

服务教育创新的学习水平性别差异分析

谷峰

(浙江经济职业技术学院, 浙江杭州, 310018)

[摘要] 研究男女新生的学习水平的量化差异能帮助我们认识教育现状, 指导教育创新, 有效提高教育水平。文章给出了高职新生三门主要课程的学习水平的性别差异的实证常规分析和标准化分析。常规分析通过原始分的平均分计算学习水平相对差距。标准化分析采用实用标准化评价模型作为评价平台。实用标准化变换的偏移系数 E 体现了试卷的相对难度, 膨胀系数 M 体现了试卷的区分度。偏移系数和膨胀系数反映了试卷的难度、区分度在可用条件下得到的最大程度的修正。据此可以超时间、超空间的给出不同群体的水平比较, 对教学的定量化管理具有实际意义。Kappa 一致性检验和理想模式比较检验从不同角度考量了标准化分析的有效性。

[关键词] 标准化评价; 学习水平; 性别差异; Kappa 检验; 理想模型

[中图分类号] G649.21 **[文献标识码]** A **[文章编号]** 1674-893X(2011)02-0056-03

一、问题的提出

创新教育已成为高职教育中最生机勃勃的部分。研究分析高职院校男女新生的学习水平差异, 对其有一个比较清楚的、量化的认识, 无疑会使我们更好的认识高职教育的实际状况, 指导教育创新, 有效地提高高职创新教育水平。

学生的学习成绩是最常用、最有效地反映学生学习状况的数据^[1,2], 我们采取一组高职新生第一年两个学期的高等数学(试卷相同)、英语(试卷相同)和高职体育三门课程的真实成绩作为评价数据, 采用常规方法和实用标准化评价方法^[3], 对男女新生的学习水平做抽样分析。原始分的统计数据如表 1 所示:

表 1 男女新生不同学科考试成绩原始分分布

原始分	平均分	标准差	90-100 分	80-89 分	70-79 分	60-69 分	0-59 分	
高等数学	男	75.17	12.95	15.22%	19.57%	34.78%	21.74%	8.70%
	女	75.46	12.59	17.86%	21.43%	20.54%	33.93%	6.25%
英语	男	82.17	6.06	13.04%	54.35%	30.43%	2.17%	0.00%
	女	83.55	5.47	9.82%	72.32%	16.07%	1.79%	0.00%
高职体育	男	77.87	11.97	4.35%	52.17%	34.78%	4.35%	4.35%
	女	82.69	4.89	12.5%	63.39%	24.11%	0.00%	0.00%

二、学习水平性别差异的定量分析

1. 学习水平性别差异的常规分析

常规分析是通过原始分的平均分计算男女学生之间的学习水平相对差距:

高等数学课程的相对差距

$$\bar{D}_{\text{math}} = \frac{75.46 - 75.17}{100} = 0.29\%$$

英语课程的相对差距

$$\bar{D}_{\text{eng}} = \frac{83.55 - 82.17}{100} = 0.38\%$$

高职体育课程的相对差距

$$\bar{D}_{\text{ath}} = \frac{82.69 - 77.87}{100} = 4.82\%$$

其中的高职体育因为男女生考试内容不

[收稿日期] 2011-02-21

[基金项目] 浙江省教育科学规划研究课题“高职学生学习水平的实证定量分析及扩展研究”(SCG270)

[作者简介] 谷峰(1959-), 男, 浙江温州人, 讲师, 主要研究方向: 计算数学。

同，可比性不大。常规分析只考虑了平均分，未将反映分数分布的区分度的标准差等影响因素考虑进去，结果有所偏颇。

2. 学习水平性别差异的标准化分析

原始分数含义不清，试卷质量影响原始分数的分布，从而影响对考生的实际水平的反映，很难据此确定考生在总体中的排序。如果将原始分数转换为相对分数，使不等值的原始分数转换为等值的标准分数，使标准分数能进行比较和运算，以此表示学生的成绩水平，可以更科学地反映学生的实际学习水平^[4,5]。

