

应用数理统计

MIAN
XIANG
21
SHI
JI
XIN
XI
GUAN
LI
YU
XIN
XI
XI
TONG
ZHUAN
YE
HE
XIN
KE
CHENG
XI
LIE
JIAO
CAI

主编：孙建军 成颖 / 东南大学出版社

应用数理统计。是信息管理与信息系统专业的
核心课程之一。

本书的主要内容包以下四部分：

- (一) 数据的收集方法；
- (二) 数理统计学中的基本概念；
- (三) 两种重要的统计推断方法：
 - (a) 参数估计，包括极大似然估计和Bayes估计
 - (b) 点估计方法，一般方法的区间估计；
- (四) 假设检验，包括各种单参数和非参数方法的
假设检验及有关分布的假设检验；
- (五) 三种重要的分析方法：回归分析、时间序列分析
以及指数分析。

本书的最后一章还介绍了常用的统计分析软件
SPSS的使用。

应用数理统计

孙建军 成颖 主编

东南大学出版社
· 南京 ·

内容提要

“应用数理统计”是信息管理与信息系统专业的核心课程之一。本书的主要内容包括以下四部分:(1)数据的收集方法;(2)数理统计学中的基本概念;(3)两种重要的统计推断方法:(a)参数估计,包括极大似然估计和 Bayes 估计等点估计方法、一般方法的区间估计;(b)假设检验,包括各种单参数和非参数方法的假设检验及有关分布的假设检验;(4)三种重要的分析方法:回归分析、时间序列分析以及指数分析。本书的最后一章还介绍了常用的统计分析软件 SPSS 的使用。

本书可作为信息管理与信息系统、情报学、图书馆学、信息资源管理、计算机科学与技术等专业的本科生教材,同时,其翔实的资料可作为国内情报学、计算机科学、传播学、社会学等领域学者从事应用统计研究时的重要参考。

图书在版编目(CIP)数据

应用数理统计/孙建军,成颖主编. —南京:东南大学出版社, 2007. 8

ISBN 978-7-5641-0710-9

I. 应... II. ①孙...②成... III. 数理统计—教材 IV. 0212

中国版本图书馆 CIP 数据核字(2007)第 041237 号

应用数理统计

主 编 孙建军 成 颖

策划编辑 张 煦

文字编辑 莫凌燕

装帧设计 王 玥

出 版 人 江 汉

出版发行 东南大学出版社

社 址 江苏省南京市四牌楼 2 号(210096)

经 销 江苏省新华书店

印 刷 南京京新印刷厂

版 次 2007 年 8 月第 1 版 2007 年 8 月第 1 次印刷

开 本 B5

印 张 21.75

字 数 438 千字

印 数 1—4000 册

书 号 ISBN 978-7-5641-0710-9/(I)·46

定 价 34.80 元

前 言

概率论与数理统计是研究随机现象内在规律性的学科。概率论旨在从理论上研究随机现象的数量规律,是数理统计的基础。数理统计则研究如何有效地收集、分析和使用随机性数据,为概率论的实际应用提供了广阔的天地。概率论和数理统计相互依存、相互推动,借助计算机技术,正在科学技术、工农业生产、经济金融、信息管理、环境保护等方面发挥重要的作用。概率统计思想渗入各个学科已成为近代科学发展的明显特征之一。

信息管理与信息系统专业培养的是具备现代管理学理论基础及计算机科学技术知识及应用能力,掌握系统思想和信息系统分析与设计方法以及信息管理等方面知识与能力,能在国家各级管理部门、工商企业、金融机构、科研单位等部门从事信息管理以及信息系统分析、设计、实施管理和评价等方面的高级专门人才。在该专业的培养目标中,信息分析是毕业生的核心能力之一,而概率论与数理统计是培养该能力的基础课程。鉴于此,教育部高等学校管理科学与工程类学科教学指导委员会信息管理与信息系统专业组在清华大学等高校多次召开了核心课程教学大纲研讨会,形成了包括“应用数理统计”等九门课程的教学大纲。大纲要求“应用数理统计”对概率论与数理统计学基础部分需要做准确、全面的论述,注意统计思想和统计方法的介绍,重视数理统计理论与实践的联系。

