

# 实用体育统计学

◎ 赵书祥 编著



北京体育大学出版社

# 实用体育统计学

赵书祥 编著

北京体育大学出版社

责任编辑 孙宇辉 毕虹  
审稿编辑 鲁牧  
责任校对 世力 毕莹  
责任印制 陈莎

图书在版编目(CIP)数据

实用体育统计学/赵书祥编著. -北京:北京体育大学出版社,2005.10  
ISBN 7-81100-429-1

I. 实… II. 赵… III. 体育统计 IV. G80-32

中国版本图书馆CIP数据核字(2005)第099577号

实用体育统计学

赵书祥 编著

---

出版 北京体育大学出版社  
地址 北京海淀区中关村北大街  
邮编 100084  
发行 新华书店总店北京发行所经销  
印刷 北京市昌平阳坊精工印刷厂  
开本 850×1168毫米 1/16  
印张 17

---

2005年10月第1版第1次印刷

定价 37.00元

(本书因装订质量不合格本社发行部负责调换)

# 前 言

随着我国全民健身计划和奥运争光计划的进一步推进,随着 2008 年北京奥运会的即将到来,体育统计在体育科研中的重要作用进一步突现。而计算机的进一步普及已使原有的以计算器为计算工具的体育统计教学方法落后于时代的发展了。所以体育统计学的教学改革也应适应信息时代的要求,借助于先进的统计分析软件,走电算化道路。

笔者从事体育统计教学十多年,强烈地感受到在教学中如果过多地讲授统计理论,会使学生感觉内容枯燥,虽耗费了许多时间,但在实际应用中还是无所适从。我校学生的专业不同,学生的数理基础也不同,对统计知识的需求不一致,但每个专业都编一套统计教材是不实际的。鉴于此,本教材从统计概念及与各个专业有关的实例来满足各个专业学生的需要。

SPSS 是社会科学用统计软件包,它具有统计功能齐全、操作简便,适合于体育类学生与体育科研工作者在实际中使用的特点,本书的实例主要用 SPSS 软件进行处理。

与同类教材相比,本教材从应用的角度出发,以数据的收集与整理、描述统计学、参数与非参数推断统计学、简单的多元统计分析为主线,结合大量体育社会科学、体育生物学及体育管理等科学中的实例,直观、系统、全面地介绍统计基础理论并与 SPSS 软件相结合。注重培养和提高体育类学生运用统计软件的能力,即如何根据实际问题选择正确的统计方法、怎样进行简单或复杂的统计计算、怎样对 SPSS 软件的计算结果给出正确的统计解释。本书以应用为主,理论结合实践,通俗易懂,培养学生的动手及实际工作能力,是一本适合于我国体育类学生用的电子教材,同时也是体育科研工作者的应用参考书。

在本书的编写过程中,参阅了许多相关的书籍与教材,引用了其中一些材料并吸收了各书的经验,同时还得到了校领导与教研室其他老师的帮助,在此一并表示衷心的感谢。

本书的不足与缺陷之处在所难免,恳请读者批评指正。

# 目 录

<b>第一章 体育科研中常用的统计概念</b> .....	(1)
<b>第一节 统计历史基本介绍</b> .....	(1)
一、统计的产生与发展 .....	(1)
二、体育统计学简介 .....	(2)
<b>第二节 体育科研中常用的统计基本概念</b> .....	(3)
一、总体和样本 .....	(3)
二、统计误差 .....	(3)
三、统计量与参数 .....	(4)
<b>第三节 体育科研中主要涉及的统计内容</b> .....	(5)
一、描述统计学 .....	(5)
二、推断统计学 .....	(5)
三、统计研究设计 .....	(5)
<b>第二章 体育科研数据的收集方法</b> .....	(7)
<b>第一节 体育科研中变量与数据的概念</b> .....	(7)
一、变量的类型 .....	(7)
二、体育科研中数据的结构与特点 .....	(8)
<b>第二节 体育科研数据的收集方法</b> .....	(9)
一、调查法 .....	(9)
二、实验法 .....	(17)
三、次级资料的收集方法 .....	(21)
<b>第三章 数据文件的建立与整理方法</b> .....	(23)
<b>第一节 SPSS 软件的安装、启动和退出方法</b> .....	(23)
一、SPSS for Windows 的安装、启动和退出 .....	(23)
二、SPSS for Windows 的窗口和对话框 .....	(24)
<b>第二节 如何使用 SPSS 软件创建数据文件</b> .....	(27)
一、数据文件的创建方法 .....	(27)
二、数据文件的操作 .....	(33)
<b>第三节 数据文件的整理</b> .....	(37)
一、数据的审核 .....	(37)
二、数据的整理 .....	(39)
<b>第四节 常用统计图制作方法</b> .....	(48)

