

《信息科学基础》试卷质量的统计与分析

廉爽, 张妍

辽宁师范大学数学学院, 辽宁 大连

收稿日期: 2024年3月17日; 录用日期: 2024年4月10日; 发布日期: 2024年4月17日

摘要

本文首先详细阐述了试卷质量的统计分析方法, 包括试题难度、区分度、信度、效度等指标的计算方法和解释; 然后对《信息科学基础》试卷进行了应用, 并根据计算结果对本次试卷进行定性分析, 进而指出其优点与不足。本文通过对试卷质量进行统计与分析, 旨在探究更为科学的评价试卷质量的方法, 为提高试卷质量提供参考, 为提升评判试卷质量的科学性提供指引。

关键词

统计分析, 难度, 区分度, 信度

Statistical Analysis of the Quality of *Fundamentals of Information Science* Text Paper

Shuang Lian, Yan Zhang

School of Mathematics, Liaoning Normal University, Dalian Liaoning

Received: Mar. 17th, 2024; accepted: Apr. 10th, 2024; published: Apr. 17th, 2024

Abstract

Firstly, this article elaborates on the statistical analysis methods of test paper quality, including the calculation and interpretation of indicators such as difficulty, discrimination, reliability and validity; then, the paper is applied to “*Fundamentals of Information Science*” test paper; conducts a qualitative analysis of the paper according to the calculation results and then points out its advantages and disadvantages. Through the statistics and analysis of the quality of test papers, this paper aims to explore a more scientific method of evaluating the quality of test papers, so as to provide a reference for improving the quality of test papers and a guide for improving the scientific judging

the quality of test papers.

Keywords

Statistical Analysis, Difficulty, Discrimination, Reliability

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

考试作为评估学生知识和技能水平的一种方式,可以激发学生的学习动力,促使他们主动学习和复习知识,提高学习效果。试卷应该涵盖教学内容中的各个知识点,以便全面评估学生的掌握情况。试卷质量的好坏直接关系到考试结果的公平性。通过对试卷质量的分析和统计,可以帮助教育工作者反思和改进教学策略,提高教学质量和学生的学习效果。

2. 衡量试卷质量的指标

试卷是实现考试功能的工具,高质量的试卷能真实、全面地反映学生的学习情况。文献[1]用灰色关联评价对试卷质量进行综合评价,并在 Excel VBA 中构建了试卷质量分析系统,实现试卷质量的分析与评价;文献[2]采用 Java 语言进行程序编写,用 Excel 作为数据统计工具,从而实现试卷与试题分析;文献[3]以试卷库和问卷调查为基础应用统计分析的方法考察试卷库质量并研究影响教学质量的因素,并以此分析学生的学习状态和学习质量;文献[4]运用经典测量理论与数理统计方法,从宏观、中观和微观三个不同层面对试卷进行定量分析;文献[5]给出了更多的检验考生成绩正态分布的方法,并建立试卷质量分析的数学模型,利用数理统计的方法对试卷质量分析常用的指标进行分析和比较,找出这些指标最合适的定量分析方法。根据教育测量学相关理论[6][7],试卷质量的主要衡量指标有难度、区分度、信度与效度四种,它们的含义、计算方法的简要介绍如下。

2.1. 难度

难度作为衡量试卷难易层次的一项指标,若将试题难度表示为 P ,则对于二分法计算得分的题目,其计算公式为

$$P = \frac{R}{N}$$

其中 R 为通过或答对题目的学生数目, N 为参加考试的全体考生数量。

对于选择题,由于允许对正确答案进行猜测,需要对分数采取矫正措施。设选择题共有 n 种选项,则其难度计算公式为

$$P = \frac{nR - N}{N(n-1)}$$

对于考生人数较多的情况,可以考虑采用极端分组的方法。所谓极端分组法,就是将学生此题的得分从高分到低分依次排出来,且分别取学生分数的前 27% 和后 27%,按顺序依次定为高分组和低分组。在此,设高分组的平均分数为 X_H , 设低分组的平均分数为 X_L , 难度定义为

$$P = \frac{X_H + X_L}{2W}$$

其中 W 为试题的总分值。

对于主观性试题, 则难度的计算公式为

$$P = \frac{\bar{X}}{W}$$

其中 \bar{X} 为学生在该题中所得的平均分, W 为此题的满分值。上式适用于非二分法记分的题目。

试卷的难度是指试卷整体的难易程度, 计算公式为

$$P = \frac{\bar{X}}{W}$$

其中 \bar{X} 为试卷的平均分, W 为试卷的满分值。

试卷的难度应根据考试目的而定, 一般认为课程的结业考试试题难度值应控制在 0.2~0.8, 而试卷的平均难度在 0.5~0.85 为宜。大规模标准化考试难度应控制在 0.4~0.7 之间。

