

脑卒中发病环境因素分析及干预

摘要

环境因素已被证实与脑卒中的诱发密切相关，本文从定量角度给出了脑卒中的发病率与环境因素之间的关系，并提出发病预警和干预的建议方案。

问题一要求对发病人群进行统计描述，我们首先对原始数据进行再加工整理，得到不同性别、不同职业及不同年龄段的发病率数据，通过计算发病人群分布的众数、四分位差、偏度、峰度等统计指标，得到了发病人群分布的特征：如发病人群的年龄呈左偏、平峰分布等。

针对问题二，为全面分析发病率与环境因素的关系，我们增加考虑温度差、和湿度差因素，通过建立统计回归模型，得到了脑卒中发病率与气压、温度、湿度、温度差和湿度差之间的量化关系，结果分析显示拟合优度和显著性检验都令人满意。

最后，根据问题一和问题二得到的结果，我们对不同的年龄层次、职业人群，气候条件等提出了相应的预警干预方案。

关键词：众数、四分位数、偏度、峰度、统计回归

问题的重述

脑卒中（俗称脑中风）是目前威胁人类生命的严重疾病之一，它的发生是一个漫长的过程，一旦得病就很难逆转。这种疾病的诱发已经被证实与环境因素，包括气温和湿度之间存在密切的关系。对脑卒中的发病环境因素进行分析，其目的是为了进行疾病的风险评估，对脑卒中高危人群能够及时采取干预措施，也让尚未得病的健康人，或者亚健康人了解自己得脑卒中风险程度，进行自我保护。同时，通过数据模型的建立，掌握疾病发病率的规律，对于卫生行政部门和医疗机构合理调配医务力量、改善就诊治疗环境、配置床位和医疗药物等都具有实际的指导意义。

数据来源于中国某城市各家医院 2007 年 1 月至 2010 年 12 月的脑卒中发病病例信息以及相应期间当地的逐日气象资料。根据题目提供的数据，回答以下问题：

1. 根据病人基本信息，对发病人群进行统计描述。
2. 建立数学模型研究脑卒中发病率与气温、气压、相对湿度间的关系。
3. 查阅和搜集文献中有关脑卒中高危人群的重要特征和关键指标，结合1，2中所得结论，对高危人群提出预警和干预的建议方案。

问题假设

- 1.脑卒中发病因素只考虑气压、温度、湿度、温度差、湿度差，不考虑其它非环境因素；
- 2.在07至10年的相应时间段上，当环境因素稳定时，脑卒中人群的发病率服从正态分布；
- 3.忽略数据统计过程中的微小误差。

符号的假设

M_0 ——脑卒中发病人群年龄分布的众数

M_e ——脑卒中发病人群年龄分布的中位数

Q_L ——脑卒中发病人群年龄分布的上四分位数

Q_U ——脑卒中发病人群年龄分布的下四分位数

V_r ——脑卒中发病人群年龄分布的异众比率

\bar{X} ——脑卒中发病人群年龄分布的均值

Q_D ——脑卒中发病人群年龄分布的四分位数差

α_3 ——脑卒中发病人群年龄分布的偏态系数

α_4 ——脑卒中发病人群年龄分布的峰态系数

β_i ——回归系数

问题分析

问题一：发病人群的统计描述

附表中提供给我们的数据信息量非常多，首先需对原始数据进行处理，得出不同性别，不同年龄段，不同职业脑卒中的人数及发病率，使得我们对脑卒中发病率分布的类型和特点有一个大致的了解。为进一步掌握数据分布的特点和规律，进行深入的分析，就要找到反映数据分布特点的各个代表值。我们考虑对脑卒中发病率从三个方面进行统计描述[4]：一是脑卒中发病人群分布的集中趋势，反映各数据向其中心值靠拢或聚集的程度；二是脑卒中发病人群分布的离散程度，反映各数据远离其中心值的趋势；三是发病人群分布的偏态和峰度，反映数据分布的形状。从这三个方面反映脑卒中发病人群的不同侧面。

问题二：发病率与气温、气压、相对湿度间的关系

在不考虑脑卒中发病的其它因素的前提下，脑卒中的发病率与环境因素密切相关，与气温、气压、相对湿度的变化紧密联系。这些环境因素推动或制约着脑卒中的发病情况。我们考虑对附表数据进行处理，得出2007-2010年中1-12月的脑卒中发病率、平均气压、平均气温、平均相对湿度、气压差、气温差和湿度差数据，然后应用多元回归统计的方法得出它们之间的定量关系。

