

文章编号: 1000-8020(2017)06-0930-05

·调查研究·

基于潜在剖面模型的广州市成年居民 健康知识知晓模式

陈小娜 赵丽¹ 郜艳晖 张瑛 喻嘉宏 李丽霞²

广东药科大学公共卫生学院, 广州 510000



摘要: 目的 基于潜在剖面模型识别广州市居民健康知识知晓模式, 从多维度识别异质性人群的特征。方法 通过分层整群抽样选取广州市 1179 名居民进行健康知识和健康行为的问卷调查, 采用潜在剖面模型依据健康知识得分对人群分类, 并与传统聚类方法进行比较。分析不同健康知识知晓模式的人口学特征和健康行为。结果 (1) 潜在剖面分析将被调查居民分成高、中、低三类健康知识知晓模式, 分别占 62.2%、27.4% 和 10.4%, 其分类效果优于传统聚类法。(2) 除性别外, 健康知识知晓高、中、低三类人群的年龄($\chi^2 = 10.431, P = 0.034$)、文化程度($\chi^2 = 49.510, P < 0.001$)、职业类型($\chi^2 = 20.781, P < 0.001$) 差异均有统计学意义。较低水平健康知识知晓人群以 18~44 岁、初中及以下文化程度、体力型劳动者居多。(3) 除吸烟外, 健康知识知晓高、中、低三类人群在是否吃早餐($\chi^2 = 25.763, P < 0.001$)、睡眠时间 > 7 h($\chi^2 = 7.483, P = 0.024$)、是否运动量足够($\chi^2 = 8.317, P = 0.016$) 和是否健康体检($\chi^2 = 6.909, P = 0.032$) 等方面差异有统计学意义。结论 潜在剖面模型应用到健康教育领域, 可有效识别出不同健康知识知晓情况的异质亚群体, 揭示健康教育的重要人群和内容。

关键词: 潜在剖面模型 健康知识 健康行为 成年居民

中图分类号: R193

文献标志码: A

Health knowledge awareness in Guangzhou adult residents based on latent profile model

Chen Xiaona, Zhao Li, Gao Yanhui, Zhang Ying, Yu Jiahong, Li Lixia

School of Public Health, Guangdong Pharmaceutical University, Guangzhou 510000, China

Abstract: Objective To explore the application of latent profile model in classification of health knowledge awareness in Guangzhou residents, and identify the characteristics of the heterogeneous population from multi dimensions. **Methods** By stratified cluster sampling, health knowledge and health behavior among 1179 residents in Guangzhou City were investigated. Latent profile model was used to population classification based on health knowledge scores, and compared with traditional clustering method, the demographic and health behaviors with different levels of health knowledge were analyzed. **Results** Community residents were divided into three classes of health knowledge awareness model crowd by latent profile model, which accounted for 62.2%, 27.4% and 10.4%, respectively, was superior to the traditional cluster method. There

基金项目: 国家自然科学基金(No. 71573059); 广东省科技厅社会发展领域科技计划项目(No. 2013B021800269)

作者简介: 陈小娜, 女, 硕士研究生, 研究方向: 疾病预防控制, E-mail: chanyil777@163.com

¹ 普瑞盛医药科技开发有限公司

² 通信作者: 李丽霞, 女, 教授, 研究方向: 流行病学与卫生统计学, E-mail: llx19@163.com

were significant differences in the demographic characteristics of different health knowledge subgroups, people who is younger, low education, manual worker with weak health knowledge. And the result showed that the higher level of health knowledge awareness with the better health behavior. **Conclusion** The application of latent profile model is extended to the field of health education. It can identify the heterogeneous subgroups of different health knowledge awareness effectively, which can indicate the key of health education programmes.

