 + \beta_1 x_i + e_i,\ i = 1, 2\cdots n. \tag{7.1.3}$$

e_i 为随机误差。若用适当的估计方法求得 β_0, β_1 的估计为 $\widehat{\beta}_0$ 和 $\widehat{\beta}_1$，代入到(7.1.2)中将误差项 e_i 用其均值 0 代替，得到

$$Y = \widehat{\beta}_0 + \widehat{\beta}_1 X, \tag{7.1.4}$$

称为经验回归方程，它是由 n 组样本观察值获得的。如果经检验，是合适的回归方程，则(7.1.4)刻画了 Y 与 X 之间的相关关系。

例7.1.1 设身高(X) 与体重(Y) 之间有近似回归关系(7.1.2)，e 表示除了身高 X，所有影响体重(Y) 的其他因素(如遗传、饮食、锻炼等)，假定调查了 n 个人的身高和体重得样本 (x_i, y_i), $i = 1, 2, \cdots, n$，估计 β_0 和 β_1 得 $\widehat{\beta}_0 = -40, \widehat{\beta}_1 = 0.6$，则经验回归方程为

$$Y = -40 + 0.6X. \tag{7.1.5}$$

如果甲身高160 cm，算得体重 $y_0 = 56\,\mathrm{kg}$，称 $y_0 = 56$ 为身高160cm的体重的预测值。

二、多元线性回归模型

实际问题中影响因变量的自变量往往不止一个，如有 $X_1, X_2, \cdots, X_{p-1}$，则它们有如下线性关系：

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_{p-1} X_{p-1} + e, \tag{7.1.6}$$

若有样本 $(x_{i1}, x_{i2}, \cdots, x_{ip-1}, y_i)$, $i = 1, 2, \cdots, n$，则有

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_{p-1} x_{ip-1} + e_i, \tag{7.1.7}$$

e_i 为随机误差，将上述方程组用矩阵表：

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1,p-1} \\ 1 & x_{21} & x_{22} & \cdots & x_{2,p-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n,1} & x_{n,2} & \cdots & x_{n,p-1} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix},$$

即：

$$y_{n\times 1} = X_{n\times p}\beta_{p\times 1} + e_{n\times 1}, \tag{7.1.8}$$

其中 y 为观测向量，X 称为设计阵(习惯称法)，β 为未知回归参数向量，e 是随机误差向量，关于 e 通常有两种假定：

(1) Gauss－Markov假定(简称G-M假定)：$E(e) = 0,\ Cov(e) = \sigma^2 I$, 即：

(a) $E(e_i) = 0,\ i = 1, 2, \cdots, n;$

(b) $Var(e_i) = \sigma^2,\ i = 1, 2, \cdots, n;$

(c) $Cov(e_i, e_j) = 0,\ i, j = 1, 2, \cdots, n,$ 且$i \neq j.$

(2) 正态假定：$e \sim N_n(0, \sigma^2 I)$, 即$e_1, \cdots, e_n$ 相互独立，具有相同分布$N(0, \sigma^2)$.

若利用样本对 $\beta_0, \beta_1, \cdots, \beta_{p-1}$ 作出估计，估计量为$\widehat{\beta}_0, \widehat{\beta}_1, \cdots, \widehat{\beta}_{p-1}$，则

$$Y = \widehat{\beta}_0 + \widehat{\beta}_1 X_1 + \cdots + \widehat{\beta}_{p-1} X_{p-1} \tag{7.1.9}$$

称为经验回归方程，它是否真正描述了 Y 和 $X_1, X_2, \cdots, X_{p-1}$ 之间的关系，还需要进行检验。

三、可化为线性模型的情形

有些模型表面上是非线性的，但是经过适当变换，可以化为线性模型，请看下例：

例7.1.2 在著名的经济学的Cobb-Duglas生产函数为：

$$Q_t = aL_t^b K_t^c, \tag{7.1.10}$$

其中 Q_t、L_t 和 K_t 分别表示为 t 年的产值、劳力投入和资金投入，a，b，c 为参数。表面上是(7.1.10)是非线性关系，若将两边取对数得

$$\ln Q_t = \ln a + b\ln L_t + c\ln K_t,$$

令 $\ln L_t = X_{t1}$, $\ln K_t = X_{t2}$, $y_t = \ln Q_t$, $\beta_0 = \ln a$, $\beta_1 = b$, $\beta_2 = c$，则有

$$y_t = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + e_t, \quad t = 1, 2, \cdots, T. \tag{7.1.11}$$

这就转化成线性模型的形式。

例7.1.3 多项式回归，因变量 Y 和自变量 X 之间具有下列关系

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \cdots + \beta_k X^k + e,$$

这是一个 k 次多项式，若令 $X_1 = X$, $X_2 = X^2, \cdots$, $X_k = X^k$，则

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + e,$$

就变为一个线性模型的形式。

注：“回归”一词的由来：英国生物统计学家Galton在研究人类遗传问题时提出“Regression”一词，他收集1078对父子身高数据，用 X —父亲身高，Y —儿子身高，单位：英寸。把(x_i, y_i)标在直角坐标纸上，大致成一直线，其规律大致：(1) 父亲身高 X 增加时，儿子身高 Y 也增加，这与常识一致；(2) 属于高个子的那类父亲的儿子的平均身高要比父亲的平均身高低，反之属于矮个子那类父亲的儿子的平均身高要比父亲的高。即反映了一个现象：身高超过平均高度(1078个父亲平均身高) $\overline{x} = 68$ 英寸的，他们的儿子的平均身高将低于父亲的身高；反之身高低于平均高度 $\overline{x} = 68$ 英寸的儿子的平均身高要高于父亲的平均身高。Golton解释：大自然有一种约束力，人的身高向中间值“回归”，不会两极分化。这就是所谓的回归效应。

四、应用

对回归模型所进行的统计分析，通常称为回归分析。回归分析的实际应用归纳起来主要有以下几个方面：

1. 描述变量之间的关系：找出对 Y 有重要相关关系的因变量，建立回归方程(变量选择—检验—诊断)；

2. 分析变量之间关系：通过对回归系数的估计，建立经验回归方程

$$Y = \widehat{\beta}_0 + \widehat{\beta}_1 X_1 + \cdots + \widehat{\beta}_{p-1} X_{p-1}.$$

回归系数 β_i的估计量$\hat{\beta}_i$ $(i=0,1,\cdots,p-1)$ 的大小在一定程度上反映了X_i 对Y_i 的影响的大小。另一方面，应用一些统计分析方法，还可以分析自变量之间存在的相关关系。

3. 预测：点预测、区间预测。

§7.2　　若 干 预 备 知 识

一、均值向量与协方差阵

定义7.2.1　设 $X=(X_1,X_2,\cdots,X_n)'$ 为随机向量，则称

$$E(X)=\big(E(X_1),E(X_2),\cdots,E(X_n)\big)'$$

为随机向量 X 的均值向量，称 $n\times n$ 阶对称阵

$$Cov(X)=E[(X-EX)(X-EX)']=\big(Cov(X_i,X_j)\big)_{n\times n}$$

为随机向量 X 的协方差阵，其中

$$Cov(X_i,X_j)=E(X_i-EX_i)(X_j-EX_j).$$

当 $i=j$ 时，$Cov(X_i,X_i)=Var(X_i)$.

定理7.2.1　设 X 和 b 分别为 $n\times 1$ 维和 $m\times 1$ 维的随机向量，A 是 $m\times n$ 阶的非随机矩阵，记 $Y=AX+b$，则

$$E(Y)=E(AX+b)=AE(X)+E(b).$$

证明： 设$A=(a_{ij})_{m\times n}$, $b=(b_1,b_2,\cdots,b_m)'$, $X=(X_1,X_2,\cdots,X_n)'$, $Y=(Y_1,Y_2,\cdots,Y_m)'$, 则由 $Y=AX+b$ 可知

$$\begin{aligned}Y_i&=\sum_{j=1}^{n}a_{ij}X_j+b_i\\ E(Y_i)&=\sum_{j=1}^{n}a_{ij}E(X_i)+E(b_i),\ \ i=1,2,\cdots,m.\end{aligned}$$

推论7.2.1　$tr[Cov(X)]=\sum\limits_{i=1}^{n}Var(X_i)$, 此处$trA$ 标识方阵A的迹。

定理7.2.2　设 $X_{n\times 1}$ 为随机向量，则有 $Cov(X)=\big(Cov(X_i,X_j)\big)_{n\times n}\geq 0$.

证明： 设 c 为任一非随机向量，按定义只要证明 $c'Cov(X)c\geq 0$. 记 $Y=c'X$，则

$$\begin{aligned}Var(Y)&=Var(c'X)=E[c'X-c'E(X)]^2\\ &=E[c'\big(X-E(X)\big)\big(X-EX\big)'c]\\ &=c'E[(X-EX)(X-EX)']c\\ &=c'Cov(X)c\geq 0,\end{aligned}$$

故知 $Cov(X) \geq 0$. □

定理7.2.3 设 A 为 $m \times n$ 阵，$X_{n\times 1}$ 为随机向量，$Y = AX$，则 $Cov(Y) = ACov(X)A'$.

证明：

$$\begin{aligned} Cov(Y) &= E[(AX - AEX)(AX - AEX)'] \\ &= AE[(X - EX)(X - EX)']A' \\ &= ACov(X)A'. \end{aligned}$$

□

定理7.2.4 设 X 和 Y 分别为 $n \times 1$ 维和 $m \times 1$ 维的随机向量，$A_{p\times n}$ 和 $B_{q\times m}$ 为常数阵，则 $Cov(AX, BY) = ACov(X, Y)B'$.

证明： 从定义出发。 □

二、随机向量的二次型

定义7.2.2 设 $X_{n\times 1} = (X_1, X_2, \cdots, X_n)'$ 为 $n \times 1$ 维随机向量，$A = (a_{ij})$ 为 $n \times n$ 对称阵，则

$$X'AX = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}X_iX_j$$

称为随机向量 X 的二次型。

如何求二次型的均值、方差, 我们有下述定理：

定理7.2.5 设随机向量 $X_{n\times 1}$ 有 $E(X) = \mu_{n\times 1}$，$Cov(X) = \Sigma_{n\times n}$，则

$$E(X'AX) = \mu'A\mu + tr(A\Sigma).$$

证明：

$$\begin{aligned} X'AX &= [(X - \mu) + \mu]'A[(X - \mu) + \mu] \\ &= (X - \mu)'A(X - \mu) + 2\mu'A(X - \mu) + \mu'A\mu, \end{aligned}$$

由于 $E[\mu'A(X - \mu)] = \mu'AE(X - \mu) = 0$, 故有

$$\begin{aligned} E(X'AX) &= E[(X - \mu)'A(X - \mu)] + \mu'A\mu \\ &= E\big[tr\big(A(X - \mu)(X - \mu)'\big)\big] + \mu'A\mu \\ &= trAE[(X - \mu)(X - \mu)'] + \mu'A\mu \\ &= tr[ACov(X)] + \mu'A\mu = tr(A\Sigma) + \mu'A\mu. \end{aligned}$$

特别：

(1) 当 $\mu = 0$ 时，$E(X'AX) = trA\Sigma$;

(2) 当 $\Sigma = \sigma^2 I$ 时，$E(X'AX) = \mu'A\mu + \sigma^2 trA$;

(3) 当 $\mu = 0, \Sigma = I$ 时，$E(X'AX) = trA$.

例7.2.1 设随机变量 X 为一维总体，$E(X) = \mu$，$Var(X) = D(X) = \sigma^2$，$X_1, X_2, \cdots, X_n$ 为从此总体中抽取的样本，求 $E(S^2)$，其中 $S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$.

解：将$Q=\sum_{i=1}^{n}(X_i-\overline{X})^2=(n-1)S^2$表示成$\mathbf{X}=(X_1,X_2,\cdots,X_n)'$的二次型，记$\mathbb{1}_n$为所有元素皆为1的$n$维向量，则$E(\mathbf{X})=\mu\mathbb{1}_n,\ Cov(\mathbf{X})=\sigma^2I_n$. 此外

$$\begin{aligned}\overline{X}&=\frac{1}{n}\sum_{i=1}^{n}X_i=\frac{1}{n}\mathbb{1}'\mathbf{X}\\ \mathbf{X}-\mathbb{1}_n\overline{X}&=\mathbf{X}-\frac{1}{n}\mathbb{1}_n\mathbb{1}_n'\mathbf{X}=\Big(I_n-\frac{1}{n}\mathbb{1}_n\mathbb{1}_n'\Big)\mathbf{X}=C\mathbf{X},\end{aligned}$$

其中$C=I_n-\frac{1}{n}\mathbb{1}_n\mathbb{1}_n'$, 这是一个对称幂等阵,即$C^2=C,\ C'=C$. 从而有

$$\begin{aligned}Q&=\sum_{i=1}^{n}(X_i-\overline{X})^2=(\mathbf{X}-\overline{X}\mathbb{1}_n)'(\mathbf{X}-\overline{X}\mathbb{1}_n)\\&=(C\mathbf{X})'C\mathbf{X}=\mathbf{X}'C^2\mathbf{X}=\mathbf{X}'C\mathbf{X},\end{aligned}$$

由定理7.2.1可得

$$\begin{aligned}E(Q)&=E(\mathbf{X}'C\mathbf{X})\\&=\mu^2\mathbb{1}'C\mathbb{1}+\sigma^2tr\Big(I_n-\frac{1}{n}\mathbb{1}\mathbb{1}'\Big)\\&=0+\sigma^2(n-1)=\sigma^2(n-1),\end{aligned}$$

从而

$$E(S^2)=\frac{1}{n-1}E(Q)=\sigma^2.$$

三、多元正态分布

1. 定义：

由一元正态和二元正态分布的定义容易推广到一般的情形，得到下列多元正态分布的定义。

定义7.2.3 设n元随机向量$\mathbf{X}=(X_1,X_2,\cdots,X_n)'$具有密度函数

$$f(\mathbf{x})=(2\pi)^{-\frac{n}{2}}|\Sigma|^{-\frac{1}{2}}\exp\Big\{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\Big\},\tag{7.2.1}$$

其中$\mathbf{x}=(x_1,\cdots,x_n)',\ -\infty<x_i<\infty,\ \mu=(\mu_1,\cdots,\mu_n)',\ \Sigma_{n\times n}>0$为正定阵，则称随机向量$\mathbf{X}$的分布为$n$元正态分布，记为$\mathbf{X}\sim N_n(\mu,\Sigma)$.

容易验证：

(1) $f(\mathbf{x})$是密度，即$f(\mathbf{x})>0$，且$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}f(x_1,\cdots,x_n)\,dx_1\cdots dx_n=1$；

(2) $E(X)=\mu,\ Cov(X)=\Sigma$.

