

Study on Linear Relation of Discharge of Major Pollutants in Wastewater

—Research on Multivariate Linear Model and Classification

Chang Liu

Yunnan University of Finance and Economics, Kunming Yunnan
Email: 245103325@qq.com

Received: Oct. 1st, 2019; accepted: Oct. 14th, 2019; published: Oct. 21st, 2019

Abstract

After consulting the data of the total amount of sewage discharged and the main pollutants discharged in different regions of the country, it is found that the pollutant composition of wastewater can be divided into two parts, one is mixed pollutants such as petroleum, the other is heavy metal elements such as "Arsenic". Through the calculation of canonical correlation analysis, the two kinds of pollutants are found to have a concomitant and interdependent relationship. For instance, chemical oxygen demand pollutants contain many other variables, so a multivariate linear model is constructed to analyze the multivariate linear relationship between these main pollutants, and calculate the size of the correlation relationship. Through model testing and diagnosis, a correct regression model is constructed, and ultimately a multivariate linear model can be obtained. Chemical oxygen demand and ammonia nitrogen are correlated. According to the magnitude of positive or negative correlation, the main pollutants are explained and some suggestions are put forward.

Keywords

Wastewater Discharge, Major Pollutants, Canonical Correlation Analysis, Multivariate Linear Model

废水主要污染物排放的线性关系研究

——基于多元线性模型和分类研究

刘 畅

云南财经大学, 云南 昆明
Email: 245103325@qq.com

收稿日期: 2019年10月1日; 录用日期: 2019年10月14日; 发布日期: 2019年10月21日

摘要

查阅了全国各地区的污水排放总量及主要污染物的排放数据发现, 废水的污染物组成大体可以分为两部分, 一类是混合型污染物如石油类等, 另一类是重金属元素如重金属“砷”, 经过典型相关分析的计算, 发现这两类污染物存在相伴相依的关系, 如化学需氧量污染物就含有多种其他变量相关, 于是构造了多元线性模型, 分析这些主要污染物之间存在的多元线性关系, 并计算相关关系的大小, 通过模型检验和诊断构造正确的回归模型, 最终可以得出化学需氧量和氨氮等变量有相关关系, 根据相关关系的大小和正负, 最后对这些主要污染物进行解释并提出相应的建议。

关键词

废水排放, 主要污染物, 典型相关分析, 多元线性模型

Copyright © 2019 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

1.1. 课题背景

废水排放一直是造成污染的主要原因。关于废水的处理, 在近些年一直被大家所重视, 习近平总书记也提出“绿水青山就是金山银山的”的科学论调。废水处理的关键在于找到主要污染物排放的规律, 现有的废水处理办法是单调的, 一般是只处理一种主要污染物, 虽然在某一个阶段处理掉了主要污染物, 但是其中的其他污染物经常是相伴而生的, 没有处理的污染物经过大江大河的汇聚将这些污染物增加, 造成了多种污染物的汇聚。

如果可以将污染物的内在规律研究清楚, 在源头就可以将这些污染物处理和中和, 这无疑是对废水处理是有很大帮助的, 查阅资料发现, 对于污水处理的资料较多, 但是关于污染物相关关系的资料比较少, 于是想到通过统计的办法将主要污染的相关关系进行研究。

1.2. 数据简介

1.2.1. 数据来源

本例选取了 2014 年度全国 31 个省、直辖市、自治区分地区废水中主要污染物的排放情况(2014 年)。数据来源于《中国统计年鉴(2015)》(数据经整理)。具体数据内容见附录一。

