§6.2 一元线性回归中的检验与预测

1. 一元线性回归中的假设检验

要检验一元线性回归模型

$$\begin{cases} y_i = a + bx_i + \varepsilon_i, \ i = 1, 2, \cdots, n, \\ \varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n \overset{i.i.d.}{\sim} N(0, \sigma^2) \end{cases}$$

是否成立是一件比较复杂的事情.

检验一元线性回归模型“三步曲”

1.检验x取定时，Y是否服从同方差的正态分布.

2.检验x取定时，Y是不是x的线性函数.

3.检验x取定时，相应Y值是否相互独立.

可见要严格地检验线性回归这一假设，其计算麻烦且冗长.

1. 回归系数假设检验

上节求得的线性回归方程是否具有实用价值,需要经过假设检验才能确定,若假设符合实际,则 $b \neq 0$, 否则 y不依赖于 x.

因此需要检验假设 $H_0: b=0;\ H_1: b\neq 0.$

(1) t检验

由§6. 1知 $\hat{b} \sim N(b, \sigma^2/S_{xx})$

$$\Rightarrow U=\frac{\hat{b}-b}{\sigma/\sqrt{S_{xx}}} \sim N(0,1),\ \text{其中}\ S_{xx}=\sum_{i=1}^{n}(x_i-\bar{x})^2.$$

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

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

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

给定显著性水平α，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1 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

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

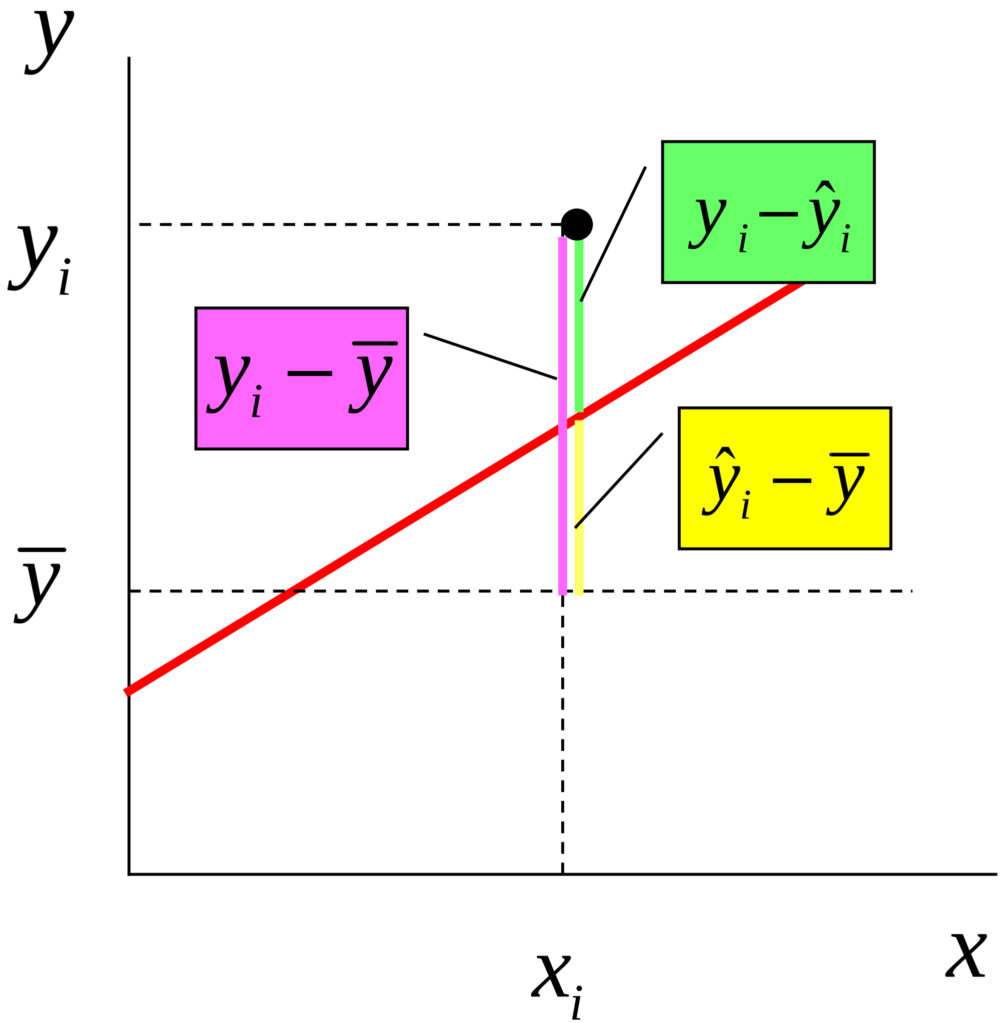
推广

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.

2) 相关系数法



2) 相关系数法

总平方和 $S_T=\sum_{i=1}^{n}(y_i-\bar{y})^2=\sum_{i=1}^{n}y_i^2-\frac{1}{n}(\sum_{i=1}^{n}y_i)^2=S_{yy}$

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

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

$\left(\begin{array}{l}\hat{y}_i=\bar{y}+\hat{b}(x_i-\bar{x}),\\ \text{于是交叉项为}0\end{array}\right)$

$$=\underbrace{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}_{S_E}+\underbrace{\sum_{i=1}^{n}(\hat{y}_i-\bar{\hat{y}})^2}_{S_R}$$

$\left(\begin{array}{l}\bar{\hat{y}}=\hat{a}+\hat{b}\bar{x}\\ \quad=\bar{y}-\hat{b}\bar{x}+\hat{b}\bar{x}=\bar{y}\end{array}\right)$

平方和分解式 $S_T = S_R + S_E$，

回 归平方和 $S_R = \sum_{i=1}^{n} (\hat{y}_i - \bar{\hat{y}})^2 = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$

$$= \sum_{i=1}^{n} (\hat{a} + \hat{b} x_i - \bar{y})^2$$

$$= \hat{b} S_{xy} = \hat{b}^2 S_{xx}$$

回归值的离散程度取决于$\{x_i\}$的离散程度以及经验回归函数的斜率的平方

平方和分解式 $S_T = S_R + S_E$，

残　差平方和　$$S_E = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}(y_i - \hat{a} - \hat{b}x_i)^2$$

S_R越大（S_E越小）：

　观测值 y_i 与经验回归值 $\hat{y}_i$ 越接近，即拟合效果越好，线性回归效果越显著

相关系数 r　$$r^2 = \frac{S_R}{S_T} = \frac{\hat{b}^2 S_{xx}}{S_{yy}} = \frac{S_{xy}^2}{S_{xx}S_{yy}}$$

3) F检验

平方和分解式 $S_T = S_R + S_E$，$f_T = f_R + f_E$，

总平方和 $S_T = \sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - \frac{1}{n}(\sum_{i=1}^{n} y_i)^2 = S_{yy}$

自由度 $f_T = n-1;$

回　归
平方和 $S_R = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$

自由度 $f_R = 1;$