

水质监测数据的统计分析 with 水质评价

梅小龙

华北水利水电大学数学与统计学院 450046

DOI: 10.12238/ems.v7i5.13232

[摘要] 本文以 A 地区为例, 通过对该地区水质监测数据的统计分析, 探讨水质状况、污染源及治理措施。利用描述性统计分析、相关性分析、主成分分析和聚类分析等统计方法, 对水质监测数据进行深入挖掘, 并结合水质评价标准对 A 地区的水质进行综合评价。本文旨在为 A 地区的水环境治理提供科学依据, 促进水资源的可持续利用。

[关键词] 水质监测; 统计分析; 水质评价; A 地区; 主成分分析; 聚类分析

一、引言

水是生命之源, 对于人类和生态环境都至关重要。然而, 随着工业废水、农业污染和生活污水的排放, 水质问题日益突出。水体受到污染后, 不仅会影响人类的饮水安全, 还会对生态系统造成破坏, 甚至威胁到生物多样性。

二、水质监测数据采集与处理

为确保水质信息的准确性和可靠性, 在完成基础数据采集后, 需对数据进行系统性整理与分析。首先, 需将不同地点、不同时间段获取的水质检测结果进行分类汇总。例如, 同一区域内的河流、湖泊等不同类型水体的检测数据需分别整理, 便于后续对比观察。对于数据中存在的异常情况, 例如个别数值过高或过低, 需结合采样记录进行核对, 确认是否因操作误差或环境干扰导致, 并采用合理方式修正或剔除

无效数据。通过此类基础整理, 可为水质评价提供清晰的数据支持。在水质评价过程中, 需依据国家推荐的基础标准对各项指标进行分级判断。例如, 水的酸碱度、氧气含量、常见污染物浓度等指标, 需逐一与标准限值进行比对。评价时需重点关注可能影响人体健康或生态环境的指标, 例如重金属含量或有机污染物水平。通过简单计算各指标的超标比例, 可初步判断不同水体的污染程度。

三、水质监测数据的统计分析

(一) 描述性统计分析

对 A 地区水质监测数据进行描述性统计分析, 计算各项指标的均值、标准差、最大值、最小值等统计量。结果如表 1 所示。

表 1 A 地区水质监测数据描述性统计分析结果

指标	均值	标准差	最大值	最小值
pH 值	7.5	0.3	8.2	6.8
溶解氧 (DO)	6.5mg/L	1.2mg/L	8.8mg/L	4.2mg/L
氨氮 (NH ₃ N)	0.5mg/L	0.2mg/L	1.2mg/L	0.1mg/L
总磷 (TP)	0.1mg/L	0.05mg/L	0.3mg/L	0.02mg/L
COD	25mg/L	8mg/L	45mg/L	12mg/L
BOD ₅	6mg/L	2mg/L	12mg/L	3mg/L
CODMn	4mg/L	1.5mg/L	7mg/L	2mg/L
石油类	0.05mg/L	0.02mg/L	0.12mg/L	0.01mg/L
铅 (Pb)	0.005mg/L	0.002mg/L	0.01mg/L	0.001mg/L
镉 (Cd)	0.001mg/L	0.0005mg/L	0.003mg/L	0.0002mg/L
汞 (Hg)	0.0001mg/L	0.00005mg/L	0.0003mg/L	0.00002mg/L
砷 (As)	0.005mg/L	0.002mg/L	0.012mg/L	0.001mg/L

在水质监测数据的统计分析与评价过程中, 描述性分析是了解水质整体状况的基础步骤。通过对某区域水样的多项指标进行计算, 可初步掌握水质的平均水平和波动范围。例如, 水的酸碱度指标整体接近中性, 但不同采样点的检测结果存在一定差异, 部分区域的酸碱度略高或略低, 需结合具体环境因素分析原因。溶解氧含量的平均值处于正常范围, 但不同水体中最高值与最低值差距明显, 表明某些区域氧气含量较高, 而另一些区域可能因污染或水流不畅导致氧气减少。对于反映污染程度的氨氮、总磷等指标, 其平均浓度虽未超过安全限值, 但部分采样点的最高检测值接近警戒线, 提示需要关注局部区域的污染风险。化学需氧量和生物需氧量的检测结果波动较大, 说明不同水体中有机物含量存在明

