



高校教材

高等学校经济与管理类系列教材



统计学学习指导

主 编 ◇ 李继根



华东师范大学出版社



高等学校经济与管理类系列教材



统计学学习指导

主 编 ◇ 李继根

 华东师范大学出版社

图书在版编目(CIP)数据

统计学学习指导/李继根主编. —上海:华东师范大学出版社, 2015. 10

ISBN 978 - 7 - 5675 - 4260 - 0

I. ①统… II. ①李… III. ①统计学—高等学校—教学参考资料 IV. ①C8

中国版本图书馆 CIP 数据核字(2015)第 258150 号

统计学学习指导

主 编 李继根

项目编辑 李 帆

特约审读 伍 平

封面设计 张 季

出版发行 华东师范大学出版社

社 址 上海市中山北路 3663 号 邮编 200062

网 址 www.ecnupress.com.cn

电 话 021 - 60821666 行政传真 021 - 62572105

客服电话 021 - 62865537 门市(邮购)电话 021 - 62869887

地 址 上海市中山北路 3663 号华东师范大学校内先锋路口

网 店 <http://hdsdcbs.tmall.com>

印 刷 者 常熟市文化印刷有限公司

开 本 787×1092 16 开

印 张 15.5

字 数 333 千字

版 次 2016 年 3 月第 1 版

印 次 2016 年 3 月第 1 次

书 号 ISBN 978 - 7 - 5675 - 4260 - 0/O · 265

定 价 31.00 元

出版人 王 焰

(如发现本版图书有印订质量问题,请寄回本社客服中心调换或电话 021 - 62865537 联系)

前言

本书完全遵循了李继根所主编的《统计学》(华东师范大学出版社,2015年10月)的编排顺序,是该教材的学习指导书,也可作为普通本科、高职高专以及继续教育学院经管和社科类专业学生学习“统计学”课程的参考书。

正如统计量的值(统计指标值)是从大量原始数据中概括出的重要数据特征一样,读者也希望能从海量的教材文本中“海选”出统计学的要点。因此我们首先在第一个模块“本章知识概要”中,概括总结了本章的主要知识点,有的还做了适度拓展,以便于读者能更好地掌握统计学的主要内容,也便于没有配套教材的读者能独立地使用本书。

在第二个模块“简答与思考题精析”中,我们编选了一些简述和思辨性的问题,它们大都涉及统计思想和统计方法的原理,是统计学的所谓“形而上”的部分,也是今后进行统计分析的“定海神针”。统计学是定性分析与定量分析的完美结合,如果缺乏对统计思想和方法的理解和掌握,那么在后续的统计分析中,就极易陷入数据的汪洋大海,从而出现形式上虽合乎规范的各种统计计算,而对统计分析结果却是主观臆断。

在第三个模块“典型例题详解”中,我们结合各章的重难点,选择了一些典型性的问题加以详细解析。鉴于统计学公式手工计算的繁琐,我们还特意选编了一些计算题,并给出了手工计算和软件计算(操作型的 Excel 和编程型的 Matlab)过程。当然,这只是与统计实验的简单糅合。要想全面地掌握统计的软件计算乃至编程,需要学习单独开设的相关课程和配套教材。

在第四个模块“习题选粹”中,我们选编了一些选择题和计算与分析题,同时也给出了详细的解析。

要特别指出的是,“统计学”是一门实践性很强的课程,仅仅通过所谓的“刷题”,是学不好统计学的。正如只有亲自下水才能学会游泳一样,只有通过一些统计实践活动,才能真正地掌握统计学的精髓。

另外希望读者在学习时注意三套符号系统之间的灵活转换。比如表示数据集中程度的均值,在概率论中称为数学期望(期望),表示为 EX ,在统计学理论中

2 统计学学习指导

称为总体均值,表示为 μ ,在统计学实践中则称为样本均值(样本平均数),表示则为 \bar{X} 。如果再算上统计软件中的函数表示(AVERAGE 或 mean),那就是四套符号系统。正如统计学中各种学派居然能和谐相处一样,希望读者脑中的各种符号系统也能和平共处。

