



中等职业教育国家规划教材
全国中等职业教育教材审定委员会审定

Statistics

抽样技术

(统计专业)

主编 吴桂英



209

中等职业教育国家规划教材
全国中等职业教育教材审定委员会审定

C811-43

W81

抽 样 技 术

(统计专业)

主 编 吴桂英
副 主 编 陈小娴
责任主审 张昌法
审 稿 张昌法 王 虹



A0977203

高 等 教 育 出 版 社

内 容 简 介

本书是根据教育部 2001 年颁发的《中等职业学校统计专业教学指导方案》中主干课程《抽样技术教学基本要求》编写的中等职业教育国家规划教材。

本书的编写立足于教育部关于“培养与社会主义现代化建设要求相适应,德、智、体、美等方面全面发展,具有综合职业能力,在生产、服务、技术和管理第一线工作的高素质劳动者和中初级专门人才”的培养目标。力求做到适应市场经济发展的需要,突出体现中等职业教育的特点,紧贴中等职业学校学生的实际水平,通俗易懂,深入浅出地论述统计抽样技术的基础理论和实践技能,注重学生能力的培养,每章都配有案例教学,使所学知识与统计实践紧密结合,以满足用人单位对相关知识层面的一线统计工作人员的要求。

本书共分六章:第一章抽样技术概述;第二章抽样调查方案设计;第三章抽样误差;第四章抽样方法;第五章参数估计;第六章抽样技术应用案例。本书在编写上,充分考虑了读者的需求,具有指导思想明确、程度适宜、可操作性强、灵活适用等特点。

本书可作为中等职业学校学生学习统计抽样技术知识的用书,也可作为培训统计抽样调查人员和日常生活中人们了解抽样技术知识的参考用书。

图书在版编目(CIP)数据

抽样技术/吴桂英主编. —北京:高等教育出版社, 2002.7

中等职业教育教材

ISBN 7-04-011008-3

I. 抽... II. 吴... III. 抽样调查—专业学校—教材 IV. C811

中国版本图书馆 CIP 数据核字 (2002) 第 044552 号

抽样技术

吴桂英 主编

出版发行 高等教育出版社

社 址 北京市东城区沙滩后街 55 号

邮政编码 100009

传 真 010-64014048

购书热线 010-64054588

免费咨询 800-810-0598

网 址 <http://www.hep.edu.cn>

<http://www.hep.com.cn>

经 销 新华书店北京发行所

印 刷 中国农业出版社印刷厂

开 本 787×1092 1/16

印 张 12

字 数 300 000

版 次 2002 年 7 月第 1 版

印 次 2002 年 7 月第 1 次印刷

定 价 14.70 元

本书如有缺页、倒页、脱页等质量问题,请到所购图书销售部门联系调换。

版权所有 侵权必究

中等职业教育国家规划教材

出版说明

为了贯彻《中共中央国务院关于深化教育改革全面推进素质教育的决定》精神,落实《面向 21 世纪教育振兴行动计划》中提出的职业教育课程改革和教材建设规划,根据教育部关于《中等职业教育国家规划教材申报、立项及管理意见》(教职成[2001]1 号)的精神,我们组织力量对实现中等职业教育培养目标和保证基本教学规格起保障作用的德育课程、文化基础课程、专业技术基础课程和 80 个重点建设专业主干课程的教材进行了规划和编写,从 2001 年秋季开学起,国家规划教材将陆续提供给各类中等职业学校选用。

国家规划教材是根据教育部最新颁布的德育课程、文化基础课程、专业技术基础课程和 80 个重点建设专业主干课程的教学大纲(课程教学基本要求)编写,并经全国中等职业教育教材审定委员会审定。新教材全面贯彻素质教育思想,从社会发展对高素质劳动者和中初级专门人才需要的实际出发,注重对学生的创新精神和实践能力的培养。新教材在理论体系、组织结构和阐述方法等方面均作了一些新的尝试。新教材实行一纲多本,努力为教材选用提供比较和选择,满足不同学制、不同专业和不同办学条件的教学需要。

希望各地、各部门积极推广和选用国家规划教材,并在使用过程中,注意总结经验,及时提出修改意见和建议,使之不断完善和提高。

教育部职业教育与成人教育司

二〇〇一年十月

前 言

本书是根据教育部 2001 年颁发的《中等职业学校统计专业教学指导方案》中主干课程《抽样技术教学基本要求》编写的中等职业教育国家规划教材。

本书的编写人员来自于教学第一线,多年来从事相关学科的教学工作,既有较广博的业务理论知识,又有丰富的教学经验。不仅对教学内容吃得透,而且熟悉教育对象的实际状况,有深入浅出讲授基本理论的能力和实践经验,又有较高的教科研水平和一定的教材编写能力。

