



管理统计 简明

(增修版)

贾怀勤 编著



中国统计出版社
China Statistics Press

014004365

F222
67
2

管理统计 简明

(增修版)

贾怀勤 编著



 中国统计出版社
China Statistics Press



北航

C1691816

F222
67-2

288400410

图书在版编目(CIP)数据

简明管理统计 / 贾怀勤编著. — 北京: 中国统计出版社, 2013. 8

ISBN 978-7-5037-6838-5

I. ①简… II. ①贾… III. ①经济统计学—研究生—教材 IV. ①F222

中国版本图书馆 CIP 数据核字(2013)第 133993 号

简明管理统计(增修版)

作 者/贾怀勤

责任编辑/徐 颖

装帧设计/黄 晨

出版发行/中国统计出版社

通信地址/北京市丰台区西三环南路甲 6 号 邮政编码/100073

电 话/邮购(010)63376909 书店(010)68783171

网 址/<http://csp.stats.gov.cn>

印 刷/河北天普润印刷厂

经 销/新华书店

开 本/710×1000 mm 1/16

字 数/260 千字

印 张/17.25

印 数/1—2000 册

版 别/2013 年 8 月第 1 版

版 次/2013 年 8 月第 1 次印刷

定 价/34.00 元

版权所有。未经许可,本书的任何部分不得以任何方式在

世界任何地区以任何文字翻印、拷贝、仿制或转载。

如有印装差错,由本社发行部调换。

增修版编写说明

为使本教材更好地适用教学需要,本版在以下几方面做了增写和修订。

一、结构更趋合理,知识脉络更为清晰

保持原书十二章结构,章上加“篇”。第一篇“统计推断的预备知识”,含第1~3章。第二篇“围绕一个变量的统计推断”,含第4~10章。第三篇“变量间联系的描述与推断”,含第11~12章。

为使读者对全书各章节的统计推断知识既有总括了解,又能具体分辨其应用场合,增写了两部分内容:(1)绘制了对全书知识起到总括作用的“统计推断的知识框架”图,并在第一章中增写第5节,对该图做出文字说明。(2)绘制了各章的章节知识点位图,在每章前面列出。

二、新增知识点

第12章第4节第3小节在原有“以双值定名数据为因变量的 LOGISTIC 回归”的基础上,增写了“以多值定名数据为因变量的 LOGISTIC 回归”和“以定秩数据为因变量的 LOGISTIC 回归”。

三、增写 EXCEL 的统计函数和统计分析工具用法

本书仍以介绍 Minitab 软件的用法为主,在一些例题解答后面附加 Minitab 软件的逐步进入指令和打印输出。为了使不使用大型统计软件的教师和学生也找到上机的解法,在每章后面增写 EXCEL 的相关函数和分析工具用法。这一点对第三章特别重要,因为大型统计软件都是一键到底出结果,不管概率分布的求解,EXCEL 正好弥补这方面不足。

四、增写系列性案例和更换例题

对一些例题进行更新,在题目所属时间和表现内容上进一步体现

当前经济和管理实践。

增写 9 个系列性综合案例,贯穿全书。这样的案例更接近经济和管理实践,有大型数据集支持。所需解决问题的知识,不限于某一章节,而是要求综合运用有关章节的方法和模型。

所有例题和系列性综合案例的数据,都可以索取到。请选用本教材的任课教师与作者直接联系。作者邮箱:jiahq@uibe.edu.cn。

作者

2013 年 5 月

前言

本书定名为“简明管理统计”，力求适应近年来管理类研究生统计教学的发展变化。

我国管理类研究生的统计学教学经历了由综合到分解的过程，先是力图将经济和管理所需用的统计知识和方法尽可能地装进一门统计课程中，后来逐渐平行地开设出多门统计分支学科课程，而把基干的统计知识和方法保留在“管理统计”中。尽管如此，“管理统计”教科书的编者还是在书中尽可能多地包括管理统计的各方面知识和各种方法，于是造成各相关课程教学内容的不适当重复。本书的编写，就是考虑到学生在修习“管理统计”之后还将修习“商务预测”、“多元统计分析”和“计量经济学”等课程，将“管理统计”的知识点限定于基干的统计知识和方法，不与后续课程的教科书重复。此乃“简”字的第一个涵义。

