

基于 SPSS 的物理试卷分析与评价

陈琪琪

(上海市大同中学 上海 200011)

(收稿日期:2017-03-23)

摘要:借助统计学软件 SPSS,介绍对物理试卷质量定量分析的方法,包括对样本资料绘制成绩分布曲线,进行难度、区分度、信度及效度,即分析试卷质量常用的“四度”指标的定量计算与分析,提高广大教师分析试卷的效率,准确得到分析结果,为今后的教学和考试提供科学依据。

关键词:试卷评价 SPSS 教育统计

考试是教育评价的有效办法与手段,随着 2017 年起,上海市高考改革“3+3”模式的深入开展,对物理教学评价的要求越来越客观,对教学结果的分析也越来越依赖于统计理论和方法。借助统计学软件 SPSS,介绍对物理试卷质量定量分析的方法,包括对样本资料绘制成绩分布曲线,进行难度、区分度、信度及效度,即分析试卷质量常用的“四度”指标的定量计算与分析。希望通过本文,使有意学习试卷定量分析的广大教师,学会对试卷进行科学的分析与评价。本次抽取的考卷分析样本为历年上海某区高考模拟考加试物理的全部 894 位学生。

1 试卷分析

1.1 建立分数段分布直方图(Histogram)

根据录入的数据表,运行菜单:Analyze → Descriptive Statistics → Frequencies,得到频数表。然后得到复选框,导入要建立频数分布表和直方图的项目(即总分),同时运行菜单:Analyze → Descriptive Statistics → Frequencies → Charts → Histograms → With normal curve,即可得到分数段分布直方图,如图 1 所示。通过学生成绩分布直方图和频数分布表可以看出,这 894 位学生的成绩接近正偏态分布,平均分为 111.56 分,标准差为 22.17 分,不及格的人数占 15.5%,100 分~130 分之间的人数占 52.7%,峰值出现在 130 分到 140 分之间的人数占 16.3%。

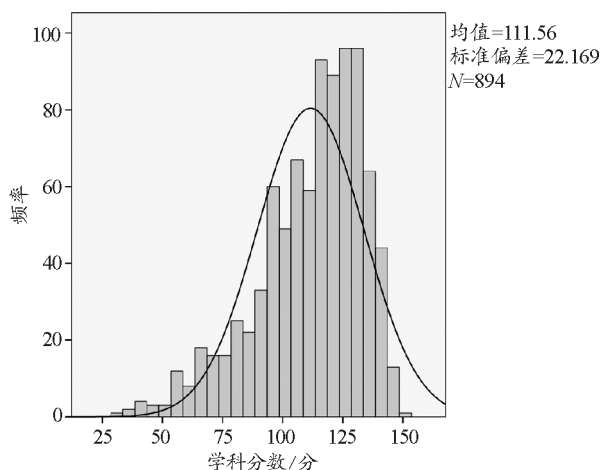


图 1 频数分布直方图

1.2 难度分析(Difficulty)

难度是指试题的难易程度。难度是由考试目标所决定的,设计和控制试题的难度是编制试卷的重要方面,也是试题质量评价的重要指标之一。对于客观性试题,一般用通过率来表示难度的大小。通过率是指被试正确回答或通过题目的人数与所有被试之比。其计算公式为: $P = \frac{R}{N}$,式中 P 为难度指标值, R 为答对某题的人数, N 为考生人数。对于主观性试题,难度的计算公式为: $P = \frac{\bar{X}}{W}$,式中 \bar{X} 为学生在该题得分的平均值, W 为该题的满分值。通过难度的定义可以看出: P 值越大,说明题目越简单; P 值越小,说明题目越难。一般来讲,对于试题的难度测量,可参照表 1 来评价。

表1 试题难度的评价

题目的难度	评价的结果
$p \geq 0.7$	较易题
$0.4 \leq p < 0.7$	中等难度题
$p < 0.4$	较难题

对于一般的常模参照测验,各试题难度值可在 0.50 ± 0.2 之间,最新上海高考等级考难度要求为 0.75.

