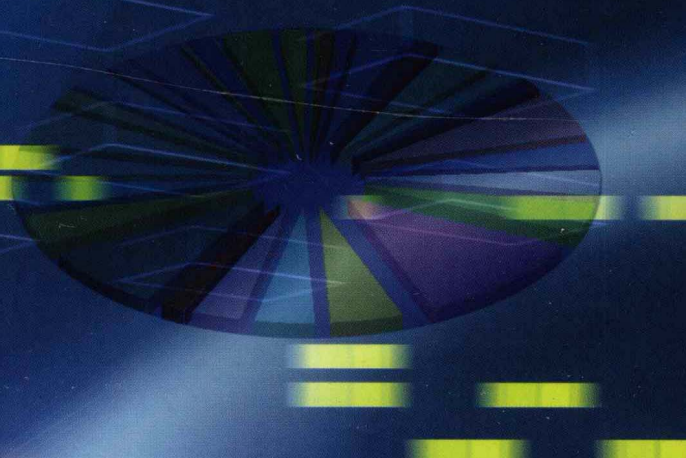




体育统计方法与运用

王路德 编著



人民体育出版社

体育统计方法与应用

王路德 编著

人民体育出版社

图书在版编目(CIP)数据

体育统计方法与应用 / 王路德编著. -北京: 人民体育出版社, 2008

ISBN 978-7-5009-3389-2

I. 体… II. 王… III. 体育统计-方法 IV. G80-32

中国版本图书馆 CIP 数据核字(2008)第 028960 号

*

人民体育出版社出版发行
三河兴达印务有限公司印刷
新华书店经销

*

787×1092 16 开本 13 印张 246 千字
2008 年 6 月第 1 版 2008 年 6 月第 1 次印刷
印数: 1—5,000 册

*

ISBN 978-7-5009-3389-2

定价: 49.50 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 邮购: 67143708

(购买本社图书, 如遇有缺损页可与发行部联系)

序

我从来没有为谁的著作写过序，真感到勉为其难。我想写序不外乎要介绍作者和有关背景，并对著作作一评价。

王路德老师是学体育专业的，原来在武汉测绘学院担任体育教师。由于“文化大革命”的干扰，统计学在我国传播较晚。但是王路德老师较早地接触和自学统计学，且学以致用。“文革”结束，万物复生。1979年，国家体委（现称国家体育总局）联络其他七个部委发起调查研究我国青少年儿童的体质现状。这是一项非常重要的大工程，涉及全国16个省市25万人的资料，共有约150人参加了调研工作。为了分析研究所获得的大量数据，必须依赖统计学。可惜当时缺乏统计学人才。经了解，借调了王路德老师来参加此项工作，从此他便同体育科研结下不解之缘，后来索性被调到湖北省体育科研所，一直从事体育统计学的研究、应用与推广工作。他是我国体育科研界的一位非常重要的统计学专家。

国家体委一贯重视科学研究，五十年前就成立了第一所体育科学研究所。如今各省、市、自治区都有专门的体育科学研究所。各个高等体育院校、师范院校以及许多大学和研究机构也都开展体育科学方面的研究。改革开放以来，国家体委组织了大量科研项目和课题，其中规模最大的有两个系列：“中国人体质”和“运动员选材”的研究。这两个项目的样本量都非常大，要建立庞大的指标体系，采集大量数据，进行长期纵向跟踪研究和复杂的统计分析。在这两个大的研究系列中，王路德老师都是主力成员，三十年如一日，倾注全部心血，尤其在统计学方面，始终发挥重要作用。直至今日，虽年逾古稀，仍然退而不休，笔耕不辍，再次整理出版对于体育科学研究有着极具实用价值的统计方法指导书，并且奔走于各地，对当前的国民体质监测与青少年运动员科学选材研究进行指导与帮助。