由于作者水平有限,书中不当乃至谬误之处,敬请各位方家高手来函批评指正。

目录

第 1 章 绪论	1
1.1 本章知识概要	3
1.2 典型例题详解	5
1.3 习题选粹	7
第 2 章 数据的收集与整理	9
2.1 本章知识概要	11
2.2 简答与思考题精析	23
2.3 典型例题详解	27
2.4 习题选粹	34
2.4.1 选择题	34
2.4.2 计算与分析题	43
第 3 章 概率论基础	57
3.1 本章知识概要	59
3.2 简答与思考题精析	69
3.3 典型例题详解	72
3.4 习题选粹	77
3.4.1 选择题	77
3.4.2 计算与分析题	82
第 4 章 统计量的抽样分布	87
4.1 本章知识概要	89
4.2 简答与思考题精析	95
4.3 典型例题详解	97

- 4.4 习题选粹 99
 - 4.4.1 选择题 99
 - 4.4.2 计算与分析题 102

第5章 参数估计和假设检验 105

- 5.1 本章知识概要 107
- 5.2 简答与思考题精析 120
- 5.3 典型例题详解 122
- 5.4 习题选粹 133
 - 5.4.1 选择题 133
 - 5.4.2 计算与分析题 140

第6章 方差分析 149

- 6.1 本章知识概要 151
- 6.2 简答与思考题精析 157
- 6.3 典型例题详解 158
- 6.4 习题选粹 163
 - 6.4.1 选择题 163
 - 6.4.2 计算与分析题 165

第7章 一元回归分析 171

- 7.1 本章知识概要 173
- 7.2 简答与思考题精析 178
- 7.3 典型例题详解 180
- 7.4 习题选粹 188
 - 7.4.1 选择题 188
 - 7.4.2 计算与分析题 190

第8章 指数分析 197

- 8.1 本章知识概要 199
- 8.2 简答与思考题精析 203
- 8.3 典型例题详解 205
- 8.4 习题选粹 210

第9章 时间序列分析	213
9.1 本章知识概要	215
9.2 简答与思考题精析	222
9.3 典型例题详解	224
9.4 习题选粹	229
9.4.1 选择题	229
9.4.2 计算与分析题	234
参考文献	240



第 1 章 绪 论

1.1 本章知识概要

一、统计学思想简史

1. 统计学的创立

(1) 国势学派与政治算术学派

统计活动古已有之。古代的统计可以说主要是人口调查,本质上是为了征兵和收税。欧洲早期最著名的统计成果是 1086 年问世的《土地赋税调查书》,又称《末日审判书》。

1660 年康令开创国势学派,主要研究“国或多数国家的显著事项”,以文字叙述为主,主要用对比分析的方法(后来又加入了图表)研究国家的组织、领土、人口、资源财富和国情国力。

1662 年,格朗特开创了以“社会数字”引导政策研究的先河。他给出的伦敦生命表为人口统计学研究奠定了基础。受其影响,威廉·配第 1676 年写成《政治算术》一书,明确宣称要“采用数字、重量、尺度等语言表达自己的想法”,从而旗帜鲜明地开创了政治算术学派,被马克思誉为“政治经济学之父”。

政治算术学派与国势学派之间后来爆发了长达百余年的争论。通过争论,人们明确意识到统计研究方法的数量特征,以及引入强有力的数学“外援”即概率论的必要性。

(2) 古典概率论

概率论处理的是随机现象,人们通常认为概率论肇始自赌金分配问题。事实上,推动概率论形成的本质性因素是社会经济实践的两个需要,即人们对公平获利的思考和对于公正的期望。

雅各布·伯努利在 1713 年给出了著名的“伯努利大数定理”,并指出:“它有如此绝妙的应用,以至于它在这个学科的每一分支都有很高的价值和地位。”与此同时,棣莫弗在 1738 年给出了“中心极限定理”的最初阐述,即二项分布收敛于正态分布,并导出了正态曲线(钟形曲线)的表达式。

总之,国势学派、政治算术学派和古典概率论,是导致统计学创立的三股主要源泉。

2. 统计学的发展

(1) 分析概率论

孔多塞主张用数理方法研究社会科学,并希望创立社会数学,从而将社会科学理性化。受其影响,拉普拉斯 1812 年贡献出了集大成之作《分析概率论》,标志着概率论从组合概率向分析概率的彻底转变。书中证明了中心极限定理,同时用大部分篇幅讨论了概率论的各种应用。

(2) 数理统计学派和社会统计学派

凯特勒将大数法则、钟形曲线、中心极限定理等成果应用到其他领域,尤其是社会科学领域,从而将概率论与统计学真正结合起来。他开创了统计学的“凯特勒时代”,并被誉为现代统计学之父。

