

大学生性别差异和年级差异对英语四级考试通过率的影响分析

Influence Analysis of Gender and Grades Difference on the Passing Rate in College English Test Band 4

卢玉桂, 刘琳

LU Yu-gui, LIU Lin

(广西大学数学与信息科学学院, 广西南宁 530004)

(School of Mathematics and Information Science, Guangxi University, Nanning, Guangxi, 530004, China)

摘要:用对数线性模型分析大学生性别差异和年级差异对英语四级考试通过率的影响, 先进行模型拟合再进行参数估计, 拟合模型时从饱和模型入手, 由高阶交互项逐步排除无统计意义的参数项, 直到选出一个能较好完成数据拟合的简约模型为止。大学生性别差异、年级差异与英语四级考试通过率都存在显著的交互效应, 女性的四级通过率要高于男性, 低年级学生比高年级学生更容易通过四级考试。

关键词:对数线性模型 logit 模型 列联表 英语 考试 通过率

中图法分类号: O211 文献标识码: A 文章编号: 1002-7378(2013)02-0147-02

Abstract: In this paper, the effect of gender and grades on the passing rate of College English Band Four is analyzed by using Log linear model, which is an effective method of multidimensional contingency table. The model is selected by using hierarchical model and parameters are estimated via the logit model. The result shows that the passing rate of girls is higher than that of boys; the passing rate of low grades is higher than that of high graders.

Key words: log-linear model, logit model, contingency Table, English, test, rate

英语四级是检验大学本科毕业生英语水平的一种有效途径, 大部分本科院校都把英语四级作为学生获得学位证书的一个必要条件。随着用人单位对大学毕业生的英语水平重视程度的增加, 分析大学生英语四级考试通过率的影响因素也越显重要。近年来, 学者们已从不同的角度分析影响大学英语四级考试通过率的因素。文献[1]以Weiner的归因理论为指导, 对大学英语四级考试成败归因进行调查分析。文献[2]运用多水平模型分析和评价大学英语四、六级考试通过率及其影响因素。文献[3, 4]从性别差异及其他相关因素分析大学生英语四级考试

通过率的原因。本文选择广西某高校2010年6月全国英语四级考试成绩作为分析数据, 运用对数线性模型的分层模型及logit模型分析大学生性别差异和年级差异对英语四级考试通过率的影响。

1 分析数据和方法

选择广西某高校2010年6月全国英语四级考试成绩, 共有3709个有效数据, 以 $x * y * z$ 三个变量建立三维列联表(表1)。其中 x 记为性别变量: 1代表男生和2代表女性; y 为年级变量: 1, 2, 3, 4, 5分别表示大学一年级至大学五年级的学生; z 为是否通过四级考试变量: 1代表通过, 2代表未通过(英语四级考试成绩大于或等于425分为通过, 否则为未通过)。

收稿日期: 2012-11-27

修回日期: 2013-01-17

作者简介: 卢玉桂(1988-), 女, 硕士研究生, 主要从事概率论与数理统计的研究。

表1 三维列联表

	一年级(个)		二年级(个)		三年级(个)		四年级(个)		五年级(个)		合计
	男	女	男	女	男	女	男	女	男	女	
通过	58	298	48	129	11	23	2	5	3	2	579
未通过	268	685	422	657	289	401	131	175	53	49	3130
合计	326	983	470	786	300	424	133	180	56	51	3709

运用 SPSS18.0 对表 1 中的数据进行分层对数线性模型选择^[5,6], 选出最优对数线性模型。拟合模型时从饱和模型入手, 从高阶交互项开始, 逐步排除无统计意义的参数项, 直到选出一个较好的完成数据拟合的简约模型为止, 其中模型拟合优度检验通过计算常用两个统计量 Pearson χ^2 和广义似然比得出。利用 logit 模型对该简约模型进行参数估计, 建立 logit 模型。假设 z 为因变量, x, y 为自变量, 则 logit 模型如下:

$$\ln(m_{ijk}) = \lambda + \lambda_k^z + \lambda_{ik}^{xz} + \lambda_{jk}^{yz}, \quad (1)$$

其中 m_{ijk} 为列联表中 ijk 格的频数期望值, $i = 1, 2, \dots, j = 1, 2, 3, 4, 5, k = 1, 2, \lambda$ 为总平均效应, λ_k^z 为变量 Z 的主效应, $\lambda_{ik}^{xz}, \lambda_{jk}^{yz}$ 为变量 x, y 与变量 z 间的交互效应。

