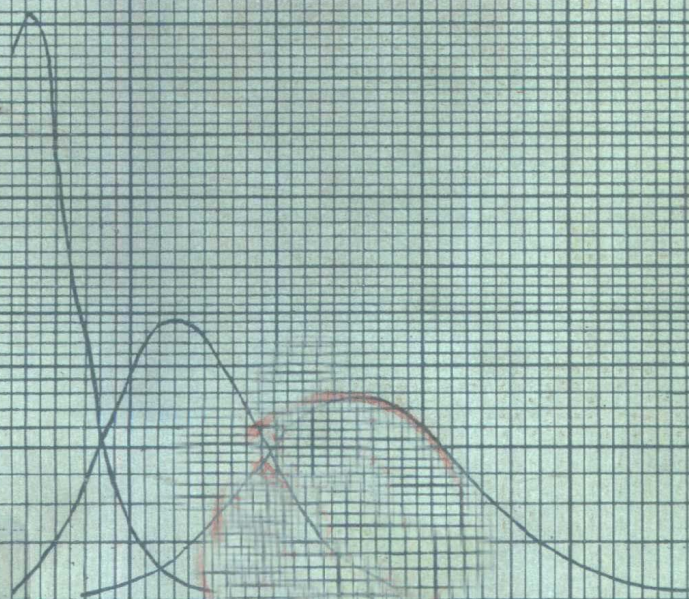


# 体育统计

施丽影 编



武汉体院教务处

15325

32  
7

# 体 育 统 计

施丽影 编

武汉体育学院教务处

一九八三年五月

## 序 言

体育科学是一门涉及多种学科的综合科学，随着体育事业的发展，训练手段的科学化，现代化，出现了一系列需要用体育统计方法解决的问题。如1)如何正确地组织学校的体育教学，合理地评定学生的成绩？2)如何合理的制定锻炼身体的标准，增强学生体质，并使他们的运动能力得到发展？3)如何科学地进行选才，制定出各级各类选材标准？4)如何确定恰当的运动量及运用合理的训练手段进行科学训练，以提高运动员的运动水平？体育统计这门学科就是为解决这些问题而产生和发展起来的。它是数理统计在体育领域里的具体应用。

近几年来，由于体育事业的不断发展，作为科学研究的一种工具——体育统计也愈来愈显示其重要性。要做到体育教学和训练的科学化、现代化，最为具体而基本的工作是要把握住体育教学与体育运动中各种各样生动活泼的信息——数据。不论是教师对学生成绩的评定，还是运动队参加比赛所获得的成绩，还是通过训练后学生体质的增长情况，这些信息都是通过数据表露出来的，而这些数据正是科学研究的原始依据。如果你想找出一种有效的训练方法，或者找出某事物的某种规律性，你就必须有目的的去获取数据。所以统计工作的第一阶段就是研究如何科学地获取数据，在这部分我们将介绍如何使试验观察次数少，而又能取得较多的，有代

AA276/12

表性的，可靠的信息，这是统计工作中一个非常重要的问题。试验设计，抽样方法都是从不同的角度来研究这个问题的。试验设计是考虑如何合理地安排试验，正交试验设计就是其中一部分，它是从农业试验需要开始的，现在更多地应用于工业试验，国防试验上。在体育中如何合理的运用，还有待大家去开拓。抽样技术则是考虑从总体中如何随机地取出部分样本，去回答总体的问题。

统计工作的第二阶段就是进行统计推断，主要包括统计假设检验与参数估计。我们将着重介绍统计假设检验。

统计工作的第三阶段就是统计预报。即依据以往某条件下发生事件所得到的数据，找出条件与事件发生的统计关系，研究如何预报将来条件下可能发生的事件。常见的有天气、地震、水文等的预报，它是统计推断的直接产物。

为了回答不同的问题，就产生了统计推断的不同分支。如对于分布族不能以有限个未知参数刻划建立起来的非参数估计；研究自变量与因变量之间关系的相关分析与回归分析；有判断一些因素对我们所考察的变量是否有显著作用的方差分析；有讨论多个总体多个指标的统计推断的多元分析等等。

本书是根据目前在体育运动及教学中经常遇到的问题所涉及的一些统计方法汇编而成的，所要求的数学基础知识是适中的。它可以作为体育院校的试用教材。为了使具有中等数学程度的同志也能掌握这方面的内容，注意避开了涉及到的一些高深数学，如雅可比行列式、矩母函数和特征函数，也没有用矩阵和 $n$ 维空间向量，使不熟悉这些内容的读者在阅读时也不会感到困难。

