

DOI:10.15923/j.cnki.cn22-1382/t.2017.2.01

PM_{2.5} 浓度影响因素的主成分回归分析

张 红, 董小刚*, 李 群

(长春工业大学 基础科学学院, 吉林 长春 130012)

摘 要: 选取全国 31 个城市, 对空气中细颗粒物(PM_{2.5})浓度的影响因素进行分析。为处理自变量之间存在的共线性, 选用主成分回归。确定主成分的个数, 将原自变量的主成分代替原自变量进行回归分析。总结出造成空气中细颗粒物(PM_{2.5})浓度上升的因素分为两方面, 直接因素中二氧化氮浓度和间接因素中汽车数量。

关键词: PM_{2.5}; 主成分分析法; 回归分析

中图分类号: O 212 **文献标志码:** A **文章编号:** 1674-1374(2017)02-0105-06

Principal component regression analysis of influencing factors of PM_{2.5} concentration

ZHANG Hong, DONG Xiaogang*, LI Qun

(School of Basic Sciences, Changchun University of Technology, Changchun 130012, China)

Abstract: For 31 cities across the country, the influential factors to concentration of air fine particulate matter (PM_{2.5}) are analyzed. Principal component regression is selected to process the collinearity among the independent variables. We choose the number of principal components and replace the original variables with principal component with regression analysis. It comes to a conclusion that the reasons for increase of PM_{2.5} are the following: increase of NO₂ concentrations and the number of cars.

Key words: PM_{2.5}; principal component; regression analysis.

0 引 言

近年来雾霾天气越来越严重, 尤其是人口聚居的城市地区, 雾霾中混杂着有害物质, 对健康产生直接危害。细颗粒物(PM_{2.5})是构成霾的主要

成分, 与较粗的大气颗粒物相比, 细颗粒物(PM_{2.5})粒径小, 面积大, 活性强, 易附带有毒、有害物质(例如重金属、细菌等), 且在大气中的停留时间长, 输送距离远。人体的鼻腔、咽喉挡不住, 它们可以一路下行, 进入细支气管、肺泡, 再通过

收稿日期: 2017-02-15

基金项目: 国家自然科学基金资助项目(11571051, 11301037)

作者简介: 张 红(1990—), 女, 汉族, 山东青岛人, 长春工业大学硕士研究生, 主要从事生存分析方向研究, E-mail: 253085825@qq.com. * 通讯作者: 董小刚(1961—), 男, 汉族, 吉林长春人, 长春工业大学教授, 博士, 主要从事数理统计方向研究, E-mail: dongxiaogang@ccut.edu.cn.

肺泡壁进入毛细血管,再进入整个血液循环系统。对人体的呼吸系统和心血管系统造成伤害。所以,治理雾霾的关键就是解决细颗粒物(PM2.5)问题^[1]。

城市空气中有毒颗粒物的来源如下:

1)汽车尾气。使用柴油的大型车大量排放PM10,使用汽油的小型车虽然排放的是气态污染物,比如氮氧化物等,但碰上雾天,也很容易转化为二次颗粒污染物。

2)北方冬季烧煤供暖所产生的废气。

3)工业生产排放的废气。比如冶金、窑炉与锅炉、机电制造业,还有大量汽修喷漆、建材生产、窑炉燃烧排放的废气。

4)建筑工地和道路交通产生的扬尘。

5)可生长颗粒,细菌和病毒的粒径相当于PM0.1~PM 2.5,空气中的湿度和温度适宜时,微生物会附着在颗粒物上,特别是油烟的颗粒物上,微生物吸收油滴后转化成更多的微生物,使得空气中的生物有毒物质生长增多。

6)家庭装修中产生的粉尘。现在很多城市的空气污染物排放水平已处于临界点,对气象条件非常敏感,空气质量在扩散条件较好时能达标,一旦遭遇不利天气条件,空气质量和能见度就会立刻下滑。