1.2.2. 变量解释

排放总量, 是指在在各省市区内废水排放的总量。

化学需氧量, 在一定条件下, 用一定的强氧化剂处理水样时所消耗的氧化剂的量, 是有机物污染的单位。

氨氮, 氨氮是指水中以游离氨(NH₃)和铵离子(NH₄⁺)形式存在的氮。

总氮, 水中各种形态无机和有机氮的总量。

总磷, 水样经消解后将各种形态的磷转变成正磷酸盐后测定的结果。

石油类，石油化工企业没有处理完全排放到废水中的污染。

挥发酚，主要污染源为煤气洗涤、炼焦、合成氨、造纸、木材防腐和化工行业的工业废水。

铅、汞、镉、六价铬、总铬和砷，都是易溶于水的微量重金属，人体吸入微量即可中毒。

1.2.3. 变量分类

观察图 1 的数据，发现这些数据大概分两类[1]，其中一类是污染物排放大量的混合类的物质，就是化学需氧量、氨氮、总氮、总磷、石油类和挥发酚，另一类是微量的重金属元素，也就是铅、汞、镉、六价铬、总铬和砷。

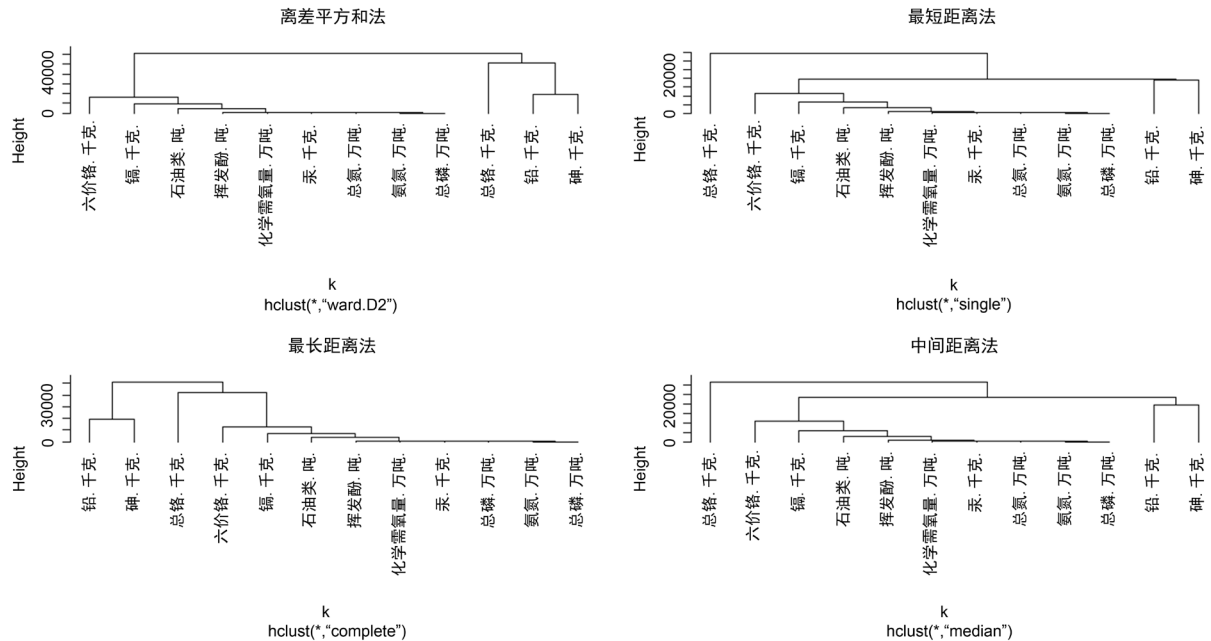


Figure 1. Tree charts of four classification methods

图 1. 四种分类方法的树状图

2. 典型相关分析

典型相关分析[2]是分析两组随机变量线性密切程度的统计方法，是两变量间线性相关分析的推广，利用的是主成分的思想进行计算，于是利用这两类污染物进行典型相关分析，通过 R 语言的运算[3]得到：

