

2012 高教社杯全国大学生数学建模竞赛

承 诺 书

我们仔细阅读了中国大学生数学建模竞赛的竞赛规则.

我们完全明白, 在竞赛开始后参赛队员不能以任何方式(包括电话、电子邮件、网上咨询等)与队外的任何人(包括指导教师)研究、讨论与赛题有关的问题。

我们知道, 抄袭别人的成果是违反竞赛规则的, 如果引用别人的成果或其他公开的资料(包括网上查到的资料), 必须按照规定的参考文献的表述方式在正文引用处和参考文献中明确列出。

我们郑重承诺, 严格遵守竞赛规则, 以保证竞赛的公正、公平性。如有违反竞赛规则的行为, 我们将受到严肃处理。

我们授权全国大学生数学建模竞赛组委会, 可将我们的论文以任何形式进行公开展示(包括进行网上公示, 在书籍、期刊和其他媒体进行正式或非正式发表等)。

我们参赛选择的题号是(从 A/B/C/D 中选择一项填写): _____ C _____

我们的参赛报名号为(如果赛区设置报名号的话): _____

所属学校(请填写完整的全名): _____

(隐去论文作者相关信息等)

日期: _____ 2012 年 9 月 10 日 _____

赛区评阅编号(由赛区组委会评阅前进行编号): _____

2012 高教社杯全国大学生数学建模竞赛

编 号 专 用 页

赛区评阅编号（由赛区组委会评阅前进行编号）：

赛区评阅记录（可供赛区评阅时使用）：

评阅人										
评分										
备注										

全国统一编号（由赛区组委会送交全国前编号）：

全国评阅编号（由全国组委会评阅前进行编号）：

脑卒中发病环境因素分析及干预

摘要: 脑卒中逐渐威胁人们的生活, 本文主要针对脑卒中发病病例信息和受病环境因素进行统计分析, 从实际数据结果加深对脑卒中的认识, 旨在对脑卒中加以预防。

针对问题一, 先主要借助于 EXCEL 编程及筛选功能、MATLAB 辅助编程对附件数据进行错误修复及标准化处理, 得到 2007~2010 年期间有效数据的发病年、月、日, 然后在 EXCEL 中分别按性别、年龄、职业、时间 (包括年、月、日) 四个字段对发病人数进行统计, 并以图、表的形式予以展示, 最后总结出脑卒中患者男女性别比为 1.17:1、集中患病年龄段为 71~80 岁、高危职业为农民、存在一定季节性等结论, 该问属于一般的数据统计分析模型。

针对问题二, 先对患者按照天来统计四年每天的发病人数 (共 1461 条数据), 再将气象数据与发病人数按天进行关联构成新的源数据, 同时计算每天的气压差、温差, 最后以发病率为因变量, 以平均气压、最高气压、最低气压、气压差、平均温度、最高温度、最低温度、温度差、平均湿度、最低湿度 10 个特征为自变量进行多元线性回归, 其步骤是先画因变量与自变量的散点图观测它们的关系, 再利用 SPSS 软件统计所有变量之间的相关性, 最后进行多元逐步回归分析。结果表明: ①发病率与这 10 个指标的相关性并不大, 但整体上与最低气压、最高温度和温差呈正相关、与平均湿度和气压差成负相关; ②发病率与平均湿度直接线性相关, 逐步回归的模型为 $y = 3.022 - 0.004x$, 且模型检验为 $F=7.555$ 、 $\text{Sig.}=0.006$, 表明该模型通过显著性检验; ③再次以平均湿度为因变量, 以气压和温度为自变量进行逐步回归发现, 平均湿度受温差、平均气压影响, 这间接地对脑卒中发病率产生影响。

针对问题三, 通过查阅资料文献得到脑卒中高危人群的重要特征和关键指标、主要诱发因素, 并结合问题一和问题二中的相关结论对脑卒中高危人群进行了预警和干预建议。

最后, 本文对模型进行了检验及评价分析, 用 2007~2010 年的发病数据进行回代检验, 两者绝对距离小于 1 的比例为 86%。同时, 本文的分析可以推广应用到其它疾病、农作物收成等受环境、气候影响的分析及预警评估中。

关键词: 脑卒中, 环境因素, 统计分析, 多元线性回归, 逐步回归, 显著性检验, 预警, 回代检验

一、 问题重述

随着社会的发展，人们生活水平不断提高，但与此同时，伴随着城市化进程加快，人口密度加大，生活节奏加快和膳食结构改变等不良现象，一些严重威胁人们身体健康的疾病发生，心脑血管疾病以其高死亡率而越来越引起人们的关注。

其中脑卒中（俗称脑中风，包括脑出血、蛛网膜下腔出血和脑梗塞，脑出血和蛛网膜下腔出血均属心脑血管疾病）是目前威胁人类生命的严重疾病之一，目前对脑卒中尚无特效治疗方法或令人满意的治疗效果，因此积极预防尤为重要。随着人们对预防疾病和保证健康生活方式的重视，气候变化对人类健康的影响也倍受关注，国内外许多研究表明气象要素的变化对心脑血管疾病有着重要影响。因此研究气象要素与心脑血管疾病之间的关系对于防病和治病具有重要的现实意义。

脑卒中的发生是一个漫长的过程，一旦得病就很难逆转。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（见 Appendix-C1）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（Appendix-C2）。

请建立数学模型，解决如下问题：

问题一：根据病人基本信息，对发病人群进行统计描述。

问题二：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。

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，结合问题一、问题二中所得出结论，对高危人群提出预警和干预的建议方案。

二、 问题分析

本文主要目标是要分析脑卒中受发病环境因素的影响关系以及对应的预防措施，其总体研究方法是通过现有数据进行统计规律分析，找出脑卒中的发病率与环境因素（温度、湿度、大气压）的关系描述，并通过查询资料文献了解脑卒中高危人群的重要特征及常见的预防脑卒中的预防措施，最后再结合第一问和第二问分析的结果对高危人群提出预警和干预的建议方案，旨在提高对脑卒中的防护能力。鉴于此目的，针对本文具体 3 个问题，可以进行如下分析：

2.1 针对问题一的分析

本问题主要根据附件(Appendix-C1)中四个文件中的脑卒中发病病例信息进行相关统计分析，这些病例信息指标主要有性别、年龄、职业、发病时间、诊断时间，为了对发病人群进行统计描述，本文主要从以下几点进行考虑：

1. 按性别统计，包括总人数、主要集中年龄段、高危职业名称、发病与诊断时间的间隔（判断该病的潜伏性）；

2. 按年龄段统计, 包括该年龄段内的性别、人数、比例、高危职业、发病与诊断时间的间隔;
3. 按职业统计, 包括该职业内的发病人的性别、集中年龄段、发病与诊断时间的间隔;
4. 分别按发病年、月统计 (发病年月和诊断年月基本一致), 包括性别、年龄段、高危职业等。

但是从附件数据中发现, 在“Time of incidence (发病时间)”和“Report time (诊断报告时间)”中存在不同的时间格式以及错误(如: #### 或空格), 因此在对数据进行统计分析前, 需要首先对数据进行修复, 根据一定修复原则将一些明显的错误信息 (如发病时间为 5008/7/31、诊断报告时间为 27/09/2008 情况下, 很明显 5008 应该是 2008)。

同时, 从附件数据中易发现, 部分诊断时间没有数据, 而且诊断时间比较混乱, 错误比较多, 因此本文将不对诊断报告时间进行分析, 进而也将不统计发病与诊断时间的间隔。

最后在修复完成后的基础上按上述思想进行脑卒中的发病信息统计, 其统计的工具主要是 EXCEL, 利用 EXCEL 丰富的公式编辑、筛选、绘图、统计等功能进行处理。

2.2 针对问题二的分析

本问题欲研究脑卒中发病率与气温、气压、相对湿度间的关系, 主要需要注意以下几点:

1. 在第一问已修复的数据基础上进行发病率统计, 主要统计方法是通过 EXCEL 的筛选功能和编写程序统计出在 2007~2010 年期间每一天的发病人数, 进而可以计算出按天及按月的发病率;
2. 对附件(Appendix-C2)中数据文件进行整理及统计计算, 先按天统计 2007~2010 年期间每一天的气象信息 (温度、湿度、大气压), 并计算出每一天的温度差、气压差, 再按月分别统计这四年中的 8 种指标 (平均气压、最高气压、最低气压、平均温度、最高温度、最低温度、平均湿度、最低湿度) 每月的各个平均值、最大值、最小值;
3. 将 1 和 2 统计或计算的数据进行一一关联, 构造后续分析的数组。

从上面的统计数据可以看出, 该问是一个多元统计问题^[1], 即分析脑卒中发病率与温度、湿度、大气压的各种指标的关系, 主要分析思想如下:

1. 先整体按天 (2007~2012 年共 1461 天) 分析, 分析过程为: ① 在 EXCEL 中画出发病率与各个统计指标的散点图, 从直观上寻求发病率与它们是否有明显的规律 (如线性相关); ② 利用 SPSS 统计软件对所有数据进行相关性分析, 分析两两之间的相关性; ③ 利用 SPSS 软件进行多元线性回归, 分析回归结果是否通过显著性检验; ④ 由于某些变量之间存在非常大的互相关 (如温度之间的三个指标互相关系数都比较大), 因此需要对多个变量进行筛选, 可用的方法为多元线性逐步回归法 (可以借助于 SPSS 统计软件中的逐步回归选项或 MATLAB 中的 stepwise 逐步回归工具箱); ⑤ 如果不存在前面操作没有求出发病率与温度、湿度、大气压的相关表达式, 则继续按后续方法进行分析处理;
2. 然后按照每月或季节的数据进行类似分析;
3. 按照温度、湿度、大气压三类进行单因素相关性分析, 先选择其中两个特征变化很小或在一个指定范围内变化的数据, 对发病率与第三个指标进行相关性分析, 通过此方法进行单因素分析。

整个过程需要做大量的统计分析, 包括绘图及数据归纳整理, 主要工具有 EXCEL、SPSS、MATLAB。

2.3 针对问题三的分析

本问题首先要通过资料文献了解脑卒中高危人群的重要特征和关键指标、脑卒中的主要诱发因素、常见的预防措施、已有的某些地区对脑卒中发病的统计信息和规律，根据这些信息最大化地提取关于脑卒中发病的指标，再结合问题一、问题二中所得结论，可以根据所查到的关键指标、气象信息、时间序列进行预测模型的建立，如多指标影响因素的多元线性或非线性回归、神经网络预测模型、时间序列预测等等，最后对高危人群提出预警和干预的建议方案。

三、 模型假设及符号说明

3.1 基本假设

1. 假设附件中的数据除空格、R####等本身有误外其它数据是合理可靠的。
2. 假设附件数据中每一位病人都属于不同的人。
3. 假设除环境因素（温度、湿度、大气压）外，影响脑卒中发病的其他因素保持不变。
4. 假设当地人口不发生较大的变动，死亡率与出生率相近。
5. 假设 2007~2010 数据四年间，没有发生重大自然灾害。
6. 假设当地医疗环境相当，数据代表整个城市数据，数据具有代表性。

3.2 符号说明

P ：某天（月或其它统计范围）的年发病率
 N ：某天（月或其它统计范围）的发病人数
 M ：某年的总发病人数
 m ：自变数个数
 Y ：因变数
 $X_1 \dots X_m$ ：自变数
 b_i ：各个自变数 x_i 对依变数 y 的各自效应；
 \hat{y} ：自效应的集合

3.3 基本定义

$$\text{发病率: } P = \frac{N}{M} \quad \text{式 (1)}$$

四、 模型建立及求解

4.1 针对问题一的模型建立及求解

由问题分析可知,这属于多信息变量的统计描述模型,该问题主要是对脑卒中发病者信息进行统计描述,其方法是分别对脑卒中患者病历信息——性别、年龄、职业、发病时间进行统计,全部操作在 EXCEL 中进行。

4.1.1 附件数据的修复处理

由于附件中的患者病例信息有许多格式错误及信息不完整,在进行统计描述前,有必要对数据做修复处理,本文的修复过程及方法如下:

1. 年龄(Age)字段中存在大于 110 岁(如 799)、0 岁的信息,本文处理方法为将区间[1 110]之间的数据作为有效值,其余的全视为该患者年龄信息缺失。
2. 职业(Occupation)字段中存在 1-8 之外的数据(如 9、工等异常),可能是数据录入错误,也可能是还有其它类的职业没在附件中说明,本文处理方法为将 1-8 之外的数据视为其他职业段。
3. 发病时间(Time of incidence)字段存在日期格式错误(如 15-06-2008、20080620 等)或与 EXCEL 标准时间格式(如 2007/1/1)不统一,需要对时间数据进行修复及标准化处理,处理原则有以下几点:
 - ① 类似“2009-0-24”的数据丢失了月份信息,此类数据认为是错误数据,不统计在 2007~2010 期间内;
 - ② 类似“发病时间为 5008/7/31、诊断报告时间为 27/09/2008”存在明显错误的数字,5008 应该修复成 2008;
 - ③ 类似“31/12/2009”的数据不是 EXCEL 标准的时间格式,为了便于在 EXCEL 中快速按年、月、日进行统计分析,有必要对非标准的日期数据进行标准化处理,其方法可以通过在一单元格中进行编写公式进行字符串处理,假设“31/12/2009”所在的单元格为“D2”,则计算标准化的日期格式(2009/12/31)的公式为“=DATE(RIGHT(D2,4),MID(D2,4,2),LEFT(D2,2))”;
 - ④ 类似“发病时间为 20110/05/09、诊断报告时间为 2010-08-08”的数据,直接视为无效数据;
 - ⑤ 类似“2009/0/24”的数据也视为无效数据。

4.1.2 脑卒中患者信息统计分析

通过上述数据修复过程后,将得到标准格式的脑卒中患者信息数据,现按照模型分析的思路对脑卒中病例信息进行统计描述,其核心方法是在 EXCEL 中利用“COUNTIF”函数对某条件进行筛选后统计患者人数、及“COUNTIFS”函数对多重条件进行筛选后统计患者人数,具体操作界面截图见附录 B-1。

4.1.2.1 按性别统计

对男女性别分别进行筛选，以年为单位，将四年的数据信息进行统计，2007~2010 年按性别的脑卒中发病人数统计如表 1 所示，2007~2010 年男女患病人数统计图如图 1 所示。

表 1 2007~2010 年按性别的脑卒中发病人数统计表

性别	年统计人数					
	附件总数据	2007-2010 年总	2007	2008	2009	2010
男	33385	33367	7302	10384	5198	10483
女	28526	28506	5940	8659	4805	9102
丢失信息	12	12	0	12	0	0
男女比	1.17:1	1.17:1	1.23:1	1.20:1	1.08:1	1.15:1

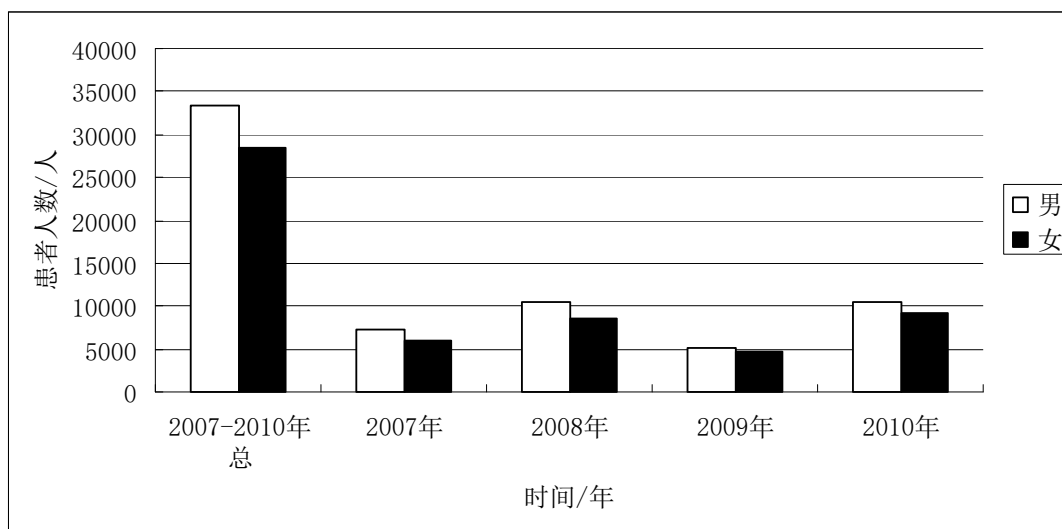


图 1 2007~2010 年男女患病人数统计图

从表 1 及图 1 可以看出，2007 年男女患者之比达 1.23:1，男性比女性更容易患脑卒中这类疾病，可能原因有以下几点：一是男性高血压多于女性；二是男性吸烟与饮酒者多于女性；三是男性从事体力劳动较多，突然用力可能诱发中风。

4.1.2.2 按职业统计

按职业字段进行筛选得到 2007~2010 年各职业患病人数统计数据如表 2 所示。

表 2 2007~2010 年各职业患病人数统计表

2007~2010 年按职业统计数据				
职业		发病人数	性别	
编号	名称		男	女
1	农民	29750	14644	15084
2	工人	4856	3108	1745
3	退休人员	6646	4126	2517
4	教师	216	163	53

5	渔民	66	43	23
6	医务人员	90	65	25
7	职工	735	513	220
8	离退人员	1751	1181	570
其它或缺失	其它或缺失	17775	9524	8268