不同城市空气中细颗粒物(PM2.5)浓度有很

$$y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \alpha_4 x_4 + \alpha_5 x_5 + \alpha_6 x_6 + \alpha_7 x_7 + \alpha_8 x_8 + \alpha_9 x_9 + \alpha_{10} x_{10} + \alpha_{11} x_{11} + \alpha_{12} x_{12} + \alpha_{13} x_{13} + \alpha_{14} x_{14} + \alpha_{15} x_{15} + \alpha_{16} x_{16} + \alpha_{17} x_{17} + \epsilon \quad (1)$$

式中: y ——因变量,细颗粒物(PM2.5)年平均浓度;

x_1 ——平均温度,℃;

x_2 ——平均湿度;

x_3 ——降水量;

x_4 ——日照时长;

x_5 ——工业二氧化硫排放量,t;

x_6 ——工业氮氧化物排放量,t;

x_7 ——工业烟(粉)尘排放量,t;

x_8 ——生活二氧化硫排放量,t;

x_9 ——生活氮氧化物排放量,t;

x_{10} ——生活烟尘排放量,t;

x_{11} ——二氧化硫年平均浓度, $\mu\text{g}/\text{m}^3$;

x_{12} ——二氧化氮年平均浓度, $\mu\text{g}/\text{m}^3$;

大的差异,这与每个城市所处的地理位置、气候以及人们的行为活动有很大的关系。选取全国 31 个主要城市,分别为北京、天津、石家庄、太原、呼和浩特、沈阳、长春、哈尔滨、上海、南京、杭州、合肥、福州、南昌、济南、郑州、武汉、长沙、广州、南宁、海口、重庆、成都、贵阳、昆明、拉萨、西安、兰州、西宁、银川、乌鲁木齐,分析不同因素对空气中细颗粒物(PM2.5)浓度的影响。

1 PM2.5 浓度影响因素的模型建立

造成细颗粒物(PM2.5)浓度上升的原因很多,有直接的因素,也有间接的因素。根据现有理论和检测结果,文中将细颗粒物(PM2.5)浓度的影响因素概括为 17 个方面。间接的因素有:平均温度、平均湿度、降水量、日照时长、工业二氧化硫排放量、工业氮氧化物排放量、工业烟(粉)尘排放量、生活二氧化硫排放量、生活氮氧化物排放量、生活烟尘排放量、汽车总量。直接的因素有:二氧化硫年平均浓度、二氧化氮年平均浓度、可吸入颗粒物(PM10)年平均浓度、一氧化碳日均值第 95 百分位浓度、臭氧(O₃)日最大 8 h 第 90 百分位浓度、空气质量达到及好于二级的天数。

确定了造成细颗粒物(PM2.5)浓度上升的因素后,可以建立多元线性回归模型^[2]:

x_{13} ——可吸入颗粒物(PM10)年平均浓度, $\mu\text{g}/\text{m}^3$;

x_{14} ——一氧化碳日均值第 95 百分位浓度, $\mu\text{g}/\text{m}^3$;

x_{15} ——臭氧(O₃)日最大 8 h 第 90 百分位浓度, $\mu\text{g}/\text{m}^3$;

x_{16} ——空气质量达到及好于二级的天数;

x_{17} ——汽车总量,万辆。

文中数据来源于文献[3]。

2 数据分析

对因变量与所有自变量做相关分析,相关系数矩阵说明自变量与因变量之间有较强的相关关系,分析结果见表 1。

表 1 相关分析结果

Prob> r under $H_0: Rho=0$																	
y	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}
1	-0.183	-0.183	-0.280	-0.008	0.232	0.371	0.217	0.401	0.525	0.583	0.548	0.760	0.753	0.675	0.420	-0.936	0.527