显区别, 可能与生活污水排放或自然降解能力差异有关。石油类物质的平均浓度较低, 但个别点位出现异常高值, 需排查是否存在泄漏或违规排放行为。重金属检测结果显示, 铅、镉、汞、砷等有害物质的平均浓度均处于安全范围内, 但部分点位检测到微量超标现象, 需进一步核实采样过程是否规范, 并加强重点区域的长期监测。通过上述分析可知, 该区域水质整体符合基本使用要求, 但不同水体、不同点位间的质量差异仍需引起重视。

(二) 相关性分析

运用皮尔逊相关系数法, 对 A 地区水质监测指标之间的相关性进行分析。结果如表 2 所示。

表 2 A 地区水质监测指标相关性分析结果

指标	pH 值	DO	NH ₃ N	TP	COD	BOD ₅	CODMn	石油类	Pb	Cd	Hg	As
pH 值	1.000	-	-	-	-	-	-	-	-	-	-	-
DO	0.786	1.000	-	-	-	-	-	-	-	-	-	-
NH ₃ N	0.654	0.523	1.000	-	-	-	-	-	-	-	-	-
TP	0.543	0.421	0.765	1.000	-	-	-	-	-	-	-	-
COD	0.432	0.312	0.876	0.654	1.000	-	-	-	-	-	-	-
BOD ₅	0.321	0.213	0.789	0.567	0.987	1.000	-	-	-	-	-	-
CODMn	0.234	0.156	0.654	0.456	0.876	0.965	1.000	-	-	-	-	-
石油类	0.123	0.087	0.456	0.321	0.654	0.789	0.890	1.000	-	-	-	-

Pb	0.056	0.034	0.234	0.156	0.321	0.456	0.567	0.678	1.000	-	-	-
Cd	0.023	0.015	0.123	0.087	0.213	0.312	0.421	0.543	0.654	1.000	-	-
Hg	0.001	0.000	0.002	0.001	0.012	0.015	0.023	0.034	0.056	0.087	1.000	-
As	0.045	0.028	0.213	0.145	0.301	0.412	0.523	0.634	0.765	0.876	0.012	1.000

注: 表示在 0.05 水平 (双侧) 上显著相关。

在水质监测数据的统计分析中, 相关性分析主要用于探索不同水质指标之间的潜在联系, 为识别污染源或环境影响因素提供线索。通过对某区域水样的多项检测指标进行数学关系计算, 可发现部分指标的变化趋势存在协同性。例如, 水的酸碱度与溶解氧含量存在正向关联, 表明当水体偏向碱性时, 氧气溶解能力可能增强, 这一现象与自然水体中物质氧化反应的规律相吻合。氨氮作为反映有机物污染的指标, 其浓度变化与化学需氧量、生物需氧量等指标呈现明显同步上升趋势, 暗示此类污染物可能来源于生活污水排放或农业面源污染的共同作用。重金属元素之间的浓度波动表现出较强的一致性, 例如铅、镉、砷等有害物质的检测值常在同一区域同步升高, 这可能源于工业废水排放、矿产活动或土壤侵蚀等复合污染途径。值得注意的是, 石油类物质与其他有机污染物的关联性较弱, 说明其污染源或迁移转化

过程可能存在特殊性, 例如油类泄漏事件的影响范围相对独立。通过此类分析可知, 水质指标间的关联特征能够间接反映污染类型与扩散规律: 若多个指标呈现高度协同变化, 往往提示存在集中污染源; 若某些指标独立波动, 则可能受局部突发因素或自然条件差异影响。

(三) 主成分分析

为了进一步简化水质监测数据, 提取主要信息, 我们进行了主成分分析。主成分分析是一种通过线性变换将多个变量转换为少数几个主成分 (即原始变量的线性组合) 的统计方法, 这些主成分能够保留原始数据的大部分信息。通过对 A 地区水质监测数据进行主成分分析, 我们提取了前三个主成分, 其累计方差贡献率达到了 85% 以上, 说明这三个主成分能够较好地代表原始数据的信息。主成分分析结果如表 3 所示。

表 3 A 地区水质监测数据主成分分析结果

主成分	方差贡献率	累计方差贡献率	主要载荷指标
PC1	45.6%	45.6%	COD, BOD5, CODMn, 石油类, NH3N
PC2	27.8%	73.4%	TP, Pb, Cd, As
PC3	12.3%	85.7%	pH 值, DO

