



普通高等院校经济管理类“十二五”应用型规划教材  
经济管理类专业基础课系列

免费提供  
授课用  
电子课件

ECONOMETRICS

# 计量经济学基础

张兆丰 主编



机械工业出版社  
China Machine Press

普通高等院校经济管理类“十二五”应用型规划教材  
经济管理类专业基础课系列

ECONOMETRICS

# 计量经济学基础

张兆丰 编著



机械工业出版社  
China Machine Press

## 图书在版编目 (CIP) 数据

---

计量经济学基础 / 张兆丰主编. —北京: 机械工业出版社, 2013. 4  
(普通高等院校经济管理类“十二五”应用型规划教材·经济管理类专业基础课系列)

ISBN 978-7-111-42076-7

I. 计… II. 张… 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 (2013) 第 071883 号

---

### 版权所有·侵权必究

封底无防伪标均为盗版

本书法律顾问 北京市展达律师事务所

计量经济学是高等院校经济管理类专业的核心课程, 为了更好地体现应用型本科教学的规律和学习特点, 我们组织编写了本书。

本书以经典计量经济学内容为主, 着重体现计量经济学理论、方法与应用的结合。通过实际案例, 深入浅出地讲解了计量经济学的基本理论和基本方法, 在各章后面设计了上机实验的内容, 一方面突出计量经济学的应用性, 另一方面方便教师组织实验课的教学。

本书共 12 章, 主要包括符合古典假设的一元和多元线性回归模型, 违反古典假设的多重共线性、异方差性和自相关性, 还介绍了滞后变量模型、虚拟变量模型、时间序列模型、联立方程模型, 较好地体现了系统性、实用性与适用性。

本书以初级水平为主, 包含了高等院校经济学科本科计量经济学课程教学基本要求的全部内容, 可以作为高等院校经济学科和管理学科专业本科生计量经济学课程的学习用书, 也可供经济管理工作者和研究人员参考。

机械工业出版社 (北京市西城区百万庄大街 22 号 邮政编码 100037)

责任编辑: 王振杰 版式设计: 刘永青

北京市荣盛彩色印刷有限公司印刷

2013 年 5 月第 1 版第 1 次印刷

185mm×260mm·14 印张

标准书号: ISBN 978-7-111-42076-7

定 价: 35.00 元

---

凡购本书, 如有缺页、倒页、脱页, 由本社发行部调换

客服热线: (010) 88379210 88361066

投稿热线: (010) 88379007

购书热线: (010) 68326294 88379649 68995259

读者信箱: hzjg@hzbook.com

# Preface 前 言

计量经济学是教育部规定的高等院校经济管理类专业核心课程之一，是人们对经济现象进行定量分析的基本方法论，是经济管理类专业重要的专业基础课。

本书根据经济管理类专业的特点，吸收国内外优秀计量经济学教材的成果，总结作者多年的计量经济学教学与科研的经验编写而成。

本书主要有以下几个特点：

(1) 实用性。根据经济管理类专业的教学目标和要求安排教材内容，力求使学生较容易地理解计量经济学的基本概念，掌握计量经济学的基本方法，并结合大量的实际问题，用计量经济学的方法进行估计和解释，引导学生分析和解决问题。

(2) 简明性。本书简明扼要，通俗易懂，便于自学。在编写过程中，作者省略了许多复杂公式的推导过程，用通俗的实例和实验引导学生得出相关的结论。

(3) 适用性。在编写过程中，本书十分注重本科学生的特点，在层次上更加适合本科学生的教学工作。本书在各章中安排了一定数量的例题和案例分析，特别是设计了上机实验的内容，以国家统计局公布的统计数据作为材料，引导学生运用计量经济学的方法解决实际问题，体现了理论与实际相结合的教学原则。

(4) 为了更好地配合教师的教学工作，以期达到理想的教学效果，本书免费提供丰富的配套教学资源，包括数据、上机实验、习题答案、PPT 课件，以上教学资源均可在 <http://www.hzbook.com/ps/> 免费下载。

本书由张兆丰主编。其中张兆丰负责本书内容的总体设计和总纂，并编写了第 3~10 章；陶芸编写了第 1 章和第 2 章；黄春花编写了第 11 章和第 12 章；张兆丰还编写了附录内容。