问题三：查阅和搜集文献中有关脑卒中高危人群的重要特征和关键指标，结合1,2中所得结论，对高危人群提出预警和干预的建议方案。

模型的建立与求解

一. 发病人群的统计描述

(一) 脑卒中发病数据统计

为了对脑卒中发病人群进行深入地统计分析，我们首先对附表数据进行统计处理，得出07至10年脑卒中发病人群中，男性33298人，女性28462，分别占总人数的54%和46%。进一步统计出不同性别各年龄段脑卒中的发病人数(见表1和图1)。

表1 不同性别各年龄段脑卒中的发病人数

年龄段	0~10	11~20	21~30	31~40	41~50	51~60	61~70	71~80
男	71	33	144	579	1969	5088	8600	11458
女	123	27	121	295	1168	3604	6286	10093

表1中数据的折线图见图1

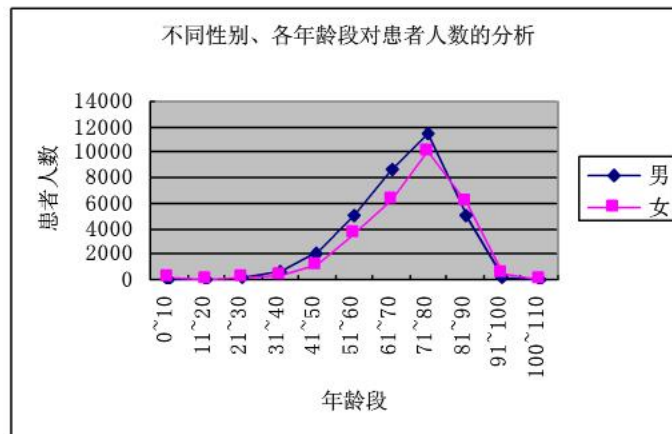


图 1 不同性别各年龄段脑卒中的发病人数

由图 1 可以看出男性和女性的脑卒中发病主要人群都是集中在 50 岁至 90 岁之间。同时我们得到不同年龄段和不同职业脑卒中发病人数（分别见表 2、表 3）。

表 2 07-10 年不同年龄段脑卒中发病人数

年龄段	0-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90	91-100	100-110
人数	194	59	265	874	3137	8692	14886	21551	11276	817	9
比例	0.31%	0.1%	0.43%	1.42%	5.08%	14%	24.1%	34.89%	18.26%	1.32%	0.01%

表 3 07-10 年不同职业脑卒中发病人数

职业	农民	工人	退休	教师	渔民	医务	职工	离退	其他
人数	29632	4844	6634	216	66	90	735	1751	17792
比例	0.4798	0.0784	0.1074	0.0035	0.001	0.0015	0.011	0.03	0.29