\$Cor

[1] 0.8422213 0.6919939 0.4832242 0.2977800 0.1298041 0.0638123

\$Sxcoef

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]
化学需氧量.万吨.	0.34878675	0.3082448	0.0739831	-0.0250541	-0.0449042	0.5322099
氨氮.万吨.	-0.36058566	-0.0846835	0.0696620	0.2243247	-0.0561619	-0.2627890
总氮.万吨.	-0.06065231	-0.4390986	0.1843135	0.4685476	-1.2018838	-0.2318784
总磷.万吨.	-0.09058587	0.0928561	-0.2908861	-0.6168346	1.2764255	0.0836312
石油类.吨.	0.00475546	0.1409063	-0.1490477	-0.0924792	-0.0597922	-0.0580233
挥发酚.吨.	0.03420705	-0.0633240	-0.0576192	0.1667018	0.0608844	0.0546728

Sycoef

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]
铅.千克.	-0.1684614	0.8257916	0.4445493	-0.2978355	0.7222736	0.2164321
汞.千克.	0.0874625	-0.0836052	-0.2475170	0.3283287	-0.0641557	-0.0757728
镉.千克.	-0.2137683	-0.3930189	0.3744792	0.1202555	-0.2767173	-0.0998673
六价铬.千克.	-0.0635944	0.1317535	0.0804124	0.0368075	-0.1247078	-0.1173750
总铬.千克.	-0.0765014	-0.1345282	-0.2174887	-0.0225775	0.0510550	0.0163813
砷.千克.	0.2542479	-0.3480042	-0.5996226	-0.0517086	-0.4483730	0.1108174

\$xcenter

化学需氧量.万吨.	氨氮.万吨.	总氮.万吨.	总磷.万吨.	石油类.吨.	挥发酚.吨.
3.14489e - 17	2.05928e - 17	-1.07441e - 17	1.65638e - 17	1.16394e - 17	4.09898e - 18

\$ycenter

铅.千克.	汞.千克.	镉.千克.	六价铬.千克.	总铬.千克.	砷.千克.
1.99213e - 17	1.34301e - 17	-4.70054e - 18	1.81866e - 17	2.95463e - 17	7.16273e - 18

\$cor 给出了典型相关系数,体现了各对典型变量的相关性从大到小的性质, $r_1 = 0.842$, $r_2 = 0.692$,因此这两组数据存在高度相关的关系,即混合类物质越多,微量的重金属元素前两对标准化的典型变量的线性组合是:

$$U_1 = 0.349X_1 + 0.308X_2 + 0.074X_3 - 0.025X_4 - 0.045X_5 + 0.5322X_6$$

$$V_1 = -0.168Y_1 + 0.826Y_2 + 0.444Y_3 - 0.298Y_4 + 0.722Y_5 + 0.216Y_6$$

$$U_2 = -0.361X_1 - 0.085X_2 + 0.069X_3 + 0.224X_4 - 0.056X_5 + 0.263X_6$$

$$V_2 = 0.088Y_1 - 0.084Y_2 - 0.248Y_3 + 0.328Y_4 - 0.064Y_5 - 0.076Y_6$$

第一对典型相关变量(U_1, V_1)中, U_1 是混合类物质的线性组合,其中 X_6 、 X_1 和 X_2 比其他的变量的贡献率要更高一些,说明化学需氧量、氨氮和挥发酚是混合物质的主要指标,占据主导地位。 V_1 是微量的重金属元素的线性组合,其中 Y_2 、 Y_5 和 Y_3 的贡献率较大,说明微量的重金属元素中汞、总铬和镉是主要指标。

第二对典型相关变量(U_2, V_2)中, U_2 是混合类物质的线性组合,其中 X_3 、比其他的变量的贡献率要更高一些,说明总氮是混合物质的主要指标,占据主导地位。 V_2 是微量的重金属元素的线性组合,其中 Y_4 和 Y_3 的贡献率较大,说明微量的重金属元素镉和六价铬是主要指标。

3. 多元线性回归

3.1. 理论基础

3.1.1. 最小二乘估计

含有 $p-1$ 个自变量的理论线性回归模型的一般形式[4]