证明 (1)作变换$\mathbf{Y}=\Sigma^{-\frac{1}{2}}(\mathbf{X}-\mu)$，则 $\mathbf{X}=\Sigma^{\frac{1}{2}}\mathbf{Y}+\mu$，$|J|=\left|\frac{\partial(x_1,\cdots,x_n)}{\partial(y_1,\cdots,y_n}\right|=|\Sigma|^{\frac{1}{2}}$，从而

$$g(\mathbf{y})=f(\Sigma^{\frac{1}{2}}\mathbf{y}+\mu)\cdot|J|=(2\pi)^{-\frac{n}{2}}e^{-\mathbf{y}'\mathbf{y}/2}=\prod_{i=1}^{n}\Big(\frac{1}{\sqrt{2\pi}}e^{-\frac{y_i}{2}}\Big)=\prod_{i=1}^{n}f(y_i).$$

此处$f(y_i)$是标准正态密度函数，因此Y的n个分量的联合密度等于每个分量的密度的乘积。于是Y的n个分量相互独立，且$Y_i \sim N(0,1),\ i=1,\cdots,n$. 因而

$$E(\mathbf{Y})=\mathbf{0},\quad Cov(\mathbf{Y})=\Sigma^{-\frac{1}{2}}Cov(\mathbf{X})'\Sigma^{-\frac{1}{2}}=\Sigma^{-\frac{1}{2}}\Sigma\Sigma^{-\frac{1}{2}}=I,$$

从而由 $\mathbf{X}=\Sigma^{\frac{1}{2}}\mathbf{y}+\mu$ 可知

$$E(\mathbf{X})=\mu,\quad Cov(\mathbf{X})=Cov(\Sigma^{\frac{1}{2}}\mathbf{Y})=\Sigma.$$

2. 多元正态分布的性质

将向量$\mathbf{X}_{n\times 1}$和$\mu_{n\times 1}$做相应的分块

$$X_{n\times 1}=\begin{pmatrix} X_{(1)} \\ X_{(2)} \end{pmatrix},\quad \mu_{n\times 1}=\begin{pmatrix} \mu_{(1)} \\ \mu_{(2)} \end{pmatrix} \tag{7.2.2}$$

其中$X_{(1)}$, $\mu_{(1)}$皆为$p\times 1$向量；$X_{(2)}$, $\mu_{(2)}$ 均为$q\times 1$向量, $p+q=n$. 将$\mathbf{X}$ 的协方差阵Σ 有如下的分块对角的形式

$$\Sigma=\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \tag{7.2.3}$$

这里Σ_{11}为$p\times 1$的子方阵。

定理7.2.6 (1) 设$\mathbf{X}\sim N_n(\mu,\Sigma)$, 且$\mathbf{X},\mu$和$\Sigma$的分块分别由(7.2.2)和(7.2.3)给出，其中$\Sigma_{12}=0,\ \Sigma_{21}=0$, 则 $X_{(1)}$ 和 $X_{(2)}$ 相互独立，且 $X_{(i)}\sim N(\mu_{(i)},\Sigma_{ii}),\ i=1,2$.

（2） 特别若$\Sigma=\sigma^2 I,\ \mathbf{X}=(X_1,\cdots,X_n)',\ \mu=(\mu_1,\cdots,\mu_n)'$, 则$X_i\sim N(\mu_i,\sigma^2), i=1,2,\cdots,n$.

证明 (1) 由于$\mathbf{X}\sim N_n(\mu,\Sigma)$,且其协方差阵$\Sigma$ 有分块对角的形式(7.2.3), 容易将$\mathbf{X}$的密度函数分解为如下形式

$$f(\mathbf{x})=f(\mathbf{x}_{(1)})f(\mathbf{x}_{(2)})$$

其中$f(\mathbf{x}_{(1)})$和$f(\mathbf{x}_{(2)})$分别为$\mathbf{X}_{(1)}\sim N(\mu_{(\mathbf{1})},\Sigma_{11})$和$\mathbf{X}_{(2)}\sim N(\mu_{(\mathbf{2})},\Sigma_{22})$.的密度函数。

(2) 设$X\sim N_n(\mu,\Sigma)$，则其特征函数(c.f.)为

$$\varphi(\mathbf{t})=\varphi(t_1,\cdots,t_n)=E(e^{it'X})=e^{i\mathbf{t}'\mu-\frac{1}{2}\mathbf{t}'\Sigma\mathbf{t}}=\prod_{j=1}^{n}\{e^{it_j\mu_j-\frac{1}{2}t_j^2\sigma^2}\}.$$

关于正态随机向量的线性变换的正态性有下列结果：

定理7.2.7 设$X\sim N_n(\mu,\Sigma)$, A为$n\times n$可逆常数阵，b为$n\times 1$常向量，记$Y=AX+b$，则

$$Y\sim N_n(A\mu+b,A\Sigma A').$$

证明 用特征函数证明。

$$\begin{aligned}E(e^{it'Y})) &= E(e^{it'(AX+b)}) = e^{it'b}\cdot E(e^{it'AX}) \qquad (\text{令} t'A=\widetilde{t}')\\ &= e^{it'b}\cdot E(e^{i\widetilde{t}'X}) = e^{it'b}\cdot e^{i\widetilde{t}'\mu-\frac{1}{2}\widetilde{t}'\Sigma\widetilde{t}}\\ &= e^{it'b}\cdot e^{it'A\mu-\frac{1}{2}t'A\Sigma A't} = e^{it'(A\mu+b)-\frac{1}{2}t'A\Sigma A't},\end{aligned}$$

即 $Y\sim N(A\mu+b, A\Sigma A')$.

推论7.2.2 设 $\mathbf{X}\sim N_n(\mu,\Sigma)$，则 $\mathbf{Y}=\Sigma^{-\frac{1}{2}}\mathbf{X}\sim N_p(\Sigma^{-\frac{1}{2}}\mu, I)$.

关于正态随机向量的边缘分布有下列结果：

定理7.2.8 设 $X\sim N_n(\mu,\Sigma)$, X, μ, Σ 分块形式如(7.2.2) 和(7.2.3) 所示，则 $X_{(1)}\sim N_p(\mu_{(1)},\Sigma_{11})$. 同理 $X_{(2)}\sim N_q(\mu_{(2)},\Sigma_{22})$. 此处$p+q=n$.

证明 在定理7.2.7中取

$$A=\begin{pmatrix} I_p & 0\\ -\Sigma_{21}\Sigma_{11}^{-1} & I_q\end{pmatrix}, \qquad b=0,$$

则 $\mathbf{Y}=A\mathbf{X}\sim N(A\mu, A\Sigma A')$，此时

$$\mathbf{Y}=A\mathbf{X}=\begin{pmatrix} X_{(1)}\\ X_{(2)}-\Sigma_{21}\Sigma_{11}^{-1}X_{(1)}\end{pmatrix}, \qquad A\mu=\begin{pmatrix}\mu_{(1)}\\ \mu_{(2)}-\Sigma_{21}\Sigma_{11}^{-1}\mu_{(1)}\end{pmatrix},$$

$$A\Sigma A'=\begin{pmatrix}\Sigma_{11} & 0\\ 0 & \Sigma_{22\cdot 1}\end{pmatrix},$$

其中 $\Sigma_{22\cdot 1}=\Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$，从而

$$\mathbf{Y}=\begin{pmatrix} Y_{(1)}\\ Y_{(2)}\end{pmatrix}=A\mathbf{X}\sim N_n\left(\begin{pmatrix}\mu_{(1)}\\ *\end{pmatrix}, \begin{pmatrix}\Sigma_{11} & 0\\ 0 & \Sigma_{22\cdot 1}\end{pmatrix}\right),$$

由定理7.2.6可知：$Y_{(1)}=X_{(1)}\sim N_p(\mu_{(1)},\Sigma_{11})$. □

注 在上述证明中，若取

$$A=\begin{pmatrix} I_p & -\Sigma_{11}^{-1}\Sigma_{12}\\ 0 & I_q\end{pmatrix}$$

用类似方法可证 $X_{(2)}\sim N_q(\mu_{(2)},\Sigma_{22})$.

定理7.2.7还可以进一步推广，获得如下结果：

定理7.2.9 设$A_{m\times n}$ 常数阵，$R(A)=m<n$, 则

$$\mathbf{Y}_{m\times 1}=A\mathbf{X}\sim N_m(A\mu, A\Sigma A').$$

证明： 因 $A_{m\times n}$ 秩为 m，在 n 维线性空间存在 $n-m$ 个向量与 $A_{m\times n}$ 的行向量拼起来构成 R_n 的一组基向量，记这 $n-m$ 个行向量矩阵为 $B_{(n-m)\times n}$，记 $C_{n\times n}=\begin{pmatrix}A\\ B\end{pmatrix}$, C 为 $n\times n$ 可逆阵，$Z=CX\sim N(C\mu, C\Sigma C')$, 从而

$$\mathbf{Z}=C\mathbf{X}=\begin{pmatrix}A\\ B\end{pmatrix}\mathbf{X}=\begin{pmatrix}A\mathbf{X}\\ B\mathbf{X}\end{pmatrix}\triangleq\begin{pmatrix}Z_1\\ Z_2\end{pmatrix}\sim N_n\left(\begin{pmatrix}A\mu\\ B\mu\end{pmatrix}, \begin{pmatrix}A\Sigma A' & A\Sigma B'\\ B\Sigma A' & B\Sigma B'\end{pmatrix}\right),$$

由定理7.2.8可知：$Z_1 = A\mathbf{X} \sim N_m(A\mu, A\Sigma A')$.

推论7.2.3 在定理7.2.9中，若取C为一个行向量即$C = c$，则$c'_{n\times 1}\mathbf{X} \sim N(c'\mu, c'\Sigma c)$，即一元正态变量的线性组合仍为正态。

关于正态变量的两个线性型的独立性有下列结果

定理7.2.10 设$\mathbf{X} \sim N_n(\mu, \Sigma)$，则当$A\Sigma B' = 0$时，$AX$和$BX$独立。

证明 $Cov(AX, BX) = ACov(X)B' = A\Sigma B' = 0$，故$AX$和$BX$不相关，由于它们是正态变量，故不相关与独立等价，因此AX和BX独立。 □

四、正态变量二次型的分布

1. χ_n^2分布的定义及性质

略，见《数理统计》§2.4.

2. 多元正态变量二次型服从χ^2分布的判别方法

定理7.2.11 (1) 设$X \sim N_n(0, \Sigma)$, $\Sigma > 0$ (正定)，则$X'\Sigma^{-1}X \sim \chi_n^2$. 当 $X \sim N(\mu, \Sigma)$，则$(X-\mu)'\Sigma^{-1}(X-\mu) \sim \chi_n^2$.

(2) 设$X \sim N_n(0, I)$，A_n为对称阵，$R(A) = r > 0$，则当A为幂等阵, 即$A^2 = A$)时，二次型$X'AX \sim \chi_r^2$.

证明 (1) 记$Y = \Sigma^{-\frac{1}{2}}X$，则$Y = (Y_1, Y_2, \cdots, Y_n)' \sim N_n(0, I) \Longleftrightarrow Y_i \sim N(0,1), i = 1, 2, \cdots, n$，从而$X'\Sigma^{-1}X = Y'Y = \sum_{i=1}^{n} Y_i^2 \sim \chi_n^2$. □

(2) 对称幂等阵特征根非0即1，即存在正交阵$Q_{n\times n}$使得：

$$Q'AQ = \begin{pmatrix} \lambda_1 & & & & & \\ & \ddots & & & & \\ & & \lambda_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix} = \begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix}$$

此即

$$A = Q\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}Q' \triangleq Q\Lambda Q',$$

因此

$$X'AX = X'Q\Lambda Q'X = Y'\Lambda Y = \sum_{i=1}^{r} Y_i^2,$$

其中$Y = Q'X \sim N_n(0, I)$，从而$X'AX = \sum_{i=1}^{r} Y_i^2 \sim \chi_r^2$. □

推论7.2.4 (1) 若$X \sim N_n(\mu, I)$, $A^2 = A$, $A' = A$, $R(A) = r$，则$(X-\mu)'A(X-\mu) \sim \chi_r^2$.

(2) 若$X \sim N_n(\mu, \Sigma)$，A对称，$R(A) = r$，且$A\Sigma A = A$，则$(X-\mu)'A(X-\mu) \sim \chi_r^2$.

3. 正态变量的两个二次型、二次型与线性型的独立性

关于一个二次型一个线性型的独立性有下列结果：

定理7.2.12 设 $\mathbf{X} \sim N_n(0,I)$ ，若B为$m \times n$阵，A为$n \times n$对称阵，若$BA = 0$ ，则 $B\mathbf{X}$与$\mathbf{X}'A\mathbf{X}$ 独立。

证明 由于 $A' = A$ ，令 $R(A) = r,$ A 对称，故存在正交阵 $Q_{n\times n}$ 使得

$$Q'AQ = \begin{pmatrix} \lambda_1 & & & & & \\ & \ddots & & & & \\ & & \lambda_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix} = \Lambda_r, \quad \text{此即 } A = Q\Lambda_r Q',$$

记 $Y = Q'X$, 则$Y \sim N_n(0,I)$，因此有

$$X'AX = X'Q\Lambda_r Q'X = Y'\Lambda Y = \sum_{i=1}^{r} \lambda_i Y_i^2 = Y_1'\Lambda_r Y_1,\ Y_i \sim N(0,1),\ i = 1,2,\cdots,r,$$

问题转化为当 $BA = 0$ 时，$BX = BQQ'X = DY$，问 DY 与 $Y'\Lambda_r Y$ 是否独立？

由于 $X \sim N_n(0,I) \implies QX = Y \sim N(0,I_n)$，即 $Y_1,\ldots,Y_n$ iid $\sim N(0,1)$，则

$$0 = BA = BQQ'AQQ' = D\Lambda_r Q' \implies D\Lambda_r = 0.$$

将 D 分块为 $D = (\ D_1 \mid D_2\ )$，其中 D_1 的阶数为 $m \times r$，D_2 的阶数为 $m \times (n-r)$，则

$$0 = \begin{pmatrix} D_1 & D_2 \end{pmatrix} \begin{pmatrix} \Lambda_r & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} D_1\Lambda_r & 0 \end{pmatrix} \implies D_1\Lambda_r = 0,$$

又 Λ_r 可逆 $\Longrightarrow D_1 = 0$. 因此$D = (\ 0 \mid D_2\ )$，从而

$$BX = DY = \begin{pmatrix} 0 & D_2 \end{pmatrix} \begin{pmatrix} Y_{(1)} \\ Y_{(2)} \end{pmatrix} = D_2Y_2,\ \text{其中} Y_{(1)} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_r \end{pmatrix}, Y_{(2)} = \begin{pmatrix} Y_{r+1} \\ \vdots \\ Y_n \end{pmatrix},$$

由于 $Y_1,\cdots,Y_r$ 与 $Y_{r+1},\cdots,Y_n$ 独立，$BX = DY = D_2Y_2$ 只与 $Y_{r+1},\cdots,Y_n$ 有关，$X'AX = Y'\Lambda_r Y = \sum_{i=1}^{r} \lambda_i Y_i^2$只与 $Y_1,\cdots,Y_r$ 有关，故二者独立。

推论7.2.5 设 $X \sim N(0,\Sigma)$, 则当 $B\Sigma A = 0$ 时, BX 与 $X'AX$ 独立。

证明： 记 $Y = \Sigma^{-\frac{1}{2}}X \sim N_n(0,I)$，则

$$BX = B\Sigma^{\frac{1}{2}}Y = \widetilde{B}Y,\ X'AX = Y'\Sigma^{\frac{1}{2}}A\Sigma^{\frac{1}{2}}Y = Y'\widetilde{A}Y,$$

此处$\widetilde{A} = \Sigma^{\frac{1}{2}}A\Sigma^{\frac{1}{2}}$, $\widetilde{B} = B\Sigma^{\frac{1}{2}}$. 由定理7.2.12的结果可知：当 $\widetilde{B}\widetilde{A} = 0$ 时,二者独立。

$$\widetilde{B}\widetilde{A} = B\Sigma^{\frac{1}{2}} \cdot \Sigma^{\frac{1}{2}}A\Sigma^{\frac{1}{2}} = B\Sigma A\Sigma^{\frac{1}{2}} = 0 \iff B\Sigma A = 0$$

即这一条件成立时, BX 与 $X'AX$ 独立。

关于两个正态变量二次型的独立性，有下列结果：

定理7.2.13 设 $X \sim N_n(0, I)$, A 和 B 皆为对称阵,且 $AB = 0$,则二次型 $X'AX$ 和 $X'BX$ 独立。

证明： 由$AB = 0$ 可知 $B'A' = BA = 0$, 即 $AB = BA = 0$ 可交换，因此存在公共正交阵 Q，使得 A, B 同时对角化，即

$$Q'AQ = \Lambda = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix},\ Q'BQ = \Delta = \begin{pmatrix} \mu_1 & & \\ & \ddots & \\ & & \mu_n \end{pmatrix},$$