本书在编写过程中，参阅了大量的国内外文献和相关资料，在此向这些文献的作者表示感谢。

在本书的出版过程中，机械工业出版社华章公司的编辑给予了很多帮助和支持，高伟编辑在其中做了大量的工作，在此一并表示感谢。

由于编者水平有限，本书中难免有疏漏和不当之处，恳请同行和读者批评指正。

编者

2012 年 12 月 7 日

# 教学建议 Suggestion

计量经济学是高等院校经济管理类专业的核心课程，是培养高素质应用型人才的重要专业基础课，任课教师要充分认识到这门课程的重要性，认真准备，切实提高这门课程的教学质量。

计量经济学作为方法论学科，是一门应用性较强的课程，学好这门课有利于提高学生的实践能力和创新能力，对培养高素质应用型人才起到非常重要的作用。为了实现这样的目标，教师在教学中需要重视案例教学，提高学生的分析能力；加强实验教学，培养学生的动手能力和创新能力；运用多媒体方式进行教学，提高教学效率。

在这样的框架下，按 64 课时进行课时分配，下面的课时分配表仅供参考。

计量经济学课时分配表

	教学内容	理论课时	实验课时
第 1 章	导论	4	
第 2 章	回归与回归分析	2	
第 3 章	一元线性回归模型	6	4
第 4 章	一元线性回归模型的扩展	2	1
第 5 章	多元线性回归模型	4	2
第 6 章	多重共线性	2	2
第 7 章	异方差	4	2
第 8 章	自相关	4	2
第 9 章	分布滞后模型与自回归模型	4	2
第 10 章	虚拟变量回归	2	1
第 11 章	时间序列计量经济模型	4	2
第 12 章	联立方程组模型	4	2
	复习	2	
	合计	44	20
	总计	64	

# Contents 目 录

前 言		
教学建议		
第 1 章 导论 .....	1	
1.1 计量经济学的产生与 发展 .....	1	
1.2 计量经济学的性质 .....	2	
1.3 计量经济学的研究方法 .....	3	
1.4 数据 .....	5	
1.5 预备知识: 统计学基础 .....	7	
本章小结 .....	11	
学习建议 .....	11	
关键术语 .....	12	
思考练习 .....	12	
第 2 章 回归与回归分析 .....	13	
2.1 回归的基本问题 .....	13	
2.2 相关分析 .....	15	
2.3 一元线性回归分析 .....	17	
本章小结 .....	22	
学习建议 .....	22	
关键术语 .....	22	
思考练习 .....	22	
第 3 章 一元线性回归模型 .....	23	
3.1 普通最小二乘法 .....	23	
3.2 样本回归直线的代数 性质 .....	29	
3.3 拟合优度的度量 .....	30	
3.4 古典假定 .....	32	
3.5 OLS 估计量的统计 性质 .....	34	
3.6 参数的检验与区间 估计 .....	36	
3.7 预测 .....	38	
3.8 案例分析 .....	40	
本章小结 .....	44	
学习建议 .....	44	
关键术语 .....	44	
思考练习 .....	44	
第 4 章 一元线性回归模型的扩展 .....	48	
4.1 过原点的回归 .....	48	
4.2 对数模型 .....	51	
4.3 倒数模型 .....	54	
4.4 模型函数形式的选择 .....	56	
4.5 案例分析 .....	57	
本章小结 .....	58	
学习建议 .....	59	
关键术语 .....	59	
思考练习 .....	59	
第 5 章 多元线性回归模型 .....	62	
5.1 多元线性回归模型的 设定 .....	62	
5.2 多元线性回归模型的 估计 .....	64	

