

# 基于MATLAB的脑卒中发病环境因素分析

王 娟

(咸阳职业技术学院, 陕西 咸阳 712046)

**摘 要:** MATLAB是一种高度集成的计算机语言, 它具有非常灵活的程序设计流程。本文充分利用MATLAB的优势, 通过对影响脑卒中病的环境因素和相关数据进行分析, 建立了一个多元回归模型, 并对不同气压、气温、湿度下的脑卒中发病率进行预测。

**关键词:** MATLAB; 脑卒中; 发病率; 回归模型

**中图分类号:** R311

**文献标志码:** A

**文章编号:** 94047-(2014)02-047-03

## 0 引言

脑卒中是我国目前常见的多发疾病之一, 它具有发病率高、死亡率高、致残率高、复发率高“四大”特点。中国每年新增脑卒中患者200万, 发病率高达万分之十二, 每年因脑卒中病死亡的人数高达120万, 少数幸存的脑卒中患者也在不同程度上丧失了劳动能力, 已康复的脑卒中患者还有可能复发, 而且每复发一次病情就会加重一次, 给患者带来了巨大的精神压力, 也给患者家庭和社会带来了沉重的经济负担<sup>[1]</sup>。脑卒中的发生是一个漫长的过程, 一旦得病就很难逆转。如何进行脑卒中病的有效预防, 是值得我们深入研究的重大课题。脑卒中病的诱发已经被证实与环境因素(气压、温度、湿度)有关, 对影响脑卒中病的环境因素进行分析, 不仅可以系统掌握疾病发病的规律, 积极探索疾病的预防措施, 进行疾病的风险评估, 而且可以让脑卒中高危人群及时了解自己得脑卒中的风险程度, 进行自我保护<sup>[1, 2]</sup>。

MATLAB是一种可用于数值分析、可视化及编程的高级语言, 它具有强大的科学运算和灵活的程序设计功能<sup>[3, 4]</sup>。本文利用MATLAB软件的优势, 通过建立多元回归模型, 对不同气压、气温、湿度下的脑卒中发病率进行预测, 并对我国某地区的脑卒中发病情况进行了一些研究。

## 1 模型准备

参照2012年全国大学生数学建模竞赛C题数据, 我们对该城市2007年1月至2009年12月的脑卒

中发病病例信息以及相应期间当地的气象资料进行了统计整理, 统计结果如表1所示。

表1 2007年1月-2009年12月某地区  
脑卒中发病率及相应气象数据

| 年 月    | 月相对发病率 | 平均气压     | 平均温度   | 平均湿度   |
|--------|--------|----------|--------|--------|
| 200701 | 0.069  | 1028.226 | 4.503  | 73.226 |
| 200702 | 0.060  | 1020.714 | 8.725  | 70.786 |
| 200703 | 0.076  | 1018.281 | 11.565 | 69.290 |
| 200704 | 0.082  | 1016.543 | 15.370 | 62.567 |
| 200705 | 0.080  | 1008.387 | 22.616 | 61.774 |
| 200706 | 0.079  | 1006.170 | 24.683 | 75.500 |
| 200707 | 0.075  | 1002.971 | 29.394 | 73.323 |
| 200708 | 0.089  | 1004.929 | 29.555 | 69.097 |
| 200709 | 0.094  | 1010.450 | 24.297 | 76.067 |
| 200710 | 0.102  | 1018.955 | 19.339 | 71.645 |
| 200711 | 0.093  | 1024.120 | 12.573 | 62.033 |
| 200712 | 0.101  | 1023.458 | 7.926  | 67.774 |
| 200801 | 0.094  | 1027.229 | 3.129  | 75.448 |
| 200802 | 0.108  | 1027.634 | 2.993  | 64.724 |
| 200803 | 0.099  | 1018.813 | 11.077 | 61.419 |
| 200804 | 0.094  | 1014.613 | 15.630 | 66.533 |
| 200805 | 0.091  | 1008.555 | 21.410 | 66.258 |
| 200806 | 0.080  | 1005.577 | 23.493 | 82.367 |
| 200807 | 0.077  | 1003.671 | 30.174 | 71.323 |
| 200808 | 0.071  | 1005.848 | 28.010 | 77.323 |
| 200809 | 0.068  | 1011.253 | 24.947 | 79.433 |
| 200810 | 0.076  | 1018.190 | 19.906 | 77.194 |
| 200811 | 0.074  | 1023.180 | 12.067 | 75.567 |
| 200812 | 0.068  | 1025.048 | 6.803  | 65.548 |
| 200901 | 0.085  | 1027.668 | 2.961  | 69.645 |
| 200902 | 0.092  | 1019.557 | 8.296  | 79.464 |
| 200903 | 0.081  | 1019.435 | 10.003 | 70.935 |
| 200904 | 0.087  | 1015.683 | 15.907 | 67.300 |
| 200905 | 0.086  | 1012.126 | 21.613 | 61.258 |
| 200906 | 0.081  | 1003.390 | 26.020 | 74.367 |
| 200907 | 0.091  | 1003.826 | 28.439 | 75.129 |
| 200908 | 0.092  | 1005.848 | 27.552 | 81.548 |
| 200909 | 0.084  | 1012.120 | 24.330 | 80.000 |
| 200910 | 0.074  | 1016.842 | 20.442 | 69.548 |
| 200911 | 0.068  | 1023.437 | 11.033 | 77.400 |
| 200912 | 0.079  | 1024.574 | 5.555  | 71.903 |

