当代经济学系列丛书
Contemporary Economics Series
主编 陈昕

国家“十三五”重点图书

金融基础

投资组合决策和证券价格

当代经济学
教学参考书系

[美] 尤金·法玛 著
王蕾 译

格致出版社
上海三联书店
上海人民出版社

当代经济学系列丛书
Contemporary Economics Series
主编 陈昕

金融基础
投资组合决策和证券价格

[美] 尤金·法玛 著
王蕾 译

当代经济学
教学参考书系

格致出版
格致出版社
上海三联书店
上海人民出版社

图书在版编目(CIP)数据

金融基础:投资组合决策和证券价格/(美)尤金·法玛著;王蕾译.—上海:格致出版社:上海人民出版社,2017.1

(当代经济学系列丛书/陈昕主编.当代经济学教学参考书系)

ISBN 978-7-5432-2696-8

Ⅰ.①金… Ⅱ.①尤… ②王… Ⅲ.①组合投资-金融决策-研究 ②证券交易-价格-研究 Ⅳ.①F830

中国版本图书馆CIP数据核字(2016)第289301号

责任编辑 钱 敏
装帧设计 敬人设计工作室
吕敬人

金融基础——投资组合决策和证券价格

[美]尤金·法玛 著
王 蕾 译

出 版 格致出版社·上海三联书店·上海人民出版社 (200001 上海福建中路193号 www.ewen.co) 编辑部热线 021-63914988 市场部热线 021-63914081 www.hibooks.cn	**印 刷** 浙江临安曙光印务有限公司
	开 本 787×1092 1/16
	印 张 20
	插 页 3
	字 数 405,000
	版 次 2017年1月第1版
发 行 上海世纪出版股份有限公司发行中心	**印 次** 2017年1月第1次印刷

ISBN 978-7-5432-2696-8/F·983 定价:56.00元

陕西师范大学优秀著作出版基金资助
陕西师范大学211工程建设项目资助

主编的话

上世纪80年代，为了全面地、系统地反映当代经济学的全貌及其进程，总结与挖掘当代经济学已有的和潜在的成果，展示当代经济学新的发展方向，我们决定出版“当代经济学系列丛书”。

“当代经济学系列丛书”是大型的、高层次的、综合性的经济学术理论丛书。它包括三个子系列：(1)当代经济学文库；(2)当代经济学译库；(3)当代经济学教学参考书系。本丛书在学科领域方面，不仅着眼于各传统经济学科的新成果，更注重经济学前沿学科、边缘学科和综合学科的新成就；在选题的采择上，广泛联系海内外学者，努力开掘学术功力深厚、思想新颖独到、作品水平拔尖的著作。“文库”力求达到中国经济学界当前的最高水平；“译库”翻译当代经济学的名人名著；“教学参考书系”主要出版国内外著名高等院校最新的经济学通用教材。

20多年过去了，本丛书先后出版了200多种著作，在很大程度上推动了中国经济学的现代化和国际标准化。这主要体现在两个方面：一是从研究范围、研究内容、研究方法、分析技术等方面完成了中国经济学从传统向现代的转轨；二是培养了整整一代青年经济学人，如今他们大都成长为中国第一线的经济学家，活跃在国内外的学术舞台上。

为了进一步推动中国经济学的发展，我们将继续引进翻译出版国际上经济学的最新研究成果，加强中国经济学家与世界各国经济学家之间的交流；同时，我们更鼓励中国经济学家创建自己的理论体系，在自主的理论框架内消化和吸收世界上最优秀的理论成果，并把它放到中国经济改革发展的实践中进行筛选和检验，进而寻找属于中国的又面向未来世界的经济制度和经济理论，使中国经济学真正立足于世界经济学之林。

我们渴望经济学家支持我们的追求；我们和经济学家一起瞻望中国经济学的未来。

陈昕

2014年1月1日

前言

在经济学的众多领域中，就理论与实践之间的对应关系而言，金融学多少有些独特。本书的目的就是介绍金融理论及其实证检验。我集中分析资本市场中与投资者的投资组合决策以及证券定价相关的金融领域的问题。