一、条形图 .....	(48)
二、饼图 .....	(50)
三、直方图 .....	(51)
<b>第四章 描述统计与基本统计分析方法 .....</b>	<b>(56)</b>
<b>第一节 单变量样本数据的描述统计分析 .....</b>	<b>(56)</b>
一、样本数据平均水平的度量指标 .....	(56)
二、样本数据变异性的度量指标 .....	(61)
三、样本数据的其他度量指标 .....	(66)
<b>第二节 二维以上变量样本数据的分类汇总方法 .....</b>	<b>(69)</b>
一、由原始数据生成列联表 .....	(70)
二、由现有的表生成列联表 .....	(70)
<b>第三节 如何用 SPSS 软件进行描述统计分析举例 .....</b>	<b>(71)</b>
<b>第五章 概率与分布 .....</b>	<b>(80)</b>
<b>第一节 概率的意义及一些相关概念 .....</b>	<b>(80)</b>
一、随机事件 .....	(80)
二、概 率 .....	(80)
<b>第二节 随机变量的分布特征 .....</b>	<b>(82)</b>
一、离散型随机变量的概率函数 .....	(82)
二、连续型随机变量的概率密度函数 .....	(83)
三、随机变量的分布函数 .....	(83)
<b>第三节 体育科研中常用的正态分布 .....</b>	<b>(84)</b>
一、正态分布的意义 .....	(84)
二、标准正态分布表的使用方法 .....	(85)
三、正态分布在体育中的应用 .....	(86)
四、体育科研中常用的数据正态性检验方法 .....	(91)
五、用 SPSS 软件进行正态分布计算举例 .....	(92)
<b>第四节 体育科研中常用的其他分布 .....</b>	<b>(96)</b>
一、 $\chi^2$ 分布 .....	(96)
二、t 分布 .....	(97)
三、F 分布 .....	(98)
<b>第六章 体育科研中怎样使用区间估计 .....</b>	<b>(100)</b>
<b>第一节 抽样误差的概念 .....</b>	<b>(100)</b>
一、抽样误差的概念 .....	(100)
二、抽样误差的计算方法 .....	(100)
<b>第二节 体育科研中总体参数的区间估计方法 .....</b>	<b>(102)</b>
一、点估计 .....	(102)
二、区间估计 .....	(103)
<b>第三节 进行区间估计所需样本量的计算方法 .....</b>	<b>(106)</b>
一、正态总体均数置信区间的样本量求法 .....	(106)

二、总体率置信区间的样本量求法 .....	(107)
第四节 用 SPSS 软件处理区间估计问题的实例 .....	(107)
<b>第七章 假设检验</b> .....	(113)
第一节 假设检验的基本概念 .....	(113)
一、假设检验的意义 .....	(113)
二、假设检验的建立 .....	(114)
三、统计显著性和实际显著性 .....	(114)
四、参数检验和非参数检验 .....	(116)
第二节 单个总体均数的假设检验方法 .....	(116)
一、单个总体均数的参数检验方法 .....	(116)
二、单个总体均数的非参数检验方法 .....	(118)
第三节 二个总体均数比较的假设检验方法 .....	(121)
一、独立组与成对组的区分 .....	(121)
二、两独立组均数比较的假设检验方法 .....	(121)
三、两成对组(相关样本数据组)均数比较的假设检验方法 .....	(129)
<b>第八章 体育科研中多组比较问题的统计处理方法</b> .....	(138)
第一节 参数检验的方差分析法 .....	(138)
一、方差分析概述 .....	(138)
二、单因素方差分析 .....	(139)
三、双因素方差分析 .....	(146)
第二节 k 个独立样本均数比较的非参数检验 .....	(161)
一、Kruskal-wallis 单向评秩方差分析 .....	(161)
二、中位数检验 .....	(163)
<b>第九章 体育科研中如何使用率的检验</b> .....	(168)
第一节 单个总体率的假设检验 .....	(168)
一、单个总体率假设检验的原理 .....	(168)
二、单个总体率假设检验的实例 .....	(168)
三、用 SPSS 软件进行单个总体率检验的实例 .....	(168)
第二节 两个总体率的假设检验 .....	(170)
一、两个总体率的正态 U 检验法 .....	(171)
二、两个总体率的 $\chi^2$ 检验 .....	(171)
第三节 二维列联表齐性和独立性的 $\chi^2$ 检验 .....	(179)
一、 $r \times c$ 列联表的独立性检验 .....	(179)
二、 $r \times c$ 表的齐性检验 .....	(180)
三、 $r \times c$ 列联表齐性和独立性检验的实例 .....	(181)
四、 $r \times c$ 列联表检验的注意事项 .....	(182)
五、用 SPSS 软件进行 $r \times c$ 列联表检验的实例 .....	(182)
<b>第十章 相关与回归</b> .....	(186)

