

21 世纪通识教育系列教材

21st Century Textbooks of General Education

统计学基础 (第3版)

Fundamental Statistics

贾俊平 © 编著

 中国人民大学出版社

21 世纪通识教育系列教材
21st Century Textbooks of General Education

统计学基础 (第3版)

Fundamental Statistics
.....▶

贾俊平 © 编著

中国人民大学出版社

• 北京 •

图书在版编目 (CIP) 数据

统计学基础/贾俊平编著. —3 版. —北京: 中国人民大学出版社, 2017. 6
21 世纪通识教育系列教材
ISBN 978-7-300-24343-6

I. ①统… II. ①贾… III. ①统计学-高等学校-教材 IV. ①C8

中国版本图书馆 CIP 数据核字 (2017) 第 094332 号

21 世纪通识教育系列教材

统计学基础 (第 3 版)

贾俊平 编著

Tongjixue Jichu

出版发行	中国人民大学出版社	
社 址	北京中关村大街 31 号	邮政编码 100080
电 话	010-62511242 (总编室)	010-62511770 (质管部)
	010-82501766 (邮购部)	010-62514148 (门市部)
	010-62515195 (发行公司)	010-62515275 (盗版举报)
网 址	http://www.crup.com.cn	
	http://www.ttrnet.com(人大教研网)	
经 销	新华书店	
印 刷	北京富泰印刷有限责任公司	版 次 2010 年 4 月第 1 版
规 格	185 mm×260 mm 16 开本	2017 年 6 月第 3 版
印 张	13.75 插页 1	印 次 2017 年 6 月第 1 次印刷
字 数	303 000	定 价 30.00 元

版权所有 侵权必究 印装差错 负责调换

· 出版说明 ·

随着信息时代的来临、经济全球化的深入与文化软实力竞争的加剧，重视大学生人文素养与创新能力的培养，提升大学生的综合素质，已成为各国教育改革与发展关注的重点和热点。人们越来越意识到：高等教育不仅要培养大学生良好的专业素质，更重要的是使得他们在走向社会之后拥有长足的自我拓展能力。只有以宽口径、厚基础、复合型为人才培养目标，才能更好地提高我国高等教育的质量，培育出适应现代社会需求的具备公民意识、社会责任感与创新精神的优秀人才。

从中外大学通识教育的实践来看，通识教育是一项系统工程，而课程体系建设始终是推进通识教育的核心任务，教材建设则又是其中的重要环节。为满足广大高校师生对高质量通识教育教材的需求，中国人民大学出版社组织多学科、多领域的专家学者，在广泛调研与深入研讨的基础上，组织编写了这套“21世纪通识教育系列教材”，为推动高等学校通识教育教材建设进行了探索。

本套教材分为人文、政法和经管三大板块，定位为非专业统开课教材，突出“通识”的特色，强调内容阐释的“基础”和“宽度”，力求突破单纯的“专业视域”或“知识视域”，引导学生调整知识结构，拓宽文化视野，以达成人才培养效果上的“宽度”，从而实现高等教育培养复合型人才的目标。

本套教材中的每一本均由该学科领域有影响力的专家学者领衔编撰。通识教材的“基础”与“宽度”，需要特别重视教材纲目与内容的适用性、可拓展性和灵活性。唯有在该领域具有丰富教学经验及精深学术水准的名家，方能“取精用弘，由博返约”，编撰出体现“通识”特色的高水平教材。

本套教材形式与内容和谐统一，教材内容基础适用，语言简洁生动，并辅以典型、有趣的案例、图表，轻松活泼的栏目和插图等，图文并茂，引人入胜，照顾到青年学生群体的阅读习惯。

作为出版者，我们特别希望通过加强通识教育教材建设，推进高校课程体系的融会贯通，提高学生跨学科、跨文化的理解能力，为学生未来的职业生涯与人生发展奠定良好的知识和能力基础。这套通识教育系列教材只是开始，期望更多的专家学者共襄此事，推进通识教育教学的改革与发展。

