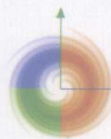


Business Management Statistics



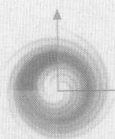
新坐标管理系列精品教材

工商管理 统计学

黄长凌 编著

清华大学出版社



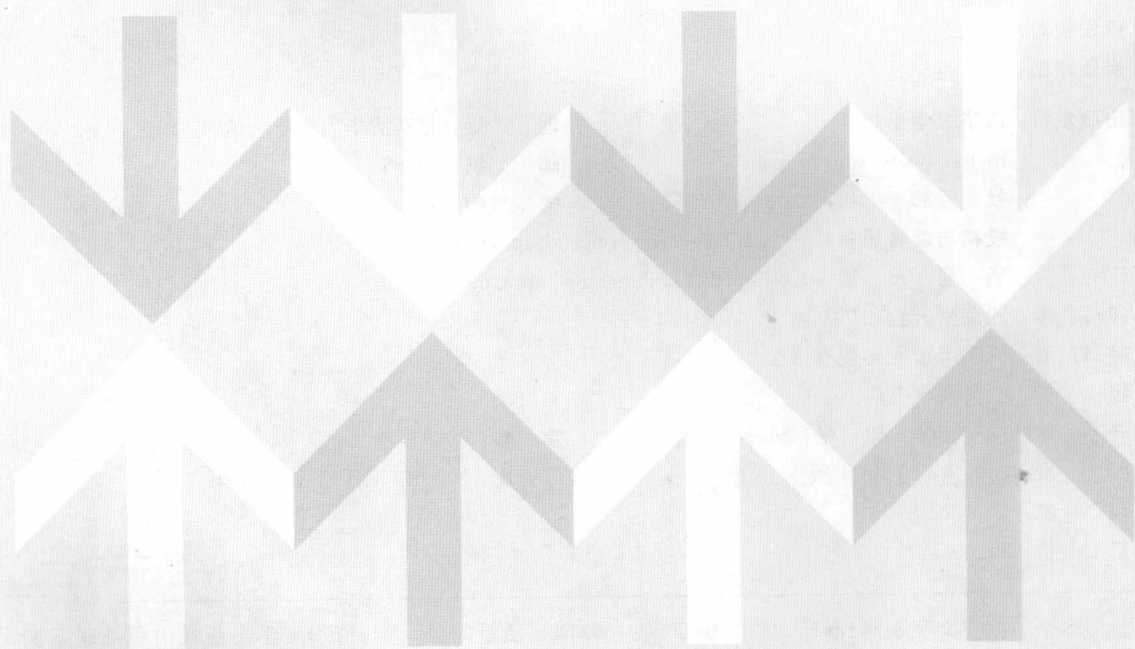


新坐标管理系列精品教材

Business Management Statistics

工商管理 统计学

黄长凌 编著



清华大学出版社
北 京

内 容 简 介

本书的主要内容由三大块构成：描述统计学、推断统计学以及统计分析方法及其应用。

本书的主要特色为：一是根据工商管理决策的需求来讲解统计分析方法及其实际应用；二是把应用统计学模式“识别模型—掌握计算—学会解释”贯穿始终；三是突出管理统计软件（比如 Excel 或 Minitab）的运用；四是汇集大量工商管理领域的实际问题。

本书可作为工商管理大类各专业本科生、研究生的教材，也可供企业各类管理人员阅读参考。

版权所有，侵权必究。侵权举报电话：010-62782989 13701121933

图书在版编目(CIP)数据

工商管理统计学/黄长凌编著. —北京：清华大学出版社，2009.4
(新坐标管理系列精品教材)

ISBN 978-7-302-16398-5

I. 工… II. 黄… III. 工商行政管理—经济统计学 IV. F203.9

中国版本图书馆 CIP 数据核字(2009)第 027967 号

责任编辑：刘志彬

责任校对：王凤芝

责任印制：李红英

出版发行：清华大学出版社

地 址：北京清华大学学研大厦 A 座

<http://www.tup.com.cn>

邮 编：100084

社 总 机：010-62770175

邮 购：010-62786544

投稿与读者服务：010-62776969, c-service@tup.tsinghua.edu.cn

质 量 反 馈：010-62772015, zhiliang@tup.tsinghua.edu.cn

印 刷 者：北京密云胶印厂

装 订 者：三河市金元印装有限公司

经 销：全国新华书店

开 本：185×260 印 张：26.5

字 数：596 千字

版 次：2009 年 4 月第 1 版

印 次：2009 年 4 月第 1 次印刷

印 数：1~4000

定 价：39.00 元

本书如存在文字不清、漏印、缺页、倒页、脱页等印装质量问题，请与清华大学出版社出版部联系调换。联系电话：010-62770177 转 3103 产品编号：019813-01

