

普通高等教育“十二五”规划教材



21世纪经济与管理应用型本科规划教材

工商管理系列

统计学

基于SPSS的应用

Statistics

Application Based on SPSS

曹慧 编著



北京大学出版社
PEKING UNIVERSITY PRESS

C15061607

C8-43
203

博雅

21世纪经济与管理应用型本科规划教材
工商管理系列

统计学

基于SPSS的应用



Statistics

Application Based on SPSS

曹慧 编著

C8-43
203



北航

C1675440



北京大学出版社
PEKING UNIVERSITY PRESS

图书在版编目(CIP)数据

统计学: 基于 SPSS 的应用/曹慧编著. —北京: 北京大学出版社, 2013. 9

(21 世纪经济与管理应用型本科规划教材·工商管理系列)

ISBN 978 - 7 - 301 - 23152 - 4

I. ①统… II. ①曹… III. ①统计学 - 高等学校 - 教材 ②统计分析 - 软件包 - 高等学校 - 教材 IV. ①C8

中国版本图书馆 CIP 数据核字(2013)第 205558 号

书 名: 统计学——基于 SPSS 的应用

著作责任者: 曹 慧 编著

责任编辑: 刘誉阳

标准书号: ISBN 978 - 7 - 301 - 23152 - 4/F · 3735

出版发行: 北京大学出版社

地 址: 北京市海淀区成府路 205 号 100871

网 址: <http://www.pup.cn>

电子信箱: em@pup.cn QQ:552063295

新浪微博: @北京大学出版社 @北京大学出版社经管图书

电 话: 邮购部 62752015 发行部 62750672 编辑部 62752926 出版部 62754962

印 刷 者: 北京大学印刷厂

经 销 者: 新华书店

787 毫米 × 1092 毫米 16 开本 15.25 印张 390 千字

2013 年 9 月第 1 版 2013 年 9 月第 1 次印刷

印 数: 0001—3000 册

定 价: 32.00 元

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有,侵权必究

举报电话:010-62752024 电子信箱:fd@pup.pku.edu.cn

013067607

丛书出版前言

《国家中长期教育改革和发展规划纲要(2010—2020年)》指出,目前我国高等教育还不能完全适应国家经济社会发展的要求,学生适应社会和就业创业能力不强,创新型、实用型、复合型人才紧缺。所以,在此背景下,北京大学出版社响应教育部号召,在整合和优化课程、推进课程精品化与网络化的基础上,积极构建与实践接轨、与研究生教育接轨、与国际接轨的本科教材体系,特策划出版《21世纪经济与管理应用型本科规划教材》。

“21世纪经济与管理应用型本科规划教材”注重系统性与综合性,注重加强学生分析能力、人文素养及应用性技能的培养。本系列包含三类课程教材:通识课程教材,如《大学生创业指导》等,着重于提高学生的全面素质;基础课程教材,如《经济学原理》《管理学基础》等,着重于培养学生建立宽厚的学科知识基础;专业课程教材,如《组织行为学》《市场营销学》等,着重于培养学生扎实的学科专业知识以及动手能力和创新意识。

本系列教材在编写中注重增加相关内容以支持教师在课堂中使用先进的教学手段和多元化的教学方法,如用课堂讨论资料帮助教师进行启发式教学,增加案例及相关资料引发学生的学习兴趣等;并坚持用精品课程建设的标准来要求各门课程教材的编写,力求配套多元的教辅资料,如电子课件、习题答案和案例分析要点等。

为使本系列教材具有持续的生命力,我们每隔三年左右会对教材进行一次修订。我们欢迎所有使用本系列教材的师生给我们提出宝贵的意见和建议(我们的电子信箱是 em@pup.cn),您的关注就是我们不断进取的动力。

在此,感谢所有参与编写和为我们出谋划策提供帮助的专家学者,以及广大使用本系列教材的师生,希望本系列教材能够为我国高等院校经管专业的教育贡献绵薄之力。

北京大学出版社
经济与管理图书事业部
2012年1月

前 言

随着信息技术的突飞猛进,统计学迎来了大发展,各种统计分析工具日新月异,在各行各业中被广泛应用,社会迫切需要既掌握基本统计学知识,又能熟练操作统计分析工具的实用型人才。本书就是为了适应这一发展而撰写的一部本科生教材,旨在对传统的统计学理论和知识体系进行重新梳理、提炼,去粗取精,化繁为简,摒弃复杂的公式推导和手工计算,代之以SPSS统计软件的操作和应用,使学生在有限的课堂教学时间里既能了解统计学的思想精髓,又能掌握一种实用的统计分析工具解决实际问题,着重于培养学生的实际操作和实际应用能力。

