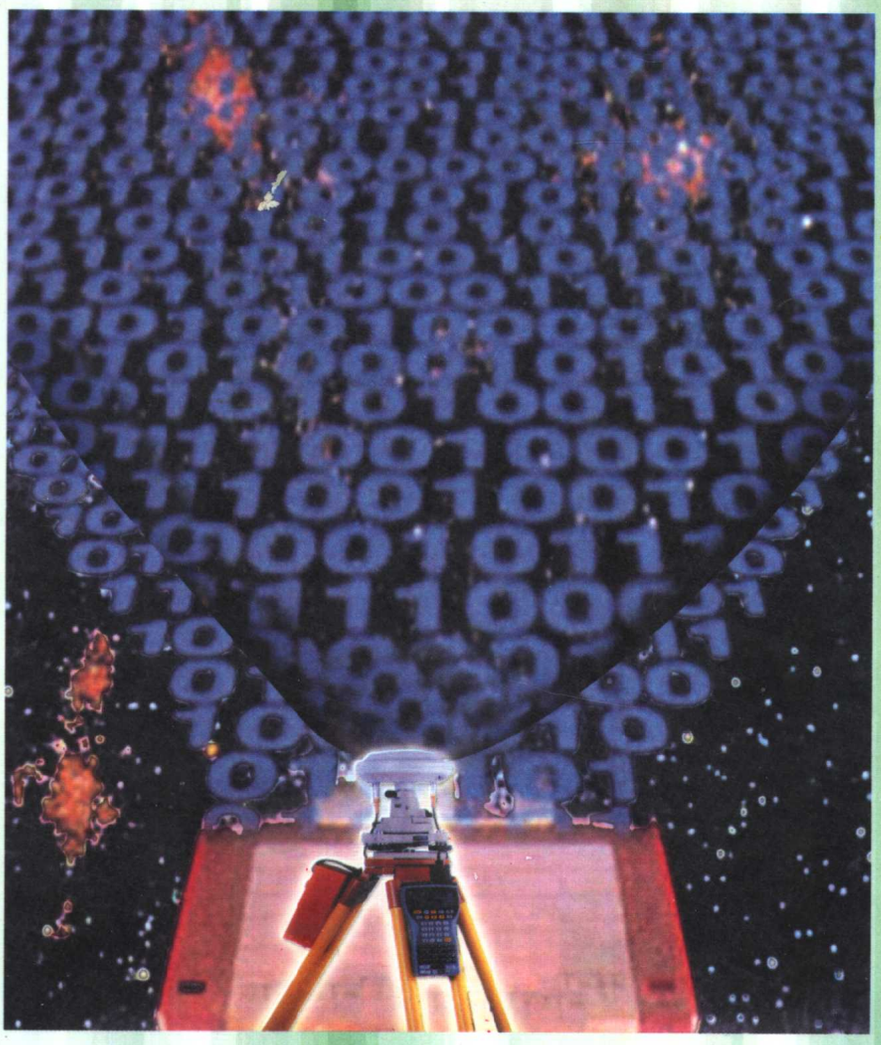
普通高等教育测绘类规划教材

实用测量数据处理方法

主编　刘大杰　陶本藻



测绘出版社

普通高等教育测绘类规划教材

实用测量数据处理方法

主编　刘大杰　陶本藻

测绘出版社

北京·2000

图书在版编目(CIP)数据

实用测量数据处理方法/刘大杰,陶本藻编. 北京:
测绘出版社,2000.3
ISBN 7-5030—0878-4

Ⅰ.实… Ⅱ.(1)刘…(2)陶… Ⅲ.测量-数据处理-
数学方法 Ⅳ.P20

中国版本图书馆CIP数据核字(2000)第17134号

测绘出版社出版发行
100054 北京宣武区白纸坊西街3号
北京市通州区次渠印刷厂印刷·新华书店总店北京发行所经销
2000年7月第一版·2000年7月第一次印刷
开本:890×1240 1/16·印张:10.125
字数:328千字·印数:0001—3000册
定价:26.00元

前　言

测量数据的处理方法，通常是指按最小二乘法进行测量平差。它是测量数据处理中最基本、最广泛应用的方法，因而这一学科得到了充分的发展。其丰富的内容已在大学测绘类专业教材中得到了充分的体现。

随着测绘科技的飞速发展，以及测绘学与其相关学科结合的需要，还要求除测量平差现有内容以外的其它数据处理方法，而且迫切需要列入测绘类专业的教学内容之中。为此，在1996年，中国测绘教育指导委员会在讨论编写21世纪重点测绘教材目录时，专家们建议在测量平差课程基础上，增开一门“测量数据处理方法”作为大学测绘类本科生的必修或选修课程，并建议由同济大学、武汉测绘科技大学牵头，组织几个大学的测量系共同编写此教材。

本书所介绍的测量数据处理方法，着眼于基础性和实用性，有利于拓宽知识面，对开拓新思路有启发性，故定名为“实用测量数据处理方法”。

本书共分七章，第一章回归分析的前五节由河海大学尹任祥教授编写；第二章插值与拟合和第一章的第六节由郑州测绘学院归庆明教授编写；第三章稳健估计由西安工程学院张勤副教授编写；第四章时间序列分析由同济大学刘大杰教授编写；第五章傅里叶分析由武汉测绘科技大学王新洲教授、陶本藻教授编写；第六章有限元方法由中南工业大学朱建军教授编写；第七章分布拟合检验由武汉测绘科技大学黄加纳教授编写。全书由陶本藻、刘大杰负责统稿。

我们希望本书的出版能够对测绘类专业大学生和测绘科技人员在拓宽测量数据处理知识面、促进测绘生产方面起到推动作用，书中不足和错误之处垦请读者批评指正。

最后要感谢中国测绘教育指导会为本书组织了审查，感谢测绘出版社为本书的出版所做的辛勤工作。

编　者

1998年9月1日

目　录

第一章　回归分析

§1.1　概　　述

在自然界中，人们常遇到一些相互依赖而又相互制约的变量。它们之间的关系不外乎两种类型：一类是确定的函数关系，例如密度均匀的物体，其质量 m 与体积 V 之间的关系可用 $m=\rho \cdot V$ 来表达（其中 ρ 是物体的密度）；另一类是非确定性的依赖关系，既相关关系，例如水库大坝的水平位移与气温这两个量的相关关系，一般来说，它们相互有联系，但这种联系又不是很确定的。即使是两个月的平均气温完全相同，它们的位移量也会有大有小。回归分析就是用数理统计的方法，找出这种变量之间的相关关系的数学表达式。利用这些数学表达式以及对这些表达式的精度估计，可以对未知变量作出预测或检测其变化，或采取适当的对策。