从表中看出农民患病人数为 29750，属于较多人群，为高危职业，而医务人员等明显较低，这与工作强度相关。

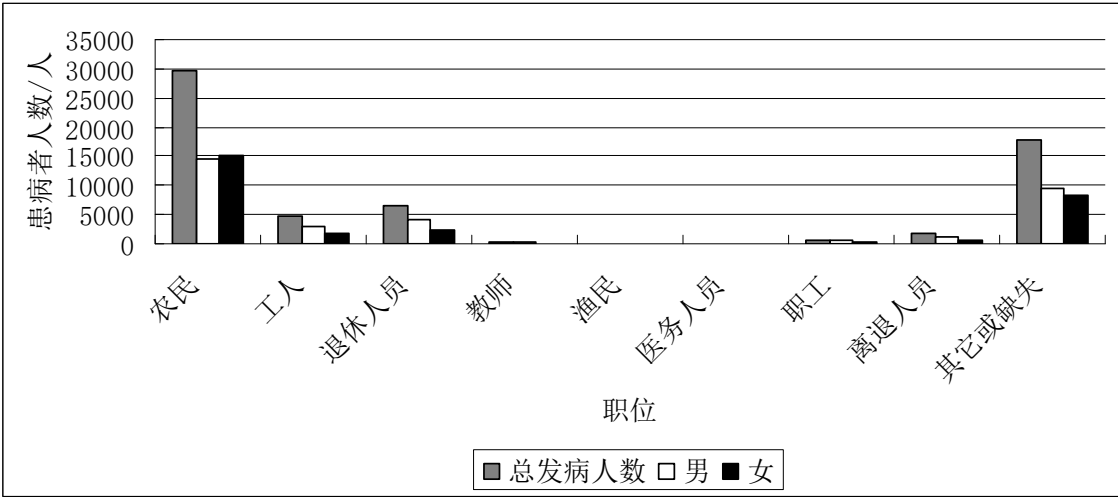


图 2 2007~2010 年各职业患病人数统计图

可以得出结论：经济收入较高的人群较收入低的人群脑卒中发病率低，户外重体力劳动者发病率较高。

4.1.2.3 按年龄统计

针对职业统计中，退休人员所占比例较大说明与年龄有关，对年龄进行筛选，将年龄分为各个阶段，统计出每年中不同年龄段的患病人数，以 2007-2008 年为例进行如表 3 所示的描述，各年详细数据见附录 A-1。

表 3 2007-2008 年各年龄段内患病人数统计表

	2007			2008		
	患病人数	男	女	患病人数	男	女
1--10	17	10	7	50	15	35
11--20	7	4	3	14	9	5
21--30	35	16	19	57	32	25
31--40	155	96	59	235	173	62
41--50	614	374	240	865	566	298
51--60	1861	1135	726	2547	1514	1033
61--70	3069	1784	1285	4669	2803	1864
71--80	4842	2678	2164	6648	3496	3147
81--90	2309	1051	1258	3549	1609	1936
91--100	170	57	113	249	82	167

101--110	3	3	0	4	2	2
其他	126	76	50	25	12	13

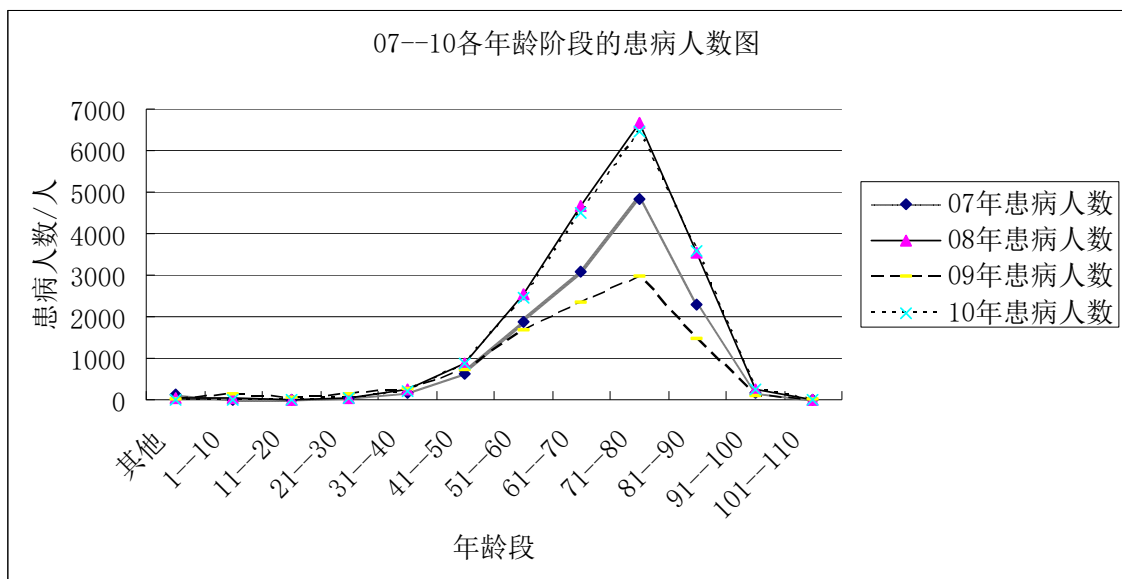


图 3 2007~2010 年各年龄阶段的患病人数图

由图 3 可见，患病人数随年龄的增加而增加，上升速度以 50 到 60 上升较快，61 岁以上的人群脑卒中的高发群体，集中年龄段在 71-80 岁之间，说明脑卒中中以老年人人居多，且脑卒中患者呈年轻化的趋势。

进一步按照各年龄段，对男女患者发病人数的进行区分，可得图 4 所示。

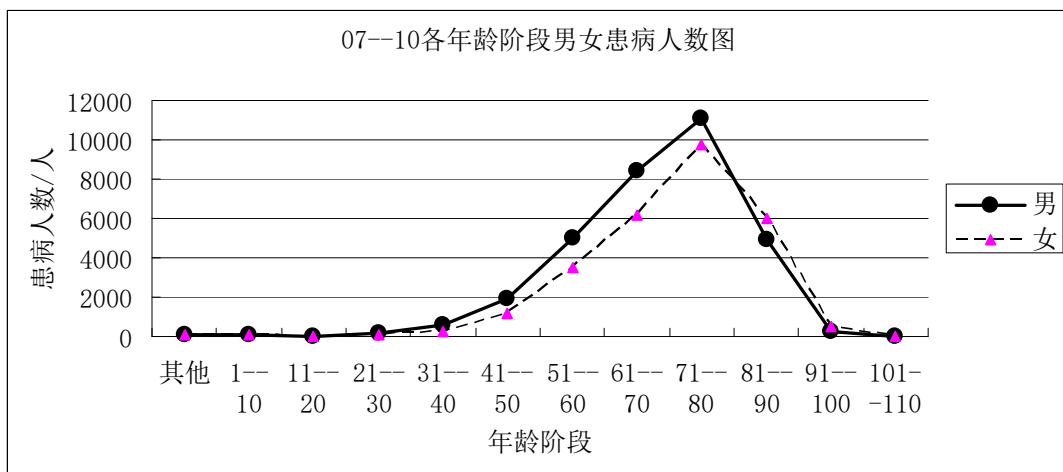


图 4 2007~2010 四年期间各年龄阶段男女患病人数图

可见，男女高峰年龄段一致；男性在 41~71 岁之间，患病人数明显高于女性；71 岁以后患病明显回落，且低于女性发病人数，可知男性发病早于女性，同时这现象可能是由于高龄组死亡率持续增高所致。但无论男女，构成随着年龄增加而增加，这与其在年龄发病相符。

4.1.2.4 按时间统计

按年份对发病人数进行统计，得到发病人数统计图如图 5 所示。

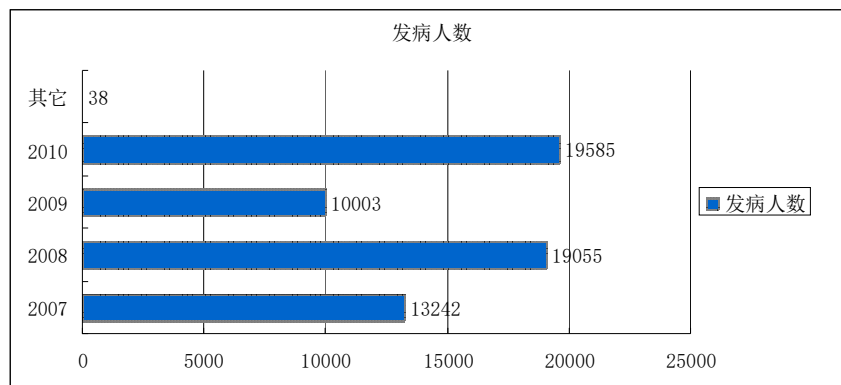


图 5 脑卒中患者按年的统计人数分布

从上图可以看出，附件总数据为 61923 条，但 2007~2010 间有效的数据为 61885 条，本文做的统计描述均是针对 2007~2010 期间内。

按月份对发病人数进行统计，得到发病人数统计表如表 4 所示。

表 4 2007~2010 年各月患病人数统计表

月份	07 年发病人数	08 年发病人数	09 年发病人数	10 年发病人数	2007-2010 年总发病人数
1	935	1827	872	1760	5394
2	732	1961	848	1487	5028
3	1019	1918	830	1724	5491
4	1069	1758	860	1699	5386
5	1072	1776	876	1882	5606
6	1032	1517	793	1610	4952
7	1014	1500	931	1757	5202
8	1197	1366	934	1680	5177
9	1221	1272	829	1632	4954
10	1374	1461	759	1718	5312
11	1208	1378	664	1565	4815
12	1369	1321	807	1071	4568

