



商业分析

Business Analytics

商务统计学

孙 蕾◎编著



华东师范大学出版社



商务统计学

孙 蕾◎编著

商业分析

Business Analytics

图书在版编目(CIP)数据

商务统计学/孙蕾编著. —上海:华东师范大学出版社,
2018

(商业分析丛书)

ISBN 978-7-5675-7494-6

I. ①商… II. ①孙… III. ①商业统计学
IV. ①F712.3

中国版本图书馆CIP数据核字(2018)第031535号

商务统计学

编 著 孙 蕾
策划组稿 孙小帆
项目编辑 孙小帆
特约审读 石 岩
版式设计 卢晓红
封面设计 俞 越

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 13.75
字 数 254千字
版 次 2018年3月第1版
印 次 2018年3月第1次
书 号 ISBN 978-7-5675-7494-6/F·409
定 价 41.00元

出版人 王 熠

(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)

丛书简介

随着信息存储和传输技术的发展，特别是在移动计算、云计算、物联网等技术的支持下，社交网络、基于位置的服务、协同服务等新型服务和信息应用模式持续拓展着人们创造和分享信息的范围和形式，数据以前所未有的速度不断地增长和积累，大数据时代已经到来。

在这个时代，数据作为最重要的生产资源，成为改变世界、推动全球经济发展的支柱。面对日益成熟的商业大数据环境，如何有效地分析和利用数据成为急需解决的基础性问题。旺盛的社会需求催生出一个新的领域——商业分析（Business Analytics）。

商业分析是在大数据时代面向现实商业环境，以系统的观点为指导，运用定性与定量相结合的方法以及相关学科的研究手段，深入探究大数据环境下数据管理、数据分析、智能决策等问题的学科领域。为推动该领域的进一步发展，特别是商业分析人才的培养，我们特组织高校相关专业教师和企业一线业务人员编写了这套商业分析丛书。

本书简介

目标

在大数据时代,统计学作为收集、分析、表述和解释数据的方法论科学,在经济、管理、法律、文学等社会科学领域,以及物理、化学、生物、医学等自然科学领域,都有着广泛和深入的应用。

本书在全面借鉴国内外商务统计学教材的基础上,立足于增强统计学在经济、管理中的实际应用性,结合案例分析和实践教学,帮助读者学以致用,培养其思维、操作和创新能力。既可作为高校经济、管理类专业的研究生教材,又可作为 MBA、EMBA 和相关培训的教材,同时也是各种层次和类型经营管理人员的理想参考读物。

内容组织

本书分为基础编、描述统计编、概率统计编和应用编。

基础编是第 1 章的内容,主要介绍了统计的概念、数据的分类和收集方法、总体和样本的概念,以及描述统计和统计推断分别包括哪些内容。

描述统计编包括第 2~3 章的内容。第 2 章介绍了数据的图表展示方法。针对不同类型的数据,如定类数据、定序数据和数值型数据,有不同的图表展示方法。两变量数据的图表展示可以采用交叉表和散点图。第 3 章分别介绍了描述数据分布的集中趋势、离散程度、偏斜程度和尖峭程度的各个指标,通过这些指标能够全面了解数据分布的特点。

概率统计编包括第 4~7 章的内容。第 4 章介绍了一些初步的概率知识,说明了如何使用概率为决策提供帮助,如何使用数量方法来解释概率。若能获得额外附加信息,还可以使用贝叶斯公式来求后验概率。第 5 章介绍了离散型和连续型两种随机变量,一些常见的概率分布,如 0—1 分布、二项分布、超几何分布、泊松分布、均匀分布、指数分布和正态分布,并由正态分布导出 χ^2 分布、 F 分布和 t 分布三个重要分布,还详细阐述了大数定理、中心极限定理以及抽样分布的相关内容。第 6 章介绍了参数估计的概念及具体的估计方法,包括点估计和区间估计。无偏性、有效性

和一致性是点估计优良性的评价标准。对于区间估计,介绍了单个总体的总体均值、总体比例和总体方差的区间估计,以及两个总体均值之差、比例之差的区间估计。第7章介绍了假设检验的基本概念、原理和主要步骤。参数的假设检验包括一个总体均值、比例、方差的假设检验,以及两个总体均值的比较、比例的比较和方差的比较的假设检验。