不同年龄段脑卒中发病人数柱状图如图 2 所示，各职业脑卒中发病人数如图 3：

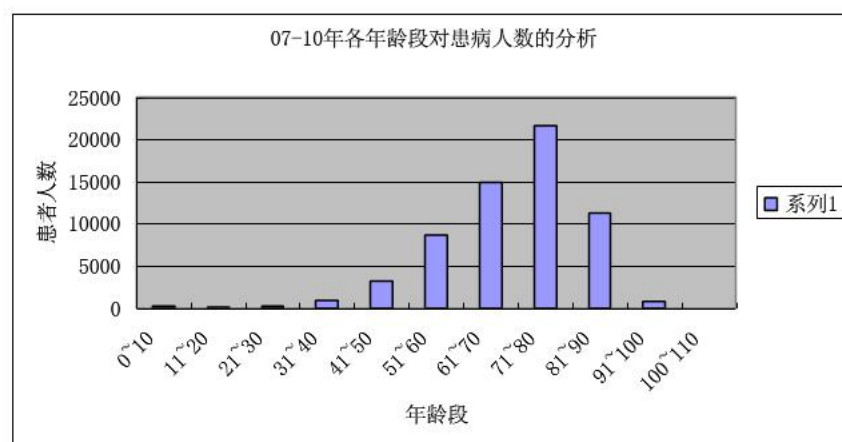


图 2 不同年龄段脑卒中发病人数

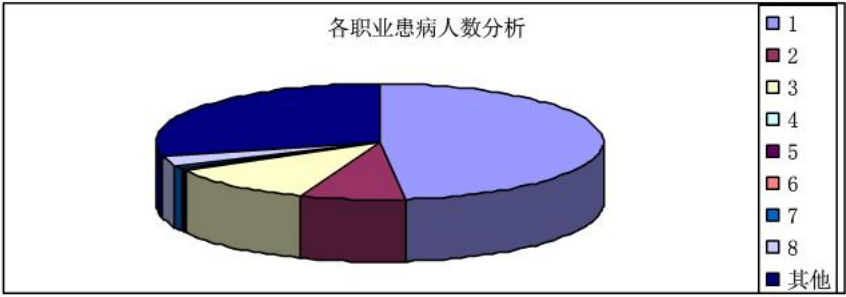


图3 不同职业脑卒中发病率

从表1和图1可以看出,脑卒中的主要发病年龄段集中在50岁到90岁之间,占了总发病人数的88.33%,因此该年龄段的人群是脑卒中发病的高危人群。进一步地,我们得到同一职业中,脑卒中各年龄段的发病情况(表3)。

表4 同一职业中,各年龄段脑卒中的发病人数

职业 年龄段	1	2	3	4	5	6	7	8	其他
0~10	2	169	2	0	4	0	1	0	16
11~20	7	4	1	0	0	0	1	0	46
21~30	37	58	2	1	0	4	12	1	150
31~40	214	220	9	7	2	2	39	1	380
41~50	1068	682	61	17	3	10	124	8	1164
51~60	3798	1124	622	32	10	17	177	149	2763
61~70	7094	1016	1864	58	25	27	133	466	4203
71~80	11069	1051	2691	79	14	18	161	690	5778
81~90	5931	513	1264	22	8	11	82	421	3024
91~100	409	42	114	0	0	1	5	14	232
101~110	3	0	4	0	0	0	0	1	1
总计	29632	4879	6634	216	66	90	735	1751	17757

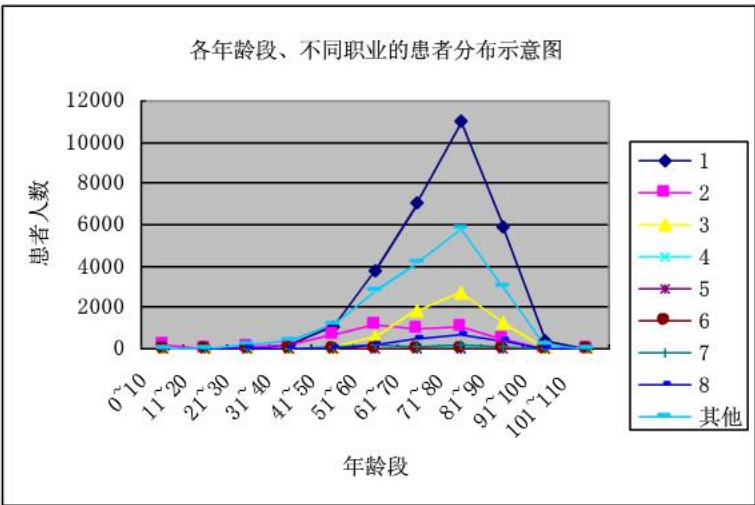


图4 同一职业中,各年龄段脑卒中的发病人数折线图

从表 4 的数据和图 4 的折线图可以看，在统计的 8 各行业中，脑卒中发病的主要人群集中在 50 岁至 90 岁之间。

通过对原始数据整理，我们得到 07 年至 10 年不同职业人群脑卒中的发病人数(表 5)，其柱状图见图 5

表 5 07-10 年不同职业发病人数数据

	农民	工人	退休人员	教师	渔民	医务人员	职工	离退休人员	其他
2007	7036	814	1898	38	18	23	62	417	9800
2008	8017	1005	2000	63	38	18	102	807	4013
2009	1491	1156	103	71	6	35	313	23	6398
2010	8150	1173	1700	44	3	13	234	500	4178
合计	24694	4148	5701	216	65	89	711	1747	24389

