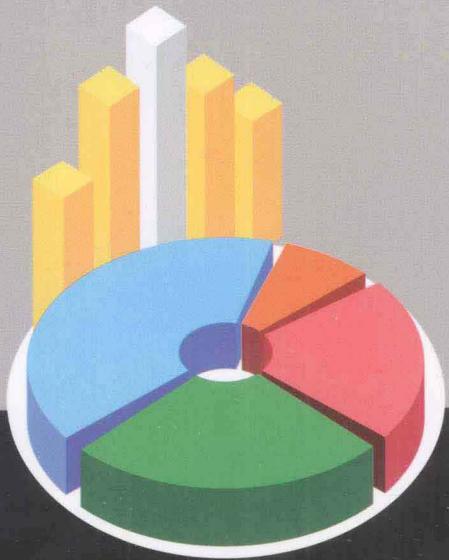


# SPSS



何国民 ◎ 编著

# 应用统计学案例教程

—以SPSS为计算工具

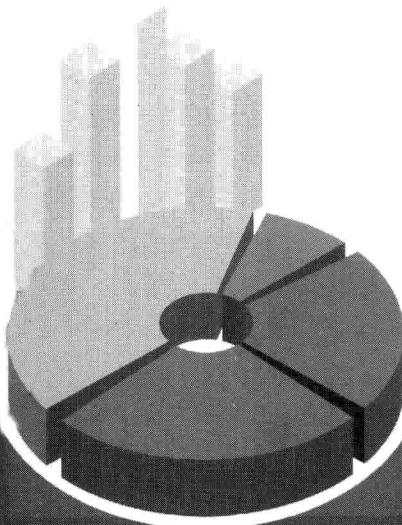
# A

Case-study Course  
for Applied Statistics



华中科技大学出版社  
<http://www.hustp.com>

SPSS



何国民 ◎ 编著

# 应用统计学案例教程

— 以SPSS为计算工具

A Case-study Course  
for Applied Statistics

## 内 容 简 介

本书将统计方法、SPSS(“统计产品与服务解决方案”软件)操作、应用案例有机结合,以统计方法应用为主线,以通俗易懂的语言对主要统计方法的核心思想进行系统介绍,以案例形式介绍各种统计方法,将SPSS作为一个“超级计算器”,以“操作示意图”的方式介绍如何利用SPSS完成各种统计方法的计算任务,得到所需的计算结果,这样就极大地减轻了统计学的学习难度,使读者能够较轻松地掌握各种统计方法的应用。本书克服了SPSS手册类教材中只注重SPSS操作说明而忽略统计原理讲解的不足,同时弥补了统计专业教材中只注重统计原理论述而缺乏数据处理实现工具的缺憾,是一本特色鲜明、使用价值广泛的教材。

### 图书在版编目(CIP)数据

应用统计学案例教程——以 SPSS 为计算工具 / 何国民 编著. — 武汉 : 华中科技大学出版社, 2013. 9  
ISBN 978-7-5609-8971-6

I . 应… II . 何… III . 应用统计学 - 教材 IV . C8

中国版本图书馆 CIP 数据核字(2013)第 100932 号

### 应用统计学案例教程——以 SPSS 为计算工具

何国民 编著

策划编辑：周小方 陈培斌

责任编辑：苏克超

封面设计：范翠璇

责任校对：李 琴

责任监印：张正林

出版发行：华中科技大学出版社(中国·武汉)

武昌喻家山 邮编：430074 电话：(027)81321915

录 排：华中科技大学惠友文印中心

印 刷：华中理工大学印刷厂

开 本：787mm×1092mm 1/16

印 张：19.75 插页：1

字 数：505 千字

版 次：2013 年 9 月第 1 版第 1 次印刷

定 价：38.00 元



本书若有印装质量问题,请向出版社营销中心调换  
全国免费服务热线: 400-6679-118 竭诚为您服务  
版权所有 侵权必究



自然界和社会上发生的现象是多种多样的,其中有一类现象在个别试验中其结果呈现出不确定性,但在大量重复试验中其结果又具有统计规律性,我们称之为随机现象。概率论与数理统计是研究和揭示随机现象统计规律的数学分支学科。