应用编是第8章的内容,主要介绍了相关分析和回归分析的区别和联系、样本相关系数的计算和检验、简单线性回归模型的估计和检验方法、判定系数的概念、残差分析、点预测和区间预测等内容。

体例特点

本书对复杂统计模型的讲授主要侧重于应用,尽量降低难度,淡化理论模型的推导,使读者真正有所收获。每章开头均由有趣的统计故事引入,厘清统计方法和思路,演示如何使用统计软件进行实际操作和分析,提高读者的动手能力。每章结尾部分均有本章小结,帮助读者掌握重要概念、方法和内容。

目录

本书简介 1

第一编 基础编 1

第1章 总论 3

- 1.1 统计 4
- 1.2 统计在商务领域的应用 4
- 1.3 数据 5
- 1.4 数据的收集 8
- 1.5 描述统计和统计推断 10
- 本章小结 13

第二编 描述统计编 15

第2章 数据的图表展示 17

- 2.1 定性数据的图表展示 17
- 2.2 数值型数据的图表展示 24
- 2.3 交叉表和散点图 31
- 本章小结 37

第3章 描述性统计量 39

- 3.1 集中趋势 40
- 3.2 离散程度 50
- 3.3 偏度与峰度 57



3.4 应用 SPSS 软件进行描述统计分析 59

本章小结 61

第三编 概率统计编 63

第4章 概率 65

4.1 概率、事件和样本空间 66

4.2 事件的并和交 68

4.3 互补事件 69

4.4 加法法则和互斥事件 69

4.5 条件概率 71

4.6 独立事件和乘法法则 73

4.7 贝叶斯定理 75

本章小结 77

第5章 随机变量及其分布 79

5.1 随机变量 80

5.2 概率分布 80

5.3 几种常见的离散型随机变量的概率分布 84

5.4 几种常见的连续型随机变量的概率分布 89

5.5 由正态分布导出的三个重要分布 97

5.6 大数定理和中心极限定理 100

5.7 抽样和抽样分布 103

本章小结 110

第6章 参数估计 112

6.1 点估计 112

6.2 区间估计 115

6.3 总体方差的区间估计 124

6.4 样本容量的确定 126

6.5 两个正态总体均值之差的区间估计 128

6.6 大样本情况下两个总体比例之差的区间估计 132

本章小结 134

第7章 假设检验 136

- 7.1 总体均值的假设检验 136
- 7.2 总体比例的假设检验 145
- 7.3 总体方差的假设检验 148
- 7.4 两个总体均值的比较 151
- 7.5 两个总体比例的比较 159
- 7.6 两个正态总体方差比的假设检验 161

本章小结 161

第四编 应用编 167

第8章 简单相关与回归分析 169

- 8.1 相关分析 170
- 8.2 简单线性回归模型 176
- 8.3 回归模型的估计 178
- 8.4 判定系数 182
- 8.5 显著性检验 184
- 8.6 残差分析 187
- 8.7 预测 190
- 8.8 案例分析 193

本章小结 199

附录 200

- 附表1 标准正态分布表 200
- 附表2 t 分布表 202
- 附表3 χ^2 分布表 204
- 附表4 F 分布表 208

主要参考文献 209



第一编
基础编

第 1 章 总 论

统计故事

弗罗伦斯·南丁格尔(1820—1910),英国护士和统计学家。她出生于名门富有之家,家境优裕。可她却顶住流言蜚语,从事当时被人瞧不起的护士这一职业。克里米亚战争之初,英国的战地战士死亡率高达 42%。南丁格尔主动申请,自愿担任战地护士。她率领 38 名护士抵达前线,在战地医院服务。她竭尽全力排除各种困难,仅仅半年左右的时间伤病员的死亡率就下降到 2.2%。每个夜晚,她都提灯巡视,伤病员们亲切地称她为“提灯女神”。