本书在内容上按照一个完整的统计分析过程的五大环节——收集数据、整理数据、分析数据、解释数据、做出决策(或得出结论)来组织材料,共划分成十个章节,有利于学生对统计分析和解释数据有一个全面、科学、系统的认识;每章前设计了引导案例,每章后又有大型应用案例,每章内又有若干练习案例,有利于学生理论联系实际,强化学生对所学知识的理解和应用。同时,在数据和案例的选择上,坚持采用最新、最具时代特征的数据,案例全部来源于经济管理领域或学生日常生活非常熟悉和感兴趣的领域,具有鲜明的经管类专业特点,能更好地满足经管类专业学生所学所用的需要;在对SPSS统计软件的讲解中,以主流的SPSS 18.0中文版软件为例,详细介绍了软件的操作步骤和统计分析结果的意义,有利于教师教学和学生自学;在写作风格上,本书力图做到简洁明快、通俗易懂、生动有趣,以便激发学生的学习兴趣。总而言之,本书适合作为经管类专业本科生“统计学”的教学使用,也可以作为各类从事经济管理工作的专业人员的培训和自学教材。

本书由北京城市学院曹慧编著和主笔,并负责全书整体框架和编写大纲的设计以及全书的修改、总纂和定稿工作。邓凯元参与了第1章和第2章的编写,黄群参与了第5章、第6章的编写;王玉华参与了第8章的编写,杨德玲参与了第9章的编写;李彬参与了部分校对工作。

本书的编写过程参考了许多国内外教材、文献及数据,若本书列出的参考文献中有所遗漏,敬请谅解。由于作者的水平有限,加之时间仓促,书中的疏漏、错误之处在所难免,恳请广大教师和读者不吝赐教,提出宝贵意见。最后,本书的编写与出版得到了北京城市学院的大力支持,在此表示深深的感谢。

曹 慧

2013年6月

目 录

第 1 章 导论	1
1.1 统计学概述	2
1.1.1 什么是统计学	2
1.1.2 身边的统计学	5
1.2 统计学中的基本概念	6
1.2.1 总体和样本	6
1.2.2 参数和统计量	6
1.2.3 变量和数据	7
1.3 SPSS 统计分析软件	9
1.3.1 SPSS 简介	9
1.3.2 SPSS 的主要窗口	11
第 2 章 数据收集与录入	18
2.1 数据的来源	18
2.1.1 数据的间接来源	19
2.1.2 数据的直接来源	20
2.1.3 抽样方案的设计	21
2.2 问卷的设计	24
2.2.1 问卷调查与问卷	24
2.2.2 问卷的基本结构	25
2.2.3 问卷设计的注意事项	27
2.3 数据的录入	27
2.3.1 问卷的编码	27
2.3.2 在 SPSS 中录入数据	30
第 3 章 数据的整理与转换	37
3.1 数据的整理	38
3.1.1 个案排序	38
3.1.2 数据文件的转置	40
3.1.3 数据文件的合并	41

3.1.4	数据文件的分类汇总	45
3.1.5	数据文件的拆分	49
3.1.6	个案的选择	50
3.1.7	个案的加权	54
3.2	数据的转换	55
3.2.1	变量的重新编码	55
3.2.2	变量的运算	57
3.2.3	变量的计数	60
3.2.4	个案的排序	62
第4章	数据的概括与描述	67
4.1	集中趋势和离散趋势	68
4.1.1	集中趋势的测量	68
4.1.2	离散趋势的测量	72
4.1.3	分布形状的测量	76
4.2	定性数据的描述性统计	77
4.3	定量数据的描述性统计	83
4.3.1	基本的描述性统计	83
4.3.2	探索性分析	85
第5章	概率基础	95
5.1	什么是概率	95
5.1.1	几个基本概念	95
5.1.2	计算概率的三种方法	96
5.2	随机变量的概率分布	97
5.2.1	随机变量及其概括性度量	97
5.2.2	离散型随机变量的概率分布	99
5.2.3	连续型随机变量的概率分布	100
5.3	抽样分布	103
5.3.1	抽样分布的基本原理	103
5.3.2	小样本抽样分布	105
5.3.3	统计量的标准误差	107
第6章	参数检验	111
6.1	参数估计	112
6.1.1	什么是参数估计	112
6.1.2	点估计与区间估计	112
6.1.3	一个总体参数的区间估计	114
6.1.4	两个总体参数比较的区间估计	116
6.2	假设检验	120

