



体育统计与SPSS

王晓芬 编著

人民体育出版社

体育统计与 SPSS

王晓芬 编著

人民体育出版社

(京)新登字 040 号

图书在版编目(CIP)数据

体育统计与 SPSS/王晓芬编著. —北京:人民体育出版社, 2002

ISBN 7-5009-2244-2

I. 体… II. 王… III. 体育统计-统计分析-软件包, SPSS IV. G80-32

中国版本图书馆 CIP 数据核字(2001)第 087445 号

*

人民体育出版社出版发行
北京昌平环球印刷厂印刷
新华书店经销

*

787×1092 16 开本 17 印张 385 千字
2002 年 2 月第 1 版 2002 年 2 月第 1 次印刷
印数:1-4,100 册

*

ISBN 7-5009-2244-2/G · 2143

定价:25.00 元

社址:北京市崇文区体育馆路 8 号(天坛公园东门)

电话:67151482(发行部)

邮编:100061

传真:67151483

电挂:9474

(购买本社图书,如遇有缺损页可与发行部联系)

前 言

体育统计是理论统计学的方法在体育领域中的应用,它以体育运动中随机现象的规律性为研究对象,提供实验设计、收集、整理和分析数据资料的方法,是从事体育科学研究必不可少的一种工具。同时,体育统计作为体育院校的专业基础课之一,它不仅直接影响学生进行科学研究的能力,而且也影响学生对相关课程的学习。

SPSS 是英文 Statistical Package for Social Science 的缩写,即“社会科学统计软件包”。目前,它是世界上通用、并具有权威性的统计分析软件包。虽然它名为“社会科学统计软件包”,实际上它除了在社会科学研究领域中发挥其巨大作用外,在自然科学的各个领域也得到越来越广泛的应用,如经济学、生物学、心理学、医疗卫生、农业、商业、金融等各领域或部门都普遍使用了该软件包。

SPSS 最初是从 DOS 环境(SPSS for Dos version 1.0 - version 4.0)开始,逐步升级发展到 Windows 环境(SPSS for Windows version 5.0 - version 10.0)。2001 年最新版本为 version 10.0。SPSS for Windows 采用窗口方式管理程序运行的全过程,简捷直观,易于操作;通过对话框,可以实现各种命令参数的指定。用户只要掌握基本的 Windows 操作方法,粗通统计分析原理,就可以应用该软件得到具有专业水准的统计分析结果。此外,SPSS 还具有强大的图表功能以及数据库的互接功能,即可以产生各种统计报表和形象直观的统计图形,也可以实现与其他应用程序的资源共享。因此,SPSS 深受广大科研工作人员的青睐和喜爱。

目前已有很多体育统计的教材和有关体育统计方法的专著,也有多种版本关于介绍 SPSS 的书籍。我们发现,体育统计的教材与专著常常因涉及一些复杂的运算过程,使一些学生和科研人员望而生畏;而 SPSS 的有关书籍则又难以很好结合体育运动训练、竞赛及科学研究的实践,针对性不强。《体育统计与 SPSS》一书尝试将体育统计的基本理论与 SPSS 的具体使用有机结合起来,使读者在运用 SPSS 进行体育统计分析时,既能了解相应统计方法的理论基础,又可快速得到统计分析的结果;对所涉及的统计过程以及输出结果也进行了详细的解释或说明;借助 SPSS 图形或表格,使抽象的问题直观化、复杂的问题简单化,同时也使简单的操作具有科学理论的指导。

笔者积多年的教学经验和使用 SPSS 的实践体会,在编写本书过程中,力求语言通俗易懂,使其既可作为 SPSS 使用者的快速查询工具,又可作为体育统计初学者的入门教材。笔者还有一个愿望就是在帮助读者提高学习体育统计效率的同时,也能使读者的科学研究的基本素养得到相应的提高。

本书在编写过程中得到了我院计算机教研室、研究生部、教务处、科研处诸位老师的帮助和支持,在此一并感谢。由于笔者水平所限,书中难免有疏漏或错误之处,诚恳期待广大读者的批评指正。

