

高等院校经济管理类核心课程系列规划教材

统计学

(第二版)

STATISTICS

主 编 张增臣 王迎春



浙江大学出版社
ZHEJIANG UNIVERSITY PRESS

高等院校经济管理类核心课程系列规划教材

统计学

(第二版)

S T A T I S T I C S

主 编 张增臣 王迎春

图书在版编目(CIP)数据

统计学/张增臣,王迎春主编. —2版. —杭州:
浙江大学出版社, 2015.7
ISBN 978-7-308-14834-4

I. ①统… II. ①张… ②王… 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 (2015) 第 149345 号

统计学(第二版)

主编 张增臣 王迎春

丛书策划 朱 玲

责任编辑 朱 玲

封面设计 周 灵

出版发行 浙江大学出版社

(杭州市天目山路 148 号 邮政编码 310007)

(网址: <http://www.zjupress.com>)

排 版 杭州中大图文设计有限公司

印 刷 浙江海虹彩色印务有限公司

开 本 787mm×1092mm 1/16

印 张 15

字 数 375 千

版 印 次 2015 年 7 月第 2 版 2015 年 7 月第 1 次印刷

书 号 ISBN 978-7-308-14834-4

定 价 32.00 元

版权所有 翻印必究 印装差错 负责调换

浙江大学出版社发行部联系方式: 0571-88925591; <http://zjdxcb.tmall.com>

前 言

统计是人类认识社会最有力的武器之一。统计信息是社会经济信息的主体。统计在认识社会经济现象的数量规律方面始终起着主导作用。

在市场经济条件下,统计信息日趋丰富,统计手段和统计信息的重要性日趋显著;政府部门制订社会经济发展规划,监督检查其执行情况,对社会经济活动过程进行管理和控制,必须充分利用统计手段,以科学充分的统计信息为依据;企事业单位和社会公众从事社会经济活动,同样离不开统计手段和统计信息。可以说,统计无处不在,无时不有。统计的基本理论和方法是从事社会经济管理工作必须具备的基本知识。只有掌握了这种基本的统计知识,及时获取各种统计信息,并进行深入分析研究,才能为做出正确的决策提供依据。

本书系统地介绍了统计学的基本理论和方法。全书共分 11 章,包括总论,统计数据的收集、整理与显示,统计数据分布特征的描述,抽样与抽样分布,参数估计,假设检验,方差分析,列联分析,时间序列分析,相关与回归分析以及统计指数分析。

本书体系合理,内容全面完整,方法简捷实用,深入浅出,通俗易懂。每章前有学习目标,后有综合练习题,可以帮助学生系统地掌握统计的基本理论,熟练地应用统计学的常用方法,提高分析问题和解决问题的能力。本书适合高等院校经济与管理类各专业本科学生使用。

本书由张增臣、王迎春任主编,并负责全书大纲的设计和全书的总纂。各章节的具体编写分工是:张汝飞(第一章、第八章),陈礼丹(第二章、第三章),吉晓东(第四章、第五章),张增臣(第六章、第十一章),谢敬(第七章),王迎春(第九章、第十章)。

本书自第一版出版以来,受到了读者的广泛好评,此次修订,对部分数据做了一些修正,同时补充了一些新的内容。

随着科技的不断进步和统计活动的发展,统计的理论和方法也在不断丰富、完善与更新,因此,对它的认识必然是一个渐进的、不断深化的过程。由于我们自身水平有限,书中难免有错误或疏漏之处,敬请读者批评指正。

编 者

2015 年 6 月

目 录

第一章 总 论	001
第一节 统计及统计学	001
一、统计的含义	001
二、统计学的产生与发展	002
三、统计学的应用	004
第二节 统计学的分科	005
一、描述统计学和推断统计学	005
二、理论统计学和应用统计学	006
第三节 统计数据的种类	006
一、定类数据、定序数据、数值型数据	007
二、截面数据与时间序列数据	007
第四节 统计学的基本概念	007
一、总体、总体单位和样本	007
二、标志和指标	008
三、参数和统计量	009
综合练习题	010
第二章 统计数据的收集、整理与显示	012
第一节 统计数据的收集	012
一、调查数据与实验数据	012
二、统计数据的来源	012
第二节 统计数据的整理	018
一、统计数据整理的概念及意义	018
二、统计整理的步骤	018
三、统计分组	019
四、分配数列	023
第三节 统计数据的显示	030
一、统计表	030
二、统计图	033
三、用 Excel 制作统计图	035
综合练习题	038