6.2.1	假设检验的基本原理	120
6.2.2	假设检验的类型	124
6.3	平均数比较与 T 检验	125
6.3.1	平均数分析过程	125
6.3.2	单样本 T 检验	127
6.3.3	独立样本 T 检验	129
6.3.4	配对样本 T 检验	132
第 7 章	交叉表和关联分析	137
7.1	定性变量与交叉表	138
7.1.1	交叉表的形式	138
7.1.2	交叉表分析	138
7.2	两个变量的关联分析	143
7.2.1	交叉表的检验	143
7.2.2	等级相关分析	147
7.2.3	其他相关分析	150
第 8 章	方差分析	157
8.1	方差分析的原理	158
8.1.1	方差分析的概念	158
8.1.2	方差分析的基本原理	158
8.1.3	方差分析的基本假定	161
8.2	单因素方差分析	161
8.3	双因素方差分析	166
8.3.1	双因素方差分析的基本原理	166
8.3.2	双因素方差分析的操作	167
第 9 章	相关与回归分析	176
9.1	相关分析	177
9.1.1	相关关系	177
9.1.2	相关关系的描述与测量	178
9.1.3	线性相关分析	182
9.2	一元线性回归分析	186
9.2.1	回归分析	186
9.2.2	一元线性回归模型	187
9.2.3	一元线性回归分析	189
9.3	多元线性回归分析	196
第 10 章	非参数检验	207
10.1	卡方检验	208

10.1.1	非参数检验概述	208
10.1.2	卡方检验	209
10.2	二项分布检验	212
10.3	游程检验	215
10.4	单个样本分布特征的检验	217
10.5	两个独立样本的检验	219
10.6	多个独立样本的检验	222
10.7	两个相关样本的检验	225
10.8	多个相关样本的检验	227
参考文献		234

第1章 导 论

学习目标

- 理解统计的含义。
- 理解统计学的研究对象和研究内容。
- 了解统计学在经济和管理领域中的应用。
- 掌握统计学中常用的基本概念。
- 了解 SPSS 统计分析软件的主要窗口。

引导案例

身边的统计学: 淘宝网趣味数据分析

淘宝数据盛典公布了有趣的 2011 年度淘宝网购趣味数据。数据清晰地描述了淘宝网顾客的消费习惯,并以地图的形式展现了中国不同地区消费者的消费偏好和特色,对于预测来年网购潮流趋势也有很好的借鉴意义,如图 1-1 所示。这些趣味数据显示:北京人最爱用安



图 1-1 淘宝中国网购地图

卓,购买此类产品的人最多;台湾人最爱摄影器材,购买单反相机的人均花费达6310元;江苏人最文艺,全年买书花了2.6亿;山西人最孝顺,30岁以下年轻人在淘宝网购买老年人用品的人数最多;浙江女性最关爱男性,浙江是女性购买男性产品最多的省份;安徽人最爱宠物,每1万个人中有16人购买过狗粮……统计就在我们身边,无时无刻不在影响着我们的生活。(资料来源:淘宝数据中心)

1.1 统计学概述

1.1.1 什么是统计学

1. 统计的含义

统计作为一种社会实践活动,历史悠久,可以追溯到古老的原始社会。而“统计学”作为一门系统的科学,却只有近300多年的历史。汉语中的“统计”就是合计或汇总计算。英语中的“统计”最早出自于拉丁语的“status”,指的是各种现象的状态和状况,后在英语中演变为“statistics”,指的是计数活动。今天“统计”一词已被人们赋予多种含义,在不同的场合、不同的语言环境中有多种不同的解释。总体来看,“统计”主要有三种含义:统计工作、统计资料和统计学。