近年来统计学教科书的另一个趋势是，随着一些大型专门统计分析软件的引进，教科书在阐述统计分析方法时，常常附带列出软件程序语言或视窗对话框截图，为的是方便教师讲授和学生演练，以至于软件操作介绍的文字和截图占据了相当大篇幅。本书作者认为，随着这些大型专门统计分析软件工具书的编印、发行和操作知识的逐渐普及，在管理统计教科书中实无必要再详列操作指南内容。因此，本书只将有关分析结果打印输出，以增强学生对统计方法的理解，而对软件工具只给出其进入路径。此乃“简”字的第二个涵义。

管理类研究生在其本科阶段有的学过统计学相关课程，有的则根本没有系统地接触过统计学知识。为此，这门课程不得不从统计基础知识讲起，但又不能完全重复本科统计学的知识点和那套路子。如何使后一种背景的研究生在有限的学时内掌握由统计描述到统计推断的基干知识和方法，同时又使前一种背景的研究生感到修习此课程有新的收获，在教材的知识展开框架上必须有双面适应的独到之处。对此，本书(1)在知识展开层次上依“经验数据的统计描述”——“数据的理论分布模型”——“分布特征的统计推断”线索渐次展开；(2)在统计推断

方法上按定标数据、定秩数据和定名数据三个级次分别演绎;(3)在数据分布基本趋势研究上遵循先集中后离散的顺序;(4)在变量数目上从单一变量到双变量再到多变量逐步推进。为全书搭建这样一个架构,无非是想避免学生接受了一大串统计方法而不辨其应用对象和场合的乱用和统计分析软件功能的滥用。此乃“明”字的涵义。

书中转引和改编的例题尽可能注明出处。另外一些例题则是在近年间 MBA 学生提供的案例素材基础上编写的。仅此向有关资料提供者致谢。

对于书中可能存在的不妥和疏漏之处,诚恳欢迎同行和读者不吝指正。



2007 年 8 月于惠园

目 录

第一篇 统计推断的预备知识

第 1 章 数据级次和数据整理 (3)

1.1 统计调查与统计推断 (3)

1.2 数据级次和分类 (5)

1.3 数据整理和频数分布展现 (7)

1.4 计量型汇总 (17)

与本章有关的 EXCEL 功能的实现途径 (18)

习题 (19)

案例研究题 (24)

第 2 章 数据分布趋势量数 (31)

2.1 众数与异众比率 (31)

2.2 五大位次量数 (33)

2.3 均值与方差 (34)

2.4 偏态系数与峰度系数 (37)

2.5 探索性工具 (39)

2.6 散点图和协方差 (43)

2.7 本书的知识结构 (46)

与本章有关的 EXCEL 功能的实现途径 (47)

习题 (48)

案例研究题 (50)

第 3 章 数据的理论分布模型 (52)

3.1 数据理论分布的概念和基本类型 (52)

3.2 概率分布参数:期望值与方差 (55)

3.3 几种简单的理论分布模型 (57)

3.4 二项分布 (59)

3.5 超几何分布 (60)

3.6 正态分布 (61)

3.7 χ^2 分布、F 分布和 t 分布	(63)
与本章有关的 EXCEL 功能的实现途径	(67)
习题	(68)
案例研究题	(70)

第二篇 围绕一个变量的统计推断

第 4 章 单一总体定标数据集中趋势量数的估计	(73)
4.1 参数与估计量	(73)
4.2 样本均值的抽样分布	(75)
4.3 区间估计的基本原理	(76)
4.4 σ^2 已知时 μ 的区间估计	(77)
4.5 σ^2 未知时 μ 的区间估计	(78)
4.6 必要样本容量的确定	(79)
与本章有关的 EXCEL 功能的实现途径	(79)
习题	(80)
案例研究题	(82)
第 5 章 单一总体定标数据集中趋势量数的假设检验	(84)
5.1 假设检验的基本思路	(84)
5.2 假设检验的基本概念	(86)
5.3 总体标准差明确统计下均值的假设检验	(89)
5.4 总体标准差不明确统计下均值的假设检验	(90)
5.5 P 值检验法	(91)
5.6 β 值的计算和必要样本容量的确定	(93)
与本章有关的 EXCEL 功能的实现途径	(94)
习题	(94)
案例研究题	(98)
第 6 章 单一总体定标数据离散趋势量数的统计推断	(99)
6.1 方差研究的重要性和方差的抽样分布	(99)
6.2 σ^2 的区间估计	(101)
6.3 σ^2 的假设检验	(101)
6.4 σ^2 的统计推断中的大样本正态近似法	(103)
习题	(105)
案例研究题	(106)
第 7 章 两总体定标数据参数间差异的统计推断	(107)