回归分析方法有着广泛的应用，例如，在工程建筑物的变形分析中可用于建立位移量与某些相关因素之间的数学相关关系，即建立所谓的回归方程，根据所建立的回归方程分析变形的某些现象，并预报位移量，这对建筑物的稳定性监测与分析是十分重要的。

回归方程应采用何种形式，在没有对所讨论的问题进行全面考察的情况下是难以完全肯定回答的。在回归分析的实际中，因受各种因素的限制而无法确定回归函数的确切形式。但为了研究变量 x 和 y 之间的数值变化规律，人们可以从统计角度对回归函数的形式作一些必要的、合理的假设，但这种假设必须能反映问题的实质。设，回归方程为线性函数类，如取为

$$y=\beta_0+\beta_1 x, \tag{1.1.1}$$

只包含一个自变量的上述回归方程，称为一元回归方程。

在实际问题中，人们往往只能在 x 取一组定值的条件下，得到 y 的一组观测值（样本），而不是 y 的所有可能值，且这组 y 的样本值带有随机性，它必然带有随机抽样误差。因此，故回归方程(1.1.1)可改写为

$$y_i=\beta_0+\beta_1 x_i+\varepsilon_i, \qquad (i=1,2,\cdots n) \tag{1.1.2}$$

并称之为一元线性回归模型，式中 ε_i 为相应于 y_i 的随机误差。

推广之，当有 m 个自变量 $x_1,x_2\cdots x_m$ 时，因变量 y 对 $x_1,x_2\cdots x_m$ 的线性函数可写为

$$y=\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_m x_m \tag{1.1.3}$$

称之为多元回归方程。

当 $x_1x_2\cdots x_m$ 取定某一组定值时，则有多元回归模型：

$$y_i=\beta_0+\beta_1 x_{1i}+\beta_2 x_{2i}+\cdots\cdots+\beta_m x_{mi}+\varepsilon_i \tag{1.1.4}$$

回归分析的主要任务，就是利用 y 的观测值（$y_1,y_2,\cdots\cdots,y_n$）和自变量之值 $x_{1i},x_{2i},\cdots x_{mi}(i=1,2,\cdots\cdots,n)$，对回归方程进行统计分析，包括对回归方程进行估计和检验。

对回归方差）的方差 σ^2 的估计，以及分析估计值显著性检验等。

本章主要归模型的选择以及非线性回归的求解……x_{mi} 实际上也是观测值，或者说它们也是随机量，但为研究简便起见，一般是将它们视为非随机量，也可以说是在“取值 x_i”或 $x_{ki}(k=1,2,\cdots m)$ 的条件下来讨论。

§1.2　一元线性回归分析

一、一元线性回归参数的最小二乘估计

考虑因变量 y 和自变量 x 的一元线性回归方程(1.1.1)，即

$$y = \beta_0 + \beta_1 x, \tag{1.2.1}$$

式中 β_0, β_1 为未知的回归参数。

设在自变量 x 分别取值为 $x_1, x_2, \cdots\cdots x_n$ 时，对 y 有观测值 $y_1, y_2, \cdots\cdots y_n$，相应的随机误差为 $\varepsilon_1, \varepsilon_2, \cdots\cdots \varepsilon_n$，则 x_i, y_i 和 ε_i 之间有一元线性回归模型

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i。 \tag{1.2.2}$$

通常对 ε_i 作假定：ε_i 服从正态分布，其数学期望为 $E(\varepsilon_i)=0$，方差为 σ^2，且 ε_i 与 $\varepsilon_j (i \neq j)$ 之间的协方差为 0(即相互独立)。

又假定 x 是非随机量，因此，观测值 y_i 的方差亦为 σ^2，且 y_i 与 $y_j (y \neq j)$ 也是互相独立的。

现在的问题是如何由(1.2.2)式求定回归参数 $\beta_0 \beta_1$ 的最佳估值 $\hat{\beta}_0$、$\hat{\beta}_1$。以下按最小二乘估计法求定 $\hat{\beta}_0$ 和 $\hat{\beta}_1$，并讨论它们的精度。

为论述方便，令

$$y = [y_1, y_2 \cdots\cdots y_n]^T,$$
$$\varepsilon = [\varepsilon_1\ \varepsilon_2 \cdots\cdots \varepsilon_n]^T$$
$$x = [x_1\ x_2 \cdots\cdots x_n]^T,$$

$$A = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix},$$

则对 n 个观测值 y_i 由(1.2.2)式可构成

$$y = A\beta + \varepsilon, \qquad \varepsilon \sim N(0, I\sigma^2) \tag{1.2.3}$$

式中 I 表示单位阵，并有

$$\hat{\varepsilon} = y - A\hat{\beta}, \tag{1.2.4}$$

其中 $\hat{\varepsilon}$ 为随机误差 ε 的估值，一般称之为残差或改正值，$\hat{\beta}$ 为回归参数 β 的最佳估值。

依最小二乘估计法由(1.2.4)式求定 $\hat{\beta}$，也就是在

$$\hat{\varepsilon}^T\hat{\varepsilon} = (y - A\hat{\beta})^T(y - A\hat{\beta}) = \text{最小}, \tag{1.2.5}$$

的要求下求定 $\hat{\beta}_0, \hat{\beta}_1$。

不难看到，(1.2.4)式与测量平差课程中的误差方程的形式一致，因此，可以参照间接平差法来阐述回归参数的求解。由

$$\frac{\partial \hat{\varepsilon}^T\hat{\varepsilon}}{\partial \hat{\beta}} = -2\hat{\varepsilon}^T A = 0,$$

转置得

$$A^T\hat{\varepsilon} = 0, \tag{1.2.6}$$

将(1.2.4)式代入上式，可得法方程(或称为正规方程)

$$A^TA\hat{\beta} - A^Ty = 0, \tag{1.2.7}$$

解得

$$\hat{\beta} = (A^TA)^{-1}A^Ty。 \tag{1.2.8}$$

若令

$$\left.\begin{aligned} &\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \\ &s_x^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2 \\ &s_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y} \end{aligned}\right\}, \tag{1.2.9}$$

则有

$$\left.\begin{aligned} A^{\mathrm{T}}A &= \begin{bmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{bmatrix} = \begin{bmatrix} n & n\bar{x} \\ n\bar{x} & s_x^2 + n\bar{x}^2 \end{bmatrix} \\ A^{\mathrm{T}}y &= \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \end{bmatrix} = \begin{bmatrix} n\bar{y} \\ s_{xy} + n\bar{x}\bar{y} \end{bmatrix} \end{aligned}\right\}, \tag{1.2.10}$$

