

公路交通事故与气象条件关系 及其气象预警模型^{* 1}

罗 慧¹⁾ 李良序¹⁾ 胡 胜²⁾ John Nairn³⁾ 刘 宇¹⁾ 郑 磊⁴⁾

¹⁾(陕西省气象局,西安 710015) ²⁾(广州中心气象台,广州 110049)

³⁾(South Australia Regional Office, Kent Town, Adelaide, Australia 5067)

⁴⁾(陕西省公安厅,西安 710049)

摘 要

以西安地区为例,分析该地区 2002 年 1 月—2004 年 12 月连续 3 年逐日公路交通事故资料 1096 个样本,以及该地区、对应的逐日 13 个气象要素资料,并将样本数据划分为春夏和秋冬两个半年,来考虑气象要素在不同时段的不同影响,建立合理有效的公路气象预警模型。通过 2005 年 3 月—2006 年 4 月交通事故发生起数共 365 个测试样本的检验,发现这个模型具有较高的预测准确性,运用该地区气象要素建立公路交通事故的预警模型是可行而有效的。研究表明:西安地区气象要素中包含着影响和可以预测当地公路交通的信息,按照在 logistic 方程中的显著性大小和因子负载的大小,在春夏半年(4—9 月),影响西安地区公路交通事故相关因素依次为:能见度因子、相对湿度因子和降水因子;而在秋冬半年(10 月—次年 3 月),依次为温度因子、能见度因子、降水因子。

关键词: 公路气象; 因子分析; 多重共线性; logistic 回归分析

引 言

对公路交通的环境气象监测,开展公路的灾害性天气预报和预警,为高速公路的运营管理提供科学信息,既是我国气象部门预报预测业务体制改革中专业化气象预报业务的一种拓展,也是交通部门发展到一定阶段提出的必然要求。

Harold 等研究认为降水天气引发的交通事故是一般天气时的 2~3 倍,并且当降水天气发生在一段干燥天气之后时,这种危险将更大^[1]。Don 等通过对美国德克萨斯州 68 条高速公路的研究,在分析潮湿天气引起事故发生率有关的预测因素的基础上,提出了一个经验公式:潮湿天气安全指数 WARs^[2]。Daniel 调查研究了 1975—2000 年发生在美国的交通事故和降水之间的关系,结果表明:每月降水量和每月重大交通事故之间显示出负相关和显著性关系,结论认为降水的滞后效应可能缘于在干燥期汽油的累积效应使得路面变得光滑,还由于在降水路况下人们

驾驶得更为谨慎的缘故^[3]。Julia 等通过研究英格兰和威尔士天气和交通事故之间的关系,比较分析了雨、大雾、大风和良好天气对交通事故的影响,并找出了它们的影响程度比率。结果表明,大雾天气严重程度比率和地理环境有明显的联系,大风对交通事故的影响现在还没有定论^[4]。Symons 等研究发现^[5],降雨和潮湿路面是引起道路偶发事故增加的一种原因,其比率高达 70%。Kevin 等研究了澳大利亚墨尔本 1989—1996 年间天气要素对交通流量的影响,发现降水是相关最大的天气要素,在冬春季的雨天,当交通流量减少时有最大的影响^[6]。冯民学等人认为在高速公路沿线布设 AMW 自动气象监测仪的基础上做出低能见度的预测是可行和可能的^[7],贺芳芳等人分析了上海地区不良天气条件与交通事故之间的关系^[8],蒋建莹等人利用诊断分析和数值模拟结果,对一次弱降雪过程引起交通大阻塞的可能因素进行了探讨^[9]。

由此可见,道路交通事故受很多因素影响,不同研究区域和研究方法均表明气象要素对公路气象交

* 陕西省气象局 2005 年重点项目“陕西省交通气象预报服务系统”资助。

2006-04-28 收到,2006-12-31 收到再改稿。

通有不同的影响作用。基于这种认识,本文试图综合运用 13 类气象要素指标来研究天气条件对交通事故的影响关系。对这些气象指标首先进行因子分析,一方面提炼出其基本结构来反映这些原始信息的本质特征,另一方面通过数据降维简化克服其多重共线性。并根据其贡献选出若干公因子变量来替代原来的观测变量进行二值多元 logistic 回归,通过评价各个自变量对因变量产生的影响作用,建立公路气象预警数学模型,通过气象信息来预警公路上发生公路交通事故的可能性。