前 言

国际上应用统计学教科书的编写越来越趋向于结合专业领域的需求来设计课程体系与教学内容,因而也就出现了生物统计学、工科统计学、理科统计学、行为科学统计学、社会科学统计学、经济统计学以及工商管理统计学等应用统计学分支学科。在瞄准国际上统计学教学研究的优秀成果、结合中国工商管理领域的实际情况、凸显专业领域的应用问题并总结我们长期教学实践的基础上,我们构想出了这本《工商管理统计学》的教学内容和课程体系。

经济全球化是企业不可回避的客观事实,因此,工商管理领域越来越需要采用统计分析方法把“数据”转变成“信息”。从这个方面来讲,工商管理类各专业的学生如果仅仅只是侧重于掌握统计学原理和统计学方法,也就远远满足不了未来工商管理决策的需求。毫无疑问,应用统计学课程及其教科书应该向学生提供一幅完整的画面:在讲清楚概念和方法的基础上,应注重讲授如何将统计学概念和统计分析方法应用到管理决策的实际问题之中,把管理者手中的“数据”转变成企业管理决策所需的有用“信息”。正是基于以上的思考,我们创设出这本教科书的基本内容和结构体系。

本书编写的基本设想是:一是注重教材内容的侧重点,在描述统计学与推断统计学的内容上,侧重推断统计学,在统计学分析方法上,侧重方差分析方法与回归分析方法;二是注重应用统计学模式——“识别模型,掌握计算,学会解释”的贯穿始终,这是提高学生解决实际问题能力的重要方式;三是注重统计分析的逻辑性,由此我们特别设计出假设检验“四个步骤”的基本程序,既方便学生记忆检验的程序,同时也提高学生逻辑表述的能力;四是注重管理统计软件(比如 Excel 或 Minitab)的运用,我们应该看到,如果没有管理统计软件,要实际处理方差分析和回归分析问题是十分困难的,因而本书始终贯穿管理统计软件的实际应用;五是注重汇集工商管理的实际问题,在例题、习题以及案例分析上,我们尽量选用生产管理、财务管理、市场营销、人力资源管理 etc 工商管理领域的实际问题,并且注重运用“识别模型,掌握计算,学会解释”的模式来解决这些统计学应用问题。

本书的编写和出版得到各方面的帮助。感谢中南民族大学给予编著者在统计学教学研究和统计学双语教学方面的大力支持,感谢统计学课程的教学同仁为本书撰写所积累的宝贵经验。另外,特别要感谢清华大学出版社的编辑约请本人撰写这本重要的统计学教科书,并对本书的编辑和出版所做的工作。

尽管本书的编著者自认为本书有所创新,但是,成书之后的疏漏或错误之处在所难免,因而恳请统计学教学同仁、工商管理专业学生以及本书读者多提宝贵意见,以便我们共同建设好适应工商管理类专业需要的应用统计学教科书。

黄长凌

2008 年 8 月于中南民族大学

目 录

第 1 章 导言：工商管理统计学	1
第 1 节 管理决策与工商管理统计学	1
第 2 节 统计学的发展简述	3
第 3 节 描述统计学与推断统计学	4
第 4 节 变量的类型与测量的尺度	7
课后习题	10
第 2 章 数据的图表描述	12
第 1 节 频数分布	12
第 2 节 数据的其他图表描述	21
课后习题	27
第 3 章 数据的数字描述	30
第 1 节 数据的数字描述(Ⅰ): 集中量数	30
第 2 节 数据的数字描述(Ⅱ): 离中量数	36
第 3 节 分类数据的统计描述	44
课后习题	48
第 4 章 数据收集与抽样方法	51
第 1 节 企业经营调研	51
第 2 节 市场统计调查	53
第 3 节 试验方法与直接观察法	57
第 4 节 二手数据	59
第 5 节 抽样调查	63
课后习题	72
第 5 章 概率与概率分布	74
第 1 节 概率: 基本术语与基本方法	74
第 2 节 概率的基本运算法则	80
第 3 节 贝叶斯定理及其应用	85
第 4 节 离散型概率分布	88
第 5 节 连续型概率分布	101
课后习题	114

第 6 章 抽样分布	121
第 1 节 样本均值的抽样分布	121
第 2 节 样本比例的抽样分布	126
课后习题	128
第 7 章 参数估计	130
第 1 节 参数估计概述	130
第 2 节 总体均值的区间估计	137
第 3 节 总体比例的区间估计	140
第 4 节 总体方差的区间估计	141
第 5 节 如何确定样本容量	142
课后习题	147
第 8 章 假设检验	151
第 1 节 假设检验概述	151
第 2 节 一个总体均值的假设检验	160
第 3 节 一个总体比例的假设检验	166
第 4 节 一个总体方差的假设检验	169
课后习题	172
第 9 章 两个总体比较的差异性检验	176
第 1 节 两个总体均值之差的假设检验	176
第 2 节 两个总体比例之差的假设检验	189
第 3 节 两个总体方差之比的假设检验	192
课后习题	193
第 10 章 方差分析	197
第 1 节 方差分析概述	197
第 2 节 单因素方差分析	199
第 3 节 随机区组设计	206
第 4 节 双因素方差分析	211
课后习题	217
第 11 章 χ^2 检验方法	225
第 1 节 拟合良好性检验与总体分布的正态性	225
第 2 节 列联表上两个变量的独立性检验	233