第一节 相关分析	(186)
一、线性相关	(186)
二、定序变量的相关分析	(191)
三、定类变量间的相关分析	(194)
四、多个连续型变量间的相关分析	(195)
五、用 SPSS 软件处理相关分析的实例	(196)
第二节 线性回归分析	(201)
一、一元线性回归分析	(201)
二、多元线性回归分析	(207)
三、线性回归分析问题的 SPSS 处理实例	(212)
第三节 路径分析	(222)
一、路径分析的基本原理	(223)
二、路径分析的基本步骤	(224)
三、应用路径分析的注意事项	(227)
<b>第十一章 试验设计</b>	(231)
第一节 试验设计简介	(231)
一、因素与水平	(231)
二、正交表	(232)
第二节 正交试验设计	(233)
一、试验安排	(233)
二、直接看与算一算	(235)
三、方差分析	(236)
<b>参考文献</b>	(240)
<b>附表</b>	(241)



# 第一章 体育科研中常用的统计概念

## 第一节 统计历史基本介绍

### 一、统计的产生与发展

统计学作为一门科学是如何产生的,又是如何发展的,这是学习这门科学首先要了解的问题。统计学是适应国家管理和社会经济发展的实践需要而产生和发展的,随着人类的进步,内容不断得到充实,逐步形成了统计学的科学体系。从历史上看,统计实践远远早于统计科学的产生,统计实践已有几千年的历史,而统计学的历史才不过二三百年来。

统计的起源极早,它萌芽于奴隶社会。当时为了赋税、徭役、征兵的需要,就开始对人口、土地等重要经济资料进行统计调查和简单的计算工作。人类进入封建社会和资本主义社会以后,统计工作在社会经济发展和国家管理中的作用越加重要,使得统计的范围不断扩大,统计方法不断改进。

随着统计工作的日益广泛发展,实践经验的不断丰富,加上数学及其他科学发展的影响,从17世纪至19世纪中叶,陆续出现了一些统计著作,并开始出现了不同的统计学派。

记述学派,也称为国势学派,这一学派的创始人物是德国的海门尔·康令(1606—1681)和其后的高特弗瑞德·阿痕瓦尔(1719—1772)。康令从1660年起就在大学中系统地讲授国势学,其内容涉及各国社会秩序、立宪、行政、人口、土地、财政、国家组织与结构等等。阿痕瓦尔于1748年起在德国格廷根(Gottingen)大学讲授国势学,并于1748年出版了《欧洲各国国势学绪论》,于1749年出版了《近代欧洲各国国势学论》两部名著。在《近代欧洲各国国势学论》中,他首先使用了“统计学”这个名词,指出统计学(即国势学)是研究一国或多国的显著事项之学,但其统计学仍采用文字记述的方法来记述显著事项,与现代统计学有很大差别(现代统计学是以现象的数量为研究内容),事实上,它虽有统计学之名,但无统计学之实。

政治算术学派,它的创始人和代表人物有威廉·配第(1623—1678)和格朗特(1620—1670)。政治算术学派以数量分析为特征,研究社会经济现象之间的联系的方法与现代统计学相吻合,被认为是具有统计学之实而无统计学之名的统计学。威廉·配第的著作《政治算术》运用大量的实际统计资料,用计量和比较的方法分析了荷兰、法国和英国的国力,是政治算术学派统计学的一部代表作,奠定了统计学的基础。

政治算术学派和国势学派在相互争论中相互影响与发展,19世纪中叶,当代著名的德国经济学家和统计学家克尼斯(K. G. A. Knies, 1821—1897)于1850年发表的论文《独立科学的统计学》中,提出将《国家论》作为《国势学》的科学命名,而将《统计学》作为政治算术的科学命名,从此统计学才名副其实。

数理统计学派,这个学派产生于19世纪中叶,创始人是比利时的凯特勒(L. A. J. Quetelet, 1796—1874),他把当时已盛行欧洲的国势学、政治算术和在对机会游戏观察与研究基础上产生的概率论融合起来,形成了以概率论为基础的对随机现象的实验数据进行处理和推断的现代统计学。凯特勒被称为

