



Study on the Relationship Between PM2.5 and Daily Consultations Volume in Respiratory Department Based on Transfer Function Model

Xiao Weiwei, Fan Meixia

School of Science, North China University of Technology, Beijing, China

Email address:

wwsunny@163.com (Xiao Weiwei), 1456522632@qq.com (Fan Meixia)

To cite this article:

Xiao Weiwei, Fan Meixia. Study on the Relationship Between PM2.5 and Daily Consultations Volume in Respiratory Department Based on Transfer Function Model. *Science Discovery*. Vol. 8, No. 6, 2020, pp. 128-133. doi: 10.11648/j.sd.20200806.12

Received: September 28, 2020; **Accepted:** October 28, 2020; **Published:** November 4, 2020

Abstract: With the development of the era, the word “haze” was born, and more and more people suffer from respiratory diseases, particularly congestive heart failure and coronary artery disease and lung cancer. In the European Union, PM2.5 cuts life expectancy by 8.6 months. Bases on the medical data relating to the hospital, in July 2018-August of respiratory medical consultations and PM2.5 value transfer function model is set up, after using the model on September 1st, 2018-September 20th respiratory medical consultations to make predictions, and comparing with the real value, the results show that compared with the ARIMA model, transfer function model predicts more accurately. In other words, the relationship between daily respiratory department visits and PM2.5 value is more similar to the transfer function model. Finally, the transfer function model is used to predict the daily number of patients in respiratory department from September 21st to September 30th, 2018. Because the management of medical treatment is one of the important indicators reflecting the level of hospital management, the prediction of daily medical treatment volume can provide a reliable basis for the allocation of out-patient medical staff, and it is of great significance for hospitals to rationally arrange human, financial, material and other resources to improve economic and social benefits.

Keywords: Daily Consultation Volume, PM2.5, Transfer Function Model, ARIMA Model

基于传递函数模型的PM2.5与呼吸科日就诊量的关系分析研究

肖维维, 范美霞

北方工业大学理学院, 北京市, 中国

邮箱

wwsunny@163.com (肖维维), 1456522632@qq.com (范美霞)

摘要: 随着时代的发展, 雾霾这两个字孕育而生, 而随之也有越来越多的人患有关呼吸方面的疾病, 尤其是充血性心力衰竭和冠状动脉等心脏疾病以及肺癌。在欧盟国家中, PM2.5导致人们的平均寿命减少8.6个月。本文基于有关医院的医疗数据, 对2018年7月-8月的呼吸科日就诊量与PM2.5值建立传递函数模型, 之后利用模型对2018年9月1日-9月20日呼吸科日就诊量进行预测, 并与真实值进行比较, 结果显示, 与ARIMA模型相比, 传递函数模型预测得更为准确, 换句话说呼吸科日就诊量与PM2.5值的关系更趋近于传递函数模型。最后, 利用传递函数模型预测2018年9月21日-9月30日呼吸科日就诊量, 因为就诊的管理是反映医院管理水平高低的重要指标之一, 日就诊量的预测能为门诊医护人员配置提供可靠依据, 对医院合理安排人、财、物等资源以提高经济效益和社会效益有重要意义。为医院相关管理人员提供参考和决策依据。

关键词: 日就诊量, PM2.5, 传递函数模型, ARIMA模型

1. 引言

雾霾中的PM2.5虽然含量少, 但是容易吸附肺内, 引发呼吸方面的疾病, 所以PM2.5成为众矢之的。随着空气污染的严重, 在北京雾霾天出现的次数增多, 从而患有呼吸方面疾病的人也变多[1-5]。

本文的主要内容, 就是根据有关医院所获得的呼吸科日就诊量和PM2.5值, 应用计算机分析计算, 来得出医院呼吸科日就诊量的走势预测。其中采用了一系列的统计方法和知识, 如ARIMA模型, 传递函数-噪声模型等, 最后在SAS中进行计算, 达到了预测趋势的目的[6]。

