线性模型分析原理

朱 军 著

科学出版社

线性模型分析原理

朱　军著

科学出版社

2000

内 容 简 介

本书系统介绍线性模型统计分析的一些新进展，主要内容包括：简单和多元回归模型分析、非线性模型分析、方差分析、混合线性模型分析等.

本书可供大学数理统计相关专业的研究生和教师阅读，也可供统计学、数量遗传学、农学等方面的科技人员阅读和参考.

图书在版编目(CIP)数据

线性模型分析原理/朱 军著.-北京：科学出版社，1999.4

ISBN 7-03-007092-5

Ⅰ.线… Ⅱ.朱… Ⅲ.线性模型-统计分析（数学）Ⅳ.O212.1

中国版本图书馆 CIP 数据核字（98）第 32393 号

科学出版社 出版

北京东黄城根北街 16 号

邮政编码：100717

北京双青印刷厂 印刷

新华书店北京发行所发行 各地新华书店经售

*

1999 年 4 月第 一 版 开本：850×1168 1/32

2000 年 2 月第二次印刷 印张：7

印数：3 001—5 000 字数：183 000

定价：18.00 元

(如有印装质量问题，我社负责调换(环伟))

序　言

80年代初，我在浙江农业大学农学系攻读硕士学位时，有幸聆听周承钥先生的生物统计课，并在他的引导下踏进生物统计学的门槛．以后从事数量遗传研究，我深切感受到统计学基础理论对于推动数量遗传学科发展的重要性．当时为了充实数理统计知识，我曾查阅了国内许多数理统计学的参考书，发现理论性的专著多数只侧重概率论原理，而应用性的专著则只限于介绍公式的应用，不涉及其推导过程．

1987年至1989年，我在美国北卡罗来纳州立大学统计系和遗传系攻读博士学位，系统地学习了统计学和数量遗传学的理论知识，并从事混合线性模型统计分析方法的理论研究，从而步入统计遗传研究领域的前沿．留学回国后，我在线性模型统计分析方法及数量遗传模型构建方面开展理论与应用研究，取得了一些成果．我还为硕士生和博士生开设了学位课程“线性模型与统计分析方法”．本书就是根据本人多年从事统计学和数量遗传学教学、科研的体验，在课程备课笔记基础上整理完成的．

这本书在介绍线性模型及其统计分析方法时，尽量兼顾理论与应用．全书共分八章．前二章简要介绍了概率统计和矩阵运算的基础知识．在第三章中，详细地介绍了简单线性回归模型的估计、推断、检验、预测等方法，并附有相应的公式推导和分析实例．在第四章和第五章中，以矩阵运算的形式介绍了多元线性回归模型的分析原理及自变量的选择方法．第六章简要地介绍了非线性模型的估计和推断方法．第七章以线性模型为基础，阐述了方差分析的原理．第八章系统地介绍了混合线性模型中方差分量估算及随机变量预测的原理和方法．

我希望书中所介绍的内容，能有助于农学或生物学领域的学

者更多地掌握、应用线性模型的基础理论及其统计分析方法.

本书得以完成，我应感谢导师季道藩教授和 Bruce S. Weir 教授对我多年的培养和教诲. 我还要感谢国家自然科学基金和国家教委跨世纪人才专项基金对我从事统计遗传基础理论研究的支持. 正是这些研究，使我们能在统计学和数量遗传学研究领域取得突破性成果，并将其总结在本书中. 这本书的编写和出版，得到了科学出版社的大力支持以及刘安编辑的鼓励. 在书稿的打印、校对过程中，我还得到研究生徐海明、吴吉祥、高用明、许自成等人的大力帮助. 在此，我向他们表示衷心的感谢.

由于本人水平有限，本书不免存在解释不妥或不完善之处，恳请专家和读者批评指正，以便今后修改补充.