5.3	多元线性模型 OLS 估计量的性质 .....	68	思考练习 .....	115
5.4	参数的检验与区间估计 .....	72	第 8 章 自相关 .....	117
5.5	多元线性回归模型的预测 .....	74	8.1 自相关概述 .....	117
5.6	案例分析 .....	75	8.2 自相关产生的原因 .....	118
	本章小结 .....	77	8.3 自相关的后果 .....	120
	学习建议 .....	77	8.4 自相关的检验 .....	122
	关键术语 .....	77	8.5 自相关的修正 .....	129
	思考练习 .....	77	8.6 案例分析 .....	131
第 6 章	多重共线性 .....	80	本章小结 .....	133
6.1	多重共线性的含义 .....	80	学习建议 .....	134
6.2	产生多重共线性的原因 .....	82	关键术语 .....	134
6.3	多重共线性对 OLS 估计量的影响 .....	83	思考练习 .....	134
6.4	多重共线性的检验 .....	86	第 9 章 分布滞后模型与自回归模型 .....	136
6.5	多重共线性的修正 .....	88	9.1 滞后变量模型 .....	136
6.6	案例分析 .....	91	9.2 分布滞后模型的估计 .....	139
	本章小结 .....	94	9.3 自回归模型的构建 .....	144
	学习建议 .....	94	9.4 自回归模型的估计 .....	148
	关键术语 .....	95	9.5 案例分析 .....	149
	思考练习 .....	95	本章小结 .....	151
第 7 章	异方差 .....	97	学习建议 .....	151
7.1	异方差概述 .....	97	关键术语 .....	152
7.2	异方差产生的原因 .....	98	思考练习 .....	152
7.3	异方差产生的后果 .....	99	第 10 章 虚拟变量回归 .....	154
7.4	异方差的检验 .....	100	10.1 虚拟变量 .....	154
7.5	异方差的修正 .....	107	10.2 虚拟变量的回归模型 .....	155
7.6	案例分析 .....	111	10.3 参数的结构变化 .....	160
	本章小结 .....	114	10.4 案例分析 .....	163
	学习建议 .....	114	本章小结 .....	164
	关键术语 .....	115	学习建议 .....	165

	关键术语·····	165			
	思考练习·····	165			
第 11 章	时间序列计量经济模型 ···	167			
	11.1 时间序列计量经济学				
	的基本概念·····	167			
	11.2 时间序列平稳性				
	检验·····	169			
	11.3 协整·····	172			
	11.4 案例分析·····	173			
	本章小结·····	178			
	学习建议·····	178			
	关键术语·····	178			
	思考练习·····	178			
第 12 章	联立方程组模型 ·····	180			
	12.1 联立方程组模型的				
	基本概念·····	180			
	12.2 联立方程组模型中				
	变量的内涵·····	181			
	12.3 联立方程组模型的				
	分类·····	182			
	12.4 联立方程组模型的				
	识别·····	184			
	12.5 联立方程组模型的				
	估计·····	189			
	12.6 案例分析·····	191			
	本章小结·····	194			
	学习建议·····	195			
	关键术语·····	195			
	思考练习·····	195			
	附录 A 标准正态分布表 ·····	197			
	附录 B $t$ 分布表 ·····	199			
	附录 C $\chi^2$ 分布表·····	201			
	附录 D $F$ 分布表 ·····	204			
	附录 E $DW$ 检验临界值表·····	208			
	参考文献 ·····	213			



**学习目标**

- 了解计量经济学的产生与发展过程；
- 理解计量经济学的含义；
- 了解计量经济学的研究过程；
- 掌握统计学的基本方法。

计量经济学是现代经济学的重要分支，在经济学、金融学、管理学、营销学以及一些相关学科的研究中，定量分析用得越来越多，而计量经济学是进行定量分析的重要工具，“在大多数大学和学院中，计量经济学的讲授已成为经济学课程表中最有权威的一部分”（R. 克莱因）。

## 1.1 计量经济学的产生与发展

在现实的经济生活中，我们经常要对诸如国民生产总值（GNP）、失业、通货膨胀、进口、出口等经济现象进行定量分析，如何对这些经济现象进行测度呢？

计量经济学是利用经济理论、数学、统计推断等工具对经济现象进行分析的一门社会科学。计量经济学运用数理统计知识分析经济数据，对构建于数理经济学基础之上的数学模型提供经验支持，并得出数量结果。

计量经济学（Econometrics）这个词是挪威经济学家、第1届诺贝尔经济学奖获得者弗里希（R. Frisch）1926年在《论纯经济问题》一文中，按照“生物计量学”（Biometrics）一词的结构仿造出来的。Econometrics 一词的本意是指“经济度量”，研究经济现象和经济关系的计量方法，因此，Econometrics 有时也译为“经济计量学”。将 Econometrics 译为计量经济学，是为了强调计量经济学是一门经济学科，不仅要研究经济现象的计量方法，而且要研究经济现象发展变化的数量规律。