我认为，当有迹象表明理论对现实世界具有一定的解释力时，学生学习理论的动机就会增强。而且，我的授课经验表明，如果金融理论具有相关经验证据的话，那么就可以避免对该理论的现实意义或对得出该理论的假设进行无意义的争论。本书采用了这种理论与实证相结合的方法。

本书的前四章是统计学基础。其目的在于：

(1) 对深入研究金融所需的统计工具进行回顾；

(2) 让读者熟悉证券价格行为的描述性证据，这是金融理论及其正规检验的实证基础。

这四章采用的方法是先介绍统计学概念，然后用这些概念来描述证券收益率行为。因此，第 1 章研究概率分布及样本特征，然后，用这些概念对普通股收益率的分布进行验证。第 2 章和第 3 章介绍在研究证券收益率和投资组合收益率之间关系时所需的统计工具。为了激发读者学习这些统计工具的积极性，第 2 章介绍了投资组合理论的一些基础知识。第 4 章利用第 2 章和第 3 章介绍的统计学概念，通过检验单个证券收益率与整个市场收益率之间的相关程度，对纽约证券交易所普通股的“市场敏感性”进行实证研究。

本书的核心部分是第 5 章到第 9 章。这几章介绍了三个相关的主题：

(1) 资本市场有效性理论及其证据；

(2) 投资组合理论；

(3) 预期收益率与风险关系的理论及其证据。

在一个有效的资本市场中，证券价格"完全反映了"可得信息。第5章和第6章讨论资本市场有效性理论及其实证；前者关注的是股票市场，后者关注债券市场。第7章详细扩展了第2章介绍的投资组合模型，并给出了分散化在降低风险方面的经验证据。然后，第8章考虑了当投资者按照第7章的模型进行投资组合决策时证券均衡价格的特征。第9章检验了由第8章中资本市场均衡模型产生的预期收益率与风险的关系。

为读者设计的习题分散在整本书中。我特意使用"分散"这个词。本书的习题并没有整齐地放在每部分末尾，而是出现在任何我认为应该强化的知识点或读者应该停下来进行思考的时候。这些习题是本书不可缺少的一部分；包含在这些习题中的结果常常会用于随后的章节中。有鉴于此，以公平性和方便性为由的读者会认为习题答案应该紧随习题之后。但这样做会让学生产生不去认真对待习题的可能性，只有具有浓厚兴趣的学生才会抵制不去看答案的诱惑自己做习题。本书的习题能让读者密切关注对材料的理解从而避免毫无来由的快意。

阅读本书需要的前期数学知识较少。仅有两章涉及超过初等代数水平的数学知识，而且这些数学知识对理解任意一章的内容都不会产生关键性的影响。除此之外，我还尽量提供一些补充材料，甚至包括关于初等数学的一些讨论；对那些可以忽略数学讨论细节的地方，都会明确说明。尽管如此，本书还是包含了大量的规范符号，读者应该尽快掌握这些符号。

尽管金融学完全被看作是经济学的一个分支，但是，有抱负的读者在没有任何经济学基础的情况下仍然可以理解本书的内容。然而，如果读者熟悉经济学分析范式，那么，也就更容易掌握金融经济学。因此，尽管我们没有具体要求，但前期的经济学基础会有助于理解本书的内容。同样的，本书回顾了用到的统计学概念，尽管并不要求必须掌握对金融学最有用的特定统计学概念，但如果读者具备统计学基础，那么就会更有效地理解本书的内容。

本书是一本金融学导论，对理论和实证进行了同等关注。作为导论性质的著作，很有必要筛选讨论的主题。我选择的都是经验证据充足的主题，能够得出与理论描述相一致的结论。但我并没有讨论所有符合这一标准的主题，因此也许有人会说我是带着个人偏见来选择讨论的内容。然而，如果通过阅读本书可以让读者熟悉金融分析的常用方法并能处理现有或未来的工作，那么，也就实现了本书的目标。