其中 $\lambda_1, \cdots, \lambda_n$ 和 $\mu_1, \cdots, \mu_n$ 分别为 A 和 B 的特征值。

令 $Y = Q'X$,则 $Y \sim N_n(0, I_n)$，因此有

$$X'AX = X'Q\Lambda Q'X = Y'\Lambda Y;$$

$$X'BX = X'Q\Delta Q'X = Y'\Delta Y,$$

由于

$$0 = AB = Q\Lambda Q'Q\Delta Q' = Q\Lambda\Delta Q' \iff \Lambda\Delta = 0$$

故知 Λ 和 Δ 中的对角线上二者非零特征值是错开的，即

若 $\Lambda = \begin{pmatrix} \lambda_1 & & & & & \\ & \ddots & & & & \\ & & \lambda_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix}$，则 $\Delta = \begin{pmatrix} 0 & & & & & \\ & \ddots & & & & \\ & & 0 & & & \\ & & & \mu_{s+1} & & \\ & & & & \ddots & \\ & & & & & \mu_n \end{pmatrix}$，且 $s \geq r$,

故有

$$X'AX = Y'\Lambda Y = \sum_{i=1}^{r} \lambda_i Y_i^2 \quad , \quad X'BX = Y'\Delta Y = \sum_{j=s+1}^{n} \mu_j Y_j^2.$$

由 $Y_1, \cdots, Y_r$ 与 $Y_{r+1}, \cdots, Y_n$ 独立, 因此有 $X'AX$ 与 $X'BX$ 独立。 □

推论7.2.6 设 $X \sim N(0, \Sigma)$, 若 $A\Sigma B = 0$, 则 $X'AX$ 与 $X'BX$ 独立。

证明： 记 $Y = \Sigma^{-\frac{1}{2}}X \sim N(0, I_n)$，则

$$X'AX = Y'\Sigma^{\frac{1}{2}}A\Sigma^{\frac{1}{2}}Y = Y'\widetilde{A}Y, \quad X'BX = Y'\Sigma^{\frac{1}{2}}B\Sigma^{\frac{1}{2}}Y = Y'\widetilde{B}Y,$$

此处$\widetilde{A} = \Sigma^{\frac{1}{2}}A\Sigma^{\frac{1}{2}}$, $\widetilde{B} = \Sigma^{\frac{1}{2}}B\Sigma^{\frac{1}{2}}$. 从而

$$\widetilde{A}\widetilde{B} = 0 \iff A\Sigma B = 0.$$ □

§7.3　　回归系数的LS估计及性质

一、模型

设 Y 为因变量，对 Y 有影响的自变量有 $p-1$ 个，$X_1,\cdots,X_{p-1}$，它们之间有线性关系

$$Y=\beta_0+\beta_1X_1+\cdots+\beta_{p-1}X_{p-1}+e, \tag{7.3.1}$$

e 为随机误差，$\beta_0,\beta_1,\cdots,\beta_{p-1}$ 为未知回归参数，β_0 称为常数项，$\beta_1,\cdots,\beta_{p-1}$ 称为回归系数。

设 $(X_1,\cdots,X_{p-1},Y)$ 的 n 组观察值 $(x_{i1},\cdots,x_{ip-1},y_i)$, $i=1,2,\cdots,n$，则有

$$y_i=\beta_0+\beta_1x_{i1}+\cdots+\beta_{p-1}x_{ip-1}+e_i\ ,\qquad i=1,2,\cdots,n. \tag{7.3.2}$$

误差 $e_1,\cdots,e_n$ 满足Gauss-Markov (G-M) 假定：

$$\begin{cases}(a)\quad E(e_i)=0;\\(b)\quad Var(e_i)=\sigma^2;\\(c)\quad Cov(e_i,e_j)=0,i\neq j.\end{cases} \tag{7.3.3}$$

将方程组(7.3.2)用矩阵表示：

$$\begin{pmatrix}y_1\\y_2\\\vdots\\y_n\end{pmatrix}=\begin{pmatrix}1&x_{11}&x_{12}&\cdots&x_{1,p-1}\\1&x_{21}&x_{22}&\cdots&x_{2,p-1}\\\vdots&\vdots&\vdots&\vdots&\vdots\\1&x_{n,1}&x_{n,2}&\cdots&x_{n,p-1}\end{pmatrix}\begin{pmatrix}\beta_0\\\beta_1\\\vdots\\\beta_{p-1}\end{pmatrix}+\begin{pmatrix}e_1\\e_2\\\vdots\\e_n\end{pmatrix},$$

即：

$$y_{n\times1}=X_{n\times p}\beta_{p\times1}+e_{n\times1}, \tag{7.3.4}$$

此处e满足G-M 假定：

$$E(e)=0,\quad Cov(e)=\sigma^2I. \tag{7.3.5}$$

此处 y 为 $n\times1$ 观察向量，X 为 $n\times p$ 设计阵，β 为$p\times1$回归参数向量，e 为$n\times1$ 随机误差向量，β 和 σ^2 未知，我们目的是求 β 和 σ^2 的估计。

二、LS估计

1. β的LS估计

参数向量 β 的估计有若干不同方法，如有MLE方法和最小二乘估计(Least Square estimation,简称LS 估计)等。下面介绍LS估计，叙述如下：在(7.3.4)中，记 $e=y-X\beta$，使得

$$\parallel e\parallel^2=e'e=\parallel y-X\beta\parallel^2=min$$

即达到最小值时 β 的取值 $\widehat{\beta}$ 称为LS估计，记

$$Q(\beta) = \| y - X\beta \|^2 = (y - X\beta)'(y - X\beta) = y'y - 2y'x\beta + \beta'X'X\beta,$$

则

$$\begin{aligned}\frac{\partial Q(\beta)}{\partial \beta} = 0 \Longleftrightarrow \frac{\partial Q(\beta)}{\partial \beta} &= -2X'y + 2X'X\beta = 0 \\ \Longleftrightarrow X'X\beta &= X'y. \end{aligned} \tag{7.3.6}$$

(7.3.6) 称为正则方程，当 $X'X$ 可逆时有唯一解

$$\widehat{\beta} = (X'X)^{-1}X'y, \tag{7.3.7}$$

称 $\widehat{\beta}$ 为 β 的LS估计。

问题是解(7.3.7)是否使 $Q(\beta) = \| y - X\beta \|^2 = min$ ？下面验证之。

$$\begin{aligned}\| y - X\beta \|^2 &= \| (y - X\widehat{\beta}) + X(\widehat{\beta} - \beta) \|^2 \\ &= \left[(y - X\widehat{\beta}) + X(\widehat{\beta} - \beta)\right]'\left[(y - X\widehat{\beta}) + X(\widehat{\beta} - \beta)\right] \\ &= \| y - X\widehat{\beta} \|^2 + (\widehat{\beta} - \beta)'X'X(\widehat{\beta} - \beta) + 2(\widehat{\beta} - \beta)'X'(y - X\widehat{\beta}) \\ &\triangleq I_1 + I_2 + 2I_3,\end{aligned}$$

由(7.3.6)式易见 $I_3 = 2(\widehat{\beta} - \beta)(X'y - X'X\widehat{\beta}) = 0$，可见

$$\| y - X\beta \|^2 = \| y - X\widehat{\beta} \|^2 + (\widehat{\beta} - \beta)'X'X(\widehat{\beta} - \beta) \geq \| y - X\widehat{\beta} \|^2 \text{，对一切 } \beta \in \mathbb{R}_p,$$

可知 $\widehat{\beta}$ 使 $Q(\beta)$ 达到极小。

2. 经验回归方程

将 $\widehat{\beta}$ 代入(7.3.1)式，e 用其均值0代替得

$$\widehat{Y} = \widehat{\beta_0} + \widehat{\beta_1}X_1 + \cdots + \widehat{\beta}_{p-1}X_{p-1}, \tag{7.3.8}$$

称为经验回归方程，它描述了 Y 与自变量 $X_1, \cdots, X_{p-1}$ 的近似关系。

例7.3.1 一元线性回归。Y 是因变量，只有一个自变量 X, 它们有线性关系

$$Y = \alpha + \beta X + e,$$

现对 (X, Y) 作了 n 次观察，得到数据 (x_i, y_i), $i = 1, \cdots, n$，则有

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix} \Longleftrightarrow y_{n\times 1} = X_{n\times 2}\beta_{2\times 1} + e_{n\times 1},$$

正则方程为

$$X'X\beta = X'y \Longleftrightarrow \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} \sum\limits_{i=1}^{n} y_i \\ \sum\limits_{i=1}^{n} x_i y_i \end{pmatrix},$$

而

$$\mid X'X \mid = \begin{vmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{vmatrix} = n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2 = n\sum_{i=1}^{n}(x_i - \bar{x})^2,$$

因此

$$(X'X)^{-1} = \frac{1}{n\sum\limits_{i=1}^{n}(x_i - \bar{x})^2} \begin{pmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{pmatrix},$$

代入到正则方程组可得

$$\begin{pmatrix} \widehat{\alpha} \\ \widehat{\beta} \end{pmatrix} = (X'X)^{-1}X'y = \frac{1}{n\sum\limits_{i=1}^{n}(x_i - \bar{x})^2} \begin{pmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{pmatrix} \begin{pmatrix} \sum\limits_{i=1}^{n} y_i \\ \sum\limits_{i=1}^{n} x_i y_i \end{pmatrix},$$

因此β和α得LS估计为

$$\widehat{\beta} = \frac{\sum\limits_{i=1}^{n} x_i y_i - \frac{1}{n}\sum\limits_{i=1}^{n} x_i \sum\limits_{i=1}^{n} y_i}{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2},$$

$$\widehat{\alpha} = \bar{y} - \widehat{\beta}\bar{x}.$$

从而 $\widehat{y} = \widehat{\alpha} + \widehat{\beta}x$ 为所要得的回归方程。 □

3. σ^2 *的 LS 估计*

$e_{n\times 1} = y_{n\times 1} - X\beta$，将 β 用 $\widehat{\beta}$ 代替得残差向量估计

$$\widehat{e} = y - X\widehat{\beta} \Longleftrightarrow \widehat{e}_i = y_i - x_i'\widehat{\beta},$$

其中 x_i' 是设计阵 X 的第 i 行的行向量。用 $\widehat{e}$ 作为 e 的估计，自然想到用

$$RSS = \| \widehat{e} \|^2 = \widehat{e}'\widehat{e} = \sum_{i=1}^{n} \widehat{e}_i^2 \tag{7.3.9}$$

作为衡量 σ^2 大小的度量，RSS 是残差平方和，它的大小反映了实际数据与理论数据的偏离程度，可以证明 $E(RSS) = (n-p)\sigma^2$，因此

$$\widehat{\sigma}^2 = \frac{1}{n-p}RSS = \frac{1}{n-p} \| y - X\widehat{\beta} \|^2 = \frac{1}{n-p}y'\big(I - X(X'X)^{-1}X'\big)y, \tag{7.3.10}$$

为 σ^2 的无偏估计，它称为 σ^2 的LS估计。

下面利用 $y \sim (X\beta, \sigma^2 I)$ 计算$E(RSS)$.

$$\begin{aligned} E(RSS) &= E\big[y'\big(I - X(X'X)^{-1}X'\big)y\big] \\ &= (X\beta)'(I - P_X)X\beta + \sigma^2 \cdot tr(I - P_X) \\ &= 0 + \sigma^2(n-p) = (n-p)\sigma^2, \end{aligned}$$

此处 $P_X = X(X'X)^{-1}X'$ 为对称幂零阵，易知 $tr(I - P_X) = R(I - P_X) = n - p$.

三、线性回归模型的中心化和标准化

1. 中心化

在回归分析的应用中，我们常常需要把原始观测数据进行中心化和标准化，这对我们进行统计分析是有益的。记

$$\bar{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}, \quad j = 1, \cdots, p-1.$$

此处$\bar{x}_j$是第j个回归自变量n次取值的平均数。将模型(7.3.2) 进行改写为

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_{i1} + \cdots + \beta_{p-1}x_{ip-1} + e_i \\ &= (\beta_0 + \beta_1\bar{x}_1 + \cdots + \beta_{p-1}\bar{x}_{p-1}) + \beta_1(x_{i1} - \bar{x}_1) + \cdots + \beta_{p-1}(x_{ip-1} - \bar{x}_{p-1}) + e_i \\ &= \alpha + \beta_1(x_{i1} - \bar{x}_1) + \cdots + \beta_{p-1}(x_{ip-1} - \bar{x}_{p-1}) + e_i,\ i = 1, \cdots, n. \end{aligned} \tag{7.3.11}$$

用矩阵表示为：

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \alpha\mathbb{1} + \begin{pmatrix} x_{11} - \bar{x}_1 & x_{12} - \bar{x}_2 & \cdots & x_{1,p-1} - \bar{x}_{p-1} \\ x_{21} - \bar{x}_1 & x_{22} - \bar{x}_2 & \cdots & x_{2,p-1} - \bar{x}_{p-1} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n,1} - \bar{x}_1 & x_{n,2} - \bar{x}_2 & \cdots & x_{n,p-1} - \bar{x}_{p-1} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{p-1} \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}$$

此即

$$y_{n\times 1} = \alpha\mathbb{1}_{n\times 1} + X_c\beta_{(p-1)\times 1} + e, \tag{7.3.12}$$

此处 α 为常数项，$\beta = (\beta_1, \cdots, \beta_{p-1})'$ 为回归系数向量，X_c 为 $n \times (p-1)$ 阶阵，称为中心化设计阵。容易验证

$$\mathbb{1}_n' X_c = 0. \tag{7.3.13}$$

这是因为 $\mathbb{1}' \begin{pmatrix} x_{1j} - \bar{x}_j \\ \vdots \\ x_{nj} - \bar{x}_j \end{pmatrix} = \sum\limits_{i=1}^{n}(x_{ij} - \bar{x}_j) = \sum\limits_{i=1}^{n} x_{ij} - n\bar{x}_j = 0,\ j = 1, \cdots, p-1.$

将(7.3.12)式改写为

$$y_{n\times 1} = (\mathbb{1} \mid X_c)\begin{pmatrix} \alpha \\ \beta \end{pmatrix} + e = \tilde{X}\begin{pmatrix} \alpha \\ \beta \end{pmatrix} + e,$$

因此正则方程为

$$\tilde{X}'\tilde{X}\begin{pmatrix}\alpha\\ \beta\end{pmatrix}=\tilde{X}'y \iff \begin{pmatrix}n & 0\\ 0 & X_c'X_c\end{pmatrix}\begin{pmatrix}\alpha\\ \beta\end{pmatrix}=\begin{pmatrix}\mathbb{1}'y\\ X_c'y\end{pmatrix}$$

故有

$$\begin{pmatrix}\widehat{\alpha}\\ \widehat{\beta}\end{pmatrix}=\begin{pmatrix}\frac{1}{n} & 0\\ 0 & (X_c'X_c)^{-1}\end{pmatrix}\begin{pmatrix}\mathbb{1}'y\\ X_C'y\end{pmatrix}$$

于是回归参数的LS估计为

$$\begin{cases}\widehat{\alpha}=\dfrac{1}{n}\sum\limits_{i=1}^{n}y_i=\bar{y}\\ \widehat{\beta}=(X_c'X_c)^{-1}X_c'y.\end{cases}\tag{7.3.14}$$

结论： 中心化模型中常数项 $\widehat{\alpha}=\bar{y}$，而 $\widehat{\beta}$ 可视为由 $y=X_c\beta+e$ 按LS方法求出的解。

