卡方检验在语言研究中的应用

刘 晓 玲

摘 要: 本文通过典型实例分析卡方检验在单因素拟合度检验和多因素独立性检验中的应用。在单因素卡方拟合度检验中,介绍了如何用卡方检验观测数据之间差异性和检验样本分布与理论分布模型的拟合度;在多因素卡方独立性检验中,讨论了用卡方进行一般列联表独立性检验和四格表独立性检验;并说明卡方检验应注意的问题。

关键词: 卡方检验;拟合度检验;独立性检验;实际频数;期望频数

中图分类号: H0 文献标识码: A

文章编号: 1001 - 490X(2006)8 - 200 - 02

作 者: 湖南大学副教授;湖南,长沙,410082

卡方检验 (Chi-square test 或 X^2 test) 作为非参数检验的一种,是语言研究中常用的一种检验方法。卡方检验通过比较两项或多项频数,检测在一定显著性水平上实际频数与以某种理论模型或分布特征假设为基础的期望频数的差异度 (Bulter, 1985, 112)。卡方检验第一步,建立原假设 H_0 和备择假设 H_1 。第二步,根据理论经验或理论分布计算期望频数。第三步,根据实际频数和期望频数计算样本卡方值。计算公式是由英国统计学家皮尔逊提出的: $X^2 = \sum [(f_o - f_e)^2 / f_e] X^2$

(1)。公式中的 f_o 代表实际频数, f_e 代表期望频数, X^2 代表卡方值。第四步,根据自由度和显著性水平在卡方分布表中查找出对应卡方临界值。如果运算出的卡方值大于卡方临界值,接受原假设,反之,接受备择假设。在语言研究中,卡方检验主要用于两种情况:拟合度检验 (Goodness of fit) 和独立性检验 (Test of independence),下面论证卡方检验这两种检验中的应用。

一 单因素卡方拟合度检验

科学研究中,影响一个事物的因素可能很多,在一项检验中,只有一个因素改变称为单因素检验,多于一个因素的检验称为多因素检验 (盛骤, 1989: 241)。我们经常要检验某一实验因素实际出现的频数与期望出现的频数是否有显著性差异,这在统计学上称作拟合度检验 (韩宝成, 2000: 155), 以下是三种常见的单因素卡方拟合度检验。

1. 检验观测数据之间差异性

对观测数据之间差异性检验是相对简单的单因素检验,只检验分成多项的单因素频数之间有没有显著性差异。下面是一个检验单因素观测数据间显著性差异的例子。

某大学英语老师为了了解学生学习英语的动机做了一次小范围的问卷调查,该问卷调查共有 6 项,调查对象是 150 名普通高校非英语专业大二学生,调查结果整理后如表 1 第二

行所示。从调查结果上看有无显著性差异? 如果有显著性差异,其主要动机是什么?

表 1 学生学习英语动机实际频数和期望频数

学习英 语动机	各科都 喜欢学	特别喜欢 学习英语	认为学习 英语有用	通过四 六级	找个好 工作	应付 考试
实际频数	13	16	28	49	24	20
期望频数	25	25	25	25	25	25

具体分析步骤如下: 建立原假设 H_0 和备择假设 H_1 。原假设认为学生学习英语的动机没有显著性差异。计算期望频数。根据原假设,如果学生学习英语的动机没有显著性差异,那么 150 名学生对 6 项的选择概率相等,所以计算期望频数的公式是:期望频数 = 总数 / 分类项数 (2)。将表 1 中数值代入公式 2,得出的结果如表 2 第二行所示。计算卡方值。将表 1 中的实际频数和期望频数代入公式 1 求出卡方值为: $X^2 = \sum [(f_o - f_e)^2 / f_e] = [(13 - 25)^2 / 25] + [(16 - 25)^2 / 25] + \dots + [(20 - 25)^2 / 25] = 33.40$ 计算自由度。根据 Woods, 自由度在某种意义上可以理解为我们检验时可以依据的独立信息的数量 (Woods, 1986), 计算此类检验的自由度公式为: $df = n - 1$ (3)。其中 df 代表自由度, n 代表分类项数。将分类项数 6 代入公式 3 得出自由度为 5。得出结论: 从卡方分布表查得到当自由度为 5, 显著性水平为 0.05 时,卡方临界值是 12.6。因为计算出卡方值 33.40 远大于临界值 12.6,说明观测数据之间存在显著性差异,所以拒绝原假设接受备择假设,即学生学习英语的动机有显著性差异。通过观察可以看出动机为通过四六级的人数最多,由此可以得出学生学习英语的主要动机是为了通过四六级考试。