中国人民大学出版社

· 前 言 ·

本书是在 21 世纪通识教育系列教材《统计学基础（第二版）》的基础上修订而成的。考虑到读者的特点和实际学习需要，本次修订保留了第二版的基本内容框架，但对所有章节都进行了重新编写。增加了一些新内容，并删去了 Excel 不能实现的一些内容。全书内容均使用 Excel 2013 版本实现，并给出了详细的操作步骤。同时，考虑到篇幅，对书后的习题及答案做了精简。

全书共包括 8 章内容。第 1 章介绍统计学的一些基本问题，目的是让读者对统计学有一个基本的了解。第 2 章介绍数据的图表展示，使读者学会数据的基本处理方法和图形展示技术。第 3 章介绍数据特征的统计描述方法，主要讨论如何使用一些统计量来对数据进行概括性度量。第 4 章介绍抽样与参数估计，讨论一些常用的概率抽样方法，并重点介绍参数估计的原理和总体参数的估计方法。第 5 章介绍假设检验的基本方法，讨论假设检验的原理及总体均值和总体比例的检验方法。第 6 章介绍相关与回归分析，讨论变量之间关系强度的度量方法，并介绍一元线性回归分析及应用。第 7 章介绍时间序列的分析和预测方法，主要讨论时间序列的描述性方法以及对时间序列的预测。第 8 章介绍指数，简要介绍加权指数的编制方法及实践中常用的几种价格指数。

本书所有例题均使用 Excel 2013 实现，所有例题和习题数据也均为 Excel 2013 格式，建议读者在使用前安装好 Excel 2013 或更高版本。希望读者于对本书使用中的问题或要求提出宝贵意见，并将您的建议及时反馈给我们，以便做出进一步修改和完善。

本书学习资源下载网址：www.rdjg.com.cn。

贾俊平
于中国人民大学

· 目 录 ·

第1章 统计学和统计数据

1.1 统计学及其应用领域	1
1.1.1 什么是统计学	1
1.1.2 统计学的应用领域	2
1.2 数据的类型	4
1.2.1 类别数据和数值数据	4
1.2.2 观测数据和实验数据	5
1.2.3 截面数据和时间序列数据	5
1.3 数据的来源	6
1.3.1 数据的间接来源	6
1.3.2 数据的直接来源	7
1.3.3 调查方案设计	10
1.3.4 数据质量	11
1.4 统计中的几个基本概念	12
1.4.1 总体和样本	12
1.4.2 参数和统计量	12
1.4.3 变量	13
1.5 Excel【数据分析】工具的安装	14
本章小结	16
练习题和答案	16

第2章 数据的图表展示

2.1 数据的预处理	20
2.1.1 数据审核	20
2.1.2 数据排序	21
2.1.3 数据筛选	22
2.2 数据的频数分布	25
2.2.1 类别数据的频数分布表	25
2.2.2 数值数据的类别化	29
2.3 类别数据的图示	33
2.3.1 条形图和帕累托图	33

2.3.2 饼图和环形图	36
2.4 数值数据的图示	37
2.4.1 直方图	38
2.4.2 散点图和气泡图	39
2.4.3 雷达图和轮廓图	41
2.4.4 时间序列图	43
2.5 合理使用图表	44
本章小结	45
练习题和答案	46

第3章 数据的概括性度量

3.1 集中趋势的度量	56
3.1.1 平均数	56
3.1.2 分位数	58
3.1.3 众数	62
3.1.4 各度量值的比较	62
3.2 离散程度的度量	63
3.2.1 极差和四分位差	63
3.2.2 平均差	64
3.2.3 方差和标准差	65
3.2.4 离散系数	67
3.2.5 标准分数	68
3.3 偏度与峰度的度量	70
3.3.1 偏度系数	70
3.3.2 峰度系数	71
3.4 Excel【数据分析】工具的应用	71
本章小结	73
练习题和答案	74