所以

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} n & n\bar{x} \\ n\bar{x} & s_x^2 + n\bar{x}^2 \end{bmatrix}^{-1} \begin{bmatrix} n\bar{y} \\ s_{xy} + n\bar{x}\bar{y} \end{bmatrix} = \frac{1}{s_x^2} \begin{bmatrix} \bar{y}s_x^2 - \bar{x}s_{xy} \\ s_{xy} \end{bmatrix},$$

取得

$$\left.\begin{aligned} \hat{\beta}_0 &= \bar{y} - \bar{x}\frac{s_{xy}}{s_x^2} = \bar{y} - \bar{x}\hat{\beta}_1 \\ \hat{\beta}_1 &= s_{xy}/s_x^2 \end{aligned}\right\}。 \tag{1.2.11}$$

在两个回归参数的估值 $\hat{\beta}_0$、$\hat{\beta}_1$ 求得后，就可得到一元线性回归方程为

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x = A\hat{\beta}, \tag{1.2.12}$$

残差 $\hat{\varepsilon}$ 可由(1.2.4)求得，且有

$$\hat{y} = y - \hat{\varepsilon}。 \tag{1.2.13}$$

二、估值的若干性质

1. 无偏性

因为由(1.2.3)式知 $E(y)=A\beta$，所以

$$E(\hat{\beta})=(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}E(y)=\beta,$$

$$E(\hat{\varepsilon})=E(y)-AE(\hat{\beta})=A\beta-A\beta=0。$$

2. 估值 $\hat{\beta}_0$，$\hat{\beta}_1$ 的方差

由 $D(\hat{\beta})=(A^{\mathrm{T}}A)^{-1}\sigma^2$，

可得

$$\left.\begin{aligned} D(\hat{\beta}_1) &= \frac{1}{s_x^2}\sigma^2 \\ D(\hat{\beta}_0) &= \left(\frac{1}{n} + \frac{\bar{x}^2}{s_x^2}\right)\sigma^2 \end{aligned}\right\}。 \tag{1.2.14}$$

3. 残差 $\hat{\varepsilon}$ 的方差为

$$D(\hat{\varepsilon}) = (I - A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}})\sigma^2 \tag{1.2.15}$$

4. 残差 $\hat{\varepsilon}$ 与 $\hat{\beta}$，$\hat{y}$ 不相关，即有

$$\left.\begin{aligned} D(\hat{\varepsilon},\hat{y}) &= 0 \\ D(\hat{\varepsilon},\hat{\beta}) &= 0 \end{aligned}\right\},$$

5. 估计值 $\hat{y}_i$ 的总和等于观测值 y_i 的总和，而残差 $\hat{\varepsilon}_i$ 的总和等于零。由(1.2.12)、(1.2.11)式和(1.2.13)式知

$$\left.\begin{aligned} \sum_{i=1}^{n}\hat{y}_i &= n\hat{\beta}_0 + \hat{\beta}_1\sum_{i=1}^{n}x_i \\ &= n(\bar{y} - \bar{x}\hat{\beta}_1) + n\bar{x}\hat{\beta}_1 = \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n}\hat{\varepsilon}_i &= \sum_{i=1}^{n} y_i - \sum_{i=1}^{n}\hat{y}_i = 0 \end{aligned}\right\}。 \tag{1.2.16}$$

6. 残差平方和(记为 Q_ε)

$$Q_\varepsilon = \hat{\varepsilon}^{\mathrm{T}}\hat{\varepsilon} = (y - A\hat{\beta})^{\mathrm{T}}\hat{\varepsilon} = y^{\mathrm{T}}\hat{\varepsilon} = y^{\mathrm{T}}(y - A\hat{\beta}) = \sum_{i=1}^{n} y_i^2 - \hat{\beta}_0\sum_{i=1}^{n} y_i - \hat{\beta}_1\sum_{i=1}^{n} x_i y_i \quad ,$$

顾及(1.2.9)式,并令

$$Q_y = s_y^2 = \sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - n\bar{y}^2 \quad , \tag{1.2.17}$$

顾及(1.2.9)式,则可得

$$\begin{aligned} Q_\varepsilon = \hat{\varepsilon}^{\mathrm{T}}\hat{\varepsilon} &= (s_y^2 + n\bar{y}^2) - (\bar{y} - \bar{x}\hat{\beta}_1)n\bar{y} - \hat{\beta}_1(s_{xy} - n\bar{x}\bar{y}) \\ &= s_y^2 - s_{xy}\hat{\beta}_1 = s_y^2 - \frac{s_{xy}^2}{s_x^2} \end{aligned} \tag{1.2.18}$$

7. 方差 σ^2 的无偏估值 $\hat{\sigma}^2$

由 $E(\hat{\varepsilon}^{\mathrm{T}}\hat{\varepsilon}) = t_r[D(\hat{\varepsilon})] = t_r[I - A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}]\sigma^2$

可得

$$\hat{\sigma}^2 = \frac{1}{n-2}\hat{\varepsilon}^{\mathrm{T}}\hat{\varepsilon} = \frac{Q_\varepsilon}{n-2} \quad 。 \tag{1.2.19}$$

三、一元回归的方差分析和线性关系的显著性检验

在采用一元回归模型(1.2.3)时,假设变量 y 与自变量 x 呈线性关系。这种假设是否恰当,需要通过统计检验来确定。一般说来,可以通过方差分析法来构造检验的统计量。

(1.2.17)式的 Q_y(即 s_y^2)是变量 y 的观测值 y_i 与其平均值 $\bar{y}$ 之间的总偏差平方和,现将它分解为两部分,即

$$Q_y = s_y^2 = \sum_{i=1}^{n}[(y_i - \hat{y}_i) + (\hat{y}_i - \bar{y})]^2 \quad , \tag{1.2.20}$$

则得 $Q_y = s_y^2 = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + 2\sum_{i=1}^{n}(y_i - \hat{y}_i)(\hat{y}_i - \bar{y})$,其中 $y_i - \hat{y}_i = \hat{\varepsilon}_i$,即为残差,顾及(1.2.16)和(1.2.6)式,则在上式中有

$$\sum_{i=1}^{n}(y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) = \hat{\varepsilon}^{\mathrm{T}}\hat{y} - \bar{y}\sum_{i=1}^{n}\hat{\varepsilon}_i = \hat{\varepsilon}^{\mathrm{T}}A^{\mathrm{T}}\hat{\beta} = 0.$$