2. 检验样本分布与理论分布模型的拟合度

卡方检验在单因素拟合度检验中的另一个重要应用是检验样本分布是否与特定分布模型或理论分布模型拟合,下面是 Woods(1986)关于样本与一个特定的分布模型的拟合检验的例子。

以英语为母语的 5 岁儿童发音标准测试在英国格拉斯哥被证实有效并被标准化,对格拉斯哥地区的 5 岁儿童这个总体的标准测试的分数呈正态分布,其均值为 50,标准差为 10。美国一学区将该测试用于 184 名 5 岁儿童,测试分数经分段整理后如表 2 所示。该学区 184 名 5 岁儿童的测试得分是否可以看作是标准测试正态分布总体的随机样本?

表 2 美国学区儿童发音测试分数的分段频数表

分数段	-30	-35	-40	-45	-50	-55	-60	-65	-70	-75	75以上
人数	2	11	17	31	32	39	22	19	6	4	1

具体分析步骤如下: 检验均值和标准差有无显著性差

异。通过加权计算得出测试分数的均值是 48.8,标准是 12.4,将两组均值和标准差排列成列联表形式。根据已知数值运算

表 3 儿童发音测试分数的均值和标准差

	均值	标准差	行合计
格拉斯哥	50	10	60
美国学区	48.8	12.4	61.2
列合计	98.8	22.4	121.2

得卡方值为 0.564,从卡方分布表查得自由度为 1,显著性水平为 0.05 时卡方临界值是 3.84。计算的卡方值 0.564 远小于临界值 3.84,所以测试分数的均值和标准差与给定模型的没有显著差异。接下来要检验的是美国学区测试分数的分布是否呈正态分布。建立原假设 H_0 和备择假设 H_1 。原假设认为该学区 184 名五岁儿童的测试得分可以呈正态分布。计算期望频数。首先根据原假设计算每个数据段在标准正态总体中所占比例。当一个正态分布的均值为 μ ,标准差为 σ ,求小于总体数的某值 X 的分布比例用如下公式: $Z = (X - \mu) / \sigma$ (4)。将表 2 中数值代入公式 4 求出各分数段的 Z 值如表 4 第三列所示。值得注意的一点是卡方检验要求每个分

表 4 美国学区儿童发音测试分数整理表

分数段	实际频数	分数段标准化 Z 值	期望概率	期望频数	相对差
0 - 30	2	2.0 - 1.5	0.067	12.3	0.04
30 - 35	11	1.5 - 1.0	0.092	16.9	0.006
35 - 40	17	1.0 - 0.5	0.150	27.6	0.42
40 - 45	31	0.5 - 0.0	0.191	35.1	0.27
45 - 50	39	0.0 - 0.5	0.191	35.1	0.43
50 - 55	22	0.5 - 1.0	0.150	27.6	1.14
55 - 60	19	1.0 - 1.5	0.092	16.9	0.26
60 - 65	6	1.5 - 2.0	0.067	12.3	0.14
65 - 70	4	2.0 - 2.5	0.067	12.3	0.14
70 - 75	1				
75 以上	1				
卡方值				2.70	

(该表格改编自 Woods, 1986)

数段的频率次数要有足够大的期望值(至少不能小于 5),否则会导致卡方值的增大,所以表中将几个小的分数段合并。有了 Z 值后查正态分布表得到各 Z 值相对应的概率值,结果见表 4 第四列。求期望频率公式是:期望频率 = 总数 $\times$ 期望概率 (5)。将表 4 中数值代入公式 4 结果见表 5 第五列。

计算卡方值。将实际频数和期望频数数据代入公式根据公式 1 求出卡方值为: $X^2 = \sum [(f_o - f_e)^2 / f_e] = [(13 - 12.3)^2 / 12.3] + [(17 - 16.9)^2 / 16.9] + \dots + [(11 - 12.3)^2 / 12.3] = 2.70$ 得出结论。此类检验自由度的计算与检验观测数据差异性自由度的方法一样。例子经过合并后的分数段共有 8 个,把 8 代入公式 3 得到自由度为: $df = n - 1 = 8 - 1 = 7$ 从卡方分布表得到自由度为 7,显著性水平为 0.05 时,卡方临界值为 14.1。计算出的卡方值 2.70 远小于临界值 14.1,意味着原假设成立,即该学区测试分数的分布是否呈正态分布,又因为已经测得美国学区测试分数的均值和标准差与标准测试总体没有显著差异,因此该地区 184 名五岁儿童的测试得分可以看作是标准测试正态分布总体的随机样本。