使用软件进行难度分析的具体操作方法为:运行菜单 Analyze \rightarrow Descriptive Statistics \rightarrow Descriptives,将全部小题分选入 Variable(s) 中,单击 OK 按钮,可得到各题的均值.然后再建立一个包含均值

表2 试卷各题的难度

题号	满分	均值	难度	题号	满分	均值	难度	题号	满分	均值	难度
1	2	1.64	0.82	15	3	1.89	0.63	29	4	1.96	0.49
2	2	1.78	0.89	16	3	2.37	0.79	29(1)	2	1.94	0.97
3	2	0.76	0.38	17	4	3.89	0.97	29(2)	4	3.05	0.76
4	2	1.21	0.61	18	4	3.68	0.92	30(1)	4	3.93	0.98
5	2	1.51	0.76	19	4	3.12	0.78	30(2)	4	3.44	0.86
6	2	1.64	0.82	20	4	2.7	0.68	30(3)	4	3.14	0.79
7	2	1.86	0.93	21	4	3.72	0.93	31(1)	4	2.67	0.67
8	2	1.27	0.64	22	4	2.77	0.69	31(2)	8	6.25	0.78
9	3	2.73	0.91	23	4	2.64	0.66	32(1)	8	6.85	0.86
10	3	2.18	0.73	24	4	2.13	0.53	32(2)	4	3.04	0.76
11	3	2.87	0.96	25	4	2	0.50	33(1)	4	3.18	0.80
12	3	1.91	0.64	26	4	2.18	0.55	33(2)	6	3.64	0.61
13	3	2.56	0.85	27	4	2.37	0.59	33(3)	4	1.37	0.34
14	3	2.81	0.94	28	6	4.9	0.82				

1.3 区分度分析(Discrimination)

区分度是指试卷题目对学业水平不同的学生的区分程度或鉴别能力.具有良好区分度的测验,实际水平高的被试应得高分,水平低的被试应得低分.它是考试是否有效的“指示器”,被作为评价试题质量、筛选试题的主要指标与依据.试题区分度的计算方法很多,在此介绍一种比较方便的方法.对于客观题来说,使用等级相关分析,使用斯皮尔曼(Spearman)等级相关分析,即求总分与每个试题得分间的相关系数;对于主观题来说,看成是非等间距测度的连续变量,并且样本数大于 30,采用皮尔逊(Pearson)相关分析来对试题进行分析,即求总分与

Mean 和各题满分值的数据文件(可以直接将刚才的结果粘贴至 Excel 文档,然后用 SPSS 软件打开),单击 Transform \rightarrow Compute,出现 Compute Variable 对话框,在 Target Variable 中输入要计算的难度系数 P,在 Numeric Expression 中输入公式 Mean / W,得到各题的难度系数.本次考试的难度统计结果如表 2 所示.通过试卷各题的难度表可以看出,全卷难度:0.74;1 卷难度:0.79;2 卷难度:0.71,难易适中.大部分题目难易程度尚可,其中第 7,9,11,14,17,18,21,29(1),30(1) 难度高得分率高,对于选拔性考试,可以适当增加些难度;第 3,25,29,33(3) 题偏难.

每个试题得分间的积差相关系数作为试题的区分度.对区分度的评价如表 3 所示.

表3 试卷区分度的评价

项目鉴别指数与评价标准(伊贝尔,1965)

鉴别指数(D)	项目标准
0.40 以上	很好
0.30 ~ 0.39	良好,修改后更佳
0.20 ~ 0.29	尚可,但需修改
0.19 以下	差,必须淘汰

在本文使用的样本中,第 1 ~ 20 题为客观题,第 21 ~ 33 题为主观题.具体的分析方法:运行菜单 Analyze \rightarrow Correlate \rightarrow Bivariate,在弹出的 Bivari-

ate Correlations对话框中选择各个客观题字段和总分进入 Variables,然后在 Correlation Coefficients中单击 Spearman,完成后即可得到客观题的区分度.主观题的区分度方法同上,选择主观题和总分字段进入,然后选择 Pearson 即可.输出结果(部分)如表4所示.