恩格尔继承了政府统计与社会调查相结合的优良传统,形成了社会统计学派。他提出了“恩格尔法则”,即家庭收入越多,则饮食费支出在家庭收入(或总支出)中所占百分比越小。

(3) 生物统计学派

高尔顿对凯特勒的发现十分着迷,同时受表哥达尔文的自然选择学说和英国文化的影响,试图证明天才和杰出是遗传的特征。通过 1875 年的豌豆实验,他提出了“向均值回归”的原理,被认为是“在我们的科学观念上创造了一场革命”,从而开创了生物统计学派。后来他又提出了“相关性”的概念,用来衡量两个不同序列的相近程度。另外,孟德尔或许也可以被看成是生物统计学派的创始人之一,他于 1866 年发表了豌豆实验的实验报告,将统计方法应用到遗传学的研究之中,但令人遗憾的是他的工作和思想在当时不为人所理解,因为读过他的报告的读者,要么懂数学但不懂植物学,要么懂植物学但不懂数学。

卡尔·皮尔逊将老师高尔顿的回归和相关推广为一般统计方法,进而提出了复相关和偏相关等概念,并贡献出了矩估计法。他的工作将统计学带上了新的高峰,特别是他主编的《生物统计》杂志,在数理统计历史上扮演了重要角色。

总之,以拉普拉斯的分析概率论为基础,凯特勒、社会统计学派和生物统计学派(尤其是卡尔·皮尔逊)建立了古典统计学(主要是描述统计学)的基本框架,极大地促进了统计学的发展。

3. 统计学的成熟

(1) 推断统计学

1908 年哥塞特提出“ t 分布”,开创了小样本理论。十余年后,费希尔提出了 t 检验法,并编制了 t 分布表,从而构建了小样本理论的数学基础。与此同时,他勇敢地向卡尔·皮尔逊等创立的“描述统计学”发起挑战,提出了 F 分布及 F 检验,并开辟了方差分析、实验设计等统计新领域,从而真正建立起了“推断统计学”,并使得统计学进入“费希尔时代”。

爱根·皮尔逊受父亲与费希尔的论战的影响,开始探索如何在两人的思想之间架设桥梁。经过几年思索,他与奈曼发展了假设检验的严格数学理论,从而在统计学中正式启动了一场“准哥白尼革命”。1937 年,奈曼正式建立了区间估计理论。至此,推断统计学的基本框架得以建立。

(2) 概率论的公理化

拉普拉斯的概率理论建基于等可能性这一概念之上,而确定等可能性则往往需要直观判断,因此他的概率理论一直受到猛烈批评。1888 年提出的“贝特朗悖论”中,对等可能性的不同理解居然产生了三种不同的正确解答。在 19 世纪末 20 世纪初的公理化浪潮中,米泽斯等开始尝试概率论的严格化工作。1933 年,柯尔莫哥洛夫最终赋予概率论演绎数学的特征。

(3) 战后的统计学

到了 1945 年,瓦尔德提出统计决策理论,其间经常用到的贝叶斯方法和逆概率思想,源自贝叶斯 1760 年的一篇遗文。

二战以后,随着计算机和统计软件的日新月异,统计学的理论、方法和应用都得到了全面发展。特别是近二十年来,随着海量数据、互联网与云计算等新科技的不断涌现,统计学迎来了大数据时代,这使得统计学与数学、哲学一样,正日渐成为所有学科的基础。如果说“21 世纪是信息的时代”,那么统计学就是推动这个信息时代的“最有发展前途的学科之一”。

二、统计学及其分类

1. 统计学的定义

统计学是关于数据的科学,涉及数据的收集、整理、分析、推断、预测和决策等环节。统计学不仅为从数量方面把握世界提供具体的方法和思维方式,而且还是处理统计数据的艺术。

2. 统计学的分类

依据数据分析所使用的方法不同,可将统计学划分为描述统计学和推断统计学。描述统计学研究的是数据的收集、整理、汇总、图表描述、数值概括与分析等方法;推断统计学则研究如何利用样本数据来推断总体特征,包括假设检验、参数估计、区间估计、预测、方差分析与回归分析等方法。

描述统计学是推断统计学的基础;推断统计学则是描述统计学的进一步发展,是现代统计学的核心和统计研究的关键环节。两者反映了统计学发展的前后两个阶段,并最终共同构成了现代统计学的整体体系。

