刘学贞　编著

TI YU YONG SHU JU CHU LI FANG FA

体育用数据处理方法

北京体育大学出版社
BEI JING TI YU DA XUE CHU BAN SHE

体育用
数据处理方法

刘学贞 编著

北京体育大学出版社

责任编辑:熊西北　　**审稿编辑**:荒　草
责任校对:长　春　　**责任印制**:长　立　陈　莎

图书在版编目(CIP)数据

体育用数据处理方法/刘学贞编著. -北京:北京体育大学出版社,2001.8
ISBN 7-81051-653-1

Ⅰ.体… Ⅱ.刘… Ⅲ.体育-数据处理-方法 Ⅳ.G80-32

中国版本图书馆 CIP 数据核字(2001)第 058954 号

体育用数据处理方法　　刘学贞　编著

北京体育大学出版社出版发行　　新华书店总店北京发行所经销
(北京·中关村北大街　邮编:100084)　　北京雅艺彩印有限公司印刷

开本:850×1168 毫米 1/32　　印张:7.875　　定价:14.00 元
2001 年 8 月第 1 版第 1 次印刷　　印数:4000 册
ISBN 7—81051—653—1/G·551
(本书因装订质量不合格本社发行部负责调换)

作者简介

刘学贞，女，副教授，生于1945年4月10日，1963年考入清华大学工程力学数学系学习，学制5年半，1968年12月毕业。大学期间一直为校体操代表队队员，通过了体操二级等级标准。1979年1月再次考入清华大学力学进修班学习，学制2年。1981年12月分配到北京体育大学运动生物力学教研室工作。

现任中国体育科学学会运动生物力学学会副主任委员。北京体育大学人体运动科学系学术委员会委员。

在北京体育大学任教期间教过：运动生物力学、理论力学、高等数学、线性代数、计算方法、概率论、计算机Basic语言等课程。同时还承担了硕士研究生的运动生物力学、线性代数、数值计算方法、研究生课程班的运动生物力学、以及运动生物力学专业的博士生和硕士生的专业课。总共教过不同类别的课程共十几门。在教学质量评估中，连年被评为优质课。

被聘为校督导组成员，并多次担任系听课组组长，对

教师的教学质量进行评估。

在搞好教学的同时，还指导了8名研究生，5名已顺利毕业并获得硕士学位。

参与了部委攻关课题："优秀青少年运动员科学训练综合研究"获1991年国家体委体育科技进步三等奖。

参与国家级课题"我国优秀运动员竞技能力状态的诊断和监测系统的研究"的研究工作。

1999年因承担国家体育总局田径运动中心亚运攻关课题"女子撑竿跳高科研攻关与科技服务"（任课题组组长），被体育总局评为亚运科研攻关与科技服务先进个人。

2000年参与的体育总局田径奥运攻关课题，获国家体育总局奥运攻关三等奖。

论文"体育运动中累进计分法的研究"入选第二届全国体育科学大会并获校科学技术成果三等奖。

参与的论文"VB3.0开发体育统计软件的研究"获校科学技术成果三等奖，并获体育总局教学成果三等奖。

编写的"数值计算方法"（16万字），获校自编讲义三等奖。

参与体育总局委管课题"运动生物力学测试规范化研究"，并出版《运动生物力学测量方法》一书。

此为试读，需要完整PDF请访问：www.ertongbook.com

前　言

在科学技术是第一生产力的时代，体育的科学研究工作空前地活跃起来了。广大体育工作者在科学研究过程中，经常遇到实验设计和数据处理等问题。如何科学地进行实验设计，如何处理和分析实验所获得的各种数据（如身体形态、机体能力、心理状态、运动能力、伤病等测验数据），不仅直接影响着科学研究的水平，而且还可能决定着科学研究的成败。因为实验设计错误，就势必把整个科学研究工作引入歧途；而实验所要获得的各项指标，是通过大量的测验数据反映出来的，大量的测验原始数据，如不经过正确处理和分析，就无法显示出有价值的信息，因而也就难以有力地论证问题。计算方法与统计学，作为两门工具课，就是为广大读者提供从事体育科学研究的方法。

