

对应分析在天津市公民科学素质调查中的应用

张磊 李婧

(天津市科学技术信息研究所, 天津 300074)

[摘要] 讨论了对应分析方法的基本特点和使用条件, 在此基础上, 以 2010 年天津市公民科学素质调查数据为研究对象, 探索了公民文化程度与未接触过科技馆等科普场馆的原因之间的关系, 用 SPSS 统计软件示范了对应分析的过程, 对主要研究结果进行了解释。

[关键词] 科学素质 科普场馆 对应分析 惯量

[中图分类号] C81 **[文献标识码]** A **[文章编号]** 1673-8357(2011)05-0033-04

The Application of Correspondence Analysis in the Survey of Tianjin Citizen's Scientific Literacy

Zhang Lei Li Jing

(Tianjin Institute of Scientific & Technical Information, Tianjin 300074)

Abstract: It analyzes the characteristics and the application of correspondence analysis and the data of the Tianjin citizen's scientific literacy survey 2010, for the purpose of understanding the relationship between education background and having no access to science venues.

Keywords: scientific literacy; science venues; correspondence analysis; inertia

CLC Numbers: C81 **Document Code:** A **Article ID:** 1673-8357(2011)05-0033-04

在进行社会调查、市场调研时经常会遇到研究两个或多个分类变量之间关系的问题, 采用对应分析方法可以得到较为理想的结果。对应分析因为其表现形式直观、简洁等优势已被广泛应用于产品定位、品牌调查、满意度研究等领域, 而在公民科学素质调查中的应用尚不多见, 本文采用对应分析方法研究公民未接触过科技馆等科普场馆的原因与公

民文化程度之间的关系, 揭示对应分析的基本使用方法, 为其在公民科学素质调查中的更多应用提供参考。

1 对应分析的主要特点

对应分析 (Correspondence Analysis) 也称为关联分析、R-Q 型因子分析, 是由法国人 Jean Paul Benzerc 于 20 世纪 60 年代创立, 到

收稿日期: 2010-10-18

作者简介: 张磊, 工程师, 供职于天津市科学技术信息研究所情报研究中心, 研究方向为科技管理, Email: tjxye@126.com。

80年代开始逐步兴起的一种多元相依变量统计分析技术。它主要对定类变量或定序变量多维频度表进行分析,探索同一变量的各个类别之间的差异以及不同变量各个类别之间的对应关系^[1]。对应分析在变量数较多或各个变量的类别较多时其优势更加明显,它可以将定类变量或定序变量转变为间距变量,将变量类别间的关系直观地表现于图形中^[2]。

对应分析分为简单对应分析(描述两个分类变量的对应关系)和多重对应分析(描述两个以上分类变量的对应关系)两种。它的使用条件包括:变量是定类或定序变量;行变量类别与列变量相互独立;行列变量构成的交叉列表中不能有0值或负数^[3];不能用于相关关系的假设检验;对极端值敏感;变量之间的关联关系理论上需要 $n-1$ 维图形才能够完全解释所有变量类别差异,其中 n 为所有变量中最小的类别数^[4]。

2 对应分析的应用

2.1 研究对象

科技馆等科普场馆是开展科普工作的基础设施,在科学教育、激发公民对科学技术的关注等方面发挥着重要作用,是影响公民科学素质水平的重要因素之一,但是在2010年天

津市公民科学素质调查中发现,多数被调查者选择2009年从未接触过科技馆等科普场馆,并说明了原因。本文以2010年天津市公民科学素质调查数据为研究样本,研究对象为选择未接触过科技馆等科普场馆并说明了原因的问卷,研究变量为公民的文化程度与原因,旨在了解公民的文化程度与其未接触过科技馆等科普场馆的原因之间的关联关系。

其中文化分为3个类别,初中及以下、高中或中专、大专及以上,分别赋值为1、2、3;原因分为门票太贵、没有时间、交通不便、不了解和内容不吸引人5个类别,分别赋值为1、2、3、4、5。

