

• 技术与方法 • doi:10.3969/j.issn.1671-8348.2014.30.023

## 趋势面分析法在艾滋病地理分布研究中的应用\*

王璐<sup>1</sup>, 罗静<sup>2</sup>, 邓丹<sup>3</sup>, 杨书<sup>4△</sup>, 张强<sup>5</sup>

(1. 荆州市第一人民医院质控办公室, 湖北荆州 434000; 2. 成都生物制品研究所质量运营部, 成都 610023;  
3. 重庆医科大学公共卫生与管理学院 400016; 4. 成都医学院公共卫生系, 成都 610083;  
5. 四川大学卫生统计教研室, 成都 610041)

**摘要:**目的 了解某市艾滋病疫情的地理分布特征。方法 以多元回归分析理论为基础, 构建趋势面回归数学模型, 将不同区县的经纬度坐标及艾滋病感染数据代入模型, 并根据模型方程计算结果, 通过 ArcGIS 软件, 绘制趋势面分析图。结果 某市艾滋病感染率趋势面回归数学模型有统计学意义 ( $P < 0.05$ ), 其拟合优度为 53.18%。存在异常残差值的地区, 应作为重点研究对象, 探讨这些地区可能存在的某种保护因素或危险因素。结论 该方法可用于分析某市艾滋病疫情的地理分布系统变异和局部变异情况, 可为当地艾滋病的疫情防控提供一些线索。

**关键词:** 获得性免疫缺陷综合征; 地理位置; 回归分析; 趋势面分析

中图分类号: R181.3

文献标识码: A

文章编号: 1671-8348(2014)30-4050-03

### Application of trend surface analysis method in geographical distribution of HIV/AIDS epidemic research\*

Wang Lu<sup>1</sup>, Luo Jing<sup>2</sup>, Deng Dan<sup>3</sup>, Yang Shu<sup>4△</sup>, Zhang Qiang<sup>5</sup>

(1. Quality Control Office, the First People's Hospital of Jingzhou, Jingzhou, Hubei 434000, China;  
2. Department of Quality Management, Chengdu Institute of Biological Products, Chengdu, Sichuan 0610023, China;  
3. School of Public Health and Management, Chongqing Medical University, Chongqing 400016, China;  
4. Department of Public Health, Chengdu Medical College, Chengdu, Sichuan 610083, China;  
5. Department of Health Statistics, School of Public Health, Sichuan University, Chengdu, Sichuan 610041, China)

**Abstract:** **Objective** To study the geographical distribution of the HIV/AIDS epidemic in C city. **Methods** Based on the theory of multiple regression analysis, trend surface regression mathematical model was constructed. The latitude and longitude coordinates of different districts and data of HIV/AIDS were collected in the model and the results by ArcGIS software were got, the trend surface analysis chart was drawn at last. **Results** HIV/AIDS infection rates trend surface regression mathematical model of C city was statistically significant difference ( $P < 0.05$ ), its goodness-of-fit was 53.18%. Regions existed abnormal residual value and should be the paid attention to explore protective or risk factors in these regions. **Conclusion** The method can be used in the analysis of C city HIV/AIDS epidemic systematic variation in the geographic distribution and local district, and provide some clues for the local epidemic prevention and controlling of HIV/AIDS.

**Key words:** acquired immunodeficiency syndrome; geographical locations; regression analysis; trend surface analysis

近年来, 空间分析的相关方法已逐步引进并应用于描述疾病的空间分布特征及引起疾病发病率变化的影响因素等方面。趋势面分析可以弥补疾病地域分布的统计地图的不足, 将地理空间位置上具有一定分布特征的数据划分为趋势值和剩余值两部分, 剔除局部和随机因素的影响, 能准确地反映出总的变化规律<sup>[1]</sup>。地理信息系统(GIS)是以地理空间数据库为基础, 对有关空间的数据按地理空间位置进行采集、管理、储存、维护、分析和校正, 输出地理信息或以位置为基础信息的计算机系统, 主要功能是以所存储的空间数据与属性数据为基础进行空间分析<sup>[2]</sup>。本研究在搭建好的 ArcGIS 数据库平台基础上, 将某市 40 个区县艾滋病新发感染率作为地理空间上连续性区域分布数据, 应用趋势面分析方法研究某市艾滋病的地理分布特征, 为当地艾滋病的疫情防控研究提供一些线索。

