21世纪财政金融系列教材

Econometrics

通过例题学习计量经济学

[日] 白砂堤津耶 著
瞿 强 译

中国人民大学出版社

21 世纪财政金融系列教材

通过例题学习计量经济学

[日] 白砂堤津耶 著

瞿强 译

中国人民大学出版社

图书在版编目（CIP）数据

通过例题学习计量经济学/（日）白砂堤津耶著；瞿强译
北京：中国人民大学出版社，2002
21世纪财政金融系列教材

ISBN 7-300-04155-8/F·1276
Ⅰ.通…
Ⅱ.①白…②瞿…
Ⅲ.计量经济学-高等学校-教材
Ⅳ.F224.0

中国版本图书馆CIP数据核字（2002）第036561号

例題で學ぶ初步からの計量經濟學
著者——白砂堤津耶
發行者——大石　進
發行所——株式会社日本評論社

21世纪财政金融系列教材
通过例题学习计量经济学
[日]白砂堤津耶　著
瞿强　译

出版发行：中国人民大学出版社
（北京中关村大街31号　邮编100080）
邮购部：62515351　门市部：62514148
总编室：62511242　出版部：62511239
E-mail：rendafx@public 3.bta.net.cn
经　　销：新华书店
印　　刷：北京东方圣雅印刷有限公司

开本：787×980毫米 1/16　印张：17.5
2002年9月第1版　2003年1月第2次印刷
字数：316 000

定价：19.00元
（图书出现印装问题，本社负责调换）

《21世纪财政金融系列教材》编委会

总　序

经过两年多的策划和论证，从写作、成稿到出版，中国人民大学财政金融学院“21世纪财政金融系列教材”终于与新世纪同步诞生！这是在黄达教授、陈共教授、王传纶教授、周升业教授、郎荣燊教授等老一辈经济学家的关心、指导下，在财政金融学院部分中青年教师的努力下，奉献给财金学子的一套全新的教科书。

中国人民大学财政金融学院成立虽然只有三年，但原有的财政金融系和投资经济系早在中华人民共和国成立之初就已具有雏形，成为新中国财政、金融、投资经济学的主要教学科研基地之一，造就了一批国内外著名的经济学家，培养了大批科研、教学与管理人才，长期以来一直享有很高的社会声誉。跨人21世纪的财政金融学院，包括财政系、金融系、投资经济系、保险系和电子应用中心，并设有金融与证券研究所、财税研究所、教育保险研究所和投资研究所，为培养新世纪的经济理论与经济管理人才奠定了坚实的基础。

一套高质量的教材是提高教学质量的前提之一。中国人民大学财政金融学院的前身就曾为中国财政金融学科的教材建设做出了重

大贡献。新中国成立伊始，财政金融学科各教研室的年轻教师，就编写出把前苏联教材与中国自身特点相结合的第一代教材，如黄达教授主编的《资本主义国家的货币流通与信用》和《货币信用学》，财政教研室主编的《财政学讲义》以及陈共和侯梦蟾教授主编的《财政学（初稿）》等。改革开放以来，财政金融学院的教材建设出现了第二次高潮，把中国财政、金融、投资专业的教材推向新的水平。黄达教授主编的《货币银行学》，陈共教授主编的《财政学》成为全国高校财经类专业核心课教材；王传纶教授主编的《国际税收》，周升业教授主编的《对外开放下的金融运行》，陈共、侯梦蟾、袁振宇教授主编的《财政学教程》，侯梦蟾教授编著的《税收经济学导论》等，都曾获得国家级优秀教材奖，在财政金融学界引起很大反响。这些教材犹如及时雨，哺育了一大批适应社会主义市场经济需要的财金理论与管理人才。

面对经济体制、财政制度和金融制度改革中所涌现出来的新问题，面对经济学的新发展，面对国内外财经教学的新挑战，中国人民大学财政金融学院的新一代教师，根据全国第三次教育工作会议对教学内容、教学方法改革所提出的具体要求，适应学科的重大调整和素质型人才的培养模式，掀起了第三次教材建设高潮。这套“21 世纪财政金融系列教材”正是经过财政金融学院学术委员会和院长办公会多次论证，由中青年学术骨干组成作者队伍精心撰写的一套本科教材。这套教材在写作方法上力求规范分析与实证分析相结合，理论与实践相结合，在内容上尽量反映国内外最新研究成果，跟踪中国的经济改革实践，做到体系完整、内容丰富、实用性强。

最后，谨向我们的作者表示诚挚的感谢，向中国人民大学出版社的领导、编辑人员的真诚合作表示感谢。同时，我们也希望同行专家和读者对这套教材提出宝贵意见和建议，今后我们将不断完善这套教材。

安体富

1999 年 11 月 15 日

译者前言

计量经济学在现代经济学的学习和研究中的重要性自不待言，近年来国内已经出现了大量翻译、影印和自编的教材，摆在读者面前的这本《通过例题学习计量经济学》仍是一本别具特色、值得一读的好教材。

大约在1996年，译者在日本一桥大学进修，深切感受到计量经济学的重要，但是苦于缺乏这方面的背景训练，急于找一本入门教材来补课，无意间在一本很有名的经济学课外辅导杂志《经济讲座》（経済セミナー）上看到连载的计量经济学辅导材料，写得深入浅出，而且富于实用性。可惜连载共分十次，当时读了一期等待下一期的心情至今记忆犹新。1999年在国家图书馆意外发现该连载已经修订出版，出版者是日本最大的出版社——日本评论社。该书出版后好评如潮，一年之内增印5次，颇有“洛阳纸贵”的感觉，这在号称“统计大国”、计量经济学十分发达的日本也不多见。看来，该书确有其独到之处。

因此，我当即联系作者和出版社，决定将其介绍到中国来，让与我有过或有着同样“困惑”的人，也能够解困。

本书的特点主要有以下几个方面：

其一，这是一本非常基础的入门教材，篇幅不大，简明扼要，读者只要具备高中数学知识，即可掌握计量经济学的一些基本知识，如一元回归、多元回归、假设检验，乃至TSP应用。目前国内引进的大多是美国教材，相对于目前国内本科教学的课程安排来说，篇幅过大，学生往往难得要领。

其二，这是一本由经济学者写的、实战性很强的入门教材。目前国内自编的计量经济学教材，作者大多是数学专业出身，侧重于数理推导，疏于经济应用。但是计量经济学毕竟是用统计方法来研究经济问题的学科，重点还在后者。本书的作者白砂堤津耶教授毕业于日本名校庆应大学，获经济学博士学位，书中巧妙地避开了各种方法的数学推导，简单介绍相关概念后，单刀直入，通过70多个例题，逐渐展开计量经济学这门看似深奥的学问，使读者在学习过程中，能够直接将理论与实际应用结合起来。而且，作者曾经研究过中国经济问题，用了不少关于中国的例子，对我们来说更是平添了一份学习的兴趣和乐趣。

本书的读者对象主要是两种人，一种是在校学习计量经济学的学生，对于计量经济学心存畏惧，而且要疲于应付英语、数学、政治等基础课；另一种是经济领域的社会人士，因工作、研究的需要补习基础计量经济学知识。对于这两种人来说，本书都是不错的入门向导。

当然，由于本书定位为“入门”，因此，对于想更全面、更深入掌握现代计量分析方法的人来说，还是不够的。原书附有不少日文参考文献，译时略去。译者一个简单的建议是，在该书的基础上，再阅读类似古扎拉蒂（D. Gujarati）的《计量经济学》(Basic Econometrics)，大致可窥全貌，再深入，就是格林（Greene）的《计量经济分析》(Econometrics Analysis) 了。

瞿强

2002年5月

作者前言

本书简明扼要地介绍了计量经济学的基本方法，并希望通过实际例题使读者能够完全掌握它。作者在写作时设想的读者对象主要是初学计量经济学的经济专业和工商管理专业的学生，以及实际经济工作者。读者只要能够大致掌握高中的数学知识，就已足够。统计学的基础知识，完全不作要求。而且，本书的范围不仅限于“狭义计量经济学”，还包括实际经济分析中经常使用的“描述统计”（第一、二章）、“投入产出分析”（第九章），因此，也可以看做是“广义计量经济学”（又称“数量经济分析”）的入门教材。

计量经济学在应用经济学课程中，占有极其重要的位置，这一点想必读者早已清楚，无须笔者在此强调。近年来，经济学论文、报告、白皮书中计量经济学方法被大量运用，我们无论是阅读还是写作这些材料，计量经济学知识都不可或缺。

另外，人们通常认为计量经济学“很难”、“不易掌握”，为什么呢？作者认为，造成这种状况的原因在于，计量经济学的理论学习和实际运用之间存在着很大的差距。

为了弥补这一差距，本书通过对取材于实际经济生活例题的解答，总结了学习计量经济学的基本方法。计量经济学是经济理论和统计学相结合的学问，也是极具实用性的学问。因此，学习计量经济学和学习统计学一样，通过对课本中的例题和经济数据的分析、运用，不断掌握“计算能力”和“应用能力”（针对何种问题，采用何种方法，如何采用等），做到真正的理解，也就是所谓的“知行合一”，是非常重要的。本书准备了70多个附有详细解答的例题（少量较为复杂的例题带有*号），通过对本书的学习，相信对读者掌握实用的计量经济学会有所帮助。

不言而喻，本书还不够完善。由于优先考虑“便于理解”和“能够使用”，自然会牺牲一些严密性。此外，概率论和时间序列分析等，未能深入探讨。这些问题希望能在将来有所改进。读者如有疑问和建议，望不吝赐教。

最后，本书是在1995年4—12月的《经济讲座》上的连载文章《通过例题学习计量经济学》的基础上，经过较大幅度的增补和修改而形成的。从文章的连载到本书的出版，日本评论社的饭冢英俊给予了很大帮助，在此深表谢意。

白砂堤津耶

1998年1月　大寒

目 录

绪　论　　什么是计量经济学

1. 计量经济学的发展

计量经济学（econometrics）是如何产生的呢？首先，让我们来简单地看一下计量经济学的发展史。计量经济学的学术渊源，最早可以上溯到英国的经济学者和统计学者威廉·配第（W. Petty）的著作《政治算术》（1690）。配第在该书的前言中，强调尽量排除主观因素，运用数（number）、重量（weight）以及尺度（measure）等现实数据进行分析的重要性。这正是计量经济学的开端，熊彼特曾经将配第称做计量经济学的始祖，对他的有关数量分析的做法给予了高度的评价。

现代计量经济学的尝试，始于 1911 年摩尔（H.L. Moore）的《工资的法则》一书。他用统计方法对工资的边际生产力理论进行了检验。1930 年，以挪威经济学家弗瑞

希（R. Frisch，计量经济学的命名者）为中心，在美国成立了计量经济学会(Econometric Society)，1933年该会的会刊《计量经济学研究》(Economitrica) 创刊。这样，计量经济学在经济学学科中正式取得了“公民权”，终于开始了真正的发展。

道格拉斯（P.H. Douglas）计算了美国制造业的生产函数，发现产量与生产要素之间具有稳定的关系，荷兰经济学家丁伯格（J.Tinbergen，与弗瑞希同获第一届诺贝尔经济学奖）开发了联立方程体系的宏观计量经济模型，克莱因（L.R. Klein）与戈德伯格（A.S. Goldberger）在凯恩斯经济学框架中完成了现代宏观计量经济模型的原型，库兹涅茨（S.S. Kuznets）和斯通（J.R.N. Stone）系统完善了作为宏观计量经济模型源泉的国民收入统计体系，哈佛尔莫（T. Haavelmo）明确了联立方程模型推定中的问题，而库普曼斯（T.C. Koopmans）克服了前人的困难，发现了新的推定方法，里昂惕夫（W.W. Leontief）创立了投入产出分析，所有这些都是在计量经济学的发展中留下光辉业绩的学者。

此外，即使在经济史的研究领域，计量经济学也得到了灵活的运用，称做“计量经济史”或“数量经济史”。其开拓者佛格尔（R.W. Fogel）和诺斯（D.C. North）荣获1993年诺贝尔经济学奖。

2. 计量经济学的研究方法与学习方法

所谓计量经济学，简单地说就是这样一门学科，它运用现实的数据对根据经济理论而建立的模型进行统计推定与检验，以用于经济预测以及政策评估和制定等，这样做的同时也深化和发展了经济理论。

经济学通常被认为是一种“实证科学”。为了更好地解释由大量的人类行为累积构成的经济现象，首先建立假说（即模型），然后对它进行检验，试图发现稳定的关系，这恐怕是计量经济学最大的任务。但是，经济学与自然科学不同，它不能经常在实验室里进行“受控实验”(controlled experiment)，因而能够得到的信息极其有限，实证分析实际上非常困难。从这一角度看，经济学的研究环境有点类似于天文学。

下面，我们根据计量分析的过程，对计量经济学的方法和学习时应当注意的问题，作一些简单的解说。

建立模型（model building）

计量经济分析通常是从建立模型开始的。它是在以经济理论为中心的“先验的信息”（a priori information）的基础上，为了将复杂的现实经济简化为假说，运用数学方法构建模型的过程。

建立模型的第一步是要建立模型的函数关系，也即进行特定化处理（specification）。这时，经济理论对模型的函数关系（例如是线性的还是非线性的）还没有具体的表示，为了使模型特定化，在按照理论指导的同时，分析者必须在分析的目的、过去同类的分析、数据的观察事实（数据的变化）的基础上，做出自身的判断。

一般说来，最常用的多是简单的函数关系。例如类似下面这样的一次函数：

$$Y = \alpha + \beta X \text{（}\alpha,\ \beta\text{ 为参数）}$$

这是因为，函数关系越简单，参数越稳定，估计的结果越容易进行经济学解释。尤其是，参数的稳定性是发现稳定的经济法则的必要条件，因而更加重要。此外，函数关系越简单，越容易进行推定和检验。不过，由于目前计算机技术已经非常发达，这些因素的重要性已经减小了。

有关模型的特定化问题暂且到此为止，下面重点说明建立模型时最重要的一个问题。这就是，如果可能的话，应该尽量建立自律性的（autonomous）模型。所谓自律性的模型，简单地说，就是由深厚的经济理论所推导出的模型（例如根据微观经济学中的效用函数、生产函数等建立的模型）。判断模型是不是自律的，对于初学者来说是一件非常困难的事情，但是对于难以做受控实验的经济学来说，通过对自律性模型的实证分析，有可能发现稳定的经济规律，提高对未来预测的准确度，并提出真正有效的政策建议。举一个例子，在分析轿车的消费需求结构时，不是单纯计算轿车的需求函数，而是更进一步考虑其背后反映消费者口味（偏好）的效用函数、预算制约下的最大化条件等。将消费者决策过程明确化、系统化，是建立自律性模型所需要的。弗瑞希、哈佛尔莫、马夏克（J. Marschak）等计量经济学的先驱们，已经在分析中考虑到这一问题，并为建立自律性的模型费尽了心思。

最后再补充一点并希望引起注意的是，不要被数据表面的变化和关系等所迷惑，陷入空洞的“缺乏理论的计算”（measurement without theory）陷阱之中。在实践中，这一类的失败和错误经常出现。回避没有理论的计算与建立自律性的模型有关，因此首先需要对微观或宏观经济学有充分的理解。如果可能的话，希望能够具备中级以上的经济学知识。如果研究课题需要的话，国际经济学、金融理

论、公共经济学、劳动经济学等经济学的各个领域也不可忽视。

数据的收集

对于计量经济分析来说，收集数据是一件极为重要的工作。不论模型构建得多么精致，如果没有赖以计算的数据，仍然是纸上谈兵。换句话说，是否有数据，也关系到模型的建立。数据的收集，看上去不难理解，但却是一件费时费力的工作。因此，说数据收集完成意味着分析完成了一半，恐怕并不为过。平常在大学的图书馆中多接触《国民经济计算年报》(经济企划厅)、《经济统计年报》(日本银行)、《家庭调查年报》(总务厅)、《工资结构基本统计调查》(劳动省)等基本经济统计资料，掌握数据的来源、性质、统计方法等基础知识，是非常必要的。①

此外，收集起来的数据并不总是能够直接运用，通常需要进行必要的加工。例如，物价水平、季节变化调整、资本存量的推算等常遇到这种情况。这样的数据加工方法，在实践学习的同时，如果能从《经济统计》的教材中以及课堂学习中事先掌握，将会更为方便。

数据到手以后，不要匆忙代人模型进行推算，可以先作成散点图 (scatter diagram) 等形式，注意养成仔细观察数据的习惯。近年来，计算机作图已经很容易了，随着经验的积累，可以从这些图形中获得很多信息，甚至可以对原先建立的模型进行修改。

模型的统计估计与检验

运用费力收集的数据，对模型的参数 (parameter) 进行具体的估算 (estimation)，然后再利用其结果对模型进行验证，就是所谓的假设检验 (test of hypothesis) 工作。对这一部分的解说构成了标准计量经济学教材的主要内容。从 20 世纪 50 年代至 60 年代初期，计量经济学界有关模型的统计估计与检验的技术以及相应的计算软件的开发，进展迅速。这一阶段主要的学习内容是，一元回归、多元回归、联立方程模型的推定方法 (间接最小二乘法、二阶段最小二乘法、有限信息最优法、完全信息最优法、三阶段最小二乘法) 与假设检验的各种方法。假设检验一般要检查以下三个方面：第一是检查显著性，即对估计出来的参数，用统计学方法检查其信赖程度；第二是检查是否符合条件，即看看推断出来的参数符号 (正或负) 是否与经济理论相一致；第三是检查拟合性，即检查推断出来的模

① 中国目前主要的经济统计资料有：《中国经济统计年鉴》、《金融统计年鉴》、《财政统计年鉴》以及其他各类专业性统计资料。——译者注

型是否能够很好地说明现实数据的变化。有时还通过模拟（simulation）方法，看看推断出来的模型在多大程度上能够再现过去的情况。

在这一阶段，我们还要学习、掌握与回归分析有关的一些较为麻烦的问题，例如不均匀离散、多重共线性、系列相关（自我相关），以及与联立方程模型有关的识别问题等。

如果被推断的模型通过了检验，就意味着新的经济法则的发现、经济预测、政策评价与制定等较为顺利；相反，如果没有通过检验，说明以前构建的模型不适当，应该放弃，并重新建立模型。以上就是计量经济分析的基本过程。

最后需要指出的是，计量经济分析，尤其是估计与检验工作，需要从各个不同的角度对大量的数据进行统计处理，因此利用计算机是必不可少的。所幸的是，目前已经出现了大量功能完备并且易于操作的软件包，例如 TSP、SAS、SPSS、GAUSS、RATS、SHAZAM、LIMDEP 等。笔者在讲课中使用的是 TSP，这是计量经济分析中具有代表性的软件，用起来非常简单，而且通过切身体验，有助于提高学生的学习兴趣和理解能力。希望本书的读者也能够灵活使用这一便利的软件，向书中的例题挑战。

第一章　统计学基础知识(一)

本书虽然不将统计学和数学的预备知识作为学习的必要前提，但是统计学的基本知识对于理解计量经济学，极为重要。本章将对计量经济分析中使用频率很高的统计知识，尤其是以数据的观察和整理为目的的描述统计学（descriptive statistics）为中心，进行初步的学习。

1. 算术平均

算术平均（arithmetic mean），就是我们日常生活中所使用的普通的平均数，其定义如下式：

$$\bar{X}=\frac{X_1+X_2+\cdots+X_n}{n}=\frac{\sum X}{n} \qquad (1-1)$$

式中，$\bar{X}$ 读做 X - ba；$\sum$ 读做 sigma，是希腊字母，与罗

马字母中的 S 相当，是求和计算的符号。

[例题 1—1]

表 1—1 表示 7 个发达国家与亚洲新兴工业国(地区)(NIES)的实际 GDP 的增长率和失业率（均为 1995 年数）。求各组中，(1) 和 (2) 的算术平均数。

表 1—1　　实际 GDP 增长率和失业率（1995 年）

国家(地区)		(1)实际 GDP 增长率(%)	(2)失业率(%)
7个发达国家	日本	0.9	3.2
	美国	2.0	5.6
	英国	2.4	8.2
	德国	1.9	9.4
	法国	2.2	11.6
	意大利	3.0	12.0
	加拿大	2.3	9.6
亚洲新兴工业国(地区)	韩国	9.0	2.0
	中国台湾	6.1	1.8
	中国香港	4.7	3.2
	新加坡	8.8	2.7

资料来源：日本银行《外国经济统计年报》。

[解答]

(1) 实际 GDP 增长率

7 个发达国家

$$\bar{X}=\frac{0.9\%+2.0\%+2.4\%+1.9\%+2.2\%+3.0\%+2.3\%}{7}$$

$$=\frac{14.7\%}{7}=2.1\%$$

亚洲新兴工业国(地区)

$$\bar{X}=\frac{9.0\%+6.1\%+4.7\%+8.8\%}{4}=\frac{28.6\%}{4}=7.15\%=7.2\%$$

(2) 失业率

7 个发达国家

$$\bar{X}=\frac{3.2\%+5.6\%+8.2\%+9.4\%+11.6\%+12.0\%+9.6\%}{7}$$

$$=\frac{59.6\%}{7}=8.5\%$$

亚洲新兴工业国(地区)

$$\bar{X}=\frac{2.0\%+1.8\%+3.2\%+2.7\%}{4}=\frac{9.7\%}{4}=2.4\%$$

2. 加权算术平均

加权算术平均 (weighted arithmetic mean), 是将各数据先乘以反映其重要性的权数 (w), 再求平均的方法。其定义式如下式:

$$\bar{X}_w=\frac{w_1X_1+w_2X_2+\cdots+w_nX_n}{w_1+w_2+\cdots+w_n}=\frac{\sum wX}{\sum w} \qquad (1—2)$$

式中, 权数的作用非常重要。

[例题 1—2]

表 1—2 是对关东 1 都 6 县女性临时工的小时工资与劳动者人数 (千人) 的调查结果。

(1) 求女性临时工每小时工资的算术平均数。

(2) 求女性临时工每小时工资的加权算术平均数。

表 1—2　　女性临时工的小时工资与劳动者人数 (1993 年)

都·县	女性临时工小时工资 (日元)	女性临时工劳动者人数 (千人)
茨　城	837	60
枥　木	809	33
群　马	807	36
埼　玉	851	152
千　叶	874	113
东　京	993	279
神奈川	890	191

资料来源: 劳动省《工资结构基本统计调查》。

[解答]

(1) $\bar{X}=\frac{837+809+807+851+874+993+890}{7}=\frac{6\ 061}{7}=866$ (日元)

(2) 将女性临时工劳动人数作为加权算术平均的权数,得

$$\bar{X}_w=\frac{837\times60+809\times33+807\times36+851\times152+874\times113+993\times279+890\times191}{60+33+36+152+113+279+191}$$

$$=\frac{781\ 120}{864}=904\ (\text{日元})$$

由此可见,加权算术平均数 (904 日元) 大于算术平均数 (866 日元)。

3. 变化率

变化率的定义如下式:

$$\frac{X_t-X_{t-1}}{X_{t-1}}\ (t=2,3,\cdots,n) \tag{1—3}$$

[例题 1—3]

表 1—3 是日本 1994 年和 1995 年对主要出口对象国的出口额 (百万美元), 按出口额的大小 (1995 年) 顺序排列。求日本对各国的出口年增长率。

表 1—3　　日本的主要出口对象国　　单位:百万美元

国家(地区)	1994 年	1995 年
(1) 美国	117 560	120 859
(2) 韩国	24 359	31 291
(3) 中国台湾	23 792	28 969
(4) 中国香港	25 740	27 775
(5) 新加坡	19 605	23 001
(6) 中国大陆	18 682	21 931
(7) 德国	17 788	20 317

资料来源:日本关税协会《外国贸易概况》。

[解答]

根据（1—3）式，进行计算。

（1）美国

$$\frac{X_t - X_{t-1}}{X_{t-1}} = \frac{1995\text{ 年出口额} - 1994\text{ 年出口额}}{1994\text{ 年出口额}}$$

$$= \frac{120\ 859 - 117\ 560}{117\ 560} = 0.028 = 2.8\%$$

（2）韩国

$$\frac{31\ 291 - 24\ 359}{24\ 359} = 0.285 = 28.5\%$$

（3）中国台湾

$$\frac{28\ 969 - 23\ 792}{23\ 792} = 0.218 = 21.8\%$$

（4）中国香港

$$\frac{27\ 775 - 25\ 740}{25\ 740} = 0.079 = 7.9\%$$

（5）新加坡

$$\frac{23\ 001 - 19\ 605}{19\ 605} = 0.173 = 17.3\%$$

（6）中国大陆

$$\frac{21\ 931 - 18\ 682}{18\ 682} = 0.174 = 17.4\%$$

（7）德国

$$\frac{20\ 317 - 17\ 788}{17\ 788} = 0.142 = 14.2\%$$

可以看出，除了美国较为特殊，暂不考虑外，日本对亚洲新兴工业国(地区)(韩国、中国台湾、中国香港、新加坡）和中国大陆的出口数额较大，而且增长较为显著。

4. 几何平均

几何平均（geometric mean）是 n 个数据连乘积的 n 次方根，其定义如下式：

$$G=\sqrt[n]{X_1\times X_2\times\cdots\times X_n} \tag{1—4}$$

运用电子计算器，就可以简单计算。此外，对（1—4）式的两边取对数，得到

$$\log G=\frac{1}{n}(\log X_1+\log X_2+\cdots+\log X_n)=\frac{1}{n}\sum\log X \tag{1—5}$$

利用这种关系虽然也是计算几何平均数的方法，但是在目前计算器和计算机已经非常普及的情况下，从（1—4）式直接计算 G，可能更好。

需要注意的是，几何平均数是有弱点的。数据中只要有一个为零，根就会变为零，几何平均数就无法计算。而且，如果数据中有负值也无法计算。

一般地，几何平均数比较适合于经济增长率、工资上升率等增长率的平均数的计算。这种情况下的数据，应采用对上一年度的比值这样的形式。例如，从 100 变到 112 时，应采用 1.12(即 112/100)作为数据。

[例题 1—4]

求以下各数据的几何平均数。

(1) 9　25

(2) 2　4　8

(3) 3　4　6　18

(4) 3　6　12　24　48

[解答]

根据（1—4）式计算得：

(1) $G=\sqrt{9\times 25}=\sqrt{225}=15$

(2) $G=\sqrt[3]{2\times 4\times 8}=\sqrt[3]{64}=4$

(3) $G=\sqrt[4]{3\times 4\times 6\times 18}=\sqrt[4]{1\ 296}=6$

(4) $G=\sqrt[5]{3\times 6\times 12\times 24\times 48}=\sqrt[5]{248\ 832}=12$

[例题 1—5]

表 1—4 显示了日本、美国、韩国、中国台湾的出口增长率（实际）。利用几何平均数，求各自的出口平均增长率。

表 1—4 各国(地区)的出口增长率(%)

年度	日本	美国	韩国	中国台湾
1991	5.4	6.3	11.8	12.8
1992	4.9	6.6	11.0	5.3
1993	1.3	2.9	11.3	7.2
1994	4.6	8.2	16.2	5.5
1995	5.0	8.9	19.0	12.9

资料来源：经济企划厅《海外经济数据》。

[解答]

将数据换算成比值，根据 (1—4)式进行计算。

(1) 日本

$$G=\sqrt[5]{1.054\times1.049\times1.013\times1.046\times1.050}=\sqrt[5]{1.230\ 1}=1.042$$

出口平均增长率 $=1.042-1=4.2\%$

(2) 美国

$$G=\sqrt[5]{1.063\times1.066\times1.029\times1.082\times1.089}=\sqrt[5]{1.373\ 9}=1.066$$

出口平均增长率 $=1.066-1=6.6\%$

(3) 韩国

$$G=\sqrt[5]{1.118\times1.110\times1.113\times1.162\times1.190}=\sqrt[5]{1.909\ 9}=1.138$$

出口平均增长率 $=1.138-1=13.8\%$

(4) 中国台湾

$$G=\sqrt[5]{1.128\times1.053\times1.072\times1.055\times1.129}=\sqrt[5]{1.516\ 6}=1.087$$

出口平均增长率 $=1.087-1=8.7\%$

[例题 1—6]

表 1—5 显示了日本实际国内生产总值（实际 GDP）的变化。

(1) 计算包括“神武景气”和“岩户景气”在内的 1955—1961 年的实际经济增长率（实际 GDP 的增长率）。

（2）计算 1965—1970 年，钢铁、汽车、机械等出口急剧增加，经济持续增长的实际经济增长率。

（3）计算 1987—1990 年，以地价、股价飞涨为中心的景气扩张，即所谓的“泡沫”时期的实际经济增长率。

（4）计算 1991—1995 年，“泡沫”崩溃以后的实际经济增长率。

表 1—5　　日本的实际国内生产总值（GDP）的变化　　单位：兆日元

年度	实际 GDP	年度	实际 GDP	年度	实际 GDP
1955	47.9	1970	190.4	1985	345.4
1956	51.0	1971	200.1	1986	356.3
1957	54.9	1972	218.2	1987	373.2
1958	58.9	1973	229.3	1988	395.5
1959	65.5	1974	228.2	1989	413.1
1960	73.5	1975	237.3	1990	436.1
1961	82.1	1976	246.3	1991	449.8
1962	88.3	1977	257.4	1992	451.4
1963	97.5	1978	271.3	1993	452.5
1964	106.8	1979	285.3	1994	454.6
1965	113.4	1980	292.7	1995	465.0
1966	125.9	1981	301.5		
1967	139.8	1982	310.8		
1968	157.1	1983	318.7		
1969	175.9	1984	331.8		

说明：按 1990 年价格计算。

资料来源：经济企划厅《国民经济计算年报》。

［提示］

跨越数年的平均变化率的计算方法。

现假设 Y 从0期到 n 期，按照同样的变化率 g 变化，则

$$Y_1=(1+g)Y_0$$

$$Y_2=(1+g)Y_1$$

$$\vdots$$

$$Y_n=(1+g)Y_{n-1}$$

这里，用 Y_0 来表示 Y_n，则

$$Y_n=(1+g)(1+g)\cdots(1+g)Y_0$$

$$Y_n=(1+g)^nY_0$$

将该式整理，得平均变化率 g 为

$$(1+g)^n=\frac{Y_n}{Y_0}$$

$$1+g=\left(\frac{Y_n}{Y_0}\right)^{\frac{1}{n}}$$

$$g=\left(\frac{Y_n}{Y_0}\right)^{\frac{1}{n}}-1$$

$$g=\sqrt[n]{\frac{Y_n}{Y_0}}-1 \tag{1—6}$$

利用（1—6)式，可以简单地计算平均变化率。

［解答］

（1）根据（1—6）式，计算1955—1961年6年间的实际经济增长率。

$$\sqrt[6]{\frac{1961\text{ 年度实际 GDP}}{1955\text{ 年度实际 GDP}}}-1=\sqrt[6]{\frac{82.1}{47.9}}-1=\sqrt[6]{1.714}-1=1.094-1$$

$$=0.094=9.4\%$$

（2）同样，根据（1—6）式，计算1965—1970年5年间的实际经济增长率。

$$\sqrt[5]{\frac{1970\text{ 年度实际 GDP}}{1965\text{ 年度实际 GDP}}}-1=\sqrt[5]{\frac{190.4}{113.4}}-1=\sqrt[5]{1.679}-1=1.109-1$$

$$=0.109=10.9\%$$

(3) 根据 (1—6) 式，计算 1987—1990 年 3 年间的实际经济增长率。

$$\sqrt[3]{\frac{1990\text{ 年度实际 GDP}}{1987\text{ 年度实际 GDP}}}-1=\sqrt[3]{\frac{436.1}{373.2}}-1=\sqrt[3]{1.169}-1=1.053-1$$

$$=0.053=5.3\%$$

(4) 根据 (1—6) 式，计算 1991—1995 年 4 年间的实际经济增长率。

$$\sqrt[4]{\frac{1995\text{ 年度实际 GDP}}{1991\text{ 年度实际 GDP}}}-1=\sqrt[4]{\frac{465.0}{449.8}}-1=\sqrt[4]{1.034}-1=1.008-1$$

$$=0.008=0.8\%$$

5. 移动平均

所谓移动平均 (moving average)，就是对时间序列数据中的前后数据求平均，将不必要的变动 (循环变动、季节变动和不规则变动) 平滑化 (smoothing)，也即剔除这些变动，从而发现长期变化方向的一种方法。每隔 3 个月的季度数据 (quarterly data)、每个月的月度数据 (monthly data) 中存在着季度和月份中固有变化的影响，利用移动平均可以消除这些季节变动，有助于理解长期变化趋势。同样，循环变动和不规则变动也可以通过移动平均来消除，计算平滑的长期变动。

通常，移动平均大多用简单的奇数项来计算，下面是 3 项移动平均和 5 项移动平均的定义。

3 项移动平均：

$$\bar{X}_t=\frac{X_{t-1}+X_t+X_{t+1}}{3} \qquad (1—7)$$

5 项移动平均：

$$\bar{X}_t=\frac{X_{t-2}+X_{t-1}+X_t+X_{t+1}+X_{t+2}}{5} \qquad (1—8)$$

另一方面，在偶数项季节数据的情况下，可以按以下的方法来计算，即首先计算两个 4 项移动平均，然后再计算这 2 项移动平均。这种方法称做移动平均的中心化，公式(1—9) 称做中心化 4 项移动平均。同样，由于月度数据也是偶数项，因而称做中心化 12 项移动平均。

时序数据　　4项移动平均　　　　　2项移动平均

$$
\begin{array}{lll}
X_{t-2} & & \\
X_{t-1} & \dfrac{X_{t-2}+X_{t-1}+X_t+X_{t+1}}{4} & \rbrace \dfrac{X_{t-2}+2X_{t-1}+2X_t+2X_{t+1}+X_{t+2}}{8} \\
X_t & \dfrac{X_{t-1}+X_t+X_{t+1}+X_{t+2}}{4} & \\
X_{t+1} & & =\dfrac{0.5X_{t-2}+X_{t-1}+X_t+X_{t+1}+0.5X_{t+2}}{4} \\
X_{t+2} & &
\end{array}
\tag{1—9}
$$

此外，在选定移动平均的项数时，如果事先已经知道像图1—1这样的周期(period)，一般要使项数与周期相一致。

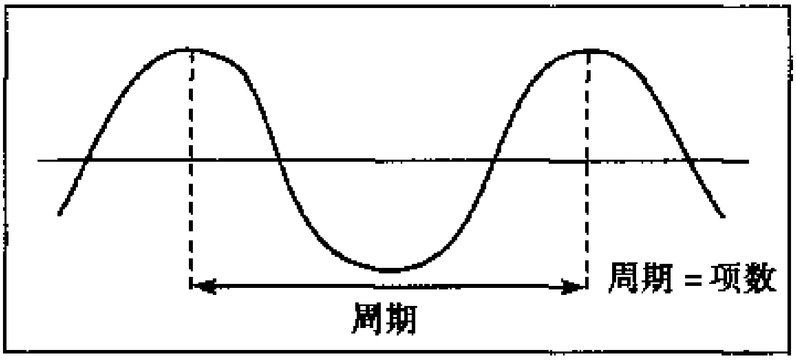


图 1—1

[例题 1—7]

表1—6显示了香港股价指数（恒生指数）的变化。求3年移动平均，并且与原数列一道画出图形。

表 1—6　　　　香港的股价指数

（1964年12月30日为100）

年度	股价指数 X
1980	1 474
1981	1 406
1982	784

续前表

年度	股价指数 X
1983	875
1984	1 200
1985	1 752
1986	2 568
1987	2 303
1988	2 687
1989	2 837
1990	3 025
1991	4 297
1992	5 512
1993	11 888
1994	8 191
1995	10 073

说明：年末值。

资料来源：中国香港政府，*Hong Kong Monthly Digest of Statistics*。

[解答]

根据（1—7）式，可计算 3 年移动平均。数据的起点和终点分别失去—项。计算结果见表 1—7 和图 1—2。

表 1—7　　香港的股价指数与 3 年移动平均

年度	股价指数	3 年之和	3 年移动平均
1980	1 474	—	—
1981	1 406	3 664	1 221.3
1982	784	3 065	1 021.7
1983	875	2 859	953.0
1984	1 200	3 827	1 275.7

续前表

年度	股价指数	3 年之和	3 年移动平均
1985	1 752	5 520	1 840.0
1986	2 568	6 623	2 207.7
1987	2 303	7 558	2 519.3
1988	2 687	7 827	2 609.0
1989	2 837	8 549	2 849.7
1990	3 025	10 159	3 386.3
1991	4 297	12 834	4 278.0
1992	5 512	21 697	7 232.3
1993	11 888	25 591	8 530.3
1994	8 191	30 152	10 050.7
1995	10 073	—	—

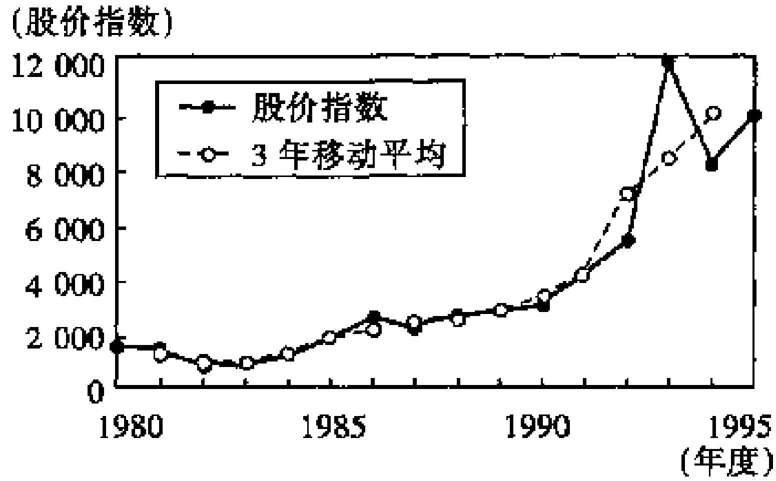


图 1—2　香港的股价指数与 3 年移动平均

[例题 1—8]

表 1—8 显示了从 1991 年第一季度开始到 1995 年第二季度为止，日本全国百货店的销售额。求中心化 4 项移动平均，并且与原数列一道画出图形。

表 1—8　　百货店销售额（全国）　　单位：100 亿日元

年·季	百货店销售额	年·季	百货店销售额
1991 年第 1 季度	219	1994 年第 1 季度	202
第 2 季度	225	第 2 季度	205
第 3 季度	232	第 3 季度	212
第 4 季度	296	第 4 季度	258
1992 年第 1 季度	221	1995 年第 1 季度	193
第 2 季度	224	第 2 季度	199
第 3 季度	227	第 3 季度	208
第 4 季度	280	第 4 季度	257
1993 年第 1 季度	208	1996 年第 1 季度	206
第 2 季度	210	第 2 季度	206
第 3 季度	215		
第 4 季度	263		

资料来源：日本百货店协会《日本百货店协会统计年报》。

[解答]

根据（1—9）式，计算中心化 4 项移动平均。计算过程中，数据的起点和终点分别失去两项，见表 1—9。

从图 1—3 可以看出，通过计算中心化 4 项移动平均，使得原数列变得平滑，也就是消除了季节变动。

表 1—9　　百货店销售额（全国）的中心化 4 项移动平均　　单位：100 亿日元

年·季	中心化 4 项移动平均	年·季	中心化 4 项移动平均
1991 年第 1 季度	—	1994 年第 1 季度	220.9
第 2 季度	—	第 2 季度	219.9
第 3 季度	243.2	第 3 季度	218.1
第 4 季度	243.4	第 4 季度	216.2
1992 年第 1 季度	242.6	1995 年第 1 季度	215.0
第 2 季度	240.0	第 2 季度	214.4

续前表

年·季	中心化4项移动平均	年·季	中心化4项移动平均
第3季度	236.4	第3季度	215.9
第4季度	233.0	第4季度	218.4
1993年第1季度	229.8	1996年第1季度	—
第2季度	226.1	第2季度	—
第3季度	223.2		
第4季度	221.9		

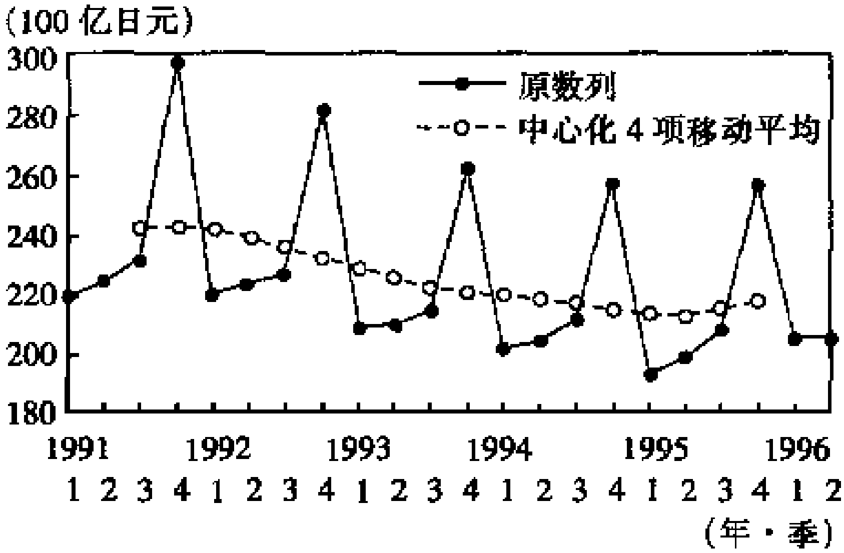


图1—3 百货店销售额原数列与中心化4项移动平均

6. 方差与标准差

为了了解数据的结构，有必要考察数据的集中趋势和分散的程度。对于集中的趋势，我们从前面学过的算术平均中已经大体有所了解，面对于分散的程度，通过对方差（variance）与标准差（standard deviation），以及下一节将要介绍的变动系数的计算，能够得到很多信息。

方差的计算方法是，先将每个数据与算术平均数之差（即离差）的平方相加求和，再除以样本数减一。面标准差是方差的正的平方根。由于方差是通过平方计算的，它与原数据的次数有所不同，而标准差由于是方差的平方根，因而又与

原数据的次数相同。因此，标准差与原数据的单位相同，而方差不附加单位。

方差 s^2 与标准差 s 的定义分别如下式：

方差：

$$s^2 = \frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2}{n-1}$$

$$= \frac{1}{\text{样本数} - 1}\sum(X - \text{算术平均})^2$$

$$= \frac{1}{n-1}\sum(X - \bar{X})^2 \tag{1—10}$$

标准差：

$$s = \sqrt{\text{方差}}$$

$$= \sqrt{s^2} \tag{1—11}$$

式中，$\sum(X - \bar{X})^2$ 称做离差的平方和。

方差与标准差越大，意味着数据的分散程度越大；相反，方差与标准差越小，则意味着数据的分散程度越小，也即向平均值（算术平均）的集中程度越高。

另外，标准差还有以下这样便利的特点。现在假定数据与多数自然现象和社会现象相似，服从统计学中最重要的正态分布（normal distribution），以算术平均值 $\bar{X}$ 为中心，左右取 $1s$ 范围，这一部分包含 68.3% 的数据。同样，取 $2s$ 范围，包含 95.4% 的数据，再取 $3s$，则包含 99.7% 的数据。图 1—4 显示了标准差的大小与所含数据比例的对应关系。

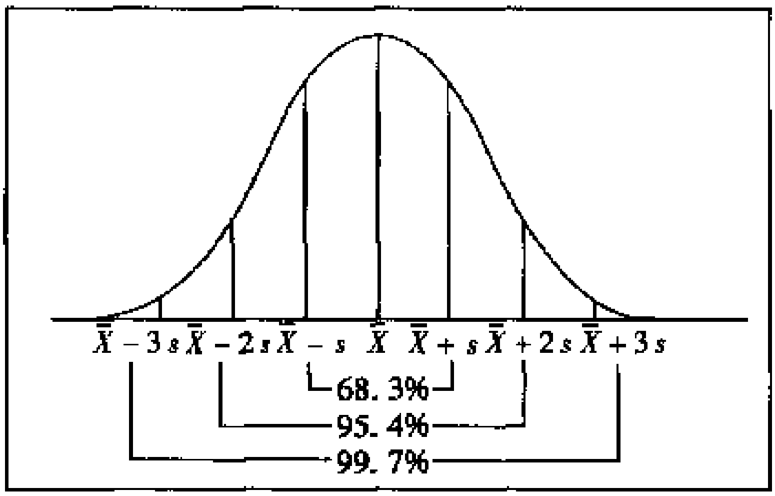


图 1—4 正态分布中标准差与其相应的范围中所含数据的比例

[例题 1—9]

表 1—10 是 1996 年 14 家电器公司的销售额中出口额所占的比重。

(1) 求方差 s^2。

(2) 求标准差 s。

表 1—10　　1996 年电器公司的销售额中出口额所占的比重

企　业	销售额中出口额所占比重(%)	企　业	销售额中出口额所占比重(%)
日立	24	松下	33
东芝	31	夏普	44
三菱	23	索尼	65
富士	15	三洋	31
NEC	20	日本 JVC	50
富士通	13	电通	19
冲电工业	19	京瓷	33

资料来源：东洋经济新报社《公社四季报》。

[解答]

(1) 将数据代入工作表中，进行计算（见表 1—11）。

表 1—11

X	$X-\bar{X}$	$(X-\bar{X})^2$
24	−6	36
31	1	1
23	−7	49
15	−15	225
20	−10	100
13	−17	289
19	−11	121
33	3	9

续前表

X	$X-\bar{X}$	$(X-\bar{X})^2$
44	14	196
65	35	1 225
31	1	1
50	20	400
19	−11	121
33	3	9
420	0	2 782
↑ $\sum X$	↑ $\sum(X-\bar{X})$	↑ $\sum(X-\bar{X})^2$

求算术平均数 $\bar{X}$。

$$\bar{X} = \frac{1}{n}\sum X = \frac{1}{14} \times 420 = 30$$

根据（1—10)式，求方差 s^2。

$$s^2 = \frac{1}{n-1}\sum(X-\bar{X})^2 = \frac{1}{13} \times 2\ 782 = 214$$

(2) 根据（1—11)式，求标准差 s。

$$s = \sqrt{s^2} = \sqrt{214} = 14.6$$

7. 变动系数

变动系数（coefficient of variation）又称变异系数，它用标准差 s 除以算术平均数 $\bar{X}$ 的商来表示。变动系数 CV 的定义如下式：

$$CV = \frac{\text{标准差}}{\text{算术平均数}} = \frac{s}{\bar{X}} \qquad (1—12)$$

对于不同的数据组来说，由于各自的算术平均数不同，因此单纯根据各自的标准差，则无法比较分散程度。在比较鲸鱼和鲣鱼的重量变化幅度时，由于鲸鱼

的标准差显然较大，因此，不能通过标准差来进行比较。在这种情况下，可以通过变动系数来对不同数据组的分散程度进行相对比较。

例如，在比较不同的数据组 A 和 B 的变动系数时，如果 A 的系数较大，说明 A 与 B 相比，数据的分散程度更大。一般地，变动系数要求所使用的数据均为正数。而且，算出的数值要按百分数形式表示。

此外，如果算术平均数为零或接近于零，变动系数无法计算，或者说，变动系数成为一种暧昧的尺度，因此，这种情况也是需要加以注意的。

[例题 1—10]

设算术平均数 $\bar{X}$ 为 56，标准差为 7，求变动系数 CV。

[解答]

根据（1—12）式，变动系数 CV 为

$$CV=\frac{s}{\bar{X}}=\frac{7}{56}=0.125=12.5\%$$

[例题 1—11]

表 1—12 列出了日本、德国和法国按美元表示的汇率变化情况。

(1) 计算各国汇率的算术平均数 $\bar{X}$ 和标准差 s。

(2) 计算各国汇率的变动系数 CV，并对结果进行比较。

表 1—12　日本、德国、法国的汇率　单位：美元

年·季	日本（日元）	德国（马克）	法国（法郎）
1991 年第 1 季度	140.55	1.697	5.746
第 2 季度	138.15	1.810	6.135
第 3 季度	132.95	1.667	5.682
第 4 季度	125.25	1.519	5.190
1992 年第 1 季度	133.05	1.643	5.572
第 2 季度	125.55	1.523	5.116
第 3 季度	119.25	1.412	4.774
第 4 季度	124.65	1.621	5.527

续前表

年·季	日本（日元）	德国（马克）	法国（法郎）
1993 年第 1 季度	115.35	1.608	5.593
第 2 季度	106.51	1.706	5.746
第 3 季度	105.10	1.634	5.695
第 4 季度	111.89	1.738	5.918
1994 年第 1 季度	104.15	1.674	5.717
第 2 季度	99.80	1.588	5.439
第 3 季度	99.45	1.551	5.298
第 4 季度	100.85	1.549	5.338
1995 年第 1 季度	90.35	1.373	4.806
第 2 季度	84.75	1.382	4.840
第 3 季度	98.15	1.426	4.921
第 4 季度	102.88	1.437	4.906
1996 年第 1 季度	106.48	1.475	5.032

资料来源：UN，*Monthly Bulletin of Statistics*.

[解答]

(1) 日本

$$\bar{X}=\frac{\sum X}{n}=\frac{2\ 365.11}{21}=112.62$$

$$s=\sqrt{\frac{\sum(X-\bar{X})^2}{n-1}}=\sqrt{\frac{5\ 042.5}{20}}=15.88\text{（日元）}$$

德国

$$\bar{X}=\frac{\sum X}{n}=\frac{33.033}{21}=1\ 573$$

$$s=\sqrt{\frac{\sum(X-\bar{X})^2}{n-1}}=\sqrt{\frac{0.306\ 5}{20}}=0.123\ 8\text{（马克）}$$

法国

$$\bar{X}=\frac{\sum X}{n}=\frac{112.991}{21}=5.381$$

$$s=\sqrt{\frac{\sum(X-\bar{X})^2}{n-1}}\doteq\sqrt{\frac{3.198}{20}}=0.399\,9\text{（法郎）}$$

(2) 根据 (1—12) 式，求变动系数。

日本

$$CV=\frac{s}{\bar{X}}=\frac{15.88}{112.62}=0.141\,0=14.10\%$$

德国

$$CV=\frac{s}{\bar{X}}=\frac{0.123\,8}{1.573}=0.078\,7=7.87\%$$

法国

$$CV=\frac{s}{\bar{X}}=\frac{0.399\,9}{5.381}=0.074\,3=7.43\%$$

由此可见，汇率变动的相对程度，由大到小分别为日本、德国和法国。

8. 标准化变量

标准化变量 (standardized variable)，又称基准化变量，它是用来测量某个数据的数值与算术平均数 $\bar{X}$ 的偏离程度，是标准差 s 的多少倍。借此可以看出该数据在全体数据中所处的位置。标准化变量 z 的定义如下式：

$$z=\frac{X-\text{算术平均数}}{\text{标准差}}=\frac{X-\bar{X}}{s} \tag{1—13}$$

通过上式进行标准化（或基准化），可以看出不管什么样的数据，由于算术平均数 $\bar{X}$ 可以变换为零，方差和标准差分别可以变换为 1，因此具有不同的算术平均的数据组，可以进行相互比较。

日本大学的入学模拟考试在公布相对的成绩时，通常利用偏差值，这种偏差值实际上就是标准化变量的一种应用，其公式如下：

$$\text{偏差值}=50+\frac{\text{自己得分}-\text{平均分}}{\text{标准差}}\times 10$$

$$=50+\frac{X-\bar{X}}{s}\times 10 \tag{1—14}$$

$$= 50 + 标准化变量 \times 10 = 50 + z \times 10 \qquad (1—15)$$

可见，这里的偏差值就是设算术平均数为 50，标准差为 10，来显示分数的分布状况。

同样，智商指数（IQ）也是标准化变量的应用，它显示的是算术平均数为 100，标准差为 15 时，分数的分布状况。智商指数的定义如下式：

$$IQ = 100 + \frac{自己得分 - 平均分}{标准差} \times 15$$

$$= 100 + \frac{X - \bar{X}}{s} \times 15 \qquad (1—16)$$

$$= 100 + 标准化变量 \times 15 = 100 + z \times 15 \qquad (1—17)$$

图 1—5 在正态分布的假设下，显示了偏差值、智商指数和算术平均数、标准差之间的相互关系。

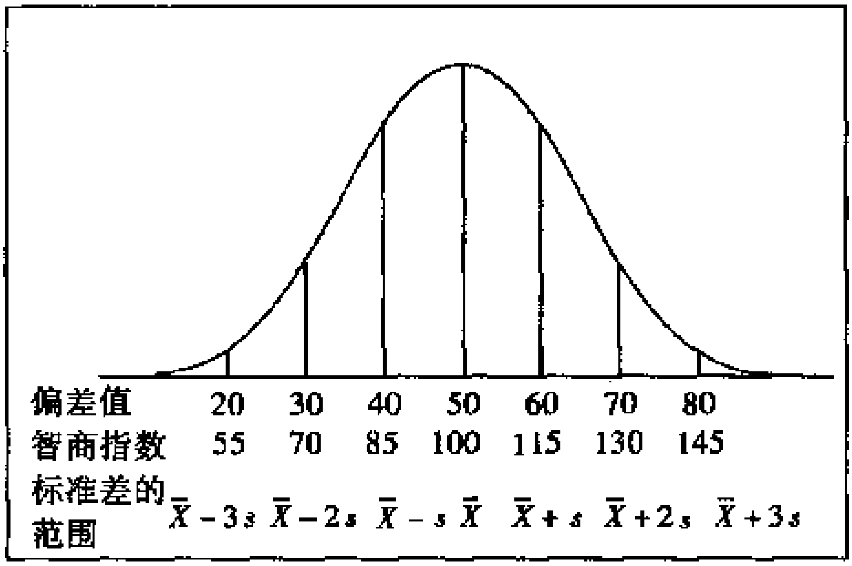


图 1—5　偏差值与智商指数

［例题 1—12］

经济系的小王，在期末考试中，宏观经济学得 82 分，微观经济学得 69 分。宏观经济学的平均成绩是 72 分，标准差是 8，微观经济学的平均成绩是 61 分，标准差是 5。

（1）计算标准化变量 z，并回答小王的宏观经济学和微观经济学成绩哪一个更好。

（2）求小王宏观经济学和微观经济学的偏差值。

[解答]

(1) 根据 (1—13) 式计算标准化变量 z。

宏观经济学为

$$z=\frac{X-\bar{X}}{s}=\frac{82-72}{8}=1.25$$

微观经济学为

$$z=\frac{X-\bar{X}}{s}=\frac{69-61}{5}=1.60$$

由于微观经济学的标准化变量比宏观经济学的标准化变量要大，因此，小王的微观经济学成绩处于上等。

(2) 根据 (1—15) 式计算偏差值。

宏观经济学为：

$$偏差值=50+z\times10=50+1.25\times10=62.5$$

微观经济学为：

$$偏差值=50+z\times10=50+1.60\times10=66.0$$

从偏差值的比较中可以看出，小王的微观经济学的成绩相对来说要好一些。

9. 相关系数

所谓相关系数 (correlation coefficient) 是用来测量诸如收入与消费、气温和啤酒的消费量、汇率与牛肉的进口价格等两个变量 X、Y 之间相互关系的大小和方向 (正或负)的系数。通过计算相关系数，可以知道 X 和 Y 之间具有多大程度的线性 (linear) 关系。相关系数 R 的定义如下式：

$$R=\frac{\sum(X-\bar{X})(Y-\bar{Y})}{\sqrt{\sum(X-\bar{X})^2\sum(Y-\bar{Y})^2}} \quad [定义式](1—18)$$

$$=\frac{n\sum XY-\sum X\times\sum Y}{\sqrt{[n\sum X^2-(\sum X)^2][n\sum Y^2-(\sum Y)^2]}} \quad [计算式](1—19)$$

相关系数 R 的取值范围为 $-1\leqslant R\leqslant1$，R 的取值具有以下不同的含义：

(1) $R=1$ → 完全正相关 (perfect correlation)

(2) $R>0$ → 正相关 (positive correlation)

(3) $R=0$ → 不相关 (no correlation)

(4) $R<0$ → 负相关 (negative correlation)

(5) $R=-1$ → 完全负相关 (perfect correlation)

正相关指的是当 X 增加时，Y 也增加；相反，负相关指的是当 X 增加时，Y 减少。这种关系如图 1—6 所示，该图称做散点图 (scatter diagram) 或相关图 (correlation diagram)。相关关系的理论是由英国遗传学者高尔顿 (F. Galton，1822—1911 年) 创始的。

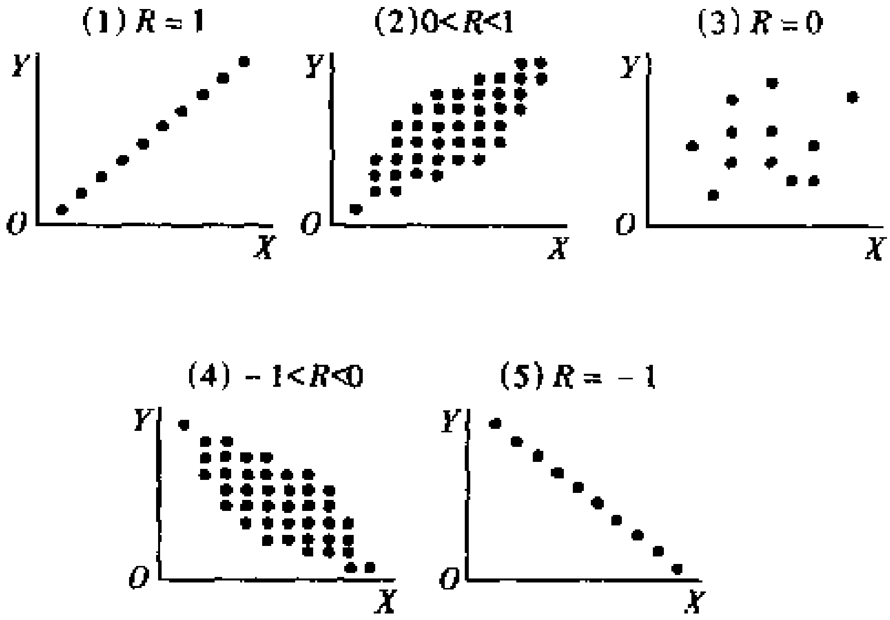


图 1—6　相关系数与散点图

顺便提一下，在相关关系中，有时有因果关系，有时则没有，请读者注意区别。所谓因果关系，指的是原因明确地存在，并且由此产生了结果。但是，即使在没有因果关系的情况下，为了看看相关关系的大小，也需要进行相关分析。

[例题 1—13]

根据下列数据，计算相关系数。

X	6	1	3	9	7
Y	5	2	4	8	6

[解答]