1 思路和方法

多元线性回归分析应用广泛,已成为标准的统计工具。但是多元线性回归中的某些统计假设非常关键,违反它将导致相当不合理的估计。比如在交通活动中涉及大量决策、行为或者意愿,需要研究哪些重要因素会影响这些决策或行动,这时模型的因变量是定性的或者作为定性观测的,再使用多元线性回归分析便不可避免地违反其许多重要假设条件,导致回归估计的推断存在严重误差^[10]。因此,本文采用 logistic 回归模型就适用于这种情况的研究,而且以往的研究也证明 logistic 回归的预测效果优于多元回归分析。

logistic 回归模型是对二分类因变量(即 $y = 1$ 或 $y = 0$)进行回归分析时最普遍使用的多元量化统计分析方法,又称增长函数,美国学者 Pearl 等在人口估计和预测中推广应用,并引起广泛注意^[11]。通过 logistic 模型将问题转化为根据样本数据使用最大似然估计法估计出各参数值,经过一定的数学推导运算,可求得响应变量取某个值的概率,即根据气象要素计算其在一定时间内影响公路交通事故的概率。当计算求出的概率大于某一设定值,就可以判断该公路交通气象安全指数的不同等级。logistic 回归模型的数学表达式为:

$$\begin{aligned} \text{logit}(y) &= \ln\left(\frac{p}{1-p}\right) \\ &= a_0 + a_1F_1 + a_2F_2 + \cdots + a_nF_n \end{aligned} \quad (1)$$

可等价地表示为:

$$p = \frac{\exp(a_0 + a_1F_1 + a_2F_2 + \cdots + a_nF_n)}{1 + \exp(a_0 + a_1F_1 + a_2F_2 + \cdots + a_nF_n)} \quad (2)$$

其中: $y = (1, 0)$ 表示某一事件发生的起数, $y = 1$ 表示发生, $y = 0$ 表示不发生。 $p = P(y = 1)$ 表示事件发

生的概率。 a_i 为待估参数, F_i 为自变量, $i = 1, \cdots, n$ 。

logistic 回归方程求解参数是采用最大似然估计的方法,因此其回归方程的整体检验通过似然函数值,即表达的是一种概率,是在假设拟合模型为真实情况时能够观察到这一特定样本数据的概率,因此所求的函数值处于 $[0, 1]$ 之间^[12]。和其他多元回归方法一样,logistic 回归模型对多重共线性敏感,当变量之间相关程度较高时,样本的较小变化将会带来系数估计的较大变化,从而降低模型的精度。而各项气象要素,比如平均温度、平均地面温度、最高最低温度,或同一天中不同时次的能见度值,它们之间的相关程度往往较高。为了有效消除变量之间的多重共线性影响,一个简单办法是从模型中删除某些变量,但这样又会损失较多的气象信息,为此,本文将首先对气象指标进行因子分析。黄嘉佑等人提出用主分量逐步筛选因子典型相关分析作短期气候的预测方法,结果表明通过物理因子提取和分析可在预测中产生一定的效果^[13]。

因子分析就是研究如何以最少的信息丢失把众多的观测变量浓缩为少数几个公因子的一种多元统计方法,其作用主要有两方面:第一是寻求基本结构,通过找到较少的几个公因子既能代表数据的基本结构,又能反映信息的本质特征;第二是数据降维简化,同时有效克服多重共线性问题。这样,通过进一步把原始观测变量的信息转换成这些因子的因子值,然后根据其贡献选出若干公因子变量来替代原来的观测变量进行 logistic 回归。

2 样本选取和方法应用

与陕西省公安厅交通管理部门合作,在全省范围内共选取了有代表性的 4 个地市,分别位于陕北黄土高原、关中平原、陕南秦巴山区,代表了不同的气候特征,收集了各地 2002 年 1 月—2004 年 12 月逐日公路交通事故数据,包括交通事故起数、交通事故伤亡人数以及交通事故造成的经济损失等。其中,交通事故是指车辆在道路上因过错或者意外造成的人身伤亡或者财产损失的事件;事故形态分为:正面相撞、侧面相撞、尾随相撞(追尾)、对向刮擦及其他,机动车碰撞是道路交通事故的主要形态;数据中所指的公路,依据《中华人民共和国公路法》中第六条规定,包括了按技术等级分为一级到四级的全部公路。在本文中主要应用了交通事故起数进行分

析。同时,选取西安地区为代表,通过对西安地区气象要素对交通事故影响规律的研究,为进一步研究陕西省公路气象问题打下基础。选取 2005 年 3 月—2006 年 4 月西安地区的交通事故发生起数作为测试样本进行检验。

