

流行性感冒爆发气象条件主成分分析

刘庚山 郭安红 安顺清

(中国气象科学研究院农业气象和遥感应用研究中心, 北京 100081)

摘要 文章对1970~1989年北京、天津和石家庄市流行性感冒爆发期间温度、湿度、降水量等9个旬平均气象条件进行主成分分析,前3个主成分的方差贡献率占总方差的81.94%。前3个主成分的特征向量表明,利于流行性感冒病毒传播流行的气象条件依次是:①旬平均湿度低、降水和蒸发量小(第一主成分方差贡献率占总方差的37.8%),②旬平均风速和蒸发量小、辐射较弱(第二主成分方差贡献率占总方差的25.37%);③气压和地表温度升高(第三主成分方差贡献率占总方差的18.77%)。

关键词 流行性感冒 旬平均气象条件 主成分分析

引言

人们生活在大气的底层,身体健康状况和感觉舒服程度均与气象条件密切相关。一年之中人体有热、炎热、酷热、温暖、凉爽、寒冷、严寒等感觉,周而复始。这些感觉主要是人体对气象场中的温、湿度及风速等的综合感受。

流行性感冒(本文指大范围的流行性感冒爆发流行,以下简称流感)的传播也与气象条件有关,对这类问题的统计分析,以往多用回归分析和相关分析,这些方法往往要事先指定待考察的因素,严格控制其他因素的均匀性,并且是在固定某几个变量的前提下进行分析其余变量的关系,得到的相互关系因不包含未指定的因素以及固定了的那些变量的作用而有局限性。而多因素分析为探索和阐明这类问题提供了新的手段,是研究相关变量共同关系的技术,由于同时考虑了诸多变量,因此能解释用单因素分析不能说明的问题,得出比单一变量对比分析更强有力的结论。

主成分分析(PCA: Principal Component Analysis)方法是被广泛采用的多因子分析方法,该方法具有两个特点:①不需要有关待分析样品或变量的任何先验信息;②以尽可能少的潜在变量表征全部测量数据。所以,PCA方法是一种变量简化技术,对变量的属性没有严格要求,只要样品达到一定数量

又具有独立性、代表性,则相同属性的变量、相近特征的样本会根据各自在潜在变量(即主成分PC)空间中的位置显示出清晰的相互关系。

文献[1~6]分别介绍了应用主成分分析方法研究作物产量、雪灾、家畜疾病等与气象条件的关系。文献[2]指出:积雪能否对牧区造成灾害是由多因子综合作用的结果,且不同雪灾因子所起作用的大小是不相同的;该文阐述了西藏那曲牧区的雪灾因子体系,并用主成分分析法对8个雪灾因子进行了分析。结果表明:一场积雪的最大积雪深度、积雪日数、降水总量和最大风速是主成分,其结果有助于雪灾判别及预测等应用模型的建立。

张绳祖等^[5]从家畜疫病生态学的研究角度出发,探讨了甘肃省平凉地区1975~1989年猪丹毒传染病与几个气候因子间的关系,并用多因子主成分分析法分别求出各个因子的主成分和方差,为研究和预测该病在此区域内的发生、发展和消亡趋势以及今后的防疫接种工作提供了科学依据。

吕厚东等^[6]研究了流感和太阳黑子之间的关系,并指出甲型流感病毒各亚型的抗原变异和流行年代与太阳黑子活动周期的极大年和极小年有密切关系。

本文通过主成分分析方法,分析华北三城市(北京、天津、石家庄)流感传播与天气条件的关系,从气象条件方面探索流感流行的规律,为流感的发生、发展趋

势预测和防疫工作提供科学依据。

1 资料与方法

在国家计划生育委员会科研所计算机室的帮助下,利用国家“八五”攻关项目子专题“我国北方人群流感规律研究”收集的部分流感样品和相应的气象资料,流感样品选自北京、天津和石家庄三城市,样本数量 17。三地样本资料时段为 1973 年至 1974 年 1 月,1976 年 11~12 月,1977 年 11 月至 1978 年 2 月,1986 年 11~12 月,1988 年 11 月至 1989 年 1 月及北京和石家庄 1982 年 11~12 月。