“现代统计学之父”，他认为无论自然和社会现象都是有规律的，认识这些规律必须通过实证的大量观察法，并且可以用概率计算表明规律的存在形式。其后经英国的高尔顿(F. Galton, 1822—1921)、卡尔·皮尔逊(K. Pearson, 1857—1936)、费歇(R. A. Fisher, 1890—1962)等很多统计学家的不断丰富和发展，逐渐形成为一门具有独立体系的应用数学学科。

凯特勒把概率论引进统计学中来，既是统计学发展到新阶段的一个标志，又是统计学一分为二的起点。也就是说，数理统计学实质上是以概率论的面貌进入社会经济统计学，后来又以数理统计学的的面貌分立出去。这就是数理统计学的由来及其与古典统计学的关系。

社会统计学派，主要代表人物是德国的恩格尔(C. L. E. Engel, 1821—1896)和梅尔(C. G. V. Mayer, 1841—1925)等。社会统计学派实际上融合了记述学派和政治算术学派的观点，又继承和发扬了凯特勒强调的研究社会现象的传统，并把政府统计与社会调查融合起来，进而形成社会统计学。最早出现的著名著作是挪威人凯尔和汉林(A. N. Kiaer & E. Hanssen)于1898年合著的《社会统计学》。

社会统计学派的前期人物坚持认为统计学是独立的实质性社会科学，强调因果关系和规律的探索，在实质论与方法论的争论中基本上是站在实质论方面的。早期的应用主要是与人口有关的问题(如早年人们关心的一个重大问题是生男生女的比例问题)。社会统计学派的后起者，逐步从实质论向方法论转化。其主要人物的观点都强调统计学是调查研究的方法。社会统计学派虽然向方法论转化，但仍强调以事物的质为前提和认识质的必要性。

随着社会的进步和科学技术的发展，现代统计学已涵盖了社会科学、自然科学的各个领域，其理论基础是数理统计学。而数理统计学是数学的一个分支，由一系列的公理、定理以及严格的数学证明组成，涉及微积分、概率论和高等代数等深奥的数学知识，人们公认数理统计是一门“收集和分析带随机性的数据的科学和艺术”，为使这门系统的科学和理论适合于各个不同领域的研究者，人们将其过程简化，变得非数学化，由此产生了一般统计学，一般统计学与不同的专业领域相结合，就产生了各种专业统计学，如：生物统计学、农业统计学、医学统计学、社会统计学、经济统计学等等。体育统计学也是其中的一门应用统计学科。

## 二、体育统计学简介

体育统计学主要是数理统计学方法在体育领域中的应用。因为在体育运动中存在着大量的随机现象，这些现象伴随着大量随机性的数据出现，对这些数据做出正确的解释并分析、预测各种体育随机现象，使得数理统计学得以渗透到体育中来，并为体育科研服务。由于体育现象的复杂性，它涉及到生物学、心理学、社会学、医学和管理学等多门学科，因此，在一定意义上说，体育统计应是生物统计学、心理统计学、社会统计学、医学统计学等多学科的综合应用。

20世纪40年代前后，我国在体育师范学校就设置了体育测验与统计课程，已故的体育教育家徐英超教授是我国开展体育统计教学与研究的先驱，并为我国的体育统计的发展做出了巨大贡献。在过去很长一段时间内，由于体育工作者认识程度不够，体育统计并没有得到足够的重视。20世纪80年代初，伴随着中国体育科学学会体育统计专业委员会的成立，体育统计才在我国体育院校中普及开来。近年来，体育生物科学、体育社会科学、运动心理学等领域的研究进一步深入，同世界各国研究者接触与交流越来越广泛，体育统计对体育科研的重要作用得到了空前重视。为能更加准确、科学地分析各种体育随机因素之间复杂的关系，体育科研工作者们对统计方法的需求越来越迫切，且层次也逐渐加深，初等统计方法已不能满足他们的要求，这些都客观地推动了体育统计的发展。现代科学技术的迅猛发展，使得计算机得以进入家庭，并且不需太多数学基础就能熟练掌握的各种统计专业软件，为体育统计应用的发展提供了强有力的支持。

由于体育领域的研究者们普遍数理基础知识水平较低，目前我国的体育统计应用还处于选用正确

的统计方法处理体育相关数据的层次上,在应用上有很多方面还有待进一步发展。

## 第二节 体育科研中常用的统计基本概念

### 一、总体和样本

#### (一) 总 体

总体和样本是统计学中最基本的概念。总体(population)就是所研究(调查)对象的全体,总体是一些值的集合。不要把一个总体看成是一些人或东西的集合,应把它看成是这些人或东西的某个指标测量值的集合(在多元统计分析中是这些人或东西的某些指标值的集合)。

