$$\sigma = \sqrt{\frac{Q}{N-2}} \tag{6-33}$$

称为残余标准差，与 σ^2 的意义相似，它可以用来衡量所有随机因素对 y 的一次性观测的平均变差的大小，σ 越小，回归直线的精度越高。当回归方程的稳定性较好时，σ 可作为应用回归方程时的精度参数。

（四）方差分析表

上述把平方和及自由度进行分解的方差分析所有结果可归纳在一个简单的表格中，这种表称为方差分析表，见表 6-3。

表 6-3

来　源	平方和	自由度	方　差	F	显著性
回　归	$U=bl_{xy}$	1		$F=\dfrac{U/1}{Q/(N-2)}$	—
残　余	$Q=l_{yy}-bl_{xy}$	$N-2$	$\sigma^2=Q/(N-2)$		
总　计	$S=l_{yy}$	$N-1$	—	—	—

例 6-2 在例 6-1 电阻对温度的回归中，由表 6-2 及表 6-3 可得表 6-4 的方差分析结果。

表 6-4

来　源	平方和/Ω^2	自由度	方差/Ω^2	F	显著性
回　归	60.574	1	—	1.18×10^3	$\alpha=0.01$
残　余	0.257	5	0.0514	—	—
总　计	60.831	6	—	—	—

显著性一栏中的 $\alpha=0.01$，表明前面所得的回归方程式（6-15）在 $\alpha=0.01$ 水平上显著，即可信赖程度为 99% 以上，这是高度显著的。

利用回归方程，可以在一定显著性水平 α 上，确定与 x 相对应的 y 的取值范围。反之，若要求观测值 y 在一定的范围内取值，利用回归方程可以确定自变量 x 的控制范围。

三、重复实验情况

应该指出，用残余平方和检验回归平方和所作出的“回归方程显著”这一判断，只表明相对于其他因素及试验误差来说，因素 x 的一次项对指标 y 的影响是主要的，但它并没有告诉我们：影响 y 的除 x 外，是否还有一个或几个不可忽略的其他因素，以及 x 和 y 的关系是否确实为线性。换言之，在上述意义下的回归方程显著，并不一定表明这个回归方程是拟合得很好的。其原因是由于残余平方和中除包括实验误差外，还包括了 x 和 y 线性关系以外的其他未加控制的因素的影响。为了检验一个回归方程拟合得好坏，可以做些重复实验，从而获得误差平方和 Q_E 和失拟平方和 Q_L（它反映了非线性及其他未加控制的因素的影响），用误差平方和对失拟平方和进行 F 检验，就可以确定回归方程拟合得好坏。

设取 N 个实验点，每个实验点都重复 m 次实验，此时各种平方和及其相应的自由度可按下列各式计算：

$$S = U + Q_L + Q_E, \quad \nu_S = \nu_U + \nu_L + \nu_E \tag{6-34}$$

$$S = \sum_{t=1}^{N}\sum_{i=1}^{m}(y_{ti} - \bar{y})^2, \quad \nu_S = Nm - 1 \tag{6-35}$$

$$U = m\sum_{t=1}^{N}(\hat{y}_t - \bar{y})^2, \quad \nu_U = 1 \tag{6-36}$$

$$Q_E = \sum_{t=1}^{N}\sum_{i=1}^{m}(y_{ti} - \bar{y}_t)^2, \quad \nu_{Q_E} = N(m-1) \tag{6-37}$$

$$Q_L = m\sum_{t=1}^{N}(\bar{y}_t - \hat{y}_t)^2, \quad \nu_{Q_L} = N - 2 \tag{6-38}$$

例 6-3 用标准压力计对某固体压力传感器进行检定，检定所得数据见表 6-5。表中 x_t 为标准压力，y_{ti}为传感器输出电压，$\bar{y}_t$ 为 4 次读数的算术平均值。试对仪器定标，并分析仪器的误差。

