

实用体育統計

成都体育学院教务处

说 明

体育统计是实现体育科学化、现代化的重要工具之一。随着我国科学技术的发展，体育统计这门新兴学科已逐渐在体育科研、教学、训练和管理中发挥出重要的作用。

本书作为成都体育学院开设的一门必修课的试用教材，是根据目前统计学在体育领域内的实际运用情况，参阅了有关专著和资料编写而成。在编写中，力求学以致用，使理论与实际相结合，书中绝大部份举例和习题都是选自体育实践中的真实数据。考虑到本学科的系统性，本书较全面地纳入了目前体育实践中常用的体育统计方法，在教学时，对书中注有※号的章节，可根据学时多少安排选学或自学。

该书也可供体育教师、教练员、体育科研及管理人员自学或参考。

本书的编写，得到了我院有关领导和教务科、图书馆情报资料室、运动医学教研室、运动生物力学教研室、田径教研室、体操教研室和测试中心计算机室部份同志的热情帮助，新学科教研室的雷利逸付教授审阅并提出了宝贵意见，特在此表示感谢。

由于本人水平有限，书中难免有缺点和错误，希望读者批评、指正。

编著者

一九八五年六月

11/18/08

目 录

第一章 绪论	(1)
第一节 体育统计的概念、对象及任务.....	(1)
第二节 体育统计在体育领域中的应用.....	(2)
第三节 几个基本概念.....	(3)
第四节 体育统计工作的基本步骤.....	(6)
第二章 资料的收集与整理	(7)
第一节 资料的收集途径.....	(7)
第二节 资料的整理.....	(8)
第三章 统计描述	(11)
第一节 集中量数.....	(12)
第二节 离中量数.....	(16)
第三节 百分位数.....	(23)
第四节 相对数和动态分析.....	(24)
第五节 统计表.....	(35)
第六节 统计图.....	(38)
第四章 正态分布	(46)
第一节 正态分布的概念和特点.....	(46)
第二节 正态分布原理在体育领域的运用.....	(49)
第三节 体育评分.....	(55)
第四节 资料的正态性检验.....	(67)
第五章 计量资料的统计推断	(70)
第一节 均数的抽样误差和标准误.....	(70)
第二节 对总体均数的估计.....	(72)
第三节 显著性检验的基本思想及一般步骤.....	(74)
第四节 双侧检验和单侧检验.....	(75)
第五节 两均数差别的显著性检验.....	(76)
第六节 进行显著性检验时的注意事项.....	(82)
※第七节 两方差的显著性检验.....	(83)
第六章 方差分析	(86)
第一节 方差分析的基本思想.....	(86)
第二节 单因素方差分析.....	(89)
第三节 多组均数间的两两比较.....	(90)
※第四节 双因素方差分析.....	(94)
※第五节 正交试验法.....	(99)
第七章 计数资料的统计推断	(104)
第一节 率的抽样误差.....	(104)
第二节 对总体率的估计.....	(105)
第三节 两率差别的显著性检验.....	(105)

※第四节 χ^2 检验	(107)
※第八章 秩和检验	(114)
第一节 两组比较的秩和检验	(115)
第二节 多组比较的秩和检验	(116)
第三节 按等级分组资料的秩和检验	(119)
第四节 配对资料的秩和检验	(120)
第五节 关于秩和检验的一些说明	(121)
第六节 正态性D检验	(121)
第九章 直线相关与一元回归	(123)
第一节 直线相关	(123)
第二节 等级相关	(126)
第三节 一元回归分析	(128)
第四节 应用相关和回归时的注意事项	(135)
※第十章 多元回归分析	(138)
第一节 二元回归方程	(138)
第二节 多元回归方程的计算	(140)
第三节 多元回归方程的方差分析	(145)
第四节 利用标准回归系数和偏回归平方和进行因素分析	(147)
第五节 多项相关分析	(151)
第六节 多元回归分析举例	(154)
第七节 运用多元回归分析的实例介绍	(159)
※第十一章 逐步回归分析	(160)
第一节 逐步回归分析的基本思想和步骤	(161)
第二节 逐步回归分析的计算过程	(161)
第三节 逐步回归分析在体育实践中的运用举例	(166)
※第十二章 曲线回归	(168)
第一节 曲线配合	(168)
第二节 几种曲线配合方法举例	(170)
第三节 曲线配合的拟合度及估计精确度	(174)
※第十三章 聚类分析	(177)
第一节 R型聚类分析	(177)
第二节 Q型聚类分析	(181)
※第十四章 判别分析	(184)
第一节 概述	(184)
第二节 两类判别的计算方法	(184)
※第十五章 因子分析法	(189)
第一节 概述	(189)
第二节 因子分析的计算过程	(190)
第十六章 研究设计	(194)

第一节 调查设计.....	(194)
第二节 实验设计.....	(198)
部分习题参考答案.....	(202)
附表1 随机数字表.....	(207)
附表2 正态分布表.....	(209)
附表3 正态分布各区间中点值表.....	(210)
附表4 t值表	(211)
附表5 方差分析表(方差齐性检验用)	(212)
附表6 方差分析表(方差分析用)	(213)
附表7 q值表.....	(217)
附表8 χ^2 值表.....	(218)
附表9 两组比较的秩和检验界值表.....	(219)
附表10 三组比较的秩和检验界值表.....	(221)
附表11 配对比较的秩和检验界值表.....	(222)
附表12 正态性D 检验界值表.....	(223)
附表13 相关系数界值表.....	(224)
附表14 等级相关系数界值表.....	(226)
主要参考书目.....	(封三)

第一章 绪 论

第一节 体育统计的概念、对象及任务

一 体育统计的概念

通过实验、调查和对积累资料进行分析的方法来探索体育领域中各种现象的客观规律性，是广大体育工作者熟知的科学研究方法。但是，研究中所获得的实验或调查数据是否可靠？怎样去正确而有效地处理所得数据？如何合理、准确地得出结论？这些重要的问题，只有在掌握了体育统计学的基本原理和方法后，才能找到全面正确的答案。

体育统计学是一门探讨如何运用数理统计方法及一般统计学方法来研究体育领域内各种活动和现象的规律性的学科。数理统计是数学的一个分支，是二十世纪二十年代兴起的一门年轻而引人入胜的学科。数理统计的主要内容是研究随机现象的规律性。

客观世界中所发生的现象是多种多样的。有一类现象，在一定的条件下必然发生（或必然不发生）。例如运动员踢出的球必然下落；在标准大气压下水热到 100°C 就会沸腾等等，这类现象就称为必然现象。除了可以事先预言其结果的必然现象外，自然界中还存在着另一类现象，其结果是事先不能预言的。例如在足球比赛之前裁判员召集双方抽签决定场地选择权时，使用圆形的塑料挑边器，它一面是红色、另一面是绿色，裁判员将它向上抛起后，落地时它是红面向上还是绿面向上，这是事先无法预知的；一个人投篮时，在球投出之前不能准确的预言该球是投中还是投不中；一个人跳远，事先也是无法准确地估计他能跳出什么成绩。这类事先无法预言其结果的现象就称为随机现象。

就随机现象每次实验或观察结果来说，它具有不确定性，似乎无规律可言。但是，通过大量的重复实验或观察，就可以发现随机现象也具有一定的内在规律性。

人体活动是相当复杂的，它既包含了物理、生化、心理等因素，又包含了各种社会因素。所以，任何一项简单的人体活动，都受着许多因素的支配和影响，任一因素的变化都会引起人体活动的变化。故对于大多数体育活动，事先都无法确定其结果。大多数体育活动现象，都属于随机现象，都需要用数理统计的方法去研究其统计规律性。

二 体育统计研究的对象

体育统计研究的对象是广泛的，概括地说，应是：体育领域内一切能用数量表现的活动和现象。如人体形态、机能，身体素质，生长发育，运动能力及体育教学、科研、训练活动中的各种指标、参数以及体育管理中的某些数据及系统工程等，凡是能用数量表现的活动和现象，都可以运用体育统计的原理和方法去进行研究。通常的数量表现方式有两种：一是定量指标，如身高、体重、100米跑的成绩等都可以由定量指标表示；一是定性指标（又称计数资料），如运动技术的优劣，某些心理素质的评定，尿蛋白的阴性、阳性等。

三 体育统计的任务

体育统计的任务是：在体育实践活动中，运用统计学的方法收集、整理和分析大量的数据，以揭示体育领域内各种现象的统计规律性，为增强人民体质和提高运动技术水平服务。

第二节 体育统计在体育领域中的应用

统计学广泛地应用于体育运动的教学、训练和科研,是近几十年内的事情。随着科学技术的发展,体育科研手段的日益现代化,数理统计被引进了体育统计的领域,用以研究许多训练手段的可靠性问题,并使许多技术因素与胜负的关系数学化,因而引起体育界的广泛重视。目前,随着整个体育队伍的科学文化素质的提高,体育统计这门科学已在体育的教学、训练和科研中发挥了很大的作用。

根据现有资料,统计学应用于体育领域大概有以下几个方面:

一 体育教育与行政管理

目前,我国在管理方面主要运用的是传统的统计学方法,主要有上级主管部门和国家统计部门下达的统计报表和因工作需要自己制定的统计表两种。这涉及统计学中的表格设计问题。近几年,在体育管理方面已开始运用数理统计的原理和方法。如招生工作中体育考试项目的选择、评分表的制订等方面已运用到正态分布原理和多因素分析;个别省份已着手体育管理系统工程的研究。

二 群体调查和体质研究

如调查某地区儿童少年身体发育情况,调查某地群众体育活动开展情况,调查学生身体素质、体育运动成绩与学习成绩的关系等等。这类调查从调查设计到结果分析都离不开统计学的知识。

三 运动技术现场统计

如篮球、排球、足球等的现场技术统计,事前规定一些项目和统计方法,组织人力在比赛中将观察的数据记录下来,供比赛后的研究使用。体操、田径等其他项目的比赛也可以应用现场统计方法进行研究。运用数理统计的方法,可对所得到的大量数据进行回归分析,求出各项技术因素与胜负关系的经验公式—回归方程,进而作出各项技术因素对胜负影响的定量分析。教练员可根据回归分析的结果和本队实际情况,改进作战方案和训练要点,从而提高本队的战术水平,增加获胜的可能性。

四 组织竞赛的统计

在体育竞赛的组织工作中,需要进行一些统计计算工作,这类统计方法在各专业理论中都有介绍,对于这些统计计算的统计学原理,本书将在有关章节涉及。

五 评定运动水平或成绩

在数学或科研工作中,往往需要用分数表示运动水平或评定成绩,由于不同项目成绩的单位不同,不能直接进行综合统计计算(例如全能运动的总成绩只能将各项成绩变为分数才能相加),因此需要将运动成绩用分数表示。在体育运动中是根据集体的水平与个人之间的差异为根据评分的,这类评分大多需要运用正态分布的原理和方法。

六 评定运动员的训练水平, 确定影响成绩的主要因素

运动员要在运动竞技中显示出卓越的运动成绩, 就必须具备多种良好的素质。这些素质包括身体的发育水平和特点, 神经肌肉的灵活性和协调性, 呼吸循环系统的工作能力, 技术和战术训练水平, 意志品质等。对这些素质的全面评价, 需要运用体育统计中的多因素分析方法。

就某一特定项目来讲, 在多种素质因素中, 哪一种因素最为重要? 相互之间关系如何? 如在长跑运动员的训练中, 改进跑的技术和发展力量应占什么位置? 培养篮球、足球、排球运动员的耐力应花多少时间? 对这类问题都只有通过统计分析才能作出较为准确的回答。

七 评定训练效果

如两种不同训练方法的比较, 需要对两种训练方法所得的结果进行差异的显著性检验, 以判断其差异是由于偶然性因素造成的还是由于两种训练方法的优劣造成的, 即判定两种训练方法的结果之间的差异是否可靠。同样, 对各种科学研究的结果也必须用统计学方法判断其可靠性。本书的统计推断部份将解决这类问题。

八 用于体育运动的基础理论及运动医学方面

研究各种生理、生化指标的变化与运动能力的关系, 各项运动技术的生物力学因素, 运动创伤治疗的疗效比较, 药物配方的改进等, 都需运用统计推断或多因素分析的方法。

统计学在体育领域的应用是广泛的, 我们还可以运用统计学的原理和方法, 研究运动训练中运动强度、密度及运动量对身体机能的影响, 运动员的选材等许多问题。

第三节 几个基本概念

必然事件和不可能事件 必然现象中存在着必然事件和不可能事件两类事件。在一定条件下必然发生的事件称为必然事件。如运动员踢出的球最终要下落, 世界杯足球赛的决赛必有胜负等, 都是必然事件。必然事件的出现是有条件的, 如足球比赛的预赛就允许出现平局。在一定条件下必然不发生的事件称为不可能事件。如一场球赛双方都输是不可能事件, 目前我国少年选手100m跑的成绩达到10秒是不可能事件。不可能事件在一定条件下也可能转化为可能事件。

随机事件 随机现象的每一种表现称为随机事件。具体地说, 在一次实验中, 如果某一事件既可能发生, 也可能不发生, 这种事件就称为随机事件。如运动员的每一次投篮能否投中, 一个人每次跳远的远度等都是随机事件。“在大量的偶然现象中, 存在着必然的规律性。”通过大量的重复实验或观察, 我们可以研究随机现象的规律性。找到在一定的条件下, 某随机事件出现的可能性有多大。如足球裁判员抛起的挑边器落地后, 出现红面向上或绿面向上都有可能, 在挑边器落地前我们不能准确地预知它哪一面朝上。但是, 通过多次的抛挑边器后, 就可得到出现红面向上和绿面向上的机会是大体相等的结论。即出现红面向上和绿面向上的可能性各占50%。若某运动员投篮100次, 投中60次, 则可以估计该运动员在同等条件下投篮命中的可能性为60%。

表1.1为迪摩根等人所作抛掷硬币实验的情况。从表中结果可见,抛掷次数愈多,硬币

表1.1 抛掷硬币实验结果				出现正面向上的次数愈接近总抛掷次数的50%。
实验者	n	正面向上 X	频率 $\frac{X}{n}$	
Demorgan	2048	1061	0.5181	该例子证明,通过大量的重复实验或观察,可研究某随机事件发生的可能性大小。
Buffon	4040	2048	0.5069	
Pearson	12000	6019	0.5016	
Pearson	24000	12012	0.5005	

概率 统计学上把事件发生的可能性大小称为概率,以符号P表示。必然事件 $P=1$ (即100%),不可能事件 $P=0$,而随机事件的概率则介于0到1之间,即 $0 < P < 1$ 。如抛边器出现红面向上的概率 $P=0.5$;上述某运动员投篮命中的概率为 $P=0.6$ 。

小概率事件 当某一事件的概率接近于0时,这个事件在大量次数的试验中出现的频率非常小,这样的事件称为小概率事件。小概率事件虽然不是不可能事件,但在一次试验中出现的可能性很小,通常就认为,在一次试验中,小概率事件几乎不会发生。如某运动员在中线外为罚球投篮,投100次可能投中1次或2次,则这种投篮命中的现象为一小概率事件,可认为它在一次投篮中一般不会发生。

总体 某一研究的全体对象称为总体。如研究四川省中小学生的身体发育状况,则全省所有中小學生就是该项研究的总体。若研究我国优秀短跑运动员的身体素质与运动成绩的关系,则全国所有的优秀短跑运动员(可定为一级以上或健将级)就为该项研究的总体。

个体 组成总体的每一个研究对象称为个体。在体育研究中个体通常指每一个人。如上例,每一名优秀短跑运动员就为该研究的个体。

样本和样本含量 从总体中抽取的对总体有代表性的一部分个体,称为样本。样本内所包含的个体数就称为样本含量,通常用符号n表示。在研究条件控制较严的情况下,一般以 $n \leq 30$ 为小样本, $n > 30$ 为大样本。由于作为体育科研的个体差异较大,所以在研究中,样本最好大一些,但其前提是要保证所测数据的准确性,不能因为追求样本大一些而使研究条件控制不严致使一些数据测得不准确,样本的含量应在研究设计中根据本课题的特点及人力、物力和时间等多方面因素而确定。

抽样 从总体中抽取部分个体的过程称为抽样。科研工作常常是通过对本体的实验或调查来研究总体的规律性。例如要研究四川省中小学生的身体发育状况,则全省各地的所有中小學生就为该项研究的总体,要对这个总体进行全面的调查是很困难的,工作量会非常大以至于办不到。所以最好的办法是进行抽样研究,从全省的中小學生中抽取一个2万人的样本,用该样本来代表总体,运用统计学的方法对该样本的数据进行处理,就可以研究全省中小學生的身体发育状况。这种由样本来研究总体的方法称为抽样研究。

运用抽样研究方法时,必须保证样本对总体的代表性,所以,要求抽样必须遵循随机化的原则,即要做到随机抽样。

随机抽样 使总体中每一个体被抽取的机会完全均等的抽样称为随机抽样。随机抽样的方法主要有以下几种:

1、单纯随机抽样法 (1)抽签法。把总体内各个体全部按顺序编号进行抽签。如某中学有16岁男生120人,要从中抽取20人。先把全部16岁男生编为0—120号,并做120个签,然后抽签20次,把抽中的20个号码所代表的人组成样本。此法通常在总体不大时采用。(2)随机数表法。运用随机数字表进行抽样,可省去做签的过程,且简便可靠。具体使用方法见本节末。

2、机械抽样法 把总体内各个体按顺序编号,然后按固定间隔抽样。如要从某校女生800人中抽80人,即抽取 $1/10$ 。现从1—10号签中抽一次签,假如抽到“6”,则6,16,26……直至796号都作为被抽中的个体,由此便得一个80人的样本。如要从400人中抽80人,则可在1—10号签中抽两次,如抽到6和3两个数,则尾数为6和3的号数就为被抽中的个体。此种方法只需做1—10号十个签,而且还减少了抽签的次数。

3、分层抽样法 先把总体按某种特征分成若干部分(称为分层),然后在每一层内采用简单抽签法或机械抽样法按比例抽样。如全校女生800人分布在初一、初二、初三和高一、高二五个年级,抽样时可以每个年级为一层,共分五层。然后在每个年级内按简单抽签法抽取 $1/10$ 的人组成样本。这种方法的抽样误差较小。

4、整群抽样法 先把总体分为若干个集体(称为群),再用简单抽签法抽出几个群组成样本。如某中学校1200人共为22个教学班,某项研究规定抽3个整班组成样本。这种方法的优点是便于组织测试,一般在大规模调研时采用。但此种方法的抽样误差较大。

在制定研究设计时,可以根据具体情况决定采用哪一种抽样方法,有时也可以把几种方法结合起来。

观察值 对某一随机事件结果的测量值称为观察值。

真值 某随机事件的真实结果为真值。

误差 统计学上所说的误差,是指测得的观察值与真值之差,以及样本指标与总体指标之差。如运动员的100米跑成绩,在精确度要求较高的情况下,由跑表计下的时间与运动员的的真正成绩往往有差别,甚至几名裁判员计下的时间都会不一致。这说明观察值与真值之间是会出现差异的。误差可分为以下几种:

1、过失误差:在收集资料的过程中,由于测试人员的一些过失造成的误差称为过失误差。如记录人员的笔误,将5写成8;计时员由于思想不集中而提前或延迟开表、停表等。

2、系统误差:在收集资料时由于仪器不准(秒表、皮尺、体重仪等)和测量方法不统一(100米跑的电动计时和手计时,弹跳力测验时的原地上跳与助跑起跳等)所造成的误差称为系统误差。在多数情况下,系统误差是一个常量。系统误差的避免,主要靠在测试前认真核对仪器,统一规定测量方法等。(有些系统误差,在测试后发现也可进行纠正。如100米跑测试后,发现某只跑表每次计时都快0.2秒,那么,就可把该跑表所计跑道的测试对象的成绩全部进行调整)。

3、随机误差:由于偶然因素引起的一类不易控制的测量误差称为随机误差。如100米跑的成绩,往往几名计时员所计的成绩不完全相同,人体血压的测定则更易出现随机误差。在研究过程中,随机误差是难免的,但随着样本的增大,随机误差的算术平均数可愈来愈小而渐趋于零。实际上,在体育运动成绩的测量中,随机误差是较小的,特别是在使用先进测量仪器的情况下,随机误差可减小到可忽略不计的程度。

4、抽样误差:由于抽样而引起的样本与总体之间,样本与样本之间的误差,称为抽样误差。抽样误差是由于个体差异的存在而造成,故抽样误差是不可避免的,但它可随着样

本的增大而减小。研究和运用抽样误差的规律，进行研究设计与资料分析，是体育统计学的重要内容之一。

随机数字表的使用方法

随机数字表是由0—9的数字随机排列而成（见附表1）。它共有6页，每页有 $50 \times 50 = 2500$ 个数字，排成50行，50列。使用时可以从任何一页的任一数字开始取数，取数的前进方向一般按以下原则决定：当被抽的总数是一位或二位数时，由起始点向右取数，到达最右端后下移一行继续由左向右取数；被抽的总数是三位或三位数以上时，由起始点向下取数，到达下端后移至下一列继续由上往下取数。下面举例介绍用随机数字表进行随机抽样和随机分组的方法。

例1.1 有80人已按顺序编成01—80号，现要求从中随机抽出20人。

设本例确定从随机数字表的第二页第3行第6列开始取数，见附表1（2），因被抽的总数80是二位数，所以，从起始点两位向右取数。把大于80的数舍去不取，重复的数也舍去不取，则本例抽中的20个号码是：6，72，17，10，25，21，31，75，49，28，24，55，65，79，78，7，63，43，36，69。

例1.2 有200人已按顺序编成001—200号，现要求从中随机抽取20人。

本例仍从第二页第3行第6列为起始点，因为被抽的总数200是三位数，所以，由起始点向下取数。把大于200的数和重复的数舍去不取，则抽中的20个号码为：67，19，11，191，135，37，200，130，62，187，82，65，118，182，80，7，175，158，111，72。

例1.3 有运动员20人已按顺序编成01—20号，现要求将该20人随机地分成甲、乙两组。

设本例确定使用随机数字表第一页，见附表1（1），并从该页第6行第1列数字开始向右取数，先决定将抽到单数的人分入甲组，抽到双数的人分入乙组。取数结果如表1.2所示。

表1.2 运用随机数字表分组举例

运动员编号	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
随机数字	91	76	21	64	64	44	91	13	32	97	75	31	62	66	54	84	80	32	75	77
组别	甲	乙	甲	乙	乙	乙	甲	甲	乙	甲	甲	甲	乙	乙	乙	乙	乙	乙	甲	甲

甲组运动员号为：1，3，7，8，10，11，12，19，20（9人）；乙组运动员号为：2，4，5，6，9，13，14，15，16，17，18（11人）。如要求两组人数相等，则应从乙组运动员中随机地抽出一人到甲组去，现仍用附表1（1），以第6行第21列为起点，向右取数的头两位数为56，将56除以（乙组的人数）11，得余数为1，所以应把乙组中的第一名运动员调整到甲组。经调整后的两组运动员号数为：甲组运动员号：1，2，3，7，8，10，11，12，19，20；乙组运动员号：4，5，6，9，13，14，15，16，17，18。

第四节 体育统计工作的基本步骤

收集资料，整理资料和分析资料是体育统计工作的三个基本步骤。

收集资料就是根据研究目的进行调查或实验设计，然后按设计要求去收集准确而完整的

原始资料。这是统计工作的基础。

整理资料就是对收集到的原始资料进行审核,并有目的有计划地进行科学的加工,使资料系统化、条理化,以便进行统计计算。这是保证统计结果的可靠性和准确性必不可少的一步。

分析资料就是对经过审核整理的资料,进一步计算相应的统计量,并结合专业知识进行比较和分析,找出其中的规律性,得出科学的结论。

必须强调,以上三个基本步骤是密切联系不可分割的,其中任何一个环节有了缺陷,都会影响研究结果的正确性。

习 题

1、举例说明什么是随机事件?

2、什么是概率?

3、为了解某市应届高中毕业生达到国家体育锻炼标准的情况,在该市32所中学中随机抽查了12个高中毕业班、共610人的“达标”情况。指出该例中的总体、个体、样本和样本含量。

4、什么是随机抽样?为什么要进行随机抽样?

5、统计资料的误差有几种?能否避免这些误差?

第二章 资料的收集与整理

第一节 资料的收集途径

收集体育统计资料主要有三个途径。

一 在日常工作中积累资料

体育工作者在日常的训练、比赛和教学中,可以经常积累许多宝贵的资料。如每次训练过程中的成绩,运动前后及运动中的心率,定期进行的各项身体素质测验和机能测验的数据,各次比赛的成绩,运动员的训练日记等。通过这些积累的资料,可以研究事物的发展过程,研究运动员的早期状况,从中发现一些现象的规律性。

在积累资料的过程中,应首先建立记录资料的制度。如要求教师或教练员测量记录一些规定的指标,要求运动员按规定写训练日记等。应明文规定各指标的测试方法,有些指标还可规定测定时间和次数等。其次应制订统一的记录表格。表格的大小和纸张的质量应便于使用和保管,表内的项目排列要便于处理,计量的单位和精确度应有规定和要求。最后应注意记载当时的环境条件和被测对象的身体状况。如特异的气候(顺风、逆风、高温等),特殊的场地条件,运动员是否伤、病等,凡影响测量指标的因素都应有所记载,否则将使积累的资料失去可靠性。

二 进行调查 research

调查研究是指所研究的因素不受研究者的支配,研究者只需到现场对被研究对象已显示的结果或条件进行调查性的测试来收集数据。调查研究可分为普查,典型调查和抽样调查三种。普查是进行大范围的全面调查,如全国人口普查,某地区中小學生达到国家体育锻炼标

准的调查等；典型调查是在对事物进行全面分析的基础上，有目的地选定典型单位或地区进行调查的方法。如对某一个或几个开展群众性体育运动的先进单位进行调查，研究其组织措施，实施方法及出勤率增高情况等；抽样调查是根据“随机化”的原则，从总体中抽取一定数量的观察单位组成样本进行调查，用以估计总体的特征。

三 进行实验性研究

实验性研究即所欲研究的因素是由研究者按自己选定的方式给予受试者的一种研究。一般地说，实验性研究的研究条件要比调查研究的研究条件控制得严格些，因而试验性研究比调查研究的效果要好些。

无论进行调查研究还是实验性研究，都必须事前根据研究工作的目的，结合统计学的要求对研究工作的全过程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。如果研究设计考虑得比较周密，完善，就能以较少的人力、物力和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，并有助于得到可靠的科学结论。反之，若事先不作研究设计，匆匆忙忙收集来一大堆数据就进行统计计算，这样不但会由于无计划的收集资料造成人力物力的浪费，还会因测试方法和条件等控制不严使所得数据缺乏可靠性，从而影响到分析结果的正确性。研究设计的详细内容将在研究设计一章中作专门介绍。

第二节 资料的整理

资料的整理包括审查资料和整理资料两大部分。

一 审查资料

统计分析工作全部建立在原始资料的基础上，研究结论的正确与否，很大程度上取决于原始资料的真实性、代表性和准确性。所以，对资料进行严格的审查是很重要的。对原始数据的审查可按以下步骤：

1、**初审** 仔细地检查全部原始记录表格和卡片。首先要核对备考项目（姓名、年龄、性别、所在单位等）是否填写清楚，然后要逐栏逐项检查测试数据是否有“缺、疑、误”三种情况。缺是指缺项未填；疑是指记录的数字书写不清楚，无法辨认；误是指明显存在的差错，如错项，错数值的单位等。

发现“缺、疑、误”应该及时处理。缺项应补填或补测，如无法补就应考虑此人其他数据是否参加统计。字迹不清的数据要经几人共同辨别确认后填写清楚，如无法确认时该数据就作废。对“缺、疑、误”数据，凡是没有绝对把握的，一律不能凭主观的估计进行涂改补正。

为了便于及时弥补收集资料中出现的差错，应当在收集资料的同时就进行初审。当受试对象完成各项测试递交卡片时，应由专人负责验收，这样，在测试对象离开现场前发现问题，可采取补测等办法弥补过失。但应注意，有些错误是不能当时弥补的，如运动前的血压、心率等指标出现的“缺、疑、误”等等。检查“缺、疑、误”应该反复进行几遍，特别是在原始资料很多时，只进行一、两遍检查往往还会有遗漏。

2、**逻辑检查** 有些毛病，初审时从卡片上一见便知，但有些错误则要从意义上根据常识进行判断和鉴别，称为逻辑检查。常用的逻辑检查方法有以下三种：（1）根据常识进

行判断。如某学生的身高为165厘米，坐高为65厘米，初看时数字记录很清楚好象没有差错，但根据常识，一般人坐高和身高的比例约大于 $\frac{1}{4}$ ，而该生相差很大，就可肯定有错。又如在一般学生的100米跑成绩中发现跑10秒的人，也必然可以怀疑该数据错了。（2）制定逻辑检查标准。在大规模研究时，为了便于进行逻辑检查，可参考过去的有关统计资料制订一个检查标准，如确定最大值至最小值的范围，当发现某一数据超出其范围时就可作为可疑数据而要求复测。（3）参考过去成绩。根据被测对象过去的统计数据进行检查。如某小学学生一年前测得身高为130厘米，现在测得的身高为128厘米，而儿童少年的身高是不会随年龄增长而下降的，故可以肯定该数据有错。

逻辑检查可配合初审同时进行。

3、抽查复核 经过初审和逻辑检查后，对较大规模的测试，还应抽查复核。如事先规定抽 $\frac{1}{10}$ ， $\frac{1}{20}$ 等。抽查中若发现错误率超过规定标准，则应该对全部资料重新检查一遍。

各种审查完毕无误后，可对原始资料编号，以便进一步整理。

二 整 理 资 料

对于大范围的调查所得的资料，其内容较多，需进行分组整理，如分性别、年龄组、项目，有的还分地区分单位等。在完成以上分类分组后，当样本较大时，还需对每一个指标内的数据进行分组。通常采用频数分布表进行指标内的分组。

1、编制频数分布表的基本思想 整理单项原始数据，一个很自然的想法，就是排队，把每一个数据按大小排列起来。但当数据太多时，排成一列队伍会太长，所以按数据大小分成几组，然后按组序排列。分组排队后的队形就构成了一个频数分布表式的结构。实际上，编制频数分布表相当于将数据进行分组排队。

2、编制频数表的基本步骤

例2.1 有一组120名12岁男孩身高（厘米）数据如下，试用频数分布表进行整理。

142.3	156.6	142.7	145.7	138.2	141.6	142.5	130.5	132.1	135.5
134.5	148.8	134.4	148.8	137.9	151.3	140.8	149.8	143.6	149.0
145.2	141.8	146.8	135.1	150.3	133.1	142.7	143.9	142.4	139.6
151.1	144.0	145.4	146.2	143.3	156.3	141.9	140.7	145.9	144.4
141.2	141.5	148.8	140.1	150.6	139.5	146.4	143.8	150.0	142.1
143.5	139.2	144.7	139.3	141.9	147.8	140.5	138.9	148.9	142.4
134.7	147.3	138.1	140.2	137.4	145.1	145.8	147.9	146.7	143.4
150.8	144.5	137.1	147.1	142.9	134.9	143.6	142.3	144.0	140.2
125.9	132.7	152.9	147.9	141.8	141.4	140.9	141.4	146.5	138.7
160.9	154.2	137.9	139.9	149.7	147.5	136.9	148.1	139.2	137.2
134.7	138.5	138.9	137.7	138.5	139.6	143.5	142.9	143.3	145.4
129.4	142.5	141.2	148.9	154.0	147.7	152.3	146.6	146.7	139.9

制表步骤如下：

（1）求极差 极差又称全距，为样本中最大数与最小数之差，即极差 = 最大值 - 最小值。本例极差 = $160.9 - 125.9 = 35$ （厘米）。

(2) 确定分组数和组距

1) 定分组数: 分多少组需根据样本含量(n)的大小而定。若分组过多, 会使计算较繁, 分组过少则计算误差较大。一般来说 n 愈大, 分组愈多。工作中可参考表 2.1 确定分组数。

表 2.1 分组数参考表

n	50	80	100	150	200	300	500	1000
分组数	6~9	7~11	8~12	10~14	12~16	13~18	15~22	18~30

本例 $n=120$, 现确定分12组。

2) 定组距: 每组的最低组限到最高组限的范围叫组距, 通常以 i 表示。

$$i = \frac{\text{极差}}{\text{组数}} \quad \text{本例 } i = \frac{35}{12} = 2.917 \approx 3 \text{ (厘米)}$$

为了整理计算方便, 组距一般取整数。组距调为整数后, 组数就可能比原计划多或少一组。

关于确定组距的方法, 有时并不采用上述计算方法, 而是事先规定一个标准。如1979年中国青少年儿童体质研究的文件规定各年龄组频数分布表一律统一组距为: 身高2厘米, 体重1公斤, 肩宽0.5厘米, 胸围1厘米, 仰卧起坐2次/分, 60米跑0.3秒, 立定跳远5厘米, 400米跑4秒等。

(3) 排列组限 组限就是一个组包括的范围。在排列组限时, 最低一组必须包括最小值在内, 故最低一组的组下限应小于或等于最小值。本例组下限从125厘米开始, 写成125~(见表2.2)。由于分组是等距离的, 以后按顺序每隔3厘米为高一组的组下限。最高一组应包括最大值。注意所排列的组限不能互包和互斥, 如:

1) 125~128 2) 125~127 第一种情况称为互包, 即两个组段中都出现了128厘米这个数; 第二种情况称为互斥, 即两个组段都不包括127~128厘米之间的频数。出现互包和互斥的现象, 都会使整理划记造成混乱。所以排列组限应按表2.2的方式进行。

(4) 划记 将原始数据用划“正”字的方法, 将120人分别划记到各组去。如第一人身高142.3厘米应划在140~一组, 第二人身高156.6厘米应划在155~一组内, 如此进行, 直到120人全部划记完毕, 划记时最好两人合作, 一人念, 一人划记, 念的人在旁检查。

(5) 求 f (各组频数)和 Σf (总频数) 每组内划记的数目为该组频数, 填在 f 栏内。如125~一组内的划记数为1, 故在 f 栏内填1。128~一组内的划记数为2, 在 f 栏内填2。各组 f 栏填写完毕后, 求 Σf , 即各组频数 f 的总和。 Σ 为希腊字母, 读['sigma:], 表示总和。习惯上读总计。

本例: $\Sigma f = f_1 + f_2 + f_3 + \cdots + f_{12} = 1 + 2 + 3 + \cdots + 1 = 120$, 若 $\Sigma f = N$, 则表明划记无遗漏, 完成分组整理。从频数分布表, 可以清楚地看出, 该组男孩的身高是中等高度的人多, 两头的人数逐渐减少。

表2.2 1973年某市120名男孩身高的频数分布表

组 限 (1)	划 记 (2)	频数 (3)
125~	—	1
128~	十	2
131~	下	3
134~	正下	8
137~	正正正正—	21
140~	正正正正正下	28
143~	正正正正—	21
146~	正正正正	20
149~	正正	9
152~	正	4
155~	丁	2
158~	—	1
Σ		120

习 题

- 1、简述积累资料的注意事项。
- 2、简述审查资料的方法和步骤。
- 3、某大学一年级男生的立定跳远成绩如下：

260 258 252 218 250 237 226 267 232 240 250
 252 248 236 239 250 235 238 239 238 240 230
 230 243 263 234 240 228 241 247 250 261 248
 230 231 250 235 260 245 240 231 202 254 252
 250 233 247 239 242 241 243 243 238 265 236
 245 263 241 231 234 237 235 235 259 240 264
 234 195 249 247 260 258 244 251 245 217 250
 256 255 250 242 256 242 253 260 234 250 258
 239 245 249 248 254 247 224 246 226 245 240
 256

单位：cm, $n=100$, 试由此资料作一频数分布表。

第三章 统 计 描 述

统计描述，就是通过综合性的统计指标来反映事物或现象的基本特征和规律性。这些综合性指标包括反映数据集中趋势的指标，反映数据离中趋势的指标，动态分析指标等。

第一节 集中量数

对一组随机变量,如果我们把它整理成频数分布表,就会发现其频数分布是中间频数多,两头频数逐渐减少,具有向中间密集的趋势(见表2.2)。用以描述随机变量集中趋势的量数(常称指标)称为集中量数。在体育统计中常用的集中量数有算术平均数,调和平均数,中位数,众数等。其中算术平均数最为常用。

一 算术平均数

算术平均数(简称均数)是分析计量资料最常用的集中量数,它说明一组观察值的平均水平和集中趋势。均数常用符号 $\bar{X}$ 表示,其计算方法如下:

1、均数的直接算法 均数等于所有观察值之和除以观察值的总个数。即:

$$\bar{X} = \frac{X_1 + X_2 + X_3 + \cdots + X_n}{n} = \frac{\sum_{i=1}^n X_i}{n} = \frac{\sum X}{n} \quad (\text{式3.1})$$

式中 $\bar{X}$ 为均数, $X_1, X_2, X_3, \dots, X_n$ 为各观察值, Σ 为总加符号, n 为观察值的个数。

当观察值的个数不多时,可按式3.1直接将所有观察值相加进行计算。

例3.1 某日径从中长跑组八名队员安静时脉搏率分别为:60, 63, 59, 58, 61, 64, 59, 60次/分,试计算其均数。

$$\text{解: } \bar{X} = \frac{\sum X}{n} = \frac{60+63+59+58+61+64+59+60}{8} = 60.5 \text{ (次/分)}$$

答:该八名队员的安静时平均脉率为60.5次/分。

2、用频数表计算均数 当观察值个数较多时,用直接算法较繁,容易出错。故在样本较大时可先编出频数分布表,再用加权法或简捷法计算均数。

(1) 加权法计算均数 现以表2.2的资料为例介绍计算方法。

1) 制定均数的加权法计算表。在频数分布表的基础上,将第(2)栏换成组中值,再加上 fX 栏即可。见表3.1。

2) 计算组中值,从表3.1第(3)栏可看出,身高在“125~”一组内有1人,在“128~”一组内有2人,……等等。因同一组内每个人的身高是不相等的,故我们可以组中值代表该组每个人的身高值。组中值 = $\frac{\text{本组下限} + \text{下组下限}}{2}$ 。本例第一组的组中值为 $\frac{125+128}{2} = 126.5$,

第二组的组中值为 $\frac{128+131}{2} = 129.5$ 余类推,见表3.1第(2)栏。

3) 计算 fX 栏,第(3)栏与表2.2第(3)栏相同。各组内第(2)栏组中值 X 与第(3)栏频数 f 的乘积即为第(4)栏 fX 。将各组的 fX 相加得 ΣfX ,见表3.1第(4)栏。

4) 按式3.2计算均数。

$$\bar{X} = \frac{f_1 X_1 + f_2 X_2 + \cdots + f_k X_k}{f_1 + f_2 + \cdots + f_k} = \frac{\sum fX}{\sum f} \quad (\text{式3.2})$$

式中 $X_1, X_2, \dots, X_k$ 分别为第一组至第 k 组的组中值; $f_1, f_2, \dots, f_k$ 分别为第一组至第 k 组的频数; ΣfX 为各组内的组中值与频数的乘积之和; $\Sigma f = n$ 。

本例 $\Sigma f = 120$ $\Sigma fX = 17163$