参数估计——最小二乘法

许多经济管理问题，常常需要研究某一现象与影响它的某一最主要因素之间的关系．例如在研究粮食产量时，在众多影响粮食产量的因素中施肥量是一个最重要的因素，需要研究粮食产量与施肥量之间的关系；在消费问题的研究中，由于国民收入是影响消费的最主要因素，需要研究消费额与国民收入之间的关系等．为了找出这类问题中两个变量之间的关系，往往根据两个变量的几组观测值或实验数据，找出这两个变量的近似表达式，一般称这样的表达式为经验公式．而一旦建立了经验公式，我们就可以将理论应用于实践，利用经验公式对自变量或因变量进行控制或预测．

下面通过具体的实例，介绍确定经验公式中未知参数的常用方法—最小二乘法.

例8 某工厂生产某种商品，为了了解这种商品的收益情况，收集了这种商品在市场上的销售量 x（单位：百件）与获得的利润额 y（单位：百元）的几组具体数据，如下表所示：

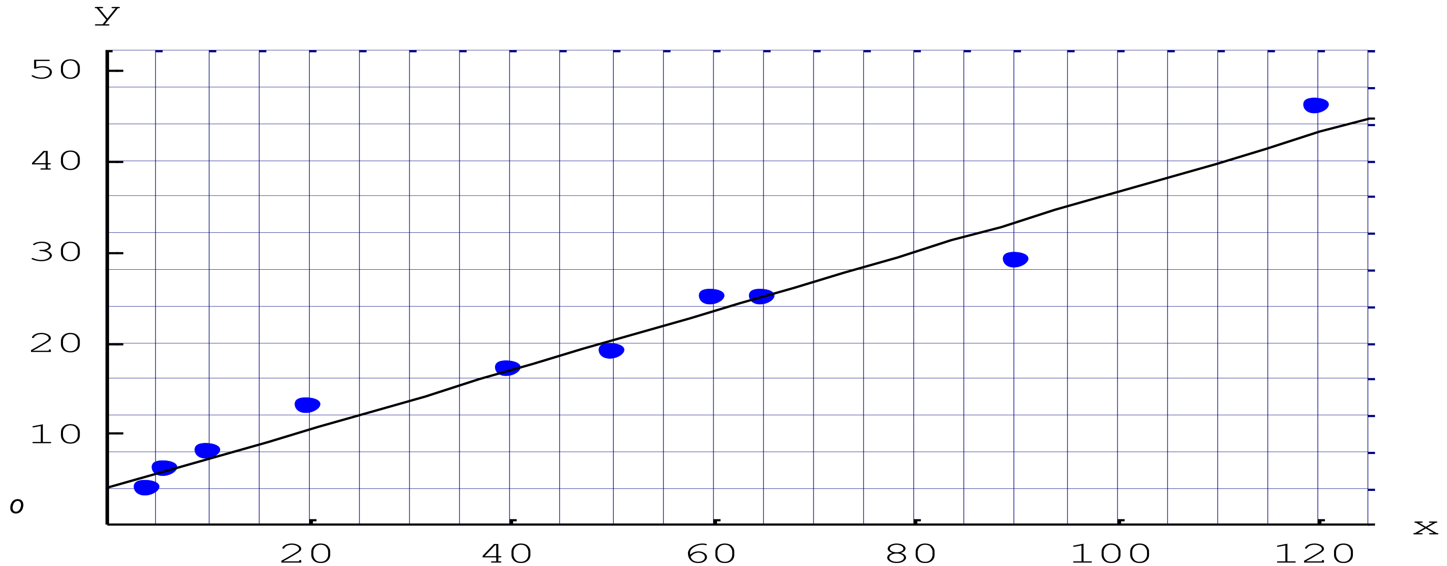
顺序编号 i	1	2	3	4	5	6	7	8	9	10
销售量 x_i	4	6	10	20	40	50	60	65	90	120
利润额 y_i	4	6	8	13	17	19	25	25	29	46

根据上表的数据建立 y 与 x 之间的经验公式 $y = f(x)$，

即要找出一个能使上述数据大体适合的函数关系 $y = f(x)$.

解 第1步 确定经验公式 $y=f(x)$ 的函数类型.

为此，在平面直角坐标系中画出这些实际数据的散点图，从图中直观分析这些点的分布规律. 如果这些点的连线大体在一条直线上，那么经验公式就用线性函数表示，否则就选取更为合适的其他非线性函数描述.



设 $f(x)=ax+b$，其中 a、b 为待定参数.

第2步估计经验公式$f(x)=ax+b$中的参数a、b，写出经验公式.

我们选取这样的a、b，使$f(x)=ax+b$在$x_i(i=1,2,\cdots,10)$处每一个函数值$f(x_i)(i=1,2,\cdots,10)$与$y_i(i=1,2,\cdots,10)$相差很小，显然，如果

$$\sum_{i=1}^{10}\left|y_i-f(x_i)\right|=\sum_{i=1}^{10}\left|y_i-(ax_i+b)\right|$$

很小，就能保证每个偏差$y_i-f(x_i)$的绝对值都很小.

在不影响问题实质的情况下，将绝对值改为平方，问题转化为：

选取适当的a、b，使$S(a,b)=\sum_{i=1}^{10}\left(y_i-(ax_i+b)\right)^2$最小，

这样就能保证每个偏差 $y_i - f(x_i)$ 的绝对值都很小.

我们把这种根据偏差的平方和最小的条件来选择经验公式 $f(x)=ax+b$ 中未知参数 a、b 的方法称为最小二乘法.

下面利用多元函数极值理论求未知参数 a、b.

为此，求 $S(a,b)$ 关于 a、b 的偏导数，得方程组：

$$\begin{cases} \dfrac{\partial S}{\partial a} = -2\sum_{i=1}^{10}[y_i-(ax_i+b)]x_i = 0 \\ \dfrac{\partial S}{\partial b} = -2\sum_{i=1}^{10}[y_i-(ax_i+b)] = 0 \end{cases}$$

记此方程组的解为 $\hat{a}$、$\hat{b}$，则得 $\hat{a} = \dfrac{\sum_{i=1}^{10} x_i y_i - 10\overline{xy}}{\sum_{i=1}^{10} x_i^2 - 10\overline{x}^2}$，$\hat{b} = \overline{y} - \hat{a}\overline{x}$，

其中 $\bar{x}=\frac{1}{10}\sum_{i=1}^{10}x_i$，$\bar{y}=\frac{1}{10}\sum_{i=1}^{10}y_i$

故变量 y 与 x 的经验公式为 $y=\hat{a}x+\hat{b}$.

本例中，代入相关数据利用mathematica软件（具体计算方法，见第8.6节技能训练）得 $\hat{a}=0.3257$、$\hat{b}=4.0569$.

因此利润额y与销售量x之间的经验公式为

$$y=0.325\ 7x+4.056\ 9$$

从而求得的经验公式近似表达原来的函数关系与实际的符合程度大小，从而求出函数值$f(x_i)$与利润额的实际值y_i有一定的偏差.

进一步计算可得上述偏差的平方和为

$$S(0.3257, 4.0569) = 39.3133$$

其中$\sqrt{S} = 6.2700$．我们称$\sqrt{S}$为均方误差，它在一定程度上反映了用所求得的经验公式近似表达原来的函数关系与实际的符合程度大小.