(1) 统计工作

统计工作即统计实践,是收集、整理和分析统计数据,探索数据内在数量规律性的活动过程。例如,为了掌握一个国家的经济社会资本,可以进行人口、自然资源和财富的统计;为了保护人类生存的自然环境,可以进行环境质量的统计。统计工作的好坏直接影响统计资料的数量和质量。统计工作的发展又为统计学的产生和发展奠定了基础。

(2) 统计资料

统计资料即统计工作的成果,是反映客观现象数量方面的统计数据和分析报告。例如,国内生产总值(GDP),反映一个国家的生产规模;消费者价格指数(CPI),反映一定时期价格水平的变动程度;外商直接投资(FDI),反映一国利用外国直接投资的规模。这些耳熟能详的数据就是统计资料。随着信息技术的发展与网络的普及,大量的电子版的统计资料还可以方便地从各国官方统计网站上免费获得。

(3) 统计学

统计学是一门收集、处理、分析、解释数据并从数据中得出结论的科学,其目的是探索数据的内在数量规律性,以达到对客观事物的科学认识。统计学源于统计工作,是统计工作经验的理论概括;理论和方法对统计工作的指导,又促使统计工作的水平越来越高,应用范围越来越广;统计工作的进一步发展,再次丰富、推动着统计学的发展。两者相互促进,相互影响,是理论与实践的关系。

统计工作、统计资料和统计学三者之间密不可分,习惯上把它们通称为“统计”。

2. 统计学的研究对象

统计学是从研究社会经济现象的数量开始的,所以,统计学的研究对象就是大量现象的数量方面,即数据。从统计学的定义也可以看出,统计学的研究对象包含两个方面:第一,收集数据;第二,分析数据。收集数据的目的是为了了解某一理论或应用上的问题。但是仅有一堆杂

乱无章的数据,是不能解决任何问题的。数据必须经过整理,从中发掘有用的信息并以适当的形式表达出来,再用科学的方法进行分析,针对问题得出一个结论或做出一项决策。下面就是一个典型的例子。

例 1-1

一个金融产品的预期收益率的变化通常可以用来衡量其投资风险。预期收益率变化越小,投资风险越低;预期收益率变化越大,投资风险越高。从市场上分别抽取 40 只商业类股票和 40 只地产类股票,根据上市公司上年派发的现金股利与本期股票价格可以计算出预期收益率。根据 40 种商业类股票的预期收益率和 40 种地产类股票的预期收益率可以画出下面两个直方图,如图 1-2 所示。如果你是投资者,你会选择投资于哪一类股票?

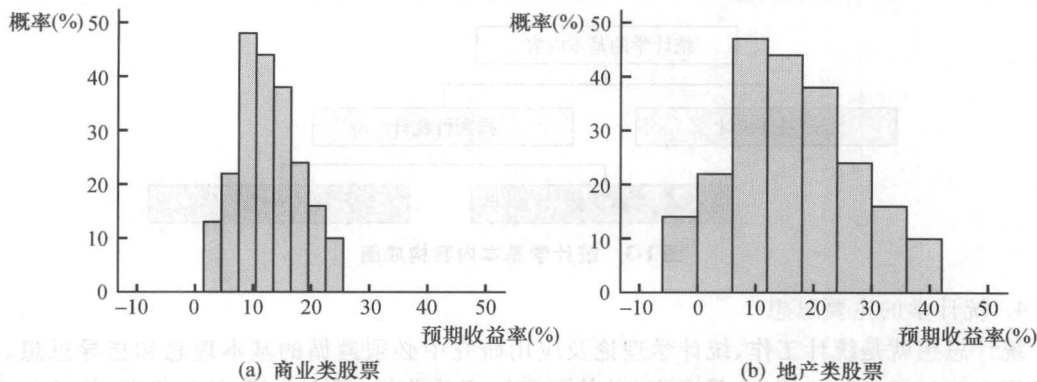


图 1-2 两类股票的预期收益率直方图

从图中可以清楚地看出:商业类股票的预期收益率波动范围较小,大约在 3% 至 24% 之间;而地产类股票的预期收益率波动范围较大,大约在 -5% 至 42% 之间。显然,商业类股票的收益较低,但其风险也较小;而地产类股票的收益较高,但其风险也较大。