根据教学指导委员会的意见,作为信息管理与信息系统专业的核心课程,应突出基础理论知识的应用和实践能力的培养。因此,在教材内容的安排上,我们选择以数理统计为主干,适当精简了概率基础部分的教学内容,主要介绍与数理统计方法密不可分的基础知识。对于数理统计部分,也努力做到在讲清基本数学思想、指明基本原理与结论的基础上,以介绍常用的统计推断方法为主,相应的例题的选择均以帮助学生理解概念、掌握常用统计方法为目的。具体来说,本书的主要内容包括以下四部分:

- (1) 数据的收集方法;
- (2) 数理统计学中的基本概念;
- (3) 两种重要的统计推断方法:(a) 参数估计,包括极大似然估计和 Bayes 估计等点估计方法、一般方法的区间估计;(b) 假设检验,包括各种单参数和非参数方法的假设检验及有关分布的假设检验;
- (4) 三种重要的分析方法:回归分析、时间序列分析以及指数分析。

本书的最后一章还介绍了常用的统计分析软件 SPSS 的使用。

本书由孙建军、成颖、苏君华、高凡、刘雅芹、刘慧及程炳娟合作编写,编写提纲由孙建军准备。本书编写任务的具体分工为:第 1 章由孙建军编写;第 2、3 章

由刘雅芹编写;第4、5、6章由高凡编写;第7、8章由苏君华、成颖、刘慧编写;第9、10章由程炳娟编写。全书最后由孙建军修改定稿,苏君华、徐美凤做了大量的校对工作,提出了很多宝贵意见。

在本书写作过程中得到了东南大学出版社张煦老师的大力支持与帮助,在此表示衷心的感谢。

国内外众多的研究为本书提供了良好的基础,本书的顺利完成也得益于参阅了大量相关作者的成果,在此我们向这些文献的作者表示诚挚的谢意。

由于我们的学识、水平有限,书中难免存在一些缺陷以及不足,恳请专家与读者批评指正。

孙建军

2007年8月

目 录

1 绪 论	(1)
1.1 统计与统计学	(1)
1.1.1 统计	(1)
1.1.2 统计学的研究对象	(1)
1.1.3 统计的特征	(2)
1.2 统计学的类别	(3)
1.2.1 理论统计学与应用统计学	(3)
1.2.2 描述统计和推断统计	(3)
1.3 统计中的几个基本概念	(4)
1.3.1 总体和样本	(4)
1.3.2 标志和指标	(5)
1.3.3 变异和变量	(6)
1.4 统计学的研究方法	(6)
2 数据的搜集与分布特征的测度	(9)
2.1 统计数据的搜集	(9)
2.1.1 专门调查	(9)
2.1.2 统计报表	(11)
2.2 调查问卷的设计	(11)
2.2.1 问卷设计的要求	(11)
2.2.2 问卷的基本结构	(12)
2.2.3 问卷中问题的主要类型	(12)
2.3 统计数据的整理	(14)
2.3.1 统计数据的预处理	(14)
2.3.2 统计分组	(14)
2.3.3 频数分布	(15)
2.3.4 统计表与统计图	(17)
2.4 集中趋势的测度	(20)
2.4.1 算术平均数	(20)
2.4.2 调和平均数	(22)
2.4.3 几何平均数	(24)
2.4.4 中位数	(25)
2.4.5 众数	(27)

