

教学研究 方法浅谈

——试卷的统计与分析

张步通 李树莉 编著



四川教育出版社

教学研究 方法浅谈

——试卷的统计与分析

说明：凡在点线之上用反面时并线，断下再粘。

责任编辑：王屏汉

封面设计：何一兵

版面设计：顾求实

教学研究方法浅谈——试卷的统计与分析

四川教育出版社出版
四川省新华书店发行

成铁二小印刷厂排版
成都培风印刷厂印刷

开本787×1092毫米 1/32

印张2.125 字数43千

1987年6月第一版

1987年5月第一次印刷

印数：1—4,100册

书号：ISBN7-5408-0150-61 G143 定价0.43元

7344·889

前 言

现代科学技术的飞速发展，向广大教师提出了新的挑战。不断充实和更新知识，尽快地改变自己的知识结构，以适应“三个面向”的要求，成了摆在每一个教育工作者面前的一项迫切任务。

很多教师与教学研究人员经常碰到的一个课题，是如何对试卷进行科学地统计与分析。这种统计与分析是教学研究的一个重要手段，在标准化考试中尤为如此。四川省教育科学研究所，为适应培训中小学教师的需要，满足广大教师的要求，特编写《教学研究方法的浅谈——试卷的统计与分析》。愿它有助于这一课题的探讨。

本书由四川教育学院张步通、四川教育科学研究所李树莉同志编著，李国新同志审稿。在编写过程中，承蒙唐瑞志、冯国卫等同志提出宝贵意见在此向他们致谢。由于时间仓促，如有不当之处敬请批评指正。

四川省教育科学研究所

1987年2月

AAZ48/04

目 录

一 导 言	(1)
二 试卷抽样	(2)
三 试卷数据整理及计算	(7)
四 绘制考试频数分布统计表和考试频数分布 曲线图	(18)
五 计算试题的难度	(31)
六 计算试题的区分度	(34)
七 计算考试的信度	(37)
八 计算考试的效度	(38)
九 写试卷分析报告	(43)
十 附 表	(44)

一 导 言

认真进行试卷的统计、分析是提高教育质量的有效手段和方法之一，也是广大的教学研究人员和各科教师的基本功，随着外语标准化考试的试点和推广，这一问题已被越来越多的教育工作者所重视。但是如何科学地统计分析试卷，才能使我们既省时、省力，又能获取更多可靠的信息呢，这是一个值得研究的重要课题。在此，愿向同行们谈谈我们在教学实践中对这一课题所作的探索和做法。

辩证唯物主义常识告诉我们，世界上的任何事物，都同时具有质与量两个方面的规定性，既不可能没有质的单纯的量，也不存在没有量的纯粹的质，任何事物都是质与量的辩证的统一。我们分析试卷也应该“质”、“量”并重，就是说既要重视定性分析，又要重视定量分析。教育改革的深入发展，更需要我们在定性分析与定量分析的结合上下功夫。马克思说过：“一切科学只有在成功运用数学时，才算达到了真正完善的地步”。我们要把定性分析与定量分析有机地结合起来，定量分析是对定性分析的深化，而在研究分析试卷反映的各种数量方面的问题时，又必须紧密联系试卷质的规定性，两者的有机结合，才有可能比较全面的反映试卷的质量，获得教学上的各种信息的反馈。

一般认为如试卷具有较高信度、效度，题目的难度、区分度又适当，这种测验所提供的信息，才是有效的、可信的。所以，在对试卷进行统计分析和必要的技术处理的时候，

首先应对试卷的效度、信度，题目的区分度和难度进行考察。那么，我们如何才能求得效度、信度和区分度呢？如何运用教育统计学的有关知识来统计分析试卷呢？下面我们将按照试卷统计分析的一般要求，从试卷的抽样、数据整理、绘制考试曲线、计算试卷的各种指标、写试卷分析报告等几方面的内容进行比较具体的分析，努力在理论与实践的结合上，谈谈试卷统计分析的一般方法和步骤。

二 试 卷 抽 样

什么叫试卷抽样？从参加考试的全部学生的试卷中，抽取一部分学生的试卷作为试卷统计分析的对象，这个过程就叫试卷抽样。

这里所指的“全部试卷”在统计学上也可称为“总体”。总体就是指具有同一性质的所有个体而言。“部分试卷”可看作是“样本”。样本就是指从总体中抽取出来的一部份性质相同的个体。