Key words: latent profile model, health knowledge, health behavior, adult residents

健康教育是提高居民健康知识和行为水平,促进健康和实施自我保健的有效手段。由于生理状况、生活条件、文化知识结构、城乡等差异,人群对健康教育的需求和提供服务的要求不同^[1-2]。因此充分了解所关注人群的健康知识知晓状况,开展以“居民健康教育需求”为导向的健康教育是健康教育领域的重要内容。目前关于居民健康知识知晓状况研究多采取划界分数方法计算知晓率,计算各个条目的总分,人为确定一个分界点,例如:答对条目的60%认为该观测为知晓^[3],分界点选择不同,结果会有很大差异,导致不同研究结果的不可比性,且该方法无法从多方面角度同时概括评价知识知晓情况。基于人群在健康知识知晓方面具有多维度特征即存在异质性方面的考虑,对多个方面的健康知识条目得分可采用传统聚类分析方法实现对异质性人群分类,但传统聚类方法基于观测间距离进行聚类,距离定义方法有多种,最终结果的选择需要主观判断,容易给研究者造成一定困扰^[4]。

近年来使用分类潜变量(categorical latent variable)统计模型研究人群不可见异质性逐渐成为许多研究领域的热点,潜在剖面模型(latent profile analysis, LPA)作为一种基于混合模型、通过找到具有相似性的个体所属的潜在类别而实现分类^[5]。该方法的优点是聚类的类别数的选择标准不再主观,不需要人为选择分类的界值点,潜在剖面分析根据后验概率的大小决定观测的分类,结果更加科学、客观^[4]。本研究对广州市1179名居民进行健康知识知晓情况调查,采用潜在剖面模型根据居民在四方面健康知识题目上的作答模式来判断个体的潜在特征来分类,并分析不同健康知识知晓模式人群的人口学特征及健康行为表现,为针对重点人群和内容开展健康教育提供科学依据,为相关研究领域提供统计分析方法支持。

1 调查对象与方法

1.1 调查对象

于2011年采用多阶段分层抽样方法,第一阶段以地理位置和经济因素为分层标准,抽取广州市七个区;第二阶段以单纯随机抽样方法按40%比例随机抽取街道,共抽取42条街道;第三阶段以偶遇抽样方法抽取受访居民,即在抽取街道中把偶遇方式抽取符合入选条件的该街道常住居民作为调查对象。此次被调查对象要求为年满18周岁、在广州居住至少满1年的社区居民,共获得有效样本1179例。

1.2 调查内容

主要调查内容包括人口学特征、健康知识、健康行为等方面。人口学特征包括性别、年龄、文化程度、职业等方面。健康知识包括合理膳食、健康生活方式、正确预防常见传染病和发生火灾时正确逃生方式四个方面,每个方面有5个备选答案(可看作5道单选题),表示5个知识点,共20题,答对1题记为1分,答错记为0分。健康行为包括5个方面,分别为近3个月是否经常吃早餐;近3个月平均每天睡眠时间是否超过7小时;近3个月运动量是否达到每周2次,每次30分钟;近2年是否到医院进行过健康体检和近6个月累计吸烟量是否超过100支。健康知识、行为调查内容均来自卫生部于2008年1月4日发布的《中国公民健康素养 基本知识及技能》^[6]。

1.3 统计学方法

采用Epidata软件进行数据录入,Mplus 7.4进行潜在剖面分析,SAS 9.4进行K-means聚类、 χ^2 检验。

1.3.1 潜在剖面模型进行健康知识知晓模式识别 潜在剖面模型分析群体的不可见异质性时,模型利用共同的一个分类潜变量来解释多个外显定量变量之间的复杂关联。该法基于模型在概率基础上对人群进行分类,即个体以一定的概率归属于各个类别,最终个体将被分到最大后验概率的那个类别^[7]。本研究以4个方面健康知识得分

作为可测(外显)变量,建立潜在剖面模型对广州市居民的健康知识知晓情况进行模式识别。研究中逐步增加潜在类别的数目,进行各模型的参数估计,同时进行适配性检验以选择最佳模型。潜在剖面模型的参数估计使用最大似然法(robust maximum likelihood, MLR)。模型评价采用赤池信息量准则(Akaike information criterion, AIC)、贝叶斯信息准则(Bayesian information criterion, BIC)和调整 BIC 等信息统计量指标,其中 AIC 考虑了模型中参数个数, BIC 则同时考虑了参数个数和样本量 N 来评价模型拟合优度。信息统计量值越小提示模型拟合越好。Lo-Mendell-Rubin 似然比检验(LMRT)用于对分别含 c 类和 $(c-1)$ 类模型拟合差异进行比较,如果 $P < 0.05$ 则提示含 c 类