我从事体育科学研究四十载，其间也做过多年的科研管理工作，

同王路德老师交往颇深，对于他的这种锲而不舍与孜孜不倦的工作精神十分钦佩，因此也就接受了写这篇序的任务。

前中国体育科学学会常务理事
体质研究分会主任委员
国家体育总局科研所副所长

高大安

2008 年 1 月

前言

体育统计方法是在体育科研、教学、训练中对数据进行正确处理的重要工具。随着体育事业的不断发展,广大体育教师、教练员和科研人员越来越感到掌握体育统计这一工具的重要性,迫切希望有一本通俗易懂而又能结合体育实际的体育统计方法参考书,供平时学习和进行数据统计计算时参考。

编写本书的指导思想是:以通过实例介绍方法为主,辅以通俗的文字讲述基本原理。因而它不是按一般的数理统计书籍的体系从抽象的数学概念出发,而是从数据的收集、审核、整理入手,一步步由浅入深地用体质研究、体育教学、运动员选材等实例来介绍体育统计的各种方法。希望读过本书的读者都能进一步感觉到体育统计并不是高不可攀的数学理论,而是从事体育教学、训练和研究工作时十分有用的工具。因而能联系自己的工作实际,用学到的统计方法去研究和分析问题。

《体育统计方法》于1983年由湖北省体育科研所出版后,受到各方面读者的肯定。不少体育学院、体育系以它作为体育统计课程的教材或教学参考书。1985年开展首次全国学生体质健康调查前,教育部邀请我对全国各省市学生体质测试工作骨干培训统计方法时,就是用这本书作教材。

后来,应人民体育出版社之约,我对书的内容又进行了补充,增加了各种统计方法的Basic计算程序,于1990年9月由人民体育出版社出版了《体育统计方法与程序》,并参加了1990年北京亚运会科学大会的体育科技书展。

1987年起国家体委科教司成立了全国运动员科学选材中心组,指导各省市开展运动员科学选材的研究和推广工作,我被选为组长直至1993年。在这7年多时间里,选材标准的研究是中心组的主要内容之

一。在制定选材标准的研究中,对统计方法在评分、评价中的应用做了许多探索。

2000年我国开展了首次全国国民体质监测,我又应邀给各省市体质工作者讲授体质研究中的统计方法,并协助许多省市体质监测中心做数据统计处理工作。

这次,我把最近十几年来在国民体质研究、运动员选材研究中对统计方法应用的实例增加了进去,改名为《体育统计方法与应用》。另外,由于1990年后我已根据《体育统计方法与程序》中的程序编制了《体育科研数据统计处理系统》软件,有许多单位和个人仍在使用,故本书取消了Basic程序的内容,增加了在Excel中进行统计计算的方法,以便读者用Excel进行数据统计分析。

教育部、国家体育总局于2007年4月联合发出了“关于实施《国家学生体质健康标准》的通知”,规定在全国各级各类学校全面实施《国家学生体质健康标准》,要求学校每学年都要对学生进行一次《标准》的测试。这就要求广大体育教师都要学会统计分析学生体质健康数据的方法。用每年测试的大批体质测试数据,开展教学研究,分析学校、班级、学生的体质健康情况,指导和改进学校体育教学工作。

希望本书能成为对广大体育教师和体育工作者有用的一本统计参考书。限于本人的水平,书中的不足甚至错误也在所难免,敬请提出批评指正。

王路德

2008.2

目 录

绪 论	(1)
第一章 统计资料的收集与整理	(4)
第一节 统计资料的来源	(4)
第二节 统计资料的收集与整理	(8)
第三节 统计资料的分组	(11)
第二章 几个常用的统计量	(17)
第一节 平均数、标准差、变异系数	(17)
第二节 正态分布规律与测试数据的检验	(22)
第三节 标准误	(23)
第四节 率、率的标准误	(25)
第五节 百分位数	(27)
第三章 概率知识	(31)
第一节 随机事件和概率	(31)
第二节 随机变量与概率分布	(32)
第三节 正态分布	(34)
第四章 假设检验	(36)
第一节 假设检验的原理简介	(36)
第二节 样本平均数的差异显著性 t 检验	(38)
第三节 样本率的差异显著性 U 检验	(41)
第四节 χ^2 检验	(43)
第五节 正态检验	(45)
第六节 秩和检验	(50)