收稿日期: 2013-12-06

作者简介: 王娟(1978—), 女, 陕西泾阳人, 讲师, 硕士, 研究方向: 统计模型及其应用。

## 2 模型建立

我们借助MATLAB软件对表1中的数据进行分  
析,通过建立数学模型来研究脑卒中发病率与平均  
气温、平均气压、平均湿度之间的关系。

MATLAB中的rstool函数是多元二项式回归函  
数,可用于研究多个变量之间的关系,其调用格式  
为rstool(x,y,'model'),其中,model为以下四种模  
型:

linear (线性):  $y = \beta_0 + \beta_1 x_1 + \dots + \beta_m x_m$  ;

interaction (交叉):  $y = \beta_0 + \beta_1 x_1 + \dots + \beta_m x_m + \sum_{1 \leq j < k \leq m} \beta_{jk} x_j x_k$  ;

quadratic (完全二次):  $y = \beta_0 + \beta_1 x_1 + \dots + \beta_m x_m + \sum_{1 \leq j, k \leq m} \beta_{jk} x_j x_k$  ;

purequadratic (纯二次):  $y = \beta_0 + \beta_1 x_1 + \dots + \beta_m x_m + \sum_{j=1}^m \beta_{jj} x_j^2$  。

这里,我们选用rstool函数建立多元二项式回归  
模型。

首先结合研究<sup>[5]</sup>,建立linear (线性)模型,在  
命令窗口输入以下程序:

```
>> x1=[1028.226 1020.714 1018.281 1016.543
1008.387 1006.170 1002.971 1004.929 1010.450
1018.955 1024.120 1023.458 1027.229 1027.634
1018.813 1014.613 1008.555 1005.577 1003.671
1005.848 1011.253 1018.190 1023.180 1025.048
```

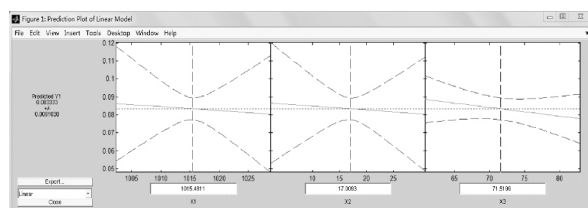


图1 多元回归

单击图框上的Export,将回归参数beta、剩余  
标准差rmse传到MATLAB工作区。在命令窗口输  
入:

```
>> beta
```

```
>> rmse
```

即得:

```
beta = 0.3455 -0.0002 -0.0002 -0.0005
```

```
rmse = 0.0112
```

故得linear (线性)模型为: , 剩余标准差为  
0.0112。

同理可得其它三个模型,通过比较四个模型

```
1027.668 1019.557 1019.435 1015.683 1012.126
1003.390 1003.826 1005.848 1012.120 1016.842
1023.437 1024.574];
```

```
>> x2=[4.503 8.725 11.565 15.370 22.616 24.683
29.394 29.555 24.297 19.339 12.573 7.926 3.129 2.993
11.077 15.630 21.410 23.493 30.174 28.010 24.947
19.906 12.067 6.803 2.961 8.296 10.003 15.907 21.613
26.020 28.439 27.552 24.330 20.442 11.033 5.555];
```

```
>> x3=[73.226 70.786 69.290 62.567 61.774
75.500 73.323 69.097 76.067 71.645 62.033 67.774
75.448 64.724 61.419 66.533 66.258 82.367 71.323
77.323 79.433 77.194 75.567 65.548 69.645 79.464
70.935 67.300 61.258 74.367 75.129 81.548 80.000
69.548 77.400 71.903];
```

```
>> y=[0.069 0.060 0.076 0.082 0.080 0.079 0.075
0.089 0.094 0.102 0.093 0.101 0.094 0.108 0.099 0.094
0.091 0.080 0.077 0.071 0.068 0.076 0.074 0.068 0.085
0.092 0.081 0.087 0.086 0.081 0.091 0.092 0.084 0.074
0.068 0.079];
```

```
>> x=[x1' x2' x3'];
```

```
>> rstool(x,y,'linear')
```

