

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。这主要有以下几个原因：

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：

$$Y = f(X) + e, \quad (7.1.1)$$

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

$$Y = \beta_0 + \beta_1 X + e. \quad (7.1.2)$$

(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 $(x_i, y_i), i = 1, 2, \dots, n$ ，将上述模型用样本表示为

$$y_i = \beta_0 + \beta_1 x_i + e_i, \quad i = 1, 2, \dots, n. \quad (7.1.3)$$

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 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ ，代入到(7.1.2)中将误差项 e_i 用其均值0代替，得到

$$Y = \hat{\beta}_0 + \hat{\beta}_1 X, \quad (7.1.4)$$

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) , $i = 1, 2, \dots, n$, 估计 β_0 和 β_1 得 $\hat{\beta}_0 = -40, \hat{\beta}_1 = 0.6$, 则经验回归方程为

$$Y = -40 + 0.6X. \quad (7.1.5)$$

如果甲身高160 cm, 算得体重 $y_0 = 56$ 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 $X_1, X_2, \dots, X_{p-1}$, 则它们有如下线性关系:

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_{p-1} X_{p-1} + e, \quad (7.1.6)$$

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

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_{p-1} x_{ip-1} + e_i, \quad (7.1.7)$$

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:

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

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

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

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

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

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

$$Y = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \dots + \hat{\beta}_{p-1} X_{p-1} \quad (7.1.9)$$

称为经验回归方程, 它是否真正描述了 Y 和 $X_1, X_2, \dots, X_{p-1}$ 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:

$$Q_t = aL_t^b K_t^c, \quad (7.1.10)$$

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

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

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

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

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

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

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) $\bar{x} = 68$ 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 $\bar{x} = 68$ 英寸的儿子的平均身高要高于父亲的平均身高。Galton 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

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

回归系数 β_i 的估计量 $\hat{\beta}_i$ ($i = 0, 1, \dots, p-1$) 的大小在一定程度上反映了 X_i 对 Y 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 $\mathbf{X} = (X_1, X_2, \dots, X_n)'$ 为 $n \times 1$ 随机向量, 则称

$$E(\mathbf{X}) = (E(X_1), E(X_2), \dots, E(X_n))'$$

为随机向量 $\mathbf{X}$ 的均值向量, 称 $n \times n$ 阶对称阵

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

为随机向量 $\mathbf{X}$ 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 $\mathbf{X}$ 和 $\mathbf{b}$ 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 $\mathbf{Y} = A\mathbf{X} + \mathbf{b}$, 则

$$E(\mathbf{Y}) = E(A\mathbf{X} + \mathbf{b}) = AE(\mathbf{X}) + E(\mathbf{b}).$$

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

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

推论7.2.1 $tr[Cov(\mathbf{X})] = \sum_{i=1}^n Var(X_i)$, 此处 trA 表示方阵 A 的迹.

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

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

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

故知 $Cov(\mathbf{X}) \geq 0$.

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

证明

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

定理7.2.4 设 $\mathbf{X}$ 和 $\mathbf{Y}$ 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 $Cov(A\mathbf{X}, B\mathbf{Y}) = ACov(\mathbf{X}, \mathbf{Y})B'$.

证明 从定义出发。

二、随机向量的二次型

定义7.2.2 设 $\mathbf{X} = (X_1, X_2, \dots, X_n)'$, $A = (a_{ij})$ 为 $n \times n$ 对称阵, 则

$$\mathbf{X}'A\mathbf{X} = \sum_{i=1}^n \sum_{j=1}^n a_{ij} X_i X_j$$

称为随机向量 $\mathbf{X}$ 的二次型。

如何求二次型的均值、方差, 我们有下列定理:

定理7.2.5 设 $\mathbf{X}$ 为 $m \times 1$ 维的随机向量, 且有 $E(\mathbf{X}) = \mu_{n \times 1}$, $Cov(\mathbf{X}) = \Sigma_{n \times n}$, 则

$$E(\mathbf{X}'A\mathbf{X}) = \mu' A \mu + tr(A\Sigma).$$

证明

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

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

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

特别:

- (1) 当 $\mu = 0$ 时, $E(\mathbf{X}'A\mathbf{X}) = trA\Sigma$;
- (2) 当 $\Sigma = \sigma^2 I$ 时, $E(\mathbf{X}'A\mathbf{X}) = \mu'A\mu + \sigma^2 trA$;
- (3) 当 $\mu = 0, \Sigma = I$ 时, $E(\mathbf{X}'A\mathbf{X}) = trA$.

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

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

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

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

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

由定理7.2.1可得

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

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 $\mathbf{X} = (X_1, X_2, \dots, X_n)'$ 具有密度函数

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

其中 $\mathbf{x} = (x_1, \dots, x_n)'$, $-\infty < x_i < \infty$, $\mu = (\mu_1, \dots, \mu_n)'$, $\Sigma_{n \times n} > 0$ 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 $\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \dots, x_n) dx_1 \cdots dx_n = 1$;
- (2) $E(\mathbf{X}) = \mu$, $Cov(\mathbf{X}) = \Sigma$.

