

全国财经专业（新课程标准）精品教材

TONGJIXUE
JICHU

统计学基础

◆ 主 编 王小毅 张继红



浙江工商大学出版社
ZHEJIANG GONGSHANG UNIVERSITY PRESS

全国财经专业(新课程标准)精品教材

统计学基础

TONGJIXUE JICHU

主 编 王小毅 张继红



浙江工商大学出版社
ZHEJIANG GONGSHANG UNIVERSITY PRESS

图书在版编目(CIP)数据

统计学基础 / 王小毅, 张继红主编. — 杭州: 浙江工商大学出版社, 2016. 11

ISBN 978-7-5178-1819-9

I. ①统… II. ①王… ②张… III. ①统计学 IV. ①C8

中国版本图书馆 CIP 数据核字(2016)第 207175 号

统计学基础

主 编 王小毅 张继红

责任编辑 李相玲

封面设计 宣是设计

责任印制 包建辉

出版发行 浙江工商大学出版社

(杭州市教工路 198 号 邮政编码 310012)

(E-mail: zjgsupress@163. com)

(网址: <http://www.zjgsupress.com>)

电话: 0571-88904980, 88831806(传真)

排 版 奥创工作室

印 刷 北京嘉实印刷有限公司

开 本 787mm×1092mm 1/16

印 张 17.5

字 数 448 千

版 印 次 2016 年 11 月第 1 版 2016 年 11 月第 1 次印刷

书 号 ISBN 978-7-5178-1819-9

定 价 38.50 元

版权所有 翻印必究 印装差错 负责调换

浙江工商大学出版社营销部邮购电话 0571-88904970

编委会

主 编 王小毅 张继红

副主编 李 兵 张学虎 高璐璐

白俊艳 何康乐 程 冀

冯 娇 李国彬

前 言

统计学是收集数据、分析数据并解释数据信息的学科,是认识社会的有用工具。它作为一门适用性很强的学科,在社会经济管理领域有着广泛的应用。从 GDP 到 CPI,从股票指数到就业率,生活中人们每天都面对大量的数字信息。学习统计学,不仅有助于学术研究,而且有助于日常应用。

本书内容上本着“实用、够用”的原则,注重实践性、应用性和技能性。全书共分 8 章,分别为统计与统计数据、统计调查、统计整理、静态分析指标、动态分析法、统计指数、抽样推断和相关与回归。除了第 1、2 章外,其他各章后都详细介绍了有关 Excel 的统计应用,以便于大家学习基本的统计电算化技术。本书在内容编排上尽量浅显易懂,一般不做过多的数学推导与证明,因此,学习时不需要太多的数学知识,尽量做到简明扼要,深入浅出。本书着重通过实例讲述统计思想,强调理论联系实际,注重能力的培养,更多的是传播统计思想、学习统计方法和研究统计应用。

在教材体例上,本书各章开篇有本章知识点和能力目标;章末有小结,对有关教学内容做扼要的总结。每章最后有技能训练项目及习题,便于学生检验和巩固所学内容。

本书可以作为高等院校经管类专业的教学用书,也可以作为广大工商企业人员及其他管理人员的参考用书。本书在编写过程中参考了大量的相关资料,也得到了多位教师的支持和协助,在此一并表示谢意。

由于时间仓促,加之作者水平有限,书中不足之处在所难免,恳请专家和学者批评指正。

编 者



目录

Contents

第1章 统计与统计数据	1
1.1 统计学的含义及应用	1
1.2 统计学中的基本概念	4
1.3 数据的类型	8
本章小结	9
本章技能训练项目及习题	10
第2章 统计调查	14
2.1 统计调查的意义和种类	14
2.2 统计调查方案	16
2.3 统计调查的组织形式	19
2.4 统计调查误差及其防止	27
本章小结	28
本章技能训练项目及习题	28
第3章 统计整理	32
3.1 统计整理的内容与技术	32
3.2 统计分组	36
3.3 分配数列	43
3.4 统计资料的表现形式	49
3.5 Excel 的运用	56
本章小结	64
本章技能训练项目及习题	65
第4章 静态分析指标	71
4.1 总量指标	71