仪器要求线性定标，故应拟合一条回归直线。可以证明，用平均值的 11 个点拟合的回归直线与用原来的 44 个点拟合的回归直线完全一样。具体计算见表 6-6 和表 6-7（表中 y_t 即为表 6-5 中的 $\bar{y}_t$）。

表 6-5

序号	x_t/N·cm^{-2}	y_{ti}/mV				$\bar{y}_t$/mV
		升压	降压	升压	降压	
1	0	2.78	2.80	2.80	2.86	2.81
2	1	9.70	9.76	9.78	9.78	9.755
3	2	16.60	16.71	16.70	16.76	16.6925
4	3	23.54	23.56	23.58	23.71	23.5975
5	4	30.44	30.51	30.54	30.64	30.5325
6	5	37.35	37.45	37.42	37.50	37.43
7	6	44.28	44.35	44.30	44.38	44.3275
8	7	51.19	51.25	51.18	51.25	51.2175
9	8	58.06	58.08	58.12	58.14	58.1
10	9	64.92	64.96	64.94	65.00	64.955
11	10	71.73	71.73	71.75	71.75	71.74

表 6-6

序号	x_t/N·cm^{-2}	y_t/mV	x_t^2/(N·cm^{-2})^2	y_t^2/(mV)^2	$x_t y_t$/mV·N·cm^{-2}
1	0	2.8100	0	7.8961	0
2	1	9.7550	1	95.1600	9.7500
3	2	16.6925	4	278.6396	33.3850
4	3	23.5975	9	556.8420	70.7925

（续）

序　号	x_t/N · cm^{-2}	y_t/mV	x_t^2/(N · cm^{-2})^2	y_t^2/(mV)^2	$x_t y_t$/mV · N · cm^{-2}
5	4	30.5325	16	932.2336	122.1300
6	5	37.4300	25	1401.0049	187.1500
7	6	44.3275	36	1964.9273	265.9650
8	7	51.2175	49	2623.2323	358.5225
9	8	58.1000	64	3375.6100	464.8000
10	9	64.9550	81	4219.1520	584.5950
11	10	71.7400	100	5146.6276	717.4000
Σ	55	411.1575	385	26601.3254	2814.4950

表 6-7

$\sum_{t=1}^{N} x_t = 55\text{N/cm}^2$	$\sum_{t=1}^{N} y_t = 411.1575\text{mV}$	$N = 11$
$\overline{x} = 5\text{N/cm}^2$	$\overline{y} = 37.3780\text{mV}$	
$\sum_{t=1}^{N} x_t^2 = 385(\text{N/cm}^2)^2$	$\sum_{t=1}^{N} y_t^2 = 20601.3254(\text{mV})^2$	$\sum_{t=1}^{N} x_t y_t = 2814.4950\text{mV} \cdot \text{N/cm}^2$
$\left(\sum_{t=1}^{N} x_t\right)^2 \Big/ N = 275(\text{N/cm}^2)^2$	$\left(\sum_{t=1}^{N} y_t\right)^2 \Big/ N = 15368.226(\text{mV})^2$	$\left(\sum_{t=1}^{N} x_t\right)\left(\sum_{t=1}^{N} y_t\right) \Big/ N = 2055.7875\text{mV} \cdot \text{N/cm}^2$
$l_{xx} = \sum_{t=1}^{N} x_t^2 - \left(\sum_{t=1}^{N} x_t\right)^2 \Big/ N$ $= 110(\text{N/cm}^2)^2$	$l_{yy} = \sum_{t=1}^{N} y_t^2 - \left(\sum_{t=1}^{N} y_t\right)^2 \Big/ N$ $= 5233.0990(\text{mV})^2$ $b = \frac{l_{xy}}{l_{xx}} = 6.89734\text{mV}/(\text{N} \cdot \text{cm}^{-2})$ $b_0 = \bar{y} - b\bar{x} = 2.8913\text{mV}$ $\hat{y} = 2.8913\text{mV} + [6.89734\text{mV}/(\text{N} \cdot \text{cm}^{-2})]x$	$l_{xy} = \sum_{t=1}^{N} x_t y_t - \left(\sum_{t=1}^{N} x_t\right)\left(\sum_{t=1}^{N} y_t\right) \Big/ N$ $= 758.7075\text{mV} \cdot \text{N/cm}^2$

