
# 基础数理统计（研究生公共课）

肖柳青 教授 博士 主讲

上海交通大學
SHANGHAI JIAO TONG UNIVERSITY

# 第4章 回归分析

# （Regress Analysis ）

本章内容

- 一元线性回归
  - 参数估计
  - 假设检验
    - 回归检验
    - 系数检验
  - 预　　测
- 一元非线性回归
- 多元线性回归
  - 参数估计
  - 假设检验
    - 回归检验
    - 系数检验
  - 预　　测
- 线性岭回归与LASSO回归

# §4.2 多元线性回归

## 一. 模型与参数估计

### 1. $p$元线性回归模型

设 $x_1, x_2, \cdots, x_p$ 是可控变量，$Y$ 是依赖 $x_1, x_2, \cdots, x_p$ 的随机变量，它们之间关系为

$$\boxed{Y = a + b_1 x_1 + \cdots + b_p x_p + \varepsilon}$$

⇨ $p$元线性回归模型

其中 $\varepsilon \sim N(0, \sigma^2)$, $a, \underbrace{b_1, \cdots, b_p}_{\text{回归系数}}, \sigma = const.$

$$\boxed{\tilde{y} = EY = a + b_1 x_1 + \cdots + b_p x_p}$$

⇨ $p$元线性回归平面方程

当 $\tilde{y}$ 为 $x_1, x_2, \cdots, x_p$ 的线性函数时，称为线性回归，否则称为非线性回归.

## 多元线性回归的主要任务

(1) 用 $n$ 组观察值 $(x_{i1}, \cdots, x_{ip}, y_i)$ $i = 1 \sim n$，对未知参数 $a, b_1, \cdots, b_p, \sigma^2$ 进行点估计;

(2) 对回归系数 $b_1, \cdots, b_p$ 作假设检验;

(3) 在点 $(x_{01}, \cdots, x_{0p})$ 处对 $Y$ 作预测.

# 模型：

$p$ 元线性回归模型

$$\begin{cases} Y_1 = a + b_1 x_{11} + b_2 x_{12} \cdots + b_p x_{1p} + \varepsilon_1 \\ Y_2 = a + b_1 x_{21} + b_2 x_{22} \cdots + b_p x_{2p} + \varepsilon_2 \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ Y_n = a + b_1 x_{n1} + b_2 x_{n2} \cdots + b_p x_{np} + \varepsilon_n \end{cases}$$

$$\Downarrow$$

简写为 $Y = X\beta + \varepsilon$ 　 $E\varepsilon = 0$，$\mathrm{cov}(\varepsilon, \varepsilon) = \sigma^2 I_n$

其中 $Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}$ $X = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}$ $\beta = \begin{pmatrix} a \\ b_1 \\ \vdots \\ b_p \end{pmatrix}$ $\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$

## 1、参数估计：未知参数 $a,b_1,\cdots,b_p,\sigma^2$的估计

选择 $a,b_1,\cdots,b_p$，使离差平方和 $Q$ 最小，

$$Q=\sum_{i=1}^{n}(y_i-\tilde{y}_i)^2=\sum_{i=1}^{n}(y_i-a-b_1x_{i1}-\cdots-b_px_{ip})^2=\min$$

令

$$\begin{cases}\dfrac{\partial Q}{\partial a}=-2\sum\limits_{i=1}^{n}(y_i-a-b_1x_{i1}-\cdots-b_px_{ip})=0\\ \dfrac{\partial Q}{\partial b_1}=-2\sum\limits_{i=1}^{n}(y_i-a-b_1x_{i1}-\cdots-b_px_{ip})x_{i1}=0\\ \cdots\cdots\cdots\cdots\cdots\cdots\\ \dfrac{\partial Q}{\partial b_p}=-2\sum\limits_{i=1}^{n}(y_i-a-b_1x_{i1}-\cdots-b_px_{ip})x_{ip}=0\end{cases}\Rightarrow$$

# 正规方程组

$$
\begin{cases}
na+(\sum x_{i1})b_1+(\sum x_{i2})b_2+\cdots+(\sum x_{ip})b_p=\sum y_i \\
(\sum x_{i1})a+(\sum x_{i1}^2)b_1+(\sum x_{i1}x_{i2})b_2+\cdots+(\sum x_{i1}x_{ip})b_p=\sum x_{i1}y_i \\
\cdots\cdots\cdots\cdots\cdots\cdots \\
(\sum x_{ip})a+(\sum x_{ip}x_{i1})b_1+(\sum x_{ip}x_{i2})b_2+\cdots+(\sum x_{ip}^2)b_p=\sum x_{ip}y_i
\end{cases}
$$

$\Downarrow$

即 $(X^TX)\beta=X^TY$ (*) 简记作 $A\beta=B$. 其中

$$
A=X^TX=\begin{pmatrix}
n & \sum x_{i1} & \sum x_{i2} & \cdots & \sum x_{ip} \\
\sum x_{i1} & \sum x_{i1}^2 & \sum x_{i1}x_{i2} & \cdots & \sum x_{i1}x_{ip} \\
\vdots & \vdots & \vdots & \cdots & \vdots \\
\sum x_{ip} & \sum x_{ip}x_{i1} & \sum x_{ip}x_{i2} & \cdots & \sum x_{ip}^2
\end{pmatrix}
\quad
B=X^TY=\begin{pmatrix}
\sum y_i \\
\sum x_{i1}y_i \\
\vdots \\
\sum x_{ip}y_i
\end{pmatrix}
$$

# 经验回归方程



当系数矩阵A可逆时，有唯一解

$$\hat{\beta} = A^{-1}B = (X^T X)^{-1}(X^T Y) \quad \cdots\cdots\cdots\cdots\cdots(1)$$