最后，我要感谢 Linda Huegel，键入了好几个版本的手稿；感谢 Agnes Farris，Vicky Longawa 和 Jane Miller 的校对工作；感谢我在芝加哥大学的同事——Nicholas Gonede 和 Harry Roberts，他们对本书提出了宝贵意见。感谢现代金融学的先驱们在金融领域所做的原创性研究，很显然，这些都是本书的基础。

目 录

▶1

股票市场收益率行为

本章在介绍金融理论时，首先回顾若干统计学概念和正态分布的某些特征。这些都是本章进行实证研究所需要的统计工具，而且会在其他章节反复使用。随后，我们对"收益率"的概念进行界定。然后，研究收益率变异性的历史和股票市场收益率的分布特征。本章的这些实证研究是后续章节的重要基础。

1.1 若干统计学概念

1.1.1 随机变量

当一个变量的观测值被看作是由概率分布决定时，该变量被称为是随机的。之所以这么说，是因为在生成一个观测值之前，要获得的该变量的值在某种程度上是未知的(随机的)，而且能够描述将被观测到的特征的唯一办法，就是按照决定该变量的概率分布来进行。

例如，下个月 IBM 的每股收益率目前来说是未知的，而且，仅能通过可能值的概率分布(也许这一分布是正态分布)来描述。这一收益率的分布形态依赖于复杂经济现象的相互影响，这些经济现象本身就是随机变量，而且，从收益率分布中得到的"抽样"是投资者之间交易的结果。不仅如此，该收益率还完全被看作是其观测值由概率分布决定的变量，因此，该收益率是一个随机变量。

为了表示随机变量，我们在变量符号的上方加一个波浪符(～)来识别该变量。当我们涉及该变量的一个特定值时，去掉波浪符。例如，IBM 公司股票下个月要观测的每股收益率记为$\tilde{R}$，而将该收益率的一个特定可能值记为 R。

1.1.2 均值

尽管在此处和其他地方通常会看到"正态分布"这个词，但"正态"这一术语实际上指的是一个概率分布族。用来区分一个正态分布与其他正态分布的两个参数是均值和标准差。首先，我们来回顾一下概率分布(无论是否为正态分布)均值的

一般定义。我们首先考虑离散变量的分布均值，因为它的解释更为简单。

离散变量 $\tilde{x}$ 的均值或“期望”可以表示为：

$$E(\tilde{x})=\sum_{x} xP(x) \tag{1.1}$$

其中，$\sum_{i=0}^{n}$ 是指“x 的所有真实值之和”，$P(x)$是指从随机变量 $\tilde{x}$ 的分布中抽取出 x 特定值的概率。因此，随机变量 $\tilde{x}$ 的期望值 $E(\tilde{x})$，就是针对 x 的所有可能值，对 x 乘以 x 的概率进行求和。也就是说，期望值就是该变量不同可能值的加权平均，其中每个权重就是每个 x 取值的概率。注意，由于这里的总和是指对 x 特定可能值的求和，那么，等式(1.1)右侧的 x 上方就没有波浪符。同时也注意，等式(1.1)中的求和结果，即随机变量 $\tilde{x}$ 的均值或期望，本身并不是随机变量。它是由 $\tilde{x}$ 的分布特征决定的唯一值。总之，$E(\tilde{x})$是 $\tilde{x}$ 分布的一个参数。