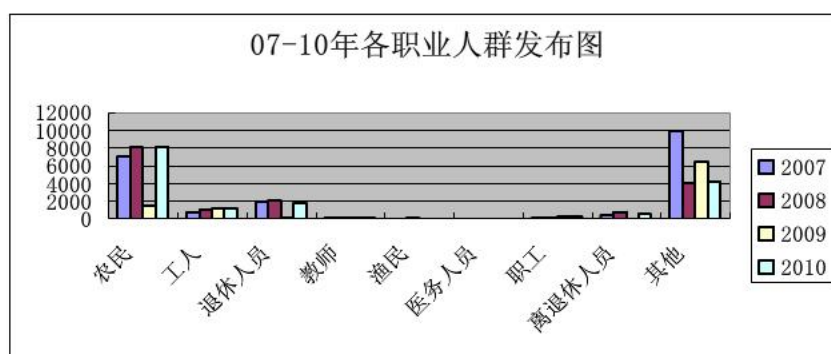


图 5 07-10 年不同职业发病人数数据柱状图

从图 5 可知，农民和其他职业脑卒中发病人数远远高于工人等另外六个职业。

(二) 脑卒中发病人群的进一步分析

在前面得到的不同性别、各年龄段及不同职业的脑卒中发病数据的基础上，我们对数据进一步分析。主要从三个方面进行描述一是脑卒中发病人群分布的集中趋势，反映各数据向其中心值靠拢或聚集的程度；二是脑卒中发病人群分布的离散程度，反映各数据远离其中心值的趋势；三是发病人群分布的偏态和峰度，反映数据分布的形状。从这三个方面反映脑卒中发病人群的不同侧面。从前面的分析可知，不同性别，不同职业，其各年龄段脑卒中发病的趋势是大致相同的，因此我们只对脑卒中发病人群的年龄分布做进一步分析。

1. 脑卒中发病人群年龄分布的集中趋势

集中趋势是指数据像某一中心值靠拢的倾向，也就是寻找数据一般水平的代表值或中心值。我们根据表 2 得到的数据，计算年龄分布的众数、中位数和(上下)分位数及均值。

(1) 年龄分布的众数

我们利用以下公式(1)[4]计算众数：

$$M_0 = L + \frac{f - f_{-1}}{(f - f_{-1}) + (f - f_{+1})} \times i \quad (1)$$

其中 L 为众数组的下限值， i 为众数组的组距(我们取定 10)。由公式 1，我们算得众数为：

$$M_0 = 71 + \frac{21551 - 14886}{(21551 - 14886) + (21551 - 11276)} \times 10 = 75$$

(2) 年龄分布的中位数和分位数

根据分组数据计算中位数，先根据 $N/2$ 确定中位数的位置，确定中位数所在的组，采用公式 2[4] 计算中位数的近似值：

$$M_e = L + \frac{\frac{N}{2} - S_{m-1}}{f_m} \times i_z \quad (2)$$

其中 N 为数据的个数(总次数)； L 为中位数所在组的下限值； S_{m-1} 为中位数所在组以前各组的累积频数； f_m 为中位数所在组的频数； i 为中位数所在组的组距(我们取定为 10)。根据(2)式算得脑卒中发病人群年龄分布的中位数约为 72。

四分位数也称四分位点，他是通过三个点将全部数据等分为四部分，其中每部分包含 25% 的数据，处在分为点上的数值就是四分位数。由公式(2)很容易算出上下四分位点：

$$Q_L = L_L + \frac{\frac{N}{4} - S_L}{f_L} \times i_L = 61 + \frac{15440 - 13221}{14886} \times 10 = 62.49$$

$$Q_U = L_U + \frac{\frac{3N}{4} - S_U}{f_U} \times i_{UL} = 71 + \frac{46320 - 28107}{21551} \times 10 = 79.45$$

(3) 年龄分布的均值

根据表 2 中的分组数据计算均值[4]。原始数据被分为 K 组(表 2 为 11 组)，各组的组中值为 X_1, X_2, \dots, X_K ，各组变量值出现的频数分别为 F_1, F_2, \dots, F_K ，则均值的计算公式为：

$$\bar{X} = \frac{X_1 F_1 + X_2 F_2 + \dots + X_K F_K}{F_1 + F_2 + \dots + F_K} = \frac{\sum_{i=1}^K X_i F_i}{\sum_{i=1}^K F_i} \quad (3)$$

由公式(3)算得年龄分布的均值为 69.78。

(4) 年龄分布的众数、中位数和均值的比较

从分布的角度看，众数始终是一组数据分布的最高峰值，中位数是出于一组数据中间位置上的值，二均值则是全部数据的算术平均。将前面计算的众数、中位数和均值标在表 2 数据的曲线图上，如图 4 所示，脑卒中发病的年龄分布为左

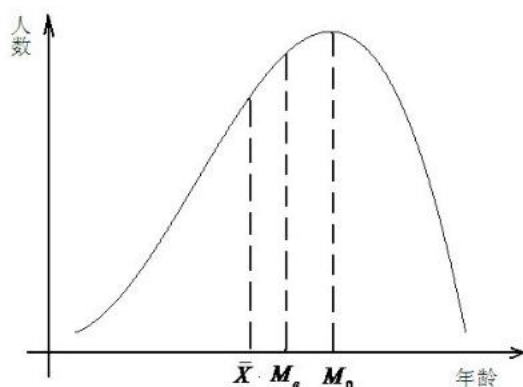


图 4 左偏分布: $\bar{X} < M_e < M_0$

偏分布,说明数据存在极小值,拉动均值像极小值一方靠,而众数和中位数由于是位置代表值,不受极值的影响,因此三者之间的关系表现为: $\bar{X} < M_e < M_0$ 。