参加本教材编写的有:吴桂英(第一、六章)、陈小娴(第二章)、何念荣(第三章)、王小侠(第四章)、娄庆松(第五章)。由吴桂英任主编,陈小娴任副主编。

在本书的编写过程中,得到了国家统计局孙继伟,北京市统计局阎明、吕凤梅,以及北京市现代职业学校领导的大力支持,中国人民大学张尧庭教授等专家提出了宝贵意见,在此谨致以诚挚的谢意。

本书通过全国中等职业教育教材审定委员会审定,由首都经济贸易大学张昌法教授担任责任主审,张昌法教授和北京工业大学王虹副教授审稿。他们对书稿提出了很多宝贵意见,在此表示衷心感谢。

由于编者水平有限,错误之处在所难免,敬请读者批评指正。

编 者

2002 年 4 月

目 录

1	第一章 抽样技术概述
1	第一节 抽样技术的特点和作用
6	第二节 抽样技术的理论依据
11	第三节 抽样技术的几个基本概念
17	技能实训
17	复习思考题
19	第二章 抽样调查方案设计
19	第一节 抽样调查的一般程序
22	第二节 抽样调查方案的内容
34	第三节 抽样调查的评价
41	案例分析
55	技能实训
56	复习思考题
57	第三章 抽样误差
57	第一节 调查误差与抽样平均误差
68	第二节 抽样极限误差
74	案例分析
77	复习思考题
78	第四章 抽样方法
78	第一节 简单随机抽样
90	第二节 分层抽样
98	第三节 整群抽样
101	第四节 系统抽样和多阶段抽样
104	案例分析
112	技能实训
125	复习思考题
131	第五章 参数估计
131	第一节 总体参数估计
135	第二节 参数的区间估计

138	第三节 几种常用抽样方法的参数估计
147	复习思考题
150	第六章 抽样技术应用案例
150	案例一 2001 年人口变动情况抽样方案
161	案例二 美国盖洛普民意测验机构的抽样设计
163	案例三 我国城市住户抽样调查方案
174	案例四 2000 年北京市限额以下批发零售贸易业餐饮业抽样调查方案
178	附录
178	附录一 随机数字表
182	附录二 正态分布概率表
184	参考文献

第一章

抽样技术概述

统计学是一门处理数据的技术,它是研究如何收集数据、整理数据、描述与分析数据,由此进行推断,并作出决策或预测的一门学科。其目的是通过数据探索事物内在的规律性。而抽样技术是统计学的重要分支,它已经成为当今最重要的统计方法。抽样技术广泛用于社会、经济、科技、自然等各个领域,成为现代统计学中发展最快、最活跃的一个分支。本章着重介绍抽样技术及与之相关的基本概念,抽样技术的特点、作用,抽样技术的发展历史及理论依据。

第一节 抽样技术的特点和作用

一、抽样技术的概念

抽样技术的完整概念应包括两个方面,一是对样本的抽取调查,一是对总体数据的估计。为了符合认识规律和叙述的逻辑性,首先介绍抽样调查,然后再介绍总体数据估计的基本理论和方法。

(一) 抽样调查

抽样调查是一种非全面调查,它是根据随机原则,从所研究的总体中抽取出部分单位进行的调查,并以这部分单位调查的结果来推断总体的特征。被抽取出的部分单位组成样本。每一个被抽取的单位叫样本单位。样本单位数目的多少,不是随心所欲任意确定的,而是根据一定的原则和要求,用科学的方法计算确定的。其原则是指随机性原则,也叫同等可能性原则,是指在抽取样本单位时,完全排除人为因素的干扰,使被研究总体中的每个单位都有同等机会被抽取,即每个单位被抽中的机会均等,而抽中与否全凭偶然。只有这样被抽取的单位所组成的样本,才能更好体现总体的分布状况,才能代表总体,也才能使我们依据样本的调查结果来推断总体的特征,完成由部分推断总体的抽样调查工作。

(二) 抽样估计

抽样估计是在抽样调查的基础上,利用根据样本数据整理计算出的一些统计量,以概率论中大数法则的理论为指导,来估计总体相应的数据的一种统计方法。例如:根据样本各单

位的数据资料计算的样本均值、样本方差等就是最常见的统计量,利用这些统计量就可以对总体数值进行估计。

(三) 假设检验

假设检验和抽样估计都是统计估计的方法,它们都是利用样本数据对总体数据进行估计,但是估计的角度不同。抽样估计是运用样本数据对总体数据估计,在估计之前总体相应的数据是未知的;而假设检验,则是先对总体某一数据提出假设,然后提取样本,运用样本数据来验证假设成立与否。例如:在抽样估计(也叫参数估计)中,用样本均值 $\bar{x}$ 去估计总体均值 $\bar{X}$,在估计前总体均值 $\bar{X}$ 是未知的;而在假设检验中,则是先对总体均值提出一个假设,如 $\bar{X} \leq 17$,而后用样本均值去判断这个假设是否成立,这就是假设检验与抽样估计的不同之处。