概率统计理论与方法的应用是很广泛的,几乎遍及所有科学技术领域,包括工农业生产、经济学、生物学、教育学、心理学、医疗卫生、体育等多个领域。

20世纪50年代以来,计算机技术的发展与普及,特别是各种统计分析软件的推出,有力地解决了统计方法在实际应用中统计计算任务繁重这一瓶颈,统计方法得到了更为广泛的应用,以前很少使用的各种多元统计分析方法现在已司空见惯。

在统计数据处理电算化的今天,学习统计学而不会使用统计软件的人,在实际应用中将寸步难行。目前介绍各种统计软件的书籍比较多,但将统计方法、统计软件的使用及应用案例相结合的书籍并不多见。有些书籍侧重对统计软件使用的介绍,对统计方法讲解很少,给不了解统计方法的读者带来诸多学习上的困难,造成读者可以“依葫芦画瓢”地对数据进行处理,但对统计结果却“目不识丁”,甚至误用统计方法的恶果;有些书籍则侧重对统计方法的论述,缺少对统计软件使用方法的介绍,这样给统计方法的实际应用带来障碍,造成学完统计学后会做题,面对实践中的大量数据却束手无策的“高分低能”现象。从快速掌握统计方法角度来看,较为合适的方法是将统计方法的学习、统计软件的使用有机结合起来,并配以统计方法应用的案例分析,这样才能使读者较快地领会统计方法的要点,掌握统计方法的使用,从而学以致用。

本书将统计方法、SPSS(“统计产品与服务解决方案”软件)操作、统计方法应用案例有机结合,以统计方法应用为主线,以通俗易懂的语言对主要统计方法的核心思想进行系统介绍,并以图示方式介绍各种统计方法的SPSS操作过程,通过案例的形式介绍各种统计方法,极大地减轻了统计学的学习难度,使读者能够较轻松地学习与掌握各种统计方法。

本书在编写过程中力求做到以下几点。

(1) 对各种统计方法所涉及的基本概念、原理的介绍力求简单明了、深入浅出,并尽量少用统计专用符号、公式,目的是减少读者阅读上的困难。

(2) 对于SPSS的介绍,只是将其作为一个“超级计算器”,以“操作示意图”的方式介绍如何利用SPSS完成各种统计方法的计算任务,读者可以用“看图识字”的方式非常简单、方便地学会用SPSS完成各种统计方法的计算工作,得到所需的计算结果。

(3) 每种统计方法都配有具体应用案例,通过案例的形式学习统计方法,极大地减轻了

读者学习统计学的难度。

(4) 精选若干文献供读者阅读,针对文献中所用统计方法、存在的不足设置若干问题。读者通过阅读所列文献,回答相关提问,可对统计方法应用有更深入的理解。

本书可作为高等院校人文社科、管理、经济等专业本科生、研究生统计学教材,也可作为从事统计分析和决策的社会各领域相关专业工作者的参考书,还可作为统计方法与 SPSS 软件的自学读本。

由于作者水平有限,书中难免有不妥与错误之处,敬请统计界同仁及读者批评指正。

本书在编写中参阅了一些教材、专著和论文,在此向相关作者致以诚挚的谢意。

编 者

2013 年 6 月



<b>绪言</b> .....	1
一、统计学的产生与发展 .....	1
二、统计学方法体系 .....	2
三、统计分析的过程 .....	2
四、统计学相关网站与书籍 .....	3
<b>第一章 概率论基础</b> .....	4
第一节 随机事件及其概率 .....	4
一、随机事件 .....	4
二、频率 .....	5
三、概率 .....	6
四、小概率事件原则 .....	6
第二节 古典概型 .....	7
第三节 条件概率 .....	8
一、条件概率、乘法公式 .....	8
二、全概率公式 .....	10
三、贝叶斯公式 .....	12
第四节 独立性 .....	13
第五节 随机变量及其概率分布 .....	15
一、随机变量的定义 .....	15
二、随机变量的分类 .....	16
三、概率分布的概念 .....	16
四、离散型随机变量的概率分布 .....	17
五、连续型随机变量的概率分布 .....	17
第六节 正态分布 .....	19
一、正态分布的概念 .....	19
二、标准正态分布 .....	20