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_{p-1} X_{p-1} + e$$

其中 $\beta_0, \beta_1, \dots, \beta_p$ 是模型的参数, e 为误差项,反映了随机因素对 Y 的影响我们将其展开成矩阵形式,

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix}, \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_n \end{pmatrix}, X = \begin{pmatrix} 1 & X_{11} & \dots & X_{1p} \\ 1 & X_{21} & \dots & X_{2p} \\ \dots & \dots & \dots & \dots \\ 1 & X_{n1} & \dots & X_{np} \end{pmatrix}, \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_n \end{pmatrix}$$

从而多元线性模型写成 $Y = X\beta + \varepsilon$ 。使用最小二乘法计算回归系数的估计值 $\hat{\beta}$ ，通过矩阵运算完成。
 β 的最小二乘估计 $\hat{\beta} = (X^T X)^{-1} X^T y$ ，残差向量为 $\hat{\varepsilon} = y - X\hat{\beta}$ ， σ^2 的最小二乘估计为 $\hat{\sigma}^2 = \frac{\hat{\varepsilon}^T \hat{\varepsilon}}{n-p-1}$ 。

3.1.2. 显著性检验

在多元线性回归方程中， t 检验和 F 检验并不一致，在 p 个自变量中，只要有一个自变量与因变量的线性关系显著， F 检验就会通过，但是这并不说明每个自变量在线性模型中都影响显著，所以 t 检验是必要的。

1) 回归方程线性关系的检验

回归方程的 F 检验是总体的显著性检验，提出假设为

$$H_0: \beta_0 = \beta_1 = \dots = \beta_p = 0, \quad H_1: \beta_0, \beta_1, \dots, \beta_p \text{ 至少有一个不为 } 0$$

$$\text{计算 } F \text{ 检验的统计量 } F = \frac{\beta_i' X_c' / (P-1)}{SSE / (n-p)}$$

2) 回归系数显著性检验

回归方程通过整体显著性检验后，我们对各回归系数单独进行检验，首先提出假设。

$$H_0: \beta_i = 0, \quad H_1: \beta_i \neq 0 (i = 1, 2, 3, \dots, p)$$

计算检验的统计量

$$t_i = \frac{\hat{\beta}_i}{s_{\hat{\beta}_i}} \text{ 服从 } t(n-p) \text{ 的 } t \text{ 分布}$$

3.2. 模型的建立

通过典型相关关系，发现了贡献率较高的数据是第一对数据，于是选取第一对 (U_1, V_1) 中的主要指标作为我们分析的主体，发现其中几个变量与其他的变量相关性较强，于是建立多元线性模型[5][6]，考虑变量之间的关系。

设立多元线性模型为：

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + \beta_8 X_8 + \beta_9 X_9 + \beta_{10} X_{10} + \beta_{11} X_{11}$$

首先，在 95% 的置信区间下，选用混合类物质变量中的化学需氧量、氨氮和挥发酚三个变量进行分析。

发现化学需氧量这个回归线性模型检验 $p = 0.337 \times 10^{10}$ 通过了模型检验而且有三个变量通过了检验，拟合优度 $R^2 = 0.941$ 说明模型的拟合效果较好，可以达到线性模型的要求，氨氮这个变量线性模型也通过了检验，拟合优度 $R^2 = 0.911$ 但是只有一个变量通过了检验，所以不选用这个变量，挥发酚这个变量发现回归模型的拟合优度只有 $R^2 = 0.012$ ，不适用线性模型。

在 95% 的置信区间下，选用重金属元素的相关变量，重金属元素镉和六价铬两个变量进行分析。

元素镉变量线性模型检验 $p = 0.197 \times 10^9$ 通过了检验，拟合优度 $R^2 = 0.929$ 达到线性模型要求，而且有三个变量通过了检验，满足线性模型要求，六价铬这个变量线性模型检验通过了检验，但是只有一个变量通过了检验，拟合优度 $R^2 = 0.393$ 也不是很高，不适用线性模型，内容见附录四。