2 结果分析

由表 2 结果可知, 维度为 1 和 2 时, 广义似然比和 Pearson χ^2 的拟合优度^[7] 显示其具有统计学意义 ($P < 0.001$), 即存在非常显著的主效应和二阶交互效应。而维度为 3 时, 广义似然比和 Pearson χ^2 拟合优度检验的 P 值分别为 0.785 和 0.777, 故其三阶交互效应不显著 ($P > 0.05$)。因此, 对于该饱和模型, 可剔除三阶交互项, 而保留主效应项和二阶交互效应项, 进而得到最优的简约模型。

表2 高阶效应分析

维度	自由度	广义似然比		Pearson χ^2		迭代次数
		计算值	P 值	计算值	P 值	
1	19	4512.994	0.000	4702.459	0.000	0
2	13	408.881	0.000	418.839	0.000	2
3	4	1.734	0.785	1.777	0.777	3

模型的拟合优度检验结果显示, 利用广义似然比时, 计算值: 1.731, 维度: 5, P 值: 0.785; 利用 Pearson χ^2 时, 计算值: 1.770, 维度: 5, P 值 0.778。由于两种方法的 P 值都大于 0.05, 故拒绝原假设, 认为该模型的拟合效果与饱和模型相比无统计学差异, 即对于表 1 中的三维列联表数据, 可用模型 (1) 对其进行模型拟合。

由表 3 结果可知, $\lambda_{11}^{yz} = -0.598 < 0$, 即变量 z

取值越大, 变量 x 的取值越小。也就是说, 在 z 取 1 时, x 取 2 的几率较大, 即女同学比男同学更容易通过全国英语四级考试; $\lambda_{11}^{yz} = -1.891 < 0$, 即变量 z 取值越大, 变量 y 的取值越小。也就是说, 低年级比高年级的学生更容易通过英语四级考试, 即年级越小, 通过四级考试的几率越高。又由 $|\lambda_{15}^{yz}| > |\lambda_{12}^{yz}| > |\lambda_{13}^{yz}| > |\lambda_{11}^{yz}| > |\lambda_{14}^{yz}|$ 可以看出, 在本次全国英语四级考试中, 大一学生的四级通过率最高, 而大三四大五次之, 大二最低。这说明各年级的四级通过率并不是逐年级递减, 只能说总的趋势递减。因为, 若某个年级学生的总英语水平较高或者较低, 则会出现奇异值, 比如本次英语四级考试中并不是大五的学生通过率最低, 而是大二学生的通过率最低。

表3 logit 模型的参数估计

效应	估计值	效应	估计值
λ_{11}^{yz}	-0.598	λ_{13}^{yz}	-0.058
λ_{11}^{yz}	1.891	λ_{14}^{yz}	-0.821
λ_{12}^{yz}	1.130	λ_{15}^{yz}	-2.142

3 结束语

本文基于对数线性模型分析大学生性别和年级对英语四级考试通过率的影响。大学生性别差异、年级差异与英语四级考试通过率都存在显著的交互效应, 而且女性四级通过率要高于男性, 低年级学生的通过率一般比高年级学生高, 但是也有例外, 在 5 个年级中, 大二学生四级通过率最低。

参考文献:

- [1] 朱剑飞. 大学英语四级考试成败归因分析[J]. 郑州航空工业管理学院学报: 社会科学版, 2010(04): 159-161, 170.
- [2] 王艳梅. 多水平模型在大学英语考试成绩影响因素分析中的应用[D]. 济南: 济南大学, 2007.
- [3] 黄燕青. 大学英语四级考试成绩的性别差异[J]. 广西教育学院学报, 2011(4): 140-142, 151.
- [4] 廖崇德. 中医大学生大学英语四级过级率及相关因素浅析[J]. 成都中医药大学学报: 教育科学版, 2000(1): 46-47.
- [5] 何晓群. 多元统计分析[M]. 北京: 中国人民大学出版社, 2004.
- [6] Goodman L A. The multivariate analysis of qualitative data: Interaction among multiple classifications[J]. J Amer Statist Assoc, 1970, 65: 226-256.
- [7] 张尧庭. 定性资料的统计分析[M]. 桂林: 广西师范大学出版社, 1991.
- [8] 张文彤. SPSS11 统计分析教程(高级篇)[M]. 北京: 希望电子出版社, 2002.

(责任编辑: 尹 闯)