

基于数学建模与数据分析的 城市居民生活用水优化策略研究

文◆上海外国语大学附属外国语学校松江云间中学 杨子聪

引言

通过数学建模和数据分析方法，系统性探索城市居民的用水模式，并为制定可行的改善措施提供理论支持。不仅提供了深入的用水模式分析，还提出了实用的优化策略，对城市水资源的有效管理具有重要意义，对于推动城市环境的可持续性和提高居民生活质量具有重要的实践价值。

1 数据收集与预处理

1.1 数据来源

为了全面理解和分析城市居民的用水行为，采用问卷调查法。发放在线问卷共计 650 份，回收的有效问卷共 500 份。

1.2 数据类型与编码

收集到的数据集涵盖了以下主要类别。

用水时间 (X_1): 反映了居民用水的具体时间段。时间以一天中的小时为单位编码，从 0 ~ 23，每个数字代表一天中的一个小时。

用水类型 (X_2): 标识了居民进行的不同用水活动，并对用水类型进行了以下编码。

(1) 饮用—表示饮用水。

(2) 洗涤—包括洗衣、洗碗等清洁活动。

(3) 冲厕—包括厕所冲洗等卫生用水。

家庭规模 (X_3): 代表每户居民的家庭成员数。家庭规模以 1 ~ 5 的整数进行编码，每个数字代表家庭的人数。

用水量 (Y): 量化了居民的总用水量，以升为单位，是一个连续的数值变量，反映了在收集期间每个家庭用水总量。

2 数据描述性分析

2.1 用水量分析

2.1.1 月均用水量分析

通过将收集到的数据按月份分组并计算每月的平均用水量，观察不同月份间用水量的差异，月均用水量如图 1 所示。4 月的平均用水量较高，1 月和 3 月的平均用水量较低。这种差异与季节性因素有关，如气温、节假日、居民生活习惯变化等。在较暖和的季节，用水量的增加与户外活动、园艺需求以及洗浴和洗衣频率增加有关^[1]。

2.1.2 用水量的分布

通过绘制直方图和箱线图分析用水量的分布情况，用水量分布如图 2 所示。图表揭示了用水量在不同范围内的频数分布。60 ~ 70 范围内的用水量频数最高，表明大多数居民的日常用水量集中在此范围，对于

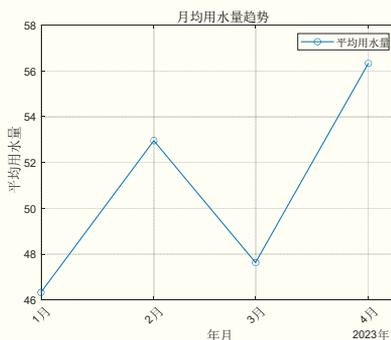


图 1 月均用水量

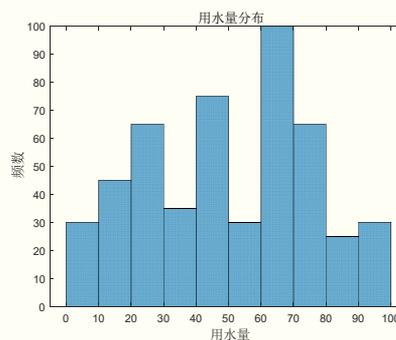


图 2 用水量分布

【作者简介】杨子聪，男，上海人。

【指导老师】王子文（1997—），男，上海人，硕士研究生，研究方向：控制科学与控制工程。

水利局在资源分配和规划方面具有重要意义，可以帮助水利局更精确地满足居民的用水需求。

2.1.3 用水量的波动性

计算用水量的标准差为 26.1947，表明用水量的波动性相对较高。这种波动与家庭规模、居民的生活习惯、季节性活动等因素有关^[2]。数据点间较大的离散程度表明居民用水行为的多样性和复杂性。

2.2 用水时间分析

2.2.1 高峰和低谷时段的识别

通过比较每个小时段的平均用水量与整体平均用水量，确定了用水量高于平均值的时段为高峰时段，低于平均值的时段为低谷时段。高峰时段的平均用水量为 2.703，高于整体平均值 2.059，而低谷时段的平均用水量为 1.158，明显低于整体平均值。