本书作者毕业于清华大学工程力学数学系，是清华大学校体操代表队队员（二级运动员），又搞过多年运动生物力学的教学与研究，并多年任研究生导师。作者总结了在北京体育大学长期任教、指导研究生论文和评审论文工作中的经验，概括出最常用的参数统计方法、非参数统计

方法和数值计算方法，并指出经常发生的错误，编著了本书。本书的写法通俗易懂、深入浅出；只介绍方法，不做原理推导，举例多为体育运动和体育科研中的实例。使本书实用性更强，更容易读懂。内容中有些有争议的地方望多多指正。

本书共分为三章：第一章 参数统计方法；第二章 非参数统计方法；第三章 数值计算方法。三章各自成为一个体系，分别介绍了几十种体育科研中常用的数据处理的方法。另外，书中对应用体育数据处理方法时常出现的错误予以了分析，对读者更好地、更准确地掌握体育用数据处理方法有很大帮助。

不当之处，请读者批评。

目录

第一章 参数统计

第一节 几个基本概念

一. 什么叫体育统计

体育统计是数理统计在体育科学中的应用。数理统计的基础理论是概率论。概率论是研究偶然事件(又叫随机事件)发生的规律的一门科学。所谓偶然事件是指在相同条件下,可能发生也可能不发生的事件。例如投掷一枚钱币,落地后恰巧正面向上就是偶然事件,又如跳高中试跳某一高度,可能成功也可能失败,所以也属于偶然事件。偶然事件的发生与否是受许多客观因素影响的,这些因素多而复杂,相互联系,相互影响,而又找不出哪一个因素是决定性的因素。因此偶然事件的发生与否事先难以确定,凡是事先没有百分之百的把握必然要发生的事件,都称为偶然事件。如果某个事件,事前就完全可以肯定它一定会发生,那么这种事件就称为必然事件了。

此为试读,需要完整PDF请访问: www.ertongbook.com

二. 为什么体育科学可以应用概率统计的方法加以研究

体育活动中需要研究的现象，绝大多数都是偶然事件。例如身体素质的提高，运动成绩的出现，以及比赛的胜负等，由于这些事件的发生是受许多客观因素影响的，情况极为复杂，尽管经过严格的科学训练，在计划时间内达到预定指标的可能性极大，但仍存在极小可能性实现不了预定指标，也仍然是属于偶然事件。因此，研究偶然事件的概率统计方法，就可以在体育科学研究中应用。

偶然事件的发生，并非是不可捉摸的，虽然我们不能准确地预见某一次偶然事件发生与否，但人们在长期的实践中发现，在偶然事件的大量重复中，往往出现几乎必然的规律，这就是大数定律。例如，抛掷钱币，就某一次来说出现正面还是反面，虽然是偶然的，但在大量重复抛掷时，出现正面和反面的次数与抛掷总次数之比，都必然接近于确定的数值，即各占 1/2。这 1/2 就是抛掷钱币时出现正面或反面的概率。那么从一付扑克牌中抽取一张牌，抽得红桃的概率为 13/52，即 1/4；抽得老 K 的概率为 4/52，即 1/13，抽得红桃 A 的概率只有 1/52。因此每次抽得哪一张牌，虽然无法事前确定，但反复多次抽取后，将会确定某一张牌出现的概率。在体育活动中某一事件的概率并不能象上述那样简单得到，需要通过大量的实验。如某一学生试跳某一高度，在 100 次试跳中有 80 次跳过去了，我们讲他有 80 %的把握能成功，这就是成功的概率。习惯上的“十之八九能成功”；“八成是成功了”……这都是根据经验指出某一事件成功的概率。以上说明偶然事件中蕴藏着必然的规律，我们应用研究偶然事件的规律的数理统计方法来研究人体的各种指标出现的规律是完全可行的。

