

硕士研究生公共课教材



多元统计与SAS应用

(第二版)

肖枝洪 余家林 编著



WUHAN UNIVERSITY PRESS

武汉大学出版社

硕士研究生公共课教材

多元统计与SAS应用

(第二版)

肖枝洪 余家林 编著



WUHAN UNIVERSITY PRESS

武汉大学出版社

图书在版编目(CIP)数据

多元统计与 SAS 应用/肖枝洪,余家林编著. —2 版. —武汉: 武汉大学出版社, 2013. 9

硕士研究生公共课教材

ISBN 978-7-307-11781-5

I. 多… II. ①肖… ②余… III. ①多元分析: 统计分析—研究生—教材 ②统计分析—统计程序—研究生—教材 IV. ①O212.4 ②C819

中国版本图书馆 CIP 数据核字(2013)第 222278 号

责任编辑: 顾素萍

责任校对: 黄添生

版式设计: 马 佳

出版发行: 武汉大学出版社 (430072 武昌 珞珈山)

(电子邮件: cbs22@whu.edu.cn 网址: www.wdp.com.cn)

印刷: 武汉中科兴业印务有限公司

开本: 720 × 1000 1/16 印张: 14.75 字数: 260 千字 插页: 1

版次: 2008 年 1 月第 1 版 2013 年 9 月第 2 版

2013 年 9 月第 2 版第 1 次印刷

ISBN 978-7-307-11781-5

定价: 28.00 元

版权所有, 不得翻印; 凡购我社的图书, 如有质量问题, 请与当地图书销售部门联系调换。

第二版前言

《多元统计与 SAS 应用》第一版自 2008 年发行以来，得到了许多读者的使用和赞扬，而且现已脱销。应广大读者的要求，我们在第一版的基础上进行修订后再出第二版。

第二版与第一版相比，基本框架不变，只是在部分章节上做了微调。本次修订了我们在教学中发现以及读者反映出来的问题。

感谢为本次修订提出宝贵意见的教师和读者。感谢武汉大学出版社的大力支持！

由于作者水平有限，书中的错误和疏漏之处可能还是存在，敬请读者提出宝贵意见，以便进一步修订和改进！

肖枝洪

于重庆理工大学花溪校区

2013 年 7 月

第一版前言

多元统计是非数学专业硕士研究生教学计划中普遍开设的一门公共基础课，各学校各专业讲授的内容大体一致。随着硕士研究生入学水平与课题研究水平的提高，急需一本相适应的教材，既能加强理论基础、帮助研究生熟悉多元统计原理，又能介绍近代流行的统计分析软件，使研究生在处理试验数据的过程中摆脱复杂计算的困扰。

由我们合编的《多元统计及 SAS 应用》是近几年来硕士研究生优质课程立项研究的一项成果。作为非数学专业硕士研究生的教材，编入了多元线性回归、多元线性相关、多元非线性回归、回归的试验设计与分析、聚类分析、判别分析、主成分分析、因子分析及 SAS 的应用等内容。讲课及上机实习可控制在 60 课时以内。

在编写中，我们特别注意说明统计方法的实际背景，详细讲述用统计方法解决实际问题的思路，对于应用 Statistical Analysis System（简称 SAS）所得到的统计分析结果，则尽可能地与实际计算步骤一一对照，使初学者能够知其所以然。考虑到专业与课时设置的不同，本教材力求简明扼要，重点突出，通俗易懂，便于自学，例题与习题都在常识的范围之内。

教材中第一章、第二章、第三章由余家林编写，第四章、第五章、第六章由肖枝洪编写。本教材的出版得到华中农业大学研究生处及武汉大学出版社的大力支持，在此表示衷心的感谢。由于编者的水平所限，不妥之处难以避免，敬请读者和使用本教材的同行学友批评指正。

