

普通高等学校财经类系列实验教材

金融统计 实验教程

JINRONG TONGJI
SHIYAN JIAOCHENG

主
编
杨
廷
干

 中国财政经济出版社

普通高等学校财经类系列实验教材



金融统计实验教程

杨廷干 主编

中国财政经济出版社

图书在版编目 (CIP) 数据

金融统计实验教程/杨廷干主编. —北京: 中国财政经济出版社, 2011. 12

普通高等学校财经类系列实验教材

ISBN 978 - 7 - 5095 - 3249 - 2

I. ①金… II. ①杨… III. ①金融统计 - 高等学校 - 教材 IV. ①F830.2

中国版本图书馆 CIP 数据核字 (2011) 第 230884 号

责任编辑: 马 真

责任校对: 胡永立

封面设计: 和易润通

版式设计: 兰 波

中国财政经济出版社 出版

URL: <http://www.cfeph.cn>

E-mail: cfeph@cfeph.cn

(版权所有 翻印必究)

社址: 北京市海淀区阜成路甲 28 号 邮政编码: 100142

营销中心电话: 88190406 北京财经书店电话: 64033436 84041336

北京财经印刷厂印刷 各地新华书店经销

787 × 1092 毫米 16 开 14.75 印张 261 000 字

2012 年 1 月第 1 版 2012 年 1 月北京第 1 次印刷

印数: 1—3 000 定价: 27.00 元

ISBN 978 - 7 - 5095 - 3249 - 2/F · 2751

(图书出现印装问题, 本社负责调换)

质量投诉电话: 88190744

普通高等学校财经类系列实验教材

编 写 委 员 会

主 任：贺 璞

副主任：刘玉平 应小陆

成 员（以姓氏笔划为序）：

丁小云	王晓光	车荣强	孙玉良	李琼飞
吴 慧	杨秋菊	施继元	查建华	龚晓莺

前 言

在实际教学过程中，统计方法、统计软件与实际问题的背景知识的教学容易脱节，如何把三者有机联系起来，如何借助统计软件，应用统计方法去解决一些经济金融领域面临的实际问题，如何培养学生的这种综合应用的能力，是我们在努力思考并积极探索和实践的一个重大课题。本书尽管集中了上海金融学院统计系全体同仁的智慧，但由于我们学习实践不够，一定还存在错讹不周之处，尚祈大家指正。编写过程中，参考了大量同类教材或专著，获益匪浅，在此向作者表示感谢，主要的参考文献附于书后。

本书是集体合作的成果，各章编著者为杨廷干（绪论、实验四）、俞雪梨（实验一）、温建宁（实验二）、刘伟（实验三、实验八）、牟嫣（实验五）、王静（实验六、实验七）。

杨廷干

2011 年 10 月

目 录

结 论	(1)
一、统计数据与统计学	(1)
二、统计学的分类及应用	(6)
实验一 统计描述：数值计算	(8)
一、集中趋势的度量	(8)
二、离散程度的度量	(9)
三、描述分布形态的统计量	(12)
四、基本操作	(13)
五、应用举例	(14)
实验二 方差分析	(16)
一、什么是方差分析	(16)
二、方差分析的基本思想	(17)
三、方差分析的应用条件	(18)
四、方差分析的主要内容	(18)
五、SPSS 应用案例	(19)
实验三 非参数检验	(34)
一、单样本非参数检验	(34)
二、两独立样本非参数检验	(37)
三、多独立样本非参数检验	(39)
四、两配对样本非参数检验	(41)