将数据代入工作表中，进行计算（见表 1—13）。

表 1—13

X	Y	XY	X^2	Y^2
6	5	30	36	25
1	2	2	1	4
3	4	12	9	16
9	8	72	81	64
7	6	42	49	36
26	25	158	176	145
↑	↑	↑	↑	↑
$\sum X$	$\sum Y$	$\sum XY$	$\sum X^2$	$\sum Y^2$

根据（1—19）式，求相关系数 R。

$$R = \frac{n\sum XY - \sum X \times \sum Y}{\sqrt{[n\sum X^2 - (\sum X)^2][n\sum Y^2 - (\sum Y)^2]}}$$

$$= \frac{(5 \times 158) - (26 \times 25)}{\sqrt{[(5 \times 176) - (26)^2][(5 \times 145) - (25)^2]}}$$

$$= \frac{140}{142.83}$$

$$= 0.98$$

可见，X 与 Y 有很大的相关性。

[例题 1—14]

表 1—14 中的数据列出了东京市区 1994 年 1—8 月的月平均气温 X 和每户平均啤酒消费量 Y。

（1）画出散点图。

（2）计算相关系数 R。

表 1—14　　东京市区月平均气温和每户平均啤酒消费量（1994 年）

月　份	1月	2月	3月	4月	5月	6月	7月	8月
月平均气温 X（℃）	5.5	6.6	8.1	15.8	19.5	22.4	28.3	28.9
啤酒消费量 Y（l）	2.38	3.85	4.41	5.67	5.44	6.03	8.15	6.87

资料来源：气象厅调查。

总务厅统计局《家庭调查报告》。

[解答]

（1）散点图见图 1—7。

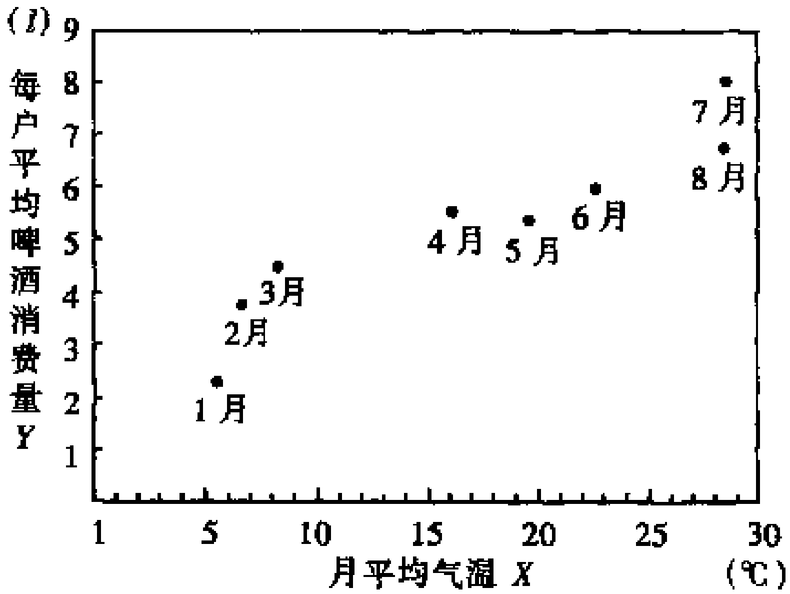


图 1—7　东京市区月平均气温和每户平均啤酒消费量（1994 年）

（2）将数据代人工作表，进行计算（见表 1—15）。

表 1—15

X	Y	XY	X^2	Y^2
5.5	2.38	13.090	30.25	5.664 4
6.6	3.85	25.410	43.56	14.822 5
8.1	4.41	35.721	65.61	19.448 1
15.8	5.67	89.586	249.64	32.148 9
19.5	5.44	106.080	380.25	29.593 6
22.4	6.03	135.072	501.76	36.360 9

续前表

X	Y	XY	X^2	Y^2
28.3	8.15	230.645	800.89	66.422 5
28.9	6.87	198.543	835.21	47.196 9
135.1	42.80	834.147	2 907.17	251.657 8
↑ $\sum X$	↑ $\sum Y$	↑ $\sum XY$	↑ $\sum X^2$	↑ $\sum Y^2$

根据（1—19）式，求相关系数 R。

$$R = \frac{n\sum XY - \sum X \times \sum Y}{\sqrt{[n\sum X^2 - (\sum X)^2][n\sum Y^2 - (\sum Y)^2]}}$$

$$= \frac{(8 \times 834.147) - (135.1 \times 42.80)}{\sqrt{[(8 \times 2\,907.17) - (135.1)^2][(8 \times 251.657\,8) - (42.80)^2]}}$$

$$= \frac{890.896}{952.934}$$

$$= 0.934\,9$$

可见，X 和 Y 之间存在很强的正相关关系。如果气温上升则导致啤酒消费增加，这种因果关系也比较明确。

[例题 1—15]

表 1—16 的数据为中国 30 个省市自治区 1994 年人均 GDP（X）与第一产业中就业者比例（Y）。根据该数据，计算两者的相关系数 R。

表 1—16　　　中国的人均 GDP 与第一产业中就业者比例（1994 年）

地　　域	X 人均 GDP （100 元）	Y 第一产业中就业者比例 （%）
1. 北京市	103	11
2. 天津市	82	18
3. 河北省	34	54
4. 山西省	28	44

续前表

地　　域	X 人均 GDP （100 元）	Y 第一产业中就业者比例 （%）
5. 内蒙古自治区	30	53
6. 辽宁省	64	31
7. 吉林省	38	46
8. 黑龙江省	44	37
9. 上海市	152	9
10. 江苏省	58	43
11. 浙江省	61	44
12. 安徽省	25	62
13. 福建省	54	51
14. 江西省	26	56
15. 山东省	45	56
16. 河南省	24	62
17. 湖北省	33	54
18. 湖南省	27	63
19. 广东省	64	41
20. 广西壮族自治区	28	68
21. 海南省	48	61
22. 四川省	25	65
23. 贵州省	16	75
24. 云南省	25	77
25. 西藏自治区	20	78
26. 陕西省	24	60
27. 甘肃省	19	59
28. 青海省	29	60
29. 宁夏回族自治区	27	60
30. 新疆维吾尔自治区	40	57

资料来源：中国国家统计局《中国统计年鉴》。

[解答]
省略工作表。

$$\sum X = 1\ 293$$

$$\sum Y = 1\ 555$$

$$\sum XY = 54\ 391$$

$$\sum X^2 = 79\ 671$$

$$\sum Y^2 = 89\ 283$$

根据（1—19）式，计算相关系数 R。

$$R = \frac{n\sum XY - \sum X \times \sum Y}{\sqrt{[n\sum X^2 - (\sum X)^2][n\sum Y^2 - (\sum Y)^2]}}$$

$$= \frac{30 \times 54\ 391 - 1\ 293 \times 1\ 555}{\sqrt{[30 \times 79\ 671 - 1\ 293^2] \times [30 \times 89\ 283 - 1\ 555^2]}}$$

$$= \frac{-378\ 885}{432\ 536}$$

$$= -0.876$$

可见，X 和 Y 之间存在着负相关关系。

随着人均 GDP 的增加，或者说随着一个国家经济的发展，就业结构也会发生相应的变化。第一产业中就业人数的比例会下降，而第二和第三产业的就业比例会上升。这一经验事实称做配第－克拉克法则。通过上述例题可以看出，即使在中国国内经济中，在省这一级，配第－克拉克法则仍然在发挥作用。

10. 相关系数的检验

上一节计算出来的相关系数在多大程度上值得信赖，需要进行检验。计算出来的相关系数，如果大于表 1—17 中所列示的相关系数，则两个变量之间存在显著的相关关系。表 1—17 将显著水平（level of signification）为 10%、5%、1%和 0.1%的相关系数，与不同的自由度（样本数 $-2 = n-2$）相对应地列示出来。显著水平越小，检验越严格。选择哪种显著水平需要由分析者根据研究内容自己决定。

表 1—17　　相关系数的检验

自由度 $n-2$	显著水平 10%	显著水平 5%	显著水平 1%	显著水平 0.1%
1	0.988	0.997	1.000	1.000
2	0.900	0.950	0.990	0.999
3	0.805	0.878	0.959	0.991
4	0.729	0.811	0.917	0.974
5	0.669	0.754	0.874	0.951
6	0.621	0.707	0.834	0.925
7	0.582	0.666	0.798	0.898
8	0.549	0.632	0.765	0.872
9	0.521	0.602	0.735	0.847
10	0.497	0.576	0.708	0.823
11	0.476	0.553	0.684	0.801
12	0.457	0.532	0.661	0.780
13	0.441	0.514	0.641	0.760
14	0.426	0.497	0.623	0.742
15	0.412	0.482	0.606	0.725
16	0.400	0.468	0.590	0.708
17	0.389	0.456	0.575	0.693
18	0.378	0.444	0.561	0.679
19	0.369	0.433	0.549	0.665
20	0.360	0.423	0.537	0.652
25	0.323	0.381	0.487	0.597
30	0.296	0.349	0.449	0.554
35	0.275	0.325	0.418	0.519
40	0.257	0.304	0.393	0.490
50	0.231	0.273	0.354	0.443
60	0.211	0.250	0.325	0.408
80	0.183	0.217	0.283	0.357
100	0.164	0.195	0.254	0.321
200	0.116	0.138	0.181	0.230
500	0.073	0.088	0.115	0.146

说明：两侧检验。

所谓显著水平，指的是很少会发生的概率，这里相当于相关系数为零（$R=0$），也即相当于不相关的概率。例如，计算出来的相关系数的绝对值，如果大于表1—17中显著水平为1%的相关系数，那就意味着，该相关系数为零的概率，也即不相关的概率，小于1%，因此存在着显著的相关。

［例题1—16］

下面（1）~（5）中，是否存在着显著的相关关系？按显著水平为10%、5%、1%、0.1%来进行检验。这里，R 是相关系数，n 是样本数。

（1）$R=0.900$（$n=5$）

（2）$R=0.612$（$n=10$）

（3）$R=0.724$（$n=19$）

（4）$R=0.269$（$n=37$）

（5）$R=0.308$（$n=102$）

［解答］

先计算自由度，然后根据表1—17进行检验。显著为○，不显著为×。

表1—18

计算出的相关系数	样本数 n	自由度 $n-2$	显著水平			
			10%	5%	1%	0.1%
(1) 0.900	5	3	○	○	×	×
(2) 0.612	10	8	○	×	×	×
(3) 0.724	19	17	○	○	○	○
(4) 0.269	37	35	×	×	×	×
(5) 0.308	102	100	○	○	○	×

［例题1—17］

表1—19列示的是“广场协议”以后，日本的汇率（1美元相当于若干日元）与汽车出口数量（万辆）之间的变化关系。

（1）计算相关系数 R。

（2）这两个变量之间是否显著相关？按显著水平为5%和1%进行检验。

表 1—19　　汇率与汽车出口数量

年度	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995
年均汇率 X（日元）	168	145	128	138	145	135	127	111	102	94
汽车出口数量 Y（万辆）	661	631	610	588	583	575	567	502	446	379

资料来源：IMF，《国际金融统计》（*International Financial Statistics*）；日本汽车工业协会《汽车统计月报》。

[解答]

（1）省略工作表。

$$\sum X = 1\ 293$$

$$\sum Y = 5\ 542$$

$$\sum XY = 732\ 776$$

$$\sum X^2 = 171\ 617$$

$$\sum Y^2 = 3\ 139\ 490$$

根据（1—19）式，求相关系数。

$$R = \frac{n\sum XY - \sum X \times \sum Y}{\sqrt{[n\sum X^2 - (\sum X)^2][n\sum Y^2 - (\sum Y)^2]}}$$

$$= \frac{(10 \times 732\ 776) - (1\ 293 \times 5\ 542)}{\sqrt{[10 \times 171\ 617 - (1\ 293)^2][10 \times 3\ 139\ 490 - (5\ 542)^2]}}$$

$$= \frac{161\ 954}{173\ 749}$$

$$= 0.932\ 1$$

（2）自由度 $= n - 2 = 10 - 2 = 8$

根据表 1—17，自由度为 8 的显著水平 5% 和 1% 的相关系数分别为 0.632 和 0.765，计算出来的相关系数（0.932 1）大于这两个系数，因此，变量之间存在显著相关。

1985 年“广场协议”之后，日元汇率急剧上升。随着日元升值的长期化，出

口价格（按美元计算）大幅上升，日本汽车的竞争力下降。从这个例题中可以看出，日本的汽车产业作为出口型产业的代表，受到了日元升值的极大影响。当然，汽车出口数量的减少，原因也不单纯是日元升值引起的价格竞争力降低，除此之外还有生产本地化的发展、出口的自主管制等方面的原因。

11. 斯皮尔曼秩相关系数

所谓斯皮尔曼秩相关系数（Spearman's rank correlation），是 1904 年提出来的一种方法，它考察的不是 X 和 Y 两组数据中的数值，而是顺序，借此来测算 X 和 Y 之间相关关系的强弱。其定义如下式：

$$R_s = 1 - \frac{6\sum(X-Y)^2}{n(n^2-1)} = 1 - \frac{6\sum d^2}{n(n^2-1)} \tag{1—20}$$

式中，n 是样本数，d 是X 和 Y 的顺序差（即 $X-Y$）。斯皮尔曼秩相关系数的取值范围和解释方法，与相关系数相同。从（1—20）式中可以看出，它的优点是计算方法非常简单，同时可靠性也很高。

当然，当数据顺序相同时，情况较为复杂，这可以用下面的公式来计算。

$$R_s = \frac{(\sum X^2 + \sum Y^2 - \sum d^2) \times \frac{n}{2} - T^2}{\sqrt{n\sum X^2 - T^2}\sqrt{n\sum Y^2 - T^2}} \tag{1—21}$$

式中，

$$d = X - Y$$

$$T = \frac{n(n+1)}{2}$$

[补充]

与上一节相关系数的检验相同，斯皮尔曼秩相关系数 R_s 的检验可以利用表 1—20。就是说，计算出来的 R_s 的绝对值，如果大于表 1—20 的数值，则存在显著相关。

表 1—20　　斯皮尔曼秩相关系数的检验

样本数 n	显著水平		
	10%	5%	1%
5	0.900	1.000	—
6	0.829	0.886	1.000
7	0.714	0.786	0.929
8	0.643	0.714	0.881
9	0.600	0.700	0.833
10	0.564	0.648	0.794
11	0.536	0.618	0.764
12	0.503	0.587	0.734
13	0.484	0.560	0.703
14	0.464	0.538	0.679
15	0.446	0.521	0.657
16	0.429	0.503	0.635
17	0.414	0.488	0.618
18	0.401	0.474	0.600
19	0.391	0.460	0.584
20	0.380	0.447	0.570

说明：两侧检验。

[例题 1—18]

表 1—21 列示了日本、美国、德国的研究者和技术人员从大学毕业后，最初的就职机会顺序。

(1) 求关于日本与美国的斯皮尔曼秩相关系数 R_s。

(2) 求关于日本与德国的斯皮尔曼秩相关系数 R_s。

(3) 将 (1) 和 (2) 的 R_s，按 10%和 5%的显著水平进行检验。

表 1—21　　研究者和技术人员从大学毕业后最初的就职机会顺序

机会	日本 X	美国 Y	德国 Z
1. 大学教师的介绍	1	3	3
2. 大学就业指导部门的介绍	3	2	8

续前表

机会	日本 X	美国 Y	德国 Z
3. 大学学长的介绍	4	7	5
4. 亲友的介绍	6	5	7
5. 自己直接到企业寻找	2	1	1
6. 公司的直接邀请	5	4	4
7. 报纸、杂志的招聘广告	7	6	2
8. 公共职业中介机构的介绍	8	9	6
9. 私人职业中介机构的介绍	9	8	9

资料来源：劳动省《劳动白皮书》。

[解答]

(1) 将数据输入工作表中进行计算（见表 1—22）。

表 1—22

X	Y	d	d^2
1	3	−2	4
3	2	1	1
4	7	−3	9
6	5	1	1
2	1	1	1
5	4	1	1
7	6	1	1
8	9	1	1
9	8	1	1
—	—	0	20

$\sum d^2$

根据（1—20）式，计算斯皮尔曼秩相关系数 R_s 得

$$R_s = 1 - \frac{6\sum d^2}{n(n^2 - 1)} = 1 - \frac{6 \times 20}{9 \times (81 - 1)}$$
$$= 0.833$$

可见，研究者和技术人员的就职方法，日本与美国非常相似。

(2) 省略工作表。根据 (1—20) 式，计算斯皮尔曼秩相关系数 R_s 得

$$R_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6 \times 62}{9 \times (81-1)}$$
$$= 0.483$$

(3) 在 (1) 中，无论在 10% 的水平，还是在 5% 的水平上，均显著相关。

在 (2) 中，两者均不显著相关。

第一章　练习题

1. 表 1—23 列出了亚洲各国和地区的人口数据。利用几何平均，计算人口年均增长率。

表 1—23　　亚洲各国和地区的人口　　单位：万人

国家和地区	1990	1995
(1) 日本	12 361	12 557
(2) 新加坡	271	299
(3) 中国香港	570	619
(4) 中国台湾	2 023	2 122
(5) 韩国	4 287	4 485
(6) 泰国	5 630	5 946
(7) 印度尼西亚	17 948	19 375
(8) 中国大陆	114 333	121 121
(9) 巴基斯坦	11 240	13 025
(10) 孟加拉国	10 812	12 043

资料来源：联合国《人口统计年鉴》。

2. 资料见表 1—24。

表 1—24　　原油价格与黄金价格的变化趋势

(1985—1995 年)　　单位：美元/每桶；美元/盎司

年　末	原油价格 (北海)	黄金价格 (伦敦)
1985	26.20	327.00
1986	18.00	390.90
1987	17.00	486.50
1988	15.50	410.15
1989	20.30	398.60

续前表

年　末	原油价格（北海）	黄金价格（伦敦）
1990	27.80	391.00
1991	17.90	353.40
1992	18.30	332.90
1993	13.25	390.65
1994	16.55	382.50
1995	18.35	386.70

说明：年末价格。

资料来源：日本银行《经济统计年报》。

(1) 计算原油价格与黄金价格的3项平均移动，并将其与原数列的图形同时画出来。

(2) 求方差 s^2 与标准差 s。

(3) 计算并比较原油价格与黄金价格的变动系数。

3. 在日本"泡沫经济"时期，以大城市为中心的地价异常高涨，其原因之一在于人们预计将来地价还会上涨，因此进行投机性购买。结果，1991年以后，随着"泡沫"的破灭，大城市地价急速下降。表1—25列出了"泡沫时期"（1990年）和"泡沫"破灭后（1995年）47个都道府县住宅用地的平均价格（1 000日元/平方米）。

表1—25　　住宅用地平均价格　　单位：1000日元/平方米

都道府县	1990年	1995年	都道府县	1990年	1995年
1. 北海道	34	33	26. 京　都	388	186
2. 青　森	28	27	27. 大　阪	571	281
3. 岩　手	32	36	28. 兵　库	283	162
4. 宫　城	61	54	29. 奈　良	212	113
5. 秋　田	25	27	30. 和歌山	91	74
6. 山　形	29	35	31. 鸟　取	34	37

续前表

都道府县	1990 年	1995 年	都道府县	1990 年	1995 年
7. 福　岛	37	38	32. 岛　根	22	26
8. 茨　城	69	75	33. 冈　山	54	55
9. 栃　木	65	71	34. 广　岛	81	74
10. 群　马	65	65	35. 山　口	38	45
11. 埼　玉	266	191	36. 德　岛	45	57
12. 千　叶	268	156	37. 香　川	67	71
13. 东　京	859	426	38. 爱　媛	59	62
14. 神奈川	351	277	39. 高　知	46	49
15. 新　潟	46	50	40. 福　冈	64	66
16. 富　山	56	56	41. 佐　贺	26	29
17. 石　川	70	78	42. 长　崎	38	43
18. 福　井	57	66	43. 熊　本	33	41
19. 山　梨	56	63	44. 大　分	31	36
20. 长　野	37	46	45. 宫　崎	29	30
21. 岐　阜	61	64	46. 鹿儿岛	31	35
22. 静　冈	129	111	47. 冲　绳	46	62
23. 爱　知	179	136			
24. 三　重	49	53			
25. 滋　贺	113	82			

资料来源：国土厅《都道府县地价调查》。

(1) 计算"泡沫时期"(1990 年) 的方差 s^2、标准差 s 和变动系数 CV。

(2) 计算"泡沫"破灭后 (1995 年) 的方差 s^2、标准差 s 和变动系数 CV。

4. 表 1—26 列出了 7 个发达工业国人均能源消耗量与人均 CO_2 排放量之间的关系。

(1) 计算相关系数 R。

(2) 这两个变量之间是否显著相关，按 5% 和 1% 的显著水平进行检验。

表 1—26　　人均能源消耗量与人均 CO_2 排放量（1992 年）

国　　家	人均能源消耗量 （换算成石油：kg）	人均 CO_2 排放量 （t）
1. 日本	3 586	8.79
2. 美国	7 662	19.11
3. 德国	4 358	10.89
4. 法国	4 034	6.31
5. 英国	3 743	9.76
6. 意大利	2 755	7.17
7. 加拿大	7 912	14.36

资料来源：World Bank，*World Development Report*。

5. 表 1—27 列出了亚洲各国和地区投资增长率 ΔI 和 GDP 增长率 ΔY（均为 1980—1994 年间的年增长率）之间的关系。

(1) 计算相关系数 R。

(2) ΔI 与 ΔY 之间是否存在显著相关，请按 5% 和 1% 显著水平进行检验。

表 1—27　亚洲各国和地区的年投资增长率和 GDP 年增长率（1980—1994 年）

国家和地区	投资增长率（%）（ΔI）	GDP 增长率（%）（ΔY）
1. 孟加拉国	4.2	1.6
2. 印度	5.2	5.7
3. 巴基斯坦	6.0	5.6
4. 中国大陆	9.6	11.1
5. 斯里兰卡	4.0	2.4
6. 缅甸	0.8	−0.7
7. 印度尼西亚	5.8	7.1
8. 菲律宾	1.4	−0.1
9. 泰国	8.2	11.4
10. 马来西亚	6.2	6.3
11. 韩国	9.1	11.8
12. 中国台湾	7.8	7.2
13. 中国香港	6.5	5.0
14. 新加坡	6.9	5.7
15. 日本	4.0	5.5

资料来源：World Bank，*World Development Report*。

6. 表1—28是对户主按年龄层次进行储蓄目的的问卷调查后依次排列的结果。

(1) 计算30年龄段和20年龄段的秩相关系数R_s。

(2) 计算30年龄段和40年龄段的秩相关系数R_s。

(3) 计算30年龄段和50年龄段的秩相关系数R_s。

表1—28　　储蓄目的（1995年）

储蓄目的	20年龄段	30年龄段	40年龄段	50年龄段
1. 防病、防灾	1	1	1	1
2. 子女教育	2	2	2	5
3. 子女结婚	8	8	6	3
4. 住宅购买与改善	5	3	5	6
5. 养老	7	5	3	2
6. 购买耐用消费品	6	6	7	8
7. 旅行、休闲	4	7	8	7
8. 没有特别目的	3	4	4	4

资料来源：日本银行《储蓄与消费意向调查》。

第二章 统计学基础知识(二)

上一章我们学习了计量经济分析中频繁使用的统计学基础知识。本章中我们将学习有关收入、财富（资产）分配状况的古典方法，例如洛伦茨曲线与基尼系数的计算，经济增长中的贡献度、贡献率的计算以及指数的计算方法等，希望通过这些学习能够进一步提高读者的数字处理能力。在计量经济学的学习过程中，数字处理作为一种基本技能是不可缺少的。

1. 洛伦茨曲线

所谓洛伦茨曲线（Lorenz curve），是用来表示收入分布、资产分布的差距、不平等程度、集中程度的一种代表性的方法，它是由美国统计学者洛伦茨 1905 年提出来的。

我们用图 2—1 来简单地说明洛伦茨曲线的作图方法。

首先，横轴表示将家庭按收入从低到高排列的累计比率。其次，纵轴表示的是与横轴的家庭累计比率相应的累计收入比率，两者的坐标可以在图中画出。最后，将各坐标相连，就形成了洛伦茨曲线。

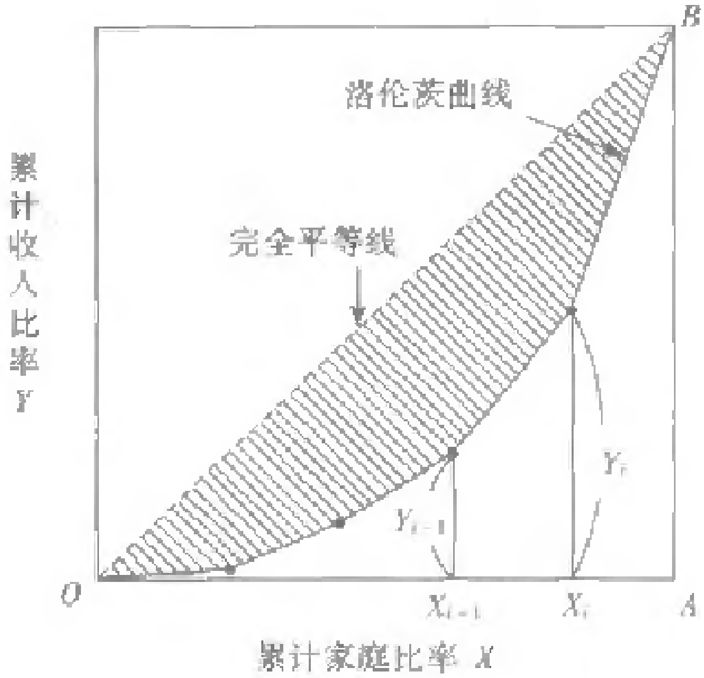


图 2—1 洛伦茨曲线与基尼系数的计算方法

直线 OB 称为完全平等线，收入分布如果趋于平等化，洛伦茨曲线就接近于完全平等线，如果收入分布完全平等，就是说所有家庭的收入完全相等，洛伦茨曲线就与完全平等线重合。另一方面，如果收入分布不平等，洛伦茨曲线就会偏离完全平等线，向右下方移动。

2. 基尼系数

所谓基尼系数（Gini coefficient），是根据洛伦茨曲线用以计算收入分布不平等程度的指数。基尼系数的大小介于 0 与 1 之间，越接近于 0，说明收入分布越平等；相反，越接近于 1，说明不平等程度越大。

基尼系数的定义如下：

$$G = 1 - [(X_1 - X_0)(Y_1 + Y_0) + (X_2 - X_1)(Y_2 + Y_1) + \cdots + (X_n - X_{n-1})(Y_n + Y_{n-1})]$$

$$= 1 - \sum (X_i - X_{i-1})(Y_i + Y_{i+1}) \qquad (2—1)$$

式中，X_i 为累计家庭比率；

Y_i 为累计收入比率；

$i=0,1,\cdots,n$。

从图形上看，图 2—1 中阴影部分面积的 2 倍就相当于基尼系数。

[例题 2—1]

表 2—1 显示了美国所有家庭（1980 年，1994 年）、白人家庭（1994 年）、黑人家庭（1994 年）中，五个等级各自所占比重。所谓五个等级，指的是将所有的家庭从低收入开始，按 20% 的幅度，依次分为五组。

（1）计算 1980 年和 1994 年关于所有家庭的基尼系数。

（2）画出 1994 年白人家庭与黑人家庭的洛伦茨曲线。

（3）计算（2）的基尼系数。

表 2—1　　美国五个等级家庭的收入比重

收入等级	1980 年	1994 年	1994 年	1994 年
	所有家庭(%)	所有家庭(%)	白人家庭(%)	黑人家庭(%)
第一等级	5.2	4.2	4.6	3.2
第二等级	11.5	10.0	10.3	8.5
第三等级	17.5	15.7	15.8	15.1
第四等级	24.3	23.3	23.0	24.6
第五等级	41.5	46.8	46.3	48.6

资料来源：美国商务部《美国统计摘要》。

[解答]

作为计算基尼系数的数据，我们先计算累计家庭比率 X 和累计收入比率 Y（见表 2—2）。

表 2—2　　美国五个等级累计家庭比率和累计收入比率

收入等级	累计家庭比率（X）	累计收入比率（Y）			
		1980 年 所有家庭	1994 年 所有家庭	1994 年 白人家庭	1994 年 黑人家庭
第一等级	0.2	0.052	0.042	0.046	0.032
第二等级	0.4	0.167	0.142	0.149	0.117

续前表

收入等级	累计家庭比率（X）	累计收入比率（Y）			
		1980 年所有家庭	1994 年所有家庭	1994 年白人家庭	1994 年黑人家庭
第三等级	0.6	0.342	0.299	0.307	0.268
第四等级	0.8	0.585	0.532	0.537	0.514
第五等级	1.0	1.000	1.000	1.000	1.000

(1) 1980 年所有家庭

根据 (2—1)式，利用表 2—2 的数据，计算基尼系数 G_{80}，得

$$
\begin{aligned}
G_{80} &= 1 - \sum(X_i - X_{i-1})(Y_i + Y_{i-1}) \\
&= 1 - [(0.2 - 0)(0.052 + 0) + (0.4 - 0.2)(0.167 + 0.052) \\
&\quad + (0.6 - 0.4)(0.342 + 0.167) + (0.8 - 0.6)(0.585 + 0.342) \\
&\quad + (1.0 - 0.8)(1.000 + 0.585)] \\
&= 1 - (0.010\,4 + 0.043\,8 + 0.101\,8 + 0.185\,4 + 0.317\,0) \\
&= 1 - 0.658\,4 \\
&= 0.341\,6
\end{aligned}
$$

1994 年所有家庭

同样，根据 (2—1) 式计算基尼系数 G_{94}，得

$$
\begin{aligned}
G_{94} &= 1 - \sum(X_i - X_{i-1})(Y_i + Y_{i-1}) \\
&= 1 - (0.008\,4 + 0.036\,8 + 0.088\,2 + 0.166\,2 + 0.306\,4) \\
&= 1 - 0.606\,0 \\
&= 0.394\,0
\end{aligned}
$$

从 1980 年（$G_{80} = 0.341\,6$）到 1994 年（$G_{94} = 0.394\,0$），基尼系数是上升的，由此可以看出，美国的收入不平等是在增加的。

(2) 见图 2—2

(3) 1994 年白人家庭

根据 (2—1) 式，利用表 2—2 的数据，计算基尼系数 G_w，得

$$
\begin{aligned}
G_W &= 1 - \sum(X_i - X_{i-1})(Y_i + Y_{i-1}) \\
&= 1 - (0.009\,2 + 0.039\,0 + 0.091\,2 + 0.168\,8 + 0.307\,4)
\end{aligned}
$$

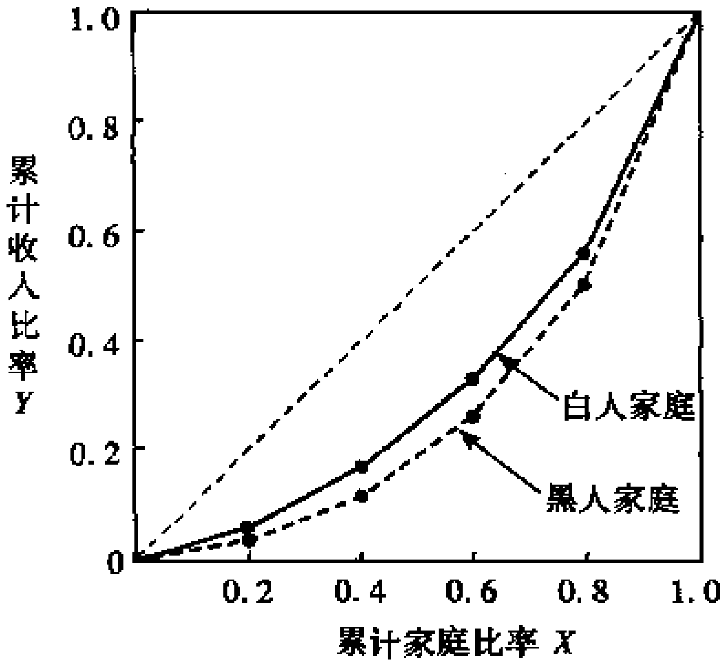


图 2—2　白人、黑人家庭的洛伦茨曲线（1994 年）

$$= 1 - 0.6156$$
$$= 0.3844$$

1994 年黑人家庭

根据（2—1）式，利用表 2—2 的数据，计算基尼系数 G_B，得

$$\begin{aligned} G_B &= 1 - \sum (X_i - Y_{i-1})(Y_i + Y_{i-1}) \\ &= 1 - (0.0064 + 0.0298 + 0.0770 + 0.1564 + 0.3028) \\ &= 1 - 0.5724 \\ &= 0.4276 \end{aligned}$$

可见，黑人家庭与白人家庭相比，收入分布的不平等程度更高。其原因可能在于黑人家庭中存在极端贫困阶层，使得收入分布扭曲，基尼系数上升。

这样，利用基尼系数，既可以像（1）那样比较分析时间序列，也可以像（3）那样进行横截面的比较分析，其应用范围非常广泛。

［例题 2—2］

表 2—3 显示了亚洲各国（地区）五个等级的收入分布状况。从低收入的个人（或者家庭）开始，按 20% 的幅度，依次分为第一等级～第五等级。

（1）计算从孟加拉国到日本，各国的基尼系数。

（2）根据“库兹涅茨倒 U 字假说”，分析（1）中的基尼系数计算结果。

表 2—3　　亚洲各国(地区)五个等级的收入分布（%）

收入等级	① 孟加拉国 1992 年 (220 美元)	② 印度 1992 年 (320 美元)	③ 中国大陆 1992 年 (530 美元)	④ 菲律宾 1988 年 (950 美元)	⑤ 泰国 1992 年 (2 410 美元)	⑥ 马来西亚 1989 年 (3 480 美元)	⑦ 韩国 1988 年 (8 260 美元)	⑧ 中国台湾 1994 年 (11 600 美元)	⑨ 日本 1995 年 (34 630 美元)
第一等级	9.4	8.5	6.2	6.5	5.6	4.6	7.4	7.3	8.1
第二等级	13.5	12.1	10.5	10.1	8.7	8.3	12.3	13.0	13.1
第三等级	17.2	15.8	15.8	14.4	13.0	13.0	16.3	17.4	17.5
第四等级	22.0	21.1	23.6	21.2	20.0	20.4	21.8	23.2	23.0
第五等级	37.9	42.5	43.9	47.8	52.7	53.7	42.2	39.1	38.3

说明：国名下面的年份，指的是收入分布调查的年份；括号内的数字是人均 GDP。①～⑥按个人分组，⑦～⑨按家庭分组。

资料来源：世界银行《世界发展报告》；总务厅统计局《家庭调查年报》。

[解答]

首先，作为计算基尼系数的数据，我们计算累计家庭比率 X 和累计收入比率 Y（见表 2—4）。

表 2—4　　累计家庭比率 X 和累计收入比率 Y

累计家庭比率（X）	累计收入比率（Y）								
	① 孟加拉国	② 印度	③ 中国大陆	④ 菲律宾	⑤ 泰国	⑥ 马来西亚	⑦ 韩国	⑧ 中国台湾	⑨ 日本
0.2	0.094	0.085	0.062	0.065	0.056	0.046	0.074	0.073	0.081
0.4	0.229	0.206	0.167	0.166	0.143	0.129	0.197	0.203	0.212
0.6	0.401	0.364	0.325	0.310	0.273	0.259	0.360	0.377	0.387
0.8	0.621	0.575	0.561	0.522	0.473	0.463	0.578	0.609	0.617
1.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

说明：①～⑥按个人分组，⑦～⑨按家庭分组。

(1) ① 孟加拉国

根据 (2—1) 式，计算基尼系数 G_B，得

$$\begin{aligned}G_B &= 1-\sum(X_i-X_{i-1})(Y_i+Y_{i-1})\\&=1-[(0.2-0)\times(0.094+0)\\&\quad+(0.4-0.2)\times(0.229+0.094)+(0.6-0.4)\times(0.401+0.229)\\&\quad+(0.8-0.6)\times(0.621+0.401)+(1.0-0.8)\times(1.000+0.621)]\end{aligned}$$

$$= 1 - (0.018\,8 + 0.064\,6 + 0.126\,0 + 0.204\,4 + 0.324\,2)$$
$$= 1 - 0.738\,0$$
$$= 0.262\,0$$

② 印度

$$\begin{aligned} G_I &= 1 - \sum (X_i - X_{i-1})(Y_i + Y_{i-1}) \\ &= 1 - (0.017\,0 + 0.058\,2 + 0.114\,0 + 0.187\,8 + 0.315\,0) \\ &= 1 - 0.692\,0 \\ &= 0.308\,0 \end{aligned}$$

③ 中国大陆

$$\begin{aligned} G_C &= 1 - \sum (X_i - X_{i-1})(Y_i + Y_{i-1}) \\ &= 1 - (0.012\,4 + 0.045\,8 + 0.098\,4 + 0.177\,2 + 0.312\,2) \\ &= 1 - 0.646\,0 \\ &= 0.354\,0 \end{aligned}$$

④ 菲律宾

$$\begin{aligned} G_P &= 1 - \sum (X_i - X_{i-1})(Y_i + Y_{i-1}) \\ &= 1 - (0.013\,0 + 0.046\,2 + 0.095\,2 + 0.166\,4 + 0.304\,4) \\ &= 1 - 0.625\,2 \\ &= 0.374\,8 \end{aligned}$$

⑤ 泰国

$$\begin{aligned} G_T &= 1 - \sum (X_i - X_{i-1})(Y_i + Y_{i-1}) \\ &= 1 - (0.012\,2 + 0.039\,8 + 0.083\,2 + 0.149\,2 + 0.294\,6) \\ &= 1 - 0.578\,0 \\ &= 0.422\,0 \end{aligned}$$

⑥ 马来西亚

$$\begin{aligned} G_M &= 1 - \sum (X_i - X_{i-1})(Y_i + Y_{i-1}) \\ &= 1 - (0.009\,2 + 0.035\,0 + 0.077\,6 + 0.144\,4 + 0.292\,6) \\ &= 1 - 0.558\,8 \\ &= 0.441\,2 \end{aligned}$$

⑦ 韩国

$$
\begin{aligned}
G_K &= 1-\sum(X_i-X_{i-1})(Y_i+Y_{i-1})\\
&= 1-(0.014\,8+0.054\,2+0.111\,4+0.187\,6+0.315\,6)\\
&= 1-0.683\,6\\
&= 0.316\,4
\end{aligned}
$$

⑧ 中国台湾

$$
\begin{aligned}
G_{TA} &= 1-\sum(X_i-X_{i-1})(Y_i+Y_{i-1})\\
&= 1-(0.014\,6+0.055\,2+0.116\,0+0.197\,2+0.321\,8)\\
&= 1-0.704\,8\\
&= 0.295\,2
\end{aligned}
$$

⑨ 日本

$$
\begin{aligned}
G_J &= 1-\sum(X_i-X_{i-1})(Y_i+Y_{i-1})\\
&= 1-(0.016\,2+0.058\,6+0.119\,8+0.200\,8+0.323\,4)\\
&= 1-0.718\,8\\
&= 0.281\,2
\end{aligned}
$$

(2)“库兹涅茨倒 U 字假说”指的是在经济发展初期，过剩劳动力的存在引起了低工资和工资差距的扩大，以及收入分布不平等的增大，等到劳动力市场供求关系紧张，人均 GDP 超过一定水平，并且进一步增加时，收入分布就会向平等化方向转变，基尼系数也随之下降。由于发展中国家收入分布的数据不足，而且相互之间差异较大，用基尼系数对国际间横截面数据进行严密的分析比较困难，但是根据（1）的计算结果（孟加拉国、印度、中国大陆、菲律宾、泰国、马来西亚的基尼系数随人均 GDP 的增加而上升，相反韩国、中国台湾、日本则是下降的），“库兹涅茨倒 U 字假说”应该是能够成立的。

此外，发达国家中，由于累进税制（所得税、资产税、遗产税）、社会保障、公共支出等再分配政策的普及，收入分布的平等化得到了进一步改善。

[例题 2—3]

表 2—5 按照不同阶层年收入数据（所有家庭，1995 年），显示了各阶层所包含的家庭数与年均收入。表 2—6 按照储蓄余额的大小（所有家庭，1995 年），显示了各阶层包含的家庭数与平均储蓄余额。

（1）分别画出年收入与储蓄余额的洛伦茨曲线。

（2）求各自的基尼系数。

表 2—5　　日本不同收入阶层的家庭数与年均收入

（所有家庭，1995 年）

年收入（万日元）	家庭数（户）	年均收入（万日元）
0～ 200	197	143
200～ 250	248	224
250～ 300	346	274
300～ 350	530	322
350～ 400	581	374
400～ 450	585	423
450～ 500	648	474
500～ 550	636	521
550～ 600	578	573
600～ 650	616	621
650～ 700	534	672
700～ 750	569	722
750～ 800	466	773
800～ 900	803	845
900～1 000	667	945
1 000～1 250	1 025	1 105
1 250～1 500	462	1 355
1 500 以上	509	2 150

资料来源：总务厅《家庭调查年报》。

表 2—6　　日本不同储蓄数额阶层的家庭数与平均储蓄余额

（所有家庭，1995 年）

当年储蓄余额（万日元）	家庭数（户）	当前平均储蓄余额（万日元）
0～ 100	465	42
100～ 150	221	127
150～ 200	257	174
200～ 250	304	222
250～ 300	295	272
300～ 400	610	349
400～ 500	565	451
500～ 600	510	547
600～ 700	499	650
700～ 800	519	751
800～ 900	453	848

续前表

当年储蓄余额（万日元）	家庭数（户）	当前平均储蓄余额（万日元）
900～1 000	331	951
1 000～1 200	662	1 091
1 200～1 500	898	1 347
1 500～2 000	980	1 736
2 000～3 000	1 059	2 438
3 000～4 000	529	3 433
4 000 以上	842	6 656

资料来源：总务厅《储蓄动向调查报告》。

[解答]

(1) 根据表 2—5、表 2—6 的数据，作成表 2—7、表 2—8。

表 2—7　　1995 年累计家庭比率与累计年收入总额的比率

年收入（万日元）	家庭			年收入			
	① 家庭数（户）	② 累计家庭数（户）① 的累计	③ 累计家庭比率（X）$\frac{②}{1\ 000}$	④ 年均收入（万日元）	⑤ 年收入总额（万日元）④×①	⑥ 累计年收入总额（万日元）⑤ 的累计	⑦ 累计年收入总额的比率（Y）$\frac{⑥}{7\ 457\ 893}$
0～200	197	197	0.019 7	143	28 171	28 171	0.003 78
200～250	248	445	0.044 5	224	55 552	83 723	0.011 23
250～300	346	791	0.079 1	274	94 804	178 527	0.023 94
300～350	530	1 321	0.132 1	322	170 660	349 187	0.046 82
350～400	581	1 902	0.190 2	374	217 294	566 481	0.075 96
400～450	585	2 487	0.248 7	423	247 455	813 936	0.109 14
450～500	648	3 135	0.313 5	474	307 152	1 121 088	0.150 32
500～550	636	3 771	0.377 1	521	331 356	1 452 444	0.194 75
550～600	578	4 349	0.434 9	573	331 194	1 783 638	0.239 16
600～650	616	4 965	0.496 5	621	382 536	2 166 174	0.290 45
650～700	534	5 499	0.549 9	672	358 348	2 525 022	0.338 57
700～750	569	6 068	0.606 8	722	410 818	2 935 840	0.393 66
750～800	466	6 534	0.653 4	773	360 218	3 296 058	0.441 96
800～900	803	7 337	0.733 7	845	678 535	3 974 593	0.532 94
900～1 000	667	8 004	0.800 4	945	630 315	4 604 908	0.617 45
1 000～1 250	1 025	9 029	0.902 9	1 105	1 132 625	5 737 533	0.769 32
1 250～1 500	462	9 491	0.949 1	1 355	626 010	6 363 543	0.853 26
1 500 以上	509	10 000	1.000 0	2 150	1 094 350	7 457 893	1.000 00
合　计	10 000				7 457 893		

表 2—8　　1995 年累计家庭比率与累计储蓄总额的比率

目前储蓄余额（万日元）	家庭 ①家庭数（户）	②累计家庭数（户）①的累计	③累计家庭比率（X）②/9 999	目前储蓄余额 ④平均储蓄（万日元）	⑤储蓄总额（万日元）④×①	⑥累计储蓄总额（万日元）⑤的累计	⑦累计储蓄总额的比率（Y）⑥/16 035 141
0～ 100	465	465	0.046 5	42	19 530	19 530	0.001 22
100～ 150	221	686	0.068 6	127	28 067	47 597	0.002 97
150～ 200	257	943	0.094 3	174	44 718	92 315	0.005 76
200～ 250	304	1 247	0.124 7	222	67 488	159 803	0.009 97
250～ 300	295	1 542	0.154 2	272	80 240	240 043	0.014 97
300～ 400	610	2 152	0.215 2	349	212 890	452 933	0.028 25
400～ 500	565	2 717	0.271 7	451	254 815	707 748	0.044 14
500～ 600	510	3 227	0.322 7	547	278 970	986 718	0.061 53
600～ 700	499	3 726	0.372 6	650	324 350	1 311 068	0.081 76
700～ 800	519	4 245	0.424 5	751	389 769	1 700 837	0.106 07
800～ 900	453	4 698	0.469 8	848	384 144	2 084 981	0.130 03
900～1 000	331	5 029	0.503 0	951	314 781	2 399 762	0.149 66
1 000～1 200	662	5 691	0.569 2	1 091	722 242	3 122 004	0.194 70
1 200～1 500	898	6 589	0.659 0	1 347	1 209 606	4 331 610	0.270 13
1 500～2 000	980	7 569	0.757 0	1 736	1 701 280	6 032 890	0.376 23
2 000～3 000	1 059	8 628	0.862 9	2 438	2 581 842	8 614 732	0.537 24
3 000～4 000	529	9 157	0.915 8	3 433	1 816 057	10 430 789	0.650 50
4 000 以上	842	9 999	1.000 0	6 656	5 604 352	16 035 141	1.000 00
合　计	9 999				16 035 141		

根据表 2—7、表 2—8 中的数据③、⑦，画出洛伦茨曲线，即图 2—3。

从图 2—3 可以看出，储蓄余额（资产）的分布与年收入的分布相比，不平等的程度要严重得多。

(2) 年收入

根据 (2—1) 式，计算基尼系数 G_I，得

$$\begin{aligned} G_I &= 1-\sum(X_i-X_{i-1})(Y_i+Y_{i-1}) \\ &= 1-[(0.019\,7-0)\times(0.003\,78+0)+(0.044\,5-0.019\,7) \\ &\quad \times(0.011\,23+0.003\,78)+\cdots+(1.000\,0-0.949\,1) \\ &\quad \times(1.000\,00+0.853\,26)] \end{aligned}$$

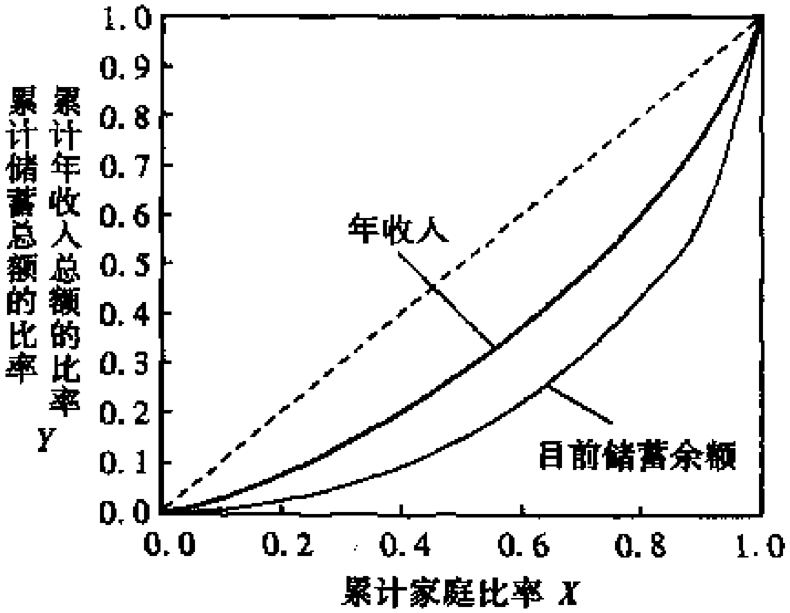


图 2—3　年收入与目前储蓄余额的洛伦茨曲线(全部家庭，1995 年)

$$= 1 - (0.000\,07 + 0.000\,37 + 0.001\,22 + 0.003\,75 + 0.007\,13$$
$$+ 0.010\,83 + 0.016\,81 + 0.021\,95 + 0.025\,08 + 0.032\,62$$
$$+ 0.033\,59 + 0.041\,66 + 0.038\,94 + 0.078\,28 + 0.076\,73$$
$$+ 0.142\,14 + 0.074\,96 + 0.094\,33)$$
$$= 1 - 0.700\,49$$
$$= 0.299\,5$$

储蓄余额

同样，计算基尼系数 G_S，得

$$G_S = 1 - \sum (X_i - X_{i-1})(Y_i + Y_{i-1})$$
$$= 1 - [(0.046\,5 - 0) \times (0.001\,22 + 0) + (0.068\,6 - 0.046\,5)$$
$$\times (0.002\,97 + 0.001\,22) + \cdots + (1.000\,0 - 0.915\,8)$$
$$\times (1.000\,00 + 0.650\,50)]$$
$$= 1 - (0.000\,06 + 0.000\,09 + 0.000\,22 + 0.000\,48 + 0.000\,74$$
$$+ 0.002\,64 + 0.004\,09 + 0.005\,39 + 0.0071\,5 + 0.009\,75$$
$$+ 0.010\,70 + 0.009\,26 + 0.022\,80 + 0.041\,75 + 0.063\,35$$
$$+ 0.096\,75 + 0.062\,84 + 0.138\,99)$$
$$= 1 - 0.477\,02$$
$$= 0.523\,0$$

可以看出，储蓄余额的分布（$G_S=0.523\ 0$）与收入分布（$G_I=0.299\ 5$）相比，更不平等。高收入阶层储蓄额很大，这是多年存量增加的结果，它与每年的流量即收入的分布相比，差距自然更大。

[建议]

洛伦茨曲线和基尼系数是非常有意思的分析方法，大家在写毕业论文时，不妨一试。日本在这方面的基本数据可以利用总务厅统计局的《家庭调查年报》、《储蓄动向调查报告》。在有经济学系、商学系以及社会学系的大学图书馆中，应该有这两种统计资料。

[例题 2—4]

表 2—9 是日本按从业者规模划分的纺织工业、纸浆与造纸工业、化学工业、钢铁工业的企业数量与从业者人数。为了比较企业的集中程度，分别计算这四个产业的基尼系数。

表 2—9　　按从业者规模划分的企业数量与从业者人数(1991 年)

从业者规模（人）	(1) 纺织工业		(2) 纸浆与造纸工业		(3) 化学工业		(4) 钢铁工业	
	企业数量	从业人数	企业数量	从业人数	企业数量	从业人数	企业数量	从业人数
5～ 9	7 710	66 617	3 322	29 514	1 690	14 614	1 962	17 144
10～ 19	4 685	76 038	2 432	40 554	1 532	24 614	1 669	27 049
20～ 29	1 872	50 777	1 110	30 487	843	22 698	737	19 906
30～ 49	1 453	60 354	869	36 895	872	36 466	653	27 452
50～ 99	1 022	75 793	673	50 616	876	65 515	485	35 536
100～199	446	64 544	274	38 019	583	85 529	243	34 993
200～299	126	32 000	75	18 522	182	45 384	67	16 951
300～499	96	37 233	51	19 904	156	60 514	60	22 532
500～999	39	24 950	28	19 662	132	92 181	42	29 314
1 000 人以上	5	8 215	11	15 633	63	107 064	41	131 610

资料来源：总务厅统计局《企业数量统计调查报告》。

[解答]

要计算基尼系数，首先需累计企业数量和从业者人数（见表 2—10），再根据该表作出相应的累计比率（见表 2—11）。

表 2—10　　企业数量与从业人数累计（1991 年）

从业者规模（人）	(1) 纺织工业		(2) 纸浆与造纸工业		(3) 化学工业		(4) 钢铁工业	
	企业数量	从业人数	企业数量	从业人数	企业数量	从业人数	企业数量	从业人数
5～ 9	7 710	66 617	3 322	29 514	1 690	14 614	1 962	17 144
10～ 19	12 395	142 655	5 754	70 068	3 222	39 228	3 631	44 193
20～ 29	14 267	193 432	6 864	100 555	4 065	61 926	4 368	64 099
30～ 49	15 720	253 786	7 733	137 450	4 937	98 392	5 021	91 551
50～ 99	16 742	329 579	8 406	188 066	5 813	163 907	5 506	127 087
100～199	17 188	394 123	8 680	226 085	6 396	249 436	5 749	162 080
200～299	17 314	426 123	8 755	244 607	6 578	294 820	5 816	179 031
300～499	17 410	463 356	8 806	264 511	6 734	355 334	5 876	201 563
508～999	17 449	488 306	8 834	284 173	6 866	447 515	5 918	230 877
1 000 人以上	17 454	496 521	8 845	299 806	6 929	554 579	5 959	362 487

表 2—11　　企业数量与从业人数的累计比率（1991 年）

从业者规模（人）	(1) 纺织工业		(2) 纸浆与造纸工业		(3) 化学工业		(4) 钢铁工业	
	企业数量	从业人数	企业数量	从业人数	企业数量	从业人数	企业数量	从业人数
5～ 9	0.441 73	0.134 17	0.375 58	0.098 44	0.243 90	0.026 35	0.329 25	0.047 30
10～ 19	0.710 15	0.287 31	0.650 54	0.233 71	0.465 00	0.070 73	0.609 33	0.121 92
20～ 29	0.817 41	0.389 57	0.776 03	0.335 40	0.586 66	0.111 66	0.733 01	0.176 83
30～ 49	0.900 65	0.511 13	0.874 28	0.458 46	0.712 51	0.177 42	0.842 59	0.252 56
50～ 99	0.959 21	0.663 78	0.950 37	0.627 29	0.838 94	0.295 55	0.923 98	0.350 60
100～199	0.984 76	0.793 77	0.981 35	0.754 10	0.923 08	0.449 78	0.964 76	0.447 13
200～299	0.991 98	0.858 22	0.989 82	0.815 88	0.949 34	0.531 61	0.976 00	0.493 90
300～499	0.997 48	0.933 21	0.995 59	0.882 27	0.971 86	0.640 73	0.986 07	0.556 06
500～999	0.999 71	0.983 45	0.998 76	0.947 86	0.990 91	0.806 95	0.993 12	0.636 92
1 000 人以上	1.000 00	1.000 00	1.000 00	1.000 00	1.000 00	1.000 00	1.000 00	1.000 00

（1）纺织工业

根据（2—1）式，利用表 2—11 的数据，计算基尼系数 G_T：

$$G_T = 1 - \sum (X_i - X_{i-1})(Y_i + Y_{i-1})$$
$$= 1 - [(0.441\,73 - 0) \times (0.134\,17 + 0) + (0.710\,15 - 0.441\,73)$$

$$\times (0.287\,31 + 0.134\,17) + \cdots + (1.000\,00 - 0.999\,71)$$
$$\times (1.000\,00 + 0.983\,45)]$$
$$= 1 - (0.059\,27 + 0.113\,13 + 0.072\,60 + 0.074\,98 + 0.068\,80$$
$$+ 0.037\,24 + 0.011\,93 + 0.009\,85 + 0.004\,28 + 0.000\,57)$$
$$= 1 - 0.452\,65$$
$$= 0.547\,4$$

(2) 纸浆与造纸业

$$G_P = 1 - \sum (X_i - X_{i-1})(Y_i + Y_{i-1})$$
$$= 1 - [(0.375\,58 - 0) \times (0.098\,44 + 0) + (0.650\,54 - 0.375\,58)$$
$$\times (0.233\,71 + 0.098\,44) + \cdots + (1.000\,00 - 0.998\,76)$$
$$\times (1.000\,00 + 0.947\,86)]$$
$$= 1 - (0.036\,97 + 0.091\,33 + 0.071\,42 + 0.078\,00 + 0.082\,61$$
$$+ 0.042\,79 + 0.013\,31 + 0.009\,79 + 0.005\,79 + 0.002\,42)$$
$$= 1 - 0.434\,43$$
$$= 0.565\,6$$

(3) 化学工业

$$G_c = 1 - \sum (X_i - X_{i-1})(Y_i + Y_{i-1})$$
$$= 1 - [(0.243\,90 - 0) \times (0.026\,35 + 0) + (0.465\,00 - 0.243\,90)$$
$$\times (0.070\,73 + 0.026\,35) + \cdots + (100\,000 - 0.990\,91)$$
$$\times (1.000\,00 + 0.806\,95)]$$
$$= 1 - (0.006\,43 + 0.021\,47 + 0.022\,19 + 0.036\,38 + 0.059\,80$$
$$+ 0.062\,71 + 0.025\,78 + 0.026\,39 + 0.027\,58 + 0.016\,43)$$
$$= 1 - 0.305\,16$$
$$= 0.694\,8$$

(4) 钢铁业

$$G_S = 1 - \sum (X_i - X_{i-1})(Y_i + Y_{i-1})$$
$$= 1 - [(0.329\,25 - 0) \times (0.047\,30 + 0) + (0.609\,33 - 0.329\,25)$$
$$\times (0.121\,92 + 0.047\,30) + \cdots + (1.000\,00 - 0.993\,12)$$
$$\times (1.000\,00 + 0.636\,92)]$$
$$= 1 - (0.015\,57 + 0.047\,39 + 0.036\,95 + 0.047\,05 + 0.049\,09$$
$$+ 0.032\,53 + 0.010\,58 + 0.010\,57 + 0.008\,41 + 0.011\,26)$$

$$= 1 - 0.269\ 40$$
$$= 0.730\ 6$$

基尼系数从高到低依次为钢铁工业（0.730 6）、化学工业（0.694 8）、造纸工业（0.565 6）、纺织工业（0.547 4），因此，从业人员数量与企业数量的集中程度，也是这种顺序。尤其是钢铁工业与化学工业基尼系数非常大，充分反映了这两种产业所要求的规模经济（economics of scale）特征。

3. 贡献度与贡献率

贡献度与贡献率反映的是在某种数据的变化中，它的各个构成要素贡献的大小或者变化的程度与方向（正、负）。

下面，我们用一个恒等式来予以说明。

$$Y = A + B + C + D \tag{2—2}$$

等式（2—2）如果每一期都能够成立，那么，关于它的变化幅度，下面的关系式能够成立：

$$\dot{Y} = \dot{A} + \dot{B} + \dot{C} + \dot{D} \tag{2—3}$$

这个公式的两边如果同时除以基准时期 Y，则有

$$\frac{\dot{Y}}{Y} = \frac{\dot{A}}{Y} + \frac{\dot{B}}{Y} + \frac{\dot{C}}{Y} + \frac{\dot{D}}{Y} \qquad [贡献度](2—4)$$