(四) 抽样设计

所谓抽样设计,就是指从一定的研究总体中抽取样本之前,为了保证样本的代表性,预先确定抽样的程序和方案,以力求取得最经济最有效的结果,即兼顾抽样效果和费用支出两方面因素。按照国外经验,在制定抽样方案时,常将下列人员——调查资料使用者、抽样专家、实际调查工作组织者和数据处理人员找到一起,共同交流意见,协商探讨各种问题,最后提出一个合理可行的抽样技术方案。交流意见的过程,常常被看作是非常重要的过程,它可以使抽样设计趋于完美。抽样设计的基本内容有:

- (1) 确定抽样的目的、任务和要求;
- (2) 确定调查对象(被研究总体)的范围、抽样框和样本单位;
- (3) 确定抽样的组织方式和抽取样本的方法;
- (4) 确定必要的抽样数目;
- (5) 提出对数据的精度要求;
- (6) 确定抽样估计方法;
- (7) 制定实施总体方案的办法和步骤。

二、抽样技术的特点

(一) 在调查单位的抽取上,遵循随机原则

根据随机原则,就可以使被研究总体中的各单位的抽中或不被抽中都不受任何主观因素的影响,而完全是纯客观的。只有这样,才能有较大的可能性使所抽取的样本变量的分布与被研究总体的变量分布状态相类似,从而保证样本的代表性,以提高对总体数据进行估计的精确度。抽样调查与其他非全面调查,如典型调查、重点调查等有明显不同,后者是有意识的确定调查单位。当然,随机取样不是随便取样或任意取样,必须遵守随机原则。只有遵守随机原则,才能保证样本对总体的代表性,才能计算抽样误差。因此,随机原则是抽样调查的根本原则。

(二) 在调查的功能上,能以样本数据推断总体数据

统计研究的目的是要认识总体的数量特征,但许多社会经济现象不能或不必要用全面调查的方法来取得总体全面资料。如何只通过对总体中部分单位的调查,来达到认识总体数量特征的目的呢?只有抽样调查独具此功能。普查虽然可以达到对总体的认识,但它是全面调查,得不到非全面调查的节省人力、物力、财力,提高时效的效果。而重点调查和典型调查虽然是非全面调查,具有相应的效果,但却又不能在数量上推断总体,只有抽样调查具备

这双重特点。因此,抽样调查广泛应用于政治、经济、科学、文化、教育、卫生等各领域。例如:根据对随机抽取的一定百分比的罐头食品的品质进行检测,来推断整批罐头食品的品质;根据对抽取的部分农户消费水平的情况推算整个地区的农户消费水平等。

(三) 在推断手段上,是以概率估计的方法对总体数据进行估计和推断

抽样推断是以概率论为基础的估计方法,用样本数据来估计总体相应数据时,其可靠性只能用一定的概率保证程度来说明,而不能完全肯定。例如,用城市居民样本计算的某电视节目的收视率估计城市居民总体的收视率,只能在一定的概率保证程度下,比如在95%或99%的概率保证程度下,作出具有一定区间范围的估计值,而不能保证以100%的把握程度给出惟一肯定的估计值。当然,概率保证程度的大小是可以根据实际情况和需要而变动的。

(四) 在推断理论上,以大数定律和中心极限定理为理论基础

大数定律和中心极限定理论证了随着样本单位数的增加,样本变量分布趋向正态分布,样本平均数接近总体平均数,样本方差接近总体方差。从而为用样本数值估计总体相应数据提供了科学的理论依据。

(五) 在推断效果上,抽样误差可以计算并加以控制

用样本数据推断总体相应数据是存在一定误差的。但是,根据前面所述的大数定律和中心极限定理及正态分布规律,抽样误差范围可以事先通过有关资料加以计算,并且可以控制,从而使抽样估计有一定的可靠程度,这是随机抽样优于其他估计方法之处。

三、抽样技术的作用

(一) 用于认识那些不能或难以采用全面调查的总体的数量特征

不能或难以采用全面调查的总体有无限总体,如:宇宙探测、大气或生态污染情况的调查中的总体;有动态总体,如:对连续生产的产品性能检测中的正在连续不断生产的产品;有范围过大、分布过散的有限总体,如:对城市居民经济状况调查时的全体城市居民、对水库中鱼苗数量调查时的整个水库、森林的木材蓄积量调查时的整片森林等。对于这些总体的相应内容进行调查时,只能用抽样技术的方法取得有关数据资料,并对总体数据进行推断。

(二) 用于认识不适宜进行全面调查,但又必须了解总体数据的事物

例如,对于电器产品耐用时间的检测、对于武器杀伤力的检测、对于食品品质的鉴定等均具有破坏性。如果对于每一个对象都进行检测的话,检测完毕时,所有产品也全部消耗掉了,检测本身也失去了意义。因此,对于这类事物总体数据的取得只有通过抽样技术来解决。