2. 脑卒中发病人群分布的离散程度

数据的分散程度是数据分布的另一个重要特征,它反映的是各变量值远离其中心值的程度,因此成为离中趋势。数据的离散程度越大,集中趋势的测度值对该组数据的代表性就越差,离散程度越小,其代表性就越好。我们主要计算异众比率和四分位差[4]。

(1) 年龄分布的异众比率

异众比率的计算公式为:

$$V_r = \frac{\sum F_i - F_m}{\sum F_i} = 1 - \frac{F_m}{\sum F_i} \quad (4)$$

有公式(4)算出年龄分布的异众比率为 $V_r = 1 - \frac{21551}{61760} = 0.6511 = 65.11\%$, 这说明在患病的人群中,71~80 岁以外的人数占 65.11%,异众比率比较大.因此,用 71~80 岁来代表患病人群的状况,其代表性不是很好。

(2) 年龄分布的四分位差

四分位差即上四分位数与下四分位之差,计算公式为: $Q_d = Q_u - Q_L$, 有前面四分位数算出的结果得到 $Q_d = 6.96$, 其值较小,说明脑卒中的发病年龄比较集中。

3. 脑卒中发病人群分布的偏态和峰度

为了得到脑卒中发病人群年龄分布的形状、偏斜程度以及分布的扁平程度等,我们计算脑卒中发病人群分布的偏态和峰度。

(1) 年龄分布的偏态系数

偏态是对分布偏斜方向和程度的测度,其计算公式为:

$$\alpha_3 = \frac{\sum_{i=1}^K (X_i - \bar{X})^3 F_i}{N\sigma^3} \quad (5)$$

式中 σ^3 为标准差的三次方。利用公式(5)，算得 $\alpha_3 = 1.236006 > 0$,说明正偏离差值较大，即年龄分布为右偏。

(2) 年龄分布的峰度系数

峰度为分布集中趋势高峰的形状，其计算公式为：

$$\alpha_4 = \frac{\sum_{i=1}^K (X_i - \bar{X})^4 F_i}{N\sigma^4} \quad (6)$$

式中 σ^4 为标准差的四次方。利用公式(6)，算得 $\alpha_4 = 0.485408 < 3$,说明脑卒中年龄分布为平峰分布，说明发病人群中，年龄大的和年龄小的人占较小比重。

二. 发病率与气温、气压、相对湿度间的关系

为了得到脑卒中的发病率与气温、气压、相对湿度的关系。我们考虑对原始数据进行整理，得出 2007-2010 年中 1-12 月的脑卒中发病率、平均气压、平均气温、平均相对湿度、气压差、气温差和湿度差数据(表 5)。

表 5 月发病率、平均气压、平均气温、平均相对湿度、气压差、气温差和湿度差数据

	发病率	平均气压	气压差	平均气温	气温差	平均湿度	湿度差
一月	0.087337377	1027.8	4.025	3.75	6.775	67.8	33.6
二月	0.08136781	1022.15	5.875	6.725	7.425	70.725	37.55
三月	0.088797327	1018.7	6.35	10.35	8.15	67.25	41.7
四月	0.087175161	1015.8	5.725	14.85	8.45	65.475	42.2
五月	0.090176	1009.3	4.525	21.575	9.4	64.45	48.5
六月	0.085472	1006.2	3.55	24.475	7.22	77.175	37.2
七月	0.081644	1004.8	7.275	29.15	7.62	73.825	36.95
八月	0.08565	1005.9	4.178	28.9	6.925	74.875	37.4
九月	0.080118742	1011.1	3.4	24.75	6.5	78.175	36.05
十月	0.085893651	1018.25	3.85	19.425	7.575	73.15	45.2
十一月	0.077766603	1023.7	4.475	12.175	8.025	70.975	44.15
十二月	0.0734192	1024.15	5.525	6.8	7.625	66.85	39.75

首先建立线性回归模型^[3]：

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 \quad (7)$$

运用 Matlab 软件得到模型(1)的回归系数估计值及置信区间（置信水平 $\alpha = 0.05$ ，程序见附录 1）、检验统计量 R^2, F, p, s^2 的结果如下表：