流感判别和大范围流行的标准以国家计划生育委员会科研所该项研究课题组制定的标准为准。

与流感样品相应的气象资料均为旬平均值(共 9 个):气温($T: ^\circ\text{C}$)、相对湿度($f: \%$)、太阳辐射能($R_s: \text{J}\cdot\text{m}^{-2}$)、降水量($P: \text{mm}$)、风速($V: \text{m}\cdot\text{s}^{-1}$)、蒸发量($E: \text{mm}$)、地表(0 cm)温度($T_0: ^\circ\text{C}$)、地下 10 cm 地温($T_{-10}: ^\circ\text{C}$)、气压($p: \text{hPa}$)。

因为 9 个变量的量纲不同,须对原始气象资料进行标准化处理。

设旬平均气象资料为 X_{ij} ,其中 $i=1,2,\dots,17$, $j=1,2,\dots,9$ 分别代表 T 、 f 、 R_s 、 P 、 V 、 E 、 T_0 、 T_{-10} 和 p 。经标准化处理后,得到新变量

$$Z_{ij} = \frac{X_{ij} - \overline{X_j}}{S_{ij}} \tag{1}$$

式中 $\overline{X_j} = \frac{1}{n} \sum_{i=1}^n X_{ij}$, $S_j^2 = \frac{1}{n-1} \sum_{i=1}^n (X_{ij} - \overline{X_j})^2$, $n=17$ 。

按主成分分析步骤^[7,8]计算 Z_{ij} 的相关矩阵 R 及 R 的特征值($\lambda_1, \lambda_2, \dots, \lambda_9$),从而计算 λ 的方差贡献率和累积方差贡献率及每个特征值对应的特征向

量(E_j)。利用公式:

$$Q_j = \sum E_j Z_{ij} \tag{2}$$

计算主成分 Q_j 。表 1 给出 R 的特征值及其方差贡献率和累计方差贡献率。

表 1 R 的特征值及其贡献率和累计方差贡献率

特征值序号	特征值	方差贡献率/ %	累计方差贡献率/ %
λ_1	5.367	37.80	37.80
λ_2	3.602	25.37	63.17
λ_3	2.665	18.77	81.94
λ_4	0.903	6.36	88.30
λ_5	0.723	5.09	93.39
λ_6	0.422	2.97	96.36
λ_7	0.201	1.42	97.78
λ_8	0.192	1.35	99.13
λ_9	0.123	0.87	100.0

2 结果分析

2.1 第一特征向量和第一主成分

Z_{ij} 的主成分分析表明:第一主成分方差贡献率占总方差的 37.80%,图 1 给出气象条件主成分分析的第一特征向量。图 1 显示旬平均湿度、降水量和蒸发量为较大的正贡献,反映了在影响流感传播的气象环境因子中,它们极可能是决定流感爆发的最主要的气象环境因子,降水和蒸发(与湿度变化具有一致性)减少,湿度也会随之降低(环境变干燥),这种气象环境条件造成了流感病毒的传播和进一步的流行。而在第一主成分中,地表温度和气压是负的贡献,表明地表温度和气压的增加,会削弱由湿度、降水和蒸发的减少造成的影响,使流感病毒的传播受到一定程度的抑制。此外,计算的第一主成分数值变化曲线和湿度随 17 个样本的变化曲线相当一致(图略),也证明湿度是第一主成分的主角。