2.2. 区分度

区分度表示了试题或试卷对参加考试学生实力的区分程度, 常用 D 表示。下面介绍两种计算方法。

对于客观性试题, 区分度的计算公式为

$$D = P_H - P_L$$

其中 P_H 为高分组的通过率, P_L 为低分组的通过率。分组方法同上。

对于主观性试题, 区分度的计算公式为

$$D = \frac{X_H - X_L}{N(H - L)}$$

其中 X_H 为高分组的总分, X_L 为低分组的总分, H 为该题的最高分, L 为该题的最低分, N 为参加考试学生总人数的 25%。

试卷的区分度是指参加考试学生水平在整张试卷上的区分程度, 其计算公式为

$$D = \frac{1}{W} \sum_{i=1}^N D_i W_i$$

其中 D_i 为第 i 题的区分度, D 为试卷的区分度。

一般认为区分度在 0.4 以上的试题为优良题, 位于 0.3~0.39 之间的为良好题, 位于 0.2~0.29 之间的为一般题, 0.2 以下的为劣等题。

2.3. 信度

试卷的信度反映了参加考试学生稳定水平的可靠性。目前为止, 使用最普遍的计算信度的方法是克朗巴赫系数[1], 其计算公式为

$$B = \frac{n}{n-1} \left(1 - \frac{\sum_{i=1}^n S_i^2}{S^2} \right)$$

其中

$$S^2 = \frac{1}{n} \sum_{j=1}^n (x_j - \bar{x})^2$$

$$\bar{x} = \frac{1}{n} \sum_{j=1}^n x_j$$

n 为试卷总的考题数目, S_i^2 为第 i 题得分的方差, S^2 为考试总分的方差。 B 的值在 0 与 1 之间, 若 B 值越接近 1, 则说明学生的分数越可信。若 B 值越接近 0, 则说明学生的分数越不可信。通常大规模标准化考试的信度在 0.9 以上为宜, 课程结业考试的信度应该在 0.5~0.9 之间。

2.4. 效度

效度是指测验结果的有效性或正确性。所谓有效性, 是指一种能够正确地测量出它所要测量的特性或功能的程度。测量学中, 效度被定义为有效分数方差与实得分数方差之比, 当随机误差足够小而且不变时, 系统误差方差越小, 则有效分数方差就越大; 即有效分数方差在实得分数方差中所占的比重越大, 则效度越高。事实上, 在教育测量中所用的工具是测验, 由于它只能间接测量, 导致测验的难度大大增加。所以在教育测量中, 必须研究是否测量到了所要测量的东西, 在多大程度上测量到了这些东西。从这个意义上来讲, 效度比信度更重要。通常, 人们把试卷的效度分为内容效度和校标关联效度两种类型。

所谓内容效度就是指测验内容对所要测量的内容的代表性程度。也就是说, 测验的内容范围、材料与所要测量的内容范围、教育目标是否相符合。评价内容效度的方法之一是观测命题是否符合双向细目表的要求。由于内容效度主要是专家和教师判断的结果, 且缺乏数量化指标, 故可能带有一定的主观性。

所谓校标关联效度就是以某一种测验分数与其校标分数之间的相关来表示效度。效标是能显示或反应所欲测量的属性的变量。效度的大小叫效度系数, 用 r_{xy} 表示, 计算公式为

$$r_{xy} = \frac{\sum_{j=1}^n (x_j - \bar{x})(y_j - \bar{y})}{nS_x S_y}$$

其中 n 为考生数, x_j 与 y_j 分别表示第 j 名考生在 x 考卷与 y 考卷中的成绩, $\bar{x}, \bar{y}, S_x, S_y$ 分别表示两次考试的均值与两次考试的标准差。一般考试的效度系数要求达到 0.4 以上。

3. 实例分析

对 2022~2023 学年第一学期《信息科学基础》科目期末考试试卷进行实例分析。该年级参加本次期末考试的人数为 29 人, 试卷共包括五道大题。其中, 第一大题为客观题, 后四道大题为主观题, 五道题分值分布如下: 20、30、16、16、18。若将试卷成绩由高至低进行排列, 最高分为 98, 最低分为 54, 平均分为 86.28, 方差为 103.35。试卷的难度系数计算为 0.8628, 试卷的区分度为 0.5464。若将前 8 人列入高分组, 最后 8 人列为低分组(各占总人数的 27%)。则计算结果如下。

Table 1. Statistical table of difficulty and discrimination of test questions

表 1. 试题的难度与区分度统计表

题号	一	二	三	四	五
分数	20	30	16	16	18
高分组平均分	19	29.75	16	16	15
低分组平均分	13.75	24.875	11.375	12.125	18.75