SAS一般用于统计方面, 可以处理大量的数据[7]。本文利用SAS来得到输入序列与输出序列的序列图、预白后的相关系数、符合模型的估计参数及有效性检验等。

2. 传递函数模型

虽然ARIMA模型在统计、建模等方面发挥着重要作用, 但在实际应用中, 使用ARIMA模型会存在一些弊端。因为想要预测一个时间序列的发展趋势, 不仅要了解其过去的行为, 还需要知道另一个与之相关的时间序列, 而且这两个序列之间的影响可能是同步的, 也可能是滞后的。所以本文采用George E.P.Box和Gwilym M.Jenkins等提出的传递函数-噪声模型[8]。

记时刻 t 的输入序列为 X_t 与输出序列为 Y_t [9], 传递函数-噪声模型为

$$Y_t = v(B)X_t + N_t \tag{1}$$

式(1)中 $v(B) = \sum_{i=0}^{\infty} v_i B^i$, $v(B)$ 叫做传递函数, v_i 为脉冲响应函数, $N_t = \frac{\Theta(B)}{\phi(B)}\alpha(t)$, $\Theta(B)$ 和 $\phi(B)$ 是滞后算

子 B 的多项式函数[8], $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$, $\Theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$, p 叫做回归阶数, q 为移动平均阶数, $\alpha(t)$ 是白噪声序列, 这个模型可以在公式上简化为

$$Y_t = \delta^{-1}(B)w(B)X_{t-b} + N_t \tag{2}$$

其中 $\delta(B)$ 和 $w(B)$ 分别是阶数为 r 和 s 的滞后算子多项式 $\delta(B) = 1 - \delta_1 B - \delta_2 B^2 - \dots - \delta_r B^r$, $w(B) = w_0 - w_1 B - w_2 B^2 - \dots - w_s B^s$, b 表示 X_t 延后 b 期影响 Y_t [8]。

3. PM2.5与日就诊量传递函数模型的建立

3.1. 数据来源

本文所用到的数据, 是广安门医院的2018年7月到8月的呼吸科日就诊量和广安门医院附近的PM2.5观测点-万寿西宫所获得的每小时PM2.5值, 因为PM2.5的浓度时时在变化, 我们使用每日的平均PM2.5浓度, 记2018年7月1日为第一日, 2日为第二日, 8月1日为第32日, 以此类推, 并记 Y_t 为 t 日的广安门医院呼吸科日就诊量, X_t 为第 t 日万寿西宫观测点观测到的PM2.5值后取平均值。

3.2. 数据预处理

图一和图二中的PM2.5 X_t 与日就诊量 Y_t 均显现出先上升后下降的趋势, 两者之间呈现出一定的正相关性, 且序列不平稳, 所以需要 X_t 和 Y_t 进行差分。

