管理预测定量方法与模型

秦德智　刘新卫　编著

科学出版社
www.sciencep.com

管理预测定量方法与模型

秦德智　刘新卫　编著

科学出版社

北京

内 容 简 介

在经济管理问题中，相当多的一类问题可以归结为在某种特定环境中的决策问题，而正确的决策离不开科学的预测。本书较为系统地、完整地从理论和应用方面介绍预测中的定量方法。全书共分十章，分别为回归分析预测方法、计量模型预测方法、时间序列分析模型预测方法、随机时间序列模型预测方法、增长曲线模型预测方法、投入产出分析预测方法、马尔柯夫预测法、非参数预测方法、分形预测方法、组合预测方法。

本书可作为高等院校管理学及相关专业师生的教学用书，也可供科研工作者参考使用。

图书在版编目（CIP）数据

管理预测定量方法与模型/秦德智，刘新卫编著.—北京：科学出版社，2007.1

ISBN 978-7-03-018230-2

Ⅰ.管… Ⅱ.①秦 ·②刘… Ⅲ.定量分析－应用－管理学 Ⅳ.C931.1

中国版本图书馆 CIP 数据核字（2006）第 148913 号

责任编辑：江 兰／责任校对：董 丽

责任印制：高 嵘／封面设计：曹 刚 陈良丽

科学出版社出版

北京东黄城根北街 16 号

邮政编码 100717

http://www.sciencep.com

武汉大学出版社印刷总厂印刷

科学出版社发行 各地新华书店经销

*

2007 年 1 月第 一 版 开本：B5（720×1000）

2007 年 1 月第一次印刷 印张：10 3/4

印数：1—1 500 字数：210 000

定价：21.80 元

（如有印装质量问题，我社负责调换）

前　言

在经济管理问题中，无论是微观意义下的企业管理、农业管理、商业管理、交通运输管理，还是宏观意义下的国民经济管理，相当多的一类问题可以归结为在某种特定环境中的决策问题——“管理即决策”. 而正确的决策离不开科学的预测，“预则立”，作为决策依据的预测方法和结果需求越来越普遍，在管理中对于预测的原理与方法的研究更是必不可少的.

随着科技革命和新兴工业的迅猛发展，生产的社会化有了巨大的发展，逐渐从封闭型的经济结构转变为开放型的结构. 控制和影响经济的因素增多了，经济结构及其运动规律变复杂了. 为了满足社会直接的迫切的需要，预测技术也有了长足的发展. 关于预测技术，目前的发展可以从广度与深度两方面来看：一是新的预测方法不断涌现，如灰色系统预测、人工神经网络预测、组合预测、智能预测等，而且预测技术的应用面越来越广；二是预测的理论深度不断加大. 传统的预测方法多以统计预测为主要手段. 目前除了统计预测的理论不断深入外，新兴的学科在预测中的应用也带来了理论上的新要求，如混沌预测就是一个例证. 大多经济现象是一种随机现象，从数据处理的角度看，可归结为时间序列. 由于大多经济现象在时间序列分布上呈极强的随机性和不确定性，形成了复杂的时序问题. 揭示经济现象时序分布的不确定性行为机制呈现了极大的困难. 传统的预测方法在解决复杂的时序问题上显得无能为力，混沌理论的出现为解决这个困难提供了有力的工具. 但混沌预测所需的理论背景对于预测工作者提出了较高的要求.

预测中有两大分析方法：定性分析和定量分析. 由于定量方法需要较为全面的数据资料和较为复杂的计算，所以在预测技术发展的初期，理论上有一些阐述而应用上极为有限. 第二次世界大战以后，西方国家随着国家垄断资本主义的发展，政府干预经济的职能日益扩大，凯恩斯的国民收入和就业理论的发展以及国民经济计算机体系的创立为预测中的定量分析提供了理论和收集资料的新途径. 特别是计算机的发展，为各种信息资料的迅速整理、储存和使用以及完成各种复杂的计算提供了有力的条件，在这种形势下，预测定量方法的理论研究及应用有了迅速的发展.