第4章 抽样与参数估计

4.1 抽样与抽样分布	78
4.1.1 概率抽样方法	79
4.1.2 抽样分布	82
4.2 参数估计的基本原理	87
4.2.1 估计量与估计值	87
4.2.2 点估计与区间估计	88

4.3 总体均值的区间估计	91
4.3.1 正态总体、方差已知，或非正态总体、大样本	91
4.3.2 正态总体、方差未知、小样本	94
4.4 总体比例的区间估计	96
4.5 样本量的确定	97
4.5.1 估计总体均值时样本量的确定	98
4.5.2 估计总体比例时样本量的确定	99
本章小结	99
练习题和答案	100

第5章 假设检验

5.1 假设检验的基本原理	105
5.1.1 假设的陈述	105
5.1.2 两类错误与显著性水平	108
5.1.3 检验统计量与拒绝域	110
5.1.4 利用 P 值进行决策	112
5.2 总体均值的检验	116
5.2.1 大样本的检验方法	116
5.2.2 小样本的检验方法	121
5.3 总体比例的检验	123
本章小结	125
练习题和答案	125

第6章 相关与回归分析

6.1 变量间关系的度量	132
6.1.1 变量间的关系	132
6.1.2 相关关系的描述与测度	133
6.1.3 相关关系的显著性检验	139
6.2 一元线性回归分析	140
6.2.1 一元线性回归模型	140
6.2.2 参数的最小二乘估计	142
6.2.3 回归直线的拟合优度	145
6.2.4 显著性检验	148
6.2.5 利用回归方程进行估计和预测	151
本章小结	155
练习题和答案	155

第7章 时间序列分析和预测

7.1 时间序列的成分	163
7.2 增长率分析	164
7.2.1 增长率与平均增长率	164
7.2.2 增长率分析应注意的问题	167
7.3 时间序列预测的程序和方法	168
7.3.1 确定时间序列的成分	168
7.3.2 预测方法的选择与评估	169
7.4 平滑法预测	170
7.4.1 移动平均法	170
7.4.2 简单指数平滑法	171
7.5 趋势外推预测	175
7.5.1 线性趋势预测	175
7.5.2 非线性趋势预测	177
7.6 多成分序列的分解预测	181
7.6.1 确定并分离季节成分	182
7.6.2 建立预测模型并进行预测	184
7.6.3 计算最终预测值	185
本章小结	186
练习题和答案	187

第8章 指数

8.1 指数及其分类	194
8.2 加权指数	195
8.2.1 加权综合指数	195
8.2.2 加权平均指数	197
8.2.3 价值指数与指数体系	198
8.3 几种常用的价格指数	199
8.3.1 消费者价格指数	199
8.3.2 生产者价格指数	200
8.3.3 股票价格指数	200
本章小结	200
练习题和答案	201

附录 用 Excel 生成概率分布表	206
参考书目	210

第1章 统计学和统计数据

内容提要

本章主要介绍统计学的一些基本问题，目的是使读者对统计学有一个初步了解和认识，具体内容包括：

(1) 统计及其应用领域。介绍统计学的含义以及统计的应用领域等。

(2) 统计数据及其类型。介绍统计数据及其类型以及统计数据与统计方法之间的关系。

(3) 统计数据的来源。介绍二手数据的来源、原始数据的调查和收集方法、调查方案的设计以及统计数据的误差等问题。

(4) 统计中的基本概念。介绍统计中常用的几个基本概念，包括总体和样本、参数和统计量以及变量等。

1.1 统计学及其应用领域

在生活中，人们经常会接触到统计数据，比如，各种报道中经常使用统计数据、图表等。本章将介绍统计学的一些基本问题，包括统计学的含义、统计数据及其分类、统计中的一些基本概念等。

1.1.1 什么是统计学

统计学是分析数据的一门科学。统计学家给统计学下的定义很多，如“统计学是收集、分析、表述和解释数据的科学”，“统计学是一门收集、分析、解释和提供数据的科学”，“统计学指的是一组方法，用来设计实验、获得数据，然后在这些数据的基础上组织、概括、演示、分析、解释和得出结论”。总结上述定义，可以将统计学的含义概括如下：