王晓芬

2001 年 9 月

责任编辑 孙静敏
封面设计 MBF 工作室
版面设计 阎建生

ISBN 7-5009-2244-2



9 787500 922445 >

定价：25.00 元

目 录

第一章 体育统计与 SPSS 概述	1
第一节 体育统计学的研究对象与内容	1
第二节 统计学研究方法	4
第三节 SPSS 统计软件及其特点	6
第二章 体育统计的有关概念	8
第一节 总体与样本	8
第二节 统计资料与测量误差	9
第三节 概率与概率分布	12
第四节 参数估计与假设检验	20
第三章 SPSS 数据输入与编辑	25
第一节 数据编辑器	25
第二节 数据输入	27
第三节 数据文件的编辑	37
第四节 数据文件的存取	41
第四章 SPSS 数据整理	45
第一节 计算生成新变量	45
第二节 数据重编码	48
第三节 产生计数新变量	51
第四节 样品排序、定秩与选择	53
第五节 数据文件的合并	59
第五章 描述统计	67
第一节 描述统计的基本概念	67
第二节 频数分布与描述统计	73
第三节 数据考察	83

第四节 多维频数分布表	90
第六章 均数比较	98
第一节 分组计算描述统计量	99
第二节 T 检验的理论背景	102
第三节 T 检验的操作程序	103
第七章 方差分析	112
第一节 方差分析概述	112
第二节 单因子方差分析的操作程序	115
第三节 多因素方差分析的操作程序	123
第四节 协方差分析的操作程序	131
第八章 非参数检验	136
第一节 卡方检验	136
第二节 样本分布检验	143
第三节 独立样本的检验	149
第四节 相关样本的检验	154
第九章 相关与回归	160
第一节 相关分析概述	160
第二节 相关分析的操作程序	163
第三节 回归分析概述	173
第四节 回归分析的操作程序	176
第十章 聚类与判别分析	189
第一节 聚类分析概述	189
第二节 快速样品聚类的操作程序	195
第三节 分层聚类的操作程序	203
第四节 判别分析概述	219
第五节 判别分析的操作程序	223
第十一章 因子分析	242
第一节 因子分析概述	242
第二节 因子分析的操作程序	245
附:词汇表	256
主要参考文献	263

第一章

体育统计与 SPSS 概述

第一节 体育统计学的研究对象与内容

一、统计学的研究对象与性质

(一) 统计学的研究对象

“统计”一词在我国古代汉语中仅为“统而计之”之意。现代汉语中“统计资料”“统计工作”和“统计学”是三个涵义不同的名词,在不会混淆涵义的情况下,人们常简称为“统计”。但这并不能理解为“统计”一词有三种涵义。在英语中 statistics 一词作为复数形式为统计资料,作为单数形式为统计学。

统计学的研究对象是指统计研究所要认识的客体,它决定着统计科学的研究领域以及相应的研究方法。一般认为,统计学的研究对象是客观事物的数量特征和数量关系。人们要认识客观事物,就必须通过试验或调查来搜集有关数据,并且加以整理、归纳和分析,以便对客观规律性的数量表现作出统计上的解释。这既是统计活动过程,也是人们对客观世界的认识过程。然而,要使统计活动过程能够适时、有效地进行,又离不开统计理论和方法的指导。例如,统计需要哪一类数据?怎样用适当、科学的方法搜集及加工这些数据?怎样从复杂纷繁的数据中获得结论,并解释这个结论?特别是在数据不完全、面对不确定的情况下,如何作出明智的判断?没有统计理论和方法的指导那是无法进行的。因此,可以说统计学是关于数据搜集、整理、归纳、分析的方法论科学。简而言之,统计学是对大量统计数据进行分析研究的科学方法。

统计学的研究对象具有以下明显的特点:

1. 数量性

统计学离不开数字,数量性是统计学对象的基本特点。统计学涉及的数字,不是随意的或单纯的数字,而是能表明客观现象某一方面特征的数字,如人口数、产量数、产品数、气温数、运动员血液中的 EPO(红细胞生成素)数等。当然,并不是任何一种数量都可以作为统计对象,比如一些虚构的数字、预测的数字等,虽然都与统计有关,但是其本身并不属于统计学研究对象的范畴。