连续型随机变量 $\tilde{x}$ 的均值或期望值为：

$$E(\tilde{x})=\int_{x} xp(x)\mathrm{d}x \tag{1.2}$$

其中，$p(x)$是随机变量 $\tilde{x}$ 的概率密度函数(也就是说，$p(x)$给 x 的不同可能值分配的权重为正，这些权重反映了随机抽样中观测到这些不同值的可能性)，而且严格意义上说，积分 $\int_{x}\mathrm{d}x$ 要求计算位于函数 $f(x)=xp(x)$ 下面那部分的面积。尽管这在某种程度上是不严谨的，但如果用与等式(1.1)相同的术语来粗略解释等式(1.2)时，不会造成什么影响，而且可以表达正确的思想。因此，我们把连续型随机变量 $\tilde{x}$ 的均值或期望值看作是该变量不同可能值的加权平均，每个值的权重大小由其概率决定。需要再次注意的是，因为期望值是根据 $\tilde{x}$ 的所有特定可能值计算出来的，因此，等式(1.2)的右侧没有出现波浪符。正如在等式(1.1)中一样，$E(\tilde{x})$ 是 $\tilde{x}$ 分布的一个参数，也就是说，它是由概率密度函数 $p(x)$的形状决定的唯一值。

1.1.3 标准差

如果 $\tilde{x}$ 是离散型随机变量，其方差被定义为：

$$\sigma^2(\tilde{x})=E([\tilde{x}-E(\tilde{x})]^2)=\sum_{x}[x-E(\tilde{x})]^2P(x) \tag{1.3}$$

因此，该方差就是函数 $g(\tilde{x})=[\tilde{x}-E(\tilde{x})]^2$ 的均值或期望值(再次由符号 E 表示)，$g(\tilde{x})$即为随机变量$\tilde{x}$与其均值 $E(\tilde{x})$离差的平方。等式(1.3)表明，离散型随机变量$\tilde{x}$的方差是$[x-E(\tilde{x})]^2$ 的不同可能值的加权平均，其每个值的权重均由其概率$P(x)$决定。

连续型随机变量 $\tilde{x}$ 的方差为：

$$\sigma^2(\tilde{x}) = E([\tilde{x} - E(\tilde{x})]^2) = \int_x [x - E(\tilde{x})]^2 p(x)\mathrm{d}x \tag{1.4}$$

等式(1.4)表明，$\tilde{x}$ 的方差是$[x-E(\tilde{x})]^2$ 的加权平均，其中$[x-E(\tilde{x})]^2$ 的权重是 $p(x)$，即 x 取特定值的可能性或概率密度。

方差是度量 $\tilde{x}$ 的概率分布离散程度的工具。它衡量了 $\tilde{x}$ 分布中的连续随机抽样与该分布均值 $E(\tilde{x})$的平均变异程度。方差是以变异的平方为单位的；也就是说，根据定义，方差测量了 $\tilde{x}$ 与其均值变异程度的平方值。给方差开平方根，就把方差转换成分散度的一个衡量指标，即标准差，其单位与 $\tilde{x}$ 的单位相同。

$$\sigma(\tilde{x}) = \sqrt{\sigma^2(\tilde{x})} \tag{1.5}$$

1.1.4 由均值和标准差描述的正态分布的特征

对于任意正态分布随机变量 $\tilde{x}$，随机抽样分布在均值的一个标准差内，即位于区间

$$E(\tilde{x}) - \sigma(\tilde{x}) \leqslant \tilde{x} \leqslant E(\tilde{x}) + \sigma(\tilde{x})$$

的概率为 0.682 6，而且对所有正态分布变量都是如此。同样的，随机抽样分布在以下区间

$$E(\tilde{x}) - 2\sigma(\tilde{x}) \leqslant \tilde{x} \leqslant E(\tilde{x}) + 2\sigma(\tilde{x})$$

的概率为 0.955 0，且对所有正态分布变量都是如此。正态分布的一个重要特征是，随机抽样分布位于区间

$$E(\tilde{x}) - \varphi\sigma(\tilde{x}) \leqslant \tilde{x} \leqslant E(\tilde{x}) + \phi\sigma(\tilde{x})$$