► **定义 1.1** 统计学 (statistics) 是收集、处理、分析、解释数据并从数据中得出结论的科学。

统计学的定义告诉我们，统计学是关于数据的科学，它提供一套有关数据收集、数据处理、数据分析、数据解释并从数据中得出结论的原则和方法。统计学研究的是来自各领域的

统计学提供的数据分析方法大致可分为描述统计和推断统计。

► **定义 1.2** 研究数据收集、处理和描述的统计学方法,称为描述统计 (descriptive statistics)。

描述统计的内容包括取得研究所需要的数据、用图表形式对数据进行处理和显示,进而通过综合、概括与分析,得出反映所研究现象的一般性特征。

► **定义 1.3** 研究如何利用样本数据来推断总体特征的统计学方法,称为推断统计 (inferential statistics)。

推断统计是研究如何利用样本数据来推断总体特征的统计学方法,内容包括参数估计和假设检验两大类。参数估计是利用样本信息推断所关心的总体特征,假设检验则是利用样本信息判断对总体的某个假设是否成立。比如,从一批灯泡中随机抽取少数几个灯泡作为样本,测出它们的使用寿命,然后根据样本灯泡的平均使用寿命估计这批灯泡的平均使用寿命,或者是检验这批灯泡的使用寿命是否等于某个假定值,这就是推断统计要解决的问题。

1.1.2 统计学的应用领域

统计方法是适用于所有学科领域的通用数据分析方法,只要有数据的地方就会用到统计方法。目前,随着定量研究的日趋重要,统计方法已应用到自然科学和社会科学的众多领域。可以说,几乎所有的研究领域都要用到统计方法。比如,政府部门、学术研究领域、日常生活、公司或企业的生产经营管理中都要用到统计。下面给出统计在工商管理中的一些应用。

1. 企业发展战略

发展战略是一个企业的长远发展方向。制定发展战略,一方面需要及时了解和把握整个宏观经济的状况及发展变化趋势,以便了解市场的变化;另一方面,还要对企业进行合理的市场定位,以便把握企业自身的优势和劣势。所有这些方面都离不开统计,它们需要统计提供可靠的数据、利用统计方法对数据进行科学的分析和预测,等等。

2. 产品质量管理

质量是企业的生命,是企业持续发展的基础。我们在质量管理中同样离不开统计的应用。在一些知名的跨国公司中,6 σ 准则已成为一种重要的管理理念。由此可知,质量控制已成为统计学在生产领域中的一项重要应用。此外,各种统计质量控制图还广泛用于监测生产过程。

3. 市场研究

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,企业就需要作广泛的市场调查,以取得所需的信息,然后对这些信息进行科学的分析,并以此作为生产和营销的依据。这些措施都需要统计的支持。

4. 财务分析

上市公司的财务数据是股民进行投资选择的重要参考依据。一些投资咨询公司主要

是根据上市公司提供的财务和统计数据进行分析,为股民提供投资参考。企业自身的投资也离不开对财务数据的分析,其中就要用到大量的统计方法。

5. 经济预测

企业要对未来的市场状况进行预测,经济学家也常常对宏观经济或某一方面进行预测,他们在进行预测时就要使用各种统计信息和统计方法。比如,企业要对产品的市场潜力作出预测,以便及时调整生产计划,此时就需要利用市场调查取得数据,并对数据进行统计分析。经济学家在预测通货膨胀时,要利用有关生产价格指数、失业率、生产能力利用等统计数据,然后利用统计模型进行预测。

6. 人力资源管理

利用统计方法对企业员工的年龄、性别、受教育程度、工资等进行分析,并以此作为企业制定工资计划、奖惩制度的依据。

当然,统计并不仅仅是为了管理,它是为自然科学、社会科学的多个领域而发展起来的,并为多个学科提供了一种通用的数据分析方法。从某种意义上说,统计仅仅是一种数据分析的方法,与数学一样,统计是一种工具,它是一种数据分析的工具。

