

全国高等院校经济和管理专业教材

新编 应用统计

XINBIAN
YINGYONG TONGJI

主编 贾怀勤

 中国商务出版社
CHINA COMMERCE AND TRADE PRESS

本书配有数据光盘一张



全国高等院校经济和管理专业教材

新编应用统计

主编 贾怀勤

中国商务出版社

图书在版编目 (CIP) 数据

新编应用统计/贾怀勤主编. —北京: 中国商务出版社,
2009. 8

全国高等院校经济和管理专业教材

ISBN 978-7-5103-0111-7

I. 新… II. 贾… III. 应用统计学-高等学校-教材
IV. C8

中国版本图书馆 CIP 数据核字 (2009) 第 148297 号

全国高等院校经济和管理专业教材

新编应用统计

XINBIAN YINGYONG TONGJI

主编 贾怀勤

出 版: 中国商务出版社

发 行: 北京中商图出版物发行有限责任公司

社 址: 北京市东城区安定门外大街东后巷 28 号

邮 编: 100710

电 话: 010—64269744 (编辑室)

010—64266119 (发行部)

010—64263201 (零售、邮购)

网 址: www.cctpress.com

邮 箱: cctp@cctpress.com

照 排: 嘉年华文排版公司

印 刷: 三河市和达印务有限公司

开 本: 787 毫米×1092 毫米 1/16

印 张: 18 字 数: 323 千字

版 次: 2009 年 8 月第 1 版 2009 年 8 月第 1 次印刷

书 号: ISBN 978-7-5103-0111-7

定 价: 33.00 元 (含光盘一张)

版权专有 侵权必究

举报电话: (010) 64242964

前 言

本书旨在为高校的非统计专业（主要指经济和管理各专业）的应用统计教学提供一部新型教科书。本书的编写指导思想，一是注重与中学数学和大学“概率论与数理统计”教学内容的衔接；二是力求更加贴近商务应用。

本书主编多年来致力于应用统计教学和教材建设，早期曾协助李志伟教授主编国内最早的区别于“社会经济统计学原理”教学体系的应用统计教科书——《统计分析概论》（1984年，1989年共两版），后来又牵头编写换代教科书《应用统计》（1994年，1998年，2002年，2006年共四版）。

1978年统计科学研讨会——峨眉会议至今已过去30年，1992年正式提出建设大统计学科也已过去16年。随着统计学研究及其在经济和社会发展中应用的历史性进步，统计学教学及教材建设也呈现一派繁荣局面。统计学不仅在高校许多专业列为重要的方法性课程，还被引入了中学的数学课。为准备编写本书，本书主编阅读了许多同行编写的教材，也参阅了初、高中的最新数学课本。

按人民教育出版社2008版课本的设计，初、高中统计知识的必修内容包括：

数据的认识——数据的搜集；数据的描述（常见统计图表）；数据的分析（数据的代表和波动）。

概率初步——排列组合和二项式定理；随机事件；基于频率（相对频数）的概率；古典概率；列举法求概率；二项分布。

选修内容包括：

抽样方法——放回抽样与不放回抽样；无限总体与有限总体；简单随机抽样；系统抽样；分层抽样。

用样本估计总体——用样本的频率分布估计总体的分布；用样本的平均数和方差估计总体的平均数和方差（点估计）。

概率分布：数学期望和方差；正态分布。

线性相关关系和回归分析——散点图；求相关系数和回归系数。

由以上信息可知，高校的统计学教学及教材建设不仅要面向社会对所需人才统计知识的需求，也面临如何与中学统计基础知识相衔接的问题。在高校内部，还有“概率论与数理统计”课与“应用统计”课的区分与衔接问题。

需要指出，本课程与先行课程在一些知识点上一定程度的重复是必要的。其理由在于：目前大学生乃至中学生，知识“吞”进去的趋多，“消化”和“吸收”