课后习题	237
第 12 章 简单线性回归与相关分析	242
第 1 节 简单线性回归模型	243
第 2 节 采用样本回归直线的区间估计	252
第 3 节 关于线性回归方程的假设检验	259
第 4 节 相关分析	266
课后习题	274
第 13 章 多元回归分析	281
第 1 节 多元线性回归模型	281
第 2 节 多元回归的显著性检验	285
第 3 节 采用样本回归方程进行估计和预测	291
第 4 节 回归模型的建构	293
课后习题	301
第 14 章 时间序列分析与预测	310
第 1 节 时间序列组成要素与模型	310
第 2 节 时间序列分析的修匀方法	318
第 3 节 应用 I：长期趋势与季节效应	325
第 4 节 应用 II：预测	338
课后习题	343
第 15 章 质量控制的统计方法	349
第 1 节 生产过程的变动偏差	350
第 2 节 统计控制图表	352
第 3 节 变量控制图表	358
第 4 节 属性控制图表	367
课后习题	372
附录 常用统计表	374

第 1 章 导言：工商管理统计学

工商管理(或称企业管理)决策需要关注数字,用数字说话正是企业管理的特征之一。对于企业管理者而言,数字可以帮助你面对现实,数字又有助于你做出有效的管理决策,而这些数字就是用一系列统计分析的工具所得出来的。正如英文“statistics”一词那样,它有两个基本的意义:一是指某一特别的数字;另一则是指某种分析的工具。企业管理者要做出有效的管理决策,也就需要数字的帮助,同时还得掌握一系列统计分析的方法,使自己能够识别现实中各种各样的数字,并且能够采用统计分析方法把企业调查所获得的“数据”转变成企业管理决策所需的“统计信息”,由此正确运用“统计资料”来帮助实施组织的管理。也就是说,工商管理领域需要适应管理需求的统计分析方法,工商管理类各专业的学生必须学习工商管理统计学。

第 1 节 管理决策与工商管理统计学

一、管理决策需要数字

工商管理领域像许多其他专业领域一样,也应有自己的专业术语,那么“数量化”就是其中最基本的术语之一。企业要做出有效的管理决策,在其管理工作中就要坚持数量化的原则,因为简单的数据可以帮助我们弄清事实真相,而仅仅只是依靠我们的直觉和经验是无法掌握的。例如,一些卓越的公司(如通用电气公司、摩托罗拉公司)在质量管理上一直坚持“六西格玛”理念,并且现在已成为众多公司搞好质量控制的信条,不过,它实质上也就是统计学中的一个比率数字,告诉企业管理者在自己的质量控制行为中没有出错的比率。

在统计学中,统计学家运用“西格玛”(σ)这一概念来表示一个标准差。“ 1σ ”表示 31% 的产品合格,“ 3σ ”表示 93.32% 的产品合格,“ 6σ ”表示 99.999 66% 的产品合格,也就是说,在每百万件产品中,不合格的产品仅仅只有 3.4 件。然而,到了 20 世纪 90 年代,还有很多公司的产品合格率只维持在“ 3.5σ ”的水平,也就是说,在每百万件产品中,不合格的产品有 2.28 万件。当企业管理者知道了这样一些数字的意义,也就明确了应该怎样来改进企业的产品质量。

从以上这个例子可以看出,统计数据对企业的表现至关重要,而一些基本的工商管理数据也并非像深奥的科学那样难懂,管理者不必害怕受制于数学知识的障碍。从哈佛商学院 MBA 专业的课程设置来看,学生在第一年的学习中,必须按规定学习统计分析

的课程,确实把握“怎样避免不懂数据”的问题。对于 MBA 的学生来说,真正的统计分析技巧是在于把握数据的意义,也就是说,使“数据”变成有意义的“统计信息”。当然,管理者有时还要通过掌握一系列统计分析的方法和技巧,才能真正理解数据或数字的意义。

总而言之,企业管理决策所要寻求的数字,是依据你自己的情况和你所想要了解的内容而确定的。一个投资公司在考虑向一个企业的某一项目投资的时候,它所最关心的问题就是评估该企业能否如期还款的信用问题。对于一个企业的 CEO 来说,他就应该十分看重负债率问题,因为从某种意义上讲,负债率说明了一个企业偿还债权人债务的能力——相对于企业的赢利能力而言,负债越多,资金缺口越大,债权人所冒的风险就会越大。这样看来,管理决策需要培养管理者从数据中发现问题的能力,当然,理解一个特别的数据需要长期的经验,但更重要的是,企业管理者需要学习统计学知识,尤其是要掌握把“数据”转换成“统计信息”的统计学分析方法和技巧。

