

· 流行病学研究 ·

# 用趋势面分析法研究我国足月 低体重儿的地理分布

温焕新

(流行病学教研室)

洪楠

(卫生统计学教研室)

**提 要** 本文介绍了趋势面分析方法,并应用于我国足月低体重儿地理分布的研究。结果显示,三阶趋势面方程为:  $\hat{Z} = 1772.6089 - 37.5849x - 4.1395y + 0.2626x^2 + 0.1518xy - 0.2326y^2 - 0.000336x^3 - 0.003572x^2y + 0.01077xy^2 - 0.009888y^3$ , 拟合优度(即复相关系数的平方)为86.32%,能解释观察变异的86.32%,可较好地描述足月低体重儿的地理分布规律。足月低体重儿具有明显的地理分布特征,其发生率以北部、东部和东北部较低,西南部较高。提示地理环境等各种因素在足月低体重儿发生中可能起着重要作用。

**关键词** 趋势面分析 出生时低体重婴儿 地理分布 地理流行病学

对疾病地理分布的研究,通常采用疾病地区分布图、条纹统计地图等方法。由于各学者所选用的等级标准不同,以致分级对比误差较大。也有人采用等值线图,但此法受主观因素影响较大<sup>[1]</sup>。以上这些方法,一般只反映疾病总的发病水平,而不能区分疾病地区分布的系统性和局部变异性。趋势面分析(trend-surface analysis)是以多元回归分析理论为基础的一种统计分析方法。它从整体出发,分析疾病地理分布系统和局部变异。趋势分析图是根据最小二乘法原理,剔除局部和随机变异的影响后求得的趋势面,能较准确地反映疾病地区分布总的变化规律性。

作者对中国出生缺陷监测协作领导小组1989年公布的全国足月低体重儿(low birth weight infant, LBWI)发生率<sup>[2]</sup>,进行趋势面分析,研究了我国LBWI发生率的直观地理分布,为探讨病因和制订预防措施提供依据。

## 材 料 与 方 法

### 资料来源

LBWI 发生率来自1986年10月~1987年9月全国围产儿出生缺陷监测结果<sup>[2]</sup>,并按我国地图册,确定各观测点的地理位置(表1)。

### 研究方法

根据各观察点的疾病频率及其相应地理位置,建立二元多项式回归,进行趋势面分析<sup>[1,3]</sup>。

1. 观察点地理位置的确定 对于一个省或地区的地理位置,可在该地区地图上,划出等距纵横线,或者用方格纸代替,人为定出该地区的横坐标  $x_i$  和纵坐标  $y_i$ 。全国范围的地理位置可用纬度  $x_i$  和经度  $y_i$  表示。

2. 趋势面方程的建立 设在观察点  $(x_i, y_i)$  处疾病的频率为  $Z_i (i=1, 2, \dots, n)$ , 则趋势面方程形式为:

$$\text{一阶趋势函数 } \hat{Z}_1 = \beta_0 + \beta_1x + \beta_2y$$

$$\text{二阶趋势函数 } \hat{Z}_2 = \beta_0 + \beta_1x + \beta_2y + \beta_3x^2 + \beta_4xy + \beta_5y^2$$

$$\text{三阶趋势函数 } \hat{Z}_3 = \beta_0 + \beta_1x + \beta_2y + \beta_3x^2 + \beta_4xy + \beta_5y^2 + \beta_6x^3 + \beta_7x^2y + \beta_8xy^2 + \beta_9y^3$$

