

Ch-09 回归分析

9.1 引言

自变量 x 与因变量 y 的关系 $y = f(x) + e$ 称为回归关系，其中 e 为误差项，是一个随机变量。

9.2 一元线性回归

定义 2.1 设 $(x_i, y_i) (i = 1, \dots, n)$ 为一组数据。用线性函数 $y = a + bx$ 拟合，系数 \hat{a} 和 \hat{b} 称为最小二乘拟合系数，若它们满足

$$\sum_{i=1}^n [y_i - (\hat{a} + \hat{b}x_i)]^2 = \min_{a,b} \sum_{i=1}^n [y_i - (a + bx_i)]^2$$

即 \hat{a}, \hat{b} 是使拟合误差的平方和 $\sum_{i=1}^n [y_i - (a + bx_i)]^2$ 达到最小的 a 和 b 。

定理 2.1 设 $(x_i, y_i) (i = 1, \dots, n)$ 为回归方程 $y = a + bx + e$ 的一组观察值，只要 $x_1, \dots, x_n (n \geq 2)$ 不完全相同，则未知参数 a, b 的最小二乘估计由下式给出

$$\begin{aligned}\hat{a} &= \bar{y} - \hat{b}\bar{x} \\ \hat{b} &= \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{l_{xy}}{l_{xx}}\end{aligned}$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$