朱　军

1998 年 8 月

第一章　概率论和统计学的基本原理

§1.1　概　　率

统计学(statistics)是一门数据分析的科学.它研究数据的取样、收集、组织、总结、表达和分析的科学方法,也研究如何根据数据的分析结果作出关于总体特性的有效推断和合理结论的科学方法. 在统计学中,总体(population)是指所要研究的对象的所有个体的总和. 在实际统计分析时,通常不可能研究所有的个体,而是在总体中选取一部分个体进行分析. 这些在实际研究中被分析的个体,称为样本(sample).

通过对样本资料的统计分析,可以推断总体的表现. 为了确保统计推断的可靠性,需要按事先设计的要求观察和收集数据. 这种过程称为试验(experiment). 实施试验所获得的任何可能结果称为试验的一次结局(outcome). 如果重复实施设计相同的试验,可以获得结果不尽相同的结果. 试验的一次或若干次结局,称为事件(event). 常用大写字母 A,B,C 等表示事件. 试验的所有可能结局称为试验的样本空间(sample space).

如果事件 A 的概率为$P(A)$,则 A 的对立事件的概率为 $P(\overline{A}) = 1-P(A)$. 如果存在二个事件 A 和B,事件 A 或者事件 B 发生,称为事件 A 与事件B 的和,其概率以 $P(A\cup B)$表示. 事件 A 与事件B 同时发生,称为事件 A 与事件B 的积. 其概率表示为 $P(A\cap B)$. 在事件 B 出现的条件下事件A 发生的概率,称为给定事件 B 时事件A 的条件概率,表示为 $P(A|B)$. 当事件 A 与事件 B 相互独立时,$P(A|B) = P(A)$ 或 $P(B|A) = P(B)$.

概率的基本运算法则是:

事件 A 发生或者事件B 发生的概率

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

事件 A 与事件 B 同时发生的概率

$$P(A \cap B) = \begin{cases} P(A)P(B \mid A) \\ P(B)P(A \mid B) \end{cases}$$

如果事件 A 和事件 B 为独立事件,则有

事件 A 发生或者事件 B 发生的概率

$$P(A \cup B) = P(A) + P(B)$$

事件 A 与事件 B 同时发生的概率 $P(A \cap B) = P(A)P(B)$

§1.2 随机变量

统计分析的目的是要推断总体的特性. 在统计学中,描述总体特性的数值称为参数(parameter). 总体的参数一般是未知的,需要经统计分析而推断. 通常的作法是在总体中抽取一个样本,由分析样本数据而获得一个可用于估计总体参数的数值. 这个数值称为总体参数的点估计(point estimate).

当从一个总体中抽取不同的样本,分析各样本所获得的点估计将不尽相同. 这种表现出变异性的特征,称为变量(variable). 在作统计试验以前,我们一般并不知道某一试验的确切结局,但是我们可以赋予试验结局以实际数量的一个函数. 因此这一变量称为随机变量(random variable). 随机变量常用大写字母表示,如 X, Y, Z. 它们可能出现的具体结果或数值则可用小写字母表示,如 x, y, z.

随机变量有二类. 一类是以计数表示的随机变量,称为离散变量(discrete variable);另一类是以任意实数表示的随机变量,称为连续变量(continuous variable).

随机变量之间常存在不同程度的关联性,这些关联性可以用数学模型或数学函数表示. 其中应用最多的是线性模型(linear model). 线性模型是描述变量之间相互关系的数学函数,它的参数只具有简单的线性关系. 在统计分析中广泛应用的回归分析、

相关分析、方差分析、协方差分析等都是建立在线性模型的理论基础上的.

在重复试验中随机变量 X 为 x 值的概率可以用概率密度函数(probability density function,简称 pdf)表示:

$$f(x) = P(X = x) \tag{1.1}$$

概率密度函数具有以下特性,

(1) 对于所有的 x 值,$0 \leq f(x) \leq 1$.

(2) 对于离散变量有 $\sum f(x) = 1$ 或连续变量则有 $\int f(x)dx = 1$.