例7.3.2 一元线性回归(续)，用中心化方法求 α 和 β 的LS估计。设中心化的一元线性回归为

$$y_i=\alpha+(x_i-\bar{x})\beta+e_i\,,\ i=1,2,\cdots,n.$$

按中心化方法

$$X_c'=(x_1-\bar{x},\cdots,x_n-\bar{x})$$
$$(X_c'X_c)^{-1}=1\Big/\sum_{i=1}^{n}(x_i-\bar{x})^2.$$

按公式(7.3.14)得到α和β的LS估计为

$$\widehat{\alpha}=\bar{y}=\frac{1}{n}\sum_{i=1}^{n}y_i,$$
$$\widehat{\beta}=(X_c'X_c)^{-1}X_c'y=\frac{\sum\limits_{i=1}^{n}(x_i-\bar{x})y_i}{\sum\limits_{i=1}^{n}(x_i-\bar{x})^2}.$$

从而得到经验回归方程

$$\widehat{y}=\widehat{\alpha}+\widehat{\beta}(X-\bar{x}).$$

2. 标准化

中心化了之后，还可以继续做标准化变换。设(7.3.12)式中刻画设计阵第 $j+1$ 列的离散程度的量为

$$s_j=\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)^2,\ j=1,2,\cdots,n.$$

而令标准化变换为：

$$z_{ij}=\frac{x_{ij}-\bar{x}_j}{s_j},\quad i=1,2,\cdots,n;\ j=1,2,\cdots,p-1.\tag{7.3.15}$$

将中心化数据模型$y_i=\alpha+\beta_1(x_{i1}-\bar{x}_1))+\cdots+\beta_{p-1}(x_{ip-1}-\bar{x}_{p-1})+e_i$ 作进一步改写，得到

$$\begin{aligned} y_i &= \alpha+\frac{x_{i1}-\bar{x}_1}{s_1}(s_1\beta_1)+\frac{x_{i2}-\bar{x}_2}{s_2}(s_2\beta_2)+\cdots+\frac{x_{ip-1}-\bar{x}_{p-1}}{s_{p-1}}(s_{p-1}\beta_{p-1})+e_i \\ &= \alpha+z_{i1}\beta_1^o+z_{i2}\beta_2^o+\cdots+z_{ip-1}\beta_{p-1}^o+e_i, \quad i=1,2,\cdots,n. \end{aligned}$$

其中$\beta_i^o=s_i\beta_i$, $i=1,\cdots,p-1$. 其矩阵表示为

$$y_{n\times 1}=\alpha\mathbb{1}_n+Z\beta^o+e.$$

其中$Z=(z_{ij})$为$n\times(p-1)$ 矩阵, z_{ij}由(7.3.15)给出。显然此时仍有

(a) $\mathbb{1}_n'Z=0$;

(b) $Z'Z=R=(r_{ij})$ 为$(p-1)\times(p-1)$ 方阵，其中$r_{ij}=\frac{1}{s_i\,s_j}\sum\limits_{k=1}^{n}(x_{ki}-\bar{x_i})(x_{kj}-\bar{x_j})$为样本相关系数，特别 $r_{jj}=1$.

当我们对这个模型求得了参数的LS估计

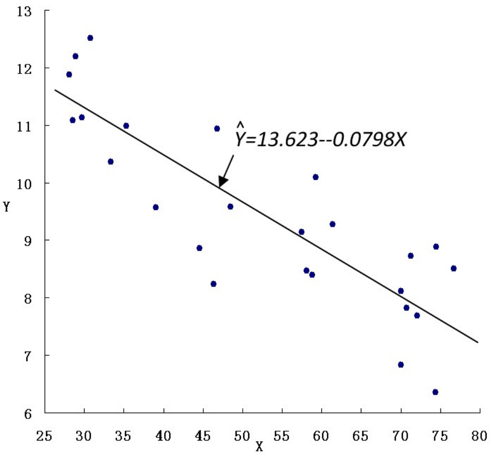
$$\widehat{\alpha}=\bar{y},\ \widehat{\beta}_1^o,\cdots,\widehat{\beta}_{p-1}^o$$

之后，则经验回归方程为

$$\begin{aligned} \widehat{y} &= \widehat{\alpha}+\Big(\frac{X_1-\bar{x}_1}{s_1}\Big)\widehat{\beta}_1^o+\cdots+\Big(\frac{X_{p-1}-\bar{x}_{p-1}}{s_{p-1}}\Big)\widehat{\beta}_{p-1}^o \\ &= \Big(\widehat{\alpha}-\sum_{i=1}^{p-1}\frac{\bar{x}_i}{s_i}\widehat{\beta}_i^o\Big)+\sum_{i=1}^{p-1}\Big(\frac{\widehat{\beta}_i^o}{s_i}\Big)X_i. \end{aligned}$$

例7.3.3 一个试验容器靠蒸汽供应热量，使其保持恒温，下表中自变量 X 表示容器周围空气中单位时间平均温度(℃)，Y表示单位时间消耗蒸汽量(L)。共有25组数据，求经验回归方程。

序号	Y(L)	X(℃)	序号	Y(L)	X(℃)
1	10.98	35.3	14	9.57	39.1
2	11.13	29.7	15	10.94	46.8
3	12.51	30.8	16	9.58	48.5
4	8.40	58.8	17	10.09	59.3
5	9.27	61.4	18	8.11	70.0
6	8.73	71.3	19	6.83	70.0
7	6.36	74.4	20	8.88	74.5
8	8.50	76.7	21	7.68	72.1
9	7.82	70.7	22	8.47	58.1
10	9.14	57.5	23	8.86	44.6
11	8.24	46.4	24	10.36	33.4
12	12.19	28.9	25	11.08	28.6
13	11.88	28.1			



蒸汽与温度的关系

蒸汽与温度的关系

解： 中心化方法由数据得到 $\bar{y}=9.424,\ \bar{x}=52.60$，从而常数项 α 和回归系数 β 的最小二乘估计分别为：

$$\widehat{\alpha}=\bar{y}=9.424,$$

$$\widehat{\beta}=-0.079\,8.$$

于是经验回归方程为

$$\widehat{Y}=9.424-0.079\,8(X-52.60),$$

即

$$\widehat{Y}=13.623-0.079\,8X.$$ □

四、LS 估计的性质

1. *回归参数 β 的 LS 估计的矩*

定理7.3.1 设 $\widehat{\beta}$ 为 β 的LS估计，则有

$$E(\widehat{\beta})=\beta,\ Cov(\widehat{\beta})=\sigma^2(X'X)^{-1}.$$

证明 由 $\widehat{\beta}=(X'X)^{-1}X'y$ 可知

$$\begin{aligned}E(\widehat{\beta})&=(X'X)^{-1}X'E(Y)=(X'X)^{-1}X'X\beta=\beta,\\ Cov(\widehat{\beta})&=Cov\big((X'X)^{-1}X'y\big)=(X'X)^{-1}X'Cov(Y)X(X'X)^{-1}\\ &=\sigma^2(X'X)^{-1}X'X(X'X)^{-1}=\sigma^2(X'X)^{-1}.\end{aligned}$$ □

2. *Gauss-Markov 定理*

设 c 为 $p\times 1$ 维常数向量，对于线性函数 $c'\beta$，称 $c'\widehat{\beta}$ 为 $c'\beta$ 的LS估计。则由定理7.3.1可知

$$E(c'\widehat{\beta})=c'\beta,\quad Cov(c'\widehat{\beta})=\sigma^2c'(X'X)^{-1}c,\tag{7.3.16}$$

$c'\beta$ 称为可估函数。

定理7.3.2 对于线性回归模型 $y=X\beta+e,\ e\sim(0,\sigma^2I)$，可估函数 $c'\beta$ 的所有线性无偏估计中，LS 估计 $c'\widehat{\beta}$ 是唯一具有最小方差者。

证明： 设 $a'y$ 为 $c'\beta$ 的任一无偏估计，于是

$$E(a'y)=a'X\beta=c'\beta,$$

此式对一切 $\beta_{p\times 1}\in\mathbb{R}_p$ 都成立，同时必有

$$a'X=c'.\tag{7.3.17}$$

我们要证明 $Var(c'\beta) \le Var(a'y)$，对一切 $\beta \in \mathbb{R}_p$成立。而

$$Var(a'y) = a'Cov(y)a = \sigma^2 a'a,$$

$$Var(c'\beta) = \sigma^2 c'(X'X)^{-1}c,$$

则

$$\begin{aligned}
Var(a'y) - Var(c'\widehat{\beta}) &= \sigma^2 a'a - \sigma^2 c'(X'X)c \\
&= \sigma^2\big(a'a - a'X(X'X)^{-1}X'a\big) \\
&= \sigma^2 a'\big(I - X(X'X)^{-1}X'\big)a,
\end{aligned}$$

这等价于证明下式成立：

$$I - X(X'X)^{-1}X' \ge 0,$$

而 $I - X(X'X)^{-1}X'$ 为对称幂等阵，其特征根非负，因此$I - X(X'X)^{-1}X' \ge 0$, 故有

$$Var(a'y) - Var(c'\beta) \ge 0,$$

一切 $\beta \in \mathbb{R}_p$ 成立，且等号成立 $\Longleftrightarrow$ $\big(I - X(X'X)^{-1}X'\big)a = 0$，即 $a = X(X'X)^{-1}c$，也就是 $a'y = c'\widehat{\beta}$，唯一性得证。

3. 正态线性模型下，β 和 σ^2 的 LS 估计的优良性

定理7.3.3 对前面的线性模型 $y = X\beta + e$，若 $e \sim N(0, \sigma^2 I)$，则

(a) $\beta \sim N(\beta, \sigma^2(X'X)^{-1})$;

(b) $RSS/\sigma^2 \sim \chi^2_{n-p}$;

(c) $\widehat{\beta}$ 与 $\widehat{\sigma}^2$ 相互独立。

证明：(a) 显然。

(b) 令$Z = (y - X\beta)/\sigma$, 则$Z_{n\times 1} \sim N(0, I)$. 由此事实和定理7.2.11可知

$$\frac{RSS}{\sigma^2} = \frac{y'(I - P_X)y}{\sigma^2} = \left(\frac{y - X\beta}{\sigma}\right)'(I - P_X)\left(\frac{y - X\beta}{\sigma}\right) = Z'(I - P_X)Z \sim \chi^2_{m-p}.$$

(c) 将RSS 和$\widehat{\beta}$ 表示成Y的二次型和线性型：

$$RSS = y'(I - P_X)y = y'Ay, \quad y \sim N(X\beta, \sigma^2 I), \quad \widehat{\beta} = (X'X)^{-1}X'y = By,$$

利用二次型和线性型独立性判定定理可知，二者相互独立 $\Longleftrightarrow BA = 0$, 而

$$\begin{aligned}
BA &= (X'X)^{-1}X'(I - P_X) \\
&= (X'X)^{-1}X' - (X'X)^{-1}X'X(X'X)^{-1}X' \\
&= (X'X)^{-1}X' - (X'X)^{-1}X' = 0,
\end{aligned}$$

故独立。 $\square$

4. 中心化回归模型的 β 和 σ^2 的 LS 估计的优良性

对中心化的回归模型：

$$y = \alpha \mathbb{1}_n + X_c\beta + e,$$

其LS 估计为$\widehat{\alpha} = \bar{y}$, $\widehat{\beta} = (X_c'X_c)^{-1}X_c'y$, 则有

(a) $E(\widehat{\alpha}) = \alpha,\ E(\widehat{\beta}) = \beta,$

(b) $Cov\begin{pmatrix} \widehat{\alpha} \\ \widehat{\beta} \end{pmatrix} = \sigma^2 \begin{pmatrix} \frac{1}{n} & 0 \\ 0 & (X_c'X_c)^{-1} \end{pmatrix}$

(c) 进一步假定 $e \sim N(0, \sigma^2 I)$, 则有

$$\widehat{\alpha} \sim N_1(\alpha,\ \frac{1}{n}\sigma^2),$$

$$\widehat{\beta} \sim N_{p-1}(\beta,\ \sigma^2(X_c'X_c)^{-1}),$$

且 $\widehat{\alpha}$ 和 $\widehat{\beta}$ 相互独立。

5. 复相关系数

记 $SS_{回} = \widehat{\beta}'X_C'y$ 称为回归平方和，$SS_{总} = \sum\limits_{i=1}^{n}(y_i - \bar{y})^2$ 称为修正的总平方和，定义

$$R^2 = \frac{SS_{回}}{SS_{总}},$$

称为判定系数，它判定回归自变量 $X_1, \cdots, X_{p-1}$ 与因变量 y 的拟合程度。显然 $0 \le R^2 \le 1$。R 称为复相关系数的，R^2 越大说明 Y 与诸 X_i 有较大的相关关系。

例7.3.4 根据经验，在人的身高相等条件，其血压收缩压 y 与体重 x_1、年龄 x_2 有关，收集了13人数据。试建立 y 与 x_1, x_2 经验回归方程。

序号	x_{i1}	x_{i2}	y_i	序号	x_{i1}	x_{i2}	y_i
1	152	50	120	8	158	50	125
2	183	20	141	9	170	40	132
3	171	20	124	10	153	55	123
4	165	30	126	11	164	40	132
5	158	30	117	12	190	40	155
6	161	50	125	13	185	20	147
7	149	60	123				

解： 利用中心化模型

$$y_i = \alpha + (x_{i1} - \bar{x}_1)\beta_1 + (x_{i2} - \bar{x}_2)\beta_2 + e_i, \quad i = 1, 2, \cdots, 13.$$

由上表中数据计算可得

$$\bar{x}_1 = \frac{1}{13}\sum_{i=1}^{13} x_{i1} = 166.8$$

$$\bar{x}_2 = \frac{1}{13}\sum_{i=1}^{13} x_{i2} = 38.85$$

$$\bar{y} = \frac{1}{13}\sum_{i=1}^{13} y_i = 130$$

中心化设计矩阵为

$$X_C = \begin{pmatrix} -14.08 & 11.15 \\ 16.92 & -18.85 \\ 4.92 & -18.85 \\ -1.08 & -8.85 \\ -8.08 & -8.85 \\ -5.08 & 11.15 \\ -17.08 & 21.15 \\ -8.08 & 11.15 \\ 3.92 & 1.15 \\ -13.08 & 16.15 \\ -2.08 & 1.15 \\ 23.92 & 1.15 \\ 18.92 & -18.85 \end{pmatrix}$$

中心化正则方程 $X_c'X_c\tilde{\beta} = X_c'y$ 为

$$\begin{cases} 2\,078.92\,\beta_1 - 1\,533.85\,\beta_2 = 1\,607.00 \\ -1\,533.85\,\beta_1 + 2\,307.69\,\beta_2 = -715.00 \end{cases}$$

解得

$$\widehat{\beta}_1 = 1.068, \qquad \widehat{\beta}_2 = 0.400.$$

又

$$\widehat{\alpha} = \bar{y} = 130,$$

故得经验回归方程为

$$\begin{aligned} \widehat{Y} &= \widehat{\alpha} + (x_1 - \bar{x}_1)\widehat{\beta}_1 + (x_2 - \bar{x}_2)\widehat{\beta}_2 \\ &= 130 + 1.068(x_1 - 166.8) + 0.400(x_2 - 38.85) \\ &= -62.963 + 1.068x_1 + 0.400x_2. \end{aligned}$$

我们还可计算出判定系数

$$SS_{\text{回}} = \widehat{\beta}'X_C'y = 1\,430.276,$$

$$SS_{\text{总}} = \sum_{i=1}^{13}(y_i - \bar{y})^2 = 1\,512,$$

故有

$$R^2=\frac{SS_{回}}{SS_{总}}=\frac{1\,430.276}{1\,512}=0.945\,9,\qquad R=0.972\,6.$$