综上所述，选出了化学需氧量和镉这两个变量进行线性模型的构建，剔除了没有通过检验的变量，化学需氧量和镉这两个变量可以写出的回归方程是：

$$Y_1 = 5.289 + 9.840X_4 - 0.003X_{11} - 0.001X_{13}$$

$$Y_2 = -36.579 + 0.889X_3 + 14.089X_4 - 0.430X_7 + 0.272X_9$$

3.3. 回归诊断

3.3.1. 残差和残差图

在进行模型假设的时候，常常假定模型误差 ε_i 满足了下列条件

- 1) $Var(e_i) = \sigma^2$
- 2) $Cov(e_i e_j) = 0$ ，其中 $i \neq j$

线性模型可以写成 $Y = X\beta + \varepsilon$ ，由回归系数估计值 $\hat{\beta} = (X^T X)^{-1} X^T Y$ 得到 Y 的拟合值为：

$$\hat{Y} = X\hat{\beta} = X(X^T X)^{-1} X^T Y = HY$$

其中 $H = (X^T X)^{-1} X^T$ 称为帽子矩阵，从而残差向量为 $\hat{\varepsilon} = Y - \hat{Y} = (1 - H)Y$ 。

因为这些假设都是关于误差项的，要分析他们的估计量(残差) [7] 的角度解决，所以要进行残差分析。于是得到了两个线性模型绘制残差图。

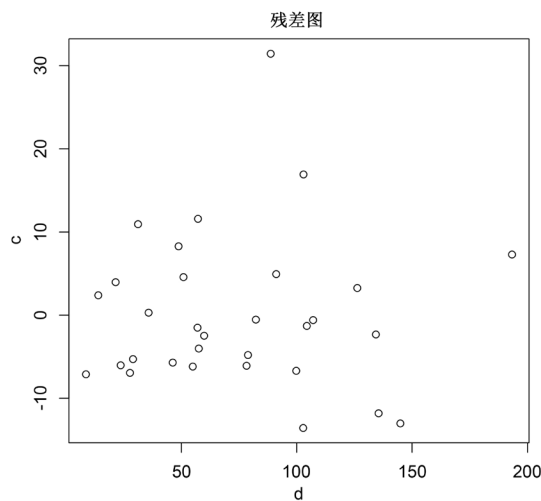


Figure 2. Residual graph of the first linear model

图 2. 第一个线性模型的残差图

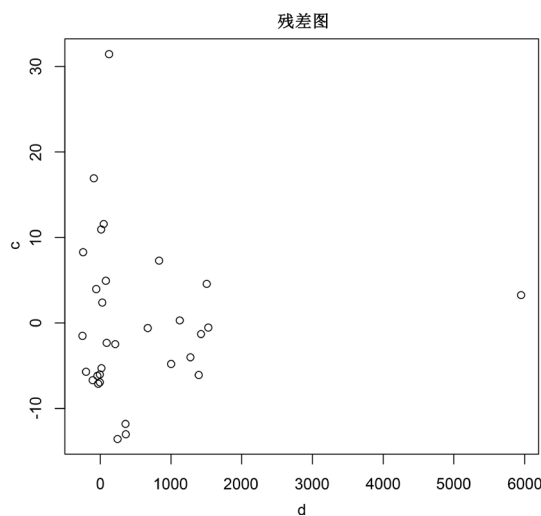


Figure 3. Residual graph of the second linear model

图 3. 第二个线性模型的残差图

通过上图明显发现，图 2 无明显违反马尔克斯假定的征兆，而图 3 则相反。所以决定弃用第二个线性回归模型。

3.3.2. 异常点检验

对参数估计或预测值有异常的数据，称为强影响数据，如果个别数据对估计有异常大的影响，当我们剔除掉这些数据，将得到和原来差异很大的经验回归方程[8]，可以怀疑原方程的正确性，从而建立更加可信的方程。