在编写过程中、学习和参考了兄弟院校的交流版本，深受教益。它对本书的成文是很有帮助的。同时我院教处务的有关负责同志及许梧雄老师也提出了宝贵意见，并给予了具体协助。在此一并致谢。

由于作者水平有限，书中有不妥之处，恳请读者批评指正。

# 目 录

|          |       |
|----------|-------|
| 序 言..... | ( 1 ) |
|----------|-------|

## 第一章 资料的收集与整理

|                  |       |
|------------------|-------|
| § 1.1 资料的收集..... | ( 1 ) |
| § 1.2 资料的整理..... | ( 3 ) |
| 习题一.....         | ( 8 ) |

## 第二章 统计特征数

|                            |        |
|----------------------------|--------|
| § 2.1 表示集中趋势量度的量——平均数..... | ( 9 )  |
| § 2.2 表示离散趋势量度的量——标准差..... | ( 17 ) |
| 习题二.....                   | ( 25 ) |

## 第三章 概率论的基本概念

|                 |        |
|-----------------|--------|
| § 3.1 随机事件..... | ( 27 ) |
| § 3.2 排列组合..... | ( 32 ) |
| § 3.3 二项式定理     |        |
| 习题三.....        | ( 44 ) |

## 第四章 随机事件的概率

|                        |        |
|------------------------|--------|
| § 4.1 频率与概率.....       | ( 49 ) |
| § 4.2 互不相容事件的概率.....   | ( 53 ) |
| § 4.3 条件概率.....        | ( 55 ) |
| § 4.4 独立事件同时发生的概率..... | ( 58 ) |

|                   |        |
|-------------------|--------|
| § 4.5 小概率事件原则及其应用 | ( 65 ) |
| 习题四               | ( 67 ) |

## 第五章 随机变量及其分布

|                 |        |
|-----------------|--------|
| § 5.1 随机变量的概率分布 | ( 71 ) |
| § 5.2 二项分布      | ( 74 ) |
| § 5.3 正态分布      | ( 78 ) |
| 习题五             | ( 96 ) |

## 第六章 抽样的基本知城

|                              |         |
|------------------------------|---------|
| § 6.1 抽样的方法                  | ( 98 )  |
| § 6.2 样本大小的预定                | ( 100 ) |
| § 6.3 几种常见的抽样分布及所依据的<br>几个定理 | ( 103 ) |
| § 6.4 抽样误差的计算                | ( 108 ) |

## 第七章 假设检验

|                 |         |
|-----------------|---------|
| § 7.1 基本概念      | ( 113 ) |
| § 7.2 平均数的假设检验  | ( 122 ) |
| § 7.3 总体率的检验    | ( 128 ) |
| § 7.4 标准差的差异性检验 | ( 131 ) |
| § 7.5 符号检验      | ( 135 ) |
| § 7.6 秩和检验      | ( 137 ) |
| § 7.7 分布的适合性检验  | ( 139 ) |
| § 7.8 独立性检验     | ( 148 ) |
| 习题七             | ( 156 ) |

## 第八章 方差分析

|                     |         |
|---------------------|---------|
| § 8.1 单因素的方差分析..... | ( 158 ) |
| § 8.2 双因素的方差分析..... | ( 167 ) |
| § 8.3 多重比较.....     | ( 177 ) |
| 习题八.....            | ( 181 ) |

## 第九章 回归分析

|                     |         |
|---------------------|---------|
| § 9.1 一元线性回归分析..... | ( 184 ) |
| § 9.2 二元线性回归分析..... | ( 197 ) |
| § 9.3 逐步回归.....     | ( 203 ) |
| 习题九.....            | ( 215 ) |

## 第十章 聚类分析

|                        |         |
|------------------------|---------|
| § 10.1 指标聚类的方法和步骤..... | ( 212 ) |
| § 10.2 样品聚类的方法和步骤..... | ( 227 ) |

## 第十一章 判别分析

|                      |         |
|----------------------|---------|
| § 11.1 两组判别分析.....   | ( 234 ) |
| § 11.2 极大似然多级判别..... | ( 242 ) |