计量经济学的产生源于对经济问题的定量研究，这是社会经济发展到一定阶段的客观需要。经济现象本来就充满数量关系，人们很早就探索用定量的方式研究经济问题。早在17世纪，英国经济学家、统计学家威廉·配第在《政治算术》中就运用统计方法研究社会经济问题，主张用“数字、重量和尺度”来阐述经济现象。以后的相当一段时间内，经济学家们也力图运用数学方法研究经济活动，用数学语言和公式表达经济范畴和经济规律。但这些都未形成计量经济学。计量经济学作为经济学的一门独立学科被正式确立，其标志一般被认为是1930年12月弗里希和丁伯根（J. Tinbergen）等经济学家在美国克里夫兰发起成立国际计量经济学会。

第二次世界大战以后，计量经济学在西方各国的影响迅速扩大，发展成为经济学的重要分支。特别是20世纪40~60年代，经典计量经济学逐步完善并得到广泛应用。美国著名经济学家、诺贝尔经济学奖获得者萨缪尔森（P. Samuelson）认为：“第二次世界大战后的经济学是计量经济学的时代。”事实上，在诺贝尔经济学奖获得者中，相当一部分都是计量经济学家。

20世纪70年代以来，计量经济学的理论和应用又进入一个新的阶段。首先是计算机的广泛应用和新的计算方法的大量提出，所使用的计量经济模型的规模越来越大。更重要的是，非经典计量经济学的理论和应用有了新的突破。微观计量经济学、非参数计量经济学、时间序列计量经济学和动态计量经济学等的提出，使计量经济学产生了新的理论体系，协整理论、面板数据、对策论、贝叶斯方法等理论在计量经济学中的应用已成为新的研究课题。

应该看到，计量经济学的发展是与现代科学技术成就结合在一起的，它反映了社会化大生产对各种经济因素和经济活动进行数量分析的客观要求。经济学从定性研究向定量分析的发展，是经济学逐步向更加精密、更加科学发展的表现。正如马克思所强调的：一种科学只有成功地运用了数学以后，才算达到了完善的地步。因此另一位获得诺贝尔经济学奖的经济学家克莱因（R. Klein）认为：“计量经济学已经在经济学科中居于最重要的地位。”

### 1.2 计量经济学的性质

计量经济学的奠基人弗里希在《计量经济学》的创刊词中说道：“用数学方法探讨经济学可以从好几个方面着手，但任何一方面都不能与计量经济学混为一谈。计量经济学与经济统计学绝非一码事；它也不同于我们所说的一般经济理论，尽管经济理论大部分都具有一定的数量特征；计量经济学也不应被视为数学应用于经济学的同义语。经验表明，统计学、经济理论和数学这三者对于真正了解现代经济生活中的数量关系来说，都是必要的，但各自并

非充分条件。而三者结合起来，就有力量，这种结合便构成了计量经济学。”

后来，美国著名计量经济学家克莱因也认为：计量经济学是数学、统计技术和经济分析的综合。也可以说，计量经济学不仅对经济现象加以测量，而且是根据一定的经济理论进行计量的。

尽管这些经济学家对计量经济学定义的表述各不相同，但可以看出，计量经济学不是对经济的一般度量，它与经济理论、统计学、数学都有密切的关系。事实上，计量经济学是以经济理论和经济数据为依据，运用数学、统计学的方法，通过建立数学模型来研究经济数量关系和规律的一门经济学科。应当注意，计量经济学所研究的主体是经济现象及其发展变化的规律，所以它是一门经济学科。计量经济学当然会运用大量的数学方法，特别是许多数理统计方法，但是数学在这里只是工具，而不是研究的主体。

计量经济学的用途或目的主要有两个方面：其一是理论检验，这是计量经济学用途最为主要和可靠的方面，也是计量经济学本身的一个主要内容；其二是预测应用。从理论研究和方法的最终目的看，预测（包括政策评价）当然是计量经济学的最终任务，必须进行学习和了解，但其预测的可靠性或有效性也是我们应该十分注意的。

### 1.3 计量经济学的研究方法

运用计量经济方法研究经济问题一般可以分为以下步骤：理论或假说的陈述→建立理论的数学模型→建立理论的计量经济学模型→抽样、收集数据→估计回归系数→参数的假设检验→模型的应用。

为了阐明计量经济学的方法论，我们首先来考虑这样一个问题：居民的消费支出与居民的收入之间有什么关系？或者说居民的收入不同对居民的消费支出会有影响吗？

#### 1. 理论或假说的陈述

根据经济学中的凯恩斯消费理论可知：随着收入水平的提高，消费也会增加；但消费的增加不及收入增加得多。那么，这个理论我们怎么来验证呢？当居民的家庭收入变动后，会引起居民的消费支出发生多大的变化呢？