编 者

2007 年 10 月 9 日

目 录

第一章

多元线性回归	1
1.1 一元线性回归	1
1.1.1 一元线性回归的概念	1
1.1.2 一元线性回归参数的确定	2
1.1.3 一元线性回归的矩阵表示	2
1.1.4 回归方程的显著性检验	4
1.1.5 相关系数与决定系数	6
1.1.6 一元线性回归方程的应用	7
1.1.7 一元线性回归的实例	8
1.1.8 应用 SAS 作一元线性回归	10
1.2 多元线性回归	15
1.2.1 多元线性回归的概念	15
1.2.2 多元线性回归的矩阵表示	16
1.2.3 回归方程的显著性检验	19
1.2.4 回归系数的显著性检验	20
1.2.5 标准回归方程及其显著性检验	21
1.2.6 多元线性回归方程的应用	26
1.2.7 多元线性回归的实例	26
1.2.8 应用 SAS 作多元线性回归	31
1.3 回归方程的比较, 逐步回归及复共线性	36
1.3.1 回归方程比较的目的	36
1.3.2 常用的比较标准	36
1.3.3 比较标准应用的实例	37
1.3.4 应用 SAS 求所有可能的回归方程并进行比较	40

1.3.5	逐步回归的基本思想	40
1.3.6	逐步回归的实例	42
1.3.7	应用 SAS 作逐步回归	44
1.3.8	复共线性与逐步回归	46

第二章

多元线性相关	53
2.1 多个变量的线性相关	53
2.1.1 简单线性相关	53
2.1.2 复线性相关	54
2.1.3 偏线性相关	55
2.1.4 三种相关系数的临界值表	58
2.1.5 三种相关系数的实例	59
2.1.6 应用 SAS 计算三种相关系数	62
2.1.7 通径系数及通径分析表	62
2.1.8 应用 SAS 计算通径系数	65
2.2 两组变量的线性相关	65
2.2.1 典型变量及典型相关系数	65
2.2.2 典型相关分析原理	66
2.2.3 典型相关系数的特例	69
2.2.4 典型变量的计算步骤	70
2.2.5 典型相关分析的实例	71
2.2.6 应用 SAS 作典型相关分析	75

第三章

多元非线性回归	78
3.1 非线性回归方程的建立	78
3.1.1 “线性化”方法	78
3.1.2 非线性回归方程拟合情况的比较	79
3.1.3 非线性最小二乘法	82
3.1.4 应用 SAS 作曲线回归	85
3.1.5 Logistic 曲线回归	85

3.1.6	多项式回归	88
3.2	一次回归的正交设计	91
3.2.1	回归设计简介	91
3.2.2	一次回归正交设计的步骤	93
3.2.3	回归系数的计算及显著性检验	94
3.2.4	零水平处的重复试验	96
3.2.5	在回归方程中引入交互效应项	96
3.2.6	一次回归正交设计的实例	97
3.3	二次回归的正交组合设计	100
3.3.1	什么是组合设计	100
3.3.2	平方项中心化及选择星号臂的意义	102
3.3.3	二次回归正交组合设计的步骤	104
3.3.4	正交组合设计的 m_0 及 γ^2 值略表	105
3.3.5	回归系数的计算及显著性检验	106
3.3.6	回归方程的失拟性检验	106
3.4	二次回归的旋转组合设计	107
3.4.1	什么是旋转设计	107
3.4.2	旋转性条件及非退化条件	108
3.4.3	二次回归组合设计的旋转性	110
3.4.4	二次回归旋转组合设计的正交性	110
3.4.5	二次回归正交旋转组合设计的实例	112
3.4.6	应用 SAS 建立正交旋转组合设计的回归方程	117
3.4.7	二次回归旋转组合设计的通用性	117
3.4.8	二次回归通用旋转组合设计的实例	120
3.4.9	应用 SAS 建立通用旋转组合设计的回归方程	122

第四章

多元聚类与判别	126
4.1 聚类的根据	126
4.1.1 观测数据矩阵	126
4.1.2 Q 型聚类的相似性统计量	127
4.1.3 R 型聚类的相似性统计量	128