的概率取决于 ϕ 和$\sigma(\tilde{x})$而非 $E(\tilde{x})$。

同样的，对任意正态分布随机变量 $\tilde{x}$ 而言，转换变量

$$\tilde{r} = \frac{\tilde{x} - E(\tilde{x})}{\sigma(\tilde{x})}$$

是以偏离其均值的标准差为单位来衡量的 $\tilde{x}$，它具有单位正态分布的特征，即它是均值为 0 标准差为 1 的正态分布。因此，如果 $\tilde{r}$ 的分布已知(如图 1.1 和本章末的表 1.8 所示)，那么，对于其他任意正态变量 $\tilde{x}$，我们仅需知道其均值和标准差即可。给定 $\tilde{x}$ 的均值和标准差，我们就可以从单位正态变量 $\tilde{r}$ 的分布中确定变量 $\tilde{x}$ 落在任意特定区间的概率。总之，正态分布是两参数分布；只要知道正态分布的均值和标准差，我们就足以完全描述该分布的特征。

对于后续章节中的投资组合模型而言，正态分布变量的这一性质非常重要。其原因是显而易见的。如果投资组合收益率的概率分布是正态分布，那么，投资组合的选择问题就被简化了，因为可以根据收益率分布的均值和标准差来对其

他投资组合进行排序。根据均值和标准差这两个参数就足以进行理性的投资组合选择。

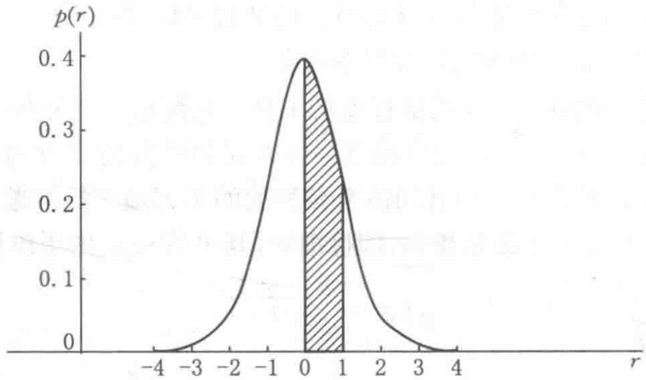

图 1.1　单位正态分布

1.1.5　样本均值和标准差

在分析现实数据中，随机变量的均值和标准差几乎是未知的，因而必须根据样本对其进行估计。例如，假定我们假设 IBM 公司每股的月度收益率是从某一概率分布(可能是正态分布)中随机抽取的。该分布的总体均值和标准差未知。如果想要得到这两个参数的信息，必须进行样本估计。接下来我们介绍如何进行这种样本估计。

在本书中，样本均值的计算公式为：

$$\bar{x}=\sum_{i=1}^{T} x_i/T \tag{1.6}$$

其中，T 为样本容量(即样本观测个数)，x_i 为样本中第 i 个观测值或第 i 个抽样，$\sum_{i=1}^{T}$ 被读为“从 $i=1$ 到 $i=T$ 的总和”。因此，样本均值就是样本中所有观测值的简单平均。

把等式(1.6)与等式(1.1)和等式(1.2)进行比较是有指导意义的，后两个等式分别是离散型变量和连续型变量的总体均值。例如，在等式(1.1)中，x 的每一个不同可能值都是由其概率赋权的，一般而言，不同的 x 有不同的概率值。然而，在计算样本均值时，每一个样本观测值都被赋予了相同的权重 $1/T$。在对观测值进行同等赋权时，并不是假定每一个不同的 x 值是等可能出现的，而是假定 x 取不同值的样本相对频率接近于总体概率。那么，对每个样本观测值赋予 $1/T$ 的权重，就会产生按照相对频率赋权给不同观测值的效应。

样本方差可以通过类似于等式(1.5)的方法得到，即：

$$s^2(x)=\sum_{i=1}^{T}(x_i-\bar{x})^2/(T-1) \tag{1.7}$$

由于分母不是 T 而是 $T-1$，因此，这个等式不是完全意义上的对 $(x_i-\bar{x})^2$ 的平均。第 2 章的问题 2.5 中第 7 个小问题讨论了这样处理的原因。样本标准差就是对样本方差开平方根，即：