第七节	进行显著性检验时应注意事项	(55)
第八节	统计检验的应用实例	(57)
第五章	动态分析	(59)
第一节	动态分析的计算方法	(60)
第二节	动态分析的统计图	(64)
第三节	体质指标变化趋势的统计处理	(66)
第六章	方差分析	(69)
第一节	方差分析原理简介	(69)
第二节	单因素方差分析	(70)
第三节	各组平均数之间的相互比较	(72)
第四节	双因素方差分析	(75)
第五节	正交设计简介	(79)
第六节	方差分析应用实例	(82)
第七章	相关与一元回归	(85)
第一节	相关与回归	(85)
第二节	散点图与回归直线	(87)
第三节	相关系数及其检验	(88)
第四节	表格式计算相关系数与一元回归方程	(91)
第五节	检验一元回归方程的效果	(93)
第六节	两个一元回归方程的比较	(96)
第七节	一元回归的应用实例	(97)
第八章	多元线性回归分析	(99)
第一节	二元线性回归方程	(100)
第二节	表格式计算多元线性回归方程	(102)
第三节	多元线性回归的方差分析	(106)
第四节	复相关系数与偏相关系数	(108)
第五节	利用标准回归系数或偏回归平方和 进行因素分析	(110)

第六节	两个多元回归方程的比较	(113)
第七节	合理地挑选指标建立回归方程	(115)
第八节	多元回归的应用实例	(117)
第九章	逐步回归分析	(120)
第一节	逐步回归分析方法的计算过程	(120)
第二节	逐步回归的应用实例	(124)
第十章	曲线回归分析	(128)
第一节	配曲线的方法	(128)
第二节	曲线拟合优度的比较	(133)
第三节	曲线拟合的应用实例	(134)
第十一章	聚类分析	(137)
第一节	R 型聚类	(137)
第二节	Q 型聚类	(141)
第三节	聚类分析的应用实例	(145)
第十二章	判别分析	(148)
第一节	两类判别	(148)
第二节	多类逐步判别	(151)
第三节	判别分析的应用实例	(155)
第十三章	主成分分析与因子分析	(156)
第一节	主成分分析	(156)
第二节	R 型因子分析	(159)
第三节	主成分分析和因子分析的应用实例	(164)
第十四章	制订评分评价标准	(166)
第一节	筛选评价指标	(166)
第二节	制订单项指标的评分标准	(168)
第三节	确定各指标的权重系数和综合评价方法	(176)
第四节	标准的试用和检验	(181)

附 录

1. 正态分布表	(183)
2. t 值表	(184)
3. χ^2 检验临界值表	(184)
4. 相关系数可靠度表	(185)
5. 等级相关系数可靠度表	(185)
6. F 值表	(186)
7. 多重比较中的 Q 值表	(188)
8. 正态性 D 检验临界值表	(190)
9. 秩和检验临界值表	(191)
10. 确定最大 (小) 秩和数的临界值系数 C 表	(193)

绪 论

体育运动的目的是要增强人民体质，提高运动技术水平。为了更好地达到这一目的，要求广大体育教师、教练员、体育科研人员不断地总结经验，开展科学研究，探索我国国民体质的现状和发展规律，推动全民健身活动；探索体育教学、运动训练的客观规律，提高体育教学和运动训练工作的水平。