预测技术有了长足的发展是一个不争的事实. 在预测中，定量方法研究与应用也在不断深入. 但对于预测中定量方法系统的、完整的阐述目前在国内的文献中尚不多见. 编撰本书的初衷是试图较为系统地、完整地从理论和应用方面介绍预测中的定量方法. 但由于预测定量方法博大精深，加之作者知识面的局限，有关资料收

集的困难，本书选取的内容远没有达到这一目标. 出版本书能达到“抛砖引玉”的效果，作者也算是聊以自慰了. 本书除了作者的部分理论研究与预测实践外，一部分内容引用了一些前辈的成果，在本书撰写的过程中得到了诸多同行的帮助，在此一并致谢！

由于作者学识有限，本书中定有缺点或疏漏，望有识之士不吝赐教！

编著者

2006 年 9 月 1 日

目　录

第1章　回归分析预测方法

在现实世界，各种变量间往往存在着相互联系、相互依存、相互制约的关系. 变量之间的这种关系一般可分为两种情况：一种是变量之间存在十分确定的关系. 例如电路中，设V表示电压、R表示电阻、I表示电流，由欧姆定律可以得到$V=IR$. 这样在三个变量中如已知两个变量的值，就可以准确算出第三个变量的值，三个变量之间的关系是确定的，这种关系称为确定性的关系. 确定性的关系其特点是指在数学上，这种变量之间的关系可以用函数来表达，一般形式上可以表示为$y=f(x_1, x_2, \cdots, x_m)$，并称$(x_1, x_2, \cdots, x_m)$为自变量，$y$为因变量. 另一种是变量之间客观存在着相互依存关系，但是变量之间的关系比较复杂，我们很难用精确的函数表达式来表达，变量之间的关系存在不确定性. 例如一般情况下，劳动生产率提高了，产量会增加，成本会下降. 劳动生产率、产量、成本之间客观存在着相互依存关系，但三者之间却不能用确定的函数关系来表示. 这种既不确定但又有联系的关系称为不确定性的关系. 从随机数学的角度来说这种不确定性的关系又称相关关系. 研究这种变量之间关系密切程度的分析称为相关分析.

相关分析通常研究的是一般变量与随机变量之间或随机变量与随机变量之间的关系. 如果在研究变量之间的关系时，将其中一些因素作为受控制的变量，而另一些随机变量作为它们的因变量，这种关系的分析称为回归分析. 由此可见，回归分析是研究变量之间相关关系的一种方法. 回归这个名词首先是由英国统计学家F. 高尔顿提出来的. 在预测中，回归分析预测方法是应用最早、理论最成熟的方法之一.

§1.1　一元线性回归预测

变量之间的相关关系从某种意义上来说，可分为线性关系和非线性关系. 在线性关系中，最简单的是一元线性相关关系. 研究回归分析预测方法一般从一元线性回归预测入手.

一、一元线性回归模型

一元线性相关关系具体表述为：设有两个变量x与y，其中x是一个可精确测量或可控制的(非随机的)通常变量，而y是一个(可观测的)随机变量. 每当变量x取一值时，变量y就有相应确定的概率分布与之对应，则称随机变量y与变量x之间有相关关系，也称x为自变量，y为因变量. 若y与x存在着线性相关关系，即当x

取固定值 $x_1,x_2,\cdots,x_n$ 时，y 有确定分布 $f_1(x_1),f_2(x_2),\cdots,f_n(x_n)$ 与之对应(如图1-1所示)，并且随机变量 y 的均值 $E(y)$ 与 x 有如下关系：

$$E(y)=u(x)=a+bx \tag{1-1}$$

称式(1-1)为 y 对 x 的理论回归直线方程.

显然式(1-1)可写成

$$y=a+bx+\varepsilon \tag{1-2}$$

其中：ε 为随机扰动项(随机误差)，且 $\varepsilon\sim N(0,\sigma^2)$；$a,b,\sigma^2$ 都是未知参数.

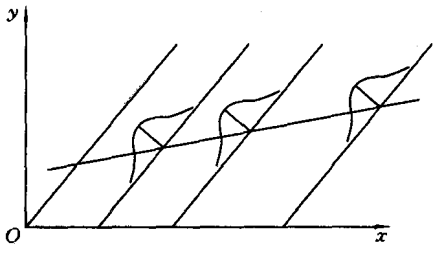

图 1-1

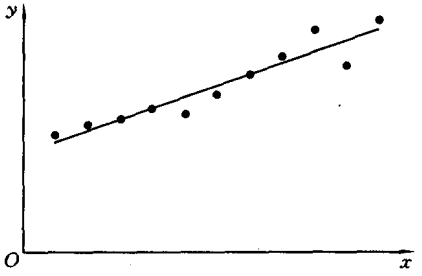

图 1-2

为了确定回归方程的类型，对给定的观测数据 $(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)$，将其描述在直角坐标系中，这种数据点在直角坐标系中的图形称为散点图(如图1-2所示). 观察它的形状，如果散点反映出直线形状，这时 $u(x)$ 为线性函数：

$$u(x)=a+bx \tag{1-3}$$

则估计 $u(x)$ 的问题称为一元线性回归问题.