统计也不同于抽象的数学运算,统计数据是对客观事物量的反映,通过数来测度事物的类型、量的顺序、量的大小、量的关系,以认识客观规律的量的表现。而且,统计定量认识必须建立在对客观事物定性认识的基础上,统计研究是密切联系现象的质来研究它的量,并通过量反映现象的质,这一点和数学研究抽象的数量关系是不同的。所以,统计数据总是客观的、具体的、也是最具说服力的。

2. 总体性

统计学是以客观现象总体的数量方面作为自己的研究对象,对现象总体中各单位普遍存在的事实进行大量观察和综合分析,从而得出反映现象总体的数量特征。恩格斯曾说:“被断定为必然的东西,是由纯粹的偶然性构成的,而所谓偶然的东西,是一种有必然性隐藏在里边的形式。”(《马克思恩格斯全集》第4卷第240页,人民出版社,1972年)他还说:“在表面上是偶然性起作用的地方,这种偶然性始终是受内部隐藏着的规律性支配着,而问题只是在于发现这些规律。”(《马克思恩格斯全集》第4卷第240页,人民出版社,1972年)客观事物的个别现象通常有其特殊性、偶然性,而总体现象则具有相对的普遍性、稳定性,是有规律可寻的。统计学研究现象总体的数量特征,将其总体的、必然的规律显示出来,有助于我们对现象规律性的认识。

3. 变异性

统计研究同类现象总体的数量特征,它的前提则是总体各单位的特征表现存在的差异,而且这些差异并不是由某种特定的原因事先给定的。例如,运动员从事训练竞赛的时间有差异,运动水平有差异,工资报酬有差异等,这才需要研究运动员的平均训练时间、运动水平等级、平均工资等指标。如果所有运动员在这些方面不存在差异,也就不需要做统计研究;如果运动员之间的差异是按已知条件事先可以推定的,也就不需要采用统计方法。例如,昼夜时间长短因季节变化而不同,这与统计无关,而江河水位高低随时间而不同则是统计研究的对象。统计上将总体各单位由于随机因素引起的某一标志表现的差异称为变异。

如果说,总体各单位的变异表现出个别现象的特殊性和偶然性,而对现象总体的数量研究,则是从各单位的变异中归纳概括出它们的共同特征,显示出现象的普遍性和必然性。这就是统计认识方法的特点。

(二)统计学的性质

统计学有自己的研究对象,也有自己的研究方法。统计学的任务就在于为统计活动提供数量研究和认识规律的科学方法,包括指导统计工作的原理原则,统计过程所应用的分析方法,以及组织方法。其核心内容是数据的搜集、整理、归纳、分析的原理和方法,这些方法论构成了统计学的科学体系。所以,统计学是一门认识方法论科学,它是研究如何搜集数据、分析数据、以便从中作出正确推断的认识方法论科学。

统计学在研究中的作用,除了对于观察数据汇总进行统计描述外,还可对于多变量之间的关系进行定量描述。而且,统计学在方法论上显得日益重要的另一个意义是,它可以用于将实际观察资料与理论假说的推论联系起来对假说进行检验。从这个角度来说,统计学并不是一种孤立的技术方法,它是理论联系实际的一种形式,也可以说是对理论进行实践检验

的一种形式。

统计学和数学都是研究数量关系的,但两个学科有不同的性质特点。数学撇开具体的对象,以最一般的研究数量的联系和空间形式,数学的分析方法主要是逻辑推理和演绎论证的方法,从严格的定义、假设的命题和给定的条件去推证有关的结论。而统计学的数据则总是与此同时客观的对象联系在一起,统计的过程就是从客观对象中抽出其数量表现,得到有关数据,然后加以适当的运算,取得一定的结果。第二步就要把这些结果又返回到客观对象中去,寻求解释这些结果的意义,提供决策的事实依据。统计分析的方法,本质是归纳的方法,根据试验或调查,观察到大量的个别情况,加以归纳来判断总体的情况,这里存在着推断的可信度,以及主观的判断能力等不确定因素。所以,统计学不仅是一门科学技术,而且是一门应用艺术。我们也要看到统计学和数学的密切关系,数学分析方法适用于一切数量分析,也包括统计的数量分析,数学为统计学提供了数量分析的方法论基础,特别是数学中的概率论,它研究随机现象的数量关系和变化规律,它从数量方面体现了偶然与必然,个别与一般,局部与总体的辩证关系,为统计科学的现代化奠定了基础。