第三章 统计数据分布特征的描述 040

第一节 统计数据分布集中趋势的测度 040

一、数值平均数 040

二、位置平均数 045

第二节 统计数据分布离散趋势的测度 048

一、极差 048

二、平均差 049

三、方差与标准差 050

四、离散系数 051

第三节 统计数据分布的偏态与峰度的测度 053

一、偏态 053

二、峰度 055

三、应用 Excel 进行统计量描述 055

综合练习题 060

第四章 抽样与抽样分布 064

第一节 常用的抽样方法 064

一、概率抽样 065

二、非概率抽样 066

第二节 一个样本统计量的抽样分布 068

一、样本均值的抽样分布 068

二、样本比率的抽样分布 071

三、样本方差的抽样分布 072

第三节 两个样本统计量的抽样分布 073

一、两个样本均值之差的抽样分布 073

二、两个样本比率之差的抽样分布 074

三、两个样本方差之比的抽样分布 074

综合练习题 075

第五章 参数估计 079

第一节 参数估计的方法 079

一、评价估计量的标准 079

二、参数估计方法 081

第二节 一个总体参数的区间估计 083

一、总体均值的区间估计 083

二、总体比率的区间估计 086

三、总体方差的区间估计 086



第三节 两个总体参数的区间估计	087
一、两个总体均值之差的区间估计	087
二、两个总体比率之差的区间估计	092
三、两个总体方差之比的区间估计	093
第四节 样本容量的确定	094
一、估计总体均值时样本容量的确定	094
二、估计总体比率时样本容量的确定	094
三、估计两个总体均值之差时样本容量的确定	095
四、估计两个总体比率之差时样本容量的确定	095
综合练习题	096
第六章 假设检验	101
第一节 假设检验的基本问题	101
一、假设检验的基本思路	101
二、假设检验的一般步骤	102
三、利用 p 值进行决策	104
四、假设检验中的两类错误	104
第二节 一个总体参数的假设检验	105
一、总体均值的假设检验	105
二、总体比率的假设检验	107
三、总体方差的假设检验	108
第三节 两个总体参数的假设检验	109
一、两个总体均值之差的假设检验	109
二、两个总体比率之差的假设检验	110
三、两个总体方差比的假设检验	111
四、用 Excel 进行假设检验	111
综合练习题	112
第七章 方差分析	115
第一节 方差分析的基本思想	115
一、方差分析的含义	116
二、方差分析的类型	117
三、方差分析的基本假定	117
第二节 单因素方差分析	118
一、单因素方差分析的步骤	119
二、单因素方差分析表及 Excel 实现	121
第三节 双因素方差分析	123
一、无交互作用的双因素方差分析及 Excel 实现	123
二、有交互作用的双因素方差分析及 Excel 实现	126

综合练习题	131
-------------	-----

第八章 列联分析	134
----------------	-----

第一节 分类数据与列联表	134
一、分类数据	134
二、列联表的构造	134
三、列联表的分布	136
第二节 拟合优度检验	138
一、 χ^2 统计量	138
二、自由度的确定	139
三、拟合优度检验	140
第三节 独立性检验	141
综合练习题	143

第九章 时间序列分析	145
------------------	-----

第一节 时间序列概述	145
一、时间序列的含义及类型	145
二、时间序列趋势因素构成及其分解	148
第二节 时间序列的描述性分析	148
一、时间序列水平分析	149
二、时间序列速度分析	151
第三节 时间序列长期趋势测定	153
一、移动平均法	153
二、线性趋势的测定	156
三、非线性趋势的测定	158
第四节 时间序列季节变动的测定	160
一、简单平均法	160
二、移动平均趋势剔除法	161
三、季节变动的调整	162
综合练习题	163