直接做最小二乘回归,得到结果见表 2。

表 2 最小二乘回归

变量	标签	自由度	参数估计	标准误差	t 值	Pr> t	方差膨胀
Intercept	Intercept	1	66.631 59	33.723 52	1.98	0.069 8	0
x_1	x_1	1	0.907 18	0.487 87	1.86	0.085 7	4.973 53
x_2	x_2	1	0.503 81	0.271 27	1.86	0.086 1	7.021 83
x_3	x_3	1	-0.014 78	0.007 46	-1.98	0.069	10.861 74
x_4	x_4	1	-0.006 43	0.005 05	-1.27	0.225 3	4.931 21
x_5	x_5	1	-5.4E-05	5.09E-05	-1.07	0.305 6	14.894 2
x_6	x_6	1	-3.5E-05	4.37E-05	-0.79	0.442	6.712 33
x_7	x_7	1	0.000 184	0.000 109	1.69	0.115 1	10.568 73
x_8	x_8	1	-0.000 36	0.000 228	-1.59	0.136 7	9.516 06
x_9	x_9	1	0.000 551	0.000 592	0.93	0.369 4	9.710 39
x_{10}	x_{10}	1	0.000 239	0.000 163	1.46	0.167 8	4.690 59
x_{11}	x_{11}	1	0.007 21	0.090 86	0.08	0.938	3.936 84
x_{12}	x_{12}	1	0.096 1	0.239 49	0.4	0.694 7	5.835 37
x_{13}	x_{13}	1	0.082 75	0.081 75	1.01	0.329 9	12.480 5
x_{14}	x_{14}	1	3.175 2	1.967 99	1.61	0.130 7	4.773 93
x_{15}	x_{15}	1	0.146 02	0.073 97	1.97	0.07	3.819 9
x_{16}	x_{16}	1	-0.237 3	0.048 06	-4.94	0.000 3	8.682 72
x_{17}	x_{17}	1	-0.008 58	0.008 21	-1.05	0.314 8	3.631 58

对回归模型的检验中 P 值小于 0.01,说明该回归模型适合该数据的分析,结果见表 3。

表 3 回归模型检验

源	自由度	平方和	均方	F 值	Pr>F
模型	17	20 717	1 218.636 39	30.54	<0.000 1
误差	13	518.729 70	39.902 28		
校正合计	30	21 236			

对数据做共线性诊断,结果见表 4。

表 4 共线性诊断

变量	特征值	条件指数
x_1	1.644 58	2.887 34
x_2	0.853 55	4.007 85
x_3	0.681 69	4.484 68
x_4	0.379 37	6.011 64
x_5	0.223 47	7.832 77
x_6	0.162 25	9.192 62
x_7	0.086 23	12.609 33
x_8	0.081 46	12.973 23
x_9	0.053 91	15.947 49
x_{10}	0.035 37	19.688 14
x_{11}	0.027 17	22.463 67
x_{12}	0.023 64	24.083 86
x_{13}	0.014 83	30.407 05
x_{14}	0.009 92	37.178 45
x_{15}	0.007 72	42.154 59
x_{16}	0.003 41	63.443 72
x_{17}	0.000 922	121.925 2

从表 4 中可以看出,很多条件指数都超过 10,说明变量之间共线性很大^[4]。

3 细颗粒物(PM_{2.5})浓度影响因素主成分回归分析

主成分回归的基本思想是将线性相关的一类变量转化为线性无关的一类新的综合变量,这些综合变量反映原来多个变量的信息,主成分回归是选取其中较少的几个新的综合变量建立模型回归方程。主成分回归的方法步骤如下:

- 1)对自变量数据进行标准化变换;
- 2)计算相关系数矩阵;
- 3)计算相关矩阵的特征值及对应的特征向量;
- 4)计算主成分贡献率并选择合适的主成分;
- 5)计算主成分载荷与得分;
- 6)利用所选主成分进行回归分析^[5-6]。

运用统计软件 SAS9.4 对影响 PM_{2.5} 浓度的各因素进行主成分分析,结果见表 5。

表 5 主成分分析结果

主成分	特征值	差分	比例/%	累积
1	5.573 148	2.177 785	32.78	0.327 8
2	3.395 364	0.581 28	19.97	0.527 6
3	2.814 084	1.128 413	16.55	0.693 1
4	1.685 671	0.796 989	9.92	0.792 3
5	0.888 682	0.293 376	5.23	0.844 5
6	0.595 306	0.075 154	3.50	0.879 5
7	0.520 153	0.191 105	3.06	0.910 1
8	0.329 048	0.026 862	1.94	0.929 5
9	0.302 186	0.079 487	1.78	0.947 3
10	0.222 699	0.072 197	1.31	0.960 4
11	0.150 502	0.001 812	0.89	0.969 2
12	0.148 690	0.009 986	0.87	0.978 0
13	0.138 704	0.025 433	0.82	0.986 1
14	0.113 271	0.053 529	0.67	0.992 8
15	0.059 742	0.026 789	0.35	0.996 3
16	0.032 953	0.003 156	0.19	0.998 2
17	0.029 797		0.18	1.000 0