三、统计学的应用

1. 统计学的威力

如果说 20 世纪以前,统计学的研究领域主要是人口统计、生物统计、社会统计和经济统计,那么统计学如今已被广泛用于自然科学、社会科学、工程技术、管理、经济、艺术和文学等各个领域。从“美国总统选举的盖洛普民意调查”、“六西格玛理论”、“足球水平排名”、“《红楼梦》与计算风格学”、“啤酒肚’与喝啤酒”、“尿布与啤酒”、“德军坦克的数量”和“流行病学的诞生”等统计学的有趣应用,就能初步地管窥到统计学的威力所在。

2. 统计思维

即便是普通大众,如果不具备基本的数量素养,在如今的信息社会,也已经被视为“数盲”(innumeracy)。这里“数盲”指的是数学无知者(mathematical illiteracy),尤其是那些不能自如地驾驭数据和概率的人。统计思维如今已经渗入到我们生活的方方面面,像读与写一样成为了一个有效率公民的必备能力。一言以蔽之,统计学如今几乎无所不包,无处不在,已成为了通用的方法论科学。

3. 统计陷阱

当然,对统计学的作用我们也要保持清醒的头脑。统计学能够帮助我们处理数据,但数据只不过是知识的原材料,对于数据及其统计分析的结果,我们一定要注意进行辩证分析,避免落入“统计陷阱”之中。

1.2 典型例题详解

例 1 统计学的发展过程中有哪些主要学派? 其学术观点主要是什么?

答: 统计学是一门很古老的科学,其学理研究可追溯至古希腊的亚里士多德时代。一般

认为从 17 世纪中叶到 18 世纪末出现的政治算术学派和国势学派,标示着统计学的创立。这个阶段也被称为古典统计学时期。

(1) 政治算术学派

政治算术学派的创始人是英国的威廉·配第和约翰·格朗特。政治算术学派被称为有实无名的统计学。该派宣称要“采用数字、重量、尺度等语言表达自己的想法”,开始将统计方法与数学计算和推理方法相结合,在分析社会经济问题的方式上更加注重运用定量分析方法。

(2) 国势学派

国势学派与政治算术学派产生于同一时期,创始人是德国的康令和阿肯瓦尔。国势学派强调用文字记述的形式,把国家的显著事项系统地整理并罗列出来,为政治家提供治国必需的国情知识。国势学派缺乏数量分析的方法和结论,与现代统计学相去甚远,被称为有名无实的统计学。

从 18 世纪末到 19 世纪末属于统计学的发展期,这一时期产生了数理统计学派、社会统计学派和生物统计学派。这个阶段也被称为近代统计学时期。

(3) 数理统计学派

数理统计学派产生于 19 世纪中叶,创始人是比利时的凯特勒,但拉普拉斯等数学家也可被视为该学派的先驱性人物。数理统计学以概率论作为理论基础,抽象掉统计学的社会经济现象内涵,使统计学变成了一门可以用于研究任何科学的方法论学科。

(4) 社会统计学派

社会统计学派产生于 19 世纪中叶,由德国的克尼斯、恩格尔以及梅尔等人创立。社会统计学派强调统计学是一门研究社会经济现象变动规律的实质性科学,也被称为社会经济统计学,包括政治统计、经济统计、人口统计、犯罪统计等多方面的内容。

(5) 生物统计学派

生物统计学派产生于 19 世纪中叶,创始人是英国的高尔顿,集大成者是卡尔·皮尔逊。该派将各种“描述”统计方法引入到生物科学问题的研究中,并将这种“生物统计学”看作一种应用统计学。生物统计学派不仅在生物科学中应用了统计方法,更为统计学贡献出“相关”、“回归”、“频数分布表与频数分布图”、“卡方检验”等重要的统计学概念和方法。卡尔·皮尔逊更是将统计学提升到通用方法论的高度。

从 20 世纪初叶到 20 世纪中叶,随着小样本理论的提出,统计学迈入成熟期,从描述统计学变革为推断统计学,开始了统计学的新纪元。这个阶段也被称为现代统计学时期。因为统计推断思想的不同,这个时期又产生了著名的三大学派:频率学派、贝叶斯学派和信念(Fiducial)学派。

例 2 什么是统计学?