的趋少，往往是提起某名词似乎记得，但深究其理则忘得差不多了，此为其一；高中数学教学内容的必修与选修（还进一步分为选修Ⅰ和选修Ⅱ）安排，文理分科和应试教育，使得大学生入校前的统计初步知识“离散度”很强，应用统计教学必须照顾到学生知识背景之一般，此为其二。另一方面，应用统计对知识点的重复必须是重复中有新意、有提高，重复外有拓展，否则就是浪费课时，挫伤学生求知欲。

本书作为本科应用统计学基本教材，知识面涵盖统计数据概述、数据采集、统计描述、参数估计、假设检验、相关回归、时间数列和指数等专题，连同绪论在内共计13章。教师可以根据课时计划和学生接受程度对内容有所取舍，目录中标题字尾加“*”的章或节，是建议选讲的内容。

有关统计描述、参数估计、假设检验和相关回归的基本知识点，或在中学数学课中已有初步讲述，或在高等数学课中已经系统展开。从经济和管理各专业人才需求的角度讲述这些知识，做到虽有术语和公式的重复，又得让学生通过本课程有新的收获，就成为“新编”的关键。为此，本书在以下五个方面有所侧重，或者说本书在这些方面与其他统计学教材相比能够体现新意：

第一，从重新认识“数据”导入，使学生明了统计学的“数据”与数学的“数”的区别。“绪论”开宗明义指出数据是进行社会经济核算、科学研究和技术设计活动所依据的数值。统计学是关于如何对总合现象的测度结果——数据进行采集、整理和分析，以便为决策提供支持的科学。第一章至第三章关于数据的采集、整理和分布态势描述，以定名数据、定秩数据和定标数据为主线展开，并突出商务数据的采集和整理、汇总特点。

第二，妥善处理由描述统计到推断统计的过渡。第四章以“从经验分布到理论分布”为题，在前三章的描述统计内容与后面的推断统计内容之间架起一座栈桥，既避免了某些教材对概率论知识的较大篇幅重复，也不采用切除概率分布知识的硬过渡。

第三，既注重例题的示范性，更注重其应用导向性。在例题编写上，本书以逼近商务实际的“示例”取代教科书通常的“算例”。所谓“算例”，即是给出简单的场景和少量数值，带入公式进行计算，说明统计学原理和公式的作用。而“示例”则是对商务场景做出描述，给出含有足够量数值的商务数据，运用所讲统计学模型和方法进行分析，导出结果，并针对商务问题做出解释。这样编写例题，目的是将学生由“读例题——对照公式——代入数字算结果”的数学应用题解题思维升华为应用统计思维，增进学生对统计方法应用于商务管理的认识。

第四，在习题的编写上，注重训练学生的统计思维。本书习题共155题，涵盖从第二章到第十二章（绪论和第一章均无习题）。平均每章14题。习题对商务

场景做出描述,给出含有足够量数值的商务数据,要求学生综合运用所学知识解决问题,突出对学生理解和运用知识的训练和考察。许多习题都是贯穿前后数章,数据多次使用,使学生认识到商务问题解决的多层次性和多角度性。

第五,在计算工具上,引导学生运用 EXCEL 的统计分析功能。由于例题和习题给出的含有足够量数值的商务数据,必须借助于计算机软件进行运算。考虑到学生就业后一般工作单位计算机上 EXCEL 的普及性和该软件的友好性和多功能性,将 EXCEL 确定为本书的数据处理和分析工具,从而拉开了中学“统计初步”运算依靠电子计算器(中学教科书大纲语),大学本科“应用统计”数据处理和分析使用 EXCEL,未来研究生阶段深入统计分析使用 SPSS、SAS 或 Minitab 等大型专门统计软件的态势。

本书并不具体讲授 EXCEL 统计函数和统计工具的使用方法和程序,只是在每章最后写有“与本章有关的 EXCEL 功能的实现途径”的附录。学生凭借已有的对 EXCEL 电子表格数据结构和一般数学运算方法的初步认识,在教师指导下按照此附录的指南联系操作,完全能够掌握与本书所讲统计方法有关的 EXCEL 函数和统计工具。