判断强影响点的方法是 DFFITS 统计量，根据矩阵对焦元素计算，公式：

$$D_i(\sigma) = \sqrt{\frac{h_{ii}}{1-h_{ii}}} \cdot \frac{\hat{\epsilon}_i}{\hat{\sigma}_{(i)}\sqrt{1-h_{ii}}}$$

对于 i 个样本，若

$$D_i(\sigma) > 2\sqrt{\frac{p+1}{n}}$$

则认为第 i 个样本是强影响点，可能是异常值。

通过计算第一个模型的异常值检验得到：

1	2	3	4	5	6	7	8	9	10	11	12
FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
13	14	15	16	17	18	19	20	21	22	23	24
FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
25	26	27	28	29	30	31					
FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE					

可以从上图中发现，第 8 个样本存在异常点，回看数据发现是黑龙江省的废水排放数据，并没有发现数据有缺失或异常。

于是运用 R 语言去掉这个异常值，重新拟合了一个回归方程：

$$Y_1 = 4.656 + 9.367X_4 - 0.003X_{11} - 0.0007X_{13}$$

两个回归方程对比发现变化并不大，为了保证数据完整性，仍然选用没有剔除异常值的回归方程。

3.3.3. 多重共线性诊断

多重共线性[9]是指线性回归模型中的解释变量之间由于存在线性关系或者近似线性的关系，而使得模型难以估计准确。

检验多重共线性我运用的是方差扩大因子的方法，判断多重共线性。

自变量 X_j 的回归系数估计值 $\hat{\beta}_j$ 的方差可以表示为：

$$\text{Var}(\hat{\beta}_j) = \frac{\sigma^2}{\sum x_j^2} \cdot \frac{1}{1-R_j^2} = \frac{\sigma^2}{\sum x_j^2} = \frac{\sigma^2}{\sum x_j^2} \cdot \text{VIF}_j$$

VIF_j 即为变量 X_j 的方差扩大因子， R_j^2 为自变量 X_j 对其余自变量作回归分析的复相关系数。方差扩大因子越大，表明自变量之间的多重共线性越严重，造成的回归估计值不稳定。

通过计算回归模型的多重共线性情况如下：

```
vif(aa)
      x4      x11      x13
```

```
1.835484 1.172172 1.619351
```

```
> sqrt(vif(aa))>2
```

```
    x4    x11    x13
```

```
FALSE FALSE FALSE
```

可以看到 X_4, X_{11}, X_{13} 的方差扩大因子都小于 10 而且接近于 1, 说明模型没有多重共线性的倾向。

4. 线性模型解释

对于回归模型进行诊断过后,发现线性模型成立[10],模型的因变量是化学需氧量,而自变量是氨氮、镉和总铬这三个变量,通过查阅资料发现,化学需氧量 COD 反映了水中还原物质污染的程度,这种还原物质包括了有机物、亚硝酸盐、硫化物等物质。因为水很容易被有机物污染且十分普遍,所以化学需氧量是用来衡量水被有机物污染程度的一个变量。氨氮是水中的营养素,可以导致水富有营养化,是水中耗氧的主要污染物,对鱼和水生物有很大的危害。镉元素的毒性很大,常用在电池中,电池的随意丢弃,是造成水质中镉元素多的主要原因。总铬产生于合金材料、染料、陶瓷等产业排放的污水中,有极强的氧化性,氧化后的铬元素有剧毒。

通过以上的资料查询,发现回归模型的现实意义与数学意义十分相近,化学需氧量和氨氮都是评价水中有机物污染的废水排放指标,所以他们的的相关性极强,而化学需氧量是评价水中强酸类排放的指标,而后两种元素和酸性物质很容易发生反应,而且产生这两种污染的企业往往不会同时产生,属于轻工业和农业的区别,我们知道各省市地区的发展经常是偏向一方的。