用什么方法去探索客观规律呢？数理统计是研究客观事物规律的一种科学方法。

在客观世界中有许多现象，人们可以事先预知它们在一定条件下是否会出现。例如，在标准大气压下把水加热到 100 摄氏度就一定会沸腾；把一个球用力向天上抛去，无论用了多大的力，由于地心引力的作用，球最后一定会落到地上。这些无论试验多少次都无一例外必然出现的现象称为必然事件。又如，对空射击，子弹是不可能飞到地球以外去的。这一类在一定条件下必然不可能出现的现象称为不可能事件。但是，除了可以事先预知的必然事件和不可能事件之外，在客观世界中更大量的现象，却是事先无法预知其结果的随机事件。这在体育领域中也是极多的。例如：球赛之前裁判员召集双方队长抽签挑选场地时，使用圆形塑料挑边器，它一面是红色，另一面是绿色，裁判员将它向上抛起后，落下来将会是红面向上还是绿面向上？这是事先无法预知的。又如，一个人投篮时，在球出手前无法准确地预言球投中篮还是不中篮？一个人跳远，事先也无法准确地预言他跳出的成绩是多少？这一类事先无法预知其结果的、带有偶然性的现象称为随机现象。随机现象的每一种表现就称为随机事件。从表面上看，随机现象完全是带有偶然性的，好像并无规律可言。但是观测大量次数以后，就可以发现随机现象也是具有一定的内在规律性的。例如，一个质地均匀的塑料挑边器，只掷一次或几次，红面向上几次是没什么规律的。但是如果掷几百、几千次，并记下每次的结果，最后就会发现红面向上和绿面向上的比例数几乎是相同的。又如，让某运动员投篮，对于他的每一次投篮来说，事先无法准确地预言投中还是不中。但是，当他投了许多次，就可以统计出他的投篮命中率。假设该运动员投了 100 次篮，投中 60 次，就可算出他的投篮命中率为 60%，根据这一命中率，当然我们也就估计说，他每一次投篮投中的可能性是 60%。假设又对另一名运动员统计了许多次投篮，计算出他的投篮命中率为 10%。那么，当他某一次投篮

前，我们就可以预言他这一次投中的可能性是 10%。数学上把事件发生的可能性大小称为概率，用符号 P 表示，对于不可能事件 $P=0$ ，对于必然事件 $P=1$ ，而随机事件的概率就介于 0 与 1 之间。

专门研究随机现象的数量规律性的学科叫做概率论，它是数学的一个分支。而数理统计就是通过对有计划地收集到的大量数据，按照概率论的法则进行统计分析，从而寻找出客观事物规律性的一种数学方法。

2007 年教育部、国家体育总局共同发出了“关于实施《国家学生体质健康标准》的通知”，要求各级大、中、小学，每年都要对全校学生的身高、体重、肺活量、台阶指数、50 米跑、立定跳远等形态、机能、素质指标和健康状况作详细的测定。对于学校体育老师、体质测定工作者、体育科研人员来说，体质测定后获得的大量反映学生体质状况的数据是十分宝贵的资料。这些数据如：身高 170.2 厘米、183.5 厘米；肺活量 3500 毫升、4850 毫升；50 米跑 7.3 秒、8.1 秒等都是用数值记录下来的随机现象的结果，数学的术语称为随机变量（简称变量）。如果能够正确地运用数理统计的原理和方法进行分析研究，就可以从一大堆看来杂乱无章的变量中，找出事物的内在规律，发现事物相互之间的关系，并为预测今后可能的发展趋势提供丰富的材料和科学依据。这对改进体育教学和增强学生的体质是会有很大用处的。

同样的，随着《国民体质测定标准》的推广、实施，以及 5 年一次的全国国民体质监测的开展，要了解某地区、某单位国民体质的综合评价情况如何？各年龄段男女的身高、体重等测试指标的平均数如何？或者要比较两个地区国民体质的情况，研究某地区市民参加体育锻炼的人数、活动项目、地点的分布以及影响人们参加锻炼的原因等等，都需要体质研究工作者正确地运用统计方法，进行分析研究，然后，把统计分析的结果，提供给各级政府和体育管理部门作为改进工作的科学依据。这是推动全民健身活动的一项十分重要的工作。

在运动员训练的过程中，也可以测试到运动成绩、身体素质、生理、生化的许多数据，需要教练员和科研人员认真分析研究，指导运动员科学训练提高成绩。

因此，体育统计是运用数学方法研究体育运动客观规律的一门学科。它应该是每一位体育工作者都应该学习和熟练掌握的工具。

统计是数学的分支，有其严谨的理论依据，各种统计计算公式都是经过推导和证明的。体育工作者学习体育统计需要学懂一些基本的统计理论知识，但是，重点应该放在学会正确地运用不同的统计方法去解决体育研究中的实际问题。因而，本书着重介绍统计方法如何在体育领域中的应用。各种统计方法和公式都尽量通过实例来介绍，关于公式的来源和推导过程，有兴趣的读者可进一步参阅有关的统计专著。