在重复试验中随机变量 X 小于或等于 x 值的累计概率可以用分布函数(distribution function)表示:

$$F(x) = P(X \leq x) \tag{1.2}$$

分布函数具有以下特性,

(1) $\lim\limits_{x \to -\infty} F(x) = 0$;

(2) $\lim\limits_{x \to \infty} F(x) = 1$;

(3) $\lim\limits_{h \to 0} F(x + h) = F(x)$;

(4) 如果 $a < b$,那么 $F(a) \leq F(b)$;

(5) $P(a < X \leq b) = F(b) - F(a)$.

如果某一随机变量 X 在试验中可能出现的具体结果 x 具有概率密度函数 $f(x)$,则随机变量 X 的期望值(expected value)定义为:

$$E(X) = \sum_x xf(x), \text{如果 } X \text{ 是离散随机变量} \tag{1.3}$$

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx, \text{如果 } X \text{ 是连续随机变量} \tag{1.4}$$

根据随机变量期望值的定义,可以进一步推算随机变量线性函数的期望值. 随机变量 X 和常数 a 和 c 的线性函数 $(a + cX)$ 的期望值为

$$\begin{aligned} E(a + cX) &= E(a) + E(cX) \\ &= a + cE(X) \end{aligned}$$

其中常数的期望值仍为原常数.

两个随机变量 X 和 Y 的函数期望值为

$$E(aX + cY) = aE(X) + cE(Y)$$

$E(XY) = E(X)E(Y)$ 如果 X 和 Y 是独立的随机变量.

如果 $X_1, X_2, \cdots, X_i, \cdots, X_n$ 是 n 个随机变量,它们的线性函数 $\sum_{i=1}^{n} a_i X_i$ 的期望值为

$$E\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i E(X_i)$$

随机变量 X 的方差(variance)定义为

$$\begin{aligned} \sigma^2(X) &= E\{[X - E(X)]^2\} \\ &= E(X^2) - [E(X)]^2 \end{aligned} \tag{1.5}$$

随机变量 X 和常数 a 和 c 的线性函数 $(a + cX)$ 的方差为

$$\begin{aligned} \mathrm{Var}(a + cX) &= \mathrm{Var}(a) + \mathrm{Var}(cX) \\ &= c^2\sigma^2(X) \end{aligned}$$

其中常数的方差为零.

随机变量 X 和 Y 的协方差(covariance)定义为

$$\begin{aligned} \sigma(X, Y) &= E\{[X - E(X)][Y - E(Y)]\} \\ &= E(XY) - E(X)E(Y) \end{aligned} \tag{1.6}$$

给定常数 a, b, c 和 d,则随机变量 X 和 Y 的线性函数的协方差为

$$\mathrm{Cov}(a + cX, b + dY) = cd\sigma(X, Y)$$

随机变量 X 与 X 的协方差即为 X 的方差

$$\begin{aligned} \sigma(X, X) &= E\{[X - E(X)][X - E(X)]\} \\ &= E(XX) - E(X)E(X) \\ &= E(X^2) - [E(X)]^2 \end{aligned}$$

如果 $X_1, X_2, \cdots, X_i, \cdots, X_n$ 是 n 个随机变量,它们的线性函数 $\sum_{i=1}^{n} a_i X_i$ 的方差为

$$\mathrm{Var}\left(\sum_{i=1}^{n} a_i X_i\right)$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \sigma(X_i, X_j)$$

$$= \sum_{i=1}^{n} a_i^2 \sigma^2(X_i) + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} a_i a_j \sigma(X_i, X_j)$$

如果 X 和 Y 是相互独立的随机变量，则它们之间的协方差为零

$$\sigma(X, Y) = 0$$

因此当 $X_1, X_2, \cdots, X_i, \cdots, X_n$ 是 n 个相互独立的随机变量时，它们的线性函数 $\sum_{i=1}^{n} a_i X_i$ 的方差为

$$\mathrm{Var}\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 \sigma^2(X_i)$$

虽然协方差可以度量不同变量之间的相互关联性，但是协方差的值受到变量度量单位的影响. 度量随机变量 X 和 Y 之间相关性，并不受变量度量单位影响的参数是相关系数(correlation coefficient)，其定义为

$$\begin{aligned} \rho &= \frac{\sigma(X, Y)}{\sqrt{\sigma^2(X)\sigma^2(Y)}} \\ &= \frac{\sigma(X, Y)}{\sigma(X)\sigma(Y)} \end{aligned} \tag{1.7}$$