4.1.4	聚类方法概述	129
4.2	系统聚类法	129
4.2.1	系统聚类法的基本思想	129
4.2.2	最短距离法(single linkage method)	130
4.2.3	最长距离法(complete linkage method)	133
4.2.4	中间距离法(median method)	136
4.2.5	重心法(centroid method)	138
4.2.6	类平均法(average linkage method)	141
4.2.7	离差平方和法(ward's mininum-variance method)	144
4.2.8	类的个数	147
4.2.9	系统聚类法小结	149
4.2.10	应用 SAS 作系统聚类	149
4.3	逐步聚类法	152
4.3.1	逐步聚类法的基本思想	152
4.3.2	成批调整法	152
4.3.3	成批调整法的 SAS 程序	154
4.3.4	离差的平方和法	157
4.4	Bayes 判别	163
4.4.1	Bayes 判别的原理	163
4.4.2	Bayes 判别的任务	164
4.4.3	正态假设下判别函数的建立	164
4.4.4	多个变量全体判别效果的检验	166
4.4.5	各变量判别能力的检验	167
4.4.6	Bayes 判别的步骤	167
4.4.7	Bayes 判别的实例	168
4.4.8	用 SAS 作 Bayes 判别	170
4.5	逐步判别	172
4.5.1	逐步判别的基本思想	172
4.5.2	逐步判别的步骤	172
4.5.3	逐步判别的实例	173
4.5.4	用 SAS 作逐步判别	176

第五章

多元试验数据的主成分分析	181
5.1 主成分分析法	181
5.1.1 什么是主成分分析	181
5.1.2 主成分分析的任务	182
5.1.3 主成分分析的原理	183
5.1.4 主成分分析的计算步骤	185
5.1.5 主成分分析的实例	185
5.1.6 用 SAS 作主成分分析	189
5.2 主成分的应用	189
5.2.1 构成综合指标	190
5.2.2 主成分聚类	190
5.2.3 主成分回归	191

第六章

多元试验数据的因子分析	194
6.1 因子分析法	194
6.1.1 什么是因子分析	194
6.1.2 因子分析的任务	195
6.1.3 因子分析的基本定理	197
6.1.4 因子分析的计算步骤	199
6.1.5 公因子得分	199
6.1.6 因子分析的实例	200
6.1.7 用 SAS 作因子分析	203
6.2 方差极大正交旋转	204
6.2.1 方差极大正交旋转的概念	204
6.2.2 正交旋转角度的计算	206
6.2.3 方差极大正交旋转的实例	206
6.2.4 正交旋转后公因子的应用	209
6.3 对应分析法	211
6.3.1 什么是对应分析	211
6.3.2 对应分析的任务	212

6.3.3	对应分析的原理.....	212
6.3.4	对应分析的计算步骤.....	216
6.3.5	对应分析的实例.....	217
6.3.6	用 SAS 作对应分析	221
参考文献.....		224

多元线性回归

多元线性回归是一元线性回归的发展, 可用来研究因变量取值与自变量取值的内在联系, 建立多元线性回归方程. 在讲述多元线性回归的计算及其应用之前, 为了承上启下, 在 1.1 节中对一元线性回归的计算及其应用作了回顾. 熟悉这些内容的读者, 不妨自 1.2 节开始读起.

1.1 一元线性回归

1.1.1 一元线性回归的概念

设自变量 x 的观测值 x_i 及因变量 y 对应的观测值 y_i 满足关系式

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, 2, \dots, n,$$

式中, $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ 是相互独立且都服从正态分布 $N(0, \sigma^2)$ 的随机变量.

根据最小二乘法, 由 n 组观测值 (x_i, y_i) 确定参数 β_0 及 β_1 的估计值 b_0 及 b_1 后, 所得到的估计式 $\hat{y} = b_0 + b_1 x$ 称为一元线性回归方程. 建立一元线性回归方程的过程以及对回归方程所作的显著性检验, 称为一元线性回归分析或一元线性回归.