(三) 用于认识不必进行全面调查的总体数据,可节省人、财、物力

有些总体虽然可以进行全面调查,但耗费的人力、物力、财力很大,花费的时间较长,并且有可能出现较多的登记误差以及抄写、计算、汇总误差等,影响资料的准确性。例如:对下岗人员再就业状况的了解,从理论上讲是可以进行全面调查的,但是,调查的范围太大,难度也太大。像这样的调查,诸如地区性、民族性、季节性消费,人体的身高与体重等都有内在的相似性与规律性,只需进行抽样调查即可。况且,抽样调查的调查单位少,调查人员经过专门训练,能取得较详细、准确度较高的资料。

(四) 用于采集灵敏度高、时效性强或时间要求紧迫的信息资料

在激烈的市场竞争中,市场信息有很强的灵敏度和时效性,若不及时取得就会贻误商机。很显然,若采用全面调查,既不可能也不允许。再如,生产线上连续不断生产的产品的

质量,要求随时掌握,一旦发生问题须及时发现及时纠正,否则,待大量生产后才发现问题,所造成的损失难以挽回。还有,商品交易中的验收、易变化波动现象的调查、特殊情况下的特殊检查(如抢险救灾、战时物资检查)等,都是时限性很强的调查,只能进行抽样调查,以争取时间。在农产品产量调查中,抽样调查的时效性也很好地体现出来了,抽样调查可以和收获工作同时进行,及时得到统计数字,以利于安排农产品的收购、存储、运输和销售工作。

(五) 与其他统计调查尤其是全面调查相结合,互相补充与核对

抽样调查与重点调查相结合。以确定好的在一定规模以上的单位作为重点单位进行逐个调查,其他单位可以进行抽样调查,形成目录抽样,有利于更好地认识总体数量特征。

抽样技术与普查相结合。普查由于调查范围广、工作量大、参加人员多,只能定期举行。我国的人口普查是每十年进行一次。普查时的短表(调查项目较少)与普查年份抽样调查的长表(调查项目较多)相配套,既可以在内容上相互补充,又可以在时间上相互衔接,而且可以满足不同的需要。

对普查或全面报表资料进行抽样检查与核对,能判断全面调查质量的高低,发现全面调查工作中的各种差错并给以纠正。例如,我国历次人口普查后都要进行差错率的抽样检查并对检查结果进行修正,以提高普查资料的准确性。

(六) 抽样调查和估计,还可以用来对总体特征的某种假设进行检验,以判断真伪,决定方案的取舍,为行动决策提供依据

例如,某项新工艺、新配方或农业新品种在生产中的推广是否具有显著价值,就可以通过抽样推断来进行假设性检查,以决定是采用还是放弃。

综上所述,由于抽样调查具备以上特点及作用,使其在实际工作中,特别是在市场经济条件下得到了越来越广泛的应用。每一项关于社会经济的抽样调查都需要我们要认真对待,要严格遵照抽样调查的基本原理,正确运用抽样方法,以获得较好的预期效果。

四、抽样技术的形成、发展及应用

(一) 抽样技术的形成与发展

统计作为认识世界的方法,已有较长的历史,而抽样技术作为一种科学方法的形成和发展只是近一个世纪的事。但不论是其理论方法,还是实践应用,抽样技术的发展都是非常迅速的,并已成为现代统计学的一个重要组成部分。

正式的抽样技术的历史大致可以从 1895 年算起。但在这以前,人类在使用样本和进行简单推算方面已有相当的发展。世界上最早采用抽样技术来研究社会经济问题的要算是法国数学家拉普拉斯。他曾在 1800 年利用人口出生统计抽样资料来估计当时法国人口总数。在 1786 年他写了一篇关于巴黎的出生、婚姻、死亡的文章,提出了依据法国特定地方的出生率来推算法国全部人口的方法。他的建议得到法国新政府的支持,法国政府选了 30 个县进行抽样调查,从 1802 年上溯到 1799 年。在当时条件下,这件事无论对人口统计来说,还是对统计方法论来说,都是一个巨大的贡献,而且,对以后统计学的发展,特别是抽样技术的运用,也产生了深远的影响。

1891 年,挪威统计局局长凯尔进行了挪威国民收入和财富的统计调查。在这次抽样调查中,可能代表全国国民收入和财富情况的许多城市和村落被选出来,从中按年龄、社会阶层、婚姻状况以及姓名的某个字母的起首标志抽选居民。通过被抽选出的居民的有关情况的详细资料,来推断出挪威全国国民收入和财富。