二、何谓工商管理统计学

工商管理统计学可定义为:针对企业管理决策,对数字数据进行收集、描述、分析并对统计结论给以报告和解释的系统性方法和技巧。简言之,工商管理统计学就是根据管理决策的需要从“数据”中获得“信息”的方法和技巧,其目的在于帮助企业管理者把握和理解统计信息,以便运用统计数字做出有效的管理决策。无论是哪个行业的企业,也不论我们工商管理类的学生未来从事哪种职业(比如会计控制、财务管理、市场营销、信息管理、企业战略、员工管理、生产运作等),各种类型的企业管理工作都需要统计学知识。后面的学习将会使我们看到,统计学不仅仅只是图表和平均数字的计算,更重要的是,统计学是一种强有力的分析工具,它能够通过最初的样本数据推断出有关某一总体的一般性结论,并且能够对其数字意义做出相关的解释。因此,统计学分析方法目前已广泛应用于工商管理的各个领域。本教科书着重要讨论的内容是,企业管理者和统计分析人员如何采用统计分析方法,何时进行统计分析以及为何要进行统计分析。

作为一个企业管理者,你将会看到,你将以两种身份与统计学相关联:(1)你是运用统计分析方法的实践者,在你的工作中,你要收集、描述、分析数据,从而编制出用于企业管理决策的统计信息(统计表、统计图以及推断总体的相关统计数字);(2)你是使用统计信息的消费者,在你的工作中,你会使用别人向你提供的统计资料,而在这些人中间,就有可能存在缺乏能力或缺乏良好职业道德的统计信息的提供者。

我们这本教科书主要向你提供“如何做”统计分析的方法和技巧,使你成为熟练运用统计学方法的“实践者”,为帮助企业管理决策提供可靠的统计信息。另外,如果你熟练掌握了收集、描述、分析、报告、解释数字数据的方法和技巧,那么,你与统计学相关联的第二种角色也能实现。你就会对别人提供的统计信息能够加以识别,保护你的公司和你个人不会受到某些统计信息的误导。

第2节 统计学的发展简述

一、昨日的统计学

收集统计数据以利于管理控制已有几千年的历史。不过早期的统计活动只不过是点计人口和财产,其目的是为了有助于提高国家税收。这种类似于今日的人口调查和经济普查的统计活动仍延续到21世纪。然而在20世纪20年代,人们发现统计学的作用已有了转变。

统计学小资料：统计推断在企业的应用^①

(1925年左右)在企业界和政治领域,运用统计学方法和进行统计思考已成为一种惯例。大型企业不仅设置了自己的统计机构,收集、描述、解释有关自己经营活动的数据,而且还是别人提供的统计信息的消费者。商业出版物和政府文献中有大量的统计资料,这是非常必要的,因为从事企业经营活动和实施政府管理需要借助统计资料来帮助做出有效的决策。

作为企业,它需要对过去的经营活动做出记载,比如销售额、成本、原材料来源、市场上的设备等。由此所反映出的企业经营状况,就可以用来考核经营的进展、财务的状况以及经济的增长。至于企业经营变化方面的记录——企业经营业务的增长和减少以及相关的影响因素——也就可以用于估计未来的发展——根据过去的历史资料来预测未来,这对于企业管理是非常必要的。

这里请注意,小资料中所指出的“估计未来的发展”就标志着统计学的作用已有了根本性的转变。统计学已从过去相对被动的记录数据和描述数据转变为日益主动并有助于企业经营的管理工具,管理者运用这一统计分析工具来帮助他们做出有效的企业决策,并且能够从样本数据推断出关于总体的相关结论。这样,有助于把企业管理决策的推断统计学方法应用到企业管理的各个领域。

二、今日的统计学

当今,统计学及其各种统计分析方法已成为我们生活的一个组成部分。政治、法律、企业、教育、科学、卫生等各个领域的活动,既受到统计方面的考察,又受到统计信息的引导。比如,我们在市场上的行为产生了经营销售的统计数字,这些数字可能会帮助企业做出有关维持产品产量、减少产品产量或增加产品产量的管理决策。同样,汽车保险公司要收集有关年龄、车型、事故率等方面的统计信息,运用这些统计资料,保险公司才可能设计出恰当合理的保险类别和保险金额。

^① Horace Secrist, An Introduction to Statistical Methods, rev. ed., New York: Macmillan Co., 1925, 1.

当今社会,统计信息已渗透到人们的工作领域和日常生活领域。政府部门有专门的统计机构(各级政府的统计局),每年发布各种各样的统计公报,有关人口、收入、职业、运输、住房等方面的统计信息是人们尤为关注的。一些统计数字是通过抽样调查,运用统计推断而得出的。政府进行人口普查和经济普查,却要很长时间才进行一次。除了政府发布的统计信息之外,还有一些独立的调查机构和管理咨询公司向市场提供的有偿和无偿的统计信息服务,这些统计信息也可能会对企业管理有着影响。