总体中的每一个具体研究对象称为个体;总体中所包含的个体数叫总体含量,一般用字母  $N$  表示;如果总体含量具有上限,称该总体为有限总体,否则称为无限总体。

例如,参加某运动会百米比赛的 32 个运动员的百米成绩就构成一个有限总体。在全国群众体育活动状况调查中,全国所有 16 岁以上的人的各种指标值就构成了一个多元有限总体。

#### (二) 样本与随机样本

##### 1. 样 本

样本(sample)也是一些值的集合,但它与总体之间是有区别的。例如,研究全国体育彩票的销售情况,总体代表了全国销售体育彩票的所有省市的销售情况的整个集合,而样本只代表上面这个总体的一部分(比如北京和宁夏),(图 1-1)揭示了总体和样本之间的关系。

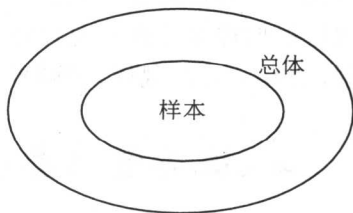


图 1-1 总体与样本之间的关系

##### 2. 随机样本

对于无限总体来说,不可能收集到总体中所有的个体数值;即使对有限总体来说,由于总体含量  $N$  有时相当大,收集到整个总体的所有个体数值也是不可能的,且也没有必要,通常只能收集到总体中一部分个体的数值,即样本,样本中所包含的个体数叫样本含量,用  $n$  表示。为了根据样本对总体做出有效的推断,这个样本必须是总体的典型代表,而这要通过抽取随机样本实现。

在抽样过程中,如果从总体中抽取的任何一个个体都有同等的机会被抽到样本中来,这个样本称为简单随机样本。

在大多数情况下,收集一个随机样本是不容易实现的。例如当总体含量  $N$  相当大时,经费、时间等都是不允许的。简单随机样本仅仅是众多抽样方法中的一种,其他抽样方法也许在某些情况下更可取,这在以后章节再进行介绍。

### 二、统计误差

测量是一种人为的作业,测量的结果常与实际情况有出入而造成测量误差。任何一项统计工作都不可能完全避免误差,关键是误差要得到有效控制。

在调查和统计过程中所得到的数据或指标,与客观实际数量特征之间存在的差别,统称统计误差。例如对北京市海淀区居民进行体育消费的调查时,由样本算出每人月均体育消费额为 30 元,而实际为 35 元(一般很难得到其实际数值),统计误差为 5 元。测量误差只是统计误差的一

种形式。

根据产生统计误差的原因不同,统计误差分为登记性误差和代表性误差两大类。

### (一) 登记性误差

登记性误差也称调查误差,它是在调查、编辑、编码和汇总过程中由于观察、测量、登记、计算上的差错或被调查者提供虚假资料而引起的误差。这种误差的直接表现就是不能真实客观地搜集或记录被调查单位的标志值或标志特征,从而使样本统计量的值偏离其真实性。登记性误差在所有统计调查中都可能存在,并且调查的范围越广,调查的单位越多,产生登记性误差的危险就越大。这类误差与调查测量工具(仪器)的精确性、调查测量技术、调查人员的工作责任心、被调查者的合作态度等密切相关。

### (二) 代表性误差

代表性误差指样本不能完全代表总体而产生的误差,其原因又有系统性与偶然性之分。系统性的代表性误差简称系统误差,是一种非随机性误差,它包括以下一些可能的情况:有意选择了总体中较特殊的单位进行调查;调查表设计不当;样本仅限于总体中易抽取的一部分等。系统性误差与登记性误差合在一起称为偏差。偶然性的代表性误差简称偶然性误差,是指按随机抽样原则下,由于偶然因素引起的样本结构不能完全代表总体结构而产生的误差,是抽样调查所不可避免的误差,这就是我们通常所说的抽样误差。全面调查对总体中每个个体都进行调查,没有代表性误差。而进行抽样调查(按一定的程序,从全体调查对象—总体中抽取一部分样本进行调查,然后根据样本数据对总体特征数进行估计)时,两种误差都可能发生,不可避免,但可以计算和控制,它有一定的规律,即随样本含量的增大而减少。

## 三、统计量与参数

总体与样本之间的另一个差别是计算概括度量值(特征数)时所用的方法不同,而且名称也不同。总体的度量值(特征数)叫参数;而样本的度量值叫统计量。

### (一) 参 数

任何一个由总体中所有数据计算得到的描述这一总体特征的数值称为参数(总体参数)。注意,当研究总体确定以后,总体参数就随之确定,一般是未知的常数,常用希腊字母表示。

### (二) 统计量

由总体中的样本数据计算得到的描述样本特征的数值称为统计量(样本统计量)。样本统计量是变量,每次抽样都能得到相应的统计量,常用英文字母表示。

本书中常用的三种概括度量值:均值、方差、标准差。均值用以描述数据值分布的中心位置,而方差和标准差是用于描述数据值相对于均值的离散程度,标准差是方差的平方根。计算方法下面章节再介绍。

