



普通高等教育“十三五”规划教材

统计学原理

马立平 张玉春 主编



中国工信出版集团



电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
<http://www.phei.com.cn>



教育“十三五”规划教材

统计学原理

马立平 张玉春 主编

电子工业出版社
Publishing House of Electronics Industry
北京·BEIJING

内 容 简 介

本书是作者在多年统计学教学工作基础上编写出来的,用通俗易懂、深入浅出的语言阐述统计学的基本思想与方法,具有较强的实用性和可操作性。读者通过学习可以掌握统计学中基本的、常用的统计方法的原理与思想,并能够在此基础上正确地使用统计方法进行统计分析与研究。全书共 16 章,分别为导论、数据收集的方法、数据的预处理与分组整理、数据特征的统计量描述、数据资料的图形显示、统计指标与多指标综合评价、概率抽样方法与抽样分布、参数估计、参数的假设检验、方差分析、列联分析与对应分析、相关与回归分析、时间数列的描述性分析、时间数列的构成与预测、聚类分析和判别分析、主成分与因子分析。

本书可作为高等学校统计学基础课程的教材,也可作为感兴趣的读者学习统计学入门知识的参考书。

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。

图书在版编目(CIP)数据

统计学原理 / 马立平, 张玉春主编. —北京: 电子工业出版社, 2018.11

ISBN 978-7-121-34216-5

I. ①统… II. ①马… ②张… 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(2018)第 099218 号

策划编辑: 王二华

责任编辑: 王二华 文字编辑: 蔡馥羽

印 刷: 北京虎彩文化传播有限公司

装 订: 北京虎彩文化传播有限公司

出版发行: 电子工业出版社

北京市海淀区万寿路 173 信箱 邮编: 100036

开 本: 787×1092 1/16 印张: 17.75 字数: 450 千字

版 次: 2018 年 11 月第 1 版

印 次: 2019 年 3 月第 2 次印刷

定 价: 46.00 元

凡所购买电子工业出版社图书有缺损问题,请向购买书店调换。若书店售缺,请与本社发行部联系,联系及邮购电话: (010)88254888, 88258888。

质量投诉请发邮件至 zltts@phei.com.cn, 盗版侵权举报请发邮件至 dbqq@phei.com.cn。

本书咨询联系方式: (010)88254532。

前 言

统计学是收集、分析、表述和解释数据的方法论科学。通过数据资料对现象进行数量方面的分析,不仅能够更客观地认识研究对象,也能使对问题的认识与研究更加深入,统计方法的应用在某种程度上影响着实际工作与理论研究的水平。统计学方法可以应用到几乎各个领域,其既可用于对社会现象数量方面的研究,也可用于对自然现象数量方面的研究,在科学研究与社会经济管理中发挥了重要的作用。相信读者在学习统计学并实际应用之后,能够更深刻地感受到这一点。

我们编写本书的目的是,让读者可以通过学习掌握统计学中最基本的、常用的统计方法的基本原理与思想,并能够在此基础上正确地使用统计方法进行统计分析与研究。本书在撰写时特别注意突出以下几方面的特点。

第一,为使初学统计学的读者尽快了解统计研究的思维方式、掌握统计分析的基本方法并进行应用,本书编写时着重用通俗易懂、深入浅出的语言阐述统计学的基本思想与方法,力求强调统计思想,略掉了一些数学证明和公式推导,在对各种具体方法做必要的阐述之后,配备有实例说明其基本思想与方法应用,并用图表等比较直观的形式进行解释。

第二,突出实际问题分析与统计研究中的工作顺序与知识逻辑关系。本书在结构上共分为5个部分,即统计数据及特征描述、单一指标的统计推断、变量之间关系的统计推断、时间数列分析与预测、统计聚类与数据降维。在内容上不仅包括描述统计,也包括推断统计;从分析方法方面不仅包括传统意义上的统计学原理的基本内容,也包括实际分析研究中常用的多元统计分析的方法与应用,力求使读者学习后能够系统地、较为完整地掌握统计方法论的总体框架,从而在学习了解统计学的基本理论、思想与方法的同时,能够把握何时、用何种方法、对何种问题进行统计分析与研究。

第三,统计学是通过对大量数据的分析与研究发现问题、得出结论的方法论科学,具有很强的应用价值。为了避免读者学完之后只掌握了统计学基本思想与方法而无法实际操作,本书编写时力求具有较强的实用性和可操作性,在介绍方法的基础上,结合统计软件全面、系统地介绍统计计算与分析过程及其技术实现,针对具体问题进行分析并对计算结果进行解读,以提高读者统计分析方法的实际操作能力与水平。