右边各项就是各要素的贡献度，它反映了在 Y 的变化中，各个要素分别做了多大贡献。

其次，我们将（2—3）式两边同时除以 $\dot{Y}$，得

$$1 = \frac{\dot{A}}{\dot{Y}} + \frac{\dot{B}}{\dot{Y}} + \frac{\dot{C}}{\dot{Y}} + \frac{\dot{D}}{\dot{Y}} \qquad [贡献率](2—5)$$

右边各项就是各要素的贡献率。如果将 Y 的变化幅度看做 100%，贡献率反映的就是各要素分别贡献了百分之多少。

如果将时点 2 相对于时点 1 的变化幅度看做 $\dot{Y}$，其定义如下：

$$\dot{Y} = Y_2 - Y_1$$

[**例题 2—5**]

表 2—12 是日本 1990 年、1995 年实际 GDP 与总支出核算的数据。分别计算实际 GDP 变化中，各项支出的贡献度与贡献率。

表 2—12 **历年日本的国民经济核算** 单位：兆日元

项　目	1990 年	1995 年
实际 GDP：Y	429.98	461.46
1. 民间最终消费支出：C	249.14	274.17
2. 政府最终消费支出：G	38.68	43.64
3. 国内总资本形成：I	139.12	138.14
4. 出口：X	46.01	56.79
5. 进口：M	−42.97	−51.28

说明：按 1990 年价格计算。

资料来源：经济企划厅《国民经济核算年报》。

[**解答**]

GDP 从支出的角度看，有如下的恒等式：

$$Y = C + G + I + X - M \tag{1}$$

同样，由于上述公式在 1990 年和 1995 年均适用，其有关变化的公式也成立，即

$$\dot{Y} = \dot{C} + \dot{G} + \dot{I} + \dot{X} - \dot{M} \tag{2}$$

将 (2) 式用 1990 年的 Y 来除，得贡献度，用 $\dot{Y}$ 来除，得贡献率，即：

贡献度

$$\frac{\dot{Y}}{Y} = \frac{\dot{C}}{Y} + \frac{\dot{G}}{Y} + \frac{\dot{I}}{Y} + \frac{\dot{X}}{Y} - \frac{\dot{M}}{Y} \tag{3}$$

贡献率

$$1 = \frac{\dot{C}}{\dot{Y}} + \frac{\dot{G}}{\dot{Y}} + \frac{\dot{I}}{\dot{Y}} + \frac{\dot{X}}{\dot{Y}} - \frac{\dot{M}}{\dot{Y}} \tag{4}$$

根据 (3) 和 (4)，贡献度与贡献率的计算结果见表 2—13。

表 2—13　　日本实际 GDP 变化中各要素的贡献度与贡献率

项　　目	变化幅度（兆日元）	贡献度（%）	贡献率（%）
实际 GDP：Y	31.48	7.32	100.00
1. 民间最终消费支出：C	25.03	5.82	79.51
2. 政府最终消费支出：G	4.96	1.15	15.76
3. 国内总资本形成：I	−0.98	−0.23	−3.11
4. 出口：X	10.78	2.51	34.24
5. 进口：M	−8.31	−1.93	−26.40

4. 拉氏指数·帕氏指数·费雪指数

拉氏指数（Laspeyres index）是德国经济学家拉斯普雷斯于 1864 年提出的指数，日本的政府统计中使用最广的统计方法（如消费者价格指数、批发价格指数等）都属于这种指数。设价格为 p，数量为 q，拉氏价格指数 P_L 定义如下：

$$P_L = \frac{\sum p_1 q_0}{\sum p_0 q_0} \tag{2—6}$$

该公式以 0 期为基期，基期时的数量为 q_0，假定它在比较期（1 期）也没有发生变化。也就是说，拉氏指数表示的是，如果在 1 期购买与 0 期相同数量的商品，两者相比较所支出的费用变化程度。但是，这个指数对于比基期数量增加的商品来说，存在过小评价的倾向，而对交易数量减小的商品则存在过大评价的倾向。如果两个时期间隔较远，计算这种价格指数就很不准确。

此外，日本在编制采矿工业生产指数时，将拉氏数量指数作如下定义：

$$Q_L = \frac{\sum p_0 q_1}{\sum p_0 q_0} \tag{2—7}$$

帕氏指数（Paasche index）是在拉氏指数提出 10 年之后，由另一位德国经济学家帕舍提出的。日本的国民收入平减指数、东京证券交易所股票价格指数（TOPIX）等采用这种方法。帕氏价格指数 P_P 定义如下：

$$P_P = \frac{\sum p_1 q_1}{\sum p_0 q_1} \tag{2—8}$$

该式与拉氏指数的区别在于，分子分母中均用比较期的交易数量 q_1 代替基期的交易数量 q_0。

一般说来，拉氏价格指数比帕氏价格指数容易编制。因为，在编制拉氏指数时，如果已经有了基期的价格与数量，只需要调查比较期的价格，而帕氏指数必须同时调查比较期的价格与数量。

顺便提一下，有一个所谓的“帕舍效应”，指的是帕氏价格指数比拉氏价格指数较小的倾向。这是因为价格上升的商品的消费量通常是减小的，因而比较期的数量 q_1 小于基期的数量 q_0。

关于帕氏数量指数，有如下定义：

$$Q_P = \frac{\sum p_1 q_1}{\sum p_1 q_0} \tag{2—9}$$

费雪指数（Fisher index）是美国经济学家与统计学家 I. 费雪提出的一种价格指数，它是拉氏指数与帕氏指数的几何平均，定义如下：

$$\begin{aligned} P_F &= \sqrt{P_L \cdot P_P} \\ &= \sqrt{\frac{\sum p_1 q_0}{\sum p_0 q_0} \cdot \frac{\sum p_1 q_1}{\sum p_0 q_1}} \end{aligned} \tag{2—10}$$

这种指数虽然又称费雪理想计算公式，但是由于它的计算过程中还必须利用帕氏指数，因此在实际中运用很少。日本的贸易价格指数采用这种方法。

费雪数量指数，定义如下：

$$\begin{aligned} Q_F &= \sqrt{Q_L \cdot Q_P} \\ &= \sqrt{\frac{\sum p_0 q_1}{\sum p_0 q_0} \cdot \frac{\sum p_1 q_1}{\sum p_1 q_0}} \end{aligned} \tag{2—11}$$

[例题 2—6]

表 2—4 列示了日本五种新鲜水果的年平均价格 p_t 和每户年均购买数量 q_t。

(1) 以 1990 年为基期，计算 1995 年的拉氏价格指数 P_L。

(2) 以 1990 年为基期，计算 1995 年的帕氏价格指数 P_P。

(3) 以 1990 年为基期，计算 1995 年的费雪价格指数 P_F。

(4) 以 1990 年为基期，计算 1995 年的拉氏、帕氏、费雪数量指数 Q_L、Q_P、Q_F。

表 2—14　　新鲜水果的年平均价格与年户均购买量　　单位：日元/100 克；100 克

新鲜水果	1990 年		1995 年	
	价格 p_0	数量 q_0	价格 p_1	数量 q_1
1. 蜜橘	30	278	36	214
2. 苹果	39	185	39	176
3. 梨子	47	71	50	66
4. 葡萄	90	39	95	34
5. 草莓	106	45	118	38

资料来源：总务厅统计局《家庭调查年报》。

[解答]

(1) 根据 (2—6) 式，计算拉氏价格指数 P_L。

$$P_L = \frac{\sum p_1 q_0}{\sum p_0 q_0}$$

$$= \frac{36 \times 278 + 39 \times 185 + 50 \times 71 + 95 \times 39 + 118 \times 45}{30 \times 278 + 39 \times 185 + 47 \times 71 + 90 \times 39 + 106 \times 45}$$

$$= \frac{10\ 008 + 7\ 215 + 3\ 550 + 3\ 705 + 5\ 310}{8\ 340 + 7\ 215 + 3\ 337 + 3\ 510 + 4\ 770}$$

$$= \frac{29\ 788}{27\ 172}$$

$$= 1.096$$

设 1990 年为 100，1995 年的拉氏价格指数为 109.6。

(2) 根据 (2—8) 式，计算帕氏价格指数 P_P。

$$P_P = \frac{\sum p_1 q_1}{\sum p_0 q_1}$$

$$= \frac{36\times 214+39\times 176+50\times 66+95\times 34+118\times 38}{30\times 214+39\times 176+47\times 66+90\times 34+106\times 38}$$

$$= \frac{7\,704+6\,864+3\,300+3\,230+4\,484}{6\,420+6\,864+3\,102+3\,060+4\,028}$$

$$= \frac{25\,582}{23\,474}$$

$$= 1.090$$

帕氏价格指数为 109.0，小于拉氏价格指数（109.6），可见存在着“帕舍效应”。

(3) 根据（2—10）式，计算费雪价格指数 P_F。

$$P_F = \sqrt{P_L \cdot P_P}$$

$$= \sqrt{\frac{29\,788}{27\,172}\times\frac{25\,582}{23\,474}}$$

$$= \sqrt{1.194\,7}$$

$$= 1.093$$

费雪价格指数为 109.3。

(4) 根据（2—7）式，计算拉氏数量指数 Q_L。

$$Q_L = \frac{\sum p_0 q_1}{\sum p_0 q_0} = \frac{23\,474}{27\,172} = 0.864$$

设 1990 年为 100，1995 年的拉氏数量指数为 86.4。

其次，计算帕氏数量指数 Q_P。

$$Q_P = \frac{\sum p_1 q_1}{\sum p_1 q_0} = \frac{25\,582}{29\,788} = 0.859$$

帕氏数量指数为 85.9。

最后，根据（2—11）式，计算费雪数量指数 Q_F。

$$Q_F = \sqrt{Q_L \times Q_P} = \sqrt{0.741\,9} \approx 0.861$$

费雪数量指数为 86.1。

第二章　练习题

1. 一般认为，日本的劳动者家庭收入分布，从国际比较的角度看，也是非常平等的。表 2—15 列示了 1975 年、1980 年、1985 年、1990 年及 1995 年大阪府劳动者家庭五个等级各自的收入（年收入）比重。计算这几年的基尼系数，并考察收入分配的变化。

表 2—15　　大阪府劳动者家庭五个等级各自的收入比重

收入等级	家庭比例	年收入（万日元）				
		1975 年	1980 年	1985 年	1990 年	1995 年
第一等级	0.2	7.6	8.5	11.8	11.6	10.8
第二等级	0.2	13.5	13.6	15.5	15.6	15.0
第三等级	0.2	17.4	17.6	18.5	19.0	18.7
第四等级	0.2	23.0	22.6	22.8	23.0	22.6
第五等级	0.2	38.5	37.7	31.4	30.8	32.9

资料来源：大阪府《大阪府统计年鉴》。

2. 中国城市的经济改革稍晚于农村的经济改革，大约从 20 世纪 80 年代中期正式开始，结果，城市家庭的平均年收入从2 913元（1985 年）飞速增加到13 851元（1995 年）。表 2—16 列示了（1）1985 年、（2）1990 年、（3）1995 年中国城市家庭按七个等级划分，各个等级中每个家庭的平均年收入。计算（1）～（3）的基尼系数，并考察随着年收入的增加，收入分布的不平等程度发生了什么样的变化。

表 2—16　　中国城市家庭按七个等级划分的户均年收入

收入等级	家庭比例	户均年收入（元）		
		（1）1985 年	（2）1990 年	（3）1995 年
第一等级	0.1	2 158	3 517	7 992
第二等级	0.1	2 510	4 179	9 697
第三等级	0.2	2 760	4 699	11 302
第四等级	0.2	3 060	5 197	13 159
第五等级	0.2	3 387	5 797	15 470
第六等级	0.1	3 822	6 506	18 049
第七等级	0.1	4 483	7 679	22 965

资料来源：中国国家统计局《中国统计年鉴》。

3. 表2—17列示了马来西亚1986年和1995年实际GDP和其中各个产业的规模。计算不同产业的贡献度与贡献率。

表2—17　　马来西亚的实际GDP和其中的各个产业规模　　单位：100万林吉特

产　　业	1986年	1995年
实际GDP	57 859	120 489
(1) 农林水产业	12 389	16 721
(2) 矿业	6 433	8 851
(3) 制造业	12 111	39 895
(4) 建筑业	2 355	5 287
(5) 电力、煤气、自来水业	1 027	2 820
(6) 批发、零售业	6 147	14 635
(7) 金融、保险、房地产业	5 073	12 884
(8) 运输、通信业	3 851	8 787
(9) 服务业与其他	1 352	2 436
(10) 政府服务业	7 121	8 173

说明：按1978年价格计算。

资料来源：财政部《经济报告》。

4. 表2—18列示了日本10种鲜鱼的年平均价格 p_t 和每个家庭年均购买数量 q_t。

(1) 以1980年为基期，分别计算1995年的拉氏价格指数、帕氏价格指数和费雪价格指数。

(2) 同样，以1980年为基期，计算1995年的拉氏数量指数、帕氏数量指数和费雪数量指数。

表2—18　　鲜鱼的年平均价格与每个家庭的购买量　　单位：日元/100克；100克

鲜　　鱼	1980年		1995年	
	价格 p_0	数量 q_0	价格 p_1	数量 q_1
金枪鱼	259	33	246	36
竹荚鱼	122	22	108	25
沙丁鱼	40	25	71	15

续前表

鲜　　鱼	1980 年		1995 年	
	价格 p_0	数量 q_0	价格 p_1	数量 q_1
鲑鱼	145	11	140	28
青花鱼	48	32	71	13
秋刀鱼	67	23	68	20
鲷	217	11	239	9
鰤鱼	215	25	190	24
墨鱼	95	79	102	46
章鱼	166	10	151	13

资料来源：总务厅统计局《家庭调查年报》。

第三章 一元回归模型

回归分析是计量经济分析中利用最多的一种方法，它是用来分析两个及两个以上的变量相互之间因果关系的统计方法。从某种意义上说，计量经济学就是回归分析的不断改进与发展。本章简要说明回归分析方法。

1. 一元回归模型

现在，我们用一个一次方程来表示两个数据 X 和 Y 之间的关系。

$$Y_i = \alpha + \beta X_i + u_i \qquad i = 1,2,\cdots,n \tag{3—1}$$

（下面，如无必要将省略 i。）

这个公式就是一元回归模型。X 表示原因（cause）的变量，称为解释变量（explanatory variable）或独立变量

(independent variable)；Y 表示结果的变量，称为被解释变量（explained variable）或从属变量（dependent variable）。u 是误差项（error term）或扰乱项（disturbance term），它是 Y 的变化中不能完全由 X 的变化来解释的部分，换句话说，它表示的是 Y（实际值）与 $\alpha+\beta X$（理论值）之间的偏差。

一般说来，回归分析是为了发现 X 作为原因，Y 作为结果时两者之间的因果关系，因此，X 和 Y 之间的关系，理论上必须能够成立。如果对没有理论意义的公式进行推算，即使能够得出良好的推算结论，也是没有意义的分析，这一点必须充分注意。

回归分析的主要目的是估计回归系数 α、β，最常用的方法就是下面将要介绍的最小二乘法（ordinary least square method，OLS）。

2. 最小二乘法

利用 OLS 来估计（3—1）式，可以得到所谓的估计回归线（最小二乘回归线），即

$$\hat{Y}=\hat{\alpha}+\hat{\beta}X \tag{3—2}$$

$\hat{\alpha}$、$\hat{\beta}$ 称为 α 和 β 的估计值（最小二乘估计值）。$\hat{Y}$ 读做 Y hat，如图 3—1 所示，它是与 X 实际值（观测值）相对应的估计回归线上的 Y 值，称为 Y 的理论值（估计值、计算值、预测值等）。

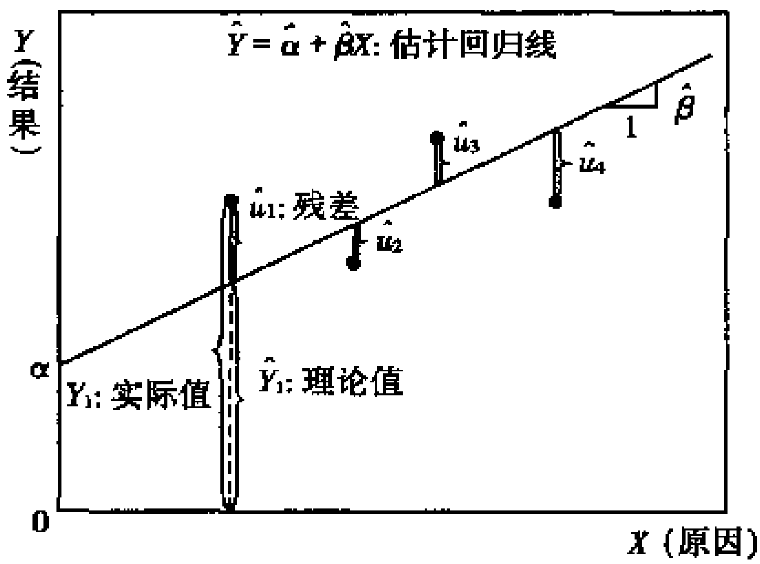


图 3—1　估计回归线与残差

根据 OLS 计算 $\hat{\alpha}$、$\hat{\beta}$ 的公式如下：

$$\hat{\beta} = \frac{n\sum XY - \sum X \sum Y}{n\sum X^2 - (\sum X)^2} \tag{3—3}$$

$$= \frac{\sum (X - \overline{X})(Y - \overline{Y})}{\sum (X - \overline{X})^2} = \frac{\sum xy}{\sum x^2} \tag{3—4}$$

$$= \frac{\frac{1}{n-1}\sum (X - \overline{X})(Y - \overline{Y})}{\frac{1}{n-1}\sum (X - \overline{X})^2} = \frac{X\text{和}Y\text{的协方差}}{X\text{的方差}} \tag{3—5}$$

$$\hat{\alpha} = \frac{\sum X^2 \sum Y - \sum X \sum XY}{n\sum X^2 - (\sum X)^2} \tag{3—6}$$

$$= \frac{\sum Y - \hat{\beta}\sum X}{n} \tag{3—7}$$

$$= \overline{Y} - \hat{\beta}\,\overline{X} \tag{3—8}$$

下面，就公式的推导方法，作一些简单的说明。

首先，设残差 (residual) 为 a，

残差 = 实际值 − 理论值

$$a = Y - \hat{Y} = Y - (\hat{\alpha} + \hat{\beta}X) \tag{3—9}$$

其次，计算残差的 2 次方的总和，即残差平方和 (residual sum of squares, RSS)，得

$$\sum a^2 = \sum [Y - (\hat{\alpha} + \hat{\beta}X)]^2 \tag{3—10}$$

寻找能够使残差平方和最小的 $\hat{\alpha}$ 和 $\hat{\beta}$ 值，就是 OLS 的基本原理 (这是 1794 年由德国数学家高斯发现的)。为了求残差平方和 $\sum a^2$ 关于 $\hat{\alpha}$ 和 $\hat{\beta}$ 的最小值，需要将 (3—10) 式对 $\hat{\alpha}$、$\hat{\beta}$ 分别求偏微分，并设其为零。即

$$\frac{\partial \sum a^2}{\partial \hat{\alpha}} = -2\sum (Y - \hat{\alpha} - \hat{\beta}X) = 0 \tag{3—11}$$

$$\frac{\partial \sum a^2}{\partial \hat{\beta}} = -2\sum X(Y - \hat{\alpha} - \hat{\beta}X) = 0 \tag{3—12}$$

将这两个方程整理，得联立方程：

$$\sum Y = n\hat{\alpha} + \hat{\beta}\sum X \tag{3—13}$$

$$\sum XY = \hat{\alpha}\sum X + \hat{\beta}\sum X^2 \tag{3—14}$$

这个联立方程称为正规方程（normal equation），这个方程关于未知数 $\hat{\beta}$、$\hat{\alpha}$ 的解，正是前面出现过的 OLS 的公式（3—3）式和（3—6）式。而且，（3—3）式变形得到（3—4）式和（3—5）式，（3—6）式变形得到（3—7）式和（3—8）式。

[例题 3—1]

利用下面的数据，对单元回归模型 $Y=\alpha+\beta X+u$ 进行最小二乘法估计。

X	6	11	17	8	13
Y	1	3	5	2	4

[解答]

将数据代人工作表中进行计算（见表 3—1）。

表 3—1

X	Y	XY	X^2
6	1	6	36
11	3	33	121
17	5	85	289
8	2	16	64
13	4	52	169
55	15	192	679
↑	↑	↑	↑
$\sum X$	$\sum Y$	$\sum XY$	$\sum X^2$

根据（3—3）式，求 $\hat{\beta}$。

$$\hat{\beta} = \frac{n\sum XY - \sum X\sum Y}{n\sum X^2 - (\sum X)^2} = \frac{5\times192 - 55\times15}{5\times679 - (55)^2} = \frac{135}{370} = 0.365$$

根据（3—7）式，求 $\hat{\alpha}$。

$$\hat{\alpha} = \frac{\sum Y - \hat{\beta}\sum X}{n} = \frac{15 - \frac{135}{370}\times 55}{5} \doteq -1.01$$

因此，估计的回归曲线为

$$\hat{Y} = -1.01 + 0.365X$$

[又解]

将数据记人工作表，进行计算（见表3—2）。

$$\overline{X} = \frac{\sum X}{n} = \frac{55}{5} = 11$$

$$\overline{Y} = \frac{\sum Y}{n} = \frac{15}{5} = 3$$

设 $x = X - \overline{X}$

$y = Y - \overline{Y}$

表3—2

X	Y	x	y	xy	x^2
6	1	-5	-2	10	25
11	3	0	0	0	0
17	5	6	2	12	36
8	2	-3	-1	3	9
13	4	2	1	2	4
55	15	0	0	27	74
↑ $\sum X$	↑ $\sum Y$	↑ $\sum x$	↑ $\sum y$	↑ $\sum xy$	↑ $\sum x^2$

根据（3—4）式，求 $\hat{\beta}$。

$$\hat{\beta} = \frac{\sum xy}{\sum x^2} = \frac{27}{74} = 0.365$$

根据（3—8）式，求回归系数（常数项）$\hat{\alpha}$。

$$\hat{\alpha} = \overline{Y} - \beta\overline{X} = 3 - \frac{27}{74}\times 11 = -1.01$$

因此，估计的回归曲线为

$$\hat{Y} = -1.01 + 0.365X$$

3. 决定系数

决定系数（coefficient of determination）是反映估计的回归曲线对观测的数据的解释能力，或者说是反映两者拟合优度（goodness of fit）的尺度，是回归分析不可缺少的统计量。我们将 Y 的实际值与平均值之差的平方和称做 Y 的全部变化，即

$$Y\text{的全部变化} = \sum(Y-\overline{Y})^2 \tag{3—15}$$

另外，将 Y 的理论值 $\hat{Y}$ 与平均值 $\overline{Y}$ 之差的平方和，称做能够由回归解释的变化（回归平方和），即

$$\text{能够由回归解释的变化} = \sum(\hat{Y}-\overline{Y})^2 \tag{3—16}$$

这样，所谓的决定系数 R^2 反映的是 Y 的全部变化中，能够由回归来说明的部分的比率，其定义公式如下：

$$R^2 = \frac{\text{能够由回归解释的变化（回归平方和）}}{Y\text{的全部变化}} \tag{3—17}$$

$$= \frac{\sum(\hat{Y}-\overline{Y})^2}{\sum(Y-\overline{Y})^2} \tag{3—18}$$

$$= 1 - \frac{\text{不能由回归解释的变化（残差平方和）}}{Y\text{的全部变化}} \tag{3—19}$$

$$= 1 - \frac{\sum(Y-\hat{Y})^2 \longleftarrow = \sum \hat{u}^2}{\sum(Y-\overline{Y})^2} \tag{3—20}$$

决定系数的计算公式可以采用下面三个公式中的任何一个，具体如下：

$$R^2 = \frac{\hat{\beta}^2\sum(X-\overline{X})^2}{\sum(Y-\overline{Y})^2} \tag{3—21}$$

$$= \frac{[\sum(X-\overline{X})(Y-\overline{Y})]^2}{\sum(X-\overline{X})^2\sum(Y-\overline{Y})^2} = \frac{(\sum xy)^2}{\sum x^2 \sum y^2} \tag{3—22}$$

$$= \frac{[n\sum XY - (\sum X)(\sum Y)]^2}{[n\sum X^2 - (\sum X)^2][n\sum Y^2 - (\sum Y)^2]} \quad (3—23)$$

此外，决定系数和相关系数（参见第一章）之间，有以下的关系：

决定系数＝相关系数的平方 (3—24)

决定系数的取值范围是 $0 \leqslant R^2 \leqslant 1$，决定系数越接近于 1，理论值 $\hat{Y}$ 和实际值 Y 越近似，说明模型的解释能力越强。例如，如果 $R^2 = 0.9$，则模型具有 90％ 的解释力，如果 $R^2 = 0.3$，则说明模型只有 30％的解释力。

［例题 3—2］

利用例题 3—1 的回归分析结果，计算决定系数 R^2。

［解答］

$$\sum X = 55 \quad \sum Y = 15 \quad \sum XY = 192$$

$$\sum X^2 = 679 \quad \sum Y^2 = 55$$

根据（3—23）式，求决定系数 R^2。

$$R^2 = \frac{[n\sum XY - (\sum X)(\sum Y)]^2}{[n\sum X^2 - (\sum X)^2][n\sum Y^2 - (\sum Y)^2]}$$

$$= \frac{[5 \times 192 - 55 \times 15]^2}{[5 \times 679 - (55)^2][5 \times 55 - (15)^2]} = \frac{18\,225}{18\,500} = 0.985$$

可见，模型的拟合状况非常良好。

［又解］

$$x = X - \overline{X} \qquad y = Y - \overline{Y}$$

$$\sum x^2 = 74 \quad \sum y^2 = 10 \quad \sum xy = 27$$

由此，可以根据（3—22）式，求决定系数。

$$R^2 = \frac{(\sum xy)^2}{\sum x^2 \sum y^2} = \frac{(27)^2}{74 \times 10} = \frac{729}{740} = 0.985$$

[例题 3—3]

表 3—3 显示了日本在 1981—1994 年的 14 年间，家庭实际可支配收入 X 与家庭实际最终消费支出 Y 的数据。

(1) 设横轴为 X，纵轴为 Y，画出数据的散点图。

(2) 对下面的宏观消费函数，进行最小二乘法估计。

$$Y = \alpha + \beta X + u$$

(3) 计算决定系数 R^2，并考察估计出来的宏观消费函数的拟合优度。

(4) 求理论值 $\hat{Y}$ 和残差 $\hat{u}$。

(5) 利用 (2) 中估算出来的宏观消费函数，对家庭实际可支配收入在 270 兆日元和 320 兆日元时，家庭实际最终消费支出分别进行预测。

表 3—3　　日本的家庭实际可支配收入与实际最终消费支出　　单位：兆日元

年　份	家庭实际可支配收入 X	家庭实际最终消费支出 Y
1981	212	173
1982	217	181
1983	223	187
1984	228	192
1985	234	198
1986	243	205
1987	248	214
1988	259	225
1989	271	236
1990	280	246
1991	290	252
1992	296	257
1993	300	260
1994	304	265

说明：1990 年价格。

资料来源：经济企划厅《国民经济计算年报》。

[解答]

(1) 见图 3—2。

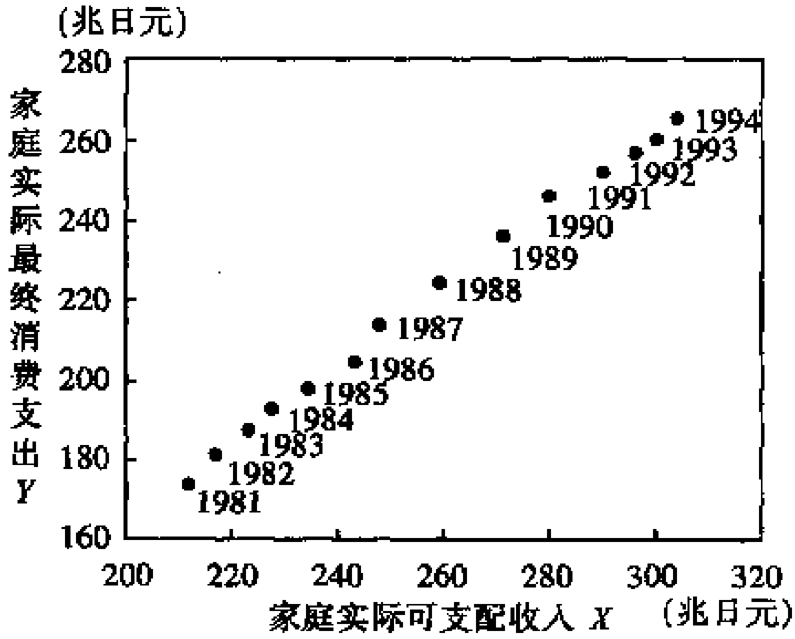


图 3—2 收入与消费的散点图

(2) 将数据代入工作表进行计算 (见表 3—4)。

表 3—4

X	Y	XY	X^2	Y^2
212	173	36 676	44 944	29 929
217	181	39 277	47 089	32 761
223	187	41 701	49 729	34 969
228	192	43 771	51 984	36 864
234	198	46 332	54 756	39 204
243	205	49 815	59 049	42 025
248	214	53 072	61 504	45 796
259	225	58 275	67 081	50 625
271	236	63 956	73 441	55 696
280	246	68 880	78 400	60 516
290	252	73 080	84 100	63 504
296	257	76 072	87 616	66 049
300	260	78 000	90 000	67 600
304	265	80 560	92 416	70 225
3 605	3 091	809 472	942 109	695 763
↑	↑	↑	↑	↑
$\sum X$	$\sum Y$	$\sum XY$	$\sum X^2$	$\sum Y^2$

首先，根据（3—3）式，求 $\hat{\beta}$。

$$\hat{\beta}=\frac{n\sum XY-\sum X\sum Y}{n\sum X^2-(\sum X)^2}=\frac{14\times 809\,472-3\,605\times 3\,091}{14\times 942\,109-(3\,605)^2}$$

$$=\frac{189\,553}{193\,501}=0.979\,60$$

这个 $\hat{\beta}$ 称做边际消费倾向（marginal propensity to consume），它的意思是如果家庭的实际可支配收入增加了 1 兆日元，将会引起家庭实际最终消费支出增加9 796亿日元。

其次，根据（3—7）式，求 $\hat{\alpha}$。

$$\hat{\alpha}=\frac{\sum Y-\hat{\beta}\sum X}{n}=\frac{3\,091-\frac{189\,553}{193\,501}\times 3\,605}{14}$$

$$=-31.461$$

因此，估计的日本宏观消费函数为

$$\hat{Y}=-31.461+0.979\,60X$$

（3）根据（3—23）式，求决定系数 R^2。

$$R^2=\frac{[n\sum XY-(\sum X)(\sum Y)]^2}{[n\sum X^2-(\sum X)^2][n\sum Y^2-(\sum Y)^2]}$$

$$=\frac{[14\times 809\,472-3\,605\times 3\,091]^2}{[14\times 942\,109-(3\,605)^2][14\times 695\,763-(3\,091)^2]}$$

$$=\frac{35\,930\,339\,809}{36\,068\,779\,901}=0.996\,2$$

这就是说，实际家庭最终消费支出 Y 的变化中，有 99.62%可以由模型($\hat{\alpha}+\hat{\beta}X$)来解释。估计出来的宏观消费函数具有极高的拟合优度。

（4）关于理论值 $\hat{Y}$，可在（2）中求出的下式中，

$$\hat{Y}=-31.461+0.927\,60X$$

代入 X 的实际值求解。

另外，可以根据（3—9）式，求残差 a。

$$a=Y-\hat{Y}$$

表 3—5 是理论值与残差的计算结果。

$a = Y - \hat{Y} < 0$ 即 $Y < \hat{Y}$→过大估计

$a = Y - \hat{Y} > 0$ 即 $Y > \hat{Y}$→过小估计

在回归分析中，考虑过大估计、过小估计的发生原因，非常重要。

表 3—5 理论值与残差

年份	实际值 Y	理论值 $\hat{Y}$	残差 a
1981	173	176.2	-3.2
1982	181	181.1	-0.1
1983	187	187.0	0.0
1984	192	191.9	0.1
1985	198	197.8	0.2
1986	205	206.6	-1.6
1987	214	211.5	2.5
1988	225	222.3	2.7
1989	236	234.0	2.0
1990	246	242.8	3.2
1991	252	252.6	-0.6
1992	257	258.5	-1.5
1993	260	262.4	-2.4
1994	265	266.3	-1.3

(5) ① $X = 270$ 兆日元时，

$$\hat{Y} = -31.461 + 0.979\,60X = -31.461 + 0.979\,60 \times 270$$
$$= 233.0\text{（兆日元）}$$

② $X = 320$ 兆日元时，

$$\hat{Y} = -31.461 + 0.979\,60X = -31.461 + 0.979\,60 \times 320$$
$$= 282.0\text{（兆日元）}$$

① 是在解释变量 X 的实际取值的范围之内（212 兆日元$\leqslant X \leqslant$304 兆日元）

预测 Y，称为内插预测（interpolation）。② 是在解释变量 X 的实际值之外预测 Y，称为外插预测（extrapolation）。

［例题 3—4］

表 3—6 反映了在东京证券交易所第一部上市的医药行业中 34 家公司 1996 年 3 月决算时的销售额与经常收益。两者之间的关系通常被认为是日本企业经营中一种很强的财务因素。

(1) 对一元回归模型 $Y=\alpha+\beta X+u$ 进行最小二乘法估计。

(2) 计算决定系数 R^2 和相关系数 R。

(3) 求理论值 $\hat{Y}$ 与残差 $\hat{u}$。

表 3—6　医药业制造商销售额与经常收益（东证第一部：1996 年 3 月）　单位：10 亿日元

公司名	销售额 X	经常收益 Y	公司名	销售额 X	经常收益 Y
三共	410	87	艾再	254	40
武田药品工业	602	92	罗特制药	28	4
山之内工业	295	61	小野药品工业	132	55
第一制药	217	43	日研化学	49	3
大日本制药	130	10	久光制药*	37	4
盐野义制药	226	22	东京田边制药	40	4
田边制药	175	9	持田制药	70	9
吉富制药	102	13	大正制药	221	62
藤泽药品制药	219	22	参天制药	65	17
川本制药	11	1	SS 制药	57	5
帝国脏器制药	24	3	扶桑药品工业	41	2
万有制药	131	22	日本开米发	21	1
日本新药	52	6	津村	98	8
富山化学工业	40	3	特鲁姆	104	16
中外制药	169	21	富士**	23	1
科研制药	70	4	北陆制药	16	1
绿十字	18	−1	气生药品工业	55	15

*久光为 1996 年 2 月。

**富士为 1995 年 12 月。

资料来源：东洋经济新报社《公司四季报》。

[解答]

(1) 省略工作表。

$$\sum X = 4\,202 \qquad \sum Y = 665 \qquad \sum XY = 176\,703$$

$$\sum X^2 = 1\,053\,192 \qquad \sum Y^2 = 33\,391$$

根据 (3—3) 式，求 $\hat{\beta}$。

$$\hat{\beta} = \frac{n\sum XY - \sum X \sum Y}{n\sum X^2 - (\sum X)^2} = \frac{34 \times 176\,703 - 4\,202 \times 665}{34 \times 1\,053\,192 - (4\,202)^2}$$

$$= \frac{3\,213\,572}{18\,151\,724} = 0.177\,04$$

根据 (3—7) 式，求 $\hat{\alpha}$。

$$\hat{\alpha} = \frac{\sum Y - \hat{\beta}\sum X}{n} = \frac{665 - \dfrac{3\,213\,572}{18\,151\,724} \times 4\,202}{34}$$

$$= -2.321\,2$$

因此，估计出来的回归曲线为

$$\hat{Y} = -2.321\,2 + 0.177\,04X$$

$\hat{\beta}$ 等于 0.177 04 意味着医药行业销售额每增加 10 亿日元，能够引起经常收益增加 1.770 4 亿日元。

(2) 根据 (3—23) 式，求决定系数 R^2。

$$R^2 = \frac{[n\sum XY - (\sum X)(\sum Y)]^2}{[n\sum X^2 - (\sum X)^2][n\sum Y^2 - (\sum Y)^2]}$$

$$= \frac{[34 \times 176\,703 - 4\,202 \times 665]^2}{[34 \times 1\,053\,192 - (4\,202)^2][34 \times 33\,391 - (665)^2]}$$

$$= \frac{10\,327\,044\,999\,184}{12\,580\,397\,200\,956} = 0.820\,9$$

即经常收益的变化中，有 82.09%可以由模型 ($\hat{\alpha} + \hat{\beta}X$) 来解释。

再求相关系数，得

$$R = \sqrt{R^2} = \sqrt{0.820\,9}$$

$=0.9060$

顺便提一下，通过对这个相关系数的检验可以看出，它在1%的水平上显著，意味着 X 和 Y 不相关的概率小于1%，也就是说销售额与经常收益之间存在着显著的相关关系（R 的检验方法，请参照第一章）。

(3) 理论值 $\hat{Y}$ 可以在（1）中求出的下式中，代入 X 的实际值来计算。

$$Y=-2.3212+0.17704X$$

残差 a 可以由（3—9）式计算，即

$$a=Y-\hat{Y}$$

表3—7是理论值与残差的计算结果。正的残差从大到小依次为小野(34.0)、大正（25.2)、三共（16.7)，说明其盈利能力值得怀疑。

表3—7　　　　理论值与残差　　　　单位：10亿日元

公司名	理论值 $\hat{Y}$	残差 a	公司名	理论值 $\hat{Y}$	残差 a
三共	70.3	16.7	艾再	42.6	-2.6
武田药品工业	104.3	-12.3	罗特制药	2.6	1.4
山之内工业	50.0	11.0	小野药品工业	21.0	34.0
第一制药	36.1	6.9	日研化学	6.4	-3.4
大日本制药	20.7	-10.7	久光制药	4.2	-0.2
盐野义制药	37.7	-15.7	东京田边制药	4.8	-0.8
田边制药	28.7	-19.7	持田制药	10.1	-1.1
吉富制药	15.7	-2.7	大正制药	36.8	25.2
藤泽药品制药	36.5	-14.5	参天制药	9.2	7.8
川本制药	-0.4	1.4	SS制药	7.8	-2.8
帝国脏器制药	1.9	1.1	扶桑药品工业	4.9	-2.9
万有制药	20.9	1.1	日本开米发	1.4	-0.4
日本新药	6.9	-0.9	津村	15.0	-7.0
富山化学工业	4.8	-1.8	特鲁姆	16.1	-0.1
中外制药	27.6	-6.6	富士	1.8	-0.8
科研制药	10.1	-6.1	北陆制药	0.5	0.5
绿十字	0.9	-1.9	气生药品工业	7.4	7.6

[例题 3—5]

表 3—8 反映了以下一些产业的大企业中（从业人数在 1 000 以上），男性大学毕业生的工作年数与相应的收入水平。这些产业分别为：①产业合计；②建筑业；③制造业；④运输、通信业；⑤批发、零售、餐饮业；⑥金融、保险业。

(1) 对①产业合计，估计其回归模型 $Y=\alpha+\beta X+u$，并计算决定系数 R^2。

(2) 同样，对②～⑥的各个产业，估计其回归模型 $Y=\alpha+\beta X+u$，并计算决定系数 R^2。

表 3—8　　男性大学毕业生的工作年数与相应的收入（大企业:1995 年）

① 产业合计		② 建筑业		③ 制造业	
工作年数 X（年）	相应收入 Y（千日元）	工作年数 X（年）	相应收入 Y（千日元）	工作年数 X（年）	相应收入 Y（千日元）
1.3	213	1.2	218	1.5	215
4.2	262	4.3	268	4.2	249
8.5	340	8.5	341	8.6	319
12.7	429	12.8	428	12.8	402
17.6	523	18.1	519	18.0	498
22.3	597	22.4	581	22.9	581
27.1	670	26.5	677	27.9	658
④ 运输、通信业		**⑤ 批发、零售、餐饮业**		**⑥ 金融、保险业**	
工作年数 X（年）	相应收入 Y（千日元）	工作年数 X（年）	相应收入 Y（千日元）	工作年数 X（年）	相应收入 Y（千日元）
1.3	207	1.3	213	1.3	207
4.1	245	4.4	259	4.6	286
7.9	308	8.7	328	9.0	404
12.7	377	13.1	401	13.7	516
17.6	458	17.6	484	18.5	598
22.1	536	22.5	567	22.8	675
26.9	625	27.3	640	27.8	715

资料来源：劳动省《工资结构基本调查统计》。

[解答]

(1) 省略工作表。

$$\sum X = 93.7 \qquad \sum Y = 3\ 034 \qquad \sum XY = 50\ 390.5$$

$$\sum X^2 = 1\ 794.33 \qquad \sum Y^2 = 1\ 492\ 492$$

根据 (3—3) 式，求 $\hat{\beta}$。

$$\hat{\beta} = \frac{n\sum XY - \sum X\sum Y}{n\sum X^2 - (\sum X)^2} = \frac{7 \times 50\ 390.5 - 93.7 \times 3\ 034}{7 \times 1\ 794.33 - (93.7)^2}$$

$$= \frac{68\ 447.7}{3\ 780.62} = 18.105$$

根据 (3—7) 式，求 $\hat{\alpha}$。

$$\hat{\alpha} = \frac{\sum Y - \hat{\beta}\sum X}{n} = \frac{3\ 034 - \frac{68\ 447.7}{3\ 780.62} \times 93.7}{7}$$

$$= 191.08$$

由此，可得估计出来的回归曲线为

$$\hat{Y} = 191.08 + 18.105X$$

根据 (3—23) 式，求决定系数 R^2。

$$R^2 = \frac{[n\sum XY - (\sum X)(\sum Y)]^2}{[n\sum X^2 - (\sum X)^2][n\sum Y^2 - (\sum Y)^2]}$$

$$= \frac{[7 \times 50\ 390.5 - 93.7 \times 3\ 034]^2}{[7 \times 1\ 794.33 - (93.7)^2][7 \times 1\ 492\ 492 - (3\ 034)^2]}$$

$$= \frac{4\ 685\ 087\ 635}{4\ 696\ 618\ 859} = 0.997\ 5$$

由此可见，模型的拟合优度极其良好。这一回归分析结果明确反映了日本年功序列制工资体系的特征。$\hat{\beta}$ 意味着工作年数每增加 1 年，月工资增加 18.105 千日元，$\hat{\alpha}$ 意味着刚开始工作时的工资水平为 191.080 千日元。

(2) 与 (1) 相同，利用 (3—3) 式、(3—7) 式和 (3—23) 式分别计算 $\hat{\beta}$、$\hat{\alpha}$ 和决定系数 R^2。

② 建筑业

$\sum X = 93.8 \qquad \sum Y = 3\ 032 \qquad \sum XY = 50\ 139.7$

$\sum X^2 = 1\ 787.64 \qquad \sum Y^2 = 1\ 484\ 064$

$\therefore\ \hat{Y} = 193.01 + 17.921X$

$R^2 = 0.998\ 1$

③ 制造业

$\sum X = 95.9 \qquad \sum Y = 2\ 922 \qquad \sum XY = 49\ 884.4$

$\sum X^2 = 1\ 884.51 \qquad \sum Y^2 = 1\ 390\ 120$

$\therefore\ \hat{Y} = 180.89 + 17.265X$

$R^2 = 0.998\ 4$

④ 运输、通信业

$\sum X = 92.6 \qquad \sum Y = 2\ 756 \qquad \sum XY = 45\ 213.6$

$\sum X^2 = 1\ 763.98 \qquad \sum Y^2 = 1\ 227\ 552$

$\therefore\ \hat{Y} = 178.83 + 16.244X$

$R^2 = 0.998\ 2$

⑤ 批发、零售、餐饮业

$\sum X = 94.9 \qquad \sum Y = 2\ 892 \qquad \sum XY = 48\ 271.1$

$\sum X^2 = 1\ 829.65 \qquad \sum Y^2 = 1\ 346\ 180$

$\therefore\ \hat{Y} = 186.88 + 16.690X$

$R^2 = 0.999\ 4$

⑥ 金融、保险业

$\sum X = 97.7 \qquad \sum Y = 3\ 401 \qquad \sum XY = 58\ 619.9$

$\sum X^2 = 1\ 926.47 \qquad \sum Y^2 = 1\ 878\ 571$

$\therefore\ \hat{Y} = 209.33 + 19.813X$

$R^2 = 0.976\ 9$

决定系数都很高，说明模型的拟合优度都非常良好。$\hat{\beta}$ 的估计值表示每年相应的工资增长额，依次分别是金融、保险业（19.813 千日元）、建筑业（17.921 千日元）、制造业（17.265 千日元）、批发、零售、餐饮业（16.690 千日元）、运输、通信业（16.244 千日元）。

4. 非线性方程的回归分析

到目前为止，我们学习的都是线性函数，或者说是一次方程的回归分析，下面将介绍非线性方程（non-linear equation）的回归分析。

表 3—9　　非线性方程的线性转换

非线性式	变　换	线性式（估计式）
(1) 指数函数 I $Y = \alpha X^{\beta}$	两边取对数 $\log Y = \log\alpha + \beta\log X$ 设 $y = \log Y$、$x = \log X$、 $a = \log\alpha$	$y = a + \beta x$
(2) 指数函数 II $Y = \alpha \cdot \beta X$	两边取对数 $\log Y = \log\alpha + X\log\beta$ 设 $y = \log Y$、 $a = \log\alpha$、$b = \log\beta$	$y = a + bX$
(3) 分数函数 I $Y = \alpha + \frac{\beta}{X}$	设 $x = \frac{1}{X}$	$Y = \alpha + \beta x$
(4) 分数函数 II $Y = \frac{1}{\alpha + \beta X}$	设 $y = \frac{1}{Y}$	$y = \alpha + \beta X$
(5) 半对数函数 $Y = \alpha + \beta\log X$	设 $x = \log X$	$Y = \alpha + \beta x$

续前表

非线性式	变　换	线性式（估计式）
(6) 2 次函数 $Y=\alpha+\beta X+\gamma X^2$	设 $Z=X^2$	$Y=\alpha+\beta X+\gamma Z$ 多元回归，参照第 4 章
(7) 柯布－道格拉斯函数 $Z=aX^{\beta}Y^{\gamma}$	两边取对数 $\log Z=\log\alpha+\beta\log X+\gamma\log Y$ 设 $z=\log Z$、$x=\log X$、$y=\log Y$、$a=\log\alpha$	$z=a+\beta x+\gamma y$ 多元回归，参照第 4 章
(8) 逻辑函数Ⅰ $Y=\dfrac{e^{\alpha+\beta X}}{1+e^{\alpha+\beta X}}$ （普及率：$0<Y<1$）	设 $y=\log\dfrac{Y}{1-Y}$	$y=\alpha+\beta X$
(9) 逻辑函数Ⅱ $Y=\dfrac{\gamma}{1+e^{\alpha+\beta X}}$ 普及率的饱和水平：γ	设 $y=\log\left(\dfrac{\gamma}{Y}-1\right)$	$y=\alpha+\beta X$

被解释变量和解释变量之间的回归关系，当数据的分散情况在一定的较小的范围时，大部分线性近似，但是随着范围的扩大，非线性近似的情况很多。一般说来，非线性方程的回归分析，通过对被解释变量和解释变量的某种变换，可以将非线性方程转换成能够用最小二乘法进行分析的线性方程。表 3—9 所列的线性变换，是非线性函数回归分析中常用的类型。计量经济分析中，对于已经给定的数据，找出具有较强说服力的函数，是一件非常重要的工作，而且，这种函数必须满足由经济理论派生出来的各种条件。

[例题 3—6]

表 3—10 反映了日本物价上涨率 $\dot{P}$ 与失业率 U 之间的关系。

(1) 设横轴是 U，纵轴是 $\dot{P}$，画出数据的散点图。

(2) 对下面的菲力普斯曲线，进行 OLS 估计。

$$\dot{P}=\alpha+\beta\frac{1}{U}+u$$

(3) 计算决定系数 R^2。

表 3—10　　日本物价上涨率与失业率的关系

年　份	物价上涨率（%）$\dot{P}$	失业率（%）U
1986	0.6	2.8
1987	0.1	2.8
1988	0.7	2.5
1989	2.3	2.3
1990	3.1	2.1
1991	3.3	2.1
1992	1.6	2.2
1993	1.3	2.5
1994	0.7	2.9
1995	−0.1	3.2

资料来源：日本银行《经济统计年报》。

[解答]

(1) 散点图如图 3—3 所示。

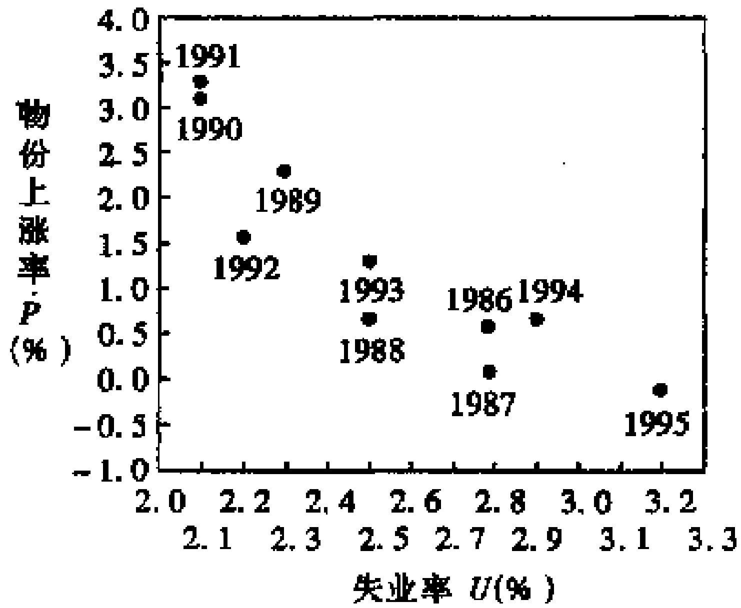


图 3—3　日本的物价上涨率和失业率

(2) 设 $X=\frac{1}{U}$、$Y=\dot{P}$，将菲力普斯曲线变换为线性曲线 $Y=\alpha+\beta X+u$。

$$\sum X = 4.013\ 32 \quad \sum Y = 13.6 \quad \sum XY = 6.035\ 02$$

$$\sum X^2 = 1.640\ 83 \quad \sum Y^2 = 31.40$$

根据 (3—3) 式，求 $\hat{\beta}$。

$$\hat{\beta} = \frac{n\sum XY - \sum X \sum Y}{n\sum X^2 - (\sum X)^2} = \frac{10\times 6.035\ 02 - 4.013\ 32\times 13.6}{10\times 1.640\ 83 - (4.013\ 32)^2}$$

$$= \frac{5.769\ 05}{0.301\ 56} = 19.1$$

根据 (3—7) 式，求 $\hat{\alpha}$。

$$\hat{\alpha} = \frac{\sum Y - \beta\sum X}{n} = \frac{13.6 - \frac{5.769\ 05}{0.301\ 56}\times 4.013\ 32}{10}$$

$$= -6.32$$

因此，

$$Y = -6.32 + 19.1X$$

则菲力普斯曲线为

$$\dot{P} = -6.32 + 19.1\frac{1}{U}$$

(3) 根据 (3—23) 式，求决定系数 R^2。

$$R^2 = \frac{[n\sum XY - (\sum X)(\sum Y)]^2}{[n\sum X^2 - (\sum X)^2][n\sum Y^2 - (\sum Y)^2]}$$

$$= 0.855$$

[例题 3—7]

表 3—11 反映了日本 VTR 的普及率。对逻辑（logistic）函数进行线性变换，并用最小二乘法估计，求 VTR 普及率的饱和水平。

表 3—11　　日本 VTR 的普及率

年　份 (t)	普及率 (%) Y
1978 (1)	1.3
1979 (2)	2.0
1980 (3)	2.4
1981 (4)	5.1
1982 (5)	7.5
1983 (6)	11.8
1984 (7)	18.7
1985 (8)	27.8
1986 (9)	33.5
1987 (10)	43.0
1988 (11)	53.0
1989 (12)	63.7
1990 (13)	66.8
1991 (14)	71.5
1992 (15)	63.8
1993 (16)	75.1
1994 (17)	72.5
1995 (18)	73.7

资料来源：经济企划厅《消费动向调查年报》。

[解答]

用以表示 VTR 普及率的逻辑函数为

$$Y=\frac{\gamma}{1+e^{\alpha+\beta t}} \tag{1}$$

式中，Y 为 VTR 普及率 (%)；

t 为年份 ($t=1,2,\cdots,18$)；

γ 为 VTR 普及率的饱和水平；

e 为自然对数的底（e=2.718 28…）；

α、β 为待估计的参数。

首先，将（1）式进行线性变换，得

$$\log\left(\frac{\gamma}{Y}-1\right)=\alpha+\beta t \tag{2}$$

其次，饱和水平中代入大于 Y 的最大值的数，编制 $\log\left(\frac{\gamma}{Y}-1\right)$的数据，对(2)式进行OLS估计，以便求决定系数 R^2。每隔一定的 γ 反复计算，R^2 最大时，取最终的参数估计结果。表 3—12 反映了估计出来的饱和水平 γ 和相应的决定系数 R^2。

表 3—12　　普及率的饱和水平与决定系数

普及率的饱和水平 γ	决定系数 R^2	
76	0.964 174	
77	0.969 687	←R^2 最大的情形
78	0.969 549	
79	0.967 987	
80	0.965 955	

根据表 3—12，普及率的饱和水平 γ 为 77% 时，决定系数 R^2 最大，此时的估计结果为

$$\log\left(\frac{77.0}{Y}-1\right)=4.429-0.457\,4t$$

$$R^2=0.969\,7$$

表示为逻辑函数为

$$Y=\frac{77.0}{1+e^{4.429-0.457\,4t}}$$

第三章　练习题

1. 表 3—13 列示了美国 1960—1995 年的 36 年间，个人实际可支配收入 X 和个人实际消费支出 Y。

表 3—13　　**美国的个人实际可支配收入和个人实际消费支出**　　单位：100 亿美元

年　份	个人实际可支配收入 X	个人实际消费支出 Y	年　份	个人实际可支配收入 X	个人实际消费支出 Y
1960	157	143	1978	326	295
1961	162	146	1979	335	302
1962	169	153	1980	337	301
1963	176	160	1981	345	305
1964	188	169	1982	348	308
1965	200	180	1983	358	324
1966	211	190	1984	384	341
1967	220	196	1985	396	357
1968	230	207	1986	409	371
1969	237	215	1987	415	382
1970	247	220	1988	432	397
1971	256	228	1989	440	406
1972	268	242	1990	448	413
1973	287	253	1991	449	411
1974	285	251	1992	461	422
1975	290	257	1993	467	434
1976	301	271	1994	478	447
1977	311	283	1995	493	458

说明：1992 年价格。

资料来源：*Economic Report of the President*。

（1）1960—1969 年

（2）1970—1979 年

（3）1980—1995 年

估计上述各个期间的消费函数 $Y=\alpha+\beta X+u$，并计算决定系数 R^2。此外，分析估计出的边际消费倾向。

2．表 3—14 反映了中国台湾第一产业就业人数比率（Y）和人均 GDP（X）的变化。

（1）以 X 为横轴，Y 为纵轴，画出散点图。

（2）对回归模型 $Y=\alpha+\beta\dfrac{1}{X}+u$ 进行 OLS 估计。

（3）解释（2）中 α 估计值的意义。

表 3—14　　中国台湾第一产业就业人数比率与人均 GDP　　单位：100 新台币/人

年　份	第一产业就业人数比率（%）Y	人均 GDP X	年　份	第一产业就业人数比率（%）Y	人均 GDP X
1971	35.1	620	1984	17.6	1 448
1972	33.0	689	1985	17.5	1 500
1973	30.5	764	1986	17.0	1 658
1974	30.9	759	1987	15.3	1 848
1975	30.4	781	1988	13.7	1 970
1976	29.0	870	1989	12.9	2 111
1977	26.7	942	1990	12.8	2 198
1978	24.9	1 050	1991	13.0	2 340
1979	21.5	1 113	1992	12.3	2 475
1980	19.5	1 172	1993	11.5	2 607
1981	18.8	1 222	1994	10.9	2 754
1982	18.9	1 243	1995	10.6	2 896
1983	18.6	1 328			

说明：1991 年价格。

资料来源：Counsil for Economic Planning and Development，*Taiwan Statistical Data Book*。

3. 表 3—15 反映了日本普通工薪家庭 CD 唱机、文字处理机及摄像机的普及程度。对逻辑函数进行线性变换和最小二乘法估计，计算各种耐用消费品普及率的饱和水平。

表 3—15　　日本普通工薪家庭 CD 唱机、文字处理机、摄像机的普及程度（%）

年　份	CD 唱机 普及率	文字处理机 普及率	摄像机 普及率
1987	10.3	—	10.8
1988	18.2	15.5	12.1
1989	29.2	20.8	15.1
1990	38.6	25.7	16.1
1991	45.5	29.8	24.5
1992	52.3	33.8	27.2
1993	59.6	38.6	28.0
1994	59.9	40.9	31.7
1995	61.8	42.1	33.7

资料来源：经济企划厅《消费动向调查年报》。

4. 表 3—16 反映了中国建国以来人口的变化。

（1）利用逻辑函数计算（OLS）中国人口的饱和水平。

（2）根据（1）中逻辑函数的估计结果，预测中国 2010 年、2020 年、2030 年、2040 年、2050 年的人口。

表 3—16　　中国建国以来人口的变化　　单位：100 万人

年　份	人　口	年　份	人　口	年　份	人　口
1949	542	1956	628	1963	692
1950	552	1957	647	1964	705
1951	563	1958	660	1965	725
1952	575	1959	672	1966	745
1953	588	1960	662	1967	764
1954	603	1961	659	1968	785
1955	615	1962	673	1969	807

续前表

年 份	人 口	年 份	人 口	年 份	人 口
1970	830	1979	975	1988	1 110
1971	852	1980	987	1989	1 127
1972	872	1981	1 001	1990	1 143
1973	892	1982	1 017	1991	1 158
1974	909	1983	1 030	1992	1 172
1975	924	1984	1 044	1993	1 185
1976	937	1985	1 059	1994	1 199
1977	950	1986	1 075	1995	1 211
1978	963	1987	1 093		