由式(1-2)对具体的观测值 $(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)$，有

$$y_i=a+bx_i+\varepsilon_i \tag{1-4}$$

ε_i 与 ε 同分布，则

$$E\varepsilon_i=0 \quad (i=1,2,\cdots,n)$$

$$E\varepsilon_i\varepsilon_j=\begin{cases}\sigma^2 & (i=j)\\ 0 & (i\neq j)\end{cases}$$

式(1-4)就称为一元线性回归模型. 对于一元线性回归的研究主要涉及下述三个问题：

(1) 用观测值 (x_i,y_i) $(i=1,2,\cdots,n)$ 对式(1-2)中未知参数 a,b,σ^2 作出估计；

(2) 对 y 和 x 的线性相关关系进行显著性检验；

(3) 在 $x=x_0$ 处对 y 作预测.

二、一元线性回归的参数估计

1. 最小二乘法

对 (x,y) 进行 n 次独立观测，得到 n 个观测如下：$(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)$，

其中：x_i 表示 x 的第 i 次观测值；y_i 表示 y 的第 i 次观测值. 在直角坐标系中描述出其散点图，很显然 $y=u(x)+\varepsilon$，其中 $u(x)$ 的估计值 $\hat{u}(x)$ 的合理取法应是这样一条曲线：使得这个点从整体上看距此曲线最近，即

$$\sum_{i=1}^{n}[y_i-\hat{u}(x_i)]^2=\min_{一切u(x)}\sum_{i=1}^{n}[y_i-u(x_i)]^2 \tag{1-5}$$

由式(1-5)确定 $u(x)$ 的估计 $\hat{u}(x)$ 的方法通常称为最小二乘法.

在式(1-5)中，要找出 $\hat{u}(x)$ 是困难的，因此我们常常要根据实际问题的性质，或根据以往的经验，或根据散点图对函数 $u(x)$ 的类型先作出假定，如假定 $u(x)=a+bx$，$u(x)=b_0+b_1x+b_2x^2$，$u(x)=be^{ax}$，$u(x)=a\log_b x$，$u(x)=b\sin ax$，等等，这样一来，在函数类型确定的先决条件下，只要确定几个未知参数就可以了，即可设 $u(x)=u(x,\alpha_1,\alpha_2,\cdots,\alpha_k)$，其中 $\alpha_1,\alpha_2,\cdots,\alpha_k$ 为未知参数，为了求 $\alpha_1,\alpha_2,\cdots,\alpha_k$ 的估计，由式(1-5)最小二乘法，即取估计量 $\hat{\alpha}_1,\hat{\alpha}_2,\cdots,\hat{\alpha}_k$，使

$$\sum_{i=1}^{n}[y_i-u(x_i;\alpha_1,\alpha_2,\cdots,\alpha_k)]^2=\min\sum_{i=1}^{n}[y_i-u(x_i;\alpha_1,\alpha_2,\cdots,\alpha_k)]^2$$

此时令

$$Q=\sum_{i=1}^{n}[y_i-u(x_i;\alpha_1,\alpha_2,\cdots,\alpha_k)]^2 \tag{1-6}$$

对 Q 求最小值点 $(\hat{\alpha}_1,\hat{\alpha}_2,\cdots,\hat{\alpha}_k)$，通常用微分法来求，即用 Q 分别对 $\alpha_1,\alpha_2,\cdots,\alpha_k$ 求偏导数，并令

$$\frac{\partial Q}{\partial\alpha_1}=\frac{\partial Q}{\partial\alpha_2}=\cdots=\frac{\partial Q}{\partial\alpha_k}=0$$