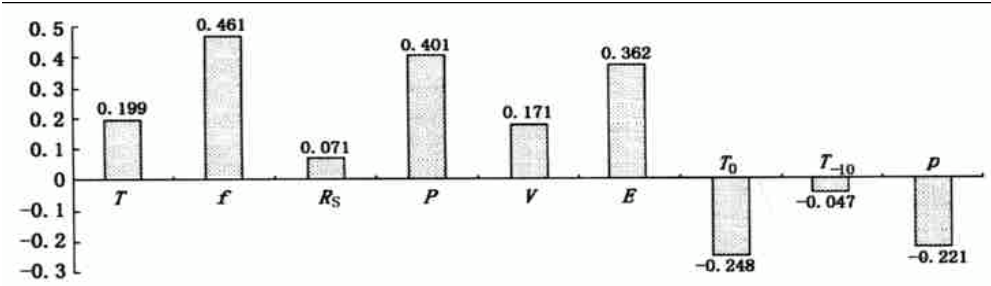


图 1 气象条件主成分分析的第一特征向量

2.2 第二特征向量和第二主成分

第二主成分方差贡献率占总方差的 25.37%,图 2 给出气象条件主成分分析的第二特征向量。从图 2 可以看出:旬平均风速、蒸发量和辐射为较大的

正贡献,它们可能是影响流感爆发的次要气象环境因子,当风速较大,辐射较强时,流感病毒不易传播;反之,则容易导致流感病毒的传播流行。

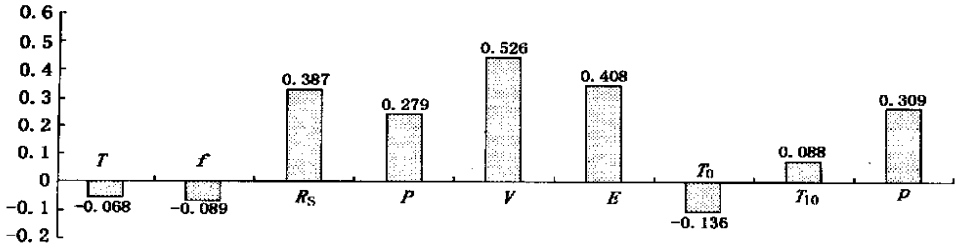


图 2 气象条件主成分分析的第二特征向量

2.3 第三特征向量和第三主成分

第三主成分方差贡献率占总方差的 18.77%,图 3 给出气象条件主成分分析的第三特征向量。图 3 显示第三特征向量的数值普遍比第一和第二主成分相对应的特

征向量的数值要低。旬平均气压和地表温度在第三主成分上表现出了较强的一致性,它们有较大的负贡献,表明当气压降低,地表面温度也随之降低的情况下,流感病毒不易传播。反之则有利于流感病毒的传播流行。

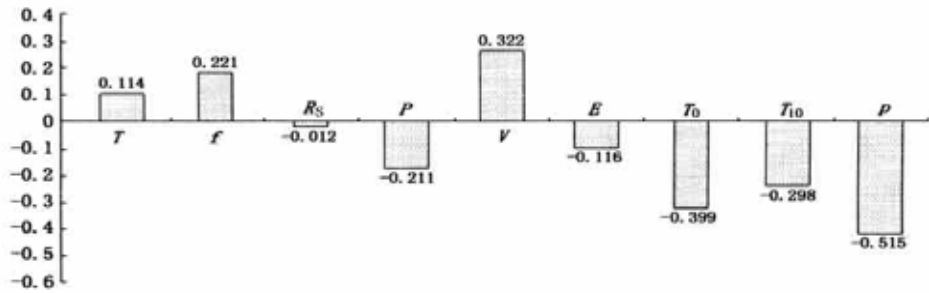


图 3 气象条件主成分分析的第三特征向量

3 结语

通过对北京、天津和石家庄三城市流感流行期间温度、湿度、降水量等 9 个气象因子的主成分分析,得到以下结论和建议:

(1) 第一主成分方差贡献率占总方差的 37.80%,其中旬平均湿度、降水量和蒸发量可能是影响流感病毒传播的最主要气象环境因子。

(2) 前 3 个主成分的方差贡献率占总方差的 81.94%,表明可以用这 3 个新的变量代替原来的 9 个变量分析和研究流感流行与气象环境条件的关系。同时又能把 9 个原始变量按照它们之间一定的相关性(相似性)进行分类,为进一步研究流感的发生、发展趋势预测和防疫工作提供科学依据。

参考文献

- 1 韩永翔,葛秉钧.甘肃省玉米产量的主成分分析.应用气象学报,1995,6(2):252-256
- 2 鲁安新,冯学智,曾群柱.西藏那曲牧区雪灾因子主成分分析.冰川冻土,1997,19(2):180-185
- 3 邵淑彩,崔承章.关于调整气象观测站问题的数学模型.工科数学,1997,13(4):14-16
- 4 张钰仁.气象现代化业务效益评价方法探讨及评估.成都气象学院学报,1997,12(1):8-17
- 5 张绳祖,牛锋.甘肃省平凉地区猪丹毒病与气候因子间的关系研究.家畜生态,1996,17(4):6-10
- 6 吕厚东,李荣华,吕厚远.太阳黑子活动周期与世界性流感流行关系的初探.微生物学通报,1991,18(1):23-26
- 7 裴鑫德.多元统计分析及其应用.北京:北京农业大学出版社,1991.196-203
- 8 美国 SAS 软件研究所上海办事处.SAS 基础教程.上海:上海科学技术文献出版社,1997.146-203

致谢:感谢国家计划生育委员会科研所赵恒研究员的帮助