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

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

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 $Y_i \sim N(0, 1)$, $i = 1, \dots, n$. 因而

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(\mathbf{x}) d\mathbf{x} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(\mathbf{y}) d\mathbf{y} = \prod_{i=1}^n \left[\int_{-\infty}^{\infty} f(y_i) dy_i \right] = 1.$$

(2) 由变换 $\mathbf{Y} = \Sigma^{-\frac{1}{2}}(\mathbf{X} - \mu)$ 可知

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

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

设 $\mathbf{X} \sim N_n(\mu, \Sigma)$, 将向量 $\mathbf{X}_{n \times 1}$ 和 $\mu_{n \times 1}$ 做相应的分块

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

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 有如下分块的形式

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

这里 Σ_{11} 为 $p \times p$ 的子方阵。

定理7.2.6 (1) 设 $\mathbf{X} \sim N_n(\mu, \Sigma)$, 且 $\mathbf{X}, \mu$ 和 Σ 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 $\Sigma = \sigma^2 I, \mathbf{X} = (X_1, \dots, X_n)', \mu = (\mu_1, \dots, \mu_n)'$, 则 $X_i \sim N(\mu_i, \sigma^2)$, 且相互独立, $i = 1, 2, \dots, n$.

证明 (1) 由于 $\mathbf{X} \sim N_n(\mu, \Sigma)$, 且其协方差阵 Σ 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

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

其中 $f(x_{(1)})$ 和 $f(x_{(2)})$ 分别为 $X_{(1)} \sim N(\mu_{(1)}, \Sigma_{11})$ 和 $X_{(2)} \sim N(\mu_{(2)}, \Sigma_{22})$ 的密度函数。

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

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

故 $X_j \sim N(\mu_j, \sigma^2)$, 且相互独立, $j = 1, \dots, n$.

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 $\mathbf{X} \sim N_n(\mu, \Sigma)$, A 为 $n \times n$ 可逆常数阵, b 为 $n \times 1$ 常向量, 记 $\mathbf{Y} = A\mathbf{X} + b$, 则

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

证明 用特征函数证明。

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

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 $\mathbf{X} \sim N_n(\mu, \Sigma)$, 其中 $\mathbf{X}$, 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

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

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

$$\begin{aligned} \mathbf{Y} = A\mathbf{X} &= \begin{pmatrix} X_{(1)} \\ X_{(2)} - \Sigma_{21}\Sigma_{11}^{-1}X_{(1)} \end{pmatrix}, \quad A\mu = \begin{pmatrix} \mu_{(1)} \\ \mu_{(2)} - \Sigma_{21}\Sigma_{11}^{-1}\mu_{(1)} \end{pmatrix}, \\ A\Sigma A' &= \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22.1} \end{pmatrix}, \end{aligned}$$

其中 $\Sigma_{22.1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$, 从而

$$\mathbf{Y} = \begin{pmatrix} Y_{(1)} \\ Y_{(2)} \end{pmatrix} = A\mathbf{X} \sim N_n\left(\begin{pmatrix} \mu_{(1)} \\ * \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22.1} \end{pmatrix}\right),$$

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 $\mathbf{X} \sim N_n(\mu, \Sigma)$, 其中 $A_{m \times n}$ 为常数阵, $R(A) = m < n$, 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, 由定理7.2.7可知 $\mathbf{Z} = C\mathbf{X} \sim N(C\mu, C\Sigma C')$, 从而

$$\mathbf{Z} = C\mathbf{X} = \begin{pmatrix} A \\ B \end{pmatrix} \mathbf{X} = \begin{pmatrix} A\mathbf{X} \\ B\mathbf{X} \end{pmatrix} = \begin{pmatrix} Z_{(1)} \\ Z_{(2)} \end{pmatrix} \sim N_n\left(\begin{pmatrix} A\mu \\ B\mu \end{pmatrix}, \begin{pmatrix} A\Sigma A' & A\Sigma B' \\ B\Sigma A' & B\Sigma B' \end{pmatrix}\right),$$

由定理7.2.8可知: $\mathbf{Y} = Z_{(1)} = A\mathbf{X} \sim N_m(A\mu, A\Sigma A')$.

推论7.2.3 在定理7.2.9中, 若取 A 为一个行向量即 $A = c$, 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变量。

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, $A\mathbf{X}$ 和 $B\mathbf{X}$ 独立。

证明 $Cov(A\mathbf{X}, B\mathbf{X}) = ACov(\mathbf{X})B' = A\Sigma B' = 0$, 故 $A\mathbf{X}$ 和 $B\mathbf{X}$ 不相关, 由于它们是正态变量, 故不相关与独立等价, 因此 $A\mathbf{X}$ 和 $B\mathbf{X}$ 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 $\mathbf{X} \sim N_n(0, \Sigma)$, $\Sigma > 0$ (正定), 则 $\mathbf{X}'\Sigma^{-1}\mathbf{X} \sim \chi_n^2$. 当 $\mathbf{X} \sim N(\mu, \Sigma)$, 则 $(\mathbf{X} - \mu)'\Sigma^{-1}(\mathbf{X} - \mu) \sim \chi_n^2$.

(2) 设 $\mathbf{X} \sim N_n(0, I)$, 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 $\mathbf{X}'A\mathbf{X} \sim \chi_r^2$.

证明 (1) 记 $\mathbf{Y} = \Sigma^{-\frac{1}{2}}\mathbf{X}$, 则 $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)' \sim N_n(0, I) \iff Y_i \sim N(0, 1)$, 且相互独立, $i = 1, 2, \dots, n$. 从而 $\mathbf{X}'\Sigma^{-1}\mathbf{X} = \mathbf{Y}'\mathbf{Y} = \sum_{i=1}^n Y_i^2 \sim \chi_n^2$.