所以

$$Q_y = \hat{\varepsilon}^{\mathrm{T}}\hat{\varepsilon} + \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 = Q_\varepsilon + Q_R, \tag{1.2.21}$$

其中残差平方和 Q_ε 可按(1.2.18)式计算,而 Q_R 是回归直线与 $y=\bar{y}$ 之间的偏离平方和,有

$$\begin{aligned} Q_R &= \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 = \sum_{i=1}^{n}[(\hat{\beta}_0 + \hat{\beta}_1 x_i) - (\hat{\beta}_0 + \hat{\beta}_1\bar{x})]^2 \\ &= \sum_{i=1}^{n}[\hat{\beta}_1^2(x_i - \bar{x})^2] = \hat{\beta}_1^2 s_x^2 \quad 。 \end{aligned} \tag{1.2.22}$$

可以看到,观测值 y_i 对其平均值 $\bar{y}$ 的总偏离平方和 s_y^2 可以分解为 Q_R 和 Q_ε 两部分,Q_R 表示了当变量 y 和 x 之间完全按照回归方程线性变化时,由于 x_i 在 $\bar{x}$ 周围变化而引起 y_i 对 $\bar{y}$ 的偏离平方和,通常称之为回归平方和;另一部分 Q_ε 也就是残差平方和,它代表除了上述线性回归模型引起的偏离以外的其它因素,如试验误差,观测误差等随机因素,以及线性模型本身所不能描述的误差所造成的偏离平方和,通常也称之为剩余平方和。因此,Q_R 越大,Q_ε 就越小,这表示 y 对 x 的线性关系显著;反之,Q_R 越小 Q_ε 就越大,表示 y 对 x 的线性关系不显著。这样,也就可以由此得到一种判别回归效果的方法。如果回归平方和 Q_R 与总的偏离平方和 Q_y 之比值(Q_R/Q_y)接近于 1,或 Q_R 与剩余平方和 Q_ε 之比值(Q_R/Q_ε)较大,则认为回归效果较好。

当 $y_i \sim N(\mu,\sigma^2)$,$\beta_1 = 0$ 时,由(1.2.22)和(1.2.18)式可以证明,Q_R/σ^2 和 Q_ε/σ^2 都服从 χ^2 分布,且它们的自由度分别为 $f_R = 1$ 和 $f_\varepsilon = n-2$,即有

$$\frac{Q_R}{\sigma^2} \sim \chi^2(1), \qquad \frac{Q_\varepsilon}{\sigma^2} \sim \chi^2(n-2) \quad ,$$

而 Q_y 的自由度为

$$f_y = f_R + f_\varepsilon = n - 1$$

同时，Q_R/σ^2 与 Q_ε/σ^2 是相互独立的。

因此，为检验回归效果的显著性，可以作原假设

$$H_0: \quad \beta_1 = 0$$

并构造统计量

$$F = \frac{\frac{Q_R}{\sigma^2}\Big/1}{\frac{Q_\varepsilon}{\sigma^2}\Big/(n-2)} = \frac{Q_R}{Q_\varepsilon/(n-2)} \sim F(1, n-2) \quad . \tag{1.2.23}$$

在给定置信水平 α 后，从 F 分布表中查取 $F_{1-\alpha}(1, n-2)$，如果 $F > F_{1-\alpha}(1, n-2)$，则拒绝 H_0。即认为回归效果是显著的；反之，则说明回归效果不显著。

又因 $\hat{\beta}_1$ 与 ε 不相关，$\hat{\beta}_1$ 也就与 Q_ε 不相关，故当 $\beta_1 = 0$ 时，有统计量

$$T = \frac{\hat{\beta}_1 \Big/ \frac{\sigma}{s_x}}{\sqrt{\frac{Q_\varepsilon}{\sigma^2}/(n-2)}} = \frac{s_x\hat{\beta}_1}{\sqrt{Q_\varepsilon/(n-2)}} = \frac{s_x\hat{\beta}_1}{\sigma} \sim t(n-2) \quad , \tag{1.2.24}$$

因此，对于给定的显著性水平 α，采用 t 检验法用 T 统计量检验 H_0，当 $|T| > t_{1-\frac{\alpha}{2}}(n-2)$ 时，拒绝 H_0，否则就接受 H_0。

四、样本相关系数 γ

前已述及，如果 Q_R/Q_y 之比值接近于 1，则回归效果显著。将该比值记为 γ^2，由(1.2.22)和(1.2.11)式得

$$\gamma^2 = \frac{Q_R}{Q_y} = \frac{\hat{\beta}_1^2 s_x^2}{s_y^2} = \left(\frac{s_{xy}}{s_x s_y}\right)^2, \tag{1.2.25}$$

则有

$$\gamma = \frac{s_{xy}}{s_x s_y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad . \tag{1.2.26}$$

如 §1-1 所述，在回归分析中，为研究方便而将自变量 x 视作非随机量。但实际上 x_i 也是观测值，或者说它们也是随机量的样本值。所以也可以说 s_x^2 是 x 的样本方差。s_{xy} 可以说是 x 与 y 的样本协方差，而 s_y^2 是 y 的样本方差。因此，通常称 γ 为 x 与 y 的样本相关系数。

由(1.2.25)可以看出，γ 的符号与 $\hat{\beta}_1$ 一致。又因为

$$\left.\begin{aligned} \gamma^2 &= \frac{Q_R}{Q_y} = \frac{Q_y - Q_\varepsilon}{Q_y} = 1 - \frac{Q_\varepsilon}{Q_y} \\ Q_\varepsilon &= (1 - \gamma^2)Q_y \end{aligned}\right\}, \tag{1.2.27}$$

而 $Q_y > Q_R$，$Q_y > Q_\varepsilon$，所以 $|\gamma| < 1$，且当 Q_y 固定时，$|\gamma|$ 越近 1，Q_ε 就越小；而当 $|\gamma| = 1$ 时，$Q_\varepsilon = 0$，$Q_R = Q_y$，即 y 的变化完全由 y 与 x 的线性关系引起。因此，可以用 γ 来描述 y 与 x 之间的线性相关程度。也可用来检验原假设 $H_0: \beta_1 = 0$，其检验方法称为 γ 检验。

因为

$$F = \frac{Q_R(n-2)}{Q_\varepsilon} = \frac{(n-2)\gamma^2 Q_y}{(1-\gamma^2)Q_y} = \frac{(n-2)\gamma^2}{1-\gamma^2} \quad , \tag{1.2.28}$$

可得

$$\gamma=\sqrt{\frac{F}{F+(n-2)}}\quad , \tag{1.2.29}$$

则由

$$P\left\{\frac{(n-2)\gamma^2}{1-\gamma^2}\geqslant F_{1-\alpha}(1,n-2)\right\}=\alpha\quad ,$$