$$s(x)=\sqrt{s^2(x)} \tag{1.8}$$

1.1.6 正态性检验：学生化极差

1. 定义

在现实数据分析中，不仅真实均值和标准差未知，而且生成样本的分布类型也未知。例如，如果样本为 IBM 每股的月度收益率，我们也许会乐于假设该收益率是某一概率分布的抽样，但概率分布的类型未知。用来判断生成样本的分布是否为正态分布的一个有用的统计量①是学生化极差。其公式为：

$$SR=\frac{\max(x_i)-\min(x_i)}{s(x)} \tag{1.9}$$

也就是说，学生化极差就是样本观测值的范围，即用最大观测值减去最小观测值，它以样本标准差来衡量。

由于学生化极差对样本极端观测值的依赖性很大，因此，它对偏离正态性的情况比较敏感，在这种情况下，离均值较远的观测值的概率大于(或小于)正态分布变量的概率。结果发现这与普通股收益率的分布相关，与正态分布相比，这些收益率的分布呈现"厚尾"现象；也就是说，这些分布中出现较大的正收益率或负收益率的概率要大于正态分布的情况。

2. 样本统计量的概率分布

给定概率分布的数据中，样本是互不相同的，而且一般而言，一个样本并不能准确无误地再现该分布的特征。由于样本之间的变异性，任何样本统计量(如样本均值，样本标准差或学生化极差)本身就是概率分布的一个抽样。因此，该样本统计量是随机变量。在统计学文献中，普遍会把样本统计量的分布看作是在特定分布下，从一个无限大样本中抽取给定规模的样本量，据此计算该统计量的值时生成的分布。正因如此，统计量的概率分布被称为该统计量的抽样分布。然而，选择夸大其原始样本的话，样本统计量的分布与任何其他概率分布没有差别。

当把统计量看作是随机变量时，我们在该统计量上标记了波浪符，但是当该统计量有特定观测值时，我们就去掉波浪符。因此，在抽样前，未知的样本均值、方差和学生化极差的表达式如下：

$$\tilde{x}=\sum_{i=1}^{T}\tilde{x}_i/T \tag{1.10}$$

① 统计量是对由样本计算出来的任意数值的总称。

$$s^2(\tilde{x}) = \sum_{i=1}^{T} (\tilde{x}_i - \bar{\tilde{x}}_i)^2 / (T-1) \tag{1.11}$$

$$\widetilde{SR} = \frac{\max(\tilde{x}_i) - \min(\tilde{x}_i)}{s(\tilde{x})} \tag{1.12}$$

这种记号方法意味着，在抽样前，因为 T 个样本观测值 $\tilde{x}_i(i=1, \cdots, T)$ 是随机变量，因此，样本均值、标准差以及学生化极差都是随机变量。当我们从样本中给这些统计量赋予特定值时，采用式(1.6)、(1.7)和(1.9)的记号方法；也就是说，式(1.10)、(1.11)和(1.12)中的波浪符在此不再出现。这再一次说明了我们区分随机变量和具有特定值变量的方法。

3. 由学生化极差进行的正态性推断

在随后的一章中我们会看到，当从一个正态分布中随机抽取容量为 T 的样本时，该样本均值的分布就已确定。学生化极差的抽样分布更难界定。幸运的是，该分布的分位数是已知的，而且在应用中会使用这些分位数表。

当我们根据正态分布的随机抽样计算学生化极差时，可以利用本章末的表 1.9 给出的学生化极差抽样分布的上尾分位数和下尾分位数。对表 1.9 中分位数的解释如下：如果 $SR(p, T)$是容量为 T 的样本中 SR 分布的第 p 分位数，那么，从正态分布中抽取的容量为 T 的样本中，观测到的一个 SR 值等于或小于 $SR(p, T)$ 的概率为 p。或者说，$1-p$ 是从正态分布中抽取的容量为 T 的样本，其学生化极差大于 $SR(p, T)$的概率。直观而言，如果从正态分布中抽取了容量为 T 的多个样本，并计算了每个样本的学生化极差，那么，样本学生化极差的 p 值将等于或小于 $SR(p, T)$，且样本学生化极差的 $1-p$ 值将大于 $SR(p, T)$。