样本均值记为  $\bar{x}$ , 总体均值记为  $\mu$ ;

样本方差记为  $S^2$ , 总体方差记为  $\sigma^2$ ;

样本标准差记为  $S$ , 总体标准差记为  $\sigma$ 。

### 第三节 体育科研中主要涉及的统计内容

统计学是关于数据资料的收集、整理、分析和推断的一门科学,它包括的内容从统计历史发展的顺序依次是:描述统计学→推断统计学→统计研究设计;从一次完整的统计活动来讲,一个完整的统计过程依次是:统计研究设计→描述统计→推断统计。

#### 一、描述统计学

描述统计学给出的是将原始数据资料加工成统计图表的方法,这些方法包括数据的收集、整理、计算得到若干能代表总体或样本特征的概括或描述统计量,并设法找出原始数据的分布特征。

#### 二、推断统计学

如果在研究中可以得到整个总体,那么描述统计学就足够了,但是,实际工作中往往只能得到总体中的一小部分个体(样本),所以必须由样本的特征去推断总体的特征,或者说由样本所带有的有限的、不确定的信息推断有关总体的信息,这就是推断统计学的内容。

推断统计学提供了由样本特征推断总体特征的方法,并且还给出推测的可靠性以及推测值接近真实值的程度。包括参数估计和假设检验两部分内容。

##### (一) 参数估计

在参数估计问题中,测量总体的未知参数是用相应测量样本的统计量来估计的,例如:用样本均值  $\bar{x}$  估计总体均值  $\mu$ ;用样本方差  $s^2$  估计总体方差  $\sigma^2$ 。包括点估计和区间估计两部分内容。

##### (二) 假设检验

统计假设是指关于一个或多个待研究总体的未知性质,特别是关于它们的参数或它们的分布形式的假设或猜测。包括参数检验和非参数检验两部分内容。

(图 1-2)给出了单个总体的某个样本的估计和假设检验的步骤。图中研究总体的均值  $\mu$  是未知的,为了估计并进行假设检验,首先必须从这个研究总体中抽取一个样本,然后进行测量并获得样本测量值,由样本均值和其他样本统计量,根据推断统计的原理去推断总体均值  $\mu$  并进行假设检验。

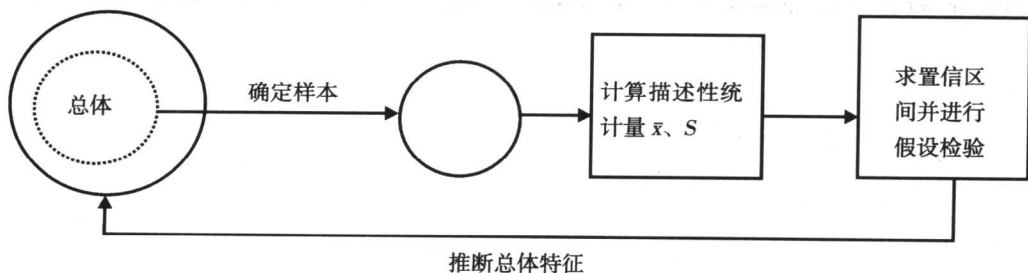


图 1-2 推断统计的步骤

### 三、统计研究设计

统计工作大致分为四个工作过程,即统计研究设计、统计调查(或实验)、统计整理、统计分析四个阶段。统计研究设计是开始阶段,是根据研究对象的性质和目的,在进行统计研究工作之前,对统计工作的各个方面和全部过程所做的通盘考虑和周密安排。通常我们把统计研究设计分为两种,一种是调查设计,一种是实验设计。调查设计一般在体育社会学研究中应用较多,而实验设计(试验设计)在体育人体科学和运动心理学研究中应用较多。

#### 本章提要

统计历史介绍:统计的产生与发展,记述学派、政治算术学派、数理统计学派、社会统计学派、体育统计学。体育统计基本概念:总体与样本、随机样本、统计误差、统计量与参数。体育统计的内容:描述统计学、推断统计学、统计设计。

#### 思考与练习

1. 举例说明什么是总体与样本。
2. 指出下列总体是有限总体还是无限总体;
  - (a)所有参加2004年雅典奥运会的男运动员;
  - (b)所有患有心脏病男子的心率。
3. 请系统地阐述误差的概念与内容。
4. 什么叫总体参数?什么叫样本统计量。
5. 简述体育统计的内容。