当 $\mathbf{X} \sim N_n(\mu, \Sigma)$ 时, 令 $\mathbf{Y} = \Sigma^{-\frac{1}{2}}(\mathbf{X} - \mu)$, 类似方法可证明 $(\mathbf{X} - \mu)'\Sigma^{-1}(\mathbf{X} - \mu) \sim \chi_n^2$.

(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

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

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

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

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

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

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

记 $\mathbf{Y} = Q'\mathbf{X}$, 则 $\mathbf{Y} \sim N_n(0, I)$, 因此有

$$\mathbf{X}'A\mathbf{X} = \mathbf{X}'Q\Lambda Q'\mathbf{X} = \mathbf{Y}'\Lambda\mathbf{Y} = \sum_{i=1}^r \lambda_i Y_i^2 = Y_{(1)}' \Lambda_r Y_{(1)}, \quad Y_i \sim N(0, 1), i = 1, 2, \dots, r,$$

此处 $\Lambda_r = \text{diag}\{\lambda_1, \dots, \lambda_r\}$, $Y_{(1)} = (Y_1, \dots, Y_r)$. 问题转化为当 $BA = 0$ 时, $B\mathbf{X} = BQQ'\mathbf{X} = D\mathbf{Y}$, 问 $D\mathbf{Y}$ 与 $\mathbf{Y}'\Lambda_r\mathbf{Y}$ 是否独立?

由于 $\mathbf{X} \sim N_n(0, I) \implies Q'\mathbf{X} = \mathbf{Y} \sim N(0, I_n)$, 即 $Y_1, \dots, Y_n \text{ iid } \sim N(0, 1)$, 则

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

将 D 分块为 $D = (D_1 | D_2)$, 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 $\implies D_1 = 0$. 因此 $D = (0 | D_2)$, 从而

$$B\mathbf{X} = D\mathbf{Y} = \begin{pmatrix} 0 & D_2 \end{pmatrix} \begin{pmatrix} Y_{(1)} \\ Y_{(2)} \end{pmatrix} = D_2 Y_{(2)}, \quad \text{其中 } Y_{(1)} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_r \end{pmatrix}, Y_{(2)} = \begin{pmatrix} Y_{r+1} \\ \vdots \\ Y_n \end{pmatrix},$$

由于 $Y_{(1)} = (Y_1, \dots, Y_r)$ 与 $Y_{(2)} = (Y_{r+1}, \dots, Y_n)$ 独立, $B\mathbf{X} = D\mathbf{Y} = D_2 Y_{(2)}$ 只与 $Y_{(2)}$ 有关, $\mathbf{X}'A\mathbf{X} = Y_{(1)}' \Lambda_r Y_{(1)}$ 只与 $Y_{(1)}$ 有关, 故二者独立。

推论7.2.5 设 $\mathbf{X} \sim N_n(0, \Sigma)$, 则当 $B\Sigma A = 0$ 时, $B\mathbf{X}$ 与 $\mathbf{X}'A\mathbf{X}$ 独立。

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

$$B\mathbf{X} = B\Sigma^{\frac{1}{2}}\mathbf{Y} = \tilde{B}\mathbf{Y}, \quad \mathbf{X}'A\mathbf{X} = \mathbf{Y}'\Sigma^{\frac{1}{2}}A\Sigma^{\frac{1}{2}}\mathbf{Y} = \mathbf{Y}'\tilde{A}\mathbf{Y},$$

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

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

即这一条件成立时, $B\mathbf{X}$ 与 $\mathbf{X}'A\mathbf{X}$ 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 $\mathbf{X} \sim N_n(0, I)$, A 和 B 皆为 $n \times n$ 对称阵, 且 $AB = 0$, 则二次型 $\mathbf{X}'A\mathbf{X}$ 和 $\mathbf{X}'B\mathbf{X}$ 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

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

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

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

$$\mathbf{X}'A\mathbf{X} = \mathbf{X}'Q\Lambda Q'\mathbf{X} = \mathbf{Y}'\Lambda\mathbf{Y};$$

$$\mathbf{X}'B\mathbf{X} = \mathbf{X}'Q\Delta Q'\mathbf{X} = \mathbf{Y}'\Delta\mathbf{Y},$$

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

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

由 $Y_1, \dots, Y_r$ 与 $Y_{r+1}, \dots, Y_n$ 独立, 因此有 $\mathbf{X}'A\mathbf{X}$ 与 $\mathbf{X}'B\mathbf{X}$ 独立。

推论7.2.6 设 $\mathbf{X} \sim N_n(0, \Sigma)$, 若 $A\Sigma B = 0$, 则 $\mathbf{X}'A\mathbf{X}$ 与 $\mathbf{X}'B\mathbf{X}$ 独立。

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

$$\mathbf{X}'A\mathbf{X} = \mathbf{Y}'\Sigma^{\frac{1}{2}}A\Sigma^{\frac{1}{2}}\mathbf{Y} = \mathbf{Y}'\tilde{A}\mathbf{Y}, \quad \mathbf{X}'B\mathbf{X} = \mathbf{Y}'\Sigma^{\frac{1}{2}}B\Sigma^{\frac{1}{2}}\mathbf{Y} = \mathbf{Y}'\tilde{B}\mathbf{Y},$$

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

$$\tilde{A}\tilde{B} = 0 \iff A\Sigma B = 0.$$

§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, $X_1, \dots, X_{p-1}$, 它们之间有线性关系

