应用线性回归模型

〔美〕约翰·内特
威廉·沃塞曼
迈克尔·H·库特纳 著
张勇 王国明 赵秀珍 译

中国统计出版社

应用线性回归模型

〔美〕 约翰·内特
威廉·沃塞曼 著
迈克尔·H·库特纳
张勇 王国明 赵秀珍 译

中国统计出版社

John Neter William Wasserman

Michael H・Kutner

APPLIED LINEAR REGRESSION MODELS

First edition

RICHARD D. IRWIN, INC., 1983

本书根据1983年版本译出

应用线性回归模型

YINGYONG XIANXING HUIGUI MOXING

〔美〕约翰·内特 威廉·沃塞曼 迈克尔·H·库特纳 著

张勇 王国明 赵秀珍 译

*

中国统计出版社出版

新华书店北京发行所发行

北京市通县永乐印刷厂印刷

*

850×1168毫米 32开本 20.125印张 50万字

1990年6月第1版 1990年6月北京第1次印刷

印数：1—3500

ISBN 7-5037-0335-0/O.6

定价：9.70元

DY1/4/1/19

译 者 序

当前，社会、经济和科学技术在高速发展，许多过去被认为是纯理论的知识，现在已被广泛应用于实际，同时在实际应用中得到提高、充实。线性回归模型就是这样一种知识。经国内外有关专家、学者推荐，我们对美国的《应用线性回归模型》一书进行了认真阅读和研究，认为它是一本有特色的教材，因此把它翻译过来介绍给我国读者。

本书的最大特点是应用性强。我国目前已出版了几种介绍线性回归模型的书，但都是侧重于数理推导，很少涉及应用问题。本书主要介绍线性回归模型在实际中的应用，以解决实际中存在的有关问题。要用模型分析社会、经济、科学技术中存在的问题，除了认识这些问题外，还要对建立模型的条件有明确的认识，清楚模型的参数意义和有关概念，本书对这些内容作了强调和认真的解释，并使用不同的模型对同一问题进行拟合，比较结果，使读者能够掌握有关建模方法，对实际问题进行分析、应用。

本书涉及的应用范围很广，介绍了线性回归模型在工商管理、经济、社会、卫生和生物科学等方面的应用，举了大量例子，分析了很多案例。此外，作为教材，它适用的层次也多，既可用作研究生之教材，也可用作本、专科学生之教材，还可以用作在职工作人员培训之教材或自学教材。读者可以根据时间和需要情况，有选择地学习有关内容。

本书另一特点是通俗易懂。本书是为应用者准备的，全书见不到繁琐的数学推导。它采用深入浅出、循序渐进的方法较系统地介绍了应用线性回归模型的知识，文理清晰，逻辑性强，叙述严谨，理论基础坚实。在重要的地方，书中多给出了直观、简练的描述证明，使读者容易领会其实质。具有不同学科背景的人通

过阅读此书都能得到裨益。如果说读者初学时尚有恐惧心理，当学完此书后，定会觉得既对有关知识有了全面的了解，又能在学习过程中轻松自如。

如何处理好学科之间的关系，使读者既能掌握本学科知识，又能了解和运用有关学科知识而不相互重复内容，是教科书中较难解决的问题。本书较好地处理了这个问题，比如计算机应用知识的介绍，有些经验是值得我们借鉴的。

书中安排了大量练习，并把它们分为三类，分别称为问题、习题和补充题。读者可以选做一些练习，以巩固各章所学的知识。

为了展现原著原貌，我们在翻译过程中忠实原文，没作改动。但由于译者水平有限，误译之处难免，欢迎读者指正。本书第一章至第五章由张勇翻译，第八、九、十一、十二、十三、十四、十五章由王国明翻译，第六、七、十章由赵秀珍翻译。张勇做了全稿的整理工作。

译者

1988.3.1

作 者 序

《应用线性回归模型》是《应用线性统计模型》回归部分的修订。有鉴于近年来回归分析的许多重要发展，亟需单独出版专书，以提供回归模型的修正和最新处理。《应用线性统计模型》一书的其他部分现在正在修订，修订以后将要出版。

线性回归模型当前在工商管理、经济、社会、卫生和生物科学等方面都有广泛应用。要成功地应用这些模型，对基础理论以及在现实中利用模型时会遇到的实际问题，都需要有充分的理解。《应用线性回归模型》基本上是一本应用的书，它试图把理论与实际有效地结合起来，避免两种极端，或者孤立地讲解理论，或者在没有必要理论基础的情况下只讲应用方法。

《应用线性回归模型》在许多重要方面与《应用线性统计模型》不同。

1.增加了一些重要的新专题，如多重共线性的检测、岭回归、强影响观察值的检测以及非线性回归。近年来，检测多重共线性和强影响观察值等方面有值得注意的新发展，现代回归分析教科书需要适当地包含这些专题。此外，我们还增加了一些特殊专题，诸如统计检验的 P 值以及最小绝对离差方法等。