在我们这本教科书中,我们将针对工商管理决策的需求来考察今日统计学四个方面的作用:(1)信息的描述作用;(2)分析的工具作用;(3)推断结论的方法作用;(4)帮助管理决策的助手作用。把握这四个方面的作用,既有助于我们成为胜任统计数据分析和统计信息编制的实践者,又能使我们成为识别市场上种类繁多的统计信息的消费者。尤其是在工商管理领域,统计分析实际上已在各个方面发挥着重要的作用,本书将主要选择财务管理、市场营销、人力资源管理、生产管理等方面的事例来讨论统计学在工商管理领域的应用问题,并且用统计学概念来解释这些应用问题的结论。

第3节 描述统计学与推断统计学

前面已经指出,统计学一词既是指某一组特别的数字或事实,又是指某种统计分析的工具。一般来讲,运用统计分析工具解决实际问题都会涉及三项任务,即收集数据、整理探究数据以及依据数据来推断结论、做出决策。根据企业管理决策的需求,我们有时侧重于只是描述一组数据的特征;有时又得注重从一组数据的特征做出一般性结论,或是根据收集的样本数据做出估计、预测及相关的推断。这样,工商管理统计学可以分为描述统计学(descriptive statistics,又称之为探索性数据分析)和推断统计学(inferential statistics,又称之为证实性数据分析)。对于今日的企业管理决策而言,这两种统计方法都是十分重要的。

一、描述统计学

对于描述统计学,我们只是对收集的数据进行概括和描述。也就是说,描述统计学是指对数据进行初步整理、基本汇总、简单编制的系统方法。例如,考察你自己所在学院的高等数学成绩,你可能会发现,有86%的学生及格。这里,“86%”这一数字就只是一个描述统计数值。你并没有试图推测,全中国大学生的高等数学成绩有86%的学生及格,你只不过对收集到的数据做出了描述而已。在企业管理中,我们经常要对企业的经营状况(如营业收入、生产成本、管理费用等)用统计数字进行描述,这样才能使我们在管理中做到“心中有数”。为了较好地概括和描述我们已收集到的数据,我们一般采用形象化的图表描述方法和统计性的数字描述方法。然而,你现在所要记住的是,描述统计学仅仅只用于对已收集到的数据进行概括和描述。

二、推断统计学

对于推断统计学,不仅是对收集到的数据进行描述,我们还会根据收集到的样本数据来对某些现象或事物得出某种推断。例如,中国中央电视台(CCTV)往往只是根据几个家庭的收视率的调查,来确定哪些节目是优秀的,哪些节目在主要时段应该取消。之所以会做出这样一些决策,管理者是根据这些少量的样本数据做出了推断:全国几亿观众都具有这样的收视行为。所以,收视率调查成为每个节目负责人都非常关注的事情。

在企业管理中,某一行业的产品市场占有率的调查也采用的是推断统计学的方法,所获得的购买行为的数据只是来自少量的消费者,但预测的却是所有消费者的购买行为。然而,企业有时却会根据产品的市场占有率来做决定,是增加产量还是减少产量。实际上,企业管理决策确实会受到推断统计结论的影响,因此,推断统计学是指根据样本数据推断出关于总体结论的系统方法。我们这本教科书将会有较大的篇幅来介绍推断统计学的概念和方法。

三、推断统计学的主要术语

在前面举例所谈到的产品市场占有率的调查中,统计分析人员是从所有消费者构成的总体(population)中抽取少量的具有代表性的消费者组成一个样本(sample),然后使用样本数据进行考察。根据对样本数据观测得出的结果,统计分析人员由此推断,所有消费者都具有样本数据所反映出来的购买行为。比如,某一电子产品A的市场占有率为32%,在这里,结论是总体性的结论,而这一总体性结论却是根据样本数据推断得出的。我们把通过观测样本数据所得出的结果称之为“样本统计量”(sample statistic),由此来估计的总体特征值称之为“总体参数”(population parameter)。推断统计学所涉及的一些主要概念或术语定义如下:

■ 总体(population)是指统计调查所针对的研究对象的全体,它是由统计调查所包括的全部个体组成的集合。这一集合中的元素既可以是人,也可以是物。它可以指某种产品的全体消费者,也可以指英特尔公司所生产的每一个CPU产品。

由于总体既涉及人又可涉及物,因此,在进行统计调研之前,一定要明确地确定总体的属性和范围。例如,我们若是要调查星期五消费者的购买行为,零售商就必须清楚地确定,所要调查的“顾客”是指营业时间进入商店的所有人。当然,我们对统计调查的总体也可采取其他的方式来确定。但是,我们必须记住,如果你首先不对总体的属性和范围有一个明确的界定,那么,你就不可能得到一个较好的样本来进行统计分析。

■ 样本(sample)是指从总体集合中抽取少量元素所组成的总体的一个真子集。在推断统计学中,我们之所以要从大量个体构成的总体中抽取少量个体作为样本来进行统计分析,其目的是为了从样本所提供的信息中,对总体特征做出推断。例如,前面所述的零售商对消费者购买行为的调查, he 可以从星期五营业时间进入商店的顾客中随机抽取1 000人构成一个样本,然后从这1 000个人的购买行为去推断全体消费者的购买行为。