4.2 相对指标	74
4.3 平均指标	81
4.4 标志变异指标	90
4.5 Excel 的运用	94
本章小结	100
本章技能训练项目及习题	101
第5章 动态分析法	111
5.1 动态数列概述	111
5.2 动态数列的水平分析指标	115
5.3 动态数列的速度分析指标	121
5.4 动态数列因素分析	128
5.5 Excel 的运用	142
本章小结	149
本章技能训练项目及习题	150
第6章 统计指数	158
6.1 统计指数的意义和种类	158
6.2 综合指数的编制	161
6.3 平均指数的编制	164
6.4 指数体系与因素分析	167
6.5 几种常用的经济指数	176
6.6 Excel 的运用	182
本章小结	185
本章技能训练项目及习题	185
第7章 抽样推断	193
7.1 抽样推断的意义和内容	193
7.2 抽样误差	198
7.3 抽样估计的方法	205
7.4 抽样组织设计	213
7.5 Excel 的运用	221
本章小结	224
本章技能训练项目及习题	225

第 8 章 相关与回归分析	232
8.1 相关分析	232
8.2 一元线性相关分析	237
8.3 估计的线性回归方程	239
8.4 回归估计标准误差	240
8.5 判定系数	241
8.6 利用回归方程进行估计和预测	243
8.7 Excel 的运用	244
本章小结	253
本章技能训练项目及习题	254
附表	264
附表一 标准正态概率较小制累计分布表	264
附表二 标准正态概率双侧临界值表	266
附表三 t 分布临界值表	267
参考文献	268

第 1 章 统计与统计数据

本章知识点

1. 统计学的含义。
2. 统计学的应用领域。
3. 统计学中的基本概念。
4. 统计数据的类型。

能力目标

1. 掌握统计学的基本概念及各有关概念之间的区别与联系。
2. 了解统计在经济生活中的作用。

1.1 统计学的含义及应用

1.1.1 什么是统计学

统计是处理数据的一门科学。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;统计是一门收集、分析、解释和提供数据的科学;统计指的是一组方法,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:

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,统计是用来处理数据的,统计学是关于数据的科学,它提供了一套有关数据收集、数据处理、数据分析、数据解释并从数据中得出结论的方法。统计研究的是来自各领域的数据。

统计提供的数据分析方法可分为描述统计和推断统计两大类。

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。

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。同时,统计学也已发展成为由若干分支学科组成的学科体系。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) 在企业发展战略中的应用。企业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) 在产品质量管理中的应用。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,六西格玛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被广泛用于监测生产过程。

3) 在市场研究中的应用。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,企业就需要做广泛的市场调查,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) 在财务分析中的应用。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。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) 在经济预测中的应用。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做出预测,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,运用统计模型进行预测。

6)在人力资源管理中的应用。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、社会科学等多个领域发展起来的,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,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。对于一名工商管理人员来说,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。

有人认为,统计的全部目的就是让人们看懂数据。实际上,这仅仅是统计的一个方面,统计更重要的功能是数据分析,统计提供了一套分析数据的方法和工具。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的前提。比如,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,否则就无从下手。但是,这种假设必须经过严格的检验后,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,不能指望统计成为解决所有问题的灵丹妙药。

1.2 统计学中的基本概念

1.2.1 总体与个体

总体(population)是指客观存在的、在同一性质基础上集合起来的许多单位的整体。构成总体的这些单位称为总体单位。在表1-1中,总体就是所有制造业企业。

表1-1 制造业上市公司行业平均市盈率统计表

行业名称	公司数量	静态市盈率	动态市盈率
农副食品加工业	39	29.52	25.07
食品制造业	28	39.01	32.89
酒、饮料和精制茶制造业	36	16.09	18.51
纺织业	40	33.55	30.41
纺织服装、服饰业	32	35.72	28.38
皮革、毛皮、羽毛及其制品和制鞋业	6	33.91	35.99
木材加工和木、竹、藤、棕、草制品业	9	85.92	78.43
家具制造业	6	30.64	32.91
造纸和纸制品业	27	50.93	46.26
印刷和记录媒介复制业	7	31.79	29.42
文教、工美、体育和娱乐用品制造业	10	78.46	71.18
石油加工、炼焦和核燃料加工业	19	48.03	68.07
化学原料和化学制品制造业	180	38.13	31.04
医药制造业	145	42.66	38.39
化学纤维制造业	24	46.5	42.25
橡胶和塑料制品业	52	35.34	33.57
非金属矿物制品业	79	23.85	17.84
黑色金属冶炼和压延加工业	31	44.28	35.38
有色金属冶炼和压延加工业	55	39.4	37.67
金属制品业	40	41.68	35.65
通用设备制造业	101	47.06	44.26