本书中含有足够量数值的所有例题和习题的数据文件,全部灌装入随书附送的光盘中。

综上所述,本书以商务数据导入,以 EXCEL 电子表格为平台和工具,突出统计方法在经济和管理中的应用。作者力求将其写成一部适合学生知识背景和培养需要的具有商务应用特色的教科书。

主编贾怀勤教授编写全书绝大多数章节和习题,并负责总纂;丁岚副教授提供第五章、第十一章、第十二章的初稿;赵军教授参与全书的审校,提出中肯的意见。

本书难免存有失误和不足之处,恳请同行和读者给予指正。

编著者

2009.5

目 录

绪 论	(1)
0.1 数据与统计学	(1)
0.2 普查与抽样调查	(3)
0.3 数据的类别	(5)
本章要点	(7)
第 1 章 数据的采集	(9)
1.1 数据采集的方法和途径	(9)
1.2 数据采集的组织方式	(10)
1.3 数据采集活动的策划	(15)
1.4 中国特色的政府统计调查	(18)
1.5 其他数据来源	(21)
本章要点	(22)
第 2 章 数据的整理	(24)
2.1 数据的分组和汇总：计数型汇总	(24)
2.2 数据的分组和汇总：计量型汇总	(30)
2.3 复合分组与交叉分组	(31)
2.4 截面数据的图形展示	(32)
2.5 数据的组织形式：截面数列、时间数列和拼纳数列	(45)
本章要点	(48)
习题	(50)
第 3 章 截面数据的分布态势描述	(64)
3.1 数据分布特征概述	(64)
3.2 算术平均数、方差与标准差：仅适于定标数据的量数	(64)
3.3 中位数和四分位距：适于定标数据和定秩数据的量数	(69)
3.4 众数和异众比率：对三个等级数据都适用的量数	(72)
3.5 关于三种集中趋势量数的讨论	(74)
3.6 偏态系数和峰度系数*	(75)
3.7 相关系数	(76)

本章要点	(79)
习题	(82)
第4章 从经验分布到理论分布	(88)
4.1 概率和概率分布	(88)
4.2 二项分布	(91)
4.3 超几何分布	(94)
4.4 正态分布	(96)
本章要点	(102)
习题	(104)
第5章 参数估计	(108)
5.1 点估计与区间估计	(108)
5.2 总体平均数区间估计的原理	(109)
5.3 总体平均数的区间估计(大样本)	(110)
5.4 总体平均数的区间估计(小样本)	(112)
5.5 比率的区间估计	(113)
5.6 总体总量的推断	(113)
5.7 必要样本容量的确定	(114)
本章要点	(116)
习题	(116)
第6章 假设检验原理	(119)
6.1 假设检验引入	(119)
6.2 总体平均数特定值的检验(σ 已知)	(119)
6.3 假设检验中的双向风险和两类错误	(124)
6.4 P值决策法	(125)
6.5 总体平均数特定值的检验(σ 未知)	(127)
本章要点	(128)
习题	(129)
第7章 假设检验应用(上)	(133)
7.1 两总体平均数差量的检验(配对样本)	(133)
7.2 两总体平均数差量的检验(独立样本)	(136)
7.3 多总体平均数齐一性检验*	(141)
本章要点	(146)
习题	(148)
第8章 假设检验应用(下)	(151)

8.1 单一总体比率的检验	(151)
8.2 两总体比率差幅的检验	(152)
8.3 单一总体中位数的检验*	(154)
8.4 独立性检验*	(156)
本章要点	(158)
习题	(159)
第9章 一元回归分析	(165)
9.1 总体简单回归模型和方程	(165)
9.2 样本简单回归方程的拟合	(166)
9.3 估计标准误	(170)
9.4 回归系数的检验	(170)
9.5 因变量数值的估计	(173)
本章要点	(176)
习题	(177)
第10章 多元回归分析	(181)
10.1 总体多元回归模型和方程	(181)
10.2 样本多元回归方程的拟合	(182)
10.3 因变量数值的区间估计*	(186)
10.4 回归方程的评价	(187)
10.5 多元回归方程的单检验和总检验	(193)
10.6 多元回归分析中自变量的选择*	(195)
10.7 非线性回归分析和变量转换*	(197)
本章要点	(199)
习题	(201)
第11章 时间数列	(204)
11.1 时间数列的动态分析	(204)
11.2 时间数列的解析	(209)
11.3 预测*	(224)
本章要点	(232)
习题	(234)
第12章 指数	(239)
12.1 指数的概念和种类	(239)
12.2 指数编制原理	(241)
12.3 指数体系和因素分析	(247)