这个问题可以用计量经济学的方法来回答。我们可以建立相应的模型来“计量”因收入变化而使消费变化的程度。这个问题中涉及两个经济变量：收入和消费，由经济理论可知，收入影响消费，即收入是“自变量”，消费是“因变量”。我们应该用什么样的函数形式来描述这两个经济变量的关系呢？

#### 2. 建立理论的数学模型

经济学理论和大量事实证明，收入与消费是线性关系。于是，我们建立数学模型：

$$Y = \beta_0 + \beta_1 X \quad (1-1)$$

式中,  $Y$  为消费;  $X$  为收入;  $\beta_0$ 、 $\beta_1$  为回归系数。

我们将式 (1-1) 称为一元线性回归方程。在这个方程中, 斜率项系数有着特定的经济学意义——边际消费倾向。根据斜率项系数的几何意义可知:  $\beta_1 = \Delta Y / \Delta X$ , 即消费的增量比收入的增量, 表示边际消费倾向。

### 3. 建立理论的计量经济学模型

由于消费是随机的, 故这两个变量之间的关系不会表现出像式 (1-1) 那样严格的函数关系。也就是说, 式 (1-1) 是“近似”地表示消费与收入的关系, 那么, 这两个经济变量之间的真实关系应该是:

$$Y = \beta_0 + \beta_1 X + u \quad (1-2)$$

式中,  $u$  为误差项。

我们将式 (1-2) 称为一元线性回归模型。与式 (1-1) 相比, 式 (1-2) 多出一个误差项, 这是因为对于同一收入水平 ( $X$ ) 的居民, 他们的消费 ( $Y$ ) 也会有差异, 有非常多的偶然因素会影响到消费行为, 这些因素都应该归结到误差项当中。

显然, 这样的误差是随机误差。

### 4. 抽样、收集数据

式 (1-1) 和式 (1-2) 描述的是总体的情形。我们知道, 总体一般来说是不可全面观测的, 虽然斜率项系数表示边际消费倾向, 但是我们相信, 总体中一部分人群的消费与收入的关系和总体人群的消费与收入的关系具有相同的特性, 可以建立相同形式的样本一元线性回归方程和模型。于是, 我们抽样并收集样本数据, 并用样本数据得到样本的斜率项系数, 即样本的边际消费倾向; 再用样本边际消费倾向推断总体边际消费倾向, 这个过程是可以实现的。

### 5. 估计回归系数

收集到样本数据后, 我们可以用相应的统计方法得到样本的截距项系数和斜率项系数。假设我们用某一个样本数据得到如下结果:

$$\hat{Y} = 235.60 + 0.65X \quad (1-3)$$

式中,  $\hat{Y}$  读作“Y 帽”, 其含义是消费 ( $Y$ ) 的估计值。

在这个结果中, 斜率项系数为 0.65, 即样本边际消费倾向是 0.65, 表示收入增加 1 元, 消费会增加 0.65 元。

如何得到式 (1-3) 的结果, 我们会在第 3 章中介绍。

### 6. 参数的假设检验

式 (1-3) 得到的结果是一个样本结果, 样本结果是带有偶然性的, 那么这样的—个结

果是偶然的吗?

为什么要提出这样的问题呢?这是因为,由样本得到的斜率项系数为0.65,是不等于0的,这是一个偶然的的结果吗?或者说,我们是不是偶然得到了一个不等于0的斜率项系数呢?如果不是,会有怎样的后果呢?

这个问题的另外一种表达方式是,由样本的这个结果能判断总体的 $\beta_1$ 也不等于0吗?我们建立模型式(1-2)的含义是“计量”X对Y影响的程度,如果 $\beta_1=0$ ,则式(1-2)变为 $Y=\beta_0+u$ ,这说明X没有对Y产生影响,这个结果显然与我们最初建立模型的意图是不相符的,或者说建立这样的模型是不可靠的。