依据西安市公路交通事故起数分月同比的时间分布特征(图 1),2004 年的事故起数明显上升,秋冬半年的事故起数普遍高于春夏半年。依据交通部门的统计显示,机动车碰撞是陕西省公路交通事故的主要形态,以 2003 年为例,发生的正面相撞事故、侧面相撞事故起数和尾随相撞事故起数分别占总数的 19.95% ,43.86% 和 16.59% 。其直接事故原因所占比例从高到低依次为:因制动失效、制动不良、转向失效等机件故障引发的交通事故占总数的 14.89% ,其相应的致死率为 31.6% ,超出事故平均致死率 13% ,致伤率为 81.4% ,超出事故平均致伤率 6% ;因超速行驶引发事故占总数的 4.54% ;因不按规定让行引发事故占总数的 9.17% ;因纵向间距不够引发事故占总数的 7.50% ;因违章占道行驶引发事故占总数的 7.39% 。公路管理部门发现当机动车在公路上行驶,遇有高温、雾、雨、雪、沙尘、冰雹等高敏感性气象条件时,超速行驶,雾灯、近光灯和示廓灯等的不当开启、没有保持必要车距等都是诱发交通事故的直接原因。

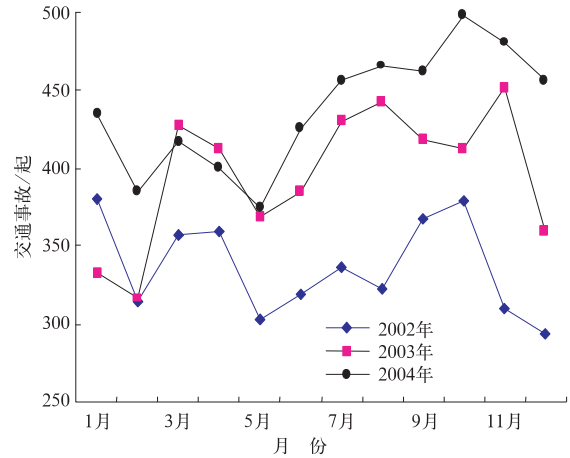


图 1 2002—2004 年西安地区公路交通事故起数分月同比

Fig. 1 Comparison of traffic crashes data of each month in Xi'an area from 2002 to 2004

西安地区 2002 年 1 月到 2004 年 12 月逐日公路交通事故起数资料有效样本共 1096 个,通过计算得出事故发生的众数,即在一系列的数据资料中出现频次最多的值(图 2),得到春夏半年事故发生起数的众数为 16 起(出现了 47 次);秋冬半年事故发生起数的众数为 14 起(出现了 48 次)。按照高于众数,被认为事故高发,赋值为 1;低于众数,被认为事故低发,赋值为 0,从而将 logistic 模型因变量 Y 变成二分定性变量。

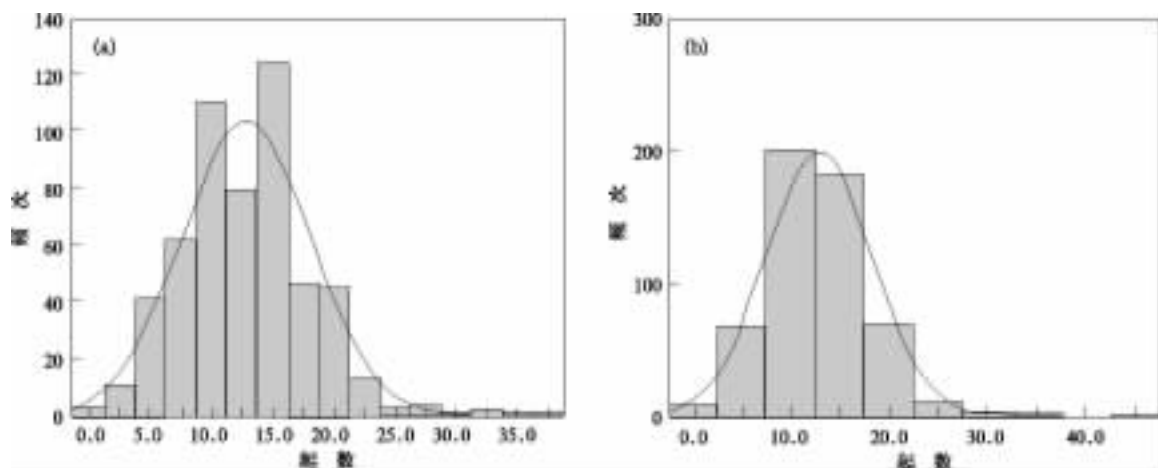


图 2 西安地区 2001 年 1 月—2004 年 12 月交通事故众数直方图

(a) 春夏半年, (b) 秋冬半年

Fig. 2 Mode of traffic crashes data in spring-summer half-year (a) and autumn-winter half-year (b) in Xi'an area from January, 2001 to December, 2004