对于样本,要特别注意的是,只有样本元素与调研对象的总体具有相同的特征(比如,购买行为、教育层次、收入、年龄等)时,才能说这一样本具有代表性。例如,某一总体的成员由45%的女性购买者组成,那么,随机抽取的任一样本也应包括45%的女性。当样本逐渐增大,以至于包括所有总体元素时,这种统计调查就称之为“全面性普查”。

■ 统计量是指描述样本特征的数字化量数。例如,在统计调查中,某一零售商发现,样本成员中有72%的购买者对商店的商品质量评价高于附近的其他商店。这里的数字72%就是一个统计量,因为它是根据样本得出的。样本统计量可以是集中量数,比如均值或平均数(mean)、中位数(median)、众数(mode)或比例(proportion);样本统计量也可以是差异量数,比如全距或极差(range)和标准差(standard deviation)。在统计学中,统计量一般用英文字母表示,比如,我们用 $\bar{x}$ 表示样本均值,用 s 表示样本标准差,用 p 表示样本比例。

样本均值(sample mean)一般是指一组数据的算术平均数(arithmetic average),即全组数据之和除以数据的个数。比如数据组2、5、8的均值可以通过 $(2+5+8) \div 3$ 得出等于5。

样本中位数(sample median)是指一组有序数据的中点值(midpoint),它把一组有序数据分成相等的两部分。比如有序数据组3、5、9、13、17的中位数为9,比它小的有两个数据3和5,比它大的也有两个数据13和17。

样本众数(sample mode)是指一组数据中出现次数最多的那个数据值。比如一组数据由9、12、15、20、15、34、15、18构成,其中15出现的次数最多,所以15就是这组数据的众数。

样本比例(sample proportion)就是表示十进制小数的一个百分数。比如36.4%所表示的就是一个十进制小数0.364。

样本全距(sample range)是指一组数据中最大值与最小值之差。比如数据组5、8、9、14的全距为 $(14 - 5) = 9$ 。

样本方差(sample variance)和样本标准差(sample standard deviation)是两个重要的差异量数,求这样两种统计量的值要根据相应的计算公式来求出,我们在后面的内容将会做出详细的介绍。

■ 参数(parameter)是指描述总体特征的数字化量数。如果在统计调查中对总体采取全面性的普查(census),那么我们就能够测量出总体的参数。然而,对于大多数企业管理调研而言,进行全面性的普查实际是不可行的。这样,对于这种存在而又未知的总体参数,我们就只能通过对样本统计量的分析来估计总体参数(比如用样本均值来估计总体均值)。在统计分析中,我们经常使用的参数有总体均值、总体中位数、总体比例、总体方差以及总体标准差。一般地,参数用希腊字母表示,比如, μ 用来表示总体均值, σ 用来表示总体标准差, π 用来表示总体比例。

在前面所述的消费者购买行为的调查中,零售商不可能获得实际的总体比例(这一实际存在但又未知的量数就是参数)。然而,该零售商可以用统计调查得出的样本统计量(72%)来对总体比例做出一个估计,以此来帮助企业进行管理决策,进而改进自己的经营。企业之所以要这样做,是因为从人力、物力、财力以及时间上考虑,对这样的统计调查进行

全面性的普查在实际上是不可行的。

■ 统计推断(statistical inference)是指根据样本数据对总体做出估计、预测、决策的系列方法。由于我们所要研究的总体几乎总是规模很大,因此要对总体的每个个体进行考察就不太切合实际,并且花费很大,在时间上也受到限制。

针对这种情况,统计推断就试图从较大的总体中抽取较小的样本来做出关于总体特性的一些推断。这就使得统计研究变得比较便捷又花费较少。然而,我们必须记住,这样做出的关于总体的结论和估计也总会发生错误。因此,对于统计推断,我们就要建构起关于统计推断的可靠性(reliability)的度量标准。一般来说,在统计推断中,有两种可靠性的度量标准:置信水平(confidence level)和显著性水平(significance level)。如果抽样过程大量地重复许多次,置信水平是指一项统计估计程序可能正确的比例值。当统计推断的意图是要得出关于总体的相关结论时,显著性水平是指统计结论将会发生错误的比例值。

第4节 变量的类型与测量的尺度

在统计调查或实验中,所谓数据(data)就是你收集得来的观测数值。一大堆孤立而枯燥的数据有待于进行统计处理,不然就毫无意义。因此,如果你要有效地运用统计分析方法,你就必须首先知道数据的具体背景、数据获得的具体方式以及数据来源的总体状况。应该知道获得样本数据的抽样方法(sampling techniques)和随机原理(random principles)对于统计推断来说,是非常重要的。

在统计研究中,有些数据组是由数字构成的(如人的体重),有些数据组是由非数字构成的(如人的性别)。我们采用定量变量(quantitative variables)和定性变量(qualitative variables)的术语来区分这两种类型的数据。有时,我们也可以采用测量尺度(scales of measurement)的方式来区分不同的数据类型。采取这种方式,我们一般把数据分成四种类型:定类尺度(nominal scale)、定序尺度(ordinal scale)、定距尺度(interval scale)以及定比尺度(ratio scale)。