进行统计分析时应注意的问题

1. 学会用概率论的思想方法分析问题，作结论防止绝对化

一定要记住，由样本统计出的结果由于有抽样误差存在，所以统计推断的结论是有可能出错的。因为所有的统计结论总是和概率相联系的，所以我们应该学会用概率论的思想方法分析问题，作结论要防止绝对化。就像应用统计方法作天气预报时，不是预报明天肯定下雨，而是说明天可能下雨，下雨的概率是多少。

2. 要注意体育领域研究的特点

统计计算的结果一定要结合体育专业知识进行分析、作结论。在工业的实验中，温度、时间、压力、配料比例等等都比较容易准确地控制，各种偶然因素对实验结果的干扰较小，所以，工业统计中有时允许以较少的观察例数就得出统计分析结论。但是，体育统计研究的对象是人，人的个体差异是很大的，同性别、同年龄的人，身高、体重、运动能力等的差异都较大。由于存在较大的个体差异，所以在进行研究时，观察人数太少就进行统计分析是不合适的。另外，统计分析时必须注意“同质”，即被研究的对象应是性质相同的，否则统计分析的结果就不能说明问题。例如，研究某运动处方对减肥的效果，虽然参加研究的对象是同性别的人，但是还必须考虑他们的年龄段是否大致相同，参加实验前的体质水平是否相同，假如把一批老年人和一批年轻人作比较就不对了。

3. 要有严肃的科学态度

对统计的数据和计算结果必须实事求是，一丝不苟，来不得半点儿虚假。有的人在取数据时，为了希望说明自己的预定结论而弄虚作假。有的人在测数据和计算时粗枝大叶，这样算出的数据不真实、不可靠，根据它所作出的结论是会误人的。

第一章 统计资料的收集与整理

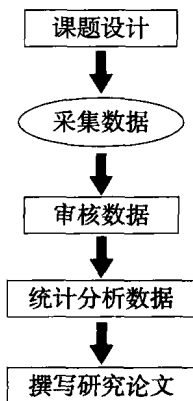
收集资料、整理资料和分析资料是统计工作的三个基本步骤。

收集资料之前必须根据研究目的制订研究设计，然后再按设计的要求去收集准确、完整的原始资料。这是进行统计工作的基础。

整理资料是对收集到的原始资料进行审核、分组、归纳汇总，以便于统计计算。这是保证统计结果准确性不可缺少的一步。

分析资料是对经过整理的原始资料计算相应的指标和绘制必要的统计图表，结合专业知识进行分析、比较，找出其中规律性，撰写研究分析论文。

以上三个基本步骤是互相有联系的，任何一步有了缺陷或差错都会影响研究结论的正确性。所以说，研究工作的全过程都离不开统计方法。



第一节 统计资料的来源

统计资料的来源主要有三方面

(1) 日常工作中积累的资料。平时进行的体质测试、上体育课、训练、开展群众性体育活动和组织运动竞赛中都可以积累许多宝贵的数据。如“国民体质测定标准”测试的数据、体育课考核的各项数据、训练时的运动成绩、运动会的比赛成绩等等。

用这一类数据进行统计前，要特别注意数据的准确性。有时测试记录的数据笔迹模糊，有时不同班级测试时条件不相同，如一个班跑 50 米时正巧刮大风，而另一个班跑 50 米时天气很好，这些数据究竟能否用于统计计算就要注意审核。

(2) 全面普查。如一年一度的“学生体质健康标准”测试。这种全面的测

试调查,工作量比较大,而学生又往往由于某种原因,容易漏测某些项目,甚至有时还有测试卡片上忘了填写姓名等情况。因此,事先要有周密的安排,做到测试时不忙乱,不出差错,并且在测完最后一项,回收卡片时要有专人负责逐项审查,发现漏填或漏测的项目就要督促学生及时补上,以免造成整理数据时的困难。

(3) 专题研究。这包括专题的调查研究 and 专题的实验研究。像 2000 年开始的每 5 年一次的全国国民体质监测,就是我国大规模的体质专题调查研究,其目的是为了摸清我国 3~69 岁国民的体质现状、特点及生长发育规律。有时为了研究某个问题,需要开展专题研究。例如,为了研究运动处方对增强体质的效果,就要组织实验组和对照组,通过一定时间实验,取得数据,对比用运动处方指导锻炼和不用运动处方的锻炼效果。