资料来源：中国国家统计局《中国统计年鉴》。

第四章 多元回归模型

1. 多元回归分析

对被解释变量（结果）与两个以上的变量（原因）之间的关系进行估计的回归分析，称为多元回归分析（multiple regression analysis）。

现在，我们用 Y 表示被解释变量，X_1、X_2 分别表示两个解释变量，构建一个多元回归模型，

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + u \tag{4—1}$$

多元回归分析的主要目的是计算参数 α、β_1、β_2（u 为误差项）。

多元回归分析中参数估计的计算原理，与上一章学过的一元回归模型相同，都是依靠最小二乘法（OLS）。也就是计算能够使残差平方和 $\sum \hat{u}^2 = \sum(Y - \hat{\alpha} - \hat{\beta}_1 X_1 - \hat{\beta}_2 X_2)^2$ 最小的参数值。

将残差平方和分别对 $\hat{\alpha}$、$\hat{\beta}_1$、$\hat{\beta}_2$ 求偏微分，并设其为零，得到下面的联立方程组：

$$\sum Y = n\hat{\alpha} + \hat{\beta}_1 \sum X_1 + \hat{\beta}_2 \sum X_2 \tag{4—2}$$

$$\sum YX_1 = \hat{\alpha} \sum X_1 + \hat{\beta}_1 \sum X_1^2 + \hat{\beta}_2 \sum X_1 X_2 \tag{4—3}$$

$$\sum YX_2 = \hat{\alpha} \sum X_2 + \hat{\beta}_1 \sum X_1 X_2 + \hat{\beta}_2 \sum X_2^2 \tag{4—4}$$

设 $\hat{\alpha}$、$\hat{\beta}_1$、$\hat{\beta}_2$ 为未知数的这个联立方程组，称为多元回归的正规方程式，只要解开就可以了。但是，匆忙解这个方程组有时较为困难，一般可按照以下的步骤求解 $\hat{\alpha}$、$\hat{\beta}_1$、$\hat{\beta}_2$。

步骤一：编制工作表，求以下统计量。

$$\sum Y \qquad \sum X_1 \qquad \sum X_2$$

$$\sum Y^2 \qquad \sum X_1^2 \qquad \sum X_2^2$$

$$\sum YX_1 \qquad \sum YX_2 \qquad \sum X_1 X_2$$

步骤二：对算术平均的离差平方和与积和，分别记做 S_{YY} 和 S_{12}，并进行计算。

$$S_{YY} = \sum (Y - \bar{Y})^2 = \sum Y^2 - \frac{(\sum Y)^2}{n} \tag{4—5}$$

$$S_{11} = \sum (X_1 - \bar{X}_1)^2 = \sum X_1^2 - \frac{(\sum X_1)^2}{n} \tag{4—6}$$

$$S_{22} = \sum (X_2 - \bar{X}_2)^2 = \sum X_2^2 - \frac{(\sum X_2)^2}{n} \tag{4—7}$$

$$S_{Y1} = \sum (Y - \bar{Y})(X_1 - \bar{X}_1) = \sum YX_1 - \frac{(\sum Y)(\sum X_1)}{n} \tag{4—8}$$

$$S_{Y2} = \sum (Y - \bar{Y})(X_2 - \bar{X}_2) = \sum YX_2 - \frac{(\sum Y)(\sum X_2)}{n} \tag{4—9}$$

$$S_{12} = \sum (X_1 - \bar{X}_1)(X_2 - \bar{X}_2) = \sum X_1 X_2 - \frac{(\sum X_1)(\sum X_2)}{n} \tag{4—10}$$

步骤三：计算 D_0、D_1、D_2。

$$D_0 = S_{11}S_{22} - S_{12}^2 \tag{4—11}$$

$$D_1 = S_{Y1}S_{22} - S_{Y2}S_{12} \tag{4—12}$$

$$D_2 = S_{Y2}S_{11} - S_{Y1}S_{12} \tag{4—13}$$

步骤四：求 $\hat{\beta}_1$、$\hat{\beta}_2$、$\hat{\alpha}$。

$$\hat{\beta}_1 = \frac{D_1}{D_0} \tag{4—14}$$

$$\hat{\beta}_2 = \frac{D_2}{D_0} \tag{4—15}$$

$$\hat{\alpha} = \frac{\sum Y}{n} - \hat{\beta}_1 \frac{\sum X_1}{n} - \hat{\beta}_2 \frac{\sum X_2}{n} \tag{4—16}$$

一元回归是估计回归曲线，而多元回归是估计像图 4—1 这样的回归平面(regression plane)。

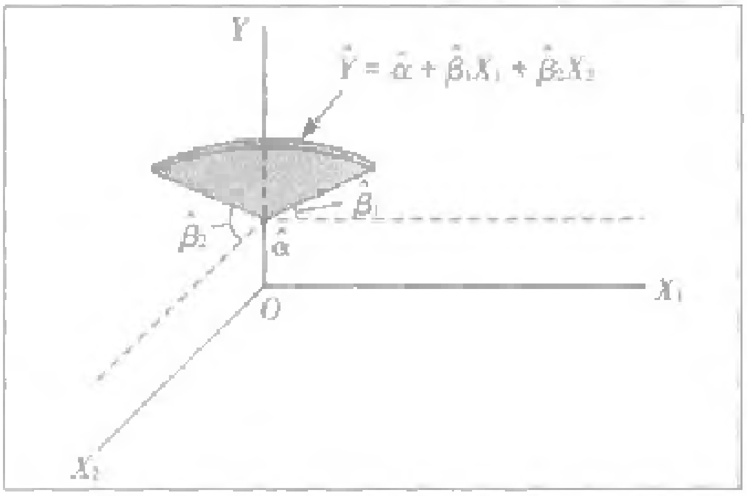


图 4—1　回归平面

$\hat{\beta}_1$、$\hat{\beta}_2$ 又称做偏回归系数（partial regression coefficient)，具有很重要的意义。$\hat{\beta}_1$ 表示的是当 X_2 既定时，由 X_1 的变化所引起的 Y 的变化，而 $\hat{\beta}_2$ 表示的是当 X_1 既定时，由 X_2 的变化所引起的 Y 的变化。

[例题 4—1]

利用下面的数据，对多元回归模型 $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + u$ 进行 OLS 估计。

Y	0	1	5	6	8
X_1	4	3	9	8	6
X_2	1	2	0	2	5

[解答]

步骤一：制作工作表，见表 4—1。

表 4—1

Y	X_1	X_2	Y^2	X_1^2	X_2^2	YX_1	YX_2	X_1X_2
0	4	1	0	16	1	0	0	4
1	3	2	1	9	4	3	2	6
5	9	0	25	81	0	45	0	0
6	8	2	36	64	4	48	12	16
8	6	5	64	36	25	48	40	30
20	30	10	126	206	34	144	54	56
↑ $\sum Y$	↑ $\sum X_1$	↑ $\sum X_2$	↑ $\sum Y^2$	↑ $\sum X_1^2$	↑ $\sum X_2^2$	↑ $\sum YX_1$	↑ $\sum YX_2$	↑ $\sum X_1X_2$

步骤二：

$$S_{YY} = \sum Y^2 - \frac{(\sum Y)^2}{n} = 126 - \frac{(20)^2}{5} = 46$$

$$S_{11} = \sum X_1^2 - \frac{(\sum X_1)^2}{n} = 206 - \frac{(30)^2}{5} = 26$$

$$S_{22} = \sum X_2^2 - \frac{(\sum X_2)^2}{n} = 34 - \frac{(10)^2}{5} = 14$$

$$S_{Y1} = \sum YX_1 - \frac{(\sum Y)(\sum X_1)}{n} = 144 - \frac{20 \times 30}{5} = 24$$

$$S_{Y2} = \sum YX_2 - \frac{(\sum Y)(\sum X_2)}{n} = 54 - \frac{20 \times 10}{5} = 14$$

$$S_{12}=\sum X_1X_2-\frac{(\sum X_1)(\sum X_2)}{n}=56-\frac{30\times10}{5}=-4$$

步骤三：

$$D_0=S_{11}S_{22}-S_{12}^2=26\times14-(-4)^2=348$$

$$D_1=S_{Y1}S_{22}-S_{Y2}S_{12}=24\times14-14\times(-4)=392$$

$$D_2=S_{Y2}S_{11}-S_{Y1}S_{12}=14\times26-24\times(-4)=460$$

步骤四：

$$\hat{\beta}_1=\frac{D_1}{D_0}=\frac{392}{348}=1.13$$

$$\hat{\beta}_2=\frac{D_2}{D_0}=\frac{460}{348}=1.32$$

$$\hat{\alpha}=\frac{\sum Y}{n}-\hat{\beta}_1\frac{\sum X_1}{n}-\hat{\beta}_2\frac{\sum X_2}{n}=\frac{20}{5}-\frac{392}{348}\times\frac{30}{5}-\frac{460}{348}\times\frac{10}{5}$$

$$=-5.40$$

据此，整理估计结果，得

$$\hat{Y}=-5.40+1.13X_1+1.32X_2$$

2. 决定系数与多元相关系数

与上一章介绍一元回归模型决定系数的思想方法相同，多元回归的决定系数是表示多元回归模型拟合优度的一种尺度。多元回归（有两个解释变量的情形）决定系数 R^2 的定义为

$$R^2=\frac{\text{能够由回归解释的变化（回归平方和）}}{Y\text{ 的全部变化}} \tag{4—17}$$

$$=\frac{\sum(\hat{Y}-\overline{Y})}{\sum(Y-\overline{Y})} \tag{4—18}$$

$$=\frac{\hat{\beta}_1S_{Y1}+\hat{\beta}_2S_{Y2}}{S_{YY}}\qquad\text{[计算式]} \tag{4—19}$$

取值范围为

$$0 \leqslant R^2 \leqslant 1$$

另外，多元相关系数（multiple correlation coefficient）主要用来测量被解释变量与两个以上的解释变量之间相关关系强弱的指标。多元相关系数是决定系数的非负的平方根，公式如下：

$$多元相关系数（R）=\sqrt{多元回归的决定系数} \tag{4—20}$$

3. 自由度调整后的决定系数

上面介绍的多元回归决定系数有一个缺陷，即随着解释变量数量的增加，系数会自动变大。这是因为即使没有解释力的解释变量，随着数量的增加，残差也会变小。为了克服这一缺陷，人们想出了一个自由度调整后的决定系数（coefficient of determination adjusted for the degree of freedom)。自由度调整后的决定系数，又称自由度修正后的决定系数，其定义如下：

$$\bar{R}^2 = 1 - \frac{n-1}{n-k-1}(1-R^2) \tag{4—21}$$

式中，R^2 为决定系数；

n 为样本数；

k 为解释变量个数。

多元回归中一般采用 $\bar{R}^2$，而不是 R^2。而且，$\bar{R}^2$ 通常小于 R^2，有时甚至可以取负值。$\bar{R}^2$ 的取值范围为

$$-\infty \leqslant \bar{R}^2 \leqslant 1$$

自由度调整后的相关系数为

$$\bar{R} = \sqrt{自由度调整后决定系数}$$

$$= \sqrt{\bar{R}^2} \tag{4—22}$$

［例题 4—2］

根据例题 4—1 的数据回答以下问题：

(1) 计算决定系数 R^2。

(2) 计算自由度调整后的决定系数 $\bar{R}^2$。

[解答]

(1) 根据 (4—19) 式，计算决定系数 R^2。

$$R^2=\frac{\hat{\beta}_1 S_{Y1}+\hat{\beta}_2 S_{Y2}}{S_{YY}}=\frac{\frac{392}{348}\times 24+\frac{460}{348}\times 14}{46}$$

$$=0.990$$

可见，模型的拟合优度极其良好。

(2) 样本数 n 为 5，解释变量数 k 为 2，因此，可以根据 (4—21) 式，计算自由度调整的决定系数 $\bar{R}^2$。

$$\bar{R}^2=1-\frac{n-1}{n-k-1}(1-R^2)=1-\frac{5-1}{5-2-1}\times(1-0.990)$$

$$=0.980$$

可见，$\bar{R}^2$ 比决定系数 $R^2(0.990)$ 稍小。

[例题 4—3]

计算下面三个自由度调整后的决定系数。这里，R^2 为决定系数，n 为样本数，k 为解释变量个数。

(1) $R^2=0.75$　$n=8$　$k=2$

(2) $R^2=0.35$　$n=9$　$k=3$

(3) $R^2=0.95$　$n=31$　$k=5$

[解答]

根据 (4—21) 式，计算自由度调整决定系数 $\bar{R}^2$。

(1) $$\bar{R}^2=1-\frac{n-1}{n-k-1}(1-R^2)=1-\frac{8-1}{8-2-1}\times(1-0.75)$$

$$=0.65$$

(2) $$\bar{R}^2=1-\frac{9-1}{9-3-1}\times(1-0.30)$$

$$=-0.12$$

这样，$\bar{R}^2$ 为负值的情况也是存在的。

(3) $$\bar{R}^2=1-\frac{31-1}{31-5-1}\times(1-0.95)$$

$$=0.94$$

[例题 4—4]

表 4—2 是以进出车站的乘客为主要服务对象的 10 家便利店的数据。

Y 是日均销售额，X_1 为店铺面积，X_2 是作为选址条件的店铺距车站的距离。

(1) 对多元回归模型 $Y=\alpha+\beta_1X_1+\beta_2X_2+u$ 进行 OLS 估计。

(2) 求决定系数 R^2 和自由度调整后的决定系数 $\bar{R}^2$。

(3) 假设其他条件不变，店铺面积增加 1 平方米，日均销售额能增加多少?

(4) 假设其他条件不变，店铺离车站比现在远 100 米，日均销售额会减少多少?

(5) 假设有人想新建一个店铺 K 店，计划店铺面积为 80 平方米，距车站 300 米，试预测其日均销售额 $\hat{Y}_K$。

表 4—2　　10 家便利店日均销售额、店铺面积以及与车站的距离

店铺	日均销售额（万日元）Y	店铺面积（平方米）X_1	离车站的距离（100 米）X_2
A	40	60	3
B	45	100	5
C	80	85	2
D	60	50	1
E	50	75	3
F	20	55	4
G	15	70	6
H	90	95	1
I	30	45	3
J	70	65	2

[解答]

(1) 步骤一：编制工作表，见表 4—3。

表 4—3

Y	X_1	X_2	Y^2	X_1^2	X_2^2	YX_1	YX_2	X_1X_2
40	60	3	1 600	3 600	9	2 400	120	180
45	100	5	2 025	10 000	25	4 500	225	500
80	85	2	6 400	7 225	4	6 800	160	170
60	50	1	3 600	2 500	1	3 000	60	50
50	75	3	2 500	5 625	9	3 750	150	225
20	55	4	400	3 025	16	1 100	80	220
15	70	6	225	4 900	36	1 050	90	420
90	95	1	8 100	9 025	1	8 550	90	95
30	45	3	900	2 025	9	1 350	90	135
70	65	2	4 900	4 225	4	4 550	140	130
500	700	30	30 650	52 150	114	37 050	1 205	2 125
$\uparrow$ $\sum Y$	$\uparrow$ $\sum X_1$	$\uparrow$ $\sum X_2$	$\uparrow$ $\sum Y^2$	$\uparrow$ $\sum X_1^2$	$\uparrow$ $\sum X_2^2$	$\uparrow$ $\sum YX_1$	$\uparrow$ $\sum YX_2$	$\uparrow$ $\sum X_1X_2$

步骤二：

$$S_{YY} = \sum Y^2 - \frac{(\sum Y)^2}{n} = 30\ 650 - \frac{(500)^2}{10} = 5\ 650$$

$$S_{11} = \sum X_1^2 - \frac{(\sum X_1)^2}{n} = 52\ 150 - \frac{(700)^2}{10} = 3\ 150$$

$$S_{22} = \sum X_2^2 - \frac{(\sum X_2)^2}{n} = 114 - \frac{(30)^2}{10} = 24$$

$$S_{Y1} = \sum YX_1 - \frac{(\sum Y)(\sum X_1)}{n} = 37\ 050 - \frac{500 \times 700}{10} = 2\ 050$$

$$S_{Y2} = \sum YX_2 - \frac{(\sum Y)(\sum X_2)}{n} = 1\ 205 - \frac{500 \times 30}{10} = -295$$

$$S_{12} = \sum X_1X_2 - \frac{(\sum X_1)(\sum X_2)}{n} = 2\ 125 - \frac{700 \times 30}{10} = 25$$

步骤三：

$$D_0 = S_{11}S_{22} - S_{12}^2 = 3\ 150 \times 24 - (25)^2 = 74\ 975$$

$$D_1 = S_{Y1}S_{22} - S_{Y2}S_{12} = 2\ 050 \times 24 - (-295) \times 25 = 56\ 575$$

$$D_2 = S_{Y2}S_{11} - S_{Y1}S_{12} = (-295) \times 3\ 150 - 2\ 050 \times 25 = -980\ 500$$

步骤四：

$$\hat{\beta}_1 = \frac{D_1}{D_0} = \frac{56\ 575}{74\ 975} = 0.754\ 58$$

$$\hat{\beta}_2 = \frac{D_2}{D_0} = \frac{-980\ 500}{74\ 975} = -13.078$$

$$\begin{aligned}\hat{\alpha} &= \frac{\sum Y}{n} - \hat{\beta}_1 \frac{\sum X_1}{n} - \hat{\beta}_2 \frac{\sum X_2}{n} \\ &= \frac{500}{10} - \frac{56\ 575}{74\ 975} \times \frac{700}{10} - \frac{-980\ 500}{74\ 975} \times \frac{30}{10} \\ &= 36.412\end{aligned}$$

据此，将估计结果整理，得

$$\hat{Y} = 36.412 + 0.754\ 58X_1 - 13.078X_2$$

(2) 根据（4—19）式，计算决定系数 R^2，得

$$\begin{aligned}R^2 &= \frac{\hat{\beta}_1 S_{Y1} + \hat{\beta}_2 S_{Y2}}{S_{YY}} \\ &= \frac{\frac{56\ 575}{74\ 975} \times 2\ 050 + \frac{-980\ 500}{74\ 975} \times (-295)}{5\ 650} \\ &= 0.956\ 6\end{aligned}$$

根据（4—21）式，计算自由度调整决定系数 $\bar{R}^2$。

$$\begin{aligned}\bar{R}^2 &= 1 - \frac{n-1}{n-k-1}(1-R^2) \\ &= 1 - \frac{10-1}{10-2-1}(1-0.956\ 6) \\ &= 0.944\ 2\end{aligned}$$

可见，$\bar{R}^2$ 稍小于前面求出的决定系数 $R^2(0.956\ 6)$。

(3) 假设其他条件不变，由 $\hat{\beta}_1$ 等于 0.754 58 可以看出，店铺面积扩大 1 平方米，日均销售额大约可以增加7 546日元。

(4) 假设其他条件不变，由 $\hat{\beta}_2$ 等于 −13.078 可以看出，店铺如果比现在地址离车站远 100 米，日均销售额大约会减少 13.078 万日元。

(5) 由于在 K 店情形中，$X_1=80$，$X_2=3$，预计其日均销售额为，

$$
\begin{aligned}
\hat{Y}_K &= 36.412 + 0.754\ 58X_1 - 13.078X_2 \\
&= 36.412 + 0.754\ 58 \times 80 - 13.078 \times 3 \\
&= 57.544\ 4 \text{（万日元）}
\end{aligned}
$$

4. 偏相关系数

在 Y、X_1、X_2 三个变量中，当 X_1 既定时（即不受 X_1 的影响），表示 Y 与 X_2 之间相关关系的指标，称为偏相关系数（partial correlation coefficient），记做 $R_{Y2\cdot1}$。$R_{Y2\cdot1}$可以通过下面的两个变量之间的相关系数（单纯相关系数）R_{Y1}、R_{Y2}、R_{12}很容易地计算出来。

此外，当 X_2 既定时，Y 与 X_1 的偏相关系数 $R_{Y1\cdot2}$也可以用同样的方法计算出来。

$$R_{Y2\cdot1} = \frac{R_{Y2} - R_{Y1}R_{12}}{\sqrt{(1-R_{Y1}^2)(1-R_{12}^2)}} \tag{4—23}$$

$$R_{Y1\cdot2} = \frac{R_{Y1} - R_{Y2}R_{12}}{\sqrt{(1-R_{Y2}^2)(1-R_{12}^2)}} \tag{4—24}$$

$R_{Y2\cdot1}$、$R_{Y1\cdot2}$的取值范围为

$$-1 \leqslant R_{Y2\cdot1} \leqslant 1$$

$$-1 \leqslant R_{Y1\cdot2} \leqslant 1$$

作为参考，Y、X_1、X_2、X_3 这四个变量之间的偏相关系数 $R_{Y1\cdot23}$可以根据以下的公式计算：

$$R_{Y1\cdot23} = \frac{R_{Y1\cdot2} - R_{Y3\cdot2}R_{13\cdot2}}{\sqrt{(1-R_{Y3\cdot2}^2)(1-R_{13\cdot2}^2)}} \tag{4—25}$$

$$= \frac{R_{Y1\cdot3} - R_{Y2\cdot3}R_{12\cdot3}}{\sqrt{(1-R^2_{Y2\cdot3})(1-R^2_{12\cdot3})}} \tag{4—26}$$

[例题 4—5]

下面是家庭消费 Y、收入 X_1 和资产 X_2 的相关分析结果，即单纯相关系数。

$$R_{Y1}=0.97 \quad R_{Y2}=0.79 \quad R_{12}=0.72$$

(1) 若收入 X_1 既定，求 Y 与 X_2 的偏相关系数 $R_{Y2\cdot1}$。

(2) 若资产 X_2 既定，求 Y 与 X_1 的偏相关系数 $R_{Y1\cdot2}$。

[解答]

(1) 根据 (4—23) 式，计算偏相关系数 $R_{Y2\cdot1}$。

$$\begin{aligned} R_{Y2\cdot1} &= \frac{R_{Y2}-R_{Y1}R_{12}}{\sqrt{(1-R^2_{Y1})(1-R^2_{12})}} \\ &= \frac{0.79-0.97\times0.72}{\sqrt{[1-(0.97)^2][1-(0.72)^2]}} \\ &= 0.543 \end{aligned}$$

(2) 根据 (4—24) 式，计算偏相关系数 $R_{Y1\cdot2}$。

$$\begin{aligned} R_{Y1\cdot2} &= \frac{R_{Y1}-R_{Y2}R_{12}}{\sqrt{(1-R^2_{Y2})(1-R^2_{12})}} \\ &= \frac{0.97-0.79\times0.72}{\sqrt{[1-(0.79)^2][1-(0.72)^2]}} \\ &= 0.943 \end{aligned}$$

[例题 4—6]

表 4—4 中的数据反映的是 1992 年亚洲各国人均寿命 Y、按购买力平价计算的人均 GDPX_1、成人识字率 X_2 和一岁儿童疫苗接种率 X_3。

(1) 对多元回归模型 $Y=\alpha+\beta_1X_1+\beta_2X_2+u$ 进行 OLS 估计。

(2) 计算 (1) 中的决定系数 R^2 与自由度调整后的决定系数 $\bar{R}^2$。

(3) 求 X_1 既定时 Y 与 X_2 的偏相关系数 $R_{Y2\cdot 1}$，以及 X_2 既定时 Y 与 X_1 的偏相关系数 $R_{Y1\cdot 2}$。

(4) 利用计量分析统计软件（如 Excel、TSP 等），对下面的多元回归模型进行 OLS 估计，并计算决定系数 R^2 和自由度调整的决定系数 $\bar{R}^2$。

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + u$$

表 4—4　　亚洲各国（地区）人的发展指标（1992 年）

国家和地区	平价寿命 Y（年）	按购买力平价计算的人均 GDP X_1(100 美元)	成人识字率 X_2(%)	一岁儿童疫苗接种率 X_3(%)
1. 日本	79	194	99	99
2. 中国香港	77	185	90	79
3. 韩国	70	83	97	83
4. 新加坡	74	147	92	90
5. 泰国	69	53	94	86
6. 马来西亚	70	74	80	90
7. 斯里兰卡	71	27	89	88
8. 中国大陆	70	29	80	94
9. 菲律宾	65	24	90	92
10. 朝鲜	71	18	95	96
11. 蒙古	63	23	95	85
12. 印度尼西亚	62	27	84	92
13. 越南	63	13	89	90
14. 缅甸	57	7	81	74
15. 巴基斯坦	58	20	36	81
16. 老挝	50	18	55	36
17. 印度	60	12	50	90
18. 孟加拉国	52	12	37	69
19. 柬埔寨	50	13	38	37
20. 尼泊尔	53	11	27	73
21. 不丹	48	6	41	85
22. 阿富汗	43	7	32	35

资料来源：联合国发展规划署《人的发展报告》。

(1) 步骤一：省略工作表。

$$\sum Y = 1\,375 \qquad \sum X_1 = 1\,003 \qquad \sum X_2 = 1\,571$$

$$\sum Y^2 = 88\,075 \qquad \sum X_1^2 = 113\,977 \qquad \sum X_2^2 = 126\,087$$

$$\sum YX_1 = 71\,448 \qquad \sum YX_2 = 102\,803$$

$$\sum X_1X_2 = 87\,118$$

步骤二：

$$S_{YY} = \sum Y^2 - \frac{(\sum Y)^2}{n} = 2\,137.500$$

$$S_{11} = \sum X_1^2 - \frac{(\sum X_1)^2}{n} = 68\,249.32$$

$$S_{22} = \sum X_2^2 - \frac{(\sum X_2)^2}{n} = 13\,903.32$$

$$S_{Y1} = \sum YX_1 - \frac{(\sum Y)(\sum X_1)}{n} = 8\,760.500$$

$$S_{Y2} = \sum YX_2 - \frac{(\sum Y)(\sum X_2)}{n} = 4\,615.500$$

$$S_{12} = \sum X_1X_2 - \frac{(\sum X_1)(\sum X_2)}{n} = 15\,494.68$$

步骤三：

$$D_0 = S_{11}S_{22} - S_{12}^2 = 708\,807\,027$$

$$D_1 = S_{Y1}S_{22} - S_{Y2}S_{12} = 50\,284\,339$$

$$D_2 = S_{Y2}S_{11} - S_{Y1}S_{12} = 179\,263\,592$$

步骤四：

$$\hat{\beta}_1 = \frac{D_1}{D_0} = 0.070\,942$$

$$\hat{\beta}_2 = \frac{D_2}{D_0} = 0.252\,91$$

$$\hat{\alpha} = \frac{\sum Y}{n} - \hat{\beta}_1 \frac{\sum X_1}{n} - \hat{\beta}_2 \frac{\sum X_2}{n} = 41.206$$

整理估计结果得

$$\hat{Y} = 41.206 + 0.070\,942 X_1 + 0.252\,91 X_2$$

(2) 根据 (4—19) 式，计算决定系数 R^2。

$$R^2 = \frac{\hat{\beta}_1 S_{Y1} + \hat{\beta}_2 S_{Y2}}{S_{YY}} = \frac{0.070\,942 \times 8\,760.500 + 0.252\,91 \times 4\,615.500}{2\,137.500}$$

$$\approx 0.836\,9$$

Y 的变化中，大约有 84％可以由模型来解释。

由于样本数 n 为 22，解释变量数 k 为 2，根据 (4—21) 式，计算自由度调整后的决定系数 $\bar{R}^2$。

$$\bar{R}^2 = 1 - \frac{n-1}{n-k-1}(1 - R^2) = 1 - \frac{22-1}{22-2-1} \times (1 - 0.836\,9)$$

$$\approx 0.819\,7$$

(3) 计算两个变量间的单纯相关系数。

$$R_{Y1} = \frac{\sum (Y - \bar{Y})(X_1 - \bar{X}_1)}{\sqrt{\sum (Y - \bar{Y})^2 \sum (X_1 - \bar{X}_1)^2}} = \frac{S_{Y1}}{\sqrt{S_{YY} S_{11}}} = 0.725\,32$$

$$R_{Y2} = \frac{\sum (Y - \bar{Y})(X_2 - \bar{X}_2)}{\sqrt{\sum (Y - \bar{Y})^2 \sum (X_2 - \bar{X}_2)^2}} = \frac{S_{Y2}}{\sqrt{S_{YY} S_{22}}} = 0.846\,65$$

$$R_{12} = \frac{\sum (X_1 - \bar{X}_1)(X_2 - \bar{X}_2)}{\sqrt{\sum (X_1 - \bar{X}_1)^2 \sum (X_2 - \bar{X}_2)^2}} = \frac{S_{12}}{\sqrt{S_{11} S_{22}}} = 0.503\,01$$

当 X_1 既定时，Y 与 X_2 的偏相关系数 $R_{Y2\cdot 1}$，可由 (4—23) 式计算得

$$R_{Y2\cdot 1} = \frac{R_{Y2} - R_{Y1} R_{12}}{\sqrt{(1 - R_{Y1}^2)(1 - R_{12}^2)}} = \frac{0.846\,65 - 0.725\,32 \times 0.503\,01}{\sqrt{[1 - (0.725\,32)^2][1 - (0.503\,01)^2]}}$$

$$\approx 0.809\,8$$

Y 与 X_2 的偏相关系数 $R_{Y2\cdot 1}$，由于剔除了 X_1 的影响，小于单纯相关系数 $R_{Y2}(0.846\,65)$。

当 X_2 既定时，Y 与 X_1 的偏相关系数 $R_{Y1\cdot2}$，可由（4—24）式计算得

$$R_{Y1\cdot2}=\frac{R_{Y1}-R_{Y2}R_{12}}{\sqrt{(1-R_{Y2}^2)(1-R_{12}^2)}}=\frac{0.725\,32-0.846\,65\times0.503\,01}{\sqrt{[1-(0.846\,65)^2][1-(0.503\,01)^2]}}$$

$$=0.651\,1$$

同样，Y 与 X_1 的偏相关系数 $R_{Y1\cdot2}$小于单纯相关系数 $R_{Y1}(0.725\,32)$。

（4）将这一多元回归模型的估计结果进行整理，得到下面的公式。Y 的变化中，大约有 91%可以由模型来解释。

$$Y=32.993+0.071\,619X_1+0.168\,73X_2+0.179\,04X_3$$

$$R^2=0.906\,5 \quad \bar{R}^2=0.891\,0$$

[例题 4—7]

表 4—5 列示了日本小企业（10 人～99 人）中每月现金工资 Y 与劳动者年龄 X 之间的关系。

（1）设横轴为 X，纵轴为 Y，画出数据的散点图。

（2）对多元回归模型 $Y=\alpha+\beta_1X+\beta_2X^2+u$ 进行估计，并计算决定系数 R^2 与自由度调整后的决定系数 $\bar{R}^2$。

表 4—5　　1995 年日本小企业每月现金工资与劳动者年龄的关系

每月现金工资额（1 000 日元） Y	劳动者年龄（岁） X
141.4	16.8
170.6	19.1
202.2	22.6
246.6	27.4
288.1	32.4
306.0	37.5
311.6	42.6
318.6	47.3
316.5	52.5
304.0	57.4
267.2	62.2
237.1	68.7

资料来源：劳动省《工资结构基本统计调查》。

[解答]

(1) 散点图见图 4—2。

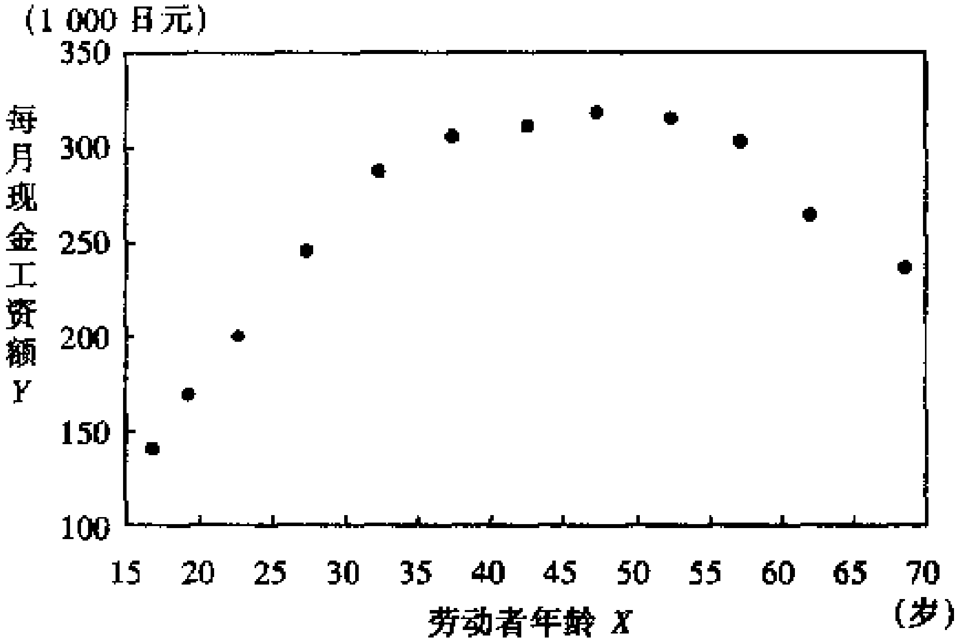


图 4—2 日本小企业每月现金工资与劳动者年龄的关系 (1995 年)

(2) 由于该模型是非线性的 (2 次式), 设 $Z=X^2$, 进行线性变换 (参照第三章表 3—9)。

$$Y=-107.46+18.198X-0.193\,17\underbrace{X^2}_{(=Z)}$$

$$R^2=0.993\,0 \quad \bar{R}^2=0.991\,4$$

可见, 模型的拟合情况极其良好。

[补充 1] OLS 中误差项 u_t 的假定

(1) $E(u_t)=0$

u_t 的期望值为 0 (u_t 的均值为 0)。

(2) $E(u_t^2)=\sigma^2$

u_t 的方差在任何时点均为既定的。这一假设称为同方差 (homoskedasticity)。

(3) $E(u_t u_s)=0 \quad (t\neq s)$

u_t 与 u_s 相互之间不相关, 即假设不存在系列相关 (自我相关)。

(4) $E(X_t u_t)=0$

解释变量 X_t 与 u_t 不相关。

(5) $u_t \sim N\ (0, \sigma^2)$

u_t 服从均值为 0，方差为 σ^2 的正态分布。

如果（1）项～(4）项的假定成立，由 OLS 得出的估计量就是最优线性无偏估计量（best linear unbiased estimator，BLUE)，又叫做 Gauss-Markov 定理。为了方便起见，我们可以将最优线性无偏估计理解为所有线性无偏估计（估计量的平均值等于实际参数）中，方差最小的估计。因此，能够满足（1）项～(4）项假定的最小二乘估计所得到的参数，其精确度非常高。而且，本书中假定，误差项根据中心极限定理，服从第(5)项的正态分布。

不过，实际经济现象并不完全能够满足（1）项～(5）项的假定，相反，这些假定大多不能成立。因此，计量经济学在这些假定不能满足的时候，会出现什么样的问题，以及如何解决这些问题，是非常重要的研究课题。

[补充 2] 异方差性

当回归模型误差项第（2）项假定不成立时，称做误差项处于异方差状态(heteroscadasticity)。例如，随着解释变量的值增大，误差项的离散现象也增大，这是异方差性的一个典型。有异方差性的 OLS 估计，所得到的估计值就不是 BULE。通常，利用横截面数据比利用时序列数据，更容易出现异方差问题。

异方差性的解决方法：

(1) 对解释变量与被解释变量进行对数变换。

(2) 用加权最小二乘法（weighted least squares，WLS)，或者最优法模型进行估计。

[补充 3] 多重共线性

在多元回归分析中，如果解释变量之间关系很密切，就会出现所谓的多重共线性（multicollinearrity）这种棘手的问题。多重共线性表现为：(1）尽管决定系数很高，但是 t 值（参照第五章）较低；(2）存在估计出的回归系数的符号（正、负）与理论不一致等问题，从而降低了多元回归模型估计结果的可靠性。

能够完全解决多重共线性问题的方法目前还没有，下面是常用的六种解决方法。

(1) 将相关性较高的解释变量消除几个。

(2) 如果数据很多的话，可以延长计算期间对回归模型进行估计。

(3) 如果年度数据不够理想的话，可以利用季度数据、月度数据对回归模型进行估计。

(4) 对解释变量和被解释变量用阶差、比率等形式进行变换，重新构建多元回归模型，进行估计。

(5) 试用岭回归分析 (ridge regression)。

(6) 根据解释变量的主成分分析 (principal component analysis)，做出相互之间没有关系的合成变量，以此为新的解释变量构建新的模型，进行估计。

(具体方法参见相应的计量经济学资料。——译者注)

第四章 练习题

1. 表4—6列示了日本家庭年消费支出 Y、年收入 X_1 以及家庭人数 X_2(年收入十等分数据)之间的关系。

(1) 对多元回归模型 $Y=\alpha+\beta_1X_1+\beta_2X_2+u$ 进行 OLS 估计。

(2) 求决定系数 R^2 与自由度调整后的决定系数 $\bar{R}^2$。

(3) 计算当 X_1 既定时,Y 与 X_2 的偏相关系数 $R_{Y2\cdot1}$。

(4) 计算当 X_2 既定时,Y 与 X_1 的偏相关系数 $R_{Y1\cdot2}$。

表4—6 日本家庭的年消费支出、年收入以及家庭人数(1995年) 单位:万日元;人

年消费支出 Y	年收入 X_1	家庭人数 X_2
225	243	2.66
279	363	2.83
313	449	3.14
332	526	3.30
366	609	3.55
403	696	3.58
430	795	3.67
466	921	3.73
509	1 101	3.84
624	1 757	3.89

资料来源:总务厅统计局《家庭调查年报》。

2. 表4—7列示了日本家庭户主的年均零用钱 Y 与年龄 X 之间的关系。

(1) 设横轴为 X,纵轴为 Y,画出数据的散点图。

(2) 对多元回归模型 $Y=\alpha+\beta_1X+\beta_2X^2+u$ 进行 OLS 估计。

(3) 求决定系数 R^2 与自由度调整后的决定系数 $\bar{R}^2$。

表 4—7　　日本家庭户主的年均零用钱与年龄之间的关系（1995 年）

单位：1 000 日元；岁

年均零用钱 Y	年龄 X
195	22.8
221	27.5
269	32.1
262	37.1
288	42.1
304	46.9
300	52.1
235	57.0
164	61.9
90	71.1

资料来源：总务厅统计局《家庭调查年报》。

3*. 根据例题 4—4，有如下的利用车站者人数 X_3(万人)：

A——8　B——10　C——17　D——11　E——7

F——6　G——14　H——5　I——9　J——23

利用计量分析统计软件，进行以下多元回归分析。

(1) 对多元回归模型 $Y=\alpha+\beta_1X_1+\beta_2X_2+\beta_3X_3+u$，进行 OLS 估计。

(2) 计算决定系数 R^2 与自由度调整后的决定系数 $\bar{R}^2$。

(3) 假设其他条件既定，如果利用车站的人数 X_3 增加 10 000，销售额能增加多少？

(4) 计算当 X_2、X_3 既定时，Y 与 X_1 的偏相关系数 $R_{Y1\cdot23}$。

(5) 计算当 X_1、X_3 既定时，Y 与 X_2 的偏相关系数 $R_{Y2\cdot13}$。

(6) 计算当 X_1、X_2 既定时，Y 与 X_3 的偏相关系数 $R_{Y3\cdot12}$。

4*. 表 4—8 反映的是某个产业的附加价值生产额 Y、劳动者人数 L 以及资本额 K 的历史变化（实际值），t 为趋势变量。设在计算期间中，人均劳动时间没有变化。

(1) 将下面的 cobb-douglas（C－D 函数）作对数变换，然后进行 OLS 估计，并计算自由度调整后的决定系数 $\bar{R}^2$。

(2) 根据 (1) 的估计结果，分析该产业的技术进步率应为多少。

(3) 对 (1) 中的 C－D 函数作一次同次假设，同样作对数变换，然后进行 OLS 估计，并计算 $\bar{R}^2$。

(4) 根据 (3) 的结果，分析该产业的技术进步率应为多少。

表 4—8　　　　某产业的附加价值生产额、劳动者人数以及资本额的变化

年　份	附加价值生产额 (10亿日元) Y	劳动者人数 (千人) L	资本额 (10亿日元) K	趋势变量 t
1986	242	36	60	1
1987	262	38	65	2
1988	299	38	74	3
1989	328	44	72	4
1990	336	41	78	5
1991	365	43	80	6
1992	378	40	83	7
1993	402	40	89	8
1994	409	38	91	9
1995	432	39	93	10

说明：1990 年价格，实际值。

第五章　回归模型的假设检验

本章介绍 t 值与 F 值，它们是用来检验估计出来的回归模型可靠性的统计量。此外，对于结构变化的 F 检验（Chow 检验）与基于回归模型的预测方法，也有简要说明。

1. t 值

t 值（t value）是用来检验根据 OLS 估计出来的回归系数是否显著的统计量。回归系数在统计学上如果被判断不为零，就是“显著的”。如果回归系数是不显著的（回归系数 $=0$），则意味着解释变量对被解释变量没有任何影响，该解释变量在模型中没有存在的必要。由此可见，t 值还具有选择解释变量的作用。

顺便提一下，关于常数项的 t 值，除非在经济理论上具有重要意义或者在进行经济预测时，一般地，即使不显

著，也没有必要在意。

下面，我们分别就一元回归模型和多元回归模型两种不同的情况，说明回归系数显著性的检验方法。

（1）一元回归模型的情形。

模型：$Y = \alpha + \beta X + u$ 设由 OLS 估计出的 α、β 分别为 $\hat{\alpha}$、$\hat{\beta}$。

步骤一：估计残差方差 s^2(残差的无偏方差)。

$$s^2 = \frac{\text{残差平方和}}{\text{样本数 - 解释变量数 - 1}} = \frac{\sum \hat{u}^2}{n-2} \tag{5—1}$$

式中，$\hat{u} = Y - \hat{Y}$。

u 上加上ˆ（读做 hat），表示样本的残差。此外，s^2 的正平方根 s，称做回归方程的标准误差。

步骤二：估计 $\hat{\alpha}$ 和 $\hat{\beta}$ 的方差。

$$s_{\hat{\alpha}}^2 = s^2\left[\frac{1}{n} + \frac{\bar{X}}{\sum(X-\bar{X})^2}\right] \tag{5—2}$$

$$= \frac{s^2\sum X^2}{n\sum(X-\bar{X})} \tag{5—3}$$

$$= \frac{s^2\sum X^2}{n\sum X^2 - (\sum X)^2} \quad [\text{计算式}] \tag{5—4}$$

$$s_{\hat{\beta}}^2 = \frac{s^2}{\sum(X-\bar{X})^2} \tag{5—5}$$

$$= \frac{s^2 n}{n\sum X^2 - (\sum X)^2} \quad [\text{计算式}] \tag{5—6}$$

计算式（5—4）和（5—6）的分母相同。$\hat{\alpha}$ 与 $\hat{\beta}$ 是从总体中多次抽取样本数据，并根据 OLS 估计回归系数情形下（实际上只估计了一次）的分布。$s_{\hat{\alpha}}^2$ 与 $s_{\hat{\beta}}^2$ 表示的是相应的离散程度。

步骤三：计算回归系数的标准误差。

$$s_{\hat{\alpha}} = \sqrt{s_{\hat{\alpha}}^2} \tag{5—7}$$

$$s_{\hat{\beta}} = \sqrt{s_{\hat{\beta}}^2} \tag{5—8}$$

现在假设 α、β 为真正的回归系数，它们与估计的回归系数 $\hat{\alpha}$、$\hat{\beta}$ 之间的误差（估计误差），即 $\hat{\alpha}-\alpha$、$\hat{\beta}-\beta$ 超过 $2s_{\hat{\alpha}}$、$2s_{\hat{\beta}}$ 的概率在5%以下，超过 $3s_{\hat{\alpha}}$、$3s_{\hat{\beta}}$ 的可能性非常小。在计量经济分析中，这种统计误差通常略小于估计的回归系数。

步骤四：计算 t 值。

$$t_{\hat{\alpha}}=\frac{\text{回归系数的估计值}}{\text{回归系数的标准误差}}=\frac{\hat{\alpha}}{s_{\hat{\alpha}}} \tag{5—9}$$

$$t_{\hat{\beta}}=\frac{\text{回归系数的估计值}}{\text{回归系数的标准误差}}=\frac{\hat{\beta}}{s_{\hat{\beta}}} \tag{5—10}$$

步骤五：对估计出来的回归系数 $\hat{\alpha}$、$\hat{\beta}$ 进行显著性检验（t 检验）。

t 检验有双侧检验（two-tailed test）与单侧检验（one-tailed test）两种，这里，我们仅说明双侧检验。

首先，建立原假设（null hypothesis）与备择假设（alternative hypothesis）。

$$\text{原假设}\begin{cases}H_0:\alpha=0\\ H_0':\beta=0\end{cases}$$

$$\text{备择假设}\begin{cases}H_1:\alpha\neq 0\\ H_1':\beta\neq 0\end{cases}$$

计量经济分析中通常希望通过放弃原假设、支持备择假设来进行假设检验。

由步骤四计算出来的 t 值服从自由度 $n-2$（自由度＝样本个数－解释变量个数－1），因此，可以根据表5—1的 t 分布表，进行显著性检验。具体说来，如果计算出来的 t 值的绝对值 $|t_{\hat{\alpha}}|$、$|t_{\hat{\beta}}|$ 比 t 分布表中找到的 t 值（判定值）大，则放弃原假设，估计的回归系数显著。

表5—1　　　　t 分布表

自由度	双侧检验		单侧检验	
	显著水平 5%	显著水平 1%	显著水平 5%	显著水平 1%
1	12.706	63.657	6.314	31.821
2	4.303	9.925	2.920	6.965
3	3.182	5.841	2.353	4.541
4	2.776	4.604	2.132	3.747
5	2.571	4.032	2.015	3.365
6	2.447	3.707	1.943	3.143

续前表

自由度	双侧检验		单侧检验	
	显著水平 5%	显著水平 1%	显著水平 5%	显著水平 1%
7	2.365	3.499	1.895	2.998
8	2.306	3.355	1.860	2.896
9	2.262	3.250	1.833	2.821
10	2.228	3.169	1.812	2.764
11	2.201	3.106	1.796	2.718
12	2.179	3.055	1.782	2.681
13	2.160	3.012	1.771	2.650
14	2.145	2.977	1.761	2.624
15	2.131	2.947	1.753	2.602
16	2.120	2.921	1.746	2.583
17	2.110	2.898	1.740	2.567
18	2.101	2.878	1.734	2.552
19	2.093	2.861	1.729	2.539
20	2.086	2.845	1.725	2.528
21	2.080	2.831	1.721	2.518
22	2.074	2.819	1.717	2.508
23	2.069	2.807	1.714	2.500
24	2.064	2.797	1.711	2.492
25	2.060	2.787	1.708	2.485
26	2.056	2.779	1.706	2.479
27	2.052	2.771	1.703	2.473
28	2.048	2.763	1.701	2.467
29	2.045	2.756	1.699	2.462
30	2.042	2.750	1.697	2.457
31	2.040	2.744	1.696	2.453
32	2.037	2.738	1.694	2.449
33	2.035	2.733	1.692	2.445
34	2.032	2.728	1.691	2.441
35	2.030	2.724	1.690	2.438
36	2.028	2.719	1.688	2.434
37	2.026	2.715	1.687	2.431

续前表

自由度	双侧检验		单侧检验	
	显著水平 5%	显著水平 1%	显著水平 5%	显著水平 1%
38	2.024	2.712	1.686	2.429
39	2.023	2.708	1.685	2.426
40	2.021	2.704	1.684	2.423
41	2.020	2.701	1.683	2.421
42	2.018	2.698	1.682	2.418
43	2.017	2.695	1.681	2.416
44	2.015	2.692	1.680	2.414
45	2.014	2.690	1.679	2.412
46	2.013	2.687	1.679	2.410
47	2.012	2.685	1.678	2.408
48	2.011	2.682	1.677	2.407
49	2.010	2.680	1.677	2.405
50	2.009	2.678	1.676	2.403
60	2.000	2.660	1.671	2.390
80	1.990	2.639	1.664	2.374
120	1.980	2.617	1.658	2.358
240	1.970	2.596	1.651	2.342
∞	1.960	2.576	1.645	2.326

说明：自由度＝样本数 n－解释变量数 k－1，这里是有常数项的回归分析。

这时，显著水平（即回归系数尽管确实为零，将其误判为非零的概率，或者说分析者出现错误判定的概率）虽然根据分析者自己的判断来决定，但是最常用的是5%的水平，其次是1%。当然，显著水平越小，检验越严格。

此外，样本个数 n 如果大到一定程度（$n \geqslant 30$），t 值只要大于2.0，计量经济学家们就习惯于将回归系数判定为显著。当待验回归系数非常多时，利用这种方法比较方便，不用特意去查 t 分布表。至于 t 值为什么在2.0以上就显著，是因为通常在利用5%的显著水平下（双侧检验），如果自由度在28以上（即一元回归的 $n \geqslant 30$），则将小数第二位四舍五人，全部等于2.0（即使当自由度＝∞时，$1.96 \approx 2.0$）。

但是，如果样本数很小，即使判定值在2.0以上，也不要使用这一规则。

[补充]单侧检验（one-tailed test）

众所周知，凯恩斯消费函数（$c=\alpha+\beta y$）中，边际消费倾向 β 事先就被假设为大于 0，像这种在理论模型中符号条件事先已经决定的情况下，一般运用以下的单侧检验。

原假设 $H_0: \beta=0$

备择假设 $H_1: \beta>0$

在双侧检验中，备择假设 H_1 为 $\beta\neq0$，而在单侧检验中，备择假设 H_1 为 $\beta>0$。同理，当符号条件为负时，备择假设 H_1 为 $\beta<0$。在进行计量经济分析时，理论模型中符号条件既定的情况很多，单侧检验经常会用到。

(2) 多元回归模型的情形（解释变量数 $k=2$）。

模型：$Y=\alpha+\beta_1X_1+\beta_2X_2+u$ 用 OLS 求 α、β_1、β_2 的估计值 $\hat{\alpha}$、$\hat{\beta}_1$、$\hat{\beta}_2$。

步骤一：估计残差的方差 s^2。

$$s^2=\frac{\text{残差平方和}}{\text{样本数}-\text{解释变量数}-1}=\frac{\sum\hat{u}^2}{n-3} \qquad (5\text{—}11)$$

步骤二：估计 $\hat{\alpha}$、$\hat{\beta}_1$、$\hat{\beta}_2$ 的方差。

$$s_{\hat{\alpha}}^2=s^2\left(\frac{1}{n}+\frac{\overline{X}_1^2S_{22}+\overline{X}_2^2S_{11}-2\overline{X}_1\overline{X}_2S_{12}}{S_{11}S_{22}-S_{12}^2}\right) \qquad (5\text{—}12)$$

$$s_{\hat{\beta}}^2=s^2\left(\frac{S_{22}}{S_{11}S_{22}-S_{12}^2}\right) \qquad (5\text{—}13)$$

$$s_{\hat{\beta}_2}^2=s^2\left(\frac{S_{11}}{S_{11}S_{22}-S_{12}^2}\right) \qquad (5\text{—}14)$$

式中，$$S_{11}=\sum(X_1-\overline{X}_1)^2=\sum X_1^2-\frac{(\sum X_1)^2}{n}$$

$$S_{22}=\sum(X_2-\overline{X}_2)^2=\sum X_2^2-\frac{(\sum X_2)^2}{n}$$

$$S_{12}=\sum(X_1-\overline{X}_1)(X_2-\overline{X}_2)=\sum X_1X_2-\frac{(\sum X_1)(\sum X_2)}{n}$$

步骤三：计算回归系数的标准误差。

$$s_{\hat{\alpha}}=\sqrt{s_{\hat{\alpha}}^{2}} \tag{5—15}$$

$$s_{\hat{\beta}_1}=\sqrt{s_{\hat{\beta}_1}^{2}} \tag{5—16}$$

$$s_{\hat{\beta}_2}=\sqrt{s_{\hat{\beta}_2}^{2}} \tag{5—17}$$

步骤四：计算 t 值。

$$t_{\hat{\alpha}}=\frac{\text{回归系数的估计值}}{\text{回归系数的标准误差}}=\frac{\hat{\alpha}}{s_{\hat{\alpha}}} \tag{5—18}$$

$$t_{\hat{\beta}_1}=\frac{\text{回归系数的估计值}}{\text{回归系数的标准误差}}=\frac{\hat{\beta}_1}{s_{\hat{\beta}_1}} \tag{5—19}$$

$$t_{\hat{\beta}_2}=\frac{\text{回归系数的估计值}}{\text{回归系数的标准误差}}=\frac{\hat{\beta}_2}{s_{\hat{\beta}_2}} \tag{5—20}$$

步骤五：对估计的回归系数 $\hat{\alpha}$、$\hat{\beta}_1$、$\hat{\beta}_2$ 进行显著性检验（t 检验）。

由于根据步骤四求出的 t 值服从自由度 $n-3$ 的 t 分布，与一元回归模型的情况相同，可以根据表 5—1 的 t 分布表，进行显著性检验。

［例题 5—1］

表 5—2 记载了(1)项～(6)项回归分析，计算各自的自由度，并对回归系数进行显著性检验（t 检验），如果显著，记做○，如果不显著，记做×。这里，假设回归模型中有定数项（常数项）。

表 5—2

计算出的 t 值	解释变量个数	样本个数	自由度	双侧检验		单侧检验	
				显著水平 5%	显著水平 1%	显著水平 5%	显著水平 1%
(1)	2	6					
(2)	5	47					
(3)	1	30					
(4)	3	54					
(5)	4	125					
(6)	2	300					

[解答]

有常数项的回归模型的自由度，可由下式计算：

自由度＝样本个数－解释变量个数－1

显著性检验指的是，计算出来的 t 值，如果大于其自由度相对应的 t 分布表中的判定值，则为显著，否则为不显著，见表 5—3。

表 5—3

	自由度	双侧检验		单侧检验	
		显著水平 5%	显著水平 1%	显著水平 5%	显著水平 1%
(1)	3	○	×	○	×
(2)	41	×	×	○	×
(3)	28	○	○	○	○
(4)	50	○	×	○	○
(5)	120	×	×	×	×
(6)	297	○	×	○	○

[例题 5—2]

根据一元回归模型 $Y=\alpha+\beta X+u$ 的估计结果，回答以下问题。括号中的数值，上一行是标准误差，下一行是 t 值。

$$Y=14.107+1.224X$$

$$(1.863)\quad(0.061)\qquad R^2=0.9760$$

$$(7.751)\quad(20.166)\qquad n=12$$

(1) 按 5%的显著水平，对回归系数进行显著性检验；

(2) 求关于 α 和 β 的 95%的置信区间 (confidence interval)。

[解答]

(1) 设样本个数为 n，解释变量个数为 k，则 t 检验的自由度为 $n-k-1=12-1-1=10$。根据表 5—1 的 t 分布表，双侧检验中显著水平为 5%，自由度为 10 的判定值为 2.228。因此，

$$|t_{\hat{\alpha}}|=7.751>2.228$$

$$|t_{\hat{\beta}}|=20.166>2.228$$

这样，原假设（$H_0:\alpha=0, H_0':\beta=0$）被放弃，估计的回归系数在5%水平上显著。

(2) 设 α 与 β 的估计值为 $\hat{\alpha}$、$\hat{\beta}$，标准误差为 $s_{\hat{\alpha}}$，$s_{\hat{\beta}}$，α 与 β 的95%的置信区间为：

$\hat{\alpha}\pm$（t 分布表双侧检验中5%显著水平上自由度 $n-2$ 的判定值）$\times s_{\hat{\alpha}}$
$\hat{\beta}\pm$（t 分布表双侧检验中5%显著水平上自由度 $n-2$ 的判定值）$\times s_{\hat{\beta}}$

因此，α 的95%的置信区间为

$$14.107\pm2.228\times1.863=(9.956,18.258)$$

β 的95%置信区间为

$$1.224\pm2.228\times0.061=(1.088,1.360)$$

这就是说，分析者对于 α 处于9.956～18.258之间、β 处于1.088～1.360之间这件事，具有95%的把握。

[例题5—3]

表5—4反映了新加坡1981—1993年间，实际GDP(X)与进口额(Y)的变化。

(1) 对进口函数 $Y=\alpha+\beta X+u$ 的回归系数 α 与 β 进行OLS估计。这里，$\beta>0$。

(2) 计算决定系数 R^2。

(3) 计算残差方差 s^2 和回归方差的标准误差 s。

(4) 计算回归系数的标准误差 s_α，s_β。

(5) 计算 t 值，并在1%水平下，对回归系数进行显著性检验。

表5—4　　新加坡的实际GDP和实际进口额的变化　　单位：10亿新加坡元

年份	实际GDP X	实际进口额 Y
1981	31	51
1982	33	54
1983	36	56
1984	39	61
1985	38	59

续前表

年份	实际 GDP X	实际进口额 Y
1986	39	65
1987	43	73
1988	47	93
1989	52	101
1990	56	116
1991	60	126
1992	64	134
1993	70	157

说明：1987 年价格。

资料来源：World Bank, *World Tables*。

[解答]