本书由马立平、张玉春编写。本书的编写是建立在作者多年本科及研究生统计学教学经验与体会的基础上的,同时也参考了许多统计学专著和教材,得到了电子工业出版社王二华老师的大力支持和帮助,受到了北京市高层次创新创业人才支持计划教学名师项目的资助,在此一并表示感谢。尽管编写此书时我们投入了许多时间和精力,但书中难免还会存在一些不尽人意之处,真诚欢迎广大读者提出宝贵意见,在此先行表示感谢。

编 者

目 录

第一章 导论	1	第三章 数据的预处理与分组整理	30
第一节 统计学的产生与发展	1	第一节 统计数据的预处理	30
一、统计学的产生	1	一、数据的审核	30
二、统计学的发展	2	二、数据的筛选	30
第二节 统计学的性质与统计方法	3	三、数据的排序	30
一、什么是统计学	3	第二节 数据的分组	31
二、统计学的性质与研究对象	3	一、单一标志的分组	31
三、统计的基本方法	4	二、两标志的交叉分组	36
四、统计学科体系	5	思考与练习	37
第三节 统计的应用领域	7	第四章 数据特征的统计量描述	38
思考与练习	8	第一节 数据集中趋势的测度	38
第一部分 统计数据及特征描述		一、平均数	38
第二章 数据收集的方法	10	二、中位数	42
第一节 数据的来源	10	三、众数	44
一、数据的间接来源	10	四、算术平均数、中位数和众数的 关系	44
二、数据的直接来源	11	第二节 数据离散程度的测度	45
第二节 统计调查方案的设计	12	一、极差与四分位差	45
一、调查目的	12	二、方差与标准差	46
二、研究对象、调查对象和调查 单位	13	三、离散系数	48
三、调查项目和调查表	13	第三节 数据分布形状的度量	49
四、调查时间	14	一、偏态系数	49
五、调查的组织实施	14	二、峰度系数	50
第三节 数据搜集的方法	14	第四节 描述数据特征的统计量的 计算与应用	51
一、数据的搜集方法	14	一、用 Excel 计算	51
二、问卷的设计	16	二、用 SPSS 软件计算	51
第四节 统计数据的质量与类型	22	思考与练习	52
一、统计数据的质量	22	第五章 数据资料的图形显示	54
二、统计数据的误差	22	第一节 定性数据的统计图示	54
三、统计数据质量的检查与要求	24	一、条形图与柱形图	54
四、数据的类型	25	二、帕累托图	55
思考与练习	28	三、饼图	55

四、环形图·····	56	三、抽样分布·····	90
第二节 数值型数据的统计图示·····	57	四、中心极限定理·····	91
一、直方图·····	57	第四节 常用的抽样分布·····	91
二、折线图·····	58	一、样本均值 $\bar{x}$ 的抽样分布·····	91
三、曲线图·····	59	二、样本比例 p 的抽样分布·····	93
四、茎叶图·····	59	三、样本方差的抽样分布·····	94
五、箱线图·····	61	四、两个样本均值之差的抽样分布·····	94
六、雷达图·····	62	五、两个样本比例之差的抽样分布·····	95
第三节 统计图应用中的几个问题·····	63	六、两个样本方差比的抽样分布·····	95
一、合理使用统计图·····	63	思考与练习·····	96
二、统计图的绘制实现·····	65	第八章 参数估计·····	97
思考与练习·····	66	第一节 样本估计量·····	97
第六章 统计指标与多指标综合评价·····	67	一、总体参数与样本估计量·····	97
第一节 统计指标概述·····	67	二、评价估计量的标准·····	97
一、统计指标的概念·····	67	第二节 区间估计的基本思想·····	99
二、统计指标的形成·····	67	一、点估计与区间估计·····	99
三、统计指标的主要类型·····	69	二、区间估计的基本思想·····	100
第二节 指标体系与多指标综合评价·····	71	第三节 一个总体参数的区间估计·····	101
一、指标体系·····	71	一、总体均值的区间估计·····	102
二、多指标综合评价方法·····	72	二、总体比例的区间估计·····	104
思考与练习·····	74	三、总体方差的区间估计·····	105
第二部分 单一指标的统计推断		第四节 两个总体参数的区间估计·····	105
第七章 概率抽样方法与抽样分布·····	76	一、两个总体均值差的区间估计·····	106
第一节 随机变量的概率分布·····	76	二、两个总体比例差的区间估计·····	109
一、随机变量·····	76	三、两个总体方差比的区间估计·····	109
二、离散型随机变量的概率分布·····	76	第五节 样本容量的确定·····	110
三、连续型随机变量的概率分布·····	79	一、影响样本容量的主要因素·····	110
第二节 概率抽样方法·····	85	二、估计总体均值时样本容量的确定·····	111
一、基本概念·····	85	三、估计总体比例时样本容量的确定·····	112
二、简单随机抽样·····	85	思考与练习·····	113
三、分层抽样·····	86	第九章 参数的假设检验·····	115
四、等距抽样·····	87	第一节 假设检验的基本问题·····	115
五、整群抽样·····	88	一、假设检验的基本概念·····	115
六、多阶段抽样·····	89	二、假设检验的基本步骤·····	117
第三节 总体、样本与抽样分布·····	89	三、假设检验中的两类错误·····	118
一、总体分布·····	89	四、假设检验结论的解读·····	119
二、样本分布·····	89	第二节 一个总体参数的假设检验·····	119