2.4.6 算术平均数与中位数、众数的比较	(28)
2.5 离散趋势的测度	(29)
2.5.1 极差	(30)
2.5.2 平均差	(30)
2.5.3 方差和标准差	(32)
2.5.4 离散系数	(34)
2.6 数据分布形状的测度	(35)
2.6.1 偏态及其测度	(35)
2.6.2 峰态及其分布	(36)
2.7 案例:某城市著名电子商场电脑、MP3 销售量对比分析	(37)
2.7.1 案例背景	(37)
2.7.2 案例目标	(37)
2.7.3 调查数据的搜集	(37)
2.7.4 数据整理、统计描述及分析	(38)
3 概率及概率分布	(41)
3.1 概率论的基本概念	(41)
3.1.1 随机试验	(41)
3.1.2 样本空间与随机事件	(41)
3.1.3 事件间的关系与事件的运算	(42)
3.1.4 随机事件的概率	(43)
3.2 条件概率与事件的独立性	(45)
3.2.1 条件概率及乘法定理	(45)
3.2.2 全概率公式与贝叶斯公式	(46)
3.2.3 事件的独立性	(47)
3.2.4 贝努利概型	(48)
3.3 随机变量及其分布	(48)
3.3.1 随机变量	(48)
3.3.2 离散型随机变量及其概率分布	(49)
3.3.3 连续型随机变量及其概率分布	(51)
3.3.4 随机变量的数学期望与方差	(55)
3.4 大数定律及中心极限定理	(57)
3.4.1 大数定律	(57)
3.4.2 中心极限定理	(58)
4 抽样分布与估计	(60)
4.1 抽样分布	(60)
4.1.1 概述	(60)

4.1.2 一个总体参数的抽样分布	(67)
4.1.3 两个总体参数的抽样分布	(70)
4.2 参数估计	(72)
4.2.1 概述	(72)
4.2.2 点估计	(74)
4.2.3 一个总体参数的区间估计	(77)
4.2.4 两个总体参数的区间估计	(82)
4.2.5 样本容量的确定	(90)
4.3 其他抽样方法	(92)
4.3.1 简单随机抽样	(92)
4.3.2 分层抽样	(97)
4.3.3 整群抽样	(104)
4.3.4 系统抽样	(110)
4.4 案例分析:大西洋两岸过关机场等级测评	(114)
5 假设检验	(116)
5.1 概述	(116)
5.1.1 统计假设	(116)
5.1.2 假设检验的基本原理	(117)
5.1.3 两类错误与显著性水平	(117)
5.1.4 检验统计量与拒绝域	(118)
5.1.5 P 值决策	(120)
5.1.6 假设检验的步骤	(121)
5.2 一个总体参数的假设检验	(121)
5.2.1 总体均值检验	(121)
5.2.2 总体比例检验	(125)
5.2.3 总体方差检验	(126)
5.3 两个总体参数的假设检验	(127)
5.3.1 两总体均值差检验	(127)
5.3.2 两总体比例差检验	(133)
5.3.3 两总体方差比检验	(135)
5.4 案例分析	(136)
5.4.1 大西洋两岸过关机场等级测评(续)	(136)
5.4.2 建筑景观设计公司人工时间检验	(136)
6 非参数检验	(138)
6.1 χ^2 检验	(138)
6.1.1 皮尔逊定理	(138)

6.1.2 拟合优度检验	(138)
6.1.3 独立性检验	(142)
6.2 符号检验	(144)
6.2.1 单一样本的中位数符号检验	(144)
6.2.2 两个样本的符号检验	(146)
6.3 秩和检验	(147)
6.3.1 威尔科克森秩和检验	(147)
6.3.2 威尔科克森带符号秩和检验	(149)
6.4 游程检验	(151)
6.4.1 检验单一样本序列的随机性	(151)
6.4.2 检验两个样本所属总体是否具有相同的分布	(153)
6.5 两个以上样本的非参数检验	(154)
6.5.1 单向方差秩分析	(154)
6.5.2 双向方差秩分析	(155)
6.6 秩相关检验	(157)
6.7 案例分析: 车祸次数与行车里程独立性检验	(159)
7 回归分析法	(161)
7.1 回归分析法概述	(161)
7.2 一元线性回归分析法	(162)
7.2.1 设定回归方程	(162)
7.2.2 确定回归系数	(164)
7.2.3 相关性检验	(165)
7.2.4 预测及区间估计	(171)
7.3 多元线性回归分析法	(172)
7.3.1 多元线性回归的基本模型	(172)
7.3.2 相关性检验	(174)
7.3.3 区间估计	(175)
7.3.4 实例	(175)
7.4 非线性回归分析法	(179)
7.4.1 可转换为线性回归的曲线回归	(179)
7.4.2 实例	(182)
7.4.3 回归模型的识别和选择	(184)
8 时间序列分析法	(193)
8.1 概述	(193)
8.2 多项式曲线法	(194)
8.2.1 一次曲线	(195)