统计的主要作用是帮助我们分析数据。比如,我们可以利用统计简化繁杂的数据,用图表重新展示数据,再建立数据模型,并进行比较分析,等等。对于一个工商管理人员来说,他会面临企业经营管理的大量数据。这些数据对该管理人员进行管理与决策起到了什么样的作用?一个决策者可能更喜欢看文字性的材料,或者某种结论性的东西,他看到数据就会头痛。也就是说,他“害怕”数字。殊不知,结论性的东西就是来自对数据的分析。

统计的目的不仅仅是让人们看懂数据,其主要功能是分析数据。统计学提供一套分析数据的通用方法和工具。不同的人对数据分析的理解也会不一样,曲解数据分析是一种常见的现象。在有些人的心目中,数据分析就是寻找支持:他们的心目中可能有了某种结论性的东西;或者说,他们希望看到一种符合他们需要的结论,然后再去找些统计数据来支持他们的结论。这恰恰歪曲了数据分析的本质:数据分析的真正目的是从数据中找出规律、从数据中寻找启发,而不是寻找支持。真正的数据分析事先是没有结论的,我们需要对数据进行分析才能得出结论。

当然,统计不是万能的,它不能解决我们所面临的一切问题。统计能帮助我们进行数据分析,并从分析中得出某种结论,但对统计结论的进一步解释,则需要分析人员具备专业知识。统计不能为管理人员提供所需的一切技巧和方法。当我们把统计用到管理领域时就会发现,统计和管理之间还有差距。许多学过统计的管理人员总觉得还缺少点什么东西,好像统计没有从根本上解决他们的问题。实际上,大部分统计方法都需要某些假设作为前提。比如,我们在建立一个线性回归模型时,首先要假设变量之间的关系是线性的,否则我们就无从下手。但是,这种假设必须经过严格的检验,之后才能应用所建立的模型。作为一个管理人员,他更关心真正有用的信息是什么,结论是什么,应采取什么样的行动,他并不那么关心严格的统计检验。当然,该管理人员也会检验信息的正确性,但他不一定是通过统计检验,而是通过某种其他方式,如定性的方法,或者

干脆就凭直觉。然而,自然科学强调应在不同条件下对非统计方法推导出的假设进行严格的检验。显然,统计在自然科学和管理科学中的作用是不同的。管理人员所需要的功能,统计未必能提供。如果我们希望在没有外界帮助的条件下找出数据的特征或规律,那么统计对此无能为力。统计是一套分析数据的方法和工具,我们不能指望统计成为解决所有问题的灵丹妙药。

1.2 数据的类型

统计数据是对现象进行观测或实验的结果。比如,对经济活动总量进行测量可以得到国内生产总值(GDP)数据,对股票价格变动水平进行测量可以得到股票价格指数的数据,对人口性别进行测量可以得到男或女这样的数据,等等。由于使用的测量尺度不同,统计数据可以分为不同的类型。下面我们从不同角度说明统计数据的分类。

1.2.1 类别数据和数值数据

按照所采用的不同计量尺度,我们可以将统计数据分为类别数据和数值数据。

► **定义 1.4** 只能归于某一类别的非数字型数据,称为类别数据(categorical data)。

类别数据也称为定性数据或品质数据(qualitative data),它是对事物进行分类的结果,数据则表现为类别,是用文字来表述的。比如,观察人的性别、上市公司所属的行业、用户对商品满意度的评价,得到的结果就不是数字,而是事物的属性。比如,观测性别的结果是“男”或“女”,上市公司所属的行业为“金融业”“地产业”“旅游业”等,用户对商品满意度的评价为“很满意”“满意”“一般”“不满意”“很不满意”。这些都是类别数据。此外,将从业人员的月收入分为5 000元以下、5 000~10 000元、10 000~15 000元、15 000~20 000元、20 000元以上5档,这里的“月收入档次”的取值也不是普通的数值,而是数值区间,这实际上是将数值转化成了类别,因此也属于类别数据。