如果将 x_i 代入一元线性回归方程, 记 $\hat{y}_i = b_0 + b_1 x_i$, 则 $\hat{y}_i$ 与 y_i 之间的偏差平方和

$$Q = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - b_0 - b_1 x_i)^2.$$

由 $\frac{\partial Q}{\partial b_0} = 0$ 及 $\frac{\partial Q}{\partial b_1} = 0$ 可得到方程组

$$\begin{cases} nb_0 + b_1 \sum_{i=1}^n x_i = \sum_{i=1}^n y_i, \\ b_0 \sum_{i=1}^n x_i + b_1 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i. \end{cases}$$

解这个方程组,即可算出 b_0 及 b_1 . 根据最小二乘法, b_0 及 b_1 的值使上述偏差平方和 Q 取最小值. 称这个方程组为一元线性回归的正规方程组, b_0 为回归常数或截距, b_1 为回归系数.

注:前面曾假设 $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ 相互独立且都服从正态分布 $N(0, \sigma^2)$. 在建立回归方程的过程中,这两个假设都没有用到. 在对回归方程作显著性检验或进行区间预测时,将根据这两个假设导出检验统计量的分布.

1.1.2 一元线性回归参数的确定

由 n 组观测值 (x_i, y_i) 确定参数 β_0 及 β_1 的估计值 b_0 及 b_1 是一元线性回归的关键. 根据一元线性回归的正规方程组可以导出

$$b_1 = \frac{l_{xy}}{l_{xx}}, \quad b_0 = \bar{y} - b_1 \bar{x},$$

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

$$l_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n x_i y_i - \frac{1}{n} \sum_{i=1}^n x_i \sum_{i=1}^n y_i,$$

$$l_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2.$$

称 l_{xy} 为 x 与 y 的离均差乘积和, l_{xx} 为 x 的离均差平方和.

记 $l_{yy} = \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n y_i^2 - \frac{1}{n} \left(\sum_{i=1}^n y_i \right)^2$, 则 l_{yy} 为 y 的离均差平方和.

进一步,由正规方程组的第一个方程可以导出

$$\sum_{i=1}^n (b_0 + b_1 x_i) = \sum_{i=1}^n y_i \quad \text{及} \quad b_0 + b_1 \bar{x} = \bar{y}.$$

因此有结论:

$$\textcircled{1} \quad \sum_{i=1}^n \hat{y}_i = \sum_{i=1}^n y_i, \quad \frac{1}{n} \sum_{i=1}^n \hat{y}_i = \bar{y};$$

$$\textcircled{2} \quad \text{当 } x = \bar{x} \text{ 时, } \hat{y} = \bar{y}.$$

这说明,将 x 的 n 个观测值 x_i 代入回归方程所得到的 n 个估计值 $\hat{y}_i$ 的平均值等于 $\bar{y}$, 将 $\bar{x}$ 代入回归方程所得到的估计值 $\hat{y}$ 也等于 $\bar{y}$.

1.1.3 一元线性回归的矩阵表示

作一元线性回归时,自变量 x 及因变量 y 的观测值 x_i 及 y_i 所满足的关系式

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, 2, \dots, n$$

又称为一元线性回归模型.

若记

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix},$$

则上述模型的矩阵表示为 $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, 且

$$E(\boldsymbol{\varepsilon}) = \begin{bmatrix} E(\varepsilon_1) \\ E(\varepsilon_2) \\ \vdots \\ E(\varepsilon_n) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \mathbf{0}, \quad \text{Cov}(\boldsymbol{\varepsilon}) = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} = \sigma^2 \mathbf{I}.$$

正规方程组的矩阵表示为

$$\begin{bmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i y_i \end{bmatrix},$$