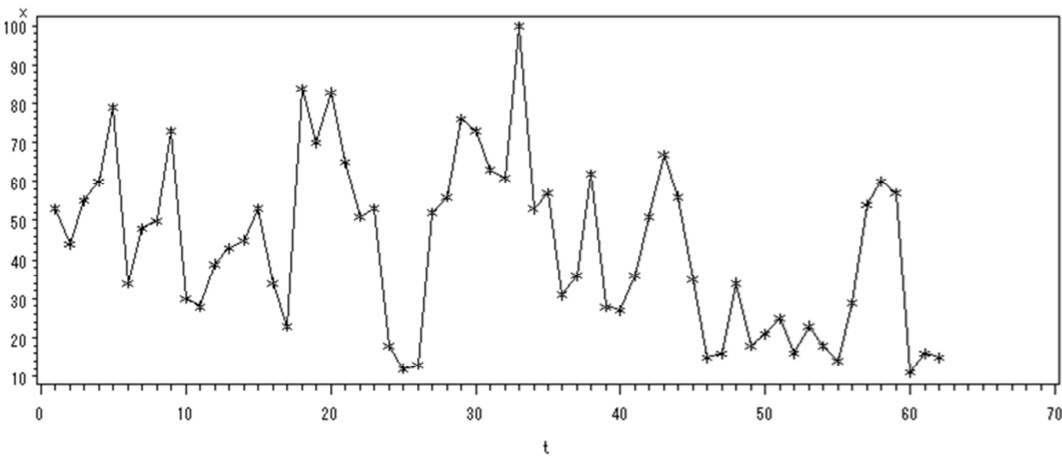


图1 PM2.5 X_t 序列图。

表3 传递函数-噪声模型中的参数估计。

Conditional Least Squares Estimation							
Parameter	Estimate	Standard Error	t Value	Pr>Approx t	Lag	Variable	Shift
MA1,1	0.584	0.12366	4.72	<.0001	2	y	0
AR1,1	0.41349	0.08773	4.71	<.0001	1	y	0
AR1,2	-0.6804	0.08898	-7.65	<.0001	3	y	0
NUM1	-1.00444	0.12845	-7.82	<.0001	0	x	6
DEN1,1	-0.80081	0.0096409	-83.06	<.0001	2	x	6
DEN1,2	0.46347	0.006607	70.15	<.0001	3	x	6

根据表2可以看出 $\nabla^2 X_t$ 与 $\nabla^2 Y_t$ 的交互相关函数在延迟6期处显著不为0, 说明 $b=6$, 并且我们取 $s=0, r=(2,3)$, 后因传递函数模型是正确的, 噪声模型是错误的[6], 且参数估计值不显著, 调整参数为 $p=(1,3), q=(2)$, 并提出如下模型:

$$\nabla^2 Y_t = \frac{1-0.41349B+0.6804B^3}{1-0.584B^2} \nabla^2 X_t + N_t$$

(3)

其中 $N_t = \frac{1}{1+0.80081B^2-0.46347B^3} \alpha_t$, 序列 α_t 仍为白噪声序列。

3.3.3. 检验 N_t 的白噪声性

表4 传递函数-噪声模型中残差的自相关检验。

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr>ChiSq	-----Autocorrelations-----					
6	5.38	3	0.1462	-0.072	0.021	0.174	-0.014	0.2	-0.13
12	9.47	9	0.3952	0.093	0.199	-0.077	-0.015	-0.1	-0.016
18	16.97	15	0.3205	0.183	-0.176	0.003	0.13	-0.132	-0.048
24	19.96	21	0.5237	-0.089	0.032	0.006	-0.161	-0.057	-0.018

表5 输入序列 $\nabla^2 X_t$ 与残差 α_t 的相关性检验。

Crosscorrelation Check of Residuals with Input x									
To Lag	Chi-Square	DF	Pr>ChiSq	-----Crosscorrelations-----					
5	5.06	3	0.1671	0.087	0.165	-0.107	-0.242	0.014	-0.086
11	9.54	9	0.3888	0.03	-0.071	0.131	0.038	-0.095	0.257
17	21.72	15	0.1152	-0.04	0.257	-0.191	0.057	0.168	-0.367
23	24.48	21	0.2704	0.22	-0.011	-0.04	-0.004	-0.103	0.018

由表4知, $0.1462>0.05$, 表明建立的传递函数-噪声模型残差显著平稳, 可认为是随机白噪声序列, 由表5知 $0.1671>0.05$, 所以可得输入序列 $\nabla^2 X_t$ 与残差相互独立, 模型是有效的[9-15]。因此, 传递函数-噪声模型为

$$\nabla^2 Y_t = \frac{1-0.41349B+0.6804B^3}{1-0.584B^2} \nabla^2 X_t + \frac{1}{1+0.80081B^2-0.46347B^3} \alpha_t$$

(4)

4. ARIMA模型与传递函数-噪声模型预测效果的比较

为了检测传递函数-噪声模型预测的准确性, 本文对日就诊量 Y_t 进行差分, 并建立了ARIMA模型, 之后将ARIMA模型、传递函数-噪声模型预测的结果与真实值进行比较。