## 第十二章 研究设计

|                    |         |
|--------------------|---------|
| § 12.1 调查设计.....   | ( 253 ) |
| § 12.2 试验设计.....   | ( 255 ) |
| § 12.3 正交试验设计..... | ( 267 ) |
| 附表.....            | ( 273 ) |
| 习题答案.....          | ( 319 ) |

# 第一章 资料的收集与整理

## § 1.1 资料的收集

在进行某类问题的分析与研究时，首先必须确定研究的目的、对象、研究课题等，然后再有的放矢的进行研究设计：一种是调查设计，包括调查目的、意义、对象、范围、纲目、调查表格、组织计划、实施时间等。另一种是试验设计，它要求用尽量少的试验次数，简单而准确的试验方法，获得足够有效的数据，而对所研究的问题可提供较多的信息，从而能得到准确可靠的结论。因此，对试验提出了以下几点要求：1. 对所取的每个试验对象的测量值，不能有系统的差异。（仪器故障，实验人员的错误，环境差异……）。如考察不同的训练手段对运动员的握力增长情况，那么第一要检查握力器正常不正常，第二排除对握力测试的干扰因素，使测试对象处在相同的条件下。2. 应该选取恰当的试验指标。在试验中如果指标选择不恰当，也很难得出合理的判断结果。如用摄影法分析投掷运动员的投掷角，就很难保证投掷角所在平面与摄影镜面是平行的，这时所测得的角度当然也不可能是准确的。3. 对所测得的数据应能找到相应的数理统计方法进行分析，使所取得的数据能满足统计分析的基本模型。

总之，收集资料的工作是科研工作中极为重要而细微的

工作，所研究的课题的真实性如何？可靠程度如果？主要是依靠这阶段的工作。做这工作时，既要根据统计的原则，又要具有专业知识，这样收集的资材才会真实可靠，且具有科学性与代表性。

为了使所收集的资料具有代表性，统计学家为我们设计了各种随机抽样方法。1) 抽签法：将所研究的对象统一编号，做成签子，随机抽取我们所需要的数目即可。如从600人中抽取20人，可先将600人统一编成1—600号，然后随意抽出20个不同的号码，如抽到了13，25，3，200，……，这20个数所对应的人，就是我们所需要的一个随机样本。2) 随机数表法：随机数表是由0—9的数字，由计算机随机的排列而成的。如表一是5000个数码字的随机数字表，分二页印成，每页上有2500个数字，排成50行50列，可以根据需要进行抽取。如从50人中抽取10人进行测试，用随机数表抽样，步骤如下：首先将50人从1—50编号，然后在随机数表上任意指定一个数字，若为奇数，用第一页；0与偶数用第二页。假定所指的数字为奇数，那么在第一页上再随意指定二个数字，若该数为39，可将3作为行的起点，9作为列的起点。即从第3行第9列开始，向右（或向左、向下、向上皆可）连续取10对数字，去掉其中大于50的数，这时所取的10对号码为50，26，07，32，13，38，14，10，还差二个，移至下一行往左取26，21，即是所抽取的10名号码。

除了上述二种单纯随机抽样方法外，还有机械抽样，整群抽样，分层抽样等，以后再介绍。

下面介绍几个统计学的基本概念：

总体——被研究对象的全体。如研究某校男生百米跑的

情况、总体就是该校全体男生的百米成绩。如对某个运动员的某项运动进行研究,那么对该运动员的某项运动指标的测定就是该总体的组成部分。所以总体可以是同一运动项目的各项指标组成,也可以是同一技术水平或同一年令组或同一性别的运动员。往往有些总体存在,但不一定能获得其全部测量值。

个体——被测总体中每个对象的测量值,如该校全体男生百公尺跑的每一个测量值。

样本——有时总体很大或不可能得到总体的全部测量值,这时需要用适当的方法,从总体中抽取一部分个体,被一次抽取的这部分个体就叫做一个样本。在同一个样本中所含有的个体的数目叫做样本容量。一般来说,选取的样本容量愈大,所得到的推断愈可靠。但事物总是一分为二的,样本愈大,所花费的人力、物力、时间愈多所以应根据问题的需要选取适当大小的样本。通常认为样本容量小于50者为小样本,大于50者为大样本。从样本特征来推断出总体特征,是统计学的主要任务之一。

## § 1.2 资料的整理

在体育运动中所取的原始资料,一般都是没有秩序的,但通过分组、归纳(可用卡片法、划记法)整理以后,它就会呈现出一种规律性,掌握了材料的规律,就可以研究问题。在统计学中,材料的规律是通过频率分布表现出来的。用统计学的术语,随机变量的变化规律,是用分布密度来描述