类别数据根据取值是否有序通常分为无序类别数据和有序类别数据两种。无序类别的各个取值是不可以排序的。例如上市公司所属的行业这一变量取值为“金融业”“地产业”“旅游业”等,这些取值之间不存在顺序关系。有序类别数据也称为顺序(ordinal)数据,其取值间可以排序。例如对商品满意度的评价这一变量的取值为“很满意”“满意”“一般”“不满意”“很不满意”,这5个值之间是有序的。

有序类别数据也是对事物进行分类的结果,但这些类别是有顺序的。比如,产品可分为一等品、二等品、三等品、次品等;考试成绩可以分为优、良、中、及格、不及格等;一个人的受教育程度可以分为小学、初中、高中、大学及以上;一个人对某一事物的态度可以分为非常同意、同意、保持中立、不同意、非常不同意,等等。同样,我们对于对有序类别数据也可以用数字代码来表示。比如,1表示“非常同意”,

2 表示“同意”，3 表示“保持中立”，4 表示“不同意”，5 表示“非常不同意”。

► **定义 1.5** 按数字尺度测量的观察值，称为数值数据 (metric data)。

数值数据 (quantitative data) 的取值为数字。例如“企业销售额”“股票价格”“月收入”“投掷一枚色子出现的点数”等，取值都用数字来表示，都属于数值数据。在现实中，我们所处理的大多数数据都是数值数据。

数值数据根据其取值的不同，可以分为离散数据 (discrete data) 和连续数据 (continuous data)。离散数据只能取有限个值，而且其取值可以一一列举，如“上市公司家数”“一个社区的居民户数”等就是离散数据。连续数据是可以在一个或多个区间中取任何值，它的取值是连续不断的，不能一一列举，如“温度”“股票价格”等都是连续数据。

1.2.2 观测数据和实验数据

按照统计数据的收集方法，可以将其分为观测数据和实验数据。

► **定义 1.6** 通过调查或观测收集到的数据，称为观测数据 (observational data)。

观测数据是在没有对事物进行人为控制的条件下得到的，有关社会经济现象的统计数据几乎都是观测数据。

► **定义 1.7** 在实验中控制实验对象收集到的数据，称为实验数据 (experimental data)。

比如，对一种新药疗效的实验数据，对一种新农物品种的实验数据。自然科学领域的大多数数据都是实验数据。

1.2.3 截面数据和时间序列数据

按照被描述的现象与时间的关系，可以将统计数据分为截面数据和时间序列数据。

► **定义 1.8** 在相同或近似相同的时间点上收集的数据，称为截面数据 (cross-sectional data)。

截面数据所描述的是现象在某一时刻的变化情况，它通常是在不同的空间上获得的数据。比如，2016 年我国各地区的国内生产总值数据就是截面数据。

► **定义 1.9** 在不同时间上收集到的数据，称为时间序列数据 (time series data)。

时间序列数据是按照时间顺序收集到的，用于描述现象随时间而变化的情况。比如，2000 年至 2016 年我国的国内生产总值数据就是时间序列数据。

图 1-1 给出了统计数据分类的框图。

区分数据的类型是很重要的。因为对不同类型的数据，需要采用不同的统计方法来处理和分析。比如，对于无序类别数据，我们通常计算出各组的频数或频率，计算其众数和异众比率，进行列联表分析和 χ^2 检验等；对于有序类别数据，我们可以计算其中位数和四分位差以及等级相关系数等；对于数值数据，我们可以用更多的统计方法进行分析，如计算各种统计量、进行参数估计和检验等。