五、多配对样本非参数检验	(43)
六、非参数检验在 SPSS 中的实现	(45)
实验四 回归分析	(52)
一、相关关系	(52)
二、回归方程的建立	(53)
三、回归方程的显著性检验	(60)
四、利用回归方程进行预报和控制	(63)
五、非线性回归	(68)
六、多元线性回归	(72)
七、回归诊断	(80)
八、回归分析在 SPSS 中的实现	(85)
实验五 聚类分析	(95)
一、聚类分析概述	(95)
二、系统聚类分析	(101)
三、动态聚类法	(115)
实验六 因子分析	(123)
一、因子分析的基本理论	(124)
二、因子分析的数学模型	(126)
三、因子载荷的求解	(131)
四、因子旋转	(140)
五、因子得分	(146)
六、因子分析的步骤及注意事项	(151)
七、因子分析在 SPSS 中的实现	(155)
八、案例：基于因子分析的中国金融风险研究	(173)
实验七 主成分分析	(180)
一、主成分分析的基本理论	(180)
二、主成分分析的模型与几何意义	(182)
三、主成分的推导与性质	(185)
四、主成分分析的步骤及注意事项	(189)

五、主成分分析在 SPSS 中的实现	(191)
六、案例：主成分分析在上市商业银行竞争力中的应用	(203)
实验八 时间序列	(209)
一、时间序列的建立和平稳化	(209)
二、指数平滑	(210)
三、自回归	(211)
四、自回归综合移动平均 (ARIMA) 模型	(212)
五、季节分解法	(212)
六、金融时间序列分析在 SPSS 中的实现	(214)
参考文献	(224)

绪 论

随着近代科学的进步和社会经济的飞速发展，人们越来越认识到，在研究和解决各种社会、经济或自然技术问题时，除进行定性分析外，还必须进行定量分析。统计学作为定量分析的重要工具，在人类认识世界和改造世界的过程中产生了越来越重要的影响。在日本，人们把战后经济腾飞的一个重要原因归结为统计方法的应用。用统计方法研究种种数据资料，从而导出某种结论，已成为国家制定政策、学者研究学问、经营者制定谋略的重要依据。正如著名经济学家马寅初先生所言：政治家不能离开统计而施政，学问家不能离开统计而究学，实业家不能离开统计而执业。

一、统计数据与统计学

（一）统计数据及其层次

日常生活中，人们常常听到、看到或用到统计学，例如，开会时主持人要统计出席会议的人数，球赛中教练员要统计各位队员的命中率、犯规次数，报刊上总是定期或不定期地公布诸如物价指数、人口增长率、国民生产总值等统计数据，每天广播里也不乏工农业生产速度、出口贸易额、居民生活费用指数等统计数字的报道。统计数据是指进行各种统计、计算、科学研究或技术设计等所依据的数值，统计数据不同于一般所讲的“数”，数据总要依附于一定的事物或现象，总是一定具体自然技术现象或社会经济现象的数量表现，而“数”则往往具有较为抽象的意义。

美国社会学家、统计学家斯蒂文斯 1968 年按照统计数据的性质和数学运算的功能特点，将统计测度划分为四个层次。

1. 定类层次的数据。定类层次是数据层次最低的，它的取值只有类别属性之分，而无大小、程度之分。根据数据值，只能知道研究对象是相同或是不同。

结合数学运算的特性来看,定类数据只具有等于或不等于($=$ 、 $\neq$)的性质。

定类测度实质是一个分类体系,即按某个品质标志将总体单位划分为若干个类型。例如,对人口按性别划分为男性和女性两类,将工业企业按经济类型划分为国有经济、集体经济、股份制经济、外商投资经济、港澳台投资经济等。每一类型都有特定的文字或数码标识,这种数码只是代号而无量的意义。定类测度要遵循两个原则:(1)互斥原则,即每一单位只能划分到某一类型中,而不能既是这一类,又是那一类;(2)穷举原则,即所有被观察的单位都可归属到适当的类型中,没有一个单位无从归属。

2. 定序层次的数据。定序数据的层次高于定类数据。它的取值除了有类别属性之分外,还有等级、次序的差别。其数学运算特性除具有等于或不等于($=$ 、 $\neq$)之外,还有大于或小于($>$ 、 $<$)之分。常见的定序数据有教育程度(文盲、小学、初中、高中、大学)、社会经济地位(上等、中等、下等)、积极性(很积极、一般、不积极)等。