现在进行方差分析。当用 $\bar{y}_t$ 求回归直线时，各平方和可按下式顺序计算：

$$\left.\begin{aligned} U &= mbl_{xy} \\ Q_L &= ml_{yy} - U \\ Q_E &= \sum_{t=1}^{N} \sum_{i=1}^{m} (y_{ti} - \bar{y}_t)^2 \\ S &= U + Q_L + Q_E \end{aligned}\right\} \tag{6-39}$$

计算结果见表 6-8。

表中 F 栏为用误差平方和对相应的平方和项进行 F 检验的数学统计量，其中

表 6-8

来　源	平方和/mV^2	自由度	方差/mV^2	F	显著性
回　归	20932.2574	1	20932.2574	7.68×10^6	$F_{0.01}=7.47$
失　拟	0.1386	9	0.0154	5.65	$F_{0.01}=3.03$
误　差	0.0899	33	0.0027	—	—
总　计	20932.4859	43	—	—	—

第一行

$$F=\frac{U/\nu_U}{Q_E/\nu_{Q_E}}=7.68\times10^6 \tag{6-40}$$

$$F>F_{0.01}(1,33)=7.47$$

第二行

$$F_1=\frac{Q_L/\nu_{Q_L}}{Q_E/\nu_{Q_E}}=5.65 \tag{6-41}$$

$$F_1>F_{0.01}(9,33)=3.03$$

由表6-8或式（6-41）可知，对失拟平方和进行 F 检验结果高度显著，说明失拟误差相对于实验误差来说是不可忽略的，这时有如下几种可能：

1）影响 y 的除 x 外，至少还有一个不可忽略的因素。

2）y 和 x 是曲线关系。

3）y 和 x 无关。

总之，所选择的一元线性回归这个数学模型与实际情况不符合，说明该直线拟合得并不好。失拟平方和 Q_L 或失拟方差反映了拟合误差，通常称为模型误差。

如果 F_1 检验结果不显著，说明非线性误差（相对于实验误差讲）很小，或者基本上是由实验误差等随机因素引起的。于是可把失拟平方和 Q_L 与误差平方和 Q_E 合并，对回归平方和进行 F 检验，即

$$F_2=\frac{U/\nu_U}{(Q_L+Q_E)/(\nu_{Q_L}+\nu_{Q_E})} \tag{6-42}$$

如果第二次 F 检验结果显著，说明一元回归方程拟合得好。

对于给定的显著性水平 α，如果 F_2 不显著，那么这时有两种可能：

1）没有什么因素对 y 有系统的影响。

2）实验误差过大。

当然所求的回归方程不理想。

现在继续对例6-3作进一步分析。

F_1 检验结果显著，回归方程是否就没有用了呢？不妨再用 Q_E 对 U 进行第二次 F 检验

$$F_2=\frac{U/\nu_U}{Q_E/\nu_{Q_E}}=7.68\times10^6\gg F_{0.01}(1,33)=7.47$$

结果高度显著。再用 $Q_E+Q_L=Q$ 对 U 进行第二次 F 检验

$$F_2=\frac{U/\nu_U}{Q/\nu_Q}=3.85\times10^6\gg F_{0.01}(1,42)=7.28$$

也高度显著。由于 F_1 检验结果显著，虽然相对于实验误差来讲，此方程不能说拟合得很好，但是由于两种 F_2 检验都高度显著，实验误差和残余误差都很小，只要残余标准差 σ 小于该仪器所要求的精度参数，就可以使用此方程对该仪器进行定标。当然，如有必要，可进一步查明原因，重做回归方程。

