实用体育统计

陆瑞当　编著

GUANGXI NORMAL UNIVERSITY PRESS
广西师范大学出版社

实用体育统计

实用体育统计

陆瑞当　编著

GUANGXI NORMAL UNIVERSITY PRESS
广西师范大学出版社
·桂林·

图书在版编目（CIP）数据

实用体育统计/陆瑞当编著.—桂林：广西师范大学出版社，2002.8

ISBN 7-5633-3659-1

Ⅰ.实· Ⅱ.陆 Ⅲ.体育统计 Ⅳ.G80－32

中国版本图书馆 CIP 数据核字（2002）第 052756 号

广西师范大学出版社出版发行

（桂林市育才路 15 号 邮政编码：541004
网址：http://www.bbtpress.com.cn）

出版人：萧启明

全国新华书店经销

广西民族印刷厂印刷

（广西南宁市明秀西路 53 号 邮政编码：530001）

开本：890mm×1 240mm 1/32

印张：4.125 字数：116 千字

2002 年 8 月第 1 版 2002 年 8 月第 1 次印刷

印数：0 001～1 500 定价：6.50 元

前　　言

本书主要针对目前国内体育院系本科生和研究生的主要课程《体育测量与评价》，补缺体育统计方法这一部分的基础知识。同时，可作为大专、中专在校生学习试用教材，也可供数理基础较薄弱的基层体育工作者、教师自学。

多年来体育统计方法采用表算法花去了教师、教练员及科研工作者的大量时间和精力。一般教师和教练员认为"用希腊字母写成的数学表达式不好理解"。为解决这一难题，本书给你一个方便、灵活地运用电子计算器即可对教学、训练等方面进行量化评分、自我评价的实用统计方法。特别是对电子计算机和数理统计方法不熟的体育工作者，只要有一个携带方便的计算器，就可根据现场统计需要，编好程序后到测试现场直接输入数据，显示结果，及时反馈教学和训练效果，因此，它具有省时、灵活、方便的特点。

本教材参考了高等学校试用教材《体育统计》和戎家增编著的《体育统计学》，并从 1992 年起作为体育系补充教材，得到师生们的大力支持并提出了宝贵意见。在此统致谢意。由于本人水平有限，不妥和错误之处，敬请读者批评指正。

编著者
2002 年 3 月 18 日

目　　录

第一章　绪　论……………………………………………………（1）

第一节　体育统计的性质和作用……………………………（1）

第二节　体育统计的基本内容和意义………………………（2）

习　　题……………………………………………………（4）

第二章　数理统计基础知识……………………………………（5）

第一节　总体与样本…………………………………………（5）

第二节　随机事件及概率……………………………………（6）

第三节　正态性检验 ………………………………………（13）

习　　题 …………………………………………………（20）

第三章　样本特征数 …………………………………………（21）

第一节　集中位置量数 ……………………………………（21）

第二节　离中位置量数 ……………………………………（24）

第三节　平均数和标准差的合成 …………………………（27）

第四节　特征数在体育中的评分方法 ……………………（29）

习　　题 …………………………………………………（33）

第四章　统计推断 ……………………………………………（35）

第一节　参数估计 …………………………………………（35）

第二节　假设检验 …………………………………………（42）

第三节　两种检验和两类错误 ……………………………（56）

习　　题 …………………………………………………（60）

第五章　方差分析 ……………………………………………（62）

第一节　方差分析的基本思想 ……………………………（62）

第二节　单因素方差分析 …………………………………（64）

第三节　多重比较 …………………………………………（68）

习　　题 …………………………………………………（71）

第六章　相关与回归 …………………………………………（72）

第一节　相关与回归的基本概念 …………………………（72）

第二节　计算相关系数与一元回归方程 …………………… (74)
第三节　回归方程效果的方差分析 ………………………… (77)
第四节　等级相关 ……………………………………… (79)
习　　题 …………………………………………… (81)
第七章　相对数 ………………………………………… (83)
第一节　率、构成比和相对比……………………………… (83)
第二节　动态数列 ……………………………………… (87)
第三节　相对数的标准化 ………………………………… (97)
习　　题………………………………………………… (101)
第八章　CASIO-fx 3600P 计算器应用指南 ……………… (102)
第一节　各键的功能与工作状态………………………… (102)
第二节　计算器程序的编写……………………………… (107)
第三节　使用注意事项…………………………………… (110)
附表 1　标准正态曲线下的面积 ………………………… (111)
附表 2　正态性 D 检验临界值表 ……………………… (112)
附表 3　t 值表 ………………………………………… (113)
附表 4　F 值表(方差齐性检验用) …………………… (114)
附表 5　F 值表(方差分析用) ………………………… (115)
附表 6　q 值表 ………………………………………… (119)
附表 7　χ^2 值表 ……………………………………… (120)
附表 8　符号检验表 ……………………………………… (121)
附表 9　秩和检验表 ……………………………………… (121)
附表 10　相关系数临界值表 …………………………… (122)
附表 11　等级相关系数临界值表 ……………………… (124)
参考文献…………………………………………………… (125)