同时选取了西安地区对应年份逐日的气象观测资料,共有 13 类气象要素指标,分别为:日平均气温 X_1 、日最高气温 X_2 、日最低气温 X_3 、日降水量 X_4 、日平均气压 X_5 、日平均相对湿度 X_6 、02:00(北京时,下同)定时能见度 X_7 、08:00 定时能见度 X_8 、14:00 定时能见度 X_9 、20:00 定时能见度 X_{10} 、日平均地面温度 X_{11} 、日最高地面温度 X_{12} 、日最低地面温度 X_{13} 。根据气象要素满足全面性、可比性以及同趋势性等原则,也根据气象要素预报的特点,即在秋冬半年不做最高气温预报而在春夏半年不做最低气温预报,将样本数据人为划分为两个半年:每年的 4—9 月,笼统称为春夏半年,共计 549 个样本,去掉变量日最低气温 X_3 和日最低地面温度 X_{13} ;10 月至次年的 3 月,笼统称为秋冬半年,共计 547 个样本,去掉变量日最高气温 X_2 和日最高地面温度 X_{12} 。同时将每天 4 个时次的能见度进行加权平均,得到综合能见度 X_{17} 。通过以上数据预处理,分别得到春夏半年的 549 个样本,8 类气象指标;秋冬半年的 547 个样本,8 类气象指标。

在实证分析中,将应用因子分析方法,其主要目的是通过研究众多变量之间的内部依赖关系,对两个半年的数据进一步降维简化分析,最终计算出少数几个公因子,夏半年 F_1 以及冬半年 H_1 ,这些整合因子并非原始变量的简单取舍,而是通过新的综合找到几个可以影响原始变量的共同因素,因子变量之间是近乎独立的,具有更高的可解释性,可以有效地克服原始变量之间的多重共线性和序列的自相关性^[14]。在数据处理工具上,主要应用了 SPSS 12.0(Statistical Product and Service Solution),这是国际上流行的视窗统计软件包之一,是一种功能强大的高级统计分析软件系统^[15]。

3 实证分析

3.1 因子分析

对 SPSS 12.0 软件进行因子分析,得到各因子的方差贡献率(表 1)。其中,对于提取出的各个因子在因子结构中的重要性用因子的方差贡献率来表示,某因子的方差贡献率是该因子可以解释全部变量的方差总和的百分率。方差贡献率越大的因子越重要,所有因子的方差贡献率的总和越大时,所提取的因子对观测变量的代表性越强。

从表 1 中可以看到,在春夏半年前 4 个因子的

累积贡献率达到 93.530%,即 4 个因子已经包含了原始指标的 93.530% 的信息。同样,在秋冬半年,前 4 个因子已经包含了原始指标的 95.242% 的信息。因此这 4 个因子就是公因子,即各个变量所共有影响的因子,可以解释变量之间的相关,用它们完全可以代替原来的指标。

表 1 春夏半年各因子的方差贡献率(单位:%)

Table 1 Total variance explained of each factor (unit: %)

| 因子 | 方差贡献率 | 累计方差贡献率 |
|----|-------|---------|
| | F_1 | 61.114 |
| 春 | F_2 | 17.286 |
| 夏 | F_3 | 10.433 |
| 半 | F_4 | 4.697 |
| 年 | F_5 | 3.746 |
| | F_6 | 1.904 |
| | H_1 | 52.271 |
| 秋 | H_2 | 23.653 |
| 冬 | H_3 | 12.328 |
| 半 | H_4 | 6.990 |
| 年 | H_5 | 3.031 |
| | H_6 | 1.137 |

3.2 因子解释

得到最后因子解后,解释因子主要是借助于因子负载矩阵,需要得到原始的 13 个气象要素对这些公因子的因子负载矩阵,因子负载矩阵见表 2 及表 3。因子负载是因子分析模型中最重要一个统计量,它不仅表示了观测变量是如何由因子线性表出的,而且反映了公因子和变量之间的相关程度,其值越大说明它们之间的关系越密切。找出在每个因子上有显著负载的变量,根据这些变量的意义给因子一个合适的名称,具有较高负载的变量对因子名称的影响更大。从表 2、表 3 中可以看到,春夏、秋冬半年各公因子与原始气象要素因子的相关系数为:

表 2 春夏半年旋转后因子负载矩阵

Table 2 Rotated component matrix of each factor in spring-summer half-year

| | F_1 | F_2 | F_3 | F_4 |
|---|----------|--------|---------|---------|
| | X_1 | 0.950 | 0.113 | 0.146 |
| | X_2 | 0.900 | 0.207 | 0.283 |
| 春 | X_4 | -0.102 | -0.074 | -0.160 |
| 夏 | X_5 | -0.893 | -0.143 | 0.09379 |
| 半 | X_6 | -0.172 | -0.488 | -0.768 |
| 年 | X_{17} | 0.183 | 0.940 | 0.235 |
| | X_{11} | 0.928 | 0.04908 | 0.291 |
| | X_{12} | 0.722 | 0.145 | 0.591 |