得方程组

$$\begin{aligned}
\frac{\partial Q}{\partial\alpha_1}&=-2\sum_{i=1}^{n}\left\{[y_i-u(x_i;\alpha_1,\alpha_2,\cdots,\alpha_k)]\frac{\partial}{\partial\alpha_1}u(x_i;\alpha_1,\alpha_2,\cdots,\alpha_k)\right\}=0\\
\frac{\partial Q}{\partial\alpha_2}&=-2\sum_{i=1}^{n}\left\{[y_i-u(x_i;\alpha_1,\alpha_2,\cdots,\alpha_k)]\frac{\partial}{\partial\alpha_2}u(x_i;\alpha_1,\alpha_2,\cdots,\alpha_k)\right\}=0\\
&\cdots\cdots\\
\frac{\partial Q}{\partial\alpha_k}&=-2\sum_{i=1}^{n}\left\{[y_i-u(x_i;\alpha_1,\alpha_2,\cdots,\alpha_k)]\frac{\partial}{\partial\alpha_k}u(x_i;\alpha_1,\alpha_2,\cdots,\alpha_k)\right\}=0
\end{aligned} \tag{1-7}$$

可求得参数 $\alpha_1,\alpha_2,\cdots,\alpha_k$ 的估计值 $\hat{\alpha}_1,\hat{\alpha}_2,\cdots,\hat{\alpha}_k$，也就得到了 y 对 x 的(经验)回归方程

$$E(y)=u(x;\hat{\alpha}_1,\hat{\alpha}_2,\cdots,\hat{\alpha}_k)=\hat{y} \tag{1-8}$$

一般来说，求解方程组(1-7)是相当困难的，但如果 $u(x;\alpha_1,\alpha_2,\cdots,\alpha_k)$ 是 $\alpha_1,\alpha_2,\cdots,\alpha_k$ 的线性函数，问题就会简化多了，而实际问题中，$u(x;\alpha_1,\alpha_2,\cdots,\alpha_k)$ 常常是可以线性化的.

2. a,b 的最小二乘估计

对式(1-1)用最小二乘法估计 a 和 b，即设

$$Q(a,b)=\sum_{i=1}^{n}[y_i-(a+bx_i)]^2=\sum_{i=1}^{n}\varepsilon_i^2 \tag{1-9}$$

令$\frac{\partial Q}{\partial a}=\frac{\partial Q}{\partial b}=0$,得方程组

$$\begin{cases}\dfrac{\partial Q}{\partial a}=-2\sum\limits_{i=1}^{n}(y_i-a-bx_i)=0\\ \dfrac{\partial Q}{\partial b}=-2\sum\limits_{i=1}^{n}(y_i-a-bx_i)x_i=0\end{cases} \tag{1-10}$$

即

$$\begin{cases}na+n\bar{x}b=n\bar{y}\\ n\bar{x}a+\sum\limits_{i=1}^{n}x_i^2b=\sum\limits_{i=1}^{n}x_iy_i\end{cases}$$

其中

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

此方程组称为正规方程组,其系数行列式为

$$\Delta=\begin{vmatrix}n & n\bar{x}\\ n\bar{x} & \sum\limits_{i=1}^{n}x_i^2\end{vmatrix}=n\sum_{i=1}^{n}(x_i-\bar{x})^2$$

解此方程组得a,b的估计量$\hat{a},\hat{b}$为

$$\begin{cases}\hat{a}=\bar{y}-\hat{b}\bar{x}\\ \hat{b}=\dfrac{\begin{vmatrix}n & \overline{ny}\\ \overline{nx} & \sum\limits_{i=1}^{n}x_iy_i\end{vmatrix}}{n\sum\limits_{i=1}^{n}(x_i-\bar{x})^2}=\dfrac{n\left(\sum\limits_{i=1}^{n}x_iy_i-n\,\overline{xy}\right)}{n\sum\limits_{i=1}^{n}(x_i-\bar{x})^2}=\dfrac{\sum\limits_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum\limits_{i=1}^{n}(x_i-\bar{x})^2}\end{cases} \tag{1-11}$$

记

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

$$L_{xy}=\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})=\sum_{i=1}^{n}x_iy_i-n\overline{xy}$$

$$L_{yy}=\sum_{i=1}^{n}(y_i-\bar{y})^2=\sum_{i=1}^{n}y_i^2-n\bar{y}^2$$

于是式(1-11)可写成

$$\begin{cases}\hat{a}=\bar{y}-\hat{b}\bar{x}\\ \hat{b}=L_{xy}/L_{xx}\end{cases} \tag{1-12}$$

式(1-12)中的$\hat{a},\hat{b}$分别是a,b的最小二乘估计量，故

$$\hat{y}=\hat{E}(y)=\hat{u}(x)=\hat{a}+\hat{b}x \tag{1-13}$$

称为y关于x的经验线性回归方程，简称为线性回归方程，$\hat{a},\hat{b}$称为回归系数，直线$\hat{y}=\hat{a}+\hat{b}x$称为回归直线.