可见二元线性回归拟合的良好。 □

§7.4 约束最小二乘估计

一、模型

设有下列线性模型 $y_{n\times1}=X_{n\times p}\beta_{p\times1}+e_{n\times1}$，$e\sim(0,\sigma^2I)$，其中参数 β 和 σ^2 的LS估计及其性质已在 §7.3 中介绍过。在假设检验问题和一些实际问题中，我们常需要求带约束条件的LS估计问题，即有约束条件

$$A\beta=b \tag{7.4.1}$$

其中 A 为 $k\times p$ 阶矩阵，b 为 $k\times1$ 维向量($k\le p$)，A、b 皆已知，$R(A)=k$，且(7.4.1)式为一个相容方程组条件(7.4.1)的一个例子是在回归模型参数的假设检验问题中常求下列检验：

$$H_0:\beta_1=\beta_2=\cdots=\beta_{p-1}=0.$$

即回归方程的显著性检验。将上式用矩阵表示

$$H_0:\left\{\begin{array}{c}\beta_1-\beta_{p-1}=0\\ \beta_2-\beta_{p-1}=0\\ \vdots\\ \beta_{p-2}-\beta_{p-1}=0\end{array}\right.\iff H_0:\begin{pmatrix}0&1&0&\cdots&0&-1\\0&0&1&\cdots&0&-1\\ \vdots&\vdots&\vdots&\ddots&\vdots&\vdots\\0&0&0&\cdots&1&-1\end{pmatrix}\begin{pmatrix}\beta_0\\ \beta_1\\ \vdots\\ \beta_{p-1}\end{pmatrix}=\begin{pmatrix}0\\0\\ \vdots\\0\end{pmatrix}.$$

此即 $H_0:A\beta=0$，此处

$$A=\begin{pmatrix}0&1&0&\cdots&0&-1\\0&0&1&\cdots&0&-1\\ \vdots&\vdots&\vdots&\ddots&\vdots&\vdots\\0&0&0&\cdots&1&-1\end{pmatrix},\quad \beta=\begin{pmatrix}\beta_0\\ \beta_1\\ \vdots\\ \beta_{p-1}\end{pmatrix},\quad b=\begin{pmatrix}0\\0\\ \vdots\\0\end{pmatrix}.$$

因此有约束的线性回归模型为

$$\left\{\begin{array}{l}y_{n\times1}=X_{n\times p}\beta_{p\times1}+e_{n\times1},\quad e\sim(o,\sigma^2I)\\ A\beta=b,\quad \text{其中 } R(A)=k,\ \text{即行满秩}.\end{array}\right. \tag{7.4.2}$$

二、约束的LS解

定理7.4.1 对约束的线性回归模型(7.4.2)，β 的有约束的LS解为

$$\widehat{\beta}_c=\widehat{\beta}-(X'X)^{-1}A'\big(A(X'X)^{-1}A'\big)^{-1}(A\widehat{\beta}-b),$$

其中 $\widehat{\beta}=(X'X)^{-1}X'y$ 是 §7.3.1 中求出的 β 的无约束的LS解。

证明： 将约束条件(7.4.1)分解，记

$$A=\begin{pmatrix}\tilde{a}_1'\\ \vdots\\ \tilde{a}_k\end{pmatrix},\quad b=\begin{pmatrix}b_1\\ \vdots\\ b_k\end{pmatrix},$$

则约束条件(7.4.1)式等价于

$$\begin{pmatrix}\tilde{a}_1'\beta\\ \vdots\\ \tilde{a}_k\beta\end{pmatrix}=\begin{pmatrix}b_1\\ \vdots\\ b_k\end{pmatrix}，即\ \tilde{a}_i'\beta=b_i,\quad i=1,2,\cdots,k. \tag{7.4.3}$$

令 $Q(\beta)=\|y-X\beta\|^2$，问题转化为在(7.4.3)约束条件下求 $Q(\beta)$ 的最小值点，用Lagrange乘子法构造函数，记 $\lambda'=(\lambda_1,\cdots,\lambda_k)$，则

$$\begin{aligned}F(\beta,\lambda)&=\|y-X\beta\|^2+2\sum_{i=1}^{n}\lambda_i(\tilde{a}_i{}'\beta-b_i)\\&=\|y-X\beta\|^2+2\lambda'(A\beta-b)\\&=(y-X\beta)'(y-X\beta)+2\lambda'(A\beta-b)\\&=y'y+\beta'X'X\beta-2y'X\beta+2\lambda'(A\beta-b)\end{aligned}$$

将函数 $F(\beta,\lambda)$ 对 β 求偏导数，整理并令它们等于0，得到

$$\begin{cases}\dfrac{\partial F(\beta,\lambda)}{\partial\beta}=2X'X\beta-2X'y+2A'\lambda=0\\ \dfrac{\partial F(\beta,\lambda)}{\partial\lambda}=2(A\beta-b)=0\end{cases}$$

此即

$$\begin{cases}X'X\beta-X'y+A'\lambda=0 & ①\\ A\beta=b & ②\end{cases}$$

由①可以得到

$$\widehat{\beta}_c=(X'X)^{-1}X'y-(X'X)^{-1}A'\widehat{\lambda}_c=\widehat{\beta}-(X'X)^{-1}A'\widehat{\lambda}_c, \tag{7.4.4}$$

将(7.4.4)式带入②得

$$\begin{aligned}&A\widehat{\beta}_c=A\widehat{\beta}-A(X'X)^{-1}A'\widehat{\lambda}_c=b\\ \Longleftrightarrow\ &A(X'X)^{-1}A'\widehat{\lambda}_c=A\widehat{\beta}-b,\end{aligned} \tag{7.4.5}$$

由 $R(A)=k$ 可知$A(X'X)^{-1}A'$ 之逆存在，故有

$$\widehat{\lambda}_c=\big(A(X'X)^{-1}A'\big)^{-1}(A\widehat{\beta}-b). \tag{7.4.6}$$

将(7.4.6)式代回(7.4.4)式可得

$$\widehat{\beta}_c=\widehat{\beta}-(X'X)^{-1}A'\big(A(X'X)^{-1}A'\big)^{-1}(A\widehat{\beta}-b). \tag{7.4.7}$$

现在我们证明 $\widehat{\beta}_c$ 确实是线性约束 $A\beta=b$ 下 β 的 LS 估计。为此我们只需证明如下两点：

(a) $\widehat{\beta}_c$ 适合约束方程，即 $A\widehat{\beta}_c=b$，显然成立。

(b) 要证明对适合 (7.4.1) 的约束条件下的一切 β，都有

$$\|y-X\widehat{\beta}\|^2\geq\|y-X\widehat{\beta}_c\|^2.$$

为此，我们将平方和 $\|y-X\beta\|^2$ 作分解。

$$\begin{aligned}\|y-X\beta\|^2&=\|(y-X\widehat{\beta}_c)+X(\widehat{\beta}_c-\beta)\|^2\\&=\|y-X\widehat{\beta}_c\|^2+(\widehat{\beta}_c-\beta)'X'X(\widehat{\beta}_c-\beta)+2(\widehat{\beta}_c-\beta)'X'(y-X\widehat{\beta}_c)\\&=I_1+I_2+2I_3,\end{aligned}$$

其中 $I_3=(\widehat{\beta}_c-\beta)'(X'y-X'X\widehat{\beta}_c)$. 因为

$$\begin{aligned}X'X\widehat{\beta}_c&=X'X\big[\widehat{\beta}-(X'X)^{-1}A'\big(A(X'X)^{-1}A'\big)^{-1}(A\beta-b)\big]\\&=X'X(X'X)^{-1}X'y-X'X(X'X)^{-1}A'\big(A(X'X)^{-1}A'\big)^{-1}(A\beta-b)\\&=X'y-A'\widehat{\lambda}_c,\end{aligned}$$

从而

$$I_3=(\widehat{\beta}_c-\beta)'A'\widehat{\lambda}_c=(A\widehat{\beta}_c-A\beta)'\widehat{\lambda}_c=(b-b)\widehat{\lambda}_c=0.$$

故有$\|y-X\beta\|^2=I_1+I_2\geq\|y-X\widehat{\beta}_c\|^2$, 定理证毕。

例7.4.1 在天文测量中，对天空中三个星位点构成的三角形的三个内角 θ_1, θ_2, θ_3 进行测量，得到的测量值分别为 y_1, y_2, y_3，由于测量存在误差，所以需要对 θ_1, θ_2, θ_3 进行估计，求 $\beta'=(\theta_1,\ \theta_2,\ \theta_3)'$ 的估计量。

解：我们利用线性模型表示有关的量：

$$\left\{\begin{array}{l}y_1=\theta_1+e_1\\y_2=\theta_2+e_2\\y_3=\theta_3+e_3\\\theta_1+\theta_2+\theta_3=\pi\quad(\text{约束条件})\end{array}\right.\iff\left\{\begin{array}{l}y=X\beta+e\\A\beta=b\end{array}\right.$$

此处 e_1, e_2, e_3 为测量误差，且

$$y=\begin{pmatrix}y_1\\y_2\\y_3\end{pmatrix},\ \beta=\begin{pmatrix}\theta_1\\\theta_2\\\theta_3\end{pmatrix},\ X=\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix},\ e=\begin{pmatrix}e_1\\e_2\\e_3\end{pmatrix},\ A=\begin{pmatrix}1&1&1\end{pmatrix},\ b=\pi.$$

注意到 $\widehat{\beta}=(X'X)^{-1}X'y=y$ ，由定理 7.4.1 可知

$$
\begin{aligned}
\widehat{\beta}_c &= \widehat{\beta}-(X'X)^{-1}A'\big(A(X'X)^{-1}A'\big)^{-1}(A\widehat{\beta}-b)\\
&= \begin{pmatrix} y_1\\ y_2\\ y_3\end{pmatrix}-\begin{pmatrix}1\\1\\1\end{pmatrix}\left[\begin{pmatrix}1&1&1\end{pmatrix}I_n\begin{pmatrix}1\\1\\1\end{pmatrix}\right]^{-1}\left(\begin{pmatrix}1&1&1\end{pmatrix}\begin{pmatrix} y_1\\ y_2\\ y_3\end{pmatrix}-\pi\right)\\
&= \begin{pmatrix} y_1\\ y_2\\ y_3\end{pmatrix}-\frac{1}{3}\Big(\sum_{i=1}^{3}y_i-\pi\Big)\mathbb{1}_3,
\end{aligned}
$$

即$\widehat{\theta_i}=y_i-\dfrac{1}{3}(y_1+y_2+y_3-\pi),\ i=1,2,3$ 为θ_i有约束的 LS 解。

§7.5　　广义最小二乘（GLS）估计

一、引言

前面讨论回归模型

$$y_{n\times 1}=X_{n\times p}\beta_{p\times 1}+e_{n\times 1},\quad e\sim(0,\sigma^2 I),$$

但在许多实际问题中$Cov(e)=\sigma^2 I$这个假定未必成立，常有 $Cov(e)=\sigma^2\Sigma,\ \Sigma>0$ (正定)，此时如何求 LS 解？($Cov(e)=\sigma^2\Sigma$ 表示 G-M 假定不成立，即它们的误差方差可能不相等，随机误差之间可能彼此相关)。此时模型为

$$y_{n\times 1}=X_{n\times p}\beta_{p\times 1}+e_{n\times 1},\quad E(e)=0,\ Cov(e)=\sigma^2\Sigma,\ \Sigma>0. \tag{7.5.1}$$

为求参数的估计量，我们希望经过适当的变换，将这一明显转化成 §7.3 中讨论过的情形。将 (3.5.1) 式两边同时左乘 $\Sigma^{-\frac{1}{2}}$，得到

$$\Sigma^{-\frac{1}{2}}y=\Sigma^{-\frac{1}{2}}X\beta+\Sigma^{-\frac{1}{2}}e$$

即

$$\tilde{y}=\tilde{X}\beta+\tilde{e},\ \tilde{e}\sim(0,\sigma^2 I). \tag{7.5.2}$$

其中 $\tilde{y}=\Sigma^{-\frac{1}{2}}y,\ \tilde{X}=\Sigma^{-\frac{1}{2}}X,\ \tilde{e}=\Sigma^{-\frac{1}{2}}e\sim(0,\sigma^2 I)$. (7.5.2) 为一般的回归模型，由 §7.3 可知 β 和 σ^2 的 LS 估计为

$$
\begin{cases}
\widehat{\beta}^*=(\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{y}=(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y\\
\widehat{\sigma}^2_*=\frac{\|\tilde{y}-\tilde{X}\widehat{\beta}^*\|^2}{n-p}=\frac{(\tilde{y}-\tilde{X}\widehat{\beta}^*)'(\tilde{y}-\tilde{X}\widehat{\beta}^*)}{n-p}=\frac{(y-X\widehat{\beta}^*)'\Sigma^{-1}(y-X\widehat{\beta}^*)}{n-p}
\end{cases} \tag{7.5.3}
$$

称$\widehat{\beta}^*$ 和$\widehat{\sigma}^2_*$为β 和σ^2的广义最小二乘估计，记为GLS 估计。

二、主要结果

定理7.5.1 对线性模型 (7.5.1) , 设β 和σ^2 的GLS 估计由(7.5.3) 给出，则有

(a) $E(\widehat{\beta}^*)=\beta$;

(b) $Cov(\widehat{\beta}^*)=\sigma^2(X'\Sigma^{-1}X)^{-1}$;

(c) 对 $\forall c_{p\times 1}$, $c'\widehat{\beta}^*$ 为 $c'\beta$ 的BLU 估计；

(d) $\widehat{\sigma}_*^2$ 为 σ^2 的无偏估计。

证明： (a) $E(\widehat{\beta}^*)=(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}E(y)=(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}X\beta=\beta.$

(b) 直接计算可得

$$\begin{aligned}Cov(\widehat{\beta}^*)&=Cov\big((X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y\big)\\&=(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}\cdot Cov(y)\cdot\Sigma^{-1}X(X'\Sigma^{-1}X)^{-1}\\&=\sigma^2(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}\cdot\Sigma\cdot\Sigma^{-1}X(X'\Sigma^{-1}X)^{-1}\\&=\sigma^2(X'\Sigma^{-1}X)^{-1}.\end{aligned}$$

(c) 将模型(7.5.1) 转换成(7.5.2), 由§7.3 G-M 定理可知，$c'\widehat{\beta}^*$ 为 $c'\beta$ 的BLU 估计。

(d) 显然。

注： 此时 $c'\widehat{\beta}^*$ 为 $c'\beta$ 的BLU 估计，而 $c'\widehat{\beta}=c'(X'X)^{-1}X'y$ 仍是 $c'\beta$ 的无偏估计，因为 $E(c'\widehat{\beta})=c'(X'X)^{-1}X'\cdot X\beta=c'\beta$，因此为无偏估计，但不再是最小方差的无偏估计，因为由定理7.5.1(3) 可知 $Var(c'\widehat{\beta}^*)\le Var(c'\widehat{\beta})$.