这样的一个问题就是参数的假设检验。如果通过检验可以证明 $\beta_1 \neq 0$ ,那么说明我们建立的模型式(1-2)是可靠的。

由于斜率项系数为0.65,也符合经济学意义,故模型的设定是合理的。

## 7. 模型的应用

如果通过了检验,说明模型是可靠的,那么我们就可以对模型进行应用了。模型的应用主要有两个方面:一是对总体的系数做估计。例如,在消费模型中,斜率项系数表示边际消费倾向,我们只能得到样本的边际消费倾向,我们也可以运用统计学的方法对总体的边际消费倾向进行估计。另外一个预测。对于样本以外的X值,我们可以通过样本方程预测其对应的Y值,例如,当收入(X)达到8000元时,对应的消费大约为5435.60元。

## 1.4 数据

由上述分析可知,要“计量”收入对消费的影响,必须要有数据,不同的现象表现出来的数据类型是不一样的。最常用的数据有时间序列数据、截面数据和面板数据。

### 1. 时间序列数据

时间序列数据(time series data)是同一总体在不同时间上的统计数据,比如不同年份的GNP、失业、就业、货币供给、政府赤字等数据就可以构成不同的时间序列。表1-1为中国2011年社会消费品零售总额与居民消费价格指数各个月份的数据,这些数据按照时间先后的原则进行排列,反映了我国(同一总体)2011年各个月份(不同时间)社会消费品零售总额与居民消费价格指数的情况。

表 1-1 中国 2011 年社会消费品零售总额与居民消费价格指数

月份	社会消费品零售总额(亿元)	居民消费价格指数(上月=100)
2011年1月	15249	101
2011年2月	13769.1	101.2
2011年3月	13588	99.8

(续)

月份	社会消费品零售总额 (亿元)	居民消费价格指数 (上月=100)
2011年4月	13 649	100.1
2011年5月	14 696.8	100.1
2011年6月	14 565.1	100.3
2011年7月	14 408	100.5
2011年8月	14 705	100.3
2011年9月	15 865.1	100.5
2011年10月	16 546.4	100.1
2011年11月	16 128.9	99.8
2011年12月	17 739.7	100.3

资料来源：中经网统计数据库。

## 2. 截面数据

截面数据 (cross-sectional data) 是不同总体在同一时间截面上的调查数据。例如, 各国或各地区的工业普查数据、人口普查数据等。表 1-2 为 2010 年我国华北地区工资与存款的相关数据, 是不同地区 (总体) 在同一时间截面 (2010 年) 的数据。

表 1-2 2010 年我国华北地区的工资与存款

地区	年平均工资 (元)	存款余额 (亿元)	地区	年平均工资 (元)	存款余额 (亿元)
北京	65 158	7 477.7	山西	33 057	4 119.7
天津	51 489	2 461.5	内蒙古	35 211	1 973.6
河北	31 451	7 084.0			

资料来源：《中国统计年鉴 2011》。

## 3. 面板数据

面板数据 (panel data) 中既有时间序列数据又有横截面数据。例如, 表 1-3 是中国、美国、日本 3 国 (总体) 在不同时间的登记失业率数据。

表 1-3 中美日 3 国不同年份登记失业率的比较

年份	中国的登记失业率	美国的登记失业率	日本的登记失业率
2000	3.1	4.0	4.7
2001	3.6	4.7	5.0
2002	4.0	5.8	5.4
2003	4.3	6.0	5.3
2004	4.2	5.5	4.7
2005	4.2	5.1	4.4
2006	4.1	4.6	4.1
2007	4.0	4.6	3.8
2008	4.2	5.8	4.0
2009	4.3	9.3	5.1
2010	4.2	9.6	5.1

资料来源：中经网统计数据库。

## 1.5 预备知识：统计学基础

### 1. 随机现象、随机试验与概率

随机现象是无法事先准确确定其结果的现象。在社会经济领域中，随机现象是普遍存在的，研究随机现象，对认识它们是非常必要的。

观察随机现象或为了观察随机现象而进行的试验称为随机试验。随机试验具有可以重复多次；可能的结果不止一个，但事先可知；每一次试验都会出现上述结果中的某一个，但事先不能预知是哪一个等特点。

随机试验的每一个可能结果称为一个样本点，全体样本点的集合称为样本空间。随机试验的结果称为随机事件。随机事件由一系列样本点组成。

某随机事件  $A$  发生的可能性称为事件  $A$  发生的概率，记为  $p(A)$  ( $0 \leq p(A) \leq 1$ )， $p(A)=0$  表示不可能发生的事件， $p(A)=1$  表示必然发生事件。

