



$X_1$ —用水人口(万人);

$X_2$ —国民生产总值(亿元);

$X_3$ —工业用水重复率(%);

$X_4$ —年降水量(亿 $m^3$ );

$X_5$ —绿化覆盖率(%);

$R$ —样本相关矩阵;

$\lambda_i$  ( $i=1, 2, \dots, 10$ )——相关矩阵 $R$ 的特征值;

$e_i$  ( $i=1, 2, \dots, 10$ )——相关矩阵 $R$ 的特征向量。

由表1可知,第一个样本主成分的贡献率达77%以上,两个样本主成分的累积贡献率达98%以上,故两个样本可包含所有变量的信息。将样本的两个主成分系数(即特征向量 $e$ )列于表2。

表1 主成分分析结果表

Component (成分序号)	Eigenvalue (特征值 $\lambda_i$ )	Cumulative 累计贡献率(%)
1	3.8885908572	77.77182
2	1.0182462123	98.13674
3	0.0622206932	99.381155
4	0.0301177477	99.983510
5	0.0008244896	100.00000000

表2 样本的主成分系数表

特征向量	$e_1$	$e_2$
$X_1$	-0.49905013	-0.06980550
$X_2$	-0.50185606	0.06372725
$X_3$	-0.49518317	-0.11082845
$X_4$	-0.03359997	0.98813040
$X_5$	-0.50274659	0.04879947

分析表2:第一个主成分信息来源于 $X_1$ 、 $X_2$ 、 $X_3$ 、 $X_5$ 四个变量,第二个主成分信息来源于 $X_4$ ,这两个主成分包含了所有变量的信息。

用R软件运行出载荷结果:

$$Y_1 = (-0.49905013)X_1 + (-0.50185606)X_2 + (-0.49518317)X_3 + (-0.50274659)X_5 \quad (2)$$

$$Y_2 = (-0.11082845)X_3 + (0.98813040)X_4 \quad (3)$$

根据主成分分析结果可以看出,所选用的五个自变量均为有效变量,可以用于建立多元线性回归模型。

### 1.3 模型的建立

回归分析根据模型中影响因素的多少及影响因素与预测对象之间的相互关系,可分为一元线性回归分析、多元线性回归分析和非线性回归分析<sup>[3]</sup>,本文采用多元——重线性回归分析。回归分析方法即根据相互影响、相互关联的两个或多个因素(又称为变量)的实测或调查资料,由不确定的函数关系建立数学模型,确定参数,从而建立函数关系的过程。而将回归分析用于预测未来和研究未来,寻找待测对象与影响因素之间的数学关系,并采用数学模型予以表达,然后通过对未来影响因素的确定,间接导出待测数据的过程,称为回归分析预测方法<sup>[4]</sup>。多元线性回归模型中系数的求解或估计是多元线性回归分析中的基本问题,采用的是统计学中最成熟也最广泛使用的最小二乘估计方法<sup>[5]</sup>。本文应用R语言编制相关程序,实现多元线性回归分析。

回归模型如下:

$$Y = \beta_0 + \beta_1 X_1 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \varepsilon$$

$$\text{假设 } \varepsilon \sim N(0, \sigma^2) \quad (4)$$

其中, $\beta_0, \beta_1, \beta_3, \beta_4, \beta_5$ 是6个有待估计的参数,称为总体回归参数; $\varepsilon$ 是随机变量,其均值为0,方差为 $\sigma^2$ ,即遵从同一正态分布 $N(0, \sigma^2)$ 。

用R语言编程对数据进行多元线性回归模型的求解及残差分析,运行结果为:

$$Y = -17.130219 + 0.034652X_1 + 0.002792X_2 - 0.090996X_3 - 0.005750X_4 + 0.148059X_5 \quad (5)$$

