

提高学校教学质量 教师应知应会教学用书
师资岗前培训教材 师范类必修课教材用书

课程考试

● 马龙友 著

的编制与评价

P I N G J I A

B I A N Z H I

KECHENG
KAOSHI

DE BIANZHI YU PINGJIA
——为学校建立课程考试分析与评价系统提供理论、方法和实例
——帮助教师科学地编制试题与试卷
——帮助教师定量分析出所教课程质量



中国轻工业出版社

提高学校教学质量
师资岗前培训教材

教师应知应会教学用书
师范类必修课教材用书

课程考试的编制与评价

马龙友 著

为学校建立课程考试分析与评价系统
提供理论、方法和实例
帮助教师科学地编制试题与试卷
帮助教师科学地评价所教课程质量

 中国轻工业出版社

图书在版编目 (CIP) 数据

课程考试的编制与评价/马龙友著. —北京: 中国轻工业出版社, 2009. 4

ISBN 978-7-5019-6679-0

I. 课… II. 马… III. 考试学 IV. G424.74

中国版本图书馆 CIP 数据核字 (2008) 第 163621 号

责任编辑: 古 倩 责任终审: 孟寿萱 封面设计: 品创设计
版式设计: 王超男 责任校对: 吴大鹏 责任监印: 张 可

出版发行: 中国轻工业出版社 (北京东长安街 6 号, 邮编: 100740)
印 刷: 三河市世纪兴源印刷有限公司
经 销: 各地新华书店
版 次: 2009 年 4 月第 1 版第 1 次印刷
开 本: 787 × 1092 1/16 印张: 8.75
字 数: 180 千字
书 号: ISBN 978-7-5019-6679-0 定价: 20.00 元
读者服务部邮购热线电话: 010-65241695 85111729 传真: 85111730
发行电话: 010-85119845 65128898 传真: 85113293
网 址: <http://www.chlip.com.cn>
Email: club@chlip.com.cn
如发现图书残缺请直接与我社读者服务部联系调换
80536J5X101ZBW

前言

考试和课堂教学是学校教学过程中相互联系的两个重要方面。

教师既要重视课堂教学的研究与质量提高,也要重视考试的研究与质量提高。考试的功能是多方面的,主要表现在以下三方面:

1. 评价教学和学习成绩

考试可以评价教师教学水平,促进师资水平提高;还可评价和鉴定教材和教学方法,提高教学质量。根据单元考试、期中与期末的综合考试等,可以学生的学习成绩作出评定,使学生了解自己的学习水平,进行自我评价,提高学习质量。

2. 因材施教

依据考试结果,教师可以比较准确地了解学生学习情况,有针对性地指导学生学习,做到有的放矢,因材施教。

3. 选拔人才

利用考试这一手段可以选拔满足实际需要的人才,且达到高效、准确的目的。如普通高等学校招生全国统一考试。

考试是需要的,在受到重视的同时,还应不断提高考试质量,朝着科学化方向发展,发挥更大作用。

作为学校的一名教师,会出考试题,会组织各类考试是不够的,还应当了解和掌握有关考试的一些基本理论与评价方法。实际上有些教师对考试学的一些基本理论与评价方法是不熟悉的,这对提高教学与考试质量及改善考试现状等方面的工作都是不利的。这种情况出现,原因是多方面的:一是考试学的基本理论与评价方法在学校教学中还没有受到应有重视,普及宣传工作也没跟上;二是很多教师在其专业学习过程中没有学过教育统计学、教育测量学与教育评价学,而考试学体现在这三门学科中,是它们的实际应用;三是缺少满足各级各类教师需要的介绍考试学基本理论与评价方法的书,以及对新教师举办的岗前培训班,应有考试学基本理论与评价方法的介绍,大多未列入培训内容。基于上述原因,需要出版介绍考试学基本理论与评价方法的书。

本书全面系统地介绍了考试的基本理论与评价方法。内容包括:考试分数的特征量与分布,考试分数的转换与作用,试题的题型与编制,考试试卷的编制,试题与试卷的质量评价,考试的质量评价,建立课程考试编制与评价系统,考试统计分析 with 考评报告等。

本书注重考试观念的更新。改变应试教育的思想观念,即改变把考试结果作为衡量教育教学质量和学生学习水平的唯一价值尺度的思想。考试是为素质教育、为培养人才这个目的服务的。