三. 总体和样本

根据研究目的所确定的研究对象的全体称为总体;从总体中抽取的有代表性的一部分称为样本。例如要调查全国学生各项运动水平如何,以便制订适合于全国大、中、小学不同年龄组的测验标准,那么这一研究的对象和内容就是总体。这么大的总体(有的总体很大或是抽象的)是很难取到的,因此,只有从总体中抽出对总体有代表性的若干个省市的部分学生进行测验,这叫"抽样"。抽出的这一部分人测验得到的数据叫"样本"。样本中的数据叫"变量",用 X 表示。样本所包含的变量个数叫"样本含量",以"n"表示。

如果研究的对象是十个学生的跳远水平,即使每人跳一次,得到的成绩仍然不是总体。因为再跳一次得到的成绩与第一次成绩不同。究竟这十个学生客观存在的水平(即总体的"真值")有多高,是取不到的,每人多次试跳所得到的测验成绩仍然是样本。

我们研究的对象是总体,但我们能得到的只是样本。因此,我们只能通过样本对总体情况做出估计。这是统计学中一个主要内容。

样本既然是估计总体的根据,就要求样本对总体有较高的代表性,样本情况愈接近总体情况,它的代表性就愈高。要想提高样本的代表性,就需要有科学的抽样方法:一是抽样的原则要随机化,即不加选择的抽取。例如,要了解某校初三男生百米水平,准备抽 50 名进行测验,这 50 名学生怎么确定呢? 可以采取抽签、抓阄的方法;也可以按学号逢五便取;或是在这一年级的几个班中各随机抽取几名;都是符合随机原则的。如果有意挑 50 名好的或差的,那就不能作为这一研究的样本。二是增加样本含量。即从总体中抽得的个体愈多,它对总体情况的代表性也就愈高。一般讲总体愈大样本含量也应愈大;总体中个体之间差别愈大,所采取的样本也应愈大。具体取

多少合适，要根据实验要求来确定。

四. 抽样误差

由于总体中的个体之间存在着差异，又由于样本仅是总体中的一部分，因此样本的统计量（如样本中各变量的平均数）与总体客观存在的"真值"（如总体中各变量的平均数）也就存在差异。即使从一个总体中随机抽取相同含量的两个样本，也不会相同。这种由于抽样引起的差别称为"抽样误差"。抽样误差是无法避免的，但这种误差有它一定的规律，研究这个规律，应用统计学方法可以计算出误差的范围和它的概率，从而对总体的"真值"做出估计。其他如测验方法存在缺陷、秒表不准、计算错误、登记错误都属于工作过失，在实验和统计中应注意防止。没有可靠的数据，无论怎样统计处理也得不到正确的结果和结论，这一点是至关紧要的。

第二节　平均数与标准差

一. 怎样收集和整理材料

当我们准备研究某一个问题之前，首先要阅读有关的资料或书刊，了解他人对此问题是否有所研究，研究到什么程度，有哪些论点和争论，已有什么结论，然后才能设计研究方案。研究问题要通过自己的实践或实验来论证才有意义，才会得到有价值、有创造性的结论。

由于研究问题是多种多样的，因此不可能提出一个统一的实验设计方案。但实验设计应遵循基本的原则和方法。

实验设计的原则是要求实验次数尽量少;尽量用简易而准确的实验方法来获得足够的、有效的资料,从而得到较可靠的结论

实验设计要从这个要求出发来制定最优的实验方案。因此实验不必过于追求复杂的仪器和繁杂的方法。在体育测验中有不少公认的方法,如田径中的比赛规则、视力测定表等。这些都可以作为实验的方法。

比较研究是科学研究中不可少的方法

有比较才能有鉴别。把实验分为实验组和对照组是比较研究中常用的基本方法。在实验过程中对两个组用不同的处理方法来影响它们,一般是对实验组施以所要研究的处理方法,而对照组则有几种:"空白对照",即不加任何处理,听其自然发展,例如要了解开展锻炼标准活动对于学生健康的影响,空白的对照组可以不开展这项活动,按原有的活动自然发展;"标准对照"是指运用已知的正常值,如正常男子的心率为72次/分,研究长跑锻炼的学生心率是否低于72次/分;"实验对照"是对实验组和对照组施以不同的处理,以比较两种处理结果的差异,例如施以两种不同的教法。