通过上面的例子,我们可以归纳总结出一个完整的统计过程所包括的五大环节:

(1) 收集数据:收集 40 只商业类股票和 40 只地产类股票的上一年派发的现金股利与本期股票价格的数据。

(2) 整理数据:计算 40 只商业类股票的预期收益率和 40 只地产类股票的预期收益率。

(3) 分析数据:用图表的形式展示数据,本例是通过直方图来比较两类股票收益率的差异。

(4) 解释数据:商业类股票的收益较低,但其风险也较小;而地产类股票的收益较高,但其风险也较大。

(5) 做出决策(或得出结论):若是稳健型投资者,可选择商业类股票;若是激进型投资者,可选择地产类股票。

3. 统计学的基本内容

随着统计学的不断发展,其研究内容越来越丰富,按照统计方法的构成来看,其内容可分为描述性统计和推断性统计两大类,如图 1-3 所示。

(1) 描述性统计

描述性统计(descriptive statistics)是说明现象的情况和特征的方法。具体来说,描述性统计就是研究如何收集反映现象的数据,并用图表的形式对数据进行加工处理和显示,再通过概

括和分析得出反映现象规律性数量特征的方法。所以,描述性统计包括统计数据收集的方法、数据加工整理的方法、数据显示的方法、数据分布特征的概括与分析的方法等。例 1-1 就是一个描述性统计的例子。

(2) 推断性统计

推断性统计(inferential statistics)是研究如何根据样本数据去推断总体参数的方法,它是在对样本数据进行描述的基础上,对总体参数做出以概率形式表述的推断。推断性统计又分为参数估计和假设检验两种类型。参数估计就是用样本统计量去估计总体参数;假设检验是先对总体参数提出一个假设值,然后根据样本信息去判断这一假设是否成立。例如,调查北京市大学全体学生的英语成绩,可以提出一个假设:北京市大学全体学生的英语平均成绩是 80 分,然后从全校学生中抽取一个样本,根据样本平均数来检验总体平均成绩是不是 80 分。

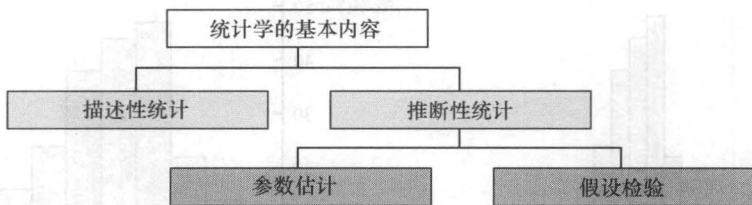


图 1-3 统计学基本内容构成图

4. 统计学的主要思想

统计思想就是统计工作、统计学理论及应用研究中必须遵循的基本理念和指导思想。统计思想主要包括定量分析、大量观察(总体思想)、平均分析、平衡分析、估计推断、相关分析等思想,如表 1-1 所示。

表 1-1 主要的统计学思想

统计思想	解释
定量分析思想	对于事物的认识和反映是通过对事物数量的搜集、整理和分析,用数量指标去说明事物发展的规模、水平、速度、趋势以及若干事物之间的联系。使人们更加准确地认识和把握事物发展的规模大小、水平高低、速度快慢、品质优劣,使人们对事物的认识由模糊变清晰,由粗略变精确。
大量观察思想 (总体思想)	观察研究对象总体中的全部或足够多的基本单位,以取得足够反映现象总体数量特征的统计资料,综合概括出反映现象本质特征和发展规律的统计指标。
平均分析思想	揭示和描述事物的主要指标是统计平均数。从统计理论与实践的发展过程来看,统计平均数可以说无处不在,如总体水平的反映、不同总体水平的比较、同一总体不同时间发展水平的比较、相关事物之间联系的程度、多种事物发展水平的综合变动情况,等等,都要使用平均指标。
平衡分析思想	最初的平衡思想是指收、支、存之间的数量对等关系。在现代统计学中,平衡思想是指相关事物存在和发展的数量联系、制约关系。例如,投入产出分析法是平衡分析思想的典型代表,在社会经济分析研究中发挥着重要作用。
相关分析思想	研究两个或两个以上的现象之间是否存在某种不确定的数量依存关系。对存在数量依存关系的,需要进一步确定其形式、程度和方向,对依存关系程度较高,需要再确定自变量和因变量,建立回归方程,并根据自变量数值推断因变量数值。
估计推断思想	根据已掌握的研究对象的规模、水平、发展变化规律,及其与其他相关现象之间的数量联系,去推断研究对象或相关现象的规模、水平及发展变化规律。