12.4 指数数列*	(249)
12.5 指数的应用*	(253)
本章要点	(257)
习题	(258)
统计数表	(262)
附表1 二项分布累积概率表	(262)
附表2 普哇松分布累积概率表	(265)
附表3 标准正态分布	(270)
附表4 随机数字 4000 个	(271)
附表5 t 分布	(274)
附表6 χ^2 分布分位数 $\chi^2_{1-\alpha}(n)$ 表	(275)
附表7 F 分布: 显著水平 1% ($\alpha=0.01$) 的临界值	(276)
附表8 F 分布: 显著水平 5% ($\alpha=0.05$) 的临界值	(277)
参考文献	(278)

绪 论

0.1 数据与统计学

在进入本课程之前，同学们已经在中学接触了统计学的初步知识，进入大学后又学习了“概率论和数理统计”。那么，统计学是一门什么样的科学呢？有人概言之：统计学就是数据科学。按着这样的概括方法，也可以说数学是关于数的科学。这两句话虽然不是统计学和数学的定义，但却点出了两者的区别。

数 (numbers) 是表示事物的量的基本概念，如自然数、整数、有理数、实数、复数等。所有这些数都体现为一定的空间形式，存在着这样或那样的关系。数学则研究数的空间形式和数量关系。数据 (data) 是进行社会经济核算、科学研究和技术设计活动所依据的数值。作为数据的数值，一定是与某一具体社会经济核算、科学研究和技术设计活动相联系的，有别于体现抽象的空间形式和数量关系的数。

数学课侧重对体现空间形式和数量关系的概念、定义、公理、定理和公式的讲授，即使有应用题，也是为了加深对所讲原理的认识。统计课关注的数据，寄寓于一定的社会经济核算、科学研究和技术设计活动中，其关于数据的研究完全是为了解决社会经济核算、科学研究和技术设计活动中的问题。

统计学是关于如何对总合现象的测度结果——数据进行采集、整理和分析，以便为决策提供支持的科学。这条定义有以下 4 层含义：第一，统计学是关于数据的科学；第二，它研究数据采集、数据整理和数据分析的方法；第三，数据采集、整理和分析的目的在于获取信息支持决策；第四，数据采集、整理和分析的活动针对总合现象，而非个别现象。在数据采集、整理和分析中需要使用许多统计方法，对这些方法的数学理论依据的论证，属于理论统计学 (theoretical statistics) 或称数理统计学的范畴。而对统计方法在社会经济核算、科学研究和技术设计的各个专门领域的应用的研究，则属于应用统计学 (applied statistics) 的范畴。本书主要讲述在商务领域应用的统计方法。

一次数据采集活动获取的未经整理的数据是原始数据 (raw data)，原始数据录入电子表格后的存在状态有向量和矩阵两种。

【例 0—1】某企业生产特殊用途电台。因为产品批量小、品种批次多，不适

合高度自动化、大规模生产作业，故而采用半自动生产线组装生产，产品质量依靠各类高精度测试仪器设备保证。在实际生产中，电台整机装配合拢后进行逐台调试，确保各项功能达到设计要求。为了提高质量管理水平，公司人事、工程技术部门在新品生产时，都要组织专题技术培训，内容涉及电子线路基础、关键技术原理、生产中可能遇到的技术问题等。