从以上分析可以看出，在一般情况下，重复实验可将误差平方和与失拟平方和从残余平方和中分离出来，这对统计分析是有好处的。同时，在精密测试仪器中，通常失拟平方和及误差平方和分别与仪器的原理误差（定标误差、非线性误差）及仪器的随机误差相对应。应用这种方法分析传感器或非电量电测仪器及其他类似需要变换参量的测量仪器的精度，可以将系统误差与随机误差分离开来，并可用回归分析方法进一步找出仪器的误差方程，从而可以对仪器的误差进行修正。不需要对仪器作任何改进，只是通过数据处理，对仪器的系统误差进行修正，就可使仪器的精度明显提高，这是提高仪器精度的一种颇为有效的方法。总之，通过重复实验的回归分析对了解这类仪器的误差来源和提高仪器的精度是有益的。如果没有条件做重复实验，只能用残余平方和对回归平方和按式（6-30）进行 F 检验，也可大致说明回归效果的好坏。习惯上，经常也把这种检验结果显著与不显著说成拟合得好与坏。但需要注意，一个方程拟合得好的真正含义应该是失拟平方和相对于误差平方和来讲是不显著的。

四、回归直线的简便求法

回归分析是以最小二乘法为基础的，因此所建立的回归直线方程误差（标准差）最小，但它的计算一般是比较复杂的。为了减少计算，在精度要求不太高或实验数据线性较好的情况下，可采用如下简便方法：

（一）分组法（平均值法）

用分组法求回归方程 $\hat{y}=b_0+bx$ 中的系数 b_0 和 b 的具体作法是：将自变量数据按由小到大的次序安排，分成个数相等或近于相等的两个组（分组数等于欲求的未知数个数）：第一组为 x_1，x_2，…，x_k；第二组为 x_{k+1}，…，x_N，建立相应的两组观测方程

$$\left.\begin{aligned} y_1 &= b_0 + bx_1 \\ &\ \vdots \\ y_k &= b_0 + bx_k \end{aligned}\right\} \qquad \left.\begin{aligned} y_{k+1} &= b_0 + bx_{k+1} \\ &\ \vdots \\ y_N &= b_0 + bx_N \end{aligned}\right\}$$

两组观测方程分别相加，得到关于 b_0 及 b 的方程组

$$\left.\begin{aligned} \sum_{t=1}^{k} y_t &= kb_0 - b\sum_{t=1}^{k} x_t \\ \sum_{t=k+1}^{N} y_t &= (N-k)b_0 + b\sum_{t=k+1}^{N} x_t \end{aligned}\right\} \tag{6-43}$$

该方程组可解得 b 及 b_0。特别当 $N=2k$ 时，回归系数

$$\left.\begin{aligned} b &= \frac{\sum_{t=1}^{N/2} y_t - \sum_{t=N/2+1}^{N} y_t}{\sum_{t=1}^{N/2} x_t - \sum_{t=N/2+1}^{N} x_t} \\ b_0 &= \frac{\sum_{t=1}^{N} y_t}{N} - \frac{\sum_{t=1}^{N} x_t}{N} = \bar{y} - b\bar{x} \end{aligned}\right\} \tag{6-44}$$

例 6-4 对例 6-1 用分组法求回归方程。

因观测数据已按自变量从小到大的次序排列，故可按实验顺序分成两组，并建立两组相应的观测方程(取 $k=(N+1)/2=4$)，然后分别相加：

$$
\begin{array}{rcl}
76.30\Omega &=& b_0+(19.1^\circ\mathrm{C})b \\
77.80\Omega &=& b_0+(25.0^\circ\mathrm{C})b \\
79.75\Omega &=& b_0+(30.1^\circ\mathrm{C})b \\
80.80\Omega &=& b_0+(36.0^\circ\mathrm{C})b \\
\hline
314.65\Omega &=& 4b_0+(110.2^\circ\mathrm{C})b
\end{array}
\qquad
\begin{array}{rcl}
82.35\Omega &=& b_0+(40.0^\circ\mathrm{C})b \\
83.90\Omega &=& b_0+(46.5^\circ\mathrm{C})b \\
85.10\Omega &=& b_0+(50.0^\circ\mathrm{C})b \\
\hline
251.35\Omega &=& 3b_0+(136.5^\circ\mathrm{C})b
\end{array}
$$