如果总体较小，如只对一个小组，一个小班的试卷进行分析，就不需要抽样，可以直接把全组或全班学生的试卷作为统计分析的对象。

如果总体很大，譬如全省的会考、高等学校的统一招生考试，考生人数往往多达十万以上，要以全体考生的试卷作为统计分析对象，工作量大得惊人，没有计算机等先进的科学设备，要在短短几天之内完成统计分析任务，人力、财力均不允许，事实上也无此必要。在实践中，往往采取在全体考生中

随机地抽取一小部份试卷——样本，作为统计分析的对象，然后运用有关的数学知识，推断出全体考生的大致情况。为什么随机抽取少量考生的试卷，就可以推断出全体考生的试卷情况呢？统计学的知识告诉我们：用样本平均数，估计总体平均数是无偏估计。在所有可能的样本配合的平均数中，接近总体平均数的，总是占大多数。如果总体很大，样本数又大于50，其平均数也总是成正态分布，或趋近于正态分布。因为正态分布是以平均数为中心的对称分布，越接近平均数的变量（变量在这里可理解为考生的成绩），被抽取的可能性就越大，越远离平均数的变量，抽样抽到的可能性就越小。所以用抽样的方法来进行统计分析是一种科学的方法。

为了使样本能代表总体，抽样是个关键问题，如果总体有几万份试卷，乃至十几万，甚至更大的数字，我们又如何来进行抽样呢？下面就来谈谈抽样的原则、类型和样本容量的确定等几个方面的问题。

1. 抽样需要遵循的原则

(1) 随机性原则

抽样时要排除人们主观因素的影响，不能按照自己的主观意志进行挑选，可以按中国科学院数学研究所概率统计室编的《常用数理统计表》中的“随机数表”进行随机抽样（见附表一）。此表是按照随机抽样的原则编制的，每两个数字为一小组，每25个小组为一个区组（即一个小方块）。每一个区组中的每一个数字可以当作任何位数（二位数字、三位数字、四位数字……）来使用。进行随机抽样时，可根据需要随意从任何一个区组、任何一个数目字开始，依次向下（或向上、向左、向右）观察随机数表中的数字。具体使用

方法见附表一后面的例题。

(2) 可行性原则

采用什么样的抽样方法，要根据当时的实际情况决定。如按照全体考生在各个分数段的比例进行分层随机抽样，是比较理想的抽样方法。因为用这种方法抽得的样本最典型、最具有代表性，但采用此种方法有一个前提条件，就是必须在作好总体分数段的频数（频数即次数）统计后，才能进行抽样。如果考生数量大、又没有现代化设备和统计技术人才，采用分层随机抽样，就不符合可行性原则。相反，如设备、人才条件具备，时间又充裕，就可采用这种方法抽样。

(3) 信息性原则

努力使抽得的样本，尽可能的反映出各种不同的信息，如优、中、差生的不同情况等。

2. 抽样的类型

抽样的类型有多种，现介绍四种：

(1) 简单随机抽样

简单随机抽样即纯随机抽样。从总体中，即从全体考生的试卷中，不需要进行分组、排队，完全按随机原则抽取样本（试卷）的方法，一般常采用现成的随机数表查出所要抽取的单位。在实际工作中，一般多是运用抽签的方法，将组成总体的每一个体逐一编号，然后用手工或机械的方法从中随机抽取。这种抽样方法，从理论上讲能确保每一份试卷都有同等被抽中的机会，但是在实际应用上有极大的局限性，如总体极大，要对每一个体进行编号就比较困难。所以在实践中较少运用。

(2) 类型抽样

类型抽样亦称分层抽样。它是将试卷按一定标志（如地区或学校类别）划分为若干个类型或层次，然后根据随机原则，在各个类型或层次中抽样的方法。类型抽样既提高了所抽样本的代表性，又遵守了随机原则，与纯随机抽样一样，每份试卷都有同等被抽取的机会，尤其是在全距（即最高分与最低分之差）较大，而试卷数量又多的情况下，采用类型抽样比较好。

（3）等距抽样

等距抽样也称机械抽样，就是将全部试卷按一定标志进行排列，然后按固定顺序和相等的间隔距离抽取样本的方法。用于排列次序的标志，一般选择与试卷分析项目、内容无关的标志，如按编号或按地名排列等等。等距离抽样的间隔距离的确定，原则上由全部试卷数量除以抽样数得出。等距抽样所抽得的样本，在总体中（全部试卷中）分布是均匀的，因此抽得的样本有较高的代表性。