§1.3 概率分布

正态分布(normal distribution)是连续变量的一个重要的理论分布，在数理统计的理论和实践中占有重要的地位. 如果随机变量 X 具有正态分布，其概率密度函数是

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], \quad -\infty < x < \infty \tag{1.8}$$

具有正态分布的随机变量 X 的期望值和方差是

$$E(X) = \mu$$

$$\sigma^2(X) = \sigma^2$$

因而随机变量 X 是具有均值为 μ 和方差为 σ^2 的正态分布，表示

为

$$X \sim N(\mu, \sigma^2)$$

均值 μ 和方差 σ^2 是正态分布的二个参数. 由于正态分布具有均值和方差二个参数,这些参数的取值不同,可以产生不同的正态分布.

如果随机变量 X 是正态分布,其线性函数 $a + cX$ 也是正态分布的随机变量,具有均值 $a + c\mu$ 和方差 $c^2\sigma^2$,

$$a + cX \sim N(a + c\mu, c^2\sigma^2)$$

设 $X_1, X_2, \cdots, X_i, \cdots, X_n$ 是取自某总体的一个随机样本,该总体具有均值为 μ 和方差为 σ^2 的正态分布. 样本均值 $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ 是正态分布的随机变量,具有均值 μ 和方差 σ^2/n, $\overline{X} \sim N(\mu, \sigma^2/n)$. 当某总体具有均值为 μ 和方差为 σ^2 的未知分布,并且样本容量 n 趋于无穷大时,样本均值 $\overline{X}$ 趋近正态分布 $N(\mu, \sigma^2/n)$. 这称为中心极限定理(central limit theorem).

具有正态分布的随机变量可以转换成其它一些重要的概率分布. 具正态分布的随机变量 $X \sim N(\mu, \sigma^2)$,经由标准化转化

$$z = \frac{X - \mu}{\sigma}$$

产生的 z 变量具有均值为零和方差为 1 的标准正态分布(standard normal distribution), $z \sim N(0,1)$.

如果 $z_1, z_2, \cdots, z_i, \cdots, z_\nu$ 是 ν 个相互独立的标准正态变量,那么

$$X = z_1^2 + z_2^2 + \cdots z_i^2 + \cdots + z_\nu^2$$

是具有自由度(degrees of freedom, df)为 ν 的 χ^2 分布(chi square distribution), $X \sim \chi^2(\nu)$. ν 是 χ^2 分布的唯一参数. 具有 χ^2 分布的随机变量 X 的期望值和方差是

$$E[X] = \nu$$

$$\sigma^2(X) = 2\nu$$

如果随机变量 X 具有正态分布 $X \sim N(\mu, \sigma^2)$,其样本标准差 S

(standard deviation)的计算公式为

$$S = \sqrt{\left[\sum_{i=1}^{n}(X_i - \overline{X})^2\right]/(n-1)}$$

其中$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$，那么标准差的以下函数具有 χ^2 分布

$$(n-1)S^2/\sigma^2 \sim \chi^2(n-1)$$

如果随机变量 $z \sim N(0,1)$和 $X \sim \chi^2(\nu)$相互独立，那么它们的函数 T

$$T = \frac{z}{\sqrt{X/\nu}}$$

是具有自由度为 ν 的 t 分布(t distribution)，表示为 $T \sim t(\nu)$. 随机变量 T 的期望值和方差是

$$E[T] = 0$$

$$\sigma^2(T) = \nu/(\nu-2)$$

如果随机变量 X 具有正态分布 $X \sim N(\mu, \sigma^2)$，那么其样本均值和样本标准差的函数具有以下分布，

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim z(0,1)$$

$$(n-1)S^2/\sigma^2 \sim \chi^2(n-1)$$

因此随机变量 T

$$T = \frac{\dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}}}{\sqrt{\dfrac{(n-1)S^2}{\sigma^2}/(n-1)}}$$

$$= \frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$$

具有自由度为$(n-1)$的 t 分布.