7.1	两总体均值差量的统计推断(独立样本)	(108)
7.2	两总体均值差量的统计推断(配对样本)	(113)
7.3	两总体方差差异的统计推断	(115)
	与本章有关的 EXCEL 功能的实现途径	(117)
	习题	(117)
	案例研究题	(121)
第 8 章	多总体定标数据集中趋势量数间差异的统计推断	(122)
8.1	方差分析的概念和作用	(122)
8.2	单因素方差分析	(123)
8.3	双因素方差分析	(131)
	与本章有关的 EXCEL 功能的实现途径	(138)
	习题	(138)
	案例研究题	(142)
第 9 章	定名数据的统计推断	(144)
9.1	引言	(144)
9.2	单一总体比率的统计推断	(145)
9.3	两总体比率间差异的统计推断	(147)
9.4	单一总体多值分类比率的假设检验	(151)
9.5	多总体多值分类齐一性的假设检验(独立样本)	(153)
9.6	多总体双值分类齐一性的假设检验(关联样本)	(156)
9.7	拟合优度检验	(158)
9.8	数据的随机性检验(游程检验)	(163)
	习题	(164)
	案例研究题	(167)
第 10 章	定秩数据的统计推断	(169)
10.1	非参数统计	(169)
10.2	单一总体位势检验——符号检验	(170)
10.3	两总体位势差异性检验(配对样本)——符号检验	(173)
10.4	两总体位势差异性检验(配对样本)——符号秩检验	(174)
10.5	两总体位势差异性检验(独立样本)——符号秩和检验	(177)
10.6	多总体位势差异性检验(独立样本)——克鲁斯卡尔— 华利思一向方差分析	(179)
10.7	多总体位势差异性检验(配对样本)——弗里德曼二向 方差分析	(180)

习题	(182)
案例研究题	(186)

第三篇 变量间联系的描述和推断

第 11 章 关联分析	(191)
11.1 定标数据的关联分析	(192)
11.2 定秩数据的关联分析	(195)
11.3 定名数据的关联分析	(200)
与本章有关的 EXCEL 功能的实现途径	(202)
习题	(203)
案例研究题	(207)
第 12 章 回归分析	(208)
12.1 线性回归的总体模型	(208)
12.2 线性回归的统计描述	(210)
12.3 线性回归的统计推断	(216)
12.4 回归模型的扩展	(224)
12.5 回归建模的若干问题	(234)
与本章有关的 EXCEL 功能的实现途径	(239)
习题	(240)
案例研究题	(246)

统计数表

附表 1 二项分布累积概率表	(248)
附表 2 标准正态分布表	(251)
附表 3 卡方(χ^2)分布表	(253)
附表 4 F 分布表	(254)
附表 5 t 分布表	(262)
附表 6 塔基的学生氏化极差 q 值表	(263)
附表 7 杜宾-沃岑检验上下界(DW 统计量临界值)表	(264)
附表 8 柯尔莫哥洛夫检验临界值表	(266)

第一篇

统计推断的预备知识

第1章 数据级次和数据整理

本章的教学目标

认识抽样调查和统计推断在管理统计中的重要作用,明了数据级次及其与传统的分类方法的关系,掌握对各种数据进行分组和汇总的方法,并能够构建统计表和统计图来展现数据分布特征。

1.1 抽样调查和统计推断

1.1.1 统计总体和统计描述

管理,无论是宏观管理还是微观管理,无论是企业管理还是公共管理,都需要借助于数据对所管理的对象的存在方式和变化趋势做出认识、判断和推测。管理统计就是使用统计手段对管理所需数据进行采集和加工,以导出对管理决策有支持作用的信息的一整套科学方法。管理者面对的是管理对象的全体,如某一级行政区政府面对的是本行政区的全体居民,企业营销策划部面对的是本企业产品现实和潜在用户全体。因此,其必须掌握能够概括管理对象全体的数据特征的信息,方能做出科学决策。