其中

$$\begin{bmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{bmatrix} = \mathbf{X}'\mathbf{X}, \quad \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i y_i \end{bmatrix} = \mathbf{X}'\mathbf{y}.$$

若记 $\mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}$, 则正规方程组可进一步用矩阵表示为 $\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{y}$, 当 $\mathbf{X}'\mathbf{X}$

的逆矩阵存在时, 正规方程组的解 $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$, 式中,

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} \frac{\sum_{i=1}^n x_i^2}{nl_{xx}} & -\frac{\sum_{i=1}^n x_i}{nl_{xx}} \\ -\frac{\sum_{i=1}^n x_i}{nl_{xx}} & \frac{1}{l_{xx}} \end{bmatrix}.$$

在统计分析软件 SAS 的输出中, 将正规方程组的增广矩阵

$$(X'X, X'y) \quad \text{或} \quad \begin{pmatrix} n & \sum_{i=1}^n x_i & \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i y_i \end{pmatrix}$$

表示为下列形式的加边增广矩阵:

$$\begin{pmatrix} X'X & X'y \\ y'X & y'y \end{pmatrix} \quad \text{或} \quad \begin{pmatrix} n & \sum_{i=1}^n x_i & \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i y_i \\ \sum_{i=1}^n y_i & \sum_{i=1}^n x_i y_i & \sum_{i=1}^n y_i^2 \end{pmatrix},$$

将 $(X'X)^{-1}$, $b = (X'X)^{-1}X'y$ 及 SSE (即剩余平方和, 其定义见 1.1.4 节) 表示为矩阵

$$\begin{pmatrix} (X'X)^{-1} & b \\ b' & \text{SSE} \end{pmatrix}.$$

1.1.4 回归方程的显著性检验

离均差平方和 $l_{yy} = \sum_{i=1}^n (y_i - \bar{y})^2$ 表示 n 个观测值 y_i 之间的差异. 当各个 y_i 已知时, l_{yy} 是一个定值, 作回归方程的显著性检验时, 称它为总平方和, 也记为 SST 或 SS_{tot} .

以下证明: $\text{SST} = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$. 因为

$$\begin{aligned} \sum_{i=1}^n (y_i - \bar{y})^2 &= \sum_i (y_i - \hat{y}_i + \hat{y}_i - \bar{y})^2 \\ &= \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \\ &\quad + 2 \sum_{i=1}^n (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}), \end{aligned}$$

最后面的一项可写为

$$\begin{aligned}
& 2 \sum_{i=1}^n (y_i - b_0 - b_1 x_i)(b_0 + b_1 x_i - \bar{y}) \\
&= 2 \sum_{i=1}^n (y_i - \bar{y} + b_1 \bar{x} - b_1 x_i)(\bar{y} - b_1 \bar{x} + b_1 x_i - \bar{y}) \\
&= 2b_1 \sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x}) - 2b_1^2 \sum_{i=1}^n (x_i - \bar{x})^2 \\
&= 2b_1 (l_{xy} - b_1 l_{xx}) \\
&= 0,
\end{aligned}$$

因此,

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

式中, $\sum_{i=1}^n (y_i - \hat{y}_i)^2$ 是 y_i 与 $\hat{y}_i$ 之间的偏差平方和, 通过回归已经达到了最

小值, 称 $\sum_{i=1}^n (y_i - \hat{y}_i)^2$ 为剩余平方和, 记为 SSE 或 SS_{res} .

而 $\sum_{i=1}^n (\hat{y}_i - \bar{y})^2$ 表示 n 个 $\hat{y}_i$ 之间的差异, 是将 x_i 代入回归方程得到 $\hat{y}_i$ 造成

的, 称 $\sum_{i=1}^n (\hat{y}_i - \bar{y})^2$ 为回归平方和, 记为 SSR 或 SS_{reg} .

由等式 $SST = SSE + SSR$ 可以对 SSR 的意义作下列分析:

如果 SSR 的数值较大, SSE 的数值便比较小, 说明回归的效果好. 如果 SSR 的数值较小, SSE 的数值便比较大, 说明回归的效果差.