利用考试资源,做好考试评价工作要做到:第一,考试要突出学生的主体地位。要考出学生的长处和优点,要能够把学生的兴趣、能力、潜质等特征性的东西表现出来。通过考试提高学生的能力和素质,鼓励每个学生发展自己的特长。考试要关注学生的发展过

程,把形成性评价和终结性评价结合起来,使发展变化的过程成为评价的组成部分。第二,要使考试具有较多的评价内涵。对学生的能力、学力和潜质作出测评。考试为学生不仅仅提供一个分数,还要通过各种数据给学生知识、能力和潜质的报告,能通过“成绩报告单”的形式体现,使学生了解自己哪些方面是强项,哪些方面是弱项,让他们知道往哪个方向发展最适合,鼓励他们向自己的专长发展。第三,建立课程考试分析与评价系统,考试朝着标准化和科学化方向发展,使考试统计数据能够形成一个客观标准,做好考试质量评价工作。建立课程考试分析与评价系统是一项系统工程,应达到试卷编制过程科学化、考试实施过程管理科学化、考试评分过程科学化、评价考试结果科学化。

本书有以下特点:

1. 在结构体系上,将教育统计、测量与评价三方面内容有机结合,应用到考试学的基本理论与评价方法上。教师不了解教育统计学、教育测量学与教育评价学也可以阅读本书。

2. 在应用上,将考试评价指标的测量方法与实际应用相结合,可实现用计算机完成考试结果统计分析,给出考试评价结果。

3. 在内容撰写上,简明扼要、条理清晰、由浅入深、循序渐进、直观性强,便于读者阅读。

总之,本书具有科学性、实用性及简明性。

本书为学校建立课程考试分析与评价系统提供了理论、方法和实例。学校建立课程考试分析与评价系统的作用是可评价出班级(年级)课程考试质量,可评价出试题、试卷和考试本身质量,可分析出教学取得成绩和存在问题。学校建立课程考试分析与评价系统的意义是可促进学校教学质量提高,可促进教师教学水平提高,可促进学生学习水平提高,可促进学校实施考试科学化。学校内每一位任课教师都应对所教课程进行考试分析与评价,故建立课程分析与评价系统具有广泛的实用性和群众性。

本书可帮助教师科学地编制试题和试卷,可帮助教师科学地评价出所教课程质量。

本书读者对象为各级各类教师,教育行政管理干部,考试管理干部,师范院校学生等。本书可作为师资岗前培训教材,师范院校的公共必修课参考教材。

本书作者从事学校教学工作40多年,深知考试工作的重要性,鉴于当前考试评价中的一些概念与测量方法还不是很成熟,可供参考的资料也不是很多,实践应用上还不够广泛,决定写一本关于考试编制与评价的书。作者虽在考试评价实现科学化、量化与计算机化方面进行了实践,对书中内容涉及到评价指标的定义、数学模型和计算公式进行了研究,写作过程做到认真严谨、反复修改,书中还是难免存在问题和不足之处,非常希望得到各位老师、专家和学校的帮助与指导。

写作本书得到许多同志的大力支持和热情帮助,在此深表谢意。

感谢阅读本书的同志,若对您的考试工作有所帮助,这就达到了作者写本书的意愿。

作 者

2001年12月完成第一次书稿撰写

2005年1月第一次书稿印刷

2007年6月完成第二次书稿撰写

2008年8月8日完成第二次书稿修改

目 录

第一章 考试分数的特征量与分布	(1)
第一节 抽样方法与原始分数的登记	(1)
一、随机抽样	(1)
二、等距抽样	(2)
三、分层抽样	(2)
四、整体抽样	(3)
五、分阶段抽样	(3)
六、原始分数的登记	(3)
第二节 考试分数的特征量	(4)
一、平均分	(4)
二、中位数	(6)
三、众数	(7)
四、方差与标准差	(7)
五、差异系数	(9)
六、相关系数	(10)
七、用电子计算器计算特征量	(12)
第三节 考试分数的频数(频率)分布直方图与正态分布	(13)
一、考试分数的频数及频率分布表	(13)
二、考试分数的频数(频率)分布直方图	(14)
三、考试分数的分布服从正态分布	(15)
四、考试分数分布的检验	(17)
第二章 考试分数的转换与作用	(21)
第一节 考试分数的转换	(21)
一、原始分数	(21)
二、绝对分数与相对分数	(22)
三、线性转换的标准分数	(22)
四、其他线性标准分数	(23)
五、正态转换的标准分数	(24)
六、高考总分的正态转换	(24)
七、百分等级	(25)
八、等级记分与合格分数	(25)
第二节 考试分数的作用	(26)
一、目标参照性考试	(26)