得

$$P\{|\gamma|\geqslant\gamma_{1-\alpha}(n-2)\}=\alpha\quad 。 \tag{1.2.30}$$

式中

$$\gamma_{1-\alpha}(n-2)=\sqrt{\frac{F_{1-\alpha}(1,n-2)}{F_{1-\alpha}(1,n-2)+(n-2)}}\quad 。 \tag{1.2.31}$$

当应用 γ 检验法对原假设($H_0:\beta_1=0$)进行检验时，可利用(1.2.29)式计算 γ，给定显著性水平 α，然后由 F 分布表查得 $F_{1-\alpha}(1,n-2)$，按(1.2.31)式计算 $\gamma_{1-\alpha}(n-2)$，或按(1.2.31)式编制出 γ 临界值表后，由相应的数表查得 $\gamma_{1-\alpha}(n-2)$，如果 $|\gamma|>\gamma_{1-\alpha}(n-2)$，则拒绝 H_0，否则接受 H_0，容易理解 γ 检验与上面的 F 检验法实质上是一致的。

【例 1.1】 某回归问题的自变量取值 x_i 和观测值 y_i 的数据如表 1-1，试求回归方程，并应用 t 检验，F 检验和 γ 检验回归效果是否显著。

解(1) 将表 1-1 中的观测数据描绘在平面坐标系中得到图 1-1。从图上可以看出，随着 x_i 的增加，y_i 值也随之增加，且大致成直线关系。但因受到一些随机因素的影响，各点不完全在一条直线上。为此，可以采用一元回归模型

$$y_i=\beta_0+\beta_1x_i+\varepsilon_i$$

表 1-1 观测数据与残差

序号	x_i	y_i	$\hat{y}_i$	$\hat{\varepsilon}_i$
1	−5	1	2.093	−1.093
2	−4	5	3.529	1.471
3	−3	4	4.965	−0.964
4	−2	7	6.401	0.599
5	−1	10	7.837	2.163
6	0	8	9.273	−1.273
7	1	9	10.709	−1.709
8	2	13	12.145	0.855
9	3	14	13.581	0.419
10	4	13	15.017	−2.017
11	5	18	16.453	1.547

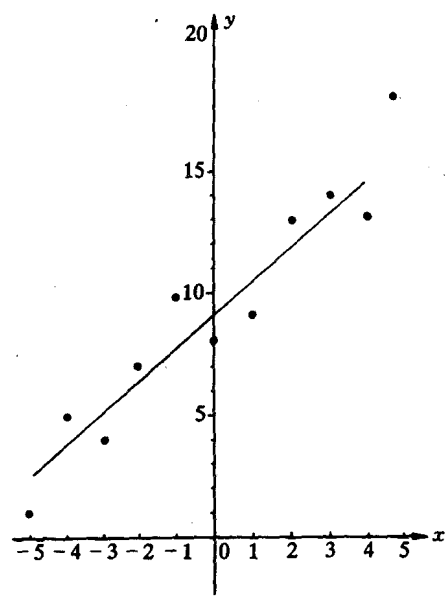


图 1.1 数据点与回归直线

(2) 按(1.2.11)求回归参数的估值 $\hat{\beta}_0$、$\hat{\beta}_1$，有

$$\bar{x}=0,\qquad \bar{y}=9.273$$

$$s_x^2=\sum_{i=1}^{n}(x_i-\bar{x})^2=110$$

$$s_{xy}=\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})=158$$

$$\hat{\beta}_1=\frac{s_{xy}}{s_x^2}=\frac{158}{110}=1.436$$

$$\hat{\beta}_0 = \bar{y} - \bar{x}\hat{\beta}_1 = 9.273$$

故回归方程为

$$\hat{y} = 9.273 + 1.436x$$

(3) 残差与方差估值 $\hat{\sigma}$

残差 $\hat{\varepsilon}$ 和回归值 $\hat{y}_i$ 按(1.2.4)和(1.2.13)式计算，也列于表 1-1 中。残差平方和 $\hat{\sigma}$ 为

$$Q_e = \hat{\varepsilon}^T\hat{\varepsilon} = 21.23$$

$$\hat{\sigma} = \sqrt{\frac{\hat{\varepsilon}^T\hat{\varepsilon}}{n-2}} = \sqrt{\frac{21.23}{9}} = \pm 1.536$$

(4) t 检验

$$T = \frac{s_x\hat{\beta}_1}{\sqrt{Q_e/(n-2)}} = \frac{\sqrt{110} \times 1.436}{1.536} = 9.805$$

取 $\alpha=0.01$，查得 $t_{1-0.005}(9)=3.2498<9.805$，故拒绝 $H_0:\beta_1=0$，即可以认为回归效果是显著的。

(5) F 检验和 γ 检验

因 $Q_R=\hat{\beta}_1^2 s_x^2=1.436^2\times 110=226.83$

$Q_e=21.23$

得 $F=\dfrac{Q_R}{Q_e/(n-2)}=\dfrac{226.83}{21.23/9}=96.16$

取 $\alpha=0.01$，查 F 分布表，得 $F_{1-\alpha}(1,9)=10.56$，因为 $F>F_{1-\alpha}(1,9)$，故也拒绝 H_0，认为回归效果显著。

又可得

$$|\gamma| = \sqrt{\frac{Q_R}{Q_y}} = \sqrt{\frac{226.83}{248.06}} = 0.9542$$

而

$$\gamma_{1-\alpha}(9) = \sqrt{\frac{F_{1-\alpha}(1,9)}{F_{1-\alpha}(1,9)+9}} = 0.7348 < |\gamma|$$

这表明，由 γ 检验也可认为回归效果显著。

§1.3 多元线性回归分析

上节讨论的是最简单的一元线性回归，回归模型中只有一个自变量。在实际问题中，影响变量 y 的因素往往不只一个，而包含多种影响的多个自变量。例如，在大坝监测中，大坝的位移量 y，包含的自变量有时间、气温、水位等。通常将研究一个因变量 y 与多个自变量之间的关系问题称为多元回归分析。当然，多元回归分析也是一元回归分析的推广。

一、多元线性回归的最小二乘估计

假定因变量 y 与 m 个自变量 $x_1,x_2,\cdots,x_m$ 有线性关系

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_m x_m, \tag{1.3.1}$$

设在 $x_1,x_2,\cdots,x_m$ 分别取值 $x_{1i},x_{2i},\cdots,x_{mi}$ 时，对 y 取得第 i 次样本观测值 y_i，则对观测值 y_i 有