根据对 $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ 中的 ϵ_i 所作的两个假设可以证明:

当原假设 H_0 为 $\beta_1 = 0$ 并且 H_0 成立时,

$$\frac{SST}{\sigma^2} \sim \chi^2(n-1), \quad \frac{SSR}{\sigma^2} \sim \chi^2(1), \quad \frac{SSE}{\sigma^2} \sim \chi^2(n-2),$$

且 SSR 与 SSE 相互独立, $F = \frac{SSR/1}{SSE/(n-2)} \sim F(1, n-2)$, $\hat{\sigma}^2 = MSE = \frac{SSE}{n-2}$ 为 σ^2 的无偏估计量.

因此, 给出显著性水平 α , 将 F 与 $F_\alpha(1, n-2)$ 进行比较, 当 $F > F_\alpha$ 时放弃 H_0 , 称回归方程显著; 否则接受 H_0 , 称回归方程不显著.

注: 对回归方程作显著性检验的基本思想与方法类似于方差分析, 在 SAS 输出的结果中检验的过程与结果将用方差分析表来显示.

计算 SSR 及 SSE 的公式为

$$\text{SSR} = b_1 l_{xy}, \quad \text{SSE} = l_{yy} - \text{SSR}.$$

这里,

$$\begin{aligned} \text{SSR} &= \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = \sum_{i=1}^n (b_0 + b_1 x_i - b_0 - b_1 \bar{x})^2 \\ &= b_1^2 \sum_{i=1}^n (x_i - \bar{x})^2 = b_1^2 l_{xx} = b_1 l_{xy}. \end{aligned}$$

1.1.5 相关系数与决定系数

由 SSR, SSE 及 b_1 的计算公式可推出

$$\text{SSE} = l_{yy} \left(1 - b_1 \frac{l_{xy}}{l_{yy}} \right) = l_{yy} \left(1 - \frac{l_{xy}^2}{l_{xx} l_{yy}} \right).$$

若记 $r = \frac{l_{xy}}{\sqrt{l_{xx} l_{yy}}}$, 则

$$\text{SSE} = l_{yy} (1 - r^2), \quad \text{SSR} = r^2 l_{yy}, \quad l_{xy} = r \sqrt{l_{xx} l_{yy}}.$$

因此, 当 $|r|$ 大时, SSE 小, SSR 大, 变量 x 与 y 的线性关系密切; 当 $|r|$ 小时, SSE 大, SSR 小, 变量 x 与 y 的线性关系不密切.

当 $r > 0$ 时, $b_1 > 0$, $\hat{y}$ 随 x 的增加而增加, x 与 y 的线性相关关系为正相关; 当 $r < 0$ 时, $b_1 < 0$, $\hat{y}$ 随 x 的增加而减少, x 与 y 的线性相关关系为负相关.

称 r 为变量 x 与 y 的相关系数.

至于 r^2 也有很重要的实际意义. 根据

$$r^2 = \frac{l_{xy}^2}{l_{xx} l_{yy}} = \frac{b_1 l_{xy}}{l_{yy}} = \frac{\text{SSR}}{\text{SST}},$$

可以将 r^2 解释为 SSR 在 SST 中所占的比率, 也就是 SST 中可以用线性关系来说明的部分在 SST 中所占的比率.

称 r^2 为变量 x 与 y 的决定系数.

对相关系数作显著性检验时, 可以由

$$F = \frac{\text{SSR}}{\text{SSE}/(n-2)} = \frac{r^2}{(1-r^2)/(n-2)}$$

作 F 检验.

也可以先查相关系数检验专用的临界值, 再将 $|r|$ 与临界值进行比较, 然后作出 r 是否显著的结论.

$|r|$ 的临界值是将上述统计量变形为 $|r| = \sqrt{\frac{F}{F + (n-2)}}$ 后, 将 F 检验的