为了评价专题技术培训的效果，公司从 300 名接受过技术培训的调试工中选取 20 名进行测验，每人在规定时间内调试 100 台，记录他们调试过的产品合格数目。具体数据如下：

87	91	86	71	94	85	72	85	36	92
94	96	90	99	90	86	86	88	73	82

将以上 20 个生产记录录入数据文件，无论是按列还是按行录入，都是一个向量。

产品合格数反映的是培训效果，而这些工人参加培训的课时数不一致。公司要研究接受技术培训的时间长短与培训效果的关系，就必须同时采集这 20 人参加培训的课时数。此外，公司还对他们有关的个人背景一并考虑在内。把这诸多内容全都录入数据文件，于是就形成了下列矩阵（见表 0.1）。

表 0.1 20 名调试工的合格产品数和背景资料

行序号	工人编号	性别	年龄	文化程度	培训课时	合格产品数
1	4	男	29	大专	24	87
2	18	女	27	大专	36	91
3	28	男	25	大专	18	86
4	47	男	22	大专	10	71
5	48	男	26	大专	36	94
6	7	女	24	大专	16	85
7	77	男	25	大专	6	72
8	87	男	25	大专	12	85
9	87	男	22	大专	5	36
10	102	男	30	中专	30	92
11	112	男	35	中专	46	94
12	129	男	32	中专	42	96
13	144	女	28	大专	30	90
14	149	女	34	中专	46	99
15	170	女	31	大专	40	90
16	175	男	26	大专	24	86
17	181	男	27	大专	20	86
18	248	女	26	大专	26	88
19	286	女	23	大专	6	73
20	297	男	24	大专	12	82

在表 0.1 中，每名工人占据一行，每项背景资料和合格产品数占据一列，构成矩阵。在电子表格和统计运算中，我们称每一行为一个个体。“性别”、“培训课时”和“合格产品数”等考察项目，具体落实到每一个体上，其文字或数值都不尽相同，因此称每一列为一个变量。因此，我们可以把前面的单纯由 20 个数目字组成的数据称为单变量数据（single-variable data），把表 0.1 这样的“个体×变量”的数据称为多变量数据（multi-variable data）。统计学首先集中对单变量数据进行研究，厘清有关基本概念和基本原理。在单变量数据研究的基础上，再开展多变量数据的研究。

0.2 普查与抽样调查

在例 0—1 中，公司关注的是全部接受技术培训的工人的产品合格率，即合格产品数对受调试产品数（100 台）的比率，以及他们培训效果与接受培训的时间长短及其他背景因素的关系。认识问题的途径是从全部接受技术培训的 300 名工人中，选取 20 名进行测试，记录有关数据。被选取 20 名工人和全部接受技术培训的 300 名工人就构成了样本与总体的关系。如果在 300 名调试工外，还有没接受培训的，则不属于公司在这个问题上的研究范围。这 300 名工人都接受过培训，我们说他们在“是否接受了培训”这个属性上具有同质性，而那些未接受培训的人则与这 300 人不具有同质性。

所谓总体（population），就是在研究中同质个体的集合。组成总体的个体称为单元（element）。每一单元都有一系列的属性，如体现在 300 名工人身上的属性有性别、年龄、婚姻状态、受教育程度、居住地域、业余爱好等。总体是在某一属性相同基础上许多单元的集合，它们在其他属性上则不尽相同，这即是变异性。单元承担的属性很多，研究者关注其中某些属性，就采集这方面的数据，于是有了变量。变量（variable）一词，指的是其数值随单元不同而变化的量，在统计学上它与数据采集工具中的问项相对应，是单元承载的属性在数据分析中所用的概念。

作为同质个体的集合的总体，可以是实实在在的，也可以仅是观念上的。其含有的单元数目可以是有限多，也可以是无限多，前一种情况称为有限总体（finite population），后一种情况称为无限总体（infinite population）。例 0—1 的 300 名工人即是有限总体，而作为环保部门监测对象的空气总体和水流总体则视为无限总体。总体含有的单元数目称为总体总量（population size）。

