



医学课程考试数据分析与评价初探（初稿）

李冬 姜声扬 程纯 （226001）

摘要： 介绍在医学课程考试数据分析中，系列质量指标、相关计算机工具软件及评价方案。

关键词： 试卷分析 质量指标 工具软件评价方案

医学课程考试是在期末或一门课程结束时进行的对课程学习和教学情况的全面检查，其自身质量非常重要。考试目的不同，对考试的控制要求也截然不同，只有符合预定要求的考试才是成功的考试。课程考试不同于水平考试、选拔考试等其他考试之处在于它必须完成对学生的学习状况进行检查和区分；而且还要对教学计划的完成程度进行检查以及对学生成绩进行解释和便于交流。由于考试质量控制缺乏统一明确的评判依据，任课老师命题组卷随意性较强，常常出现一些偏差。因此，建立一套实用规范的考试数据分析与质量评价方案以规范课程考试意义重大。几年的医学教学实践中我们根据教育测量学原理，结合实际情况在试卷数据分析与质量评价方面进行了一些有益尝试。

1 质量指标及实用意义

1-1 平均分： 表示考试分数集中趋势的统计量，通常利用平均分可以估计学生成绩分布的密集区域和考试的难易程度，并且对分数解释有重要影响，平均分应控制在70-80之间。

1-2 标准差： 描述一组分数离散情况的最常用、最可靠的统计量，利用标准差可以观察学生学习能力与考试结果是否一致，判断考试分数的集中程度，标准差理想控制范围在7-10。

1-3 合格率： 考试成绩达到或超过60分的学生比例，它是观察教学目标完成情况的一个统计指标，合格率偏低说明教学目标过高或出题存在偏差，而100%的合格率同样不应该出现。

1-4 优良率： 考试成绩达到或超过80分学生的比例，它是观察高分区成绩分布、成绩解释能力的统计指标。考试优良率偏低将减少中等偏上学生的社会认可机会，而优良率过高又导致对成绩优异者的不公和成绩交流的价值。

1-5 成绩分布统计： 以10分为一个分数段，统计每个分数段包含的学生数，组成成绩的频数分布（frequency distribution）统计表。将该统计表各分数段的中值和该组学生数对应点依次制成条形图，形成中间高两边低的成绩频数分布图。

频数分布统计表和频数分布图各具特色，前者数据准确而后者形象直观。成绩分布的统计是考试数据统计的一项重要工作，它可以初步了解分数分布的特征，粗略判断考试运作是否理想；对学生成绩进行初步分档，帮助合理解释学生成绩。理想考试的成绩应该呈对称分布，高分或低分考生数量较少，中间部分数量较多，且其分档情况在成绩解释上比较合理。

1-6 成绩分布的正态性检验： 人群个体智力分布属于正态分布（normal distribution），学生学习成绩的分布受智力影响较大，也应呈正态分布。当考试成绩分布呈正态分布时，可以认为测量效果符合自然分布规律，试卷的编制质量也较好，考试比较成功，近年一些学者认为负偏态分布也属合理的分布。正态性检验的经典方法是矩法（method of moment） [1] 。

1-7 难度： 有试题难度和考试难度 [2] 之分，前者是分析试题使用的量化指标，在题库建设中对试题做难易定位的测试参数，后者则是整个考试难易程度的综合反映，在考试分数控制、成绩解释上十分重要。

1-8 区分度： 指考试本身对学业水平不同学生的区分程度或鉴别能力，具备良好的区分度的考试才会满足考试效度要求；考试的信度随考试区分度的提高而增加，且前者的增加速率更快；试题的难度也与区分度关系密切，在难度接近于0.5时，区分度量值接近最大。试卷区分度最常用的计算方法是极端分组法 [2]，即通过比较两个极端效标组（高分组和低分组）的差异来估计试题区分度的。

1-9 信度： 评价考试结果的稳定性和可信程度的量化指标，实际测量中多采用分半信度或内在一致性信度算法 [3]，前者适合于由选择题组成的标准化考试，而后者使用上无任何限制。

1-9-1分半信度： 方法是奇数题组成一个部分，偶数题构成奇数题的复本，计算两半考试得分的皮尔逊相关系数（Pearson's product moment correlation），再采用 Spearman-brown公式矫正。

1-9-2内在一致性信度：把考试分成若干部分，用克伦巴赫（Cronbach） α 系数计算各部分一致性程度并做为考试信度值的估计值。

1-10 效度：用于考试误差的估计，取同一批学生的某次权威性考试 [4] 作为校标，与待测考试进行数据相关性分析，该方法也称做效标效度。

1-11 极值：指该次考试出现的最高分数和最低分数，极值既是数据完整性的体现，又是数据的关键观点。课程考试不应出现0或满分的成绩，0分属于异常数据，在数据分析前就应将其剔除，而出现多个满分则提示较难考题数量严重不足。

2 计算工具与评价方案

2-1 计算工具：数据分析指标繁多、计算复杂，但其数学模型都来源于数理统计方法，所以各类统计学软件都是很好的工具。

2-1-1 SPSS (Statistics Package for Social Science)：一种目前十分流行的社会科学用统计软件包，采用了窗口化操作界面，统计功能齐全，只是使用该软件包需要一定的统计学相关知识和对该软件的操作十分了解。

2-1-2 Stata：美国计算机资源中心（Computer Resource Center）研制的功能强大又小巧玲珑的统计分析软件，采用命令或编程方式操作，该软件的使用对一些专业的要求更强。