续表

行业名称	公司数量	静态市盈率	动态市盈率
专用设备制造业	139	13.26	49.71
汽车制造业	87	19.08	16.4
铁路、船舶、航空航天和其他运输设备制造业	34	44.25	38.11
电气机械和器材制造业	171	30.14	23.18
计算机、通信和其他电子设备制造业	217	53.17	44.36
仪器仪表制造业	28	53.78	48.69
其他制造业	16	39.25	34.73
废弃资源综合利用业	1	91.03	71.77

资料来源:《中国证券报》,2014年12月12日。

确定总体与总体单位,需注意以下两个方面:

1) 构成总体的单位必须是同质的,不能把不同质的单位混在总体之中。例如,研究职工的工资水平,就只能将靠工资收入的职工列入统计总体的范围。同时,也只能对职工的工资收入进行考察,对职工由其他方面取得的收入就要加以排除,这样才能正确反映职工的工资水平。

2) 总体与总体单位具有相对性,随着研究任务的改变而改变。同一单位可以是总体,也可以是总体单位。例如,要了解全国工业企业职工的工资收入情况,那么全部工厂是总体,各个工厂是总体单位。如果旨在了解某个企业职工的工资收入情况,则该企业就成了总体,每位职工的工资就是总体单位。

个体(element)是指构成总体数据的每一个子体。在表1-1中,每一个行业可以分割成一个个体,数据集由29个个体构成。

1.2.2 变量与参数

变量(variable)是统计学研究对象的特征,变量值就是变量的具体表现。例如,表1-1的数据集中有以下3个变量:公司数量;静态市盈率;动态市盈率。

变量按其变量值是否连续,可以分为离散变量和连续变量两种。离散变量的取值是按一定次序逐一列举,通常取整数形式。可以用计数的方法取得数值的表现。如学生数、设备台数、企业数等。连续变量的取值是连续不断的,相邻两值之间可以做无限分割,它必须用测量或度量的方法取得数值的表现。如身高、体重、粮食亩产量、平均温度等都属于连续变量。变量按其性质可以分为确定性变量和随机变量。如果某一变量值能够被另一个变量或若干个变量(因素)的值按一定的规律唯一地确定,则该变量就可以称为确定性变量。例如,在销售价格 P 一定的条件下,某商品的销售额 Y 的变动完全由销售量 X 所确定, Y 就称为确定性变量。所谓随机变量,其数值的变动受到许多种因素的影响,在相同条件下进行观测,由于影响因素的作用不同,其可能的实现值(或观测值)不止一个。数值的大小随机波动,带

有偶然性,事前无法确定。例如,除了正常的、决定性的因素外,影响某工厂生产的同一批次灯泡的质量波动还受许多偶然因素的影响,如果抽取一部分灯泡进行检验。这种灯泡的寿命值不尽相同,数值的大小带有偶然性的波动,检验前是不能预先确定的,则灯泡寿命就是随机变量。随机变量具有随机性或偶然性,但它的数值变动却有一定的规律性,通过大量观察,应用统计技术方法,可以揭示和描述其数量特征以及变动的规律性。

参数(parameter)是描述总体特征的概括性数字度量,它是研究者想要了解的总体的某种特征值。在表1-1中,描述“汽车制造业”公司数量有87家,动态市盈率是16.4,静态市盈率是19.08,这些数字度量都是参数。在一项研究中,收集的个体变量测量值形成的集合,称为一个观测值。例如表1-1中,我们看到的第一个观测值的测量值集合是39,29.52,25.07。29个个体的数据集有29个观测值。

1.2.3 统计标志与指标

1. 统计标志

统计标志简称标志,是指统计总体各单位所具有的共同特征的名称。从不同角度考察,每个总体单位可以有許多特征。例如,每个职员可以有性别、年龄、国籍、服务年限等特征。这些都是职员的标志。