2.重新组织和扩充了一些专题，包括加权回归、自变量的选择以及正态概率图等。同时，我们还根据教学经验，大量地修订了另一些内容，以便讲解时更加清楚明了。

3.例子的范围，除工商管理、经济和社会科学的应用外，还扩大到卫生和生物科学的应用。在所有的情况下，一般读者都能容易理解。

4.介绍了大量的计算机图，以说明回归分析中计算机图示的用途。

5.增加了两大组实际数据资料，它们可以不同方式加以使

用。

6.最后，大大地增加了每章结束后的练习材料，并把它们分为三类，分别称为问题、习题和补充题。问题部分是基本的问题和难题；习题部分是概念性和理论性的问题；补充题部分是利用大量数据，进行大量计算和分析的问题。

这本书不仅包括较传统的回归专题，而且包括实际中通常被忽视的重要专题。所以，我们专门用一章讲指示变量，包括应变量指示变量和自变量指示变量。另有一章专讲用计算机选择回归模型中一组“最佳”自变量的方法。应用残差分析检验模型的适度在全书中是个重复的主题。同时还讨论了模型不合适时所用的补救方法。在分析研究结果时，我们强调使用估计方法，而不是检验，因为在实际中，估计通常更有意义。另外，因为实际问题很少只与单个估计有关，所以我们强调使用联合估计方法。

我们根据很好理解应用问题所需要的程度介绍理论概念。在那些我们认为能说明重要方法的例子中给出证明。本书主要强调对模型的透彻理解，特别是理解模型的参数意义，因为只有这样才能进行适当的应用。我们用许多不同的案例来说明理论原理的运用，以显示回归模型应用的广泛多样性以及对不同问题进行分析的方法。

我们在每一章中，以“注”和“说明”的形式，给出问题进展主流的有关补充讨论情况和材料。这样，每章的基本观点可以得到简明的表述而不被打乱。同样，供选择的“专题”章节补充了那些讲述主要发展的章节，提供了各种附加专题，在多数情况下，这些附加专题可以省略而不致破坏内容连续性。

回归模型的应用通常需要进行大量计算，我们设想在大多数应用工作中可以利用计算机，而且几乎每个计算机使用者都接触过回归分析的程序软件。所以，我们解释拟合回归模型的基本数学步骤，而不详细讲述计算过程。这样，我们可以避免许多复杂的计算公式，而把主要精力放在基本原理上。在本书中，我们大量地使用计算机进行计算，说明了各种计算机的打印输出结果，

并解释了如何使用这些结果进行分析。

每章（第一章除外）的后面附有选题，供读者加深对方法的理解，并利用学过的概念来分析资料。我们还仔细地提供了反映真实应用的资料分析问题。对大多数问题，最好使用计算器或计算机进行处理，我们主张只要有可能就要利用这条途径。

我们设想《应用线性回归模型》的读者已具有统计推断的入门知识，即第一章列出的内容。读者可以根据自己的情况，选读该章的有关部分；任课教师也可以使用一些补充材料以补不足。第一章主要是当作基本统计结果的一个参考章节，读者在阅读全书时会不断地使用它们。

学习《应用线性回归模型》不需要微积分知识。在某些例子中，我们使用微积分证明一些重要结果是如何得到的，但这些证明都放在补充的“说明”和“注”中，可以省略掉而不影响本书的连续性。具有微积分知识的读者，将按自然的顺序发现这些“说明”和“注”，从而可以直接得到数学推导的益处。在多元回归中需要一些矩阵代数的知识，为了便于学习，第六章针对简单回归介绍了这些矩阵代数的基本知识。

《应用线性回归模型》可用作本科生或研究生的回归分析教程，也可以用作应用统计的第二教程。不同的课程可根据时间的安排和本课程的目的选用本教科书的内容。第二、三、四、五（只包括第一节至第四节）、六、七、八和十二章介绍回归的基本知识。有兴趣者如果时间允许可学习第九、十、十一、十三、十四和十五章。

从事工商管理、经济、社会、卫生和生物科学的人员如想获得应用回归模型的知识，也可以把本书作为自学课本。

写这样一本书，没有其他人的大力帮助是不可能完成的。我们感谢许多为发展本书理论和实际作出过贡献的人，也感谢我们的学生用不同的方式帮助我们，以改进本书所介绍的各种方法。我们感谢许多本书的使用者，他们根据教学经验，给我们提供了意见和建议。我们还要感谢密苏里大学的J.E.霍尔斯坦教授和

西佛罗里达大学的D.L.谢里教授，他们仔细地审阅了全书并提出了建议。R.L.沃格尔孜孜不倦地帮助我们校核了书稿，我们向他表示感谢。M.J.林使用Zeta型3600作图机画出了计算机图，G.柯特森尼斯和山本志筑（Shizuki Yamamoto）用其他方法帮助我们检查了计算结果，R.巴格特完成了几乎所有的打字任务，能干地处理了困难的书稿准备工作。我们向所有给我们帮助的人表示感谢。