$$\text{四阶趋势函数 } \hat{Z}_4 = \beta_0 + \beta_1x + \beta_2y + \beta_3x^2$$

$$\begin{aligned}
 & + \beta_4xy + \beta_5y^2 + \beta_6x^3 + \\
 & \beta_7x^2y + \beta_8xy^2 + \beta_9y^3 + \\
 & \beta_{10}x^4 + \beta_{11}x^3y + \beta_{12}x^2y^2 \\
 & + \beta_{13}xy^3 + \beta_{14}y^4 \\
 \text{五阶趋势函数 } \hat{Z}_5 = & \beta_0 + \beta_1x + \beta_2y + \beta_3x^2 \\
 & + \beta_4xy + \beta_5y^2 + \beta_6x^3 + \\
 & \beta_7x^2y + \beta_8xy^2 + \beta_9y^3 + \\
 & \beta_{10}x^4 + \beta_{11}x^3y + \beta_{12}x^2y^2 \\
 & + \beta_{13}xy^3 + \beta_{14}y^4 + \beta_{15}x^5
 \end{aligned}$$

$$\begin{aligned}
 & + \beta_{16}x^4y + \beta_{17}x^3y^2 + \beta_{18}x^2y^3 \\
 & + \beta_{19}xy^4 + \beta_{20}y^5 \\
 & \dots\dots
 \end{aligned}$$

$$\text{K阶趋势函数为 } \hat{Z}_k = \beta_0 + \beta_1x + \beta_2y + \beta_3x^2 + \beta_4xy + \beta_5y^2 + \dots + \beta_k y^k$$

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

根据最小二乘法原理解方程组，得到趋势函数回归系数估计值： $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ 。将回归系数估计值代入方程即得相应阶次的趋势面方程。

3. 对趋势面方程作显著性检验 假设各回归系数为 0,  $H_0: \beta_0 = 0, \beta_1 = 0, \beta_2 = 0,$

$\dots, \beta_p = 0$ 。总离均差平方和  $SS_{\text{总}} = \sum_{i=1}^n (Z_i - \bar{Z})^2,$

其中  $\bar{Z} = \frac{\sum Z_i}{n}$ ,  $SS_{\text{总}}$  反映观察值  $Z_i$  关于  $\bar{Z}$  总的离散程度，自由度为  $n-1$ , 其中  $n$  为观察点数。

$$\text{剩余平方和 } SS_{\text{剩}} = \sum_{i=1}^n (Z_i - \hat{Z}_i)^2$$

$SS_{\text{剩}}$  说明观察值  $Z_i$  偏离趋势面的程度，自由度为  $n-m-1$ , 其中  $m$  为方程中  $\beta$  的项数。

$$\text{回归平方和 } SS_{\text{回}} = \sum_{i=1}^n (\hat{Z}_i - \bar{Z})^2 = SS_{\text{总}} - SS_{\text{剩}}$$

$SS_{\text{回}}$  表明由于  $x, y, \dots, y^k$  的改变所引起趋势面上  $\hat{Z}_i$  的变化，自由度为  $m$ 。

$$\text{计算 } F \text{ 统计量: } F = \frac{SS_{\text{回}}/m}{SS_{\text{剩}}/(n-m-1)}$$

它服从  $F$  分布，自由度为  $(m, n-m-1)$ 。  $F \geq F_{\alpha}(m, n-m-1), P \leq \alpha$ 。

如果  $F \geq F_{0.05}(m, n-m-1), P \leq 0.05$ , 就拒绝  $H_0$ , 说明趋势方程有显著性意义。如果  $F < F_{0.05}(m, n-m-1), P > 0.05$ , 接受  $H_0$ , 则趋势方程无显著性意义。

4. 计算趋势面拟合优度 趋势面方程对观察值拟合得“好”与“不好”，取决于  $SS_{\text{回}}$  在  $SS_{\text{总}}$  中所占的比重。  $SS_{\text{回}}$  越大，拟合程