可以证明回归直线始终是通过点$(\bar{x},\bar{y})$的，因此有时为了计算简便，常常可利用平移坐标的方法适当地选择邻近$(\bar{x},\bar{y})$的点(x_0,y_0)作为新的坐标原点.

设$x'=x_i-x_0,y'=y_i-y_0$，则有

$$\bar{x}=\bar{x}'+x_0 \qquad \bar{y}=\bar{y}'+y_0$$

$$L_{xx}=L_{x'x'} \qquad L_{xy}=L_{x'y'} \qquad L_{yy}=L_{y'y'}$$

于是有

$$\begin{cases}\hat{b}=L_{x'y'}/L_{x'x'}\\ \hat{a}=\bar{y}+y_0-\hat{b}(\bar{x}'+x_0)\end{cases} \tag{1-14}$$

例1-1　以家庭为单位，某商品的需求量y与该商品价格x之间的一组调查数据如表1-1所示，求y对x的回归直线方程.

表1-1

商品价格x/元	1.0	2.0	2.0	2.3	2.5	2.6	2.8	3.0	3.3	3.5
商品需求量y/千克	5.0	3.0	3.5	2.7	2.4	2.5	2.0	1.5	1.2	1.2

解　从观测值的散点图(如图1-3所示)上来看，一些点分布在直线附近，作一元线性回归比较合适.

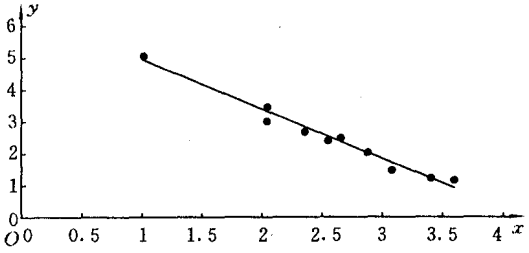

图1-3

为求一元线性回归方程$\hat{y}=\hat{a}+\hat{b}x$，将所求计算列表(如表1-2所示).

表1-2

i	x_i	y_i	x_i^2	y_i^2	x_iy_i
1	1.0	5.0	1.00	25.00	5.00
2	2.0	3.0	4.00	9.00	6.00
3	2.0	3.5	4.00	12.25	7.00

续表

i	x_i	y_i	x_i^2	y_i^2	x_iy_i
4	2.3	2.7	5.29	7.29	6.21
5	2.5	2.4	6.25	5.76	6.00
6	2.6	2.5	6.76	6.25	6.50
7	2.8	2.0	7.84	4.00	5.60
8	3.0	1.5	9.00	2.25	4.50
9	3.3	1.2	10.89	1.44	3.96
10	3.5	1.2	12.25	1.44	4.20
$\sum$	25.0	25.0	67.28	74.68	54.97

由此得

$$\bar{x}=2.5 \qquad \bar{y}=2.5 \qquad n=10$$

$$L_{xx}=\sum_{i=1}^{10}x_i^2-n\bar{x}^2=67.28-2.5^2=4.78$$

$$L_{xy}=\sum_{i=1}^{10}x_iy_i-n\bar{x}\bar{y}=54.97-10\times2.5\times2.5=-7.53$$

$$L_{yy}=\sum_{i=1}^{10}y_i^2-n\bar{y}^2=74.68-10\times25^2=12.18$$

$$\hat{b}=L_{xy}/L_{xx}=-\frac{7.53}{4.78}=-1.58$$

$$\hat{a}=\bar{y}-\hat{b}\bar{x}=2.5-(-1.58)\times25=6.45$$

故所求回归方程

$$\hat{y}=6.45-1.58x$$

这里回归系数$\hat{b}=-1.58$表示商品的价格每增加1元，该商品的需求量平均减少1.58千克.