三、非标准正态分布 .....	23
<b>第七节 正态分布的应用 .....</b>	<b>28</b>
一、利用正态分布制定考核标准 .....	28
二、统一计分标准 .....	29
三、累进计分法 .....	31
思考与练习 .....	33
<b>第二章 数据收集 .....</b>	<b>37</b>
第一节 调查设计 .....	37
一、问卷设计 .....	37
二、抽样设计 .....	43
第二节 实验设计 .....	51
一、实验设计的内容 .....	51
二、实验设计的基本原则 .....	52
三、实验设计的方法 .....	53
思考与练习 .....	56
<b>第三章 SPSS 基本操作 .....</b>	<b>58</b>
第一节 SPSS 概述 .....	58
第二节 SPSS 的启动及其数据编辑器界面介绍 .....	58
一、SPSS 的启动方式 .....	58
二、SPSS 数据编辑器界面介绍 .....	60
第三节 数据的输入 .....	61
一、定义变量 .....	61
二、输入数据 .....	63
第四节 数据文件的编辑 .....	64
一、修改单元格中的数据 .....	64
二、变量的插入与删除 .....	64
三、样品的插入与删除 .....	65
第五节 数据文件的存取 .....	65
一、保存数据文件 .....	65
二、读取数据文件 .....	66
第六节 数据重编码 .....	67
第七节 SPSS 的运行方式 .....	72
一、完全窗口菜单运行方式 .....	72
二、命令语句运行方式 .....	72
三、混合运行方式 .....	73
思考与练习 .....	74
<b>第四章 描述统计 .....</b>	<b>76</b>
第一节 统计特征数 .....	76
一、表示整体水平的统计特征数 .....	77



二、表示差异程度的统计特征数 .....	78
三、表示分布形态的统计特征数 .....	80
第二节 图表描述 .....	82
一、频数分布表 .....	83
二、统计图 .....	84
第三节 描述统计命令的 SPSS 操作步骤 .....	86
一、Descriptives 命令的操作步骤 .....	86
二、Frequencies 命令的操作步骤 .....	90
第四节 多维频数分析 .....	96
思考与练习 .....	98
<b>第五章 假设检验 .....</b>	<b>100</b>
第一节 假设检验的原理及基本概念 .....	100
一、假设检验预备知识 .....	100
二、问题的提出 .....	100
三、样本推断总体的依据、原理、方法、步骤 .....	102
四、假设检验的相关概念 .....	105
第二节 假设检验的步骤 .....	106
一、分析问题,选用合适的假设检验方法 .....	106
二、提出假设 $H_0$ .....	107
三、计算检验统计量的值,得抽样结果 .....	107
四、计算相伴概率 .....	107
五、根据小概率事件原则作出结论 .....	108
第三节 参数检验 .....	108
第四节 参数检验的 SPSS 操作步骤及结果分析 .....	112
一、单样本 T 检验的 SPSS 操作步骤及结果分析 .....	112
二、F 检验独立样本 T 检验独立样本校 T 检验的 SPSS 操作步骤与结果分析 .....	113
三、配对样本 T 检验的 SPSS 操作步骤与结果分析 .....	116
第五节 非参数检验 .....	119
一、单样本非参数检验 .....	120
二、两独立样本非参数检验 .....	122
三、多独立样本非参数检验 .....	123
四、两配对样本非参数检验 .....	124
五、多配对样本非参数检验 .....	127
第六节 非参数检验的 SPSS 操作步骤与结果分析 .....	130
一、卡方检验的 SPSS 操作步骤与结果分析 .....	130
二、游程检验的 SPSS 操作步骤与结果分析 .....	132
三、K-S 分布适合性检验的 SPSS 操作步骤与结果分析 .....	133
四、二项式检验的 SPSS 操作步骤与结果分析 .....	136
五、两独立样本非参数检验的 SPSS 操作步骤与结果分析 .....	137