一、关于总体均值 μ 的假设检验	119
二、关于总体比例 P 的假设检验	123
三、关于总体方差 σ^2 的假设检验	124
第三节 两个总体参数的假设检验	125
一、两个总体均值之差的检验	125
二、两个总体比例之差的假设检验	130
三、两个总体方差之比的假设检验	131
思考与练习	132

第三部分 变量之间关系的统计推断

第十章 方差分析	136
第一节 方差分析的基本原理	136
一、问题的提出	136
二、方差分析的原理及应用条件	137
三、方差分析中的基本概念	138
第二节 一元单因素方差分析	139
一、数据结构形式	139
二、数据分析与计算步骤	139
三、关系强度的测量与多重比较	142
四、方差分析的软件操作与实现	143
第三节 双因素方差分析	145
一、双因素方差分析及类型	145
二、无交互作用的双因素方差分析	146
三、有交互作用的双因素方差分析	148
四、双因素方差分析的软件操作与实现	150
五、包含协变量的多因子方差分析	151
思考与练习	152

第十一章 列联分析与对应分析	153
第一节 交叉分组与列联表	153
一、交叉分组	153
二、列联表	153
第二节 变量独立性的检验与相关测量	155
一、 χ^2 检验统计量	155
二、变量独立性检验	155
三、列联表中变量的相关度测量	157

四、应用中的准则	158
第三节 对应分析	158
一、对应分析的原理	159
二、对应分析计算机实现与输出结果解读	159
三、对应分析应用——失业原因与教育程度关系分析	162
思考与练习	166

第十二章 相关与回归分析	168
第一节 相关分析	169
一、相关关系	169
二、相关关系的描述——散点图	170
三、相关程度的测定——相关系数的计算	171
第二节 线性回归分析	176
一、线性回归模型	176
二、模型参数估计	177
三、回归系数的含义	178
四、回归方程的评价与检验	179
五、利用回归方程进行预测	183
第三节 可线性化的非线性回归	185
一、可线性化的非线性回归模型	185
二、主要模型及参数估计	185
第四节 相关与回归分析应用中的几个问题	188
一、建立回归模型的基本过程	188
二、解释变量的确定与筛选方法	189
三、带有定性解释变量的回归模型	190
四、回归分析应用——交通事故状况与机动车情况相关分析	191
五、回归分析应用——交通事故损失影响因素的回归分析	193
思考与练习	197

第四部分 时间数列分析与预测

第十三章 时间数列的描述性分析	201
第一节 时间数列及其种类	201
一、时间数列	201
二、编制时间数列的基本原则	201

三、时间数列的种类	202
第二节 时间数列的水平特征分析	202
一、发展水平	202
二、平均发展水平	203
三、增长量	205
四、平均增长量	205
第三节 时间数列的速度特征分析	206
一、发展速度	206
二、增长速度	207
三、平均发展速度与平均增长速度	207
思考与练习	209
第十四章 时间数列的构成与预测	210
第一节 时间数列的构成要素与 模式	210
一、时间数列的构成要素	210
二、时间数列的构成模式	211
第二节 时间数列的长期趋势与 预测分析	214
一、长期趋势的确定——时间数列 的修匀	214
二、长期趋势模型的建立——趋势 线配合	216
三、长期趋势模型的选择	222
第三节 时间数列的季节变动分析	222
一、包含有季节变动的时间数列构成 模型	223
二、季节指数的计算	223
第四节 复合型时间数列的分析与 预测	226
思考与练习	227

第五部分 统计聚类与数据降维

第十五章 聚类分析和判别分析	230
第一节 统计聚类分析	230
一、聚类的基本思想	230
二、距离与相似性度量	232
三、聚类的基本方法	237
第二节 判别分析	243
一、判别分析的基本思路	243
二、判别分析的基本模型	244
三、判别分析的 SPSS 实现	244
思考与练习	248
第十六章 主成分分析与因子分析	250
第一节 主成分分析	250
一、主成分分析的降维思路	251
二、主成分分析的一般模型	252
第二节 因子分析	253
一、因子分析的目的	253
二、因子旋转	254
三、因子得分	255
四、主成分和因子分析的一些注意 事项	255
五、因子分析的 SPSS 实现与输出 结果解读	256
第三节 应用案例：我国工业企业 经济效益评价	260
思考与练习	265
附录 A 常用统计表	266
参考文献	276