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_{p-1} X_{p-1} + e, \quad (7.3.1)$$

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

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

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_{p-1} x_{i,p-1} + e_i, \quad i = 1, 2, \dots, n. \quad (7.3.2)$$

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

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

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:

$$y_{n \times 1} = X_{n \times p} \beta_{p \times 1} + e_{n \times 1}, \quad (7.3.4)$$

此处 e 满足 G-M 假定:

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

此处 y 为 $n \times 1$ 观察向量, X 为 $n \times p$ 设计阵, β 为 $p \times 1$ 回归参数向量, e 为 $n \times 1$ 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

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

即随机误差的平方和达到最小值时 β 的取值 $\hat{\beta}$ 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

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

(7.3.6) 称为正则方程, 当 $X'X$ 可逆时有唯一解

$$\hat{\beta} = (X'X)^{-1}X'y. \quad (7.3.7)$$

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

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

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

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

可知 $\hat{\beta}$ 使 $Q(\beta)$ 达到极小, 故称 $\hat{\beta}$ 为 β 的 LS 估计。

2. 经验回归方程

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

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1X_1 + \cdots + \hat{\beta}_{p-1}X_{p-1}, \quad (7.3.8)$$

称为经验回归方程, 它描述了 Y 与自变量 $X_1, \dots, X_{p-1}$ 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 $(x_i, y_i), i = 1, \dots, n$, 则有

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

正则方程为

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

而

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

因此

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

代入到正则方程组可得

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

因此 β 和 α 的LS估计为

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

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

从而 $\hat{Y} = \hat{\alpha} + \hat{\beta}X$ 为所要得的经验回归方程。

3. σ^2 的LS估计

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

$$\hat{e} = y - X\hat{\beta} \iff \hat{e}_i = y_i - x'_i \hat{\beta},$$

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

$$RSS = \|\hat{e}\|^2 = \hat{e}'\hat{e} = \sum_{i=1}^n \hat{e}_i^2 \quad (7.3.9)$$

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

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

为 σ^2 的无偏估计, 它称为 σ^2 的LS估计。

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

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

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

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

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

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

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

用矩阵表示为:

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

此即

$$y_{n \times 1} = \alpha \mathbf{1}_{n \times 1} + X_c \beta_c + e, \quad (7.3.12)$$

此处 α 为常数项, $\beta_c = (\beta_1, \dots, \beta_{p-1})'$ 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

$$\mathbf{1}'_n X_c = 0. \quad (7.3.13)$$

这是因为 $\mathbf{1}' \begin{pmatrix} x_{1j} - \bar{x}_j \\ \vdots \\ x_{nj} - \bar{x}_j \end{pmatrix} = \sum_{i=1}^n (x_{ij} - \bar{x}_j) = \sum_{i=1}^n x_{ij} - n\bar{x}_j = 0, \quad j = 1, \dots, p-1.$

将(7.3.12)式改写为

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

因此正则方程为

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

故有

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\beta}_c \end{pmatrix} = \begin{pmatrix} \frac{1}{n} & 0 \\ 0 & (X_c'X_c)^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{1}'y \\ X_c'y \end{pmatrix}$$

于是回归参数的LS估计为

$$\begin{cases} \hat{\alpha} = \frac{1}{n} \sum_{i=1}^n y_i = \bar{y} \\ \hat{\beta}_c = (X_c'X_c)^{-1} X_c'y. \end{cases} \quad (7.3.14)$$

结论： 中心化模型LS估计为： $\hat{\alpha} = \bar{y}$ ，而 $\hat{\beta}_c$ 可视为由 $y = X_c\beta_c + e$ 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

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

按中心化方法

$$\begin{aligned} X_c' &= (x_1 - \bar{x}, \dots, x_n - \bar{x}) \\ (X_c'X_c)^{-1} &= 1 / \sum_{i=1}^n (x_i - \bar{x})^2. \end{aligned}$$

按公式(7.3.14)得到 α 和 β 的LS估计为

$$\begin{aligned} \hat{\alpha} &= \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i, \\ \hat{\beta} &= (X_c'X_c)^{-1} X_c'y = \frac{\sum_{i=1}^n (x_i - \bar{x})y_i}{\sum_{i=1}^n (x_i - \bar{x})^2} \end{aligned}$$

从而得到经验回归方程

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

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 X_c 的第 j 列离散程度的量为

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

而令标准化变换为：

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

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

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

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

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

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

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

(b) $Z'Z = R = (r_{ij})$ 为 $(p-1) \times (p-1)$ 方阵, 其中 $r_{ij} = \frac{1}{s_i s_j} \sum_{k=1}^n (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)$ 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

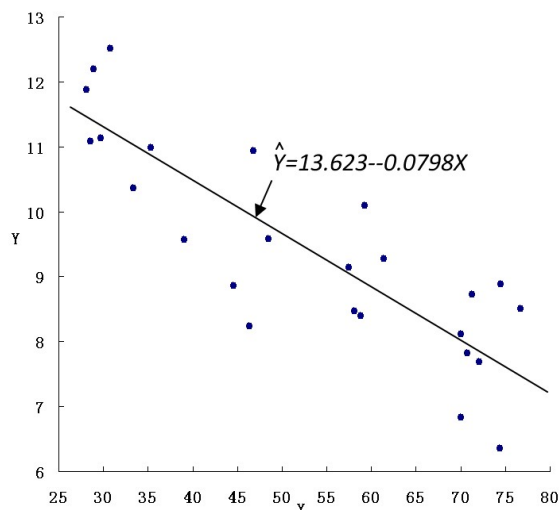
$$\hat{\alpha} = \bar{y}, \quad \hat{\beta}^o = (\hat{\beta}_1^o, \cdots, \hat{\beta}_{p-1}^o)' = (Z'Z)^{-1} Z'y$$