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_m x_{mi} + \varepsilon_i, \tag{1.3.2}$$

式中 ε_i 为相应于 y_i 的随机误差，通常设 $\varepsilon_i \sim N(0,\sigma^2)$，即设它们为互相独立的服从数学期望为 0，方差为 σ^2 的正态随机变量。并且与一元线性回归一样，样本观测值 y_i 的方差也为 σ^2，y_i 与 $y_j(i\neq j)$ 也是相互独立的。需要注意的是，这里也假定 $x_1,x_2,\cdots,x_m$ 是非随机量，虽然 $x_{1i},x_{2i},\cdots,x_{mi}$ 实际上往往也是观测值，但还是将它们当作是非随机量的取定值。

现在由 $(y_i;x_{1i},x_{2i},\cdots,x_{mi})$，根据(1.3.2)式求 $(m+1)$ 个未知的回归参数 $\beta_0,\beta_1,\cdots,\beta_m$ 的最小二乘估值

$\hat{\beta}_0,\hat{\beta}_1,\cdots,\hat{\beta}_m$。

若记

$$y=\begin{bmatrix}y_1\\y_2\\\vdots\\y_n\end{bmatrix},\quad \varepsilon=\begin{bmatrix}\varepsilon_1\\\varepsilon_2\\\vdots\\\varepsilon_n\end{bmatrix},\quad \beta=\begin{bmatrix}\beta_0\\\beta_1\\\vdots\\\beta_m\end{bmatrix}=\begin{bmatrix}\beta_0\\\beta_1\\\vdots\\\beta_m\end{bmatrix},$$

$$A=\begin{bmatrix}1&x_{11}&x_{21}&\cdots&x_{m1}\\1&x_{12}&x_{22}&\cdots&x_{m2}\\\vdots&\vdots&\vdots&&\vdots\\1&x_{1n}&x_{2n}&\cdots&x_{mn}\end{bmatrix},$$

则对 n 个观测值由(1.3.2)式可得

$$y=A\beta+\varepsilon, \tag{1.3.3}$$

或写为

$$\hat{\varepsilon}=y-A\hat{\beta}。 \tag{1.3.4}$$

(1.3.3)式和(1.3.4)式与§1.2中的(1.2.3)式和(1.2.4)式的形式相同，式中 $\hat{\varepsilon}$ 也是随机误差 ε 的估值，称为残差或改正值，而称(1.3.4)式为误差方程。

按最小二乘估计法由(1.3.4)式求未知的多元回归参数的估值 $\hat{\beta}$，也就是在

$$\hat{\varepsilon}^{\mathrm{T}}\hat{\varepsilon}=(y-A\hat{\beta})^{\mathrm{T}}(y-A\hat{\beta})=\text{最小} \tag{1.3.5}$$

的要求下求定 $\hat{\beta}$。

同样可以参照测量平差法求解 $\hat{\beta}$，由

$$\frac{\partial\hat{\varepsilon}^{\mathrm{T}}\hat{\varepsilon}}{\partial\hat{\beta}}=-2\hat{\varepsilon}^{\mathrm{T}}A=0$$

经转置可得

$$A^{\mathrm{T}}\hat{\varepsilon}=0, \tag{1.3.6}$$

将(1.3.4)式代入上式，得法方程

$$A^{\mathrm{T}}A\hat{\beta}-A^{\mathrm{T}}y=0, \tag{1.3.7}$$

式中

$$A^{\mathrm{T}}A=\begin{bmatrix}n&\sum x_{1i}&\cdots\sum x_{mi}\\\sum x_{1i}&\sum x_{1i}^2&\cdots\sum x_{1i}x_{mi}\\\vdots&\vdots&\vdots\\\sum x_{mi}&\sum x_{1i}x_{mi}&\cdots\sum x_{mi}^2\end{bmatrix},\quad A^{\mathrm{T}}y=\begin{bmatrix}\sum y_i\\\sum x_{1i}y_i\\\vdots\\\sum x_{mi}y_i\end{bmatrix},$$

可解得

$$\hat{\beta}=(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}y。 \tag{1.3.8}$$

在回归参数求得后，也可得到多元回归方程和误差方程

$$\hat{y}=A\hat{\beta}, \tag{1.3.9}$$

$$\hat{\varepsilon}=y-A\hat{\beta}=y-\hat{y}。 \tag{1.3.10}$$

则对于 $\hat{y}_i$ 有

$$\hat{y}_i=\hat{\beta}_0+x_{1i}\hat{\beta}_1+x_{2i}\hat{\beta}_2+\cdots+x_{mi}\hat{\beta}_m, \tag{1.3.11}$$

$$\hat{\varepsilon}_i=y_i-\hat{y}_i=y_i-\hat{\beta}_0-x_{1i}\hat{\beta}_1-\cdots x_{mi}\hat{\beta}_m。 \tag{1.3.12}$$

二、参数的中心化解

为便于分析，对以上求解方程作适当变换，将自变量值域空间的原点移至它的 n 次取值的中心点处，称为中心化。

记

$$\bar{x}_k = \frac{1}{n}\sum_{i=1}^{n} x_{ki} \quad (k = 1,2\cdots,m),$$

$$\bar{x} = [\bar{x}_1\ \bar{x}_2\ \cdots\ \bar{x}_m]^{\mathrm{T}},$$

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

则由(1.3.7)式的第一式可解得

$$\begin{aligned}\hat{\beta}_0 &= \bar{y} - \bar{x}^{\mathrm{T}}\hat{\beta}_x \\ &= \bar{y} - \bar{x}_1\hat{\beta}_1 - \bar{x}_2\hat{\beta}_2 - \cdots - \bar{x}_m\hat{\beta}_m。\end{aligned} \tag{1.3.13}$$

将(1.3.13)式代入(1.3.7)式中的其它各式，则可得到中心化法方程

$$A_x^{\mathrm{T}}A_x\hat{\beta}_x - A_x^{\mathrm{T}}y_x = 0, \tag{1.3.14}$$

式中

$$A_x = \begin{bmatrix} x_{11}-\bar{x}_1 & x_{21}-\bar{x}_2 & \cdots & x_{m1}-\bar{x}_m \\ x_{12}-\bar{x}_1 & x_{22}-\bar{x}_2 & \cdots & x_{m2}-\bar{x}_m \\ \vdots & \vdots & & \vdots \\ x_{1n}-\bar{x}_1 & x_{2n}-\bar{x}_2 & \cdots & x_{mn}-\bar{x}_m \end{bmatrix} = \begin{bmatrix} A_{x1} \\ A_{x2} \\ \vdots \\ A_{xn} \end{bmatrix}, \quad y_x = \begin{bmatrix} y_1-\bar{y} \\ y_2-\bar{y} \\ \vdots \\ y_n-\bar{y} \end{bmatrix}。$$