解方程组

$$
\left.\begin{array}{l}
314.65\Omega=4b_0+(110.2^\circ\mathrm{C})b \\
251.35\Omega=3b_0+(136.5^\circ\mathrm{C})b
\end{array}\right\}
$$

得

$$b_0=70.80\Omega$$

$$b=0.2853\Omega/^\circ\mathrm{C}$$

故所求的回归方程为

$$\hat{y}=70.80\Omega+(0.2853\Omega/^\circ\mathrm{C})x$$

这与用最小二乘法求得的回归方程式（6-15）比较接近。

此法简单明了，拟合的直线就是通过第一组重心和第二组重心的一条直线，这是工程实践中常用的一种简单方法。

（二）图解法（紧绳法）

把 N 对观测数据画出散点图于坐标纸上，假如画出的点群形成一直线带，就在点群中画一条直线，使得多数点位于直线上或接近此线，并均匀地分布在直线的两边。这条直线可以近似地作为回归直线，回归系数可以直接由图中求得。利用此直线也可在坐标纸上直接进行预报。

例 6-5 用 X 射线机检查镁合金焊接件及铸件内部缺陷时，为达到最佳灵敏度，透照电压 y 应随被透照件厚度 x 而改变。经实验得如下一组数据：

x/mm	12	13	15	16	18	20	22	24	25	26
y/kV	53	55	60	65	70	75	80	85	88	90

把这组数据点在坐标纸上，然后通过点群作一直线（见图 6-3）。在接近直线的两端各取一点，如（12，53）和（25，88），回归系数 $b=(53-88)\mathrm{kV}/(12-25)\mathrm{mm}=2.7\mathrm{kV/mm}$。将回归直线上任一点代入回归方程可求出 b_0，如 $53\mathrm{kV}=b_0+(2.7\mathrm{kV/mm})\times 12\mathrm{mm}$，$b_0=20.6\mathrm{kV}$，故所求回归方程为

$$\hat{y}=20.6\mathrm{kV}+(2.7\mathrm{kV/mm})x$$

图解法由于作图时完全凭经验画直线，故主观性较大，精度较低，但此法非常简单，精度要求不高时可采用。

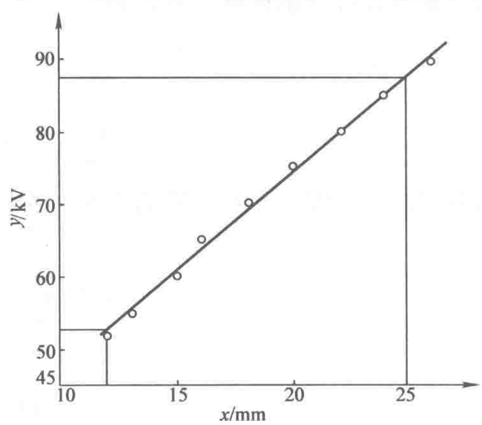

图 6-3

第三节 两个变量都具有误差时线性回归方程的确定

一、概述

上面用最小二乘法求得的回归方程一般认为是最佳的，但它是假设 x 没有误差或误差是可以忽略的，其所有误差都归结在 y 方向。然而，x 的测量也可能是不精确的，存在实验误差。现在考察另一种极端情况，即 y 没有误差，而所有误差都归结于 x。在这种情况下，一元线性回归方程的数学模型是

$$x_t = \alpha_0 + \alpha y_t + \varepsilon_t', \quad t = 1,2,\cdots, N \tag{6-45}$$