例如，表 1.9 表明，从正态分布中抽取的 100 个样本中，学生化极差等于或小于 6.36 的概率为 0.99。从直观上看，当样本学生化极差是从正态分布中重复抽取 100 个样本计算而来时，99%的样本学生化极差将等于或小于 6.369，1%的样本学生化极差将大于 6.36。

问题 1.1

1. 表 1.9 表明，样本容量 T 越大，学生化极差分布的任意给定分位数，即对于任意指定的 p 的 $SR(p, T)$值就越大。请解释这一现象。

答案

1. 正态随机变量从任一方向上偏离均值的概率都较小。大样本比小样本更容易出现这种极端观测值。由于学生化极差直接依赖于样本中观测值的范围，即最大观测值与最小观测值之差，因此，当样本来自正态分布时，对大样本而言，SR 的分布会移向更大值。

现在假定有一个随机样本数据，如 IBM 每股收益率的月度数据，我们希望判断该样本来自于正态分布的可能性有多大。假定我们计算出了样本的学生化极差，且表 1.9 表明，它对应的是 0.1—0.9 之间的某个分位数。如果样本来自正态分

布,那么这种学生化极差就极可能出现。在从正态分布的重复抽样中,对于80%的样本而言,学生化极差可能位于表1.9中 *SR* 相关分布的0.1—0.9分位之间。另一方面,如果计算出来的IBM公司的 *SR* 值所对应的分位数处在表1.9中 *SR* 分布的尾部,那么,如果该样本来自正态分布,该样本的学生化极差就不可能处于上述区间。在正态分布的重复抽样中,仅有一小部分样本会出现 *SR* 的极端值。如果IBM的样本 *SR* 值太大,我们也许会拒绝正态性假设,反而会断定,该样本来自于这样一个分布,其观测值远离均值的概率要大于正态分布的情况。另一方面,较小的 *SR* 值也许会让我们认为,生成样本的分布,其远离均值的观测值的概率大于正态分布的情况。另一方面,较小的 *SR* 值也许会让我们认为,相对于正态分布的情况而言,该样本来自于一个"薄尾"分布。

拒绝样本来自正态分布这一假设往往会导致某种偏误。如果分布是正态的,那么不可能出现过大或过小的 *SR* 值,但并不是说一定不会出现这样的数值。另一方面,接受这一假设也会产生某种偏误。非正态分布也会碰巧生成看起来像是来自正态分布的样本。

对实证研究而言,在某种程度的不确定性情况下进行统计推断是正常的。一个研究假设永远都不可能被确定性的证实或证伪。相反,细心的研究者总是会采用在某个置信水平上(通常用概率来表示)接受或拒绝某个假设这种说法。例如,如果包含100个观测值的样本学生化极差为6.4,那么研究者会做如下表述:

> 来自正态分布的容量为100的样本中,学生化极差等于或大于6.4的概率小于1%。据此,拒绝该数据来自正态分布这一假设。

如果研究人员详细描述了获取样本的条件、明确提出已使用统计方法的潜在假设,并详细报告了检验结果,那么,如果拒绝了一个为真的假设,或接受了错误的假设,读者可以根据自己的损失评价方法对该研究结果和结论进行重估。

1.1.7 统计模型和现实情况

既然我们打算对某个股票市场数据进行分析,那么,很有必要在结束本节内容之前介绍另一个方法要点。当一个假设或模型被看作是对数据的描述时,该模型并不意味着是对现实情况的完全再现。相反,它被看作是比其他模型能够更好地解释现实数据,能够便捷地对现实情况进行粗略估计。实际上,"模型"的说法要表达的是对现实的粗略估计这一概念。

例如,我们假定正态分布是纽交所普通股月度收益率的模型。通过衡量该模型在多大程度上代表了收益率样本,以及该模型是否比其他模型更好地描述了这些样本来正确判断该模型的功效。这是我们要解决的问题之一。然而,我们首先必须定义"收益率"这一概念。