之后, 则经验回归方程为

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

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($^{\circ}\text{C}$), 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($^{\circ}\text{C}$)	序号	Y(L)	X($^{\circ}\text{C}$)
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

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_c 的最小二乘估计分别为：

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

$$\hat{\beta}_c = -0.0798.$$

于是经验回归方程为

$$\hat{Y} = 9.424 - 0.0798(X - 52.60),$$

即

$$\hat{Y} = 13.623 - 0.0798X.$$

□

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 $\hat{\beta}$ 为 β 的 LS 估计, 则有

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

证明 由 $\hat{\beta} = (X'X)^{-1}X'y$ 可知

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

$$\begin{aligned} Cov(\hat{\beta}) &= Cov((X'X)^{-1}X'y) = (X'X)^{-1}X'Cov(Y)X(X'X)^{-1} \\ &= \sigma^2(X'X)^{-1}X'X(X'X)^{-1} = \sigma^2(X'X)^{-1}. \end{aligned}$$

□

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 $c'\hat{\beta}$ 为 $c'\beta$ 的 LS 估计。则由定理7.3.1可知

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

$c'\beta$ 称为可估函数。

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

证明： 设 $a'y$ 为 $c'\beta$ 的任一线性无偏估计, 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

$$a'X = c'. \quad (7.3.17)$$

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

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

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

则

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

这等价于证明下式成立:

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

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

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

一切 $\beta \in \mathbb{R}_p$ 成立, 且等号成立 $\iff (I - X(X'X)^{-1}X')a = 0$, 即 $a = X(X'X)^{-1}c$, 也就是 $a'y = c'\hat{\beta}$, 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) $\hat{\beta} \sim N(\beta, \sigma^2(X'X)^{-1})$;
- (b) $RSS/\sigma^2 \sim \chi_{n-p}^2$;
- (c) $\hat{\beta}$ 与 $\hat{\sigma}^2$ 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

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

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

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

利用二次型和线性型独立性判定定理可知, 二者相互独立 $\iff AB' = 0$, 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。

□

4. 中心化回归模型的 β_c 和 σ^2 的 LS 估计的优良性

对中心化的回归模型：

$$y = \alpha \mathbf{1}_n + X_c \beta_c + e,$$

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

- (a) $E(\hat{\alpha}) = \alpha$, $E(\hat{\beta}_c) = \beta$,
- (b) $Cov \begin{pmatrix} \hat{\alpha} \\ \hat{\beta}_c \end{pmatrix} = \sigma^2 \begin{pmatrix} \frac{1}{n} & 0 \\ 0 & (X_c' X_c)^{-1} \end{pmatrix}$
- (c) 进一步假定 $e \sim N(0, \sigma^2 I)$, 则有

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

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

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

5. 复相关系数

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

$$R^2 = \frac{SS_{\text{回}}}{SS_{\text{总}}},$$

称为判定系数, 它判定回归自变量 $X_1, \dots, X_{p-1}$ 与因变量 Y 的拟合程度。显然 $0 \leq R^2 \leq 1$. R 称为复相关系数, 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 的经验回归方程。

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

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

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 $X_c' X_c \beta_c = X_c' y$ 为

$$\begin{cases} 2078.92 \beta_1 - 1533.85 \beta_2 = 1607.00 \\ -1533.85 \beta_1 + 2307.69 \beta_2 = -715.00 \end{cases}$$

解得

$$\hat{\beta}_1 = 1.068, \quad \hat{\beta}_2 = 0.400.$$

又

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

故得经验回归方程为

$$\begin{aligned} \hat{Y} &= \hat{\alpha} + (X_1 - \bar{x}_1) \hat{\beta}_1 + (X_2 - \bar{x}_2) \hat{\beta}_2 \\ &= 130 + 1.068(X_1 - 166.8) + 0.400(X_2 - 38.85) \\ &= -62.963 + 1.068X_1 + 0.400X_2. \end{aligned}$$

我们还可计算出判定系数

$$SS_{\text{回}} = \hat{\beta}' X_c' y = 1430.276,$$

$$SS_{\text{总}} = \sum_{i=1}^{13} (y_i - \bar{y})^2 = 1512,$$

故有

$$R^2 = \frac{SS_{\text{回}}}{SS_{\text{总}}} = \frac{1430.276}{1512} = 0.9459, \quad R = 0.9726.$$

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 $y_{n \times 1} = X_{n \times p} \beta_{p \times 1} + e_{n \times 1}$, $e \sim (0, \sigma^2 I)$, 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, 我们常要求带约束条件的 LS 估计问题, 即有约束条件

$$A\beta = b \quad (7.4.1)$$

其中 A 为 $k \times p$ 阶矩阵, b 为 $k \times 1$ 维向量 ($k \leq p$), 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