第十章 相关分析与回归分析	168
---------------------	-----

第一节 相关分析	168
一、相关关系与函数关系	169
二、相关关系的种类	169
三、相关关系的描述与测度	171
第二节 一元线性回归分析	174
一、一元线性回归模型的形式	175
二、最小二乘法的基本思想	175

三、回归系数的最小二乘估计	176
四、最小二乘估计量的性质	177
第三节 一元线性回归方程的统计检验	177
一、回归方程的拟合优度检验	178
二、回归方程的显著性检验	180
第四节 一元线性回归模型的预测	182
一、总体被解释变量均值预测值的置信区间	182
二、总体被解释变量个别预测值的置信区间	182
第五节 多元线性回归	184
一、多元线性回归模型的形式	184
二、多元线性回归模型的基本假定	184
三、多元线性回归模型的参数估计	185
四、多元线性回归方程的统计检验	187
第六节 曲线相关与曲线回归分析	190
一、曲线相关与曲线回归分析的含义	190
二、非线性回归方程的主要函数形式	190
三、曲线回归方程的线性变换	190
四、相关指数	192
五、利用 Excel 进行相关与回归分析	192
综合练习题	196
第十一章 统计指数分析	203
第一节 统计指数的概念、作用与分类	203
一、统计指数的概念	203
二、统计指数的作用	203
三、统计指数的种类	204
第二节 总指数的编制	205
一、综合指数的编制	205
二、综合指数的类型	206
三、平均数指数的编制	208
第三节 平均指标指数	212
第四节 指数体系与因素分析	213
一、指数体系的概念	213
二、指数体系的作用	214
第五节 统计指数的应用	216
一、工业生产指数	216
二、股票价格指数	217
三、居民消费价格指数	218
综合练习题	220

附表 1 标准正态分布表 224

附表 2 t 分布表 225

附表 3 卡方分布表 227

附表 4 随机数字表 228

参考文献 229



第一章 总 论



学习目标

1. 掌握统计的含义,了解统计的产生和发展。
2. 掌握统计数据的种类和适用场合,能够区分定类数据、定序数据、数值型数据、截面数据与时间序列数据。
3. 掌握总体、总体单位、标志、指标、样本、参数和统计量等统计学基本概念。

第一节 统计及统计学

一、统计的含义

学习统计,首先要了解什么是统计,即统计的含义有哪些?根据人们理解的角度不同,“统计”一词一般有三种含义,即统计工作、统计资料和统计学。

统计工作,即统计活动,是对现象总体的数量方面进行统计设计、统计调查、统计整理、统计分析、统计预测和统计决策活动过程的总称。社会经济统计的基本任务是对国民经济和社会发展情况进行统计调查和统计分析,提供统计资料,实行统计监督。

统计资料,即统计信息,是统计工作过程中所取得的各项数据以及与之相联系的情况说明等资料的总称,是统计工作的成果,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。

统计学,即统计理论,是阐述统计工作的理论和方法的科学。

统计工作、统计资料和统计学三者关系密切。统计工作和统计资料是工作过程和工作成果的关系。统计工作的好坏直接影响着统计资料的数量多少和质量好坏。统计工作做好了,才能取得反映社会经济各方面情况的统计资料。统计工作的现代化是关系向社会提供丰富统计信息、提高决策可行性和工作效率的问题。统计学和统计工作的关系是理论与实践的关系。统计理论是统计工作的经验总结和理论概括,只有统计工作有了一定的发展,人们积累了相当的经验,并加以总结提高,才能形成一门独立的统计科学。统计工作的

发展,需要统计理论的指导。统计科学的研究会大大促进统计工作的发展,统计工作的现代化是与统计科学技术进步分不开的。

二、统计学的产生与发展

统计的起源很早,是随着社会生产的发展和国家管理的需要而逐步产生和发展起来的。统计学的产生只有 300 多年的历史。它是在长期统计实践活动的基础上,伴随着资产阶级的古典哲学、古典政治经济学和空想社会主义的产生和发展而产生和发展的。1690 年,英国的威廉·配第《政治算术》一书的问世,标志着古典政治经济学的诞生,同时也标志着统计学的产生。

在统计理论的发展过程中,由于各国的历史背景、经济发展水平与思想渊源的不同,对统计的研究内容和表达方式也互异,因此后来陆续形成了不同的统计学派。

(一)古典统计学(17 世纪末至 18 世纪末)