8.2.2 二次曲线	(197)
8.2.3 三次曲线	(200)
8.3 指数曲线法	(202)
8.3.1 一次指数曲线	(202)
8.3.2 二次指数曲线	(205)
8.3.3 修正指数曲线	(207)
8.4 生长曲线	(211)
8.4.1 逻辑曲线(Logistic 曲线)	(211)
8.4.2 Compertz 曲线(戈珀资曲线)	(215)
8.4.3 生长曲线模型的选择	(218)
8.5 包络曲线	(219)
8.6 移动平均法	(220)
8.6.1 移动平均法概述	(220)
8.6.2 一次移动平均法	(221)
8.6.3 二次移动平均法	(223)
8.6.4 移动平均模型	(225)
8.7 指数平滑法	(226)
8.7.1 一次指数平滑	(226)
8.7.2 二次指数平滑	(228)
8.7.3 三次指数平滑	(229)
8.7.4 初始值的估算	(231)
8.8 周期变动预测	(232)
8.8.1 同月(季)平均法	(232)
8.8.2 季节系数预测法	(234)
9 指数分析	(238)
9.1 指数概述	(238)
9.1.1 指数的概念	(238)
9.1.2 指数的作用	(238)
9.1.3 指数的分类	(238)
9.2 总指数的编制原理及方法	(240)
9.2.1 综合指数的编制原理及方法	(240)
9.2.2 加权平均数指数的编制原理及方法	(243)
9.3 指数体系与因素分析	(245)
9.3.1 总量指数和指数体系	(245)
9.3.2 加权综合指数体系因素分析	(246)
9.3.3 加权平均数指数体系因素分析	(247)

9.4 价格指数的编制与应用	(248)
9.4.1 价格指数的编制	(248)
9.4.2 几种常用的价格指数	(250)
9.4.3 价格指数应用	(258)
9.4.4 价格指数的发展	(260)
9.5 指数应用的实例	(261)
9.5.1 我国经济的宏观形势分析	(261)
9.5.2 南京市居民的生活消费状况分析	(263)
附录 A SPSS 统计分析软件	(265)
A.1 SPSS 软件介绍	(265)
A.2 SPSS 软件的使用	(265)
A.2.1 SPSS 软件的启动、主窗口及退出	(265)
A.2.2 系统主菜单项介绍	(268)
A.3 统计功能及实例分析	(269)
A.3.1 频数(Frequencies)分布图表的绘制及统计分析	(269)
A.3.2 参数检验	(274)
A.3.3 非参数检验	(280)
A.3.4 回归分析	(289)
A.3.5 时间序列的曲线估计(Curve Estimation)及分析预测	(297)
A.4 统计图的绘制及编辑	(300)
A.4.1 条形图(Bar Charts)	(302)
A.4.2 三维柱状图(3-D Bar Charts)	(311)
A.4.3 直方图	(313)
附录 B 常用统计数表	(315)
参考文献	(337)



绪 论

1.1 统计与统计学

1.1.1 统计

人类的统计实践是随着记数活动而产生的。“统计”一词在不同的场合有着不同的含义,分别是统计工作、统计数据、统计学。

统计工作是一种社会实践活动,是对社会、经济、自然现象等的数据进行搜集、整理和分析的活动过程,有着悠久的历史。

统计数据是统计工作的成果,是统计工作过程中所取得的各种有关数字资料以及与之相关的其他资料的总称,它反映现象的规模、水平、结构、比例、速度及预测的数字和文字信息等,表明现象发展的特征。它的表现形式有各种统计报表、统计图、统计报告、统计年鉴和其他有关的统计数字信息载体等。统计工作的质量直接影响着统计数据的质量。

统计学即统计理论,是研究如何收集数据、整理数据和进行数据分析推断的一门方法论科学,其目的是探索事物数据的内在数量规律性,以达到对客观事物的科学认识。它是统计工作的科学总结和理论概括,反之,统计学又以其理论和方法指导统计工作。统计数据的收集是进行统计分析的基础,如何取得准确、可靠的统计数据是统计学研究的内容之一;统计数据的整理是对数据进行加工处理,使得统计数据系统化、条理化、符合统计分析的需要;统计数据的分析是统计学的核心内容,其目的是探索数据的内在规律。