最后，我们的家庭成员耐心地承担了因为我们修订此书而带来的负担，我们感谢他们的理解。

约翰.内特（John Neter）

威廉·沃塞曼（William Wasserman）

迈克尔·H·库特纳（Michael H.Kutner）

目 录

第一章　概率和统计的基本结果

本章介绍概率和统计的基本结果，作为学习本书的参考知识，供阅读本书时查考。有时读者也可以根据需要，自行参考本章的某些结果。

最好在学习第二章之前先浏览一下本章的概率和统计推断，也可以直接就学习第二章。

1.1　求和算符与乘积算符

求和算符

求和算符 Σ 定义如下：

$$\sum_{i=1}^{n} Y_i = Y_1 + Y_2 + \cdots + Y_n \tag{1.1}$$

这个算符具有下述重要性质：

当 k、a 和 c 是常数时，有

$$\sum_{i=1}^{n} k = nk \tag{1.2a}$$

$$\sum_{i=1}^{n} (Y_i + Z_i) = \sum_{i=1}^{n} Y_i + \sum_{i=1}^{n} Z_i \tag{1.2b}$$

$$\sum_{i=1}^{n} (a + cY_i) = na + c\sum_{i=1}^{n} Y_i \tag{1.2c}$$

双重求和算符 $\Sigma\Sigma$ 定义如下：

$$\sum_{i=1}^{n} \sum_{j=1}^{m} Y_{ij} = \sum_{i=1}^{n} (Y_{i1} + Y_{i2} + \cdots + Y_{im})$$

$$=Y_{11}+\cdots+Y_{1m}+Y_{21}+\cdots+Y_{2m}+\cdots+Y_{nm} \tag{1.3}$$

双重求和算符的一个重要性质是：

$$\sum_{i=1}^{n}\sum_{j=1}^{m}Y_{ij}=\sum_{j=1}^{m}\sum_{i=1}^{n}Y_{ij} \tag{1.4}$$

乘积算符

乘积算符 Π 定义如下：

$$\prod_{i=1}^{n}Y_i=Y_1\cdot Y_2\cdot Y_3\cdots Y_n \tag{1.5}$$

1.2 概　　率

加法定理

设 A_i、A_j 是某样本空间的两个事件，则

$$P(A_i\cup A_j)=P(A_i)+P(A_j)-P(A_i\cap A_j) \tag{1.6}$$

其中 $P(A_i\cup A_j)$表示 A_i 或 A_j 发生，或二者同时发生的概率；$P(A_i)$、$P(A_j)$分别表示 A_i 和 A_j 发生的概率；$P(A_i\cap A_j)$表示 A_i 和 A_j 同时发生的概率。

乘法定理

设 $P(A_i|A_j)$为给定 A_j 已发生时 A_i 发生的条件概率，定义如下：

$$P(A_i|A_j)=\frac{P(A_i\cap A_j)}{P(A_j)},\quad P(A_j)\neq 0 \tag{1.7}$$

则乘法定理为：

$$\begin{aligned}P(A_i\cap A_j)&=P(A_i)P(A_j|A_i)\\&=P(A_j)P(A_i|A_j)\end{aligned} \tag{1.8}$$

对立事件

A_i 的对立事件记为 $\bar{A}_i$。对立事件的下列结果很有用：

$$P(\bar{A}_i)=1-P(A_i) \tag{1.9}$$

$$P(\overline{A_i\cup A_j})=P(\bar{A}_i\cap\bar{A}_j) \tag{1.10}$$

1.3 随机变量

本节假设随机变量 Y 只取有限个值。（如果 Y 是连续随机变量，可以用积分代替求和过程。）

期望值

设随机变量 Y 的取值为 Y_1，…，Y_k，相应的概率由下列概率函数给出：

$$f(Y_s) = P(Y = Y_s) \quad s = 1,\cdots,k \tag{1.11}$$

则 Y 的期望值定义为：

$$E(Y) = \sum_{s=1}^{k} Y_s f(Y_s) \tag{1.12}$$

期望算符 E 具有一个重要性质，即当 a 和 c 是常数时，有

$$E(a + cY) = a + cE(Y) \tag{1.13}$$

由此可得：

$$E(a) = a \tag{1.13a}$$

$$E(cY) = cE(Y) \tag{1.13b}$$

$$E(a + Y) = a + E(Y) \tag{1.13c}$$

方差

随机变量 Y 的方差用 $\sigma^2(Y)$表示，定义如下：

$$\sigma^2(Y) = E\{[Y - E(Y)]^2\} \tag{1.14}$$

它的一个等价表达式为：

$$\sigma^2(Y) = E(Y^2) - [E(Y)]^2 \tag{1.14a}$$

经常遇到 Y 的线性函数的方差。我们把 $a + cY$ 的方差记为 $\sigma^2(a + cY)$，则当 a 和 c 是常数时，有：

$$\sigma^2(a + cY) = c^2\sigma^2(Y) \tag{1.15}$$

由此可得：

$$\sigma^2(a + Y) = \sigma^2(Y) \tag{1.15a}$$

$$\sigma^2(cY) = c^2\sigma^2(Y) \tag{1.15b}$$

联合、边际和条件概率分布