

体育统计套题命题考试信度、效度检验及质量分析*

刘 华 刘富萍

(云南师范大学体育学院, 云南 昆明 650092)

摘 要: 通过对体育统计学套题命题考试信度、效度、正态性检验等综合量化分析, 为现有套题的不断改进和完善, 进一步提高命题科学性, 建立合理的控制模式, 以及将来实现计算机题库命题考试提供了必要的参考依据。

关 键 词: 体育统计学; 套题考试; 测量信度; 测量效度; 命题科学性

中图分类号: G807 文献标识码: C 文章编号: 1007- 9793(2000)01- 0067- 05

1 研究目的:

教学考核是整个教学工作中一个十分重要的环节。提高考试与命题的科学性对现代教育的改革与发展以及对学生综合能力的培养都有着十分重要的意义。

体育统计学是体育基础理论学科的主要课程之一。作为我校的一类课程建设, 自96年起, 施行了套题命题考试。本文通过对体院97、98届毕业生使用的第三、第五、第一、第四套考试结果所作信度、效度、正态性等检验以及对试题质量的综合分析, 旨在进一步对现有套题的考试、编制工作等进行更深入的研究和探讨, 使学科考试工作更加科学、合理、规范, 亦为将来实现计算机题库命题考试提供必要的参考依据。^{[1][6]}

2 研究方法

2.1 试卷选取

试卷由学科教师统一编制, 学校教务处备案, 并于考试前在十套题中随机抽取。根据研究需要, 每次考试使用A、B两套试卷, 其中97届使用的是第三、第五套, 98届使用的是第一、第四套。

2.2 试卷检验

2.2.1 采用逻辑分析方法, 请有关学科三位以上专家利用编题时拟定的双向细目表, 确定考试内容效度。

2.2.2 运用数理统计方法对考试结果作正态性检验、方差分析以及对考试信度、效度、区分度等进行估价。有关数据处理采用SPSS/PC(V. 3. 2)进行分析, 在486计算机上完成。

3 结果与分析

3.1 成绩统计及正态性检验

* 收稿日期: 1999- 11- 03

© 1994-2012 China Academic Journal Electronic Publishing House. All rights reserved. <http://www.cnki.net>

第一作者: 刘 华(1958-), 女(满族), 辽宁省绥中县人, 副教授, 主要从事体育统计与测量及体质研究

两个年级考试总人数101人, 考试平均成绩及各分数段人数统计结果见表3- 1- 1。

表3- 1- 1 考试平均成绩及各分数段人数统计

Tab. 3- 1- 1 average score of examination and unmdr of people of different fraction

试 题	样本数 (n)	$\bar{X} \pm S$	各分数段人数(n)				
			60分以下	60- 69分	70- 79分	80- 89分	90分以上
第三套	25	78.56±9.20	1	3	7	11	3
第五套	28	79.15±11.18	2	3	7	11	5
第一套	23	77.70±11.52	2	3	5	11	2
第四套	25	73.96±12.84	2	6	8	5	4
总 计	10	77.63±11.31	7	15	27	38	14
总百分比		100%	6.93%	14.85%	26.73%	37.62%	13.68%

从表中可以看出, 各套试卷的平均成绩、各分数段人数、百分比均与总平均成绩、各分数段人数、总百分比呈相似的分布类型。以总考试成绩为例, 70- 79分和80- 89分分数段人数最多, 合计约占总人数的64.73%; 60分以下, 90以上分别约8.6%和7%左右。经正态性 χ^2 检验, $\chi^2=4.92 < \chi^2_{0.05(2)}=5.90$, 故 $P > 0.05$, 总考试成绩近似正态分布, 曲线峰值向次高分偏移, 基本符合期末考试常用的控制模式。^{[3][6]}

3.2 题型及题量分析

试题编制基本为客观性试题, 六种主要题型及其分值见表3- 2- 1。

表3- 2- 1 考试题型及其分值

Tab. 3- 2- 1 theme type of examination and more or less of score

题 型	名词解释	是非判断	单、多项选择	简答题	填空题	计算分析
分 值	10- 12分	8- 10分	10- 12分	10- 12分	28- 26分	34- 28分
百分比	11%	9%	11%	11%	27%	31%