综上所述,对于废水排放的处理[11]提出两点建议:1) 在有机物、亚硝酸盐、硫化物等物质存在较多的废水处理过程中,不但要注意对这些物质的处理,同时我们要关注氨氮污染物的处理,因为这两类污染物常常是相伴而生的。2) 在混合废水中,有机物、亚硝酸盐、硫化物等物质存在,镉和总铬污染物往往存在比较少,这可能是两类物质相互发生了反应,也可能是各省市经济侧重点不同,当然造成这两种废水不会同时产生的原因,还有待研究。

5. 总结

本文首先对数据进行四种办法的分类,发现数据答题可以分为两类,将两类污染物进行典型相关分析,观察发现第一对相关关系的贡献率十分高,于是我们选取了第一对中的主要指标作为我们分析的主体,之后为了分析他们的线性相关性,建立了多元线性模型,通过 R 语言的计算,得出适当的回归方程,再通过模型的诊断和检验,判断出最正确的线性回归模型。

通过查阅资料,对线性模型的现实意义进行解释,发现与线性模型的数学意义相符,从而再次确定了模型的正确性,之后根据回归方程变量系数的大小,判断变量的正负相关性和他们的相关性大小,得出了相应的结论。回归方程的建立解释了废水主要污染物排放规律,根据变量的相关性大小,我们给出具体的建议,希望对今后的废水处理提供一些帮助。

参考文献

- [1] 王禹. 基于线性回归分析的特征抽取及分类应用研究[D]: [硕士学位论文]. 扬州: 扬州大学, 2016.
- [2] 孙权森, 曾生根, 王平安, 等. 典型相关分析的理论及其在特征融合中的应用[J]. 计算机学报, 2005, 28(9): 1524-1533.
- [3] 李诗羽, 张飞, 王正林. 数据分析: R 语言实战[M]. 北京: 电子工业出版社, 2014.
- [4] 吕伏, 梁冰, 孙维吉, 等. 基于主成分回归分析法的回采工作面瓦斯涌出量预测[J]. 煤炭学报, 2012, 37(1): 113-116.

- [5] 王荣本, 徐友春, 李兵, 等. 基于线性模型的导航路径图像检测算法研究[J]. 公路交通科技, 2001, 18(2): 47-51.
- [6] 娄思佳, 王书博, 王奇. 基于多元线性回归的任务定价规律模型[J]. 中国高新区, 2018(4): 5-15.
- [7] 周朴, 王小林, 马阎星, 等. 基于随机并行梯度下降算法自适应光学系统的校正残差分析[J]. 光学学报, 2010, 30(3): 613-617.
- [8] 王志高, 姚德利. 多元线性回归模型在江苏建筑业总产值预测中的效果分析[J]. 太原城市职业技术学院学报, 2018(5): 58-60.
- [9] 赵松山. 关于多重共线性检验方法的研究[J]. 山东工商学院学报, 2001, 15(4): 296-300.
- [10] 朱念, 贺嘉, 秦安莉, 等. 基于灰色线性回归组合模型的我国家用陶瓷出口预测研究[J]. 数学的实践与认识, 2018, 48(10): 23-30.
- [11] 段汉明, 张刚. 西安城市地域空间结构发展框架和发展机制[J]. 地理研究, 2002, 21(5): 627-634.