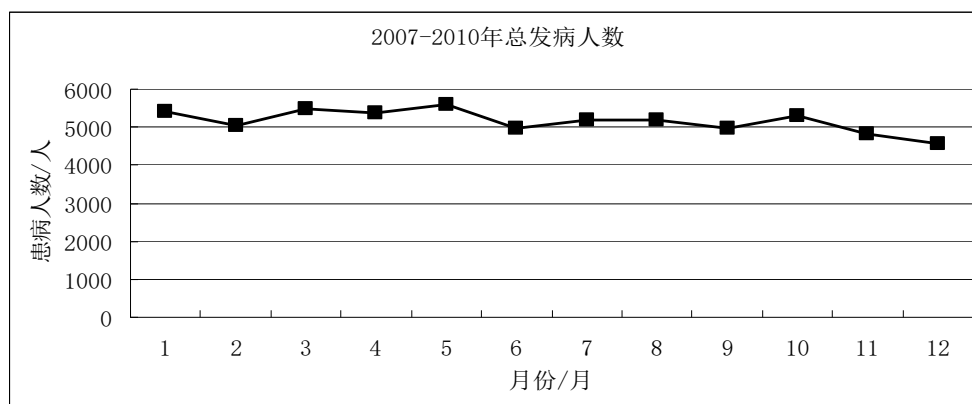


图 6 2007-2010 年总发病人数随月份的变化曲线

从 2007~2010 年逐年脑卒中发病人数的月分布发现，该病以春节多发，高峰出现在 3~

5 月, 1 月为次高峰, 6~9 月发病较为平缓, 12 月出现低谷期。由此可见发病存在一定的季节差异, 脑卒中春季高于其他季节, 而夏、秋、冬三季发病差异不大。

利用 EXCEL 中的“COUNTIFS”函数对脑卒中病例数据进行多重筛选统计患者数量, 得到 2007~2010 四年每天的发病人数, 其曲线如图 7 所示。

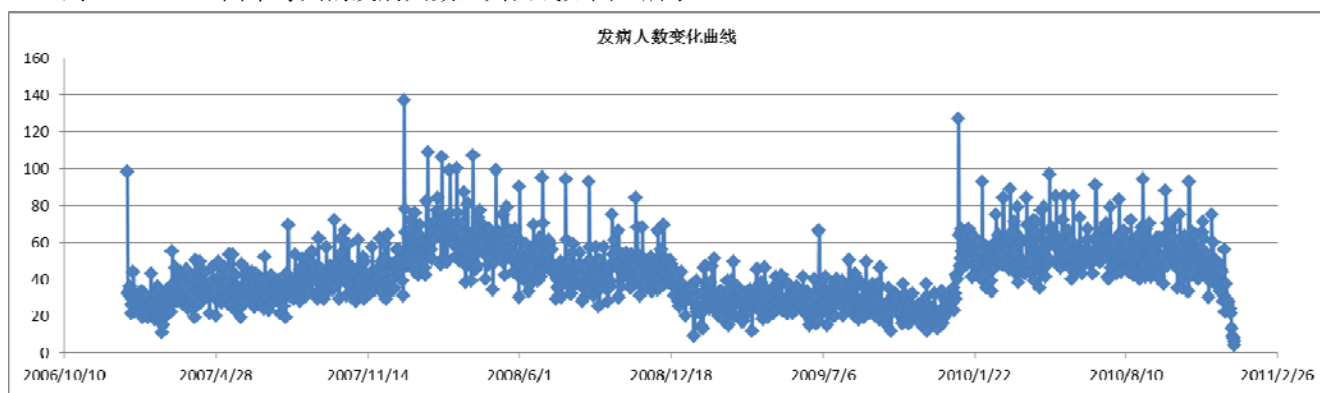


图 7 2007-2010 年总发病人数随天的变化曲线

根据式 (1) 求出 2007~2010 四年内每天的发病率, 其发病率随时间的变化曲线如图 8 所示。

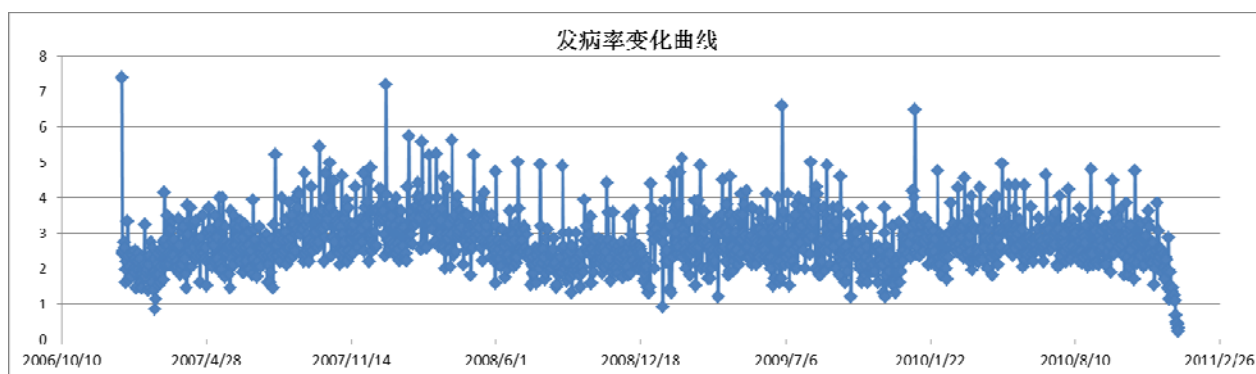


图 8 2007-2010 年发病率随天的变化曲线

从图 7 和图 8 可以看出, 2007~2010 四年内每天发病人数变化不大, 每天的发病率基本保持不变。但是如果按天进行统计分析, 每天的随机误差容易对结果造成影响, 再每月的均值作为统计对象进行分析, 四年内每月的发病率百分比曲线如图 9 所示。

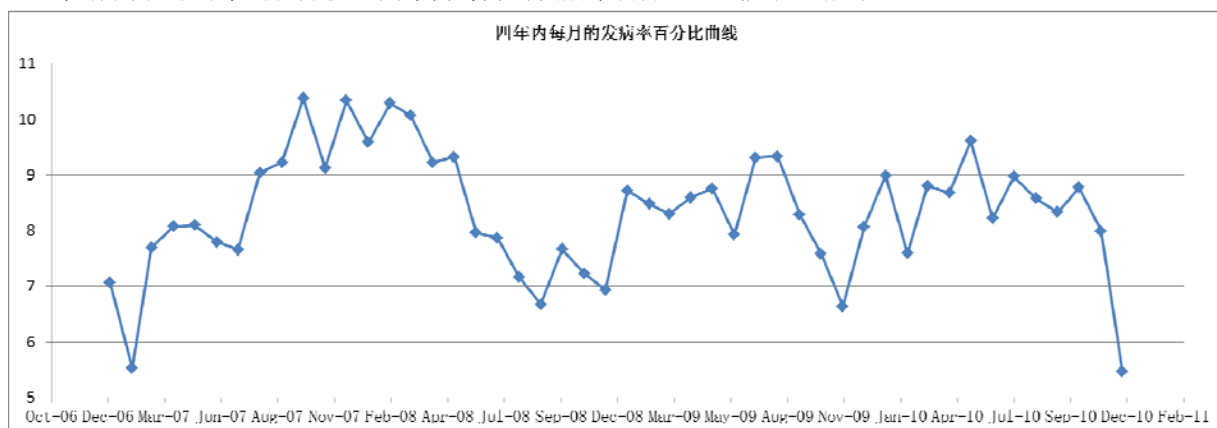


图 9 2007-2010 年发病率随月的变化曲线

从图 9 可以看出，月发病率随时间呈周期性波动，具有一定的季节性。

4.1.2.5 重要结论

- (1) 脑卒中的发病有年集中趋势，更呈增长趋势；
- (2) 发病存在时间差异，春节为高发季，1 月为高峰月；
- (3) 患者人数男性多于女性，性别比重为 1.17:1；
- (4) 工作性质对脑卒中发病有直接影响，农民为高危职业；
- (5) 脑卒中发病处于老年阶段，集中年龄段为 71~80，且逐年呈年轻化发展。

4.2 针对问题二的模型建立及求解

由问题分析可知，问题二属于一个多元统计分析模型，目标是研究因变量发病率与自变量温度（包括平均温度、最高温度、最低温度、温度差）、湿度（包括平均湿度、最低湿度）、气压（平均气压、最高气压、最低气压、气压差）之间的关系，本文主要从多元线性或非线性回归模型上进行分析。

4.2.1 数据归纳与统计

附件(Appendix-C2)中的数据已经给出了 2007-2010 年每天对应的气象数据，可以在这基础上对气象数据进行进一步细化：

(1)计算每天的气压差与温差,最终得到 2007~2010 年期间每一天的气象特征信息——平均气压、最高气压、最低气压、气压差、平均温度、最高温度、最低温度、温度差、平均湿度、最低湿度等 10 个特征变量；

(2) 按月份统计所有数据中每月的最大值及最小值情况。

最后将第一问进行统计出的发病率情况与气象数据信息进行一一关联，得到最终待分析的数据集，其数据形式如表 5 所示。

表 5 数据归纳统计形式

按天统计													
时间	发病人 数	发病率	发病率千 分比	平均 气压	最高 气压	最低 气压	平均 温 度	最 高 温 度	最 低 温 度	平均 湿 度	最 低 湿 度	气 压 差	温 度 差
2007/1/1	98	0.007401	7.40069476	1025.1	1028.5	1023.3	8.1	9.9	7.4	86	71	5.2	2.5
2007/1/2	32	0.002417	2.41655339	1025.2	1026.7	1023.5	6.5	7.4	6	84	73	3.2	1.4
2007/1/3	33	0.002492	2.49207068	1026.1	1027.8	1025.1	5	6.9	4.2	86	77	2.7	2.7
2007/1/4	36	0.002719	2.71862256	1027.1	1029.2	1025.7	5.9	7.4	4.2	82	78	3.5	3.2
2007/1/5	34	0.002568	2.56758798	1027.1	1029	1025.2	5	6.5	4.3	84	76	3.8	2.2

