

体育统计学

● 梁荣辉 顾培仁 编
● 杨书鹏 张 杰
● 谷世菁 审

河北教育出版社

273040



序 言

随着体育事业的飞速发展，愈来愈需要用体育统计学理论来指导教学与训练，并为体育科研服务。但是，目前尚缺乏完备的阐述体育统计学理论与方法的书籍，编者试图从理论与实践相结合、全面论述体育统计学的理论与方法的角度上作一个尝试，编出本书奉献给读者。

本书结合体育实例较系统地介绍了体育领域中常用的统计方法。在编撰过程中，既非单纯地追求理论上的阐述，也非简单地罗列方法，而是作到了理论与方法的结合，使读者既能学会方法的应用，又能了解其理论依据。本书文字简练，通俗易懂，凡具有高中数学基础知识的读者都能学懂，会用。

全书共分十五章。第一章阐述了体育统计学研究的对象、任务及体育统计学中一些常用的基本概念。第二章介绍了常用的统计图与统计表。第三章到第十四章，较完整地介绍了常用的统计方法。第十五章介绍了研究设计，以供读者在撰写特别是设计论文时参考。

本书可作为体育院校（系、科）学生、助教班及研究生的体育统计学课教材，也可作为体育教师、教练员以及科研工作者的参考书。

由于编者水平所限，书中不足之处在所难免，敬请读者批评指正。

编者

1989年

目 录

序言

第一章 绪论.....(1)

第一节 体育统计学研究的对象与任务.....(1)

一、对象(1)

二、任务(1)

第二节 体育统计学的作用.....(2)

一、阅读资料.....(2)

二、教学与训练(2)

三、培养科学思维能力(2)

第三节 几个常用的基本概念(3)

一、随机现象(3)

二、事件(4)

三、概率(4)

四、随机变量.....(6)

五、总体与样本.....(7)

第二章 统计表与统计图.....(9)

第一节 统计表.....(9)

一、统计表的构造与种类(9)

二、制表时应注意的问题(10)

第二节 统计图.....(12)

一、绘图时应遵守的一般原则.....(12)

二、图形的种类(12)

第三章 资料的收集与整理	(18)
第一节 资料的收集	(18)
一、经常性资料	(18)
二、一时性资料	(19)
第二节 资料的审查	(19)
一、影响资料正确和完整的主要因素	(19)
二、数据的误差	(20)
三、资料的审查	(21)
第三节 资料的整理	(23)
一、频数分布表	(24)
二、频数分布图	(27)
第四章 几个常用的统计量	(30)
第一节 算术平均数	(30)
一、平均数的定义	(30)
二、平均数的性质	(31)
三、平均数的简便计算	(32)
四、平均数的合并	(35)
第二节 中位数和众数	(36)
一、中位数 M_d	(36)
二、众数 M_o	(37)
第三节 标准差	(38)
一、标准差的概念	(38)
二、为何用 $n-1$ 代替 n	(39)
三、标准差的简便计算	(40)
四、标准差的合并	(42)
第四节 变异系数	(43)
第五节 标准误	(44)
第五章 相对数	(46)

第一节 率、构成比以及相对比	(46)
一、率	(46)
二、构成比	(47)
三、相对比	(48)
第二节 动态分析	(48)
一、定基比	(49)
二、环比	(49)
第三节 应用相对数时应注意的几个问题	(51)
第六章 正态分布	(52)
第一节 正态分布的定义与性质	(52)
第二节 正态分布表	(53)
第三节 正态分布的应用	(57)
一、估计某一范围内的人数或百分比	(57)
二、制订考核标准	(60)
三、T 标准分	(62)
四、用“累进计分法”制评分表	(62)
五、制订“学生身体发育与运动技能”对照表	(65)
第四节 正态性检验	(67)
第五节 百分位数法	(70)
第七章 区间估计	(73)
第一节 正态总体均值 μ 的区间估计	(74)
第二节 正态总体方差 σ^2 的区间估计	(76)
第三节 总体率 π 的区间估计	(78)
第八章 假设检验	(80)
第一节 问题与想法	(80)
一、假设检验问题	(80)
二、假设检验的基本思想	(81)
第二节 假设检验的步骤	(82)

第三节 总体均值 μ 的假设检验	(84)
一、一个正态总体	(84)
二、两个正态总体	(85)
三、总体分布未知	(89)
第四节 总体方差 σ^2 的假设检验	(90)
一、一个正态总体	(90)
二、两个正态总体	(92)
第五节 总体率 π 的假设检验	(94)
一、一个总体	(94)
二、两个总体	(95)
第六节 单侧检验	(95)
第七节 审查数据的舍弃检验	(97)
一、拉依达准则	(97)
二、斯米尔诺夫法	(98)
三、格拉布斯准则	(99)
四、汤姆松法	(99)
第八节 正态总体分布函数的假设检验	(101)
一、皮尔逊 χ^2 检验法	(101)
二、正态性 D 检验法	(103)
第九节 进行假设检验时应注意的几个问题	(105)
第十节 小结	(105)
第九章 方差分析	(109)
第一节 单因素方差分析	(110)
一、基本思想	(110)
二、各水平试验次数相等	(111)
三、各水平试验次数不等	(115)
第二节 双因素方差分析	(117)
一、无交互作用	(117)

二、有交互作用	(121)
第三节 应用方差分析时应注意的几个问题	(126)
第四节 多重比较法	(127)
一、T 法	(127)
二、S 法	(129)
第十章 回归分析	(132)
第一节 回归分析及其所研究的对象	(132)
第二节 求一元线性回归方程	(133)
一、散点图与回归直线	(133)
二、确定回归直线的原则	(134)
三、具体计算格式	(136)
第三节 一元线性回归方程的显著性检验	(137)
一、相关系数检验法	(138)
二、方差分析法	(140)
第四节 利用回归方程进行预报和控制	(143)
一、预报问题	(143)
二、控制问题	(145)
第五节 回归方程的稳定性和两条回归直线的比较	(146)
一、回归方程的稳定性	(146)
二、两条回归直线的比较	(147)
第六节 二元线性回归分析	(149)
一、二元线性回归方程的求法	(150)
二、二元线性回归方程的显著性检验	(153)
三、偏相关系数、标准回归系数以及偏回归平方和	(155)
第七节 一元非线性回归分析	(159)
第八节 多元线性回归分析	(165)
一、多元线性回归方程的求法	(165)
二、多元线性回归方程的显著性检验	(167)

三、回归系数的显著性检验	(168)
四、偏相关系数	(169)
五、应用举例	(169)
第九节 逐步回归分析	(178)
一、方法步骤	(179)
二、应用举例	(184)
第十节 多项式回归分析	(190)
第十一节 多对多双重筛选逐步回归分析	(192)
第十二节 进行回归分析时应注意的几个问题	(196)
第十一章 聚类分析	(198)
第一节 R型聚类分析	(199)
一、方法步骤	(199)
二、应用举例	(200)
第二节 Q型聚类分析	(214)
一、方法步骤	(214)
二、应用举例	(215)
第十二章 判别分析	(220)
第一节 费歇判别法	(220)
一、方法步骤	(220)
二、应用举例	(223)
第二节 逐步判别法	(227)
一、方法步骤	(228)
二、应用举例	(231)
第十三章 主成分分析	(242)
一、方法步骤	(242)
二、应用举例	(244)
第十四章 因子分析	(248)
一、方法步骤	(248)

二、应用举例·····	(249)
第十五章 研究设计·····	(254)
第一节 试验的目的、对象、指标、方法、组织 工作以及样本含量的确定·····	(254)
一、试验的目的·····	(254)
二、试验的对象与指标·····	(254)
三、试验的方法·····	(255)
四、样本含量 n 的确定·····	(258)
五、组织工作·····	(264)
第二节 试验设计·····	(264)
一、完全随机化试验设计·····	(264)
二、配对比较试验设计·····	(266)
三、随机区组试验设计·····	(266)
四、拉丁方试验设计·····	(270)
思考题与习题·····	(275)
附表·····	(294)

第一章 绪 论

体育统计学是将数理统计的理论与方法应用于体育科学领域中所形成的一门学科，是进行体育科学研究、教学与训练必不可少的工具。

第一节 体育统计学研究的对象与任务

一、对象

马克思主义认为，任何一门科学，都有其特定的研究对象以及适合于这种对象的研究方法。

体育统计学所研究的对象是体育科学领域中随机现象（定义见第三节）的数量规律性。如体质评价，运动能力评价，如何改进、提高教学或训练方法等都是它所研究的对象。通过研究分析，揭示出体育科学领域中的一些规律性，从而指导人们的实践。

二、任务

体育统计学的任务是和体育的目的紧密相联的。我国体育的目的是增强人民体质，提高运动技术水平，丰富文化生活，为社会主义的两个文明建设服务。由此提出了体育统计学的任务：运用数理统计的理论与方法整理、分析数据，揭示随机现象所呈现的必然规律，为教学、训练与科研提供理论依据和科学的方法，从而为增强人民体质、提高运动技术水平服务。

第二节 体育统计学的作用

随着体育科学的飞速发展,体育领域中的某些问题仅仅用定性的描述或是凭借教师、教练员的经验来描述就显得很不够了。前者不能揭示出事物的本质,后者往往或多或少地要掺杂一些主观上的因素、缺乏科学性。为了克服这些弊病,即能给予定量、客观地、科学地分析,就需要采用相应的方法——体育统计方法。用体育统计方法能将从表面上看来是杂乱无章、无规律可循的数据中找出事物发展变化的原因以及规律性,从而使教学、训练等各项工作趋于科学化。下面具体谈谈体育统计学的作用。

一、阅读资料

目前,国内外一些有关体育教学、训练以及基础理论等方面的资料中已大量采用了统计分析方法,特别是对一些尖端项目的研究方面。如果我们不很好地学习和掌握这些统计方法,就不能从中汲取有益的东西,在阅读资料方面,就会成为“盲人”。

二、教学与训练

现代体育科学的发展,要求每一个体育工作者都能运用科学的方法于教学或训练之中。而若不具备一定的统计知识,就难于针对教学或训练中遇到的问题采取科学的对策。如制订合理的教学或训练计划,探索新的教学或训练方法以及评价其优劣,制订各个运动项目的达标标准,估计达标人数,研究青少年儿童的生长发育情况,探讨和掌握某运动项目的训练最佳期,从影响某项成绩的众多因素中找出主要因素,由已知的、易测试的项目去推测未知的、难测试的项目等,所有这些都要用到统计知识。因此,不学习和掌握它,就难于适应时代发展的需要、作好本职工作。

三、培养科学思维能力

体育统计学以数理统计为其理论基础,故它也具有数学的严

谨性、系统性和科学性。因此，它可以锻炼人们的科学思维，使人们学会科学的推理与分析问题和解决问题的方法，有助于培养人们的科学态度、提高思维能力。

总之，学习和掌握体育统计学理论和方法是我们更好地学习与吸取国内外先进的教学、训练方法以及基础理论成果的前提，也是我们在教学、训练与科研等方面作开创性工作的基础。

这门学科还很年轻，体育科学领域中的许多问题还未涉及到，因此，它有着相当广阔的发展前景。

第三节 几个常用的基本概念

为了以后各章节的需要，本节介绍几个最常用的基本概念，读者要很好地掌握它们，否则，将会影响以后内容的学习。

一、随机现象

在同样条件下，多次进行同一观测或试验，所得结果并不完全一样，往往存在着差异，而且在每次观测或试验之前不能确切预料将会出现什么结果，这样的现象称为随机现象。

随机现象在体育领域中是常见的。如几次跑、跳、投的成绩不会完全相同，且事先不能预言。每次射击，子弹不会完全落在同一点上，且事先不能预言。再如，抛掷一枚均匀的硬币，结果可能是正面向上，也可能是反面向上，抛掷前也不能作出确切的预测。

随机现象是不是就没有什么规律可循了呢？回答是否定的。正如恩格斯所指出的：“在表面上是偶然性在起作用的地方，这种偶然性始终是受内部的隐蔽着的规律支配的，而问题只是在于发现这些规律。”（《马克思恩格斯选集》中译本第四卷243页）。人们经过长期的实践发现，所谓不可预言，只是对一次或少数几次试验而言，当进行大量重复试验时，随机现象就会呈现出某种规律性，因而也是可以预言的。这种规律性称为随机现象的统计规律性。

如进行射击，当射击次数不多时，从几个零星的弹着点看不出什么规律性来，但当多次射击时，就会看到弹着点的分布呈现出一定的规律性：弹着点差不多关于目标中心对称；愈靠近中心弹着点愈密集等，而且射击次数愈多，这种规律性就愈明显。再如，抛掷硬币，只抛少数几次，看不出正面向上与反面向上的次数的比例，但当多次抛掷时，就会看到正面向上与反面向上出现的次数的比例总是近似 1:1，而且，抛掷次数愈多，愈接近该比值。

二、事件

试验的每一个可能结果称为一个随机事件，简称为事件，用字母 $A, B, C, \dots$ 表示。

例如，“百米成绩 12 秒”、“跳高成绩 1.50m”就是事件。抛掷硬币时，“正面向上”、“反面向上”也是事件。为了了解某校的达标情况，从该校随机抽取 100 人，结果可能是：“无一人达标”，“一人达标”，“二人达标”，…，这每一个结果都是一个事件。另外，“达标人数少于 3”，“达标人数在 10 与 20 之间”等也都是事件。

事件可分为：

1. 基本事件：不可能再分的事件称为基本事件。例如，事件“二人达标”就是一个基本事件。

2. 复合事件：由若干个基本事件组合而成的事件称为复合事件。例如，事件“达标人数少于 3”就是一个复合事件，它由“无一人达标”、“一人达标”与“二人达标”三个基本事件组合而成。

3. 必然事件：在试验条件下必然发生的事件称为必然事件。例如，事件“达标人数在 0 与 100 之间”就是一个必然事件。

4. 不可能事件：在试验条件下不可能发生的事件称为不可能事件。例如，事件“达标人数不在 0~100 之间”就是一个不可能事件。

三、概率

我们只粗略地介绍一下古典概率和统计概率。

一般地说，一个事件在每次试验中可能发生也可能不发生，这是具有偶然性的。但是，实践告诉我们，事件发生可能性的大小则是事件本身所具有的一种确定属性，不受我们主观上的支配。不同的事件，发生的可能性的大小是有差别的，有的发生的可能性大，有的发生的可能性小。例如，一箱乒乓球中如果坏球很多，则从中随机取一个得到坏球的可能性就大；如果坏球很少，则得到坏球的可能性就小。为了研究事件发生的可能性的大小，就需要用一个数值将这种可能性的大小表示出来，赋予事件发生的可能性大小的一个数值表示。表示事件发生可能性大小的数值称为事件的概率。通常用符号 $P(A), P(B), P(C), \dots$ 表示事件 $A, B, C, \dots$ 发生的概率。

1. 古典概率

设试验的一切基本事件有 n 个，且每个基本事件出现的可能性都相等。若事件 A 包含的基本事件有 k 个，则事件 A 的古典概率定义为：

$$P(A) = \frac{k}{n}$$

例 1.1. 1000 个乒乓球中有 8 个坏球，从中随机取一个，取得坏球的概率为：

$$P(\text{坏球}) = \frac{K}{n} = \frac{8}{1000}$$

古典概率要求试验的一切基本事件是等可能的，但在许多实际问题中，这一要求是难于得到保证的，于是引出了统计概率。

2. 统计概率

一般说来，在同一组条件下，重复进行 n 次试验，若事件 A 出现了 m 次，则称 m 为事件 A 出现的频数， $\frac{m}{n}$ 称为事件 A 出现的频率。当试验次数 n 不断增加时，事件 A 的频率呈现出

一种稳定的趋势, 即 $\frac{m}{n}$ 围绕某一常数摆动, 且 n 愈大, 摆动的幅度就愈小, 这个常数就定义为事件 A 出现的概率, 称为统计概率。

在大量重复试验中, 可用事件 A 出现的频率近似地表示该事件的概率, 即

$$P(A) \approx \frac{m}{n}$$

例 1.2. 一名射手射击 500 次, 中靶 200 次, 我们就说他中靶的概率是 $\frac{2}{5}$ 。

例 1.3. 历史上曾有不少人作过抛掷硬币的试验。如皮尔逊抛掷了 24000 次, 有 12012 次正面向上, 则正面向上的频率为:

$$\frac{12012}{24000} = 0.5005$$

我们就用 0.5 作为正面向上的概率。

概率的性质:

- (1) 对于任何事件 A , 有 $0 \leq P(A) \leq 1$ 。
- (2) 必然事件的概率等于 1, 不可能事件的概率等于 0。

四、随机变量

我们通过一个具体例子来阐明随机变量的概念。

例 1.4. 设袋中装有 5 个同型号的乒乓球, 其中 3 个为正品, 2 个为次品, 现每次从中随机抽取 2 个, 我们来考察抽出的 2 个乒乓球中次品个数的分布规律。为表达方便, 用 ①, ②, ③ 表示 3 个正品, 用 ④、⑤ 表示 2 个次品, 则所有结果可由表 1.1 给出。

由表 1.1 可以看出, 次品个数是一个变量, 它的取值是具有随机性的, 而它的随机性来源于试验结果的随机性, 即这个变量

的取值是由试验结果来确定的，它是试验结果的函数。象这种随着试验结果而变的变量称为随机变量。如本例随机抽取 2 个乒乓球中所含次品的个数就是一个随机变量，它的所有可能取值为 0, 1, 2，且该随机变量取值为 0 的概率为 $\frac{3}{10}$ ，取值为 1

的概率为 $\frac{3}{5}$ ，取值为 2 的概率为 $\frac{1}{10}$ 。

体育领域中随机变量是大量存在的，如引体向上、仰卧起坐次数，跑、跳、投成绩等。

表 1.1 次品个数表

试验结果	次品个数
①, ②	0
①, ③	0
①, ④	1
①, ⑤	1
②, ③	0
②, ④	1
②, ⑤	1
③, ④	1
③, ⑤	1
④, ⑤	2

我们引入了随机变量的概念之后，就使得事件数量化了，即使得每一个事件与一个数相对应。

最重要的随机变量有离散型随机变量和连续型随机变量。离散型随机变量是指所有可能取值为有限个，或虽不为有限个但可以一一列举出来(可数个)的随机变量。例如，脉搏、引体向上、仰卧起坐次数以及例 1.4 中的次品个数都是离散型随机变量。连续型随机变量是指所有可能取值连续地充满某一区间的随机变量。例如，身高、体重以及跑、跳、投成绩都是连续型随机变量。

五、总体与样本

我们把所要研究的对象的全体称为总体。组成总体的每个基本单位称为个体。从总体中抽取的部分个体称为样本。样本中所包含的个体的个数称为样本含量，通常用字母 n 表示。例如，我们要研究河北省大学生的体质情况，那么全省所有的大学生就构成了总体，而其中每一个大学生就是一个个体；从这个总体中抽取一部分大学生，比如抽取 10000 名，这 10000 名大学生就是一个样本含量 $n=10000$ 的样本。

我们对总体的研究，总是要归结为对总体的某个或若干个特性指标的研究。如上述问题要研究的总体是全省所有的大学生，即要归结为对大学生的身高、体重等特性指标的研究。而身高、体重等特性指标都是随机变量，因此，我们以后就把总体与某个随机变量可能取值的全体等同起来。

我们所要研究的是总体，但在实际工作中往往是办不到的，只能抽取样本来进行研究，通过样本所提供的信息，对总体的特性作出估计与推断。如上面的例子，我们不可能对全省每一个大学生都进行研究，而只能抽取一部分大学生来加以研究。为使我们的研究工作有价值，就必须使样本能很好地代表总体。形象地说，样本应是总体的“缩影”。要作到这一点，就必须严格按照“随机抽样”的原则进行抽样，所谓随机抽样就是使总体内的每一个个体都有同等的概率被抽取，抽样时不可掺杂自己主观上的因素。此外，样本的代表性还与样本含量 n 的大小有关， n 愈大，样本的代表性就愈强； n 愈小，代表性就愈弱。但 n 不能太大，太大了，人力、物力等开支就要受到限制，特别是一些损坏性的试验更是如此。也不能太小，太小了，样本的代表性就要差，不能说明问题。究竟 n 取多大为宜呢？要视具体问题而定，比如要看问题中对精度的要求、个体之间差异的大小、抽样时采用的抽样方法以及选用的统计方法等。