### 1 资料与方法

**1.1 一般资料** 艾滋病疫情数据资料来源于某市疾病控制部

门, 该市各区县常住人口数据来自当地统计年鉴, 地图资料来源于中国国家地理地图, 以此作为数据基础搭建了 ArcGIS 数据库平台。

**1.2 方法** 趋势面分析(trend surface analysis, TSA)是一种多元非线性的统计分析方法, 它以多元回归理论分析为基础, 通过线性函数模型, 运用最小二乘法原理, 通过数学曲面拟合样本数据的统计分析方法, 利用不同阶次趋势面方程模型, 研究多个影响因素对观测变量的影响作用<sup>[3-5]</sup>。

**1.2.1 趋势面方程的建立** 以当地各区县疾病控制部门在地理地图上的经纬度定为各研究点的坐标, 即横坐标( $x_i$ )和纵坐标( $y_i$ )<sup>[6-7]</sup>。其理论模型为:

$$Z = XB + e$$

令  $Z_i$  表示所研究区县在观察点  $X_i(x_i, y_i)$  处的 HIV/AIDS 的新发感染率,  $B$  是趋势面参数矩阵,  $e$  为残差。用回归

\* 基金项目: 国家自然科学基金青年基金资助项目(81001295)。  
研究。 △ 通讯作者, E-mail: lttys-1983@163.com。

作者简介: 王璐(1987-), 住院医师, 硕士, 主要从事流行病学与卫生统计

分析的方法求出回归面  $\hat{Z} = f(x, y)$ , 通过这些区县内艾滋病感染率建立空间变化在总体范围内区域性变化的光滑曲面, 即为趋势面 (trend surface)。其模型方程形式为:

一阶趋势面:  $f(x, y) = b_0 + b_1x + b_2y$

二阶趋势面:  $f(x, y) = b_0 + b_1x + b_2y + b_3x^2 + b_4xy + b_5y^2$

三阶趋势面:  $f(x, y) = b_0 + b_1x + b_2y + b_3x^2 + b_4xy + b_5y^2 + b_6x^3 + b_7x^2y + b_8xy^2 + b_9y^3$

.....

n 阶趋势面:  $f(x, y) = b_0 + b_1x + b_2y + b_3x^2 + b_4xy + b_5y^2 + \dots + b_p y^n$

$$p = \frac{(n+1)(n+2)}{2} - 1$$

其中,  $b_0$  表示常数项,  $b_1, b_2, \dots, b_p$  表示理论偏回归系数。根据  $Z_i(x_i, y_i)$ , 利用最小二乘法原理, 解方程组, 得到趋势面回归系数的估计值  $b_1, b_2, \dots, b_p$ , 将估计值代入方程, 即可得到相应阶次的趋势面回归方程。

**1.2.2 确定趋势面方程的阶次** 根据研究目的和观测对象的地理流行病学特征, 对趋势面方程进行显著性检验, 趋势面函数方程对观测值的拟合程度优劣, 主要取决于回归平方和在总的离差平方和中所占到的比重  $R^2$ , 来表示趋势面函数方程的拟合优度<sup>[8]</sup>。一般而言, 趋势面函数的次数越高, 拟合程度也越高。但并非趋势面方程阶次拟合越高越好, 阶次高的多项式所表示的曲面过于复杂, 有时反而会偏离实际趋势<sup>[9]</sup>。在模型的拟合过程中, 趋势面次数的选择常结合模型检验结果、拟合优度和标准误差的大小、剩余均方等指标<sup>[10]</sup>。