3. 定距层次的数据。定距数据的层次又高于定序数据。在定距数据测度中,一般要求建立某种物理的量度单位。例如,学生某门课程的考分可以从高分到低分排序,形成90分、80分、70分,直到零分的序列。它们不仅有明确的高低之分,而且可以计算差距,90分比80分高10分,比70分高20分等等。类似地,如某种产品按质论价,一等品100元,二等品80元,三等品50元。一等品比二等品、三等品质优,差价分别为20元和50元。定距数据可以进行加或减的运算,其原因是在等级序列中没有固定的、有确定意义的零位。例如,学生甲得分90分,学生乙的得分是0分,可以说甲比乙多的90分,却不能说甲的成绩是乙的90倍或无穷大。因为0分在这里不是一个绝对的标准,并不意味着乙毫无知识。

4. 定比层次的数据。定比数据是最高层次的数据,它不仅可以进行加减运算,而且可以进行乘除运算。绝大多数的经济变量都可以进行定比测度,如银行存款余额、固定资产投资、产值等。正因如此,对社会经济现象不仅可以用总量指标测定总规模、总水平,而且还可以用相对指标测定数据关系,用平均指数测度一般水平或集中趋势等。相对指标和平均指标都是加减乘除运算的结果。

应该指出,数据层次的划分并不是唯一的。如果数据是最高层次的,它也可以作为低层次来使用。但降低层次的使用,一般会使资料的信息使用不完全。例如,收入按实际数填写时是定距数据,但如果按低薪、一般水平和高薪水平来填,则是定距数据,而如果只问有无收入,则为定类数据。

在实际问题的研究中,研究者往往还根据研究的需要,对数据做出不同的划

分,工业产品的质量检验中往往把数据分为计量型和计数型,数学上则习惯把数据分为离散型和连续型。

四个层次的统计数据可概括如下:

数据层次	运算功能	特 征	举 例
1. 定类数据	计算	分类	产业分类
2. 定序数据	计数 排序	分类 排序	企业等级
3. 定距数据	计数 排序 加减	分类 排序 有基本的测量单位	产品质量差异
4. 定比数据	计数 排序 加减 乘除	分类 排序 有基本的测量单位 有绝对零点	商品销售额

(二) 统计学及其学科性质

统计学及其学科性质的表述,是一个存在争议的理论问题,较为一致的看法是:统计学是研究“数据”采集、整理、描述、分析以及对之进行预测、决策的方法论科学。

统计学首先是一门方法论,它是研究现象(包括自然技术现象和社会经济现象)的数据特征和数量关系,具有明确对象的方法论科学,统计方法论性质是指它作为一门认识方法论科学,为人们提供一套从不确定的现象中探索规律性的理论与方法。

统计学的研究对象——数据所具有的不同特点,明显地制约着统计方法,使统计学色彩缤纷,各具特色。数据中的试验数据主要来自于自然技术现象,这类数据大多在可控条件下通过物理测量取得,这类数据的收集、整理工作并不复杂,研究的重点在于数据分析。另一类是观察数据,它主要来自社会经济现象,由于社会经济现象的复杂性,尤其是不能通过一定条件下的物理或化学试验进行研究,致使观察数据的收集往往十分困难。统计学不仅要研究观察数据的整理、分析技术,而且还要花很大力气研究观察数据的调查收集技术。正因为试验数据与观察数据有不同的特点,所以用试验数据作为研究对象的自然技术统计学和以观察数据作为研究对象的社会经济统计学表现出很不相同的特点。社会经济统计

学利用统计指标、统计分组方法不厌其详地研究数据收集的技术、资料来源、指标口径和计算方法；自然技术统计学则与社会经济统计学相反，它的研究重点往往放在对数据所作的各种分析上。自然技术统计学和社会经济统计学没有不可逾越的鸿沟，两者只是由于研究数据所具有的不同特点，才产生了各有侧重的理论体系和学科特色^①。