1859 年,南丁格尔被选为英国皇家统计学会的第一个女成员。在 1863 年,英国的疾病命名与分类混淆不清,各地医院各自为政。南丁格尔制定了医疗统计标准模式,英国各医院相继采用,被公认是了不起的贡献。她发展出一种色彩缤纷的图表形式,让人们对数据的印象更加深刻,这种图表叫做南丁格尔玫瑰图,又被南丁格尔自己称为鸡冠花图(coxcomb diagram),是一种圆形的直方图,用来描述野战医院内,病人死亡率在不同季节的变化,同时向不会阅读统计报告的国会议员,说明克里米亚战争的医疗条件。

后来,南丁格尔对于印度的农村生活,做了全面的卫生统计研究,并在印度改善医疗和公共卫生服务。她成功地游说成立了一个皇家委员会,研究印度的情况。在 1873 年,南丁格尔报道,“卫生改革 10 年后,印度士兵的死亡率已经从每千名 69 人,降低到每千名 18 人”。

南丁格尔是护理战线上的白衣天使,她以强烈的事业心和高度的责任感,把真诚的爱心无私奉献给了每一位患者,燃烧自己,照亮别人,同时也为统计学的发展做出了不可磨灭的贡献。为了纪念这位近代护理事业的创始人,将她的生日 5 月 12 日定为国际护士节。

1.1 统计

“统计”(statistics)这个词是我们生活和工作中经常听到的一个词。例如,一个班级要统计男女生人数、要统计期末考试成绩;证券交易所要统计每日的交易量、交易额;商店要统计每日的销售额等等。但学科意义上的“统计”含义远非如此简单,确切地说,统计是对客观现象数量方面进行调查研究,对资料进行收集、整理,以及对所收集、整理的资料进行理论分析、推断等一系列活动的总称。

统计包含统计实践活动、统计资料和统计学三个方面。统计实践活动是指搜集、整理和分析客观事物总体数量方面资料的工作过程。统计资料是在统计实践活动中所取得的各项数字资料及有关文字资料,一般反映在统计表、统计图、统计手册、统计年鉴、统计资料汇编和统计分析报告中。统计学是指研究如何搜集、整理和分析统计资料的理论与方法。这三者间有密切的关系:统计资料既是在统计工作中对客观现象进行分析和研究的基础,又是对客观现象进行研究工作所得的成果;而统计学则是对客观现象统计资料的数量特征进行定量描述和分析推断的方法和科学。

统计学成为重要的学科,始于17世纪中期,由实践上升为理论,更进一步指导实践。经历了300余年的发展,从一开始在自然科学中的研究和应用,到不断进入社会经济管理、心理学、医学等多个学科的应用中,并与各个专业结合,逐渐形成了诸如人口统计学、经济统计学、卫生统计学、医学统计学、生物统计学等多个学科分支。而计算机的迅速发展,更使得统计学在各个领域中被广泛应用,使用方法也从初级向高级转变。

1.2 统计在商务领域的应用

统计学被广泛应用于经济、商业管理等许多方面,诸如统计报表的分析、财务预测、投资风险的预测、市场预测、销售预测、质量管理的评估和投资效益的分析与决策等等。让我们看几个统计在商务领域应用的具体例子:

一家公司在准备投入新产品时,通常需要估计相关市场上的消费者偏好。而这可以通过对随机抽取的若干家庭进行市场调查,用调查的结果作为对总体消费者偏好的估计。

一家会计事务所要确定某一客户的资产负债表中所示的应收账款余额是否公允地反映了其真实的应收账款余额。鉴于应收账款的数目往往较大,逐一检查和判断账户会浪费时间。通常

做法是取一个样本账户,审查这个样本账户的正确性后,审计人员就可以得出相关结论。

就股票投资来说,分析师们仔细检查包括市盈率和红利在内的一系列金融数据,通过比较某只个股和股票市场的平均数,大致就可判断该股票的价值是被高估还是低估。

经济学家在对通货膨胀率进行预测时,用生产者价格指数、失业率和制造业设备利用率等方面的统计信息,将这些统计指标输入计算机化的预测模型中来预测通货膨胀率。

随着社会经济的发展,广泛而深入的市场研究和理论探索长期持续进行,海量数据席卷着各行各业,管理方法日新月异但更加依赖于统计分析,客户满意度测评及客户关系管理如火如荼,这些应用于社会经济领域的理念,无一不与统计学有着极大的关联。