通过对天津市用水总量与用水人口、国民生产总值、工业用水重复率、年降水量、绿化覆盖率这五个变量的关系做多元统计分析,得出了用水总量的回归模型。回归模型的结果见式(5)所示。从回归方程中可以看出,用水总量与 $X_3$ 、 $X_4$ 成负相关关系,其中年降水量难以人为控制,所以要控制用水量必须提高工业用水重复率,节约工业用水,提高水的利用效率<sup>[6]</sup>;用水总量与 $X_1$ 、 $X_2$ 、 $X_5$ 成正相关,从用水人口这方面看,控制人口增长也是减少用水量的重要手段,而在五个变量中,绿化覆盖率对用水总量的影响最大,这也说明公共事业用水在用水总量中占了相当大的比例,如何既提高绿化覆盖率又做到节约用水是一个很重要的问题。在水资源短缺条件下要使经济快速发展,社会不断进步,污水资源化,中水回用,提高水资源利用率就成为了解决问题的关键。同时,运行结果中, $R^2=0.991$ ,指示数据间有较强的回归关系,该回归方程是适合的。

### 1.4 模型检验

通过残差图判断该模型的有效性,如图1所示,模型回归效果基本良好。

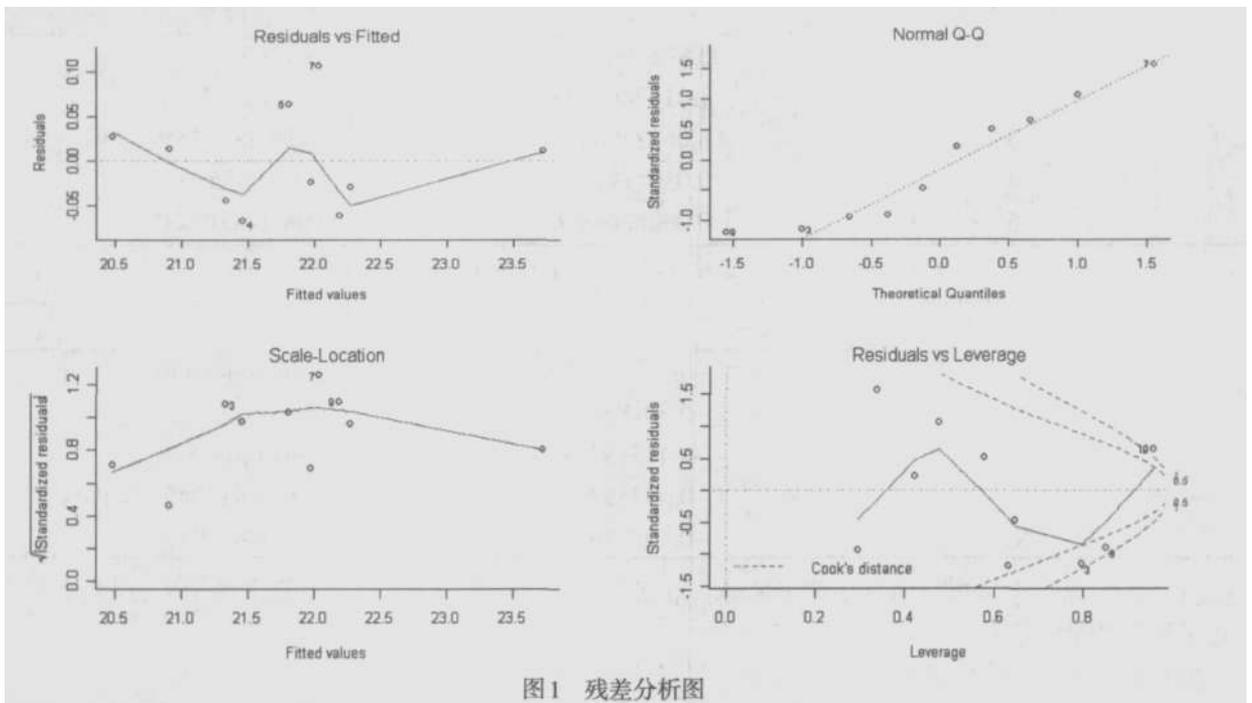


图1 残差分析图

用2001~2005年的数据对模型进行检验,其相对误差见表3:

表3 预测值的相对误差

年份	原始值	预测值	相对误差%
2001	23.84	23.77582	-0.27
2002	24.06	23.78394	-1.15
2003	24.33	23.67038	-2.71
2004	24.17	23.40055	-3.18
2005	24.22	22.75236	-6.06

结果分析:如表3和图2所示,预测误差均小于10%,说明模型有效。从图中也可看出虽然人口和国民生产总值不断提高,公共事业用水不断增加,但是由于近些年来采取了一定的节水措施,提高了水资源利用率,用水量基本可以保持稳定<sup>[7]</sup>。

## 2 结论

本文通过对天津市用水总量与用水人口、国民生产总值、工业用水重复率、年降水量、绿化覆盖率这五个变量的关系做多元统计分析,建立了用水总量的回归模型,运用R软件对模型进行求解和残差分析,并用5年数据进行模型检验,说明所建立的模型有较好的效果,有一定的理论意义和实用价值。  
21994-2016 China Academic Electronic Journal Publishing House. All rights reserved. http://www.cnki.net

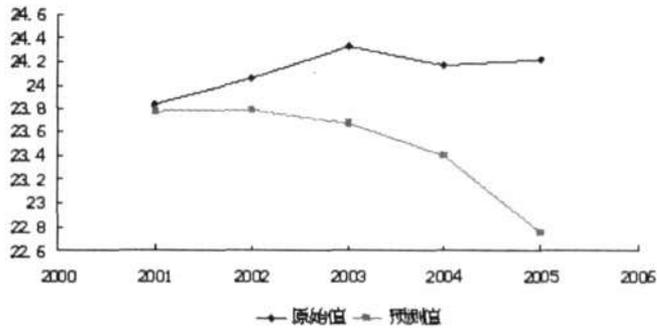


图2 原始值与预测值对比图

## 注释及参考文献:

- [1] 李林. 多元线性回归模型预测城市用水量[J]. 塔里木大学学报, 2007, 19(2): 40-42.
- [2] Richard A. JOHNSON, Dean W. Wichern. 实用多元统计分析[M]. 北京: 清华大学出版社, 2001.
- [3] 丁士水. 城市供水水量预测模型研究及案例分析[J]. 山西建筑, 2007, 33(14): 164-165.
- [4] 吕谋, 赵洪宾, 李红卫. 城市日用水量预测的组合动态建模方法[J]. 给水排水, 1997(11): 13-14.
- [5] 李红艳, 崔建国, 张星全. 城市用水量预测模型的优选研究[J]. 中国给排水, 2004(2): 27-28.
- [6] 张雄, 党志良. 西安市日用水量预测模型研究[J]. 机械工程与自动化, 2005(5): 23-24.
- [7] 张杰明. 试论用多元线性回归法预测城市用水量[J]. 城市给排水, 2004, 18(4): 23-25.

## Research on Forecasting the Water Consumption of Tianjin by Multiple Linear Regression Model

SHA Zhi-jie<sup>1</sup>, ZHOU Jin-feng<sup>2</sup>

(1. Environmental Protection Bureau of Jingyang Town, Deyang, Sichuan 618000; 2. Tianjin Planning, Design & Research Institute of Automotive Industry, Ministry of Machinery Industry, Tianjin 300162)

**Abstract:** The water consumption of Tianjin was forecasted by multiple linear regression model. Firstly, using the main composition analysis method, the independent variables were chosen as population, GDP, rate of industrial water reuse, the average rainfall and the green cover percentage. Then established the dependent program by the R language and made a Mathematical Model to do the multiple linear regression. It can be seen that the purpose of the regression model is good through researching the validity of the residual analysis model. This article provides the reference for accurately forecasting the urban water consumption by a simple method.

**Key words:** City water use; Multiple linear regression; R language