参数	参数估计值	参数置信区间
β_0	1.0222	[-2.2231 4.2675]
β_1	-0.0008	[-0.0039 0.0022]

β_2	-0.0014	[-0.0065 0.0036]
β_3	-0.0004	[-0.0030 0.0022]
β_4	-0.0026	[-0.0326 0.00274]
β_5	-0.0008	[-0.0036 0.0019]
β_6	0.0003	[-0.0038 0.0044]
$R^2 = 0.3950 \quad F=0.5440 \quad p=0.7602 \quad s^2=0$		

得到的 $R^2 < 0.6, P > 0.05$ 故模型(1)不合理，即发病率与平均气压、平均气温、平均相对湿度、气压差、气温差和湿度差呈非线性关系。为此我们反复进行分析，引入交叉变量 $x_2 * x_3$ 和 $x_1 * x_3$ ，最终确定回归模型为：

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_1 x_2 + \beta_6 (x_2 x_3)^2 + \beta_7 x_1 x_3 \quad (8)$$

运用 Matlab 软件得到模型（2）的回归系数估计值及置信区间（置信水平 $\alpha = 0.05$, 程序见附录 2）、检验统计量 R^2, F, p, s^2 的结果如下表：

参数	参数估计值	参数置信区间
β_0	-7.1711	[-9.6355 -4.7067]
β_1	0.0072	[0.0047 0.0096]
β_2	2.2260	[1.7118 2.7401]
β_3	0.0680	[0.0241 0.1120]
β_4	-0.0155	[-0.0231 -0.0079]
β_5	0.0025	[0.0012 0.0039]
β_6	0.0022	[-0.0027 -0.0017]
β_7	-0.0000032	[-0.000004166267057 -0.00000230927687]
β_8	-0.00068	[-0.000110821086 -0.000025095048]
$R^2 = 0.98987599, F= 36.66567079, p= 0.0065640042, s^2 = 0.00000090562$		

得出非线性回归方程

$$y = -7.1711 + 0.0072x_1 + 2.226x_2 + 0.068x_3 - 0.0155x_4 + 0.0025x_5 - 0.0022x_1x_2 - 0.0000032(x_2x_3)^2 - 0.00068x_1x_3 \quad (9)$$

结果分析:

其中 $R^2 = 0.98987599$ 指 y (发病率) 的 98.987599% 可由该模型确定, F 指超过了 F 检验的临界值。 $p < \alpha$, 因而该模型从整体上看是可用的。

上表的回归系数给出了模型(2)中 $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8$ 的估计值. 检查他们的置信区间发现, 没有零点, 表明回归变量(对因变量 y 的影响)都是显著的. 所以模型是可用的. 由(9)式可知如果知道了平均气压, 气压差, 平均气温, 气温差, 平均湿度, 湿度差, 就可以计算预测值 y (发病率). 这对于某个地方预测脑卒中人数具有很好的参考价值. 从中我们也可以发现气压, 温度等变量是无法控制的, 所以对于脑卒中的发病率我们无法进行有效的控制.

三. 提出预警和干预的建议方案

脑中风在现在的生活中越来越普遍, 人们对于脑中风都有种非常恐惧的心理. 其实大多数人对脑中风并没有深入的了解. 脑中风是一种突然起病的脑血液循环障碍性疾病, 又叫脑血管意外。

重要特征: 作息不规律, 高血压, 高血脂, 糖尿病, 高度肥胖, 吸烟与酗酒, 高度紧张等是引起脑中风的危险因素. 脑卒中高危人群的重要特征: 血压显著增高, 发热感, 多汗, 口干, 寒战. 胸闷, 胸痛, 心悸. 多食, 多饮, 多尿。

关键指标: 通过收集数据我们得出了 07-10 年 5 年间的脑卒中的发病情况, 用单因素相关多元逐步回归分析 8 项主要月气象指标, 即月平均气温 $T_a(^{\circ}\text{C})$, 月平均最高气温 $H_h(^{\circ}\text{C})$ 、最低气温 $T_l(^{\circ}\text{C})$, 月平均气压 $P_a(\text{hpa})$ 、月平均最高气压 $P_h(\text{hpa})$ 、最低气压 $P_l(\text{hpa})$ 、最低气压 $P_l(\text{hpa})$, 月平均相对湿度 $H_a(\%)$ 、最低湿度 $H_l(\%)$ (最大湿度为 100%, 通常不记录) 对于近 5 年脑卒中的发病的影响. 男性较女性更易患脑中风. 这是因为所有与脑中风有关的危险因素均与男性关系更为密切, 例如, 男性承担着更多的压力, 男性吸烟、饮酒的比例较高, 男性患高血压、糖尿病等疾病较女性早且比例高等. 脑中风的发生与季节有一定的关系, 秋、冬季脑中风发病率比夏季高, 所以, 脑中风患者在冬季寒冷天气应避免过多户外活动, 坚持在医生指导下服用必要药物, 及时控制血压波动。