采用实用标准化评价模型，将原始分数 $\{x_1, x_2, \dots, x_n\}$ 转换成实用标准分数

$$\{y_1, y_2, \dots, y_n\}: y_i = M(x_i + E), i = 1, 2, \dots, n。$$

其中 E 体现了试卷相对受试学生的相对难度。 $E=0$ 表示相对难度适当； $E<0$ 表示试卷相对难度较低； $E>0$ 表示试卷相对难度较高。膨胀系数 M 体现了试卷相对受试学生的区分度。 $M=1$ 表示试卷区分度与受试学生学习水平匹配； $M<1$ 表示为了使平均分达到 μ ，同时按照实用要求将分数总体保持在 $[0,100]$ 内，分布曲线有所收缩； $M>1$ 表示分布曲线扩展，区分度改善。偏移系数和膨胀系数反映了两份试卷的难度、区分度差别在可用条件下得到的最大程度的修正。得到的标准分数既能在同一个平台上评价，又保证了可用性。即其平均分给定，包含于 $[0,100]$ ，且区分度达到最大。在比较两组学生的学习水平相对差距时，我们可取 $\frac{M_A E_A - M_B E_B}{2\mu} + \text{试卷难度相对差}$ 作为两组

学生的学习水平相对差距的标准化评价。这个水平相对差距系数综合考虑了试卷的难度和区分度，以及学生水平分布的具体情况，具有实用性和可行性。据此可以超时间、超空间的给出不同群体的水平比较，对教学的定量化管理具有实际意义^[3]。

对上述标准化评价模型，取标准分平均分 $\mu=75$ ，对男女新生样本组计算如表 2 所示。

其中的 σ 就是标准分数的标准差。与表 1 比较，显然标准分的区分度比原始分要好，分数分布更均匀。从三门课程的标准变换的膨胀

系数和偏移系数可以看出，高等数学试卷的难度和区分度适中，与学生的学习水平般配。英语试卷难度和区分度都偏低。高职体育的男生组测试与学生的水平较般配，女生组测试偏容易。女生与男生的高等数学学习水平相对差距为 -2.65%，男生微弱领先。女生与男生的英语学习水平相对差距为 40.64%，女生比男生好。以下会分析这个数据的可靠性不高。从 σ 值可以看出女生中的英语水平落差也较大。女生与男生的高职体育学习水平相对差距为 62.71%。因为男女生测试内容不同，这个比较意义不大。从 σ 值可以看出男生中的体育水平差距较大。

表 2 男女新生样本组标准分分布

		σ	M	E	女生组对男生组的相对差异
高等数学	男	14.50	1.12	-8.16	-2.65%
	女	13.37	1.06	-4.86	
英语	男	11.81	1.95	-43.70	40.64%
	女	14.47	2.65	-55.21	
高职体育	男	13.63	1.14	-12	62.71%
	女	10.81	2.21	-48.75	

3. 学习水平性别差异标准化分析的一致性检验

将学生成绩的原始分和标准分按 A (90~100 分)、B (80~89 分)、C (70~79 分)、D (60~69 分)、F (0~59 分) 分等级后，对上述标准化评价做 Kappa 一致性检验，分析其可靠性。Kappa 一致性系数 (K coefficient of agreement) 即内部一致性系数，是判断一致性和信度评价的统计学重要指标。当 Kappa 一致性系数 $K<0$ 时，一致性强度极差；0.0~0.2 时，一致性强度微弱；0.21~0.40 时，一致性强度弱；0.41~0.60 时，一致性强度为中度；0.61~0.80 时，一致性强度为高度；0.81~1.00 时，一致性强度极强^[6]。

高等数学课程的男生组的原始分与标准分的观察一致率为 0.91，机遇一致率为 0.23，Kappa 值为 0.89。女生组的观察一致率为 0.96，机遇一致率为 0.23，Kappa 值为 0.94。

英语课程的男生组的原始分与标准分的观察一致率为 0.37，机遇一致率为 0.24，Kappa 值为 0.17。女生组的观察一致率为 0.43，机遇一致率为 0.29，Kappa 值为 0.19。