答:统计学是关于数据的科学,涉及数据的收集、整理、分析、推断、预测和决策等环节。数据收集就是取得统计数据;数据整理就是用图表和数值指标等形式来展示数据;数据分析是选择适当的统计方法来研究和描述数据;数据推断研究的是如何利用样本数据来推断出总体特征;数据预测则是指通过对现有数据的整理和分析,对未来的事情进行预先的推测或测定;数据决策则是通过对现有数据的整理、分析和推断,从拟定的多个可行方案中确定出

问题的最终解决方案。统计学不仅为从数量方面把握世界提供具体的方法和思维方式,而且还是处理统计数据的艺术。

值得注意的是,结合统计学的思想发展简史,不难看出统计学的学科内容和研究方法是从无到有、“与时俱进”的,不同时期对统计学的认识也是不相同的。

例3 如何理解描述统计学与推断统计学之间的关系?

答:依据数据分析所使用的方法不同,可将统计学划分为描述统计学和推断统计学。

描述统计学研究的是数据的收集、整理、汇总、图表描述、数值概括与分析等方法,即用直观的图形、汇总的表格和概括性的数字来表示数据的分布和形状等特征,是统计学的基础和统计研究工作的第一步。

推断统计学则研究如何利用样本数据来推断总体特征,包括假设检验、参数估计、区间估计、预测、方差分析与回归分析等方法,是现代统计学的核心和统计研究的关键环节。

描述统计学是推断统计学的基础,因为描述统计中收集、整理和分析的多为样本数据;推断统计学则是描述统计学的进一步发展,因为如果没有描述统计收集的可靠数据及有效的样本信息,即使“巧妇也难为无米之炊”,这里“巧妇”指统计学家,“米”指数据和样本信息,“炊”指的是总体的数量规律性。两者反映了统计学发展的前后两个阶段,并最终共同构成了现代统计学的整体体系。

1.3 习题选粹



1. 【 】被称为有名无实的统计学。

- A. 政治算术学派 B. 国势学派 C. 数理统计学派 D. 社会统计学派

【答案】B

【解析】政治算术学派无统计学之名却有统计学之实,国势学派则有名无实。故选 B。

2. 被马克思誉为“政治经济学之父”的是【 】。

- A. 康令 B. 威廉·配第 C. 凯特勒 D. 约翰·格朗特

【答案】B

【解析】马克思的原话是:“威廉·配第——政治经济学之父,在某种程度上也是统计学的创始人。”故选 B。

3. 下列学者中属于社会统计学派的有【 】。

- A. 克尼斯 B. 康令 C. 恩格尔 D. 梅尔

【答案】A, C, D

【解析】社会统计学派的代表人物是德国的克尼斯、恩格尔及梅尔等人。故选 A, C, D。

4. 推断统计的主要功能是【 】。

- A. 应用总体信息描述样本 B. 描述样本中包含的信息
C. 描述总体中包含的信息 D. 应用样本信息描述总体

【答案】D

【解析】推断统计的最大威力就在于能够根据一定量的样本信息和概率论理论,对总体的相应信息进行推断。故选 D。

5. 下列不属于描述统计问题的是【 】。

- A. 根据样本信息对总体进行推断
- B. 分析感兴趣的总体特征
- C. 使用图、表或其他数据汇总工具分析数据
- D. 了解数据分布的特征

【答案】A

【解析】“根据样本信息对总体进行推断”是推断统计的最重要的特征,是推断统计与描述统计最重要的区别。故选 A。

6. 某报告指出:“通过对 150 部新车的价格调查,发现进口车的价格明显高于国产车。”这个结论属于【 】。

- A. 对样本的描述
- B. 对样本的推断
- C. 对总体的描述
- D. 对总体的推断

【答案】D

【解析】这个结论中明显地隐含了普遍性的含义,即通过“150 部新车的价格”这个样本,推断出了“所有的新车中,进口车的价格明显高于国产车”。这里的总体是“所有新车的价格”。故选 D。

7. 下列叙述中,错误的是【 】。

- A. 统计推断区别于算命的一个主要特点是前者可以进行可靠性度量
- B. 根据样本推断总体,无法确定可靠性
- C. 用样本推断总体,总是存在一定程度的不确定性
- D. 可靠性是关于统计推断不确定性的度量