(1.3.14)式是一组含 m 个回归参数 $\hat{\beta}_1,\hat{\beta}_2,\cdots,\hat{\beta}_m$ 和 m 个方程的对称性方程组，由此式可解得

$$\hat{\beta}_x = (A_x^{\mathrm{T}}A_x)^{-1}A_x^{\mathrm{T}}y_x, \tag{1.3.15}$$

且可将(1.3.11)式改写为

$$\left.\begin{aligned}\hat{y}_i &= \bar{y} + (x_{1i}-\bar{x}_1)\hat{\beta}_1 + \cdots + (x_{mi}-\bar{x}_m)\hat{\beta}_m = \bar{y} + A_{xi}\hat{\beta}_x \\ \hat{\varepsilon}_i &= (y_i-\bar{y}) - (x_{1i}-\bar{x}_1)\hat{\beta}_1 + \cdots + (x_{mi}-\bar{x}_m) = A_{xi}\hat{\beta}_x\end{aligned}\right\} \tag{1.3.16}$$

误差方程(1.3.10)也就可以改写为

$$\hat{\varepsilon} = y_x - A_x\hat{\beta}_x。 \tag{1.3.17}$$

三、多元线性回归统计性质

按照类似于测量平差的推证方法，还可以得到以下结果：

(1) 方差 σ^2 的无偏估值为

$$\hat{\sigma}^2 = \frac{Q_\varepsilon}{n-m-1} = \frac{\hat{\varepsilon}^{\mathrm{T}}\hat{\varepsilon}}{n-m-1}, \tag{1.3.18}$$

式中残差平方和 Q_ε 为

$$Q_\varepsilon = \hat{\varepsilon}^{\mathrm{T}}\hat{\varepsilon} = (y - A\hat{\beta})^{\mathrm{T}}\hat{\beta},$$

顾及(1.3.6)式得

$$Q_\varepsilon = y^{\mathrm{T}}\hat{\varepsilon} = y^{\mathrm{T}}y - (A^{\mathrm{T}}y)^{\mathrm{T}}\hat{\varepsilon}。 \tag{1.3.19}$$

又由(1.3.16)式和(1.3.14)式可得

$$A_x^{\mathrm{T}}(A_x\hat{\beta}_x + y_x) = A_x^{\mathrm{T}}\hat{\varepsilon} = 0, \tag{1.3.20}$$

$$Q_\varepsilon = y_x^{\mathrm{T}}\hat{\varepsilon} = y_x^{\mathrm{T}}y_x - (A_x^{\mathrm{T}}y_x)^{\mathrm{T}}\hat{\beta}_x。 \tag{1.3.21}$$

(2) $\hat{\beta}$ 和 $\hat{\beta}_x$ 是 β 和 β_x 的无偏估值，即

$$\left.\begin{aligned}E(\hat{\beta}) &= \beta \\ E(\hat{\beta}_x) &= \beta_x\end{aligned}\right\}。$$

(3) $\hat{\beta}$ 的方差为

$$D(\hat{\beta}) = \sigma^2(A^{\mathrm{T}}A)^{-1}, \tag{1.3.22}$$

而 $\hat{\beta}_x$ 的方差为

$$D(\hat{\beta}_x) = \sigma^2(A_x^{\mathrm{T}}A_x)^{-1}。 \tag{1.3.23}$$

(4) $\hat{\beta}$、$\hat{y}$ 与残差 $\hat{\varepsilon}$ 不相关，即

$$\left.\begin{aligned} D(\hat{\beta},\hat{\varepsilon}) &= \operatorname{cov}(\hat{\beta},\hat{\varepsilon}) = 0 \\ D(y,\hat{\varepsilon}) &= \operatorname{cov}(y,\hat{\varepsilon}) = 0 \end{aligned}\right\}。$$

(5) $\hat{y}$、$\hat{\varepsilon}$ 的方差为

$$\left.\begin{aligned} D(\hat{y}) &= \sigma^2 A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}} \\ D(\hat{\varepsilon}) &= \sigma^2[I - A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}] \end{aligned}\right\}。 \tag{1.3.24}$$

(6) 由(1.3.6)式和(1.3.20)式展开，可得

$$\left.\begin{aligned} &\sum_{i=1}^{n}\varepsilon_i = 0 \\ &\sum_{i=1}^{n}x_{ki}\varepsilon_i = 0 \qquad (k = 1,2,\cdots,m) \\ &\sum_{i=1}^{n}(x_{ki} - \bar{x}_k)\varepsilon_i = 0 \end{aligned}\right\}。 \tag{1.3.25}$$

(7) $\bar{y}$ 与 $\hat{y}$ 的方差为

$$\left.\begin{aligned} D(\bar{y}) &= \frac{1}{n}\sigma^2 \\ D(\hat{y}_i) &= D(\bar{y}) + \sigma^2 A_{si}(A_s^{\mathrm{T}}A_s)^{-1}A_{si}^{\mathrm{T}} \\ &= \sigma^2\{\frac{1}{n} + A_{si}(A_s^{\mathrm{T}}A_s)^{-1}A_{si}^{\mathrm{T}}\} \end{aligned}\right\}。 \tag{1.3.26}$$

(8) y_s、$\hat{\beta}_s$ 与 $\bar{y}$ 不相关，即有

$$D(y_s,\bar{y}) = 0, \quad D(\hat{\beta}_s,\bar{y}) = 0 \quad 。$$

四、多元线性回归的方差分析和显著性检验

与一元回归分析一样，按上面的方法求得的回归方程是否能较好地描述 y 与 $x_1,x_2,\cdots,x_m$ 之间的变化规律，还需要进一步作方差分析和回归效果的显著性检验。

对变量 y 相对于平均值 $\bar{y}$ 的总偏离平方和进行分解，即有

$$\begin{aligned} Q_y &= \sum_{i=1}^{n}(y_i - \bar{y})^2 = y_s^{\mathrm{T}}y_s \\ &= \sum_{i=1}^{n}[(y_i - \hat{y}_i) + (\hat{y}_i - \bar{y})]^2 \\ &= \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + 2\sum_{i=1}^{n}(y_i - \hat{y}_i)(\hat{y}_i - \bar{y})。 \end{aligned} \tag{1.3.27}$$