就一个管理课题而言,承担其研究内容的全部事物或现象的集合,即构成统计总体,简称总体(population)。而该集合中的每一个单元,就称为总体单位。含有无数多个总体单位的总体称为无限总体,否则即是有限总体。总体单位在各方面的属性称为标志,如企业的行业、规模、所有制类别、赢利状态等等。对某个统计总体单位标志的测量获得标志值。总体单位的标志值集合在一起,就构成数据(data)。当对数据进行定量分析时,该属性标志即称为变量(variable),而构成数据的每个标志值,即是变量值。

对于总体的研究,按所接触的全体单位的全面性与否,存在两种组织方式。一种是全面调查,即采集全部总体单位的数据,直接认识总体。政府所组织的大型国情、国力的调查,涵盖和接触到所研究现象的全部总体单位,如人口普查、农业普查

和经济普查等。另一种是非全面调查,即采集部分总体单位的数据,通过部分来认识总体。开展非全面调查时,部分总体单位的集合,即构成样本(sample)。从总体中选取这部分单位的过程称为抽样。如果遵循随机原则从总体中抽取样本,就是随机抽样(random sampling),否则就是非随机抽样。

无论是全面调查还是非全面调查,在从所接触的的总体单位采集到数据后,都需要对原始数据进行整理,制成统计表和统计图,并对数据本身的分布特征做概略分析。以上这个过程称为统计描述(statistical description)。如果进行的是全面调查,统计描述针对的就是总体,到此即形成了一个完整的认识过程。如果进行的是非全面调查,统计描述并非针对总体做出,如何认识总体的分布特征,则需要进一步考虑。

1.1.2 随机抽样和统计推断

随机原则根据概率原理,按一定法则和程式从总体中抽取一部分总体单位。在每一个总体单位被抽出来之前,总体中所有单位都有同等被抽取的机会。由于随机抽样的操作特点,使得样本对于总体有充分的代表性,样本数据的分布特征与总体的分布特征之间有规律可循,就可以从样本数据的概略分析结果来推测和判定总体分布特征。这个过程称为统计推断(statistical inference)。

总体按其所含单位是否有可数性确定数目,分为无限总体和有限总体,前者含有无限多的单位数目,后者所含单位数目即便众多,但终归有限。

随机抽样调查的必要性和重要性在于:

- (1)对于无限总体,不可能进行全面调查,只能做非全面调查;
- (2)对于有限总体,在理论上全面调查行得通。政府旨在摸清国情国力的调查,可以采用普查方式。除此之外,一般还是做非全面调查,原因如下:其一,全面调查费用巨大,与决策的潜在收益相比,不划算;其二,全面调查耗时较长,中短期决策等不及;其三,耗竭性数据采集调查(如产品检验)经全面调查使得产品全部报废。
- (3)无论是全面调查还是非全面调查,都会遭遇作业误差。全面调查涉及的单位众多,作业误差控制难度很大。开展非全面调查,可以较为严格地培训访员、取得被访者配合,因而大幅度减少作业误差。
- (4)随机抽样调查结果,根据抽样概率分布原理,可以用来推断总体特征。尽管这种推断存在抽样误差,但是可以将之控制在已知的一定范围内。非随机抽样调查结果尽管也具有认识总体数量特征的作用,但是不宜进行统计推断。
- (5)即便是政府普查,也只是隔固定间隔年头进行一次,普查年份之间数据的取得,还是要以原普查数据为基础,配以随机抽样调查数据进行推算。

综上所述,企业和非政府组织若需要直接采集管理数据,往往采用随机抽样调查的方式。政府的统计调查以周期性普查为基础,以经常性抽样调查为主体,以其他手段为补充。

1.2 数据级次和分类

数据是对所研究的事物、现象的属性的定量具体描述。从不同的角度看问题,数据就有不同的分类方法。管理统计将数据按其所蕴涵的内容深度分为3个等级:定名数据、定秩数据和定标数据。

【例 1-1】 设想某一天你收到的关于空调市场调查的问卷中含有下列4个问题:

一、您家主卧室的空调品牌是

1. 品牌 A 2. 品牌 B 3. 品牌 C 4. 其他

二、您对该品牌空调售后服务的评价是

1. 非常满意 2. 满意 3. 一般 4. 不满意 5. 非常不满意

三、在夏天最热的时间里,您家主卧室的空调夜间设定的温度是_____℃。

四、您购买主卧室空调的实际支出金额是人民币_____元。

您在回答问题时考虑的是问题答案的实质,诸如空调是什么品牌的,花多少钱买的,对售后服务满意不满意,再就是空调夜间设定的温度数。你不会关心这些问题的答案在形式上有什么区别,调查组织者面对采集到的这4个问题的答案数据所采用的数据整理和分析方法有什么不同。但是,从管理统计的角度看,这几个问题的答案恰是不同层级数据的代表。

1.2.1 定名数据、定秩数据和定标数据

(1) 定名数据

定名数据(nominal data)仅对事物、现象所属类别的具体描述。它们不表示顺序,没有多少大小之分。

关于空调品牌的答案属于定名数据。为了便于进行数据处理,需要给每个具体答案赋予一个数字代码,例如该问卷中以“1”代表“品牌 A”,“2”代表“品牌 B”,“3”代表“品牌 C”,“4”代表除以上3种列名品牌之外的其他品牌。当然,也可以交换代码的含义,这丝毫不影响对数据的研究,因为它们仅仅是代码类别而已。

对于性别,可以用“1”代表“男”,“0”代表“女”,同样也可以交换代码的含义,或者用“1”和“2”作代码。关于行业和职业的代码,从理论上说完全可以自由赋予代码,但是在实践上国家已经规定了统一的行业代码和职业代码,一般就不要“另起

炉灶”了。

(2) 定秩数据

定秩数据(ordinal data)是对事物、现象所具有的属性顺序的具体描述。

关于对空调售后服务评价的答案属于定秩数据。为了便于进行数据处理,需要给序列中每个级次赋予一个数字代码,例如本问卷中以“1”代表“非常满意”,“2”代表“满意”,“3”代表“无所谓”,“4”代表“不满意”,“5”代表“非常不满意”。既然为有序排列,就不能任意交换级次的位置。此外,也不能说某两个相邻级次之间的差异与另外两个相邻级次之间的差异必然相等。

(3) 定标数据

定标数据(metric data)是对事物、现象所具有的属性数量规模的具体描述。

关于空调夜间设定的温度数的答案和购买空调的支出金额的答案统归为定标数据。它们分别精确地测度空调设定的温度和购买空调的支出金额。

如果深究起来,温度数和支出金额的测度层级还有细微差别。支出金额的 0 表示没有支出。由于有绝对意义上的零存在,不仅可以说 4000 元与 2000 元的差额,和 2000 元与 0 的差额相等,还可以说 4000 元是 2000 元的 2 倍。温度计的“0”点只是一个参照基准数值,并不代表“没有”。可以说从 0°C 到 40°C 的差距相对于从 0°C 到 20°C 的差距的 2 倍,但是不可说 40°C 温度是 20°C 温度的 2 倍。由于没有绝对意义上的零存在的数据很少见,以上所说的细微差别在以后的数据采集、整理和分析中可以忽视不问。

定名数据、定秩数据、定标数据依次强化对事物属性的测度。定名数据仅区分事物的类别,定秩数据能够区分事物属性程度的差异,定标数据以等距间隔测定事物属性的差异。

1.2.2 质别数据和量别数据

传统的统计学概念将数据区分为质别数据(qualitative data)和量别数据(quantitative data)。凡是从定性角度考察事物、现象属性所获数据称为质别数据,它与定名数据和定秩数据相对应。凡是从定量角度考察事物、现象属性所获数据称为量别数据,它与定标数据相对应。

1.2.3 离散型数据和连续型数据

数据除按测度层次划分外,还可以按其数值的离散与否划分。凡其数值呈现为有限个整数的,是为离散型数据(discrete data)。凡其数值可以无限细分成小数的,是为连续型数据(continuous data)。离散型数据在数轴上体现为有限个整数点,互相不接连,如例 1-1 中的对空调售后服务评价的答案 1,2,3,4,5;又如某路