表4 用斯皮尔曼相关系数计算区分度

题号	1	2	3	4	5	
1	相关系数	1.000	0.082*	0.173**	0.131**	0.171**
	Sig. (双频)		0.014	0.000	0.000	0.000
	N	894	894	894	894	894
2	相关系数	0.082*	1.000	0.059	0.048	0.061
	Sig. (双频)	0.014		0.077	0.155	0.070
	N	894	894	894	894	894
3	相关系数	0.173*	0.059	1.000	0.138**	0.083*
	Sig. (双频)	0.000	0.077		0.000	0.013
	N	894	894	894	894	894
4	相关系数	0.131**	0.048	0.138**	1.000	0.059
	Sig. (双频)	0.000	0.155	0.000		0.080
	N	894	894	894	894	894
5	相关系数	0.177**	0.061	0.083*	0.059	1.000
	Sig. (双频)	0.000	0.070	0.013	0.080	

输出结果的最后一行或最后一列每小題与总分之间的相关系数即为区分度.输出整理结果如表5所示.通过试卷各題的区分度表可以看出,如作为选拔性考试第2,17,30(1)題的区分度不够,需要修改.

表5 试卷各題的区分度

题号	区分度	题号	区分度
1	0.426	11	0.298
2	0.182	12	0.292
3	0.362	13	0.307
4	0.261	14	0.351
5	0.295	15	0.362
6	0.342	16	0.376
7	0.221	17	0.148
8	0.443	18	0.334
9	0.272	19	0.497
10	0.306	20	0.226

续表

题号	区分度	题号	区分度
21	0.240	29(2)	0.445
22	0.472	30(1)	0.170
23	0.370	30(2)	0.576
24	0.480	30(3)	0.597
25	0.433	31(1)	0.472
26	0.530	31(2)	0.494
27	0.499	32(1)	0.560
28	0.432	32(2)	0.595
29	0.486	33(1)	0.585
29(1)	0.246	33(2)	0.679
		33(3)	0.626

1.4 信度分析(Reliability)

信度是指同一个测验对同一组被试,施测两次或多次得分的一致程度.一个好的测验必须是稳定可靠的,多次使用所获得的结果是前后一致的.为了能够真实、准确地反映测量对象的实际水平,必须重视对试卷信度的研究,从而正确地判断测量结果的价值.信度是反映测量中随机误差大小的指标,由于造成误差的方式和来源多种多样,所以信度的估计方法也是多种多样.在试卷信度的检验中一般采用的是同质性信度.同质性信度是衡量测验内部所有题目间一致性程度的指标.内部一致性系数的一种粗略计算方法是求测验的分半信度,在SPSS软件中一般采用科隆巴赫(Cronbach) α 系数,它可以计算任何测验的内部一致性系数,而不要求测验题目必须是记分型的.对信度的评价看信度系数值,应该在0~1之间,如果量表的信度系数在0.9以上,表示量表的信度很好;如果量表的信度系数在0.8~0.9之间,表示量表的信度可以接受;如果量表的信度系数在0.7~0.8之间,表示该量表应进行较大修订,但仍不失其价值;如果量表的信度系数在0.7以下,表示量表有些项目需要抛弃.具体的分析方法:运行菜单 Analyze \rightarrow Scale \rightarrow Reliability Analysis 打开信度分析对话框,在弹出的对话框中将所有表示题目的字段全部选择进入 Items 框中,在

Model 栏后的下拉列表采用默认的 Alpha 方法 (Cronbach α 系数). 得到分析结果如表 6 所示. 从分析结果可以看出, 本次考试的信度系数为: $\text{Alpha} = 0.876$, 信度可以接受. 为了提高试卷的信度, 可以从以下几个方面改进: 适当增加试题的数量; 保持所有试题的难度接近正态分布; 努力提高试题的区分度; 严格监考和按评分标准给分.