表6 ARIMA模型的参数估计。

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Pr>Approx t	Lag
AR1,1	-0.81201	0.06414	-12.66	<.0001	2
AR1,2	-0.33862	0.06665	-5.08	<.0001	3
AR1,3	-0.37316	0.06822	-5.47	<.0001	4
AR1,4	-0.75227	0.06666	-11.29	<.0001	5

表7 ARIMA模型残差的自相关检验。

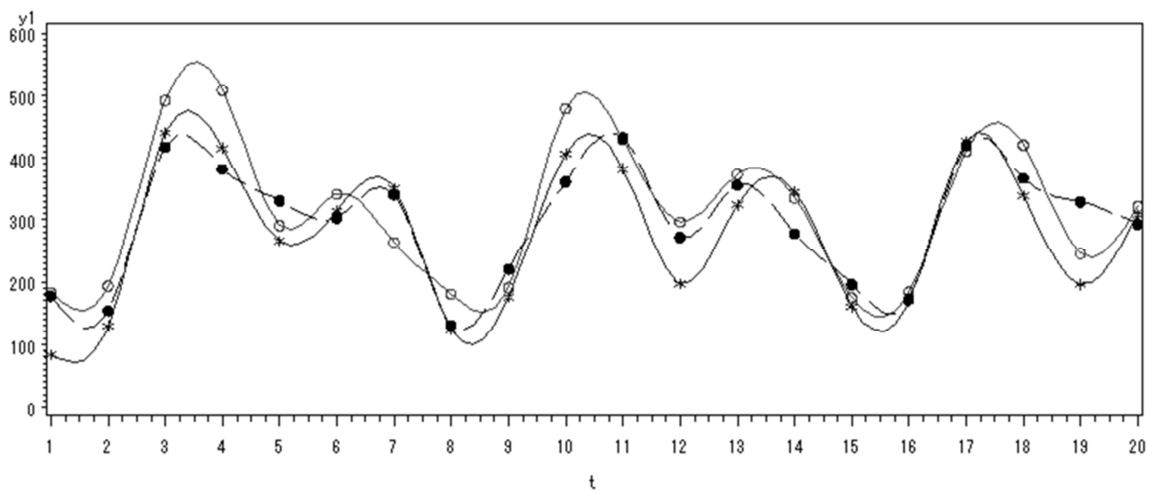
Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr>ChiSq	-----Autocorrelations-----					
6	9.4	2	0.0091	-0.181	-0.01	-0.018	-0.173	0.162	-0.227
12	11.37	8	0.1818	0.075	0.007	0.093	-0.042	-0.091	-0.047
18	16.05	14	0.31	-0.021	0.11	-0.092	0.105	0.019	0.151
24	21.22	20	0.3842	-0.021	-0.151	0.031	-0.131	0.095	-0.056

根据表6和表7最终确定ARIMA模型为

$$\nabla^2 Y_t = (1 + 0.81201B^2 + 0.33862B^3 + 0.37316B^4 + 0.75227B^5) \nabla^2 X_t + \varepsilon_t \quad (5)$$

其中 ε_t 为白噪声序列。

我们将传递函数-噪声模型与ARIMA模型对2018年9月1日-9月20日的预测结果与医院呼吸科日就诊量真实值用图来直观地进行比较，突出了传递函数-噪声模型较好的预测结果。所以在实际生活当中，传递函数-噪声模型较为准确地预测未来的发展及趋势。



(注：带“*”线代表ARIMA模型的预测值，带“o.”线代表真实值，带“x.”线代表传递函数-噪声模型的预测值)

图3 两种模型的预测结果和真实值的比较。

5. 传递函数-噪声模型对日就诊量的预测

PM2.5与日就诊量之间存在着一定的关系，用PM2.5作为输入序列的传递函数-噪声模型在预测呼吸科日就诊量未来的数据相比更为准确，表8是对2018年9月21-30日