(1) 将数据填入工作表中，进行计算（见表 5—5）。

表 5—5 工作表

年份	X	Y	XY	X^2	Y^2	$\hat{Y}$	u	u^2
1981	31	51	1 581	961	2 601	43.761 42	7.238 58	52.397 0
1982	33	54	1 782	1 089	2 916	49.391 68	4.608 32	21.236 6
1983	36	56	2 016	1 296	3 136	57.837 07	−1.837 07	3.374 8
1984	39	61	2 379	1 521	3 721	66.282 46	−5.282 46	27.904 4
1985	38	59	2 242	1 444	3 481	63.467 33	−4.467 33	19.957 0
1986	39	65	2 535	1 521	4 225	66.282 46	−1.282 46	1.644 7
1987	43	73	3 139	1 849	5 329	77.542 98	−4.542 98	20.638 6
1988	47	93	4 371	2 209	8 649	88.803 50	4.196 50	17.610 6
1989	52	101	5 252	2 704	10 201	102.879 15	−1.879 15	3.531 2
1990	56	116	6 496	3 136	13 456	114.139 67	1.860 33	3.460 8
1991	60	126	1 560	3 600	15 876	125.400 19	0.599 81	0.359 8
1992	64	134	8 576	4 096	17 956	136.660 71	−2.660 71	7.079 4
1993	70	157	10 990	4 900	24 649	153.551 48	3.448 52	11.892 3
—	608	1 146	58 919	30 326	116 196	—	—	191.087 3
	↑ $\sum X$	↑ $\sum Y$	↑ $\sum XY$	↑ $\sum X^2$	↑ $\sum Y^2$			↑ $\sum u^2$

$$\hat{\beta} = \frac{n\sum XY - \sum X \sum Y}{n\sum X^2 - (\sum X)^2}$$

$$= \frac{13 \times 58\ 919 - 608 \times 1\ 146}{13 \times 30\ 326 - (608)^2} = \frac{69\ 179}{24\ 574}$$

$$= 2.815\ 13$$

$$\hat{\alpha} = \frac{\sum Y - \hat{\beta}\sum X}{n}$$

$$= \frac{1\ 146 - \frac{69\ 179}{24\ 574} \times 608}{13}$$

$$= -43.507\ 61$$

因此，新加坡的进口函数为

$$\hat{Y} = -43.507\ 61 + 2.815\ 13X$$

边际进口倾向为2.815 13，即每一单位GDP的增加，相应地有2.8单位进口额的增加。由此可见，新加坡经济的特征之一是进口贸易依存度极高。

(2) 决定系数 R^2 为

$$R^2 = \frac{[n\sum XY - (\sum X)(\sum Y)]^2}{[n\sum X^2 - (\sum X)^2][n\sum Y^2 - (\sum Y)^2]}$$

$$= \frac{(13 \times 58\ 919 - 608 \times 1\ 146)^2}{[13 \times 30\ 326 - (608)^2][13 \times 116\ 196 - (1\ 146)^2]}$$

$$= \frac{4\ 785\ 734\ 041}{4\ 846\ 779\ 168}$$

$$= 0.987\ 4$$

估计出的进口函数的拟合度非常良好。

(3) 根据 (5—1) 式，求残差方差 s^2。

$$s^2 = \frac{\sum \hat{u}^2}{n - 2}$$

$$= \frac{191.087\ 3}{13 - 2}$$

$$= 17.371\ 57$$

回归方程的标准误差 s 为

$$s=\sqrt{s^2}$$

$$=\sqrt{17.371\ 57}$$

$$=4.167\ 92$$

(4) 根据 (5—4) 式，(5—6) 式计算回归系数估计值 $\hat{\alpha}$、$\hat{\beta}$ 的方差 $s_{\hat{\alpha}}^2$、$s_{\hat{\beta}}^2$。

$$s_{\hat{\alpha}}^2=\frac{s^2\sum X^2}{n\sum X^2-(\sum X)^2}$$

$$=\frac{17.371\ 57\times 30\ 326}{13\times 30\ 326-(608)^2}$$

$$=21.437\ 7$$

$$s_{\hat{\beta}}^2=\frac{s^2 n}{n\sum X^2-(\sum X)^2}$$

$$=\frac{17.371\ 57\times 13}{24\ 574}$$

$$=0.009\ 189\ 81$$

标准误差 $s_{\hat{\alpha}}$、$s_{\hat{\beta}}$ 为 $s_{\hat{\alpha}}^2$、$s_{\hat{\beta}}^2$ 的正的平方根，即

$$s_{\hat{\alpha}}=\sqrt{s_{\hat{\alpha}}^2}$$

$$=\sqrt{21.437\ 7}$$

$$=4.630\ 1$$

$$s_{\hat{\beta}}=s_{\hat{\beta}}^2$$

$$=\sqrt{0.009\ 189\ 81}$$

$$=0.095\ 864$$

(5) 根据 (5—9) 式，(5—10) 式求 t 值。

$$t_{\hat{\alpha}}=\frac{\hat{\alpha}}{s_{\hat{\alpha}}}=\frac{-43.507\ 6}{4.630\ 1}$$

$$=-9.397$$

$$t_{\hat{\beta}}=\frac{\hat{\beta}}{s_{\hat{\beta}}}=\frac{2.8151}{0.095864}$$

$$=29.366$$

t 检验的自由度为

$$n-k-1=13-1-1=11$$

$t_{\hat{\alpha}}$为双侧检验，另一方面，$t_{\hat{\beta}}$由于存在$\beta>0$这一符号条件，为单侧检验。根据表5—1的t分布表，得

$$|t_{\hat{\alpha}}|=9.397>3.106$$

$$|t_{\hat{\beta}}|=29.366>2.718$$

放弃原假设（$H_0:\alpha=0$，$H'_0:\beta=0$），估计出来的回归系数在1%水平上显著。

2. F 值

t 值用于检验单个回归系数的显著性，而 F 值（F value）是在多元回归分析中对多个回归系数进行综合检验（F 检验）时采用的。F 检验也称为决定系数 R^2 或重相关系数 R 的显著性检验。

步骤一：建立原假设和备择假设。

原假设 H_0：常数项以外所有的回归系数为零。

备择假设 H_1：H_0 不成立。

上述原假设意味着除常数项以外所有的解释变量对被解释变量 Y 没有任何影响，也即在估算出的多元回归模型中没有意义。相反，如果原假设被放弃，备择假设得到支持，则可以判断解释变量的全部或部分对被解释变量 Y 有影响。但是，哪一个解释变量是有效的还无法判定。与 t 检验一样，F 检验也是希望放弃原假设、支持备择假设。

步骤二：计算 F 值。

$$F=\frac{\dfrac{\text{回归平方和}}{\text{解释变量数}}}{\dfrac{\text{残差平方和}}{\text{样本数}-\text{解释变量数}-1}}$$

$$
= \frac{\dfrac{\sum(\hat{Y}-\overline{Y})}{k}}{\dfrac{\sum \hat{u}^2}{n-k-1}} \tag{5—21}
$$

$$
= \frac{决定系数}{1-决定系数} \times \frac{样本数-解释变量数-1}{解释变量数}
$$

$$
= \frac{R^2}{1-R^2} \times \frac{n-k-1}{k} \quad [计算式] \tag{5—22}
$$

步骤三：计算出来的 F 值，服从自由度（分子，分母）=（$k, n-k-1$）的 F 分布，将其与 F 分布表（见表 5—6）中得到的 F 值（判定值）相比较，进行显著性检验。如果计算出的 F 值大于判定值，放弃原假设，结果为显著。通常，显著水平采用 5% 或 1%。顺便说一下，在 F 分布表中，横向为分子，纵向为分母。

表 5—6 **F 分布表**

自由度（纵向分母）	自由度（横向：分子）									
	显著水平 5%					显著水平 1%				
	1	2	3	4	5	1	2	3	4	5
1	161	200	216	225	230	4 052	5 000	5 403	5 625	5 764
2	18.5	19.0	19.2	19.2	19.3	98.5	99.0	992	99.2	99.3
3	10.1	9.55	9.28	9.12	9.01	34.1	30.8	29.5	28.7	28.2
4	7.71	6.94	6.59	6.39	6.26	21.2	18.0	16.7	16.0	15.5
5	6.61	5.79	5.41	5.19	5.05	16.3	13.3	12.1	11.4	11.0
6	5.99	5.14	4.76	4.53	4.39	13.7	10.9	9.78	9.15	8.75
7	5.59	4.74	4.35	4.12	3.97	12.2	9.55	8.45	7.85	7.46
8	5.32	4.46	4.07	3.84	3.69	11.3	8.65	7.59	7.01	6.63
9	5.12	4.26	3.86	3.63	3.84	10.6	8.02	6.99	6.42	6.60
10	4.96	4.10	3.71	3.48	3.33	10.0	7.56	6.55	5.99	5.64
11	4.84	3.98	3.59	3.36	3.20	9.65	7.21	6.22	5.67	5.32
12	4.75	3.89	3.49	3.26	3.11	9.33	6.93	5.95	5.41	5.06
13	4.67	3.81	3.41	3.18	3.03	9.07	6.70	5.74	5.21	4.86
14	4.60	3.74	3.34	3.11	2.96	8.86	6.51	5.56	5.04	4.70
15	4.54	3.68	3.29	3.06	2.90	8.68	6.36	5.42	4.89	4.56

续前表

自由度（纵向：分母）	自由度（横向：分子）									
	显著水平 5%					显著水平 1%				
	1	2	3	4	5	1	2	3	4	5
16	4.49	3.63	3.24	3.01	2.85	8.53	6.23	5.29	4.77	4.44
17	4.45	3.59	3.20	2.96	2.81	8.40	6.11	5.19	4.67	4.34
18	4.41	3.55	3.16	2.93	2.77	8.29	6.01	5.09	4.58	4.25
19	4.38	3.52	3.13	2.90	2.74	8.19	5.93	5.01	4.50	4.17
20	4.35	3.49	3.10	2.87	2.71	8.10	5.85	4.94	4.43	4.10
21	4.32	3.47	3.07	2.84	2.68	8.02	5.78	4.87	4.37	4.04
22	4.30	3.44	3.05	2.82	2.66	7.95	5.72	4.82	4.31	3.99
23	4.28	3.42	3.03	2.80	2.64	7.88	5.66	4.76	4.26	3.94
24	4.26	3.40	3.01	2.78	2.62	7.82	5.61	4.72	4.22	3.90
25	4.24	3.39	2.99	2.76	2.60	7.77	5.57	4.68	4.18	3.86
30	4.17	3.32	2.92	2.69	2.53	7.56	5.39	4.51	4.02	3.70
40	4.08	3.23	2.84	2.61	2.45	7.31	5.18	4.31	3.83	3.51
60	4.00	3.15	2.76	2.53	2.37	7.08	4.98	4.13	3.65	3.34
120	3.92	3.07	2.68	2.45	2.29	6.85	4.79	3.95	3.48	3.17
∞	3.84	3.00	2.60	2.37	2.21	6.63	4.61	3.78	3.32	3.02

说明：该表可用于 F 检验和 Chow 检验。

[例题 5—4]

表 5—7 是 10 个家庭的月均储蓄 Y、月收入 X_1 以及家庭人数 X_2 的数据。

(1) 对下列多元回归模型进行 OLS 估计。

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + u$$

(2) 计算决定系数 R^2 和自由度调整后的决定系数 $\bar{R}^2$。

(3) 计算 F 值，并对估计出的回归系数的显著性进行综合检验；设显著水平为 1%。

(4) 计算 t 值，并对估算出的回归系数的显著性进行检验；设显著水平为 1%。

表 5—7　　10 个家庭的月均储蓄、月收入以及家庭人数

家庭 i	月均储蓄（万日元） Y	月收入（万日元） X_1	家庭人数（人） X_2
1	7	40	4
2	6	32	3
3	9	48	4
4	5	26	3
5	4	35	5
6	7	30	2
7	4	27	4
8	5	41	6
9	9	37	2
10	4	44	7

［解答］

（1）步骤一：省略工作表。

$$\sum Y = 60 \qquad \sum X_1 = 360 \qquad \sum X_2 = 40$$

$$\sum Y^2 = 394 \qquad \sum X_1^2 = 13\ 444 \qquad \sum X_2^2 = 184$$

$$\sum YX_1 = 2\ 206 \qquad \sum YX_2 = 223 \qquad \sum X_1X_2 = 1\ 497$$

$$\sum a^2 = 0.558\ 145$$

步骤二：

$$S_{YY} = \sum Y^2 - \frac{(\sum Y)^2}{n} = 34$$

$$S_{11} = \sum X_1^2 - \frac{(\sum X_1)^2}{n} = 484$$

$$S_{22} = \sum X_2^2 - \frac{(\sum X_2)^2}{n} = 24$$

$$SY_1 = \sum YX_1 - \frac{(\sum Y)(\sum X_1)}{n} = 46$$

$$S_{Y2} = \sum YX_2 - \frac{(\sum Y)(\sum X_2)}{n} = -17$$

$$S_{12} = \sum X_1 X_2 - \frac{(\sum X_1)(\sum X_2)}{n} = 57$$

步骤三：

$$D_0 = S_{11} S_{22} - S_{12}^2 = 8\ 367$$

$$D_1 = S_{Y1} S_{22} - S_{Y2} S_{12} = 2\ 073$$

$$D_2 = S_{Y2} S_{11} - S_{Y1} S_{12} = -10\ 850$$

步骤四：设 α、β_1、β_2 的估计值为 $\hat{\alpha}$、$\hat{\beta}_1$、$\hat{\beta}_2$，

$$\hat{\beta}_1 = \frac{D_1}{D_0} = 0.247\ 759$$

$$\hat{\beta}_2 = \frac{D_2}{D_0} = -1.296\ 761$$

$$\hat{\alpha} = \frac{\sum Y}{n} - \hat{\beta}_1 \frac{\sum X_1}{n} - \hat{\beta}_2 \frac{\sum X_2}{n} = 2.267\ 718$$

因此，估计出的多元回归模型为

$$\hat{Y} = 2.267\ 718 + 0.247\ 759 X_1 - 1.296\ 761 X_2$$

回归系数的符号条件也得到满足。

(2) 根据 (4—19) 式，求决定系数 R^2。

$$R^2 = \frac{\hat{\beta}_1 S_{Y1} + \hat{\beta}_2 S_{Y2}}{S_{YY}} = \frac{0.247\ 759 \times 46 + (-1.296\ 761) \times (-17)}{34}$$

$$= 0.983\ 58$$

由于样本数 n 为 10，解释变量 k 为 2，根据 (4—21) 式，自由度调整后的决定系数 $\bar{R}^2$ 为

$$\bar{R}^2 = 1 - \frac{n-1}{n-k-1}(1-R^2) = 1 - \frac{10-1}{10-2-1} \times (1-0.983\ 58)$$

$$= 0.978\ 89$$

(3) 根据 (5—22) 式求 F 值。

$$F = \frac{R^2}{1-R^2} \times \frac{n-k-1}{k} = \frac{0.983\ 58}{1-0.983\ 58} \times \frac{10-2-1}{2}$$

$$= 209.7$$

根据表 5—6（F 分布表），1% 显著水平 F 自由度（分子，分母）＝（$k=2$，$n-k-1=7$）的 F 检验的判定值 $F_0=9.55$，由于估计出来的 F 值（209.7）大于 F_0，因此放弃原假设（$H_0:\beta_1=\beta_2=0$），可见解释变量全部或部分对 Y 有影响。

（4）步骤一：根据（5—11）式求残差方差 s^2。

$$s^2=\frac{\sum\hat{u}^2}{n-3}=\frac{0.558\ 145}{10-3}=0.079\ 735$$

步骤二：根据（5—12）式、（5—13）式、（5—14）式分别计算回归系数估计值 $\hat{\alpha}$、$\hat{\beta}_1$、$\hat{\beta}_2$ 的方差 $s^2_{\hat{\alpha}}$、$s^2_{\hat{\beta}_1}$、$s^2_{\hat{\beta}_2}$。

$$s^2_{\hat{\alpha}}=s^2\left(\frac{1}{n}+\frac{\overline{X}_1^2S_{22}+\overline{X}_2^2S_{11}-2\overline{X}_1\overline{X}_2S_{12}}{S_{11}S_{22}-S_{12}^2}\right)$$

$$=0.079\ 735\times\left(\frac{1}{10}+\frac{36^2\times24+4^2\times484-2\times36\times4\times57}{484\times24-57^2}\right)$$

$$=0.221\ 744$$

$$s^2_{\hat{\beta}_1}=s^2\left(\frac{S_{22}}{S_{11}S_{22}-S_{12}^2}\right)$$

$$=0.079\ 735\times\frac{24}{8\ 367}$$

$$=0.000\ 228\ 712$$

$$s^2_{\hat{\beta}_2}=s^2\left(\frac{S_{11}}{S_{11}S_{22}-S_{12}^2}\right)$$

$$=0.079\ 735\times\frac{484}{8\ 367}$$

$$=0.004\ 612\ 37$$

步骤三：由于回归系数的标准误差 $s_{\hat{\alpha}}$、$s_{\hat{\beta}_1}$、$s_{\hat{\beta}_2}$ 是 $s^2_{\hat{\alpha}}$、$s^2_{\hat{\beta}_1}$、$s^2_{\hat{\beta}_2}$ 的正平方根，则

$$s_{\hat{\alpha}}=\sqrt{s^2_{\hat{\alpha}}}=\sqrt{0.221\ 744}=0.470\ 897$$

$$s_{\hat{\beta}_1}=\sqrt{s^2_{\hat{\beta}_1}}=\sqrt{0.000\ 228\ 712}=0.015\ 123$$

$$s_{\hat{\beta}_2}=\sqrt{s^2_{\hat{\beta}_2}}=\sqrt{0.004\ 612\ 37}=0.067\ 914$$

步骤四：根据（5—18）式、（5—19）式、（5—20）式计算 t 值。

$$t_{\hat{\alpha}}=\frac{\hat{\alpha}}{s_{\hat{\alpha}}}=\frac{2.267\,718}{0.470\,897}=4.816$$

$$t_{\hat{\beta}_1}=\frac{\hat{\beta}_1}{s_{\hat{\beta}_1}}=\frac{0.247\,759}{0.015\,123}=16.383$$

$$t_{\hat{\beta}_2}=\frac{\hat{\beta}_2}{s_{\hat{\beta}_2}}=\frac{-1.296\,761}{0.067\,914}=-19.094$$

步骤五：t 检验的自由度为

$$n-k-1=10-2-1=7$$

设 $t_{\hat{\alpha}}$ 为双侧检验，而 $t_{\hat{\beta}_1}$、$t_{\hat{\beta}_2}$存在 $\hat{\beta}_1>0$，$\hat{\beta}_2<0$ 的符号条件，设为单侧检验，根据表 5—1（t 分布表），得

$$|t_{\hat{\alpha}}|=4.816>3.499$$

$$|t_{\hat{\beta}_1}|=16.383>2.998$$

$$|t_{\hat{\beta}_2}|=19.094>2.998$$

放弃原假设（$H_0:\alpha=0$、$H'_0:\beta_1=0$、$H''_0:\beta_2=0$），估计出来的回归系数在 1%水平上显著。

3. 结构变化的 F 检验

结构变化的 F 检验，也称为 Chow test，用于调查、检验经济分析中一个极其重要的问题，即“是否存在结构变化”。结构变化的 F 检验的基本步骤如下：

步骤一：在利用时间序列所做的回归分析中，找出估算期间内发生结构变化的时点（分界点），以此时点为标准，将期间分为前期和后期。

步骤二：对前期、后期、全部期间进行回归分析，求各自的残差平方和 $SSR1$、$SSR2$、SSR。

步骤三：根据结构变化的 F 检验公式，计算 F 值。

$SSR1$：前期的残差平方和　　　n_1：前期的样本数

$SSR2$：后期的残差平方和　　　n_2：后期的样本数

SSR：全部期间的残差平方和　　k：解释变量数

（1）$n_1>k+1$ 而且 $n_2>k+1$ 的情形。

结构变化的 F 检验为

$$F=\frac{SSR-(SSR1+SSR2)}{SSR1+SSR2}\times\frac{n_1+n_2-2(k+1)}{k+1} \tag{5—23}$$

(2) $n_2\leqslant k+1$ 的情形（以及 $n_1\leqslant k+1$ 的情形）。

结构变化的 F 检验为

$$F=\frac{SSR-SSR1}{SSR1}\times\frac{n_1-(k+1)}{n_2} \tag{5—24}$$

步骤四：利用表 5—6 的 F 分布表，对步骤三计算出的 F 值进行检验。在检验时，分别就上述（1）的情形中，自由度(分子，分母) = $(k+1, n_1+n_2-2k-2)$，(2)的情形中，自由度 (n_2, n_1-k-1) 进行 F 检验。

如果计算出的 F 值大于F 分布表中的判定值，放弃"前期的回归系数与后期的回归系数完全相等"的假设，说明出现了结构性变化。相反，如果计算出的 F 值小于F 分布表中的判定值，不放弃"前期的回归系数与后期的回归系数完全相等"的假设，说明没有发生结构变化。

[例题 5—5]

表 5—8 为中国综合零售物价指数 P 的变化。自 20 世纪 70 年代末经济改革以来，物价水平趋势是否发生了变化？下面用一元回归模型 $Y=\alpha+\beta t+u$ 进行估算，并作结构变化的 F 检验（Chow test）。

此处设 1978 年为分界点，前期为 1971—1978 年，后期为 1979—1985 年。

表 5—8　　中国综合零售物价指数的变化　　单位：1950 年为 100

年份 (t)	综合零售物价指数 P	年份 (t)	综合零售物价指数
1971 (1)	130.5	1979 (9)	138.6
1972 (2)	130.2	1980 (10)	146.9
1973 (3)	131.0	1981 (11)	150.4
1974 (4)	131.7	1982 (12)	153.3
1975 (5)	131.9	1983 (13)	155.6
1976 (6)	132.3	1984 (14)	160.0
1977 (7)	135.0	1985 (15)	174.1
1978 (8)	135.9		

资料来源：中国国家统计局《中国统计年鉴》。

[解答]

步骤一：

分界点为 1978 年

(1) 前期：1971—1978 年

(2) 后期：1979—1985 年

步骤二：

(1) 前期 (1971—1978 年)

$$P=128.78+0.784\ 52t$$

$$R^2=0.8586 \qquad SSR1=4.258\ 7$$

(2) 后期 (1979—1985 年)

$$P=95.029+4.9250t$$

$$R^2=0.913\ 5 \qquad SSR2=64.316\ 8$$

(3) 全部期间 (1971—1985 年)

$$P=120.16+2.791\ 8t$$

$$R^2=0.855\ 7 \qquad SSR=368.090\ 4$$

步骤三：

由于本例属于 $n_1>k+1$ 而且 $n_2>k+1$ 的情形，可以由 (5—23) 式计算 F 值。

$$\begin{aligned}F&=\frac{SSR-(SSR1+SSR2)}{SSR1+SSR2}\times\frac{n_1+n_2-2\ (k+1)}{k+1}\\&=\frac{368.090\ 4-\ (4.258\ 7+64.316\ 8)}{4.258\ 7+64.316\ 8}\times\frac{8+7-2\times\ (1+1)}{1+1}\\&=\frac{299.514\ 9}{68.575\ 5}\times\frac{11}{2}\\&=24.022\end{aligned}$$

步骤四：

从表 5—6 的 F 分布表中可以找到，显著水平 5%下自由度 ($k+1=2$，$n_1+n_2-2k-2=11$) 的 F 检验的判定值 $F_0=3.98$，计算出来的 F 值 (24.022) 大于 F_0，“前期的回归系数与后期的回归系数完全相等”的假设被放弃，因此可以判断

出现了结构性变化。也就是说，受20世纪70年代末期开始的经济体制改革的影响，中国的物价水平趋势也发生了变化。

4. 预测

下面利用估算出来的回归模型，说明预测置信区间的计算方法。所谓预测置信区间，指的是被解释变量 Y 的预测值 Y_0，在某个概率（例如95%）下取值的上限与下限范围。具体说来，预测置信区间的计算方法有以下几个步骤。

步骤一：

将给定的观测数据用于一元回归模型，进行OLS估算。

$$\hat{Y} = \hat{\alpha} + \hat{\beta}X$$

在这一步骤中，将 s（回归方程式的标准误差）、n（样本数）、$\bar{X}$（平均值）、$\sum(X-X)^2$ 计算出来，留待步骤五使用。

步骤二：

预测者自己决定作为预测基础的 X_0，代入步骤一中计算出来的回归模型中，计算 $\hat{Y}_0$(预测值)。这一工作称为点预测。

$$\hat{Y}_0 = \hat{\alpha} + \hat{\beta}X_0 \tag{5—25}$$

式中，X_0 不限于未来值，观测期间以外的过去值也可以。

步骤三：

决定预测的信赖度（信赖系数）A%。通常设信赖度为95%或99%。

步骤四：

根据 t 分布表（表5—1），查找自由度为 $n-2$，显著水平（双侧）为 $(100-A)$% 的 $\boxed{t\text{ 值}}$。

步骤五：

将上述各步骤中计算出来的数值代入下面预测置信区间的计算公式（5—26）中，预测区间的计算即告完成。

$$\hat{Y}_0 \pm \boxed{t\text{ 值}} \times s \times \sqrt{1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum(X-\bar{X})^2}} \tag{5—26}$$

预测的准确程度下降的原因主要有以下四点：

(1) 样本数 n 较小。

(2) 回归方程的标准误差 s 较大。

(3) $\sum(X-\bar{X})^2$ 较小，也即 X 在平均值 $\bar{X}$ 的附近过于集中。

(4) 作为预测基准的 X_0 与平均值 $\bar{X}$ 离得较远。

预测的置信区间的图示可参见图 5—1。从图中可以看出，X_0 离 $\bar{X}$ 越远（上述（4）的情形），置信区间的幅度越大（预测的准确性也就越低）。

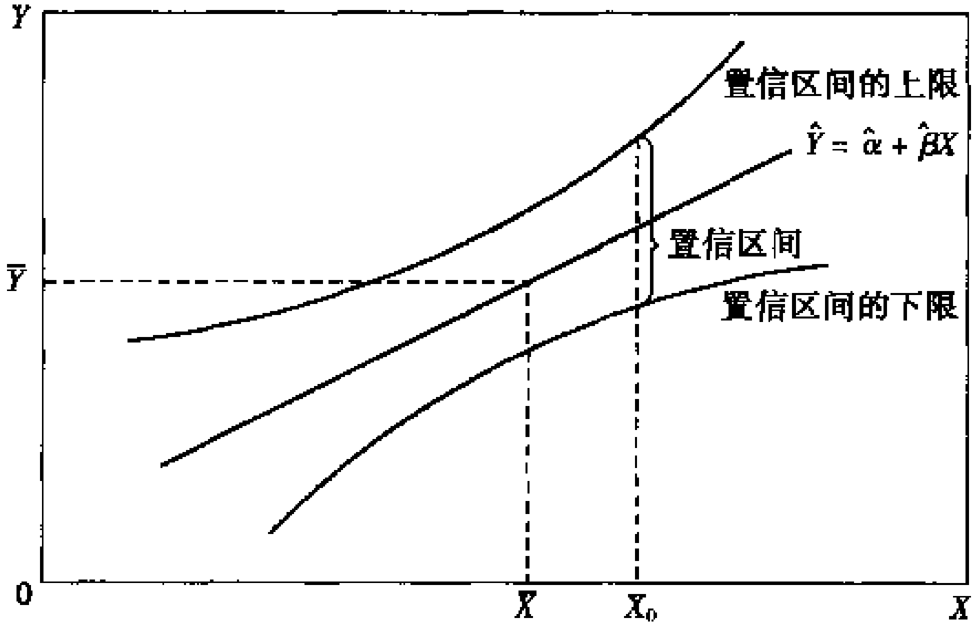


图 5—1　预测的置信区间的上限与下限

根据多元回归模型计算预测置信区间的方法，请参见岩田（1982），Madala（1992）。

[例题 5—6]

根据例题 5—3 的回归分析结果，计算新加坡的实际 GDP $X_0=80$ 时，实际进口额的预测置信区间。设预测的信赖度为 95%。

[解答]

步骤一：新加坡进口函数的 OLS 估计结果为

$$\hat{Y}=-43.507\ 61+2.815\ 13X$$

同时，计算得：

$$\left.\begin{array}{l} s=4.167\ 92 \\ n=13 \\ \bar{X}=46.769\ 23 \\ \sum(X-\bar{X})^2=1\ 890.307\ 71 \end{array}\right\}\text{留待步骤五使用}$$

步骤二：将 $X_0=80$ 代入步骤一估计的进口函数中，计算 $\hat{Y}_0$。

$$\hat{Y}_0=-43.507\ 61+2.815\ 13\times 80$$

$$=181.702\ 80$$

步骤三：现在，设预测的信赖度为 95%。

步骤四：根据 t 分布表（表 5—1），查找自由度 $n-2=11$、显著水平（两侧）为 100% − 95% = 5% 的 $\boxed{t\text{ 值}}$。

$$\boxed{t\text{ 值}}=2.201$$

步骤五：将上述各步骤计算出来的数值代入（5—26）式，计算预测的置信区间。

$$\hat{Y}_0\pm\boxed{t\text{ 值}}\times s\times\sqrt{1+\frac{1}{n}+\frac{(X_0-\bar{X})^2}{\sum(X-\bar{X})^2}}$$

$$=181.702\ 80\pm 2.201\times 4.167\ 92\times\sqrt{1+\frac{1}{13}+\frac{(80-46.769\ 23)^2}{1\ 890.307\ 71}}$$

$$=181.702\ 80\pm 11.823$$

$$=(169.9,\ 193.5)$$

由此可见，新加坡的实际进口额 Y_0 在实际 GDP 为 80 时，在信赖度为 95% 的情况下，位于（下限，上限）=（169.9，193.5）的范围中。

第五章 练习题

1. 表 5—9 反映了英国、法国、意大利的实际 GDP 与实际进口额的变化。

(1) 用 OLS 对下列一元回归模型按国别估算，并计算决定系数 R^2、标准误差以及 t 值。

$$Y=\alpha+\beta X+u$$

(2) 关于计算出来的各国边际进口倾向 $\hat{\beta}$，求 95% 的置信区间。

(3) 当各国的实际 GDP 为以下数值时，计算实际进口额的预测值 $\hat{Y}_0$。

英国 $X_0=600$

法国 $X_0=400$

意大利 $X_0=150$

(4) 对 (3) 中的各项，计算实际进口额预测的置信区间。

表 5—9 英国、法国、意大利的实际 GDP 与实际进口额

单位：10 亿英镑；100 亿法郎；10 兆里拉

年份	英国		法国		意大利	
	GDP X	进口额 Y	GDP X	进口额 Y	GDP X	进口额 Y
1985	468	106	303	67	113	17
1986	488	112	310	72	116	18
1987	512	122	317	77	120	20
1988	537	137	332	84	125	21
1989	549	148	346	91	128	23
1990	551	148	355	96	131	25
1991	540	141	357	99	133	26
1992	537	150	361	101	133	27
1993	549	154	357	97	132	25
1994	570	163	367	103	135	27
1995	583	169	375	109	139	30

资料来源：日本银行《外国经济统计》。

2. 表5—10列示了北海道、东北、关东等14个都道县农业总产值、农民家庭数以及耕地面积。

(1) 对下列多元回归模型进行OLS估计，计算决定系数 R^2、自由度调整后的决定系数 $\bar{R}^2$。

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + u$$

(2) 计算标准误差、t 值和 F 值。

表5—10　　1994年北海道、东北、关东等14个都道县的农业总产值、农民家庭数以及耕地面积

都道县	农业总产值（亿日元）Y	农民家庭数（10户）X_1	耕地面积（100公顷）X_2
1. 北海道	11 646	7 671	12 040
2. 青森	3 575	7 133	1 675
3. 岩手	3 516	8 681	1 709
4. 宫城	2 955	8 121	1 468
5. 秋田	2 953	8 148	1 585
6. 山形	3 015	6 593	1 340
7. 福岛	3 531	10 429	1 729
8. 茨城	4 907	12 104	1 913
9. 枥木	3 031	7 533	1 376
10. 群马	2 579	6 100	895
11. 埼玉	2 681	7 416	932
12. 千叶	5 109	9 534	1 444
13. 东京	364	1 200	103
14. 神奈川	1 004	2 347	249

资料来源：农林水产省《农业结构动态调查报告》。
　　　　　农林水产省《生产农业收入统计》。

3. 将例题5—5中的数据分成两个部分，对所有的期间进行Chow test，然后进行所谓的逐次Chow test（stepwith Chow test），计算 F 值。

第六章 虚拟变量

本章介绍多元回归分析中特殊的解释变量，即虚拟变量（dummy variable）。虚拟变量用于以下这些场合：(1) 需要排除数据中的异常值、季节性因素等；(2) 存在结构性变化；(3) 需要对难以量化的数据进行处理。计量经济中的虚拟变量，在明确其引入理由基础上，被用于很多的多元回归模型。

1. 临时虚拟

临时虚拟，也称为突发虚拟。为了更好地对模型进行估算，经常需要在回归模型中排除一些由突发事件产生的异常值（outlier），及其对模型的影响，例如地震、战争、内乱、罢工等。以下是引入一个临时虚拟变量的多元回归模型。

$$Y = \alpha + \beta_1 X + \beta_2 D + u \tag{6—1}$$

$$D = \begin{cases} 1 & \text{异常时期} \\ 0 & \text{平时} \end{cases}$$

式中，$\beta_2 D$ 是临时虚拟变量。一般地，如果只有一个异常值，就采用一个虚拟变量。如果异常值有两个（例如地震的年份和水灾的年份），那就像下面的多元回归模型一样，引入两个临时虚拟变量。

$$Y = \alpha + \beta_1 X + \beta_2 D_1 + \beta_3 D_2 + u \tag{6—2}$$

$$D_1 = \begin{cases} 1 & \text{发生地震的年份} \\ 0 & \text{其他年份} \end{cases}$$

$$D_2 = \begin{cases} 1 & \text{发生水灾的年份} \\ 0 & \text{其他年份} \end{cases}$$

像上面这样的临时虚拟变量，对于异常值的处理来说是非常便利的变量，但是在引入之前，对于引入的理由、根据必须做出清晰的说明。因此，如果没有理由，不应该单纯为了提高模型的适应性，而使用虚拟变量。

[例题 6—1]

表 6—1 反映了日本 1985—1995 年 11 年间水稻产量 Y 和耕种面积 X 的变化。

(1) 以 X 为横轴、Y 为纵轴，画出数据的散点图。

(2) 对下面的单元回归模型进行 OLS 估计，并计算 t 值和决定系数 R^2。

$$Y = \alpha + \beta X + u$$

(3) 受 1993 年冻害的影响，水稻的收成指数为战后最低水平（74①），出现了前所未有的歉收。因此，设 1993 年为 $D = 1$，其他年份为 $D = 0$，引入临时虚拟变量，对下面的多元回归模型进行估算。并计算 t 值和自由度调整后的决定系数 $\bar{R}^2$。

$$Y = \alpha + \beta_1 X + \beta_2 D + u$$

表 6—1　　日本水稻产量与耕种面积的变化

年份	产量（10 万吨）Y	耕种面积（万公顷）X
1985	116	232
1986	116	228

① 74 应为 78，原文有误。——译者注

续前表

年份	产量（10 万吨）Y	耕种面积（万公顷）X
1987	106	212
1988	99	209
1989	103	208
1990	105	206
1991	96	203
1992	105	209
1993	78	213
1994	120	220
1995	107	211

资料来源：农林水产省《作物统计》。

[解答]

（1）散点图见图 6—1。

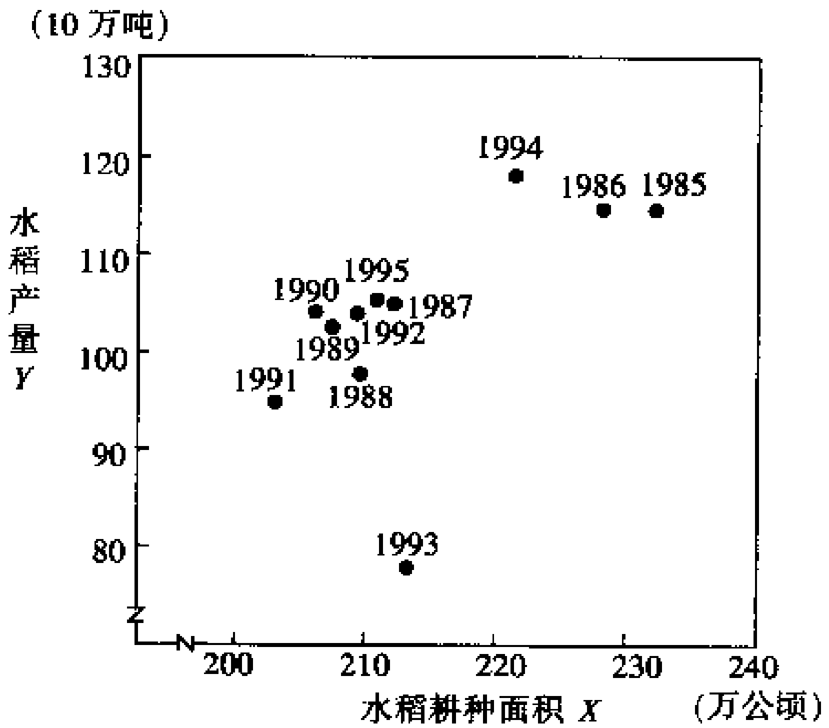


图 6—1　水稻产量与耕种面积的散点图

（2）省略工作表与计算过程。

$\sum X = 2\ 351$　　$\sum Y = 1\ 151$　　$\sum XY = 246\ 601$

$$\sum X^2 = 503\ 313 \qquad \sum Y^2 = 121\ 757 \qquad \sum \hat{u}^2 = 890.767$$

$n = 11$

$$\hat{y} = \underset{(-0.657)}{-48.224} + \underset{(2.084)}{0.715\ 21}X$$

$R^2 = 0.325\ 5$

可见，决定系数较低，模型的拟合优度不太好。

(3) 省略工作表与计算过程。

$$\sum Y = 1\ 151 \qquad \sum X = 2\ 351 \qquad \sum D = 1$$

$$\sum Y^2 = 121\ 757 \qquad \sum X^2 = 503\ 313 \qquad \sum D^2 = 1$$

$$\sum YX = 246\ 601 \qquad \sum YD = 78 \qquad \sum XD = 213$$

$$\sum \hat{u}^2 = 139.985\ 5$$

$$\hat{Y} = \underset{(-1.304)}{-40.292} + \underset{(4.782)}{0.690\ 33}X - \underset{(-6.550)}{28.748}D$$

$\bar{R}^2 = 0.867\ 5$

可见，不但 $\bar{R}^2$ 提高了，而且估算出来的回归系数除常数项均在 1% 水平显著。这样，对于歉收年景(1993 年)通过引入临时虚拟变量，消除了异常值的影响。

[补充]

去除歉收的 1993 年数据，$Y = \alpha + \beta X + u$[(2)的模型]的估算结果为

$$Y = \underset{(-1.304)}{-40.292} + \underset{(4.782)}{0.690\ 33}X$$

$R^2 = 0.740\ 8 \quad \bar{R}^2 = 0.708\ 4$

与引入虚拟变量 (3) 的模型估算结果相比较可以看出，常数项与回归系数以及相应的 t 值相同 (但是 $\bar{R}^2$ 不同)。也就是说，引入临时虚拟变量的估算结果与去除异常值的估算结果、回归系数及 t 值相等。因此，对于异常年份 (1993 年) 来说，理论值与实际值相一致，就例题而言，$\hat{Y}_{1993} = Y_{1993}$(两边同为 78)。

2. 季度虚拟

季度虚拟是通过回归模型的常数项变化 (斜率回归系数一定) 来掌握季度和月

度等季节变化，因此，从“技术的角度”称为“常数项虚拟”。关于消除季节变化，第一章已经学过移动平均方法，这里所说的季度虚拟也是一种简单而且有效的方法。

例如，利用季度数据对消费 Y 与收入 X 进行回归分析时，在夏季和冬季（第二与第四季度）发半年奖与年终奖时，收入会显得异乎寻常的高，为了消除这种季节变化，应该设立下列这样的多元回归模型，引入季度虚拟变量 D_1、D_2、D_3。

$$Y=\alpha+\beta_1X+\beta_2D_1+\beta_3D_2+\beta_4D_3+u \tag{6—3}$$

$$D_1=\begin{cases}1 & 第一季度\\ 0 & 其他\end{cases}$$

$$D_2=\begin{cases}1 & 第二季度\\ 0 & 其他\end{cases}$$

$$D_3=\begin{cases}1 & 第三季度\\ 0 & 其他\end{cases}$$

虽然季度数据共有四个，但是由于虚拟变量是以第四季度为基准（$D_1=0$、$D_2=0$、$D_3=0$），仅用三个就可以了。而且，从第一季度到第四季度的常数项为：

第一季度：$\alpha+\beta_1$

第二季度：$\alpha+\beta_2$

第三季度：$\alpha+\beta_3$

第四季度：α （基准期）

β_1、β_2、β_3 分别表示第四季度与各季度之差。

[补充]

季度虚拟变量从理论上说，季度数据情形下要用三个，月度数据情形下要用11个。但是在实际计量分析时，根据 t 检验的结果，将不显著的季度虚拟变量从模型中消除，用剩下的显著的虚拟变量对模型进行估算就足够了。

[例题 6—2]

表 6—2 中的数据反映了日本饮食消费总额（食品、饮料、烟草）Y 与国内家庭最终消费支出 X 的变化（实际值）。

(1) 用 OLS 对下面的模型（宏观恩格尔函数）进行估算。同时计算 t 值和决定系数 R^2。

$$Y=\alpha+\beta X+u$$

(2)利用计量经济分析软件，对下面的引入季度虚拟变量 D_1（第一季度）、

D_2(第二季度)、D_3(第三季度) 的模型进行估算。同时计算 t 值与自由度调整后的决定系数 $\bar{R}^2$。

$$Y = \alpha + \beta_1 X + \beta_2 D_1 + \beta_3 D_2 + \beta_4 D_3 + u$$

(3) 根据 (2) 的估算结果，计算各期的模型，并将计算结果用图表示出来。

表 6—2　日本的饮食消费总额与国内家庭最终消费支出　单位：兆日元

年·季度	饮食消费总额	国内家庭最终消费支出
1990 年1—3 月 (1)	10.0	53.5
4—6 月 (2)	11.0	54.4
7—9 月 (3)	12.2	56.4
10—12 月 (4)	13.3	60.6
1991 年1—3 月 (1)	10.2	54.7
4—6 月 (2)	11.0	55.4
7—9 月 (3)	12.3	57.6
10—12 月 (4)	13.2	62.4
1992 年1—3 月 (1)	10.5	56.5
4—6 月 (2)	11.1	56.4
7—9 月 (3)	12.3	58.3
10—12 月 (4)	13.4	62.6
1993 年1—3 月 (1)	10.4	56.7
4—6 月 (2)	11.2	56.8
7—9 月 (3)	12.2	58.9
10—12 月 (4)	13.4	63.7
1994 年 1—3 月 (1)	10.4	58.2

说明：1985 年价格，实际值。

资料来源：经济企划厅《国民经济计算年报》。

[解答]

(1) 省略工作表以及计算过程。

$$\sum X = 983.10 \qquad \sum Y = 198.10 \qquad \sum XY = 11\,505.46$$

$$\sum X^2 = 56\,993.83 \qquad \sum Y^2 = 2\,331.77$$

$$\sum \hat{u}^2 = 6.067\,29$$

$$\hat{Y} = -8.5246 + 0.34891X$$
$$(-2.756) \quad (6.531)$$

$$R^2 = 0.73985$$

(2) 以第四季度为基准，编制季度虚拟变量 D_1、D_2、D_3 的数据表，见表 6—3。

表 6—3　季度虚拟变量数据表

年·季度	D_1	D_2	D_3
1990 年1—3 月 (1)	1	0	0
4—6 月 (2)	0	1	0
7—9 月 (3)	0	0	1
10—12 月 (4)	0	0	0
1991 年1—3 月 (1)	1	0	0
4—6 月 (2)	0	1	0
7—9 月 (3)	0	0	1
10—12 月 (4)	0	0	0
1992 年1—3 月 (1)	1	0	0
4—6 月 (2)	0	1	0
7—9 月 (3)	0	0	1
10—12 月 (4)	0	0	0
1993 年1—3 月 (1)	1	0	0
4—6 月 (2)	0	1	0
7—9 月 (3)	0	0	1
10—12 月 (4)	0	0	0
1994 年 1—3 月 (1)	1	0	0

模型估计结果为

$$\hat{Y} = 9.0681 + 0.068301X - 2.5875D_1 - 1.8009D_2 - 0.76594D_3$$
$$(7.766) \quad (3.649) \quad (-19.088) \; (-12.866) \quad (-7.105)$$

$$\bar{R}^2 = 0.99390$$

$\bar{R}^2$ 极高，说明模型的拟合优度良好。同时，估算出来的回归系数，在 1% 的水平下全部显著。

(3) 在 (2) 中估算出来的模型中，代入第一季度 $D_1=1$、$D_2=0$、$D_3=0$，第二季度 $D_1=0$、$D_2=1$、$D_3=0$，第三季度 $D_1=0$、$D_2=0$、$D_3=1$，第四季度 $D_1=$

0、$D_2=0$、$D_3=0$，分别计算各期的模型。

第一季度：

$$Y=9.068\,1-2.587\,5+0.068\,301X$$
$$=6.480\,6+0.068\,301X$$

第二季度：

$$Y=9.068\,1-1.800\,9+0.068\,301X$$
$$=7.267\,2+0.068\,301X$$

第三季度：

$$Y=9.068\,1-0.765\,94+0.068\,301X$$
$$=8.302\,2+0.068\,301X$$

第四季度：

$$Y=9.068\,1+0.068\,301X$$

各计算结果中，X 的回归系数（0.068 301）均相同。

上述计算结果如图 6—2 所示。

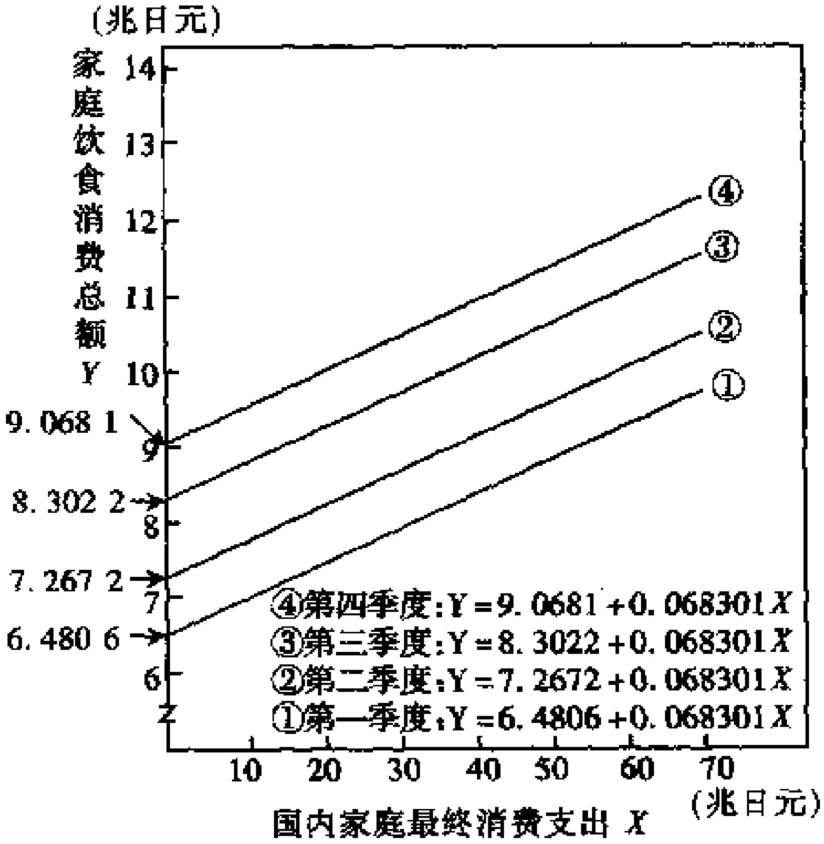


图 6—2　季度虚拟（常数项虚拟）

3. 定性数据的虚拟处理

计量经济分析中经常需要处理一些定性的差异，例如学历、性别、人种、城市与农村、自有住宅与租借住宅、制造业与非制造业等。在这种情况下，利用虚拟变量，可以将这些定性数据纳入多元回归模型。例如，可以进行下面这样的虚拟处理。

学历 $D_1 = \begin{cases} 1 & \text{大学毕业} \\ 0 & \text{其他} \end{cases}$

性别 $D_2 = \begin{cases} 1 & \text{男性} \\ 0 & \text{女性} \end{cases}$

人种 $D_3 = \begin{cases} 1 & \text{白人} \\ 0 & \text{其他} \end{cases}$

城市与农村 $D_4 = \begin{cases} 1 & \text{城市} \\ 0 & \text{农村} \end{cases}$

自有住宅与租借住宅 $D_5 = \begin{cases} 1 & \text{自有住宅} \\ 0 & \text{租借住宅} \end{cases}$

制造业与非制造业 $D_6 = \begin{cases} 1 & \text{制造业} \\ 0 & \text{非制造业} \end{cases}$

[例题 6—3]

表 6—4 列示了 15 个工人的月收入以及相应的性别、年龄层（30 多岁与 40 多岁）、学历（大学毕业、高中毕业、初中毕业）、企业规模（大型企业、中型企业、小型企业）之间的关系。根据这些定性数据，通过下面的问题分析收入差距的原因。

(1) 为了将定性数据作为解释变量纳入模型，引入下面六个虚拟变量。根据表 6—4，制作虚拟变量的数据表。

性别 $S = \begin{cases} 1 & \text{男性} \\ 0 & \text{女性} \end{cases}$

年龄 $A = \begin{cases} 1 & \text{40 多岁} \\ 0 & \text{30 多岁} \end{cases}$

学历① $E_1 = \begin{cases} 1 & \text{大学毕业} \\ 0 & \text{其他} \end{cases}$

学历② $E_2 = \begin{cases} 1 & \text{高中毕业} \\ 0 & \text{其他} \end{cases}$

企业规模① $F_1 = \begin{cases} 1 & 大型企业 \\ 0 & 其他 \end{cases}$

企业规模② $F_2 = \begin{cases} 1 & 中小型企业 \\ 0 & 其他 \end{cases}$

(2) 利用计量经济分析软件对多元回归模型（工资函数）进行 OLS 估算。同时，计算 t 值和自由度调整后的决定系数 $\bar{R}^2$。

$$Y = \alpha + \beta_1 S + \beta_2 A + \beta_3 E_1 + \beta_4 E_2 + \beta_5 F_1 + \beta_6 F_2 + u$$

$$\alpha > 0,\ \beta_1 > 0,\ \beta_2 > 0,\ \beta_3 > 0,\ \beta_4 > 0,\ \beta_5 > 0,\ \beta_6 > 0$$

(3) 解释估算出来的常数项的意义。

(4) 计算下列属性所对应的月收入。

a) 大型企业中 40 多岁男性大学毕业工人的月收入 $\hat{Y}_a$。

b) 中型企业中 30 多岁女性高中毕业工人的月收入 $\hat{Y}_b$。

c) 小型企业中 30 多岁男性初中毕业工人的月收入 $\hat{Y}_c$。

表 6—4　　月收入与性别、年龄层、学历、企业规模之间的关系

月收入（万日元）	性别	年龄层	学历	企业规模
25	女性	40 多岁	初中毕业	小企业
26	男性	30 多岁	初中毕业	小企业
28	女性	40 多岁	高中毕业	小企业
30	女性	40 多岁	高中毕业	小企业
31	男性	30 多岁	初中毕业	中企业
32	男性	30 多岁	高中毕业	小企业
34	女性	30 多岁	大学毕业	中企业
36	男性	30 多岁	高中毕业	中企业
39	女性	30 多岁	大学毕业	大企业
40	男性	30 多岁	高中毕业	中企业
43	男性	30 多岁	大学毕业	小企业
46	男性	30 多岁	大学毕业	中企业
52	男性	40 多岁	初中毕业	大企业
54	女性	40 多岁	大学毕业	大企业
55	男性	40 多岁	高中毕业	大企业

[解答]

(1) 制作虚拟变量的数据表，如表 6—5 所示。

表 6—5　　定性数据虚拟处理后的数据表

月收入(万日元) Y	性别 S	年龄层 A	学历		企业规模	
			大学毕业 E_1	高中毕业 E_2	大型企业 F_1	中型企业 F_2
25	0	1	0	0	0	0
26	1	0	0	0	0	0
28	0	1	0	1	0	0
30	0	1	0	1	0	0
31	1	0	0	0	0	1
32	1	0	0	1	0	0
34	0	0	1	0	0	1
36	1	0	0	1	0	1
39	0	0	1	0	1	0
40	1	0	0	1	0	1
43	1	0	1	0	0	0
46	1	0	1	0	0	1
52	1	1	0	0	1	0
54	0	1	1	0	1	0
55	1	1	0	1	1	0

(2)
$$\hat{Y} = \underset{(7.061)}{11.966} + \underset{(11.612)}{14.385S} + \underset{(8.320)}{12.643A} + \underset{(10.821)}{15.873E_1} + \underset{(4.541)}{5.083E_2} + \underset{(9.163)}{12.152F_1} + \underset{(4.635)}{5.544F_2}$$

$$\bar{R}^2 = 0.970\,8$$

估计出来的回归系数满足所有的符号条件，同时在 1%水平(单侧检验)均显著。

(3) 估计出的常数项 (11.966 万日元) 表示的是小型企业 30 多岁女性初中毕业工人的月收入。

(4) a)
$$\begin{aligned}\hat{Y}_a &= 11.966 + 14.385\times1 + 12.643\times1 + 15.873\times1 + 5.083\times0 \\ &\quad + 12.152\times1 + 5.544\times0 \\ &= 67.190\ (\text{万日元})\end{aligned}$$

b)
$$\begin{aligned}\hat{Y}_b &= 11.966 + 14.385\times0 + 12.643\times0 + 15.873\times0 + 5.083\times1 \\ &\quad + 12.152\times0 + 5.544\times1 \\ &= 22.593\ (\text{万日元})\end{aligned}$$

c) $\hat{Y}_c = 11.966 + 14.385 \times 1 + 12.643 \times 0 + 15.873 \times 0 + 5.083 \times 0$
$+ 12.152 \times 0 + 5.544 \times 0$
$= 26.351$（万日元）

4. 系数虚拟

所谓系数虚拟，是为了反映结构变化之前与之后的回归系数(斜率)的差异(而不是常数项)而采取的虚拟变量处理方法。利用系数虚拟变量的回归模型，形式如下：

$$Y = \alpha + \beta_1 X + \beta_2 DX + u \tag{6—4}$$

系数虚拟　$D = \begin{cases} 0 & \text{结构变化之前} \\ 1 & \text{结构变化之后} \end{cases}$

也就是说，结构变化之后的回归系数为 $\beta_1 + \beta_2$，结构变化之前的回归系数为 β_1。不论在哪种情形下，常数项均为 α。

此外，如果结构变化引起回归系数和常数项双方变化，可以用下面的模型引入系数虚拟变量和常数虚拟变量。

$$Y = \alpha + \beta_1 X + \beta_2 DX + \beta_3 D + u \tag{6—5}$$

[例题 6—4]

表 6—6 用指数的方式（1965 年为 100）列示了包括 1973 年石油危机在内的 1965—1979 年间，某一国家初次能源需求量 Y 与实际 GDPX 的变化。

(1) 以 X 为横轴、Y 为纵轴，画出数据的散点图。

(2) 对下面的回归模型进行 OLS 估算，并计算 t 值与决定系数 R^2。

$$Y = \alpha + \beta X + u \qquad \beta > 0$$

(3) 考虑 1973 年石油危机以后，该国能源需求结构的变化，对下面引入系数虚拟变量的多元回归模型进行 OLS 估算，同时画出数据表。

$$Y = \alpha + \beta_1 X + \beta_2 DX + u \qquad \beta_1 > 0 \quad \beta_2 < 0$$

$$D = \begin{cases} 0 & \text{石油冲击前(1965—1972 年)} \\ 1 & \text{石油冲击后(1973—1979 年)} \end{cases}$$

式中，设 $\beta_2 < 0$，是因为考虑到石油冲击后，出现了节能性的经济增长。

表 6—6　　初次能源需求量与实际 GDP 的变化　　指数：1965 年为 100

年　份	初次能源需求量 Y	实际 GDP X
1965	100	100
1966	106	108
1967	115	117
1968	122	123
1969	129	132
1970	136	141
1971	141	145
1972	143	154
1973	114	150
1974	117	156
1975	121	161
1976	123	169
1977	129	174
1978	130	177
1979	134	183

[解答]

（1）散点图如图 6—3 所示。

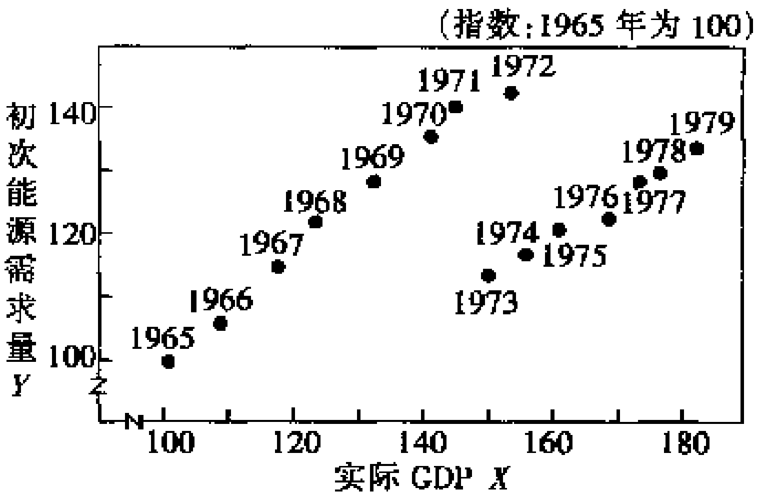


图 6—3　实际 GDP 与初次能源需求量的散点图

（2）省略工作表和计算过程。

$$\sum X = 2\ 190 \quad \sum Y = 1\ 860 \quad \sum XY = 274\ 178$$

$$\sum X^2 = 328\ 940 \quad \sum Y^2 = 232\ 764 \quad \sum a^2 = 1\ 379.01$$

$$\hat{Y} = \underset{(5.185)}{82.453} + \underset{(2.650)}{0.284\ 57}X$$

$$R^2 = 0.350\ 75 \quad \bar{R}^2 = 0.300\ 81 \quad s = 10.299$$

R^2 较小，说明模型的拟合优度不够好。

(3) 省略工作表和计算过程。依据题意做数据表，见表 6—7。

表 6—7　　使用系数虚拟变量的数据表

年　份	初次能源需求量 Y	实际 GDP X	系数虚拟变量 D	含 D 的解释变量 DX
1965	100	100	0	0
1966	106	108	0	0
1967	115	117	0	0
1968	122	123	0	0
1969	129	132	0	0
1970	136	141	0	0
1971	141	145	0	0
1972	143	154	0	0
1973	114	150	1	150
1974	117	156	1	156
1975	121	161	1	161
1976	123	169	1	169
1977	129	174	1	174
1978	130	177	1	177
1979	134	183	1	183

$$\sum DX = 1\ 170 \qquad \sum (DX)^2 = 196\ 412$$

$$\sum X(DX) = 196\ 412 \qquad \sum Y(DX) = 145\ 598$$

$$\sum a^2 = 31.617\ 7$$

$$\hat{Y} = \underset{(4.469)}{17.095} + \underset{(28.163)}{0.838\ 64}X - \underset{(-22.614)}{0.199\ 18}DX$$

$$\bar{R}^2 = 0.982\ 63 \qquad s = 1.623\ 21$$

$\bar{R}^2$ 明显提高，估算出来的回归系数满足符号条件，而且均在1%水平显著。石油冲击前的系数为0.838 64，石油冲击后的系数为0.639 46，可见石油冲击后，经济增长模式向节能化方向转变。

第六章　练习题

1．表 6—8 列示了日本 1980—1995 年的 16 年间，某种植物的收获量 Y 与耕种面积 X 的变化。

(1) 对下面的一元回归模型进行 OLS 估算，并计算 t 值和决定系数 R^2。

$$Y = \alpha + \beta X + u$$

(2) 受 1991 年特大型台风 19 号所造成的风灾水灾的影响，这种植物严重歉收。引入临时虚拟变量 D，对下面的多元回归模型进行 OLS 估算，并计算 t 值和自由度调整后的决定系数 $\bar{R}^2$。

$$Y = \alpha + \beta_1 X + \beta_2 D + u$$

$$D = \begin{cases} 1, & 1991\text{ 年} \\ 0, & \text{其他年份} \end{cases}$$

表 6—8　　日本某种植物的收获量与耕种面积的关系　单位：1 000 吨，100 公顷

年份	收获量 Y	耕种面积 X
1980	94	94
1981	90	85
1982	86	75
1983	82	78
1984	87	75
1985	81	74
1986	81	74
1987	83	78
1988	84	84
1989	96	86
1990	90	85
1991	65	71
1992	78	68
1993	67	65
1994	66	61
1995	68	60