2.2 分析过程与结果

首先对基础资料按照SPSS数据记录格式进行整理,并进行逻辑纠错检查,然后采用SPSS软件对数据进行对应分析,主要结果如下。

2.2.1 对应分析表

对应分析表是行变量类别与列变量类别的交叉汇总,从表中频度分布看,没有数据是零或负数,满足对应分析的要求,样本总数为1485份。从对应分析表来看,研究样本中文化变量中大专及以上学历的样本最多,原因变量中选择没有时间的样本最多(见表1)。

表1 对应表

文 化	原 因					
	门票太贵	没有时间	交通不便	不了解	内容不吸引人	有效边际
初中及以下	55	176	81	90	46	448
高中或中专	57	285	42	38	51	473
大专及以上	63	367	50	38	46	564
有效边际	175	828	173	166	143	1485

2.2.2 结果汇总表

结果汇总表显示对应分析最终汇总统计量,包括维数、奇异值、惯量、卡方检验值、惯量解释比例等数据。依据对应分析的使用条件,本研究中文化变量的类别数3为最小,因此共获得了两个维度的惯量;卡方值为103.433, $p < 0.05$,说明文化和原因两个变量

存在关联关系;奇异值为惯量的平方根,惯量表示每个维度对各个变量类别之间差异的解释量,解释比例即每一维度惯量在总特征值中的比例。在结果汇总表中,第一维度的解释比例为97.8%,说明第一维度能够解释所有变量类别差异的97.8%;第二维度的解释比例为2.2%,说明第二维度仅能够解释所有变量类

别差异的 2.2%。由此可以得出结论,二维图形可以完全解释两个变量间的所有信息,并且在观察时以第一维度为主(见表 2)。

2.2.3 概述行点表和概述列点表

概述行点表显示的是文化变量在两个维度中的分值、坐标值以及对两个维度的贡献率,从中可以看出文化 3 个类别在第一维度上分散得比较好,概述列点表显示的是原因变量在两个维度中的分值、坐标值以及对两个维度的贡献率,可以看出 5 个类别原因在第一维上分散得比较好(见表 3、4)。

2.2.4 对应分析图

对应分析图是对应分析中最重要的部分,

对它进行分析可以达到两个目的:首先,处于大致相同区域内的不同变量的类别之间有关联,因此可以更直观地反映行变量与列变量之间的关系;其次,在某个维度上靠得较近的同一样本的不同类别区别较小,借此可以分析变量分类的合理性。

从对应分析图可看出,大专及以上学历与没有时间距离较近,高中或中专与门票太贵和内容不吸引人距离较近,初中及以下与不了解和交通不便距离较近,从而说明以上 3 组之间联系紧密。

综合以上分析可以得出如下结论:大专及以上学历文化程度的被调查者未接触过科技馆等

表 2 摘要

维数					惯量比例		置信奇异值	
	奇异值	惯量	卡方	Sig.	解释	累积	标准差	相关
1	0.261	0.068	—	—	0.978	0.978	0.027	0.019
2	0.039	0.002	—	—	0.022	1.000	0.026	—
总计	—	0.070	103.433	0.000a	1.000	1.000	—	—

a. 8 自由度

表 3 概述行点^b

文化		维中的得分			贡献				
					点对维惯量		维对点惯量		
	质量	1	2	惯量	1	2	1	2	总计
初中及以下	0.302	-0.773	0.033	0.047	0.690	0.009	1.000	0.000	1.000
高中或中专	0.319	0.260	-0.271	0.007	0.082	0.599	0.860	0.140	1.000
大专及以上	0.380	0.396	0.201	0.016	0.228	0.392	0.963	0.037	1.000
有效总计	1.000	-	-	0.070	1.000	1.000	-	-	-

b. 对称标准化

表 4 概述列点^b

原因		维中的得分			贡献				
					点对维惯量		维对点惯量		
	质量	1	2	惯量	1	2	1	2	总计
门票太贵	0.118	-0.060	-0.141	0.000	0.002	0.060	0.549	0.451	1.000
没有时间	0.558	0.386	0.072	0.022	0.317	0.075	0.995	0.005	1.000
交通不便	0.116	-0.706	0.200	0.015	0.222	0.120	0.988	0.012	1.000
不了解	0.112	-1.030	0.050	0.031	0.454	0.007	1.000	0.000	1.000
内容不吸引人	0.096	-0.109	-0.547	0.001	0.004	0.738	0.211	0.789	1.000
有效总计	1.000	-	-	0.070	1.000	1.000	-	-	-

b. 对称标准化

(下转第 51 页)