表 1 观察点地理位置和  
低体重儿发生率

观察点	省市	纬度(x)	经度(y)	发生率(%)
S <sub>1</sub>	西藏	88.8	31.6	85.2
S <sub>2</sub>	青海	96.2	35.6	62.1
S <sub>3</sub>	广西	108.3	23.9	54.2
S <sub>4</sub>	云南	101.8	25.1	53.5
S <sub>5</sub>	贵州	106.4	27.0	47.7
S <sub>6</sub>	陕西	108.4	35.6	46.2
S <sub>7</sub>	四川	103.8	30.2	45.5
S <sub>8</sub>	山西	112.3	37.6	44.1
S <sub>9</sub>	广东	113.3	22.9	39.4
S <sub>10</sub>	安徽	117.6	32.9	38.6
S <sub>11</sub>	福建	117.9	25.9	37.4
S <sub>12</sub>	宁夏	106.0	37.9	37.1
S <sub>13</sub>	甘肃	99.7	37.7	35.6
S <sub>14</sub>	江西	116.0	27.3	34.4
S <sub>15</sub>	天津	117.2	39.2	33.6
S <sub>16</sub>	江苏	119.1	32.9	32.3
S <sub>17</sub>	湖南	111.8	27.4	32.2
S <sub>18</sub>	吉林	126.2	43.5	32.0
S <sub>19</sub>	黑龙江	128.8	48.2	31.5
S <sub>20</sub>	辽宁	122.3	41.5	29.6
S <sub>21</sub>	河北	116.9	39.4	29.5
S <sub>22</sub>	新疆	85.1	41.6	29.2
S <sub>23</sub>	河南	113.5	33.8	28.7
S <sub>24</sub>	内蒙古	112.5	42.0	28.6
S <sub>25</sub>	浙江	120.2	29.1	28.1
S <sub>26</sub>	北京	116.4	39.9	27.5
S <sub>27</sub>	上海	121.5	31.3	26.8
S <sub>28</sub>	湖北	112.3	31.2	26.4
S <sub>29</sub>	山东	118.8	36.3	26.2

度越好； $SS_{回}$ 越小，拟合程度越差。用  $C$  表示趋势面拟合优度，即  $SS$  在  $SS$  中所占的比例（即复相关系数的平方）

$$C = \frac{SS_{回}}{SS_{总}} \times 100\% = \left(1 - \frac{SS_{剩}}{SS_{总}}\right) \times 100\%$$

$C$  越接近100，趋势面拟合越好。

5. 选择趋势面方程 趋势面方程并不是阶次越高越好。阶次高的多项式虽然从逼近观察点数据方面效果较好，但其所表示的曲面过于复杂，有时反而偏离实际趋势，在内插或外推上效果不好。低次趋势面较适用于区域性变化的研究，高次趋势面则更能反映较小范围的变异。一般来说，应根据趋势面方程显著性检验结果、拟合优度和标准误的大小等，结合实际情况，综合考虑，选择适当阶次的趋势函数。

6. 绘制趋势面分析图 在固定趋势值(等高)的件下，设定  $X_i$ ，求出  $Y_i$ 。按照  $X_i$  和  $Y_i$  绘制趋势面分析图。

作趋势面分析，一般需借助于计算机软件来完成。本项研究数据是在 IBM PC/XT 微型电子计算机上，用 SAS 软件包进行整理和分析的。

### 结 果

对表 1 数据进行趋势函数拟合，结果一阶趋势面拟合优度为43.17%，二阶为76.87%，三阶为86.32%和四阶为89.03%。一阶和二阶趋势面的拟合优度都较差，而三阶和四阶却有很大的提高。经检验，均有高度显著性意义（ $P < 0.001$ ）。三阶趋势面拟合优度提高幅度较大，四阶比三阶只增加2.71%，但计算量增加很多。按趋势面阶次从小到大，它们的标准误分别为10.1945, 6.9143; 5.8505和6.1044，三阶趋势面标准误最小。另外，在实际工作中，如果能用较低阶趋势面函数拟合实际资料，则应采用较低阶趋势面函数来拟合。因此，本文选择三阶趋势面函数作分析较为适宜。经运算，得到三阶趋势面方程为  $\hat{Z} = -772,6089 - 37,5849x - 4,1395y + 0,2626x^2 +$

表 2 低体重儿发生率趋势面方差分析

变异来源	自由度	平方和	均方	F值	P值
回归	9	4109.3025	456.0336	13.32	0.0001
剩余	19	650.3279	34.2278		
校正总和	28	4754.6303			

$$C = 86.32\% \quad CV = 15.3792 \quad \text{Root MSE} = 5.8505 \\ Z \text{ Mean} = 38.0414$$