表3 秋冬半年旋转后因子负载矩阵

Table 3 Rotated component matrix of each factor in autumn-winter half-year

| | H_1 | H_2 | H_3 | H_4 |
|----------|---------|----------|----------|----------|
| X_1 | 0.928 | 0.141 | -0.019 | -0.293 |
| X_3 | 0.973 | 0.008885 | 0.07027 | -0.174 |
| 秋 | | | | |
| X_4 | 0.101 | -0.132 | 0.980 | 0.005357 |
| 冬 | | | | |
| X_5 | -0.426 | 0.08519 | 0.008481 | 0.888 |
| 半 | | | | |
| X_6 | 0.0204 | -0.899 | 0.262 | 0.128 |
| 年 | | | | |
| X_{X7} | 0.06874 | 0.922 | 0.05518 | 0.193 |
| X_{11} | 0.945 | 0.115 | 0.005211 | -0.243 |
| X_{13} | 0.940 | -0.172 | 0.187 | -0.048 |

① 公因子 F_1 在 X_1, X_2, X_{11}, X_{12} 上有较大的负载值,所以 F_1 主要由日平均气温、日最高气温、日平均地面温度、日最高地面温度以及日平均气压这几个气象要素指标解释,它可以代表温度类要素对公路交通的影响,可命名为“温度因子”。公因子 H_1 在 X_1, X_3, X_{11}, X_{13} 上有较大的负载,所以 H_1 主要由日最低气温、日平均地面温度、日最低地面温度和日平均气温这几个气象要素指标解释,它代表的是温度类要素(最低温度)对公路交通的影响,因此因子 H_1 也可以命名为温度因子。

② 公因子 F_2 在 X_{X7} 上有较大的负载值,所以 F_2 主要由4个时次能见度的平均值这个气象指标来反映,它代表的是能见度要素对交通的影响,可命名为“能见度因子”。公因子 H_2 也在 X_{X7} 上有较大的负载,所以由4个时次能见度的平均值这个气象指标来反映,它代表的是能见度要素对交通的影响,因此因子 H_2 也是能见度因子。

③ 公因子 F_3 在 X_6 上有较大的负载值,所以 F_3 主要由日平均相对湿度这个气象指标来反映,它代表的是相对湿度对交通的影响,可命名为“相对湿度因子”。公因子 H_3 在 X_4 上有较大的负载,所以 H_3 主要由日降水量这个气象指标来反映,它代表的是秋冬半年降水要素对公路交通的影响,因子 H_3 是降水因子。

④ 公因子 F_4 在 X_4 上有较大的负载值,所以 F_4 主要由日降水量这个气象指标来反映,它代表的是降水要素对公路交通的影响,可命名为“降水因子”。公因子 H_4 在 X_5 上有较大的负载,所以 H_4 主要由日平均气压这个气象指标来反映,它代表的是气压要素对公路交通的影响,因子 H_4 是“气压因子”。

因此,在春夏半年的影响公因子,按照影响大小依次为温度因子、能见度因子、相对湿度因子和降水

因子。而在秋冬半年的影响公因子,按照影响大小依次为温度因子、能见度因子、降水因子和气压因子。

在确定了各公因子所代表的气象信息意义之后,需要得到 $F_i (i=1, \dots, 4), H_i (i=1, \dots, 4)$ 因子关于原始气象要素数据的线性表达式,可以从公因子得分系数矩阵中得到。从表4、表5中分别可以看到春夏、秋冬半年各公因子得分系数。比如 F_1 的表达式为

$$F_1 = 0.29X_1 + 0.223X_2 + 0.034X_4 - 0.346X_5 + 0.209X_6 - 0.047X_{X7} + 0.237X_{11} + 0.06X_{12}。$$

表4 春夏半年因子得分系数矩阵

Table 4 Component score coefficient matrix of each factor in spring-summer half-year

| | F_1 | F_2 | F_3 | F_4 |
|----------|--------|--------|--------|--------|
| X_1 | 0.290 | -0.002 | -0.164 | 0.000 |
| X_2 | 0.223 | 0.013 | 0.020 | 0.036 |
| 春 | | | | |
| X_4 | 0.034 | 0.001 | 0.294 | 1.071 |
| 夏 | | | | |
| X_5 | -0.346 | -0.240 | 0.604 | 0.079 |
| 半 | | | | |
| X_6 | 0.209 | -0.022 | -0.868 | -0.100 |
| 年 | | | | |
| X_{X7} | -0.047 | 1.074 | -0.432 | 0.019 |
| X_{11} | 0.237 | -0.210 | 0.168 | 0.067 |
| X_{12} | 0.060 | -0.318 | 0.668 | 0.090 |