的模型拟合效果更好^[8]。通过熵(Entropy)来反映模型的分类型质量,Entropy 值越接近 1 表明分类越明确。

1.3.2 K-means 聚类 根据居民在四个方面健康知识题目上的作答情况使用传统聚类方法(K-means 聚类)划分不同健康知识知晓水平亚人群,并将分类结果与潜在剖面分析结果做比较。

2 结果

2.1 调查人群的基本情况

调查共收集 1179 份有效问卷,其中 4 人缺失学历信息,1 人缺失职业信息。由表 1 可见,在 1179 例受访居民中,男女比例接近 1:1。平均年龄 42.61 岁,按年龄分组^[9],18~44 岁超过 50%。

表 1 广州市成年居民人口学特征

人口学特征	n	构成比/%	人口学特征	n	构成比/%
性别			年龄/岁		
男	589	50.0	18~44	677	57.4
女	590	50.0	45~59	283	24.0
文化程度			≥60	219	18.6
小学及以下	233	19.8	职业类型		
初中	359	30.6	脑力型	429	36.4
高中/中专/技校	311	26.5	体力型	370	31.4
大专及以上	272	23.2	其他	379	32.2

2.2 潜在剖面分析识别居民健康知识知晓模式

以四个方面健康知识得分作为 4 个连续型显变量,采用 LPA 方法对广州市居民进行分类,为了选择最优的潜在类别模型,依次拟合 1 至 5 类 LPA 模型拟合的结果见表 2。从拟合信息可以看出,虽然随着类别数的增加 AIC、BIC 信息统计量值变小,但 LMRT 值、 P 值提示模型 3 拟合效果较好,综合考虑 Entropy 和模型的简洁性,提示潜在类别数为 3 类的模型较佳,模型对应的条件均值和类别概率见表 3,三个类别在四个方面的平均得分均不相同且呈现一定的规律,第一类(class 1)四题平均得分普遍较低,第二类(class 2)得分高于一类,第三类(class 3)得分最高,因此结合实际情况可将社区居民分为低、中和高三种健康知

识知晓模式,三类人群所占比例分别 10.4%、27.4% 和 62.2%。本文同时采用传统的 K-means (快速)聚类方法将人群分为 3 类,三类观测的比例分别为 8.5%、14.6% 和 76.9%,总体变异中仅 37.7% 被类成分解释,提示不能很好地解释群体异质性原因,三类人群的各项健康知识得分见表 3。

2.3 不同健康知识知晓模式人群的人口学特征

由表 4 可见,除性别外,健康知识知晓高、中、低三类人群的年龄($\chi^2 = 10.431, P = 0.034$)、文化程度($\chi^2 = 49.510, P < 0.001$)、职业类型($\chi^2 = 20.781, P < 0.001$)差异均有统计学意义,较低水平健康知识知晓人群以 18~44 岁、初中及以下、体力型劳动者居多。

表 2 广州市成年居民健康知识的潜在剖面分析模型拟合结果

潜在类别数	AIC	BIC	调整 BIC	LMRT	P 值	熵
1 类	13609.590	13650.170	13624.759			
2 类	12872.311	12938.253	12896.960	747.279	0.0000	0.879
3 类	11878.098	11969.402	11912.228	1004.213	0.4269	0.997
4 类	9599.761	9716.426	9643.370	2288.338	0.4141	0.997
5 类	9446.033	9588.061	9499.123	163.727	0.0801	0.996

注: AIC: 赤池信息量准则; BIC: 贝叶斯信息准则; LMRT: Lo-Mendell-Rubin 似然比检验

表 3 潜在剖面分析(3 类模型)、K-means 聚类法中各类的条件均值及类别概率

健康知识	潜在剖面分析			K-means 聚类		
	低	中	高	低	中	高
合理膳食	2.605	3.570	4.159	1.545	3.337	4.182
健康生活方式	2.750	4.000	5.000	3.716	3.828	4.710
传染病预防	3.387	4.056	4.565	2.920	3.715	4.567
火灾避险	2.553	3.093	3.540	3.270	1.273	3.707
类别概率	0.104	0.274	0.622	0.085	0.146	0.769