六、多独立样本非参数检验的 SPSS 操作步骤与结果分析	139
七、两配对样本非参数检验的 SPSS 操作步骤与结果分析	142
八、多配对样本非参数检验的 SPSS 操作步骤与结果分析	145
思考与练习	148
<b>第六章 方差分析</b>	<b>150</b>
第一节 方差分析的基本原理与相关概念	150
一、方差分析的基本原理	150
二、方差分析相关概念	152
第二节 单因素方差分析	153
一、单因素方差分析的基本思想	153
二、单因素方差分析的步骤与实例	154
三、多重比较	155
第三节 单因素方差分析的 SPSS 操作步骤及结果分析	156
第四节 多因素方差分析	159
一、多因素方差分析的基本思想	159
二、多因素方差分析的基本步骤	160
三、多因素方差分析的实例	161
第五节 多因素方差分析的 SPSS 操作步骤及结果分析	162
第六节 协方差分析	171
一、协方差分析的基本概念	171
二、协方差分析的 SPSS 操作步骤及结果分析	174
思考与练习	176
<b>第七章 相关与回归分析</b>	<b>178</b>
第一节 相关分析	178
一、散点图	179
二、相关系数	180
三、偏相关分析	185
第二节 相关分析的 SPSS 操作步骤与结果分析	187
一、绘制散点图的 SPSS 操作步骤	187
二、计算相关系数进行线性相关性检验的 SPSS 操作步骤与结果分析	189
三、计算偏相关系数的 SPSS 操作步骤与结果分析	191
第三节 一元线性回归	193
一、一元线性回归的数学模型	193
二、一元线性回归模型显著性检验	194
三、一元线性回归方程的求法	196
四、一元线性回归方程的应用	197
第四节 一元线性回归分析的 SPSS 操作步骤及结果分析	198
第五节 多元线性回归分析	200
一、多元线性回归分析的数学模型	200



二、多元线性回归模型显著性检验 .....	200
三、多元线性回归模型的求解 .....	202
四、回归自变量的检验 .....	203
五、多元线性回归分析中回归自变量的选取方法 .....	204
第六节 多元回归分析的 SPSS 操作步骤及结果分析 .....	208
一、多元线性回归分析的 SPSS 操作步骤 .....	208
二、对多元回归分析 SPSS 输出结果的解释 .....	214
思考与练习 .....	216
<b>第八章 聚类分析 .....</b>	<b>219</b>
第一节 聚类分析概述 .....	219
第二节 快速聚类分析 .....	220
一、快速聚类举例 .....	220
二、快速聚类的基本概念 .....	223
第三节 快速聚类分析的 SPSS 操作步骤及结果分析 .....	224
一、快速聚类的 SPSS 操作步骤 .....	224
二、快速聚类的输出结果及说明 .....	227
第四节 分层聚类 .....	229
一、分层聚类举例 .....	230
二、分层聚类的有关概念 .....	231
第五节 分层聚类的 SPSS 操作步骤与结果分析 .....	233
一、分层聚类的 SPSS 操作步骤 .....	233
二、分层聚类的输出结果及其说明 .....	236
思考与练习 .....	239
<b>第九章 判别分析 .....</b>	<b>240</b>
第一节 判别分析概述 .....	240
第二节 两类判别分析 .....	241
一、两类判别分析的数学模型 .....	241
二、两类判别分析实例 .....	241
三、判别分析中有关统计量的意义 .....	244
四、判别分析的步骤 .....	246
第三节 多类判别分析 .....	247
一、判别函数的个数 .....	247
二、判别函数的显著性检验 .....	247
第四节 判别分析的 SPSS 操作步骤及结果分析 .....	248
一、判别分析的 SPSS 操作步骤 .....	248
二、输出结果及说明 .....	255
思考与练习 .....	262
<b>第十章 因子分析 .....</b>	<b>264</b>
第一节 因子分析概述 .....	264

一、因子分析的基本原理与数学模型 .....	264
二、因子分析模型中各统计量的意义 .....	266
第二节 因子分析实例 .....	266
一、因子分析的 SPSS 操作步骤 .....	268
二、因子分析 SPSS 输出结果解读 .....	274
第三节 因子分析应用 .....	278
一、多变量结构分析 .....	278
二、基于因子分析的综合评价 .....	279
思考与练习 .....	282
<b>附录 A 文献阅读 .....</b>	<b>285</b>
文献一 体育锻炼对研究生健康影响的实验研究 .....	285
1 前言 .....	285
2 研究假设 .....	285
3 研究对象 .....	286
4 实验设计和运动处方 .....	286
5 研究结果 .....	287
6 分析与讨论 .....	292
7 结论与建议 .....	293
文献二 竞技体育与经济发展水平的相关分析 .....	295
1 竞技体育、经济发展水平评价指标的选取 .....	295
2 竞技体育成绩与经济发展水平指标间的相关分析 .....	296
3 结论 .....	299
<b>附录 B 常用统计表 .....</b>	<b>300</b>
<b>参考文献 .....</b>	<b>306</b>