式中，α_0、α 为特定参数；ε_t'为误差项。

此时应该求 x 对 y 的回归方程

$$\hat{x} = a_0 + ay \tag{6-46}$$

式中，a_0、a 为 α_0、α 的最小二乘估计。

应用最小二乘原理，使 $\sum\limits_{t=1}^{N}(x_t - \hat{x}_t)^2$ 为最小，求得

$$a = \frac{l_{xy}}{l_{yy}} \tag{6-47}$$

$$a_0 = \bar{x} - a\bar{y} \tag{6-48}$$

式中，$\bar{x}$、$\bar{y}$、l_{xy}、l_{yy}分别由式（6-9）、式（6-10）、式（6-12）、式（6-13）确定。

为便于与式（6-2）比较，将式（6-46）改写为

$$y = b_0' + b'\hat{x} \tag{6-49}$$

式中

$$b' = \frac{1}{a} = \frac{l_{yy}}{l_{xy}} \tag{6-50}$$

$$b_0' = \bar{y} - b'\bar{x} \tag{6-51}$$

一般情况下，$b_0' \neq b_0$，$b' \neq b$。

例 6-6 若例 6-1 中电阻值 y 是可以精确测定的，而把所有误差都归结于温度 x，则根据 $\sum\limits_{t=1}^{N}(x_t - \hat{x}_t)^2$ 为最小的原则，由式（6-50）和式（6-51）得另一回归方程

$$y = 70.86\Omega + 0.2836\hat{x} \tag{6-52}$$

将其与式（6-15）比较，两个方程的回归系数是不一样的。

在这里我们发现，通过一批点能够作两条最佳直线，原因是，假定有一批试验点是均匀分布在一个椭圆域内。如果假定误差是在 y 方向，我们作一系列细的垂直线，找出每一条细的垂线的中点，并作一条线通过各点（见图 6-4）。如果假定误差是在 x 方向，则作一系列水平线，找出各水平线的中点，并作一条

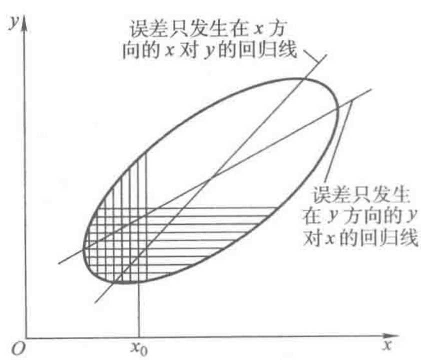

图 6-4

线通过各点。用这个方法得到两条直线，一条通过椭圆水平方向极值点，另一条通过垂直方向极值点。前者就是 y 对 x 的回归直线，回归值 $\hat{y}$ 表示当任取 $x=x_0$ 时 y 的平均数；后者即是 x 对 y 的回归直线，其回归值 $\hat{x}$ 表示当任取 $y=y_0$ 时 y 的平均数。

两个最小二乘解的存在，提出了在实验中需要判断的问题，并可能有下列三种情况：

1）如果两变量中一个变量的误差可以忽略，则应采用另一变量对该变量的回归线。

2）如果两变量的误差大体相当，则采用图 6-4 所示两条回归线的平均线。

3）如果两个变量中的一个变量误差比另一个大，则所采用的中间线应偏向于误差大的变量对另一变量的回归线。

在实践中，以上三种情况都是可能存在的，求回归直线时应加以区别。随着两个变量之间线性关系的加强，即相关系数越接近于 1，两条最小二乘直线越接近。当相关系数为 1 时，这两条直线将重合。本例中两个回归方程差别不大，是线性相关强之故。