### 2. 随机变量

以随机试验的结果为取值的变量称为随机变量。一个随机变量具有下列特性：可以取许多不同的数值，取这些数值的概率为  $p$ 。

样本就是一个随机变量，所谓“样本容量为  $n$  的样本”，就是  $n$  个相互独立且与总体有相同分布的随机变量  $X_1, \dots, X_n$ 。每一次具体抽样所得的数据，就是  $n$  元随机变量的一个观察值，记为  $(X_1, \dots, X_n)$ 。

随机变量可以分为离散随机变量和连续随机变量。一个离散随机变量只能取有限（或可数无穷）多个值。例如，投掷骰子的所有可能点数为  $1 \sim 6$  中的任何一个，我们就可以定义随机变量为点数  $X=1, 2, 3, 4, 5, 6$ ，从而可知它是一个离散随机变量。连续随机变量可以取某一区间的任何值，例如人的身高、体重、学生的分数等都是连续随机变量。

若  $X$  为一随机变量，对任意实数  $x$ ，称  $F(x)=p(X < x)$  为随机变量  $X$  的分布函数。对于连续随机变量如果有：

$$F(x) = \int_{-\infty}^x f(t) dt \quad (1-4)$$

其中， $f(x) \geq 0$ ，则称  $f(x)$  为  $X$  的概率分布密度函数，简称分布密度。

分布密度函数具有以下性质：

$$\textcircled{1} f(x) \geq 0; \textcircled{2} \int_{-\infty}^{+\infty} f(x) dx = 1; \textcircled{3} p(a < X \leq b) = \int_a^b f(x) dx; \textcircled{4} F'(x) = f(x).$$

如果  $X$  的分布密度为  $f(x)$ ，则记为  $X \sim f(x)$ 。

### 3. 随机变量的数字特征

(1) 数学期望。数学期望也称均值，它描述随机变量（总体）的一般水平，从计算方法上来看，它是一个加权平均值。

离散随机变量  $X$  的数学期望记为  $E(X)$  或  $\mu$ , 定义如下:

$$E(X) = \sum xp(x) = \mu \quad (1-5)$$

式中,  $p(x)$  为取  $x$  值的概率。

连续型随机变量数学期望的定义如下:

$$E(X) = \int_{-\infty}^{+\infty} xf(x)dx = \mu \quad (1-6)$$

式中,  $f(x)$  为分布密度。

数学期望具有如下性质:

①如果  $a, b$  为常数, 则  $E(aX+b) = aE(X)+b$ , 特别地  $E(b) = b$ ; ②如果  $X, Y$  为两个随机变量, 则  $E(X+Y) = E(X)+E(Y)$ ; ③如果  $g(x)$  和  $f(x)$  分别为  $X$  的两个函数, 则  $E[g(X)+f(X)] = E[g(X)] + E[f(X)]$ ; ④如果  $X, Y$  是两个独立的随机变量, 则  $E(XY) = E(X)E(Y)$ 。

(2) 方差。如果随机变量  $X$  的数学期望  $E(X)$  存在, 称  $[X-E(X)]$  为随机变量  $X$  的离均差或离差, 显然, 随机变量离均差的数学期望是 0, 即  $E[X-E(X)] = 0$ 。

随机变量离差平方的数学期望称为随机变量的方差, 记作  $\text{Var}(x)$  或  $\sigma^2$ , 即

$$\text{Var}(X) = E[X-E(X)]^2 = E[X-\mu]^2 = \sigma^2 \quad (1-7)$$

方差的算术平方根称为标准差, 即

$$\sqrt{\text{Var}(X)} = \sigma \quad (1-8)$$

方差和标准差刻画了随机变量取值相对于均值的分散程度, 方差或标准差的值越大, 说明随机变量的取值越分散。

方差具有以下性质 ( $c$  是常数):

①  $\text{Var}(c) = 0$ ; ②  $\text{Var}(c+x) = \text{Var}(x)$ ; ③  $\text{Var}(cx) = c^2 \text{Var}(x)$ ; ④  $x, y$  为相互独立的随机变量, 则  $\text{Var}(x+y) = \text{Var}(x) + \text{Var}(y) = \text{Var}(x-y)$ ; ⑤  $\text{Var}(x) = E(x^2) - (E(x))^2$ 。