如果随机变量 $X \sim \chi^2(\nu_1)$和 $Y \sim \chi^2(\nu_2)$是相互独立的具有 χ^2 分布的变量，随机变量 F 是 X 和 Y 的函数，

$$F = \frac{X/\nu_1}{Y/\nu_2} \sim F(\nu_1, \nu_2)$$

并且具有 F 分布(F distribution) $F \sim F(\nu_1, \nu_2)$. F 分布具有两个参数,分子自由度 ν_1 和分母自由度 ν_2.

如果有二个相互独立的 χ^2 分布的随机变量

$$(n_1 - 1)S_1^2/\sigma_1^2 \sim \chi^2(n_1 - 1)$$

$$(n_2 - 1)S_2^2/\sigma_2^2 \sim \chi^2(n_2 - 1)$$

可以得到检验方差分量的 F 分布变量

$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n_1 - 1, n_2 - 1)$$

当假定 $\sigma_1^2 = \sigma_2^2$ 成立时

$$F = \frac{S_1^2}{S_2^2} \sim F(n_1 - 1, n_2 - 1)$$

二个正态分布的随机变量 X 和 Y 具有二元正态分布(bivariable normal distribution).

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim BVN\left(\begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}, \begin{bmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{bmatrix}\right) \tag{1.9}$$

这二个变量之间的相关系数为 $\rho = \sigma_{XY}/(\sigma_X\sigma_Y)$. 每一个变量具有边缘分布(marginal distribution),它们也是正态分布

$$X \sim N(\mu_X, \sigma_X^2)$$

$$Y \sim N(\mu_Y, \sigma_Y^2)$$

对于二元正态分布的二个变量,给定一个变量的赋值,另一个变量的条件分布(conditional distribution)也是正态分布. 例如给定 $Y = y$,X 的条件分布是具有均值为 $\mu_X + \rho\sigma_X/\sigma_Y(y - \mu_Y)$和方差为 $\sigma_X^2(1 - \rho^2)$的正态分布

$$X \mid Y = y \sim N(\mu_X + \rho\sigma_X/\sigma_Y(y - \mu_Y), \sigma_X^2(1 - \rho^2)). \tag{1.10}$$

如果给定 $X = x$,Y 的条件分布是具有均值为 $\mu_Y + \rho\sigma_Y/\sigma_X(x - \mu_X)$和方差为 $\sigma_Y^2(1 - \rho^2)$的正态分布

$$Y \mid X = x \sim N(\mu_Y + \rho\sigma_Y/\sigma_X(x-\mu_X), \sigma_Y^2(1-\rho^2)) \qquad (1.11)$$

§1.4 统计估计量

统计分析的一项重要内容是通过对样本的分析，估计总体的参数. 如果 $X_1, X_2, \cdots, X_i, \cdots, X_n$ 是总体的样本，统计量(statistic)是样本的已知函数 $g(X_1, X_2, \cdots, X_i, X_n)$，它不包含总体分布的任何未知参数. 用于估计总体参数的统计量是参数的估计量(estimator). 可以有若干个估计量估计总体的某个参数. 无偏性(unbiasedness)和有效性(efficiency)是评价估计量优劣的二项指标. 如果估计量 $\hat{\theta}$ 是参数 θ 的点估计，并有 $E[\hat{\theta}] = \theta$，则 $\hat{\theta}$ 是 θ 的无偏估计(unbiased estimation). 对于某个参数，可能存在若干个无偏估计量. 这些无偏估计量并不都是等效的.