**1.2.3 绘制趋势面分析图** 根据选定的趋势面方程, 可以在 ArcGIS 中绘制等值线图<sup>[11]</sup>。等值线图是在固定趋势值(等高)和条件下, 并将经度和纬度按一定的步长进行取值, 为 Z 值取定一系列的特定值, 设定 X 值, 代入最优趋势面方程, 从而求出 Y 值。再在行政地图上, 绘制出等高线, 即为趋势面方程的等值线图<sup>[12-13]</sup>。可以根据等值线图直观地发现不同地理位置疾病发生的严重程度的变化状态, 并可以观察研究区域中疾病感染高发区和低发区的分布状况及变化形式。

**1.2.4 残差值分析** 残差是指观测值与趋势值之差, 即  $e_i = z_i - \hat{z}_i$ 。在做疾病的趋势面分析研究时, 正残差值所代表的区县提示在扣除了地域空间因素的作用后, 可能存在着促进疾病发生、发展的危险因素; 负残差值所代表的区县提示在扣除了地域空间因素的作用后, 可能存在着抑制疾病发生、发展的保护因素<sup>[12-14]</sup>。得出残差值, 计算正、负残差值的界值, 进行残差值分析, 来区分各区县的局部综合因素的作用。采用下述公式:

$$L^{(\pm)} = \sum_{i=1}^m e_i^{(\pm)} / k$$

式中,  $L^{(\pm)}$  表示正负残差值的界值,  $e_i^{(\pm)}$  表示正负的残差值,  $m$  表示正负残差的个数,  $k = n/2$ , 用  $L_1$  表示正残差值上限,  $L_2$  表示负残差值下限; 各观测点残差值与  $L_1$  或  $L_2$  相比, 如观测点的正残差值大于  $L_1$  或负残差值小于  $L_2$ , 则认为该观测值的残差部分的变异主要是由该区县的局部综合因素的作用。

**1.3 统计学处理** 采用 ArcGIS9.0 软件搭建地理信息系统平台, 完成对数字地图的矢量化, 搭建地理信息系统平台的个体查询、分区查询、时空扩散等功能。采用 SAS9.1 软件进行

统计分析, 对数据进行趋势面函数拟合, 以  $P < 0.05$  为差异有统计学意义。

**2 结 果**

**2.1 趋势面函数方程的确定** 拟合结果如下表 1 所示。从表 1 中的拟合结果, 计算求得的一至四阶趋势面函数方程, 经检验, 均有统计学意义 ( $P < 0.05$ )。由上表可知, 随着方程阶次的增加, 决定系数稍有增加, 三阶和四阶趋势面方程有统计学意义, 同时二阶与三阶和四阶趋势面方程比较, 误差均方根和变异系数都相对的小, 故认为二阶趋势面方程为最优方程, 表明模型的拟合较为适度, 其趋势面方程表达式为:

$$\hat{z} = 1681.785 - 28.614x - 8.679y + 0.120x^2 + 0.088xy - 0.015y^2$$

表 1 2009 年某市艾滋病感染率趋势面分析结果

趋势面模型	Root MSE	F	P	R <sup>2</sup>	变异系数
一阶	0.435 1	17.72	<0.01	0.489 1	62.115 6
二阶	0.434 5	7.72	<0.01	0.531 8	62.035 7
三阶	0.440 6	6.27	0.000 2	0.532 7	62.906 9
四阶	0.440 6	6.27	0.000 2	0.532 7	62.906 8

**2.2 趋势面函数的等值线图 and 趋势面分析图** 根据二阶趋势面方程函数, 在 ArcGIS 软件中画出二阶趋势面等值线图 (图 1) 和趋势面分析图 (图 2)。图 1 中颜色由浅到深表示艾滋病感染率由少到多, 提示艾滋病感染存在高发区和低发区, 高发区域可能存在着潜在的危险因素。整体趋势上病情大致呈现由中西部地区向周围逐渐减少, 由西向东先增加后减少的趋势。某市东北、东南地区发病较少, 西部和中部地区发病相对较多。表明高发地带主要集中在某市西部和中部地区, 这种分布模式在一定程度上反映了空间集聚的存在。以地理位置的经度和纬度分别为自变量 x、y, 艾滋病疫情为因变量进行趋势面分析。如图 2 所示, 结果表明, 2009 年某市艾滋病疫情总体分布在南北方向上呈较明显的倒 U 形, 中间高两边低, 东西方向分布西部地区高于东部地区。趋势面分析结果表明艾滋病在某市西部和中部地区发病较多, 和等值线图结果基本吻合, 表明存在着区域聚集性。