【答案】B

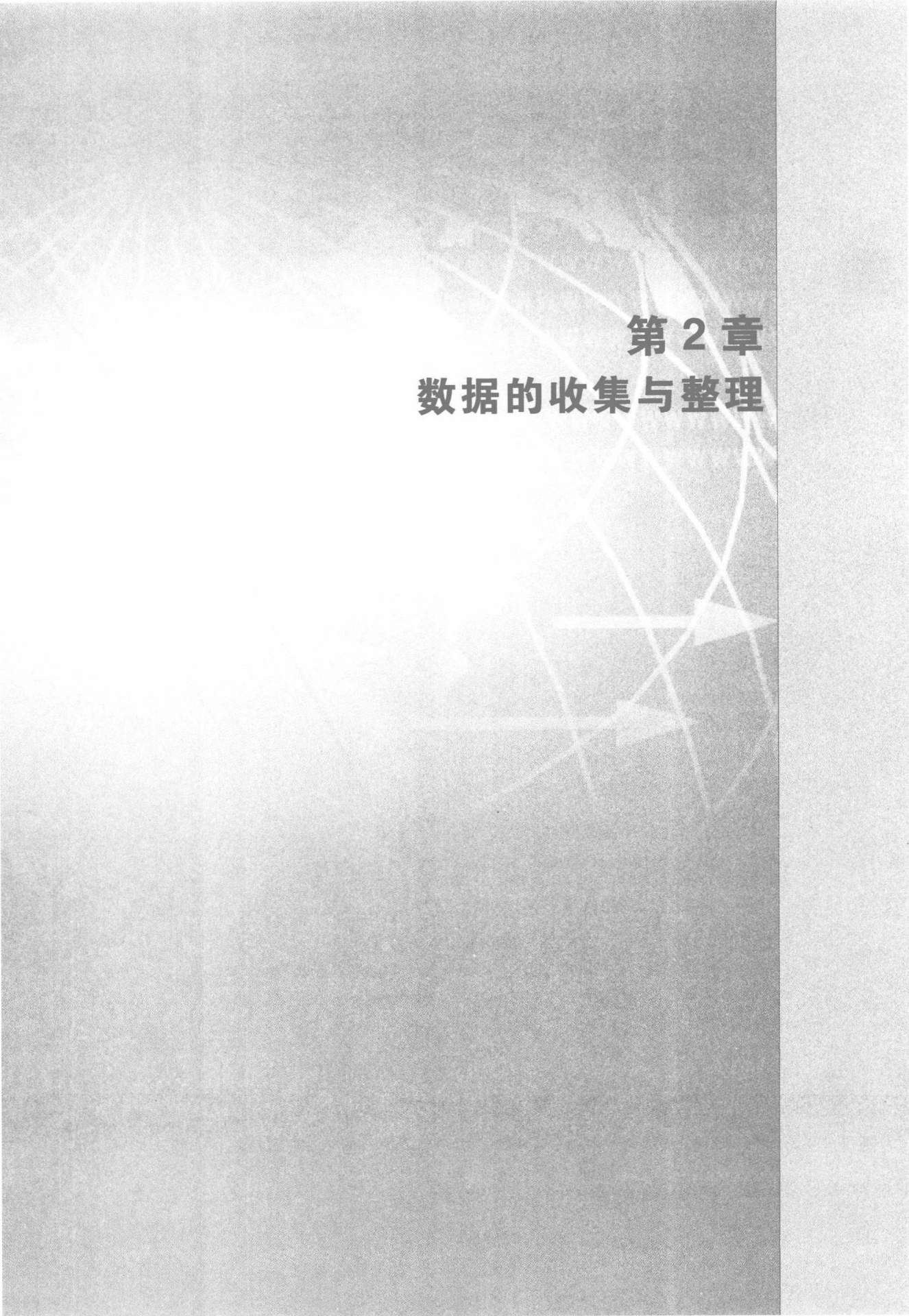
【解析】用样本推断总体,当然存在一定的不确定性(风险),但这种不确定性(风险)并非不可度量。事实上,这正是为这种推断提供理论基础的概率论理论的强大威力所在。例如,在区间估计理论中,对于推断出的置信区间的确定性,就是通过置信水平 $1-\alpha$ 来度量的。故选 B。

8. 下列叙述中,不属于推断统计的描述是【 】。

- A. 从总体中抽取一个样本
- B. 总体必须是已知的
- C. 需要调查一个或若干个变量
- D. 关于可靠性的度量无法确定

【答案】B, D

【解析】正因为总体是未知的,才需要推断统计。它通过考察与样本数据有关的一个或若干个变量,来推断出总体的特征。一句话,统计推断就是用已知的样本来推断未知的总体。另外,统计推断的可靠性是可以度量的。故选 B 和 D。



第 2 章 数据的收集与整理