其输出如图1。

的剩余标准差 (比较结果如表2所示),可知  
linear (线性)模型的显著性最好。

表2 四种模型的剩余标准差

| 函 数   | linear | interaction | quadratic | purequadratic |
|-------|--------|-------------|-----------|---------------|
| 剩余标准差 | 0.0112 | 0.0115      | 0.0114    | 0.0114        |

由上,选取linear (线性)模型为回归模型,  
即: ,

其中, 为平均气压, 为平均温度, 为平均湿  
度, 为脑卒中发病率。

## 3 模型应用

下面,我们用此模型对2009年各月的发病率进  
行计算,并与真实值进行比较,结果如表3所示。  
由于所给脑卒中发病病例信息缺失严重,故导致部  
分月份的测算值与实际值出现偏差,但整体还比较  
理想。

表 3 2009年各月脑卒中发病率

| 月 份    | 月相对发病率 (计算值)            | 月相对发病率 (实际值) |
|--------|-------------------------|--------------|
| 200901 | $0.084532 \pm 0.011688$ | 0.085        |
| 200902 | $0.080446 \pm 0.016332$ | 0.092        |
| 200903 | $0.084231 \pm 0.009564$ | 0.081        |
| 200904 | $0.085561 \pm 0.007677$ | 0.087        |
| 200905 | $0.088054 \pm 0.014028$ | 0.086        |
| 200906 | $0.082713 \pm 0.013247$ | 0.081        |
| 200907 | $0.081735 \pm 0.010864$ | 0.091        |
| 200908 | $0.078377 \pm 0.012904$ | 0.092        |
| 200909 | $0.078425 \pm 0.013056$ | 0.084        |
| 200910 | $0.083257 \pm 0.012794$ | 0.074        |
| 200911 | $0.080006 \pm 0.012151$ | 0.068        |
| 200912 | $0.083573 \pm 0.010836$ | 0.079        |

研究中我们还发现：当平均气压为1023、平均气温为11、平均湿度为77时，脑卒中发病率最低。因此，我们建议脑卒中高危人群（老年人）应积极营造这一最优气象组合，预防脑卒中病的发生。而当平均气压为1027、平均气温为3、平均湿度为64时，脑卒中发病率最高。此时，脑卒中高危人群应提高警惕，保护好自我，将脑卒中发病率降到最低。

表 4 2010年任意7天脑卒中发病率

| 日 期      | 平均气压   | 平均气温 | 平均湿度 | 发病率 (计算值) | 发病率 (实际值) |
|----------|--------|------|------|-----------|-----------|
| 20100504 | 1008.7 | 22.8 | 67   | 0.085     | 0.084     |
| 20100506 | 1007.4 | 20.9 | 69   | 0.085     | 0.085     |
| 20100602 | 1014.8 | 20.3 | 65   | 0.085     | 0.082     |
| 20100605 | 1013.3 | 22.4 | 64   | 0.086     | 0.089     |
| 20100607 | 1012.3 | 22.8 | 67   | 0.084     | 0.082     |
| 20100608 | 1012.1 | 22.8 | 65   | 0.086     | 0.085     |
| 20100611 | 1008.9 | 22.1 | 78   | 0.081     | 0.082     |

当然，该模型还可对具体某一天的发病率进行预测，但前提是必须知道这一天的平均气压、平均气温、平均湿度，具体应用如表4所示。

#### 4 结语

本文所建模型比较简单，操作起来比较容易，预测结果较为精确，有较高的可信度。脑卒中发病率跟环境因素（气温、气压、湿度）有着密不可分的联系，所以建议大家关注天气变化和季节变化，积极做好防护工作，将脑卒中带给人们的伤害降到最低。

#### 参考文献

- [1] 焦云芳.脑卒中发病环境因素分析模型[J].湖北民族学院学报(自然科学版),2013(3):51-56.
- [2] 凌佳.基于半参数回归模型的脑卒中发病率影响因素分析[J].数理医药学杂志,2013(1):21-24.
- [3] 冯三营. MATLAB软件在多元统计分析教学中的应用研究[J].洛阳师范学院学报,2010(4):23-26.
- [4] 陈永胜.基于MATLAB和SPSS的非线性回归分析[J].牡丹江大学学报,2009(05):102-104.
- [5] 唐家德.基于MATLAB的商品销售额预测模型[J].商场现代化,2007(25):65-66.

[责任编辑、校对：王军利]

## The Analysis for Environmental Factors of Cerebral Apoplexy on MATLAB

WANG Juan

(Xianyang Vocational Technical College, Xianyang, Shaanxi 712000)

**Abstract:** MATLAB is a highly integrated computer language. It provided a flexible program design process. This paper takes full advantage of MATLAB, analyzes the environmental factors which influence the morbidity of cerebral apoplexy, constructs a multiple regression model, then uses the related data to calculating morbidity of cerebral apoplexy.

**Keywords:** MATLAB; cerebral apoplexy; morbidity; regression analysis