

应用数理统计方法分析试卷质量

◆刘蓉¹ 唐婷² 李兴杰³ 杨佳²

(1 北京服装学院基础教学部; 2 北京服装学院材料学院; 3 北京服装学院商学院)

摘要: 本文主要应用数理统计相关知识,通过试卷的难度、效度、区分度、信度和覆盖度等质量指标,来分析试卷。并在研究探讨课程考试试卷质量分析的测评指标及其计算公式的基础上,结合评价方法,对我校2017级高等数学考试试卷质量进行了分析。

关键词: 质量分析; 数理统计; 分析试卷

试卷质量的分析是评价考试质量,检验考试有效性的基本方法。运用数理统计的方法,通过分析学生的成绩,计算此套试卷的难度,效度,区分度,信度,覆盖度,从而推断此套试卷的合理性。

一、试卷质量分析指标

(一)难度

难度是指试卷或试题的难易程度。可以用学生在某一题目的错误率来反应试题的难度,即难度系数($0 < \mu < 1$)。难度系数越大,即错的人数越多,该题难度越大。数学试卷上的选择填空题答案唯一属于客观题, $\mu = \text{答错该题人数} / \text{总人数}$;计算题解法多样,一般按步骤计分属于主观题: $\mu = 1 - \frac{\bar{x}}{x}$ ($\bar{x}$ 指考生在某一考题得分的平均值; x 指该题的满分值)。

一套试卷试题偏难,应该用于选拔考试,难题的比例可以拉开分值。试卷试题偏简单,难以判断学生对知识的掌握情况,也无法评估教学效果。试题的难度直接影响考试成绩的分布形态和离散度。过难,总体考试成绩偏低,呈正偏态分布;过易,总体考试成绩偏高,呈负偏态分布。一般应用于学校测试的高质量试卷的难易程度应该适中,平均难度系数控制在0.5左右,即难、中、易试题的比例2:6:2。

(二)效度

效度是指评价考试结果的有效性、正确性的指标。

我们可以用矩相关系数法计算效度,用2次考试成绩的相关

系数表示, $r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n S_x S_y}$,其中 x_i 、 y_i 分别指第 i 个

考生在 x 、 y 考卷中的分数, $\bar{x}$ 、 $\bar{y}$ 、 S_x 、 S_y 分别指两次考试中的平均值与标准差。一般好的试题要求 r_{xy} 在0.4~0.7之间,

$r_{xy} < 0.4$ 视为无效考试。有时考试不预先进行模考,可使用前几次考试的平均成绩作为参照标准计算 r_{xy} 。

(三)区分度

区分度是区别被测对象对于该测试学科理解掌握程度高低的指标,它反映出被测对象水平的高低。如果对于某道题,水平高的同学得高分,水平低的同学得低分,那么就说明该题目的区分度高,如果所有的同学全得满分或者全不得分,说明该题目区分度低。一般区分度高于0.3,说明该题目质量不错。下面我们用法准确计算。

计算区分度的方法有很多,不同的方法所得结果不同,其中最常用的是极端分组法,即将学生按照第 j 题所得的分数从高到低排列,前27%的学生组成高分组,后27%的学生组成低分组,那么第 j 题的区分度为 $I_j = \frac{A_j - B_j}{N_j}$ 。其中 A_j 表示高分组在第 j 题的总得分, B_j 表示低分组在第 j 题的总得分, N_j 表示第 j 题应得满分乘以单组人数之和。一般区分度 $I_j > 0.4$ 说明区分度高,

题目质量非常好, I_j 在0.2~0.4说明题目质量一般, $I_j < 0.2$ 说明题目区分度非常差,需要修改或淘汰。试卷的区分度

$I = \frac{1}{100} \sum_{j=1}^n I_j \cdot N_j$,其中当 $I > 0.4$,试卷整体区分度为优; $I < 0.2$ 为差。

(四)信度

信度是评价试卷质量的一项重要指标。信度表示的是该试卷测试结果的可靠性,即能否真实地反映应试者的真实水平。当一份试卷多次测验,应试者所答相同时,那么我们说该试卷的可信度高,质量是可靠的,能反应出考生的真实水平。