2.4 不同健康知识知晓模式人群的健康行为

由表 5 可见,除在吸烟行为差异无统计学意义外,在是否吃早餐($P < 0.001$)、睡眠时间 > 7 h ($P < 0.05$)、是否运动量足够($P < 0.05$)和是否健康体检($P < 0.05$)等方面表现为健康知识知晓水平越高的人群健康行为形成率越高。

3 讨论

目前潜在剖面模型多用于心理及教育学领

表 4 不同健康知识知晓模式人群的人口学特征 [$n(r/\%)$]

人口学特征	健康知识知晓模式			χ^2 值	P 值
	低	中	高		
性别					
女	58(47.2)	163(50.5)	369(50.3)	0.459	0.795
男	65(52.8)	160(49.5)	364(49.7)		
年龄/岁					
18~44	87(70.7)	180(55.7)	410(55.9)	10.431	0.034
45~59	18(14.6)	82(25.4)	183(25.0)		
≥ 60	18(14.6)	61(18.9)	140(19.1)		
文化程度					
小学及以下	31(25.4)	85(26.4)	117(16.0)	49.510	< 0.001
初中	52(42.6)	109(33.9)	198(27.1)		
高中/中专/技校	23(18.9)	77(23.9)	211(28.9)		
大专及以上	16(13.1)	51(15.8)	205(28.0)		
职业类型					
脑力型	45(36.6)	100(31.1)	284(38.7)	20.781	< 0.001
体力型	54(43.9)	112(34.8)	204(27.8)		
其他	24(19.5)	110(34.2)	245(33.4)		
合计	123(10.4)	323(27.4)	733(62.2)		

表 5 不同健康知识知晓模式人群的健康行为表现 [$n(r/\%)$]

健康行为	健康知识知晓模式			χ^2 值	P 值
	低	中	高		
经常吃早餐					
是	81(65.9)	262(81.1)	622(84.9)	25.763	< 0.001
否	42(34.1)	61(18.9)	111(15.1)		
睡眠时间 > 7 h					
是	89(72.4)	237(73.4)	586(79.9)	7.483	0.024
否	34(27.6)	86(26.6)	147(20.1)		
运动量足够					
是	74(60.2)	208(64.4)	520(70.9)	8.317	0.016
否	49(39.8)	115(35.6)	213(29.1)		
参加健康体检					
是	49(39.8)	160(49.5)	385(52.5)	6.909	0.032
否	74(60.2)	163(50.5)	348(47.5)		
吸烟					
是	86(69.9)	238(73.7)	564(76.9)	3.435	0.179
否	37(30.1)	85(26.3)	169(23.1)		

域,如苏斌原等^[10]依据大学生心理行为问题应用潜在剖面分析将人群划分为三个亚类:风险组、困

扰组和健康组,发现潜在剖面分析优于传统的量表划界分类方法。在公共卫生及预防医学领域研

究中可测变量为连续变量的情况比较常见,潜在剖面模型也有很多成功的应用,FITZPATRICK等^[11]利用四个临床指标将822名非裔美国青少年分为低、中、高三类代谢综合征发病风险亚人群。MARTINSON等^[12]依据家庭饮食、生活习惯采用潜在剖面模型定义不同肥胖风险家庭类型并提出相应的公共卫生干预措施。

本研究应用潜在剖面模型将居民分成高、中、低三类健康知识知晓水平亚人群,大部分人知识知晓水平较高(62.2%),但结果显示三类人群伤害类知识即火灾避险方面知识掌握均不理想,建议加强伤害类健康宣传工作。人口学分析结果显示,青年人、低学历者、体力型劳动者健康知识掌握较薄弱,这些人群应该列为健康教育的重点对象。此外,本研究还显示健康知识知晓水平越高,健康行为表现越佳,健康知识与健康行为关联密切,证实居民的健康知识在一定程度上会影响其健康行为,凸显了健康教育的重要性。