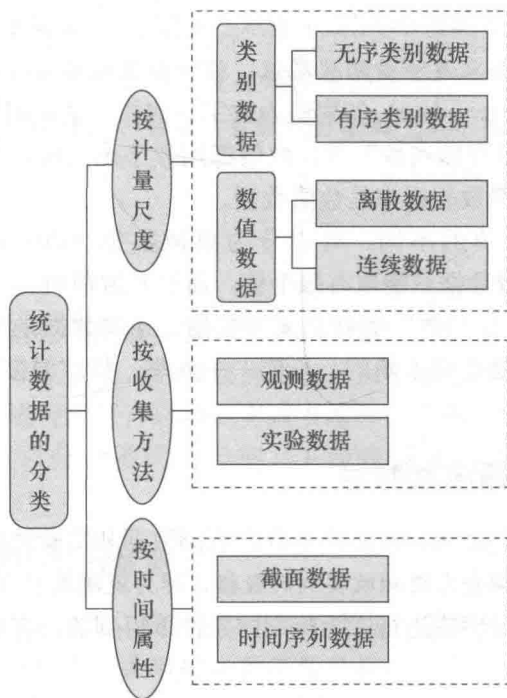


图 1-1 统计数据的分类

1.3 数据的来源

应用统计方法分析问题时首先要取得数据。从使用者的角度看，数据主要来源于两种渠道：一是来源于直接的调查和科学实验的数据，对使用者来说，这是数据的直接来源，称为一手数据或直接数据；二是来源于别人调查或实验的数据，对使用者来说，这是数据的间接来源，称为二手数据或间接数据。

1.3.1 数据的间接来源

对大多数使用者来说，亲自去作调查往往是不可能的或不必要的。他们所使用的数据大多是别人调查或科学实验的数据，这些数据对使用者来说是二手数据。

二手数据主要是公开出版或公开报道的数据，当然有些是尚未公开出版的数据。公开出版或报道的社会经济统计数据主要来自国家和地方的统计部门以及各种报刊媒介。除了公开出版的统计数据外，还可以通过其他渠道使用一些尚未公开的统计数据以及广泛分布在各种报纸、杂志、图书、广播、电视媒体中的数据。现在，随着互联网的发展，也可以在网络上获取所需的各种数据，例如，登录国家统计局网站（www.stats.gov.cn）即可得到有关宏观经济和社会发展的各种统计数据。

利用二手数据对使用者来说既经济又方便，但使用时应注意统计数据的含义、计算

口径和计算方法,以避免误用或滥用。同时,使用者在引用二手数据时,一定要注明数据的来源,以尊重他人的劳动成果。

1.3.2 数据的直接来源

数据的直接来源主要有两个渠道:一是调查或观察;二是实验。调查是取得社会经济数据的重要手段,其中有统计部门进行的调查,也有其他部门或机构为特定目的而进行的调查,如市场调查等;实验是取得统计数据的重要方法之一,它是取得自然科学数据的主要手段。在此,我们着重介绍取得社会经济数据的主要方式和方法。

1. 统计调查方式

统计调查是取得社会经济数据的主要来源,也是获得直接统计数据的重要手段。我们在实际中常用的统计调查方式主要有抽样调查、普查、统计报表等。

(1) 抽样调查。

► **定义 1.10** 从总体中随机抽取一部分单位作为样本进行调查,并根据样本调查结果来推断总体特征的数据收集方法,称为抽样调查(sampling survey)。

抽样调查是实践中应用最广泛的一种调查方式和方法,它具有以下几个特点。

1) 经济性。这是抽样调查的一个最显著优点。由于调查的样本单位通常是总体单位中的很小一部分,调查的工作量小,因而可以节省大量的人力、物力、财力和时间,其调查费用较低。

2) 时效性强。抽样调查可以迅速、及时地获得所需要的信息。由于工作量小,调查的准备时间、调查时间、数据处理时间等都可以大大缩减,从而提高数据的时效性。与普查等全面调查相比,抽样调查可以频繁地进行,并随事物的发生和发展及时取得有关信息,以弥补普查等全面调查的不足。比如,在两次人口普查之间各年份的人口数据都是通过抽样调查取得的。