第一章 导 论

第一节 统计学的产生与发展

人类的统计活动最早可追溯到远古的原始社会，历史源远流长。可以说，自从人类有了数的概念，有了计数的需求，也就有了统计活动。但是，将统计活动、统计实践上升到理论并形成“统计学”这门科学却是近代的事。虽然对于统计学产生于什么年代这个问题人们的看法不尽一致，但多数人认为，统计学兴起于 17 世纪，距今有 300 多年的历史。

一、统计学的产生

17 世纪中叶，西方社会出现了人们有意识地、比较系统地用数字语言表述问题，并从数量的角度探索客观事物变化规律的研究活动。当时最著名的、最具有代表性的是政治算术学派和国势学派。

政治算数学派的代表人物是英国的学者威廉·配第(W. Petty)。17 世纪 70 年代，威廉·配第的著作《政治算术》问世，书中威廉·配第以劳动价值论理论为基础，对当时的英国、荷兰、法国之间的国情、国力(主要是经济实力)进行了数量上的对比分析，并以此为依据，为当时英国的社会经济的发展出谋划策。这是历史上首次明确地从数量的角度、用大量的数据资料分析问题，揭示了一些经济学的科学原理，研究了许多经济范畴中的经济关系。无论是在古典政治经济学学科，还是在统计学学科的发展史上，《政治算术》都可以称得上是一部具有奠基意义的重要著作，马克思曾称威廉·配第是政治经济学之父，是统计学的创始人。在关于统计史的研究中，人们一般把以威廉·配第为代表的关于社会经济现象的“算术”式的研究，称为“政治算术”。从统计学科看《政治算术》的意义主要表现在人们所研究、关注的问题这方面。威廉·配第在其著作的序言里写道：我进行这种工作所使用的方法在目前还不是常见的，因为我不采用比较级或最高级的词语进行思辨式的议论，相反地，采用用数字、重量和尺度来表达自己的问题的方法。自威廉·配第后的 200 年内，以用数量方法研究社会经济问题为基本特征的政治算术模式，是统计学发展的主流。

除威廉·配第外，政治算术还有一位重要的人物约翰·格朗特(J. Graunt)，他的主要工作是对伦敦市 50 多年的人口出生和死亡数进行计算分析和研究。1662 年约翰·格朗特写出了代表性的著作《关于死亡表的自然观察与政治观察》，该著作通过对人口变动数据的分析，揭示了一系列的人口变化规律。从此，统计的含义从由记述为主，转变为从量的方面说明并分析国家的重要事项，为统计学作为从数量方面认识事物的一种方法开辟了广阔的研究与应用前景。政治算术学派第一次有意识地运用可度量的方法，力求把自己的论证建立在具体的数字基础上，其在统计发展史上具有重要的地位。但是，其毕竟还处于统计核算的初始阶段，从现在的视角看，它只是用最简单的算术方法对社会经济现象进行计量和比较。

统计学的另一个重要起源是概率论。概率论虽然起源于赌博游戏，但真正意义上的概率

论,是从 17 世纪开始的。在早期从事概率论研究的众多学者中,拉普拉斯是古典概率的集大成者,他给出了概率数学的古典解释,建立了严密的概率数学体系。与“政治算术”研究的是简单的、确定的数量关系不同,概率统计研究的则是复杂的、随机性的现象。概率论的出现极大地充实和深化了数量问题研究的内容,以概率论为基础的统计学也很快进入了一个新的发展时期。

二、统计学的发展

凯特勒(A. Quetelet)是统计学发展史上承前启后的重要人物,人们称他为“近代统计学之父”。他在统计学上做出的突出贡献是把概率论全面引进“政治算术”,引入到各种社会经济问题的研究当中,大大推动了概率论和数学方法的应用,促进了数量研究由“算术”水平向“数理”阶段的迅速转化。

自文艺复兴以后,人们已经注意到,当诸如玩纸牌、掷骰子等赌博活动大量进行之后,会出现某种类型的规律性,而概率论最早就是研究这种规律性的产物。当然,概率论的产生和形成在 16~18 世纪,当时与统计学关联性不强,统计学也很少将概率论应用到自己的领域,真正将统计学与概率论结合起来的是凯特勒。凯特勒在自己的研究工作中,首次在社会科学的范畴提出了大数律思想,把统计学的理论建立在大数律的基础上,认为一切社会现象都会受到大数律的支配,他不仅把概率统计的方法引入人口、领土、政治、农业、工业、商业、道德等社会领域,还把概率统计的方法引入天文、气象、地理、动物、植物等自然领域。他的关于概率统计的方法是可以应用于任何事物数量研究的最一般方法的思想,对统计学的发展具有重大意义。