表 3 较体重儿发生率趋势值和剩余值

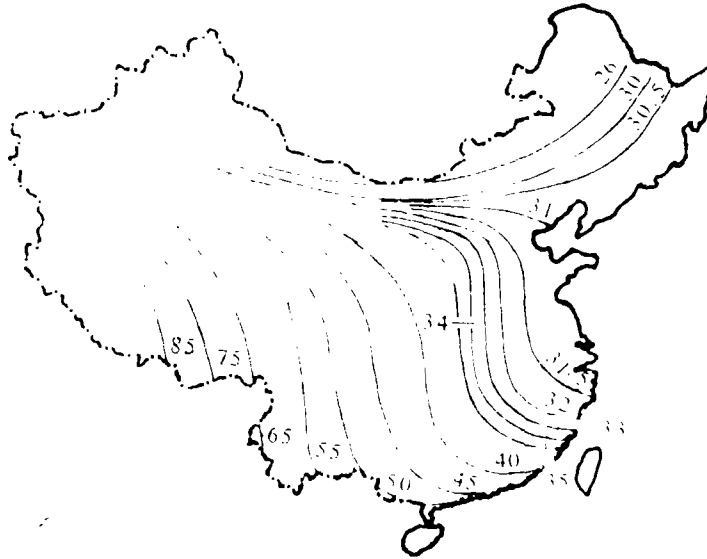
观察点	观察值	趋势值	剩余值
$S_1$	85.2	85.9467	-0.7467
$S_2$	62.1	55.4038	6.6962
$S_3$	54.2	46.1220	8.0780
$S_4$	53.5	55.7272	-2.2272
$S_5$	47.7	45.0444	2.6556
$S_6$	46.2	38.4859	7.7141
$S_7$	45.5	48.1413	-2.6413
$S_8$	44.1	34.0964	10.0036
$S_9$	39.4	44.3987	-4.9987
$S_{10}$	38.6	30.1212	8.4788
$S_{11}$	37.4	34.8483	2.5518
$S_{12}$	37.1	39.2613	-2.1613
$S_{13}$	35.6	43.8097	-8.2097
$S_{14}$	34.4	33.5547	0.8453
$S_{15}$	33.6	31.5402	2.0598
$S_{16}$	32.3	29.1914	3.1086
$S_{17}$	32.2	37.2023	-5.0023
$S_{18}$	32.0	32.0038	-0.0038
$S_{19}$	31.5	30.9625	0.5375
$S_{20}$	29.6	31.0134	-1.4134
$S_{21}$	29.5	31.5275	-2.0275
$S_{22}$	29.2	29.0794	0.1206
$S_{23}$	28.7	33.8970	-5.1972
$S_{24}$	28.6	27.8320	0.7680
$S_{25}$	28.1	29.3186	-1.2186
$S_{26}$	27.5	31.3359	-3.8359
$S_{27}$	26.8	27.6136	-0.8136
$S_{28}$	26.4	35.0881	-8.6881
$S_{29}$	26.2	30.6329	-4.4329

$$t^2 = 0,0039 \quad df = 28 \quad P > 0,90$$

$0.1518xy - 0.2326y^2 - 0.000336x^3 - 0.003572x^2y + 0.01077xy^2 - 0.009888y^3$ 。方差分析见表 2。