例3.6.1 广义LS 估计最简单的情形是取 $\Sigma=diag\big(\sigma_1^2,\cdots,\sigma_n^2\big)$, 此处 $\sigma_1^2,\sigma_2^2,\cdots,\sigma_n^2$ 不全相等，求 β 的GLS 估计。

解： GLS 估计为 $\widehat{\beta}^*=(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y.$

记

$$X=\begin{pmatrix}\tilde{x}_1'\\ \vdots\\ \tilde{x}_n'\end{pmatrix},\quad \tilde{x}_i'=(x_{i1},\cdots,x_{ip}),\quad y=\begin{pmatrix}y_1\\ \vdots\\ y_n\end{pmatrix},$$

则有

$$X'\Sigma^{-1}X=(\tilde{x}_1,\cdots,\tilde{x}_n)\begin{pmatrix}\sigma_1^{-2}&&\mathbb{0}\\&\ddots&\\ \mathbb{0}&&\sigma_n^{-2}\end{pmatrix}\begin{pmatrix}\tilde{x}_1'\\ \vdots\\ \tilde{x}_n'\end{pmatrix}=\sum_{i=1}^n\frac{\tilde{x}_i\tilde{x}_i'}{\sigma_i^2},$$

$$X'\Sigma^{-1}y=(\tilde{x}_1,\cdots,\tilde{x}_n)\begin{pmatrix}\sigma_1^{-2}&&\mathbb{0}\\&\ddots&\\ \mathbb{0}&&\sigma_n^{-2}\end{pmatrix}\begin{pmatrix}y_1\\ \vdots\\ y_n\end{pmatrix}=\sum_{i=1}^n\frac{y_i\tilde{x}_i}{\sigma_i^2},$$

从而

$$\widehat{\beta}^*=(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y=\left(\sum_{i=1}^n\frac{\tilde{x}_i\tilde{x}_i'}{\sigma_i^2}\right)^{-1}\cdot\left(\sum_{i=1}^n\frac{y_i\tilde{x}_i}{\sigma_i^2}\right),\tag{7.5.4}$$

从这个表达式我们可以看出，上式中两个和式分别是$\tilde{x}_i\tilde{x}_i'$ 和$y_i\tilde{x}_i$ 加权和，所用的“权”都是$1/\sigma_i^2$. 故 $\widehat{\beta}^*$ 也称为加权LS 估计。

§7.6　一般线性假设的检验

一、引言

考虑正态线性回归模型：

$$y_{n\times 1}=X_{n\times p}\beta_{p\times 1}+e_{n\times 1},\quad e\sim N_n(0,\sigma^2 I),\tag{7.6.1}$$

其中 $R(X)=p$，本节讨论一般线性假设：

$$H_0:\ A\beta=b,\tag{7.6.2}$$

其中A和β分别为$m\times p$和$p\times 1$ 矩阵，$R(A)=m$ (即，行满秩).

设 $\beta=(\beta_0,\beta_1,\cdots,\beta_{p-1})'$, 常见的检验问题，皆可用 (7.6.2) 式表示。例如：

1. 检验常数项 $H_0:\ \beta_0=0$, 则检验问题可表示为：

$$H_0: A\beta=0,\ \ \text{其中}\ A=(1,0,0,\cdots,0),\ b=0\,;$$

2. 检验k个回归系数为零 $H_0:\beta_1=\cdots=\beta_k=0,\ 1\le k<p-1$, 则检验问题可表示为：

$$H_0: A\beta=0,\ \ \text{其中}\ A=\left(\mathbb{0}_{k\times 1}\vdots I_k\vdots \mathbb{0}_{k\times(p-1-k)}\right),\ b=0;$$

3. 检验所有回归系数为零 $H_0:\beta_1=\cdots=\beta_{p-1}=0$, 则检验问题可表示为：

$$H_0: A\beta=0,\ A=(\mathbb{0}_{(p-1)\times 1}\vdots I_{p-1}),\ b=0\,.$$

下面导出假设检验思想：对模型 (7.6.1) 应用 LS 法，设 $\widehat{\beta}$ 为 β 的 LS 估计，则残差平方和为

$$RSS=(y-X\widehat{\beta})'(y-X\widehat{\beta})=y'\big(I-X(X'X)^{-1}X'\big)y,\tag{7.6.3}$$

其中 $\widehat{\beta}=(X'X)^{-1}X'y$. RSS 反映了实验数据与模型的拟合程度。RSS 越小，拟合越好。将 (7.6.1)式 加上约束条件(7.6.2)，获得有约束条件的 LS 估计

$$\widehat{\beta}_H=\widehat{\beta}-(X'X)^{-1}A'\big(A(X'X)^{-1}A'\big)^{-1}(A\widehat{\beta}-b),\tag{7.6.4}$$

相应的残差平方和为：

$$RSS_H=(y-X\widehat{\beta}_H)'(y-X\widehat{\beta}),\tag{7.6.5}$$

数据与新模型 $y=X\beta+e,\ A\beta=b$ 之间拟合好坏用 RSS_H 来衡量。此模型中 β 的变化范围缩小了 (因为假设 $A\beta=0$，即假设一部分自变量不起作用，如果上述假设不成立，即 $A\beta\neq 0$ 实际成立，则这些变量的作用归入误差平方和，使 RSS_H 增大)。故新模型的拟合不会比旧的好，即 $RSS_H\ge RSS$。如果真正的参数确实满足线性假设 (7.6.2)，则加上约束和不加约束条件本质上是一样的，这时对无约束条件和有约束的模型，数据的拟合会是一样的，RSS_H-RSS 应当较

小。反过来，若真正参数不满足 (7.6.2)式，即 $H_0: A\beta = 0$ 不成立，则 $RSS_H - RSS$ 应当较大，因此拒绝 H_0，故用 $(RSS_H - RSS)/RSS$ 相对大小来决定 (7.6.2) 式是否成立是合适的。

注： RSS 的计算公式:

$$
\begin{aligned}
RSS &= (y - X\widehat{\beta})'(y - X\widehat{\beta}) \\
&= y'y + \widehat{\beta}'X'X\widehat{\beta} - 2\widehat{\beta}'X'y \\
&= y'y + \widehat{\beta}'X'y - 2\widehat{\beta}'X'y = y'y - \widehat{\beta}'X'y.
\end{aligned}
$$

二、F 检验

定理7.6.1 对于正态线性回归模型 (7.6.1) 有

(a) $\dfrac{RSS}{\sigma^2} \sim \chi^2_{n-p}$;

(b) 若 (7.6.2) 式成立，则 $\dfrac{RSS_H - RSS}{\sigma^2} \sim \chi^2_m$;

(c) RSS 与 $RSS_H - RSS$ 独立;

(d) 当 (7.6.2) 成立时，

$$
F_H = \frac{(RSS_H - RSS)/m}{RSS/(n-p)} \sim F_{m,n-p},
$$

这里 $F_{m,n}$ 表示自由度分别为 m, n 的 F 分布。

证明： (a) $\dfrac{RSS}{\sigma^2} = \left(\dfrac{y}{\sigma}\right)'\big(I - X(X'X)^{-1}X'\big)\left(\dfrac{y}{\sigma}\right) \sim \chi^2_{n-p}$ 在定理7.3.3中证过。

(b) 对 RSS_H 进行分解

$$
\begin{aligned}
RSS_H &= \|y - X\widehat{\beta}_H\|^2 = \|y - X\widehat{\beta} + X(\widehat{\beta} - \widehat{\beta}_H)\|^2 \\
&= \|y - X\widehat{\beta}\|^2 + \|X(\widehat{\beta} - \widehat{\beta}_H)\|^2 + 2(\widehat{\beta} - \widehat{\beta}_H)'X'(y - X\widehat{\beta}) \\
&\triangleq RSS + \|X(\widehat{\beta} - \widehat{\beta}_H)\|^2 + 2I,
\end{aligned}
$$

其中 $I = (\widehat{\beta} - \widehat{\beta}_H)'(X'y - X'X\widehat{\beta}) = 0$. 这是由于 $\widehat{\beta}$ 满足正则方程 $X'X\widehat{\beta} = X'y$，故 $I = 0$ 成立。故有

$$
RSS_H - RSS = (\widehat{\beta} - \widehat{\beta}_H)'X'X(\widehat{\beta} - \widehat{\beta}_H).
$$

将 $\widehat{\beta} - \widehat{\beta}_H = (X'X)^{-1}A'\big(A(X'X)^{-1}A'\big)^{-1}(A\widehat{\beta} - b)$ 代入到上式得

$$
\begin{aligned}
&RSS_H - RSS \\
&= (A\widehat{\beta} - b)'\big(A(X'X)^{-1}A'\big)^{-1}A(X'X)^{-1}X'X(X'X)^{-1}A'\big(A(X'X)^{-1}A'\big)^{-1}(A\widehat{\beta} - b) \\
&= (A\widehat{\beta} - b)'\big(A(X'X)^{-1}A'\big)^{-1} \cdot A(X'X)^{-1}A' \cdot \big(A(X'X)^{-1}A'\big)^{-1}(A\widehat{\beta} - b) \\
&= (A\widehat{\beta} - b)'\big(A(X'X)^{-1}A'\big)^{-1}(A\widehat{\beta} - b), \qquad (7.6.6)
\end{aligned}
$$

因此令 $Z = A\widehat{\beta} - b$，则在H_0成立时

$$
Z \sim N_m\big(0, \sigma^2 A(X'X)^{-1}A'\big) = N_m(0, \Sigma).
$$

因为 $R(A(X'X)^{-1})=m$, 故由定理7.2.11可知

$$Z'\Sigma^{-1}Z=\frac{(A\widehat{\beta}-b)'\big(A(X'X)^{-1}A'\big)^{-1}(A\widehat{\beta}-b)}{\sigma^2}\sim\chi^2_m.$$

(c) 因为

$$RSS_H-RSS=(A\widehat{\beta}-b)'\big(A(X'X)^{-1}A'\big)^{-1}(A\widehat{\beta}-b),$$

将

$$\begin{aligned}A\widehat{\beta}-b&=A(X'X)^{-1}X'(X\beta+e)-b\\&=(A\beta-b)+A(X'X)^{-1}X'e\end{aligned}$$

代入上式得

$$\begin{aligned}&RSS_H-RSS\\&=\big[e'X(X'X)^{-1}A'+(A\beta-b)'\big]\big(A(X'X)^{-1}A'\big)^{-1}\big[A(X'X)^{-1}X'e+(A\beta-b)\big]\\&=e'X(X'X)^{-1}A'\big(A(X'X)^{-1}A'\big)^{-1}A(X'X)^{-1}X'e\\&\quad+(A\beta-b)'\big(A(X'X)^{-1}A'\big)^{-1}(A\beta-b)\\&\quad+2(A\beta-b)'\big(A(X'X)^{-1}A'\big)^{-1}A(X'X)^{-1}X'e\\&=e'Me+\Delta+2c'e,\end{aligned}$$

此处

$$\begin{aligned}&M=X(X'X)^{-1}A'\big(A(X'X)^{-1}A'\big)^{-1}A(X'X)^{-1}X',\\&\Delta=(A\beta-b)'\big(A(X'X)^{-1}A'\big)^{-1}(A\beta-b),\\&c'=(A\beta-b)'\big(A(X'X)^{-1}A'\big)^{-1}(X'X)^{-1}X'.\end{aligned}$$

由于 Δ 与误差向量 e 无关，故其是非随机的。若记 $N=I-P_X$，$P_X=X(X'X)^{-1}X'$，注意到 $NX=0$，则

$$RSS=y'(I-P_X)y=(X\beta+e)'N(X\beta+e)=\beta'X'NX\beta+2e'NX\beta+e'Ne=e'Ne.$$

因此，要证 RSS_H-RSS 与 RSS 独立，只要证 $e'Me$ 与 $e'Ne$ 独立、$c'e$ 与 $e'Ne$ 独立即可。因为 $e\sim N(0,\sigma^2I)$，由 §7.3.2 关于两个二次型独立性和线性型与二次型的独立性的判别方法，只需验证 $MN=0$、$c'N=0$. 而根据 M, N 和 c 的定义及 $NX=0$，容易得到：

$$\begin{aligned}&M\cdot N=X(X'X)^{-1}A'\big(A(X'X)^{-1}A'\big)^{-1}A(X'X)^{-1}X'(I-P_X)=0;\\&c'N=(A\beta-b)'\big(A(X'X)^{-1}A'\big)^{-1}(X'X)^{-1}X'N=0,\end{aligned}$$

故 RSS_H-RSS 与 RSS 独立。

(d) 按 F 分布的定义，当 (7.6.2) 成立，则 (a)、(b)、(c) 成立，故有

$$F_H = \frac{(RSS_H - RSS)/m}{RSS/(n-p)} \sim F_{m,n-p},$$

因此当 $F_H > F_{m,n-p}(\alpha)$ 时否定 H_0，其中 α 为检验水平。由 (7.6.6) 式可知可将上述检验统计量改写为

$$F_H = \frac{(A\beta - b)'\big(A(X'X)^{-1}A'\big)^{-1}(A\beta - b)/m}{RSS/(n-p)}. \tag{7.6.7}$$

三、RSS_H 计算的方法

考虑假设检验：

$$H_0 : A\beta = b$$

将它融入到原来模型中去，化成一个无约束的回归模型，称为约简模型。例如把模型 (7.6.1) 写成分量形式：

$$y_i = \beta_0 + x_{i1}\beta_1 + \cdots + x_{ip-1}\beta_{p-1} + e_i. \tag{7.6.8}$$

若要检验 $H_0 : \beta_1 = \beta_2 = \beta_3$，这时线性假设具有形式 $A\beta = 0$，其中

$$A = \begin{pmatrix} 0 & 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & 0 & -1 & \cdots & 0 \end{pmatrix},\ \beta' = (\beta_0, \beta_1, \cdots, \beta_{p-1}).$$

将 $H_0 : A\beta = 0$ 代入到 (7.6.8) 式中，得

$$y_i = \beta_0 + (x_{i1} + x_{i2} + x_{i3})\beta_1 + x_{i4}\beta_4 + \cdots + x_{i,p-1}\beta_{p-1} + e_i,\ i = 1, \ldots n.$$

用矩阵表示

$$y_{n\times 1} = \widetilde{X}_{n\times(p-2)}\alpha + e\text{，其中}\alpha' = (\beta_0, \beta_1, \beta_4, \cdots, \beta_{p-1}),$$

则 α 的 LS 估计为

$$\widehat{\alpha} = (\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'y,$$

其残差平方和为

$$RSS_H = y'y - \widehat{\alpha}'\widetilde{X}'y. \tag{7.6.9}$$

又因为

$$RSS = y'y - \widehat{\beta}X'y,$$

因此

$$F_H = \frac{(RSS_H - RSS)/m}{RSS/(n-p)} = \frac{(\widehat{\beta}'X'y - \widehat{\alpha}'\widetilde{X}'y)/m}{(y'y - \widehat{\beta}'X'y)/(n-p)}.$$

例7.6.1 假设

$$\begin{aligned} y_1 &= \beta_1 + e_1 \\ y_2 &= 2\beta_1 - \beta_2 + e_2 \\ y_3 &= \beta_1 + 2\beta_2 + e_3, \end{aligned}$$

其中 $e = (e_1, e_2, e_3)' \sim N_3(0, \sigma^2 I)$. 求 $H_0 : \beta_1 = \beta_2$.

解： 将数据模型写成矩阵形式如下

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 2 & -1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ e_3 \end{pmatrix}.$$

线性假设 H_0 等价于

$$H_0 : \ (\ 1 \quad -1 \ ) \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = 0.$$

于是，对现在的情形 $A = (\ 1 \quad -1 \ )$, $n = 3$, $p = 2$, $m = 1$. 易求原模型中 β 的LS估计为

$$\widehat{\beta} = \begin{pmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{pmatrix} = (X'X)^{-1}X'y = \begin{pmatrix} 6 & 0 \\ 0 & 5 \end{pmatrix}^{-1} \begin{pmatrix} y_1 + 2y_2 + y_3 \\ -y_2 + 2y_3 \end{pmatrix} = \begin{pmatrix} \dfrac{1}{6}(y_1 + 2y_2 + y_3) \\ \dfrac{1}{5}(-y_2 + 2y_3) \end{pmatrix},$$

所以

$$RSS = y'y - \widehat{\beta}'X'y = \sum_{i=1}^{3} y_i^2 - 6\widehat{\beta}_1^2 - 5\widehat{\beta}_2^2.$$

为求 $H_0 : \beta_1 = \beta_2$ 下的残差平方和 RSS_H，将 $\beta_1 = \beta_2 = \alpha$ (即记它们的公共值为 α) 代入原模型，得到约简模型

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix} \alpha + \begin{pmatrix} e_1 \\ e_2 \\ e_3 \end{pmatrix} = \widetilde{X}\alpha + e.$$

从这个模型我们容易得到 α 的LS估计 (也就是 β_1 和 β_2 的约束LS估计)为

$$\widehat{\alpha} = (\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'y = \frac{1}{11}(y_1 + y_2 + 3y_3),$$

于是残差平方和为

$$RSS_H = y'y - \widehat{\alpha}'\widetilde{X}'y = \sum_{i=1}^{3} y_i^2 - \frac{1}{11}(y_1 + y_2 + 3y_3)^2.$$

从而 H_0 的检验统计量为

$$F_H = \frac{(RSS_H - RSS)/1}{RSS/1} = \frac{RSS_H - RSS}{RSS}.$$

当 H_0 成立时，$F_H \sim F_{1,1}$. 当$F_H > F_{1,1}(\alpha)$ 是否定H_0.