2-1-3 试卷分析专用工具：根据以往的实践和相关资料，我们研制开发了《通用试卷数据处理分析系统》 [5]（下载地址：http://s.jy.ntu.edu.cn/test_page_0.exe）。该工具以教育测量学经典理论和方法为依据，专为试卷数据分析工作量身定做，操作简单，输入数据后可一次完成各项指标的计算，极大减低了试卷分析工作的难度。

2-2 考试数据综合评价方案 本着积极引导、科学评价、规范管理、逐步提高的主导思想，我们采用加权定量评定法 [4] 设计了这套方案。方法是在完成上述各项指标测量后，按下量表确定被测考试各项指标的得分，计算质量指数值，最后根据质量指数值做出评价，质量指数数值范围在0 - 1之间，数值越大考试越规范。

表 1: 课程考试质量评价量表

指标	标准	权重系数	达标值	未达标值	达标得分	未达标得分
平均分	70 ≤ 范围 ≤ 80	0.1	1	0.30	0.100	0.030
标准差	7 ≤ 范围 ≤ 10	0.12	1	0.30	0.120	0.036
优良率 %	20% ≤ 范围 ≤ 50%	0.1	1	0.30	0.100	0.030
合格率 %	80% ≤ 范围 ≤ 100%	0.1	1	0.30	0.100	0.030
成绩统计图 / 表	图形呈两端低，中间高	0.05	1	0.30	0.050	0.015
分布正态性检验	正态或负偏态	0.1	1	0.30	0.100	0.030
难度	0.70 ≤ 范围 ≤ 0.80	0.04	1	0.30	0.040	0.012
区分度	> 0.2	0.1	1	0.30	0.100	0.030
考试信度	> 0.7	0.12	1	0.30	0.120	0.036
考试效度	> 0.4	0.12	1	0.30	0.120	0.036
极值	90< 最高分 <100, 0< 最低分 <70	0.05	1	0.30	0.050	0.015

质量指数计算：质量指数 = $\sum X_i$

X_i 为各检测指标的得分，未测项目得分记为 0（未测项过多将产生较大误差），在多批次考试分析中，当出现相同项目未测情况，可进行数据修正。

修正系数： $\sigma = 1 / (1 - \text{未测项权重系数})$

修正后的质量指数 = $\sigma \cdot \text{原质量指数}$

考试评价：0 < 不合格 < 0.6 ≤ 合格 < 0.7 ≤ 良好 < 0.85 < 优秀

2-3 评价实例 下面以公共卫生学院 6 门课程考试数据为实例，用本方案对其进行评价，数据见下表。

表 2 公共卫生学院 6 门课程考试数据

课程	劳卫	营卫	卫化	统计	毒理	少儿
----	----	----	----	----	----	----

指标						
平均分	73.1	76.5	83.9	80	80.5	91
标准差	9.74	7.19	7.35	9.27	12.75	5.13
优良率	25.5%	46.4%	72.7%	60.7%	67.9%	98.2%
合格率	92.7%	98.2%	100.0%	100.0%	92.9%	100.0%
统计图 / 表	合格	合格	合格	合格	合格	不合格
正态性	负偏态陡峭 .	负偏态陡峭 .	正态	负偏态 .	负偏态 .	负偏态 .
难度	0.73.1	0.7654	0.8385	0.7996	0.8046	0.9102
区分度	0.245	0.1729	0.1877	0.2321	0.3171	0.1279
效度	0.4973	0.5438	0.5117	0.4205	0.521	0.6343
极值	89/43	92/51	96/66	93/60	98/41	97/78
质量指数	0.775	0.74	0.572	0.74	0.628	0.418
矫正后	0.881	0.841	0.65	0.841	0.714	0.475
结论	优秀	良好	合格	良好	良好	不合格

说明：计算使用我们研制的工具软件（3-1-3）：信度未测，故对其矫正。

3 结束语

数据分析可以帮助教师发现考试中存在的不足，引导教师提高试卷的编制质量，使课程考试更加科学规范，学生成绩更加有价值。医学类考试知识点多、考题量大，对试题进行逐一分析是不现实的，因此本方案淡化了对考题的检测工作，而立足于充分挖掘成绩数据自身隐含的信息，采用增加和细化评价指标，强调的是考试的总体质量评价。配套计算机工具软件的使用极大地降低了开展这项工作的难度，加之合理可行的评价方案，为推动考试质量评价工作的开展创造了必要的条件。

参考文献

- 1 杨树勤主编. 卫生统计学 [M], 北京: 人民卫生出版社, 1990.:56
- 2 罗方亮, 张绍文, 韦宁. 现代教育评价技术在考试分析中的应用 [J], 河北能源职业技术学院学报, 2001, (1):79
- 3 张敏强. 教育测量学 [M], 北京: 人民教育出版社, 1998.:106-108
- 4 戚焕林, 邱坤荣. 高校教学管理基础 [M], 南京: 东南大学出版社, 1990.:223,217
- 5 李冬, 姜声扬. 试卷处理分析软件的设计 [J], 交通高教研究, 1999, (3):71

An Approach for Evaluating and Analyzing the Data in Medical Course Examinations Abstract This paper introduces the series targets for quality, the related tool software of computer techniques and their programs for evaluation .

Key words: test paper analysis, quality target, tool software, programs for evaluation