二、常模参照性考试	(27)
第三章 试题的题型与编制	(28)
一、客观题编制的一般原则	(28)
二、编制客观题的题型	(28)
三、主观题编制的一般原则	(30)
四、编制主观题的题型	(31)
第四章 考试试卷的编制	(34)
一、确定考查知识内容及百分比	(34)
二、确定考查学科能力要求及百分比	(35)
三、编制命题双向细目表	(37)
四、编制考题	(38)
五、组卷与试题分值的配置	(40)
六、编制考卷标准答案与评分细则	(41)
七、审查考试试卷	(41)
第五章 试题与试卷的质量评价	(43)
第一节 试题的质量评价	(43)
一、试题难度	(43)
二、试题区分度	(46)
第二节 试卷的质量评价	(51)
一、试卷难度	(52)
二、试卷区分度	(53)
第六章 考试的质量评价	(55)
第一节 考试信度	(55)
一、考试信度的经典理论	(55)
二、考试信度的评估方法	(57)
三、评分者信度	(63)
第二节 考试效度	(69)
一、考试效度的概念	(69)
二、内容效度	(72)
三、效标关联效度	(73)
四、结构效度	(75)
五、提高考试效度的途径	(77)
第七章 建立课程考试编制与评价系统	(78)
一、试卷编制过程科学化	(78)
二、考试实施过程管理科学化	(79)
三、考试评分过程科学化	(81)
四、评价考试结果科学化	(81)

第八章 考试统计分析与考评报告	(83)
一、考试评价总结工作	(83)
二、班级课程考试成绩统计分析表及说明	(84)
三、班级课程考试成绩统计分析汇总表及说明	(86)
四、年级课程考试成绩统计分析表及说明	(86)
五、评价试题、试卷和考试质量统计分析表及说明	(87)
六、使用教务网络管理系统进行考试评价应注意的几个问题	(94)
附论文 1 建立课程考试分析与评价系统的意义和作用	(99)
附论文 2 分析统计数据 提高教学质量	(102)
附论文 3 科学地解释与利用大学英语四、六级考试分数	(107)
附表 1 班级课程考试成绩统计分析表	(114)
附表 2-1 班级课程考试成绩统计分析汇总表	(115)
附表 2-2 年级课程考试成绩统计分析表	(116)
附表 3 评价试题、试卷和考试质量统计分析表	(117)
附表 4 《班级课程考试成绩统计分析表》实例	(119)
附表 5-1 《班级课程考试成绩统计分析汇总表》实例	(121)
附表 5-2 《年级课程考试成绩统计分析表》实例	(122)
附表 6 《评价试题、试卷和考试质量统计分析表》实例	(124)
跋 在教育科学中准确运用数学知识	(129)

第一章 考试分数的特征量与分布

考试是对学习结果的检查。考试分数是评定学习成果所记的成绩,就学校课程考试来说,它是试卷中所有试题上所获得的成绩的总和。当考试实施后,参加考试的学生都会获得考试分数。对所有这些考试分数进行整理与分析,具有重要的意义。考试分数是评定学生学习成绩的重要依据,也是科学地评价教学质量和改进教学工作的重要依据。

本章,我们要研究考试分数的抽样方法,考试分数的特征量及直方图与正态分布。

第一节 抽样方法与原始分数的登记

在统计学中,总体(population)是指研究对象的某一项数量指标的全部可能的观察值,个体(individual)是指每一个观察值。

例如,研究一所学校某年级某门课程的教学质量,则该年级该门课程的所有学生的考试分数就是一个总体,而每个学生的考试分数则是一个个体。

从总体中抽出若干个体的集合,叫做该总体的一个样本(sample),而样本中所包含的个体的个数称为样本的容量(size of a sample)。样本是总体的一部分;我们通过这一部分(局部)去认识总体(整体)。一般来说,样本的容量越大,样本的性质就越接近总体性质,至于样本的容量大到什么程度为好,应根据具体问题来定。