于是得经验回归平面方程

$$\hat{y} = \hat{a} + \hat{b}_1 x_1 + \cdots + \hat{b}_p x_p$$

正规方程组第一个方程两边除以$n$，再移项得

$$\hat{a} = \bar{y} - \hat{b}_1 \bar{x}_1 - \cdots - \hat{b}_p \bar{x}_p$$

为计算方便，将$Q_{min}$改写成另一形式

# 推导的另一形式：



$$Q_{\min} = \sum_{i=1}^{n}(Y_i - \hat{y}_i)^2 = \left\| Y - X\hat{\beta} \right\|^2$$

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

$$= Y^T Y - \underline{Y^T X\hat{\beta}} - \underline{\hat{\beta}^T X^T Y} + \hat{\beta}^T X^T X\hat{\beta}$$

$$= Y^T Y - 2\hat{\beta}^T X^T Y + \hat{\beta}^T X^T X\hat{\beta}$$

$$\frac{\partial Q}{\partial \hat{\beta}} = -2X^T Y + 2X^T X\hat{\beta} = 0$$

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

**例1** 某化工产品的得率 $y$ 与反应温度 $x_1$，反应时间 $x_2$ 及某反应物浓度 $x_3$ 有关. 试验结果如下（$x_1, x_2, x_3$ 均为二水平且以编码形式表示）

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $x_1$ | -1 | -1 | -1 | -1 | 1 | 1 | 1 | 1 |
| $x_2$ | -1 | -1 | 1 | 1 | -1 | -1 | 1 | 1 |
| $x_3$ | -1 | 1 | -1 | 1 | -1 | 1 | -1 | 1 |
| $y$ | 7.6 | 10.3 | 9.2 | 10.2 | 8.4 | 11.1 | 9.8 | 12.6 |

若以模型 $y = a + b_1 x_1 + b_2 x_2 + b_3 x_3$ 来拟合它，试求 $y$ 的三元线性回归方程.

**解一**

$$X=\begin{pmatrix}1&1&1&1&1&1&1&1\\-1&-1&-1&-1&1&1&1&1\\-1&-1&1&1&-1&-1&1&1\\-1&1&-1&1&-1&1&-1&1\end{pmatrix}^T \qquad X^TY=\begin{pmatrix}79.2\\4.6\\4.4\\9.2\end{pmatrix}$$

$$Y=\begin{pmatrix}7.6&10.3&9.2&10.2&8.4&11.1&9.8&12.6\end{pmatrix}^T$$

$$X^TX=\begin{pmatrix}8&0&0&0\\0&8&0&0\\0&0&8&0\\0&0&0&8\end{pmatrix} \qquad (X^TX)^{-1}=\frac{1}{8}\begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{pmatrix}$$

$$X^T Y = \begin{pmatrix} 79.2 \\ 4.6 \\ 4.4 \\ 9.2 \end{pmatrix}, \quad \hat{\beta} = \begin{pmatrix} \hat{a} \\ \hat{b}_1 \\ \hat{b}_2 \\ \hat{b}_3 \end{pmatrix} = (X^T X)^{-1}(X^T Y) = \begin{pmatrix} 9.9 \\ 0.575 \\ 0.55 \\ 1.15 \end{pmatrix}.$$

所以 $y$ 的三元线性回归方程为

$$\hat{y} = 9.9 + 0.575\, x_1 + 0.55\, x_2 + 1.15\, x_3.$$

# 2、系数估计量的分布

## $\hat{\beta}$ 的分布

由 (1) 式知 $\hat{\beta}$ 服从 $p+1$ 维正态分布，因为 $Y=\begin{pmatrix} Y_1 & Y_2 & \cdots & Y_n \end{pmatrix}^T$ 的所有分量是独立正态变量，正态变量在线性变换下具有不变性.

为得到 $\hat{\beta}$ 的概率分布，需分别计算它的数学期望矢量和协方差矩阵.

# 3、最小二乘估计量的性质：（p. 104）

（1）$\hat{\beta}$是线性估计且$\hat{\beta}=\left(X^TX\right)^{-1}X^Ty\sim N_{p+1}(\beta,(X^TX)^{-1}\sigma^2)$

证：$\hat{\beta}=\left(X^TX\right)^{-1}X^Ty$服从$p+1$维正态分布，$\hat{\beta}$是线性估计

$$E\hat{\beta}=\left(X^TX\right)^{-1}X^TEy=\left(X^TX\right)^{-1}X^TX\beta=\beta$$

$$D\hat{\beta}=D\left[(X^TX)^{-1}X^Ty\right]=\left((X^TX)^{-1}X^T\right)Dy\left((X^TX)^{-1}X^T\right)^T$$

$$=(X^TX)^{-1}X^TDyX(X^TX)^{-1}=(X^TX)^{-1}\sigma^2$$

$$\hat{\beta}=\left(X^TX\right)^{-1}X^Ty\sim N_{p+1}(\beta,(X^TX)^{-1}\sigma^2)$$

# （1–1）无偏性

$$E\hat{\beta} = E(A^{-1}B) = E[(X^TX)^{-1}(X^TY)]$$
$$= (X^TX)^{-1}X^TEY = (X^TX)^{-1}X^TE(X\beta+\varepsilon)$$
$$= (X^TX)^{-1}X^TX\beta = \beta.$$

即

$$E\hat{\beta} = E\begin{pmatrix} \hat{a} \\ \hat{b}_1 \\ \vdots \\ \hat{b}_p \end{pmatrix} = \begin{pmatrix} E\hat{a} \\ E\hat{b}_1 \\ \vdots \\ E\hat{b}_p \end{pmatrix} = \begin{pmatrix} a \\ b_1 \\ \vdots \\ b_p \end{pmatrix} = \beta.$$