2.2.2 结果解释

(1) 高峰时段的特点

时段：早上 8 点至下午 5 点（8:00—17:00）。

用水量：平均用水量相对较高，特别是在早上 8 点至下午 2 点（8:00—14:00），用水量最高。

原因分析：这个时段是家庭成员进行早餐准备、清洁、洗漱等日常活动的高峰时段。午餐准备和清洁活动也导致这个时间段用水量增加。

(2) 低谷时段的特点

时段：晚上 10 点至早上 7 点（22:00—7:00）。

用水量：平均用水量最低。

原因分析：这段时间大多数家庭成员正在睡眠，因此用水需求显著减少。除了个别夜间活动，如洗澡或喝水，晚间的用水量通常较低。

(3) 其他注意事项

午后减少：下午 2 点至 5 点（14:00—17:00）用水量略有下降，但仍然高于日均水平。

晚间小幅升高：晚上 6 点至 9 点（18:00—21:00）用水量略有上升，家庭成员进行晚餐准备和洗漱等活动导致用水量上升。

3 数学模型建立

3.1 定义变量

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \epsilon$$

在数学模型中，定义因变量 Y 表示居民的用水量或用水行为，自变量 X_1, X_2, \dots, X_n 代表影响用水量或行为的各种因素，如用水时间、用水类型、家庭规模等。模型参数 $\beta_0, \beta_1, \dots, \beta_n$ 是需要估计的系数，表明自变量对因变量的影响程度。误差项 ϵ 代表模型无法解释的随机变异。

3.2 模型拟合

3.2.1 最小二乘法原理

在数学建模中，最小二乘法是一种常用的参数估计方法^[3]。它的核心思想是通过最小化残差的平方和估计模型中的参数^[4]。残差是指实际观察值与模型预测值之间的差异。数学上，可以表述为寻找一组参数 $(\beta_0, \beta_1, \dots, \beta_n)$ ，使所有数据点的残差平方和最小。

3.2.2 拟合过程

(1) 构建目标函数。首先，构建一个目标函数，即残差平方和 (RSS)。公式为 $RSS = \sum_{i=1}^n (Y_i - (\beta_0 + \beta_1 X_{i1} + \dots + \beta_n X_{in}))^2$ ，其中 Y_i 是第 i 个观测值， X_{i1}, \dots, X_{in} 是与 Y_i 相关的自变量的值。

(2) 参数估计。使用数学或统计软件最小化 RSS。通常涉及求解一组非线性方程，以找到最佳的参数值 $\beta_0, \beta_1, \dots, \beta_n$ 。

(3) 模型的拟合。得到参数估计值后，代入模型中，得到拟合模型，用来预测或解释因变量 Y 。

3.2.3 优化和迭代

(1) 迭代过程。在某些情况下，需要通过迭代过程逐步逼近最优解。通常是在更复杂的模型中进行，如当模型包含大量变量或非线性关系时。(2) 优化技术。为了提高拟合过程的效率和准确性，使用各种优化技术，如梯度下降、牛顿法等。

3.2.4 模型拟合的评估

(1) 残差分析。完成拟合后，进行残差分析评估模型的适用性。残差分布的检查（如是否接近正态分布）可以提供模型是否恰当的重要线索。(2) 诊断检测。进行各种统计诊断，如多重共线性检测、异方差性检测等，满足模型假设。

3.3 残差 / 误差

模型的残差或误差是衡量模型预测精度的重要指标^[5]。残差平方和 (RSS) 是评估模型拟合好坏的关键指标，计算公式为 $RSS = \sum_{i=1}^n (Y_i - (\beta_0 + \beta_1 X_{i1} + \dots + \beta_n X_{in}))^2$ 。RSS 越小，表示模型对数据的拟合越好。