1.1.2 身边的统计学

在日常生活中,我们每个人都离不开统计,这是因为统计分析方法作为一种为决策提供依据的工具,广泛应用于所有领域,只要有数据的地方就会用到统计。今天,统计已经被应用到自然科学和社会科学的很多领域,统计学也发展成为由若干分支学科组成的学科体系,如社会统计学、经济统计学、管理统计学、生物统计学、体育统计学,等等。下面是几个典型的统计学在经济管理领域应用的例子。

(1) 统计学在经济领域的应用

例 1-2

出口、投资和消费是拉动经济增长的三驾马车。但出口和投资受国内外各种因素的影响较大,只有消费完全由本国居民自己控制,是最重要的马车。据统计,我国居民消费对 GDP 的贡献率在 50%—60% 左右,国外居民消费对 GDP 的贡献率在 70%—80% 左右,英美等国甚至高达 85% 以上。因此,解决经济困境的关键就是增加消费。一项根据 1978—2011 年北京城镇居民人均可支配收入和消费性支出的研究发现,北京城镇居民人均可支配收入每增加 1 元,消费性支出增加 0.679 元。所以,切实提高居民收入,藏富于民是发展内需经济,推动经济增长的必由之路。

例 1-3

上市公司的财务数据是股民投资的重要参考依据。一些投资公司根据上市公司提供的财务和统计数据进行分析,为股民提供参考。例如,市盈率是股票的价格和每股收益的比率,投资者通常会利用市盈率来评价股票的投资价值。2012 年 6 月 7 日上证 50 指数的平均市盈率是 9.97 倍,北京金隅股份有限公司的市盈率是 18.40 倍。这一统计信息说明,与上证指数股票的平均收入相比,北京金隅股份有限公司的股票价格较高,即北京金隅股份有限公司现行股价被高估了。投资顾问在综合分析公司其他有关的信息之后可以做出买入、卖出还是持股的建议。

(2) 统计学在管理领域的应用

例 1-4

企业要在激烈的市场竞争中取得优势,首先必须了解市场,要了解市场就需要进行广泛的市场调查,取得所需信息,并对这些信息进行统计分析,以便作为生产和营销的依据。例如,一家生产减肥茶的企业想要调查新研发的减肥茶到底有没有减肥效果,于是对 35 名肥胖志愿者进行了减肥茶试用的跟踪调查。先将 35 名肥胖志愿者喝减肥茶前的体重记录下来,三个月后再记录其喝茶后的体重,然后进行均值比较,发现新研发的减肥茶的确有比较明显的减肥作用,于是企业做出了大规模生产这一新型减肥茶的决策。

例 1-5

利用统计方法对员工的年龄、性别、教育程度、态度、行为等进行分析,可以作为企业设计

薪酬体系、福利政策的依据。例如,某企业调查了员工福利需求后发现,薪酬水平在1 000—3 000元的员工最重视的福利是免费工作餐,选择率达54.5%。薪酬水平在3 001—5 000元的员工最重视的是免费单身宿舍,选择率达63.2%。薪酬水平在5 001—7 000元及7 000元以上人群因为薪水较高,所以偏向于文体设施的福利,选择率分别达54.5%和93.3%。可见,薪酬水平的高低是影响员工福利需求的重要因素,员工在收入不足5 000元时主要考虑的还是食宿问题。因此,企业应该针对不同收入水平的员工,提供多样化的福利政策。