(1 – 2) $cov(\widehat{\beta},\widehat{\beta}) = \sigma^2(X^TX)^{-1}$

$\hat{\beta}$ 的协方差矩阵为

$$\begin{pmatrix} D\hat{a} & \operatorname{cov}(\hat{a},\hat{b}_1) & \cdots & \operatorname{cov}(\hat{a},\hat{b}_p) \\ \operatorname{cov}(\hat{b}_1,\hat{a}) & D\hat{b}_1 & \cdots & \operatorname{cov}(\hat{b}_1,\hat{b}_p) \\ \vdots & \vdots & \vdots & \vdots \\ \operatorname{cov}(\hat{b}_p,\hat{a}) & \operatorname{cov}(\hat{b}_p,\hat{b}_1) & \cdots & D\hat{b}_p \end{pmatrix}$$

$$= E\left[(\hat{\beta} - E\hat{\beta})(\hat{\beta} - E\hat{\beta})^T\right]$$

$$=E[(X^TX)^{-1}X^T(\mathrm{Y}-E\mathrm{Y})(\mathrm{Y}-E\mathrm{Y})^T\mathrm{X}(X^TX)^{-1}$$

$$= \sigma^2(X^TX)^{-1} = \sigma^2A^{-1} \triangleq = \sigma^2C.$$

推导过程详见教材P. 105页.

（2） $\hat{\beta}$ 是 $\beta$ 的最小方差线性无偏估计

证：设 $T$ 是 $\beta$ 的任一线性无偏估计，则 $T=Ay$，且 $ET=AEy=AX\beta=\beta$，

$DT=D(Ay)=ADyA^T=\sigma^2AA^T$。由 $\beta$ 的任意性知 $AX=I_{p+1}$，而

$$\begin{aligned}\left[A-(X^TX)^{-1}X^T\right]\left[A-(X^TX)^{-1}X^T\right]^T &= AA^T-AX(X^TX)^{-1}-(X^TX)^{-1}X^TA^T+(X^TX)^{-1}\\ &= AA^T-AX(X^TX)^{-1}-\left[AX(X^TX)^{-1}\right]^T+(X^TX)^{-1}\\ &= AA^T-(X^TX)^{-1}-(X^TX)^{-1}+(X^TX)^{-1}\\ &= AA^T-(X^TX)^{-1}\end{aligned}$$

非负定。$DT-D\hat{\beta}=\sigma^2AA^T-(X^TX)^{-1}\sigma^2=\left[AA^T-(X^TX)^{-1}\right]\sigma^2$ 非负定，即 $\hat{\beta}$ 是 $\beta$ 的最小方差线性无偏估计。□

(3) 残差向量 $\hat{\varepsilon} = y - \hat{y} = y - X\hat{\beta}$ 与 $\hat{\beta}$ 互不相关

证：$\hat{\varepsilon} = y - X\hat{\beta} = \left(I - X(X^T X)^{-1} X^T\right) y$

$$Cov(\hat{\varepsilon}, \hat{\beta}) = Cov\left(\left[I - X(X^T X)^{-1} X^T\right] y, (X^T X)^{-1} X^T y\right) = \left[I - X(X^T X)^{-1} X^T\right] Dy \left[(X^T X)^{-1} X^T\right]^T$$

$$= \left[I - X(X^T X)^{-1} X^T\right]\left[X(X^T X)^{-1}\right]\sigma^2 = \left[X(X^T X)^{-1} - X(X^T X)^{-1}\right]\sigma^2 = 0$$

即残差向量 $\hat{\varepsilon} = y - \hat{y} = y - X\hat{\beta}$ 与 $\hat{\beta}$ 互不相关。□

（4） $E\hat{\varepsilon}=0$，$D\hat{\varepsilon}=\left[I_n-X(X^TX)^{-1}X^T\right]\sigma^2$

证：$E\hat{\varepsilon}=Ey-XE\hat{\beta}=X\beta-X\beta=0$

$$D\hat{\varepsilon}=D(y-X\hat{\beta})=D\left[\left(I-X(X^TX)^{-1}X^T\right)y\right]=\left(I-X(X^TX)^{-1}X^T\right)Dy\left(I-X(X^TX)^{-1}X^T\right)^T$$

$$=\left(I-X(X^TX)^{-1}X^T\right)\left(I-X(X^TX)^{-1}X^T\right)^T\sigma^2=\left(I-X(X^TX)^{-1}X^T\right)\sigma^2$$

注：投影阵：$A^2=A$，$A^T=A$。

（5）残差平方和$Q(\hat{\beta})=(y-X\hat{\beta})^T(y-X\hat{\beta})$的期望值为$(n-p-1)\sigma^2$，$\sigma^2$的无偏估计为

$$\hat{\sigma}^2=\frac{Q(\hat{\beta})}{n-p-1}=\frac{1}{n-p-1}y^T\left[I_n-X(X^TX)^{-1}X^T\right]y。$$

证：残差平方和$Q(\hat{\beta})=\hat{\varepsilon}^T\hat{\varepsilon}=(y-X\hat{\beta})^T(y-X\hat{\beta})$

$$\begin{aligned}EQ(\hat{\beta})&=E\hat{\varepsilon}^T\hat{\varepsilon}=Etr(\hat{\varepsilon}^T\hat{\varepsilon})=Etr(\hat{\varepsilon}\hat{\varepsilon}^T)=tr(E\hat{\varepsilon}\hat{\varepsilon}^T)=tr(D\hat{\varepsilon})=tr(\left[I_n-X(X^TX)^{-1}X^T\right]\sigma^2)\\&=\sigma^2\left[n-tr\left(X(X^TX)^{-1}X^T\right)\right]=\sigma^2\left[n-tr\left(X^TX(X^TX)^{-1}\right)\right]=\sigma^2\left[n-tr\left(I_{p+1}\right)\right]\\&=(n-p-1)\sigma^2\end{aligned}$$