六种题型中测量知识性的题型约占总分值的1/3弱; 测量领会、运用层次的题型约占总分值的1/3; 测量计算及综合分析能力的约占总分值的1/3强。^{[3][2]}从学生考试交卷时间统计可以看出, 约57%的学生交卷时间在2小时以上, 到时被抽卷的学生约20%左右, 见表3- 2- 2。

表3- 2- 2 不同交卷时间人数统计表

Tab. 3- 2- 2 Unmdr of people give in test paper in different time

试 题	交 卷 时 间			
	1- 1.5小时(人)	1.5- 2小时(人)	2- 2.5小时(人)	到时抽卷(人)
第三套	1	4	15	6
第五套		6	18	3
第一套		3	12	8
第四套		4	14	7

备注: 总考试时间为2小时30分。

3.3 内容效度检验

对内容效度的检验, 考前请有关学科的专家对根据教学目标而制订的“体育统计考试双向细目表”进行检查, 并依专家意见在双向细目表的内容、目标等权重上作了适当修改, 最后形成表3- 3- 1。

表3- 3- 1 “体育统计”考试双向细目表

Tab. 3- 3- 1 the direction of double and detailed form of sport statistics examination

教 学 内 容	教 学 目 标				
	知 识	领 会	运 用	分 析	合 计
诸论		0. 25			0. 25
样本特征数	0. 25	0. 25	0. 25		0. 75
概率分布	0. 25	0. 25	0. 50		1. 00
参数估计、假设检验	1. 00	1. 00	1. 50	0. 50	4. 00
方差分析			0. 50	0. 50	1. 00
相关与回归分析	0. 50	0. 50	1. 00	1. 00	3. 00
合 计	2. 00	2. 25	3. 75	2. 00	10. 00

然后按表3- 3- 1形式再列出“双向细目检查表”,请三位专家分别对第三、第五套试题就内容(目标)、代表性、容量、难度等与‘双向细目表’作对照,进行逻辑分析判断,按其吻合程度的高低作出内容效度的估计。^[3]

表3- 3- 2 专家检查评定表

Tab. 3- 3- 2 examine and evaluate by expert

专 家 (n= 3)	第 三 套				第 五 套			
	目标(内容)	代表性	容量	难度	目标(内容)	代表性	容量	难度
A	++	+	0	-	++	+	+	-
B	++	+	+	0	+	+	+	0
C	+	++	0	-	++	+	0	-

表中“+”表示合、强、多、难,“-”则与之相反,“0”表示属中。从专家对第三、第五套试卷检查评定结果的统计可以看出,在对试题与教学内容和目标吻合程度的评价上,三位专家的意见较一致,对试题取样代表性,三位专家的意见也基本一致。

3. 4 信度系数检验

对信度的检验,采用了分半积差相关法和库理法分别对其内部一致性进行估价^{[5][3]}。以第三套试卷为例,用分半相关,斯一布公式。经检验,信度系数为 $r_{1.1} = 0. 68 > r_{0. 01(23)} = 0. 537$, 故 $p < 0. 01^{**}$, 相关具有显著性。见表3- 4- 1。

表3- 4- 1 分半积差相关系数表

Tab. 3- 4- 1 half a divideing with coefficient of correlation test

奇数题得分偶数题得分						公式:
题号	X	Y	X ²	Y ²	XY	
1	45	47	2025	2209	2115	(1) $r_{\frac{1}{2}, \frac{1}{2}} = \frac{\sum XY - \frac{(\sum X)(\sum Y)}{n}}{\left[\sum X^2 - \frac{(\sum X)^2}{n} \right] \left[\sum Y^2 - \frac{(\sum Y)^2}{n} \right]}$
2	46	46	2116	2116	2116	
⋮	⋮	⋮	⋮	⋮	⋮	
26	24	34	576	1156	816	
Σ	973	1022	38535	42450	40077	(2) $r_{1.1} = \frac{2 \cdot r_{\frac{1}{2}, \frac{1}{2}}}{1 + r_{\frac{1}{2}, \frac{1}{2}}}$