1894年,凯尔和汉森应用抽样调查进行了挪威国会劳工委员会委托举办的年龄、收入和职业的相关研究。他们将可能代表各阶层居民的某些街道作为挑选对象,对每个第十户的成年居民的情况作详细的记录,最后根据这些调查结果来推断全国人口的年龄、收入和职业的相关情况。

1895年,在第五届国际统计学会伯尔尼会议上,凯尔提出了所谓“代表性调查”的抽样方法,认为“调查结果的准确性不是取决于观察数量的多少,而是取决于是否采用正确和有代表性的方法”。1903年,第九届国际统计学会建议各国运用抽样技术。1924年,国际统计学会成立“抽样技术应用研究委员会”。1925年在罗马举行的第十六次国际统计学会上,“抽样技术应用研究委员会”提出了鲍莱的《抽样精度的确定》和詹森的《代表性方法的实践》这两个报告,从理论和实践上充分肯定了抽样技术的科学性。在这一阶段及其以后的日子里,抽样技术相继出现了典型抽样、随意抽样、定额抽样等有意抽样形式。

但是,有意抽样毕竟不够完善、不够科学。为此,随机抽样逐步产生并为人们所接受。早在1901年,德国统计学家波特基维茨提出,应用概率理论来计算样本与总体之间的差异究竟在多大程度上属于偶然因素引起的。1906年,鲍莱也提出,随机抽样可以得到我们所愿意要的完善的结果。后来,英国统计学家蒂配特编制了历史上第一张《随机数字表》,解决了随机取样问题。与此同时,俄国数学家里亚普诺夫给出了中心极限定理的严格数学证明,英国的皮尔生提出了著名的“卡方检验”。1908年,英国的戈塞特提出了小样本思想和“ t 分布”。1923年,费喧提出了“方差分析法”,完善了小样本理论,阐明了著名的实验设计原理。1925年,詹森和鲍莱在第十六次国际统计学会上第一次提出了随机抽样要按照概率原理使每个单位有同等被抽中的机会。到了20世纪30年代,随机抽样的概念和原则得到了确立,逐步发展成为完善的抽样理论。

20世纪30年代以后,抽样调查实践得到了迅速的发展。首先是波兰的统计学家尼曼在1934年根据概率的原理,提出了置信区间的推断理论。他还提出了类型抽样的样本最优分配原理,导出了最佳分配方法,改进了整群抽样设计。印度的马哈拉诺比斯从1930年开始,采用了简单随机抽样、等距抽样、类型抽样、整群抽样、多阶段抽样等多种抽样方式,研究解决了许多方法技术问题。在美国,在1937年失业情况调查中,采用了置信区间和比估计方法;在1938年失业调查中采用了多阶段调查抽样方法;1942年各阶段抽样进一步采用非等概率抽样,并提出控制非抽样误差设计。美国著名的统计学家科克伦在20世纪50年代出版的《抽样技术》一书成了这一领域的权威性著作。另一位美国著名的统计学家Kish在20世纪60年代出版的《抽样调查》一书至今仍享有盛誉。1960年联合国统计局出版了《抽样调查理论基础》一书,向世界各国政府推荐22种社会经济抽样调查方法。进入20世纪80年代,计算机在抽样技术中的广泛应用,给抽样技术带来了新的突破。

(二) 抽样技术在我国的研究和应用

我国对抽样技术的研究和应用始于20世纪三四十年代,以北京大学许宝禄教授为代表的中国统计学家,为抽样理论和方法的创建作出了杰出贡献。许教授一系列的抽样推断和多元分析等领域的论著达到了国际领先水平。他所著的《抽样论》,为开展大规模的抽样调查,如人口调查、能源调查、社会经济调查、森林木材调查、草原和农田估产、昆虫数量估计等,提供了方法和依据,也为培养下一代统计人才提供了良好的教材。1943~1944年,在清华大学陈达教授的主持下,我国进行了云南户籍示范人口普查,并以现代抽样方式抽取部分