抽样(sampling)是从总体中抽取一部分个体的过程,抽取的那一部分个体就是样本。抽样的目的在于从样本性质估计总体性质,要保证这种估计的正确性,首先必须保证所抽定的样本有较好的总体的代表性。样本的代表性与样本的容量有关外,还与抽样方法有非常重要的关系。

在实际工作中,常用的抽样方法有:随机抽样、等距抽样、分层抽样、整体抽样和分阶段抽样等。

一、随机抽样

1. 随机抽样的定义

随机抽样(random sampling)就是从总体中抽取样本时,总体中的每一个个体都是等可能抽到的。

2. 随机抽样的操作方法

采用抽签的方式就是属于随机抽样。

例如,要从650名学生中,随机抽取容量为50人的样本,其步骤是:

(1) 将全体学生编号,每人都有一个号码。这样,则有1~650号。

(2) 把每一个号码写在大小相同的小纸卡片上,并放入一个小箱子里。

(3) 把箱子里的卡片搅匀,然后用手每次抽取一张卡片,并记录下卡片上的号码,然后再放回箱子中,如果抽取相同的号码,也放回箱中,不予记录,再继续抽取,直至抽

到 50 个不同号码为止, 就为该总体的样本。

二、等距抽样

1. 等距抽样的定义

等距抽样 (equidistant sampling) 是把总体中各个个体按一定顺序排列后, 按相等的间隔来抽取个体构成随机样本。

2. 等距抽样的操作方法

我们要在 150 名学生中抽取 30 名为样本, 其步骤是:

(1) 确定等距的距离

$$\text{间距} = \frac{\text{总数}}{\text{样本容量}} = \frac{150}{30} = 5$$

(2) 从间距以下的数目中随机确定抽样的起点号码

本例从 1 至 5 中随机确定一个数为起点号码, 如 5。

(3) 从起点号码开始, 每隔一个间距抽取一人, 达到样本容量为止

本例, 从 5 开始, 依次抽取号码构成集合是:

$$\{5, 10, 15, 20, \dots, 145, 150\}$$

为该总体的样本, 样本容量为 30 人。

三、分层抽样

1. 分层抽样的定义

先按一定标准把总体分成几个部分 (即若干层), 然后从各部分 (即各层) 中按一定的比进行随机抽样, 这种抽样方法称为分层抽样 (stratified sampling)。

2. 分层抽样的操作方法

总体分层的基本原则是: 各层内部的差异要小, 层与层之间的差异要大。否则就会失去分层的意义。

当总体分好层, 且样本容量也已确定之后, 从各层中所抽取人数的比都应等于样本的容量 n 与总数 N 之比。如总体中各层人数分别为 $N_1, N_2, \dots, N_k$, 而从每层中所抽的人数为 $n_1, n_2, \dots, n_k$, 则满足下式:

$$\frac{n}{N} = \frac{n_1}{N_1} = \frac{n_2}{N_2} = \dots = \frac{n_k}{N_k} \quad (1-1)$$

例如, 某校有四个年级 1600 个学生, 一年级有 640 人, 二年级有 480 人, 三年级有 320 人, 四年级有 160 人, 拟抽取 80 个学生作为样本。

$$\text{按年级抽取, 各年级抽取} \quad \frac{n}{N} = \frac{80}{1600} = \frac{1}{20}$$

$$\text{从一年级中抽取} \quad 640 \times \frac{1}{20} = 32 \text{ (人)}$$

$$\text{从二年级中抽取} \quad 480 \times \frac{1}{20} = 24 \text{ (人)}$$

$$\text{从三年级中抽取} \quad 320 \times \frac{1}{20} = 16 \text{ (人)}$$

$$\text{从四年级中抽取} \quad 160 \times \frac{1}{20} = 8 \text{ (人)}$$

组成的集合： $\{32 \text{ 人}, 24 \text{ 人}, 16 \text{ 人}, 8 \text{ 人}\}$ 为一个随机样本。

分层抽样应注意两个问题：一是确定分层的标准，即根据什么原则进行分层；二是各层之间要避免重叠现象。

四、整体抽样

1. 整体抽样的定义

从总体中抽出来的研究对象，不是以个体为单位，而是以整体为单位的抽样方法，称为整体抽样（integral sampling）。

2. 整体抽样的操作方法

为了增强样本对总体的代表性，弥补整体抽样的不均匀性，整体抽样可以与分层抽样相结合使用。先按一定标准以整体为单位，把总体分成几部分，然后从每部分中再按一定的比抽取若干整体，组成整体样本。