如果参数 θ 的二个无偏估计量 $\hat{\theta}_1$ 和 $\hat{\theta}_2$ 的方差分别为 $\sigma^2(\hat{\theta}_1)$ 和 $\sigma^2(\hat{\theta}_2)$，并且 $\sigma^2(\hat{\theta}_1) < \sigma^2(\hat{\theta}_2)$，那么无偏估计量 $\hat{\theta}_1$ 比无偏估计量 $\hat{\theta}_2$ 更有效. 无偏估计量 $\hat{\theta}_1$ 对于无偏估计量 $\hat{\theta}_2$ 的相对有效性可由二个方差的比值算得

$$\text{相对有效性} = \frac{\sigma^2(\hat{\theta}_1)}{\sigma^2(\hat{\theta}_2)} \qquad (1.12)$$

如果 $\hat{\theta}$ 是 θ 的一个点估计，θ^* 是 θ 的所有其它点估计，并有 $\sigma^2[\hat{\theta}] \leq \sigma^2[\theta^*]$，则 $\hat{\theta}$ 是 θ 的最小方差估计(minimum variance estimation). 具有最小方差的无偏估计量是最好的估计量.

在线性模型统计分析中，常用的参数估计方法是最小二乘法(method of least squares)和最大似然法(method of maximum likelihood).

如果观察值 Y_i 是参数 θ 的函数 $f_i(\theta)$ 和随机机误 ε_i 的和

$$Y_i = f_i(\theta) + \varepsilon_i, \quad i = 1, 2, \cdots, n \qquad (1.13)$$

通常假设 $E(\varepsilon_i) = 0$. 因而平方和

$$Q = \sum_{i=1}^{n} [Y_i - f_i(\theta)]^2 \tag{1.14}$$

也可视为参数 θ 的函数. 最小二乘估计量(least squares estimator)是使 Q 值为最小值的参数 θ 估计量.

现以一个简单的线性模型为例,说明运用最小二乘法估计参数的过程. 如果随机变量 Y_i 与变量 X_i 存在如下的线性关系

$$Y_i = b_0 + b_1 X_i + \varepsilon_i, i = 1,2,\cdots,n \tag{1.15}$$

其中回归系数 b_0 和 b_1 是未知的参数. 平方和 Q 的计算公式是

$$Q = \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_i)^2 \tag{1.16}$$

要估计 b_0 和 b_1 并使 Q 值为最小,可以分别对 b_0 和 b_1 求偏导,

$$\begin{aligned} \partial Q/\partial b_0 &= -2\sum_{i=1}^{n} (Y_i - b_0 - b_1 X_i) \\ \partial Q/\partial b_1 &= -2\sum_{i=1}^{n} X_i (Y_i - b_0 - b_1 X_i) \end{aligned} \tag{1.17}$$

然后用估计量 $\hat{b}_0$ 和 $\hat{b}_1$ 替代以上式子中的参数,并设以上偏导为零. 解下列正规方程组

$$\begin{aligned} -2\sum_{i=1}^{n} (Y_i - \hat{b}_0 - \hat{b}_1 X_i) &= 0 \\ -2\sum_{i=1}^{n} X_i (Y_i - \hat{b}_0 - \hat{b}_1 X_i) &= 0 \end{aligned} \tag{1.18}$$

便可获得二个参数的最小二乘估计量

$$\begin{aligned} \hat{b}_1 &= \left(\sum_{i=1} X_i Y_i - n\overline{X}\,\overline{Y}\right) / \left(\sum_{i=1} X_i^2 - n\overline{X}^2\right) \\ \hat{b}_0 &= \overline{Y} - \hat{b}_1 \overline{X} \end{aligned} \tag{1.19}$$

如果观察值 $Y_1, Y_2, \cdots, Y_i, \cdots, Y_n$ 是 n 个独立的随机变量,这一组观察值的联合概率函数为

$$g(y_1, y_2, \cdots, y_i, \cdots, y_n) = \prod_{i=1}^{n} f(y_i; \theta) \tag{1.20}$$

这个联合概率函数是给定一组观察值时参数 θ 的函数,常被称为

似然函数(likelihood function)，

$$L(\theta) = \prod_{i=1}^{n} f(y_i;\theta) \tag{1.21}$$

最大似然法估计量(maximum likelihood estimator)是使 $L(\theta)$值为最大值的 θ 估计量.