3. 分解公式及σ^2的估计

下面来讨论回归分析中具有重要意义的分解公式，进而给出σ^2的估计值. 对于任意n组数据$(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)$，恒有

$$\sum_{i=1}^{n}(y_i-\bar{y})^2=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2+\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2 \tag{1-15}$$

其中：$\hat{y}_i=\hat{a}+\hat{b}x_i$ $(i=1,2,\cdots,n)$.

我们将y_i与其均值$\bar{y}$之间的差称为离差，将离差分解为

$$y_i-\bar{y}=(y_i-\hat{y}_i)+(\hat{y}_i-\bar{y})$$

故

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

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

又

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

$$=\sum_{i=1}^{n}[(y_i-\bar{y})-\hat{b}(x_i-\bar{x})]\hat{b}(x_i-\bar{x})$$

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

$$=\hat{b}(L_{xy}-\hat{b}L_{xx})=0$$

故

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

式(1-15)称为平方和分解式. 该式中三个平方和的意义如下：$L_{yy}=\sum_{i=1}^{n}(y_i-\bar{y}_i)^2$ 是 $y_1,y_2,\cdots,y_n$ 这 n 个数据的离差平方和，它的大小描述了这 n 个数据的分散程度，称总离差平方和，记为 $Q_{总}$. 注意到

$$\bar{\hat{y}}=\frac{1}{n}\sum_{i=1}^{n}\hat{y}_i=\frac{1}{n}\sum_{i=1}^{n}(\hat{a}+\hat{b}x_i)=\hat{a}+\hat{b}\bar{x}=\bar{y}$$

即 $\hat{y}_1,\hat{y}_2,\cdots,\hat{y}_n$ 这 n 个数的平均值也是 $\bar{y}$，所以 $\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2$ 就是 $\hat{y}_1,\hat{y}_2,\cdots,\hat{y}_n$ 这 n 个数的离差平方和，它反映了 $\hat{y}_1,\hat{y}_2,\cdots,\hat{y}_n$ 的分散程度. 又由于

$$\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2=\sum_{i=1}^{n}[(\hat{a}+\hat{b}x_i)-(\hat{a}+\hat{b}\bar{x})]^2=\hat{b}^2\sum_{i=1}^{n}(x_i-\bar{x})^2=\hat{b}^2L_{xx}$$

所以我们说 $\hat{y}_1,\hat{y}_2,\cdots,\hat{y}_n$ 的分散性来源于 $x_1,x_2,\cdots,x_n$ 的分散性，通过 x 对 y 的线性相关性反映出来，为此称 $\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2$ 为回归平方和，记为 $Q_{回}$. $\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$，就是 $Q(a,b)$，它反映了观测值 y_i 偏离回归直线的程度，称剩余平方和或残差平方和，记为 $Q_{残}$. $Q_{残}$ 是除了 x 对 y 的线性影响之外的其他因素如实验误差、观测误差等随机因素所造成的离差平方和. 这样式(1-15)可写成

$$Q_{总}=Q_{残}+Q_{回} \tag{1-16}$$

显然 $Q_{回}/Q_{残}$ 较大，则表明 x 对 y 的线性影响也较大，可以认为 x 与 y 之间有线性相

关性，反之，没有理由认为 x 与 y 之间有线性相关关系.

在分解公式的基础上可以证明 $\hat{\sigma}^2$ 的估计值

$$\hat{\sigma}^2 = Q_{残}/(n-2) \tag{1-17}$$

4. 估计量 $\hat{b},\hat{a}$ 和 $\hat{\sigma}^2$ 的统计性质

由式(1-12)知 $\hat{b}=L_{xy}/L_{xx}$，$\hat{a}=\bar{y}-\hat{b}x$ 分别是 b,a 的最小二乘估计量，由式(1-17)给出 $\hat{\sigma}^2=Q_{残}/(n-2)$ 是 σ^2 的估计量. 从统计的角度来说 $\hat{b},\hat{a},\hat{\sigma}^2$ 分别是 b,a,σ^2 的无偏估计量. 事实上，由于

$$E(\hat{b}) = E(L_{xy}/L_{xx}) = E(L_{xy})/L_{xx} = \frac{1}{L_{xx}}E\Big[\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})\Big]$$

$$= \frac{1}{L_{xx}}\sum_{i=1}^{n}(x_i-\bar{x})E(y_i-\bar{y})$$

又

$$E(y_i) = E(a+bx_i+\varepsilon_i) = a+bx_i$$

$$E(\bar{y}) = E\Big(\frac{1}{n}\sum_{i=1}^{n}y_i\Big) = \frac{1}{n}E\Big(\sum_{i=1}^{n}y_i\Big) = \frac{1}{n}\Big(\sum_{i=1}^{n}y_i\Big) = a+b\bar{x}$$