统计学和相关的实质性学科,如经济学、社会学、医学、生物学等,有共同的研究对象,但不能认为统计学是一门实质性的科学。实质性的科学是研究该领域现象的本质关系和发展变化规律。而统计学则是为研究认识这些关系和规律提供适当的方法,特别是数量分析的方法。因此,两个学科的性质是不同的。但是,我们也必须认识到,统计学与实质性学科的关系也是非常密切的。首先,由于统计的方法在很大程度上决定于对象的性质,通常是实质性的学科提出了问题,同时社会上有此需要,才有相应的方法,并且有用武之地。其次,在统计分析中往往数字本身未必能够得到解释,这又需要实质性学科的理论研究指导。反之,统计的实证研究又可以检验实质性学科理论的真实性和完善程度。统计归纳分析所获得的新知识往往又为实质性学科开辟新的领域,这在哲学、经济学的历史上都屡见不鲜。

二、体育统计学的性质与研究对象

统计学自身的发展沿着两个不同的方向,形成理论统计学与应用统计学两大类,它们之间的学科性质也有所差异。

理论统计学(亦称数理统计学)将研究对象一般化、抽象化,以概率论为基础,从纯理论的角度,对统计方法加以推导论证,中心的内容是统计推断问题,实质是以归纳方法研究随机变量的一般规律。包括抽样理论、实验设计、参数估计、假设检验、复变数分析(回归分析、相关分析、方差分析、因素分析、判别分析)等。这些方法无论对自然现象还是对社会现象的随机变量都是适用的。

应用统计学是从所研究的领域或专门问题出发,视研究对象的性质采用适当的指标和统计方法,以解决所需研究的问题。如:生物统计学、教育统计学、工业统计学、心理统计学、体育统计学等。应用统计学需要有关的专业实质性学科的理论作指导,它通常具有边缘交叉学科的性质。在统计科学发展的道路上,理论统计学和应用统计学总是相互促进,共同提高。理论统计学为应用统计学提供理论与方法,可以大大提高分析认识的能力;而应用统计

学又为理论统计学提供新的研究内容和资料,提出新的问题,开拓理论统计的研究领域,从而促进理论统计学的深入发展。

体育作为一门科学有其自身的发展规律。体育工作者在实际工作中要想使自己发现认识体育运动的规律,必须有科学的方法,体育统计学正是由于这种需要而产生的。体育运动中的许多现象都是不确定的,这正是以概率论为基础的数理统计所研究的对象。同时,体育运动又是一种极复杂的人体活动,许多问题需依靠实验才能解决,而实验的个体只能是抽取部分受试者进行,这正是统计学所能解决的由部分个体对总体推断。因此,体育领域是理论统计发展应用的一块沃土。

体育统计学作为应用统计学的一个分支,是以体育运动中随机现象的规律性为研究对象,它的内容与应用统计学的内容一样常划分为三个部分:1. 描述统计(descriptive statistics),主要研究如何对搜集到的大量数据进行既能描述该组数据全貌,又能反映所要研究现象的内容和本质的各种简缩数据的方法。其内容包括统计分组,编制统计表,绘制统计图,计算各种统计量。2. 推断统计(inferential statistics),主要研究如何依据样本所提供的信息,对总体的有关特征进行推断的理论和方法,包括总体参数的估计与假设检验、参数检验和非参数检验,还包括方差分析、因素分析、主成分分析、回归分析、聚类分析等。3. 实验设计(design of experiments),主要研究如何按照预定目标制定适当的实验方案,以利于运用数学原理对实验结果进行有效的统计分析。一个实验的设计,即对实验的一个安排,需要考虑实验所要解决的问题类型,对结论赋予何种程度的普遍性,希望以多大功效作检验,试验单元的齐性,选取适当的因子和相应的水平,从而给出实验实施的具体程序和数据分析的框架。