从 17 世纪末至 18 世纪末,是统计学的萌芽时期,即古典统计学时期。当时有两大学派,即政治算术学派和国势学派。

1. 政治算术学派

政治算术学派产生于英国。主要创始人是英国的经济学家威廉·配第(William Petty, 1623—1687),他的名著有《政治算术》(写于 1671 年,1690 年出版)与《爱尔兰的政治解剖》(1672 年完稿,1691 年出版)。这是配第以数字资料为基础,独创地在研究政治经济问题的同时,简述统计方法论的两部代表作。在著作中,配第还以一系列分析手段和大量计算手段清晰地描述了英格兰、荷兰、法兰西和爱尔兰等地在经济、军事、政治等方面的情况,为英国称霸世界提供了各种有说服力的实证性分析。《政治算术》一书的问世,标志着统计学的初步创立。他所说的“政治”是指政治经济学,“算术”是指统计方法。

配第在典型调查、分组法、平均数、相对数、统计推算与估计、统计分析报告等方面,对统计学都有重要的贡献。例如,他强调统计调查的重要性;初步形成了一套分组方法;在著作中广泛地使用“平均数”来分析现象总体之间的数量关系。这些在当时的确是一种创见。为此,马克思评价他是“政治经济学之父,在某种程度上也可以说是统计学的创始人”。

政治算术学派的另一位创始人是约翰·格朗特(John Graunt, 1620—1674),他的代表作是《关于死亡公报的自然和政治观察》。这是一本关于人口数量变动规律以及如何处理人口统计分析生命表的书,同时在书中提出了统计人口的推算公式。人口统计中有名的性比率,就是他第一个提出来的。他是最早利用统计资料查明统计规律性的人。这是他对统计学的重要贡献。

政治算术学派以后还有一系列的统计学家和统计学著作,但一直未正式采用“统计学”这一科学命名。因此,被认为有统计学之实而无统计学之名。

2. 国势学派

国势学派亦称记述学派,产生于 17 世纪封建制的德国。这个学派把统计学理解为国家重要事实的记述。国势学派的创始人是德国著名学者海尔曼·康令(Hermann Conring, 1606—1681)。在《康令国势学著作集》(1730 年出版)中,他第一次把国势叙述上升到系统化和理论化的高度,以致后来逐渐形成了以国家为研究对象、以记载国家重大事项来形成新知识为目的的国势学。

国势学派继承者主要有戈特弗里德·阿亨瓦尔(Gottfried Achenwall, 1719—1772)和奥古斯特·路德维格·斯勒兹(August Ludwig Von Schlozer, 1735—1809)。阿亨瓦尔对于统计学的研究对象和研究方法,继承和发展了康令的观点,认为统计学是研究“关于国家组织、人口、军队、领土、财产、地面和地下资源等事实的学问”。他在1749年出版的《欧洲最主要各国新国势学概要》中,首创了一个新的德文词汇 Atatistik,即统计学,并用这个名称代替了国势学。斯勒兹的主要著作是于1804年出版的《统计学理论及包括有关政治学研究的想法》。在书中,他企图把统计学与政治学区分开,认为统计学是研究有关国家及其设施、人口及经济等知识的科学,而政治学则是以评价和利用这些知识为任务的科学。主张统计学不仅要研究事物的现状,而且要研究事物的过去。他有句名言:“统计学是静态的历史,历史是动态的统计学。”

国势学派对统计学的创立和发展所做出的贡献有:为统计学这门学科起了至今仍为世界所公认的名词“统计学”,并提出了至今仍为统计学者所采用的“统计数字资料”、“数字对比”等一些术语;认为“对比乃统计方法之母”,即只有在对比分析中才能更清楚地看出事物的规模、结构及其发展变动等。但是这个学派始终没有把数量对比分析作为这门科学的基本特征,而主要以文字来记述国家的显著事项。正因为如此,国势学派被认为有统计学之名而无统计学之实。这个学派也就逐渐成为历史的一个概念而传于后世。

(二)近代统计学(18世纪末至19世纪末)