19 世纪后半期,统计学在生物遗传学、农业田间试验等领域都取得了创新性的成果。例如,生物统计学的主创者高尔顿(F. Galton)利用正态法则研究优生学、遗传学等问题,先后提出了“百分位数”、“中位数”、“四分位数差”、相关与回归等概念及计算方法。而皮尔逊(K. Pearson)则系统地发展了高尔顿的相关与回归理论,研究复相关和偏相关,研究出极大似然估计方法,导出了卡方分布。以皮尔逊为代表的统计学家,通过大量观察和以正态分布为基础的关于总体分布曲线的研究,确立了大样本统计理论,从而奠定了描述统计学的框架体系。

进入 20 世纪后,随着新的统计思想和统计方法的大量涌现,带有归纳性质的统计推断逐渐占据了统计学的主流地位。从苏歇米尔斯(J. Sussmilch)提出大数法则开始到 20 世纪初的这段时期,大量观察法一直是统计的核心思想,直到 1908 年戈塞特(W. Gosset)导出了重要的 t 分布,统计学逐渐实现了由描述统计阶段发展到推断统计阶段,实现了由大样本统计向小样本统计理论的转变。费雪(R. Fisher)开辟了方差分析、试验设计等统计分枝,论证了相关系数的抽样分布,提出了 t 检验、 F 检验、相关系数检验等理论与方法,因而在统计学发展史上有着很高的地位。此后,内曼(J. Neyman)和皮尔逊共同完善了现代统计学的核心内容,即区间估计和假设检验的理论,瓦尔德(A. Wald)提出的统计决策理论和质量检验的“序贯分析”,进一步推动了统计学研究和应用的范围。到 20 世纪五六十年代后,随着统计学的发展,稳健统计、时间数列、抽样理论、统计诊断、探索性分析、贝叶斯统计等取得了重要的进展。随着网络信息技术的发展,面对大数据时代,统计学与统计方法将面临着又一次的革命和飞跃。

20 世纪以来,统计学的发展表现出三个明显的趋势:(1)随着数学的发展,统计学依赖和吸收数学方法的程度越来越深入,发展越来越迅速;(2)统计学方法应用领域越来越广泛,向其

他学科领域的渗透越来越深入，以统计学为基础的边缘学科不断形成；(3)随着应用的日益广泛和深入，特别是计算机的发展，统计学发挥了重要的作用，而且将会发挥越来越大的作用。

第二节 统计学的性质与统计方法

一、什么是统计学

提到“统计”这个词，人们自然会将其和统计数字、统计数据联想到一起，如人口总数、国内生产总值、物价指数、收入水平、消费支出等。从统计学的产生和发展可以看到，统计学是随着人类社会的发展和社会管理的需要而不断发展起来的，它是一门以大量现象的数量方面为研究对象的认识方法论科学。

随着统计方法在各个领域的应用，统计学已发展成为具有多个分支学科的大家族，对统计学的认识由于应用和研究的角度不同而不尽相同。正因为如此，即使是统计学家，也对统计学的定义给出了不同的回答，有着不同的表述。在这之中，具有代表性的、被普遍接受的是不列颠百科全书的定义：“统计学是收集、分析、表述和解释数据的科学。”统计学的定义告诉我们，统计学是关于数据的方法论科学，这意味着统计学离不开数据，其可以表现在统计研究过程的各个环节中。

从统计研究的整个过程来看，进行统计研究首先需要搜集能够反映或说明客观现象特征的数据资料，这是统计活动的首要的、也是最基本的环节。没有统计数据，统计方法就失去了用武之地，而如何取得可靠的、高质量的统计数据则是统计学研究的重要内容之一。

有了统计数据后，如何进行数据的分析则是统计学的核心内容。统计分析是对已有的数据资料，通过科学的统计方法探索数据的内在规律、提取有价值信息的过程，其目的是形成一个对研究对象具有概括性的、全面的、整体的数量描述。分析数据所用的方法可分为描述性统计方法和统计推断方法，其中描述性统计是对数据的分布形态、数量特征和随机变量之间的关系等进行估计和描述的方法，主要包括对数据的集中趋势、离中趋势和变量间相关关系等内容的概括性描述方法。而推断统计主要研究如何根据样本数据去推断总体数量特征，它是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推断。当然，分析数据之前，需要对统计数据加工和整理，即数据整理，目的是使统计数据系统化、条理化，以符合统计分析的要求并得到正确的分析结果。统计整理在统计活动中处于承上启下的位置，一方面它是统计搜集资料工作的延续，另一方面又是统计分析的前提，是统计工作的必要环节。数据整理的内容包括：对所搜集的数据进行甄别与筛选、对数据进行审核与修正、对数据进行分组分类、编制数据频数分布表、计算统计指标、数据的图示表现与可视化等，通过数据整理可以帮助分析研究人员发现数据的初步特征，或者方便他人看懂数据所要表达的问题。