$\sigma^2$的无偏估计$\hat{\sigma}^2=\dfrac{Q(\hat{\beta})}{n-p-1}=\dfrac{1}{n-p-1}y^T\left[I_n-X(X^TX)^{-1}X^T\right]y$。□

（6）**定理 2** ① $\hat{\beta}$ 和 $\hat{\varepsilon}$ 相互独立，且服从正态分布；

② $\hat{\beta}$ 与 $Q(\hat{\beta})$ 相互独立，即 $\hat{\beta}$ 与 $\hat{\sigma}^2$ 相互独立；

③ $\dfrac{Q(\hat{\beta})}{\sigma^2} \sim \chi^2(n-p-1)$

# 二. 多元线性回归中的检验与预测

## 1. 多元线性回归的显著性检验p. 108

要检验多元线性回归模型

$$\begin{cases} Y = a + b_1 x_1 + \cdots + b_p x_p + \varepsilon \\ \varepsilon \sim N(0, \sigma^2),\ a, b_1, \cdots, b_p, \sigma = const. \end{cases}$$

变量间有无线性联系，只要检验假设

$$H_0:\ b_1 = b_2 = \cdots = b_p = 0.$$

若成立，则认为线性回归不显著；反之显著.

下面采用方差分析法作检验.

将观察值代入经验回归平面方程有

$$\hat{y}_i = \hat{a} + \hat{b}_1 x_{i1} + \cdots + \hat{b}_p x_{ip}, \quad i = 1, 2, \cdots, n.$$

总离差平方和

$$Q_T = \sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum_{i=1}^{n}[(Y_i - \hat{y}_i) + (\hat{y}_i - \bar{Y})]^2$$

$$= \sum_{i=1}^{n}(Y_i - \hat{y}_i)^2 + \sum_{i=1}^{n}(\hat{y}_i - \bar{Y})^2 + 2\underbrace{\sum_{i=1}^{n}(Y_i - \hat{y}_i)(\hat{y}_i - \bar{Y})}_{=0}$$

见附证

$$= \underbrace{\sum_{i=1}^{n}(Y_i - \hat{y}_i)^2}_{\text{剩余离差}} + \underbrace{\sum_{i=1}^{n}(\hat{y}_i - \bar{Y})^2}_{\text{回归离差}} \triangleq Q_{\text{剩}} + Q_{\text{回}}.$$

**附证：** 将 $\hat{y}_i = \hat{a} + \hat{b}_1 x_{i1} + \cdots + \hat{b}_p x_{ip}$ 代入整理得

$$\sum_{i=1}^{n}(Y_i - \hat{y}_i)(\hat{y}_i - \bar{Y}) = \sum_{i=1}^{n}(Y_i - \hat{a} - \hat{b}_1 x_{i1} - \cdots - \hat{b}_p x_{ip})$$

$$[(\hat{a} - \bar{Y}) + \hat{b}_1 x_{i1} + \cdots + \hat{b}_p x_{ip}]$$

$$= (\hat{a} - \bar{Y})\sum_{i=1}^{n}(Y_i - \hat{a} - \hat{b}_1 x_{i1} - \cdots - \hat{b}_p x_{ip})$$

$$+ \hat{b}_1 \sum_{i=1}^{n}(Y_i - \hat{a} - \hat{b}_1 x_{i1} - \cdots - \hat{b}_p x_{ip})x_{i1} + \cdots$$

$$+ \hat{b}_p \sum_{i=1}^{n}(Y_i - \hat{a} - \hat{b}_1 x_{i1} - \cdots - \hat{b}_p x_{ip})x_{ip} = 0.$$

正规方程组的约束条件

在 $H_0$成立的前提下(即 $Y_i = a + \varepsilon_i,\ i = 1 \sim n$)

构造统计量并求其分布.

$$Q_T / \sigma^2 = Q_{剩} / \sigma^2 + Q_{回} / \sigma^2 .$$

$$\frac{Q_T}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n}(a + \varepsilon_i - a - \bar{\varepsilon})^2$$

$$= \frac{1}{\sigma^2}\sum_{i=1}^{n}(\varepsilon_i - \bar{\varepsilon})^2 \sim \chi^2(n-1)$$

由分解定理 $\frac{Q_{剩}}{\sigma^2} \sim \chi^2(n-p-1)$, $\frac{Q_{回}}{\sigma^2} \sim \chi^2(p)$

正规方程组有 $p+1$ 个约束条件

由$F$分布的定义,

$$F=\frac{Q_{回}/p}{Q_{剩}/(n-p-1)}\sim F(p,n-p-1)$$

给定显著性水平$\alpha$,当

$$F\geq F_{\alpha}(p,n-p-1)$$

时拒绝$H_0$,即认为线性回归显著;反之,接受$H_0$,即认为线性回归不显著.