# 绪言

## 一、统计学的产生与发展

statistics(统计学)一词起源于法语 status(状态),大概兴起于 17 世纪。1676 年英国经济学家威廉·配第的《政治算术》一书的问世,标志着统计学的诞生。统计学的发展过程大致可划分为以下三个阶段。

第一阶段:统计学初创阶段(17 世纪中叶到 19 世纪末)。

统计学的产生源于国势学派、政治算学术派的贡献。

17 世纪中叶,国势学派诞生于德国,代表人物是康令,他主张用记述的方法记录国家的重大事项,几乎不用数字资料。到 18 世纪,德国人阿亨瓦尔首次在大学开讲《国势学》课程,并将“统计”定义为“把国家的显著事项全部记录下来的学科”,并称此学科为 statistik(德文:统计学)。国势学派对社会经济现象进行研究时,只注重文字分析,完全不用数据,对图形表格、数学公式十分蔑视,因而被称为有名无实的统计学。

17 世纪中叶,政治算学术派在英国兴起,代表人物是威廉·配第,他在代表作《政治算术》(见图 0-1)一书中第一次用计量和比较的方法,将英国的国力与法国、意大利、荷兰等国的国力进行比较研究,以论证英国的国际地位。由于最早提出了一套较为系统的用于对社会经济现象进行数量性描述和分析比较的方法,威廉·配第被称为“统计学的创始人”。政治算学术派对社会经济现象进行分析时,注重数量分析,这与排斥数量只讲观念的国势学派不同,但由于在其所有著述中并没有提到“统计学”,因此这个学派被称为有名无实的统计学。

第二阶段:推断统计方法体系基本确定阶段。

19 世纪中叶,以比利时统计学家阿道夫·凯特勒为代表的数理统计学派,将概率论等数学方法引入统计分析中,完成了统计学与概率论的结合,开创了推断统计的先河。在这个阶段,估计理论、样本分布理论、方差分析理论、假设检验理论等方面都获得了重大进展。

第三阶段:多元统计方法应用全面发展阶段。

从 20 世纪 50 年代起,统计学受计算机、信息论等现代科学技术的影响,统计学的应用领域不断扩展,新的研究分支不断增加,如多元统计分析、探索性数据分析、数据挖掘技术、现代时间序列分析方法等等。据有关学者统计,多元统计方法的应用是以指数式加速发展的。