4.2.2 多元回归分析过程

多元回归分析包括多元线性回归及多元非线性回归，判断方法主要通过绘制因变量与各个自变量之间的散点图，首先直观分析因变量与自变量的关系，如果从散点图可以看出明显

的线性关系，那么可以考虑通过多元线性回归进行分析；如果从散点图并不能发现明显的线性规律，可能是呈非线性，也可能是多个自变量之间的耦合关系的影响，需要进一步分析才能决定。

4.2.2.1 多元线性回归数学模型

若依变数 Y 同时受到 m 个自变数 X_1, X_2, \dots, X_m 的影响，且这 m 个自变数皆与 Y 成线性关系，则这 $m+1$ 个变数的关系就形成 m 元线性回归。因此，一个 m 元线性回归总体的线性模型为：

$$Y_j = \beta_0 X_0 + \beta_1 X_{1j} + \beta_2 X_{2j} + \dots + \beta_m X_{mj} + \varepsilon_j \quad \text{式 (2)}$$

其中， $\varepsilon_j \sim N(0, \sigma_\varepsilon^2)$ 。相应的，一个 m 元线性回归的样本观察值组成为：

$$y_j = b_0 + b_1 x_{1j} + b_2 x_{2j} + \dots + b_m x_{mj} + e_j \quad \text{式 (3)}$$

在一个具有 n 组观察值的样本中，第 j 组观察值($j=1, 2, \dots, n$)可表示为($x_{1j}, x_{2j}, \dots, x_{mj}, y_j$)，便是 $M=(m+1)$ 维空间中的一个点。

同理，一个 m 元线性回归方程可给定为：

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_m x_m \quad \text{式 (4)}$$

式(3)中， b_0 是 x_1, x_2, \dots, x_m 都为 0 时 y 的点估计值； b_1 是 $b_{y1 \cdot 23 \dots m}$ 的简写，它是在 x_2, x_3, \dots, x_m 皆保持一定时， x_1 每增加一个单位对 y 的效应，称为 x_2, x_3, \dots, x_m 不变(取常量)时 x_1 对 y 的偏回归系数(partial regression coefficient)； b_2 是 $b_{y2 \cdot 13 \dots m}$ 的简写，它是在 x_1, x_3, \dots, x_m 皆保持一定时， x_2 每增加一个单位对 y 的效应，称为 x_1, x_3, \dots, x_m 不变(取常量)时 x_2 对 y 的偏回归系数；依此类推， b_3 是 x_3 对 y 的偏回归系数；……； b_m 是 x_m 对 y 的偏回归系数。

在多元回归系统中， b_0 一般很难确定其专业意义，它仅是调节回归响应面的一个参数； $b_i(i=1, 2, \dots, m)$ 表示了各个自变数 x_i 对依变数 y 的各自效应，而 \hat{y} 则是这些各自效应的集合，代表着所有自变数对依变数的综合效应。

多元线性回归模型的求解可以直接通过 SPSS 软件和 MATLAB 相应的工具求解。

4.2.2.2 按天的数据分析

显然，本文中的因变量 Y 为脑卒中发病人数或发病率或发病率千分比，自变量 X 有平均气压、最高气压、最低气压、气压差、平均温度、最高温度、最低温度、温度差、平均湿度、最低湿度等 10 个变量，首先按照每天的统计数据进行多元线性回归分析。

(1) 观测发病率与自变量的散点图

以平均气压为例，绘制发病率千分比与平均气压的散点图如图 10 所示。

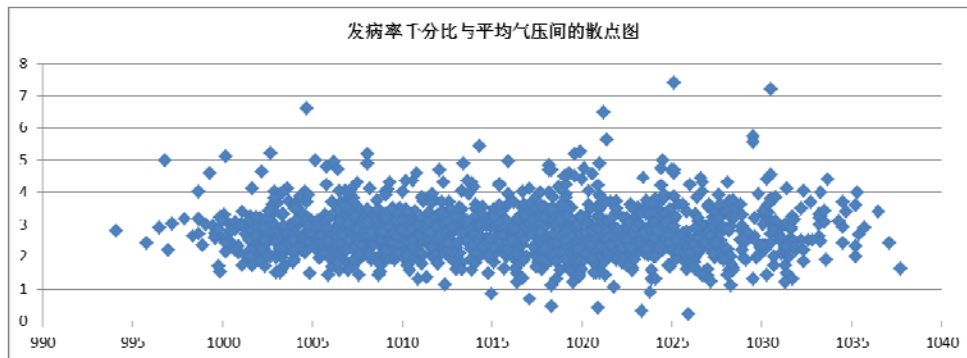


图 10 2007-2010 年每天发病率千分比与平均气压间的散点图

从该图可以看出发病率与平均气压并没有明显的线性变化关系,可能原因是发病率与平均气压的相关性不强,也可能是受其它自变量耦合关系的影响,因为要分析发病率与平均气压的关系必须要在其它指标保持不变或变化很小范围内研究才具有可靠性,因此需要进一步分析。

(2) 所有变量两两相关性分析

将表 5 所示的数据导入 SPSS 软件中^[1], 进行变量之间的相关性分析, 所得结果如表 6 所示。

表 6 发病率与 10 个自变量之间的相关性统计

		发病率千分比
平均气压	Pearson 相关性	-.001
	显著性 (双侧)	.965
	N	1461
最高气压	Pearson 相关性	-.003
	显著性 (双侧)	.910
	N	1461
最低气压	Pearson 相关性	.001
	显著性 (双侧)	.968
	N	1461
平均温度	Pearson 相关性	.006
	显著性 (双侧)	.834
	N	1461
最高温度	Pearson 相关性	.009
	显著性 (双侧)	.742
	N	1461
最低温度	Pearson 相关性	.001
	显著性 (双侧)	.977
	N	1461
平均湿度	Pearson 相关性	-.072
	显著性 (双侧)	.006
	N	1461
最低湿度	Pearson 相关性	-.028
	显著性 (双侧)	.290
	N	1461
气压差	Pearson 相关性	-.015
	显著性 (双侧)	.554

	N	1461
温度差	Pearson 相关性	.024
	显著性（双侧）	.368
	N	1461

从表 6 可以看出，发病率只与平均湿度能通过显著性检验（ <0.05 ），且相关系数都非常低，但总体上发病率与最低气压呈正相关、与最高温度成正相关、与平均湿度成负相关、与温差呈正相关、与气压差呈负相关。

（3）逐步回归分析

从（1）和（2）分析可知发病率从单因素上讲，它与其它自变量的相关性非常小，且无规律可行，需要进行多因素分析，可以通过多元线性回归进行尝试。但是，部分自变量之间又存在很强的相关性，如关于温度的四个指标之间的相关系数都大于 0.9，因此发病率肯定不是所有这 10 个特征变量的函数表达式，因此本文采用逐步回归法进行分析。

逐步回归分析的基本原理为：在建立多元回归方程的过程中，按偏相关系数的大小次序将自变量逐个引入方程，对引入方程中的每个自变量偏相关系数进行统计检验，效应显著的自变量留在回归方程内，循此继续遴选下一个自变量。如果效应不显著，停止引入新自变量。由于新自变量的引入，原已引入方程中的自变量由于变量之间的相互作用其效应有可能变得不显著者，经统计检验确证后要随时从方程中剔除，只保留效应显著的自变量，直至不再引入和剔除自变量为止，从而得到最优的回归方程。

对于本文中的逐步回归分析仍借助于 SPSS 软件的“回归”功能进行分析，设定显著性水平为 0.05，逐步回归的模型检验为 $F=7.555$ ， $\text{Sig.}=0.006$ ，具体结果如下：

表 7 逐步回归模型结果

模型		非标准化系数		标准系数	t	Sig.
		B	标准 误差	试用版		
1	(常量)	3.022	.105		28.741	.000
	平均湿度	-.004	.001	-.072	-2.749	.006