因此,统计工作、统计数据、统计学三者密切联系。统计工作是对社会进行调查研究的过程,统计工作的成果是统计数据。而统计学和统计工作的关系是理论和实践的关系,统计学是统计工作的科学总结和理论概括;反过来,统计学所论述的理论和方法又指导着统计工作的开展。

1.1.2 统计学的研究对象

辩证唯物主义告诉我们,不论是自然现象还是社会现象,都存在质与量的两个方面,二者是辩证统一、密切联系的。事物的质是通过量表现出来的,量的积累达到一定程度将引起质的变化。因此要认识客观事物、研究事物的发展规律,必须从事物的量的角度出发,研究事物发展的规律性在一定时间、地点、条件下的数量表现。

统计学研究事物发展过程中表现在数量方面的特征,通过数量方面的特征

进而认识事物本身的特征。其特征是在质与量的辩证统一中研究大量现象的总体数量方面,反映现象发展变化的规律性在具体时间、地点和条件下的数量表现,揭示事物的本质、相互联系、变动规律性和发展趋势。因此,利用统计研究事物的发展,必须把量和质两方面结合起来。即首先要明确事物的质的特征,然后才能正确地研究其量的表现。例如要研究人口增长率,就必须首先明确什么是人口增长率、涉及哪些数据、应该如何计算等质的因素。

1.1.3 统计的特征

统计学的研究对象主要是社会经济现象的数量方面,它具有数量性、总体性、具体性等特征。

1) 数量性

统计学的研究对象包括社会经济现象的规模、水平,现象之间的数量关系以及决定现象质量的数量界限。它是运用大量的数据资料来综合反映事物的数量方面的。这主要是由于客观事物中个体与个体之间的差异是普遍存在的。例如在同一环境下,拥有相同物资、资金等条件的企业,其经营效果各不相同。同时,事物是不断发展的,在其发展过程中,随着各种条件的发展变化,影响其发展的因素也在不断变化。因此,少量的数据并不能充分表现事物的一般特征,不能说明事物发展的规律性,必须收集大量的数据资料进行综合分析,才能消除事物发展过程中偶然因素的影响,得到事物发展的一般规律。

2) 总体性

由于社会经济现象错综复杂,各个个体处在不同的环境之中,它们的发展过程既要受到共同因素的影响,又受到某些个别、偶然因素的影响。因此,只利用个别个体现象的数量特征和变动趋势难以说明社会经济现象总体的发展规律,必须以社会经济现象的总体作为研究对象,才能消除各种偶然因素的影响,从而正确地反映社会经济现象的发展规律。可见,统计研究对象不是个体现象的数量方面,而是由多个个体构成的总体现象的数量方面。

总体是由个体组成的,要认识社会经济现象的总体,就必须从每个个体现象着手,获取数据,然后综合、分析,最后获得总体现象的规律性。

3) 具体性

统计学所研究对象的量是具体的而不是抽象的数量,这是统计学和数学的重要区别。即统计学的数量是具体的,它反映社会经济现象在一定时间、地点、条件下的具体的数字表现;数学则完全撇开研究对象的具体内容和质的特征,主要研究抽象的数量关系。

1.2 统计学的类别

1.2.1 理论统计学与应用统计学

理论统计学是指统计学的数学原理,主要研究统计学的一般理论和统计方法,是关于数据的收集、整理、分析的最基本的原理、原则和方法。应用统计学是运用于某一特定领域的统计理论和方法。由于各个领域在研究和实际工作中都要通过数据来分析和解决问题,统计理论和方法的应用就自然而然地扩展到了几乎所有的研究领域,形成了各种应用统计学,如生物统计学、经济统计学、金融统计学、管理统计学等等。

理论统计学的目的是为适应对现实的客观资料进行统计分析研究的需要,完善、发展和创新统计理论和方法。而应用统计学与被研究现象的客观数据资料以及其内在的规律性有关,因此,应用统计学研究人员不仅必须具备统计学的一般原理及分析研究方法的知识,还必须同时具备所研究现象方面的专业知识。