1.3 数据

要进行统计分析,离不开统计数据。统计数据是对客观现象的总体进行计量的结果。在某个特定的研究中搜集的所有数据,称为数据集。表 1-1 是某项调查提供的 13 个楼盘的均价。

表 1-1 13 个楼盘均价

楼盘	均价(元/平)	楼盘	均价(元/平)
御珑湾	13 000	绿地海珀	45 000
润华庭	38 000	江湾九里	43 000
嘉誉湾	43 000	尚浦领世	42 000
建珑庭	38 000	香景园	40 000
仓玺园	60 000	七星府邸	35 000
公馆 1088	33 000	东外滩	38 000
御江湾	37 000		

1.3.1 元素、变量和观察值

元素(elements)是收集数据所依赖的实体。对表 1-1 中的数据集来说,每个楼盘就是一个元素,13 个楼盘就是 13 个元素。

变量(variable)是元素相关的特征。表 1-1 只有 1 个变量,就是楼盘均价。根据不同研究目的搜集的数据集不一样,所含的变量个数也不一样。

数据通过收集每个元素的各个变量的测度值而得到。为某个元素而收集的测度值的集合称为观察值(observation)。表 1-1 中,可以看到一共有 13 个观察值,第 1 个元素(御珑湾)的观察值为 13 000。

1.3.2 测量尺度

数据可以根据测量尺度进行分类。资料的测量尺度通常规定了计算及呈现资料的方式。测量的尺度可以分为名义尺度、顺序尺度、等距尺度和比例尺度。

(1) 名义尺度

名义尺度(nominal scale)按照事物总体的某种属性对其进行平等的分类、分组、计数和汇总。例如,对人口按照性别划分为男性和女性两类,按省份划分为江苏省、广东省等多个分类。每一类型虽然有特定的文字甚至数量标识,但这种数量标识只是平等的代号或标签。例如,可用1表示男性,2表示女性,也可用1表示女性,2表示男性。这些数字只是给不同类别的一个代码,并不意味着可以区分大小或进行任何数学运算。定类测量形成各种类型,它们都具有同等的重要性,具备计数功能,即计算每一个类型的频数或频率。有时为了方便了解,我们会将计数转换成百分比。

(2) 顺序尺度

顺序尺度(ordinal scale)是对序列变量的测量。顺序尺度不仅具有名义尺度的特点,而且具有排序功能。各类型之间按照既定标志形成一种确定的顺序,但是这种排序不能具体测量各等级之间的间距大小。例如,每位同学都对某教师的教学做出评价,我们将评价的等级分为“非常好”、“良好”、“一般”、“差”和“非常差”,并且用数值5、4、3、2、1分别表示这5个等级。可以分辨“非常好”比“良好”等级高,“良好”又比“一般”等级高。但是我们不能将“非常好”对应的数值5加上“良好”对应的数值4,加总之后没有任何意义。也不能说“良好”(数值是4)比“差”(数值是2)好两倍,只能说“良好”比“差”要好,具体好多少无法计算。

(3) 等距尺度

等距尺度(interval scale)是比顺序尺度高一层次的定量测量。它不仅能将事物总体区分为不同类型、进行排序,还可以测量类型的间距大小。例如,上海市连续3天的最高温度分别为28摄氏度、25摄氏度和14摄氏度。这些温度值除了能够进行排序外,还可以计算不同温度间的差距,即等距尺度的量可以进行加和减的运算。但是它不能进行乘或除的运算,就如此例中,不能认为28摄氏度比14摄氏度暖和2倍,也不能认为“0”摄氏度代表“没有温度”。