脑中风的发病与死亡都与年龄有十分密切的关系。

(1)、根据问题 1 的结论中我们发现 80 岁之前随着年龄的增长, 脑中风的发病率显著上升, 基本规律是年龄每增加 5 岁, 脑中风的死亡率增加近 1 倍, 所以, 年龄大的人容易患脑中风. 另外, 职业差异也影响着脑中风的发病率, 农民和工人患病比例较高, 这可能是因为农民和工人的劳动强度比较大, 压力较大, 身体负荷较重。

气象因素也是诱发脑中风患者发病的重要因素。

(2)、由问题二我们对于脑中风的发病率与气象因素的关系进行了分析, 得出脑卒中的发病率与平均气压、平均气温、气压差、气温差呈正相关关系, 与湿度差呈负相关关系, 平均相对湿度对发病率的影响甚微. 对此我们建议人们应该尽量少到温度差、气压差较大、气温、气压较高的地方活动. 此外, 对于脑中风患者还要控制饮食, 要少摄入高脂肪, 高盐, 油腻的食物. 经常的体育锻炼, 培养健康、稳定、乐观的情绪. 可使精神放松, 血压平稳, 生活松紧有度, 积极干预极端的气象变化会减轻对人体的不良影响, 从而减少脑中风的发生。

综上所述,冬季及早春气候变化比较大,温度,湿度及气压处于大幅度波动状态,这种变化多端的气象条件使人体生理失去平衡,引起微循环、血液动力学及血凝等方面的变化,从而诱发急性心脑血管病的发生。由此,我们得出一下结论:(1)脑梗塞及脑出血的发病存在季节性。(2)脑梗塞与脑出血的年发病有集中趋势,集中于12月和1月。(3)脑卒中的发生主要与气温和气压的变化及其幅度有关。

干预措施:

1、预防脑中风,就要把中风的危险因素尽可能降到最低。控制高血压是预防中风的重点。高血压病人要遵医嘱按时服用降压的药物,有条件者最好每日测1次测血压,特别是在调整降压药物阶段,以保持血压稳定。要保持情绪平稳,少做或不做易引起情绪激动的事,如打牌、搓麻将、看体育比赛转播等;饮食须清淡有节制,戒烟酒,保持大便通畅;适量运动,如散步、打太极等。防治动脉粥样硬化,关键在于防治高脂血症和肥胖。建立健康的饮食习惯,多吃新鲜蔬菜和水果,少吃脂肪高的食物如肥肉和动物内脏等;适量运动增加热量消耗;服用降血脂药物。控制糖尿病与其它疾病如心脏病、脉管炎等。

2、注意中风的先兆征象:一部分病人在中风发作前常有血压升高、波动,头痛头晕、手脚麻木无力等先兆,发现后要尽早采取措施施加以控制。

3、有效的控制短暂性脑缺血发作,当病人有短暂性脑缺血发作先兆时,应让其安静休息,并积极治疗,防止其发展为脑血栓。

4、注意气象因素影响:季节与气候变化会使高血压病人情绪不稳,血压波动,诱发中发,在这种时候更要防备中风发生。

参考文献

- [1] 钟南山 崔丽英 脑中风防治 科学出版 2010.12
- [2] 赵静 但琦 数学建模与数学实验(第三版) 高等教育出版社 2007.6
- [3] 姜启源 谢金星 叶俊 数学模型(第四版) 高等教育出版 2010.9
- [4] 贾俊平 何晓群 金进勇 统计学 中国人民大学出版社 2000.8