二 多因素卡方独立性检验

除了进行拟合度检验,卡方检验还用于分析两个或两个以上的因素之间有无关联和判断因素之间的差异是由于样本抽样造成的随机差异还是由于来自不同总体造成的真正差异。下面分别说明卡方检验在其他类型列联表独立性检验和四格表独立性检验中的应用。

1. 列联表独立性检验

为了说明列联表卡方独立性检验,我们引用来自 Ferris 和 Politzer (Ferris & Politzer, 1981) 报告例子说明。

A、B 两组儿童,每组 30 人,A 组儿童出生于美国,直接接受英语教育,B 组儿童出生于墨西哥,早期学习西班牙语,后来接受英语教育。被观测的 60 名儿童要求写一篇不少于 100 字的影评,对作文中时态错误的个数统计如表 5(a) 所示。问 A、B 两组儿童作文中时态错误个数与早期接受的英语教育有无关系?

表 5 儿童作文中时态错误个数列联表

(a) 实际频数		时态错误个数			
		0	1	2 - 6	行合计
A 组	7	7	16	30	
B 组	13	11	6	30	
列合计	20	18	22	60	
(b) 期望频数		时态错误个数			
		0	1	2 - 6	行合计
A 组	10	9	11	30	
B 组	10	9	11	30	
列合计	20	18	22	60	

(此表摘自 Ferris & Politzer (1981))

分析步骤为:建立原假设 H_0 和备择假设 H_1 。原假设认为 A、B 两组儿童作文中时态错误个数与早期接受的英语教育无关。计算期望频数。根据原假设计算出每个单元格的期望频数的公式是:期望频数 = (行合计 $\times$ 列合计) / 总人数 (6)。将表 5(a) 中的数值代入公式 6 求出的期望频数如表 5(b) 所示。计算卡方值。将表 5 中的实际频数和期望频数代入公式 1 求出卡方值: $X^2 = \sum [(f_o - f_e)^2 / f_e] = [(7 - 10)^2 / 10] + [(7 - 9)^2 / 9] + \dots + [(6 - 11)^2 / 11] = 7.22$ 计算自由度。对于列联表检验,自由度的计算公式是: $df = (\text{列数} - 1) \times (\text{行数} - 1)$ (7)。将表 5 中的数据代入公式 8 得出自由度为 1。得出结论。从卡方分布表得到自由度为 1,显著性水平为 0.05 时,卡方临界值为 5.99。计算的卡方值 7.22 大于临界值 5.99,所以否定原假设接受备择假设,即 A、B 两组儿童作文中时态错误个数与早期接受的英语教育有关。

2. 检验四格表数据独立性

四格表又称相依表,是一种 2×2 形式的列联表格,计算 2×2 的四格表卡方的方法与计算其他类型的列联表方法一样,但是除此之外,计算四格表卡方值还有一个不需要计算期望频数的特殊公式,相对常规的卡方公式计算更简单方便,而且结果相同。四格表卡方的特殊计算公式(下转 101 页)

想一直占统治地位,儒家漠视自然、回避自然科学的态度对中国人的世界观产生了十分消极的影响,对近代数学的产生非常不利,因为数学是对自然现象和规律的量的反映,只有深入地探索自然,才能得出准确的数学关系。在中国古代,科学技术被视为“雕虫小技”,数学被讥笑为“九九贱技”,儒家认为一切专门学问都是“玩物丧志”,这种对科学技术的鄙夷态度给数学的发展同样投下了阴影。^[9]另外,在儒家思想和封建制度的影响下,形成了等级森严的“君臣、父子、官民、师生”关系,在这种关系下,人们的创造性、怀疑精神会受到极大程度的约束。相反在西方,崇尚理性、追求真理的精神一直影响着西方科技的发展,古希腊的智者柏拉图的“理念论”成为倡导理性的一部重要著作。“吾爱吾师,吾更爱真理”更是古希腊智者们追求真理、崇尚理性的真实写照。

中国古代数学曾经受到有机论自然哲学的影响,这种自然哲学过分强调系统的复杂性,认为自然就是“自然之势也”,“万物固以自然”^[10],强调“天人合一”、“天人感应”,正如有的学者所说:中国有机论之箭射过了近代,射到了卢瑟福那里,受有机论自然哲学的影响,中国数学错过了近代科学分门别类的发展良机。相反,在西方,在牛顿和笛卡尔时期形成的机械论自然观对数学和其他学科的蓬勃发展在一定时期内起到了重要的推动作用。另外,中国数学也曾经受到《周易》及玄学的影响,刘徽在《九章》注文的序中说:昔在包牺氏始画八卦,以通神明之德,以类万物之情,作九九之术,以合六爻之变。^[11]另一位数学家秦九韶则宣称他的重大发明是研究《周易》中的“大衍术”而获得的。部分研究者认为:中国古代数学不仅在写作形式上,而且在思维方法上都刻意模仿《周易》。