第二节 统计学的研究方法

统计学有自己特定的研究对象,也有专门的研究方法。理解并掌握统计方法对于认识统计对象的规律性和完成统计任务都有重要意义。

一、大数定律的方法论意义

统计研究现象总体的数量特征,所用的基本方法都与数量的总体性有关,其数学依据是大数定律。

大数定律又称大数法则,它是说明大量的随机现象平均结果具有稳定性质的法则,即说明如果被研究的总体数量特征是由大量的相互独立的随机变量形成的,每个变量对总体的影响都相对地小,那么对大量随机变量加以综合平均的结果,变量的个别影响相互抵消,而显现出它们共同作用的倾向,使总体数量特征具有稳定的性质。大数定律正是从数量方面表现了偶然与必然的辩证关系,因此,我们可通过大量随机现象的综合概括,以消除偶然性的误差,发现必然性的趋势,认识规律的表现形式。

大数定律对于认识现象规律性的方法论意义表现在以下几个方面:

1. 现象的某种总体规律只有当具有这些现象的足够多数的单位汇总综合在一起的时候,才能显示出来。因此,只有从大量现象的总体中才能研究这些现象的规律性。

2. 现象的总体性规律,通常是以平均数的形式表现出来的。

3. 所研究的现象总体包含的单位愈多,平均数也就能够正确反映出这些现象的规律性。

4. 各单位共同倾向(这些表现为主要的、基本的因素)决定着平均数的水平,而单位对平均数的离差(这些表现为次要的、偶然的因素)则由于足够多数单位的汇总综合的结果,而相互抵消,趋于消失。

统计研究对象,不论自然现象还是社会现象,都具有随机性。这些现象的出现取决于多种多样的原因,既有必然的因素,也有偶然的因素,它们对于各具体单位所起的作用,在程度上、甚至在方向上都可能不同,这就导致同一现象在每个单位的数量表现具有随机性。统计研究运用大量观察法对总体中的全部或是足够的单位进行调查,并运用综合指标法对各单位变量加以综合,通过平均化的结果以抵消偶然因素的个别影响,描述现象的典型水平,并进一步推断总体的数量特征,这些都是在大量定律的作用下进行的。

应该指出,大量定律的作用是帮助我们通过偶然性达到必然性,认识现象规律的表现形式,但它并不能说明现象的本质,这须借助相关实质性学科的知识来解释现象的本质及其内在联系。

二、统计研究的基本方法

统计研究过程的各个阶段,从统计资料的搜集、数据的整理汇总,以至统计分析检验,都运用着各种专门的方法。在资料搜集时,对于无法从科学试验取得资料的现象,特别如社会经济现象,应运用大量观察法。对于可通过科学试验取得的资料,则应建立在合理的实验设计基础上。在数据的整理汇总时,运用统计描述法,而统计分析与检验主要运用统计推断法。

(一)实验设计

统计研究首先需要考察的是数据的来源是否合适,是否符合分析的目的要求。如果实验设计本身不科学,实验数据不能反映现象的真实情况,或不能用以估计总体的数量特征,那么此后一系列的分析工作只能是一种数字游戏,得出的结论肯定是不客观或毫无意义的。

(二)大量观察

许多统计对象,特别是运动训练和竞赛的现象是已发生的事情,并且无法进行重复实验的。这是因为体育运动的实践本质上是反映人与人之间的关系,这种关系客观地存在于训练、竞赛的现实生活中,要研究这种关系就不能用实验的方法或推理的方法,而必须到体育运动的实践中去作调查研究,即采用大量观察的方法就总体中的全部或足够多数单位进行调查研究,并加以综合研究。复杂的训练、竞赛现象是在诸多因素作用下形成的,个别现象

往往受各种偶然因素的影响,使不同群体或个体的特征和数量表现有很大的差别,所以不能任意抽取个别或少数单位进行观察。必须在对所研究对象的定性分析的基础上,确定调查对象和总体范围,并对总体中的所有单位或足够多数单位的变量进行登记和计算,然后把观察得来的个别数量加以整理汇总,计算相应的综合指标来反映总体现象的数量特征,这就是现象规律性的表现形式。普查、抽样调查、统计报表调查、重点调查等都是大量观察的组织形式。

(三)统计描述

统计描述是指对实验或调查而得到的数据进行登记、审核、整理、归类、计算出各种能反映总体数量特征的综合指标,并加以分析从中抽出有用的信息,用表格或图像形式把它表示出来。统计描述是统计研究的基础,它为统计推断、统计咨询、统计决策提供必要的事实依据。统计描述也是对客观事物认识的不断深化过程。它通过对分散无序的原始资料的整理归纳,运用分组法和综合指标法得到现象总体的数量特征,揭示客观事物内在的数量规律性,达到认识的目的。