**例2** 某化工产品的得率 $y$ 与反应温度 $x_1$，反应时间 $x_2$ 及某反应物浓度 $x_3$ 有关. 试验结果如下 ($x_1, x_2, x_3$ 的水平以编码形式表示)

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **x1** | -1 | -1 | -1 | -1 | 1 | 1 | 1 | 1 |
| **x2** | -1 | -1 | 1 | 1 | -1 | -1 | 1 | 1 |
| **x3** | -1 | 1 | -1 | 1 | -1 | 1 | -1 | 1 |
| **y** | 7.6 | 10.3 | 9.2 | 10.2 | 8.4 | 11.1 | 9.8 | 12.6 |

前面例1中已求出 $y$ 的线性回归方程为

$$\hat{y} = 9.9 + 0.575x_1 + 0.55x_2 + 1.15x_3.$$

检验回归的显著性, 显著水平取5%.

**解** 作假设 $H_0: b_1 = b_2 = b_3 = 0.$ 计算

$$\sum_{i=1}^{8} y_i = 79.2, \qquad \sum_{i=1}^{8} y_i^2 = 801.1,$$

$$\sum_{i=1}^{8} x_{i1} y_i = 4.6, \quad \sum_{i=1}^{8} x_{i2} y_i = 4.4, \quad \sum_{i=1}^{8} x_{i3} y_i = 9.2,$$

$$Q_T = \sum_{i=1}^{8} y_i^2 - \frac{1}{8}\left(\sum_{i=1}^{8} y_i\right)^2 = 801.1 - \frac{1}{8} \times 79.2^2 = 17.02$$

$$Q_{剩} = \sum y_i^2 - \left(\sum y_i\right)\hat{a} - \left(\sum x_{i1} y_i\right)\hat{b}_1 - \left(\sum x_{i2} y_i\right)\hat{b}_2 - \left(\sum x_{i3} y_i\right)\hat{b}_3$$

$$= 801.1 - 79.2 \times 9.9 - 4.6 \times 0.575 - 4.4 \times 0.55 - 9.2 \times 1.15$$

$$= 1.375$$

$$Q_{回} = Q_T - Q_{剩} = 17.02 - 1.375 = 15.645.$$

# F检验的方差分析表

| 来源 | 离　差 | 自由度 | 均方离差 | $F$　值 |
|---|---|---|---|---|
| 回归 | $Q_{回}=15.645$ | 3 | 5.215 | $F=15.16$ |
| 剩余 | $Q_{剩}=1.375$ | 4 | 0.344 | |
| 总和 | $Q_T=17.02$ | 7 | | |

查表$F_\alpha(3,4)=6.59$，由于$F=15.16>6.59$

故拒绝$H_0$，即认为线性回归显著.

# 2. 局部回归系数的显著性检验 p. 108

若线性回归显著，则$b_1, b_2, \cdots, b_p$不全为零. 此时，系数为零的自变量对$Y$的值不起作用. 因此检验某个回归系数 $b_j$ 是否为零，相当于检验相应的 $x_j$ 对$Y$ 的值是否起作用.

对假设 $H_0: b_j = 0$, $j$ 固定，$1 \le j \le p$ 可用$\hat{b}_j$作检验.

$\hat{b}_j$是矢量 $\hat{\beta} = (X^T X)^{-1}(X^T Y)$ 的第$j$ 个分量，且是$Y_1, Y_2, \cdots, Y_n$的线性组合，故$\hat{b}_j \sim N(E\hat{b}_j, D\hat{b}_j)$.

由性质1知 $E\hat{b}_j = b_j,\ D\hat{b}_j = c_{jj}\sigma^2.$

其中 $c_{jj}$ 是矩阵 $C=(X^TX)^{-1}$ 主对角线上第 $j$ 个元素.

$$U = \frac{\hat{b}_j - b_j}{\sqrt{c_{jj}}\,\sigma} \sim N(0,1),\quad \frac{Q_{剩}}{\sigma^2} \sim \chi^2(n-p-1),$$

在 $H_0$ 成立的前提下(即 $b_j=0$) 有

$$T = \frac{\hat{b}_j}{\sqrt{c_{jj}}\sqrt{Q_{剩}/(n-p-1)}} \sim t(n-p-1).$$

给定显著性水平$\alpha$，当

$$\left| T \right| \geq t_{\alpha/2}(n-p-1)$$

时拒绝 $H_0$，即认为$b_j$显著不为零; 反之，接受 $H_0$，认为 $b_j$显著地等于零.

$$E(Q_{剩}/\sigma^2) = n-p-1, \text{故} \quad \hat{\sigma}^{*2} = \frac{Q_{剩}}{n-p-1} \text{是}$$

$\sigma^2$ 的无偏估计. 于是检验统计量为

$$T = \frac{\hat{b}_j}{\sqrt{c_{jj}}\,\hat{\sigma}^*}$$

$$\chi^2 = (n-2)\hat{\sigma}^{*2}/\sigma^2 \sim \chi^2(n-2)$$

且$\hat{b}$与$\hat{\sigma}^{*2}$相互独立，由 $t$ 分布定义得

$$T = \frac{\hat{b}-b}{\hat{\sigma}^*}\sqrt{S_{xx}} \sim t(n-2) \quad \cdots\cdots\cdots\cdots \quad (1)$$

给定显著性水平$\alpha$，取检验统计量 $T = \frac{\hat{b}}{\hat{\sigma}^*}\sqrt{S_{xx}}$

当$|T| \geq t_{\alpha/2}(n-2)$时，拒绝 $H_0$，认为线性回归显著，否则不显著.