例如，为了评价某年级某学科考试质量，以班为单位进行抽样。按考试平均分数把全年级所有班排队，分成好、中、差三个部分，然后从每部分抽出一个班，三个班的试卷放在一起进行整体分析，这就是整体抽样。

五、分阶段抽样

1. 分阶段抽样的定义

分阶段抽样（stage-by-stage sampling）就是采取两阶段或多阶段抽样。

2. 分阶段抽样的操作方法

例如，为了分析统考英语课程考试质量，计划抽取容量为 50 的样本，分为两阶段进行：先从五位任课教师的任课班级中抽取一个班，再从五个班中各抽取 10 份试卷，为该总体的样本。这就是分阶段抽样。

上述五种常用抽样方法，各有特点，各种方法可以单独使用，也可以互相结合使用。

六、原始分数的登记

将抽取学生的原始分数，如实地登记在“考试分数登记表”内，如表 1-1 所示，并录入计算机。

表 1-1

考试分数登记表

考试课程_____ 抽样班级_____ 抽样人数_____

试题序号	1	2	...	j	...	m	总分
试题得分	a_1	a_2	...	a_j	...	a_m	b
抽样序号							
1	a_{11}	a_{12}	...	a_{1j}	...	a_{1m}	x_1
2	a_{21}	a_{22}	...	a_{2j}	...	a_{2m}	x_2
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$

续表

试题序号	1	2	...	j	...	m	总分
试题得分	a_1	a_2	...	a_j	...	a_m	b
抽样序号							
i	a_{i1}	a_{i2}	...	a_{ij}	...	a_{im}	x_i
⋮	⋮	⋮		⋮		⋮	⋮
⋮	⋮	⋮		⋮		⋮	⋮
n	a_{n1}	a_{n2}	...	a_{nj}	...	a_{nm}	x_n

表中 a_j 表示第 j 题的满分值, b 表示试卷满分值, a_{ij} 表示第 i 个考生在第 j 题上的得分, x_i 表示第 i 个考生的试卷得分。

第二节 考试分数的特征量

为了通过考试分数获得有用信息, 分析出考试质量, 需要通过考试分数求得一些特征量, 以此反映出考试分数的集中趋势、离散程度和相关程度等各项特点。

特征量 (characteristic measure) 是反映一组考试分数分布特征的具体量化指标, 是描述考试分数整体水平特征的量, 是对考试分数的整理和概括。

有反映一组考试分数集中趋势或代表一组考试分数水平的特征量, 叫做集中量。集中量反映整个考试分数分布中多数考试分数的集结点的位置。常用集中量有三种: 平均分、中位数和众数。

有反映一组考试分数彼此之间差异、分散程度的特征量, 叫做差异量。差异量有方差、标准差、差异系数等。

有反映两组考试分数之间相关程度的特征量, 叫做相关量。相关量有相关系数。

一、平均分

1. 平均分的定义

平均分 (average) 是所有考试分数总和除以分数总个数后所得之商, 简称为平均分, 用 $\bar{x}$ 表示。

2. 平均分的计算公式

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i$$

(1-2)

在这里 $\bar{x}$ 表示平均分

n 表示考生的人数

$x_1, x_2, \cdots, x_n$ 分别表示 n 个考试分数

x_i 表示第 i 个考试分数, $i=1, 2, \cdots, n$

3. 平均分的作用

平均分是反映考试分数整体水平的代表值, 它代表全体考生考试成绩的平均水平,

它是对考试分数集中趋势的定量描述。

例如,要想比较两个平行班某门课程的考试分数,不能将两个班每个学生的分数一一列举出进行比较。因为每个学生的考试分数由于受多种因素的影响,大多是不相同的,用个别学生的分数进行比较是得不出什么结果的,如果将两个班的平均分数加以比较,就会既经济又简明。