遵照一定的程序或法则，从总体中选取出来一部分单元，就构成了样本（sample）。通常是选取样本，对其中单元的属性进行测度和研究，间接地认识总

体，这就是抽样调查（sampling）。样本含有的单元数目称为样本容量（sample size）。在特别情况下，对总体所含全部单元的有关属性进行测度和研究，直接地认识总体，这就是普查（census）。

无论是普查还是抽样调查，它们都是从所研究的对象的数量表现出发，经过一系列统计操作，达到认识总体数量特征的目的，只不过是直接认识和间接认识的不同。

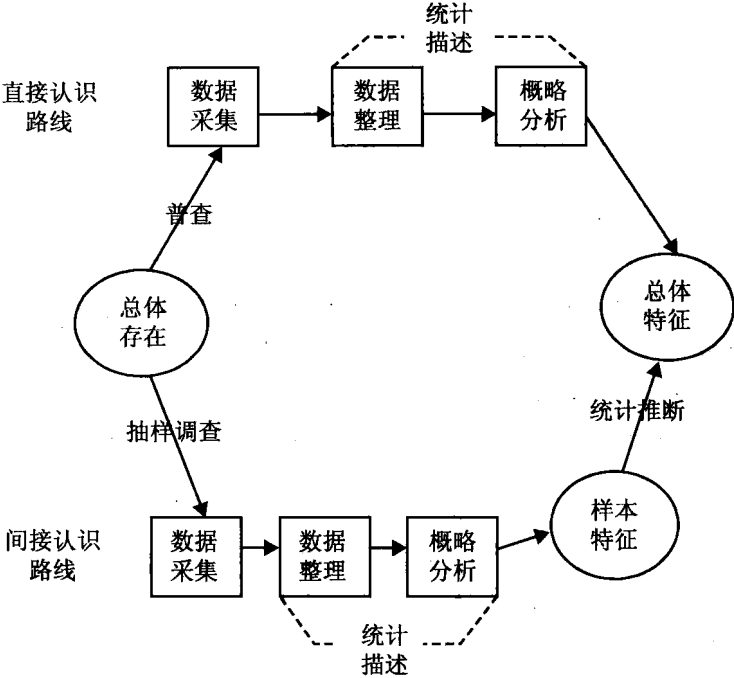


图 0.1 普查与抽样调查的认识程序

就面对的数据——无论是总体数据还是样本数据——进行研究，使用典型数值描述其分布特征，是概略分析（summary analysis）。数据整理和概略分析共同构成统计描述（statistical description）。采用普查的方法，只需经过统计描述，即可达到认识总体特征的目的。而针对样本数据的统计描述，其直接结果是样本特征；从样本特征到总体特征，还需要经过统计推断（statistical inference）。从认识程序看，似乎抽样调查更有难度。但是在实际执行上，由于总体总量庞大，普查的作业量远大于抽样调查，且其中更难于控制作业过程误差，一般不宜采用普查方法。况且耗竭性测度，使普查对象失去存在价值；而对于无限总体，则根本无法实施普查。相反，抽样调查作业量小，便于控制作业过程误差；而抽样的代表性误差，又可以通过其分布规律加以控制。这条间接认识路线有两个关键，

一个是如何抽取样本；另一个是怎样从样本特征推断总体特征，它们都需要以概率和概率分布的原理为保障。

0.3 数据的类别

无论是统计描述还是统计推断，都不是只有一种方法，而是可以根据数据的类别采用不同的方法。

0.3.1 定名数据、定秩数据和定标数据

区分数据类别的一个标准是属性变异的测度层级，有的属性变异只适于简单区分，有的则可以作较为精细的区分。以例 0—1 的属性为例，工人的性别分为男和女，只是划类，属于简单区分。工人的受教育程度有中专和大专之分，这不仅是类的划分，而且有程度的强弱。合格产品数，以具体数值形式出现，不仅可以区分出多少，还精细地计量出谁比谁多几个、少几个。工人的年龄和培训课时数也属于最后这种情况。