把握统计学的学科性质，需要注意以下四个特点：

1. 统计学是数据科学。统计学的英文是 Statistics。这个英文名词有两个含义：作为复数，它是指通过观察或调查所收集的一定时间、空间下反映客观现象的特征的数字资料；作为单数，指的是作为一门科学的“统计学”。统计学的这一名词特性，很直观地说明了统计学和统计数字之间紧密相连的特点。数据是统计的语言，统计学是用数字来说明现象特征的。

统计研究客观事物的数量方面还必须是有对象的具体数量，而不是抽象的量。例如，某省 2009 年国民生产总值为 1724.10 亿元，职工月平均工资为 2497 元，这些都是特定总体在一定时间、地点、条件下的具体的数量。数学研究抽象的数量关系和空间几何形式，是舍去了具体对象质的规定性的抽象的量。因此，尽管统计学有许多数学公式和应用各种数学的方法，但它不同于数学，他必须在质与量的密切结合中研究事物现象的数量方面。

2. 统计学研究群体现象。统计学研究客观事物的数量方面，指的是总体的数量方面，而不是个别事物的数量方面。最初的统计学，如国势学派所研究的“国家显著事项”，就明显地带有总体特征。最初的统计学还同人口问题的研究结合在一起，人口就是一个群体。一定范围内人口的出生率、死亡率、平均寿命等都是人口总体的数量特征，而不是个别人的数量特征。虽然统计必须对人口总体中的每一个人的数量特征和属性进行观察，但这不是统计研究的目的，而是统计研究的必要阶段，为的是从个体特征向总体数量特征过渡。

3. 大量观察的统计规律性。统计研究同类现象总体的数量特征，它的前提则是总体各单位的特征表现存在着差异，而且这种差异并不是由某种固定的原因事先给定的。商品的价格有高有低，时涨时落，这种波动是对价值的背离，但经过长时间或大范围的调查观察，就会发现总价格和总价值、平均价格和价值基本趋于一致，在价格的偶然性波动中存在着价值规律的必然性的支配作用。

婴儿的性别比，从各家各户看，生男生女纯属偶然，但在下面的资料中，清

^① 对不同层次统计数据的研究，形成了特有的统计方法的统计学门类，如对定类数据的研究，形成了“定性资料的统计分析”这一活跃的统计学分支。

楚地表明了男、女的性别比在大量观察的基础上一直在 50% 左右摆动。

某市 1956—1975 年婴儿出生数及所占百分比^①

出生年月	男性		女性	
	人数	百分比	人数	百分比
1956	68688	51.0	66102	49.0
1957	71803	50.0	71768	50.0
1958	59931	48.8	62868	51.2
1959	55675	49.7	56441	50.3
1960	59564	49.1	61831	50.9
1961	50464	49.6	51346	50.4
1962	76986	51.8	71628	48.2
1963	92953	51.5	87598	48.5
1964	60033	51.6	56250	48.4
1965	43089	51.4	40811	48.6
1966	37461	51.8	34892	48.2
1967	37066	51.7	34585	48.3
1968	53194	51.5	50006	48.5
1969	46385	51.7	43418	48.3
1970	41504	51.8	38653	48.2
1971	37507	51.4	35476	48.6
1972	35742	51.5	33635	48.5
1973	33046	51.5	31098	48.5
1974	24219	51.6	22729	48.4
1975	22062	51.8	20547	48.2

在证券投资中，风险与收益相伴而生，如影随形，那么风险与收益之间是否具有某种必然的数量联系呢？通过大量的调查观察，人们可以找出风险与收益之间的数量规律性，目前广泛应用的资本资产定价模型反映的正是这个规律。

可见，表面杂乱无章的随机现象，实际上是有其内在规律性的。恩格斯说

^① 卢淑华：《社会统计学》，北京大学出版社 1997 年版。

过：“在表面上是偶然性在起作用的地方，这种偶然性是受内部的隐藏着的规律支配的，而问题只是在于发现这些规律。”