（4）整群抽样

在总体中根据随机原则，成群成批地抽样的方法叫整群抽样。各群（批）的抽取，可采用纯随机抽样，等距抽样或类型抽样等形式。各群或各批样本被抽出以后，要对样本中间的所有个体进行统计、分析。这种方法抽样，样本比较集中，但分布不易均匀。其优点是组织较简单。

以上介绍了四种抽样方法。在实际工作中往往把各种抽样方法结合起来运用。如1986年政治学科高考试卷抽样就是把类型抽样与等距抽样结合起来进行的。先按编号的类型随机抽取试卷袋，而每袋中的25份试卷，则按固定的间隔距离抽取样本的方法来进行的。

3. 样本容量的确定

样本容量是指从总体中抽出来的一部份性质相同的个体的数量。

样本容量的确定有两种方法：

(1) 通过计算来确定

按统计学的要求，样本容量上千的总体，一般可用总体的 $1/20 \sim 1/30$ 进行取样。

(2) 根据具体情况确定

当总体很大又无计算机时，样本可取370。为什么确定370这个数字呢？主要是为了计算上的方便。因为进行试卷统计分析时，往往要计算高分组和低分组的难度值，而高分组或低分组的难度值的计算，是以占样本总量27%的高分考生和27%的低分考生的成绩作为计算范围的。而370这个数的27%，是99.9，四舍五入，正好是100，便于计算百分比。如果排除上述原因，样本容量的大小，可由自己根据当时情况确定。

从理论上分析，样本容量的确定，受到抽样误差、考生分数的差异程度、全距的大小及时间、人力等各种条件的制约。样本愈大，愈接近总体，抽样误差愈小，但工作量愈大；样本愈小，抽样误差愈大，但工作量愈小。在这里可明显看出样本的多少与抽样误差成反比，而与工作量成正比。样本容量的确定与考生全距的大小和差异程度也有关，若全距大，差异程度大，样本的容量应大一点，反之，则可少抽一点。明确了它们之间的辩证关系，我们就可根据当时的具体情况，如人力、时间、财力和考生的情况来确定样本的容量。

三 试卷数据整理及计算

通过抽样，我们获得了部分考生的原始分数，这些分数往往是杂乱无章的，因此，需要对这些原始分数按照分数段，从高分到低分（当然也可以从低分到高分）进行加工、整理和归类，使这些表面看来是杂乱无章的分数为我所用，提供有用的信息；并通过计算分析，进一步认识这些数字之间的相互关系和内在联系，从而得出比较合乎实际的科学结论。

数据正确与否，关系到试卷的分析质量和科学性，因此必须十分重视数据的可靠性、准确性和完整性。尤其是可靠性和准确性，是进行试卷统计分析的前提条件，关系到试卷分析的可信程度，我们务必持严肃、认真的态度。

数据整理及计算的具体步骤和方法如下：

1. 造表

根据试题大小题目的多少和排列次序，设计一张各题成绩登记表。前面是大题，一题一格，有几大题则设计几格。“总计”后面又按小题依次设计成表格，这样设计便于以后的数据统计。现以1986年高考政治试题为例，设计一份成绩登记表。（1986年高考政治共有九个大题，其中第一至第五大题各有若干小题，第六至九大题无小题，共有大小试题四十个）

参

2.逐一登记抽样样本中每一考生各大、小题所得的原始分数，每人一张。然后根据总分得分情况，按高分到低分排列好，待用。

3.计算样本平均分 (Mean Score)

平均分是试卷统计分析中的一个重要的数据。它能代表考生的平均水平，但它只能说明共性，不能说明个性。一般用符号“ $\bar{X}$ ”代表，“ $\bar{X}$ ”读作X拔(平均分也可用“M”代替)。

平均分的求法：

将样本中的每一个考生的原始分数相加，除以样本总数，即可得到平均分。计算方法有多种，现介绍实用价值最大，人们又经常使用的两种方法：

(1)用各个考生的原始分直接求样本平均分

计算公式如下：

$$\bar{X} = \frac{X_1 + X_2 + X_3 + \cdots + X_n}{N} = \frac{\sum X_i}{N} \quad (1)$$

式中：

$\bar{X}$ 为样本平均分

X_1 、 X_2 、 X_3 、…… X_n 分别为第一个、第二个、第三个、第 n 个考生的原始成绩。

N 为样本总数

Σ 为连加符号 (Σ 读作“西格玛”)

若不是计算样本平均分，而是计算总体平均分，则要把“ $\bar{X}$ ”换成“ μ ” (“ μ ”读作“缪”)