续表

难度	0.84	0.94	0.89	0.90	0.70
区分度	0.50	0.51	0.59	0.49	0.67

由表 1 可以看出, 这 5 个题均是较低难度题目, 各题的区分度均良好。

Table 2. Statistical table of variance for test questions

表 2. 试题的方差统计表

题号	一	二	三	四	五
分数	20	30	16	16	18
试题的方差	10.40	10.93	9.67	6.68	5.15

由表 2 可以算出此次考试的信度为

$$B = \frac{n}{n-1} \left(1 - \frac{\sum_{j=1}^n S_j^2}{S^2} \right) = 0.59$$

从试卷上可以看出, 试卷中的题型设计多样, 包括选择题、填空题、解答题等, 这样的设计能够全面考察学生的不同能力, 并培养他们的解题技巧和思维方式。从基础概念到应用都有涉及, 能够全面考察学生对知识的掌握程度。各章内容所占比例分别为: 第一章(绪论)、第七章(保真度准则下的信源编码) 0%, 第二章(离散信源及其信息测度) 28%, 第三章(离散信道及其信道容量) 26%, 第四章(波形信源和波形信道) 16%, 第五章(无失真信源编码定理)、第六章(有噪信道编码定理)、第八章(无失真的信源编码) 各 10%。由此可以看出, 试题基本上覆盖了教学计划的主要内容, 其中第二章和第三章为试卷主体部分。试卷考核内容与大纲中课程目标相对应, 对基本知识的掌握占 50%, 对信息科学各种应用技术与方法的掌握占 60%, 熟悉国内外信息理论发展趋势, 了解本学科理论前沿知识占 42%。但因为没有找到可靠的效标, 所以此次考试的效度系数未能计算得到。

本文使用衡量试卷质量的四项主要指标对《信息科学基础》课程期末考试状况进行了严谨的分析, 研究结果显示试卷难度系数为 0.8628, 超出试卷难度标准 0.5~0.85, 这表明本次考试试卷难度总体偏低, 使得学生更容易取得较为不错的成绩; 试卷区分度为 0.5464 > 0.4, 说明本次试卷题目为优良题, 能够较为完整地体现出学生的不同水平; 试卷信度为 0.59, 说明本次考试学生受偶然因素影响较大, 可靠性较低。基于以上结果, 可判断本次期末考试试卷与高质量试题仍有较大差距, 具体来讲, 本次考试结果虽然能体现出学生的不同水平, 但学生成绩整体偏高, 无法体现学生之间真实差距。此外, 本次试题通过难度、区分度、信度与效度四项指标对试卷进行整体性分析, 仍存在一定的局限性。当今社会, 学生“高分低能”现象屡见不鲜, 在教育教学中, 应注重培养学生创新思维能力, 注重教学知识与现实生活中的具体实际相结合。但本文采取衡量试卷质量的四项指标无法衡量试题的创新程度及与现实生活关联程度。针对以上分析, 在后续的研究中将采用更为精确的统计方法, 并创建一种新的指标——创新度, 以衡量试题创新性。

基于此, 教师和出题人应该认真分析每次考试的试卷质量, 总结经验和教训, 针对试卷的缺点进行改进, 提高试卷设计质量, 进一步提高教学效果。通过定期的试卷分析, 关注学生在各个知识点和题型上的得分情况, 并根据分析结果及时调整教学内容和方式, 以提高学生的学习效果和应试能力。

基金项目

教育部产学研合作协同育人项目(230815093007023)。

参考文献

- [1] 倪海儿, 裘晓华, 魏丹毅. 试卷质量评估与分析系统的构建与实现[J]. 宁波大学学报(理工版), 2016, 29(3): 118-122.
- [2] 徐丽敏, 钱晓耀. 试卷质量分析评价方法与工具的设计[J]. 价值工程. 2018, 37(30): 188-190.
- [3] 努尔古丽·艾力, 张艳, 张瑜. 基于统计分析的教学质量的分析评价——以高等数学教学为例[J]. 南昌教育学院学报, 2012, 27(5): 54-55.
- [4] 司俊峰. 高校课程考试质量评价统计分析模式研究——以《心理统计学》课程考试质量评价为例[J]. 高师理科学刊, 2007(2): 113-117.
- [5] 洪冉. 用数理统计方法分析试卷质量[D]: [硕士学位论文]. 武汉: 华中师范大学, 2012.
- [6] 王孝玲. 教育测量[M]. 上海: 华东师范大学出版社, 1989.
- [7] 刘新平, 刘存侠. 教育统计与测评导论[M]. 北京: 科学出版社, 2003.