统计学研究事物现象要从总体上加以考察，就总体中的全部或足够多数单位进行调查观察并加以综合研究。统计研究要运用大量观察法是由于研究对象的大量性和复杂性所决定的，大量复杂的事物现象是在诸多因素的综合作用下形成的，各单位的特性及其数量表现有很大的差别，不能任意抽取个别或少数单位进行观察，而必须观察全部或足够多数的单位，借以从中认识客观现象的规律性。

4. 完整的统计认识过程。一个完整的统计认识过程，既包括计量描述（描述统计学），也包括分析推论（推断统计学）。

统计学研究的一个方面是收集、整理、描述统计数据的方法，这部分内容就形成描述统计学。例如，要对下岗职工再就业问题进行分析，首先要通过一定的行之有效的调查方法，收集反映下岗职工再就业的有关数据，然后再对数据进行汇总、归纳和计算，将原始数据整理成有条理的能够说明被研究现象特征的科学指标，最后再以相应的统计图、统计表将这些结果表现出来。

统计学研究的另外一个方面则是利用样本数据，对总体数量特征作出分析、预测，提供科学的决策依据。统计研究中常常存在这种情况，人们观察的只是部分或有限的单位，而所需要判断的总体范围却是大量的，甚至无限的。这就产生根据局部的样本资料对整个总体数据特征推断的置信度问题。以一定的置信标准要求，根据样本数据来判断总体数据特征的归纳推理方法，称为统计推断法。统计推断是逻辑归纳法在统计推理中的应用。从某种意义上来说，统计所观察的资料都是一种样本资料，因而归纳推断法也就广泛地应用于统计研究的许多领域。

二、统计学的分类及应用

统计学历经 300 余年的发展，各学派之间既相互渗透、不断融合，又相互借鉴、不断分化，表现出很不相同的特点。在当代科学既分化又综合的发展趋势下，边缘科学不断涌现，不同学科的交叉是一个普遍的现象。统计学作为一门方法论科学应用于各种实质性科学，同它们相结合，产生了一系列专门的或专业的统计学，从而使统计学成为一门多科性的学科“家族”。

20 世纪以来，统计学被广泛应用于解决自然科学和社会科学各个领域中的问题，形成了应用于解决各学科问题的统计学。在自然科学方面，诸如在空间科学试验、气象观察与预测、生物试验、药理试验、工程机械试验、环境保护等问题的研究中，都广泛地应用统计理论与方法；在社会科学方面，国民经济计划与

预测、市场调查与商情分析、产品抽样调查与质量监控、财务管理和投资决策、人类心理与行为分析、民意测验、历史与教育研究等，也常常应用统计学作为分析问题和解决问题的工具。

经济管理实践或经济管理科学的研究中，越来越广泛地应用统计理论与统计方法。经济学是研究资源配置和资源利用的科学，效用、生产函数、成本、收益、利润、总投资、总消费、总储蓄，这些变量的测度及其相关问题的研究，都不能只停留在定性的理论分析上，还需要有经验性的定量方法。统计学在工商管理领域的运用，正像统计学在经济学领域的运用一样，无论何时何地都可以找到应用的实例。例如，商业部门需要了解某种商品的需求状况，就可以选择一个有代表性的顾客样本进行销售调查，征求他们对该商品的意见。根据样本资料，判断该商品是否具有充分的需求。如果存在充分需求，商业部门就必须包装设计最佳销售价格和市场范围，所有这些问题都能够根据样本观察资料中所导出的信息得到解答。如此等等，不一而足。20 世纪 70 年代以来，随着电子计算机这一现代化计算工具的普及，产生了一系列旨在加强企业内部管理的定量管理方法，如模型分析、库存决策、需求预测、实验设计、综合评判等。从某种意义上说，这些都是统计与管理融合的结果，它把统计的功能从记数、反映和监督推广到预测、决策和控制。管理的重点在经营，经营的中心在于决策，而决策离不开统计方法的支持，统计决策方法是现代决策学说的重要组成部分。