信度可采用克朗巴赫系数公式计算, $B = \frac{n(1 - \sum_{i=1}^n \frac{S_i^2}{S^2})}{n-1}$ 其中,

S_i^2 表示第 i 题的方差, S^2 表示试卷的总方差, n 表示试卷中的题目个数。因此,信度的最大值是1,最小值是-1。反映信度大小的量是方差,各题的方差小,则试卷的信度高,反之则信度低。一般标准化考试要求 $S > 0.9$ 。

(五)覆盖度

覆盖度是指该试卷的考点与检测周期内知识点之比,一份试卷考察的知识点越广泛,覆盖度则越高,试卷的质量越好。覆盖度是难易度、区分度、置信度、校度作为衡量试卷质量指标的有效前题,覆盖度越高,指标的有效性越好。一份试卷一定要有广泛的覆盖度,对于重要的知识点要面面俱到,这样考察才具有代表性。

覆盖度不能直接计算出来,一般用各题得分之间的相关系数

矩阵 R_{mn} 表示,公式如下: $R_{mn} = \frac{1}{N} \sum_{j=1}^N (x_{mj} - \bar{x}_m)(x_{nj} - \bar{x}_n)$,其中 R_{mn}

越大,说明题目之间的相关性越大,考点越相似,试卷的覆盖度越小,所以 R_{mn} 越小,试卷的覆盖度越广泛,质量越好。一般要求 $R_{mn} < 0.2$,反之说明试卷考察范围太单一,应及时修改试卷内容。

二、试卷指标实例分析

应用上述指标对某班高等数学期末试卷进行分析,该班参加测试人数为59人,试卷共5大题,通过计算结果如下:

表一 试卷运用所给公式计算所得数据

题号	1	2	3	4	5	总分
满分值	16	20	42	10	12	100
平均值	10.17	12.68	28.54	6.85	7.81	66.05
方差	15.21	33.74	136.87	16.37	11.78	495.67
标准差	3.90	5.81	11.70	4.05	3.43	22.26
难度	0.36	0.37	0.32	0.32	0.35	
区分度	0.60	0.71	0.67	0.91	0.67	$I=0.57$

表二 覆盖度系数矩阵

1	0.1357	-0.0131	0.1583	0.0496
0.1357	1	0.1621	0.1304	0.2044
-0.0131	0.1621	1	0.0164	-0.0986
0.1583	0.1304	0.0164	1	0.1958
0.0496	0.2044	-0.0986	0.1958	1

在所选实例试卷的各项数据系数中,①在难度这一栏,各大题难度系数的平均值为 0.344,说明该试卷试题难度偏易,同学们错误率较低,难度系数在 0.3 附近其中第三、四、五大题相对比较容易,第一、二题难度相对较大。②此次所选试卷效度为 0.7542,达到 0.4 以上,说明此试题能够真实有效的反应出学生的学习水平。③从各大题的区分度来看,都大于 0.4,各大题的区分度有明显差距,但总体来看是合理的,这些试题的质量对很好,可以很好的区分出学生的水平④此次试卷的信度为 0.71,信度一般,小于 0.9 没有达到标准试卷的信度要求,信度受到影响的因素有很多,比如考场监考情况、学生考试态度等等。⑤从覆盖度的矩阵中看, $R_{mn} < 0.2$ 的居多说明此份试题的覆盖面大,所涉及到的知识点很全面。

综上所述,该套试卷整体难度偏低,试卷试题偏简单,可以适当提高试卷难度,有利于更好地评估教学效果。其它指标良好,说明该试卷考点分布合理,考查面广,考核较全面。

参考文献:

- [1]盛聚,等.概率论与数理统计[M].北京:高等教育出版社,1989.
- [2]郭述平,王景英.教育测量[M].长春:东北师范大学出版社,1987
- [3]张玲,基于数理统计方法的试题,质量的科学评价与分析[J],理论与实践 2018.(1):23-25
- [4]万金凤.关于试卷质量的分析方法[J].山西师范大学学报:自然科学报,2006,20(1):22-24.
- [5]熊德云,张东军.概率论与数理统计及其应用[M].科学出版社,2005.
- [6]朱德全,宋乃庆.现代教育统计与测评技术[M].重庆:西南师范大学出版社,1998.

北京服装学院 2018 年大学生创新创业训练计划项目,项目编号: NHFZ20180122/003。