资料来源：农林水产省《作物统计》。

2. 表 6—9 列示了日本 1896—1905 年，政府支出 Y 与 GNPX 实际值的变化。

(1) 以 X 为横轴、Y 为纵轴，画出数据的散点图。

(2) 对下面的回归模型进行 OLS 估算，并计算 t 值与决定系数 R^2。

$$Y = \alpha + \beta X + u$$

(3) 设日俄战争时 (1904—1905 年) $D = 1$ 为临时虚拟变量，引入下列多元回归模型并进行估算。同时计算 t 值和自由度调整后的决定系数 $\bar{R}^2$。

$$Y = \alpha + \beta_1 X + \beta_2 D + u$$

表 6—9　　日本明治年间政府支出与 GNP 的关系　　单位：100 万日元

年份	政府支出 Y	GNP X
1896	459	5 773
1897	383	5 701
1898	421	5 907
1899	476	6 318
1900	538	6 232
1901	603	6 469
1902	584	6 358
1903	671	6 390
1904	1 428	7 084
1905	1 574	6 769

说明：1934—1936 年价格。

资料来源：大川一司等：《长期经济统计国民收入》，东洋经济新报社。

3. 表 6—10 中的季度数据反映的是日本服务消费支出 Y（实际值）的变化。同时利用例题 6—2 中国内家庭最终消费支出 X 的数据，回答以下问题：

(1) 对下面的回归模型进行 OLS 估算，并计算 t 值与决定系数 R^2。

$$Y = \alpha + \beta X + u$$

(2) 引入季度虚拟变量 D_1(第一季度)、D_2(第二季度)、D_3(第三季度)，对下面的多元回归模型进行 OLS 估算。同时计算 t 值和自由度调整后的决定系数 $\bar{R}^2$。

$$Y = \alpha + \beta_1 X + \beta_2 D_1 + \beta_3 D_2 + \beta_4 D_3 + u$$

表 6—10　　日本的服务消费支出　　单位：兆日元

年·季度	服务消费支出 Y	年·季度	服务消费支出 Y
1990 年1—3 月（1）	27.8	1992 年1—3 月（1）	29.4
4—6 月（2）	28.0	4—6 月（2）	29.8
7—9 月（3）	28.5	7—9 月（3）	30.3
10—12 月（4）	29.0	10—12 月（4）	30.7
1991 年1—3 月（1）	28.3	1993 年1—3 月（1）	29.9
4—6 月（2）	28.4	4—6 月（2）	29.9
7—9 月（3）	29.3	7—9 月（3）	30.7
10—12 月（4）	30.3	10—12 月（4）	31.7
		1994 年 1—3 月（1）	31.0

说明：1985 年价格，实际值。

资料来源：经济企划厅《国民经济计算年报》。

4*. 表 6—11 列示了 15 位工人每月额外劳动时间（加班时间）与性别、产业（制造业、非制造业）、职位（管理工作与非管理工作）、企业规模（大型企业、中型企业、小型企业）之间的关系。

（1）利用计量经济分析软件，对下面的多元回归模型进行 OLS 估算，并计算 t 值和自由度调整后的决定系数 $\bar{R}^2$。

$$Y = \alpha + \beta_1 S + \beta_2 I + \beta_3 M + \beta_4 F_1 + \beta_5 F_2 + u$$

性别　$S = \begin{cases} 1 & \text{男性} \\ 0 & \text{女性} \end{cases}$

产业　$I = \begin{cases} 1 & \text{制造业} \\ 0 & \text{非制造业} \end{cases}$

职位　$M = \begin{cases} 1 & \text{管理工作} \\ 0 & \text{非管理工作} \end{cases}$

企业规模(1)　$F_1 = \begin{cases} 1 & \text{大型企业} \\ 0 & \text{其他} \end{cases}$

企业规模(2)　$F_2 = \begin{cases} 1 & \text{中型企业} \\ 0 & \text{其他} \end{cases}$

(2) 试解释估算出的常数项的意义。

(3) 计算在中型规模制造业中工作的男性管理人员的加班时间。

表 6—11　　　　每月额外劳动时间与性别、产业、职位、企业规模的关系

每月额外劳动时间（小时）	性别	产业	职位	企业规模
4	女性	非制造业	管理工作	中型企业
7	女性	制造业	非管理工作	小型企业
8	女性	非制造业	管理工作	大型企业
11	女性	非制造业	非管理工作	大型企业
13	男性	非制造业	非管理工作	小型企业
14	男性	非制造业	管理工作	大型企业
15	男性	非制造业	非管理工作	中型企业
15	女性	制造业	非管理工作	大型企业
16	男性	制造业	管理工作	大型企业
17	男性	制造业	非管理工作	中型企业
17	男性	制造业	非管理工作	中型企业
18	男性	制造业	非管理工作	中型企业
18	男性	制造业	管理工作	大型企业
20	男性	非制造业	非管理工作	大型企业
21	男性	制造业	非管理工作	大型企业

5. 表 6—12 列示了某发展中国家汽车进口额 Y 与 GDPX（实际值）之间的关系。可以看出，该国自 1991 年国内生产汽车以后，汽车的进口额开始下降。

(1) 对①～④的回归模型进行 OLS 估算，并计算 t 值、决定系数 R^2 和自由度调整后的决定系数 $\bar{R}^2$。此处，将虚拟变量定义为：

$$D=\begin{cases}0 & 1985\text{—}1990\text{ 年}\\ 1 & 1991\text{—}1995\text{ 年}\end{cases}$$

① 一元回归模型。

$$Y=\alpha+\beta X+u$$

② 利用常数项虚拟构建的多元回归模型。

$$Y=\alpha+\beta_1 X+\beta_2 D+u$$

③ 利用系数虚拟构建的多元回归模型。

$$Y = \alpha + \beta_1 X + \beta_2 DX + u$$

④ 利用系数虚拟与常数项虚拟构建的多元回归模型。

$$Y = \alpha + \beta_1 X + \beta_2 DX + \beta_3 D + u$$

(2) 利用①的一元回归模型，以 1990 年为分界点（前期：1985—1990 年；后期：1991 年—1995 年），对结构变化进行 F 检验（Chow test）。

表 6—12　　汽车进口额与 GDP 的关系　　单位：100 万美元

年份	汽车进口额 Y	GDP X
1985	6.1	52
1986	6.5	54
1987	6.9	58
1988	7.0	61
1989	7.3	63
1990	7.6	69
1991	7.1	71
1992	6.8	76
1993	6.6	80
1994	6.5	82
1995	6.3	85

说明：1990 年价格，实际值。

第七章 系列相关

本章讨论时间系列回归分析中一个麻烦的问题，即系列相关。首先说明系列相关的含义和检验方法杜宾－沃特森比检验；然后简要介绍解决方法，如 Cochrane-Orcutt 法、基于 Prais-Winsten 变换的一般化最小二乘法。

1. 什么是系列相关

所谓系列相关（serial correlation），又称自我相关（autocorrelation），是在用时间系列数据进行回归分析时经常出现的问题，意指误差项之间存在着相互关系。尤其是在经济活动中，当期的活动经常对以后时期的活动有影响，容易出现系列相关。例如，在最频繁发生的一阶正的系列相关中，利用 OLS 计算的 t 值、F 值和决定系数偏大，原

本不显著的东西，被误认为显著。这样，如果无视系列相关进行 OLS 估算，估算出的结果虽然是无偏的，但是已经不再是 BLUE 了（参见第四章的补充 1）。

误差项 u 中存在一阶系列相关的情形可以表示如下：

$$u_t = \rho u_{t-1} + \varepsilon_t, -1 < \rho < 1 \quad (7\text{—}1)$$

$0 < \rho < 1$ → 一阶正的系列相关

$-1 < \rho < 0$ → 一阶负的系列相关

式中，ρ 称为自我相关系数，ε 是平均为 0、同方差、无系列相关的误差项。在一阶正的系列相关情形下，存在若 $u_{t-1}>0$，则 $u_t>0$；若 $u_{t-1}<0$，则 $u_t<0$ 趋势（见图 7—1）；在一阶负的系列相关情形下，存在若 $u_{t-1}>0$，则 $u_t<0$；若 $u_{t-1}<0$，则 $u_t>0$ 趋势（见图 7—2）。

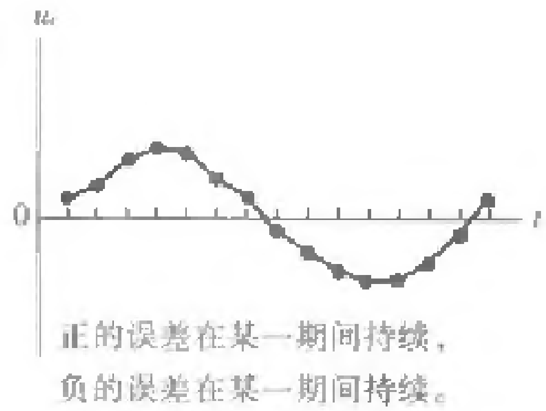


图 7—1　正的系列相关

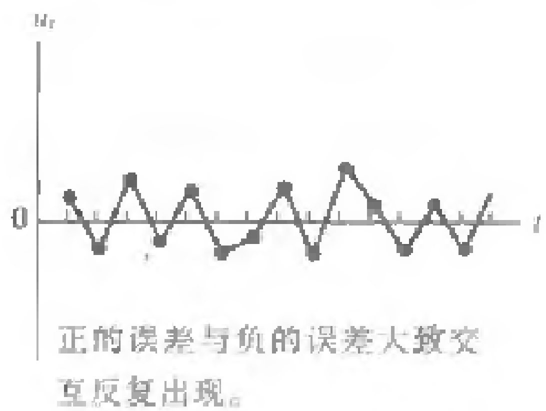


图 7—2　负的系列相关

误差项中出现系列相关的主要原因有以下五点：

(1) 模型中遗漏了重要的解释变量。

(2) 经济行为（消费、储蓄、投资、进口、出口）的惯性。

(3) 某种冲击（石油危机、战争等）造成的经济影响在该时期没有结束，还会波及到以后时期。

(4) 函数形式具体化的失败。

(5) 时间系列回归分析的时间单位越短（年→季度→月→周），越容易受到上一时期的影响而出现系列相关。

2. 杜宾－沃特森比

所谓杜宾－沃特森比（Durbin-Watoson ratio，DW），是用来检验是否存在一阶系列相关的统计量，设 OLS 的残差为 $\hat{u}_t$，可以定义如下：

$$DW = \frac{\sum_{t=2}^{n}(\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{n}\hat{u}_t^2} \tag{7—2}$$

$0 \leqslant DW \leqslant 4$

如果 DW 在样本数量足够大时（$n \geqslant 30$），可以用下式进行近似计算。

$$DW = 2(1 - \hat{\rho}) \tag{7—3}$$

这里，$\hat{\rho}$ 是 ρ 的估计值。

$$\hat{\rho} = \frac{\sum_{t=2}^{n}\hat{u}_t\hat{u}_{t-1}}{\sum_{t=2}^{n}\hat{u}_{t-1}^2} \tag{7—4}$$

如果 $\hat{\rho}=0$，则 $DW=2$；如果 $\hat{\rho}=1$，则 $DW=0$；如果 $\hat{\rho}=-1$，则 $DW=4$。就是说，误差项之间没有相关（即不存在系列相关）时，DW 大约为 2.0，误差项之间有正相关（正系列相关）时，DW 近似于 0，如果误差项之间有负相关（负系列相关）时，DW 大约为 4。

下面具体说明 DW 的检验方法。DW 检验根据备择假设 H_1 的建立，共有三种方式。

（1）出现一阶正的系列相关的情形（单侧检验）。

原假设	$H_0: \rho = 0$
备择假设	$H_1: \rho > 0$
$DW < d_L$	→放弃 H_0（存在一阶正的系列相关）
$d_L \leqslant DW \leqslant d_U$	→无法判定
$DW > d_U$	→不能放弃 H_0（不存在一阶正的系列相关）

(2) 出现负的系列相关的情形（单侧检验）。

原假设	$H_0: \rho = 0$
备择假设	$H_1: \rho < 0$
$DW > 4 - d_L$	→放弃 H_0（存在一阶负的系列相关）
$4 - d_U \leqslant DW \leqslant 4 - d_L$	→无法判定
$DW < 4 - d_U$	→不能放弃 H_0（不存在一阶负的系列相关）

(3) 出现一阶正或负的系列相关的情形（双侧检验）。

原假设	$H_0: \rho = 0$
备择假设	$H_1: \rho \neq 0$
$DW < d_L$	→放弃 H_0（存在一阶正的系列相关）
$d_L \leqslant DW \leqslant d_U$	→无法判定
$d_U < DW < 4 - d_U$	→不能放弃 H_0（不存在一阶系列相关）
$4 - d_U \leqslant DW \leqslant 4 - d_L$	→无法判定
$DW > 4 - d_L$	→放弃 H_0（存在一阶负的系列相关）

图 7—3 简要列示了 DW 检验中的各个领域，可将计算出的 DW 代入各区域进行检验。

图 7—3　DW 检验

DW 检验与 t 检验、F 检验不同，存在着无法判断、得不出结论的领域，d_L 与 d_U 是该领域下限与上限的临界值，它们取决于：样本数量 n；解释变量的数量 k；检验的显著性水平；单侧检验或者双侧检验。表 7—1 与表 7—2 用数字表的形式给出了各个单侧检验中显著水平为 5% 与 2.5% 的 d_L 与 d_U。

表 7—1 **DW** 单侧检验：显著水平 5%

n	$k=1$		$k=2$		$k=3$		$k=4$		$k=5$	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
15	1.08	1.36	0.95	1.54	0.82	1.75	0.69	1.97	0.56	2.21
20	1.20	1.41	1.10	1.54	1.00	1.68	0.90	1.83	0.79	1.99
25	1.29	1.45	1.21	1.55	1.12	1.66	1.04	1.77	0.95	1.89
30	1.35	1.49	1.28	1.57	1.21	1.65	1.14	1.74	1.07	1.83
35	1.40	1.52	1.34	1.58	1.28	1.65	1.22	1.73	1.16	1.80
40	1.44	1.54	1.39	1.60	1.34	1.66	1.29	1.72	1.23	1.79
45	1.48	1.57	1.43	1.62	1.38	1.67	1.34	1.72	1.29	1.78
50	1.50	1.59	1.46	1.63	1.42	1.67	1.38	1.72	1.34	1.77
60	1.55	1.62	1.51	1.65	1.48	1.69	1.44	1.73	1.41	1.77
70	1.58	1.64	1.55	1.67	1.52	1.70	1.49	1.74	1.46	1.77
80	1.61	1.66	1.59	1.69	1.56	1.72	1.53	1.74	1.51	1.77
90	1.63	1.68	1.61	1.70	1.59	1.73	1.57	1.75	1.54	1.78
100	1.65	1.69	1.63	1.72	1.61	1.74	1.59	1.76	1.57	1.78

说明：n 为样本数。
k 为除常数项的解释变量数。
本表与显著水平 10% 的双侧检验表相同。

表 7—2 **DW** 单侧检验：显著水平 2.5%

n	$k=1$		$k=2$		$k=3$		$k=4$		$k=5$	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
15	0.95	1.23	0.83	1.40	0.71	1.61	0.59	1.84	0.48	2.09
20	1.08	1.28	0.99	1.41	0.89	1.55	0.79	1.70	0.70	1.87
25	1.18	1.34	1.10	1.43	1.02	1.54	0.94	1.65	0.86	1.77
30	1.25	1.38	1.18	1.46	1.12	1.54	1.05	1.63	0.98	1.73
35	1.30	1.42	1.25	1.48	1.19	1.55	1.13	1.63	1.07	1.70
40	1.35	1.45	1.30	1.51	1.25	1.57	1.20	1.63	1.15	1.69
45	1.39	1.48	1.34	1.53	1.30	1.58	1.25	1.63	1.21	1.69
50	1.42	1.50	1.38	1.54	1.34	1.59	1.30	1.64	1.26	1.69
60	1.47	1.54	1.44	1.57	1.40	1.61	1.37	1.65	1.33	1.69
70	1.51	1.57	1.48	1.60	1.45	1.63	1.42	1.66	1.39	1.70
80	1.54	1.59	1.52	1.62	1.49	1.65	1.47	1.67	1.44	1.70
90	1.57	1.61	1.55	1.64	1.53	1.66	1.50	1.69	1.48	1.71
100	1.59	1.63	1.57	1.65	1.55	1.67	1.53	1.70	1.51	1.72

说明：n 为样本数。
k 为除常数项的解释变量数。
本表与显著水平 5% 的双侧检验表相同。

顺便指出，由于双侧检验的显著水平是单侧检验的2倍，因此，表7—1和表7—2也可以看做是双侧检验5%和10%显著水平的数字表。

[例题7—1]

对下面a项～f项的DW比，回答以下问题。n为样本数，k是解释变量数，ρ是自我相关系数。

(1) 关于a～c，对$H_0:\rho=0$、$H_1:\rho>0$按5%显著水平进行单侧检验。

(2) 关于d～f，对$H_0:\rho=0$、$H_1:\rho<0$按5%显著水平进行单侧检验。

(3) 关于a～f，对$H_0:\rho=0$、$H_1:\rho\neq0$按5%显著水平进行双侧检验。

a. $DW=0.92(n=15, k=1)$

b. $DW=1.60(n=40, k=3)$

c. $DW=1.81(n=90, k=5)$

d. $DW=2.75(n=20, k=2)$

e. $DW=2.54(n=70, k=4)$

f. $DW=2.27(n=100, k=1)$

[解答]

(1) 由于是5%显著水平的单侧检验，可以利用表7—1。

a. $DW(0.92)<d_L(1.80)$，放弃原假设$H_0:\rho=0$，选择备择假设$H_1:\rho>0$，也就是说，可以认为存在一阶正的系列相关。

b. $d_L(1.34)\leqslant DW(1.60)\leqslant d_U(1.66)$，这种情况无法进行假设检验，得不出结论。也就是说，一阶正的系列相关可能存在也可能不存在。

c. $DW(1.81)>d_U(1.78)$，原假设$H_0:\rho=0$不能放弃。也就是说，不存在一阶正的系列相关。

(2) 与(1)一样，由于是5%显著水平的单侧检验，可以利用表7—1。

d. $4-d_U(2.46)\leqslant DW(2.75)\leqslant 4-d_L(2.90)$，这种情况无法进行假设检验，得不出结论。也就是说，一阶负的系列相关可能存在也可能不存在。

e. $DW(2.54)>4-d_L(2.51)$，放弃原假设$H_0:\rho=0$，选择备择假设$H_1:\rho<0$，也就是说，可以认为存在一阶负的系列相关。

f. $DW(2.27)<4-d_U(2.31)$，原假设$H_0:\rho=0$不能放弃。也就是说，不存在一阶负的系列相关。

(3) 由于是5%显著水平的双侧检验，可以利用表7—2。

a. $DW(0.92)<d_L(0.95)$，放弃原假设$H_0:\rho=0$，选择备择假设$H_1:\rho\neq0$，

也就是说，可以认为存在一阶正的系列相关。

b. $d_U(1.57) < DW(1.60) < 4 - d_U(2.43)$，原假设 $H_0: \rho = 0$ 不能放弃。也就是说，不存在一阶系列相关。

c. $d_U(1.71) < DW(1.81) < 4 - d_U(2.29)$，原假设 $H_0: \rho = 0$ 不能放弃。也就是说，不存在一阶系列相关。

d. $4 - d_U(2.59) \leqslant DW(2.75) \leqslant 4 - d_L(3.0)$，这种情况无法进行假设检验，得不出结论。也就是说，一阶负的系列相关可能存在也可能不存在。

e. $4 - d_U(2.34) \leqslant DW(2.54) \leqslant 4 - d_L(2.58)$，这种情况无法进行假设检验，得不出结论。也就是说，一阶负的系列相关可能存在也可能不存在。

f. $d_U(1.63) < DW(2.27) < 4 - d_U(2.37)$，原假设 $H_0: \rho = 0$ 不能放弃。也就是说，不存在一阶系列相关。

[例题 7—2]

表 7—3 列示了 1970—1994 年 25 年间，日本工薪家庭实际消费支出 Y 与实际可支配收入 X 的变化。

(1) 对下面的回归模型进行 OLS 检验，并计算 t 值和决定系数 R^2。

$$Y_t = \alpha + \beta X_t + u_t$$

(2) 利用 (1) 的估计结果，计算 OLS 的残差，并求 DW。

(3) 对 (2) 的 DW，进行 $H_0: \rho = 0$、$H_1: \rho > 0$ 的单侧检验，显著水平为 5%。

表 7—3　　日本工薪家庭实际消费支出与实际可支配收入　　单位：1 000 日元

年　份	家庭实际消费支出 Y	家庭实际可支配收入 X	年　份	家庭实际消费支出 Y	家庭实际可支配收入 X
1970	239	300	1978	285	370
1971	248	311	1979	293	378
1972	258	329	1980	291	374
1973	272	351	1981	294	371
1974	268	354	1982	302	381
1975	280	364	1983	304	384
1976	279	360	1984	308	392
1977	282	366	1985	310	400

续前表

年　份	家庭实际消费支出 Y	家庭实际可支配收入 X	年　份	家庭实际消费支出 Y	家庭实际可支配收入 X
1986	312	403	1991	334	449
1987	314	411	1992	336	451
1988	324	428	1993	334	449
1989	326	434	1994	330	449
1990	332	441			

说明：1990 年价格。

资料来源：日本银行《经济统计年报》。

[解答]

(1) 省略计算过程。

$$\sum X_t = 9\,700 \qquad \sum Y_t = 7\,455 \qquad \sum X_tY_t = 2\,921\,268$$

$$\sum X_t^2 = 3\,808\,668 \qquad \sum Y_t^2 = 2\,241\,861$$

$$\sum \hat{u}_t^2 = 467.717$$

$$\hat{Y}_t = \underset{(6.136)}{50.875} + \underset{(30.008)}{0.637\,44}X_t$$

$$R^2 = 0.975\,1$$

(2) 根据下式求 OLS 残差 $\hat{u}_t$，并制作用于计算 DW 的工作表（见表 7—4）。

$$\hat{u}_t = Y_t - \hat{Y}_t = Y_t - 50.875 - 0.637\,44X_t$$

表 7—4　　　　DW 的计算

年　份	$\hat{u}_t = Y_t - \hat{Y}_t$	$\hat{u}_t - \hat{u}_{t-1}$	$\hat{u}_t^2$	$(\hat{u}_t - \hat{u}_{t-1})^2$
1970	−3.105 6	—	9.644 5	—
1971	−1.117 4	1.988 2	1.248 5	3.952 9
1972	−2.591 2	−1.473 9	6.714 5	2.172 3
1973	−2.614 8	−0.023 6	6.837 4	0.000 6
1974	−8.527 2	−5.912 3	72.712 3	34.955 4
1975	−2.901 5	5.625 6	8.418 8	31.647 7

续前表

年　份	$\hat{a}_t = Y_t - \hat{Y}_t$	$\hat{a}_t - \hat{a}_{t-1}$	$\hat{a}_t^2$	$(\hat{a}_t - \hat{a}_{t-1})^2$
1976	−1.351 8	1.549 7	1.827 3	2.401 7
1977	−2.176 4	−0.824 6	4.736 7	0.680 0
1978	−1.726 1	0.450 3	2.979 6	0.202 7
1979	1.174 4	2.900 5	1.379 1	8.412 9
1980	1.724 1	0.549 7	2.972 6	0.302 2
1981	6.636 4	4.912 3	44.042 1	24.130 8
1982	8.262 1	1.625 6	68.261 6	2.642 7
1983	8.349 7	0.087 7	69.718 3	0.007 7
1984	7.250 3	−1.099 5	52.566 2	1.208 9
1985	4.150 8	−3.099 5	17.228 8	9.606 9
1986	4.238 4	0.087 7	17.964 4	0.007 7
1987	1.139 0	−3.099 5	1.297 2	9.606 9
1988	0.302 5	−0.836 4	0.091 5	0.699 6
1989	−1.522 1	−1.824 6	2.316 8	3.329 2
1990	0.015 9	1.537 9	0.000 3	2.365 3
1991	−3.083 6	−3.099 5	9.508 9	9.606 9
1992	−2.358 5	0.725 1	5.562 6	0.525 8
1993	−3.083 6	−0.725 1	9.508 9	0.525 8
1994	−7.083 6	−4.000 0	50.178 0	16.000 0
合计	−	−	467.717	164.993

$$\uparrow \sum_{t=1}^{25}\hat{a}_t^2 \qquad\qquad \uparrow \sum_{t=2}^{25}(\hat{a}_t - \hat{a}_{t-1})^2$$

将表 7—4 的计算结果代入（7—2）式，求 DW。

$$DW = \frac{\sum_{t=2}^{25}(\hat{a}_t - \hat{a}_{t-1})^2}{\sum_{t=1}^{25}\hat{a}_t^2} = \frac{164.993}{467.717}$$

$$= 0.352\,8$$

（3）由于是 5% 显著水平的单侧检验，可以利用表 7—1。因为 $n=25$，$k=1$，

$DW(0.352\,8) < d_L(1.29)$，放弃原假设 $H_0: \rho = 0$，选择备择假设 $H_1: \rho > 0$，也就是说，可以认为存在一阶正的系列相关。

[补充] 杜宾 h 统计量

在诸如 $Y_t = \alpha + \beta_1 X_t + \beta_2 Y_{t-1} + u_t$ 这样的回归模型的解释变量中，如果引入了滞后的被解释变量（Y_{t-1}），存在 DW 偏向 2 这样的缺陷。也就是说，尽管存在一阶系列相关，也可能做出相反的错误判断。在这种情况下，就不能利用 DW 检验，而要根据杜宾 h 统计量进行检验。详细情况请参见山本（1995）。

3. Cochrane-Orcutt(CO) 法

根据 DW 检验法，如果检验出一阶系列相关，应该在增加新的解释变量、改变模型的函数形式或受重大意外冲击的影响时，引入虚拟变量，努力设法消除系列相关。但是，如果利用这些方法还不能消除系列相关时，就应该利用参数估计法来代替 OLS 法。以下是三个代表性的估计方法。

(1) Cochrane-Orcutt(CO) 法。

(2) 基于 Prais-Winsten(PW) 变换的一般化最小二乘法（Generalized Least Squares，GLS）。

(3) 极大似然法（Maximum Likelihood，ML）。

本书为方便计量经济学初学者理解，主要介绍（1）与（2）。首先，说明 CO 变换，然后就基于该变换的模型，介绍具体的估计步骤。

设有以下一阶系列相关模型。

$$Y_t = \alpha + \beta X_t + u_t \tag{7—5}$$

$$t = 1, 2, \cdots, n$$

$$u_t = \rho u_{t-1} + \varepsilon_t \tag{7—6}$$

式中，ε_t 是平均为 0、同方差、没有自我相关的误差项。将（7—5）式用上一时期的形式表示为

$$Y_{t-1} = \alpha + \beta X_{t-1} + u_{t-1} \tag{7—7}$$

$$t = 1, 2, \cdots, n$$

两边同乘以 ρ，得

$$\rho Y_{t-1} = \rho\alpha + \rho\beta X_{t-1} + \rho u_{t-1} \tag{7—8}$$

$t=1,2,\cdots,n$

(7—5) 式减去 (7—8) 式，得

$$Y_t-\rho Y_{t-1}=\alpha-\rho\alpha+\beta X_t-\rho\beta X_{t-1}+u_t-\rho\mu_{t-1}$$

$$Y_t-\rho Y_{t-1}=\alpha(1-\rho)+\beta(X_t-\rho X_{t-1})+\varepsilon_t \tag{7—9}$$

该式的误差项 ε_t 中没有系列相关。这里再将 (7—9) 式用新的变量和回归系数进行替换，即

$$Y_t^*=Y_t-\rho Y_{t-1} \tag{7—10}$$

$$X_t^*=X_t-\rho X_{t-1} \tag{7—11}$$

$$\alpha^*=\alpha(1-\rho) \tag{7—12}$$

$$\beta^*=\beta \tag{7—13}$$

$t=2,3,\cdots,n$

则 (7—9) 式为

$$Y_t^*=\alpha^*+\beta^*X_t^*+\varepsilon_t \tag{7—14}$$

$t=2,3,\cdots,n$

这就是 CO 变换，CO 法最终是对 (7—14) 式进行 OLS 估计。

在实际中，运用 CO 法对模型进行估计，要采取以下步骤。

步骤一：对 (7—5) 式进行 OLS 估计，求残差 $\hat{u}_t$。

步骤二：对 (7—6) 式进行没有常数项的 OLS 估计，求 ρ 的估计值$\hat{\rho}$。具体的计算公式为

$$\hat{\rho}=\frac{\sum_{t=2}^{n}\hat{u}_t\hat{u}_{t-1}}{\sum_{t=2}^{n}\hat{u}_{t-1}^2} \tag{7—15}$$

此外，ρ 的估计方法还有很多，例如蓑谷 (1992) 介绍了五种方法。

步骤三：将 $\hat{\rho}$ 代人 (7—10) 式和 (7—11) 式，编制 Y_t^* 与 X_t^* 的数据。

步骤四：将 OLS 用于 (7—14) 式，估计 α^* 与 β^*。

步骤五：利用下面的公式，求 α、β 新的估计值 $\tilde{\alpha}$、$\tilde{\beta}$ 。

$$\tilde{\alpha}=\frac{\alpha^*}{1-\hat{\rho}} \tag{7—16}$$

$$\tilde{\beta}=\beta^* \tag{7—17}$$

一般的 CO 法到这一步就可以结束了，在迭代方法（iterative method）的情况下，还要继续以下的步骤。根据迭代计算的 Cochrane-Orcutt 法的步骤有：

步骤六：根据以下的公式，计算新的残差 $\tilde{\varepsilon}_t$。

$$\tilde{\varepsilon}_t = Y_t - \tilde{\alpha} - \tilde{\beta}X_t \qquad (7—18)$$

$$t = 1, 2, \cdots, n$$

步骤七：与步骤二一样，利用新的残差 $\tilde{\varepsilon}_t$，求 ρ 的新的估计值 $\tilde{\rho}$。

步骤八：考察步骤七计算出来的 $\tilde{\rho}$，是否满足下面的条件。

$$\left|\frac{\tilde{\rho} - \hat{\rho}}{\hat{\rho}}\right| \leqslant \delta \qquad (7—19)$$

式中，δ 为判断收敛的条件值。例如 $\delta = 0.005$。

如果满足上述条件，迭代计算就结束了，采取步骤五的估计结果。如果上述条件不能满足，再回到步骤三，进行反复迭代计算，直到上述条件满足为止。

[例题 7—3]

根据例题 7—2 的数据和模型（$Y_t = \alpha + \beta X_t + u_t$），回答下面的问题。

(1) 利用 Cochrane-Orcutt（CO）法，对模型进行估计，并计算 t 值、R^2 以及 DW。

(2) 利用迭代计算 CO 法（迭代两次），对模型进行估计，并计算 t 值、R^2 以及 DW。

[解答]

(1) 步骤一：

残差 $\hat{u}_t$ 依据例题 7—2 中的表 7—4。

步骤二：

根据（7—15）式，求自我相关系数的估计值 $\hat{\rho}$，

$$\hat{\rho} = \frac{\sum_{t=2}^{25} \hat{u}_t \hat{u}_{t-1}}{\sum_{t=2}^{25} \hat{u}_{t-1}{}^{①}} = \frac{355.309\ 1}{417.538\ 7} = 0.850\ 961$$

① $\sum_{t=2}^{25} \hat{u}_{t-1}$ 应为 $\sum_{t=2}^{25} \hat{u}_{t-1}^2$，原文有误。—— 编者注

步骤三：

将 $\rho=0.850\ 961$ 代入（7—10）式和（7—11）式，根据 CO 变换制作新的变量 Y_t^* 和 X_t^* 的数据。

$$Y_t^* = Y_t - 0.850\ 961\, Y_{t-1}$$

$$X_t^* = X_t - 0.850\ 961\, X_{t-1}$$

表 7—5 为新的数据，1970 年的数据丢失（这是 CO 法的缺点），$n=24$。

表 7—5　　基于 CO 变换的 Y_t^* 和 X_t^* 数据

t 年份	Y_t^*（消费支出）	X_t^*（可支配收入）	t 年份	Y_t^*（消费支出）	X_t^*（可支配收入）
1970	—	—	1983	47.009	59.783
1971	44.620	55.711	1984	49.307	65.230
1972	46.961	64.351	1985	47.904	66.423
1973	52.452	71.033	1986	48.202	62.615
1974	36.538	55.312	1987	48.500	68.062
1975	51.942	62.759	1988	56.798	78.255
1976	40.730	50.250	1989	50.288	69.788
1977	44.581	59.654	1990	54.586	71.682
1978	45.028	58.548	1991	51.480	73.726
1979	50.476	63.144	1992	51.779	68.918
1980	41.668	52.336	1993	48.077	65.216
1981	46.370	52.740	1994	45.779	66.918
1982	51.817	65.293			

说明：舍去小数点后第四位。但是计算中，四位以后也使用。

步骤四：

利用步骤三制作的 Y_t^* 和 X_t^* 数据（见表 7—5），对（7—14）式进行 OLS 估计。

$$\underset{(2.918)}{Y_t^* = 13.973} + \underset{(7.155)}{0.535\ 13X_t^*}$$

$$R^2=0.699\ 4 \qquad DW=2.378$$

可以看出 DW 有所改善，误差项中不存在系列相关。此外，虽然 t 值在 1% 水平显著，但是小于例题 7—2 的 OLS 估计结果。同样，R^2 也变小了。

步骤五：

根据（7—16）式、（7—17）式求 α 与 β 的估计值 $\tilde{\alpha}$、$\tilde{\beta}$。

$$\tilde{\alpha} = \frac{\alpha^*}{1-\hat{\rho}} = 93.756$$

$$\tilde{\beta} = \beta^* = 0.535\ 13$$

据此，整理 CO 法的估计结果如下。R^2 是由下面的模型计算出来的。

$$\underset{}{Y_t} = \underset{(2.918)}{93.756} + \underset{(7.155)}{0.535\ 13}X_t$$

$$R^2 = 0.990\ 6 \qquad DW = 2.378$$

（2）步骤六：

由下式求新的残差 $\tilde{\varepsilon}_t$。

$$\tilde{\varepsilon}_t = Y_t - 93.756 - 0.535\ 125X_t$$

$$t = 1, 2, \cdots, n$$

步骤七：

利用步骤六求出的新的残差 $\tilde{\varepsilon}_t$，根据（7—15）式，计算 ρ 的新的估计值 $\hat{\rho}$。

$$\hat{\rho} = \frac{\sum_{t=2}^{25}\tilde{\varepsilon}_t\tilde{\varepsilon}_{t-1}}{\sum_{t=2}^{25}\tilde{\varepsilon}_{t-1}^2} = \frac{980.157\ 5}{1\ 176.831\ 0} = 0.832\ 879$$

步骤八：

在本例中不需要，因此跳过步骤八，回到步骤三。

步骤三：

将 $\hat{\rho} = 0.832\ 879$ 代入（7—10）式和（7—11），制作新变量 Y_t^{**} 和 X_t^{**} 的数据。

$$Y_t^{**} = Y_t - 0.832\ 879Y_{t-1}$$

$$X_t^{**} = X_t - 0.832\ 879X_{t-1}$$

表 7—6 是制作的数据，其中 1970 年的数据丢失，这是 CO 法的缺点。

表 7—6　　基于 CO 变换的 Y_t^{**} 和 X_t^{**} 数据

（迭代次数＝2）

t 年份	Y_t^{**}（消费支出）	X_t^{**}（可支配收入）	t 年份	Y_t^* ①（消费支出）	X_t^* ②（可支配收入）
1970	-	-	1983	52.470	66.673
1971	48.941	61.136	1984	54.804	72.174
1972	51.446	69.974	1985	53.473	73.511
1973	57.117	76.982	1986	53.807	69.848
1974	41.456	61.659	1987	54.141	75.349
1975	56.788	69.160	1988	62.476	85.686
1976	45.793	56.832	1989	56.147	77.527
1977	49.626	66.163	1990	60.481	79.530
1978	50.128	65.166	1991	57.484	81.700
1979	55.629	69.834	1992	57.818	77.037
1980	46.966	59.171	1993	54.152	73.371
1981	51.632	59.503	1994	51.818	75.37
1982	57.133	72.002			

说明：舍去小数点后第四位。但是计算中，四位以后也使用。

①② Y_t^* 应为 Y_t^{**}；X_t^* 应为 X_t^{**}。原文有误。——编者注

步骤四：

利用表 7—6 的数据，对模型进行 OLS 估计。

$$Y_t^{**} = \underset{(3.073)}{15.387} + \underset{(7.634)}{0.538\,30}X_t^{**}$$

$$R^2 = 0.726\,0 \quad DW = 2.330$$

步骤五：

反复计算的 CO 法最终估计结果整理如下：

$$Y_t = \underset{(3.073)}{92.073} + \underset{(7.634)}{0.538\,30}X_t$$

$$R^2 = 0.990\,6 \quad DW = 2.330$$

4. 基于 Prais-Winsten(PW)变换的一般化最小二乘法

基于 PW 变换的一般化最小二乘法与 CO 法虽然非常相似，但也存在几点差异：(1) 编制第一期（t 期）的变换数据（CO 法中没有第一期数据）；(2) 编制常数项数据；(3) 估计方法采用“没有常数项的 OLS”。一般说来，基于 PW 变换的一般化最小二乘法与 CO 法相比，精确度更高，是一种很好的估计方法。

基于 PW 变换的一般化最小二乘法的步骤如下所述。

步骤一
步骤二 } 与 CO 法相同。

步骤三：

利用 $\hat{\rho}$，$t=1$ 期与 $t=2,\cdots,n$ 期的解释变量 Y_t^*，常数项 C_t^* 以及解释变量 X_t^* 的数据，由下面的 PW 变换来制作。

$t=1$ 时，

$$Y_1^* = \sqrt{1-\hat{\rho}^2}\,Y_1 \tag{7—20}$$

$$C_1^* = \sqrt{1-\hat{\rho}^2} \tag{7—21}$$

$$X_1^* = \sqrt{1-\hat{\rho}^2}\,X_1 \tag{7—22}$$

$t=2,\cdots,n$ 时，

$$Y_t^* = Y_t - \hat{\rho}Y_{t-1} \tag{7—23}$$

$$C_t^* = 1-\hat{\rho} \tag{7—24}$$

$$X_t^* = X_t - \hat{\rho}X_{t-1} \tag{7—25}$$

步骤四：

对下式进行无常数项 OLS 估计，并求 α 和 β。

$$Y_t^* = \alpha C_t^* + \beta X_t^* + u_t^* \tag{7—26}$$

到这一步虽然结束了，但是在使用迭代法时，须按照和 CO 法相同的步骤来进行。

[例题 7—4]

利用基于 Prais-Winsten 变换的一般化最小二乘法，估计例题 7—2 的模型。

[解答]

步骤一 | 步骤二 | 与 CO 法相同，$\rho=0.850\ 961$

步骤三：

将 $\rho=0.850\ 961$ 代入（7—20）式至（7—25）式中，制作基于 PW 变换的新变量 Y_t^*、C_t^* 和 X_t^* 的数据见表 7—7。

表 7—7　　基于 PW 变换的 Y_t^*、C_t^*、X_t^* 数据

t 年	Y_t^*（消费）	C_t^*（常数项）	X_t^*（收入）	t 年	Y_t^*（消费）	C_t^*（常数项）	X_t^*（收入）
1970	125.529	0.525 22	157.568	1983	47.009	0.149 03	59.783
1971	44.620	0.149 03	55.711	1984	49.307	0.149 03	65.230
1972	46.961	0.149 03	64.351	1985	47.904	0.149 03	66.423
1973	52.452	0.149 03	71.033	1986	48.202	0.149 03	62.615
1974	36.538	0.149 03	55.312	1987	48.500	0.149 03	68.062
1975	51.942	0.149 03	62.759	1988	56.798	0.149 03	78.255
1976	40.730	0.149 03	50.250	1989	50.288	0.149 03	69.788
1977	44.581	0.149 03	59.654	1990	54.586	0.149 03	71.682
1978	45.028	0.149 03	58.548	1991	51.480	0.149 03	73.726
1979	50.476	0.149 03	63.144	1992	51.779	0.149 03	68.918
1980	41.668	0.149 03	52.336	1993	48.077	0.149 03	65.216
1981	46.370	0.149 03	52.740	1994	45.779	0.149 03	66.918
1982	51.817	0.149 03	69.293				

说明：Y_t^*、X_t^* 舍去小数点后第四位；C_t^* 舍去小数点后第六位。但是计算时小数点四位或六位以下也使用。

步骤四：

利用步骤三制作的 Y_t^*、C_t^*、X_t^* 数据（见表 7—7），对（7—26）式进行无常数项 OLS 估计。

$$Y_t^*=55.468C_t^*+0.621\ 31X_t^*$$
$$(3.503)\qquad(15.314)$$

$$\bar{R}^2=0.975\ 1\qquad DW=2.007$$

关于无常数项 OLS，参见蓑谷（1997a）。

[补充]

作为参考，这里也介绍一下根据极大似然法估计的结果（利用 TSP）。

$$Y_t = \underset{(3.658)}{55.161} + \underset{(16.102)}{0.622\,56}X_t$$

$$\bar{R}^2 = 0.977\,25 \qquad DW = 1.958\,5$$

极大似然法和基于 PW 的一般化最小二乘法一样，在对具有一阶系列相关的模型进行估计时，是非常优良的方法。这里可以看出，两者的估计结果非常相近。另一方面，例题 7—3 中 CO 法的估计结果和基于 PW 的一般化最小二乘法以及极大似然法的估计结果相差很大，因此带来了一些问题。顺便指出，蓑谷（1992）认为，CO 法是有问题的估计法，最好不要采用。

第七章 练习题

1. 表 7—8 是日本 1971—1990 年的 20 年间，税收（国税 + 地税）T 与国民生产总值 Y 的数据。

(1) 对下面的回归模型进行 OLS 估计，并计算 t 值、R^2 以及 DW。

$$T_t = \alpha + \beta Y_t + u_t$$

(2) 对 (1) 中计算出来的 DW 进行 $H_0: \rho = 0$、$H_1: \rho > 0$、显著水平为 5% 的单侧检验。

(3) 对 (1) 中的税收函数，按下列ⓐ～ⓒ的方法进行估计，并计算 t 值、R^2 以及 DW。

ⓐ Cochrane-Orcutt 法。

ⓑ 迭代计算的 Cochrane-Orcutt 法（迭代次数为 2）。

ⓒ 基于 PW 变换的一般化最小二乘法。

表 7—8　日本的税收与国民生产总值　单位：兆日元

年度 t	税收 T	国民生产总值 Y	年度 t	税收 T	国民生产总值 Y
1971	27.9	181.9	1981	50.9	277.4
1972	31.6	198.3	1982	53.2	287.2
1973	36.6	207.7	1983	55.9	295.8
1974	36.0	207.3	1984	58.9	309.1
1975	32.1	215.6	1985	62.2	324.0
1976	34.6	224.3	1986	66.2	333.3
1977	36.4	235.0	1987	73.6	349.8
1978	42.0	247.1	1988	80.4	370.6
1979	45.1	260.6	1989	84.9	387.5
1980	48.5	268.8	1990	90.0	407.2

说明：1985 年价格，实际值。年度为会计年度。

资料来源：经济企划厅《国民经济计算年报》。

2. 利用第三章的练习题 1（94 页），根据美国个人实际可支配收入 X 与个人实际消费支出 Y（1961—1995 年）的数据，回答以下问题。

(1) 对下面的回归模型消费函数进行 OLS 估计，并计算 t 值、R^2 以及 DW。

$$Y_t = \alpha + \beta X_t + u$$

(2) 对 (1) 中计算出来的 DW 进行 $H_0: \rho = 0$、$H_1: \rho \neq 0$、显著水平为 5% 的双侧检验。

(3) 根据下列ⓐ～ⓒ的方法，对 (1) 中的消费函数用 Cochrane-Orcutt 法进行估计，并计算 t 值、R^2 以及 DW。

ⓐ 没有迭代计算的 CO 法。

ⓑ 迭代计算法。

判断收敛的条件值 → $\left|\dfrac{\tilde{\rho} - \rho}{\rho}\right| \leqslant 0.001$

ⓒ 迭代计算法。

判断收敛的条件值 → $\left|\dfrac{\tilde{\rho} - \rho}{\rho}\right| \leqslant 0.000\,1$

(4) 对 (1) 中的消费函数，进行基于 PW 变换的一般化最小二乘法估计，并计算 t 值、R^2 以及 DW。

第八章　联立方程模型

1. 联立方程模型

联立方程模型（multi-equations model），又称同时方程模型（simultaneous equations model），它是利用多个方程表达经济变量之间复杂关系的模型。例如，让我们设想以下苹果的供给与需求关系的模型。

（需求函数）　$Q_t = \alpha_0 + \alpha_1 P_t + \alpha_2 Y_t + u_{1t}$ (8—1)

（供给函数）　$Q_t = \beta_0 + \beta_1 P_t + \beta_2 T_t + u_{2t}$ (8—2)

（内生变量）　Q_t：苹果的数量

P_t：苹果的市场价格

（外生变量）　Y_t：需求者收入

T_t：日照时间

在这个联立方程模型中，Q_t 与 P_t 的取值是由模型内

部相互关系所决定的，因此称为内生变量（endogenous variable），而 Y_t 与 T_t 的取值是由模型外部所决定的，因此称为外生变量（exogenous variable）。外生变量影响内生变量，但是外生变量不受内生变量的影响。

此外，还有一些在上述模型中没有表示的变量，例如，Q_{t-1}、P_{t-1} 这样的内生变量的滞后变量（lagged variable），由于这些变量的值在 t 期中是被决定的，因此被称为先决内生变量（pre-determined endogenous variable）。这样，外生变量与先决内生变量合在一起称为先决变量（pre-determined variable），见图 8—1。

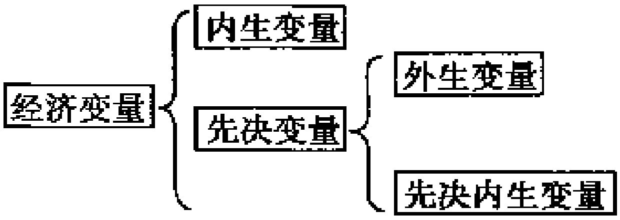


图 8—1 经济变量的类型

内生变量与外生变量的区分并没有绝对的标准，往往是根据分析的目的或者模型涵盖的经济活动范围，由分析者自己决定。但是，需要注意的是，变量的这种区分也是基于经济理论的一种常识。

2. 结构型与诱导型

所谓结构型（structural form，S.F.），是按原样记述由经济理论推导出来的经济变量之间的关系，反映了经济的结构，所以称为结构型。上述苹果的供求模型（8—1）式、（8—2）式就属于结构型。结构型中各种关系式称为结构方程式（structural equation）。

所谓诱导型（reduced form，R.F.），是指内生变量仅由外生变量、先决内生变量、误差项来表现，由于它是由结构型推导出来的，因此称为诱导型。诱导型中的各种关系式称为诱导方程式（reduced equation），只有内生变量需要设定。

结构型与诱导型的分类也没有绝对的标准，而是根据分析的目的、范围，由分析者自己决定。

下面，我们从结构方程式（8—1）、（8—2），推导关于内生变量 Q_t 与 P_t 的诱导型方程式。

$$Q_t = \pi_{10} + \pi_{11} Y_t + \pi_{12} T_t + v_{1t} \tag{8—3}$$

$$P_t = \pi_{20} + \pi_{21} Y_t + \pi_{22} T_t + v_{2t} \tag{8—4}$$

式中，
$$\pi_{10} = \frac{\alpha_1\beta_0 - \alpha_0\beta_1}{\alpha_1 - \beta_1} \qquad \pi_{11} = \frac{-\alpha_2\beta_1}{\alpha_1 - \beta_1} \qquad \pi_{12} = \frac{\alpha_1\beta_2}{\alpha_1 - \beta_1}$$
$$\pi_{20} = \frac{-\alpha_0 + \beta_0}{\alpha_1 - \beta_1} \qquad \pi_{21} = \frac{-\alpha_2}{\alpha_1 - \beta_1} \qquad \pi_{22} = \frac{\beta_2}{\alpha_1 - \beta_1}$$
$$v_{1t} = \frac{\alpha_1 u_{2t} - \beta_1 u_{1t}}{\alpha_1 - \beta_1} \qquad v_{2t} = \frac{-u_{1t} + u_{2t}}{\alpha_1 - \beta_1}$$

诱导方程式的参数 π_{10}、π_{11}、π_{12}、π_{20}、π_{21}、π_{22}称为诱导型参数，结构方程式的参数 α_0、α_1、α_2、β_0、β_1、β_2 称为结构参数。

3. 间接最小二乘法

如果对 (8—1)、(8—2) 这样的结构方程式直接应用普通最小二乘法 (OLS)，估计出来的参数会出现偏差，理想的统计性质无偏性 (unbiasedness)、一致性 (consistency) 均得不到保证。这种偏差叫做联立方程偏差 (同时方程偏差)，由于它是哈维莫 (Haavelmo，1989 年诺贝尔经济学奖得主) 首先提出的，因此又称为“哈维莫偏差” (Haavelmo bias)。

联立方程偏差产生的原因在于，由于解释变量中包含内生变量，解释变量与误差项之间会出现相关 (通常，OLS 中假设解释变量与误差项之间不相关。参见第四章[补充 1]之(4)。)

因此，如果只用外生变量、先决内生变量来解释左边的内生变量，也就是说只推导诱导方程式，并且对其使用 OLS 估计，联立方程偏差不会发生，问题可以得到解决。所谓间接最小二乘法 (indirect least squares method，ILS) 就是根据这种思想来估计联立方程式的方法。间接最小二乘法的步骤如下所示。

步骤一：由结构型推导诱导型。

步骤二：对推导出来的各诱导方程式应用 OLS 分析，并估计诱导型参数。

步骤三：根据诱导型参数，对原先的结构参数进行逆运算。

上述这种利用 OLS 估计出来的诱导型参数、间接地计算结构参数的方法，就称为间接最小二乘法。估计出来的诱导型参数，同时满足无偏性与一致性。与此相应的结构参数，不满足无偏性，只满足一致性。

上述步骤三中，根据诱导型参数不仅限于计算一个结构参数，有时还要计算两个以上的参数 (过剩识别的情况)，这是间接最小二乘法的一个较大的弱点。也

就是说，间接最小二乘法，只是在研究对象结构方程存在适度识别（参见下一节）的情形下，才有可能利用，因此在实际运用中存在一定的局限性。

[补充 1]

一致性：所谓一致性指的是这样一种性质，即样本数越大，估计出来的参数越接近真值。

无偏性：所谓无偏性指的是这样一种性质，即用同样的方法、不同的数据对同样的参数反复估计，估计出来的参数的平均值应该等于真值。

[例题 8—1]

利用表 8—1 中的数据，对苹果的供求模型（8—1）、（8—2）进行间接最小二乘法估计。

表 8—1　　苹果的供求模型数据

年序 t	数量（千个） Q	市场价格（日元） P	需求者收入（万日元） Y	日平均日照时间（小时） T
1	57	78	28	7.0
2	55	96	29	4.1
3	66	87	32	7.2
4	65	98	33	5.4
5	71	104	35	5.8
6	74	105	36	6.7
7	71	110	36	5.0
8	77	113	38	6.3

[解答]

步骤一：

由结构方程式（8—1）、（8—2）推导诱导方程式（8—3）、（8—4）。

步骤二：

对诱导方程式（8—3）、（8—4）应用 OLS 法，并估计诱导型参数。

$$\hat{Q}_t = -13.1105 + 2.13142\,Y_t + 1.51150\,T_t$$
$$(-7.274)\quad(44.618)\quad(9.546)$$
$$\bar{R}^2 = 0.9968 \quad DW = 2.004$$

$$\hat{P}_t = 25.1196 + 3.06024\,Y_t - 4.77983\,T_t$$
$$(3.156)\quad(14.506)\quad(-6.836)$$
$$\bar{R}^2 = 0.9720 \quad DW = 2.080$$

步骤三：

由估计出的诱导参数，求结构参数 $\hat{\alpha}_0$、$\hat{\alpha}_1$、$\hat{\alpha}_2$、$\hat{\beta}_0$、$\hat{\beta}_1$、$\hat{\beta}_2$。

$$\hat{\alpha}_0 = \hat{\pi}_{10} = \frac{\hat{\pi}_{12}\hat{\pi}_{20}}{\hat{\pi}_{22}} = -5.167\ 1 \quad \hat{\alpha}_1 = \frac{\hat{\pi}_{12}}{\hat{\pi}_{22}} = -0.316\ 22$$

$$\hat{\alpha}_2 = \frac{\hat{\pi}_{12}\hat{\pi}_{21}}{\hat{\pi}_{22}} - \hat{\pi}_{11} = 3.099\ 1$$

$$\hat{\beta}_0 = \hat{\pi}_{10} - \frac{\hat{\pi}_{11}\hat{\pi}_{20}}{\hat{\pi}_{21}} = -30.606 \quad \hat{\beta}_1 = \frac{\hat{\pi}_{11}}{\hat{\pi}_{21}} = 0.696\ 49$$

$$\hat{\beta}_2 = \hat{\pi}_{12} - \frac{\hat{\pi}_{11}\hat{\pi}_{22}}{\hat{\pi}_{21}} = 4.840\ 6$$

整理间接最小二乘法估计的结果，得到如下的苹果供求模型（结构型）。

需求函数：$\hat{Q}_t = -5.167\ 1 \underbrace{-0.316\ 22}_{负} P_t + \underbrace{3.099\ 1}_{正} Y_t$

供给函数：$\hat{Q}_t = -30.606 \underbrace{+696\ 49}_{正} P_t + \underbrace{4.840\ 6}_{正} T_t$

估计出的结构参数的符号条件也得到满足，可见是良好的估计结果。

[例题 8—2]

以下我们给出一个模型，将家庭的全部消费分为南瓜消费（P_1, Q_1）和其他消费（P_1, Q_2）两大类型。

贝努利-拉普拉斯型效用函数：

$$U = b_1 \log(a_1 + Q_1) + b_2 \log(a_2 + Q_2) \tag{8—5}$$

收支等式：

$$Y = P_1 Q_1 + P_2 Q_2 \tag{8—6}$$

式中，U——效用指标；

Q_1——每户南瓜年均消费量；

Q_2——其他商品年均消费量；

P_1——南瓜价格；

P_2——其他商品价格（消费物价指数）；

Y——每户年均消费支出；

a_1、a_2、b_1、b_2——结构参数。

(1) 求各商品的边际效用，并推导边际效用等式（效用最大化的一阶条件）。

(2) 根据边际效用等式和收支等式，推导相当于诱导方程式的南瓜需求函数。

(3) 对 (2) 中推导出的南瓜需求函数，利用表 8—2 日本的数据（1980—1993 年），进行 OLS 估计。

(4) 设正规化（normalize）$b_1+b_2=1$，根据 (3) 中估计出来的诱导型参数，求结构参数 a_1、a_2、b_1、b_2。

(5) 根据 (3) 中估计出来的需求函数，求南瓜消费量的理论值 $\hat{Q}_t$，并将其与实际值 Q_t 一道画出图形。

表 8—2　日本每户南瓜的年均消费量及其价格

年份	南瓜消费量（100 克）Q	南瓜价格（日元/100 克）P_1	消费者物价指数（1990 年为 100）P_2	户均年消费支出（日元）Y
1980	46.81	26.94	81.7	2 767 000
1981	48.56	25.71	85.6	2 880 000
1982	52.42	23.33	88.0	3 038 000
1983	44.87	30.31	89.6	3 114 000
1984	51.83	26.37	91.7	3 196 000
1985	54.21	26.43	93.5	3 277 000
1986	52.39	27.73	94.1	3 316 000
1987	55.18	27.03	94.2	3 371 000
1988	58.82	28.37	94.9	3 493 000
1989	57.12	26.46	97.0	3 592 000
1990	56.90	28.90	100.0	3 734 000
1991	54.27	35.92	103.3	3 925 000
1992	59.15	30.19	105.0	4 004 000
1993	59.81	31.82	106.4	4 023 000

资料来源：总务厅统计局《家庭调查年报》，日本银行《经济统计年报》。

[解答]

(1) 南瓜的边际效用：

$$\frac{\partial U}{\partial Q_1}=\frac{b_1}{a_1+Q_1}$$

其他商品的边际效用：

$$\frac{\partial U}{\partial Q_2}=\frac{b_2}{a_2+Q_2}$$

因此，边际效用等式为

$$\frac{b_1}{(a_1+Q_1)P_1}=\frac{b_2}{(a_2+Q_2)P_2} \tag{8—7}$$

(2) (8—7)式与(8—6)式联立求解，推导出南瓜的需求函数(诱导方程式)：

$$Q_1=\pi_{10}+\pi_{11}\frac{Y}{P_1}+\pi_{12}\frac{P_2}{P_1} \tag{8—8}$$

式中， $\pi_{10}=\dfrac{-a_1b_2}{b_1+b_2}$

$$\pi_{11}=\frac{b_1}{b_1+b_2}$$

$$\pi_{12}=\frac{a_2b_1}{b_1+b_2}$$

(3) 对南瓜的需求函数 (8—8) 式进行 OLS 估计。

$$\hat{Q}_1=\underset{(4.197)}{26.028\ 1}+\underset{(8.062)}{0.000\ 732\ 622}\ \frac{Y}{P_1}-\underset{(-4.831)}{18.086\ 0}\ \frac{P_2}{P_1}$$

$$\bar{R}^2=0.864\ 6\quad DW=1.987$$

(4) 根据估计出的诱导型参数，求结构参数 a_1、a_2、b_1、b_2。

$$\hat{a}_1=\frac{\hat{\pi}_{10}}{\hat{\pi}_{11}-1}=-26.047$$

$$\hat{a}_2=\frac{\hat{\pi}_{12}}{\hat{\pi}_{11}}=-24\ 687$$

$$\hat{b}_1=\hat{\pi}_{11}=0.000\ 732\ 62$$

$$\hat{b}_2=1-\hat{\pi}_{11}=0.999\ 27$$

这里，从理论上讲，南瓜的边际效用 $\dfrac{\partial U}{\partial Q_1}=\dfrac{b_1}{a_1+Q_1}$ 必须为正数（而且 $0<b_1<1$），将 Q_1 数据代入检验，确认所有值均为正值。满足这一条件，则贝努利-拉普拉斯型一定是边际效用递减，因而满足效用最大化的充分条件。