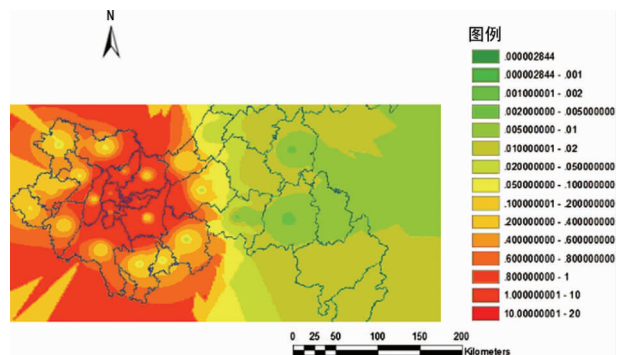


图 1 2009 年某市部分区县艾滋病感染率等值线图

**2.3 残差值分析** 根据所拟合的趋势函数计算得出各观察区县在 2009 年艾滋病感染率的趋势值, 计算出其与观察值之间存在的残差值。根据公式计算出正负残差值界值。其中正残差值有 16 个, 总和为 5.43; 负残差值有 24 个, 总和为 -5.43,

按前述公式求出  $L_1=0.2715$ ,  $L_2=-0.2715$ ; 16 个正残差值中有 5 个高于  $L_1$ , 分别是表示在扣除了地理因素的作用后, 这些区县存在着使 HIV 感染水平相对增加的危险因素; 24 个负残差值中有 9 个小于  $L_2$ , 表示在扣除了地理因素的作用后, 这些区县存在着抑制 HIV 感染水平的保护因素。

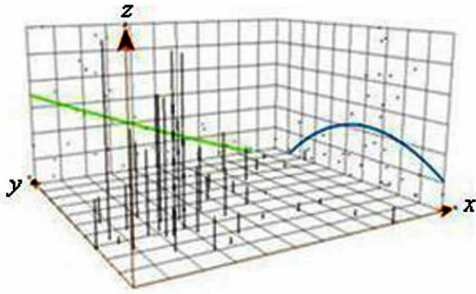


图 2 2009 年某市艾滋病感染率趋势面分析图

### 3 讨论

由趋势面分析的结果可见, 艾滋病的趋势方程的拟合优度为 53.18%, 表明所取的趋势面方程能说明实际变异的 53.18%, 提示地理环境因素在艾滋病疫情发展中可能起着重要作用<sup>[15]</sup>, 并且存在着除地理因素以外可能对应变量有重要作用的其他因素。趋势面分析法能排除局部不稳定因素的干扰, 以三维图像直观展示艾滋病疫情在大范围上的变化趋势, 并显示出局部变异区域。结果显示, 艾滋病疫情存在东部感染率低、中西部感染率高、南北方向呈较明显倒 U 形分布的规律。综合研究结果, 得出造成某市艾滋病聚集区域的可能原因有 (1) 经济因素: 某市中部地区为该市的经济、商业和信息中心, 是市都市发达经济圈的重要组成部分, 为该市的经济强区; 而西部地区有一个重型工业带动当地的经济, 这些可能是致使其艾滋病感染率高于周边的主要因素。(2) 人口因素: 中西部地区有两大江流交汇, 为水陆客运的交通枢纽, 人口密集, 流动人口占到 20% 以上; 据其相关卫生部门提供的资料显示, 近几年来该区域内的吸毒人数存在着显著地增加; 而东北大部分地区处于高山地带, 其人口基数较小, 交通不便和人口流动均较低。