3.4 模型评价

使用决定系数 R^2 评价模型的解释能力。 R^2 表示模型能解释

数据变异的比。将数据分为训练集（用于模型拟合）和测试集（用于模型验证），例如，70%的数据用于训练，30%的数据用于测试，评估模型在未知数据上的表现。

3.5 参数估计

利用最小二乘法 and R^2 统计，估计模型中的参数 β 。参数的估计值反映了每个自变量对因变量的影响程度，便于理解不同因素对居民用水行为的影响。

4 模型分析与结果

4.1 结果解释

4.1.1 模型的具体形式

使用 Matlab 进行计算，得到了线性回归模型来描述居民的用水行为，即

$$Y=62.6737-1.0996*X_1-$$

$$1.928*X_2+1.6939*X_3+24.3358$$

其中， Y 代表用水量， X_1 ， X_2 和 X_3 分别代表不同的影响用水量的因素。

4.1.2 参数解释

β_1 (对应 X_1): 系数为 -1.0996 。意味着在其他条件保持不变的情况下， X_1 每增加一个单位，用水量 Y 预计将减少 1.0996 单位。表明 X_1 是一个与用水量负相关的因素，例如，代表节水措施的效果或者某种减少用水的活动。

β_2 (对应 X_2): 系数为 -1.928 。表明 X_2 对 Y 的影响更为显著， X_2 每增加一个单位，用水量 Y 预计将减少 1.928 单位。意味着 X_2 是用水量减少的另一个重要因素，如更高效用水设备或节水技术的应用。

β_3 (对应 X_3): 系数为 1.6939。表明 X_3 的增加与用水量 Y 的增加呈正相关， X_3 每增加一个单位，用水量 Y 预计将增加 1.6939 单位。

4.1.3 模型的应用

(1) 用水管理策略。模型可以帮助城市水务管理者和政策制定者理解影响居民用水量的因素，制定更有效的用水管理策略和节水措施。(2) 预测未来用水趋势。通过分析这些参数，可以预测在特定条件下的用水趋势，如在提高节水意识或引入新技术后用水量的变化。

4.2 模型验证

4.2.1 R^2 (判定系数)

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \mu)^2}$$

(1) 解释力度。模型的 R^2 值为 0.111，意味着模型仅解释了约 11.1% 的因变量（用水量）的变异性。这是一个相对较低的值，表明模型在解释用水量变异性方面的能力有限。(2) 原因。较低的 R^2 值表明还有其他重要的未考虑变量对用水量有显著影响，或者用水量与所选自变量之间存在非线性关系。

4.2.2 MSE (均方误差)

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

(1) 预测误差。MSE 值为 614.52，提供了关于模型预测误差大小的信息。较高的 MSE 值表明模型在预测用水量方面可能存在较大误差。(2) 改进空间。高 MSE 值暗示模型需要进一步优化，如通过添加更多的解释变量、考虑非线性关系或交互作用。

结语

通过数学建模和数据分析方法深入探讨城市居民的用水行为，旨在提出有效的用水优化策略。构建线性回归模型，考虑用水时间、用水类型和家庭规模等因素对居民用水量的影响。模型结果显示，用水时间的增加未必导致用水量的增加，而用水类型和家庭规模对用水量的影响在统计上并不显著。因此，除了考虑的上述因素外，还有其他未考虑的因素影响着用水行为，如经济状况、文化习惯、气候条件等。

从优化策略的角度来看，推动非高峰时段用水、提高用水效率和考虑家庭内部的用水共享等措施，都是减少用水浪费和提高用水效率的有效途径。这些策略的实施需要政府、社区和个人的共同努力，通过制定合理的政策、开展教育和宣传活动以及采用先进的节水技术实现。

引用

- [1] 赵彤彤,侯文.基于LMDI模型的山东省生活用水变化及驱动因素分析[J].科学与管理,2021,41(5):88-94.
- [2] 汪林,张立影,贾玲,等.家庭生活用水精细化管理研究与应用[J].水利发展研究,2021,21(8):7-14.
- [3] 梁博尧.浅析最小二乘法及其在数学建模中的应用[J].中国新通信,2018,20(3):197-199.
- [4] 韩加坤.数学建模中的常用最小二乘方法[J].科教导刊(上旬刊),2018(7):38-39.
- [5] 颜焱,王鑫.残差在数学建模中的应用[J].安徽职业技术学院学报,2010,9(2):15-16.