从18世纪末至19世纪末,是近代统计学时期。这个时期的一个重大成就是大数法则和概率论被引入统计学。这一时期有两大学派,即数理统计学派和社会统计学派。

1. 数理统计学派

数理统计学派的创始人是比利时统计学家、数学家和天文学家阿道夫·凯特勒(Adolphe Jacques Quetelet, 1796—1874)。他一生的著作很多,其中有关统计方面的就有65种之多。其中,流传最广、影响最大的有4种:《论人类》(1835年出版)、《关于概率论》(1846年出版)、《论社会制度》(1848年出版)和《社会物理学》(1869年出版)。《社会物理学》是他的代表作。书中指出,他的社会物理学“是要给政治科学和精神科学加上一种以观察和计算为基础的方法,而支配着社会现象的法则和方法则是概率论”。他强调,统计学或社会物理学的任务,就在于显现与研究这些规律性,也就是要探索在纷纭杂乱的大量偶然现象的背后所隐藏着的必然规律。

凯特勒在统计学中最重要的贡献,是把法国的古典概率引入统计学,提高了统计计量上的准确性,使统计学产生了质的飞跃,为近代统计学奠定了基础。他运用这一统计方法既研究自然,又研究社会,使统计方法发展为既可应用于社会现象研究,又可应用于自然现象研究的通用方法。他是第一个明确提出研究规律性问题的人。他把德国的国势学、英国的政治算术和法国的古典概率论加以统一,改造和融合成具有近代意义的统计学,促使统计科学向新的境界发展。到了凯特勒时代,“统计”一词才由国家统计资料的含义,扩充为近代的统计工作、统计资料和统计学三种意义的“统计”。

凯特勒对于统计学的发展有巨大的贡献,但在理论上也有严重的错误,一是混淆了自然现象与社会现象之间的本质差别,二是扩大了概率论的作用。

在凯特勒的研究基础上,后经高尔顿(F. Galton, 1822—1911)、皮尔逊(K. Pearson, 1857—1936)等统计学家的不断丰富和发展,逐渐形成一门独立的应用数学。高尔顿受凯

特勒的影响和启发,进行了许多研究,相关和回归的概念就是他首先提出来的;皮尔逊发展了相关和回归的理论,又提出假设检验和 χ^2 分布。1867年,威特斯坦(T. Wittstein)发表了题为《关于数理统计学及其在政治经济学和保险学中的应用》的论文,把既是数学又是统计学的新型科学,定名为统计学,也就是今天我们所说的数理统计学。

2. 社会统计学派

社会统计学派的主要代表人物是德国统计学家、政府统计学家和统计教育家恩格尔(C. L. E. Engel, 1821—1896)。恩格尔的主要著作有《人的价值》(1833年出版)、《比利时工人家庭生活费的过去和现在》(1895年出版)和《关于统计学是独立科学还是方法问题》(1851年出版)。恩格尔在统计方面的主要贡献有:在人口调查中提倡用个人调查卡片法,在消费计量上首创消费权数,发明恩格尔法则和恩格尔系数,建设社会统计学体系。

另一位主要代表人物是德国统计学家和政府统计学家梅尔(C. V. Magr, 1841—1925)。梅尔的主要著作有《社会生活中的规律性》(1877年出版)、《理论统计学》(1895年出版)、《人口统计学》(1897年出版)、《伦理统计学》(1909年出版)和《一般的统计文献》(到1906年已出版七卷)。梅尔在《社会生活中的规律性》一书中,有个很精辟的统计学定义,他说:“统计学是在对总体现象大量观察的基础上,对人类社会生活实际状态及其所产生的规律性,做有系统地表述和说明。”他认为,统计学是社会科学中的一门独立科学,是研究规律性的实质性科学。对社会现象表现出来的规律性形式及其性质进行系统阐述,梅尔还是第一人。他对统计方法也很有研究,把统计方法看作是统计学理论部分的一个重要内容。如他认为大量观察法是认识社会唯一可能的统计方法。总之,如果认为恩格尔是社会统计学派的中坚的话,那么梅尔是当之无愧的社会统计学体系的完成者。