二、回归方程的求法

两个变量都具有误差时，比较精确的计算回归系数的方法是戴明（Deming）解法。

若 x_t、y_t 分别具有误差 $\delta_t \sim N(0, \sigma_x)$，$\varepsilon_t \sim N(0, \sigma_y)$，$t=1, 2, \cdots, N$，假定 x、y 之间为线性关系，则其数学模型为

$$y_t=\beta_0+\beta(x_t-\delta_t)+\varepsilon_t, \quad t=1,2,\cdots,N \tag{6-53}$$

所求的回归方程为

$$\hat{y} = b_0 + b\hat{x} \tag{6-54}$$

式中，$\hat{x}$、$\hat{y}$、b_0、b 分别为 x、y、β_0、β 的估计值。

为使 x、y 的误差在求回归方程时具有等价性，令 $\sigma_x^2/\sigma_y^2=\lambda$，$y'=\sqrt{\lambda}y$，则式（6-54）可写成

$$\hat{y}' = b_0' + b'\hat{x} \tag{6-55}$$

式中，$\hat{y}'=\sqrt{\lambda}\hat{y}$；$b_0'=\sqrt{\lambda}b_0$；$b'=\sqrt{\lambda}b$。

根据戴明推广的最小二乘原理，点（x_t，y_t'）到回归直线的垂直距离 d_t'（见图 6-5）的平方和 $\sum_{t=1}^{N} d_t'^2$ 为最小条件下所求得的回归系数 b_0、b 是最佳估计值。

由解析几何可知，点（x_t，y_t'）到回归直线式（6-55）的距离 d_t'为

$$d_t' = \frac{y_t'-b_0'-b'x_t}{\sqrt{1+b'^2}} = \frac{\sqrt{\lambda}}{\sqrt{1+\lambda b^2}}(y_t-b_0-bx_t) = \frac{\sqrt{\lambda}}{\sqrt{1+\lambda b^2}}d_t \tag{6-56}$$

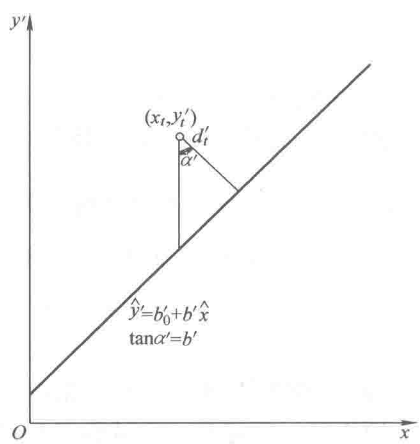

图 6-5

式中

$$d_t = y_t - b_0 - bx_t \tag{6-57}$$

根据最小二乘原理，为使 $\sum_{t=1}^{N} d_t'^2$ 为最小，即求解

$$\left.\begin{aligned}\frac{\partial\left(\sum_{t=1}^{N} d_t'^2\right)}{\partial b_0} = 0 \\ \frac{\partial\left(\sum_{t=1}^{N} d_t'^2\right)}{\partial b} = 0\end{aligned}\right\} \tag{6-58}$$

得

$$\left.\begin{aligned} b &= \frac{\lambda l_{yy} - l_{xx} + \sqrt{(\lambda l_{yy} - l_{xx})^2 + 4\lambda l_{xy}^2}}{2\lambda l_{xy}} \\ b_0 &= \bar{y} - b\bar{x}\end{aligned}\right\} \tag{6-59}$$

式中，$\bar{x}$、$\bar{y}$、l_{xx}、l_{xy}、l_{yy}含义同前［见式（6-9）~式（6-13）］。

变量 x、y 的方差可用下式估计：

$$\left.\begin{aligned}\sigma_x^2 &= \frac{1}{N-2} \frac{\lambda}{1+\lambda b^2} \sum_{t=1}^{N} d_t^2 \\ \sigma_y^2 &= \frac{1}{N-2} \frac{1}{1+\lambda b^2} \sum_{t=1}^{N} d_t^2 = \frac{\sigma_x^2}{\lambda}\end{aligned}\right\} \tag{6-60}$$