因此，结构方程贝努里-拉普拉斯型效用函数为

$$\hat{U}=0.000\,732\,62\log\left(-26.047+Q_1\right)+0.999\,27\log\left(-24\,687+Q_2\right)$$

（5）由（3）中南瓜需求函数计算的理论值 Q，见表 8—3。理论值 Q_1 与实际值 Q_1 的图形见图 8—2。可见，理论值很好地反映了实际值的拐点（turning point），而且近年来的误差（$Q_1-\hat{Q}_1$）在缩小，显示了良好的结果。

表 8—3　　南瓜的消费量与理论值　　单位：100 克

年份	理论值 Q_1
1980	46.43
1981	47.88
1982	53.21
1983	47.83
1984	51.93
1985	52.88
1986	52.26
1987	54.37
1988	55.73
1989	59.18
1990	58.10
1991	54.07
1992	60.29
1993	58.18

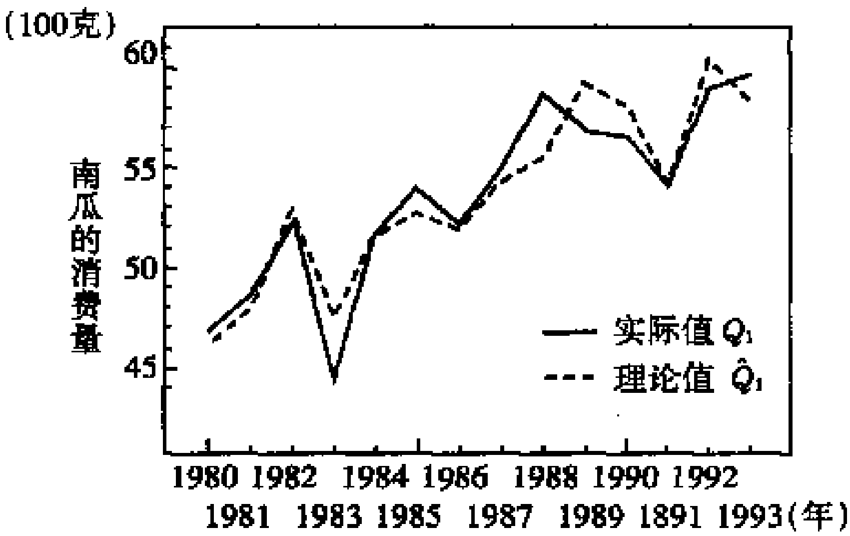


图 8—2　南瓜消费量的实际值 Q_1 与理论值 $\hat{Q}_1$

[补充 2]

效用函数，在各种商品、服务的消费需求的计量分析中经常采用。迁村（1981）介绍了一些实证的例子。

4. 识别问题

所谓识别问题（identification problem），指的是根据诱导型参数能否求出结构参数的问题。关于识别共有下面三种情形。

（1）适度识别（just-identifiable）。

指的是根据诱导型参数只计算一个结构参数的情形。可以应用间接最小二乘法和下面将要介绍的二阶段最小二乘法。

（2）过剩识别（over-identifiable）。

指的是根据诱导型参数计算两个以上结构参数的情形。间接最小二乘法不适用，二阶段最小二乘法适用。

（3）无法识别（not-identifiable）。

指的是根据诱导型参数不能计算一个结构参数的情形。间接最小二乘法和二阶段最小二乘法均不适用。

各结构方程式是否可以识别的判断条件，分为作为必要条件的阶条件（order condition）和作为充分必要条件的秩条件（rank condition）。本书只介绍阶条件，而秩条件由于需要行列计算的知识，略去不讲。有兴趣的读者可以参见蓑谷（1997b）。需要注意的是，即使是阶条件，在很多情况下可以作可能识别的判断。用于识别的阶条件可以整理如下：

（1）$K-J=H-1$	适度识别
（2）$K-J>H-1$	过剩识别
（3）$K-J<H-1$	无法识别

K——模型中先决变量（外生变量、先决内生变量）的个数

J——结构方程式中先决变量的个数

H——结构方程式中内生变量的个数

识别问题不是关于估计方法的问题，而是关于建模的问题，如果发现无法识别的结构方程式，为了使其能够识别，需要重新建立模型。

[例题 8—3]

下面 a～d 的结构型中，利用阶条件分别考察各结构方程式的识别可能性。这里（与下标无关）Y 为内生变量，X 为先决变量，α、β、γ 为结构参数，u 为误差项。

a. $Y_1 = \alpha_0 + \alpha_1 Y_2 + \alpha_2 X_1 + u_1$ ①

$Y_2 = \beta_0 + \beta_1 X_1 + u_2$ ②

b. $Y_1 = \alpha_0 + \alpha_1 Y_2 + \alpha_2 X_1 + \alpha_3 X_2 + u_1$ ①

$Y_2 = \beta_0 + \beta_1 X_2 + u_2$ ②

c. $Y_1 = \alpha_0 + \alpha_1 Y_2 + \alpha_2 X_1 + \alpha_3 X_2 + u_1$ ①

$Y_2 = \beta_0 + \beta_1 Y_1 + \beta_2 X_1 + \beta_3 X_2 + \beta_4 X_3 + u_2$ ②

d. $Y_1 = \alpha_0 Y_3 + \alpha_1 X_1 + \alpha_2 X_2 + u_1$ ①

$Y_2 = \beta_0 Y_1 + \beta_1 Y_3 + \beta_2 X_1 + \beta_3 X_2 + \beta_4 X_3 + u_2$ ②

$Y_3 = \gamma_0 Y_2 + \gamma_1 X_2 + \gamma_2 X_3 + \gamma_3 X_4 + u_3$ ③

[解答]

设模型中的先决变量数为 K，结构方程式中的先决变量数为 J，结构方程式中的内生变量数为 H。

a. ① $K - J = 1 - 1 = 0$

$H - 1 = 2 - 1 = 1$

$K - J < H - 1$

因此，①式无法识别。

② $K - J = 1 - 1 = 0$

$H - 1 = 1 - 1 = 0$

$K - J = H - 1$

因此，②式适度识别。

b. ① $K - J = 2 - 2 = 0$

$H - 1 = 2 - 1 = 1$

$K - J < H - 1$

因此，①式无法识别。

② $K-J=2-1=1$

$H-1=1-1=0$

$K-J>H-1$

因此，②式过剩识别。

c. ① $K-J=3-2=1$

$H-1=2-1=1$

$K-J=H-1$

因此，①式适度识别。

② $K-J=3-3=0$

$H-1=2-1=1$

$K-J<H-1$

因此，②式无法识别。

d. ① $K-J=4-2=2$

$H-1=2-1=1$

$K-J>H-1$

因此，①式过剩识别。

② $K-J=4-3=1$

$H-1=3-1=2$

$K-J<H-1$

因此，②式无法识别。

③ $K-J=4-3=1$

$H-1=2-1=1$

$K-J=H-1$

因此，③式适度识别。

5. 二阶段最小二乘法

上面学习的间接最小二乘法，只能适用于结构方程式适度识别的情形，在过剩识别的情况下，如果要计算两个以上的结构参数，就无法适用。

因此，利用泰尔（H. Theil）提出的**二阶段最小二乘法**（two-stage least squares method，2SLS 或 TSLS），即使在过剩识别的情形下，也可以只计算一个结构参数。当然，在适度识别的情形中也可以应用，在这种情况下，二阶段最小二乘法与间接最小二乘法的估计值是一致的。根据二阶段最小二乘法作出的估计值，满足一致性，不满足无偏性。

下面，我们再一次利用苹果的供求模型，说明二阶段最小二乘法的步骤。

步骤一：

从结构方程式（8—1）、(8—2）推导出诱导方程式（8—3）、(8—4）。

$$Q_t = \pi_{10} + \pi_{11} Y_t + \pi_{12} T_t + v_{1t} \tag{8—3}$$

$$P_t = \pi_{20} + \pi_{21} Y_t + \pi_{22} T_t + v_{2t} \tag{8—4}$$

步骤二：

对诱导方程式（8—4）使用 OLS 法，并估计诱导型参数。这时，(8—3）式没有必要进行估计（参照下面的步骤法）。

步骤三：

对结构方程式中成为解释变量的内生变量，计算其理论值。因此，利用步骤二中的估计结果，求 P_t 的理论值 $\hat{P}_t$。这里，由于 Q_t 没有成为结构方程式中的解释变量，没有必要计算其理论值。

步骤四：

用 $\hat{P}_t$ 代替 P_t，对结构方程式（8—1）、(8—2）进行 OLS 估计。即对下面的方程进行 OLS 估计。

$$Q_t = \alpha_0 + \alpha_1 \hat{P}_t + \alpha_2 Y_t + u'_{1t} \tag{8—9}$$

$$Q_t = \beta_0 + \beta_1 \hat{P}_t + \beta_2 T_t + u'_{2t} \tag{8—10}$$

根据这一步计算，$\hat{P}_t$ 与 u'_{1t}，$\hat{P}_t$ 与 u'_{2t}不相关，因而满足 OLS 中重要的假设——“解释变量与误差项不相关”。

这样，由于使用了两次 OLS，因而称为二阶段最小二乘法。

[例题 8—4]

对例题 8—1 中的苹果供求模型，进行二阶段最小二乘法（2SLS）估计。

[解答]

$\left.\begin{array}{l}\text{步骤一}\\\text{步骤二}\end{array}\right\}$ 已由例题 8—1 计算。

步骤三：

求 P_t 的理论值 $\hat{P}_t$（见表 8—4）。

$$\hat{P}_t = 25.1196 + 3.06024 Y_t - 4.77983 T_t$$

表 8—4　　苹果市场价格的理论值

t	$\hat{P}_t$
1	77.347 6
2	94.269 4
3	88.632 6
4	100.296 6
5	104.505 1
6	103.263 5
7	111.389 2
8	111.295 9

步骤四：

用 $\hat{P}_t$ 代替 P_t，对结构方程式（8—9）、（8—10）进行 OLS 估计，

需求函数：

$$\begin{aligned}\hat{Q}_t = &-5.1671 \quad - 0.31622 \ \hat{P}_t + 3.0991 \ Y_t \\ &(-1.779) \qquad (-5.255) \qquad (15.634) \\ &\bar{R}^2 = 0.9893 \quad DW = 2.293\end{aligned}$$

供给函数：

$$\begin{aligned}\hat{Q}_t = &-30.606 \quad + 0.69649 \ \hat{P}_t + 4.8406 \ T_t \\ &(-4.583) \qquad (14.335) \qquad (9.125) \\ &\bar{R}^2 = 0.9691 \quad DW = 1.895\end{aligned}$$

由此可见，本例的估计结果与例题 8—1 用间接最小二乘法估计的结果是一致的。

[补充 1]

需要注意的是，二阶段最小二乘法中的决定系数、t 值等，与 OLS 中不同，它们并不严密反映模型的适合程度、参数的显著性，只是表示模型的一些近似的相关指标。

[例题 8—5]

下面是由两个方程构成的结构型，它是 A 国消费函数的模型，表 8—5 是用于估计该模型的数据。

消费函数：$C_t = \alpha_0 + \alpha_1 Y_t + \alpha_2 C_{t-1} + u_t$ （8—11）

定义式：$Y_t = C_t + Z_t$ （8—12）

内生变量 $\begin{cases} C_t\text{：消费支出} \\ Y_t\text{：国内总产值} \end{cases}$ 先决变量 $\begin{cases} Z_t\text{：投资等（其他支出）} \\ C_{t-1}\text{：}(t-1)\text{ 期的消费支出} \end{cases}$

（1）利用阶条件，考察结构方程式（8—11）的识别可能性。

（2）利用二阶段最小二乘法对结构方程式（8—11）进行估计。

表 8—5 A 国的宏观经济数据 单位：10 亿美元

年份 t	国内总产值 Y_t	消费支出 C_t	投资等 Z_t
(1984)	—	70	—
1985	100	76	24
1986	108	82	26
1987	110	84	26
1988	117	87	30
1989	116	87	29
1990	124	91	33
1991	131	95	36
1992	136	98	38
1993	134	97	37
1994	142	102	40
1995	149	105	44

说明：1990 年价格。

[解答]

（1）由于模型中的先决变量数（K）为 2，结构方程式（8—11）中的先决变量数（J）为 1、内生变量数（H）为 2，即

$$K - J = 2 - 1 = 1$$
$$H - 1 = 2 - 1 = 1$$

因此，$K-J=H-1$

(8—11) 式为适度识别。

(2) 步骤一：

根据方程式 (8—11)、(8—12) 推导诱导方程式。

$$C_t=\pi_{10}+\pi_{11}Z_t+\pi_{12}C_{t-1}+v_{1t} \tag{8—13}$$

$$Y_t=\pi_{20}+\pi_{21}Z_t+\pi_{22}C_{t-1}+v_{2t} \tag{8—14}$$

式中， $\pi_{10}=\dfrac{\alpha_0}{1-\alpha_1}\quad \pi_{11}=\dfrac{\alpha_1}{1-\alpha_1}\quad \pi_{12}=\dfrac{\alpha_2}{1-\alpha_1}$

$$\pi_{20}=\frac{\alpha_0}{1-\alpha_1}\quad \pi_{21}=\frac{1}{1-\alpha_1}\quad \pi_{22}=\frac{\alpha_2}{1-\alpha_1}$$

$$v_{1t}=v_{2t}=\frac{u_t}{1-\alpha_1}$$

步骤二：

对诱导方程式 (8—14) 应用 OLS 法，估计诱导型参数。

$$\hat{Y}_t=\underset{(7.198)}{32.171\,6}+\underset{(12.217)}{1.866\,94}\;Z_t+\underset{(3.403)}{0.346\,142}\;C_{t-1} \tag{8—15}$$

$$\bar{R}^2=0.995\,9$$

(8—13) 式没有必要进行估计 (因为 C_t 不是结构方程式中的解释变量)。

步骤三：

利用步骤二的估计结果，计算 Y_t 的理论值 $\hat{Y}_t$ (见表 8—6)。

表 8—6　　A 国国内生产总值的理论值

年份 t	国内总产值的理论值 $\hat{Y}_t$
1985	101.208 2
1986	107.019 0
1987	109.095 8
1988	117.255 9
1989	116.427 3
1990	123.895 1
1991	130.880 5
1992	135.999 0
1993	135.170 5
1994	140.425 1
1995	149.623 6

步骤四：

用 $\hat{Y}_t$ 代替 Y_t，对结构方程式（8—11）进行 OLS 估计。

$$\hat{C}_t = \underset{(10.943)}{17.232} + \underset{(10.591)}{0.464\,37}\ \hat{Y}_t + \underset{(2.684)}{0.185\,41}\ C_{t-1} \qquad (8\text{—}16)$$
$$\bar{R}^2 = 0.996\,6$$

[补充 2]

联立方程模型的估计方法，除了上面所介绍的间接最小二乘法、二阶段最小二乘法以外，还有有限信息极大似然法（LIML）、完全信息极大似然法（FIML）、三阶段最小二乘法等，这些方法均超出了本书的范围，有兴趣的读者可以参照岩田（1982）。

6. 全体检验与最终检验

联立方程模型估计之后，为了判断模型的拟合优度是好是坏，或者说判断模型的表现（performance）（即理论值对实际值的刻划能力），通常还需要进行总体检验或最终检验。

所谓总体检验（total test），首先由二阶段最小二乘法估计出来的结构方程式推导诱导方程式，将其右边的外生变量与先决内生变量代入实际值，求左边内生变量的理论值，然后与实际值进行比较。

与总体检验相比，最终检验（final test）更为严格，它要求诱导方程式右边的先决内生变量中也要代入由模型中解出的理论值（即由前期计算出来的理论值），计算左边内生变量的理论值，并将其与实际值进行比较。只有最终检验结果良好的模型，才能最终用于预测和政策模拟。而且，预测和政策模拟时，根据分析者自己的判断，必须将外生变量的未来值设定为前期条件。

此外，无论是总体检验还是最终检验，最后都要检验求出的内生变量的理论值对实际值的刻划程度，检验的方法有两种，一是统计学的定量方法，二是画出图形的视觉方法。

定量方法中常用的有三种，即均方根误差（root mean square error，RMSE）、均方根误差率（RMSE 率）、泰尔的 U（Theil -U）。不管是哪种方法，数值越小，模型的适应性也越好。三种方法的计算公式如下：

$$\text{均方根误差} = \sqrt{\frac{1}{t}\sum(Y-\hat{Y})^2} \qquad (8\text{—}17)$$

$$均方根误差率=\sqrt{\frac{1}{t}\sum\left(\frac{Y-\hat{Y}}{Y}\right)^2} \tag{8—18}$$

$$泰尔的\ U=\sqrt{\frac{\frac{1}{t}\sum(Y-\hat{Y})^2}{\frac{1}{t}\sum Y^2}} \tag{8—19}$$

式中， Y_t——内生变量的实际值；

$\hat{Y}_t$——内生变量的理论值；

t——比较期间的样本数。

视觉方法中，重视理论值是否能很好地反映实际值的拐点（turning point）。这种方法在定量方法不能正确评价检验效果时，是一种极为重要的检验方法。而且，离现在较近的期间的误差（$Y-\hat{Y}$）是否较小，也很重要。

[例题 8—6]

（1）利用例题 8—5 中二阶段最小二乘法的估计结果，进行总体检验。

（2）同样，进行最终检验。

（3）根据（1）的最终检验结果，对 C_t 与 Y_t 分别求均方根误差、均方根误差率、泰尔的 U。

（4）假设外生变量 Z_t 在 1996—2000 年间按年均 10% 的速度增长，求 C_t 与 Y_t 的预测值 $\hat{C}_t$ 与 $\hat{Y}_t$，并进行模拟。

[解答]

（1）步骤一：

根据由 2SLS 估计出来的结构方程式（8—16）与定义式（8—12），求诱导方程式。

$$C_t=32.171\ 6+0.866\ 944Z_t+0.346\ 142C_{t-1} \tag{8—20}$$

$$Y_t=32.171\ 6+1.866\ 944Z_t+0.346\ 142C_{t-1} \tag{8—21}$$

步骤二：

将（8—20）式、（8—21）式右边的外生变量 Z_t 和先决内生变量 C_{t-1} 中代入实际值，求左边内生变量 C_t、Y_t 的理论值 $\hat{C}_t$、$\hat{Y}_t$（见表 8—7）。

表 8—7　　总体检验的结果

年份 t	消费支出理论值 $\hat{C}_t$	实际值 (C_t)	国内总产值的理论值 $\hat{Y}_t$	实际值 (Y_t)
1985	77.208 2	(76)	101.208 2	(100)
1986	81.019 0	(82)	107.019 0	(108)
1987	83.095 8	(84)	109.095 8	(110)
1988	87.255 9	(87)	117.255 9	(117)
1989	87.427 3	(87)	116.427 3	(116)
1990	90.895 1	(91)	123.895 1	(124)
1991	94.880 5	(95)	130.880 5	(131)
1992	97.999 0	(98)	135.999 0	(136)
1993	98.170 4	(97)	135.170 4	(134)
1994	100.425 1	(102)	140.425 1	(142)
1995	105.623 6	(105)	149.623 6	(149)

(2) 步骤一：

与 (1) 中的步骤一相同。

步骤二：

将 (8—20) 式、(8—21) 式右边的外生变量 Z_t 中代入实际值，先决内生变量 C_{t-1} 中代入由前期模型中计算出来的理论值，求左边内生变量 C_t、Y_t 的理论值 $\hat{C}_t$、$\hat{Y}_t$（见表 8—8)。

表 8—8　　最终检验的结果

年份 t	消费支出理论值 $\hat{C}_t$	实际值 (C_t)	国内总产值的理论值 $\hat{Y}_t$	实际值 (Y_t)
1985	77.208 2	(76)	101.208 2	(100)
1986	81.437 2	(82)	107.437 2	(108)
1987	82.901 0	(84)	108.901 0	(110)
1988	86.875 4	(87)	116.875 4	(117)
1989	87.384 2	(87)	116.384 2	(116)
1990	91.028 1	(91)	124.028 1	(124)
1991	94.890 2	(95)	130.890 2	(131)
1992	97.961 0	(98)	135.961 0	(136)
1993	98.156 9	(97)	135.156 9	(134)
1994	100.825 6	(102)	140.825 6	(142)
1995	105.217 1	(105)	149.217 1	(149)

(3) ⓐ利用 (8—17) 式计算得:

$$C \rightarrow \text{均方根误差} = \sqrt{\frac{1}{t}\sum(C-\hat{C})^2} = \sqrt{\frac{1}{11}\times 5.926\ 8} = 0.734\ 0$$

$$Y \rightarrow \text{均方根误差} = \sqrt{\frac{1}{t}\sum(Y-\hat{Y})^2} = \sqrt{\frac{1}{11}\times 5.926\ 8} = 0.734\ 0$$

ⓑ 利用 (8—18) 式计算得:

$$C \rightarrow \text{均方根误差率} = \sqrt{\frac{1}{t}\sum\left(\frac{C-\hat{C}}{C}\right)^2} = \sqrt{\frac{1}{11}\times 0.000\ 773\ 27} = 0.008\ 384$$

$$Y \rightarrow \text{均方根误差率} = \sqrt{\frac{1}{t}\sum\left(\frac{Y-\hat{Y}}{Y}\right)^2} = \sqrt{\frac{1}{11}\times 0.000\ 430\ 97} = 0.006\ 259$$

ⓒ 利用 (8—19) 式计算得:

$$C \rightarrow \text{泰尔的 } U = \sqrt{\frac{\frac{1}{t}\sum(C-\hat{C})^2}{\frac{1}{t}\sum C^2}} = \sqrt{\frac{\frac{1}{11}\times 5.926\ 8}{\frac{1}{11}\times 92\ 442}} = 0.008\ 007$$

$$Y \rightarrow \text{泰尔的 } U = \sqrt{\frac{\frac{1}{t}\sum(Y-\hat{Y})^2}{\frac{1}{t}\sum Y^2}} = \sqrt{\frac{\frac{1}{11}\times 5.926\ 8}{\frac{1}{11}\times 172\ 263}} = 0.005\ 866$$

(4) 步骤一:

与 (1) 中的步骤一相同。

步骤二:

先设定外生变量 Z_t 的未来值。由于 Z_t 的年增长率假设为 10%，未来值见表 8—9。

表 8—9 **Z_t 的未来值**

(增长率为 10% 的情形)

年份 t	Z_t 的未来值
1996	48.4
1997	53.24
1998	58.564
1999	64.420 4
2000	70.862 44

步骤三：

将（8—20）式、（8—21）式右边的外生变量 Z_t 中代入表8—9设定的未来值。先决内生变量 C_{t-1} 中，与最终检验一样，代入由前期模型计算出来的理论值（这里为预测值）。然后，求左边内生变量 C_t、Y_t 的理论值 $\hat{C}_t$、$\hat{Y}_t$。注意，在计算预测的初始点1996年的预测值时，C_{t-1} 中请使用实际值，也就是将1995年消费支出的实际值（105）代入。

表8—10　　预测的结果

年份 t	C_t 的预测值 $\hat{C}_t$	Y_t 的预测值 $\hat{Y}_t$
1996	110.476 6	158.876 6
1997	116.568 3	169.808 3
1998	123.292 5	181.856 5
1999	130.697 2	195.117 6
2000	138.845 2	209.707 6

第八章　练习题

1. 回答以下与例题 8—2 相关的问题。

(1) 用啤酒代替南瓜，推导啤酒的需求函数（诱导方程式），利用表 8—11、表 8—2 的数据（1980—1993 年），进行 OLS 估计。

(2) 求贝努利-拉普拉斯型效用函数（结构方程式）。

(3) 求啤酒消费量的理论值 $\hat{Q}_1$，并将其与实际值一道画出图形。

表 8—11　　日本每个家庭年均啤酒消费量及啤酒价格

年份	啤酒消费量（L） Q_B	啤酒的价格（日元/L） P_B
1980	46.76	387
1981	45.91	425
1982	46.87	440
1983	47.04	446
1984	44.48	507
1985	42.45	524
1986	45.08	525
1987	47.15	526
1988	49.14	528
1989	49.59	522
1990	55.39	541
1991	56.43	548
1992	54.32	547
1993	54.99	541

资料来源：总务厅统计局《家庭调查年报》。

2. 以下是由四个方程构成的结构型，它是台湾的宏观经济模型，表 8—12 给出了用于估计该模型的数据。

消费函数：$C_t = \alpha_0 + \alpha_1 Y_t + \alpha_2 C_{t-1} + u_1$　　①

投资函数：$I_t = \beta_0 + \beta_1 Y_t + \beta_2 I_{t-1} + u_2$　　②

进口函数：$M_t = \gamma_0 + \gamma_1 Y_t + u_3$　　③

定义式：$Y_t = C_t + I_t + G_t + X_t - M_t$　　④

$$
\text{内生变量}\begin{cases} Y_t\text{：总产值} \\ C_t\text{：民间最终消费支出} \\ I_t\text{：总固定资本形成} \\ M_t\text{：进口} \end{cases}
$$

$$
\text{先决变量}\begin{cases} G_t\text{：政府最终消费支出等} \\ X_t\text{：出口} \\ C_{t-1}\text{：1 期前民间最终消费支出} \\ I_{t-1}\text{：1 期前总固定资本形成} \end{cases}
$$

（1）利用阶条件考察结构方程式①、②、③的识别可能性。

（2）利用二阶段最小二乘法，对结构方程式①、②、③进行估计。

表 8—12　　台湾的宏观经济数据　　单位：10 亿新台币

年份 t	总产值 Y_t	民间最终消费支出 C_t	总固定资本形成 I_t	政府最终消费支出等 G_t	出口 X_t	进口 M_t
(1984)	—	(139)	(56)	—	—	—
1985	289	148	53	52	119	83
1986	323	159	58	53	153	100
1987	365	177	69	64	182	127
1988	392	201	79	73	191	152
1989	424	227	91	74	201	169
1990	447	246	98	80	202	179
1991	481	264	107	88	228	206
1992	514	287	121	97	240	231
1993	546	310	132	97	257	250
1994	582	337	139	93	272	259
1995	617	354	148	92	307	284

说明：1991 年价格，实际值。

资料来源：Council for Economic Planning and Development，*Taiwan Statistical Data Book*.

3．与例题 8—6 相关，在以下的前提条件（1）和（2）的基础上，计算内生变量的预测值，并进行模拟。模拟期间为 1996—2000 年的 5 年间。

（1）设外生变量 Z_t（投资等）按每年 5% 的比率增长。

（2）设外生变量 Z_t 按每年 3% 的比率减少。

第九章 投入产出分析

1. 什么是投入产出表

本章介绍由俄裔美国经济学家里昂惕夫（W. W. Leontief，1973 年诺贝尔经济学奖获得者）最先提出的投入产出表（input-output table，IO 表），以及由此而发展起来的分析方法。

所谓投入产出表，简单说来，就是将一定期间（通常为一年），一个国家的产业之间（有时也将地区之间、国际之间）产业活动的相互关系，也即商品或服务的流动状况，用一览表的形式表现出来。这种方法一个突出的特点是将国民收入统计中没有涉及的产业之间的相互关系，也即中间商品（原材料、燃料）等的交易情况，明确地记录下来。

日本最早编制的全国性投入产出表是“昭和 26 年表”

（1951 年），第二次是“昭和三十年表”（1955 年），此后每隔 5 年，由各相关省厅共同编制。最近的“平成二年表”（1990 年，由 11 省厅共同编制），基本分类是 527 个行部门×411 个列部门，以及将其与 13、32、91、187 个部门综合起来的投入产出表，同时公布。此外，日本全国九大板块（北海道、东北、关东等）的地区投入产出表，47 个都道府县的投入产出表，政府特定城市的投入产出表也要编制、公布。

自从里昂惕夫 1936 年提出投入产出表以来，由于其优越的理论体系和操作性，不但用于实际经济计划和经济预测，而且从经济结构分析到环境问题，在极其广泛的范围中得到灵活运用。以联合国为中心，投入产出表的统一化不断推进，现在世界上约有 80 多个国家编制、公布投入产出表。

2. 投入产出表的阅读方法

表 9—1 给出了一个最简单的投入产出表，一个国家的经济由工业和农业两个部门构成。下面，我们利用这个表，简要介绍投入产出表的阅读方法。

表 9—1　　由两个部门构成的投入产出表

部　门		中间需求		最终需求	总产值
		农业	工业		
中间投入	农业	X_{11}	X_{12}	F_1	X_1
	工业	X_{21}	X_{22}	F_2	X_2
毛增加值		V_1	V_2		
总 产 值		X_1	X_2		

从横向（行，row）看第一行的农业，可以看出农业产品的销售去向。X_{11}为农产品在农业内部作为中间产品（原材料）而销售的金额，X_{12}是农产品向工业部门（也是作为原材料）销售的金额。F_1 表示农产品不是本期的原材料，而是作为最终产品用于消费、投资或者出口的金额，因此，它反映的是最终需求（final demand）。由此可见，从横向方面阅读投入产出表，可以看出各种产业的产品如何分配给各种产业以及最终需求。这样，由于投入产出表的各行中，中间需求（intermediate demand）与最终需求之和等于总产值 X_i，因而有以下的供求等式。

$$\begin{cases} X_{11}+X_{12}+F_1=X_1 \\ X_{21}+X_{22}+F_2=X_2 \end{cases} \tag{9—1}$$

现在，我们再从纵向（列，column）来看表 9—1，它反映的是各种产业如何从各种产业购入中间产品（原材料）。而且，毛增加值 V_j 中，包括了各种产业中雇员收入（工资、奖金）、营业剩余（企业的营业利润）、家庭外部消费支出（交际费、福利健康费用等）、资本损耗（折旧）、间接税（但不包括消费税、关税）等。因此，第一列（农业）之和为 X_1，第二列（工业）之和为 X_2，下面这样的供求等式，关于列也是成立的。

$$\begin{cases} X_{11}+X_{21}+V_1=X_1 \\ X_{12}+X_{22}+V_2=X_2 \end{cases} \tag{9—2}$$

［例题 9—1］

填充表 9—2 的投入产出表的空栏ⓐ～ⓕ。

表 9—2

部门		中间需求		最终需求	总产值
		农业	工业		
中间投入	农业	ⓐ	15	ⓑ	40
	工业	8	ⓒ	32	ⓓ
毛增加值		20	ⓒ		
总产值		ⓕ	100		

［解答］

利用各行与各列的供求等式（9—1）、（9—2），求ⓐ～ⓕ的值，见表 9—3。

表 9—3

部门		中间需求		最终需求	总产值
		农业	工业		
中间投入	农业	ⓐ12	15	ⓑ13	40
	工业	8	ⓒ60	32	ⓓ100
毛增加值		20	ⓐ25		
总产值		ⓕ40	100		

3. 投入系数

所谓投入系数（input-coefficient），反映的是生产的技术关系，在生产技术没有重大变化的期间中（主要是短期），可以假设它是稳定的系数。这种性质是投入产出分析的重要基础。下面，我们在此基础上展开投入产出分析。

投入产出系数可以用下面的公式来表示：

$$a_{ij}=\frac{X_{ij}}{X_j} \tag{9—3}$$

式中，a_{ij}表示的是第 j 个产业生产 1 个单位产品时，需要第 i 个产业中的多少产品作为原材料。表 9—4 是根据表 9—1 的投入产出表而编制的投入系数表。

表 9—4　　投入系数表

部门	农业	工业
农业	$a_{11}\left(=\frac{X_{11}}{X_1}\right)$	$a_{12}\left(=\frac{X_{12}}{X_2}\right)$
工业	$a_{21}\left(=\frac{X_{21}}{X_1}\right)$	$a_{22}\left(=\frac{X_{22}}{X_2}\right)$

将农业的投入系数纵向排列的向量 $\begin{pmatrix}a_{11}\\a_{21}\end{pmatrix}$，反映了一个国家农业的生产技术。同样，工业的向量可以表示为 $\begin{pmatrix}a_{12}\\a_{22}\end{pmatrix}$。

下面，将（9—3）式变形得

$$X_{ij}=a_{ij}X_j \tag{9—4}$$

这种技术关系式，规模收益一定（例如，投入增加 2 倍，产量也增加 2 倍），而且可以被看做是生产要素之间不能相互替代的生产函数。利用（9—4）式，将表 9—4 变形得

$$\begin{aligned}X_{11}&=a_{11}X_1\\X_{12}&=a_{12}X_2\\X_{21}&=a_{21}X_1\\X_{22}&=a_{22}X_2\end{aligned} \tag{9—5}$$

将 (9—5) 式代人 (9—1) 式，可以得到下面这样的利用投入系数的供求等式：

$$\begin{aligned} a_{11}X_1 + a_{12}X_2 + F_1 = X_1 \\ a_{21}X_1 + a_{22}X_2 + F_2 = X_2 \end{aligned} \tag{9—6}$$

上述供求等式用行列式表示为

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} + \begin{pmatrix} F_1 \\ F_2 \end{pmatrix} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \tag{9—7}$$

$$AX + F = X \tag{9—8}$$

附加价值系数可以表示为

$$v_j = \frac{V_j}{X_j} \tag{9—9}$$

式中，v_j 表示的是第j 个产业生产 1 个单位产品时，会产生多少单位的附加价值。

[补充]　预测投入系数的资料

在利用投入产出分析进行经济预测和经济计划时，必须考虑投入系数在预测期间中的变化。关于投人系数的预测方法，有斯通开发的 RAS 方法。关于 RAS 方法，可参见宫泽 (1995) 的简要解说。

[例题 9—2]

根据例题 9—1 的投入产出表 (见表 9—3)，计算投入系数 a_{ij}与附加价值系数 v_j。

[解答]

设投入系数矩阵为 A，附加价值行向量为 V。

$$(1)\ A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \begin{pmatrix} \dfrac{X_{11}}{X_1} & \dfrac{X_{12}}{X_2} \\ \dfrac{X_{21}}{X_1} & \dfrac{X_{22}}{X_2} \end{pmatrix} = \begin{pmatrix} \dfrac{12}{40} & \dfrac{15}{100} \\ \dfrac{8}{40} & \dfrac{60}{100} \end{pmatrix} = \begin{pmatrix} 0.3 & 0.15 \\ 0.2 & 0.6 \end{pmatrix}$$

$$(2)\ V = (v_1 \quad v_2) = \left(\frac{V_1}{X_1} \quad \frac{V_2}{X_2}\right) = \left(\frac{20}{40} \quad \frac{25}{100}\right) = (0.5 \quad 0.25)$$

4. 里昂惕夫逆矩阵

所谓里昂惕夫逆矩阵（Leontief inverse matrix），指的是这样一种矩阵，已知某产业一个单位的最终需求，它对各个产业有多少生产波及效果。该矩阵在投入产出分析中具有核心作用。

里昂惕夫逆矩阵可以先从总产值中求解横向的供求等式（9—8），然后按以下方法推导出来。

$$AX + F = X$$

$$(I - A)X = F$$

均衡产出余额模型：
$$\underset{\substack{\uparrow\\ \text{总产值}}}{X} = \underset{\substack{\uparrow\\ \text{里昂惕夫逆矩阵}}}{(I - A)^{-1}} \underset{\substack{\uparrow\\ \text{最终需求}}}{F} \tag{9—10}$$

式中，I 是单位向量。如果是两部门投入产出表，则

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$(I - A)^{-1}$为里昂惕夫逆矩阵，包含该矩阵的（9—10）式称为均衡产出余额模型。利用该模型，如果外生给定任意最终需求 F，则与该需求相应的直接、间接必要产量 X 可以计算出来。

进一步，里昂惕夫逆矩阵 B 为

$$B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \tag{9—11}$$

b_{ij}表示如果对第 j 个产业的最终需求增加 1 个单位，由最终需求的增加或者由此派生的中间需求的增加，将会使得第 i 个产业的产量增加多少单位。因此，在例题 9—1 的案例中，里昂惕夫逆矩阵 B 从纵向来看，b_{11}表示的是对农业的最终需求增加 1 个单位时，农业的生产量直接、间接增加多少单位；b_{21}表示的是对农业

的最终需求增加 1 个单位时，工业的产量间接地增加多少单位。同样，b_{12}表示的是对工业的最终需求增加 1 个单位时，农业的生产量间接增加多少单位；b_{22}表示的是对工业的最终需求增加 1 个单位时，工业的生产量直接、间接增加多少单位。

顺便提一下，里昂惕夫逆矩阵主对角线上的系数（例如 b_{11}、b_{22}），一定大于 1，其他的系数通常小于 1。主对角线上的 b_{ij} 表示对自己部门直接就有 1 个单位的最终需求，此外还要加上派生的中间需求，因而自然要大于 1。

[例题 9—3]

根据例题 9—1 的投入产出表，回答下面的问题。

(1) 计算里昂惕夫逆矩阵。

(2) 如果现在对工业的最终需求只增加 10 个单位，计算农业与工业的总产量各增加多少单位。这里，设农业的最终需求不变。

(3) 设农业与工业的雇用人数分别为 200 万人和 400 万人，计算在 (2) 中，经济总体雇用人数的增加量。

[解答]

$$(1)\ B=(I-A)^{-1}=\left[\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}-\begin{pmatrix}0.3 & 0.15\\ 0.2 & 0.6\end{pmatrix}\right]^{-1}=\begin{pmatrix}0.7 & -0.15\\ -0.2 & 0.4\end{pmatrix}^{-1}$$

$$=\frac{1}{0.28-0.03}\begin{pmatrix}0.4 & 0.15\\ 0.2 & 0.7\end{pmatrix}=4\cdot\begin{pmatrix}0.4 & 0.15\\ 0.2 & 0.7\end{pmatrix}$$

$$=\begin{pmatrix}1.6 & 0.6\\ 0.8 & 2.8\end{pmatrix}$$

2×2 矩阵的逆矩阵，可以求解如下：

$$\begin{pmatrix}a & b\\ c & d\end{pmatrix}^{-1}=\frac{1}{ad-bc}\begin{pmatrix}d & -b\\ -c & a\end{pmatrix}$$

(2) 设最终需求的增加额为 ΔF，总产值的增加额为 ΔX，则

$$\Delta X=B\cdot\Delta F=\begin{pmatrix}1.6 & 0.6\\ 0.8 & 2.8\end{pmatrix}\begin{pmatrix}0\\ 10\end{pmatrix}=\begin{pmatrix}6\\ 28\end{pmatrix}$$

农业与工业的总产值分别增加 6 个和 28 个单位。

(3) 设农业和工业的就业系数为 e_1、e_2。

$$e_1=\frac{\text{农业就业量}}{\text{农业总产值}}=\frac{200}{40}=5$$

$$e_2=\frac{\text{工业就业量}}{\text{工业总产值}}=\frac{400}{100}=4$$

设经济总体的就业增加量为 ΔE。

$$\begin{aligned}\Delta E &= e_1\times\Delta X_1+e_2\times\Delta X_2\\ &=5\times6+4\times28=142(\text{万人})\end{aligned}$$

[例题 9—4]*

表 9—5 是中国由六个产业部门构成的投入系数表（1992 年），根据该表，计算里昂惕夫逆矩阵。

表 9—5　　中国由六个产业部门构成的投入系数表（1992 年）

部门	农业	工业	建筑	运输、通信	商业、饮食业	服务
农业	0.139 3	0.078 6	0.003 5	0.000 1	0.032 5	0.005 3
工业	0.156 6	0.502 5	0.563 7	0.339 6	0.240 5	0.263 8
建筑	0.000 1	0.000 5	0.006 9	0.001 4	0.009 0	0.014 8
运输、通信	0.011 2	0.017 7	0.029 8	0.014 2	0.110 8	0.041 2
商业、饮食业	0.021 6	0.074 7	0.085 9	0.048 8	0.040 1	0.045 4
服务	0.026 9	0.040 5	0.014 3	0.035 8	0.108 8	0.098 1

资料来源：国家统计局《1992 年度中国投入产出表》，中国统计出版社。

[解答]

利用 Lotus 1-2-3 或者 Excel 等计算机软件，就可以进行简单的计算，计算结果见表 9—6。作者在这里使用的是 TSP 中的矩阵演算功能（参照第十章）。

表 9—6　　里昂惕夫逆矩阵（1992 年）

部门	农业	工业	建筑	运输、通信	商业、饮食业	服务
农业	1.208 2	0.218 2	0.141 9	0.084 2	0.116 0	0.082 9
工业	0.474 0	2.299 4	1.412 5	0.861 7	0.793 0	0.777 8
建筑	0.002 0	0.005 1	1.011 4	0.004 5	0.013 5	0.019 0
运输、通信	0.032 9	0.071 5	0.086 2	1.049 4	0.149 8	0.078 0
商业、饮食业	0.069 1	0.194 4	0.213 3	0.127 4	1.123 5	0.123 1
服务	0.067 0	0.136 1	0.112 9	0.098 3	0.180 8	1.164 4

5. 影响力系数与感应度系数

下式表示里昂惕夫逆矩阵的列和（纵向合计）。

$$c_j = \sum_{i=1}^{n} b_{ij} \tag{9—12}$$

式中，c_j 表示当对第 j 个产业的最终需求增加 1 个单位时（但是，其他产业的最终需求增加为零），全部产业的产量增加多少单位。据此就可以知道第 j 个产业对全部产业影响力的大小。为了将该影响力与其他产业的影响力进行比较，需要将各产业的列和 c_j 除以 c_j 的平均值 $\bar{c}$，得出的数值称为影响力系数（index of the power of dispersion，α_j）。

$$\alpha_j = \frac{c_j}{\bar{c}} = \frac{\sum_{i=1}^{n} b_{ij}}{\frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{n} b_{ij}} \tag{9—13}$$

如果 $\alpha_j>1$，意味着影响力相对较大；如果 $\alpha_j<1$，则影响力相对较小。一般说来，原材料对其他产业的供给比率极高，因此影响力系数也较大。

另一方面，里昂惕夫逆矩阵的行和（横向合计），可以表示为

$$d_i = \sum_{j=1}^{n} b_{ij} \tag{9—14}$$

式中，d_i 表示当对全部产业的最终需求每增加 1 个单位时，第 i 个产业的产量增加多少单位。也就是说，根据 d_i 就可以知道第 i 个产业受全部产业的影响的程度。因此，将该行和 d_i 除以 d_i 的平均值 $\bar{d}$，得出的值称为感应度系数（index of the sensitivity of dispersion，β_i）。据此可以相对比较产业间受影响程度的大小。

$$\beta_i = \frac{d_i}{\bar{d}} = \frac{\sum_{j=1}^{n} b_{ij}}{\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} b_{ij}} \tag{9—15}$$

如果 $\beta_i>1$，说明所受的影响相对较大；如果 $\beta_i<1$，则所受的影响相对较小。如果某个产业对其他产业的原材料供应依赖较大，则该产业的感应度系数通常也

较高。

[例题 9—5]

根据例题 9—3 求出的里昂惕夫逆矩阵，计算：

(1) 列和 c_j 和影响力系数 α_j。

(2) 行和 d_i 和感应度系数 β_i。

[解答]

(1) 列和

根据 (9—12) 式，得

$$c_1 = b_{11} + b_{21} = 1.6 + 0.8 = 2.4(\text{农业})$$
$$c_2 = b_{22} + b_{22} = 0.6 + 2.8 = 3.4(\text{工业})$$

影响力系数

根据 $\bar{c} = \frac{1}{2}(c_1 + c_2) = \frac{5.8}{2} = 2.9$，利用 (9—13) 式，得

$$\alpha_1 = \frac{c_1}{\bar{c}} = \frac{2.4}{2.9} = 0.827\ 6(\text{农业})$$
$$\alpha_2 = \frac{c_2}{\bar{c}} = \frac{3.4}{2.9} = 1.172\ 4(\text{工业})$$

(2) 行和

根据 (9—14) 式，得

$$d_1 = b_{11} + b_{21} = 1.6 + 0.6 = 2.2(\text{农业})$$
$$d_2 = b_{21} + b_{22} = 0.8 + 2.8 = 3.6(\text{工业})$$

感应度系数

根据 $\bar{d} = \frac{1}{2}(d_1 + d_2) = \frac{5.8}{2} = 2.9$，利用 (9—15) 式，得

$$\beta_1 = \frac{d_1}{\bar{d}} = \frac{2.2}{2.9} = 0.758\ 6(\text{农业})$$
$$\beta_2 = \frac{d_2}{\bar{d}} = \frac{3.6}{2.9} = 1.241\ 4(\text{工业})$$

[例题 9—6]

根据例题 9—4 求出的里昂惕夫逆矩阵（见表 9—6），计算列和 c_j 和影响力系数 α_j，以及行和 d_j 和感应度系数 β_j。

[解答]

表 9—7 1992 年的影响力系数和感应度系数

	列和	影响力系数	行和	感应度系数
农业	1.853 1	0.761 4	1.851 4	0.760 7
工业	2.924 6	1.201 6	6.618 4	2.719 3
建筑	2.978 2	1.223 7	1.055 5	0.433 7
运输、通信	2.225 5	0.914 4	1.467 7	0.603 0
商业、饮食业	2.376 6	0.976 5	1.850 8	0.760 4
服务	2.245 3	0.922 5	1.759 5	0.722 9

由此可见，工业的特征是感应度系数极高，达到 2.719 3（见图 9—1)。这反映了中国的工业部门受其他产业的影响很大，而且原材料或能源不足导致工业产品的生产不能满足需求，形成通货膨胀和经济增长的瓶颈。

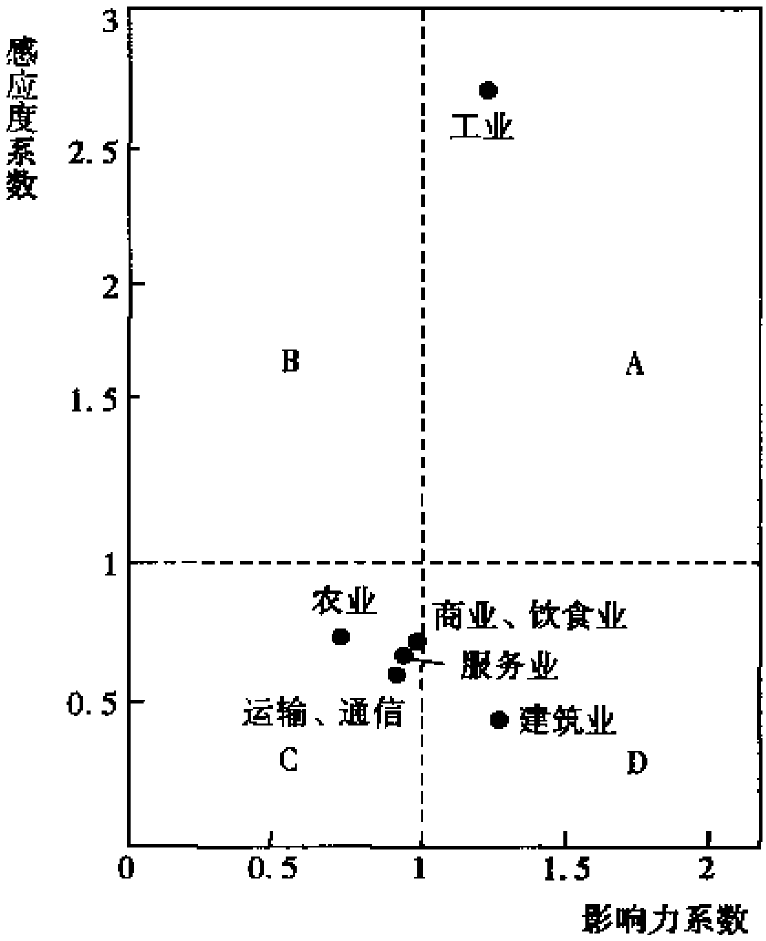


图 9—1 1992 年的影响力系数和感应度系数

图 9—1 中，A 区：对其他产业的影响力较大，而且容易受影响。

B 区：对其他产业的影响力较小，但是容易受影响。

C 区：对其他产业的影响力较小，而且不容易受影响。

D 区：对其他产业的影响力较大，但是不容易受影响。

6. 进口如何处理

到目前为止，在我们的投入产出模型分析中，为了解释的方便，省略了与进口有关的问题。但是，国内总需求中［中间需求和国内最终需求（出口以外的最终需求）之和］总有一部分要依赖于进口，因此，为了构建更现实的模型，必须考虑如何将进口纳入到模型中。

投入产出表在考虑进口时，基本上有两种方法，一种是竞争进口方式（competitive import），另一种是非竞争进口方式（complementary import）（还有一种竞争与非竞争进口的折衷方式），这里主要说明竞争进口方式，因为这种方式的投入系数稳定，因而经常被采用。

表 9—8 为竞争进口型投入产出表。这里第 i 个产业的进口额为 M_i，假设它不是与总产量 X_i，而是与国内总需求（$\sum_{j=1}^{n} X_{ij} + F_{Ci} + F_{Ii}$）成比例，则进口系数 m_i 可以表示如下：

$$m_i = \frac{M_i}{\sum_{j=1}^{n} X_{ij} + F_{Ci} + F_{Ii}} \tag{9—16}$$

表 9—8　　竞争进口型投入产出表

部门		中间需求		最终需求			（扣除）进口	总产值
		农业	工业	消费	投资	出口		
中间投入	农业	X_{11}	X_{12}	F_{C1}	F_{I1}	F_{E1}	M_1	X_1
	工业	X_{21}	X_{22}	F_{C2}	F_{I2}	F_{E2}	M_2	X_2
毛增加值		V_1	V_2					
总产值		X_1	X_2					

根据 (9—16) 式，国产自给率可以表示为

$$r_i = 1 - m_i \tag{9—17}$$

在表 9—8 的例子中，农业的进口系数 m_1 为

$$m_1 = \frac{M_1}{X_{11} + X_{12} + F_{C1} + F_{I1}} \tag{9—18}$$

工业的进口系数 m_2 为

$$m_2 = \frac{M_2}{X_{21} + X_{22} + F_{C2} + F_{I2}} \tag{9—19}$$

设进口系数的对角矩阵为 $\overline{M}$，国内最终需求为 F_D(消费 + 投资)，出口为 F_E，根据供求等式 (9—20)，按下面的方式可以推导出竞争进口方式中的里昂惕夫逆矩阵：

$$X = AX + F_D + F_E - M \leftarrow \text{供求均衡式} \tag{9—20}$$

$$X = AX + F_D + F_E - \overline{M}(AX + F_D)$$

$$[I - (I - \overline{M})]X = (I - \overline{M})F_D + F_E$$

$$\underbrace{X}_{\text{总产值}} = [\underbrace{I - (I - \overline{M})A}_{\text{里昂惕夫逆矩阵}}]^{-1}[\underbrace{(I - \overline{M})F_D}_{\text{对本国产品的国内最终需求}} + \underbrace{F_E}_{\text{出口}}] \tag{9—21}$$

包含里昂惕夫逆矩阵$[I - (I - \overline{M})A]^{-1}$的 (9—21) 式，相当于竞争进口方式中的均衡产量模型。而且，$(I - \overline{M})A$ 称做国内产品投入系数，根据投入系数 A 扣除进口部分，则意味着国产部分的投入系数。同样，$(I - \overline{M})F_D$ 表示的是对国内产品的国内最终需求。

[补充]　竞争进口方式与非竞争进口方式的“适当”区分

(1) 竞争进口方式：像经济预测、经济计划那样计算波及效果的情形。

(2) 非竞争进口方式：分析结构、产业结构现状的情形。

对非竞争进口方式有兴趣的读者可以参见宫泽 (1995)。

[例题 9—7]

表 9—9 的竞争进口型投入产出表是将例题 9—1 表中的最终需求与消费、投资、出口相区分，并纳入了进口的模型。

(1) 计算进口系数 m_i 与国产自给率 r_i。

(2) 求 $[I-(I-\overline{M})A]^{-1}$ 型的里昂惕夫逆矩阵 B，并将其与例题 9—3 中计算出的里昂惕夫逆矩阵相比较。

表 9—9　　由两部门构成的竞争进口型投入产出表

部门		中间需求		最终需求			（扣除）进口	总产值
		农业	工业	消费	投资	出口		
中间投入	农业	12	15	18	5	5	−15	40
	工业	8	60	12	10	28	−18	100
毛增加值		20	25					
总产值		40	100					

[解答]

(1) 根据 (9—18) 式，农业的进口系数 m_1 为

$$m_1=\frac{M_1}{X_{11}+X_{12}+F_{C1}+F_{I1}}=\frac{15}{12+15+18+5}$$
$$=\frac{15}{50}=0.3$$

根据 (9—17) 式，国产自给率 r_1 为

$$r_1=1-m_1=1-0.3=0.7$$

另一方面，根据 (9—19) 式，工业进口系数 m_2 为

$$m_2=\frac{M_2}{X_{21}+X_{22}+F_{C2}+F_{I2}}=\frac{18}{8+60+12+10}$$
$$=\frac{18}{90}=0.2$$

根据 (9—17) 式，国产自给率 r_2 为

$$r_2=1-m_2=1-0.2=0.8$$

(2) 单位矩阵：$I=\begin{pmatrix}1&0\\0&1\end{pmatrix}$

进口系数(对角矩阵)：$\overline{M}=\begin{pmatrix}0.3&0\\0&0.2\end{pmatrix}$

投入系数：$A=\begin{pmatrix}0.3 & 0.15\\ 0.2 & 0.6\end{pmatrix}$

因此，$[I-(I-\overline{M})A]^{-1}$型里昂惕夫逆矩阵为

$$B=[I-(I-\overline{M})A]^{-1}$$
$$=\left[\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}-\left[\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}-\begin{pmatrix}0.3 & 0\\ 0 & 0.2\end{pmatrix}\right]\begin{pmatrix}0.3 & 0.15\\ 0.2 & 0.6\end{pmatrix}\right]^{-1}$$
$$=\begin{pmatrix}0.79 & -0.105\\ -0.16 & 0.52\end{pmatrix}^{-1}$$
$$=\frac{1}{0.394}\begin{pmatrix}0.52 & 0.105\\ 0.16 & 0.79\end{pmatrix}$$
$$=\begin{pmatrix}1.3198 & 0.2665\\ 0.4061 & 2.0051\end{pmatrix}$$
$$(I-A)^{-1}=\begin{pmatrix}1.6 & 0.6\\ 0.8 & 2.8\end{pmatrix}$$

可见，它略小于 $(I-A)^{-1}$型。其原因在于，对于单纯面向进口的国内产业来说，最终需求的生产诱发波及效果（repercussion effect）较小。

7. 生产诱发额与生产诱发系数

所谓生产诱发额指的是，给定消费 F_C、投资 F_I、出口 F_E 这些最终需求，由此可以诱发出多少产量。利用竞争进口型均衡产量模型（9—21)式，消费、投资、出口可以分别表示如下：

$$\text{消费的生产诱发额}=[I-(I-\overline{M})A]^{-1}(I-\overline{M})F_C$$
$$=B\Gamma F_C \tag{9—22}$$
$$\text{投资的生产诱发额}=[I-(I-\overline{M})A]^{-1}(I-\overline{M})F_I$$
$$=B\Gamma F_I \tag{9—23}$$
$$\text{出口的生产诱发额}=[I-(I-M)A]^{-1}\cdot F_E$$
$$=BF_E \tag{9—24}$$

式中，　$[I-(I-\overline{M})A]^{-1}=B$（里昂惕夫逆矩阵）

$I-\overline{M}=\Gamma$（国产自给率的对角矩阵）

Γ 为希腊字母，读做 Gama。各产业的生产诱发额之和，等于各产业的总产量之和。

其次，利用（9—22）式、（9—23）式、（9—24）式中求出的消费、投资、出口的生产诱发额，除以各自的最终需求合计额，可以得到生产诱发系数。具体的计算方法为：

$$\text{消费的生产诱发系数}=\frac{\text{消费的生产诱发额}}{\text{消费合计额}}=\frac{B\Gamma F_C}{iF_C} \tag{9—25}$$

$$\text{投资的生产诱发系数}=\frac{\text{投资的生产诱发额}}{\text{投资合计额}}=\frac{B\Gamma F_I}{iF_I} \tag{9—26}$$

$$\text{输出的生产诱发系数}=\frac{\text{出口的生产诱发额}}{\text{出口合计额}}=\frac{BF_E}{iF_E} \tag{9—27}$$

式中，i 为单位行向量。$i=(1,1,1,\cdots,1)$

通过计算生产诱发系数，可以知道某一最终需求（消费、投资、出口）增加 1 个单位时，能够诱发各产业（农业、工业）产量增加多少单位。

进一步看，将各产业的消费、投资、出口的生产诱发额，除以相应各产业的生产诱发额之和（= 各产业的总产量），可以计算每个产业中，消费、投资、出口的生产诱发额的构成比。该比值称为各最终需求项目的生产诱发依存度，或生产的最终需求依存度，通过计算这一系数，可以判断各个产业是消费依赖型、投资依赖型还是出口依赖型。

[例题 9—8]

根据例题 9—7 的竞争进口型投入产出表（见表 9—9），回答以下问题。

(1) 消费、投资、出口的生产诱发额。

(2) 消费、投资、出口的生产诱发系数。

(3) 消费、投资、出口的生产诱发依存度。

[解答]

设 $[I-(I-\overline{M})A]^{-1}$ 型里昂惕夫逆矩阵为 B，国产自给率 $(I-\overline{M})$ 为 Γ。

$$B=\begin{pmatrix}1.3198 & 0.2665\\ 0.4061 & 2.0051\end{pmatrix}$$

$$\Gamma=\begin{pmatrix}0.7 & 0\\ 0 & 0.8\end{pmatrix}$$

(1)消费的生产诱发额为

$$B\Gamma F_C=\begin{pmatrix}1.3198 & 0.2665\\0.4061 & 2.0051\end{pmatrix}\begin{pmatrix}0.7 & 0\\0 & 0.8\end{pmatrix}\begin{pmatrix}18\\12\end{pmatrix}$$
$$=\begin{pmatrix}19.19\\24.37\end{pmatrix}$$

投资的生产诱发额为

$$B\Gamma F_I=\begin{pmatrix}1.3198 & 0.2665\\0.4061 & 2.0051\end{pmatrix}\begin{pmatrix}0.7 & 0\\0 & 0.8\end{pmatrix}\begin{pmatrix}5\\10\end{pmatrix}$$
$$=\begin{pmatrix}6.75\\17.46\end{pmatrix}$$

出口的生产诱发额为

$$BF_E=\begin{pmatrix}1.3198 & 0.2665\\0.4061 & 2.0051\end{pmatrix}\begin{pmatrix}5\\28\end{pmatrix}$$
$$=\begin{pmatrix}14.06\\58.17\end{pmatrix}$$