从表 5 中可知,第 1 主成分的特征根为 5.573,解释了总变异的 32.78%;第 2 主成分的特征根为 3.395,解释了总变异的 19.97%;第 3 主成分的特征根为 2.814,解释了总变异的 16.55%;第 4 主成分的特征根为 1.685,解释了总变异的 9.92%。前 4 个特征根的累计贡献率达到 79.23%,说明前 4 个主成分已经反映原来 17 个指标 79.23%的信息,因此确定选择前 4 个主成分建立模型。提取的前 4 个主成分 F_1, F_2, F_3, F_4 见表 6。

以 4 个主成分 F_1, F_2, F_3, F_4 为自变量进行多元线性回归分析,建立回归模型:

$$y = 8.63F_1 + 3.79F_2 - 6.14F_3 + 3.87F_4 \quad (2)$$

表6 主成分表

变量	F_1	F_2	F_3	F_4
x_1	-0.236 410	0.343 047	-0.168 250	-0.046 590
x_2	-0.233 270	0.354 897	-0.016 860	0.148 691
x_3	-0.306 230	0.300 967	-0.113 080	0.121 125
x_4	0.134 224	-0.436 510	0.002 320	-0.065 030
x_5	0.176 964	0.315 544	0.215 866	-0.428 490
x_6	0.254 371	0.241 128	0.177 108	-0.302 370
x_7	0.216 189	0.249 281	0.259 131	-0.325 670
x_8	0.172 284	0.200 938	0.405 351	0.195 214
x_9	0.160 549	0.106 591	0.369 978	0.413 267
x_{10}	0.169 652	-0.020 690	0.320 887	0.496 327
x_{11}	0.328 085	-0.030 270	-0.140 130	-0.016 370
x_{12}	0.270 315	0.184 061	-0.225 750	0.269 682
x_{13}	0.350 431	-0.001 050	-0.233 650	-0.014 660
x_{14}	0.328 427	-0.144 770	-0.185 040	-0.007 870
x_{15}	0.143 030	0.267 756	-0.214 860	0.014 577
x_{16}	-0.330 950	-0.124 360	0.256 680	-0.081 380
x_{17}	0.097 680	0.245 597	-0.371 290	0.200 952

注:表中各变量同模型(1)。

方差分析结果见表7。

表7 方差分析结果

源	自由度	平方和	均方	F值	$Pr>F$
模型	4	17 864	4 466.081 86	34.44	<0.000 1
误差	26	3 371.220 97	129.662 34		
校正合计	30	21 236			

R^2 为 0.841 2,调整 R^2 为 0.816 8,说明模型拟合较好。最后,将 4 个主成分表达式代入,得到

$$y = 0.112 1x_1 + 0.010 5x_2 - 0.339 1x_3 - 0.761 1x_4 - 0.263 7x_5 + 0.843 9x_6 - 0.043 6x_7 + 0.514x_8 + 1.117 5x_9 + 1.337 4x_{10} + 3.514 8x_{11} + 5.462 4x_{12} + 4.399 6x_{13} + 3.392 9x_{14} + 3.625 4x_{15} - 5.220 1x_{16} + 4.833 3x_{17} \quad (3)$$

4 结论及政策建议

从主成分回归方程可以看出,直接影响因素对细颗粒物(PM2.5)浓度影响比较大。空气中污染物浓度的上升直接导致细颗粒物(PM2.5)浓度上升,其中二氧化氮的浓度(x_{12})对细颗粒物(PM2.5)浓度影响最大,可吸入颗粒物(PM10)的

最终的主成分回归方程:

浓度(x_{13})对细颗粒物(PM 2.5)浓度影响次之,而二氧化硫年平均浓度(x_{11})、一氧化碳日均值第 95 百分位浓度(x_{14})、臭氧(O_3)日最大 8 h 第 90 百分位浓度(x_{15})对细颗粒物(PM 2.5)浓度影响较小。空气质量达到及好于二级的天数与对细颗粒物(PM2.5)浓度是负相关的。

间接影响因素对细颗粒物(PM2.5)浓度影响

比较小。但是汽车总量(x_{17})作为间接影响因素对细颗粒物(PM2.5)浓度影响非常大,废气的排放中,生活氮氧化物排放量(x_9)和生活烟尘排放量(x_{10})对细颗粒物(PM2.5)浓度影响比较大,工业二氧化硫排放量(x_5)、工业氮氧化物排放量(x_6)和生活二氧化硫排放量(x_8)对细颗粒物(PM2.5)浓度影响较小,工业烟(粉)尘排放量(x_7)对细颗粒物(PM2.5)浓度影响很小,甚至有了负相关。天气因素中温度(x_1)、湿度(x_2)、降水量(x_3)、日照时长(x_4)对细颗粒物(PM2.5)浓度影响都很小。温度(x_1)越高,湿度(x_2)越大,细颗粒物(PM2.5)浓度越大,降水量(x_3)越小,日照时长(x_4)越短,细颗粒物(PM2.5)浓度越小。

根据实证分析结果,温度、湿度对细颗粒物(PM2.5)有一定的影响,温度越高,湿度越大,越有利于空气中颗粒物的形成。但这个因素是人为无法控制的。降低细颗粒物(PM2.5)浓度应该降低空气中二氧化氮、可吸入颗粒物(PM10)、二氧化硫、一氧化碳、臭氧的浓度。措施如下:

1)控制汽车尾气对空气的污染,应该减少汽车的数量或者改善汽车的尾气排放系统,将尾气进行处理再排放,这是减少废气排放的主要措施。

2)生活氮氧化物来源主要是居民生活燃煤燃油和生物质的燃烧,应该提高天然气的普及程度,减少小型煤炭炊事或者蜂窝煤的使用,避免秸秆的焚烧处理。

3)交通工具的排气,垃圾焚烧、取暖锅炉和家庭炉灶排出的烟气都是生活烟尘污染的主要来源。应该加强城市绿化,集中供暖,改造锅炉,改进燃料的燃烧方法,安装净化除尘设备^[7-9]。

4)工业二氧化硫、工业氮氧化物主要来源于

钢铁、有色金属冶炼、火力发电、水泥和石油化工企业的生产过程。应该加强管理,使工厂排放的废气在标准控制范围内。把排放烟尘的企业安排在居住区常年风向频率最小的上风侧,以减少烟尘对居住区的污染。

通过以上分析,可以知道影响空气中细颗粒物(PM 2.5)浓度的主要原因,并根据各种原因给出相应的解决方案,希望可以为各城市空气质量的改善提供一些参考。

参考文献:

- [1] 顾昊,张亚男,刘欣,等.哈尔滨市春季PM(2.5)污染状况及来源调查研究[J].环境科学与技术,2013(S2):94-98.
- [2] 李毛侠.安徽省消费需求影响因素的主成分回归分析[J].现代经济:现代物业中旬刊,2010,9(2):48-51.
- [3] 国家统计局城市社会经济调查司.中国城市统计年鉴[M].北京:中国统计出版社,2014.
- [4] Mcdonald G C, Schwing R C. Instabilities of regression estimates relating air pollution to mortality [J]. Technometrics, 1973, 15(3): 463-481.
- [5] 何晓群.应用回归分析[M].北京:中国人民大学出版社,2001.
- [6] 哈德勒.应用多元统计分析[M].北京:北京大学出版社,2011.
- [7] 杨天智.长沙市大气颗粒物PM2.5化学组分特征及来源解析[D].长沙:中南大学,2010.
- [8] 李东海,何彩霞.浅谈雾霾天气的识别及预警策略[J].安徽农学通报,2011,17(18):165-166.
- [9] 董小刚,王淑影,王纯杰.基于动态因子的经济水平差异分析[J].长春工业大学学报,2015,36(2):125-129.