表5 秋冬半年因子得分系数矩阵

Table 5 Component score coefficient matrix of each factor in autumn-winter half-year

| | H_1 | H_2 | H_3 | H_4 |
|----------|--------|--------|--------|--------|
| X_1 | 0.241 | 0.056 | -0.074 | -0.033 |
| X_3 | 0.306 | -0.025 | -0.044 | 0.157 |
| 秋 | | | | |
| X_4 | -0.111 | 0.135 | 0.998 | -0.161 |
| 冬 | | | | |
| X_5 | 0.212 | -0.030 | -0.110 | 1.101 |
| 半 | | | | |
| X_6 | 0.085 | -0.522 | 0.039 | 0.253 |
| 年 | | | | |
| X_{X7} | 0.038 | 0.561 | 0.214 | 0.180 |
| X_{11} | 0.268 | 0.038 | -0.068 | 0.044 |
| X_{13} | 0.342 | -0.126 | 0.015 | 0.324 |

3.3 logistic 模型

将上面分别得到的4个公因子作为变量,利用SPSS软件进行logistic回归。这里选择全部纳入法作为变量进入模型的方式,即将所有的自变量全部纳入回归^[16]。最终得到如下结果(表6),其中, B 表示计算得到的回归系数估计值; S_E 表示回归系数的标准误差,反映的不是变量的实际误差,而是对变量列做出可靠性的估计,反映了数据的离散程度^[17];Wald表示Wald检验值,该统计量是用来检验偏回归系数显著程度的,值越大表明该自变量的作用越显著;Sig则表示显著性水平,给定显著性概率 $\alpha=0.1$,则 $1-\alpha$ 表示置信水平,表示以90%的可靠度保证真值落在某一区间;给定显著性概率 $\alpha=$

0.05, 则 95% 表示置信水平, 表示以 95% 的可靠度保证真值落在某一区间。

通过表 6 可以看到, F_1 的显著性为 0.443, 没有通过检验, 说明在 logistic 方程中的该温度因子变量与因变量之间没有显著关系。同样, H_4 的显著性为 0.204, 也没有通过检验, 说明在 logistic 方程中该气压因子变量与因变量之间没有显著关系。因此, 在构建方程时这两个公因子予以舍弃。从而得

表 6 进入 logistic 回归方程的变量
Table 6 Variables in the logistic equation

| | | B | S_E | Wald | Sig |
|----------|-------|--------|-------|--------|---------|
| 春夏 半年 | F_1 | 0.074 | 0.096 | 0.589 | 0.443 |
| | F_2 | 0.164 | 0.093 | 3.123 | 0.077* |
| | F_3 | -0.198 | 0.096 | 4.269 | 0.039** |
| | F_4 | 0.179 | 0.090 | 3.929 | 0.047** |
| | 常数 | -0.872 | 0.095 | 84.403 | 0.000** |
| 秋冬 半年 | H_1 | 0.288 | 0.089 | 10.382 | 0.001** |
| | H_2 | 0.155 | 0.089 | 3.012 | 0.083* |
| | H_3 | 0.405 | 0.127 | 10.170 | 0.001** |
| | H_4 | 0.114 | 0.090 | 1.615 | 0.204 |
| | 常数 | -0.281 | 0.089 | 9.881 | 0.002** |

注: ** 表示通过 0.05 的显著性检验, * 表示通过 0.1 的显著性检验。

到预警模型为:

$$\ln\left(\frac{p_1}{1-p_1}\right) = -0.872 + 0.164F_2 - 0.198F_3 + 0.179F_4 \quad (3)$$

$$\ln\left(\frac{p_2}{1-p_2}\right) = -0.281 + 0.288H_1 + 0.155H_2 + 0.405H_3 \quad (4)$$

即

$$p_1 = \frac{\exp(-0.872 + 0.164F_2 - 0.198F_3 + 0.179F_4)}{1 + \exp(-0.872 + 0.164F_2 - 0.198F_3 + 0.179F_4)}$$

$$p_2 = \frac{\exp(-0.281 + 0.288H_1 + 0.155H_2 + 0.405H_3)}{1 + \exp(-0.281 + 0.288H_1 + 0.155H_2 + 0.405H_3)}$$

式(3), (4)中, p_1 及 p_2 分别表示春夏和秋冬半年由于气象条件影响发生交通事故的概率, F_i ($i=1, \dots, 4$) 和 H_j ($j=1, \dots, 4$) 分别表示用来拟合模型的公因子得分。在模型拟合之前, 高于交通事故发生众数值, p_1 和 p_2 赋值为 1, 否则取 0。根据所得到 logistic 方程, 以 0.5 为最佳判定点对原始数据进行回判判定, 如果得到的 p_1 和 p_2 值大于 0.5, 判定交通事故发生, 否则判定交通事故不发生。