总体单位是统计标志的直接承担者,是载体;统计标志依附于总体单位并说明总体单位的属性和特征。依附于某个总体单位的标志可以有多个。当一个统计标志在各个单位的具体表现都相同时,这个标志称为不变标志;当一个标志在各个单位的具体表现有可能不同时,这个标志称为可变标志或变异标志。

例如,中国第五次人口普查规定:“人口普查的对象是具有中华人民共和国国籍并在中华人民共和国境内常住的人。”按照这一规定,在作为调查对象的人口总体中,国籍和在国境内居住是不变标志,而性别、年龄、民族、职业等则是变异标志。不变标志是构成统计总体的基础,因为至少必须有一个不变标志将各总体单位联结在一起,才能使它具有“同质性”,从而构成一个总体。变异标志是统计研究的主要内容,因为如果标志在各总体单位之间的表现都相同,那就没有进行统计分析研究的必要了。

标志根据其性质分类,可以分为品质标志和数量标志。品质标志表示事物的质的特性,是不能用数值表示的。例如雇员的性别、国籍、工种等。数量标志表示事物的量的特性,是可以数值表示的,如雇员年龄、工资、服务年限等。品质标志主要用于分组,将性质不相同的总体单位划分开来,便于计算各组的总体单位数、计算结构和比例指标。数量标志既可用于分组,也可用于计算标志总量以及其他各种质量指标。

2. 统计指标

统计指标简称指标,是反映同类社会经济现象总体数量特征的范畴及其具体数值。统计指标通常有两种理解和使用方法。一是用来反映总体现象、总体数量状况的基本概念。例如,年末全国人口总数、全年国内生产总值、国内生产总值年度增长率等。二是反映现象

总体数量状况的概念和数值。例如,2010年我国年末总人口数为137 627万人、全社会固定资产投资增长率为13%等。

1) 统计指标按照其反映的内容或数值表现形式,可以分为总量指标、相对指标和平均指标三种。总量指标是反映现象总体规模的统计指标,通常以绝对数的形式来表现,因此又称为绝对数。如土地面积、国内生产总值、财政收入等。总量指标按其反映的时间状况不同又可以分为时期指标和时点指标。时期指标又称时期数,它反映的是现象在一段时期内的总量,如产品产量、能源生产总量、财政收入、商品零售额等。时期数通常可以累积,从而得到更长时期内的总量。时点指标又称时点数,它反映的是社会经济现象在某一时刻上的总量,如年末人口数、科技机构数、公司员工数、股票价格等。时点数通常不能累积,各时点数累计后没有实际意义。

相对指标又称相对数,是两个绝对数之比,如经济增长率、物价指数、全社会固定资产投资增长率等。相对数的表现形式通常为比例和比率两种。

平均指标又称平均数或均值,它反映的是社会经济现象在某一空间或时间上的平均数量状况,如人均国内生产总值、人均利润等。

2) 统计指标按其所反映总体现象的数量特性的不同可分为数量指标和质量指标。数量指标是反映社会经济现象总规模水平和工作总量的统计指标,一般用绝对数表示。如职工人数、工业总产值、工资总额等。

质量指标是反映总体相对水平或平均水平的统计指标,一般用相对数或平均数表示。如计划完成程度、平均工资等。

由于统计指标反映一定社会经济范畴的内容,因此,统计指标的确定,一方面,必须和经济理论对范畴所做的一般概括相符合,要以经济理论为指导,设置科学的统计指标;另一方面,统计指标又必须是对社会经济范畴的进一步具体化,这样才能确切地反映社会经济现象的数量关系。如经济学对劳动生产率做了一般的概括说明,即劳动生产率是表明单位劳动时间所创造的使用价值。但劳动生产率作为一个统计指标时,必须明确规定其劳动时间是指雇员的劳动时间。

3) 统计指标按管理功能作用不同,可以分为描述指标、评价指标和预警指标。描述指标主要是反映社会经济运行的状况、过程和结果,提供对社会经济总体现象的基本认识,是统计信息的主题。例如,反映社会经济条件的土地面积指标、自然资源拥有量指标、社会财富指标、劳动资源指标、科技力量指标;反映生产经营过程和结果的国民生产总值指标、工农业总产值指标、国民收入指标、固定资产指标、流动资金指标、利润指标;反映社会物质文化娱乐设施指标、医疗床位数指标等。