在进行专题研究前应该根据研究工作的目的和要求,结合统计原理的要求,对研究工作的全过程事先制订一个周密的工作计划,通常把制订这样的计划称为研究设计(或课题计划)。如果研究设计考虑得比较周密、完善,就能以较少的人力、物力、时间获得可靠的资料,使误差减小到最低限度,提高研究工作的质量和效率,并有助于得到科学的可靠的结论。反之,事先不作研究设计,随意收集一堆数据就进行统计计算,往往是不可能得到可靠的结论的。

研究设计的基本内容大致应包括以下几方面

① 研究目的。围绕着增强体质和提高运动水平,有许多大大小小的研究课题可以研究。一般来说,一项研究工作不宜涉及过多的问题,最好只要求解决一两个问题,否则容易影响工作的质量。

② 研究对象和人数,包括抽样的方法及分组的办法等。如果本研究课题是为了进行不同锻炼方法的效果比较,就应该设立实验组和对照组,而且这两组的人数和实验前的水平应该大体相仿。

③ 研究时间和方法。包括研究工作开始和预计结束的时间,在整个研究时间内是否划分成几个阶段,研究工作开始时要搜集哪几项统计指标,研究过程中每隔多少时间再次搜集哪几项统计指标,对这几项指标用什么方法、什么仪器设备进行测试也应同时订出详细的测试细则。应该特别重视指标的设计,要使每一项指标所测得的结果都能比较客观、真实地反映出测试对象某一方面的情况。

④ 各种原始记录表格的设计、制订。

⑤ 初步规定数据的整理、统计分析方法。

与制订研究设计有关的统计学概念

总体

某项研究的全体属性相同的对象。

个体

是组成总体的最小研究对象。在体质研究中，通常就是指人。如要研究某市 7~18 岁青少年儿童的身高，总体就是某市 7~18 岁的全部青少年儿童的身高，某市 7~18 岁的青少年儿童就是属性相同的本研究的对象，其中每一个人的身高就是一个个体。

样本

有时总体很大，要对其中每一个个体进行观察是不可能的，有时由于人力、物力、时间的限制不允许对总体内的每一个体进行观察，因此，要用适当的方法从总体中抽出一部分个体来进行观察、研究。被抽出的这一部分就叫做样本。

例如，某市 7~18 岁的全部青少年儿童可能有几十万人，如果对每个人都进行测定，工作量太大，还可能因人数过多，不能控制测试的精度。所以，一般只抽取部分青少年进行测试。这里被抽取的部分青少年就是样本。

2000 年起，全国 5 年一次的国民体质监测，就是通过监测样本，用样本的统计量来描述总体的体质特征。

为了使抽取的样本能正确地反映总体的特征，必须严格按照“随机抽样”的原则，就是在抽样时要保证每个个体被抽入的机会均等，抽样时不能掺杂抽样者的主观因素。例如身高，有的人很高，有的人很矮，很多人是中等高度，那么样本里就应该是高的、矮的、中等高度的都有，不能只挑选高的人测试。形象地说，样本应该是总体的“缩影”才有研究价值。

在国民体质测试工作中，我们在某地测试了几百、几千人的体质数据。但是，我们的目的并不仅仅是要了解这几百、几千人的体质情况，而是要通过计算分析他们的数据，了解他们所在的社区、城市总的情况。

从统计学的角度说，就是通过计算分析样本的数据来了解总体的情况。这是我们分析数据，写分析报告和体质研究论文时必须牢牢记住的概念。

样本的含量

样本内所包括的个体数称为样本的含量，通常用 n 表示。习惯上认为 $n < 30$ 为小样本， $n > 30$ 为大样本。体质研究测得的是人的数据，人的个体差异较大，所以在条件许可时，样本最好尽量大一些，但前提是要保证所测数据的准确性。如果单纯追求样本大一些，结果测试的人数增加了，却使得一些数据测得不够准确，那么这样的大样本统计出来的结论也是没有什么意义的。倒不如样本略微小