类似于一元回归的情况，顾及(1.3.6)式和(1.3.25)式，知上式最后一项中的交叉乘积和为零，即

$$\begin{aligned} \sum_{i=1}^{n}(y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) &= \sum_{i=1}^{n}(y_i - \hat{y}_i)\hat{y}_i - \sum_{i=1}^{n}(y_i - \hat{y}_i)\bar{y} \\ &= \hat{\varepsilon}^{\mathrm{T}}\hat{y} - (\sum_{i=1}^{n}\hat{\varepsilon}_i)\bar{y} = \hat{\varepsilon}^{\mathrm{T}}A\hat{\beta} = 0, \end{aligned}$$

亦可得

$$Q_y = Q_e + Q_R,$$

其中残差平方和 Q_e 可按(1.3.19)式或(1.3.21)式计算，而在回归平方和 $Q_R = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$ 中因为

$$\begin{aligned} \hat{y}_i - \bar{y} &= (\beta_0 + x_{1i}\hat{\beta}_1 + \cdots + x_{mi}\hat{\beta}_m) - (\beta_0 + \bar{x}_i\hat{\beta}_1 + \cdots + \bar{x}_m\hat{\beta}_m) \\ &= (x_{1i} - \bar{x}_1)\hat{\beta}_1 + \cdots + (x_{mi} - \bar{x}_m)\hat{\beta}_m, \end{aligned}$$

所以

$$\begin{aligned} Q_R &= (A_s\hat{\beta}_s)^{\mathrm{T}}(A_s\hat{\beta}_s) = \hat{\beta}_s^{\mathrm{T}}A_s^{\mathrm{T}}A_s\hat{\beta}_s \\ &= \hat{\beta}_s^{\mathrm{T}}(A_s^{\mathrm{T}}y_s)。 \end{aligned} \tag{1.3.28}$$

可见，对于多元回归，也可以将 y 对 $\bar{y}$ 的偏离平方和分解为回归平方和 Q_R 和残差平方和 Q_ε 两部分。Q_R 越大，Q_ε 越小，表示 y 对自变量 $x_1,x_2,\cdots,x_m$ 的线性关系显著；否则，表示它们的线性关系不显著。因此，当比值 (Q_R/Q_y) 接近于 1 时，就可以认为回归效果好。

因为 $\hat{\beta}_x$ 的方差为 $D(\hat{\beta}_x)=\sigma^2(A_x^{\mathrm{T}}A_x)^{-1}$，所以，由(1.3.28)式知，当 $y_i\sim N(\mu,\sigma^2)$ $\beta_i=0$ 时，有

$$\frac{Q_R}{\sigma^2}\sim\chi^2(m),\quad \frac{Q_\varepsilon}{\sigma^2}\sim\chi^2(n-m-1),$$

且 Q_R/σ^2 与 Q_ε/σ^2 之间相互独立，因此，可以构造统计量

$$F=\frac{\frac{Q_R}{\sigma^2}/m}{\frac{Q_\varepsilon}{\sigma^2}/(n-m-1)}=\frac{Q_R/m}{Q_\varepsilon/(n-m-1)}\sim F(m,n-m-1)。\tag{1.3.29}$$

用自由度为 m 和 $n-m-1$ 的 F 变量作 F 检验，对于原假设

$$H_0:\quad \beta_1=\beta_2=\cdots\cdots=\beta_m=0$$

取显著性水平 α，若从样本值按(1.3.29)式计算的 $F>F\alpha(m,n-m-1)$，则拒绝原假设 H_0，说明回归效果显著，即建立的多元回归模型基本上反映了变量 y 与自变量 $x_1,x_2,\cdots,x_m$ 之间的变化规律。

另一方面，对 y 与自变量 $x_1,x_2,\cdots,x_m$ 的回归平方和 Q_R 及总偏离平方和 Q_y，记

$$R=\sqrt{\frac{Q_R}{Q_y}}=\sqrt{1-\frac{Q_\varepsilon}{Q_y}}\quad 。\tag{1.3.30}$$

因 $Q_R>Q_\varepsilon$，所以

$$0<R\leqslant 1$$

称 R 为 y 与 $x_1,x_2,\cdots,x_m$ 的变相关系数。R 也是评价多元线性回归模型对变量之间线性关系代表性的一个指标，它的值越大，说明回归效果越好。容易证明

$$\left.\begin{aligned}R&=\sqrt{\frac{F}{F+\frac{n-m-1}{m}}}\\F&=\frac{(n-m-1)R^2}{m(1-R^2)}\end{aligned}\right\},\tag{1.3.31}$$

故用 R 检验 H_0，与用 F 检验是等价的。

五、回归系数的显著性检验

在经 F 检验或 R 检验确认 y 与自变量 $x_1,x_2,\cdots,x_m$ 之间在总体上有显著的线性关系后，还需要检验每个自变量 x_k 的显著性。如果 x_k 对 y 的作用不显著，x_k 的回归参数 β_k 就应该等于零。也就是说，要检验的原假设是

$$H_{0k}:\quad \beta_k=0\quad (k=1,2,\cdots,m)。$$

由 $\hat{\beta}_x$ 的无偏性和(1.3.23)式知

$$\hat{\beta}_x\sim N[\beta_x,\sigma^2(A_x^{\mathrm{T}}A_x)^{-1}]。\tag{1.3.32}$$

设 q_k 为 $(A_x^{\mathrm{T}}A_x)^{-1}$ 的第 k 个对角元素，则有

$$(\hat{\beta}_k-\beta_k)/\sigma\sqrt{q_k}\sim N(0,1),\tag{1.3.33}$$

又因为残差 $\hat{\varepsilon}$ 与 $\hat{\beta}_x$ 不相关，当然 $\hat{\beta}_k$ 与残差平方和 Q_ε 也不相关。因此，当 H_0 成立时，有

$$\left.\begin{aligned}&\hat{\beta}_k^2/\sigma^2q_k\sim\chi^2(1)\\&Q_\varepsilon/\sigma^2(n-m-1)\sim\chi^2(n-m-1)\end{aligned}\right\},\tag{1.3.34}$$

则可以构成统计量

$$\left.\begin{aligned}F_k&=\frac{\hat{\beta}_k^2/q_k}{Q_\varepsilon/(n-m-1)}=\frac{\hat{\beta}_k^2}{\sigma^2q_k}\sim F(1,n-m-1)\\T_k&=\frac{\hat{\beta}_k/\sqrt{q_k}}{\sqrt{Q_\varepsilon/(n-m-1)}}=\frac{\hat{\beta}_k}{\hat{\sigma}\sqrt{q_k}}\sim t(n-m-1)\end{aligned}\right\}。\tag{1.3.35}$$