社会统计学派主张统计学是实质性的研究社会现象的社会科学,目的在于明确社会现象内部的联系和相互关系。统计方法应当包括社会统计调查中的资料收集、资料整理以及对统计资料的分析研究。在社会统计中,全面调查,包括人口普查和工农业普查,居于重要地位;以概率论为根据的抽样调查,在一定范围内具有实际意义和作用。该学派的前期人物都坚持认为统计学是独立的、实质性的社会科学,数理统计是一门应用数学。该学派的后起者,尽管仍然以研究社会领域的规律为主,但逐步从实质论向方法论转化,强调统计学是调查研究方法,促使德国社会统计学派在20世纪上半叶解体。

(三)现代统计学(19世纪末至今)

从19世纪末到现在,是现代统计学时期。在这一时期,统计学家哥赛特(1876—1937)首创小样本 t 分布理论;统计学家费希尔(1890—1962)提出 F 分布理论和实验设计方法等;波兰统计学家尼曼(1894—1981)提出统计假设检验理论等。现代统计学时期是统计学发展最辉煌的时期。

三、统计学的应用

目前,统计方法已被应用到自然科学和社会科学的众多领域,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。表1-1列出了统计学的一些应用领域,目的是让我们通过简单浏览形成这样一个概念——统计学非常有用!

表 1-1 统计的应用领域

actuarial work(精算)	hydrology(水文学)
agriculture(农业)	industry(工业)
animal science(动物学)	linguistics(语言学)
anthropology(人类学)	literature(文学)
archaeology(考古学)	manpower planning(劳动力计划)
auditing(审计学)	management science(管理科学)
crystallography(晶体学)	marketing(市场营销学)
demography(人口统计学)	medical diagnosis(医学诊断)
dentistry(牙医学)	meteorology(气象学)
ecology(生态学)	military science(军事科学)
econometrics(经济计量学)	nuclear material safeguards(核材料安全管理)
education(教育学)	ophthalmology(眼科学)

统计学是一门应用性很强的学科,由于几乎所有的学科都要研究和分析数据,因而统计学与几乎所有的学科领域都有着或多或少的联系。这种联系表现为,统计方法可以帮助其他学科探索学科内在的数量规律性,而对这种数量规律性的解释并进而研究各学科内在的规律,还需要由各学科的研究来完成。例如,利用统计方法对吸烟和不吸烟者患肺癌的数据进行分析,得出吸烟是导致肺癌的原因之一,但为什么吸烟能导致肺癌?这就需要医学进行解释了。由此可以看出,统计学能做什么和不能做什么。可以这样说,统计方法仅仅是一种有用的定量分析工具,它不是万能的,不能解决你想要解决的所有问题。能否用统计方法解决各学科的具体问题,首先要看使用统计工具的人是否能正确选择统计方法,其次还要在定量分析的同时进行必要的定性分析,也就是要在用统计方法进行定量分析的基础上,应用各学科的专业知识对统计分析的结果做出合理的解释和分析,才能得出令人满意的结论。尽管各学科所需要的统计知识不同,所使用统计方法的复杂程度大不相同,统计学也不能解决各学科的所有问题,但统计方法在各学科研究中发挥的作用已越来越重要。

第二节 统计学的分科

从统计方法的构成来看,统计学可以分为描述统计学和推断统计学;从统计理论研究和统计方法的应用角度来看,统计学可以分为理论统计学和应用统计学。

一、描述统计学和推断统计学

描述统计学(descriptive statistics)研究如何取得反映客观现象的数据并以文字或图表形式对所收集的数据进行加工、处理和显示,通过综合概括与分析描述客观现象的数量特征和规律性。具体内容包括统计数据的收集方法、整理方法、显示方法、数据分布特征的描述与统计分析方法等。

推断统计学(inductive statistics)研究如何根据样本数据去推断总体数量特征,在对样本数据进行描述的基础上,对统计总体的未知数量特征做出以概率形式表述的科学推断。