故

$$E(\hat{b}) = \frac{1}{L_{xx}}\sum_{i=1}^{n}(x_i-\bar{x})[(a+bx_i)-(a+b\bar{x})]$$

$$= \frac{1}{L_{xx}}\sum_{i=1}^{n}(x_i-\bar{x})b(x_i-\bar{x}) = b \tag{1-18}$$

$$E(\hat{a}) = E(\bar{y}-\hat{b}\bar{x}) = E(\bar{y}) - \bar{x}E(\hat{b}) = a+b\bar{x}-\bar{x}b = a \tag{1-19}$$

可见 $\hat{b},\hat{a}$ 分别是 b,a 的无偏估计. 又 $y_1,y_2,\cdots,y_n$ 相互独立，且 $y_i\sim N(a+bx_i,\sigma^2)$ $(i=1,2,\cdots,n)$，所以 $\hat{b},\hat{a}$ 均服从正态分布.

因为

$$D(\hat{b}) = D(L_{xy}/L_{xx}) = D\Big[\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})\Big/\sum_{i=1}^{n}(x_i-\bar{x})^2\Big]$$

$$= \frac{1}{L_{xx}^2}\sum_{i=1}^{n}D((x_i-\bar{x})y_i) = \frac{1}{L_{xx}^2}\sum_{i=1}^{n}(x_i-\bar{x})^2D(y_i) = \frac{\sigma^2}{L_{xx}}$$

故

$$\hat{b}\sim N\Big(b,\frac{\sigma^2}{L_{xx}}\Big) \tag{1-20}$$

同理

$$\hat{a}\sim N\left(a,\frac{\sigma^2\sum_{i=1}^{n}x_i^2}{nL_{xx}}\right) \tag{1-21}$$

由于

$$Q_{回}=\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2=\hat{b}^2\sum_{i=1}^{n}(x_i-\bar{x})^2=\hat{b}^2L_{xx}$$

而

$$E(\hat{b}^2)=D(\hat{b})+[E(\hat{b})]^2=\frac{\sigma^2}{L_{xx}}+b^2$$

于是

$$E(Q_{回})=E(\hat{b}^2)L_{xx}=\sigma^2+b^2L_{xx} \tag{1-22}$$

由式(1-22)可看出$Q_{回}$的大小不仅与σ^2有关，而且与b^2的大小有关. 因为$y_i\sim N(a+bx_i,\sigma^2)$ $(i=1,2,\cdots,n)$且相互独立，故

$$E(y_i^2)=D(y_i)+[E(y_i)]^2=\sigma^2+(a+bx_i)^2$$

且

$$E(\bar{y}^2)=D(\bar{y})+[E(\bar{y})]^2=\frac{1}{n^2}\sum_{i=1}^{n}D(y_i)+(a+b\bar{x})^2=\frac{1}{n}\sigma^2+(a+b\bar{x})^2$$

故

$$\begin{aligned}
E(Q_{总})&=E\Big[\sum_{i=1}^{n}(y_i-\bar{y})^2\Big]=E\Big[\sum_{i=1}^{n}y_i^2-n\bar{y}^2\Big]=\sum_{i=1}^{n}E(y_i^2)-nE(\bar{y}^2)\\
&=\sum_{i=1}^{n}[\sigma^2+(a+bx_i)^2]-n\Big[\frac{1}{n}\sigma^2+(a+b\bar{x})^2\Big]\\
&=n\sigma^2+\sum_{i=1}^{n}(a^2+2abx_i+b^2x_i^2)-\sigma^2-(na^2+2nab\bar{x}+nb^2\bar{x}^2)\\
&=n\sigma^2+na^2+2nab\bar{x}+b^2\sum_{i=1}^{n}x_i^2-\sigma^2-na^2-2nab\bar{x}-nb^2\bar{x}^2\\
&=(n-1)\sigma^2+b^2\sum_{i=1}^{n}x_i^2-b^2n\bar{x}^2\\
&=(n-1)\sigma^2+b^2\Big[\sum_{i=1}^{n}(x_i)^2-2\bar{x}\sum_{i=1}^{n}x_i+n\bar{x}^2\Big]\\
&=(n-1)\sigma^2+b^2\sum_{i=1}^{n}(x_i-\bar{x})^2\\
&=(n-1)\sigma^2+b^2L_{xx}
\end{aligned}$$