用库理法检验,信度系数为 $r_{K-R20} = 0. 918$, 内部一致性较高, 见表3- 4- 2。

表3- 4- 2 第三套体育统计考试各题通过情况(n= 25)

Tab. 3- 4- 2 situation of every question pass in examination of the third test paper

试题号	答对人数	p_i	q_i	p_iq_i	
1	19	0. 76	0. 24	0. 18	公式: $r_{KR20} = \frac{k}{k-1} \left(1 - \frac{\sum p_i q_i}{S_i^2} \right)$ K 为试题数目; $p_i q_i$ 分别为每题的通过率 和未通过率; S_i^2 为考试成绩的方差
2	22	0. 88	0. 12	0. 11	
⋮	⋮	⋮	⋮	⋮	
26	12	0. 48	0. 52	0. 25	
Σ				9. 89	

分半信度系数低于库理法信度系数, 作者认为主要有两个方面的原因。一是分半法要求的两半试题的平均数($\bar{X}$)和标准差(S)在实际编题过程中难以完全满足大致相同的条件; 二是本文用库理法对答题通过率的统计是以答对半数以上分值即答对进行的。

3. 5 考试结果的方差分析

套题间的等价性也是试题质量的一个方面, 一般可以通过对复本信度系数的检验提供时间稳定性信息。本文由于一些客观原因, 未能做复本信度检验。而用方差分析方法对不同试卷考试结果平均成绩间差异显著性检验能在一定程度上反映试卷的等价性。^{[4][5]} 方差分析结果表明, 四套试题考试平均成绩间无显著性差异。见表3- 5- 1。

表3- 5- 1 四套考试成绩方差分析表

Tab. 3- 5- 1 variance analysis for four suite examination

变异来源	平方和	自由度	均方	F 值	P
组 间	494. 13	3	164. 71		
组 内	12295. 36	97	126. 77	1. 299	> 0. 05
总 计	12798. 45	100			

注: $P> 0. 05$ 差异不显著

4 小结

4. 1 对二套随机抽检试卷内容效度的检验, 专家定性分析, 试题与教学内容和目标的吻合程度, 三位专家的评价意见较一致, 内容有效性可接受。
4. 2 考试信度检验, 用分半相关法和库理法进行定量分析, 结果显示内部一致性显著。本次研究未对复本信度作检验。
4. 3 对考试结果作正态性检验, 成绩分布近似正态分布。70- 90分数段人数较多, 说明试题区分度不够。
4. 4 采用单因素方差分析对套题间考试成绩差异的显著性检验表明, 差异不显著, 说明套题间具有一定的等价性。
4. 5 对试题质量进行综合分析, 提示对现有套题的改进应以提高试题区分度, 缩减计算量为主, 并在不断改进的基础上最终实现计算机题库命题考试。

参 考 文 献

- 1 桂诗春编著. 标准化考试- 理论、原则与方法. 广州: 广东高等教育出版社, 1986. 21- 26, 118- 222.
- 2 于信风著. 考试学引论. 沈阳: 辽宁人民出版社, 1987.
- 3 郭述平等编. 新编教育统计与测量. 长春: 东北师范大学出版社, 1991. 186- 206, 294- 296, 309- 313.

© 1994-2012 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

4 全国体育学院《体育统计》教材编写组. 体育统计. 北京: 人民体育出版社, 1995.

5 全国体育学院《体育测量评价》教材编写组. 体育测量评价. 北京: 人民体育出版社, 1995.

6 李静等. 提高体育硕士生入学命题科学性 & 建立合理控制模式的研究. 北京体育大学学报, 1998, (3) :

Testing and analysing the reliability and ralidity
of suite examination in sport statistics

Liu Hua Liu Fupin

(Yunnan Normal University, Kunming 650092)

ABSTRACT By testing and analysing the reliability, ralidity and normality et al. In the suite examination of sport statistics for students of physicaleducation. The results of the study offered refernces for the reform of the examination, improve of scientificalness of the proposition and realize of select problems by computer in the future.

KEY WORDS Sport statistics; Suite examination; Scientificalness of the proposition; Reliability; Validity