的。

下面我们介绍一下什么叫频率分布？

在统计学中，频率是指在一次试验过程中，某事件发生的次数与样本容量的比值。

例1.1 掷钱币50次，出现正面次数为27次，则其频率为  $\frac{27}{50}$ 。

一个射手射击10次，则其命中的频率为  $\frac{9}{10}$ 。

频率分布是研究事件发生在某个范围内所占的比例的大小。

为了研究频率分布情况，我们需要对所取的样本数据进行整理。

下面举例说明频数分布表及频数分布图的作法。

例1.2 为了了解中学生身体发育情况，对某中学同年的女同学，随机抽取60名进行身高测量，结果如下（单位厘米）。

|     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 167 | 154 | 159 | 166 | 169 | 159 | 156 | 166 | 162 | 158 |
| 159 | 156 | 166 | 160 | 164 | 160 | 157 | 156 | 157 | 161 |
| 158 | 158 | 153 | 158 | 164 | 158 | 163 | 158 | 153 | 157 |
| 162 | 162 | 159 | 154 | 165 | 166 | 157 | 151 | 146 | 157 |
| 158 | 160 | 165 | 158 | 163 | 163 | 162 | 161 | 154 | 165 |
| 162 | 162 | 159 | 157 | 159 | 149 | 164 | 168 | 159 | 153 |

作其频率分布，并绘制频数直方图。

整理步骤：

一、求极差（计算最大值与最小值之差）。

此例最大值为169，最小值为146

$$\text{极差 } R = 169 - 146 = 23$$

## 二、确定组距与组数

一般数据越多，分的组数也越多。根据苏联H. A. 马萨利金博士提出的组数估计如下：

| 观 察 个 数  | 可分组数   |
|----------|--------|
| 30——60   | 5——8   |
| 60——100  | 7——10  |
| 100——200 | 9——12  |
| 200——500 | 11——16 |

这样，本例60个数据分为8组较适宜。再确定组距，组距 =  $\frac{\text{极差}}{\text{组数}}$ 。本例组距 =  $\frac{23}{8} \approx 3$ （在这里为将每个数据都包含进去，需取过剩近似值，一般取成整数值）。

## 三、确定分组点

上述60个数据等，如从最小的一个数值开始确定每一组的分点（或组限）。可分为下述八组。

146——149   149——152   152——155   155——158  
158——161   161——164   164——167   167——170

如果按上述分点分组，其中149，152，167等几个数据均在分组点上，到底归那一组好呢？为了避免这种情况，我们将分组点的精度提高一位。上述八组的分组点如下：

145.5—148.5   148.5—151.5   151.5—154.5

154.5—157.5

157.5—160.5    160.5—163.5    163.5—166.5

166.5—169.5

或者加以说明，这些数据是属于上组限还是下组限。

#### 四、列频数（频率）分布表（表一）

频数——各小组内含数据的个数

频率——频数与样本容量的比值

频 数 分 布 表

表一

| 分 组         | 划 记  | 频 数 | 频 率   |
|-------------|------|-----|-------|
| 145.5—148.5 | —    | 1   | 0.017 |
| 148.5—151.5 | ┐    | 2   | 0.033 |
| 151.5—154.5 | 正一   | 6   | 0.100 |
| 154.5—157.5 | 正正   | 9   | 0.150 |
| 157.5—160.5 | 正正正┐ | 18  | 0.300 |
| 160.5—163.5 | 正正一  | 11  | 0.183 |
| 163.5—166.5 | 正正   | 10  | 0.167 |
| 166.5—169.5 | 正    | 3   | 0.050 |
| 合 计         |      | 60  | 1.000 |

#### 五、绘制频数分布直方图

作一平面直角坐标，以身高的各分组点作为横坐标各分点，以频数作为纵坐标分点。根据表一中各组频数绘出下列

图形。(见图一)