1.2 统计学中的基本概念

1.2.1 总体和样本

1. 总体与个体

总体(population)就是所要研究的事物或现象的全体,即统计调查的全部个体的集合。例如,想要了解北京市场上消费者对手机品牌的偏好,则北京市场上所有的手机消费者就是研究的总体;想要调查北京城市大学全体学生的英语成绩,那么各院系的所有学生就是研究总体。总体根据其包含的元素数目是否可数分为有限总体和无限总体。有限总体中的元素数目是有限可数的,如北京城市大学的所有学生就是有限总体;而无限总体中的元素数目是无限不可数的,如市场上所有可能购买手机的消费者就是无限总体。

组成总体的每一个元素称为个体(item unit),个体是构成总体的个别事物。例如在上例中,北京市场上每一个手机消费者就是个体;北京城市大学各院系的每一个学生就是个体。对于不同的研究对象,个体可能是人、物,也可能是企业、机构,甚至可能是时间、地域。

2. 样本与样本量

样本(sample)是从总体中抽取的一部分个体的集合。抽取样本的目的是用来推断总体,这就要求样本能够代表总体。样本代表总体的程度越高,由样本计算的指标与总体指标的误差就越小。因此,我们总是希望样本具有较高的代表性。遵循随机原则的抽样,能够排除主观因素的影响,保证取样的客观性;采用非随机原则的抽样,有时会更快捷、更经济,但是抽出的样本无法计算误差。

构成样本的个体的数目称为样本量(sample size)或样本容量。通常我们用 N 表示总体,用 n 表示样本量。相对于 N 而言, n 一般只是一个很小的数。例如,为了了解北京市场上消费者对手机品牌的偏好,从市场上随机访问50名手机消费者,则这50名消费者就构成了一个样本,样本容量是50;为了调查北京城市大学全体学生的英语成绩,从全校学生中随机抽取了300名学生,则这300名学生就构成了一个样本,样本容量是300。

1.2.2 参数和统计量

1. 参数

参数(parameter)也称为总体参数或总体指标,是用来描述总体的数量特征,如总体平均数、总体标准差、总体比例等。我们习惯用希腊字母来表示总体参数,如总体平均数用 μ 表示,总体标准差用 σ 表示,总体比例用 π 表示。参数是一个未知的常数,是我们不知道但希望知道的。例如,我们不知道北京市场上偏好iPhone手机的消费者比例,不知道北京城市大学学

生的英语平均成绩,但它们都是我们想要知道的数据,为得到这些数据可以采用全面调查,即对北京市场上每个手机消费者都进行调查,对北京城市大学的所有学生都进行统计,但如果调查所有人是不可能的或不经济的,我们会采用抽样的方式,根据从抽取的样本中所获得的资料来推断总体参数。

2. 统计量

统计量(statistic)也称为样本统计量或样本指标,是用来描述样本的数量特征,如样本平均数、样本标准差、样本比例等。注意,统计量是样本的函数,会随着样本的不同而变化,所以统计量是一个随机变量。我们习惯用英文字母来表示样本统计量,如样本平均数用 \bar{x} 表示,样本标准差用 s 表示,样本比例用 p 表示。由于样本是我们从总体中抽取的,所以统计量总是可以计算得到的。抽样的目的就是要用统计量来推算总体参数。两者之间的关系可见图1-4。

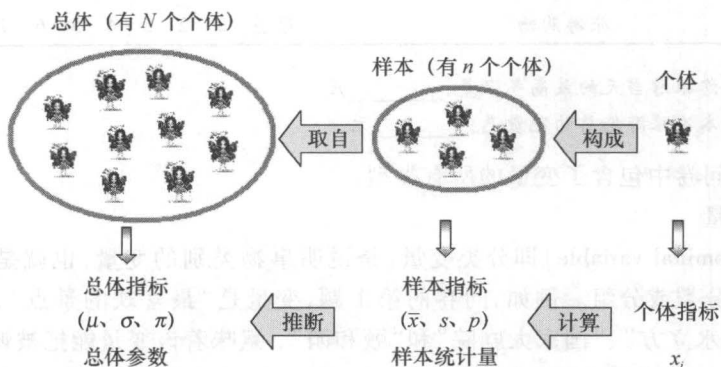


图 1-4 参数与统计量的关系

1.2.3 变量和数据

1. 变量和数据

变量(variable)是用来说明现象某种特征的概念,其特点是没有固定的值,每次观测的结果都会发生变化。例如,观察北京城市大学每个学生的英语成绩,每次观测结果都不一样,所以“英语成绩”是一个变量;而“英语成绩”的每个观测结果就是变量的取值,称为数据(data)。类似的还有:“手机品牌”、“企业每月销售收入”、“员工的学历水平”等都是变量,这些变量及其取值(数据)如表1-2所示。