一、定性变量

有些变量只与人或物相关,我们自然也就只需划分不同的类别来区分这些非数字特征的变量,这些定性变量仅仅表明某个人或某个物属于某一类别。例如:(1)你是男性还是女性;(2)你上周消费的饮料是可口可乐还是百事可乐;(3)你购买的计算机是清华同方计算机还是联想计算机;(4)你的统计学考试成绩是优秀还是良好、中等、及格或不及格。定性变量往往涉及人或物的属性(attributes),比较典型的是点计有多少人或多少物落入了所划分的每一个类别或每一个组别。

我们在表示定性变量的结果时,往往用百分数或频数来描述所划分的每一类别的特征。例如,调查大学生的饮水习惯时发现,有64%的学生饮用白开水,另外36%的学生却饮用瓶装的纯净水。再如,在某一次200人的调查中发现,喜欢统计学课程的有76人,不

喜欢的有 124 人。

统计学小资料：“statistics”一词的来历^①

英语“statistics”一词起源于拉丁语的“status”(意即“国家”之意)。早期的统计学,仅仅只是数据和图表的汇编,用以描述一个国家状况的各个方面。1662 年,英国统计学家约翰·格朗特(John Graunt)发表了关于人口出生率和人口死亡率的统计信息。紧随格朗特其后,一些统计学家又接着研究了死亡率、疾病发生率、人口规模、居民收入以及失业率等问题。这样,就开始了家庭(或个人)、政府以及企业运用统计信息来帮助决策。例如一些统计数字,如失业率、通货膨胀率、居民消费价格指数、出生率、死亡率等,往往被企业管理者拿来使用,以帮助他们尽可能做出有效的企业决策,进而又影响到企业未来的员工招聘、生产规模以及新产品研发和新工厂的建立。

二、定量变量

定量变量是由数字构成的数据变量,用以表示点计或测量的结果。这样的变量能够使我们确定,对于某种东西我们究竟拥有多大的数量。一般来说,我们把定量变量分成两种类型:离散型变量(discrete variables)和连续型变量(continuous variables)。

离散型变量是指在某一区间内只能取某些有限的数值,或所取的数值可以逐一列举。这样一些所有可能的变量值之间存在着间隙。比如,某一 IT 企业工资总表上员工的人数、某一制造业企业 A 型生产线上产品的次品数量、某一次义演的赞助人数等,就是离散型变量的例子。在统计学中,离散型数据变量往往是由调查点计的观测值构成,并且一般是整数。但是,观测值也可能是分数值。例如,我们在超市购物时,所购买的牛奶(单位: L)就有 0.25、0.50、0.75、1.00、1.25、1.50 等几种规格。这些观测值就是离散型变量的分数值。我们必须记住的是,离散型变量的特征是什么观测值之间都存在着间隙。

连续型变量是指在某一区间内能取任何的数值。例如,某一型号电热水器中水的容量就可能是 0~50(L)之间的任一数值。其他的例子还有:石油公司所生产的石油产量、乳品公司所生产的牛奶量、某一人群的身高、汽车行驶的里程数等,这些也都是连续型变量。连续型变量的特征是能取无限的数值,所取值的区间没有间隙、没有中断、没有跳跃。

三、测量尺度

针对某一变量指定一个数值,就称之为测量。例如,我们看着一个温度计,可以观测到一个温度(变量),其读数为 28℃。这里的数值 28,就是对温度(变量)测量的结果。在统计学中,当我们要对变量进行测量时,以四种层次来区分测量的结果,它们分别是:定类

^① Mario F. Triola, Elementary Statistics, 8th, ed. Boston: Addison Wesley, 2001, 5.

尺度、定序尺度、定距尺度和定比尺度。它们之间的特征区分如图 1-1 所示。在应用统计分析方法于实际问题时,明确测量尺度是十分重要的,因为它决定了描述和分析数据的恰当方法。

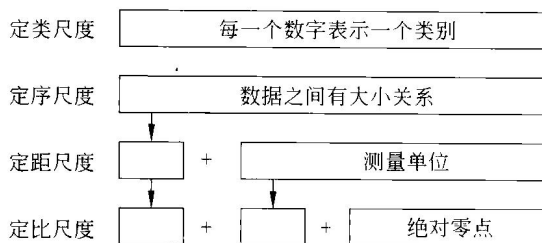


图 1-1 四种测量尺度的特征

定类尺度(nominal scale)测量的数据仅仅由名称、标号或类别构成,数据不能按大小顺序排列。定类尺度在使用数字时,也只是为了识别每一个类别的成员,尤其是在计算机的统计分析中,往往不是用名称来表示类别,而是用数字来表示类别。例如,GE 能源公司可能会列出五种类型的客户,当他们用计算机来记录时,就会用数字“1”来表示居民住宅客户,用数字“2”来表示商业客户,用数字“3”来表示工业客户,用数字“4”来表示政府客户,用数字“5”来表示其他客户。由于定类尺度的数据仅能识别类别,它们就缺乏次序或算术意义,因而也就不能用于计算。