平均分数高的班,考试成绩水平高;平均分数低的班,考试成绩水平低。

平均分数的缺点是易受极端考试分数的影响。比如,比较两个班考试成绩,甲班有2~4个成绩突出的学生,而乙班有2~4个成绩偏差的学生,由于受极端分数的影响,使两个班的平均分数拉开,甲班平均分是82分,乙班平均分是78分。在这种情况下,平均分受几个最高分或最低分的极端值的影响,不能真实地反映两个班的整体水平,即大多数考生整体水平。

4. 加权平均分数

(1) 加权平均分数 (weighted mean) 的定义

用 $W_1, W_2, \dots, W_n$ 分别表示考试分数 $x_1, x_2, \dots, x_n$ 的权重,那么,这组考试分数的加权和及加权平均分数分别是:

$$\text{加权和} \quad W_1x_1 + W_2x_2 + \dots + W_nx_n = \sum_{i=1}^n W_ix_i \quad (1-3)$$

$$\text{加权平均分数} \quad \bar{x}_w = \frac{\sum_{i=1}^n W_ix_i}{\sum_{i=1}^n W_i} \quad (1-4)$$

在这里 $\bar{x}_w$ 表示加权平均分数

n 表示一组考试分数的总个数

x_i 表示第 i 个考试分数, $i=1, 2, \dots, n$

W_i 表示考试分数 x_i 的权重, $i=1, 2, \dots, n$

例如,学生某学科期末成绩由两部分组成,即平时成绩 x_1 和期末考试成绩 x_2 。假定学校规定这两部分成绩一律按百分制考评,同时两部分成绩的权重分别是30%和70%,那么该学生某学科期末成绩综合考评公式是:

$$\bar{x}_w = \frac{0.3x_1 + 0.7x_2}{0.3 + 0.7} = 0.3x_1 + 0.7x_2$$

(2) 多班考试分数平均分数的合成

某学科考试甲班35名学生的平均分数为82分,乙班37名学生的平均分数为86分,两班全体同学某学科考试平均分数可以利用加权平均分数公式来计算:

$$\bar{x} = \frac{35 \times 82 + 37 \times 86}{35 + 37} = \frac{6052}{72} = 84.06$$

注意:本书将数学中的分数形式化为小数形式,小数点后取两位数,为实用方便,用等于符号表示。

若一总体考试分数由 k 个班组成,这总体考试分数的平均分数可由各个班的平均分数而求得。

设1班, $\dots$, i 班, $\dots$, k 班的平均分数分别为:

$$\bar{x}_1 = \frac{1}{n_1} \sum_{j=1}^{n_1} a_{1j}, \quad \dots, \quad \bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} a_{ij}, \quad \dots, \quad \bar{x}_k = \frac{1}{n_k} \sum_{j=1}^{n_k} a_{kj}$$

在这里 $\bar{x}_i$ 表示第 i 班的平均分数, n_i 表示第 i 班的人数, a_{ij} 表示第 i 班第 j 个学生的分数。

有下式成立:

$$\begin{aligned} \sum_{j=1}^{n_1} a_{1j} + \cdots + \sum_{j=1}^{n_i} a_{ij} + \cdots + \sum_{j=1}^{n_k} a_{kj} \\ = n_1 \bar{x}_1 + \cdots + n_i \bar{x}_i + \cdots + n_k \bar{x}_k \end{aligned}$$

又设 $\bar{x}$ 表示 k 个班所有学生的平均分数, 则有多班平均分数的合成计算公式:

$$\begin{aligned} \bar{x} &= \frac{1}{\sum_{i=1}^k n_i} \left(\sum_{j=1}^{n_1} a_{1j} + \cdots + \sum_{j=1}^{n_i} a_{ij} + \cdots + \sum_{j=1}^{n_k} a_{kj} \right) \\ &= \frac{1}{\sum_{i=1}^k n_i} (n_1 \bar{x}_1 + \cdots + n_i \bar{x}_i + \cdots + n_k \bar{x}_k) \end{aligned}$$