数据的解释是对分析的结果进行说明和进一步的分析，阐明分析结果所隐含的事物的特征，从数据中或分析结果里得出关于研究对象发展变化规律性的结论等。

二、统计学的性质与研究对象

统计学的发展历史是从社会经济现象的数量开始的，随着统计方法的不断完善和应用领域的不断拓展，统计学得到了快速的发展。

总体来看,统计学的性质主要表现在以下几个方面。

第一,统计学是一门关于数据的科学,它对客观事物的研究是通过对客观事物数量方面的研究来进行的,包括研究对象的数量状态、数量关系和数量变化规律等,可以说统计学的研究对象是数据。无论是社会科学还是自然科学,只要出现大量数据的地方,就需要统计学。从认识论的角度,任何事物都是数量和质量的统一体,如果数据资料真实、准确、可靠,且统计分析方法运用得当,那么通过客观事物的数量方面特征就可以正确地认识客观事物的特征与其发展变化的规律。

第二,统计学的研究对象是总体,而不是个体;是客观事物的规律,而不是偶然出现的某一现象。统计学是对大量同类现象的数量方面进行描述并对总体进行推断,对单个个体现象的分析与研究不是统计的研究对象。只有通过大量的现象,或对某一现象进行多次重复的观察,才有可能找到统计关系和统计规律。按认识论由个体数量到总体数量的认识逻辑,统计研究的对象是总体的数量方面,但需要从个体的数量方面入手。

第三,统计学的研究对象是不确定现象,表现为随机变量。不确定现象的存在是因为在一些偶然的、随机因素的影响下,客观事物的实际数量表现存在一定程度的不可确知性。现实中,太多的现象都是不确定的现象,如总体上良好的生活习惯可以延长我们的寿命,但具体到某一个具体的人来说,其寿命则是很难预先确定的,可能会出现一个生活习惯不好经常吸烟、喝酒并且不锻炼的人比一个生活习惯良好的人活得还长的现象,当然也有可能短。可以说,人的寿命就是一个随机变量,它除了受到遗传基因的影响外,还会受到生活习惯、生活质量等很多随机因素的影响。但同时我们也知道,虽然一个人的寿命有一定的随机性,但从总体上看,我国公民的预期寿命是非常稳定的,且女性的预期寿命高于男性的预期寿命,这就是在随机性之中的规律性。

第四,统计学的基本方法是归纳推断。统计对总体的认识有两种途径,一是在掌握关于研究对象的全部数据资料的基础上,运用算术方法和统计描述手段就可达到认识总体的目的;另一个是随机抽取样本,并依据样本的数据,利用归纳推断的方法,对总体特征进行推断。从应用的经济性、时效性、实用性和可行性等方面考虑,利用样本数据进行总体推断的优势比较明显。

三、统计的基本方法

统计方法是指统计学研究和认识客观事物总体数量方面的各种方法。从研究主体或工作过程的角度来说,统计学研究要经历资料的搜集、整理和分析的工作过程,或称为统计调查工作、统计整理工作和统计分析工作。在这一过程中的各个阶段,都会有专门的统计学方法。从总体上说,统计学的基本方法有大量观察法、综合指标法和归纳推断法。

(一)大量观察法

大量观察法是统计学研究的基本方法,它是一种从总体出发对研究对象的全部数据或足够多的数据进行观察和分析研究的方法。

大量观察法之所以是统计学研究的基本方法,这是由统计学的研究对象及研究目的所决定的。统计学的研究对象是总体的数量方面,是由大量数据构成的。研究对象的数量方面会受到诸多因素的影响,我们可以将这些因素分为性质不同的两大类:一类源于研究对象的基本性质及一般条件的共同性因素,这类因素对所有个体单位都会发生作用,是研究对象总体

数量规律性存在的基础。另一类源于研究对象的次要性质或偶然因素、随机因素在个体单位上发生的作用,正是由于这类因素的作用,使得各个单位在数量表现上存在差异,各不相同,或多或少地掩盖了研究对象的规律性。大量观察法的意义就在于,在全部或足够多的数据的基础上,去除掉偶然或随机因素的作用,突出共同性因素的作用,从而显示出总体相对稳定的数量特征和数量关系,即数量规律性。

(二)综合指标法

综合指标法是统计学研究中直接表现研究对象总体数量特征的最基本的统计方法、统计手段或工具。综合指标按其一般表现形式可分为总量指标、相对指标和平均指标三大类。其中总量指标的基本来源是对原始数据的整理汇总,以其为基础,可利用多种方法计算出各种派生的相对指标和平均指标。

在统计学的研究中,综合指标有着重要的意义。它不仅可以概括地表明研究对象的规模、总量,也可以表现研究对象的一般水平、内在结构和比例特征,它是对研究对象数量方面的一种测度。在指标的基础上,我们还可以进行统计指标及指标关系的分析,也就是对数量特征、数量关系、数量界限及数量规律性的分析。可以说综合指标在统计分析与研究承担着重要的作用。