(四)统计推断

常存在这种情况,我们所掌握的数据只是部分单位的数据或有限单位的数据,而我们所关心的却是整个总体甚至是无限总体的数量特征。例如,体育锻炼对增强心脏功能是否有益?某种新药是否能提高治疗效果?等等。这时,仅靠部分数据的描述就无法获得总体特性的知识,而只能利用统计推断的方法来解决。所谓统计推断就是以一定的置信标准要求,根据样本数据来判断总体数量特征的归纳推理方法。统计推断是逻辑归纳法在统计推理的应用,所以称为归纳推理的方法。统计推断既可用于总体数量特征的估计,也可用于对总体某些假设的检验。

当然在统计学的研究中,我们必须强调定量分析与定性分析的有机结合。所有的统计方法只是从事物外在的数量上去研究问题,而不是涉及事物质的规定性。单纯表面上的数量关系是否反映事物的本质,它的本质究竟如何,必须依靠专门学科的研究才能下结论。这样才能保证统计方法在客观现实中的正确运用。

第三节 SPSS 统计软件及其特点

SPSS 是英文 Statistical Package for Social Science 的缩写,即“社会科学统计软件包”。目前,它是世界上通用并具有权威性的统计分析软件包。虽然它名为“社会科学统计软件包”,但实际上它除了在社会科学研究领域中发挥其巨大作用外,在自然科学的各个领域中也越来越广泛的运用,如经济学、生物学、心理学、医疗卫生、农业、商业、金融等各领域都在使用该软件包。

SPSS 最初是从 DOS 环境(SPSS for Dos version 1.0 ~ version 4.0)逐步升级发展到 Win-

dows 环境 (SPSS for Windows version 5.0 ~ version 10.0) 的, 2001 年最新版本为 version 10.0。

本书将以 SPSS for Windows version 10.0 版本为例, 介绍 SPSS 的应用。由于该版本是 SPSS 的 Windows 版本, 除了具有 Windows 软件的共同特点之外, 还具有以下特点。

1. 采用窗口方式管理程序运行的全过程, 简捷直观, 易于操作。除数据输入工作需要使用键盘完成外, 大多数操作是通过“菜单”“图标按钮”“对话框”来完成的, 易于学习, 易于使用。

2. SPSS 的命令语句、子命令及选择项的选择, 绝大部分由“对话框”的操作完成, 用户无须花大量时间记忆统计分析的各种命令、过程、选择项等。

3. 对于常见的统计分析方法完全可以通过对“菜单”“对话框”的操作完成, 用户无须自己编写程序。对于熟悉 SPSS 语言的用户也可在语句窗中直接输入程序语句, 然后用鼠标按“RUN”按钮提交运行。用对话框指定命令、子命令和选择项后, 用“Paste”按钮可将与选择对应的语句自动置于“Syntax”窗中, 并可以文件形式保存。因此, 该版本的 SPSS 适用于 SPSS 的新老用户。

4. 具有第四代语言的特点, 只要通过菜单的选择、对话框的操作告诉系统要做什么, 无须告知怎样做。只要粗通统计分析原理, 无需通晓统计分析的各种算法, 就可得到具有专业水准的统计分析结果。

5. 该软件系统的组合结构使用户有可能根据自己分析工作的需要, 根据计算机设备的实际情况选择、装配模块, 灵活方便。

6. 与其他软件有数据转换接口, 能够与其他软件实现资源共享。其他软件生成的数据文件, 例如关系数据库生成的 DBF 文件, 或用文本编辑软件生成的 ASCII 码数据文件均可方便地转换成可供分析的 SPSS 数据文件; SPSS 文件也可以相应转换成其他形式的数据文件。