人口来核对普查结果,这是我国首次抽样调查。新中国成立后,于1952年8月成立了国家统计局。1955年,新中国第一次在全国范围内,按统一方案、统一计划进行了抽样调查——农民家计抽样调查,以等距抽样方法抽取调查乡,再按类型比例与等距抽样相结合的方法在全国抽取1.5万~2万农户进行抽样调查。1956年,国家统计局在“关于1956年全国统计工作的指示”中意识到了统计抽样的重要性,上面写道:“我们统计调查方法太机械,不灵活,只注意建立各种定期统计报表而不注意利用各方面现有的统计资料,重视普查而忽视抽样调查,更不愿进行有科学依据的估计推算……以致许多统计调查工作,费力过大,费时过长,赶不上党政领导机关的需要。”同年,国家统计局对10个工业部门进行了职工家计抽样调查,抽样方法是类型比例与等距抽样相结合。该年12月以王思华为团长的国家统计局代表团到印度考察农产量抽样工作。1957年6月,《统计工作通讯》杂志发表社论《积极在全国范围内开展抽样调查,广泛收集统计资料》。与此同时,前苏联和西方国家的抽样文献也引进到国内。1957年6月,印度的马哈拉诺比斯教授来华讲学。1963年,国务院批准成立了我国第一支抽样调查队伍——全国农业产量调查队。1964年对全国主要农作物播种面积和农作物产量进行了抽样调查。“文化大革命”十年动乱,抽样调查工作被迫中断。党的十一届三中全会以后,抽样调查工作得到恢复。1981年9月15日,国务院在批转国家统计局的“关于加强和改革统计工作的报告”中要求:“积极推广抽样调查,做到统计调查方法灵活适用……今后布置新的统计任务,增加新的统计指标,凡是适合用抽样调查的,就不用全面报表;现在已经建立的全面报表制度,适合改用抽样调查的,要逐步用抽样调查代替。”同时,国务院批准成立农村和城市两支抽样调查队。1984年7月国家统计局将农村经济抽样调查的方法由半距等距抽样改为对称等距抽样。1994年,经全国统计工作会议和国务院审批,我国确立了以周期性普查为基础,以经常性抽样调查为主体,以必要的统计报表、重点调查、综合分析等为补充,搜集、整理基本统计资料的统计工作基本方针,抽样技术的主体地位自此得到正式确立,并经国务院批准,成立了第三支抽样调查队——企业调查队。

随着社会主义市场经济体制的进一步完善,一方面,为满足党和政府及时了解社会经济发展变化的各方面情况,需要经常进行抽样调查,为宏观决策和调控提供依据。另一方面,为了使企业在市场竞争中处于战略主动地位,需要开展各种抽样调查,随时了解市场动态,以使企业的预测决策有据可循,减少盲目性,具有超前性。现在,抽样技术在工业、商业、交通运输、通讯邮电、建筑、社会调查、民意测验、环境保护、生态平衡、生命科学、医疗卫生等许多方面得到广泛应用,抽样技术发展空间日趋广泛。

第二节 抽样技术的理论依据

一、随机事件及其概率

(一) 随机事件的几个基本概念

在事物的联系和发展过程中,随机现象是客观存在的。在社会生活实践中,人们常常会

遇到随机现象。例如,跳伞运动员进行定点跳伞训练时,每次落地可能踩在点上,也可能踩不到点上,即每次跳伞踩点的结果是随机的;体育彩票抽奖,中奖号码的每位数字是随机产生的,等等。为了揭示随机现象,反映客观事物的规律性,我们必须对现象进行观察和科学实验。在同一组条件下,对某一事物或现象所进行的观察或实验,我们称之为试验,把观察或实验的结果叫做事件。如果在相同的条件下,试验可以重复进行,而且每次试验的结果不能事先确定,则称这样的试验为随机试验。

例如随意抛掷一个骰子(一个质地均匀、式样对称的正六面体,六面分别刻有1、2、3、4、5、6六个数字)就是一次试验。骰子落地,可能出现1点,2点……6点;或为奇数点,或为偶数点,或为点数大于3等等结果。每一种结果都是一个事件,而且这些事件都是在一次试验中可能出现,也可能不出现的事件。与此不同的还有两种事件,即在一次试验中,点数小于7这一事件,在每次抛掷时是一定出现的,而点数大于6这一事件,在每次抛掷时是一定不出现的。这样可以引出下面三个概念,通过互相比较,对随机事件的概念会更清楚些。

(1) 随机事件:在相同的条件下,每次试验可能出现也可能不出现的事件,或叫偶然事件;

(2) 必然事件:在相同的条件下,每次试验一定出现的事件;

(3) 不可能事件:在相同的条件下,每次试验一定不出现的事件。

在掷骰子的活动中,点数小于7这一事件就是一个必然事件;点数大于6这一事件就是一个不可能事件。

在概率论里研究的总是随机事件,并且把必然事件与不可能事件包括在随机事件内作为两个极端事件来看待。

随机事件也可简称为事件,如果一个事件不能分解成两个或更多个事件,则这个事件称为基本事件或简单事件。例如掷骰子观察点数的试验中,点数为1点,点数为2点,点数为3点,点数为4点,点数为5点,点数为6点是这个试验中的6个基本事件。

基本事件具有很重要的性质。在一次试验中,只能观察到一个且仅有一个简单事件。一个试验的全部可能结果,或者说所有的简单事件的全体称为样本空间或基本空间,记为 Ω 。如在掷硬币试验中, $\Omega = \{\text{正}, \text{反}\}$;在掷骰子试验中, $\Omega = \{1, 2, 3, 4, 5, 6\}$ 。