选取 2005 年 3 月—2006 年 4 月的交通事故发生起数作为测试样本, 其中 2005 年 4—9 月为春夏半年, 共计 183 个样本, 2005 年 10 月—2006 年的 3 月为秋冬半年, 共计 182 个样本。得到 logistic 回归

的判定结果。在春夏半年中, 在 96 个低于事故发生平均水平的样本中, 有 2 个被错判, 判定正确率为 96.2%, 在 48 个高于事故发生平均水平的样本中, 有 37 个被错判, 判定正确率为 56.47%, 总体正确率达到 79.12%。而在秋冬半年中, 在 99 个低于事故发生平均水平的样本中, 有 4 个被错判, 判定正确率为 96.12%, 在 50 个高于事故发生平均水平的样本中, 有 29 个被错判, 判定正确率为 63.29%, 总体正确率达到 81.87%。道路交通受很多因素影响, 比如路况、车流量、车况、司机驾驶技术及身体状态等, 交通事故的发生具有复杂性和随机性, 本文考察的这些气象变量还不足以描述交通事故的本质特征, 但该模型仍然显示出具有较好的预测正确率, 其中, 对秋冬半年的预测准确率要高于春夏半年的预测准确率。

4 小结与讨论

1) 通过研究发现西安地区的气象要素指标中包含着影响和可以预测公路交通安全的信息含量, 即可以运用该地区气象要素建立公路交通事故的预警模型, 根据得到的模型, 可以有效提升公路交通在恶劣天气条件下的安全能力, 为进一步实现恶劣天气条件下公路运营决策管理提供科学依据。

2) 因子负载的大小反映了公因子和各个变量之间的相关程度, 在春夏半年, 影响西安地区公路交通事故相关因素依次为: 温度因子、能见度因子、相对湿度因子和降水因子; 而在秋冬半年, 依次为温度因子、能见度因子、降水因子和气压因子。但由于春夏半年的温度因子和秋冬半年的气压因子均没有通过显著性检验, 最终在构建 logistic 方程时, 对这两个因子予以舍弃。

3) 通过分类表中样本的判定结果可知, 本文所构建的公路气象预警模型对低发事故的准确率要高于对高发事故的准确率; 同时对秋冬半年的预测准确率要高于春夏半年的预测准确率, 说明气象要素对交通事故发生的影响, 在秋冬半年其相关性较春夏半年强一些。

在交通气象业务应用和对公众作等级发布中, 可以按照得到的 p_1 值以及 p_2 值分为 5 个等级进行预警, 即以 0.2 为判定间距进行判定。在与交通部门合作联合发布预警时, 当机动车在公路上行驶, 遇有高温、雾、雨、雪、沙尘、冰雹等高敏感性气象条件

时,建议公路管理部门应当通过显示屏等方式发布速度限制、开启雾灯、近光灯、示廓灯和前后位灯、保持车距等必要的提示信息。

另外,温度因子主要影响在秋冬半年,地面结冰则严重影响路况,会引发制动失效、制动不良等机件故障,油箱一旦冻结也会影响车况,进而影响交通安全;能见度因子对全年都有影响,大雾、沙尘天气、大风等导致能见度降低,影响驾驶员的视线视野,增大交通事故发生的概率;相对湿度因子则主要影响在春夏半年,当最低气温高、日较差很小和相对湿度大时,人体通过汗液蒸发调节体温的能力急剧下降,造成驾驶员身体不适乃至中暑,进而严重影响安全驾驶;降水因子在全年都有影响,毛毛雨或者降雪结冰会导致路面摩擦系数大大减少、影响路况,而持续降水、阴雨连绵也会影响驾驶员的视线以及精神状态。当天气变化时(大致分为周期性缓慢变化和非周期性剧烈变化两种),上述气象因子不仅能影响路况、车况,还将通过影响驾驶员的精神状态和行为进而影响交通安全。有关定量的影响机制的研究,还需要进一步探讨。