附录

原始数据

1	2	4	5	6	7	8	9	10	11	12	13	14
地区	排放总量 (万吨)	氨氮 (万吨)	总氮 (万吨)	总磷 (万吨)	石油类 (吨)	挥发酚 (吨)	铅 (千克)	汞 (千克)	镉 (千克)	六价铬 (千克)	总铬 (千克)	砷 (千克)
北京	150714	1.90	3.71	0.48	51.1	0.4	41.2	0.1	0.6	157.4	266.6	8.0
天津	89361	2.45	3.68	0.47	58.8	1.1	95.5	5.4	2.9	67.3	299.4	12.5
河北	309824	10.27	38.54	4.72	964.1	33.9	321.7	2.5	14.5	2619.3	5650.0	52.4
山西	145033	5.37	9.25	1.03	957.6	653.5	299.3	42.8	52.0	19.3	802.2	264.5
内蒙古	111917	4.93	18.88	2.15	1226.5	183.5	7057.9	44.0	760.5	37.1	88.7	15637.9
辽宁	262879	10.01	20.78	2.89	789.7	8.4	130.5	6.8	24.3	199.2	684.8	71.1
吉林	122171	5.31	12.83	1.59	231.9	3.9	165.1	6.3	30.2	135.0	216.4	889.8
黑龙江	149644	8.49	27.56	2.84	223.3	4.3	44.0	1.6	3.9	40.2	100.3	12.8
上海	221160	4.46	1.50	0.20	656.0	1.7	131.6	8.2	6.9	1108.8	2523.9	71.4
江苏	601158	14.25	17.42	1.86	1160.1	44.8	1204.1	3.9	26.5	3553.8	9676.5	307.6
浙江	418262	10.32	9.59	1.16	506.2	5.0	454.3	6.5	244.4	4708.7	12902.6	185.5
安徽	272313	10.05	18.62	2.00	709.8	5.4	1345.4	6.4	144.9	221.4	757.9	2244.6
福建	260579	8.93	9.32	1.29	373.4	1.9	3727.2	12.6	633.9	2145.6	10390.5	3754.0
江西	208289	8.60	11.24	1.52	687.5	14.2	6145.3	74.8	1769.5	570.7	922.3	7355.3
山东	514423	15.50	57.42	6.37	507.4	35.2	876.6	11.2	1050.7	509.0	7859.6	2400.0
河南	422832	13.90	42.72	5.08	1069.7	116.2	3138.0	20.3	784.2	966.0	27844.0	1083.9
湖北	301704	12.04	19.28	2.38	939.6	12.9	5881.8	35.6	1039.0	10223.9	13538.7	12448.0
湖南	309960	15.44	22.29	2.71	542.7	18.4	21609.3	151.7	6536.5	1252.5	5918.9	35794.2
广东	905082	20.82	18.82	2.83	450.5	10.0	2277.9	20.3	396.3	3278.8	13355.1	855.8
广西	219304	7.93	11.43	1.41	269.9	10.0	5009.3	90.6	953.8	132.2	1353.4	5012.2
海南	39351	2.29	4.09	0.52	47.3	0.1	2.5	3.7	0.5	0.3	101.0	13.2
重庆	145822	5.13	5.44	0.68	328.8	8.5	112.7	0.5	4.8	352.2	706.0	63.4
四川	331277	13.47	22.49	2.64	545.5	2.8	1208.8	15.3	86.2	751.2	1790.4	1678.7
贵州	110912	3.80	4.70	0.49	330.8	0.6	396.6	9.9	197.4	70.3	9351.4	276.3
云南	157544	5.65	7.67	0.77	319.3	2.1	4846.4	16.4	845.7	105.5	189.7	7354.1
西藏	5450	0.34	0.62	0.05	0.8	5.4	5.0	0.1	1.0	0.6	2.0	5203.5
陕西	145785	5.82	10.00	1.00	621.6	5.3	1426.1	28.9	512.1	165.5	696.0	1017.5
甘肃	65973	3.81	5.12	0.46	281.0	4.4	4382.2	88.1	824.4	498.7	3406.1	3265.5
青海	23001	0.98	0.79	0.08	339.3	1.0	692.7	8.3	244.1	5.5	12.6	1479.0
宁夏	37277	1.66	3.08	0.33	165.5	156.9	30.5	1.8	2.5	20.1	105.3	70.0
新疆	102748	4.59	17.27	1.44	847.9	26.8	125.1	21.6	56.9	1009.3	1285.2	847.1