(二) 概率

一个随机试验有许多可能的结果。我们不仅想知道它们有哪些可能的结果,而且还想知道某些结果出现的可能性的的大小,并希望将这一可能性用数值表示出来。为了定量描述随机事件,人们引入了一个描述随机事件发生可能性大小的数量指标——随机事件的概率。在一个试验中,某一随机事件发生的次数占所有随机事件发生次数的比率就是该事件的概率。概率的研究来自于欧洲中世纪的赌博或机会游戏。在数百年的发展中,许多数学家、统计学家对概率及其计算做出了巨大的贡献。随着集合论、测度论等数学理论的发展,20世纪30年代前苏联数学家柯尔莫哥洛夫提出了概率论的公理化体系。概率论,就是研究随机现象(偶然事件)的规律性的科学。

试验概率的特征是:

(1) 大量性,即试验次数必须足够多,量必须足够大;

(2) 重复性,即在同一组条件下每次试验的条件和程序都必须相同;

(3) 误差性,即频率只是概率的接近值或估计值。

二、概率分布的概念与种类

(一) 概率分布的概念

随机变量在一次试验中出现的结果是不确定的,既可能出现这个数值,也可能出现另一数值,只有在大量试验中才能发现其规律性。概率分布由随机变量的取值 x 及其相应的概率 $P(x)$ 构成,它可以用表或图表示。将随机变量的取值与其相应的概率依序排列,即是概率分布。

任何概率分布都满足以下两个条件:

第一,随机变量任一取值的概率都界于 0 和 1 之间,即:

$$0 \leq P(x) \leq 1$$

其中: x 表示某一样本空间中的每一事件; $P(x)$ 表示在一个试验中 x 事件出现的可能性,即概率。

$0 \leq P(x) \leq 1$ 表明对于任一事件 x ,其概率值是一个不超过 1 的非负实数。

第二,随机变量所有可能的取值的概率之和等于 1,即:

$$\sum P(x) = 1$$

$\sum P(x) = 1$ 表明必然事件,即每次试验中都必然发生的事件的概率值为 1。前面已讲述过,若一事件在任何一次试验中都不发生,称为不可能事件。按上述两个条件推理可得:不可能事件的概率为零。 ϕ 表示不可能事件,则 $P(\phi) = 0$ 。

概率分布就是对某种随机变量总体规律性的描述。因此,对随机现象进行研究,首先要找出它的概率分布。当我们掌握了某个随机变量的具体分布时,反过来就能对某一具体事件出现的可能性作出判断。

(二) 常用的两种类型的随机变量

按照随机变量的特性,通常可以把随机变量分为两类,即离散型随机变量和连续型随机变量。

1. 离散型随机变量

如果随机变量 x 的所有取值都可以逐个列举出来,则称 x 为离散型随机变量。例如“在一批产品中取到次品的个数”、“单位时间内某交换台收到的呼叫次数”等都是离散型随机变量。

常见的离散型随机变量的概率分布有(0—1)分布和二项分布,这里仅介绍(0—1)分布。

设随机变量 x 只可能取 0 和 1 两个值,它的概率分布是:

$$P(x=1) = p, P(x=0) = 1 - p = q$$

其中, $0 < p < 1, p + q = 1$ 。则称 x 服从(0—1)分布,或称 x 具有(0—1)的分布,也称两点分布。(0—1)分布可以表示成表 1-1。

表 1-1 (0-1)分布

x	1	0
$P(x)$	p	q

现举例说明：设已知一批产品的废品率为 $p = 0.05$ ，合格率为： $q = 1 - p = 1 - 0.05 = 0.95$ ，并指定废品用 1 代表，合格品用 0 代表，则任抽一件或为废品或为合格品，即 1 或 0，这一离散型随机变量的概率分布如表 1-2 所示，其图示法见图 1-1。

表 1-2 (0-1)概率分布举例

x	1	0
$P(x)$	0.05	0.95

在日常生活中，我们经常可以遇到(0-1)分布。例如，垂钓是否能钓上鱼来，检验产品质量是否合格，对新生入学的性别登记，等等，我们都可以用(0-1)分布的离散型变量来描述。

2. 连续型随机变量

如果随机变量 x 的所有可能取值不可以逐个列举出来，而是取数轴上某一区间内的任一点，则称该随机变量为连续型随机变量。例如一批电子元件的寿命，成人的身高、体重，儿童的营养状况等都是连续型随机变量。