散点图

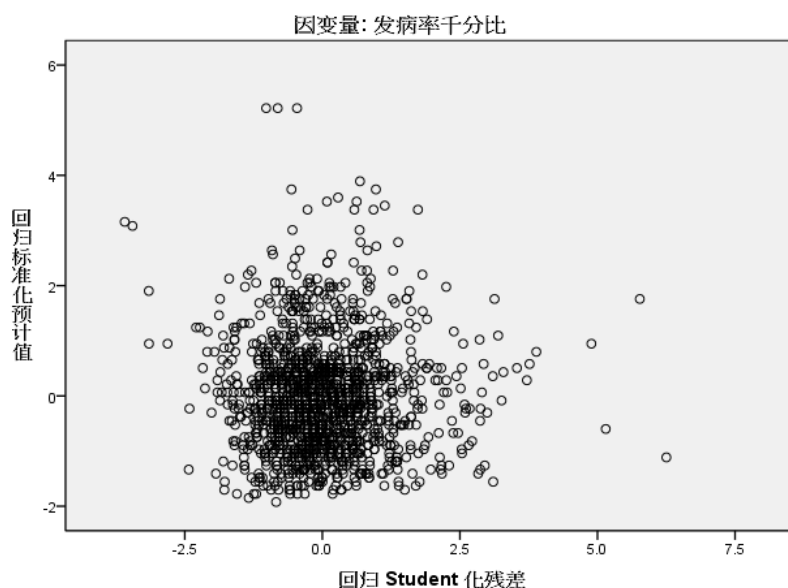


图 11 逐步回归的 Student 化残差图

从表 7 可以看出，所计算的参数的显著性水平 Sig.均小于 0.05，表示计算回归模型通过了显著性检验，且标准误差也比较小；从图 11 也可以看出残差图基本落在[-2.5 2.5]内，表明回归模型较好，该逐步回归的多元线性模型为：

$$y = 3.022 - 0.004x \quad \text{式 (5)}$$

其中 y 代表发病率千分比， x 代表平均湿度，这也说明发病率与平均湿度条件直接相关。

(4) 气象变量间的相关性分析

由于湿度与温度、气压密切相关，脑卒中发病率虽然与温度、气压没有直接关联度，但温度和气压因素却影响湿度参数，因此以平均湿度为因变量，以平均气压、最高气压、最低气压、气压差、平均温度、最高温度、最低温度、温度差 8 个特征为自变量进行逐步回归分析，通过已有数据分析平均湿度与温度、气压因素的函数关系。具体分析方法与前面一致，分析结果如下：

表 8 平均湿度与温度、气压的逐步回归模型结果

模型		非标准化系数		标准系数	t	Sig.
		B	标准 误差	试用版		
1	(常量)	84.455	.837		100.939	.000
	温度差	-1.796	.102	-.419	-17.605	.000
2	(常量)	448.227	35.594		12.593	.000
	温度差	-1.768	.099	-.412	-17.931	.000
	平均气压	-.358	.035	-.235	-10.223	.000

从表 8 可以看出湿度与温度、压强的线性关系具有两种模型：

模型 1：

$$y = 384.455 - 1.796x_1 \quad \text{式 (6)}$$

模型 2：

$$y = 448.227 - 1.768x_1 - 0.358x_2 \quad \text{式 (7)}$$

其中， y 代表平均湿度， x_1 代表温度差， x_2 代表平均气压，这说明平均湿度与温度差和平均气压相关。

4.2.2.3 按月或季度的数据分析

根据表 5 所示的数据，很容易统计出每月或季度的数据，其中，每月（季）的发病率=每月（季）的发病总人数/该年的发病总人数，每月的气象数据为该月的平均值或最大最小值，关于发病率与气象环境关系的分析与前面所描述的过程基本一致。

4.3 针对问题三的求解

4.3.1 脑卒中高危人群的重要特征和关键指标

4.3.1.1 重要特征

根据脑动脉狭窄和闭塞后，神经功能障碍的轻重和症状持续时间，分三种类型：

(1) 短暂性脑缺血发作颈内动脉缺血表现为，突然肢体运动和感觉障碍、失语，单眼短暂失明等，少有意识障碍。椎动脉缺血表现为，眩晕、耳鸣、听力障碍、复视、步态不稳和吞咽困难等。症状持续时间短，可反复发作，甚至一天数次或数十次。可自行缓解，不留后遗症。脑内无明显梗死灶。

(2) 可逆性缺血性神经功能障碍(RIND)与 TIA 基本相同，但神经功能障碍持续时间超过 24 小时，有的病人可达数天或数十天，最后逐渐完全恢复。脑部可有小的梗死灶，大部分为可逆性病变。

(3) 完全性卒中(CS)症状较 TIA 和 RIND 严重，不断恶化，常有意识障碍。脑部出现明显的梗死灶。神经功能障碍长期不能恢复，完全性卒中又可分为轻、中、重三型。

4.3.1.2 关键指标

脑卒中关键指标主要有：职业为农民或工人，年龄在 71-80 岁之间，并患有基础性疾病（高血压、糖尿病等）。

4.3.2 脑卒中的主要诱发因素

脑卒中的主要诱发原因有：

(1) 具有不良生活方式

①高危人群的饮食结构不良，多盐、油腻；②体力活动不足；③爱吸烟饮酒。

(2) 患有基础性疾病

①高血压：患高血压史在脑卒中住院病例中，有80%以上的病例有高血压病史，其中脑梗塞82.66%，脑出血80.27%有高血压史

②心梗、房颤心梗、房颤是引起脑中风的一个独立的强有力的危险因素，与健康人相比，增加脑中风危险度5倍以上，尤其是老年人房颤相当多。

③糖尿病明显增加缺血性脑中风发生率，单纯糖尿病者，严格控制血糖可减少微血管病变，减缓动脉硬化，从而减少脑中风。

(3) 职业性质的影响；

文化程度高的病人自我护理能力强，这可能由于文化程度高的病人，具有更好的学习理解能力，能更好地查阅书籍、报纸、文化程度高的病人更容易接受治疗、康复计划。

(4) 年龄阶段因素

中风发生最常见的基本条件就是动脉硬化，随着年龄的增长，生理变化和多种病理性因素相互作用使动脉硬化逐渐产生。

(5) 男女性别差异

一是男性高血压多于女性；二是男性吸烟与饮酒者多于女性；三是男性从事体力劳动较

多，突然用力可能诱发中风。

（6）季节性差异

脑卒中病人因其机体代谢差，肢体功能活动障碍而活动减少，神经传递障碍使温痛觉减弱或消失，患侧血液循环较健侧差，与季节气温相关。

4.3.3 对高危人群的预警和干预建议

（一）结合问题一：

针对 1：高危群众发病有年集中趋势，并逐年增长；

建议 1：

（1）舒缓压力：随着社会经济快速发展，社会各阶层压力过大，社会经济的发展直接影响人们的心理状态，这便是该病集中年患病的体现；文献报道^{〔5〕}血管疾病的发生与生活事件密切相关，在紧张和过多应激环境中，应激反应可通过垂体引起交感神经兴奋和肾上腺皮质激素增加，使血管强烈收缩血压突然升高而致脑出血。所以减少疾病，应保持身心愉悦，舒缓压力较为重要因素。

（2）合理饮食结构：饮食对身体，当今快节奏的快餐生活方式严重影响了我们的饮食结构，这也是疾病逐年增长的因素。为此建议平日应多摄入低盐低脂类食物，减少油脂摄入，改善血脂异常，改掉不良饮食习惯，因为肥胖与超重均为缺血性中风的危险因素。

针对 2：男性比女性更易患病

建议 2：

（1）戒烟：日常生活中男性一般要吸烟及大量酗酒，而尼古丁可刺激神经系统增快心率及脉率，血管收缩，血压升高长期的血管收缩和血液循环减慢，可使血中胆固醇、低密度脂蛋白沉积于动脉壁，导致动脉硬化^{〔8〕}。

（2）戒酒：酒精具有增压作用，每日酒精摄入量超过 78g 的重度饮酒者的高血压患病率是没有饮酒的 2 倍，酒量与血压水平呈明显的剂量依赖关系。小量饮酒有保护作用，大量饮酒可增加危险性^{〔10〕}。

针对 3：工作性质影响发病

建议 3：

（1）学会适当休息：即文化程度高的病人自我护理能力强。这可能由于文化程度高的病人，具有更好的学习理解能力，能更好地查阅书籍、报纸、文化程度高的病人更容易接受治疗、康复计划。

（2）增强自身学习能力。

针对 4：脑卒中处于老年阶段；

建议 4：

（1）每天运动 30 分钟降低 4 成患病几率：适当的运动会减少糖尿病^{〔13〕}在于运动能加速对摄入热量的消耗，这样可以将糖份积极转化为有力量的肌肉。每天坚持 30 分钟的运动就可以降低患糖尿病的风险 35%~40%。

（2）提高自我护理能力：脑卒中病人的自我护理能力受到年龄的影响，年龄增长带给病人的是神经功能缺损程度的加重和致残率增加，从而护理能力与年龄呈负相关^{〔14〕}。

（二）结合问题二：

问题 2 针对气温、气压、相对湿度等气候因素进行的干预建议：

针对 1：气温

脑卒中病人因其机体代谢差，肢体功能活动障碍而活动减少，神经传递障碍使温痛觉减弱或消失，患侧血液循环较健侧差，室温要求高。

建议 1:

(1)季节交替时期注意预防; (2) 注意保暖防寒, 早晚适量添衣, 而在白天活动时不要穿戴太多, 以防出汗多反而受凉; (3) 流感时期可佩挂防感冒香袋, 预防感冒发病。

针对 2: 气压;

低气压使各种过敏原和空气污染物、粉尘等刺激物不容易向高处扩散, 而易于向低处散落吸入呼吸道, 而且气压聚然降低可使支气管粘膜上的细小血管扩张, 气管分泌物增加, 支气管管腔变得狭窄容易诱发哮喘。研究发现平均气压偏低(840~907hpa)或偏高(926~971hpa)时, 呼吸系统疾病发病率高。