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

此即 $H: A\beta = 0$, 此处 $R(A) = k = p - 2$,

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

因此有约束的线性回归模型为

$$\begin{cases} y_{n \times 1} = X_{n \times p} \beta_{p \times 1} + e_{n \times 1}, & e \sim (0, \sigma^2 I) \\ A\beta = b, & \text{其中 } R(A) = k, \text{ 即行满秩.} \end{cases} \quad (7.4.2)$$

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

$$\hat{\beta}_r = \hat{\beta} - (X'X)^{-1}A'(A(X'X)^{-1}A')^{-1}(A\hat{\beta} - b),$$

其中 $\hat{\beta} = (X'X)^{-1}X'y$ 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

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

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 $\lambda' = (\lambda_1, \dots, \lambda_k)$ ，则

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

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

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

此即

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

由 ① 可以得到

$$\hat{\beta}_r = (X'X)^{-1}X'y - (X'X)^{-1}A'\hat{\lambda}_r = \hat{\beta} - (X'X)^{-1}A'\hat{\lambda}_r, \quad (7.4.4)$$

将(7.4.4)式带入 ② 得

$$\begin{aligned} A\hat{\beta}_r &= A\hat{\beta} - A(X'X)^{-1}A'\hat{\lambda}_r = b \\ \iff A(X'X)^{-1}A'\hat{\lambda}_r &= A\hat{\beta} - b, \end{aligned} \quad (7.4.5)$$

由 $R(A) = k$ 可知 $A(X'X)^{-1}A'$ 之逆存在，故有

$$\hat{\lambda}_r = (A(X'X)^{-1}A')^{-1}(A\hat{\beta} - b). \quad (7.4.6)$$

将(7.4.6)式代回(7.4.4)式可得

$$\hat{\beta}_r = \hat{\beta} - (X'X)^{-1}A'(A(X'X)^{-1}A')^{-1}(A\hat{\beta} - b). \quad (7.4.7)$$

现在我们证明 $\hat{\beta}_r$ 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) $\hat{\beta}_r$ 适合约束方程，即 $A\hat{\beta}_r = b$ ，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

$$\|y - X\beta\|^2 \geq \|y - X\hat{\beta}_r\|^2.$$

为此，我们将平方和 $\|y - X\beta\|^2$ 作分解。

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

其中 $I_3 = (\hat{\beta}_r - \beta)' (X'y - X'X\hat{\beta}_r)$ 。因为

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

从而

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

故有 $\|y - X\beta\|^2 = I_1 + I_2 \geq \|y - X\hat{\beta}_r\|^2$ ，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 $\theta_1, \theta_2, \theta_3$ 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 $\theta_1, \theta_2, \theta_3$ 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：

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

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

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

即 $\hat{\theta}_i = y_i - \frac{1}{3}(y_1 + y_2 + y_3 - \pi)$, $i = 1, 2, 3$ 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

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

为求参数的估计量，我们希望经过适当的变换，将这一模型转化成 §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

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

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

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

称 $\hat{\beta}^*$ 和 $\hat{\sigma}_*^2$ 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) $E(\hat{\beta}^*) = \beta$;
- (b) $Cov(\hat{\beta}^*) = \sigma^2(X'\Sigma^{-1}X)^{-1}$;
- (c) 对 $\forall c_{p \times 1}$, $c'\hat{\beta}^*$ 为 $c'\beta$ 的BLU 估计;
- (d) $\hat{\sigma}_*^2$ 为 σ^2 的无偏估计。

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

(b) 直接计算可得

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

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

(d) 显然。

注: 此时 $c'\hat{\beta}^*$ 为 $c'\beta$ 的BLU 估计, 而 $c'\hat{\beta} = c'(X'X)^{-1}X'y$ 仍是 $c'\beta$ 的无偏估计, 因为 $E(c'\hat{\beta}) = c'(X'X)^{-1}X' \cdot X\beta = c'\beta$, 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 $Var(c'\hat{\beta}^*) \leq Var(c'\hat{\beta})$.

例7.5.1 广义LS估计最简单的情形是取 $\Sigma = diag(\sigma_1^2, \dots, \sigma_n^2)$, 此处 $\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2$ 不全相等, 求 β 的GLS估计。

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

记

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

则有

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

从而

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

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, 所用的“权”都是 $1/\sigma_i^2$. 故 $\hat{\beta}^*$ 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:

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

其中 $R(X) = p$, 本节讨论一般线性假设:

$$H: A\beta = b, \quad (7.6.2)$$

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

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

1. 检验常数项 $H: \beta_0 = 0$, 则检验问题可表示为:

$$H: A\beta = 0, \quad \text{其中 } A = (1, 0, 0, \dots, 0), \quad b = 0;$$

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

$$H: A\beta = 0, \quad \text{其中 } A = (\mathbf{0}_{k \times 1} : I_k : \mathbf{0}_{k \times (p-1-k)}), \quad b = 0;$$

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

$$H: A\beta = 0, \quad A = (\mathbf{0}_{(p-1) \times 1} : I_{p-1}), \quad b = 0.$$

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

$$RSS = (y - X\hat{\beta})'(y - X\hat{\beta}) = y'(I - X(X'X)^{-1}X')y, \quad (7.6.3)$$

其中 $\hat{\beta} = (X'X)^{-1}X'y$. 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

$$\hat{\beta}_H = \hat{\beta} - (X'X)^{-1}A'(A(X'X)^{-1}A')^{-1}(A\hat{\beta} - b), \quad (7.6.4)$$

相应的残差平方和为:

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

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, 因而残差平方和就要变大, 于是总有 $RSS_H \geq RSS$. 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