§7.7 回归方程和回归系数的显著性检验

一、回归方程的显著性检验

将正态线性模型 (7.6.1) 写成下列形式：

$$y_i = \beta_0 + x_{i1}\beta_1 + \cdots + x_{ip-1} + e_i,\ i = 1, 2, \ldots, n. \tag{7.7.1}$$

其中 $e_i \sim N(0, \sigma^2)$ 且相互独立. 求下列检验

$$H_0:\ \beta_1 = \beta_2 = \cdots = \beta_{p-1} = 0. \tag{7.7.2}$$

当否定 H_0 时说明至少有一个 $\beta_i \neq 0$，说明 因变量Y 至少与一个自变量有关。若 H_0 被接受，说明相对于误差而言，所有自变量对 因变量Y 的影响是不重要的，回归方程无意义，重要的变量未选入。

检验方法如下：若取 $A_{(p-1)\times p} = \begin{pmatrix} \mathbb{0} & I_{p-1} \end{pmatrix}$，$\beta' = (\beta_0, \beta_1, \cdots, \beta_{p-1})'$. 显然 $R(A) = p - 1$，则

$$H_0:\ A\beta = 0.$$

由上节可知检验统计量为

$$F = \frac{(RSS_H - RSS)/(p-1)}{RSS/(n-p)},$$

其中 $RSS = y'y - \widehat{\beta}'X'y,\ \widehat{\beta} = (X'X)^{-1}X'y$，如何求 RSS_H？将 H_0 代入到模型 (7.7.1) 变为

$$y_i = \beta_0 + e_i,\ i = 1, 2, \cdots, n \Longleftrightarrow y_{n\times 1} = \beta_0 \mathbb{1}_n + e. \tag{7.7.3}$$

从而 β_0 的 LS 估计为

$$\widehat{\beta}_0^* = (\mathbb{1}'\mathbb{1})^{-1}\mathbb{1}'y = \frac{1}{n}\sum_{i=1}^{n} y_i = \overline{y},$$

对应的残差平方和

$$RSS_H = y'y - (\widehat{\beta}_0^*)'\mathbb{1}'y = y'y - n\overline{y}^2 = \sum_{i=1}^{n}(y_i - \overline{y})^2, \tag{7.7.4}$$

故有

$$RSS_H - RSS = \widehat{\beta}'X'y - (\widehat{\beta}_0^*)'\mathbb{1}'y = SS_{\text{回}}, \tag{7.7.5}$$

它表示回归平方和的减少量。于是检验统计量为

$$F = \frac{SS_{\text{回}}/(p-1)}{RSS/(n-p)}. \tag{7.7.6}$$

记 TSS 表示总的平方和，则有

$$TSS = RSS + SS_{\text{回}}, \tag{7.7.7}$$

其中 $SS_{回} = \widehat{\beta}'X'y - (\widehat{\beta}_0^*)'\mathbb{1}'y$ 反应回归自变量的改变引起残差平方和的改变量。RSS 是误差平方和，包刮随机误差、模型误差、重要自变量的遗漏以及非线性等对平方和的影响。TSS 是总的平方和，(7.7.6) 是将回归平方和与误差平方和进行比较，当相对误差较大时就否定 H_0，通常将平方和除以自由度称为均方和，得下列方差分析表

方差源	平方和	自由度	均　方	$F_{比}$	显著性(p值)
回归	$SS_{回}$	$p-1$	$MS_{回} = SS_{回}/(p-1)$	$F_{回} = MS_{回}/MS_e$	
误差	RSS	$n-p$	$MS_e = RSS/(n-p)$		
总计	TSS	$n-1$			

查 F 表可得出 $F_{p-1,n-p}(\alpha)$ 的值，当 $F > F_{p-1,n-p}(\alpha)$ 时拒绝原假设，等价于检验的 p 值满足

$$p = P(F_{p-1,n-p} > F_{回}|_{H_0}) < \alpha$$

时，否定 H_0.

结论： (1) 当拒绝 H_0 时，说明回归方程有意义，Y 与自变量 $X_1, X_2, \cdots, X_{p-1}$ 整体相关，即至少有一个自变量被认为是重要的，下一步是逐个检验，看看哪些自变量是重要的，哪些是不重要的。

(2) 当 H_0 被接受，即认为 $\beta_1 = \beta_2 = \cdots = \beta_{p-1} = 0$，此时有两种可能：① 模型误差太大，即 F 统计量的分母太大，即使自变量对 Y 有一定影响，也显示不出来。此时造成分母太大 (即误差平方和太大) 可能有重要自变量遗漏，它的影响进入到 RSS 中，或者Y对某些自变量有非线性相依关系。② 回归自变量对 Y 影响确实很小，此时应当认为回归方程无意义。

例7.7.1 煤净化问题。

下表给出煤净化过程的一组数据，Y 表示净化后煤溶液中所含杂质的量，这是衡量净化效率的指标。X_1 表示输入净化过程的溶液所含的杂质与煤之比，X_2 表示溶液 PH 值，X_3 表示溶液的流量，目的通过数据建立 Y 与 X_1, X_2, X_3 的经验关系，通过控制某些自变量来提高净化效率。

考虑线性关系

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + e,$$

应用最小二乘法得到经验回归方程

$$\widehat{Y} = 397.087 - 110.750X_1 + 15.583X_2 - 0.058X_3.$$

为了检验回归方程的显著性，令

$$H_0 : \beta_1 = \beta_2 = \beta_3 = 0,$$

编　　号	x_1	x_2	x_3	y
1	1.50	6.00	1 315	243
2	1.50	6.00	1 315	261
3	1.50	9.00	1 890	244
4	1.50	9.00	1 890	285
5	2.00	7.50	1 575	202
6	2.00	7.50	1 575	180
7	2.00	7.50	1 575	183
8	2.00	7.50	1 575	207
9	2.50	9.00	1 315	216
10	2.50	9.00	1 315	160
11	2.50	6.00	1 890	104
12	2.50	6.00	1 890	110

计算出下列方差分析表

方差源	平方和	自由度	均　方	$F_{比}$	p　值
回归	31 156.02	3	10 385.33	23.82	0.000 2
误差	3 486.89	8	435.85	／	／
总计	34 642.91	11	／		

给定水平 $\alpha = 0.05$，查表得 $F_{3,8}(0.05) = 4.07$，p值 $= 0.002 < \alpha$，则我们否定 H_0. 认为 Y 与 X_1, X_2, X_3 有一定关系。X_1 的系数为负值，当 X_1 增加时 Y 减少，与 X_1 实际意义矛盾，当杂质比增加时，Y 也应该增加，认为回归方程有问题，要分析原因，数据的记录有无问题，12 个数据量太小，要增加数据量再重新计算。 □

二、回归系数的显著性检验

1. 引言

由上述对回归方程的显著性检验的讨论可知对下列方程

$$y_i = \beta_0 + X_{i1}\beta_1 + X_{i2}\beta_2 + \cdots + X_{i,p-1}\beta_{p-1} + e_i,\ i = 1, 2, \cdots n.$$

的检验问题

$$H_0:\ \beta_1 = \beta_2 = \cdots = \beta_{p-1} = 0,$$

若拒绝 H_0，表明回归方程有意义，Y 至少与 $X_1, \cdots, X_{p-1}$ 之一有重要关系。Y 具体与哪些自变量有关，与哪些自变量关系不重要？需要进一步作下列检验

$$H_i:\ \beta_i = 0,\ i = 1, 2, \cdots, p-1.$$

2. 检验的求法

模型可写为：

$$Y_{n\times 1}=X_{n\times p}\beta_{p\times 1}+e_{n\times 1},\quad e\sim N_n(0,\sigma^2 I),$$

其中 $\beta'=(\beta_0,\beta_1,\cdots,\beta_{p-1)})'$，则 β 的LS估计为 $\widehat{\beta}=(X'X)^{-1}X'y$，且 $\widehat{\beta}\sim N_p\big(\beta,\sigma^2(X'X)^{-1}\big)$. 若记

$$(X'X)^{-1}_{p\times p}=C_{p\times p}=(c_{ij})=\begin{pmatrix} c_{00} & c_{01} & \cdots & c_{0,p-1}\\ c_{10} & c_{11} & \cdots & c_{1,p-1}\\ \vdots & \vdots & \ddots & \vdots\\ c_{p-1,0} & c_{p-1,1} & \cdots & c_{p-1,p-1}\end{pmatrix},$$

则 $\widehat{\beta}'=(\widehat{\beta}_0,\widehat{\beta}_1,\cdots,\widehat{\beta}_{p-1)})'$ 之分量满足

$$\widehat{\beta}_i\sim N_1(\beta_i,\sigma^2 c_{ii}),\ i=1,2,\cdots,n.$$

而 σ^2 的LS估计为

$$\widehat{\sigma}^2=\frac{\|y-X\widehat{\beta}\|^2}{n-p}=\frac{y'y-\widehat{\beta}'X'y}{n-p},$$

则将 $\widehat{\beta}_i$ 标准化得到：

$$\frac{\widehat{\beta}_i-\beta_i}{\sigma\sqrt{c_{ii}}}\sim N(0,1).$$

又将上式中的σ 用$\widehat{\sigma}$代替，再由$\widehat{\sigma}^2$与 $\widehat{\beta}$ 独立(因而与 $\widehat{\beta}_i$ 独立)，可知当H_i给定时有

$$t_i=\frac{\widehat{\beta}_i/(\sigma\sqrt{c_{ii}})}{\widehat{\sigma}/\sigma}=\frac{\widehat{\beta}_i}{\widehat{\sigma}\sqrt{c_{ii}}}\sim t_{n-p},$$

则 t_i 可作为检验统计量，当 $|t_i|=\left|\dfrac{\widehat{\beta}_i}{\widehat{\sigma}\sqrt{c_{ii}}}\right|>t_{n-p}\Big(\dfrac{\alpha}{2}\Big)$ 时否定 H_i. 方差分析表如下 (标准误差 $\sqrt{Var(\widehat{\beta}_i)}=\sigma\sqrt{c_{ii}}$)：

变量或系数	LS 估计	标准误差的估计	t_i	$p_i=P(t_{n-p}>\|t_i\|\|H_i)$
β_0	$\widehat{\beta}_0$	$\widehat{\sigma}\sqrt{c_{00}}$	t_0	
β_1	$\widehat{\beta}_1$	$\widehat{\sigma}\sqrt{c_{11}}$	t_1	
β_2	$\widehat{\beta}_2$	$\widehat{\sigma}\sqrt{c_{22}}$	t_2	
$\vdots$	$\vdots$	$\vdots$	$\vdots$	
β_{p-1}	$\widehat{\beta}_{p-1}$	$\widehat{\sigma}\sqrt{c_{p-1p-1}}$	t_{p-1}	

当 $p_i<\alpha$ (或$|t_i|>t_{n-p}(\frac{\alpha}{2})$)) 时，否定 $H_i:\beta_i=0$.

例7.7.2 (续例7.7.1). 求检验 $H_i:\beta_i=0,\ i=1,2,3$ (检验水平 $\alpha=0.05$).

解：按公式 $t_i=\dfrac{\widehat{\beta}_i}{\widehat{\sigma}\sqrt{c_{ii}}}$ 算得 $t_1=-7.50,\ t_2=3.17,\ t_3=-2.27,\ n=12,\ p=4,\ n-p=8,\ \alpha=0.05,\ t_{n-p}(\alpha)=t_8(0.05)=2.306$，得下列方差分析表：

变 量	回归系数估计值	标准误差 $\hat{\sigma}\sqrt{c_{ii}}$	t_i	$p_i = P(t_{n-p} > \|t_i\| \| H_i)$
X_1	-110.750	14.762	-7.502	0.000 1
X_2	15.583	4.921	3.167	0.013 3
X_3	-0.058	0.026	-2.27	0.052 6

对 $i = 1, 2$, 有 $p_i < 0.05$ (或$|t_i| > t_8(0.025) = 2.306$) ，否定 H_i，即认为 $\beta_1 \neq 0$ 和 $\beta_2 \neq 0$ ，它们是有重要影响的变量。

三、回归系数 β_i 的区间估计问题

由前面推导可知：$\widehat{\beta}_i \sim N_1(\beta, \sigma^2 c_{ii})$，故知：

$$\frac{(\widehat{\beta}_i - \beta_i)/(\sigma\sqrt{c_{ii}})}{\widehat{\sigma}/\sigma} = \frac{\widehat{\beta}_i - \beta_i}{\widehat{\sigma}\sqrt{c_{ii}}} \sim t_{n-p},$$

从而有

$$P\left(\left|\frac{\widehat{\beta}_i - \beta_i}{\widehat{\sigma}\sqrt{c_{ii}}}\right| \leq t_{n-p}\left(\frac{\alpha}{2}\right)\right) = 1 - \alpha.$$

解挂号中的不等式得到β_i的置信系数 $1-\alpha$ 的置信区间

$$\left[\widehat{\beta}_i - \widehat{\sigma}\sqrt{c_{ii}} t_{n-p}(\alpha/2),\ \widehat{\beta}_i + \widehat{\sigma}\sqrt{c_{ii}} t_{n-p}(\alpha/2)\right].$$

§7.8 预 测 问 题

一、引言

设所要研究的模型为

$$y_i = \beta_0 + X_{i1}\beta_1 + \cdots + X_{i,p-1}\beta_{p-1} + e_i = x_i'\beta + e_i, \quad i = 1, 2, \cdots, n\ ,. \tag{7.8.1}$$

其中$e_1, \cdots, e_n\ iid\ \sim\ (0, \sigma^2)$, $\beta = (\beta_0, \beta_1, \cdots, \beta_n)'$, $x_i = (1, x_{i1}, \cdots, x_{ip-1})'$, $i = 1, \cdots, n$. 前面讨论了 β 和 σ^2 的LS估计、检验和区间估计问题。另一个问题是若已知

$$x_0' = x_{n+1}' = (1, x_{n+1,1}, \cdots, x_{n+1,p-1}),$$

问 $y_0 = y_{n+1}$ 的取值能否通过 x_0' 而获得？这有两种情形：

(1) 根本就没有做试验，想弄清楚在试验点外，y 会取什么值？

(2) 试验是做了，但有一定周期，如小麦还未收获，要预测小麦产量。

y_0 可表示为

$$y_0 = x_0'\beta + e_0,\ e_0 \sim (0, \sigma^2). \tag{7.8.2}$$

要预测随机变量 y_0 之值，y_0 有两部分：(1) $x_0'\beta$ 未知，用 $\widehat{\beta}$ 代替 β，因而用 $x_0'\widehat{\beta}$ 预测 $x_0'\beta$ 之值; (2) $E(e_0) = 0$, 用 0 预测 e_0 之值，故知用

$$\widehat{y}_0 = x_0'\widehat{\beta} \tag{7.8.3}$$

作为 y_0 的预测值称为点预测。

预测与估计得不同点如下：

(1) 估计问题中待估的是未知参数。如 $\beta_0, \beta_i, \sigma^2$ 或其函数 $x'\beta$ 等，它们是非随机的。而预测量 y_0 是一个随机变量，而不是未知参数。