(2) 用简捷法求平均分

规模较大的考试，考生多，取的样本也大，若用公式(1)的方法求取，太繁琐了，计算过程中也容易出差错。为

便于计算，实际工作中，可将考生的原始成绩按分数段进行分组整理，采用5或10为组距（所谓组距即每组最高分数与最低分数之间的距离），把各组的组中点值看成是各组的平均分。这样计算起来比较简单、方便，也省时、省力。但由于用组中点值代表各组平均分，计算出来的得数是一个近似的平均值，不如用公式(1)计算的准确。

为了便于理解，这里需要弄清什么叫组中点，组中点值及精确的上限与下限的确定方法。

所谓组中点 (Class midpoint)，就是一组分数中（即一个分数段中），最高数值，即上限 (Upper Limit) 与最低数值，即下限 (Lower Limit) 相加除以2的得数所在的位置。而组中点值的确定，是将一个分数段的精确的上、下组限相加除以2。

计算精确的上、下限的公式是：

精确的上、下限 = 组中点值 $\pm 0.5 \times$ 组距

如果已知组中点值为 62，组距为 5

则精确的上限 = $62 + 0.5 \times 5 = 64.5$

精确的下限 = $62 - 0.5 \times 5 = 59.5$

用简捷法计算平均分 $\bar{X}$ 的公式如下：

$$\bar{X} = A + \frac{\sum fd}{N} \quad (i) \quad (2)$$

式中：

A 为假定平均数

f 为人数 (或次数)

d 为假定离差

i 为组距

N 为样本总数

利用公式(2) 求平均数, 首先要确定假定平均数 “A” 所在的组。那么, 假定平均数和假定离差应该如何确定呢? 从理论上讲, 任何一个组的组中点值均可充任, 但在实际运用中, 往往是选用包含累积次数约占样本总次数一半左右的那一组的组中点值为假定平均数。假定平均数所在组的离差 “d” 等于零, 在组距相等的条件下, 从零起, 往上各组的离差分别为 +1、+2、+3、……; 往下各组的离差则分别为 -1、-2、-3、……。现以四川教育学院外语系英语本科某年级41人的原始成绩为例, 介绍一下用简捷法求平均分数的方法。具体数据如下:

80、93、73、78、80、94、91、78、80、80、78、91、91、78、80、83、81、81、57、81、82、86、82、65、82、83、83、73、83、91、75、84、84、85、85、90、91、75、88、85、90、85、88、86、87、88、88、80、89。

表 二

分数分组	组中点值	次数	假定离差	次数×假定离差
	X_c	f	d	$f \times d$
[90~95]	92.5	10	2	20
[85~90)	87.5	11	1	11
[80~85)	82.5	17	0	0
[75~80)	77.5	6	-1	-6
[70~75)	72.5	4	-2	-8
[65~70)	67.5	1	-3	-3

续 表

[60~65)	62	1	-4	-4
[55~60)	57	1	-5	-5
Σ		49		11

将表二中各种有关的数据代入公式(2) 得出

$$\begin{aligned}\mu &= A + \frac{\Sigma fd}{N} \quad (i) = 82 + \frac{11}{49} \times 5 = 82 + \frac{55}{49} \\ &= 82 + 1.12 = 83.12 \text{分}\end{aligned}$$

(注: 用简捷法求平均分, 由于用各组的组中点值代表该组的每一个平均分数, 这就要求组成各组的分数都集中在组中点值上, 但实际上, 不可能做到这一点, 因此, 不可避免地会产生一些误差, 误差的大小与组距大小成正比。我们用CASIO fx-180p 计算器计算, 精确的平均分应为82.49分)

4. 计算标准差 (Standard deviation)

符号S为样本标准差, σ 为总体标准差, “ σ ”是希腊字母“ Σ ”的小写, 仍读作西格玛。

标准差是试卷分析中的又一个最重要的数据之一, 它是反映研究对象(如一个班、一个年级、一个市、省的考生分数)的差异程度的指标量, 用来说明考生分数的波动情况。它是以平均分为依据, 求得的差异量。平均分只能反映考生的平均水平, 但它把考生之间的分数差异掩盖了起来, 反映不出这个平均分的代表性究竟如何? 因此要反映出考生的全貌, 除了要计算平均分之外, 还必须把考生的分数差异反映出来, 而标准差就能反映考生分数的差异程度。标准差越大, 平均分的代表性就越小, 标准差越小, 平均分的代表性