(三)归纳推断法

由于种种主、客观方面的原因,我们经常会碰到所研究的对象的范围大于实际可能掌握的范围的情况,这时要认识总体的数量特征,就需要应用统计推断法。统计推断法是指以一定的置信水平,根据样本数据来估计总体数量特征的归纳推断方法,它是现代统计的重要方法。统计推断的一个重要特点是它不能对问题做出绝对肯定的结论,只能在一定的置信水平下,做出能满足研究精度的弹性结论。

四、统计学科体系

目前统计学已经形成了由理论统计学、应用统计学、统计学史等若干分支组成的完整的学科体系。

(一)理论统计学

理论统计学是侧重于研究统计学的方法论和基础理论,以解决统计学学科发展中的重大问题为目标的统计学分支。其最基本内容包括以下几个方面。

(1)统计估计。统计估计是统计学的核心内容之一。它包括两个方面的内容,一是在总体分布已知时,对总体未知参数或参数组合的函数进行估计;二是在总体分布未知时,对有关分布的特征数字及分布密度进行估计;其研究的重点内容包括估计量的确定和对估计量的评价等。

(2)假设检验。假设检验是统计学的另一重要内容。它是指根据样本资料,对总体参数的某种假设命题进行检验和推断。其研究的重点在于检验统计量的构造、假设检验的原理和检验效率等问题。

(3)抽样调查。抽样调查是搜集统计资料的基本手段与方法之一,也是统计学的一个重要分支。其研究的重点在于抽样方案的设计、样本的抽样方法、抽样分布、抽样效果与抽样误差等问题。

(4) 试验设计。试验设计主要研究如何安排试验方案、获取试验数据, 以及如何对试验结果进行分析等问题。

(5) 非参数统计。非参数统计主要研究总体分布未知或不依赖于总体分布及非总体参数的各种统计问题。

(6) 时间数列。时间数列是指按时间顺序排列的一组数据, 时间数列方面的主要研究内容包括有时间数列的基本结构、时间数列的分解、自回归过程与参数估计、非线性系统模型和空间序列分析等。

(7) 统计决策。统计分析与研究的最终目的是在认识客观规律的基础上作出科学的决策, 统计决策部分中的主要内容包括风险函数、损失函数、决策标准和决策函数等。

(8) 序贯分析。序贯分析是指在得出分析结论之前, 视具体的观察结果决定决策方案的选择, 其包括抽样方案、序列检验统计量、判别风险等内容。

(9) 多元统计。多元统计是针对多维随机变量的统计分析方法, 是从经典统计学中发展起来的一个分支。多元统计是当总体的分布是多维(多元)概率分布时, 处理该总体的理论和方法, 它能够在多个对象和多个指标互相关联的情况下分析它们的统计规律, 是一种综合分析方法。多元统计主要包括多元正态分布及其抽样分布、多元正态总体的均值向量和协方差矩阵的假设检验、多元方差分析、判别分析、典型相关分析、主成分分析、因子分析、聚类分析、多元回归等内容。

(10) 统计诊断。统计诊断是近几十年来发展起来的一个统计学的新的领域, 统计诊断主要研究观察数据、统计模型、统计推断方法的合理性问题, 并对诊断中发现的缺陷进行治理和改进。

(11) 稳健统计。稳健统计主要研究当理论分布与实际分布不一致时, 如何确定不敏感的统计方法等。

(12) 探索性分析。探索性分析就是通过对观察数据进行详细的考察, 力求挖掘出数据本身的结构和特征, 然后在此基础上建立分析模型。

(二) 应用统计学

应用统计学就是运用统计思想和方法, 处理实践中遇到的各种问题的理论与方法。应用统计学大体可以分为以下几类。

(1) 统计计算方法。它把统计方法、数学计算方法和计算机应用结合起来, 重点解决数据处理过程中所碰到的各类计算问题。

(2) 应用统计学理论基础。应用统计学理论基础是站在理论统计学角度上的应用统计学, 同理论统计学相比, 应用统计学带有较强的应用背景。

(3) 统计学应用。从实际问题的背景出发, 与具体研究对象所在领域学科紧密结合, 重点在于如何应用统计方法于实践之中。应用统计学按应用的学科性质不同, 可区分为理工科的应用统计学(如统计力学、生物统计学、医学统计学、气象统计学、地理统计学等)和社会科学类的应用统计学(如人口统计学、经济统计学、管理统计学、教育统计学、社会统计学等)。