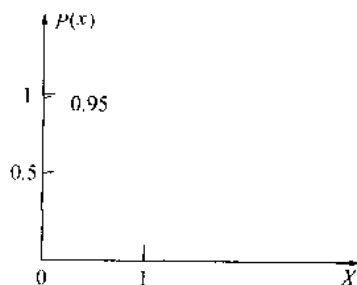


图 1-1 (0-1)分布

常用的连续型随机变量的概率分布，有均匀分布和正态分布。这里仅介绍正态分布。

正态分布是概率论与数理统计中最重要的分布。这一分布最早是由法国数学家德·莫哇佛尔在求二项分布的极限分布时得出来的，以后由德国数学家高斯和法国数学家拉普拉斯分别于 1809 年、1812 年进一步推导，所以也称为高斯-拉普拉斯分布。高斯在观察星体运动时总结了误差的分布规律并提出了正态分布，他把这一分布称为误差法则。后来，高斯在天文、物理以及其他领域中，把搜集到的资料加以整理，发现资料中的数据分布都呈现中间大、两头小且对称的性质，因此他把正态分布看作是支配一切现象的规律，称之为 normal law，过去曾译为常态分布。这种观点在很长时间内支配着统计理论，直到以后数理统计学取得进一步发展，提出了各种不同的分布，人们才认识到正态分布只是随机现象分布中的一种。但是直到现在，正态分布在数理统计中仍然具有十分重要的作用，是很多统计理论的基础。抽样理论证明，大量独立的随机事件的总和，可以看成是由许多基本的随机事件共同发生作用的结果，都趋于正态分布，所以在抽样推断中正态分布也居于特别重要的地位，其应用十分广泛。正态分布的几何形状如同一口古钟，因而也称为钟形分布(如图 1-2 所示)。

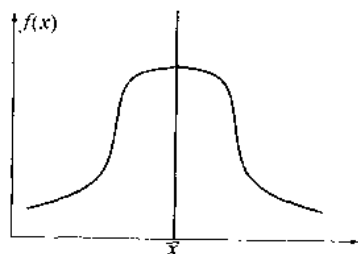


图 1-2 正态分布

三、大数定律与中心极限定理

(一) 大数定律

我们前面所讲述的随机事件的概率是由大量现象中的客观规律性引申而来。在大量的随机现象中,我们看到了随机事件频率的稳定性,如在观察大量的掷硬币的结果后,我们发现正面向上的次数比率,在次数不断增大时,其比率趋于0.5,并逐渐稳定。同时,我们还看到了一般的平均结果的稳定性。这表明,无论个别随机现象的结果以及它们在进行过程中的个别特征如何,大量随机现象的结果实际上是与每个个别随机现象的特征无关,并且几乎不再是随机的了。概率论中用来阐明大量随机现象平均结果的稳定性的一系列定理统称为大数定律,也叫大数法则。它是一种表现必然性与偶然性之间辩证联系规律的法则。它包括以下几个方面的内容:

(1) 在一定总体条件下,只有当具有某种表现的足够多数的单位综合在一起的时候,统计规律才能表现出来;

(2) 统计规律通常是以平均数的形式表现的;

(3) 所研究现象总体包含的单位数越多,平均数也就越能够正确地反映出这些现象的规律性;

(4) 各单位的共同倾向(表现为主要的、基本的因素)决定着平均数的水平,而各单位对平均数的离差(表现为次要的、偶然的因素)则会由于足够多数单位的综合汇总的结果,而相互抵消,趋于消失。

大数法则表明:如果随机变量总体存在着有限的平均数和方差,则对于充分大的抽样单位数 n ,可以用几乎趋近于1的概率,来期望抽样平均数 $\bar{x}_i$ 和总体平均数 $\bar{X}_i$ 的绝对离差为任意小,即对于任意的正数 α 有:

$$\lim_{n \rightarrow \infty} P(|\bar{x}_i - \bar{X}_i| < \alpha) = 1$$

这就从理论上揭示了样本和总体之间的内在联系,即随着抽样单位数 n 的增加,抽样平均数有接近于总体平均数的趋势,或者说,抽样平均数 $\bar{x}_i$ 在概率上收敛于总体平均数 $\bar{X}_i$ 。

(二) 中心极限定理

中心极限定理是有关论证大量的随机变量和的极限分布是正态分布的一类定理。该定理表明:一个随机变量,如果它是很多个相互独立的随机变量之和,而其中每一个变量对总和只产生微小的影响,那么当求和项数无限增加时,这一总和的分布将趋于正态分布。

我们可以将中心极限定理的内容概括为以下四个性质:

(1) 大量的社会经济现象总体是正态总体或近似于正态总体;

(2) 在大样本的条件下,抽样平均数的分布是或近似地是正态分布,抽样成数的分布是或近似地是正态分布;

(3) 抽样平均数的平均数等于总体平均数,抽样成数的平均数等于总体成数;

(4) 设总体标准差为 σ ,成数为 P ,当总体单位数 N 相当大时,抽样平均数和抽样成数的标准差分别是 $\frac{\sigma}{\sqrt{n}}$ 和 $\sqrt{\frac{P(1-P)}{n}}$ (n 为样本单位数)。

根据以上性质,我们可以按正态分布来估计抽样平均数(成数)落在一定的范围内的概率,进行抽样推断。