表 1-2 变量与数据

变量	数据
英语成绩	85,76,82,64,95,70……
手机品牌	诺基亚,三星,苹果,摩托罗拉,HTC……
企业每月销售收入	50万元,51万元,50.5万元,52万元,51.7万元……
员工的学历水平	高中及以下,大专,本科,研究生及以上

2. 变量的类型

数据对于统计来说是非常重要的,如果没有数据,那么统计如同“巧妇难为无米之炊”。问卷调查是获得数据的重要途径,现在就让我们从一份调查问卷入手来认识变量的类型。北京是著名的旅游城市,为了促进北京旅游业的发展,北京市消费者协会与北京质量协会联合进

行了《北京奥运精品一日游问卷调查》,许多游客参与了调查,下面是调查问卷的一部分:

北京奥运精品一日游问卷调查

1. 【单选题】本次旅游您最喜欢的景点是:

(1) 鸟巢	(2) 水立方	(3) 国家大剧院	(4) 颐和园
--------	---------	-----------	---------

2. 【单选题】在旅游过程中,请您对此次旅游的服务质量进行评价:

编号	问题	评价
1	旅游交通	很差 1 2 3 4 5 6 7 很好
2	旅游餐饮	很差 1 2 3 4 5 6 7 很好
3	旅游住宿	很差 1 2 3 4 5 6 7 很好
4	旅游购物	很差 1 2 3 4 5 6 7 很好

3. 【填空题】您旅游当天的最高气温是:_____度。

4. 【填空题】本次旅游您总的花费是:_____元。

在这份调查问卷中包含了变量的所有类型:

(1) 定类变量

定类变量(nominal variable)即分类变量,是说明事物类别的变量,也就是说,它的主要作用是对事物进行分类或分组。例如,问卷的第1题,变量是“最喜欢的景点”,变量值有四个,分别是“鸟巢”、“水立方”、“国家大剧院”和“颐和园”,意味着该变量能把被调查者分成四类,喜欢“鸟巢”的是一类,喜欢“水立方”的是一类,喜欢“国家大剧院”的是一类,喜欢“颐和园”的是一类。类似的变量很多,如变量“性别”,其变量值为“男”和“女”;变量“手机品牌”,其变量值为“诺基亚”、“三星”、“苹果”、“摩托罗拉”和“HTC”等;变量“地区”,其变量值为“东部地区”、“中部地区”和“西部地区”。定类变量的取值就是定类数据。

(2) 定序变量

定序变量(ordinal variable)即顺序变量,是说明事物顺序类别的变量,也就是说,它的作用是可以对事物进行分类或分组,还可以说明事物顺序或等级关系。例如,问卷第2题的第一个小题变量是“旅游交通服务质量”,从“很差”到“很好”可以把被调查者分为七类,且每一类被调查者对服务质量的评价有等级、好坏的差异,这就是定序变量。类似的变量很多,如变量“学生成绩”,其变量值可以分为“优秀”、“良好”、“中等”、“及格”和“不及格”五个等级;变量“产品的质量”,其变量值可以分为“优等品”、“合格品”和“不合格品”三个等级;变量“用户的满意度”可以分为“很满意”、“满意”、“一般”、“不满意”和“很不满意”五个等级。定类变量和定序变量都属于定性变量,但定序变量显然比定类变量含有更多的信息量,因为定序变量不仅能够起到分类分组的作用,还能进行等级、层次和好坏的比较。定序变量的取值就是定序数据。

(3) 数值型变量

数值型变量(metric variable)是说明事物数字特征的变量,其结果表现为具体数值。例如,问卷的第3题和第4题,变量“温度”和“花费”的取值都是具体的一个数值。表1-2中的变量“英语成绩”和“企业每月销售收入”的取值也都是具体的数值,它们就是数值型变量。数值型变量根据其取值不同又可以分为离散型变量(discrete variable)和连续型变量(continuous