$\sum_{t=1}^{N} d_t^2$ 可以通过式（6-57）来计算，或者利用计算回归系数时的中间结果，即

$$\sum_{t=1}^{N} d_t^2 = l_{yy} - 2bl_{xy} + b^2 l_{xx} \tag{6-61}$$

下面讨论两种特殊情况：

1）当 x 无误差时，$\lambda=0$，$b=l_{xy}/l_{xx}$，$\sigma_y^2=(l_{yy}-bl_{xy})/(N-2)$，这就是我们在第二节中所讨论的一般回归问题的情况。

2）当 y 无误差时，$\lambda=\infty$，$b=l_{yy}/l_{xy}=1/a$，$\sigma_x^2=(l_{xx}-l_{xy}/b)/(N-2)$，这就是本节概述中提到的另一种情况。

例 6-7 通过实验测量某量 x、y 的结果如下：

x	2.560	2.319	2.058	1.911	1.598	0.548
y	2.646	2.395	2.140	2.000	1.678	0.711

由重复测量已估计出 $\sigma_x=\sigma_y$，即 $\lambda=1$，试求 y 对 x 的回归直线方程。

根据有关公式，计算结果见表 6-9。所求回归方程为

$$\hat{y}=0.1350+0.9747\hat{x}$$

表 6-9

$\bar{x} = 1.8398$ $l_{xx} = 2.4495468$ $b = 0.9747$	$l_{yy} = 2.3273297$ $\sum_{i=1}^{N} d_i^2 = 3.1382 \times 10^{-4}$ $\sigma_x^2 = \sigma_y^2 = 4.023 \times 10^{-5}$ $\sigma_x = \sigma_y = 0.0063$	$\bar{y} = 1.9283$ $l_{xy} = 2.3874952$ $b_0 = 0.1350$

一般情况下，亦可用第二节、四中的分组法作近似计算。

第四节　一元非线性回归

在实际问题中，有时两个变量之间的内在关系并不是线性关系，而是某种曲线关系，这时若求所需的回归曲线，一般地说，可以分两步进行：

1）确定函数类型。

2）求解相关函数中的未知参数。用最小二乘法直接求解非线性回归方程是非常复杂的，通常是通过变量代换把回归曲线转换成回归直线，继而用前面给出的方法求解；或者把回归曲线展成回归多项式，直接用回归多项式来描述两个变量之间的关系，这样就把解曲线回归问题转化为解多项式回归的问题。

一、回归曲线函数类型的选取和检验

（一）直接判断法

根据专业知识，从理论上推导或者根据以往的经验，可以确定两个变量之间的函数类型，如化学反应物质的浓度 y 一般与时间 x 有指数关系，即 $y=y_0e^{kx}$，其中 y_0 及 k 为待定系数和指数。

（二）观察法

将观测数据作图，将其与典型曲线（见图6-6）比较，确定其属于何种类型。所选择的曲线类型是否合适，可用如下的方法来检验。

（三）直线检验法

当函数类型中所含参数不多，例如只有一个或两个时，用此法检验较好。其步骤如下：

1）将预选的回归曲线 $f(x, y, a, b)=0$ 写成

$$Z_1 = A + BZ_2 \tag{6-62}$$

式中的 Z_1 和 Z_2 是只含一个变量（x 或 y）的函数，A 和 B 是 a 和 b 的函数。

2）求出几对与 x、y 相对应的 Z_1 和 Z_2 的值，这几对值以选择与 x、y 值相距较远为好。

3）以 Z_1 和 Z_2 为变量画图，若所得图形为一直线，则证明原先所选定的回归曲线类型是合适的。

例6-8　用此法说明下列一组数据是否可用 $y=ae^{bx}$ 表示。

x	1	2	3	4	5	6	7	8	9
y	1.78	2.24	2.74	3.74	4.45	5.31	6.92	8.85	10.97
$\log y$	0.250	0.350	0.438	0.573	0.648	0.725	0.840	0.947	1.040