1.2.2 描述统计和推断统计

描述统计研究的是统计数据的搜集、整理、分析、概括和表述的方法。具体地说,描述统计的内容包括:获取反映客观现象的统计资料,通过统计图表的形式对所采集的统计资料进行加工处理和显示,进而通过综合、概括与分析等过程,得出反映客观现象的规律性数量特征。

推断统计研究和论述的是如何根据样本数据去推断总体数量特征,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的以概率形式表述的判断。

对于调查总体而言,进行全面调查就可以得到总体的资料,经过对总体资料的统计描述,就可以反映出总体的数量特征。如果没必要对总体进行全面调查,或者总体范围过大不适于进行全面调查,就可以通过抽样调查获得样本数据,通过计算样本的统计量,运用推断统计的方法对总体的参数进行估计和检验。如果总体中样本的信息已经取得,所要做的推断是对总体的某个数值做出估计,这样的问题属于估计这一类型;如果所做的推断是在几个可供选择的行动方案中进行选择,这样的问题属于检验这一类型。

描述统计的目的是通过对客观事物的统计数据进行调查采集、加工整理、分析概括,从而研究其总体的数量规律性。推断统计的目的是通过科学的方法从总体中抽取样本,应用科学的计算方法得到样本统计量,利用样本统计量及相应的统计分析方法对统计总体参数及数量特征、统计规律性等进行估计和检验。

由于样本中所包含的总体信息并不完备,据此推断总体必然产生一定的误

差,而且由于样本是随机抽取的,推断的结论是否可靠也不确定。为了保证推断的精确度和可靠度,推断统计就需要研究样本的抽取方式、样本的大小、样本的分布、样本估计量的选择、对总体特征进行推断的方法、误差的计算和控制等问题。

描述统计学和推断统计学是统计方法的两个组成部分。描述统计学是整个统计学的基础,推断统计学是现代统计学的主要内容。由于在现实问题的研究中,所获得的许多统计数据是样本数据,因此,推断统计学在现代统计学中占有重要的地位。但我们不能把描述统计学与推断统计学截然分开,没有描述统计搜集到的准确的统计数据、没有描述统计提供有效的样本信息,推断统计不可能得出准确的结论。描述统计不仅为推断统计提供统计信息,而且描述统计中也时常有推断统计的内容。

1.3 统计中的几个基本概念

1.3.1 总体和样本

统计学研究大量自然和社会经济现象的数量方面,它是从现象的整体着眼,研究现象总体的数量特征,并通过对总体数量特征的分析,进而认识现象的规律性。所以总体是统计工作的基本概念。

总体就是统计所要研究对象的全体,它是由客观存在的、具有某种共同性质的许多个体所组成。

总体具有三方面的特征:

(1) 同质性

同质性是指构成总体的各个个体具有某一方面的共性,否则就不能成为一个统计总体。

(2) 大量性

总体是由多个个体所组成的,少数个体或个别事物不能构成统计总体。统计的大量性与个体的差异性密切联系。由于统计的目的是要揭示隐藏在事物表象中的内在规律性,我们必须通过对大量个体的观察、分析和研究,才能发现其中的规律性。同时,总体的大量性也与研究目的的要求有关,精确度要求越高,个体数也应该越多。

(3) 差异性

总体的差异性是指构成总体的个体除了在某一方面具有共性外,在其他方面还存在差异,这些差异是统计研究的基础。

总体单位随着研究目的的不同而不同,可以是人、物,也可以是组织、地域,甚至可以是状态、长度和时间等等。总体根据其所包含的个体数目是否可数可

以分为有限总体和无限总体。有限总体的范围确定,而且个体的数目是有限的。比如,一定时期内城市的人口总数、建筑物总量,一批待检验的计算机等构成的总体都是有限总体。无限总体是指总体所包括的个体是无限的、不可数的。例如连续大量生产的螺丝钉的总产量。