图 0-1 威廉·配第和他的著作

## 二、统计学方法体系

统计学属于收集、整理、分析数据的知识领域,要解决的根本问题是:从总体随机抽取样本,再由样本推断总体数量特征,如图 0-2 所示。

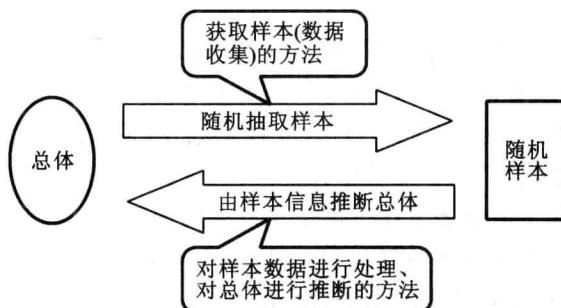


图 0-2 统计学要解决的根本问题

围绕样本推断总体这一统计学要解决的根本问题,提出了各种统计方法,这些统计方法包括数据收集、描述统计、推断统计、多元统计四大类方法。

(1) 数据收集:围绕研究问题科学合理地收集样本数据。

主要方法:调查设计,实验设计,间接数据收集。

(2) 描述统计:对抽样数据进行处理,得出一些特殊的图形、表格、数字,用它们来描述样本数据分布特点。

主要方法:数字特征描述,图表描述。

(3) 推断统计:根据样本数据对总体的数量关系作出某种推断。

主要方法:参数估计,假设检验,方差分析。

(4) 多元统计:分析多个变量间的数量关系。

主要方法:相关分析,回归分析,因子分析,判别分析,聚类分析。

## 三、统计分析的过程

用统计方法进行科学的研究的步骤可用图 0-3 表示。

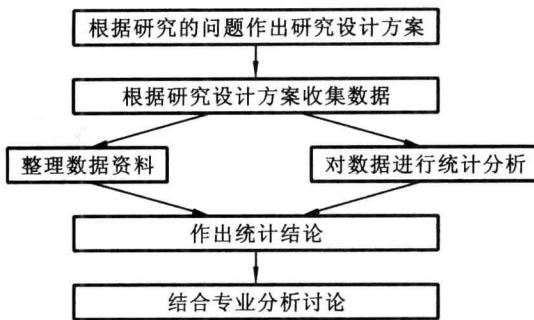


图 0-3 统计分析过程

## 四、统计学相关网站与书籍

以下是统计学一部分网络信息与书籍。

在线测试:厦门大学曾五一教授网站;暨南大学统计学精品课程。

统计学人:邱东博客;郑来铁博客;小蚊子乐园。

统计学习:中华统计学习网;中国统计师考试网;中国统计教育学会教学资源。

统计应用:数据草堂;统计家园;统计之都;统计菁英工作室;中国统计网。

统计书籍:每年都有大量的统计学教材新鲜上市,如《经济统计学》、《体育统计学》、《心理统计学》、《传播统计学》、《医学统计学》、《生物统计学》、《旅游统计学》、《教育统计学》、《社会统计学》等。还有大量统计读物可供漫游,如《看漫画,学统计》、《爱上统计学》、《统计,让数字说话》、《统计使人更聪明》、《生活中的统计学》、《统计连着你和我》、《漫游数据王国》、《统计学的世界》、《统计思想》、《漫话信息时代的统计学:兼话诺贝尔经济学奖与统计学》等。

图 0-4 是一些统计书籍的封面。



图 0-4 部分统计书籍的封面

# 第一章

## 概率论基础

自然界和人类社会中存在着两类现象——必然现象与随机现象。

必然现象就是指在一定条件下必然发生的现象。例如：在标准大气压力下，水加热到100℃必然会沸腾；抛出的物体其初速度只要小于第一宇宙速度，必然会落回地面。这些现象都是必然现象。

随机现象是指在一定条件下有时发生有时不发生的现象。如掷一枚硬币，“出现正面”这一现象有时发生，有时不发生；篮球投篮中“投中”这一现象有时发生，有时不发生。这些都是随机现象。

概率论是研究随机现象的数量规律的数学分支之一，本身具有丰富的内容与广泛的应用，但我们的主要目的是为以后的统计方法学习打下理论基础，本章所涉及的内容也就围绕着这一目的来展开。

### 第一节 随机事件及其概率

#### 一、随机事件

对随机现象进行观察，会观察到不同的结果，如观察掷硬币这一随机现象就可能看到“出现正面”或“出现反面”这两种不同的结果。“出现正面”是掷硬币这一随机现象的一种观察结果，我们称之为随机事件。同样，“出现反面”也是随机事件。

**随机事件：**对随机现象进行观察，其观察结果称为随机事件，简称事件，用大写英文字母A、B、C等表示。

作为随机事件的特例，若某事件在每次试验中总是发生，则称该事件为必然事件，一般用字母 $\Omega$ 表示；反之若某事件在每次试验中都不发生，则称该事件为不可能事件，一般用符号 $\emptyset$ 表示。

对于一个随机事件（除必然事件和不可能事件以外）来说，它在一次试验中可能发生，也可能不发生，但通过大量观察可以发现随机事件发生可能性大小具有一定的规律性，这种规律性称为统计规律性。如掷硬币，虽然掷1次无法确定是“正面向上”还是“反面向上”，但通过大量的试验可以发现，“正面向上”与“反面向上”的次数比较接近，而且随着试验次数的增加，接近程度愈来愈高，各占50%。历史上曾有不少数学家做过这类试验，如表1-1所示。



表 1-1 历史上数学家们掷硬币的试验数据

实验者	掷硬币次数	出现正面次数	出现正面频率
蒲丰	4040	2048	0.5069
皮尔逊	12 000	6019	0.5016
皮尔逊	24 000	12 012	0.5005

对于随机事件,我们常常希望知道它们在一次试验中发生的可能性究竟有多大。例如,为了确定水坝的高度,就要知道河流在造水坝地段每年最大洪水达到某一高度这一事件发生的可能性大小;再如,为了评价一位射击运动员水平的高低,我们就需要知道该运动员命中各环的可能性大小。