**例3** 检验例1中线性回归的显著性，取 $\alpha = 0.05.$

**解** 作假设 $H_{0i}: b_i = 0,\ \ i = 1, 2, 3.$

例1中已算出

$$\hat{b}_1 = 0.575,\ \ \hat{b}_2 = 0.55,\ \ \hat{b}_3 = 1.15.$$

此时 $C = A^{-1} = (X^T X)^{-1} = I_4 / 8$

从而 $c_{jj} = 1/8,\ j = 0, 1, 2, 3$

由例1 $\hat{\sigma}^{*2} = \dfrac{Q_{剩}}{n-p-1} = 0.344,\qquad \hat{\sigma}^{*} = 0.587.$

拒绝域 $W_j$: $$|T_j| = \frac{\hat{b}_j}{\sqrt{c_{jj}}\,\hat{\sigma}^*} \geq t_{0.025}(4) = 2.7764$$

$$|T_1| = \frac{0.575}{0.587\sqrt{0.125}} = 2.7706 < 2.7764 \notin W_1$$

$$|T_2| = \frac{0.55}{0.587\sqrt{0.125}} = 2.6501 < 2.7764 \notin W_2$$

$$|T_3| = \frac{1.15}{0.587\sqrt{0.125}} = 5.5412 > 2.7764 \in W_3$$

接受 $H_{01},H_{02}$，拒绝 $H_{03}$，即 $b_1,b_2$ 显著为零，$b_3$ 显著不为零．所以前面例1配的自变量 $x_1,x_2,x_3$ 的经验回归平面方程是不合适的．

在回归系数显著为零的情形可以剔除自变量．

一般，对回归系数进行一次检验后，只能剔除其中一个因子，这个因子是所有不显著因子中 $F$ 值或 $t$ 值为最小的，然后重

新建立新的回归方程，再对新的回归系数逐个进行检验，直到余下的回归系数都显著为止.

本例 $T_3 > T_1 > T_2$ 只能将$x_2$剔除，此时设新的回归方程为

$$y = a + b_1 x_1 + b_3 x_3$$

一切又重新开始.

**例4** 为研究某一化学反应过程中，温度 $x$ ($^{0}$C) 对产品收率 $y$ (%) 的影响，测得数据如下

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 温度 $x$ | 100 | 110 | 120 | 130 | 140 | 150 | 160 | 170 | 180 | 190 |
| 收率 $y$ | 45 | 51 | 54 | 61 | 66 | 70 | 74 | 78 | 85 | 89 |

上节已求出经验回归直线方程

$$\hat{y} = a + bx$$

现检验线性回归是否显著，显著水平取5%.

**解** 作假设 $H_0: b=0$; $H_1: b\neq 0.$

取统计量 $T=\dfrac{\hat{b}}{\hat{\sigma}^*}\sqrt{S_{xx}}\sim t(n-2)$

拒绝域 $W$: $|T|\geq t_{0.025}(8)=2.306$

$$\hat{\sigma}^{*2}=\frac{n}{n-2}\hat{\sigma}^2=\frac{10}{8}\times 0.7345=0.9181$$

$$|T|=\frac{0.483}{\sqrt{0.9181}}\sqrt{8250}=45.7844>2.306\in W$$

拒绝 $H_0$, 认为线性回归效果显著.

注　当回归效果显著时，常需要对回归系数 $b$ 作区间估计. 事实上，由 (1) 式得到 $b$ 的置信度为 $1-\alpha$ 的置信区间为

$$\left(\hat{b}-t_{\alpha/2}(n-2)\frac{\hat{\sigma}^*}{\sqrt{S_{xx}}},\ \hat{b}+t_{\alpha/2}(n-2)\frac{\hat{\sigma}^*}{\sqrt{S_{xx}}}\right)$$

如例1中回归系数 $b$ 的置信区间是

$$\left(0.483 \pm 2.306\frac{0.9181}{\sqrt{8250}}\right)=(0.460,\ 0.506)$$

# 推广

若检验假设 $H_0: b=b_0$; $H_1: b\neq b_0$.

取检验统计量 $T=\dfrac{\hat{b}-b_0}{\hat{\sigma}^*}\sqrt{S_{xx}}$

当 $|T|\geq t_{\alpha/2}(n-2)$ 时, 拒绝 $H_0$, 认为回归系数与 $b_0$ 有显著差异, 否则无显著差异.

# 注 上面介绍的t检验与F检验等价

因为两种检验所用统计量

$$T = \frac{\hat{b}}{\hat{\sigma}^*}\sqrt{S_{xx}} \sim t(n-2)$$

$$F = \frac{S_R}{S_E/(n-2)} \sim F(1, n-2)$$

存在关系 $T^2 = F.$

# 线性回归效果不显著的可能原因

1. 影响 $Y$ 取值的除 $x$ 外还有其它不可忽略的因素.
2. $Y$ 与 $x$ 的存在非线性关系.
3. $Y$ 与 $x$ 无关.

若要对以上三种情形配线性回归模型，都有 $b=0$，即

$$Y = a + \varepsilon$$

# 2. 预测

回归方程的一个重要应用是预测. 即当 $x = x_0$ 时, 以一定的置信度对 $Y$ 作区间估计. 气象台的气温预报便是典型的预测.