样本是从总体中抽取出来作为这一总体代表的部分个体所构成的集合。从总体中抽取样本的目的就是根据样本所提供的有关信息去推断总体的特征。

样本也具有三个特点:

(1) 代表性

抽取样本的目的是用来推断总体的,所以所抽取的样本必须能够充分地代表总体。

(2) 客观性

抽取样本的过程中,必须排除主观因素的影响。也就是说不管采用何种抽取方式,都必须保证取样的客观性,即样本的选取不受调查者或被调查者的主观影响。

(3) 随机性

在一个总体中可以抽取多个样本,每一个个体是否被选中都是随机的。

1.3.2 标志和指标

标志是说明总体单位属性和特征的名称。标志根据其性质可以分为品质标志和数量标志。品质标志表明个体质的特性,只能用文字说明。数量标志表示个体量的特性,用各种数值来表示。一个总体可能既有品质标志又有数量标志。

标志按变异情况可分为不变标志和可变标志。不变标志指总体中各个个体该标志的具体表现都相同,是总体同质性的依据。可变标志是指总体中各个个体之间该标志的具体表现具有差别,是统计所要研究的内容。

指标即统计指标,是说明现象总体数量特征的概念。通过对总体中各个体的标志值进行采集、整理、汇总,形成反映总体数量特征的综合指标,是统计反映事物数量表现的基本工具。

指标具有以下三个特点:

(1) 具体性

指标是总体在一定条件下客观存在的数量反映,它表明总体在一定条件下所达到的规模和水平。

(2) 综合性

总体的数量特征是通过个体的标志表现来反映的,各个个体的数量标识可以直接汇总成总体的统计指标。

(3) 数量性

指标反映的是各种现象的量,所以指标必须有具体的指标数值,不能用数字

表现的概念不能作为统计指标。

一个指标只能反映总体的某个方面的特征,要全面认识总体,就必须用一系列的指标从不同侧面来反映总体的数量特征。由一系列相互联系的指标组成的整体称为指标体系。

标志和指标是两个既有区别又有联系的概念。两者的区别主要是:① 指标说明的是统计总体特征,而标志说明的是构成统计总体的个体特征;② 标志有不能用数值表示的品质标志与能用数值表示的数量标志,而指标都是用数值表现的。其联系表现在:① 统计的数量标志是许多统计指标赖以建立的基础。有些统计指标的数值是从总体单位的质量标志值汇总得到的;② 由于研究目的的不同,许多统计指标与统计数量标志之间存在着变换关系。

1.3.3 变异和变量

可变标志在各个个体之间的具体表现通常是不同的,这种可变标志的差异统称为变异,包括属性的差别和数量的变化。如性别表现为男、女;教师职称表现为助教、讲师、副教授、教授;年龄表现为25岁、30岁、32岁、41岁等。变异是统计的前提条件,统计正是要对这些变异情况进行具体研究,从中得出规律性的认识,没有变异就用不着统计了。

变量就是可变的数量标志。比如产品销售额、人均工资水平等。变量的具体表现称为变量值。比如产品销售额可以是10万元、15万元、21.5万元等,人均工资水平表现为1000元、1200元等。

变量按照其取值是否连续,可分为离散性变量和连续性变量。离散性变量的变量值一般只表现为整数,如在校学生数、图书销售量等。连续变量的数值是连续不断的,两个相邻的整数之间可以插入任意的小数,如长度、高度、面积等。

变量按照所表现现象的类型不同,可以分为分类变量、顺序变量和数值变量。分类变量是说明事物类型的一个名称,如性别,其变量值表现为“男”或“女”。顺序变量是说明事物有序类别的一个名称,如产品等级,其变量值表现为“一等品”、“二等品”、“三等品”等。数值变量是说明事物数字特征的一个名称,如产品销售额、仓库面积、城市人口数等。上述的离散性变量和连续性变量都属于数值变量。

1.4 统计学研究方法

统计学的研究对象和性质决定了统计学研究方法。而正确的统计研究方法又是完成统计任务的重要条件。方法问题在统计研究中具有重要的地位。没有一套科学的统计方法,就不可能准确、及时、全面、系统地揭露社会经济现象的数量方面,更不可能由此反映社会经济现象发展的规律性。