建议 2: (1)注意口腔, 鼻的清理, 保持呼吸通畅; (2)平时要加强体质锻炼, 多进行户外活动; (3)注意营养均衡, 对食欲不振患儿要注意脾胃调理; (4)勤洗手, 养成良好的卫生习惯^[12]。

针对 3: 相对湿度;

湿度过高, 蒸发减少, 抑制出汗, 使病人感到潮湿憋闷。湿度过低, 室内空气干燥, 人体水分蒸发增加, 引起干渴、咽痛、鼻衄等症状, 对呼吸及气管切开者尤为不利。

建议 3: (1)当湿度低时多喝水补充人体水分; (2)保持室内空气流通, 每天至少通风半小时以上。

五、 模型检验

由于第一问和第三问都不涉及到模型检验, 因此模型检验主要针对第二问。本文第二问的模型是多元线性回归模型, 所用的回归方法是逐步自回归法, 该模型的检验方法为将用 2007~2010 年的发病数据按照式(5)进行回代检验, 计算回代结果与原来统计的发病率千分比数据的绝对距离, 在总共 1461 条数据中, 绝对距离小于 1 的有 1257 条数据, 所占比例为 86%, 表明该模型较为合理。

六、 模型评价

6.1 模型优点

1、本文模型比较注重数据的处理和存储方式, 以表格、图形形式呈现, 大大提高了查询效率。

2、论文给出了大量图形, 条分缕析, 虽直观易懂, 但推理严谨, 深入浅出, 结果准确。模型可操作性强, 推广应用起来也很方便。

6.2 模型缺点与改进

1、在建模与编程过程中, 使用的数据只是现实数据的一种近似, 因而得出的结果可能与现实情况有一定的差距。

2、脑卒中的发病受多种因素综合作用的影响, 包括不可抗力事件, 这些自然灾害和社会重大事件均会对人民的身心乃至脑血管意外的发生产生影响。虽然样本较大, 但是时间不

够长，对揭示规律有其局限性，有待今后更进一步的观察研究。

3、数据不知地区等准确位置，无法得到更多较为精确的相关气象，节气等相关信息。

4、在前面的统计过程中都是在理想条件下进行的，实际上，还有人口迁移、经济景气变动、死亡率变动等因素对脑卒中的影响。模型还有诸多变量因素在内。考虑以上因素进行模型的建立。

七、 模型改进及推广

7.1 问题二的多元统计分析模型改进

本文仅仅做了多元线性回归，但从以上分析过程可以看出，发病率与气象环境因素的线性相关性并不特别明显，需要对数据进行合理统计，找出发病率与环境因素的关系，该模型具有一定的局限性。模型改进可以从以下几个方面改进：

1、查阅资料考虑多元非线性回归方法；

2、气象环境指标的进一步细化统计，如考虑温度与湿度比、最高温度与最低湿度一起统计等；

3、针对多变量的选择问题，可以考虑主成分分析法、因子分析、典型相关分析等方法^[1]。

7.2 推广及应用

本文中的多元统计分析模型可以推广到任一有关某变量与其它多因素的相关关系，如冠心病、心脏病等疾病受环境、气象因素的影响分析，更加贴近生活，为我们的日常生活服务。

参考文献

- [1] 于秀林，任雪松，多元统计分析，北京：中国统计出版社，1999
- [2] 杜强，贾丽艳，SPSS 统计分析从入门到精通，北京：人民邮电出版社，2011
- [3] 岳海燕，气象条件对南京地区慢性支气管炎和脑卒中发病的影响研究，硕士学位论文，南京信息工程大学，2009
- [4] 彭扬君脑卒中发作的诱因分析，现代保健·医学创新研究，2008(8)
- [5] 李芸，郑虹，陈红信息支持对老年高血压病人生活方式的影响，现代护理，2004(6)
- [6] 现代中西医结合杂志2002年第11卷第6期3月号
- [7] 刘春岭，黄如训，朱良付，曾进胜，苏镇培卒中患者高血压危险因素及并发症3059例20年回顾性分析，中国脑血管病杂志，2005(3)
- [8] 荆淑敏，魏淑芳，史丽珍，对心血管疾病病人开展健康教育的必要性，] 山西护理杂志，1999，13(3)：105-1576
- [9] 饶明俐，中国脑血管病防治指南编写委员会治指南，北京：卫生部疾病控制司，中华医学会神经病学分会，2005
- [10] 耿贯一，流行病学第三卷，北京：人民卫生出版社，1996

- [11] 赵文慧，脑卒中病人的自我护理，
http://www.365heart.com/tabloid/2010/06/temp_38155.shtml 2010-6-18
- [12] 国际糖尿病联盟发布糖尿病预防新宣，2007，5-19
- [13] 文章来源：放心医苑，<http://www.fx120.net/disease1/200810/79219.html>
- [14] 郭建民，张淑娥，初发脑中风的危险因子及预防，现代中西医结合杂志，11 卷 6 期：
524，2002

附录 A 重要数据

表 A-1 2007-2010 年各年龄段内患病人数统计表年的预测值

	2007			2008			2009		
	患病人数	男	女	患病人数	男	女	患病人数	男	女
其他	126	76	50	25	12	13	2	0	2
1--10	17	10	7	50	15	35	116	40	76
11--20	7	4	3	14	9	5	21	9	12
21--30	35	16	19	57	32	25	119	63	56
31--40	155	96	59	235	173	62	252	163	89
41--50	614	374	240	865	566	298	729	437	292
51--60	1861	1135	726	2547	1514	1033	1669	931	738
61--70	3069	1784	1285	4669	2803	1864	2336	1259	1077
71--80	4842	2678	2164	6648	3496	3147	2951	1483	1468
81--90	2309	1051	1258	3549	1609	1936	1469	643	826
91--100	170	57	113	249	82	167	100	38	62
101--110	3	3	0	4	2	2	1	0	1
	2010			2007~2010					
	患病人数	男	女	患病人数	男	女			
其他	2	0	2	0	0	0			
1--10	116	40	76	8	5	3			
11--20	21	9	12	16	10	6			
21--30	119	63	56	53	32	21			
31--40	252	163	89	224	144	80			
41--50	729	437	292	880	558	322			
51--60	1669	931	738	2465	1426	1039			
61--70	2336	1259	1077	4510	2570	1940			
71--80	2951	1483	1468	6471	3463	3008			
81--90	1469	643	826	3572	1607	1965			
91--100									
101--110									

表 A-2 2007 年发病人数统计

月 天	1	2	3	4	5	6	7	8	9	10	11	12
1	98	43	55	50	49	48	52	69	49	72	61	62
2	32	28	37	39	41	30	30	35	55	41	39	33
3	33	23	27	44	41	27	41	35	54	46	30	49
4	36	25	28	35	38	28	37	31	39	48	33	44
5	34	26	46	33	29	46	28	31	44	48	36	43

6	21	18	35	49	27	35	23	39	45	41	45	51
7	26	28	32	43	37	43	34	37	41	29	48	59
8	29	31	28	27	35	35	28	29	31	38	52	29
9	44	19	31	35	29	38	29	39	29	53	29	39
10	31	36	41	36	46	32	42	53	62	62	49	64
11	22	21	30	34	30	38	34	32	33	43	33	37
12	27	34	44	43	38	42	27	37	32	46	43	33
13	23	18	36	44	30	33	32	40	43	45	31	39
14	23	25	31	30	43	29	34	33	45	48	37	36
15	32	30	37	31	53	44	31	40	39	66	40	35
16	31	11	26	36	26	34	38	28	33	49	32	48
17	26	15	27	36	25	25	35	45	29	31	42	40
18	25	24	29	37	28	38	37	34	37	38	43	34
19	33	28	27	21	26	25	39	37	34	38	38	35
20	30	32	45	38	53	43	39	51	57	59	57	48
21	19	20	35	31	36	34	21	34	33	36	39	45
22	26	24	24	34	32	29	31	30	33	47	38	56
23	29	21	28	46	23	40	28	38	36	46	38	50
24	32	27	26	32	42	26	33	46	46	43	34	44
25	26	33	28	32	36	35	21	31	47	44	38	50
26	30	34	40	30	32	40	25	43	39	43	44	56
27	21	22	42	34	34	26	37	34	40	33	45	47
28	19	36	31	20	35	32	31	33	46	32	41	45
29	24		30	29	24	33	19	39	34	28	37	48
30	26		24	40	35	24	43	53	36	42	36	39
31	27		19	0	19		35	41		39		31
月总和	935	732	1019	1069	1072	1032	1014	1197	1221	1374	1208	1369
年总和	13242											

表 A-3 2008 年发病人数统计

月 天	1	2	3	4	5	6	7	8	9	10	11	12
1	137	109	99	107	99	90	95	94	93	75	84	66
2	78	57	76	65	61	61	70	61	57	53	41	49
3	65	52	67	47	54	46	60	46	50	49	68	52
4	53	56	62	45	58	43	48	41	38	61	40	35
5	47	69	59	67	62	53	57	48	41	53	53	56
6	48	56	66	47	65	48	47	48	41	47	43	46
7	62	51	69	73	64	43	53	44	31	56	31	42
8	53	65	52	59	56	39	47	32	45	46	50	69