参考文献

- [1] Harold Brodsky, Hakkert A Shalom. Risk of a road accident in rainy weather. *Accident Analysis and Prevention*, 1988, 20(3): 161-176.
- [2] Don L Ivey, Griffin I Lindsay, Newton M Tommy, et al. Predicting wet weather accidents. *Accident Analysis and Prevention*, 1981, 13: 83-99.
- [3] Eisenberg Daniel. The mixed effects of precipitation on traffic crashes. *Accident Analysis and Prevention*, 2004, 36: 637-647.
- [4] Edwards B Julia. The relationship between road accident severity and recorded weather. *Journal of Safety Research*, 2002, 29(4): 249-262.
- [5] Symons L, Perry A. Predicting road hazards caused by rain, freezing rain and wet surfaces and the role of weather radar. *Meteorol Appl*, 1997, 4: 17-21.
- [6] Kevin Keay, Ian Simmonds. The association of rainfall and other weather variables with road traffic volume in Melbourne, Australia. *Accident Analysis and Prevention*, 2005, 37: 109-124.
- [7] 冯民学,袁成松,卞光辉,等.沪宁高速公路无锡段春季浓雾的实时监测和若干特征. *气象科学*, 2003, 23(4): 435-445.
- [8] 贺芳芳,房国良,吴建平,等.上海地区不良天气条件与交通事故之关系研究. *应用气象学报*, 2004, 15(1): 127-128.
- [9] 蒋建莹,史历,倪允琪.一次"高影响天气"的弱降雪过程的数值研究. *应用气象学报*, 2005, 16(2): 231-237.
- [10] 郭志刚. 社会统计分析方法——SPSS 软件应用. 北京:中国人民大学出版社, 2004: 54; 177-197.
- [11] Shryock S Henry, Siegel S Jacob. *The Methods and Materials of Demography*. New York: Academic Press, 1976: 215-216.
- [12] Varela G, Novoa N, Jimenez M F, et al. Applicability of logistic regression (LR) risk modeling to decision making in lung cancer resection. *Interactive Cardiovascular and Thoracic Surgery*, 2003, (2): 12-15.
- [13] 黄嘉佑,黄茂怡.主分量逐步筛选因子典型相关分析及其预报试验. *应用气象学报*, 2000, 11(增刊): 73-79.
- [14] 洪楠. SPSS for windows 统计产品和服务方案教程. 北京:清华大学出版社;北京:北方交通大学出版社, 2003: 217-294.
- [15] 何晓群. 现代统计分析方法与应用. 北京:中国人民大学出版社, 1998: 32; 316-342.
- [16] Ingo Hary. Analysis of survival curves in seasonally mated pastoral goat herds in northern Kenya using logistic regression techniques. *Journal of Arid Environments*, 2002, 50: 621-640.
- [17] 宁敏东. 标准误差的整体认识及运用. *国防科技大学学报*, 1994, 16(3): 131-134.

The Relationship Between Road Traffic Crashes and Meteorological Condition with Construction of Its Road Weather Warning Model

Luo Hui¹⁾ Li Liangxu¹⁾ Hu Sheng²⁾ John Nairn³⁾ Liu Yu¹⁾ Zheng Lei⁴⁾

¹⁾ (Shaanxi Provincial Meteorological Bureau, Xi'an 710015)

²⁾ (Guangzhou Central Meteorological Office, Guangzhou 110049)

³⁾ (South Australia Regional Office, Kent Town, Adelaide, Australia 5067)

⁴⁾ (Shaanxi Provincial Public Security Department, Xi'an 710049)

Abstract

In order to construct a valid road weather warning model, the effect of weather on the daily number of all reported successive traffic accidents day by day from January 2002 to December 2004 (1096 valid samples) in the area of Xi'an city is examined. Thirteen meteorological elements sampled over the same area day by day are applied for the corresponding three years. The 1096 sample days are divided into two half-year parts given different seasonal regimes of the meteorological elements. Data are processed using statistical product and service solution (SPSS) version 12.0.

Factor analysis methods are utilized to summarize four public factors, which are $F_i (i = 1, \dots, 4)$ and $H_i (i = 1, \dots, 4)$, integrated by temperature, visibility, relative humidity and rainfall in spring-summer half-year period and temperature, visibility, rainfall and pressure respectively in autumn-winter half-year period. Multi-collinearity is effectively overcome by this process and the meteorological variables are reduced from thirteen elements to four public factors.

Logistic regression analysis based on these four public factors is applied. Initially the model is calculated, the number in the set that occurs most frequently in all traffic crashes data is obtained. Then it is supposed that if the raw crash data are greater than the mode value 1 is set to this variable and targets the high crashes frequency; if the raw data are smaller than the mode value 0 is applied to this variable and targets low crashes frequency. Binary logistic by entrance option is applied to get the road weather warning model over the two half-year periods. But F_1 (temperature factor in spring-summer period) and H_4 (pressure factor in autumn-winter period) are shown to be insignificant, so they are abandoned when the logistic regression equations are constructed.

The model is evaluated for preferable prediction accuracy through samples tests. The results show that the meteorological factors used to indicate their influences and forecast traffic crashes in Xi'an city result in a suitable and effective road weather warning model.

In summary, according to significance level in the logistic regression equation and the value of factor loadings, the factors are respectively integrated by visibility, relative humidity and rainfall in spring-summer half-year period and temperature, visibility and rainfall in autumn-winter half-year period.

Key words: road weather; factor analysis; multi-collinearity; logistic regression