现仍以一个简单的线性模型为例，说明运用最大似然法估计参数的过程. 如果随机变量 Y_i 与变量 X_i 存在如下的线性关系($i = 1,2,\cdots,n$)

$$Y_i = b_0 + b_1X_i + \varepsilon_i \sim N(b_0 + b_1X_i,\sigma^2) \tag{1.22}$$

其中回归系数 b_0 和 b_1 以及方差 σ^2 是未知的参数. 似然函数是

$$\begin{aligned} L(b_0,b_1,\sigma^2) &= \prod_{i=1}^{n} f(Y_i;b_0,b_1,\sigma^2) \\ &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{(Y_i - b_0 - b_1X_i)^2}{2\sigma^2}\right] \end{aligned} \tag{1.23}$$

使 L 值为最大值亦即使 L 的自然对数值 $\ln L$ 为最大值，

$$\begin{aligned} \ln L(b_0,b_1,\sigma^2) = &-\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 \\ &-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i - b_0 - b_1X_i)^2 \end{aligned} \tag{1.24}$$

要估计回归系 b_0和 b_1以及方差 σ^2并使 $\ln L$ 值为最大，可以分别对 b_0和 b_1以及方差 σ^2求偏导，

$$\begin{aligned} \frac{\partial[\ln L(b_0,b_1,\sigma^2)]}{\partial b_0} &= \frac{1}{\sigma^2}\sum_{i=1}^{n}(Y_i - b_0 - b_1X_i) \\ \frac{\partial[\ln L(b_0,b_1,\sigma^2)]}{\partial b_1} &= \frac{1}{\sigma^2}\sum_{i=1}^{n}X_i(Y_i - b_0 - b_1X_i) \\ \frac{\partial[\ln L(b_0,b_1,\sigma^2)]}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(Y_i - b_0 - b_1X_i)^2 \end{aligned} \tag{1.25}$$

然后用估计量$\hat{b}_0$和$\hat{b}_1$以及 $\hat{\sigma}^2$替代以上式子中的参数，并设以上偏

导为零. 解下列正规方程组:

$$\begin{aligned}
&\frac{1}{\hat{\sigma}^2}\sum_{i=1}^{n}(Y_i-\hat{b}_0-\hat{b}_1X_i)=0\\
&\frac{1}{\hat{\sigma}^2}\sum_{i=1}^{n}X_i(Y_i-\hat{b}_0-\hat{b}_1X_i)=0\\
&-\frac{n}{2\hat{\sigma}^2}+\frac{1}{2\hat{\sigma}^4}\sum_{i=1}^{n}(Y_i-\hat{b}_0-\hat{b}_1X_i)^2=0
\end{aligned}\tag{1.26}$$

便可获得三个参数的最大似然法估计量

$$\begin{aligned}
\hat{b}_1&=\sum_{i=1}(X_iY_i-n\bar{X}\bar{Y})\Big/\sum_{i=1}(X_i^2-n\bar{X}^2)\\
\hat{b}_0&=\bar{Y}-\hat{b}_1\bar{X}\\
\hat{\sigma}^2&=\frac{1}{n}\sum_{i=1}^{n}(Y_i-\hat{b}_0-\hat{b}_1X_i)^2
\end{aligned}\tag{1.27}$$

§1.5 假设检验

在数理统计分析中,只能由估计量估计总体的参数. 尽管能获得总体参数的无偏估计,总体的参数始终是不可知的. 只能通过统计检验,由统计量推断总体的参数. 在统计推断过程中,需要对参数提出一定的假设,然后对提出的假设进行假设检验(hypothesis testing). 对于总体参数提出的假设有二类,一类是原假设(null hypothesis),以 H_0 表示,另一类是备择假设(alternative hypothesis),以 H_1 表示. 这二类假设是相互对立的. 原假设是将用统计的方法进行检验的假设,通常假设不存在真实的差异,

$$H_0:\theta=0$$

备择假设是当原假设被统计检验否定时准备接受的假设,常假设存在真实的差异,

$$H_1:\theta\neq 0$$

这时的检验称为二尾检验(two-tailed test). 如果原假设是

$$H_0:\theta\leq 0$$