评价指标是用于对社会经济运行的结果进行比较、评估和考核,以检查工作质量或其他定额指标的结合作用。包括国民经济评价指标和企业经济活动评价指标。

预警指标一般是用于对宏观经济运行进行监测,对国民经济运行中即将发生的失衡、失控等进行预报、警示。通常选择国民经济运行中的关键性、敏感性经济现象,建立相应的监测指标体系。

1.3 数据的类型

1.3.1 分类数据、顺序数据、数值型数据

按照所采用的不同计量尺度,我们可以将统计数据分为分类数据、顺序数据和数值型数据。

只能归于某一类别的非数字型数据,称为分类数据(categorical data)。

分类数据是对事物进行分类的结果,该数据表现为类别,是用文字来表述的。分类数据主要是由分类尺度计量形成的。例如,按照性别,人口可分为男性、女性两类;按行业属性,企业可分为医药企业、家电企业、纺织品企业等,这些均属于分类数据。为便于统计处理,我们对于分类数据可以用数字代码来表示各个类别。比如,用“1”表示“男性”,“0”表示“女性”;用“1”表示“医药企业”,“2”表示“家电企业”,“3”表示“纺织品企业”,等等。

只能归于某一有序类别的非数字型数据,称为顺序数据(rank data)。

顺序数据也是对事物进行分类的结果,但这些类别是有顺序的,它是由顺序尺度计量形成的。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,顺序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”。

按数字尺度测量的观察值,称为数值型数据(metric data)。

数值型数据是使用自然或度量衡单位对事物进行测量的结果,其结果表现为具体的数值。在现实中,我们所处理的大多数数据都是数值型数据。

分类数据和顺序数据说明的是事物的品质特征,通常是用文字来表述的,其结果均表现为类别,因而也可统称为定性数据或品质数据(qualitative data);数值型数据说明的是现象的数量特征,通常是用数值来表现的,因此也可称为定量数据或数量数据(quantitative data)。

1.3.2 观测数据和实验数据

按照收集方法的不同,统计数据可以将其分为观测数据和实验数据。

通过调查或观测而收集到的数据,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。

在实验中控制实验对象而收集到的数据,称为实验数据(experimental data)。

比如,对一种新药疗效的实验数据,对一种农作物新品种的实验数据。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

按照所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。

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,2010年我国各地区的国内生产总值数据就是截面数据。

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—2010年我国的国内生产总值数据就是时间序列数据。

下面的图1-1给出了统计数据分类的框图。

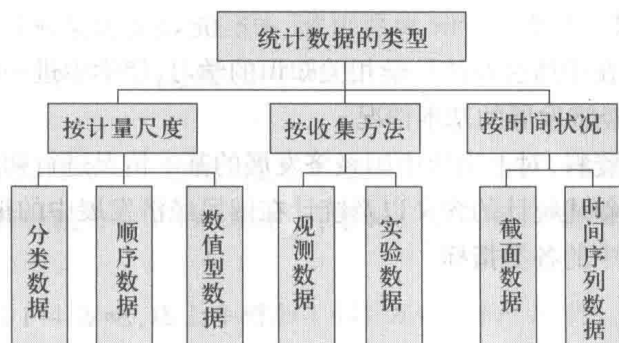


图1-1 统计数据的分类

区分数据的类型是很重要的。因为对不同类型的数据,我们需要采用不同的统计方法来处理和分析。比如,对于分类数据,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;对顺序数据,我们可以计算其中位数和四分位差以及等级相关系数等;对于数值型数据,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。

本章小结

统计学是收集、处理、分析、解释数据并从数据中得出结论的科学。描述统计是指研究数据收集、处理和描述的统计学方法。推断统计是指研究如何利用样本数据来推断总体特征的统计学方法。

按照所采用的不同计量尺度,统计数据可分为分类数据、顺序数据和数值型数据;按照统计数据的收集方法,可分为观测数据和实验数据;按照所描述的现象与时间的关系,统计数据可分为截面数据和时间序列数据。

总体是指客观存在的,由在一个方面或多个方面具有相同性质的许多个别事物构成的整体。总体单位就是构成总体的每一个别事物。总体和总体单位之间会随着研究目的的改变而变换角色。标志是说明总体单位特征的名称,变量是指可变的数量标志。指标是说明