即

$$\bar{x} = \frac{\sum_{i=1}^k n_i \bar{x}_i}{\sum_{i=1}^k n_i} \quad (1-5)$$

在这里 $\bar{x}$ 表示多班的平均分数

n_i 表示第 i 班的实考人数, $i=1, 2, \cdots, k$

$\bar{x}_i$ 表示第 i 班的平均分数, $i=1, 2, \cdots, k$

k 表示班的个数

二、中位数

1. 中位数的定义

中位数 (median) 是指一组考试分数依递增或递减次序排列, 位于正中间位置上的那个分数, 用 M_d 表示。

2. 中位数的计算方法

将全部考试分数从高到低 (或从低到高) 顺序进行排列, 若总人数为奇数, 位于正中间的那个分数就是中位数; 若总人数为偶数, 就取最中间的两个分数的平均值作为中位数。

例 1 7 名学生某学科考试成绩分别为 80, 72, 69, 90, 92, 83, 76, 求中位数。

解 先把 7 个分数从大到小排序为:

$$92, 90, 83, 80, 76, 72, 69$$

显然正中间的 80 为中位数, 即 $M_d = 80$ 。

例 2 8 名学生某学科考试成绩分别为 58, 75, 62, 83, 96, 70, 90, 87, 求中位数。

解 先把 8 个分数从大到小排序:

$$96, 90, 87, 83, 75, 70, 62, 58$$

则位于正中间的两个分数是 83 与 75。

因而这时中位数为: $M_d = \frac{83+75}{2} = 79$

3. 中位数的应用

中位数也是反映考试分数集中趋势的量，也是反映考试分数整体水平的数值。

中位数的计算不是由每个分数都参加运算求得，而是由中间位置相邻的部分分数求得，故中位数一般不受两极端分数的影响。当一组考试分数中出现特大或特小两极端分数时，可用中位数代表整体水平。

由于中位数仅利用了考试分数中相对位置的信息，故在充分利用考试分数整体的信息上，中位数不如平均分数。所以在一般情况下，中位数的集中代表性不如平均分数，中位数的应用也不如平均分数广泛。

三、众数

1. 众数的定义

众数 (mode) 是指一组考试分数中出现次数最多的那个分数，用 M_o 表示。

2. 众数的计算方法

观察法：由定义，一组分数中出现次数最多的那个分数为众数。

例如，考试分数分布为：69, 72, 72, 85, 85, 85, 92, 92，观察发现，出现次数最多的那个分数是 85，出现 3 次，于是 85 就是这组分数的众数，即 $M_o = 85$ 。

3. 众数的应用

众数与平均分数一样，同样也是表示一组考试分数集中趋势的量，指出一组分数有代表性的分数。众数的概念简单明了，容易理解，也不受极端分数的影响。

有时会遇到在一组分数中，出现次数最多的分数的数值不止一个，这组分数中的众数就不止一个。如果一组分数中，每个分数出现一次或每个分数出现次数相同，这组分数就没有众数。

众数仅利用了考试分数出现次数最多的信息，因而在一般情况下，反映考试分数集中代表性上不如平均分数和中位数。

四、方差与标准差

1. 方差与标准差的定义

一组考试分数的方差 (variance)，指的是各个分数与它们平均分数之差的平方的算术平均数。其定义用公式表示为：

$$S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} \quad (1-6)$$

在这里 S^2 表示方差

n 表示考生人数

x_i 表示第 i 个考试分数， $i = 1, 2, \dots, n$

$\bar{x}$ 表示平均分数

标准差 (standard deviation) 是指方差的算术平方根。其定义用公式表示为：

$$S = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}} \quad (1-7)$$

在这里 S 表示标准差

n 表示考生人数

x_i 表示第 i 个考试分数, $i=1, 2, \dots, n$

$\bar{x}$ 表示平均分数

注意: 本书表示标准差的符号采用了传统与多数教育统计学教材中使用的符号, 用 S 表示标准差, 未使用有些教材用希腊字母 σ 来表示标准差; 另外, 在理工类概率论与数理统计教材中, 有样本标准差概念, 也用符号 S 表示, 它的计算公式是 $\sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$, 与

本书公式 (1-7) 是不相同的, 本书称它为修正标准差, 用 S^* 来表示。在 n 很大时, 本书公式 (1-7) 表示的标准差 S 与修正标准差 S^* 相比较, 两者数值非常接近, 都可以用来估计总体标准差 σ 。读者应分清标准差 S 、修正标准差 S^* 及总体标准差 σ 的区别与联系, 及在各种教材中不同的名称, 其所指的是哪一个。