我们希望找到一个合适的数来表示随机事件在一次试验中发生的可能性大小。为此,首先引入描述事件发生频繁程度的量——频率。

## 二、频率

**频率:**在相同的条件下,进行了  $n$  次试验,在这  $n$  次试验中事件  $A$  出现了  $m$  次,则称比值  $m/n$  为事件  $A$  的频率,记为  $F(A)=m/n$ 。

显然任一事件  $A$  都有

$$0 \leq F(A) \leq 1$$

随机事件是否发生事先是不能确定的,但经过多次观察,随机事件发生的频率是有一定规律的。如:民兵射击 100 次射中 95 次,我们说其命中率为 95%;抽查某工厂的产品 100 件,结果有 98 件合格,我们说该工厂的产品合格率为 98%。

经验表明,随着试验次数的增多,随机事件频率的波动会越来越小,且会在一个固定的常数附近作微小的波动。以掷硬币为例,记正面向上为随机事件为  $A$ ,抛掷总次数为  $n$ ,出现正面向上的次数为  $m$ ,比值  $F=m/n$  为事件  $A$  的频率,所得结果如表 1-2 所示。

表 1-2 掷硬币试验的数据

$n=5$		$n=50$		$n=500$	
$m$	$F$	$m$	$F$	$m$	$F$
2	0.4	22	0.44	251	0.502
3	0.6	27	0.54	249	0.498
1	0.2	21	0.42	256	0.512
4	0.8	26	0.52	245	0.490
1	0.2	24	0.48	251	0.502

由表 1-2 可知:当  $n=5$  时,事件  $A$  发生的频率波动相当剧烈,频率变动无规律可循;当  $n=50$  时,频率波动明显减少,大致围绕 0.5 上下波动;当  $n=500$  时,频率波动已相当小,可明显看出其波动中心为 0.5。

不难想象,当抛掷次数再增多时,频率会更加稳定在 0.5 附近,这些试验的结果是很有启发性的,它们表明虽然事件  $A$  在一次试验中可能发生也可能不发生,但在大量重复试验中,它出现的频率趋于稳定,而且试验次数越多,频率越接近某一常数(在上述掷硬币的实例

中,此常数为 0.5),频率的这种随着试验次数增多而趋于稳定的情况称为频率的稳定性。

频率的稳定性说明随机事件发生可能性的大小是随机事件本身固有的一种客观属性,而不是由人的主观意志可随意改变的,因此可以对它进行度量,我们可以用一个数来描述随机事件在一次试验中发生的可能性大小,该数就是概率。

### 三、概率

直观地说,概率是描述随机事件发生可能性大小的度量,根据频率的稳定性,下面给出概率的定义,一般称其为概率的统计定义。

**随机事件的概率:**在  $n$  次重复试验中随机事件  $A$  发生的次数记为  $m$ ,当  $n$  很大时,频率  $m/n$  会稳定地在某一数值  $p$  的附近摆动,而且随着试验次数  $n$  的增加,其摆动的幅度越来越小,称  $p$  为随机事件  $A$  的概率,记为

$$P(A) = p$$

例如,在掷硬币的试验中,“出现正面”这一随机事件发生的频率在 0.5 附近摆动,且随着试验次数的增多,摆动的幅度会越来越小,因此,可以认为“出现正面”这一随机事件的概率为 0.5。

频率与概率之间的关系是非常密切的,也正因为如此,它们具有一些相同的性质。

对频率,由于事件发生的次数总满足  $0 \leq m \leq n$ ,因此有

$$0 \leq m/n \leq 1$$

而对不可能事件  $\phi$  必有  $m=0$ ,对必然事件  $\Omega$  一定有  $m=n$ ,可知它们的频率为

$$W(\Omega)=1, \quad W(\phi)=0$$

对概率,类似地有

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

以及

$$P(\Omega)=1, \quad P(\phi)=0$$

### 四、小概率事件原则

一般地,若  $P(A) \leq 0.05$ ,则称事件  $A$  为小概率事件。小概率事件在一次试验中几乎不可能发生,这一原则称为小概率事件原则。

小概率事件原则是统计学中由样本推断总体的重要原则,在以后的学习中将会多次用到此原则,图 1-1 所示的摸球模型是典型的由样本推断总体的问题。



图 1-1 摸球模型

试根据所取 10 球的结果推断:盒中最多只有 8 个黑球这一说法能否成立?

上述问题的解决就需用到小概率事件原则,该问题将在第五节给予回答。