综上可知

$$\begin{aligned}
E(Q_{残})&=E(Q_{总})-E(Q_{回})\\
&=(n-1)\sigma^2+b^2L_{xx}-\sigma^2-b^2L_{xx}\\
&=(n-2)\sigma^2
\end{aligned}$$

故

$$\sigma^2=E(Q_{残})/(n-2) \tag{1-23}$$

这表明 $\hat{\sigma}^2=Q_{残}/(n-2)$ 是 σ^2 的无偏估计，且 $Q_{残}$ 的取值只与 σ^2 的大小有关，而与 b^2 无关.

另外，由于 $\hat{a},\hat{b}$ 是正规方程的解，满足 $\sum_{i=1}^{n}(y_i-\hat{y}_i)=0$，$\sum_{i=1}^{n}(y_i-\hat{y}_i)x_i=0$，所以 $\hat{\sigma}^2$ 的自由度是 $n-2$，可以证明

$$\frac{(n-2)\hat{\sigma}^2}{\sigma^2}=\frac{Q_{残}}{\sigma^2}\sim\chi^2(n-2) \tag{1-24}$$

三、一元线性回归的显著性检验

在求回归方程的计算过程中，并不需要事先假定 x 与 y 之间具有线性关系，即不管 (x,y) 在坐标平面上多么杂乱无章，总可以给它配一条回归直线 $\hat{y}=\hat{a}+\hat{b}x$，显然所配直线是否有意义，即 y 与 x 是否确有线性关系需进一步进行检验. 下面来讨论如何检验 $y=a+bx+\varepsilon,\varepsilon\sim N(0,\sigma^2)$ 这一假设是否合适，亦即检验 $H_0:b=0$，$H_1:b\neq0$ 的真假问题.

1. 一元线性回归的 F 检验

总离差平方和 $Q_{总}$ 的自由度 $f_{总}=n-1$，回归平方和 $Q_{回}$ 是由 1 个普通变量 x 对 y 的线性影响决定的，所以它的自由度 $f_{回}=1$，前述已知残差平方和 $Q_{残}$ 的自由度为 $f_{残}=n-2$. 所以有

$$f_{总}=f_{回}+f_{残} \tag{1-25}$$

在 H_0 为真的前提条件下，前述已知 $\frac{Q_{残}}{\sigma^2}\sim\chi^2(n-2)$，可以证明 $\frac{Q_{回}}{\sigma^2}\sim\chi^2(1)$. 由于 $Q_{回}$ 与 $Q_{残}$ 相互独立，由 F 分布的定义知 F 统计量

$$F=\frac{Q_{回}/\sigma^2}{\frac{Q_{残}}{\sigma^2}/(n-2)}=\frac{Q_{回}}{Q_{残}/(n-2)}\sim F(1,n-2) \tag{1-26}$$

由 $E(Q_{回})=\sigma^2+b^2L_{xx}$ 知在 H_0 为真的前提条件下，$Q_{回}$ 是 σ^2 的无偏估计；在 H_0 不真的前提条件下，$Q_{回}$ 的期望大于 σ^2，但不管对 b 的假设如何，$Q_{残}/(n-2)$ 都是 σ^2 的无偏估计，这说明在 H_0 不真时，比值 F 有偏大的倾向，可见此统计量对假设 H_0 真与否是敏感的.

给定显著性水平 α 查 F 分布表，由样本值计算 F 统计量的值 F，如果 $F\geqslant F_\alpha$ 则拒绝 H_0，认为回归效果显著，反之 $F<F_\alpha$ 则接受 H_0 认为回归效果不显著.

2. 一元线性回归的 t 检验

由 $\hat{b}\sim N\left(b,\frac{\sigma^2}{L_{xx}}\right)$ 知 $\frac{\hat{b}-b}{\sigma\sqrt{\frac{1}{L_{xx}}}}\sim N(0,1)$. 又由式(1-24)及 t 分布的定义统计量