都由自然辩证法教师、科学技术史教师和科学教育教师等构成)等,这些都直接或间接制约了研究生综合素质培养。为此,需要改革目前的研究生教育体制、借鉴外国的先进经验、开展 STS 教育的理论与实践研究,以便有效地培养研究生的综合创新能力,为实现科教兴国和人才强国战略、建设创新型国家做贡献。

参考文献

- [1] 龚育之. 自然辩证法在中国[M]. 北京: 北京大学出版社, 1996: 196.
- [2] 岛原健三. 日本化学家获诺贝尔奖的社会背景[J]. 北京化工大学学报, 2007(3).
- [3] 张明国. 耗散结构理论与青年科技素质结构[J]. 哈尔滨师范学院学报, 1994(3).
- [4] 张明国. 技术文化论[M]. 北京: 同心出版社, 2004.

(上接第 35 页)

科普场馆的主要原因为没有时间; 高中或中专文化程度的被调查者未接触过科技馆等科普场馆的主要原因包括门票太贵和内容不吸引人; 初中及以下文化程度的被调查者未接触过科技馆等科普场馆的主要原因包括交通不便和不了解(见图 1)。这一结果提示我们在改进科普场馆的过程中可以根据不同文化程度的公众采取不同的策略。

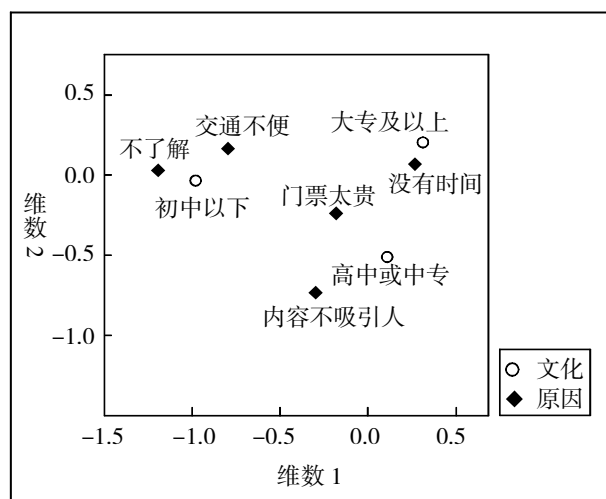


图 1 对应分析图

3 结论

与卡方检验、相关分析等统计方法相比, 对应分析是一种更为有效的对定类或定序变量进行分析的工具, 它对此类数据的分析更准确、深入和直观, 能够更加清晰地表示变量类别之间的本质关联, 特别适合用于各类社会调查、市场调研等领域。本文结合具体案例对对应分析过程进行了研究, 随着研究的不断深入, 其在公民科学素质调查中的应用将会得到更多体现。

参考文献

- [1] Bartholomew D, Steele F, Moustaki I, Galbraith J. The Analysis and Interpretation of Multivariate Data for Social Scientists [M]. London: Chapman & Hall, 2002: 8.
- [2] 郭志刚. 社会统计分析方法——SPSS 软件应用[M]. 北京: 中国人民大学出版社, 1999: 459-460.
- [3] 汪顺玉. SPSS 对应分析在外语学习需求分析中的应用[J]. 重庆邮电学院学报, 2004(5): 153-155.
- [4] 张文彤. SPSS 11.0 统计分析教程(高级篇)[M]. 北京: 希望电子出版社, 2002: 202-211.