按照三阶趋势方程算得各观察点 LBMI 发

生率的趋势值和剩余值，剩余值差异无显著性意义（表 3）。根据观察点地理位置，参考趋势值，绘出趋势面分析图（附图）。



附图 我国足月低体重儿发生率 (%) 趋势面分析

### 讨 论

疾病的地理分布特征往往不仅包含区域（或系统性）因素引起的变化，而且也包含着局部性因素和随机性因素所致的变异。趋势面分析法将每一观察值分解为趋势值和剩余值两部分。趋势值所组成的趋势面表示研究区域的系统性变异，这种变异一般被认为是由于环境总的变化或人群分布的系统性变化所致<sup>[1]</sup>。剩余包括随机误差和局部因素所引起的变化。如果趋势面能完全代表观察值，则剩余是由随机误差所致。在实际工作中，拟合往往是不完全的，所以剩余既包括随机误差，也反映了局部范围的变异。趋势面分析不但能从参差不齐的数据中剔除局部和随机变异的影响，显示大范围区域的系统性变化，研究总的变化规律，而且也能反映出疾病局部地区变异，为病因探讨提供更为重要的线索。

本文应用三阶趋势面函数，对实际资料进行趋势面分析，观察值与预测值的差异无显著

性意义 ( $P > 0.90$ )，说明剩余是由随机误差所致。趋势方程的拟合优度（即复相关系数的平方）为 86.32% ( $P < 0.001$ )，表明所取的趋势面方程能反映实际变异的 86.32%，可较好地描述我国 LBWI 发生率的地理分布规律性。研究结果显示，LBWI 有明显的地区分布特征。在全国范围内，其发生率以北部、东部和东北部较低，西南部最高（附图）。提示地理环境，医疗卫生及经济等因素在 LBWI 发生中可能起着重要的作用。

湖南医学院肿瘤研究室等报道过趋势面分析及其在肿瘤死亡回顾性调查中的应用，国内外一些学者也曾将此法用于其他疾病（如新生儿先天性畸形、脑血管病死亡等）发生率的分析<sup>[4,5]</sup>。用此法对我国 LBWI 发生率进行地理流行病学研究，迄今作者尚未见研究报道。文章所用的数据来自国家七·五重点科技攻关项目结果，资料可靠。但是，仅为 1 年的监察资料，未能确保其稳定性。如果能用历年资料，并且各年各地 LBWI 发生率变化不大时，进

行趋势面分析, 结果将会更可靠。

(本文利用了已公布的中国出生缺陷监测资料, 在此对中国出生缺陷监测协作组的全体成员, 以及对本教研室花肇猛副教授和陈维清老师的帮助, 表示感谢)

**参 考 文 献**

1. 湖南医学院肿瘤研究室, 等. 趋势面分析及其在肿瘤死亡回顾性调查中的应用. 湖南医学院学报 1980; 5(1):18
2. 肖坤则, 等. 中国围产儿素质现状的调查研

- 究. 中华医学杂志 1989; 69(4):185
3. 成都地质学院概率论与数理统计编写小组. 概率论与数理统计. 北京: 地质出版社, 1981: 269~284
4. 赵跃进. 趋势面分析法在疾病监测中的应用. 中国卫生统计学 1988; 5(3):7
5. Kagami M. Regional variance of cerebrovascular mortality in Japan. Ecol Dis 1983; 2(4):277

## STUDY OF GEOGRAPHIC DISTRIBUTION OF LBWI IN CHINA WITH TREND-SURFACE ANALYSIS

Wen Huanxin

(Department of Epidemiology)

Hong Nan

(Department of Medical Statistics)

In this article authors introduced the method of Trend-Surface Analysis and used it to study the geographic distribution of Low Birth Weight Infant at Term(LBWI) in China. The results revealed that the third order trend equation was  $\hat{Z} = 1772.6089 - 37.5849x - 4.1395y + 0.2626x^2 + 0.1518xy - 0.2326y^2 - 0.000336x^3 - 0.003572x^2y + 0.01077xy^2 - 0.009888y^3$ . Its goodness of fit (the square of multiple correlation coefficient) equaled 86.23%. The equation could explain 86.23% of variation of observation and fairly reflect the trend of the geographic distribution for LBWI. The disease showed its obvious characteristic of geographic distribution. The incidence of the disease in the north east and northeast parts in China was lower while higher in the southwest. It was suggested that the factors of geographic environment be responsible for the occurrence of the disease.

**Key words** Low birth weight infant Trend-surface analysis Geographical distribution Geographical epidemiology