(2) 从精度上讲预测值的精度要比估计的精度差。$y_0 = x_0'\beta + e$ 和 $\mu = x_0'\beta$ 二者的预测量和估计量皆用 $x_0'\widehat{\beta}$ 表示，但精度不同。用 $x_0'\widehat{\beta}$ 估计 $\mu = x_0'\beta$ 有误差，用 0 去预测 e_0 也有误差，它比估计问题多一重预测误差。

二、预测值的性质

1. $\widehat{y}_0 = x_0'\widehat{\beta}$ 为 y_0 的无偏预测，即 $E(\widehat{y}_0 - y_0) = 0$，这是因为

$$E(\widehat{y}_0) = x_0'\beta\ ,\ E(y_0) = x_0'\beta \implies E(\widehat{y}_0 - y_0) = x_0'\beta - x_0'\beta = 0.$$

2. 用 G-M 定理，可证 $\widehat{y}_0$ 为 y 的一切线性无偏预测中方差最小者。事实上假设$a'y$ 是y_0的任一线性无偏预测，则$E(a'y) = E(y_0) = x_0'\beta$, 因此$a'y$可以看作$x_0'\beta$的一个线性无偏估计。而预测量 $\widehat{y}_0 = x_0'\widehat{\beta}$ 也可视为 $x_0'\beta$ 的一个线性无偏估计，故根据 G-M 定理，总有

$$Var(a'y) \geq Var(x_0'\widehat{\beta}).$$

这就证明了我们的结论。

注： y_0 的预测量与参数函数 $\mu_0 = x_0'\beta$ 的 LS 估计 $\widehat{\mu}_o = x_0'\widehat{\beta}$ 表达式相同，它们的实际意义是不同的。引进预测偏差 $d_1 = \widehat{y}_0 - y_0$ 和估计偏差 $d_2 = \widehat{\mu}_0 - \mu_0 = \widehat{\mu}_0 - x_0'\beta$ 计算它们的方差

$$Var(d_1) = Var(\widehat{y}_0) + Var(y_0) = \sigma^2 x_0'(X'X)^{-1}x_0 + \sigma^2 = \sigma^2\big(1 + x_0'(X'X)^{-1}x_0\big); \qquad (7.8.4)$$

另一方面

$$Var(d_2) = Var(\widehat{\mu}_o) = \sigma^2 x_0'(X'X)^{-1}x_0 < Var(d_1).$$

这样的差别来源于被预测量y_0是随机变量，而被估计量$x_0'\beta$是非随机变量。

三、区间预测

若进一步假定 $e_1, e_2, \cdots, e_n$ 和 $e_0 \triangleq e_{n+1}\ \ iid \sim N(0, \sigma^2)$，则由前面的计算可知

$$\widehat{y}_0 - y_0 \sim N\big(0, \sigma^2(1 + x_0'(X'X)^{-1}x_0)\big),$$

且 $\widehat{y}_0 - y_0$ 和 $\widehat{\sigma}^2 = \|y - X\widehat{\beta}\|^2/(n-p)$ 相互独立。将 $\widehat{y}_0 - y_0$ 标准化得

$$\frac{\widehat{y}_0 - y_0}{\sigma\sqrt{1 + x_0'(X'X)^{-1}x_0}} \sim N(0, 1),$$

将上式中的 σ 用 $\widehat{\sigma}$ 代替，得到

$$\frac{\widehat{y}_0 - y_0}{\widehat{\sigma}\sqrt{1 + x_0'(X'X)^{-1}x_0}} \sim t_{n-p}, \qquad (7.8.5)$$

对给的的α有

$$P\left(\frac{|\widehat{y}_0-y_0|}{\widehat{\sigma}\sqrt{1+x_0'(X'X)^{-1}x_0}}\le t_{n-p}\left(\frac{\alpha}{2}\right)\right)=1-\alpha.$$

由此可得 y_0 的 $1-\alpha$ 的预测区间为

$$\left[\widehat{y}_0-t_{n-p}\left(\frac{\alpha}{2}\right)\widehat{\sigma}\sqrt{1+x_0'(X'X)^{-1}x_0},\ \widehat{y}_0+t_{n-p}\left(\frac{\alpha}{2}\right)\widehat{\sigma}\sqrt{1+x_0'(X'X)^{-1}x_0}\right].$$

例7.8.1 (续例7.7.1). 对煤净化问题中已建立经验回归方程，求$y_0=x_0'\beta+e$ 的概率为95%的预测区间，其中 $x_0'=(1.5,7.50,1\,315)$.

解： 代入计算可得

$$\widehat{y}_0=397.087-110.750\times 1.5+15.583\times 7.50+(-0.058)\times 1\,315=271.564,$$

$$\widehat{\sigma}=(435.682)^{\frac{1}{2}}=20.88,\ n-p=8,\ t_8(0.025)=2.306,$$

则 y_0 的预测区间为

$$\begin{aligned}&\left[\ \widehat{y}_0-t_8(0.025)\widehat{\sigma}\sqrt{1+x_0'(X'X)^{-1}x_0},\ \widehat{y}_0+t_8(0.025)\widehat{\sigma}\sqrt{1+x_0'(X'X)^{-1}x_0}\ \right]\\&=[215.756,\ 326.609].\end{aligned}$$

习题

1. 设$X_1,...,X_n$为随机变量，$Y_1=X_1,\ Y_i=X_i-X_{i-1},\ i=2,3,...,n.$记$X=(X_1,...,X_n)',\ Y=(Y_1,...,Y_n)'$
   (a) 若Cov(X)=I,其中I 为n 阶单位阵，求Cov(Y).
   (b) 若Cov(Y)=I, 求Cov(Y).
2. 设X和Y为具有相同方差的两个独立随机变量，证明：Cov(X+Y,X-Y)=0.
3. 设$X_{n\times 1},Y_{m\times 1}$ 均为随机向量，$Cov(X)>0$ (即为正定阵), 求常数矩阵$A_{n\times m}$，使得$Cov(X,Y-AX)=0$.
4. 设随机变量$X_1,...,X_n$ 相互独立，具有公共均值θ 和方差σ^2, 定义$Q=(X_1-X_2)^2+...+(X_{n-1}-X_n)^2$, 求E(Q).
5. 设$X_i\sim N_p(\mu_i,\Sigma_i)$ $i=1,2$, 相互独立，设a, b 为常数，证明

$$aX_1+bX_2\sim N_p\big(a\mu_1+b\mu_2,a^2\Sigma_1+b^2\Sigma_2\big).$$

6. 设$X_1,...,X_n$相互独立，且均服从$N(0,\sigma^2)$,证明

$$X=\frac{1}{n}\sum_{i=1}^{n}X_i,\quad 与\quad \sum_{i=1}^{n-1}(X_i-X_{i+1})^2$$

相互独立。

7. 设$X=(X_1,X_2,X_3)'\sim N_3(\mu,\Sigma)$, 其中

$$\mu=(2,1,2)',\quad \Sigma=\begin{pmatrix}2&1&1\\1&3&0\\1&0&1\end{pmatrix}$$

求$Y_1=X_1+X_2+X_3$ 与$Y_2=X_1-X_2$ 的联合分布。

8. 设$X=(X_1,X_2,X_3)'\sim N_3(\mu,\Sigma)$, 其中

$$\Sigma=\begin{pmatrix}1&\rho&0\\\rho&1&\rho\\0&\rho&1\end{pmatrix}$$

试问ρ 取什么值的时候$X_1+X_2+X_3$与$X_1-X_2-X_3$ 独立?

9. 已知$X\sim N_n(\mu,I)$.

   (a) 求$Y_1=\alpha'X$ 和$Y_2=\beta'X$ 的联合分布，其中α,β 皆为$n\times 1$ 的非随机向量($\alpha\neq k\beta$);

   (b) 若$\alpha'\beta=0$, 证明Y_1 和Y_2 独立.

10. 证明《数理统计》第二章习题16.

11. 设$X\sim N_n(\mu,\Sigma)$，A是$n\times n$对称阵，R(A)=r,证明当$A\Sigma A=A$时，$(X-\mu)'A(X-\mu)\sim\chi^2_r$,其中$r$ 为矩阵A 的秩。

12. 设$X\sim N_n(\mu,\Sigma)$, 试给出二次型$(X-\mu)'A(X-\mu)$ 与$(X-\mu)'B(X-\mu)$ 独立的条件。

13. 设X为二元正态分布，其密度函数为

$$f(x_1,x_2)=k^{-1}exp\{-\frac{1}{2}(x_1^2+2x_2^2-x_1x_2-3x_1-2x_2+4)\},$$

求E(X) 及Cov(X).

14. 利用二次型的均值及二元正态分布, 计算积分

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x^2+xy+3y^2)exp\{-(x^2+2xy+2y^2)\}dxdy.$$

15. 为了考察某种维尼伦纤维的耐水性，安排了一批试验，测得甲醛浓度x 及“缩醛化度”y 的数据如下

x	18	20	22	24	26	28	30
y	26.86	28.35	28.75	28.87	29.75	30.00	30.36

由实际经验和理论可知二者近似线性关系。

(a) 求y 关于x 的经验回归方程，画出原始数据及回归直线的图形。

(b) 对$x_0 = 23$ 求y 的预测值y_0。

16. 设$y_1 = \theta + e_1$, $y_2 = 2\theta - \varphi + e_2$, $y_3 = \theta + 2\varphi + e_3$, 其中$\theta$ 和φ 是未知参数，$E(e_i) = 0$, $Var(e_i) = \sigma^2$, $i = 1, 2, 3$. 且e_1, e_2, e_3 相互独立。

(a) 求θ, φ 的最小二乘估计$\hat{\theta}$, $\hat{\varphi}$.

(b) 求$cov\binom{\hat{\theta}}{\hat{\varphi}}$.

17. 设

$$\begin{cases} y_i = \theta + e_i, & i = 1, 2, \cdots, m. \\ y_{m+i} = \theta + \varphi + e_{m+i}, & i = 1, 2, \cdots, m. \\ y_{2m+i} = \theta - 2\varphi + e_{2m+i}, & i = 1, 2, \cdots, n. \end{cases}$$

其中θ 和φ 是未知参数，各e_i 相互独立，且服从$N(0, \sigma^2)$.

(a) 写出设计阵X.

(b) 求θ, φ 的LS 估计$\hat{\theta}$, $\hat{\varphi}$.

(c) 证明当$m = 2n$ 时，$\hat{\theta}$ 与$\hat{\varphi}$ 不相关。

18. 对正态线性模型$y = X\beta + e$, $e \sim N(0, \sigma^2 I)$, 证明β 的LS 估计与极大似然估计一致。

19. 设$y = X\beta + e$, $E(e) = 0$, $cov(e) = \sigma^2 I_n$, X 为$n \times p$ 的设计阵，其秩为p, 将X 和β 分块成：$X\beta = (X_1, X_2)\binom{\beta_1}{\beta_2}$.

(a) 证明β_2 的LS 估计$\hat{\beta}_2$ 由下式给出

$$\hat{\beta}_2 = [X_2'X_2 - X_2'X_1(X_1'X_1)^{-1}X_1'X_2]^{-1}[X_2'y - X_2'X_1(X_1'X_1)^{-1}X_1'y].$$

(b) 求$cov(\hat{\beta}_2)$.

20. 设$y = \beta + e$, $E(e) = 0$, $cov(e) = \sigma^2 I$, 直接用Lagrange 乘子法证明：在约束条件$A\beta = 0$ 下，使$\| y - \beta \|^2$ 达到极小的β 值为$\hat{\beta} = (I_n - A'(AA')^{-1}A)y$, 其中$A$ 是已知的$q \times n$ 阵，其秩为q.

21. 设$y_i \sim N(i\theta, i^2\sigma^2)$, $i = 1, 2, \cdots, n$, 且相互独立，用GLS 方法求θ 的最小方差线性无偏估计$\hat{\theta}$, 并求 $Var(\hat{\theta})$.

22. 对线性回归模型$y = X\beta + e,\ E(e) = 0,\ cov(e) = \sigma^2 V,\ V > 0,\ R(X) = p.$

   (a) 证明此时LS估计$\hat{\beta} = (X'X)^{-1}X'y$ 仍然是β 的一个无偏估计。

   (b) 证明$Cov(\hat{\beta}) = \sigma^2(X'X)^{-1}X'VX(X'X)^{-1}.$

   (c) 记$\hat{\sigma}^2 = y'(I - X(X'X)^{-1}X')y/(n-p)$, 证明

$$E(\hat{\sigma}^2) = \frac{\sigma^2}{n-p}tr[V(I_n - X(X'X)^{-1}X')].$$

23. 设$y_1 = \beta_1 + e_1,\ y_2 = 2\beta_1 - \beta_2 + e_2,\ y_3 = \beta_1 + 2\beta_2 + e_3$, 这里$e' = (e_1, e_2, e_3) \sim N_3(0, \sigma^2 I)$, 导出$H_0 : \beta_1 = 2\beta_2$ 的检验统计量。

24. 设$y_{1i} \sim N(\beta_{10} + \beta_{11}x_{1i}, \sigma^2),\ i = 1, 2, \cdots, n,\ y_{2i} \sim N(\beta_{20} + \beta_{21}x_{2i}, \sigma^2),\ i = 1, 2, \cdots, m$, 且相互独立，导出检验$H_0 : \beta_{11} = \beta_{21}$ 的检验统计量。

25. 设$y_i = \beta_0 + \beta_1 x_i + e_i,\ i = 1, 2, \cdots, n.\ e_i$ 相互独立$\sim N(0, \sigma^2)$. 对于假设$H_0 : \beta_0 = 0$ (也就是回归直线经过原点) 导出检验统计量。

26. 对线性模型$y = X\beta + e,\ e \sim N_n(0, \sigma^2 I)$, 回归系数的显著性检验$H_i : \beta_i = 0,\ i = 1, 2, \cdots, p-1$ (将检验问题表成一般线性假设$H_0 : A\beta = 0$的形式)，导出检验统计量。

27. 对习题15 中的问题：

   (a) 求回归方程及回归系数的显著性检验，并列出相应的方差分析表，取显著性水平$\alpha = 0.05$.

   (b) 对$x_0 = 27.5$, 求y_0 的包含概率为0.95 的预测区间。

28. 从空中对地面上一四边形的四个角$\theta_1,\ \theta_2,\ \theta_3,\ \theta_4$ 进行测量，得测量值分别为$Y_1,\ Y_2,\ Y_3,\ Y_4$.

   (a) 将此问题表示成线性模型的形式。

   (b) 对于假设：四边形是平行四边形导出检验统计量。

29. 设回归直线通过原点，即一元线性回归模型为

$$y_i = \beta x_i + e_i,\ i = 1, 2, \cdots, n.$$

   其中$E(e_i) = 0,\ var(e_i) = \sigma^2,\ e_i$ 互不相关。

   (a) 写出β 和σ^2 的最小二乘(LS) 估计。

   (b) 记因变量Y 在x_0 处的值y_0 的预测值为$\hat{y}_0 = \hat{\beta}x_0$, 求$Var(\hat{y}_0 - y_0)$.

   (c) 记$\mu_0 = \beta x_0$，$\hat{\mu}_0 = \hat{\beta}x_0$。求$Var(\hat{\mu}_0)$。

30. 对习题15，求回归系数β_1 的置信系数为0.95 的置信区间。

参考文献

[1] 王松桂，陈敏等. 线性统计模型——线性回归与方差分析，北京：高等教育出版社，1999.

[2] 陈希孺，王松桂. 近代回归分析——原理、方法及应用，合肥：安徽教育出版社，1987.

[3] Weisberg, S. Applied Linear Regression, New York: John Wiley and sons, 1985.