表8 2018年9月呼吸科日就诊量的预测值。

21日	22日	23日	24日	25日	26日	27日	28日	29日	30日
380	150	200	409	342	160	334	371	174	187

6. 结论

本文采用医院的有关医疗数据，对2018年7月-8月的呼吸科日就诊量与PM2.5值建立传递函数模型，之后利用模型对2018年9月1日-9月20日呼吸科日就诊量进行预测，并与真实值进行比较，结果显示，与ARIMA模型相比，传递函数模型预测得更为准确，也就是说呼吸科日就诊量与PM2.5的关系更趋近于传递函数模型。利用传递函数模型预测出2018年9月21日- 9月30日呼吸科日就诊量，可以

的呼吸科日就诊量的预测，结果显示，在9月22日日就诊量人数最低，而后两天有明显的上升趋势，25-26日有所下降，27日又回升，28日上升达到新高，29日有大幅度的下降，而30日有所回升，幅度较小。

看出23-25日和27-28日这段时间呼吸科日就诊量高，所以医院相关管理人员可以在这几天安排更多的医生护士出诊，而对于其他时间呼吸科日就诊量低，医院相关管理人员可以给一些医生安排休息放假。

致谢

本文为北方工业大学人才专项资助(207051360020XN140/004)的阶段成果之一。

参考文献

- [1] 林在生,王恺,林少凯,苏玲.PM_{2.5}对小学生肺功能影响的调查研究[J].中国预防医学杂志,2019(12):1167-1170.
- [2] 吴筱音,李国星,王旭英,梁凤超,潘小川.北京市大气污染与呼吸系统疾病死亡的相关性——基于卫星遥感数据的时空分析[J].北京大学学报(医学版),2017(3):409-417.
- [3] 王艳杰.气象因素与乌鲁木齐市医院儿童呼吸系统疾病日门诊人次的关联性研究[D].新疆医科大学,2018.
- [4] 尹延震,王苗,王静远,王景,樊庆生.南阳市PM₁₀、PM_{2.5}污染特征及其与气象因子的关系[J].干旱环境监测,2018,32(1):12-18.
- [5] 李宁,彭晓武,张本延,殷文军,余素君,大气污染与呼吸系统疾病日门诊量的时间序列分析[J].环境与健康杂志,2009,26(12):1077-1080.
- [6] 王淑英.基于传递函数模型的商品房投资与销售额关系的计算机算法研究[J].学术研究,2010(7):143-145.
- [7] 汪海波,罗莉,吴为,孟玲,杨世宏,汪海玲.SAS统计分析与应用从入门到精通[M].北京:人民邮电出版社,2013.
- [8] 余云彩,刘玲,胡宏昌,肖劲光.基于传递函数模型的白银价格分析与预测[J].当代经济,2017(1):115-117.
- [9] 肖枝洪,郭明月.时间序列分析与SAS应用[M].武汉:武汉大学出版社,2012.
- [10] 李运明,吴凡,郑驰.某三甲综合医院门诊量ARIMA预测分析[J].中国案例,2014(8):53-55.
- [11] 徐艳龙,黄发源,王志强.2014-2015年合肥市大气PM_{2.5}污染对儿童门诊量影响的时间序列分析[J].环境与健康杂志,2017(4):1001-5914.
- [12] 殷永文,程金平,段玉森.上海市霾期间PM_{2.5}、PM₁₀污染与呼吸科、儿呼吸科门诊人数的相关分析[J].环境科学,2011(7):1894-1898.
- [13] 崔亮亮,李新伟耿兴义,等.2013年济南市大气PM_{2.5}污染及雾霾事件对儿童门诊量影响的时间序列分析[J].环境与健康杂志,2015(6):489-493.
- [14] 曾允萱,蔡旭娜.基于时间序列分析方法医院月门诊量预测模型[J].中国医院统计,2009(4):312-315.
- [15] 涂宏.南昌市城区PM_{2.5}的污染特征及其与儿童健康关系的研究[D].南昌:南昌大学,2016.