第一章　绪　论

第一节　体育统计的性质和作用

体育统计是跨学科的综合体，它是以辩证唯物主义思想为指导，应用统计的理论和方法，通过对体育现象的数量描述，揭示体育领域中事物发展变化的内在规律。而数理统计是以概率为基础，专门研究数据的搜集、整理、分析和推断的一门学科，数理统计内容包括：数据的整理和样本统计量的研究，统计推断，方差分析，回归分析，抽样理论，质量控制，试验设计等。数理统计运用的方法是以样本为依据，运用数学模型来推断总体。

首先，在体育运动中许多现象是不确定的，有随机性的特点。这正是以概率论为基础的数理统计研究的对象。例如：某一短跑运动员在一次 100 m 跑测验中，运动成绩为 10.5 s。如果对他再测，运动成绩可能是 10.7 s。假如在不同的时间、地点、心理等环境下再测，也许他的运动成绩为 10.4 s。就是说这位运动员 100 m 跑的运动成绩是不确定的，具有一定的随机性。同时，要解决体育运动中的许多问题，必须依靠调查或经过实验才能获得结果，但由于调查或实验的人(次)数总是有限的，所以只能通过样本资料来推断总体。例如：对全国青少年的体质调查，不可能花费大量的人力、财力、物力对所有的青少年都进行测试。只能通过整群随机抽样的方法进行测试，获得各种数据资料。然后，用样本资料来推断全国青少年的体质、健康情况。这种利用局部调查得到的资料，经过整理、分析来推断总体就是数理统计的方法。

其次，体育现象本身所具有的随机性来源于各方面因素，比如：技术掌握程度，生理条件，心理状况，社会历史背景和地理环境等，这些因

素造成成绩具有一定的随机性。例如:在同一班中,抽取同年龄、同性别的学生在同一时间进行跳远测试,获得的成绩将会有一定的区别。即使同一运动员,重复多次测试,成绩都会呈现随机波动。但是,依概率论,当测试次数达到一定的程度后,出现有规律性的波动。这种规律性的东西正是我们需要研究的目的之一。

在体育教学、训练、科研中,为做到"心中有数",或想获得一些参数,应用体育统计方法是很有必要的。但是,体育统计不是万能的,它不可能完全解决复杂的体育问题,须紧密结合专业理论、运动实践进行综合、系统的分析研究,才能得到比较满意的结果。同时,体育统计毕竟是一门数学工具,并不能说明或表明所研究对象的本质。只有对体育统计进行正确运用时,它才会成为你的有力助手。假如在实际工作中不分场合、盲目搬用,就会产生严重错误,甚至会歪曲事实的真相。

第二节　体育统计的基本内容和意义

一、体育统计的基本内容

体育统计按内容可分为三大部分:描述统计,推断统计,试验或调查设计。

(一)描述统计

描述统计是指在杂乱无章的原始资料中提取有意义的信息。它主要是将通过调查或实验所获得的大量数据,经过整编、归纳后计算出一些具有概括性的统计数字,如平均值、标准差、相关系数等。然后可以借助这些数据的分布特征,如集中趋势、离中趋势、相关程度等,对不同总体进行分析比较,做出合乎客观规律的结论。

(二)推断统计

推断统计是指在描述统计基础上,根据样本数据所传递的信息来推断总体的性质,并说明判断可能产生误差的范围。其主要内容包括参数估计和假设检验。

(三)试验或调查设计

试验或调查设计是根据研究的目的,用最简便的方法,取得原始数

据达到科学的效果。这种效果应包括有效性、可靠性和客观性。

综上所述，描述统计、推断统计、试验或调查设计构成了体育统计的整体。它好像是一条连环锁，无论在哪个环节上出了问题，都将使整个统计结果失去意义。

二、学习体育统计的意义和要求

统计作为一种数学方法，渗透到体育领域中来，并形成体育统计学科，是现代科学与技术的密切结合，具有重要的实践意义和实际应用价值。

（一）有助于提高教学、训练和科研水平

教学方面：用因素分析方法研究各种教材对增强学生体质的效果；根据统计获得的数字分布特征，制定出合理的体育考核标准和计分标准；利用整理的成绩，对教学进行纵向观察和横向对比；利用方差分析方法分析、研究多种教学方法的综合效果。

训练方面：在获得统计处理的基础上，拟定合理的训练方案，选择最优化的训练方法；掌握运动员专项技能形成的阶段，了解不同训练手段对运动成绩的影响；利用回归分析方法对运动员成绩进行预测等。

体育教师是在教学、训练的第一线，掌握大量的原始数据，而且要搜集数据也比较容易，如果不掌握统计方法，就无法从中获得有价值的信息，同时，也无法去发现、解决现代体育教学与训练中所出现的新问题。

（二）有助于培养科学思维能力与实事求是的科学态度

数学本身就具有严谨性的特点，它可以锻炼人的推理和逻辑思维能力，培养实事求是地对待一切事物的唯物主义态度。无论是描述统计或是推断统计都需要依客观事实去伪存真、由表及里的工作态度，这样不仅能发现别人未发现的事实，同时，更重要的是让他人能得到同样的事实。总之，学习和应用体育统计，能培养我们的科学思维和科学态度。

（三）有助于学习国内外先进经验

学习体育统计方法有助于我们阅读资料，学习前人、他人成功经验和科研成果，取长补短。为加快我国体育事业的发展，认真学习别人的先进经验是非常必要的。

三、学习要求与方法

学习时，要求着重理解基本原理和概念。只有在理解概念的基础上，才能对统计方法加以灵活应用。平时搜集资料，要重视原始资料的完整性与可靠性；对数据的整理、分析须持严肃认真和实事求是的科学态度；不必深究数学公式的原理、证明过程，但要弄清公式的适用条件和使用范围，并对计算结果能做出正确、客观的分析；理论联系实际，结合专业知识，多做练习，反复实践，学以致用。

习　　题

1. 简述体育统计的性质及基本内容。
2. 举例说明体育统计中的随机性特点。
3. 简述学习体育统计的意义。

第二章　数理统计基础知识

体育统计以概率论为基础，应用统计方法研究体育随机现象，从而找出相应的随机变量分布规律和它的数字特征。为有助于掌握本课程的基本思想、内容和方法，本章特地介绍必要的数理统计基本知识，也是体育统计的基本知识，如随机事件及概率、随机变量及其分布、正态分布等一些基本概念。

第一节　总体与样本

总体就是根据研究目的确定研究同质对象的全体；把总体中的每个研究对象称为个体；样本是指从总体中随机抽取的对总体有代表性的一部分个体；把样本中所包含的个数称为样本含量。如果要研究广西初中一年级男生身高，那么，广西初中一年级男生的身高都是被研究的对象，即总体。每一个男生的身高就是个体（即是组成总体的最小研究单位）。总体平均数、标准差等习惯称为总体参数。在实际的统计研究中，由于大范围的直接调查有困难，有时甚至不可能进行，故只能对一部分学生的身高进行观测。这种从总体中抽取部分个体的过程称为抽样，让每个个体都有同等的可能被抽取到的抽取方法称为随机抽样。

总体和样本的概念是相对的。如：当选定研究的观测指标是从一个县范围推断一个省的范围时，省的范围为总体，县的范围为样本；当选定研究的观测指标是从一个学校范围推断一个市范围时，市的范围便是总体。因此，样本是对总体而言，总体是随研究课题而定。总体可分为有限总体和无限总体。

此外，统计的任务是直接观测样本推断总体。一般地说，总体参数

(即平均值 μ,标准差 σ)是未知的,可直接测知并计算的均系样本值。样本统计量反映了样本的数量特征,虽然它能推断总体情况,但它不等于总体的参数。样本质量如何,将直接影响统计结论。所谓样本的质量,指样本数据的真实性、可靠性。由于统计推断是以随机理论为基础,因此必须保证抽样的随机性,离开这个前提,样本便失去推断意义。同时,研究对象范围大、内容复杂时,还必须考虑到样本的代表性。如上例研究广西初中一年级男生身高问题时,由于广西有十二个民族,抽样时应考虑各民族、城市、乡村学校等都要有一定的比例,这样抽得的样本对总体才有一定的代表性。

第二节 随机事件及概率

一、随机事件

在一定条件下,可能发生、也可能不发生的事件称为随机事件。如果随机事件用数量指标 x(x 是一种符号)来表示,那么数量指标 x 便是一个随机变量,统计中的样本指标都具有这种性质就可以看做随机变量。

例如,对某市初中一年级男生身高进行调查,样本含量为 100。凡属该总体中的任何一个人都有同等条件可能被抽测或不被抽测,抽测是一个随机事件。样本指标身高 x,被测对象个子高的 x 值较大,个子矮的 x 值就较小。指标 x 随抽测对象的不同而随机地变化,故 x 是一个随机变量。但应该注意的是:统计指标的随机性只有当做样本时,才体现出来,若把总体中所有的个体都测完,就不存在什么可能被测或不被测的问题,也就不存在随机性。因此,总体指标、总体参数不是随机变量,它是客观存在的"真实数"。

样本指标的随机性主要是由于抽样的随机性质决定的。因此,由样本数据演算得出的样本统计量,如平均数、标准差以及 x_i 加减某一常数 k 或两个样本数据相加减等,也还是一个随机变量。

二、概率及其分布

随机事件的规律,主要体现在它的概率及概率分布两个方面,概率

和概率分布在数理统计中都具有严格的定义。

（一）概率

概率的定义式分为古典的和统计的两种形式。

1. 概率的古典定义式。

设在实验中全部等可能的、独立的基本结果有 N 个，其中有 M 个属于事件 A，则在此实验中，称事件 A 出现的概率 P 等于 M 与 N 之比。即

$$P(A)=\frac{M}{N} \tag{2.1}$$

这里分别解释等可能、独立的、基本结果三个概念。如球类比赛使用的挑边器只有两面，设一为蓝面，一为白面。每抛落一次，不是蓝面向上就是白面向上，没有第三种可能，因此，出现两种基本结果（即蓝面向上为一种基本结果，白面向上又是另一种基本结果）。所谓独立的是指基本事件之间均无因果或相联关系。不存在抛第一次出现蓝面向上导致抛第二次出现白面向上，即使同一种色面各次之间也无联系，所以说，它们两种不同的基本事件间均是相互独立的。所谓等可能的是指在一抛落中，两种基本事件出现的可能性大小程度同等。显然，要满足等可能条件，要求挑边器的制作、结构和质量均匀，两面不存在孰轻孰重现象。

根据以上解释得知：每抛落一次的全部基本结果 $N=2$。其中出现蓝面向上的基本结果 $M_1=1$，出现白面向上的基本结果 $M_2=1$，代入概率的古典定义式，得：

蓝面向上的概率

$$P(A=\text{蓝})=\frac{M_1}{N}=\frac{1}{2}=0.5$$

白面向上的概率

$$P(A=\text{白})=\frac{M_2}{N}=\frac{1}{2}=0.5$$

2. 概率的统计定义。

在统计的实际研究中，由于总体情况往往不知道，M 和 N 具体数值不知，运用古典定义式有了困难，因而产生了概率的统计定义式。即

概率是指在重复无穷多次的条件下，该事件发生频率的稳定值。如抽样调查或实验重复 n 次，事件 A 出现 m 次，则称 $\frac{m}{n}$ 为事件 A 在 n 次出现的频率。当 n 不断扩大时，频率的取值逐渐稳定在一个常数 p 附近摆动，则称事件 A 有概率，且定义：

$$P(A)=\frac{m}{n}=p \tag{2.2}$$

古典定义式和统计定义式都有相同的结论，如抛落挑边器，当抛落次数逐渐增大时，就会发现出现蓝面向上和出现白面向上的频率均趋于 0.5，并在 0.5 附近左右摆动。

在运用概率定义式时，须注意分析条件是否同等。如在挑边器质量不均匀，一面为铜质另一面为铝质的条件下，虽然仍属于两个基本事件且相互独立，但不能满足等可能，这时就不能代入概率的古典定义式。一个运动员的一次投篮，虽然投中或未投中也是两个基本事件，但由于人的技术结构不一样，也不能套入古典定义式得出投中或投不中的概率均为 0.5。在以上两种情况下，只能用实验结果，代入统计定义式求解。因此，任何一种非等可能事件，不能用古典定义式求解。同时，必须注意，随机性是样本质量的重要条件。

3. 概率的重要性质。

从以上两个定义式，可得出概率的重要性质：

(1)对任意随机事件 A 都有 $0\leqslant P(A)\leqslant 1$，且不为负值，其中有两个特例。

当 $P(A)=1$ 时，概率为 100% 的事件，说明次次必定出现，称为必然事件。

当 $P(A)=0$ 时，概率为零，说明一次也不会出现，故称为不可能事件。

(2)概率的可加性。如果事件 A 与事件 B 互斥，则事件 $A+B$ 的概率等于事件 A 的概率加事件 B 的概率，即 $P(A+B)=P(A)+P(B)$。

(3)取遍所有可能值时，诸概率之和等于 1。

这是从大量的实践经验中概括出来的，成为我们研究概率的基础与出发点。

（二）随机变量的概率分布

随机变量的概率规律体现于它的分布。知道了它的分布，一方面可掌握它可能取哪些值，或者说，取哪一个区间的值；另一方面还可以知道能以多大的概率取这些（区间里）值。

根据随机变量可能取得的值，将随机变量分为离散型随机变量和连续型随机变量。

1．离散型随机变量。

离散型随机变量是指在量尺上任意两点间(a,b)，只能读取有限数值。例如：2 和 4 间只有一个正整数 3，再无别的正整数。如引体向上运动成绩 5 次和 7 次之间惟有 6 次。还有比赛名次、运动等级等，也称计数数据。

2．连续型随机变量。

连续型随机变量是指在量尺上任意两点间(a,b)，均可加以细分，并取得无限多个大小不同的数值。如跳远成绩 4 m 和 5 m 之间有无限多个取值。

3．正态分布。

随机变量的分布类型较多，着重介绍正态分布。正态分布在数理统计的理论和应用中，占有特别重要的地位。第一，它在实践中常见，如在体育领域中，身高、体重、立定跳远成绩、短跑成绩等，一般都认为服从或近似服从正态分布。第二，很多分布的极限分布是正态分布，如二项分布在一定条件下可用正态分布或近似正态分布来表示。第三，在统计推断中，很多方法以总体正态分布为前提。

（1）正态分布的概念。

若随机变量 x 的分布密度函数是

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}\cdot e^{\frac{(x-\mu)^2}{2\sigma^2}} \qquad (2.3)$$

其中，μ,σ 都是常数，且 $\sigma>0$，$-\infty<x<+\infty$，则称随机变量 x 服从参数为 μ,σ 的正态分布，记为：$x\sim N(\mu,\sigma^2)$。

（2）密度分布曲线（图 2-1）。

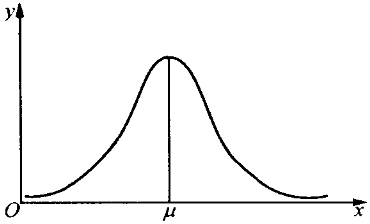

图 2-1　正态分布曲线

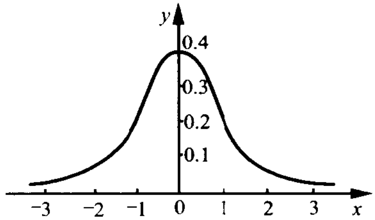

图 2-2　标准正态分布曲线

(3) 正态分布曲线有以下几个性质：

① 正态分布曲线都在横轴的上方，它和横轴间的总面积为 1，反映了随机变量取值在 $-\infty$ 与 $+\infty$ 之间，是一个必然事件，其概率为 1，即 100%。

② 正态分布曲线有一极大值的单峰曲线，具有对称性，对称轴是平行于 y 轴的直线，μ 是对称轴上所有点的横坐标。当 $x=\mu, \sigma=1$ 时，曲线的纵坐标最高，其值为 $\frac{1}{\sqrt{2\pi}}=0.3989\cdots\cdots$ 如图 2-2 所示。同时，由图 2-3 可见，在正态曲线下不仅 μ（总体平均数）两侧的面积相等，而且离 μ 的距离相等的部分，其面积也相等，曲线的高度也相等。

在曲线下最常用到的：

$\mu\pm\sigma$ 的面积占总面积的 68.27%；

$\mu\pm2\sigma$ 的面积占总面积的 95.45%；

$\mu\pm3\sigma$ 的面积占总面积的 99.73%；

$\mu\pm1.96\sigma$ 的面积占总面积的 95%；

$\mu\pm2.58\sigma$ 的面积占总面积的 99%。

③ 正态曲线由中央最高点对称地向两侧单调下降，在 $\mu\pm\sigma$ 处有拐点；当 $x\to\pm\infty$ 时，曲线以横坐标为渐近线。

有了平均数 μ 和标准差 σ，就可以把正态分布曲线完全确定下来，故对平均数为 μ、标准差为 σ 的正态分布常表示为 $N(\mu,\sigma)$，对于任一平均数为 μ，标准差为 σ 的观测值 x 的正态分布，都可以通过变换式

$$u=\frac{x-\mu}{\sigma} \tag{2.4}$$