聚类分析是探讨群体异质性的主要方法,传统的聚类分析(如K-means、系统聚类)主要的障碍是对类的结构和内容很难给出一个统一的定义^[13],分类结果常受主观因素影响。而与传统聚类法相比,潜在剖面分析将对聚类数的选择转为对模型的选择,通过统计量客观地选择合适的模型,衡量分类的准确性和有效性,以保证组间异质和组内同质的最大化,使分类更有客观依据。课题组前期对潜在剖面模型和系统聚类法的聚类效果进行了模拟研究^[4],结果显示潜变量为两类和三分类,均值或方差变化时,潜在剖面分析的平均错分率均低于系统聚类法。而本文中采用K-means聚类法将人群同样分为三类时,总体变异中只有37.7%被类成解释,说明使用该聚类方法将人群聚为三类不能很好地解释群体异质性的原因。此外,潜变量分析方法相对于传统聚类法还有数据无需标准化、可纳入协变量及不受可测变量间等方差、零相关约束等优点^[14]。潜在剖面模型的应用也需要注意几点,当研究中可测变量为定性变量时,应改用潜在类别分析。其次,当这种基于最大后验概率的归类看作外显变量进入后续的回归分析时,容易忽视分类的测量误差,常导致潜类别变量与其他变量关系的低估,研究者应用时需要考虑是否采用纳入式分类分析法^[15]。

参考文献

- [1] 陈跃辉,冉茂琴,徐凡凡,等.四川省居民健康教育服务供给与需求的定性研究[J].现代预防医学,2016,43(9):1626-1629,1633.
- [2] 谭秋蓉,申向科.浅谈提高居民对社区健康教育满意度的干预措施[J].医学信息,2010(1):124-125.
- [3] 于雪静.滕州市农村居民健康知识知晓情况调查及健康教育策略研究[D].济南:山东大学,2008.
- [4] 赵丽,李丽霞,周舒冬,等.潜在剖面分析和系统聚类法比较的模拟研究[J].心理发展与教育,2013,29(2):206-209.
- [5] NEELY-BARNES S. Latent class models in social work [J]. Soc Work Res, 2010, 34(2): 114-121.
- [6] 中华人民共和国卫生部.中国公民健康素养:基本知识与技能(试行)[J].中国实用乡村医生杂志,2008(5):1-2.
- [7] 邱皓政.潜在类别模型的原理与技术[M].北京:教育科学出版社,2008:39-41.
- [8] WANG Jichuan, WANG Xiaoqian. Structural equation modeling: applications using mplus [M]. Beijing: Higher Education Press, 2012: 293.
- [9] 李立明,饶克勤,孔灵芝,等.中国居民2002年营养与健康状况调查[J].中华流行病学杂志,2005,26(7):478-484.
- [10] 苏斌原,张洁婷,喻承甫,等.大学生心理行为问题的识别:基于潜在剖面分析[J].心理发展与教育,2015,31(3):350-359.
- [11] FITZPATRICK S L, LAI B S, BRANCATI F L, et al. Metabolic syndrome risk profiles among African American adolescents: national health and nutrition examination survey, 2003-2010 [J]. Diabetes Care, 2013, 36(2): 436-442.
- [12] MARTINSON B C, VAZQUEZBENITEZ G, PATNODE C D, et al. Obesogenic family types identified through latent profile analysis [J]. Ann Behav Med, 2011, 42(2): 210-220.
- [13] 万崇华,罗家洪.高级医学统计学[M].北京:科学出版社,2014:126.
- [14] MAGIDSON J, VERMUNT J K. Latent class models for clustering: a comparison with K-means [J]. Can J Marketing Res, 2002, 20(1): 36-43.
- [15] BRAY B C, LANZA S T, TAN X. Eliminating Bias in classify-analyze approaches for latent class analysis [J]. Struct Equ Modeling, 2015, 22(1): 1-11.

收稿日期:2016-09-13