当 $x = x_0$ 时, $Y$ 的值为 $Y_0 = a + bx_0 + \varepsilon_0$, $\varepsilon_0 \sim N(0, \sigma^2)$

取 $x_0$ 处的回归值 $\hat{y}_0 = \hat{a} + \hat{b}x_0$ 作为 $Y_0$ 的预测值, 则两者之差仍服从正态分布, 即

$$Y_0 - \hat{y}_0 = Y_0 - \hat{a} - \hat{b}x_0 \sim N\left(0, \left[1 + \frac{1}{n} + \frac{(x_0 - x)^2}{S_{xx}}\right]\right)$$

事实上，

$$E(Y_0 - \hat{y}_0) = E(a + bx_0 + \varepsilon_0 - \hat{a} - \hat{b}x_0)$$

$$= a + bx_0 - E\hat{a} - bx_0 = a - E(\bar{Y} - \hat{b}\bar{x})$$

$$= a - E\bar{Y} + b\bar{x} = a - (a + b\bar{x}) + b\bar{x} = 0,$$

$$D(Y_0 - \hat{y}_0) = D(a + bx_0 + \varepsilon_0) + D(\hat{a} + \hat{b}x_0)$$

$$= \sigma^2 + D[\bar{Y} + \hat{b}(x_0 - \bar{x})] = \sigma^2 + D\bar{Y}$$

$$+ D[\hat{b}(x_0 - \bar{x})] + 2\operatorname{cov}(\bar{Y}, \hat{b}(x_0 - \bar{x}))$$

$$= \sigma^2 + \frac{\sigma^2}{n} + \frac{(x_0 - x)^2 \sigma^2}{S_{xx}} + 2(x_0 - x)E[(\bar{Y} - E\bar{Y})(\hat{b} - E\hat{b})]$$

$$E[(\bar{Y} - E\bar{Y})(\hat{b} - E\hat{b})]$$ P. 183第4、7行

$$= E\{\frac{1}{n}\sum_{i=1}^{n}(Y_i - EY_i)[\sum_{i=1}^{n}(x_i - \bar{x})(Y_i - EY_i)]/S_{xx}\}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i - \bar{x})E[(Y_i - EY_i)(Y_j - EY_j)]/S_{xx}\}$$

$$= \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})\sigma^2 / S_{xx} = 0.$$

$$D(Y_0-\hat{y}_0)=[1+\frac{1}{n}+\frac{(x_0-x)^2}{S_{xx}}]\sigma^2$$

由 $$U=\frac{Y_0-\hat{a}-\hat{b}x_0}{\sqrt{1+\frac{1}{n}+\sigma(x_0-\bar{x})^2/S_{xx}}}\sim N(0,1)$$

又 $$\chi^2=(n-2)\hat{\sigma}^{*2}/\sigma^2\sim\chi^2(n-2)$$

得 $$T=\frac{Y_0-\hat{a}-\hat{b}x_0}{\sqrt{1+\frac{1}{n}+(x_0-\bar{x})^2/S_{xx}}\ \hat{\sigma}^*}\sim t(n-2)$$