9	66	61	52	55	60	51	60	35	41	30	36	49
10	72	71	100	77	75	59	61	54	57	66	68	51
11	51	62	76	56	49	46	59	59	36	30	39	40
12	46	72	66	58	53	50	49	39	47	48	51	49
13	49	77	50	62	42	50	45	37	33	44	46	40
14	61	84	57	69	64	33	56	41	25	36	38	38
15	76	75	57	60	79	53	47	45	50	53	48	47
16	46	64	56	66	52	54	43	38	35	46	35	38
17	51	65	64	40	57	37	47	45	39	43	50	50
18	48	76	70	53	62	45	46	47	37	42	43	48
19	42	48	50	52	61	57	29	48	42	42	42	39
20	64	106	87	67	66	69	40	54	57	54	47	46
21	60	70	49	48	51	52	35	40	39	41	44	40
22	69	51	38	56	52	42	36	41	36	47	52	31
23	59	73	47	67	50	57	44	28	37	41	33	35
24	44	73	56	50	44	56	35	29	34	49	47	38
25	50	67	81	59	50	39	40	35	44	43	45	40
26	60	49	53	56	67	53	49	34	28	35	45	42
27	43	69	54	34	53	44	30	38	46	53	45	39
28	46	71	62	61	50	62	45	36	38	49	41	28
29	42	62	48	53	42	41	41	44	39	39	39	25
30	82	0	39	49	48	44	48	49	36	43	34	25
31	57	0	56	0	30	0	38	36	0	47	0	28
月 总 和	182 7	196 1	1918	1758	1776	1517	1500	1366	1272	1461	1378	1321
年 总 和	19055											

表 A-4 2009 年发病人数统计

月 天	1	2	3	4	5	6	7	8	9	10	11	12
1	44	47	25	33	35	36	66	39	49	35	32	32
2	37	36	39	34	21	31	34	20	34	24	19	16
3	30	37	26	30	23	34	36	29	29	23	20	15
4	32	37	15	12	23	27	25	30	25	12	19	22
5	28	30	34	25	41	26	21	33	27	30	21	13
6	20	27	21	33	28	30	25	35	21	29	19	19
7	31	24	39	32	24	26	23	28	29	30	19	22
8	31	38	26	30	35	25	29	30	34	31	25	21
9	31	26	21	35	30	23	41	24	32	25	26	19

10	35	47	49	45	38	41	27	50	37	32	22	33
11	29	32	20	25	29	22	15	29	26	30	16	16
12	29	24	20	31	26	32	24	32	29	23	26	24
13	34	51	28	26	42	22	21	33	20	27	22	19
14	36	23	27	31	23	23	30	36	33	24	22	29
15	34	31	30	22	26	33	34	39	20	24	25	22
16	29	29	36	38	23	20	29	20	29	20	21	25
17	9	27	32	30	23	23	22	43	37	21	21	28
18	27	25	33	18	28	31	25	41	23	25	13	25
19	30	20	20	19	29	15	20	19	18	16	30	31
20	39	24	27	46	37	30	35	37	46	37	37	27
21	22	33	17	35	21	17	38	27	27	17	12	28
22	24	24	25	26	37	27	20	18	22	32	18	24
23	25	33	26	33	24	22	26	18	24	28	26	27
24	22	18	17	26	25	31	40	27	25	21	24	27
25	22	24	32	19	31	40	40	22	21	18	22	35
26	14	28	20	22	28	21	31	28	16	23	19	23
27	14	28	25	24	21	16	27	37	26	26	24	32
28	32	25	22	30	27	30	33	25	25	16	16	42
29	13	0	23	25	31	22	32	19	26	19	16	29
30	46	0	33	25	24	17	37	35	19	19	32	42
31	23	0	22	0	23	0	25	31	0	22	0	40
月总和	872	848	830	860	876	793	931	934	829	759	664	807
年总和	10003											

表 A-520010 年发病人数统计

月 天	1	2	3	4	5	6	7	8	9	10	11	12
1	127	93	84	84	97	85	91	83	94	88	93	75
2	64	45	61	69	66	47	62	64	68	63	52	43
3	47	45	52	60	59	43	63	61	65	47	52	60
4	51	41	58	61	68	44	55	44	47	70	42	46
5	67	58	61	44	65	54	61	63	41	54	65	50
6	55	58	49	69	63	47	49	51	67	57	55	45
7	46	35	56	51	58	55	53	66	59	64	53	50
8	58	57	62	56	64	57	54	61	60	66	46	49
9	54	46	45	56	49	56	50	58	42	51	54	36
10	64	46	89	72	85	73	67	57	70	63	64	47

11	63	45	51	39	58	54	53	58	52	57	54	39
12	66	57	65	61	62	47	57	49	41	56	56	33
13	48	39	53	65	61	43	52	47	57	72	40	45
14	54	33	47	71	48	51	59	44	48	52	41	44
15	67	55	71	46	58	55	70	48	54	55	56	38
16	60	45	53	55	46	48	59	72	60	46	57	28
17	56	58	57	45	61	52	41	54	49	35	47	56
18	60	58	43	35	72	56	40	48	51	56	65	38
19	41	52	48	51	50	49	61	42	42	54	71	22
20	60	75	79	77	85	61	79	58	62	75	62	37
21	63	52	38	47	62	67	46	46	41	58	43	29
22	48	61	63	55	56	56	57	45	41	45	49	26
23	42	64	61	79	60	51	52	56	62	35	47	28
24	48	54	44	47	58	48	54	53	47	40	42	22
25	56	55	60	41	52	54	56	63	48	61	42	24
26	51	61	47	56	66	50	53	47	52	63	53	21
27	60	51	44	48	52	43	58	46	52	49	30	13
28	48	48	49	60	46	51	48	40	37	56	39	9
29	55	0	47	48	61	61	43	41	60	43	46	8
30	41	0	44	51	40	52	63	68	63	54	49	6
31	40	0	43	0	54	0	51	47	0	33	0	4
月 总 和	176 0	148 7	1724	1699	1882	1610	1757	1680	1632	1718	1565	1071
年 总 和	19585											

附录 B 重要图片

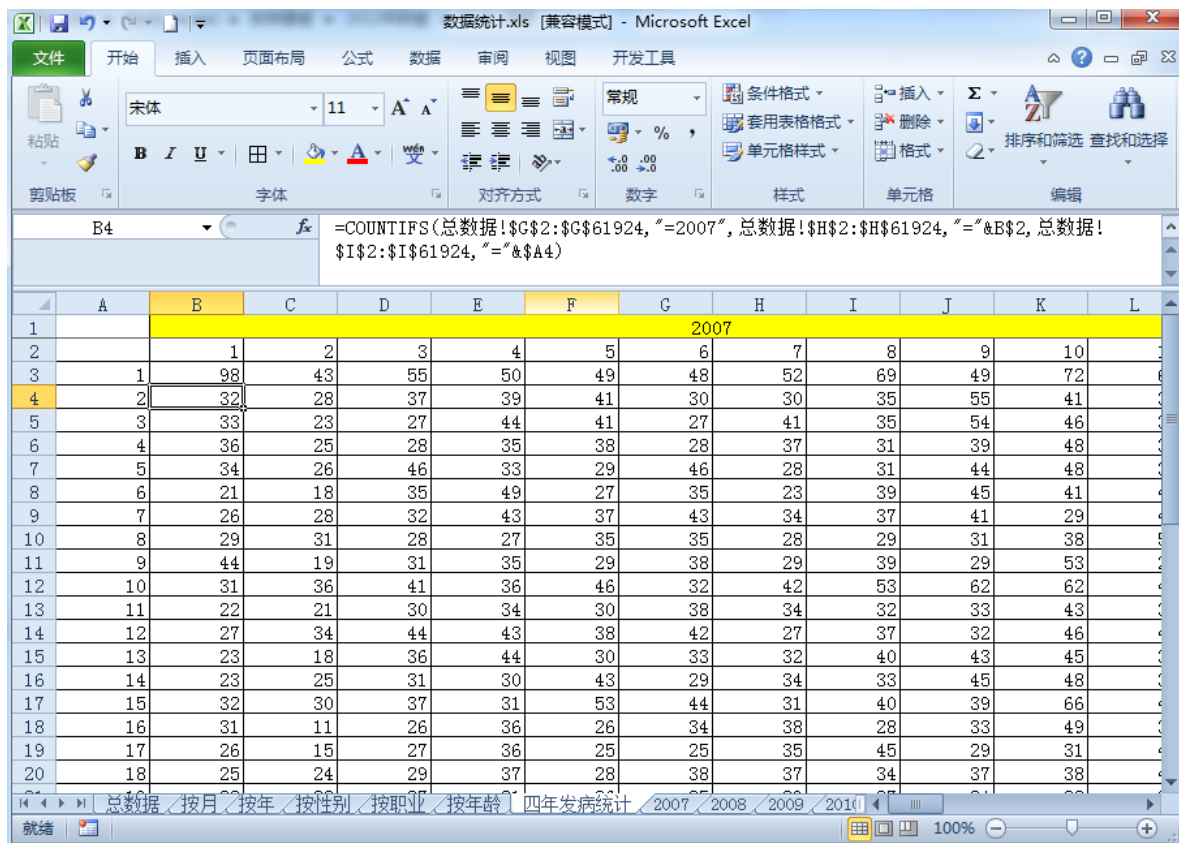


图 B-1 利用 EXCEL 进行编程处理数据的截图