描述统计学和推断统计学的划分,一方面反映了统计方法发展的前后两个阶段,另一方面也反映了应用统计方法探索客观事物数量规律性的不同过程和地位。

统计研究过程的起点是统计数据,终点是探索出客观现象内在的数量规律性。在这一过程中,如果收集到的是总体数据(如普查数据),则经过描述统计之后就可以达到认识总体数量规律性的目的;如果所获得的只是研究总体的一部分数据(样本数据),要找到总体的数量规律性,则必须应用概率论的理论并根据样本信息对总体进行科学的推断。描述统计学是整个统计学的基础,如果没有描述统计收集可靠的统计数据并提供有效的样本信息,即使再科学的统计推断方法也难以得出切合实际的结论。由于对现实问题的研究获得的数据主要是样本数据,所以推断统计学成为现代统计学的主要内容,在统计学中占据了核心地位。从描述统计学发展到推断统计学,既反映了统计学发展的巨大成就,也是统计学发展成熟的重要标志。

二、理论统计学和应用统计学

理论统计学(theoretical statistics)主要研究统计学的一般理论和统计方法的数学理论。现代统计学用到了几乎所有方面的数学知识,从事统计理论和方法研究的人员需要有扎实的数学基础。此外,由于概率论是统计推断的数学和理论基础,因而从广义上讲,理论统计学也是应该包括概率论在内的。理论统计学是统计方法的理论基础,没有理论统计学的发展,统计学不可能发展成为像今天这样一个完善的科学知识体系。

应用统计学(applied statistics)是研究如何应用统计理论和方法去解决自然、社会领域的实际问题。由于在自然科学及社会科学研究领域中,都需要通过数据分析来解决实际问题,因而统计方法的应用几乎扩展到了所有的科学研究领域。例如,统计方法在生物学中的应用形成了生物统计学;在医学中的应用形成了医疗卫生统计学;在农业试验、育种等方面的应用形成了农业统计学;在经济和社会科学研究领域的应用形成了若干分支学科,如在经济领域的应用形成了经济统计学及其若干分支,在管理领域的应用形成了管理统计学,在社会学研究和社会管理中的应用形成了社会统计学,在人口学中的应用形成了人口统计学等。以上这些应用统计学的不同分支所应用的基本统计方法都是一样的,即都是描述统计和推断统计的主要方法。但由于各应用领域具有其特殊性,统计方法在应用中又形成了一些特点。

第三节 统计数据的种类

统计数据是对现象进行计量的结果。比如,对经济活动总量的计量可以得到国内生产总值(GDP)数据;对股票价格变动水平的计量可以得到股票价格指数数据;对人口性別的计量可以得到男或女这样的数据;等等。由于使用的计量尺度不同,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。

一、定类数据、定序数据、数值型数据

按照所采用的计量尺度不同,统计数据可以分为定类数据、定序数据和数值型数据。

定类数据是对事物进行分类的结果,数据则表现为类别,用文字来表述。它是由分类尺度计量形成的。例如,人口按照性别分为男、女两类;企业按照经济性质分为国有、集体、私营、合资、独资企业等,这些均属于分类数据。为了便于统计处理,对于分类数据我们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表示国有企业,2表示集体企业,3表示私营企业;等等。

定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,产品可以分为一等品、二等品、三等品等;考试成绩可以分为优、良、中、及格、不及格等;一个人的受教育水平可以分为小学、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,对定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。

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

定类数据和定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;数值型数据说明的是现象的数量特征,通常是用数值来表述的,因此也可以称为定量数据或数量数据。

二、截面数据与时间序列数据

按照被描述的对象与时间的关系,可以将统计数据分为截面数据和时间序列数据。

截面数据是在相同或近似相同的时间点上收集的数据,它所描述的是现象在某一时刻的变化情况。比如,2009年我国各地区的国内生产总值数据就是截面数据。

时间序列数据是在不同时间上收集到的数据,它所描述的是现象随时间而变化的情况。比如,1996—2009年我国的国内生产总值数据就是时间序列数据。

第四节 统计学的基本概念

统计中的概念众多,其中有几个概念是我们经常要用到的,在此专门作一下介绍。这些概念包括总体、总体单位、样本、标志、指标、参数和统计量等。

一、总体、总体单位和样本

(一)总体

凡是客观存在的,在同一性质基础上结合起来的许多个别事物的整体,就是统计总体。统计总体简称总体,是从数理统计中借用来的名词,在数理统计中又称母体,与样本相对应。总体是指统计研究所确定的客观对象。例如,我们要研究某市居民户的生活水平,那