发病率的残差值反映实测值与趋势值之间的差异, 如果某地的残差值高于正的残差界值, 则可考虑该地是否存在艾滋病的危险因素; 若低于负的残差界值, 则可考虑该地是否存在抵制艾滋病感染的保护因素。本研究计算出正负残差界值, 若比较研究这些正负残差值地区之间某些因素的差异, 可以为研究艾滋病的危险因素提供一些线索, 并可以反映其防控情况。本次趋势面分析直观地描述了某市艾滋病的地理分布特征, 给关于艾滋病疫情流行因素的地理流行病学研究提供了一些线索。若结合其他统计方法, 对艾滋病的其他发病因素及预防等做进

一步的研究, 将对艾滋病的研究将得到进一步的深入。

### 参考文献:

- [1] 成月佳, 鲍昌俊, 梁祈, 等. 江苏省病毒性肝炎流行趋势 GIS 空间分析[J]. 中国公共卫生, 2011, 27(7): 873-876.
- [2] 郭伦, 刘瑜, 张晶, 等. 地理信息系统: 原理, 方法和应用 [M]. 北京: 科学出版社, 2001: 289-339.
- [3] 薛付忠, 王洁贞, 马吉祥, 等. 疾病空间分布趋势面模型的共线性偏倚及其测量与控制[J]. 中国卫生统计, 2004, 21(2): 81-84.
- [4] 郭凤云, 路紫. 基于空间分析方法的疾病地理研究进展 [J]. 地理信息世界, 2009, 6: 22-27.
- [5] 赵玉婉, 陈坤, 马新源, 等. 大肠癌发病地理特征的趋势面分析[J]. 生物数学学报, 2005, 20(1): 101-106.
- [6] 王晓燕, 沈毅, 陈坤, 等. 趋势面分析法在肺癌死亡率地理分布研究中的应用[J]. 中华流行病学杂志, 2007, 28(6): 608-612.
- [7] Joyce Fan Z, Lackland DT, Lipsitz SR, et al. Geographical patterns of end-stage renal disease incidence and risk factors in rural and urban areas of South Carolina[J]. Health & Place, 2007, 13: 179-187.
- [8] 裘炯良, 郑剑宁, 周健, 等. 趋势面分析法在传染病地理分布研究中的应用[J]. 中国热带医学, 2004, 4(5): 689-691, 702.
- [9] 殷俊华, 李光强, 陈冀. 利用 GIS 挖掘湖南省结核病时空格局 [J]. 中国卫生统计, 2011, 28(6): 677-679.
- [10] 薛付忠, 王洁贞, 张际文, 等. 疾病空间分布趋势面模型阶次确定方法的研究[J]. 山东大学学报: 医学版, 2004, 42(2): 125-130.
- [11] 罗盛, 马俊岭, 陈景武. 恶性肿瘤死亡率地域分布的趋势面分析[J]. 中国卫生统计, 2008, 25(4): 357-360.
- [12] 仇丽霞, 陈利民, 肖琳. 趋势面和残差分析法在研究死亡水平地域分布中的应用[J]. 实用预防医学, 2004, 11(4): 708-710.
- [13] 魏善波, 任汉英. 残差值在趋势面分析中的应用[J]. 中国卫生统计, 1994, 1(1): 52-53.
- [14] 彭仙娥, 史习舜. 应用趋势面分析探索食管癌死亡率的地理分布特征[J]. 海峡预防医学杂志, 2003, 9(2): 66-67.
- [15] 李德云, 邓佳云, 李津蜀, 等. 四川省碘缺乏病趋势面分析模型[J]. 中国地方病学杂志, 2004, 23(4): 344-346.

(收稿日期: 2014-06-15 修回日期: 2014-07-17)

## 2014 年本刊投稿须知

尊敬的广大读者, 本刊一律接受网上投稿, 不再接受纸质和电子邮箱投稿! 请您直接登陆网站 <http://cqyx.journalserv.com/> 进行注册投稿以及稿件查询。咨询电话: 023-63604477。

来稿须将审稿费 50 元通过邮局或支付宝汇至本刊编辑部, 编辑部若未收到审稿费, 稿件将不予处理。

感谢您对本刊工作的支持!