高职体育课程的男生组的原始分与标准分

的观察一致率为 0.72，机遇一致率为 0.34，Kappa 值为 0.57。女生组的观察一致率为 0.36，机遇一致率为 0.23，Kappa 值为 0.16。

可以看到，高等数学课程的原始分与标准分一致性很强。英语课程的原始分与标准分一致性较弱，试卷的难度和区分度失当影响了使用这组成绩判断学生学习水平的可靠性；高职体育课程男生组原始分与标准分一致性中等，女生组原始分与标准分一致性较弱。因此，高等数学课程学习水平的性别差异的比较评价的可靠性更高。

4. 理想模式比较分析

我们建立一个测评分数分布的理想模式，假设理想的分数分布是：服从正态分布，平均

分 75 分，不及格率 10%。这个理想条件考虑了实际教学要求，是否妥当可以做进一步探讨。可得出理想分布模型中各个等级的比例分布如表 3 所示：

表 3 分数分布的理想模式

成绩等级	A	B	C	D	F
占比	8.40%	23.89%	33.83%	23.89%	10%

以下将原始分分布、标准分分布与理想模式分布做比较分析，据此评价原始分分布和标准分分布的优劣，以及标准化比较评价的有效性（如表 4 所示）。

表 4 原始分分布、标准分分布和理想模式分布的比较

成绩等级		A	B	C	D	F	
高等数学	男生组	原始分	15.22%	19.57%	34.78%	21.74%	8.70%
		标准分	17.39%	17.39%	28.36%	28.26%	8.70%
	女生组	原始分	17.86%	21.43%	20.54%	33.93%	6.25%
		标准分	17.86%	21.43%	19.64%	31.25%	9.82%
英语	男生组	原始分	13.04%	54.35%	30.44%	2.17%	0.00%
		标准分	13.04%	23.91%	30.43%	19.57%	13.04%
	女生组	原始分	9.82%	72.32%	16.07%	1.79%	0.00%
		标准分	9.82%	33.04%	25.00%	18.75%	13.39%
高职体育	男生组	原始分	4.35%	52.17%	34.78%	4.35%	4.35%
		标准分	4.35%	34.78%	43.48%	10.87%	6.52%
	女生组	原始分	12.50%	63.39%	24.11%	0.00%	0.00%
		标准分	12.50%	23.21%	28.57%	31.25%	4.46%

表 4 显示，对 Kappa 值偏低的几组原始分，经过标准变换的标准分分布与理想模型的一致性明显高于原始分分布。说明用这几组原始分来评价学生学习水平的可靠性较低。这个方法从另一角度量度了标准化比较评价的有效性。

三、结论

本文实证给出了高职一年级学生三门主要课程的学习水平的性别差异的常规分析和标准化分析。标准化分析依据实用标准化评价模型，根据两组成绩的变换参数和试卷相对难度相对差给出男女生的学习水平相对差距评价。Kappa 一致性检验和理想模式比较检验从不同角度考量了标准化分析的有效性。与不考虑试卷区分度的常规分析相比，标准化分析的分析结果可检验可靠性，更具有效性。

参考文献：

- [1] Daniel L. Stufflebeam, George F. Madaus, Thomas Kellaghan, Evaluation Models: Viewpoints on Educational and Human Services Evaluation Kluwer Academic Publisher, 2000.
- [2] 石卫林. 学生终身学习能力的评价与比较--OECD 国际学生评估项目简介[J]. 比较教育研究, 2004, (1): 83-87.
- [3] 谷峰. 学习水平的标准评价模型及应用[J]. 数学的实践与认识, 2010, (12): 15-21.
- [4] 吴维宁, 高凌飏. 学业评价模式的初步研究[J]. 华南师范大学学报(社会科学版), 2005, (6): 107-111.
- [5] 曾桂兴. 标准化考试理论和方法的应用与发展[J]. 华南师范大学学报(社会科学版), 1989, (2): 1-3.
- [6] 戴维·K. 希尔德布兰德, 等. 社会统计方法与技术[M]. 北京: 社会科学文献出版社, 2005.

[编辑：胡兴华]