【例 1-1】 表 1-2 是美国女性服装的尺码。服装尺码属于等距尺度。尺码由 8 到 10,和尺码由 20 到 22,皆是增加 2 个号,即差距是相同的。0 号的尺码,也不代表 0 英寸,而是代表最小号,即胸围 30 英寸、腰围 24 英寸和臀围 33 英寸。此外,在等距测量中,比例并不是合理的数值,即不能用尺码 28 除以尺码 14,简言之,若数字的差距有意义,而比例没有意义,那就属于等距尺度。

表 1-2 美国女性服装尺码表

尺码	胸围(英寸)	腰围(英寸)	臀围(英寸)
0	30	24	33
2	31	25	34
4	33	26	35
6	34	27	37
8	35	28	38
10	36	30	39
12	37	31	40
14	39	32	41
16	40	33	43
18	41	34	44
20	42	35	45
22	43	37	46
24	44	38	47
26	46	39	48
28	47	40	50

(1) 比例尺度

比例尺度(ratio scale)是比等距尺度更高一级的定量尺度,它不仅可以进行加减运算,还可以进行乘除运算。比例尺度同等距尺度的显著区别是它有绝对零点,即在数值序列中,零值是有实际意义的,零表示真实的“无”。例如,人的年龄、体重,都以零表示“无”。因此,我们既可以认为某甲 30 岁,比 10 岁的某乙年长 20 岁,也可以认为甲的年龄是乙的 3 倍。几乎所有的统计量都可以进行定比测量;绝大多数的经济变量也可以进行定比测量,如产值、工人数、居民可支配收入、银行存款余额等。

1.3.3 定性数据和定量数据

数据按照所采用的测量尺度进行分组,可以分为定类数据、定序数据、定距数据和定比数据。定类数据和定序数据统称为定性数据。定性数据是一组表示事物性质、规定事物类别的文字表述型数据,用于识别单个元素特征的标签或名称。定性数据用名义尺度或顺序尺度,可以是非数字或者数字,不能将其量化,只能将其定性。定距数据和定比数据统称为数值型数据,又称定量数据,是以数字形式表现出来的研究资料,用于表示多少或者多少个。

对于定性数据,可以通过计算每种类型定性数据的数量和比例来描述定性数据。即使定性

数据使用数字代码表示,加减乘除等算术运算也都没有任何意义。我们将会在第 2.1 节讨论定性数据的图表展示方法。对于定量数据,可供选择的统计分析方法较多,各数值可以加总起来求和,也可以除以数值的个数得到算术平均数。第 2.2 节和第 3 章将会介绍定量数据的描述方法。

1.4 数据的来源

1.4.1 数据的来源

统计数据主要来源于两个方面,一是通过统计调查的方式得到的第一手数据(primary data);二是搜集国家和地方的各级统计部门以及其他有关部门所公布的数据资料,这种资料称为二手数据(secondary data)。

一手数据即原始数据的收集并不轻松。它的获得要通过统计调查来实现。某些企业每年至少要花三四千万元来收集和分析数据。他们调查其产品目前在市场上的状况和地位;调查不同地区、不同性别、不同阶层的民众对其产品的认知程度和购买意愿等。常见的统计调查方式有以下几种:

① 普查

普查(census)是为了某种特定的目的而专门组织的一次性的全面调查。它一般是调查属于一定时点上的经济现象的总量,也可以调查某些时期现象的总量。例如 1954 年进行的黑色金属、有色金属和木材库存普查,2010 年进行的全国第六次人口普查等。普查涉及面广、指标多、工作量大、时间性强,并且需要由政府部门动员组织各方面力量配合才能进行。其优点是资料全面、系统、详尽。

② 定期统计报表

定期统计报表是收集统计资料的一种主要方式。它是按照国家有关规定,自上而下布置调查任务,自下而上逐级按照统一要求提供统计资料的一种报告制度。定期统计报表要求以一定的原始记录为依据,按照统一的指标、分组、报送时间和报送程序填报。

③ 抽样调查

抽样调查(sample survey)是从全部调查研究对象中,抽选一部分单位进行调查,并据以对全部调查研究对象作出估计和推断的一种调查方法。抽样调查是一种非全面性的调查,其分析的理论基础是概率论。这是一种节省人力、财力、物力而且具备一定的可靠性水平的科学方法,被