#### 推荐阅读参考文献

1. 陈及治等(2002),《体育统计》第一章,人民体育出版社。此书对体育统计的历史作了详细的介绍。
2. 赵振伦等(1997),《社会经济统计学原理》第一章,东北财经大学出版社。此书简要介绍了统计学的历史。
3. 高惠璇等(1995),《SAS系统与基础统计分析》,北京大学出版社。此书简明介绍了总体和样本的概念。
4. 金勇进(1996),《非抽样误差分析》,中国统计出版社。此书对统计误差的概念作了透彻的分析。
5. 隗斌贤等(1997),《调查统计学》,中国统计出版社。此书对抽样误差的概念作了透彻的分析。

## 第二章 体育科研数据的收集方法

### 第一节 体育科研中变量与数据的概念

#### 一、变量的类型

在某一过程中保持同一数值的量称为常量；而在某一过程中可以取不同数值的量称为变量。

例：在计算圆面积公式  $S = \pi r^2$  中， $\pi$  是常量， $S$  与  $r$  是变量。

在统计学中，变量是指所研究的对象（人，时间，距离等）的属性特征，这些特征是可以测量的，并且对于一组对象中的不同对象可取不同的值。例如，研究 100 米运动员的素质，则其体重、立定跳远成绩都是变量。

变量的每一测量值称为观测值（或指标值），一般用大写字母表示变量，而用同一字母的小写表示它的观测值（例如  $X$  表示变量， $x$  表示它的观测值）。

变量的测量结果就构成了数据，数据可按不同的属性分成不同的类别，也就是说变量有不同的类型。变量的类型不同，其分布的规律也不同，对它们作统计处理的方法也有所不同。处理数据之前按变量分清它们的类型是很重要的。

#### （一）变量的测量尺度

为说明变量的类型，首先要分清变量的测量尺度（或者说量化尺度）。变量的量化尺度有四种：即名义尺度（定类尺度）、次序尺度（定序尺度）、区间尺度（定距尺度）、比例尺度（定比尺度）。

##### 1. 名义尺度

名义尺度或称定类尺度，它将事物分到唯一的类中，这些类别之间是平等的并列关系，无法区分优劣或大小，且在所分的全部类别中，必须保证每一事物都能归属于某一类别，且只能归于一个类别。

名义尺度是最粗略、测量层次最低的计量尺度。例如，按性别将运动员分成男子组、女子组；按比赛项目可分成：跳远、跳高、100 米、200 米等。在统计处理时，名义尺度的变量可以用数字表示，比如用“1”表示男性，“2”表示女性，但只能计算每一类别中每个个体出现的频数或频率，而不能区分大小或进行任何数学运算。

##### 2. 次序尺度

次序尺度又称定序尺度，它不仅保留了名义尺度的性质（任何事物都能归属于某一类别，且只能归于一个类别），而且这些类可以根据被测量特征等级进行排序，每一类都可以与其相邻类进行比较，来判断是大于还是小于。例如，运动员按运动级别分成国际运动健将、运动健将、一级、二级及三级运动员。

次序尺度对事物的计量水平要比定类尺度精确一些，但它只是度量了类别间顺序，而未测量出类

别之间的准确差值。因此其计量结果只能比较大小,不能进行加、减、乘、除等数学运算。

### 3. 区间尺度

区间尺度又称定距尺度,它不仅能将事物区分为不同类别并进行排序,而且可以准确地计量出它们之间的差距是多少。使用区间尺度时,通过尺度值之间的加减运算,可以得到两个事物间关于被测量特征的确切差别,但是该测量尺度有任意的但不是绝对的零点,不能进行乘除运算。

例如,考试成绩用百分制度量,温度用摄氏或华氏度来度量;10摄氏度比5摄氏度高5摄氏度,但不能说前者比后者热一倍。

### 4. 比例尺度

比例尺度又称定比尺度,它在测量中的量化水平是最高的,它除了具有上述三种尺度的全部特性外,还有一个特性,那就是可以计算两个测量值之间的比值。它与定距尺度的唯一区别是具有一个绝对固定的、非任意确定的零点。在比例尺度的测量中,零值是所测量的物体的特征不能出现或不能观测到的一点,在该尺度中,数值表示的是离绝对零点的距离。