2. 直接用原始分数计算方差与标准差

上述两个公式可以化为直接用原始分数计算方差和标准差的公式, 具有简便、准确的特点。其计算公式分别为:

$$\text{方差 } S^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i \right)^2}{n^2} \quad (1-8)$$

$$\text{标准差 } S = \frac{\sqrt{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2}}{n} \quad (1-9)$$

在这里 n 表示考生人数

x_i 表示第 i 个原始分数, $i=1, 2, \dots, n$

3. 标准差的应用

标准差是反映一组分数离散程度的重要指标。它给出一组分数偏离平均分数程度的大小, 标准差越小, 表明考试分数偏离平均分数离散程度小, 即考试分数集中、整齐, 分布范围小; 标准差大, 表明考试分数偏离平均分数离散程度大, 即考试分数参差不齐, 分布范围广。

究竟标准差大好还是小好, 这需要看所分析的具体问题。

例如, 某班某学科考试成绩标准差极大, 表明该班学生学习水平差异程度大, 这对教师的教学极其不利, 而且该科的平均分数也失去了意义。某一试题的标准差为零或极小, 表明该题对学习好和学习差的学生没有区分的能力, 是一个质量较差的试题。

通常用平均分和标准差来表示考生整体水平, 构成一对从集中到分散两个角度来描述一组考试分数特征的统计指标。整体水平平均分高, 说明考生整体水平高。标准差小, 说明考生成绩间的差别小。因此, 平均分高, 而标准差小的考生整体为真正成绩高。

4. 多班考试分数方差的合成

某一总体考试分数由 k 个班组成, 这总体考试分数的方差可以由各个班的方差而求得。

多班考试分数方差的合成:

$$S^2 = \frac{\sum_{i=1}^k n_i [S_i^2 + (\bar{x}_i - \bar{x})^2]}{\sum_{i=1}^k n_i} \quad (1-10)$$

在这里 S^2 表示多班的方差

S_i^2 表示第 i 个班的方差, $i=1, 2, \dots, k$

$\bar{x}$ 表示多班的平均分数

$\bar{x}_i$ 表示第 i 个班的平均分数, $i=1, 2, \dots, k$

k 表示班的个数

n_i 表示第 i 个班的考生人数, $i=1, 2, \dots, k$

多班考试分数标准差的合成:

$$S = \sqrt{\frac{\sum_{i=1}^k n_i [S_i^2 + (\bar{x}_i - \bar{x})^2]}{\sum_{i=1}^k n_i}} \quad (1-11)$$

在这里 S 表示多班的标准差

五、差异系数

学校教学班之间学习水平的比较,可应用平均分数和标准差。如果标准差大体相等,则可以看平均分数的大小,平均分数大的好;如果平均分数大体相等,则可以看标准差的大小,标准差小的好,说明离散程度小,学习水平比较整齐。如果平均分数和标准差均有明显的差异,如何正确比较考试分数离散程度大小呢?可用差异系数来比较。

差异系数 (coefficient of variation) 是指标准差占其平均分数的百分之几。其计算公式为:

$$CV = \frac{S}{\bar{x}} \times 100\% \quad (1-12)$$

在这里 CV 表示差异系数

S 表示标准差

$\bar{x}$ 表示平均分数

由上式可见,差异系数就是以平均分数为单位,用标准差占平均分数的百分之几,来衡量考试分数离散的程度,它是没有单位的相对差异量。差异系数越大,表明分数离散程度越大;差异系数越小,表明分数离散程度越小。

经验表明:在正常情况下,差异系数的值常在 5% ~ 35% 之间;如果差异系数的值大于 35% 时,可怀疑所求得平均分数是否失去了意义;如果差异系数的值小于 5% 时,可怀疑平均分数和标准差是否计算有误。

通常会出现:一组分数的平均数较大,其标准差也较大;平均数较小,其标准差也较小。因此,比较平均数相差较大的两组分数的离散程度时,若直接用标准差比较可能会得出不科学的结论。请看下面例题。

例 3 某门课程进行考试,一班的平均成绩为 82.41,标准差为 14.86;二班的平均成绩为 72.96,标准差为 13.09,问这两个班的学习整齐程度是否相同?

解 学习程度是否整齐就是指学习成绩离散性大小。这两个班的平均分数相差较大,