中国“八股取士”的选拔体制也是导致社会不重视数学乃至整个自然科学的重要原因。在唐朝,在教育制度和科举制度中还保留了数学的地位,唐朝的最高学府——国子监虽然设有明算科,把数学作为一个专业,但数学的官秩才是“从九下品”。宋代程朱理学盛行,在理学家们看来,数学毫无用处。到了元朝,新恢复的科举制度干脆将数学砍掉了。对“四书五

经”的执着和对“儒学”的推崇窒息了中国数学的发展。

第三,中国数学衰落的政治原因。华罗庚曾经指出:“明末以后,中国数学开始落后了,这是中国长期停留在生产力落后的封建社会所决定的”^[12]。陕西师范大学黄秦安教授在《论封建政治皇权对中国古代数学发展的影响》一文中指出:“经济与文化只能在政治一维坐标轴中充当参数,在封建专制下的政治文化所追求的是一种在不平等、缺乏自由的社会等级结构中被强化了了的稳定的、静态的、封闭的统治模式。可以推断被皇权泛化了的官本位意志所渲染的政治心理在多大程度上摧残、扼杀、消蚀了中国人对自然真理与奥秘的探索精神。从这些描述中可以清楚地看出:封建政治思想与数学精神在本质上是対立的、背道而驰的。”

参考文献:

- [1][2]李春华:《中国古代数学的辉煌成就》《通化师范学院学报》,2003年第2期。
- [3][11]欧阳维诚:《浅论《周易》对中国古代数学模式化道路形成及发展的影响》,《周易研究》,1999年第4期。
- [4]黄秦安:《中国古代数学的社会文化透视》,《陕西师范大学学报(哲学社会科学版)》,1998年第4期。
- [5]朱水林:《形式化:现代逻辑的发展》,人民出版社,1987年版,第293页。
- [6]曹胜斌:《论中国古代科学技术终结的数学原因》,《长安大学学报(社会科学版)》,2003年第2期。
- [7]爱因斯坦:《爱因斯坦文集》(第1卷),商务印书馆,1983年第574页。
- [8]孟庆云:《中国古代数学与中医学》,《中国中医基础医学杂志》,1997年第5期。
- [9][12]黄秦安:《论封建政治皇权对中国古代数学发展的影响》,《陕西师范大学学报(哲学社会科学版)》,1997年第4期。
- [10]沈南山:《古希腊数学流派哲学思想比较》,《金陵职业大学学报》,2001年第2期。

(责任编辑:南 桥)

(上接 201页)

为: $X^2 = N(AD - BC)^2 / (A + B)(C + D)(A + C)(B + D)$

(8)。四格表有一个重要特征,就是如果已知一个单元格内的数值和边际总数就可以计算出其他三个单元格的值,所以在四格表中只有一个值是自由变动的,即四格表卡方检验的自由度是 1 (Bulter, 1985, 119)。

三 小 结

卡方检验针对的主要是计数数据的统计,非常适合语言研究,但是卡方检验也有局限性,使用时应注意的问题主要有三点:(1)理论频数不能太小。通常,卡方检验要求单个理论频数不小于 5,四格表检验则要求理论频数不小于 10,才能避免卡方值出现严重错误(Brown, 2001, 168)。(2)保证观测数据的独立性。(3)卡方检验必须使用原始数据进行计算。如果把原始频数转换成百分比计算,得出的结果不准确(Woods, 1986; 桂诗春, 宁春岩, 1997)。

参考文献:

- [1]Woods, A., Fletcher, P. & Hughes, A. Statistics in Language Studies[M]. Cambridge: Cambridge University Press, 1986.
- [2]Bulter, C. Statistics in Linguistics[M]. Oxford: Basil Blackwell, 1985.
- [3]Ferris, M. R. & Politzer, R. L. Effects of early and delayed second language acquisition: English composition skill of Spanish-speaking junior high school students [J]. TESOL Quarterly 15, 1981.
- [4]桂诗春、宁春岩:《语言学方法论》,外语教学与研究出版社,1997年版。
- [5]韩宝成:《外语教学科研中的统计方法》,外语教学与研究出版社,2000年版。
- [6]盛骤、谢式千、潘承毅:《概率论与数理统计》,高等教育出版社,1989年版。

(责任编辑:南 桥)