附 录

附录一： 问题 2 模型 1 的程序

```
Y=[0.087337377 0.08136781 0.088797327 0.087175161 0.090176 0.085472  
0.081644 0.08565 0.080118742 0.085893651 0.077766603 0.0734192]';  
X1=[1027.825 1022.15 1018.7 1015.825 1009.375 1006.155 1004.838333  
1005.966905 1011.1 1018.25 1023.7 1024.15]';  
X2=[4.025 5.875 6.35 5.725 4.525 3.55 7.275 4.178571429 3.4 3.85 4.475  
5.525]';  
X3=[3.75 6.725 10.35 14.85 21.575 24.475 29.15 28.9 24.75 19.425 12.175  
6.8]';  
X4=[6.775 7.425 8.15 8.45 9.4 7.22 7.62 6.925 6.5 7.575 8.025 7.625]';  
X5=[67.8 70.725 67.25 65.475 64.45 77.175 73.825 74.875 78.175 73.15 70.975  
66.85]';  
X6=[33.6 37.55 41.7 42.2 48.5 37.2 36.95 37.4 36.05 45.2 44.15 39.75]';  
X=[ones(12,1),X1,X2,X3,X4,X5,X6];  
[b,bint,r,rint,stats] = regress(Y,X)
```

b =

```
1.0222  
-0.0008  
-0.0014  
-0.0004  
-0.0026  
-0.0008  
0.0003
```

bint =

```
-2.2231    4.2675  
-0.0039    0.0022  
-0.0065    0.0036  
-0.0030    0.0022  
-0.0326    0.0274  
-0.0036    0.0019  
-0.0038    0.0044
```

r =


```

0.0043
0.0005
0.0049
0.0009
-0.0008
-0.0002
0.0004
-0.0003
-0.0022
0.0041
-0.0018
-0.0097

```

```
rint =
```

```

0.0001    0.0085
-0.0112    0.0122
-0.0058    0.0156
-0.0131    0.0148
-0.0087    0.0071
-0.0085    0.0081
-0.0049    0.0057
-0.0115    0.0109
-0.0149    0.0105
-0.0040    0.0122
-0.0099    0.0063
-0.0151   -0.0042

```

```
stats =
```

```
0.3950    0.5440    0.7602    0.0000
```

```
>>
```

附录二 问题2中模型2的程序

```

Y=[0.087337377 0.08136781 0.088797327 0.087175161 0.090176 0.085472
0.081644 0.08565 0.080118742 0.085893651 0.077766603 0.0734192];
X1=[1027.825 1022.15 1018.7 1015.825 1009.375 1006.155 1004.838333
1005.966905 1011.1 1018.25 1023.7 1024.15];
X2=[4.025 5.875 6.35 5.725 4.525 3.55 7.275 4.178571429 3.4 3.85 4.475

```

```

5.525];
X3=[3.75 6.725 10.35 14.85 21.575 24.475 29.15 28.9 24.75 19.425 12.175
6.8];
X4=[6.775 7.425 8.15 8.45 9.4 7.22 7.62 6.925 6.5 7.575 8.025 7.625];
X5=[67.8 70.725 67.25 65.475 64.45 77.175 73.825 74.875 78.175 73.15 70.975
66.85];
X6=[33.6 37.55 41.7 42.2 48.5 37.2 36.95 37.4 36.05 45.2 44.15 39.75];
X=[ones(12,1),X1,X2,X3,X4,X6,X1.*X2,(X2.*X3).^2,X1.*X3];
[b,bint,r,rint,stats] = regress(Y,X)
b =

```

```

-7.171059307768187
0.007156116987215
2.225980764982258
0.068019909415769
-0.015493590134576
0.002531807827427
-0.002182450586350
-0.000003237771962
-0.000067958067192

```

bint =

```

-9.635464546947359 -4.706654068589014
0.004749645330254 0.009562588644175
1.711816220110520 2.740145309853995
0.024079742791028 0.111960076040511
-0.023051554941926 -0.007935625327227
0.001190539062601 0.003873076592253
-0.002685706512738 -0.001679194659963
-0.000004166267057 -0.000002309276867
-0.000110821086175 -0.000025095048210

```

r =

```

1.0e-003 *

-0.126254381651647
0.799077270144680
-0.146644840359408
-0.080143742809069
-0.243059608576762

```

0.370362314180431
0.002566784798208
0.165936794739752
-0.505525881028271
-0.084606881633045
0.800638827777214
-0.952346655557892

rint =

-0.000186348970451	-0.000066159792852
-0.001372683963462	0.002970838503752
-0.002453917128720	0.002160627448001
-0.002122362977833	0.001962075492214
-0.001774289929567	0.001288170712414
-0.001416568657828	0.002157293286189
-0.000142597256317	0.000147730825913
-0.000424066975914	0.000755940565393
-0.002351028852824	0.001339977090767
-0.001238427398453	0.001069213635187
-0.000598277012941	0.002199554668496
-0.001789630046276	-0.000115063264840

stats =

0.989875993279864	36.665670790347853	0.006564004198176	0.000000905627211
-------------------	--------------------	-------------------	-------------------