表 6 全卷信度分析结果

Cronbach's Alpha	项数
0.876	41

1.5 效度分析 (Validity)

效度是指测验能够测出它所欲测特质的程度, 即试卷准确地测量了考试目的欲测内容的多少. 具体地讲就是覆盖面和权重在多大程度上体现了课程标准的要求, 能否有效检验考试能力水平和反映教学基本要求的完成情况. 效度与信度的区别是: 信度反映了考试中随机误差的大小, 而效度反映了考试中系统误差的大小. 效度的取值范围在 0 至 1 之间. 一般来说, 效度系数在 0.4 ~ 0.7 之间, 值越大效度越高. 有的效度系数达到 0.35 就符合要求, 有的要达到 0.65 才能被认为是有效的测验. 常用的效度检验方法, 一是效标关联效度法, 这种方法首先是寻求一种可靠的效标, 然后求出测试结果与效标的相关系数, 该相关系数则为效标关联效度; 二是构想效度法, 这种方法是根据研究内容形成一些理论假设, 然后检验测试结果对这些假设的验证程度, 验证程度越高, 说明构想程度越高. 在此利用效标关联效度法进行分析, 把学生的平时成绩作为效度分析的效标, 利用积差相关法求效度 (Pearson 法). 本文使用本校高三加试学生为计算效度的样本: 平时成绩一般由 3 部分构成: 考勤 + 作业 + 课堂表现, 总分 30 分. 具体操作方法为先建立考试成绩与平时成绩表, 其中 Total 为考试总分, PS 为平时成绩. 在 SPSS 中单击 Analyze \rightarrow Correlate \rightarrow Bivariate, 选择 Total 和 PS 字段进入表达式, 然后在 Correlation Coefficients 中单击 Pearson. 执行结果如表 7 所示. 得出总分与平

时成绩的相关系数为 0.762. 说明本次考试与平时成绩相关性较好, 即本次考试反映了学生的真实水平, 符合效度要求.

表 7 效度分析结果

		总分	PS
总分	Pearson Correlation	1	0.762**
	Sig. (2 - Tailed)		0.000
	N	90	90
PS	Pearson Correlation	0.762**	1
	Sig. (2 - Tailed)	0.000	
	N	90	90

** Correlation is significant at the 0.01 level

2 分析结论与改进建议

通过 SPSS 软件对试卷考后的成绩进行分析, 可得出该次高考模拟试卷与成绩的初步评价. 试卷基本上覆盖了高中物理教材中的主要知识点, 涉及各个章节的内容所占的分数分布合理. 总体来说该套试卷的设计比较科学, 能较好地检测出学生对知识掌握的程度. 本次考试平均分为 111.56 分, 标准差为 22.17, 成绩基本符合正态分布. 信度系数为 0.876, 信度可以接受; 个别试题区分度不高; 试卷整体难度指数为 0.74, 属于正常难度范围; 效度系数为 0.762, 考试成绩能够反映学生的真实水平. 大部分学生对知识掌握较好, 但学生成绩多集中在平均分上方, 建议在试题的区分度上有所调整, 适当加大一些试题的难度和区分度, 来区分学生实际水平的差距, 有利于对学生的选拔, 同时可以提高考试的信度. 学会 SPSS 软件的操作, 可以提高广大教师分析试卷的效率, 准确得到分析结果, 为今后的教学和考试提供科学依据.

参考文献

- 薛薇. SPSS 统计分析方法及应用 (第 3 版). 北京: 电子工业出版社, 2013
- 张文彤. IBM SPSS 数据分析与挖掘实战案例精粹. 北京: 清华大学出版社, 2013