7. 分析方法丰富, 提供了从简单描述统计到多因素统计分析等各种不同的统计分析方法, 还有强大的图表生成和编辑功能, 可产生各种统计报表和形象直观的统计图形。

以上特点也正是 SPSS 软件之所以深受广大科研工作人员的青睐和喜爱得以快速推广的原因。

第二章

体育统计的有关概念

第一节 总体与样本

一、总体

总体(population)亦称“母体”“全域”。总体是根据研究目的确定的同质的研究对象的全体。更确切地说,是性质相同的所有观察单位某种变量值的集合。构成总体的每一个对象称为个体(individual),亦称“元素”“个案”。由于研究对象不同,个体可以是人,也可以是事或物。总体所包含的个体数常用符号 N 表示。总体与个体都是客观存在的,但它也是根据研究目的而确定的。若总体中所包含的个体数目是有限的,叫做有限总体;若是无限的、抽象的,就叫做无限总体。

例如研究某地某年正常成人的血压值,研究对象是该地该年的正常成人,观察单位是个人,变量是血压,变量值是测得的血压值,该地该年全部正常成人的血压值就构成一个总体。它的同质基础是同一地区、同一年份、同为正常成人。这里的总体只包括有限个观察单位,则为有限总体(finite population)。

有时总体是设想的,如研究某种训练方法对某运动项目的运动员的竞赛成绩的影响,这里总体的同质基础是同运动项目的运动员,同采用某种训练方法,包括设想采用该训练方法的所有该项目的运动员,其观察单位数显然是不确定的、无限的,则为无限总体。

二、样本

体育研究中很多是无限总体,要直接研究总体的情况是不可能的。即使对有限总体来说,若包括的观察单位数过多,直接研究总体也是耗费人力、财力很大的,有时也是不可能和不必要的。所以在实际工作中,常需要按一定方法从总体中抽取有代表性的一部分个体。这样被抽取的这部分个体的集合叫做一个样本(sample)。在一个样本里可以含有不同的个体数,样本所含的个体数目叫做样本容量(sample size),常用符号 n 表示。科学研究往往是运用样本信息来推断总体特征。

在样本的抽取中,若是独立随机抽取、数据间不存在对应关系的样本,称为独立样本

(independent sample)。在样本的抽取中,若数据间存在一一对应关系的样本,称为相关样本(correlative sample)。相关样本来自两种操作方式:1. 同一组被试先后接受两种不同的处理,得到两组具有一一对应关系的数据;2. 将同质的被试两两配对,各对中的两个被试分别接受处理 A 和 B(例如十对双生子,每对中的一个做实验 A,另一个做实验 B),得到 A、B 两组数据。这种用配对方式得到的相关样本又称做配对样本(paired sample)。

要做好抽样研究,须满足两个方面的要求:1. 样本的可靠性(reliability),即明确规定总体的同质范围,也就是确定被研究对象的同质基础,包括时间、空间、条件等。2. 样本的代表性(representation),即所抽取的样本应能代表所研究的总体。为此,必须遵循随机抽样的原则,必须保证样本中有足够的观察单位数,即样本含量,或称样本大小。通常情况下,样本容量大于 30 即称为大样本,小于 30 则为小样本。

三、参数与统计量

总体参数(population parameter),简称参数,是反映一个总体特征的数值。如总体平均数、总体方差和总体相关系数等。一般用希腊字母表示,用 μ 表示总体平均数,用 σ^2 表示总体方差,用 ρ 表示总体相关系数等。这些参数往往是未知的,通常以相应的样本统计量对参数进行估计与推断。

样本统计量(sample statistic),简称统计量,一般指由样本计算出来的指标。如:样本平均数、样本方差、样本相关系数等都是统计量。用 $\bar{x}$ 表示样本平均数,用 S^2 表示样本方差。

参数虽然常常是未知的,但它是一个客观存在的数值,因此,它是个常数。统计量是由样本而得来,故可看做是已知的,它是随抽样而变的,因而是个随机变量。

第二节 统计资料与测量误差

一、统计资料

统计资料(statistic data)就是统计数据,一般分为计量资料、计数资料和等级资料三类。实际上,等级资料是介于计量资料和计数资料之间的一种类型。

(一)计量资料(measurement data & quantitative data)

计量资料亦称“定量数据”或“测量数据”,它是对某个观察单位测量某个指标量大小所得的资料。这种资料大都是用定量仪器从观察单位逐一测量得来的,一般都带有度量衡或其他单位。每个观察单位的测量值又称变量值,如调查某地儿童的身体发育状况,是以儿童个体为观察单位,分别测量每个人的身高(cm)、体重(kg)和血压(kPa),得出不同的变量值。