在实验中要控制非实验因素的干扰

即除实验因素不同外,其他因素要一致,否则无法判断两组的差别是实验因素不同引起的还是其他因素引起的。例如研究两种跳高教学的效果,不但两个组原来跳高的水平应要求一致,身体素质基本相同,并且教学时间、条件也要一样,这样,经过一段时间后产生不同的效果,才利于确认效果的差异是由于不同的教学方法引起的。

实验的数据要符合统计要求才能应用统计方法处理

前边讲过取得数据要求随机化,并且要考虑有足够的数量。实验数据愈多说明实验次数也愈多,得到的实验结果可靠性就高。如某个学生的跳远测验以试跳十次的平均成绩比较只跳一次的成绩更能代表他的跳远水平,从这个角度考虑数据自然是愈多愈好,但测验

工作量和统计工作量都加大，因此取得实验数据要根据实际可能，结合专业知识和统计学的要求来确定。

实验数据取到后，需要整理成有条理的有系统的资料

“频数分布”是最实用和简便的方法。频数分布是将全部数据根据大小分为若干组，其步骤如下：

1. 求两极差：例如 100 名初三男生仰卧起坐成绩中最少的做一次，最多的做 19 次，两极差为 19 − 1 = 18。

2. 分组和定组距：分组的多少是根据变量的个数。一般 100 人左右可分为 7～9 组，上例可以分 6 组左右，那么每组包括的范围应为 18÷6 = 3。

3. 排列组限：组限即一个组的上、下限，在下表中第一组为 1—3 次，也可写做 1—3。第二组为 4—6 次。习惯上数值由小到大从上到下排列。

4. 填数：和选举唱票一样，将每个变量按数值大小填在所属的组内，写成“正”字，便于计数 。变量填完后将每组频数和累积频数以及总频数 n 逐栏填写(见表 2—1)。

表 2－1　某校初三男生仰卧起坐成绩频数分布表

组　距	记　数	组 频 数	累积频数
1—3	一	1	1
4—6	正正丁	12	13
7—9	正正正正正丁	27	40
10—12	正正正正正正一	31	71
13—15	正正正正下	23	94
16—18	正	5	99
19—21	一	1	100

n = 100

从频数分布表中可以掌握样本的大致情况：最低成绩不低于1次；最高成绩不超过21次；从累积频数栏中可以知道达到某一水平的频数，例如9次和9次以下者有40人；10 — 12次者（即11次左右）的频数最多，我们叫11次为“众数”，是平均数的一种。

从频数分布表中我们发现频数的分布情况是中间多两头少，这是偶然事件分布的普遍规律，是一个很重要的性质，在下一讲中再详述。

二. 平均数和标准差

平均数是分析数据的基本指标，它反映了各变量平均的水平，对各变量的总情况具有代表性。平均数有几种，最常用的是“算术平均数”。如将某班25人百米跑的成绩加起来被25除，得到的就是25人百米跑成绩的“算术平均数”，这是大家早已熟悉的方法。在统计学中算术平均数又简称为“均数”，代表符号为（读作“杠”x），计算公式为：

$$\overline{x}=\frac{\sum x}{n} \qquad (1)$$

其中：$\overline{x}$— 平均数　　$\sum$ — 表示求和，即把全部成绩加起来

x—每个成绩数　n—样本个数

例如：测得5个人50米跑的成绩为8.8″、8.6″、8.2″、9.4″、9.5″，则：

$$\overline{x}=\frac{\sum x}{n}=\frac{8.8+8.6+8.2+9.4+9.5}{5}=8.9''0$$

即这5个人50米跑的平均成绩为8.9″。

此为试读，需要完整PDF请访问：www.ertongbook.com