$H: A\beta = b$ 不成立, 则这些变量的作用归入残差平方和, 故 $RSS_H - RSS$ 应当较大, 因此拒绝 H . 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:

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

二、 F 检验

定理7.6.1 对于正态线性回归模型 (7.6.1) 有

- (a) $\frac{RSS}{\sigma^2} \sim \chi_{n-p}^2$;
- (b) 若 (7.6.2) 式成立, 则 $\frac{RSS_H - RSS}{\sigma^2} \sim \chi_m^2$;
- (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) $\frac{RSS}{\sigma^2} = \left(\frac{y}{\sigma}\right)'(I - X(X'X)^{-1}X')\left(\frac{y}{\sigma}\right) \sim \chi_{n-p}^2$ 在定理7.3.3中证过。

(b) 对 RSS_H 进行分解

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

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

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

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

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

因此令 $Z = A\hat{\beta} - b$, 则在 H 成立时

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

因为 $R(A(X'X)^{-1}A') = m$, 故由定理7.2.11可知

$$Z'\Sigma^{-1}Z = \frac{(A\hat{\beta} - b)'(A(X'X)^{-1}A')^{-1}(A\hat{\beta} - b)}{\sigma^2} = \frac{RSS_H - RSS}{\sigma^2} \sim \chi_m^2.$$

(c) 因为

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

将

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

代入上式得

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

此处

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

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 $e \sim N(0, \sigma^2 I)$, 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:

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

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 , 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

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

三、 RSS_H 计算的方法

考虑假设检验:

$$H: A\beta = b$$

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:

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

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

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

将 $H: A\beta = 0$ 代入到 (7.6.8) 式中, 得

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

用矩阵表示

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

则 α 的 LS 估计为

$$\hat{\alpha} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'y,$$

其残差平方和为

$$RSS_H = y'y - \hat{\alpha}'\tilde{X}'y. \quad (7.6.9)$$

又因为

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

因此

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

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 $H: \beta_1 = \beta_2$.

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 等价于

$$H: \begin{pmatrix} 1 & -1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = 0.$$

于是, 对现在的情形 $A = \begin{pmatrix} 1 & -1 \end{pmatrix}$, $n = 3$, $p = 2$, $m = 1$. 易求原模型中 β 的LS估计为

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

所以

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

为求 $H: \beta_1 = \beta_2$ 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

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

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

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

于是残差平方和为

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

从而 H 的检验统计量为

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

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

§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:

$$y_i = \beta_0 + x_{i1}\beta_1 + \cdots + x_{ip-1}\beta_{p-1} + e_i, \quad i = 1, 2, \dots, n. \quad (7.7.1)$$

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

$$H: \beta_1 = \beta_2 = \cdots = \beta_{p-1} = 0. \quad (7.7.2)$$

当否定 H 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 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 $A_{(p-1) \times p} = (\mathbf{0} \quad I_{p-1})$, $\beta' = (\beta_0, \beta_1, \dots, \beta_{p-1})'$. 显然 $R(A) = p-1$, 则

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

由上节可知检验统计量为

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

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

$$y_i = \beta_0 + e_i, \quad i = 1, 2, \dots, n \iff y_{n \times 1} = \beta_0 \mathbf{1}_n + e. \quad (7.7.3)$$

从而 β_0 的 LS 估计为

$$\hat{\beta}_0^* = (\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'y = \frac{1}{n} \sum_{i=1}^n y_i = \bar{y},$$

对应的残差平方和

$$RSS_H = y'y - (\hat{\beta}_0^*)'\mathbf{1}'y = y'y - n\bar{y}^2 = \sum_{i=1}^n (y_i - \bar{y})^2, \quad (7.7.4)$$

故有

$$RSS_H - RSS = \hat{\beta}'X'y - (\hat{\beta}_0^*)'\mathbf{1}'y = SS_{\text{回}}, \quad (7.7.5)$$

它表示回归平方和的减少量. 于是检验统计量为

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

记 TSS 表示总的平方和, 则有

$$TSS = RSS + SS_{\text{回}}, \quad (7.7.7)$$

其中 $SS_{\text{回}} = \hat{\beta}'X'y - (\hat{\beta}_0^*)'1'y$ 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，包括随机误差、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 ，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	$F_{\text{比}}$	显著性(p 值)
回归	$SS_{\text{回}}$	$p - 1$	$MS_{\text{回}} = SS_{\text{回}}/(p - 1)$	$F_{\text{回}} = MS_{\text{回}}/MS_e$	
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

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

时，否定 H 。

结论： (1) 当拒绝 H 时，说明回归方程有意义， Y 与自变量 $X_1, X_2, \dots, X_{p-1}$ 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 被接受，即认为 $\beta_1 = \beta_2 = \dots = \beta_{p-1} = 0$ ，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

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

为了检验回归方程的显著性，令

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

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	$F_{\text{比}}$	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 < α , 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

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

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

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

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

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

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

$$\hat{\beta}_i \sim N_1(\beta_i, \sigma^2 c_{ii}), \quad i = 1, 2, \dots, p-1.$$

而 σ^2 的LS估计为

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

则将 $\hat{\beta}_i$ 标准化得到:

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

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

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

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

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	$\hat{\beta}_0$	$\hat{\sigma}\sqrt{c_{00}}$	t_0	
β_1	$\hat{\beta}_1$	$\hat{\sigma}\sqrt{c_{11}}$	t_1	
β_2	$\hat{\beta}_2$	$\hat{\sigma}\sqrt{c_{22}}$	t_2	
$\vdots$	$\vdots$	$\vdots$	$\vdots$	
β_{p-1}	$\hat{\beta}_{p-1}$	$\hat{\sigma}\sqrt{c_{p-1,p-1}}$	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 $t_i = \frac{\hat{\beta}_i}{\hat{\sigma}\sqrt{c_{ii}}}$ 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: $\hat{\beta}_i \sim N_1(\beta, \sigma^2 c_{ii})$, 故知:

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

从而有

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

解括号中的不等式得到 β_i 的置信系数 $1 - \alpha$ 的置信区间

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

§7.8 预 测 问 题

一、引言

设所要研究的模型为

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

其中 $e_1, \cdots, e_n \text{ iid } \sim (0, \sigma^2)$, $\beta = (\beta_0, \beta_1, \cdots, \beta_n)'$, $x_i = (1, x_{i1}, \cdots, x_{i,p-1})'$, $i = 1, \cdots, n$. 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某处, 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

$$y_0 = x'_0\beta + e_0, \quad e_0 \sim (0, \sigma^2). \quad (7.8.2)$$

要预测随机变量 y_0 之值, y_0 有两部分:

- (1) $x'_0\beta$ 未知, 用 $\hat{\beta}$ 代替 β , 因而用 $x'_0\hat{\beta}$ 预测 $x'_0\beta$ 之值;
- (2) e_0 未知, 由于 $E(e_0) = 0$, 故用 0 预测 e_0 之值。

所以用

$$\hat{y}_0 = x'_0\hat{\beta} \quad (7.8.3)$$

作为 y_0 的预测值称为点预测。

预测问题与估计问题的不同点如下:

- (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) 从精度上讲预测量的精度要比估计量的精度差。 $y_0 = x'_0\beta + e_0$ 和 $\mu_0 = x'_0\beta$ 二者的预测量和估计量皆用 $x'_0\hat{\beta}$ 表示, 但精度不同。 y_0 的预测量有两部分组成: 用 $x'_0\hat{\beta}$ 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. $\hat{y}_0 = x'_0\hat{\beta}$ 为 y_0 的无偏预测, 即 $E(\hat{y}_0 - y_0) = 0$, 这是因为

$$E(\hat{y}_0) = x'_0\beta, \quad E(y_0) = x'_0\beta \implies E(\hat{y}_0 - y_0) = x'_0\beta - x'_0\beta = 0.$$

2. 用 G-M 定理, 可证 $\hat{y}_0$ 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 $\hat{y}_0 = x'_0\hat{\beta}$ 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

$$\text{Var}(a'y) \geq \text{Var}(x'_0\hat{\beta}).$$

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 $\hat{\mu}_0 = x'_0\hat{\beta}$ 表达式相同, 它们的实际意义是不同的。引进预测偏差 $d_1 = \hat{y}_0 - y_0$ 和估计偏差 $d_2 = \hat{\mu}_0 - \mu_0 = \hat{\mu}_0 - x'_0\beta$ 计算它们的方差

$$\text{Var}(d_1) = \text{Var}(\hat{y}_0) + \text{Var}(y_0) = \sigma^2 x'_0(X'X)^{-1}x_0 + \sigma^2 = \sigma^2(1 + x'_0(X'X)^{-1}x_0); \quad (7.8.4)$$

另一方面

$$\text{Var}(d_2) = \text{Var}(\hat{\mu}_0) = \sigma^2 x'_0(X'X)^{-1}x_0 < \text{Var}(d_1).$$

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 $e_1, e_2, \dots, e_n$ 和 $e_0 = e_{n+1} \text{ iid } \sim N(0, \sigma^2)$, 则由前面的计算可知

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

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

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

将上式中的 σ 用 $\hat{\sigma}$ 代替, 得到

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

对给定的 α 有

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

由此可得 y_0 的 $1 - \alpha$ 的预测区间为

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

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 $x'_0 = (1, 1.5, 7.50, 1315)$.

解: 代入计算可得

$$\hat{y}_0 = 397.087 - 110.750 \times 1.5 + 15.583 \times 7.50 + (-0.058) \times 1315 = 271.564,$$

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

则 y_0 的预测区间为

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

习题

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

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

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

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

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, $(X - \mu)'A(X - \mu) \sim \chi_r^2$, 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

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

求 $E(X)$ 及 $\text{Cov}(X)$ 。

14. 利用二次型的均值及二元正态分布, 计算积分

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

15. 为了考察某种维尼纶纤维的耐水性, 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 θ 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 $cov(\hat{\theta})$.

17. 设

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

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: $X\beta = (X_1, X_2) \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$.

(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 $y_i \sim N(i\theta, i^2\sigma^2)$, $i = 1, 2, \dots, n$, 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 $y = X\beta + e$, $E(e) = 0$, $\text{cov}(e) = \sigma^2 V$, $V > 0$, $R(X) = p$.

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

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

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

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

25. 设 $y_i = \beta_0 + \beta_1 x_i + e_i$, $i = 1, 2, \dots, n$. 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 $H_i: \beta_i = 0$, $i = 1, 2, \dots, p - 1$ (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

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

其中 $E(e_i) = 0$, $\text{var}(e_i) = \sigma^2$, 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 $\text{Var}(\hat{y}_0 - y_0)$.

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

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.