如果一个属性只是作类别的划分，那么其相应变量就称为定名变量 (nominal variable)，所测度的数据称为定名数据。属于这类情况的，除了例 0—1 的性别，还有人员的民族、企业的行业类别、就业人员的职业，等等。

如果一个属性不仅可以作类别划分，还可以按强弱、大小排序，那么其相应变量就称为定秩变量 (ordinal variable)，所测度的数据称为定秩数据。属于这类情况的，除了例 0—1 的“受教育程度”，还有教师的职级序列、人们对某一表述从“强烈支持”到“强烈反对”的态度序列，等等。

如果一个属性不仅可以按强弱、大小、多少排序，还可以精确地测度其多少，那么其相应变量就称为定标变量 (metric variable)，所测度的数据称为定标数据。例 0—1 的合格产品数、工人的年龄和培训课时数，即属于这类情况。重量、长度、时间等基本物理量和货币金额等是典型的定标变量。

【例 0—2】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. 格力 2. 美的 3. 海尔 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、在夏天最热的日子里，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。

你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，从统计学的角度看，这几个问题的答案恰是不同层级数据的代表。

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“格力”，“2”代表“美的”，“3”代表“海尔”，“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。

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。

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4 000元与2 000元的差额和2 000元与0的差额相等，还可以说4 000元是2 000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℃到40℃的差距相当于从0℃到20℃的差距的2倍，但是不可说40℃温度是20℃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。定秩数据界于定名数据和定标数据之间，有的时候把定秩数据和定名数据视做同一类型数据，称为质别数据（qualitative data）。相对于质别数据的称谓，定标数据也可以称为量别数据（quantitative data）。在某些情况下，也可以按特定的分界点把定标数据切割成定距数据，如将以金额计量的收入按特定的分界点划分成“低收入”、“中等收入”和“高收入”。

0.3.2 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 0—2 中的对空调售后服务评价的答案 1；2，3，4，5；又如某路口周一至周五每日 14：00~14：05 通过的汽车流量。连续型数据在数轴上体现为无限多的连续点，它们可以充满数轴的某个（些）区间。对于两个离散型数值，由于最小位数到个位，可以判断它们是否相等；对于两个连续型数值，其最小位数不限于个位，而是取决于测度手段精细化程度，因此只可以按四舍五入原则保留至某个数位，然后判断其是否大致相等。

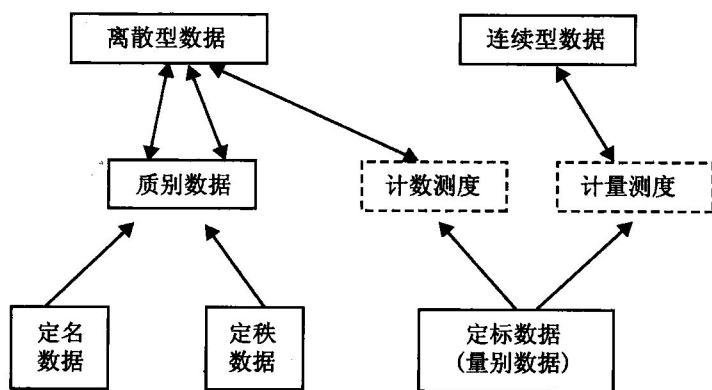


图 0.2 三种数据分类之间的对应关系

由以上关于离散型数据和连续型数据的划分标准可知，定名数据和定距数据均可判定为离散型数据。而定标数据则需要分成两种情况：凡是可计数（be counted）的，如人数、台数、企业数等，都属于离散型数据；凡是须计量（be measured）的，如重量、长度、货币金额等，都属于连续型数据。

本章要点

数据：进行社会经济核算、科学研究和技术设计活动所依据的数值。作为数据的数值，一定是与某一具体社会经济核算、科学研究和技术设计活动相联系的，有别于体现抽象的空间形式和数量关系的数。

原始数据：一次数据采集活动获取的未经整理的数据，其录入电子表格后的存在状态有向量和矩阵两种。