(4) 统计学与其他应用数学学科的结合, 形成了新的应用数学方法的基础学科, 如博弈论、多目标决策、随机规划等。

第三节 统计的应用领域

目前,统计学的理论与方法已被广泛应用到自然科学和社会科学的众多领域,如在工农业生产和商业活动方面,在社会学和政治学方面,在史学和考古学方面,在物理、化学和生物学方面,在天文地理方面,在交通运输和能源供应方面,在医学和保健方面,在教育和文化方面,在保险和社会福利方面,在自然科学和实验方面等,基本上都要用到统计工具。统计学的理论、方法与相关学科的结合,逐步形成了相关学科的统计学分支,统计学也已经发展成为由若干分支学科组成的学科体系,如经济统计学、生物统计学、物理统计等。可以说,当今年代,无论我们从事什么活动,大多都离不开数据,离不开统计学。即使对于一些复杂现象,我们一时可能难以掌握其变化规律,统计方法也未必是认识和处理问题的唯一途径,但它却可以帮助我们发现随机现象中的必然性,探究隐藏在表面现象背后的规律性,或许也可以给我们提供一些线索,引导我们把研究、分析活动深入下去。

在统计学的应用实践中,经济、管理等领域是统计方法得到较早应用和较多应用的一个领域。经济与管理统计应用中不仅包括宏观领域的经济统计分析,也包括微观领域中的企业商务管理统计的应用等。

在全球商务和经济环境中,有大量的统计信息是可利用的。成功的管理人员和决策制订者一定是那些能够理解信息并有效利用信息的人。而管理统计学,正是运用统计方法,分析和解决企业经营和管理活动中遇到的各种需要解决的问题,帮助做出正确的选择与决策。一般,管理统计中研究的问题主要有市场统计分析、产品试验设计、人员调度、成本预算管理、库存管理、生产控制与管理、风险防范、财务管理等。

统计方法在会计工作中也得到了较好的应用。如会计师事务所想确定列在客户资产负债表上的应收账款是否真正地反映了应收账款的实际金额,就可以用到统计方法。一般应收账款的数量是庞大的,以至于查看和验证每一账户要花费太多的时间和费用。这时,就需要审计人员从账户中选择部分账户作为样本,在对样本账户审查其准确性后,推断并得出有关列在客户资产负债表上的应收账款金额是不是可以接受的结论。在这过程中选取样本、得出结论的过程和方法就需要使用统计学的随机抽样方法和参数估计方法等。

统计方法在理财中的应用。理财顾问是目前一个相对比较新兴的职业,其主要是利用各种各样的统计信息来指导客户、根据客户的实际情况给其相应的投资建议。如在股票投资中,理财顾问要查询阅览各种财务数据,包括股票的价格、盈余比率和股息率,通过对比单个股票的信息和股票市场平均状况信息,得出单个股票的价值是被高估还是被低估,给客户提出买、卖或持有股票的建议等。

统计方法在市场分析中的应用。随着电子技术的发展,超市、专卖店、百货商场等零售企业收款台的电子扫描仪在收款的同时,也起到了收集各种市场调研所需要的各种数据的作用。通过对这些数据的统计分析,可以为商家提供消费者消费行为的基本特征与规律,也可以向制造商出售数据的统计分析摘要,从而提出建议帮助商家和制造商制订关于各种促销的活动计划,诸如特价销售和店内各种商品的陈列方式及位置等。此外,产品经理可以通过对扫描资料和促销活动相关数据资料的统计分析,了解促销活动和销售额之间的关系,掌握促销活动的效果,等等,类似于这种统计分析将有助于各类产品发展战略的制订。

此外,统计方法还可以在产品质量控制中发挥其作用,用于监测生产过程的产出。还有

在对未来的经济状况或对未来经济的某一方面进行预测、判断时，需要使用各种统计信息和统计方法。例如，在预测通货膨胀率时，可以利用有关诸如生产价格指数、失业率、生产能力利用率等统计信息，通过统计模型进行预测。

可以看到，统计学理论与方法是一种重要的、定量分析的方法论与工具，在社会、经济、自然等各领域的研究与实践活动中具有重要的作用，今后也将会发挥出更加重要的作用。但是它不是万能的，不能解决我们想要解决的所有问题，而且如果选用了错误的方法、角度或数据，就可能会得到错误的结论。

能否用统计方法解决具体的问题，首先要看使用统计方法的人是否对所研究的对象有基本的、初步的认识，即定性认识；其次，是否能选择正确的统计方法对可靠的数据进行分析；最后，能否应用研究对象所在学科领域的专业知识对统计分析的结果做出合理的解释和分析。

思考与练习

1. 统计学的研究对象是什么？
2. 统计学的性质是什么？
3. 请举例说明统计方法的作用。
4. 统计学的研究方法主要是什么？

第一部分 统计数据及特征描述

统计的研究对象是数据，数据是基础，数据的类型与数据的质量在统计分析中具有相当重要的作用。针对不同类型的数据，我们可以选用不同的分析方法。数据的分布与特征不仅反映了研究对象的基本状况，而且不同的分布特征也决定了不同的分析方法。以下几章将分别介绍数据的收集方法、数据的预处理与分组整理、数据特征的统计量描述、统计图显示、统计指标与多指标评价等方法。