频数分布直方图

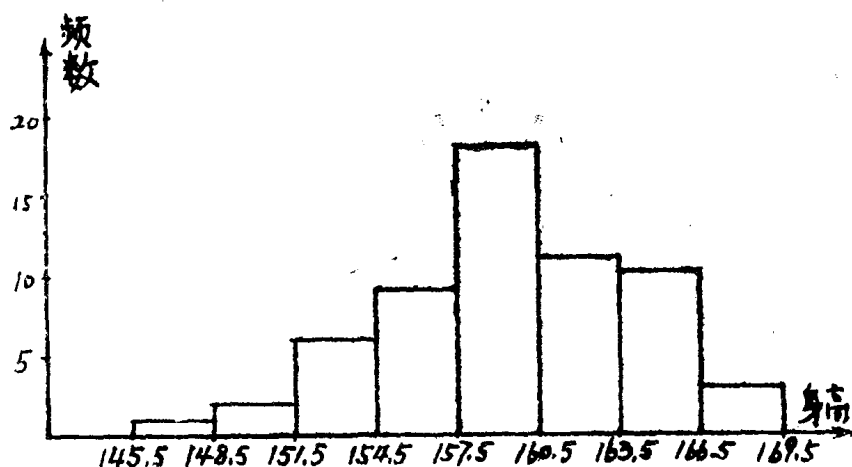


图 一

上述直方图也可绘制成多边形图(或折线图)。这种图形的优点是在同一坐标系内,可以同时绘制几种不同的折线,便于进行比较。

如果将纵轴以各组频率作分组点,这样得出来的图形就称为频率直方图。

如果组距取得相当小,而样本的容量又相当大,这时多边形图又可演变成为一条光滑的曲线。这种曲线称总体分布曲线,(如图二)。它随着总体的不同而具有各种不同的类型。而不同的研究对象可以用不同的曲线去拟合它。如水文

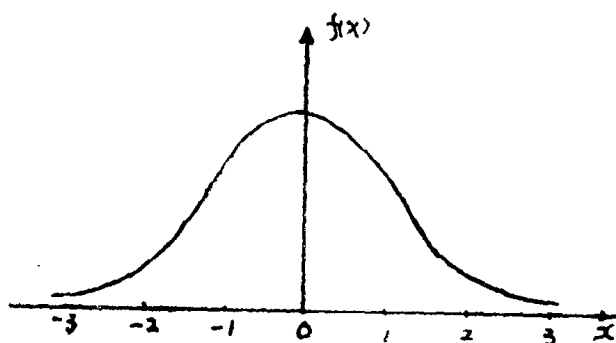


图 二

统计中研究出一种皮尔逊曲线，可以反映出水文资料的特点。到目前为止，在体育方面，到底能用什么样的曲线去拟合它能反映出体育训练的特点，这还是有待研究的问题。

### 习 题 一

1. 为了考察武汉体院77级毕业生考试情况，从中抽查了200名考生的成绩。在这个问题中，总体、个体、样本、样本容量各指什么？

2. 某校初中一年级30名男生进行手球投掷，结果如下：  
(单位：米)

|      |      |      |      |      |      |      |      |
|------|------|------|------|------|------|------|------|
| 22.8 | 15.2 | 18.3 | 17.9 | 25.6 | 12.7 | 20.8 | 21.6 |
| 19.7 | 24.5 | 21.9 | 31.5 | 19.0 | 23.5 | 22.1 | 24.5 |
| 16.0 | 22.3 | 20.3 | 19.3 | 16.9 | 23.1 | 18.8 | 21.7 |
| 20.4 | 18.1 | 14.8 | 22.3 | 24.0 | 21.0 |      |      |

作频数分布表，并绘频数直方图。

## 第二章 统计特征数

当我们获得了大量的观测数据以后，必须由实施者对大量的资料加以概括和取舍，这样才能取得有意义的数据——称为描述性的值。即能将个体与个体，组与组之间进行对比的那种值。这种值通常有二种形式：一种是表示集中趋势量度的量——平均数；一种是表示离散趋势量度的量——标准差，变异系数。

### §2.1 表示集中趋势量度的量——平均数

一、平均数的意义：它是表示大量过程的水平，也就是对总体或样本的各观测值作概括地说明，把所研究对象的特征表现出来。平均数又包含算术平均数、中位数、众数、调和平均数、几何平均数等等，后面二种在我们的讲义中用得不多就不介绍了。

1. 算术平均数：将总体资料的各所有测量值之和除以总体中所有个体的个数所得的商，称为总体的算术平均数，将某一样本的每个测量值之和除以样本容量所得的商，称为样本的算术平均数。通常简称为均值，用符号 $\bar{X}$ 表示。

设在总体 $X$ 中，取一个样本 $(x_1, x_2, \dots, x_n)$ ，则样本的均值为