金融统计学是经济、管理统计的重要组成部分，它以经济、金融理论为基础，运用统计技术解决金融领域各种各样的理论和实践问题，包括金融经济学理论检验、金融决策模拟、风险管理、金融模型的建立预估计、资产定价、衍生产品定价、资产配置策略、波动率估计等等。金融统计学以其独有的研究对象和研究方法，成为统计学中相对独立、颇具特色和最为活跃的研究领域之一。

实验一

统计描述：数值计算

在对分析对象的研究过程中，收集到大量的资料数据，但从这些杂乱无章的资料中，很难对其总体水平与分布状况作出评价判断。因此，必须采用一些适当的方法对这些资料进行处理，使之简约化、分类化、系统化，从中发现它们的分布规律，掌握总体的特征，以便对其水平作出客观的评价。为实现上述分析，往往采用数值计算的方法，计算所得数据的各种统计量，如平均数、标准差、峰度和偏度等。通过这些统计量，能准确反映数据的基本统计特征及各变量统计特征上的差异。

常用的统计量可分为三类：描述集中趋势的统计量、描述离散程度的统计量和描述变量分布情况的统计量。

一、集中趋势的度量

集中趋势是指一组数据向某一中心值靠拢的倾向。计算刻画集中趋势的描述统计量正是要寻找到能够反映数据一般水平的“代表值”或“中心值”。

（一）平均数（Mean）

平均数，又称均值，是一组观察数据中所有数据之和再除以数据的个数。平均数在统计学中具有重要地位，是进行统计分析和统计推断的基础。一般用 $\bar{x}$ 表示平均数，设 $x_1, x_2, \dots, x_n$ 为各次观察的结果，则有： $\bar{x} = \sum_{i=1}^n x_i / n$ ，其中 n 表示总数据个数。

可见平均数描述了变量值的平均水平。用平均数表示一组数据的情况，有直

观、简明的特点，所以在日常生活中经常用到，如平均速度、平均身高、平均产量、平均成绩等等。显而易见，平均数容易受到极端值的影响。

（二）众数（Mode）

众数是指观察数据中出现次数最多的数据。当一组数据有较多的重复数据时，众数往往是人们所关心的一个统计量。众数不受数据极端值的影响，从而在一定程度上提高了平均水平的代表性。例如，制衣厂可以根据消费者所需服装尺码的众数来安排生产。此外，如果众数的值出现的频数或频率越大，那么说明众数的代表性就越高，数据的集中趋势也就越显著。

观察数据中如果有两个数据以相同的最大频数出现，则该观察数据有两个众数，如果所有数据以相同频数出现，则该观察数据没有众数。

（三）中位数（Median）

中位数是将观察数据的各个数值按大小顺序排列，居于中间位置的那个数。中位数是一个反映数据集中趋势的位置代表值，通过中位数，我们可以知道这组数据中，约有一半的数据大于（或小于）中位数。中位数的确定，仅仅取决于它在数列中的位置，它也不受极端值的影响，因此用它表示总体的一般水平比平均数要更稳定。

在实际应用中，应根据平均数、众数和中位数的不同特点及实际问题选择合适的指标。例如，生产鞋的厂商在制订各种型号鞋的生产计划时应运用众数；在评价社会的老龄化程度时，可以用中位数等。这三个表示数据集中趋势的指标在SPSS中的表示及数学含义见表1-1。

表 1-1 集中趋势的度量指标

名 称	SPSS 中的表示	数学含义
平均数	Mean	一组数据中所有数据之和再除以数据的个数
众数	Mode	一组数据中出现频数最多的数据
中位数	Median	将观察数据的各个数值按大小顺序排列，居于中间位置的那个数

二、离散程度的度量

离散程度是指一组数据远离其“中心值”的程度。单纯以平均数等“中心