3) 适应面广。抽样调查可以获得更广泛的信息,它适用于对各个领域、各种问题的调查。从适用的范围和问题来看,抽样调查可用于调查全面调查能够调查的现象,也能调查全面调查所不能调查的现象。抽样调查特别适合对一些特殊现象的调查,如产品质量检验、农产品实验、医药的临床实验等。从调查的项目和指标来看,抽样调查的内容和指标可以更详细、深入,能获得更全面、更广泛和更深入的数据。

4) 准确性高。抽样调查的数据质量有时比全面调查更高,因为全面调查的工作量大、环节多,登记性(或调查)误差往往很大;而抽样调查由于工作量小,可使各环节的工作做得更细致,误差往往很小。当然,用样本数据去推断总体时,不可避免地会有推断误差,但这种误差的大小是可以计算并加以控制的,因此推断的结果通常是可靠的。

抽样调查有不同的样本抽取方法,这一问题将在第4章中详细介绍。

(2) 普查。

► **定义 1.11** 为特定目的而专门组织的全面调查,称为普查(census)。

世界各国一般都定期进行各种普查,以便掌握有关国情国力的基本统计数据。普查

是适合特定目的、特定对象的一种调查方式,它主要用于收集处于某一时点状态的社会经济现象的数量,目的是掌握特定社会经济现象的基本全貌,为国家制定有关政策或措施提供依据。

目前,我国进行的普查主要有人口普查、农业普查和经济普查。其中,经济普查的内容包括工业普查、第三产业普查、基本单位普查和建筑业普查等。由于普查涉及的面广、调查单位多,需要耗费大量的人力、物力、财力和时间,通常需要间隔较长的时间进行一次。在我国的普查中,经济普查每5年进行一次,其他普查每10年进行一次。每逢年份的末尾数字为“0”的年份进行人口普查,每逢“6”的年份进行农业普查,每逢“3”和“8”的年份进行经济普查。

(3) 统计报表。

► **定义 1.12** 按照国家有关法规的规定,自上而下地统一布置、自下而上地逐级提供基本统计数据的调查方式,称为统计报表 (statistical report forms)。

统计报表是收集统计数据的一种重要方式,在我国几十年的政府统计工作中,已形成了一套比较完备的统计报表制度,它已成为国家和地方政府部门统计数据的主要来源。

2. 数据的收集方法

不论采取何种方式进行调查,在取得统计数据时,都有一些具体的数据收集方法。数据的收集方法归纳起来可分为询问调查、观察和实验三大类。

(1) 询问调查。询问调查是调查者与被调查者直接或间接接触以获得数据的一种方法,具体包括访问调查、邮寄调查、电话调查、电脑辅助调查、座谈会、个别深访等。

1) 访问调查。访问调查又称派员调查,它是调查者与被调查者通过面对面的交谈从而得到所需资料的调查方法。访问调查的方式有标准式访问和非标准式访问两种。标准式访问又称结构式访问,它是按照调查者事先设计好的有固定格式的标准化问卷或表格,有顺序地依次提问,并由被调查者作出回答。其优点是能够对调查过程加以控制,从而获得比较可靠的调查结果。非标准式访问又称非结构式访问,它事先不制作统一的问卷或表格,没有统一的提问顺序,调查者只是给一个题目或提纲,由调查者和被调查者自由交谈,以获得所需的资料。在市场调查和社会调查中常采用访问调查。

2) 邮寄调查。它是通过邮寄或宣传媒体等方式将调查表或调查问卷送至被调查者手中,由被调查者填写,然后将调查表寄回或投放到指定收集点的一种调查方法。邮寄调查是一种标准化调查,其特点是:调查者和被调查者没有直接的语言交流,信息的传递完全依赖于调查表。邮寄调查的问卷或表格发放方式有邮寄、宣传媒介传送、专门场所分发三种。统计部门进行的统计报表及市场调查机构进行的问卷调查中经常使用邮寄调查。

3) 电话调查。它是调查者利用电话同被调查者进行语言交流,从而获得信息的一种调查方式。电话调查具有时效快、费用低等特点。随着电话的普及,电话调查的应用