比例尺度的实际例子如:体重(千克)、身高(厘米)、百米跑时间(秒)、速度(米/秒)等。

上述四种测量尺度对客观事物的度量层次或水平是由低级到高级、由粗略到精确逐步递进的。高层次的测量尺度可以度量低层次计量尺度能够度量的事物,但不能反过来。在实际进行统计计算时,适用于低层次度量水平数据的统计方法也适用于高层次度量水平的数据(将高层次测量水平的数据转化成低层次测量水平的数据),但不能反过来。

## (二)变量的类型

在这里我们说的变量的类型也就是指数据的类型。变量的测量结果就构成了统计数据,不同的测量尺度会得到不同类型的统计数据。从上述四种测量尺度形成的结果来看,我们将变量大体上分为两种类型:即连续型变量和离散型变量。

### 1. 连续型变量

连续型变量理论上可取无穷多个数值。诸如身高、体重、百米跑成绩、脉搏等均可经测量取得数值的变量,限于测量精度,身高、体重、百米跑成绩并不能取任意位小数,脉搏只能取正整数,在实际中为了研究方便,人们将这类变量视为连续型变量。有时也把连续型变量的测量数据称为计量资料。

### 2. 离散型变量

某些属性只能定性地划分成少数几个互相排斥的类别,离散型变量在任意给定范围内取有限个数值。如:性别作为变量,“男”取值“1”,“女”取值“0”;考试得分作为变量,其范围通常从0到100,那么该变量的取值可能为0,1,2,⋯,100;篮球比赛中的进球数也是一个离散变量等。有时也把一些离散型资料称为计数资料。

### 3. 变量的类型与测量尺度之间的对照

用名义尺度和次序尺度测量的变量,通常作为离散型变量;用区间尺度和比例尺度测量的变量,既可以作为离散变量,也可以作为连续型变量,要结合实际,具体问题具体分析。

## 二、体育科研中数据的结构与特点

任何实验和调查的结果必须转变为数据后才能进行统计分析。(表2-1)中列出了2000年北京市海淀区城市居民家庭体育消费结构调查的部分数据,有7个记录项目,这种数据结构可列成一个60行7列的一个数据矩阵,SPSS等统计软件都以这种形式作为数据录入的基本格式。



表 2-1 60 名海淀区城市居民家庭体育消费结构调查表

性别	年龄	职业	周余暇时间 (小时)	周活动时间 (小时)	家庭月均收入 (元)	年体育消费额 (元)
男	19—24	工人	18	8	2400	600
男	19—25	教师	22	12	3600	1500
女	35—44	职员	24	4	3000	3000
女	15—18	学生	30	10	5000	1000
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
女	45—49	教师	16	6	3000	3800

### (一) 基本观察单位

基本观察单位是按研究需要确定的收集数据的基本单位。研究对象本身可以是一个基本观察单位,也可以同时具有若干个基本观察单位。以 60 名海淀区城市居民家庭体育消费结构调查为例,如果以居民 2000 年体育消费额与周体育活动时间为研究指标,则一个居民是一个基本观察单位;如果以 2000—2003 年居民体育消费额与周体育活动时间为研究指标(纵向调查),每个居民具有 4 个基本观察单位。

### (二) 数据指标

用于统计分析的数据指标(变量)通常由分组因素、反应变量和协变量组成。同一份数据,不同的研究目的也会出现一定的差别。如(表 2-1)为 60 行 7 列的数据矩阵,如要研究不同性别居民体育消费结构,则性别是分组因素,家庭月均收入、年体育消费额为反应变量,年龄、职业、周余暇时间及周体育活动时间为协变量;如要研究不同年龄居民的体育消费结构,则年龄是分组因素,家庭月均收入、年体育消费额为反应变量,性别、职业等为协变量。

## 第二节 体育科研数据的收集方法

体育科研数据的收集方法主要有调查法和实验法两种主要渠道,使用这两种方法都可获得直接的第一手统计数据。无论采用哪种方法收集数据,在进行之前都需要制定一个周密、完整的调查(实验)方案,以指导整个工作按计划进行,使调查(实验)得以顺利实施和完成。另外,对别人已收集到的数据资料进行收集和整理,使之符合研究的需要,也是一种重要的收集数据的方法。

### 一、调查法

调查法是体育科研中最常见的方法,现在已普遍应用于体育社会科学研究和体育人体科学研究中。1979 年和 1985 年进行的两次全国青少年体质研究中,主要采用了这种方法;1996 年进行的中国群众体育现状的调查与研究中也使用了此法。

#### (一) 统计调查方案的内容

统计调查方案是指导整个调查过程的纲领性文件,因此在设计调查方案时,对调查的目的、任务、对象和采用的方法应做到心中有数。不同的调查在内容上虽有一定的差别,但主要应包括以下内容: