

体育数理统计方法



人民体育出版社

[苏] H·A·马萨利金 著

体育数理统计方法

上海体育学院

钱仲炎 刘 饶 何志敏 译

章钜林 陈及治 刘厚生 校

人民体育出版社

原 版 说 明

书 名: МАТЕМАТИКО-СТАТИСТИЧЕСКИЕ МЕТОДЫ В СПОРТЕ

作 者: Н.А.МАСАЛЬГИН

出版日期: 1974

出版单位: «ФИЗКУЛЬТУРА И СПОРТ»
(МОСКВА)

体育数理统计方法

〔苏〕Н.А.马萨利金 著

钱仲炎 刘 饶 何志敏 译

章炬林 张及治 刘厚生 校

人民体育出版社出版

妙峰山印刷厂印刷 新华书店北京发行所发行

787×1092毫米 1/32 87千字 6.625印张

1984年10月第一版 1984年10月第一次印刷

印数1—8,530册

统一书号: 7015·2201 定价: 0.72元

责任编辑: 卢 伟 骆勤方

序 言

现在，如果把体育运动排斥在科学之外，那是不可设想的。正确地组织学校体育，能增强学生的体质，并能使他们的运动能力得到协调发展。只有对运动员进行有效的训练，才能提高他们的运动成绩。但这一切都必须以科学为基础。

许多不同学科的专家：如医学专家、生物学家和社会学家，以及从事实际体育教学和训练任务的中小学教师、教育工作者和教练员，他们都是体育这门科学的创造者。

科学，首先要求具有正确的观点，占有可靠的资料，能提出具有客观依据的结论。但要想做到这一点就离不开数据。不论是教师评定学生的成绩，还是教练员评定本运动队参加比赛的成绩，也都离不开数据，这些数据正是科学研究的原始依据，如果，我们所搜集的数据，能揭示某种具体规律，那么，工作就更具有科学意义了。例如，把跑、游泳、滑冰、竞走和自行车等项目的运动成绩系统地进行整理，就可以建立运动员的成绩与动作速度之间相关的一般数学表达式。又如，把举重运动员在训练课上所举的各种重量(公斤数)加以统计，并把统计数字与他的运动成绩进行比较，就可以确定一种能获得最好运动成绩的训练负荷。

现在，凡是其工作与体育运动有关的人，都在运用各种具有实践意义或科学价值的数据。但也应当指出，一定要善于运用数据才行。本书能使你学会如何运用数据，如何系统地整理数据，并从中找到它们的规律性和主要联系，同时它还能使你学会如何查明所收集的数据是否可靠。

本书作者是生物学副博士、列宁格勒体育学院生理教研

组副教授。他担任体育学院本科学生、研究生和教师的数理统计课已有多年。在运用数理统计方法处理体育运动数据方面积累了丰富的经验，因此他在本书中所提出的一些数理统计方法能很好地满足广大体育工作者的需要。

可以预料，如果对本书所提出的体育统计方法进行系统的研究，那就会给教师、教练员、学生科研小组、研究生以及那些为体育运动提供科学依据的各邻近学科的科研人员带来很大的帮助。

生物学博士、教授B.C.法尔费利

引 言

在我们分析科学研究的调查结果时，当教练员分析训练负荷或对照试验的材料时，都会碰到构成某总体的一系列测试数据，这些数据取自测定指标，或用作质量估计。通常我们把单个运动员作为某一总体的组成部分，但对同一运动员身上的研究指标进行单独测定时，这些指标也可作为以个人为某一总体的组成部分。比如说，在分析个人训练的负荷量时，可以把对该运动员在不同时间记录下来的某些负荷量或强度作为研究总体的各个组成部分。总体的每一组成部分都可能具有一系列指标值，而且这些组成部分的某些指标值往往是完全相同的，但另一些指标值则是随着组成部分的变化而变化，即产生变异。例如，总体的各个组成部分可能是同一运动项目的、同一技术水平的、同一年龄和同一性别的一些运动员，但这些运动员的肌肉力量、动作速度以及呼吸循环系统活动的指标则各有差异。

这种变化的指标通常正是我们所要研究的对象。因此，一个教练员或体育科研人员必须经常从各种不同的人或同一个人身上多次测定这种变量。由于这种指标与许多随机因素有关，故在每一具体情况下，它的取值通常是无法预先确切知道的。这种称为随机变量所固有的规律可以通过概率论和数理统计来加以研究。采用数理统计的方法能帮助我们在分析体育活动时作出有科学依据的客观结论。为了进一步提高运动训练水平，教练员和从事运动生理、运动心理、运动训练理论与方法和医务监督等研究的科学工作者必须深入学习有关概率论和数理统计方面的知识。

本书对在科研工作和实践活动中分析训练负荷和评定训练水平时最常采用的一些基本的数理统计方法作了介绍。为了更好地理解上述方法的理论基础及其概率分析问题，本书还简明扼要地介绍了概率论的基本概念。

目 录

序 言

引 言

第一章 经验数列的分布及其特点	(1)
第一节 观察值的分组.....	(1)
第二节 数列分布的图示法.....	(4)
第三节 数列分布的位置特征.....	(6)
第四节 观察值的离散指标.....	(7)
第五节 用“假定均数”(以 X_0 表示)的方法 计算 均 数和方差.....	(7)
第二章 概率论的基本知识	(14)
第一节 随机事件.....	(14)
第二节 随机变量.....	(21)
第三章 正态分布	(25)
第一节 正态分布的定义.....	(25)
第二节 标准正态分布.....	(27)
第四章 均数的假设检验	(29)
第一节 制订统计检验标准的一般准则 与《u》检 验.....	(29)
第二节 t(Student)检验.....	(35)
第五章 两均数差的假设检验	(38)
第一节 t(Student)检验.....	(38)
第二节 X(万·德尔·瓦尔登)检验.....	(43)
第六章 观察值相关两总体的比较	(48)

第一节	t (Student) 检验	(48)
第二节	符号检验 和 Wilcoxon 检验	(51)
第七章	正态分布的假设检验	(55)
第一节	利用偏态和峰态指标的假设检验	(55)
第二节	W (Wilcoxon) 检验	(63)
第八章	简单线性相关与回归	(66)
第一节	函数关系和相关关系	(66)
第二节	关于两项相关的形式和相关密切程度的分 析	(73)
第三节	两项相关指标和回归指标的计算方法	(81)
第四节	因果关系与随机关系	(93)
第九章	多元线性相关与回归	(96)
第一节	有两个自变量的线性回归分析	(96)
第二节	有 $m-1$ 个自变量的线性回归分析	(102)
第三节	偏相关	(108)
第十章	简单非线性相关与回归	(112)
第十一章	相关与回归指标的假设检验	(125)
附 录		(136)

第一章

经验数列的分布及其特点

第一节 观察值的分组

正如引言中所提到的，一个研究人员（教练、科学工作者）在他的研究过程中，会取得一系列数据，这些数据正反映了他所研究的总体中样本的特征。上述所取得的数据通常称之为观察值或变量。为了对这些数据进行分析，就必须将他们进行排列。如果观察的数据不多，那就只要按递增或递减的顺序把它们排列起来就行了。如果观察的数据有几十个，甚至几百个，那就得把这些数据先进行分组。这可以举下面的例子来说明。

例一：为了了解35名自行车运动员的身体训练水平，我们测定他们每个人原地纵跳的高度是：53, 48, 48, 53, 52, 53, 48, 56, 50, 52, 50, 49, 51, 62, 58, 52, 56, 54, 45, 58, 47, 44, 50, 52, 59, 45, 60, 49, 39, 55, 48, 47, 55, 51, 54（单位：厘米），注。为了进行分组，必须把观察值的最大值与最小值之间的全距分成若干区间，或通常称之为组段。本例的最小观察值为39，最大观察值为

注：取自B·A·巴拉诺夫斯基的资料

62, 即所有的观察值都分布在39—62全距内。在确定组段的界限时, 不一定要把第一个组段的下限定为最小观察值, 把最后一个组段的上限定为最大观察值。在选择组段的界限时, 最好是使最小观察值大约处在第一个组段的中点, 最大观察值处在最末一个组段的中点。

在进行分组时, 必须先确定组段数、组距和第一个组段的下限。组段数可根据观察值个数的多少来确定。表1所示, 可作为确定组段数的一种参考。

确定组段数后, 便可按下列公式求出组距:

$$C = \frac{X_{\max} - X_{\min}}{p - 1}$$

公式中的C表示组距, p表示组段数, $X_{\max}$ 代表观察中记录下来的最大观察值, $X_{\min}$ 代表最小观察值。

表 1

观 察 个 数	组 段 数
30—60	6—8
60—100	7—10
100—200	9—12
200—500	11—16

第一个组段的下限(t_0)可按下列公式求得:

$$t_0 = X_{\min} - \frac{c}{2}$$

本例中的观察个数(n)是35(参考表1), 所以在分组时最好

是确定为6—7个组段。假如我们设 $p = 7$, 那么 $C = \frac{62 - 39}{7 - 1}$

$= 3.84$ 。最好把求得的价值化为整数, 即取4.0为组距, 然后按下列公式: $t_0 = 39 - 4/2 = 37$ 求出第一组段的下限。于是, 凡符合于不等式 $37 < X \leq 41$ 的观察值 X 就放在第一个组段; 符合于不等式 $41 < X \leq 45$ 的观察值放在第二个组段……。

在登记了落在各个组段的观察值以后, 就可把结果象表2所示那样记录下来。把各个组段的界限填入第1栏, 把用 n_i (见译者注) 表示的频数填入第2栏。每个组段的频数是指该组段内观察值的个数。

表 2

组 限 (厘米)	n_i
37—41	1
41—45	3
45—49	8
49—53	12
53—57	6
57—61	4
61—65	1
总 数	35

这个表叫做频数分布表, 它可告诉我们, 被测指标的各种不同数值是怎样分布的, 即它们在各组段出现的次数有多少。第一栏中的组限还可以用组中值来表示; 第二栏中的频

译者注: n 表示观察值的个数, 也称频数; i 为角标, 表示组序, 可以取1, 2, … 等自然数; n_i 表示第 i 组的观察值个数。

数也可改用频率，即用该组段观察值个数的比率来表示。频率通常记为 W_i 。它表示频数和观察值总数之比定义为：

$$W_i = \frac{n_i}{n}$$

表 3 所示为例一的频数分布。第 1 栏列举的是各个组段的组中值 (X_i)，第 2 栏列举的是频率 (W_i)。

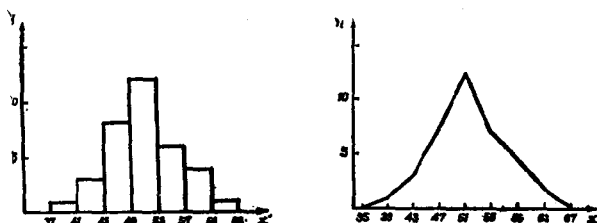
表 3

X_i	W_i
39	0.029
43	0.086
47	0.228
51	0.343
55	0.171
59	0.114
63	0.029
总 数	1.000

由此可见，数列分布能为说明研究结果提供一个简便而又严密的方法。

第二节 数列分布的图示法

为了加强直观性，研究结果还可以用图示法来表示。一般采用的图有两类，即：(1)直方图(图一,A)，(2)多边形图(图一,B)。在作直方图时，观察值 X (各个组限值)是沿着横坐标轴分布的，而所谓分布的相对密度是沿着纵坐标轴



图一 自行车运动员纵跳高度的分布
A—直方图，B—多边形图

分布的,这种相对密度 $f_{\text{相}}$ 等于组段频率除以组距所得的商。
即

$$\text{相对密度 } f_{\text{相}} = \frac{W_i}{C_i}$$

比值 f_i 反映了在各组段 i 中,平均每单位区间内的观察值个数(频数)。

在每一组段的底边上画一个高度为该组段的相对密度 f 的直方四边形。这样,所求出的四边形面积值就是各对应组段的频率值,即该组观察值的个数与观察值总数的比率。但如果对观察值进行分组时,各组段的组距相同,那么只要沿纵坐标简单地标出频率或频数,就可以作出直方图。这样作图比较容易,但在说明直方图时,必须记住,与组段频率对应的是该四边形的面积(而不是它的高度)。有时我们也采用各组段不同的组距。在这种情况下作直方图时就得沿纵坐标轴标出分布的相对密度 $f_{\text{相}}$ ($f_{\text{相}} = \frac{W_i}{C_i}$),或分布的绝对密度 $f_{\text{绝}}$ ($f_{\text{绝}} = \frac{n_i}{c_i}$)。

在作多边形图时,观察值也是沿着横坐标轴分布的,但在横坐标轴上标出的不是组限,而是各组段的组中值,沿着

纵坐标轴标出的则是频数。可对准每个组段的组中值，及在相应的频数水平上取点，然后再用直线把这些点连起来。为了绘成多边形图，还需在左右两端组段，各标上一个频数为0的组中值，并用直线分别把这两个组中值与相邻组段的点连接起来。

如果图中只需表示一种数列分布，那么选用直方图或者多边分布图都可以。如果研究人员想用一个图来同时说明几个总体的、同一指标的分布状态，那当然最好是采用多边形图。

第三节 数列分布的位置特征

观察值分组的主要目的是要以明显、简便而又严密的形式来表示观察值，以便进行分析。对此，最重要的是要定义出几个统计指标。这些指标能用数量(数的形式)来定量地反映所研究的总体分布的基本特征，同时还能通过这些指标对各个不同总体进行比较。数列在数轴上的分布状态是数理统计对总体所认定的重要特征之一。这种特征可以用各种统计指标予以反映，最常采用的是算术平均数，简称为均数。表示均数的符号与指标本身所用的符号相同，只需在符号上另加一横线。

例如，设指标是用 X 来表示，那么它的均数就用 $\bar{X}$ 来表示。如果我们取得某一指标的观察值为 n 个： $X_1, X_2 \dots X_n$ ，那么

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{\sum X_i}{n} \quad (1.1)$$

这里的 Σ (大写的希腊字母)是总和的符号， i 为角标， $i=1, 2, 3 \dots n$ ， X_i 表示第 i 个 x 值， $\sum X_i$ 表示 $x_1 + x_2 + x_3 \dots + x_n$ 。

分子有时可直接写成 ΣX 。算术平均数可按上述公式求出，下面将要谈的，是在许多情况下有可能简化计算程序的某些数学性质(参看本章第五节)。

第四节 观察值的离散指标

均数是反映分布集中位置的一种重要特征，但在均数相同的条件下，观察值偏离均数的分布情况却不一定相同。因此还必须定义反映观察值离散程度的指标，方差就是这样的一种重要指标，它用 S^2 表示，并可用下列公式求得：

$$S^2 = \frac{\Sigma (X - \bar{X})^2}{n - 1} \quad (1.2)$$

分子 $\Sigma (X - \bar{X})^2$ 代表各观察值偏离均数的离差平方和，分母为自由度，它等于观察值总个数减1。在制订各种指标时，都广泛运用方差，比如说，可以用它来比较两个不同总体的观察值的离散程度。方差的平方根叫做均方差，或称标准差，以 S 来表示。方差和标准差与度量单位有关，因此在估计观察值的离差时，有时采用无量纲的指标，即变异系数或变差系数 V 。它是用百分比来表示的，其计算公式为：

$$V = \frac{S}{\bar{X}} \times 100\%$$

变异系数可用来鉴别观察值的同类性。如果它不超过10%的话，那就可以把这些观察值看成是同一类的。

第五节 用“假定均数”(以 X_0 表示)

的方法计算均数和方差

计算方差和标准差除了直接按上述公式以外，还可以采

用其它一些计算方法。由于这些方法运用了指数的下述一些数学性质，因此可大大简化计算的程序。

如果从所有的观察值中减去(或加上)同一个数，那么，其均数也等于减去(或加上)与此相同的一个数，而这时的方差仍然不变，即如果 $Y = X - a$ (a 表示常数)，那么 $\bar{y} = \bar{x} - a$ ，而 $S_y^2 = S_x^2$ ；如果所有的观察值除以(或乘以)同一个常数 a ，那么新均数就等于原均数除以(或乘以) a ，而这时的方差也等于原方差除以(或乘以) a 的平方，即如果 $Y = \frac{X}{a}$ ，那么 $\bar{Y} = \frac{\bar{X}}{a}$ 而这时的 $S_y^2 = \frac{S_x^2}{a^2}$ (S_y^2 表示每数除以 a 后的方差——译者注)。

方差分子中的离差平方和，可以用去括号的方法进行换算(可获得离差平方和的简算公式——译者注)：

$$\sum (X - \bar{X})^2 = \sum X^2 - \frac{(\sum X)^2}{n} = \sum X^2 - n\bar{X}^2$$

由于上述数学特性，因此在计算 $\bar{X}$ 和 S^2 时就有可能运用简算公式，以简化计算中的原始数字。在计算离差平方和时，也不一定要运用分数。

下面是用“假定均数”(以 X_0 表示)的方法计算均数和方差的程序：

第一种情况 采用不分组的数据。如果原始观察值是按下列公式进行换算：

$$X' = \frac{X - X_0}{k}$$

这里的 X_0 是代表假定均数(一般是从观察值中任意取一个接近均数的整数)， k 是代表所有观察值的公约数(如果这样的公约数存在的话)，那么根据上述 1、2 两个性质，