(3) 协方差。设  $X, Y$  是两个随机变量, 定义  $X, Y$  的协方差为:

$$\text{Cov}(X, Y) = E[X-E(X)][Y-E(Y)] \quad (1-9)$$

如果  $X=Y$ , 则有  $\text{Cov}(X, Y) = E[X-E(X)]^2 = \text{Var}(X) = \sigma^2$ 。

(4) 相关系数。描述  $X$  与  $Y$  线性相关程度可以用相关系数度量,  $X$  与  $Y$  的相关定义为:

$$\rho = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} \quad (1-10)$$

相关系数的取值为  $[-1, 1]$ ,  $\rho > 0$  说明  $X$  与  $Y$  是正相关, 反之为负相关;  $|\rho|$  越接近 1, 说明  $X$  与  $Y$  的相关程度越高; 反之则越低。

#### 4. 重要的理论分布

(1) 正态分布。分布密度为

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (1-11)$$



分布密度的图像如图 1-1 所示。

正态分布取决于两个参数：均值  $\mu$  和方差  $\sigma^2$ ，如果  $X$  服从正态分布，则记为  $X \sim N(\mu, \sigma^2)$ 。

如果正态分布  $\mu=0, \sigma^2=1$ ，则称其为标准正态分布，记为  $Z \sim N(0, 1)$ 。标准正态分布的分布密度为：

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \quad (1-12)$$

标准正态分布的分布函数记为  $\Phi(Z)$ ，即  $\Phi(Z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^Z e^{-\frac{x^2}{2}} dx$ ，它有 3 个重要的性质：

$$\textcircled{1} p(a < Z < b) = \Phi(b) - \Phi(a); \textcircled{2} \Phi(-a) = 1 - \Phi(a); \textcircled{3} p(|Z| \leq a) = 2\Phi(a) - 1.$$

利用这 3 个性质，可以查标准正态分布表得到相应的概率。

可以证明，对于任意一个正态分布都可以通过标准化变换为标准正态分布：

$$Z = \frac{X - \mu}{\sigma} \quad (1-13)$$

这样，我们可以求出任意一个正态分布所对应的概率。

关于正态分布还有一个重要的结论：如果  $X_1, X_2, \dots, X_n$  都是服从  $X_i \sim N(\mu_i, \sigma_i^2)$  的独立随机变量，那么其线性组合  $X = \sum_{i=1}^n k_i X_i$  也服从均值为  $\sum_{i=1}^n k_i \mu_i$ 、方差为  $\sum_{i=1}^n k_i^2 \sigma_i^2$  的正态分布，即

$$X \sim N\left(\sum_{i=1}^n k_i \mu_i, \sum_{i=1}^n k_i^2 \sigma_i^2\right) \quad (1-14)$$

(2)  $\chi^2$  分布。设  $Z_1, Z_2, \dots, Z_k$  是互相独立的标准化的正态分布变量，则  $Z = \sum_{i=1}^k Z_i^2$  服从自由度为  $k$  的  $\chi^2$  分布，记为  $Z \sim \chi^2(k)$ 。

$\chi^2$  分布取决于自由度  $k$ 。 $\chi^2$  分布的分布密度图像是一个右偏分布（见图 1-2），当  $k$  的值越来越大时， $\chi^2$  分布的分布密度图像会越来越趋于对称。一般认为，当自由度超过 100 时， $\chi^2$  分布近似为正态分布。

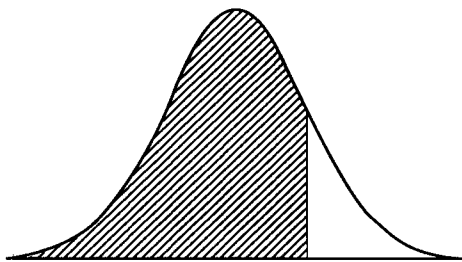


图 1-1 正态分布

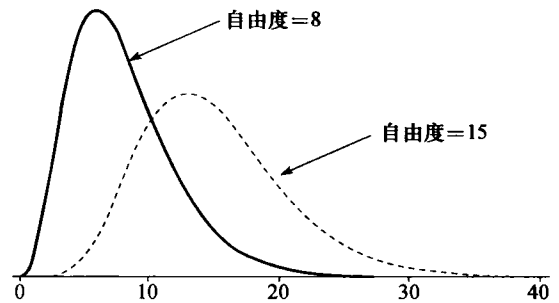


图 1-2  $\chi^2$  分布图