定序尺度(ordinal scale)测量的数据能按某种次序进行排列,但数据值之间的差却不能确定,因而也不能用于计算。我们经常用到的定序尺度的数据是学生学业成绩考核的“五分等级制”:优秀、良好、中等、及格和不及格。这些考核的等级分数能够按次序进行排列,但是,我们却不能确定每个等级之间的差。比如,我们知道等级“优秀”优于等级“良好”(其他等级之间也是如此),但是,我们不能求得(优秀-良好)的差。另外,我们还常常见到新闻媒体对城市的“居住质量”进行排序评定。这种排序评定是根据多项指标综合评比出来的(第一,第二,第三,……),我们只能知道次序的比较,要决定每个等级之间的差异则是毫无意义的,因为定序尺度没有设置测量的单位。

定距尺度(interval scale)测量的数据不仅指定了数据之间的大小关系,而且还规定了测量单位,这就能够使我们对两个数据之间差异的多少做出准确的描述。例如,我们常见的人的体温的测量数据就是定距尺度的数据,体温 37.2℃ 和 38.6℃ 不仅可以知道它们之间的大小关系,而且还可以求出它们的差(有时又称之为“两个数值之间的距离”)。这是因为摄氏度数的标记指定了测量单位。在定距尺度的测量中,测量单位是任意的,因而也就没有具体指定绝对零点。这样,测量值的倍数关系也就没有意义——比如摄氏 2 度并不是摄氏 1 度的两倍。

定比尺度(ratio scale)测量的数据与定距尺度的数据类似,但它指定了绝对零点,因而数据之间的倍数关系就有了意义。在企业管理决策中,统计分析人员经常会接触到一些定比尺度的数据,比如投资收益率、产品的市场占有率、石油消耗量、生产线的装配速度、道琼斯工业股票指数、居民消费价格指数等,就是定比尺度测量的一些例子。

统计学小资料：统计学与计算机^①

在统计学的各种实际应用中,统计分析人员必须处理大量的数据。比如,为了估计每年的利润,统计分析人员不得不进行大量的数据计算才能做出这样的估计。即使作这样的计算并不需要复杂的数学技能,但是,就是大量的算数计算也使得这方面的统计方法要消耗相当长的时间,并且冗长乏味。幸好,我们目前可以获得进行这样一些计算所需要的管理统计软件。比如,我们可以选用微软公司所开发的 Excel,这是一种电子数据表软件;还可以选用另一种统计软件 Minitab,这是一种统计软件包。Excel 之所以在工商管理领域应用比较广泛,那是因为管理人员认为,它是目前使用便捷、应用广泛又最受欢迎的管理统计软件。但是,Excel 还是存在一些缺陷,它所提供的统计方法显得少了一些。因此,如果针对 Excel 再添加一些功能,则统计分析所需要的各种统计方法也就都可以采用 Excel 来解决大量的实际问题。目前有一种“Data Analysis Plus[®]”程序就可以使 Excel 的功能扩大。另外,Minitab 也是在企业管理领域应用比较广泛的一种统计软件包,这一管理统计软件同样具有多样化统计方法的特点。所以说,有了计算机之后,统计学课程学习中必不可少的数据处理问题与统计分析问题才得以获得解决。

现在,我们就要开始工商管理统计学的学习了!虽然统计学在企业管理决策中起着重要的作用,不论是生产规模扩展、新产品研发以及产品质量控制,还是人力资源管理、业务流程设计以及供应商选择,企业管理决策都会面临着不确定性因素,在这种情况下,运用统计分析方法就可以帮助管理者进行有效的企业决策。但是,在运用统计信息帮助我们决策时,千万要记住伟大的科学家爱因斯坦的名言:“并非所有可以统计的东西都已经进行了统计,并非所有统计过的东西都可以实施统计。”^②这就是说,统计方法也可能会因为个人或组织为了获利的缘故,而进行统计上的滥用或乱用。这样的话,一些统计信息就会误导或操纵我们的行为。因此,我们不仅要学习统计学的知识、方法和技能,还应该增强我们识别市场上统计信息的能力。一般来讲,当你获得的是其他人向你提供的统计信息时,你应该自问一下:谁在进行这项统计调查?谁在做统计分析?从得出的统计结论中,他们处于什么样的获利地位?如果在学习、掌握了统计学之后,有了这样的思考头脑,那么,统计学方法的运用才可以有助于我们做出有效的管理决策。

**课后习题**

1.1 用你自己的语言,对下列统计术语给出定义,并列举一个例子:

- (1) 总体;
- (2) 样本;

^① Gerald Keller, Statistics for Management and Economics. Belmont: Duxbury. 2005.6-7.

^② 琼·玛格丽塔,南·斯通. 什么是管理[M]. 北京:电子工业出版社,2003年7月. 135.