给定置信概率 $1-\alpha$，由

$$P(|T|<t_{\alpha/2}(n-2))=1-\alpha$$

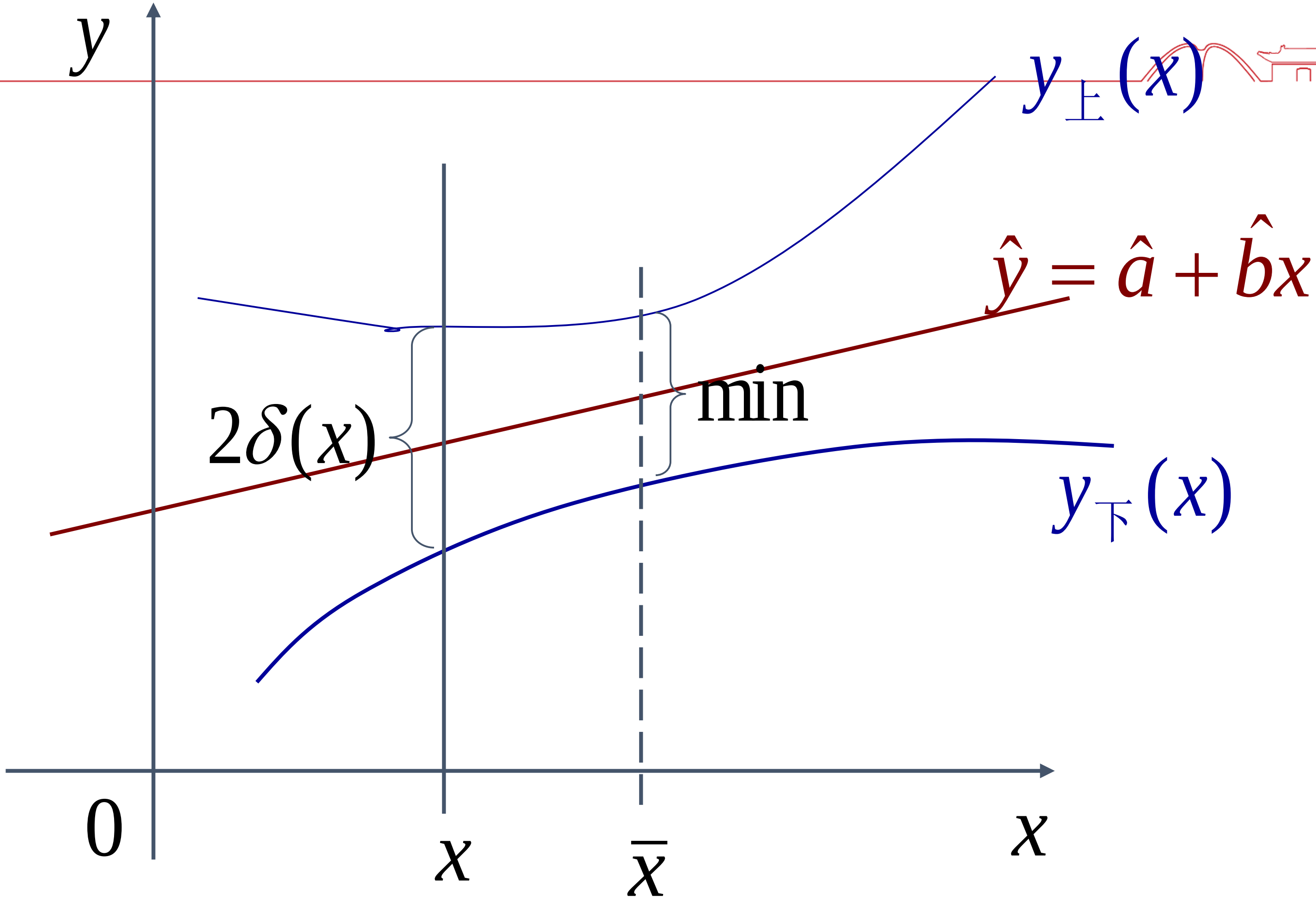
得 $Y_0$ 的置信区间

$$(y_{下}(x),\ y_{上}(x))=(\hat{y}-\delta(x),\ \hat{y}+\delta(x)) \quad (*)$$

其中 $\hat{y}=\hat{a}+\hat{b}x_0$

$$\delta(x)=t_{\alpha/2}(n-2)\sqrt{1+\frac{1}{n}+\frac{(x_0-\overline{x})^2}{S_{xx}}}\hat{\sigma}^*$$

# 置信区间的几何意义




$x$ 离 $\bar{x}$ 愈近，置信区间愈短，估计精度愈高.

**例2** 求例1中温度 $x_0 = 135\ (^0\text{C})$ 时产品收率 $y(\%)$ 的预测区间，置信度取95%.

**解** $\hat{y} = \hat{a} + \hat{b}x_0 = -2.735 + 0.483 \times 135 = 62.470$

$$\delta(x) = t_{\alpha/2}(n-2)\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}\,\hat{\sigma}^*$$

$$= 2.306\sqrt{1 + \frac{1}{10} + \frac{(135-145)^2}{8250}} \times 0.9582 = 2.330$$

由 (*) 式得预测区间为

$$(\hat{y} - \delta(x),\ \hat{y} + \delta(x)) = (60.14,\ 64.80)$$

**注** 在实际回归问题中，子样容量常很大. 此时对于在 $\bar{x}$ 附近的 $x$，不仅能得到较短的预测区间，还可简化(*)式

$$n > 45 \Rightarrow t_{\alpha/2}(n-2) \approx u_{\alpha/2}, \quad \sqrt{1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}} \approx 1,$$

于是(*)式变为

$$(\hat{y}-u_{\alpha/2}\hat{\sigma}^*,\ \hat{y}+u_{\alpha/2}\hat{\sigma}^*)$$

若取 $u_{0.025}=1.96\approx 2$，则 $(\hat{y}-2\hat{\sigma}^*,\ \hat{y}+2\hat{\sigma}^*)$.

# 例5 设线性回归模型为：

$$\begin{cases} Y_1 = -a + b + \varepsilon_1 \\ Y_2 = a + \varepsilon_2 \\ Y_3 = b + \varepsilon_3 \\ Y_4 = a - b + \varepsilon_4 \end{cases}$$

其中 $\varepsilon_1, \ldots, \varepsilon_4 \overset{i.i.d.}{\sim} N(0,\sigma^2)$，未知 $\sigma^2$，

- （1）求参数 $a, b$ 的最小二乘估计 $\hat{a}, \hat{b}$
- （2）求 $\sigma^2$ 的无偏估计

# 解： （1）将线性回归模型写出矩阵标准型

$Y = X\beta + \varepsilon$,其中

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{pmatrix}, \quad X = \begin{pmatrix} -1 & 1 \\ 1 & 0 \\ 0 & 1 \\ 1 & -1 \end{pmatrix}, \quad \beta = \begin{pmatrix} a \\ b \end{pmatrix}, \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{pmatrix}$$

且$X^T X = \begin{pmatrix} 3 & -2 \\ -2 & 3 \end{pmatrix}, \quad (X^T X)^{-1} = \frac{1}{5}\begin{pmatrix} 3 & 2 \\ 2 & 3 \end{pmatrix}$

$$\hat{\beta} = (X^T X)^{-1} X^T Y = \begin{pmatrix} -\frac{1}{5}Y_1 + \frac{3}{5}Y_2 + \frac{2}{5}Y_3 + \frac{1}{5}Y_4 \\ \frac{1}{5}Y_1 + \frac{2}{5}Y_2 + \frac{3}{5}Y_3 - \frac{1}{5}Y_4 \end{pmatrix}$$

即$a, b$的最小二乘估计为

$$\hat{a} = -\frac{1}{5}Y_1 + \frac{3}{5}Y_2 + \frac{2}{5}Y_3 + \frac{1}{5}Y_4$$

$$\hat{b} = \frac{1}{5}Y_1 + \frac{2}{5}Y_2 + \frac{3}{5}Y_3 - \frac{1}{5}Y_4$$

# (3)

$$Q_e = Y^T(I_4 - X(X^TX)^{-1}X^T)Y = \frac{1}{5}Y^T\begin{bmatrix} 3 & 1 & -1 & 2 \\ 1 & 2 & -2 & -1 \\ -1 & -2 & 2 & 1 \\ 2 & -1 & 1 & 3 \end{bmatrix}Y$$

$\sigma^2$ 的无偏估计为

$$\hat{\sigma}^2 = \frac{Q_e}{n-k} = \frac{1}{2}Q_e = \frac{1}{10}Y^T\begin{bmatrix} 3 & 1 & -1 & 2 \\ 1 & 2 & -2 & -1 \\ -1 & -2 & 2 & 1 \\ 2 & -1 & 1 & 3 \end{bmatrix}Y$$

# 第11次作业:

- 孙 p.117
  习题四
- 8. 9. 10. 11.

# 谢 谢！