农业的生产诱发额之和为

$$19.19+6.75+14.06=40$$

工业的生产诱发额之和为

$$24.37+17.46+58.17=100$$

两个产业均与各自的总产值相一致。

(2) 消费的生产诱发系数为

$$\frac{B\Gamma F_C}{iF_C}=\frac{\begin{pmatrix}19.19\\24.37\end{pmatrix}}{(1\quad 1)\begin{pmatrix}18\\12\end{pmatrix}}=\frac{1}{30}\begin{pmatrix}19.19\\24.37\end{pmatrix}=\begin{pmatrix}0.6397\\0.8123\end{pmatrix}$$

投资的生产诱发系数为

$$\frac{B\Gamma F_I}{iF_I}=\frac{\begin{pmatrix}6.75\\17.46\end{pmatrix}}{(1\quad 1)\begin{pmatrix}5\\10\end{pmatrix}}=\frac{1}{15}\begin{pmatrix}6.75\\17.46\end{pmatrix}=\begin{pmatrix}0.450\\1.164\end{pmatrix}$$

出口的生产诱发系数为

$$\frac{BF_E}{iF_E}=\frac{\begin{pmatrix}14.06\\58.17\end{pmatrix}}{(1\quad 1)\begin{pmatrix}5\\28\end{pmatrix}}=\frac{1}{33}\begin{pmatrix}14.06\\58.17\end{pmatrix}=\begin{pmatrix}0.426\ 1\\1.762\ 7\end{pmatrix}$$

工业部门出口的生产诱发系数极高，达到1.762 7，由此可见，出口的生产诱发效果非常显著。

(3) 农业的最终需求项目的生产诱发依存度为 (1) 中求出的农业消费、投资、出口的生产诱发额，除以农业的总产量 $X_1(40)$。同样，工业部门的最终需求项目的诱发依存度，是工业消费、投资、出口的生产诱发额，除以工业的总产量 $X_2(100)$ (见表 9—10)。

表 9—10　　最终需求项目的生产诱发依存度

部门	消费	投资	出口	总终需求合计
农业	$\frac{19.19}{40}$ $=0.479\ 75$	$\frac{6.75}{40}$ $=0.168\ 75$	$\frac{14.06}{40}$ $=0.351\ 5$	$\frac{40}{40}$ $=1.000\ 0$
工业	$\frac{24.37}{100}$ $=0.243\ 7$	$\frac{17.46}{100}$ $=0.174\ 6$	$\frac{58.17}{100}$ $=0.581\ 7$	$\frac{100}{100}$ $=1.000\ 0$

从计算的结果可以看出，日本的农业为消费依赖型，工业为出口依赖型，而且这种倾向非常明显。

[补充 1]　进口诱发额、进口诱发系数以及附加价值诱发额、附加价值诱发系数

下面，给出进口以及关于附加价值的诱发额与诱发系数。解释的方法与生产诱发额、生产诱发系数相同。

$$\text{消费的进口诱发额}=\overline{M}(AB\Gamma+I)F_C=\overline{M}\Gamma^{-1}B\Gamma F_C \tag{9—28}$$

$$\text{消费的进口诱发系数}=\frac{\overline{M}\Gamma^{-1}B\Gamma F_C}{iF_C} \tag{9—29}$$

$$\text{投资的进口诱发额}=\overline{M}(AB\Gamma+I)F_I=\overline{M}\Gamma^{-1}B\Gamma F_I \tag{9—30}$$

$$\text{投资的进口诱发系数}=\frac{\overline{M}\Gamma^{-1}B\Gamma F_I}{iF_I} \tag{9—31}$$

$$出口的进口诱发额 = \overline{M}ABF_E \tag{9—32}$$

$$出口的进口诱发系数 = \frac{\overline{M}ABF_E}{iF_E} \tag{9—33}$$

$$消费的附加价值诱发额 = \overline{V}B\Gamma F_C \tag{9—34}$$

$$消费的附加价值诱发系数 = \frac{\overline{V}B\Gamma F_C}{iF_C} \tag{9—35}$$

$$投资的附加价值诱发额 = \overline{V}B\Gamma F_I \tag{9—36}$$

$$投资的附加价值诱发系数 = \frac{\overline{V}B\Gamma F_I}{iF_I} \tag{9—37}$$

$$进口的附加价值诱发额 = \overline{V}BF_E \tag{9—38}$$

$$进口的附加价值诱发系数 = \frac{\overline{V}BF_E}{iF_E} \tag{9—39}$$

式中，$\overline{V}$ 为附加价值的对角矩阵。

[补充 2] 均衡价格模型

投入产出表中，根据列（纵向）的数据，可以推导出每一单位产品的价格决定模型。借用在本章所采用的工业、农业两部门情形，该模型可以表示为

$$p_1a_{11} + p_2a_{21} + v_1 = p_1 \tag{9—40}$$

$$p_1a_{12} + p_2a_{22} + v_2 = p_2 \tag{9—41}$$

式中，p_1——农产品单位价格；

p_2——工业制成品单位价格；

v_1——农业附加价值系数；

v_2——工业附加价值系数；

a_{11}、a_{12}、a_{21}、a_{22}——投入系数。

(9—40) 式、(9—41) 式可以表示为如下矩阵：

$$\underbrace{\begin{pmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{pmatrix}}_{\text{投入系数矩阵的转置矩阵}} \begin{pmatrix} p_1 \\ p_2 \end{pmatrix} + \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} p_1 \\ p_2 \end{pmatrix} \tag{9—42}$$

即 $$A'P + V = P \tag{9—43}$$

(9—43) 式对价格列向量 P 求解，可以导出如下的均衡价格模型：

$$P=(1-A')^{-1}V \tag{9—44}$$

$$=[\underbrace{(I-A)^{-1}}_{\text{里昂惕夫逆矩阵的转置矩阵}}]'V \tag{9—45}$$

$$=B'V$$

可见，均衡价格模型与均衡产出余额模型 (9—10) 式有着密切的关系，称做双对体系。

第九章 练习题

1. 根据表 9—11 的竞争进口型投入产出表，回答以下问题。

(1) 填充空栏ⓐ～ⓕ。

(2) 投入系数矩阵 A。

(3) 进口系数与国产自给率。

(4) $[I-(I-\overline{M})A]^{-1}$型里昂惕夫逆矩阵 B。

(5) 消费、投资、出口的生产诱发额、生产诱发系数以及生产诱发依存度。

(6) 消费、投资、出口的进口诱发额与进口诱发系数。

(7) 消费、投资、出口的附加价值诱发额与附加价值诱发系数。

表 9—11　　由两个部门构成的竞争进口型投入产出表

部门		中间需求		最终需求			(扣除)进口	总产值
		农业	工业	消费	投资	出口		
中间投入	农业	60	ⓐ	50	10	10	ⓑ	ⓒ
	工业	ⓓ	120	ⓔ	20	30	30	200
毛增加值		20	50					
总产值		100	ⓕ					

2、根据表 9—12 的竞争进口型投入产出表，回答以下问题。

(1) 投入系数矩阵 A。

(2) $(I-A)^{-1}$型里昂惕夫逆矩阵。

(3) 进口系数 M。

(4) $[I-(I-\overline{M})A]^{-1}$型里昂惕夫逆矩阵 BB。

(5) 根据$[I-(I-\overline{M})A]^{-1}$型里昂惕夫逆矩阵 BB，计算列和 G、行和 H、影响力系数 GG 和感应度系数 HH。

(6) 消费、投资、出口的生产诱发额 $P1$、$P2$、$P3$ 和生产诱发系数 $PP1$、$PP2$、$PP3$。

(7) 消费、投资、出口的进口诱发额 $M1$、$M2$、$M3$ 和进口诱发系数 $MM1$、$MM2$、$MM3$。

(8) 消费、投资、出口的附加价值诱发额 $V1$、$V2$、$V3$ 和附加价值诱发系数 $VV1$、$VV2$、$VV3$。

(9) 设各产业的就业人数分别为：农业 220 万人，制造业 480 万人，建筑业 210 万人，服务业 560 万人。现在政府对建筑业实施 18 个单位的公共投资，计算各产业的总产量与就业量增加多少。

表 9—12　　由四个部门构成的竞争进口型投入产出表

		中间需求					最终需求			(扣除)出口	总产值
		农业	制造业	建筑业	服务业	国内部门合计	消费	投资	出口		
中间投入	农　业	6	21	3	2	32	17	2	1	12	40
	制造业	8	120	35	24	187	58	33	45	23	300
	建筑业	2	18	1	6	27	0	73	0	0	100
	服务业	4	36	16	28	84	105	6	8	3	200
	国内部门合计	20	195	55	60	330	180	114	54	38	640
毛增加值		20	105	45	140	310					
总　产　值		40	300	100	200	640					

第十章 计算机计量经济分析——TSP基础

1. 什么是 TSP

《通过例题学习计量经济学》终于迎来了最后一章。如果读者通过本书的学习对计量经济学产生了兴趣，并大致灵活掌握了其分析方法，则作者甚感欣慰。

计量经济学和统计学一样，虽然理解了基本现论，但在实际中如何灵活应用还有很多问题。因此本书希望通过大量计量经济学例题的解说，体会计算能力和应用能力，并以此作为理论与实际应用的桥梁。

本章作为本书的最后一章，简明扼要地介绍计量经济学中广泛使用的软件 TSP（time series processor）的基本操作方法。该软件是以美国经济学者为主、为计量经济分析而开发的，自 1967 年发布第一版以来，被全世界众多大学和研究咨询机构采用。TSP 的特点在于，与其他统

计软件不同，它是立足于经济学而开发的，因此计量经济分析的菜单非常丰富，而且操作极其简便。

TSP 的使用方法（输入方式）有对话型和编程处理型（batch processing system）两种，本书主要介绍对初学者来说操作简便的对话型方法。如果程序比较复杂，则编程处理型方法更方便。

2. 描述统计与最小二乘法

[例题 10—1]

以下利用对例题 3—1（一元回归模型）的数据所做的提问，显示 TSP 的程序和输出结果。

X	6	11	17	8	13
Y	1	3	5	2	4

（1）输入 X、Y 的数据，为了确认输入的数据，显示输出结果。

（2）求 X、Y 的描述统计量（算术平均、标准偏差等）。

（3）以 X 为横轴、Y 为纵轴，画出数据的散点图。

（4）对一元回归模型 $Y=\alpha+\beta X+u$ 进行 OLS 估计。

（5）标出（4）的残差（0）。

[解答]（1）TSP 程序（数据的输入和输出）。

```
OPTIONS CRT;        ←指定显示画面的输出，限一行 80 位。
FREQ N;             ←指定非时序数列数据（N）的期限类型。
SMPL 1 5;           ←指定数据的期间。开始期为 1，结束期为 5。
LOAD X;             ←输入变量 X 的数据。也可用READ X。
6 11 17 8 13;
LOAD Y;             ←输入变量 Y 的数据。
1 3 5 2 4;
PRINT X Y;          ←输出 X、Y 的数据。
```

输出结果。

	X	Y
1	6.000 00	1.000 00
2	11.000 00	3.000 00
3	17.000 00	5.000 00
4	8.000 00	2.000 00
5	13.000 00	4.000 00

[解说]

1) TSP 的每一语句均用分号（;）表示结束。

2) 如果一个语句较长，需要多行，可用斜线（\）或符号（¥）。在投入产出分析时，变量较多，这种方法很有用处。

(例)

```
LOAD   X1   X2   X3   X4   X5\
X6   X7   X8   X9   X10;
```

3) 根据 FREQ 指令，指定期间的方法。

FREQ A;	←	每年度数据
FREQ Q;	←	季度数据
FREQ M;	←	月度数据
FREQ W;	←	周数据
FREQ N;	←	非时序数据

4) 根据 SMPL 指令，指定期间的方法。

(例 1) 从 1945 年到 1996 年年度数据的情形。

```
FREQ   A;
SMPL   45   96;
```

(例 2) 1985 年第 2 季度到 1989 年第 4 季度的季度数据情形。

```
FREQ   Q;
SMPL   85:2   89:4;
```

(例 3) 从 1990 年 4 月到 1995 年 12 月月度数据的情形。

```
FREQ   M;
```

SMPL 90:4 95:12;

(例 4) 1991 年到 1994 年，1996 年到 1999 年，2001 年到 2004 年的年度数据的情形。

```
FREQ  A;
SMPL  91  94  96  99  101  104;
```

5) 根据 LOAD 指令输入数据。

数据既可以按变量分别输入，也可以用下面的方法集中输入。

```
LOAD    X  Y;
 6   1
11   3
17   5
 8   2
13   4;
```

变量的名称可以用 8 位以内的英文字母和数字来表示，最初一位须从字母开始。但是"C"被约定为回归分析中的常数项，不能用于变量名称。

6) 数据的修正方法。

如果输入错误，需要修正数据，在 TSP（限于对话型）中，可以使用UPDATE 指令。

具体地，首先输入 $\boxed{\text{UPDATE　变量名称;}}$，指定想修正的变量名称；其次指定修正期间（即第几位数据）；最后输入正确的数据，数据的修正就完成了。

(2) 根据 MSD 指令，求描述统计量。

```
MSD    X   Y;
       变量名称
```

输出结果。

NUMBER OF OBSERVATIONS: 5

	Mean	std Dev	Minimum	Maximum
X	11.000 00	4.301 16	6.000 00	17.000 00
Y	3.000 00	1.581 14	1.000 00	5.000 00
	Sum	Variance	Skewness	Kurtosis
X	55.000 00	18.500 00	0.377 02	−0.629 66
Y	15.000 00	2.500 00	0.000 00	−1.200 00

[解说]

1）输出结果的意义如下。

Mean	：	算术平均
Std Dev	：	标准差
Minimum	：	最小值
Maximum	：	最大值
Sum	：	总和
Variance	：	方差
Skewness	：	歪度
Kurtosis	：	尖度

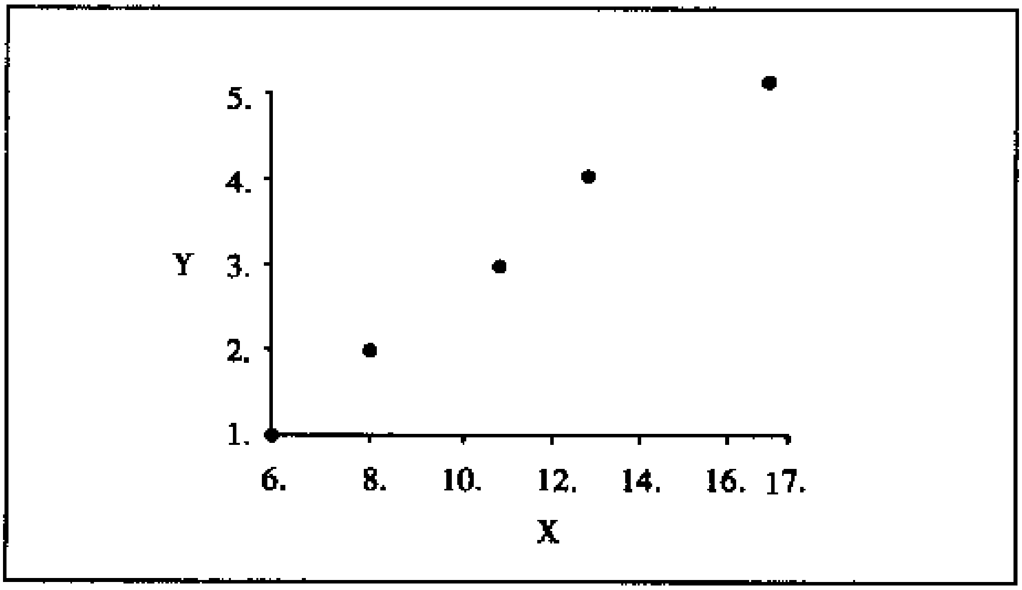
2）使用 CORR 指令，可以求多个变量之间的相关矩阵。

$$\text{CORR}\quad \underbrace{\text{X}\quad \text{Y}}_{\text{变量名称（2个以上）}};$$

(3）根据 GRAPH 指令，画出散点图。

GRAPH	X	Y
	横轴	纵轴

输出结果。



（4）根据 OLSQ 指令，对 $Y=\alpha+\beta X+u$ 进行 OLS 估计。

OLSQ	$\underset{\sim}{\mathrm{Y}}$	$\underset{\sim}{\mathrm{C}}$	$\underset{\sim}{\mathrm{X}}$;
	被解释变量	常数项	解释变量

输出结果。

```
                 Method of estimation = Ordinary Least Squares
Dependent variable: Y
Current sample: 1 to 5
Number of observations: 5
    Mean of dependent variable = 3.00000
    Std.dev. of dependent var. = 1.58114
      Sum of squared residuals = .148649
         Variance of residuals = .049550
      Std.error of regression = .222597
                     R-squared = .985135
            Adjusted R squared = .980180
      Durbin-Watson statistic = 1.19779
      F-statistic(zero slopes) = 198.818
  Schwarz Bayes.Info.Crit. = -2.87183
   Log Of likelihood function = 1.69433

            Estimated      Standard
Variable    Coefficient    Error        t-statistic

C           -1.01351       .301546      -3.36106
X           .364865        .025876      14.1003
```

上述估计结果简单整理如下：

$$
\begin{aligned}
Y = & \ -1.013\,51 \quad + \ 0.364\,865X \\
& (-3.361\,06) \quad\ (14.100\,3) \\
& R^2=0.985\,135 \quad s=0.222\,597
\end{aligned}
$$

[解说]

1）输出结果的意义如下。

Mean of dependent variable	：	被解释变量（从属变量）的平均值
Std. dev. of dependent var.	：	被解释变量的标准差
Sum of squared residuals	：	残差平方和
Variance of residuals	：	残差方差
Std. error of regression	：	回归方程式的标准差
R-squared	：	决定系数
Adjusted R-squared	：	自由度调整决定系数
Durbin-Watson statistic	：	*DW*
F-statistic (zero slopes)	：	*F* 值
Schwarz Bayes. Info. Crit.	：	Schwarz Bayes 信息量基准
Log of likelihood function	：	对数似然函数
Estimated Coefficient	：	估计出来的回归系数
Standard Error	：	回归系数的标准误差
t-statistic	：	*t* 值

2）对多元回归模型 $Y=\alpha+\beta_1X_1+\beta_2X_2+u$ 进行 OLS 估计时，同样可以使用 OLSQ 指令。

OLSQ	Y	C	X_1	X_2
	被解释变量	常数项	解释变量	解释变量

3）如果想增加输出结果的有效位数，可以用下面的指令。

```
OPTIONS SIGNIF = 希望有效位数
```

（例）如果想将有效位数更改为 8 位，则，

OPTIONS SIGNIF = 8;

（5）如要标出残差，在 OLS 运算之前，使用 PLOTS。

```
PLOTS;
OLSQ    Y  C  X;
```

如要解除，则用

```
NOPLOT;
```

输出结果。

	实际值	理论值	实际值与理论值的标记	残差	残差的标记
ID	ACTUAL(*)	FITTED(+)		RESIDUAL(0)	0
1	1.0000	1.1757	* +	−0.1757	0 : +
2	3.0000	3.0000	+	0.0000	+ 0 +
3	5.0000	5.1892	* +	−0.1892	0 : +
4	2.0000	1.9054	+ *	0.09459	+ : 0 +
5	4.0000	3.7297	+ *	0.2703	+ : + 0

[解说]

1）TSP（对话型）结束的指令。

```
END;
```

2）将所有变量的数据保存在软盘中的指令。

```
SAVE ' A: 文件名';
           ↑
    8 位以内（英文字母与数字）
```

调出已保存数据的指令。

```
RESTORE ' A : 文件名';
```

3）一些使用方便的指令。

```
REVIEW    最初一行   最后一行;
```

可以重新表示已经指定的行的范围。在自已编制的程序遗忘时，非常有用。

```
EXEC    最初一行    最后一行;
```

重新运行已经指定的行的范围。在对较长的指令重新运行时，非常方便。

```
SHOW   SERIES;
```

显示已经输入的是何种变量（或样本数）。这在偶然忘记已输入的变量名称

时，较有用处。

```
SELECT    条件式;
SMPLIF    条件式;
```

根据条件式（例如 $X>0$），可以抽出一部分数据。SELECT 与 SMPLIF 的区别在于，SELECT 指令需要每次独立地设定条件，而 SMPLIF 指令可以利用以前设定的条件，而且可以追加条件，充分利用数据。利用 SELECT 1; ，可以回到原先的数据范围（全部数据使用的状态）。

3. 数据的变换

TSP 中，使用 GENR 指令，可以将右边的算式变换成左边的变量(GENR 也可以省略)。请看下面的例子，

```
GENR  A  =X+Y;
GENR  B  =X/Y; ←X÷Y
GENR  D  =X**2; ←X²
GENR  E  =LOG (X);
```

下面是一些经常使用的算式符号与函数。

+	：加法	LOG (　　)	：自然对数
−	：减法	LOG10 (　　)	：常用对数
*	：乘法	ABS (　　)	：绝对值
/	：除法	SQRT (　　)	：平方根
**	：幂乘	EXP (　　)	：指数函数

计算顺序依次为：①（　　）内的计算；②幂乘；③乘法、除法；④加法、减法。(　　)可以多次使用。

[例题 10—2]

通过求解例题 1—8（求全国百货店销售额的中心化 4 项移动平均，并与原数列一道画出图形），显示 TSP 程序。

[解答]

TSP 程序（中心化 4 项移动平均与图形）。

```
OPTIONS  CRT;
FREQ  Q;    ←指定数据的类型为季度数据（Q)。
SMPL  91:1  96:2; ←指定数据的期间。开始期为 1991 年第一季度，结束
                    期为 1996 年第 2 季度。
LOAD  X; ←输入百货店销售额 X 的数据。
219  225  232  296
221  224  227  280
208  210  215  263
202  205  212  258
193  199  208  257
206  206
SMPL  91:3  95:4; ←由于中心化 4 项移动平均中，计算后，前后每隔二
                    个都要减少数据，需要重新指定期间。
GENR  MA=(0.5*X(-2)+X(-1)+X+X(1)+0.5*X(2))/4;
            ↑___ 设 MA 求中心化 4 项移动平均。
PRINT  MA; ←输出中心化 4 项移动平均 MA。
PLOT X MA; ←画出原数列 X 与中心化 4 项移动平均 MA。
```

[解说]

1）附加时滞（过去值）变量的情形。

变量名称（-n）

例如，

X_{t-1}的情形→X（-1）

X_{t-8}的情形→X（-8）

2）附加领先（未来值）变量的情形。

变量名称（n）

例如，

X_{t+1}的情形→X（1）

X_{t+5}的情形→X（5）

[例题 10—3]

通过求解例题 4—7（画出 X 与 Y 的散点图，并对多元回归模型 $Y=\alpha+\beta_1X+\beta_2X^2+u$ 进行 OLS 估计），显示 TSP 程序。

[解答]

TSP 程序（对 $Y=\alpha+\beta_1X+\beta_2X^2+u$ 进行 OLS 估计）。

```
OPTIONS  CRT;
FREQ  N; ←指定非时序数据（N）的期间类型。
SMPL  1  12; ←开始期为 1，结束期为 12。
LOAD  Y; ←输入变量 Y 的数据。
141.4  170.6  202.2  246.6  288.1  306.0
311.6  318.6  316.5  304.0  267.2  237.1;
LOAD  X; ←输入变量 X 的数据。
16.8  19.1  22.6  27.4  32.4  37.5
42.6  47.3  52.5  57.4  62.2  68.7;
GRAPH  X  Y; ←画出横轴为 X、纵轴为 Y 的散点图。
GENR  X2=X**2; ←设 X2 求 X²。
OLSQ  Y  C  X  X2  ←对多元回归模型进行 OLS 估计。
```

4.Cochrane-Orcutt 法与极大似然法

在第七章的学习时我们已经知道，在对时间序列数据进行回归分析时，误差项之间的相关性所引起的序列相关是非常麻烦的问题。TSP 中对存在序列相关的模型的估计，准备了以下几种方法：（1）Cochrane-Orcutt 法；（2）极大似然法（maximum likelihood method）；（3）系统网络搜索的极大似然法（maxiumum likelihood with grid search）；（4）希尔德雷斯-卢法（Hildneth-Lu method）。下回我们通过例题 10—4，介绍方法（1）和方法（2）。

[例题 10—4]

根据例题 7—2（日本工薪家庭消费函数的估计）的数据，通过对模型（$Y=\alpha+\beta X+u$）的 Cochrane-Orcutt 法和极大似然法估计，显示 TSP 程序。

[解答]

TSP 程序（Cochrane-Orcutt 法与极大似然法）。

```
OPTIONS  CRT;
FREQ  A; ←指定数据的类型为年度数据（A）。
SMPL  70  94; ←指定数据的期间。开始期为 1970 年，结束期为 1994 年。
LOAD  Y; ←输入变量 Y 的数据。
239  248  258…334  330;
LOAD  X; ←输入变量 X 的数据。
300  311  329…449  449;
AR1（METHOD=CORC）  Y  C  X; ←对 Y=α+βX+u 进行 Cochrane-
                                        Orcutt 估计。
AR1（METHOD=ML）Y  C  X; ←对 Y=α+βX+u 进行极大似然估计。
```

[解说]

以下是对误差项中存在一阶系列相关的模型进行估计的指令。

AR1 （METHOD= 估计法)被解释变量 C 解释变量名单

对存在一阶系列相关的模型进行估计时所用的指令

常数项

Cochrane-Orcutt 法：CORC

极大似然法：ML

Grid 法：MLGRID

Hildreth-Lu 法：HILU

这四种方法中，最常用的是极大似然法。

此外，也可以指定收敛判定条件以及最大迭代次数。在 METHOD = 估计法之后指定即可。

TOL = 收敛判定条件（既定值 = 0.005）
MAXIT = 最大迭代次数

5. 二阶段最小二乘法

第八章已经学习过联立方程模型的估计法，这里对其中较为典型的二阶段最小二乘法，利用 TSP 进行估算。

[例题 10—5]

通过求解例题 8—4（对苹果的供求函数进行二阶段最小二乘法估计），显示 TSP 程序。

[解答]

TSP 程序（二阶段最小二乘法）。

```
OPTIONS  CRT;
FREQ  N; ←指定数据类型为非公历、非时序数据（N）。
SMPL  1  8; ←指定期据期间。开始期为 1，结束期为 8。
LOAD  Q; ←输入苹果的数量数据 Q。
57  55  66  65  71  74  71  77;
LOAD  P; ←输入苹果的市场价格 P 的数据。
78  96  87  98  104  105  110  113;
LOAD  Y; ←输入需求者收入 Y 的数据。
28  29  32  33  35  36  36  38;
LOAD  T; ←输入每天平均日照时间 T 的数据。
7.0  4.1  7.2  5.4  5.8  6.7  5.0  6.3;
```

```
2SLS  Q  C  P Y  INVR  C Y T
```

指令：

- 2SLS ← 二阶段最小二乘法
- Q ← 被解释变量
- C ← 常数项
- P Y ← 解释变量
- INVR C Y T ← ZNVR 下面，列出模型中所有先决变量(包括常数项 C)

```
2SLS  Q  C  P T  INVR  C Y T
```

如果是有限信息极大似然法，二阶段最小二乘法的指令应该用 LIML 代替 2SLS。

6. 投入产出分析

由第九章可以看出，投入产出分析方法既有优良的理论体系和操作性，又有广泛的应用领域，由于其计算量非常大，因此利用计算机是必不可少的。下面利用 TSP 的矩阵演算功能，学习投入产出分析的一些基本计算方法，例如投入系数、里昂惕夫逆矩阵、生产诱发系数、进口诱发系数等。

[例题 10—6]

利用 TSP 的矩阵演算功能，求解第九章练习题 2（由四个部门构成的竞争进口型投入产出表)。

[解答]

(1) TSP 程序——求投入系数矩阵 A。

```
OPTIONS CRT;
FREQ N;                 ←在非时系列数据(N)中指定数据的期限种类。
SMPL 1 4;               ←内生部门数量。
LOAD X1;                ←按列（垂直）输入农业 X1 的数据。
6 8 2 4;
```

```
LOAD X2;                    ←按列输入制造业 X2 的数据。
21  120  18  36;
LOAD X3;                    ←按列输入建筑业 X3 的数据。
3  35  1  16;
LOAD X4;                    ←按列输入服务业 X4 的数据。
2  24  6  28;
GENR A1 = X1/40;            ←求农业投入系数 A1。
GENR A2 = X2/300;           ←求制造业投入系数 A2。
GENR A3 = X3/100;           ←求建筑业投入系数 A3。
GENR A4 = X4/200;           ←求服务业投入系数 A4。
MMAKE A A1 A2 A3 A4;        ←由 A1～A4 四个系列，编制矩阵 A。
                            由几个系列编制矩阵的指令。
PRINT A;                    ←输出投入系数矩阵 A。
```

(2) TSP 程序——求 $(I-A)^{-1}$ 型里昂惕夫逆矩阵 B。

```
MFORM (TYPE = DIAG, NCOL = 4) I = 1;
```

编制 4 次单位矩阵。

$$I = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

也可用 MAT I = IDENT(4);

指定列数为 4。

指定矩阵类型为对角矩阵。

变更矩阵类型和次数的指令。

```
MAT IA = I - A;             ←对矩阵实行 I - A 计算，并设其为 IA。
                            矩阵计算的指令。
INV IA B;                   ←求 IA 的逆矩阵，设其为 B。
                            求逆矩阵的指令。
```

```
PRINT B;                    ←输出 (I－A)⁻¹型里昂惕夫逆矩阵。
```

(3) TSP 程序——求进口系数 M。

```
LOAD X5;                          ←按列(垂直)输入内生部门合计 X5 的数据。
32  187  27  84;
LOAD F1;                          ←按列输入消费 F1 的数据。
17  58  0  105;
LOAD F2;                          ←按列输入投资 F2 的数据。
2  33  73  6;
LOAD F4;                          ←按列输入 F4 的数据。
12  23  0  3;
GENR M=F4/(X5+F1+F2); ←求进口系数,设其为 M。
PRINT M;                          ←输出进口系数 M。
```

(4)TSP 程序——求$[I-(I-\overline{M})A]^{-1}$型里昂惕夫逆矩阵 BB。

```
MFORM(TYPE=DIAG,NCOL=4)M;←将进口系数 M 作对角矩阵变换。
MAT IIMA=I-(I-M)*A;          ←对 I-(I-M)A 作矩阵计算,设其
                               为 IIMA。
INV IIMA BB;                 ←求 IIMA 的逆矩阵,设其为 BB。
PRINT BB;                    ←输出[I-(I-M)A]⁻¹型里昂惕夫
                               逆矩阵。
```

(5)TSP 程序——(求列和 G、行和 H、影响力系数 GG、感应度系数 HH)。

```
MAT BBB=BB';                  ←将矩阵 BB 转置,设其为 BBB。
UNMAKE BBB G1 G2 G3 G4;       ←由矩阵 BBB 形成 G1～G4 四个系列。
  ↑
  └──────────────────────由矩阵形成数个系列的指令。
UNMAKE BB H1 H2 H3 H4;        ←由矩阵 BB 形成 H1～H4 四个系列。
GENR G=G1+G2+G3+G4;           ←求列和,设其为 G。
GENR H =H1+H2+H3+H4;          ←求行和,设其为 H。
MSD G H;                      ←求列和 G 与行和 H 的平均值(1.90767)。
```

```
GENR GG=G/1.90767;              ←求影响力系数,设其为 GG。
GENR HH=H/1.90767;              ←求感应度系数,设其为 HH。
PRINT G H GG HH;                ←输出 G、H、GG、HH。
```

(6)TSP 程序——求生产诱发额与生产诱发系数。

```
LOAD F3;                        ←按列(垂直)输入出口 F3。
1   45   0   8;
MAT P1=BB*(I-M)*F1;             ←计算消费的生产诱发额 P1。
MAT P2=BB*(I-M)*F2              ←计算投资的生产诱发额 P2。
MAT P3=BB*F3;                   ←计算出口的生产诱发额 P3。
MAT PP1=P1/180;                 ←计算消费的生产诱发系数 PP1。
MAT PP2=P2/114;                 ←计算投资的生产诱发系数 PP2。
MAT PP3=P3/54;                  ←计算出口的生产诱发系数 PP3。
PRINT P1 P2 P3 PP1 PP2 PP3;     ←输出 P1、P2、P3、PP1、PP2、PP3。
```

(7)TSP 程序——求进口诱发额与进口诱发系数。

```
MAT R=I-M;                      ←计算国产自给率(对角矩阵)R。
INV R RR;                       ←计算 R 的逆矩阵,设其结果为 RR。
MAT M1=M*RR*P1                  ←计算消费的进口诱发额 M1。
MAT M2=M*RR*P2;                 ←计算投资的进口诱发额 M2。
MAT M3=M*A*P3;                  ←计算出口的进口诱发额 M3。
MAT MM1=M1/180;                 ←计算消费的进口诱发系数 MM1。
MAT MM2=M2/114;                 ←计算投资的进口诱发系数 MM2。
MAT MM3=M3/54;                  ←计算出口的进口诱发系数 MM3。
PRINT M1 M2 M3 MM1 MM2 MM3;     ←输出 M1、M2、M3、MM1、MM2、
                                  MM3。
```

(8)TSP 程序——附加价值诱发额与附加价值诱发系数。

```
LOAD V;                         ←输入附加价值 V 的数据。
20 105 45 140;
LOAD X;                         ←输入总产值 X 的数据。
```

```
40 300 100 200;
GENR VV = V/X;                              ←求附加价值系数,设其为 VV。
MFORM(TYPE = DIAG,NCOL = 4)VV;              ←对附加价值系数 VV 作对角矩阵
                                              变换。
MAT V1 = VV * P1;                           ←计算消费的附加价值诱发额 V1。
MAT V2 = VV * P2;                           ←计算投资的附加价值诱发额 V2。
MAT V3 = VV * P3;                           ←计算出口的附加价值诱发额 V3。
MAT VV1 = V1/180;                           ←计算消费的附加价值诱发系数
                                              VV1。
MAT VV2 = V2/114;                           ←计算投资的附加价值诱发系数
                                              VV2。
MAT VV3 = V3/54;                            ←计算出口的附加价值诱发系数
                                              VV3。
PRINT V1 V2 V3 VV1 VV2 VV3;                 ←输出 V1、V2、V3、VV1、VV2、
                                              VV3。
```

(9)TSP 程序——求公共投资形成的产量与就业的增加。

```
LOAD K;                                     ←输入公共投资 K 的数据。
0 0 18 0;
MAT KK = BB * (I - M) * K;                  ←计算公共投资的生产诱发额 KK。
LOAD L;                                     ←输入就业量 L。
220 480 210 560;
LL = L/X;                                   ←求就业系数,设其为 LL。
MFORM(TYPE = DIAG,NCOL = 4)LL;              ←对就业系数 LL 作对角矩阵变换。
MAT LLL = LL * KK;                          ←计算因公共投资而增加的就业量
                                              LLL。
PRINT KK LLL;                               ←输出 KK、LLL。
```

第十章 练习题

1. 请列出用于求解例题 3—6(对日本菲力普斯曲线的 OLS 估计)的 TSP 程序。

2. 请列出用于求解例题 3—7(对 VTR 普及率进行逻辑函数近似和 OLS 估计)的 TSP 程序。

3. 请列出用于求解例题 6—1(对导入虚拟变量的多元回归模型进行 OLS 估计)的 TSP 程序。

4. 例题 8—5 是对 A 国的消费函数(8—11)(联立方程模型)进行的二阶段最小二乘法估计,请列出求解该例题的 TSP 程序。

TSP 的问询地址

美国:TSP International
P.O. Box 61015 station A
Palo Alto,CA 94306 U.S.A.
phone:(415) 326 - 1927　FAX:(415) 328 - 4163
e-mail:chint@leland.stanford.edu
http://www.crl.com/~tspintl/

练习题解答

第一章

1.（1）	日本	0.32%
（2）	新加坡	1.99%
（3）	中国香港	1.66%
（4）	中国台湾	0.96%
（5）	韩国	0.91%
（6）	泰国	1.10%
（7）	印度尼西亚	1.54%
（8）	中国大陆	1.16%
（9）	巴基斯坦	2.99%
（10）	孟加拉国	2.18%

2.(1) 单位:美元/桶;美元/盎司

年末	原油价格 (北海油田现货价格)	黄金价格 (伦敦)
1985	—	—
1986	20.40	401.47
1987	16.83	429.18
1988	17.60	431.75
1989	21.20	399.92
1990	22.00	381.00
1991	21.33	359.10
1992	16.48	358.98
1993	16.03	368.68
1994	16.05	386.62
1995	—	—

图省略。

(2) 原油价格 $s^2=18.94$

$s=4.35$(美元/桶)

黄金价格 $s^2=1\ 828.73$

$s=42.76$(美元/桶)

(3) 原油价格 $CV=22.89\%$

黄金价格 $CV=11.07\%$

3.(1) 1990 年 $s^2=25\ 085.99$

$s=158.39$(1 000 日元/平方米)

$CV=139.6\%$

(2) 1995 年 $s^2=6\ 067.91$

$s=77.90$(1 000 日元/平方米)

$CV=92.69\%$

4.(1) $R=0.890\ 3$

(2) 在 5%以及 1%水平上显著

5.(1) $R=0.932$

(2) 在 5%、1%以及 0.1%水平上显著

6.(1) $R_s=0.786$ (2) $R_s=0.833$ (3) $R_s=0.333$

第二章

1. 1975年　0.285 2

   1980年　0.269 6

   1985年　0.186 0

   1990年　0.183 2

   1994年　0.207 2

2. 1985年　0.111 9

   1990年　0.117 4

   1995年　0.163 5

3.

产　　业	贡献度（%）	贡献度（%）
实际 GDP	108.25	100.00
(1) 农林水产业	7.49	6.92
(2) 矿业	4.18	3.86
(3) 制造业	48.02	44.36
(4) 建筑业	5.07	4.68
(5) 电力、煤气、自来水业	3.10	2.86
(6) 批发零售业	14.67	13.55
(7) 金融、保险、房地产业	13.50	12.47
(8) 运输、通信业	8.53	7.88
(9) 服务业与其他	1.87	1.73
(10) 政府服务业	1.82	1.68

4. (1) 拉氏价格指数　102.3

   帕氏价格指数　98.6

   (2) 费雪价格指数　100.4

   拉氏数量指数　96.5

   帕氏数量指数　93.1

   费雪数量指数　94.8

第三章

1. (1) $\hat{Y}=3.0107+0.886\ 61X$

   $R^2=0.998\ 3$

(2) $\hat{Y} = -15.498 + 0.948\ 72X$

$R^2 = 0.990\ 6$

(3) $\hat{Y} = -49.243 + 1.030\ 8X$

$R^2 = 0.997\ 0$

2.(1) 省略

(2) $\hat{Y} = 3.702\ 6 + 20\ 391\ \frac{1}{X}$

$R^2 = 0.985\ 6$

(3) 现台湾第 1 产业就业比率将来可能会降至 3.7%

3.(1) 63%

$$Y = \frac{63.0}{1 + e^{2.298\ 3 - 0.683\ 85X}}$$

$R^2 = 0.990\ 1$

(2) 46%

$$Y = \frac{46.0}{1 + e^{1.116\ 0 - 0.445\ 45X}}$$

$R^2 = 0.996\ 0$

(3) 40%

$$Y = \frac{40.0}{1 + e^{1.493\ 5 - 0.350\ 94X}}$$

$R^2 = 0.974\ 5$

4.(1) $Y = \frac{1\ 970}{1 + e^{1.032\ 1 - 0.032\ 173X}}$

$R^2 = 0.995\ 3$

饱和水平　19.7 亿人

(2) 2010 年　14.26 亿人

2020 年　15.43 亿人

2030 年　16.41 亿人

2040 年　17.20 亿人

2050 年　17.82 亿人

第四章

1.(1) $\hat{Y} = -62.999 + 0.189\ 80X_1 + 92.457X_2$

(2) $R^2=0.994\ 7$　$\bar{R}^2=0.993\ 1$

(3) $R_{Y2\cdot1}=0.931\ 8$

(4) $R_{Y1\cdot2}=0.983\ 7$

2. (1) 省略

(2) $\hat{Y}=-194.28+22.754X-0.265\ 82X^2$

(3) $R^2=0.927\ 4$　$\bar{R}^2=0.906\ 7$

3. (1) $\hat{Y}=25.532+0.757\ 80X_1-12.926X_2+0.927\ 36X_3$

(2) $R^2=0.999\ 1$　$\bar{R}^2=0.998\ 7$

(3) 增加 9 274 元

(4) $R_{Y1\cdot23}=0.998\ 6$

(5) $R_{Y2\cdot13}=-0.999\ 4$

(6) $R_{Y3\cdot12}=0.989\ 8$

4. (1) $\log Y=1.301\ 7+0.503\ 32\log L+0.571\ 65\log K+0.033\ 714t$

$\bar{R}^2=0.995\ 9$

(2) 技术进步率=3.4%（根据 t 的系数）

(3) $\log\dfrac{Y}{L}=1.6038+0.508\ 46\log\dfrac{K}{L}+0.036\ 705t$

$\bar{R}^2=0.9957$

(4) 技术进步率=3.7%（根据 t 的系数）

第五章

1. (1) 英国

$\hat{Y}=-168.34+0.578\ 30X$

[22.228] [0.041 48] ←标准误差

(−7.573) (13.942) ←t 值

$R^2=0.955\ 7$

法国

$\hat{Y}=-101.83+0.559\ 81X$

[4.181 7] [0.012 141] ←标准误差

(−24.350) (46.108) ←t 值

$R^2=0.995\ 8$

意大利

$\hat{Y} = -38.931 + 0.489\ 14X$

$[3.475\ 6]\ [0.027\ 159]$ ←标准误差

$(-11.201)\ (18.010)$ ←t 值

$R^2 = 0.973\ 0$

(2) 英国 (0.484, 0.672)

法国 (0.532, 0.587)

意大利 (0.428, 0.551)

(3) 英国 $\hat{Y}_0 = 178.64$

法国 $\hat{Y}_0 = 122.10$

意大利 $\hat{Y}_0 = 34.44$

(4) 英国 (166.5, 190.8)

法国 (119.4, 124.8)

意大利 (32.3, 36.6)

2. (1)、(2) $\hat{Y} = 68.767 + 0.266\ 25X_1 + 0.789\ 82X_2$

$[403.02]\ [0.052303]\ [0.050\ 723]$

$(0.171)\ (5.091)\ (15.571)$

$R^2 = 0.966\ 1 \quad \bar{R}^2 = 0.959\ 9 \quad F = 156.5$

3.

前期	后期	F 值
1971—1972	1973—1985	3.150
1971—1973	1974—1985	5.898
1971—1974	1975—1985	10.375
1971—1975	1976—1985	15.865
1971—1976	1977—1985	19.804
1971—1977	1978—1985	24.001
1971—1978	1979—1985	24.022
1971—1979	1980—1985	22.522
1971—1980	1981—1985	11.894
1971—1981	1982—1985	8.612
1971—1982	1983—1985	7.883
1971—1983	1984—1985	7.785

第六章

1. (1) $\hat{Y} = 11.008 + 0.924\ 87X$

$(1.105)\ (7.088)$

$R^2=0.782\ 1$

(2) $\hat{Y}=15.208+0.879\ 94X-12.683D$

$(1.969)\ (8.741)\qquad(-3.310)$

$\bar{R}^2=0.863\ 5$

2. (1) 省略

(2) $\hat{Y}=-4\ 522.4+0.831\ 11X$

$(-3.801)\qquad(4.410)$

$R^2=0.708\ 6$

(3) $\hat{Y}=-868.59+0.225\ 52X+807.55D$

$(-1.482)\quad(2.366)\qquad(8.293)$

$\bar{R}^2=0.965\ 4$

3. (1) $\hat{Y}=12.298+0.298\ 98X$

$(3.474)\quad(4.890)$

$R^2=0.614\ 5$

(2) $\hat{Y}=-18.015+0.777\ 22X+3.833\ 1D_1+3.710\ 2D_2+2.791\ 9D_3$

$(-6.020)\ (16.201)\qquad(11.034)\quad(10.343)\quad(10.106)$

$\bar{R}^2=0.954\ 7$

4. (1) $\hat{Y}=5.344\ 1+7.155\ 1S+2.156\ 7I-4.378\ 4M+6.669\ 8F_1+2.713\ 5F_2$

$(7.871)\ (14.626)\ (4.700)\ (-8.387)\ (9.143)\quad(3.661)$

$\bar{R}^2=0.969\ 4$

(2) 小规模非制造业中非管理职务女性的每月额外劳动时间

(3) 13.0 时间

5. (1) ① $\hat{Y}=6.938\ 6-0.002\ 162\ 9X$

$(7.791)\quad(-0.167\ 93)$

$R^2=0.003\ 1\ \ \bar{R}^2=-0.107\ 6$

② $\hat{Y}=5.040\ 5+0.031\ 252X-0.843\ 16D$

$(3.402)\qquad(1.264)\qquad(-1.545)$

$R^2=0.232\ 3\quad\bar{R}^2=0.040\ 4$

③ $\hat{Y}=4.253\ 7+0.044\ 840X-0.014\ 632DX$

$(2.893)\qquad(1.820)\qquad(-2.128)$

$R^2=0.363\ 5\quad\bar{R}^2=0.204\ 4$

④ $\hat{Y}=1.857\ 1+0.084\ 75X-0.140\ 65DX+9.207\ 2D$

$(4.295)\quad(11.715)\quad(-11.990)\quad(10.863)$

$R^2=0.9644\quad \bar{R}^2=0.9491$

(2) F 值 $=94.384$ 可见存在结构变化

第七章

1.(1) $\hat{T}=-26.093+0.28073Y$

$(-10.793)\quad(33.314)$

$R^2=0.9840\quad DW=0.610$

(2) 存在一阶正的系列相关

(3) ⓐ $\hat{T}=-31.402+0.29711Y$

$(-5.910)\quad(17.629)$

$R^2=0.9481$（根据 CO 变换后的公式计算）

$DW=1.878$

ⓑ $\hat{T}=-31.536+0.29749Y$

$(-5.806)\quad(17.314)$

$R^2=0.9463$（根据 CO 变换后的公式计算）

$DW=1.893$

ⓒ $\hat{T}=-25.574+0.28025Y$

$(-6.443)\quad(20.687)$

$R^2=0.9435$（根据 PW 变换后的公式计算）

$DW=1.676$

2.(1) $\hat{Y}=-9.4287+0.93587X$

$(-3.765)\quad(125.341)$

$R^2=0.9978\quad DW=0.523$

(2) 存在一阶正的系列相关

(3) ⓐ $\hat{Y}=-13.936+0.94841X$

$(-2.022)\quad(50.168)$

$R^2=0.9871$（根据 CO 变换后的公式计算）

$DW=2.097$

ⓑ $\hat{Y}=-13.853+0.94820X$ ←迭代次数 $=2$ 次

$(-2.069)\quad(51.508)$

$R^2=0.9877$（根据 CO 变换后的公式计算）

$DW=2.081$

ⓒ $\hat{Y}=-13.850+0.948\ 19X$ ←迭代次数=3 次

$(-2.071)\quad(51.556)$

$R^2=0.987\ 7$ （根据 CO 变换后的公式计算）

$DW=2.080$

（4）$\hat{Y}=-7.897\ 7+0.933\ 43X$

$(-1.515)\ (61.071)$

$R^2=0.986\ 8$（根据 PW 变换后的公式计算）

$DW=1.977$

第八章

1.（1）$\hat{Q}=10.668\ 9+0.015\ 007\ 4\dfrac{Y}{P_1}-336.896\dfrac{P_2}{P_1}$

$(1.239)\qquad(9.605)\qquad\qquad(-5.868)$

$\bar{R}^2=0.875\ 1\quad DW=2.055$

（2）$\hat{U}=0.015\ 007\log(-10.831+Q_1)+0.984\ 99\log(-22\ 449+Q_2)$

（3）户均啤酒年消费量的理论值

单位：l

年份	理论值 $\hat{Q}_1$
1980	46.85
1981	44.51
1982	46.91
1983	47.77
1984	44.34
1985	44.41
1986	45.07
1987	46.51
1988	49.40
1989	51.34
1990	51.98
1991	54.65
1992	55.85
1993	56.01

图略

2. (1) ①过剩识别

②过剩识别

③过剩识别

(2) ① $\hat{C}_t = -18.199 + 0.279\,66Y_t + 0.608\,20C_{t-1}$

$(-1.863)\quad(3.123)\qquad(4.344)$

$\bar{R}^2 = 0.997\,6$

② $\hat{I}_t = -27.584 + 0.215\,75Y_t + 0.323\,04I_{t-1}$

$(-5.037)\quad(6.076)\qquad(2.652)$

$\bar{R}=0.996\,4$

③ $\hat{M}_t = -98.205 + 0.626\,56Y_t$

$(-13.983)\quad(41.387)$

$R^2 = 0.994\,8$

3.

年份 t	(1) Z_t 的增加率为 5% 的情形		(2) Z_t 的减少率为 3% 的情形	
	C_t 的预测值 $\hat{C}_t$	Y_t 的预测值 $\hat{Y}_t$	C_t 的预测值 $\hat{C}_t$	Y_t 的预测值 $\hat{Y}_t$
1996	108.569 3	154.769 3	105.517 7	148.197 7
1997	111.807 5	160.317 5	104.586 8	145.986 4
1998	115.031 1	165.966 6	103.187 9	143.345 5
1999	118.354 8	171.837 1	101.659 2	140.612 1
2000	121.823 6	177.980 0	100.117 0	137.901 3

第九章

1. (1) ⓐ30　ⓑ60　ⓒ100　ⓓ20　ⓔ40　ⓕ200

(2) $A = \begin{pmatrix} 0.6 & 0.15 \\ 0.2 & 0.6 \end{pmatrix}$

(3) ①0.4（农业）　0.15（工业）　②0.6（农业）　0.85（工业）

(4) $B = \begin{pmatrix} 1.642\,6 & 0.301\,7 \\ 0.569\,9 & 2.145\,5 \end{pmatrix}$

（5）生产诱发额

部门＼项目	消　费	投　资	出　口
农　业	59.54	14.98	25.48
工　业	90.04	39.89	70.06

生产诱发系数

部门＼项目	消　费	投　资	出　口
农　业	0.661 5	0.449 5	0.636 9
工　业	1.000 5	1.329 8	1.751 6

生产诱发依存度

部门＼项目	消　费	投　资	出　口
农　业	0.595 4	0.149 8	0.254 8
工　业	0.450 2	0.199 5	0.350 3

（6）进口诱发额

部门＼项目	消　费	投　资	出　口
农　业	39.69	9.99	10.32
工　业	15.89	7.04	7.07

进口诱发系数

部门＼项目	消　费	投　资	出　口
农　业	0.441 0	0.333 0	0.258 0
工　业	0.176 6	0.234 7	0.176 8

（7）附加价值诱发额

部门＼项目	消　费	投　资	出　口
农　业	11.91	3.00	5.10
工　业	22.51	9.97	17.52

附加价值诱发系数

部门＼项目	消费	投资	出口
农业	0.132 3	0.099 9	0.127 4
工业	0.250 1	0.332 4	0.437 9

2.(1) $A=\begin{bmatrix}0.15 & 0.07 & 0.03 & 0.01\\ 0.20 & 0.40 & 0.35 & 0.12\\ 0.05 & 0.06 & 0.01 & 0.03\\ 0.10 & 0.12 & 0.16 & 0.14\end{bmatrix}$

(2) $B=\begin{bmatrix}1.225\,0 & 0.1610 & 0.100\,5 & 0.040\,2\\ 0.513\,2 & 1.856\,4 & 0.718\,7 & 0.290\,1\\ 0.100\,0 & 0.129\,8 & 1.068\,2 & 0.056\,5\\ 0.232\,7 & 0.301\,9 & 0.310\,7 & 1.218\,5\end{bmatrix}$

(3) $M=\begin{bmatrix}0.235\,3\\ 0.082\,7\\ 0.000\,0\\ 0.015\,4\end{bmatrix}$

(4)、(5)

供给部门＼需要部门	农业	制造业	建筑业	服务业	行和	感应度系数
农业	1.158 9	0.109 2	0.066 5	0.026 5	1.361 2	0.713 5
制造业	0.417 3	1.717 8	0.605 6	0.244 1	2.984 7	1.564 6
建筑业	0.090 1	0.117 8	1.058 8	0.052 7	1.319 3	0.691 6
服务业	0.206 0	0.269 4	0.284 0	1.206 0	1.965 5	1.030 3
列和	1.872 3	2.214 2	2.014 9	1.529 3		
影响力系数	0.981 5	1.160 7	1.056 2	0.801 7		

(6)

部门＼项目	生产诱发额			生产诱发系数		
	消费	投资	出口	消费	投资	出口
农业	23.62	10.09	6.29	0.13 12	0.088 5	0.116 4
制造业	122.05	98.28	79.67	0.678 0	0.862 1	1.475 4
建筑业	12.88	81.30	5.81	0.071 6	0.713 2	0.107 6
服务业	141.69	36.33	21.98	0.787 2	0.318 7	0.407 0

(7)

部门＼项目	进口诱发额			进口诱发系数		
	消费	投资	出口	消费	投资	出口
农业	7.268	3.105	1.627	0.040 38	0.027 24	0.030 13
制造业	11.008	8.865	3.127	0.061 16	0.077 76	0.057 91
建筑业	0.000	0.000	0.000	0.000 00	0.000 00	0.000 00
服务业	2.214	0.568	0.218	0.012 30	0.004 98	0.004 04

(8)

部门＼项目	附加价值诱发额			附加价值诱发系数		
	消费	投资	出口	消费	投资	出口
农业	11.81	5.05	3.14	0.065 6	0.044 3	0.058 2
制造业	42.72	34.40	27.88	0.237 3	0.301 7	0.516 4
建筑业	5.80	36.59	2.62	0.032 2	0.320 9	0.048 4
服务业	99.18	25.43	15.38	0.551 0	0.223 1	0.284 9

(9)

部门	总产量的增加比例	就业量的增加比例
农业	1.197	6.584
制造业	10.900	17.440
建筑业	19.058	40.021
服务业	5.113	14.316

第十章

1. OPTIONS CRT;

```
FREQ A;
SMPL 86 95;
LOAD P;
0.6  0.1  0.7  2.3  3.1  3.3  1.6  1.3  0.7  -0.1;
LOAD U;
2.8  2.8  2.5  2.3  2.1  2.1  2.2  2.5  2.9  3.2;
GRAPH U P;
GENR UU=1/U;
```

OLSQ P C UU;

2. OPTIONS CRT;

FREQ A;

SMPL 78 95;

TREND T; ←用 TREND 指令，制作趋势变量 T。

LOAD Y;

1.3 2.0 2.4 … 72.5 73.7;

GENR Y76 = LOG (76/Y − 1);

OLSQ Y76 C T;

GENR Y77 = LOG (77/Y − 1);

OLSQ Y77 C T;

GENR Y78 = LOG (78/Y − 1);

OLSQ Y78 C T;

GENR Y79 = LOG (79/Y − 1)

OLSQ Y79 C T;

GENR Y80 = LOG (80/Y − 1)

OLSQ Y80 C T;

3. OPTIONS CRT;

FREQ A;

SMPL 85 95;

LOAD Y;

116 116 106 99 103 105 96 105 78 120 107;

LOAD X;

232 228 212 209 208 206 203 209 213 220 211;

GRAPH X Y;

OLSQ Y C X;

LOAD D;

0 0 0 0 0 0 0 0 1 0 0;

OLSQ Y C X D;

4. OPTIONS CRT;

FREQ A;

SMPL 84 95;

LOAD Y;

```
.100 108 110 117 116 124 131 136 134 142 149;
LOAD CC;
70 76 82 84 87 87 91 95 98 97 102 105;
LOAD Z;
.24 26 26 30 29 33 36 38 37 40 44;
SMPL 85 95;
2SLS CC C Y CC (-1) INVR C CC (-1) Z;
```

—缺省值用句点“.”输入。

参考文献

石村貞夫（1992）：『すぐわかる多変量解析』東京図書。
岩田曉一（1982）：『計量經済学』有斐閣。
荏開津典生（1985）：『農業統計学』明文書房。
岡澤宏（1980）：『計量經済学概説』啓文社。
小尾惠一郎（1972）：『計量經済学入門』日本評論社。
加納悟・浅子和美（1992）：『入門|經済のための統計学』日本評論社。
刈屋武昭監修、日本銀行調査統計局編（1985）：『計量經済分析の基礎と応用』東洋經済新報社。
刈屋武昭・勝浦正樹（1994）：『統計学』東洋經済新報社。
國友直人（1992・94）：『現代統計学（上）（下）』日經文庫。
黒田昌裕（1984）：『実證經済学入門』日本評論社。
佐和隆光（1980）：『数量經済分析の基礎』筑摩書房。
J. Johnston（1984）：*Econometric Methods*, 3rd ed., Mc-

Graw Hill（竹内啓他訳『計量經済学の方法（上）（下）』東洋經済新報社、1975年、訳は第2版）。
高橋一編（1993）：『計量經済学』八千供出版。
田中勝人（1996）：『經済統計』岩波書店。
千田亮吉（1989）：『数量經済分析入門』文眞堂。
辻村江太郎（1981）：『計量經済学』岩波書店。
土居英二・浅利一郎・中野親徳編（1996）：『はじめよう地域産業連関分析』日本評論社。
鳥居泰彦（1994）：『はじめての統計学』日本經済新聞社。
東京大学教養学部統計学教室編（1994）：『人文・社会科学の統計学』東京大学出版会。
中村隆英・新家健精・美添泰人・豊田敬（1992）：『經済統計入門（第2版）』東京大学出版会。
縄田和満（1997）：『TSPりよる計量經済分析入門』朝倉書店。
畠中道雄（1991）：『計量經済学の方法』創文社。
馬場正雄編（1970）：『計量經済学入門』有斐閣。
伴金美・中村二朗・跡田直澄（1988）：『エコノヌトリックス』有斐閣。
G. S. Maddala（1992）：*Introduction to Econometrics*，2nd ed，Macmillan（和合肇訳著『計量經済分析の方法（第2版）』シーエーピー出版、1996年）。
溝口敏行（1985）：『經済統計論（第3版）』東洋經済新報社。
蓑谷千凰彦（1985）：『回帰分析のはなし』東京圖書。
蓑谷千凰彦（1992）：『計量經済学の新しい展開』多賀出版。
蓑谷千凰彦（1997a）：『計量經済学』多賀出版。
蓑谷千凰彦（1997b）：『計量經済学（第3版）』東洋經済新報社。
宮川公男（1991）：『基本統計学』（新版）有斐閣。
宮訳一編（1995）：『産業連開分析入門（新版）』日經文庫。
森田優三・久次智雄（1993）：『新統計概論（改訂版）』日本評論社。
森棟公夫（1990）：『統計学入門』新世社。
山本拓（1995）：『計量經済学』新世社。
吉野直行・高橋徹（1990）：『パソコン計量經済入門』多賀出版。
和合肇・伴金美（1995）：『TSPによる經済データの分析（第2版）』東京大学出版会。