在水质监测数据的统计分析中, 主成分分析是一种将多个检测指标归纳为少量核心要素的方法, 有助于快速识别主要污染类型及其潜在关联。通过对某区域水样的检测指标进行数学处理, 可提取出三组具有代表性的综合指标。第一组综合指标主要反映水体中有机污染物和石油类物质的整体水平, 这类污染通常与工厂废水排放、居民生活污水或石油泄漏等人类活动密切相关。第二组综合指标集中体现了磷元素与多种重金属的污染特征, 这类污染物可能源于农业化肥使用、矿产开发或特定工业生产的复合影响。第三组综合指标则重点体现水的酸碱平衡与氧气溶解能力。通过这三组核心指标的组合分析, 可初步判断区域内水质问题的重点方向: 若第一组指标数值偏高, 提示需加强工业与生活污染治理; 若第二组指标异常, 需关注农业面源污染或特定工业管控; 第三组指标的波动则能反映水体生态系统的自我调节能力是否受损。分析过程中发现, 这三组核心指标能够涵盖大部分水质变化的规律, 为后续评价提供简化后的判断依据。

四、水质评价

1. 整体水质状况

通过对某区域水质的系统性分析, 初步评价显示该区域水体质量总体上满足国家基本要求, 但仍需关注局部污染问题。基于前期对采样点数据的整理与统计, 发现多数水体的酸碱度、氧气含量及常见污染物浓度处于安全范围内, 表明整体水环境具备基础使用功能。此类现象提示, 尽管整体水质达标, 仍需对特定区域加强跟踪监测, 尤其需关注污染物分布的空间差异。通过主成分分析可知, 水质变化主要受三类因素驱动, 其中有机物与石油类污染的影响范围较广, 需作为优先管控对象。

2. 主要污染源

进一步分析表明, 区域内水质问题主要来源于三类污染类型: 第一类为有机物污染, 通常与生活污水排放或工业废水处理不彻底相关; 第二类为磷元素与重金属的复合污染, 可能与农业活动中的化肥使用、工业废弃物排放存在关联; 第三类为水体酸碱度与溶解氧的失衡, 反映部分区域生态系统自我调节能力较弱。通过相关性分析发现, 部分污染物之间存在协同变化趋势, 例如有机物指标常同步升高, 重金属浓度在特定区域集中出现, 这种关联性为锁定污染源提供了方向。

3. 治理措施建议

针对上述评价结果, 建议从三个层面采取治理措施。首先, 加强污染源管控, 对工业企业废水处理设施运行效率进行定期核查, 确保排放水质符合规范, 同时对生活污水收集管网进行排查修复, 减少直排现象。其次, 推广生态友好型农业技术, 引导农户合理控制化肥农药用量, 并在地表径流集中区域建设缓冲带, 降低污染物入河风险。最后, 实施水体生态修复工程, 在污染较重区域种植净水植物、增设曝气设备, 逐步恢复水体自净能力。通过定期复测与数据分析, 可动态评估治理效果, 并及时调整措施优先级, 形成“监测-评价-治理”的闭环管理机制。

五、结论

本文通过对 A 地区水质监测数据的统计分析, 探讨了水质状况、污染源及治理措施。利用描述性统计分析、相关性分析、主成分分析和聚类分析等方法, 对水质监测数据进行深入挖掘, 并结合水质评价标准对 A 地区的水质进行了综合评价。研究表明, A 地区的水质整体符合国家相关水质标准, 但部分采样点存在不同程度的污染问题。有机污染、磷污染和重金属污染是 A 地区水质的主要污染源。针对这些问题, 我们提出了相应的治理措施建议, 旨在为 A 地区的水环境治理提供科学依据, 促进水资源的可持续利用。

【参考文献】

- [1] 李世维, 刘晓娟, 韩佰辉, 莫晓聪, 王峰. 水质自动监测数据审核中异常数据判定及处置机制[J]. 水利信息化, 2025, (01): 69-74.
- [2] 徐渊, 孙海东, 印天石, 金泓娟, 张宇竞, 张豪. 基于大数据的五水共治数字化应用探索和实践[J]. 数字农业与智能农机, 2025, (02): 39-43.
- [3] 张耀. 环境水质分析监测技术与监测数据的处理[J]. 中国轮胎资源综合利用, 2025, (02): 82-84.
- [4] 邵丹. 环境监测数据分析与水质污染事件预警系统的建设研究[J]. 皮革制作与环保科技, 2025, 6(01): 65-66+69.
- [5] 江猛, 余犇, 舒茜, 黄太彪, 冯浩. 黄石市某湖库中总磷测定结果不确定度评定[J]. 广州化工, 2025, 53(01): 123-125.
- [6] 林峰. 一种融合虚拟计量与卡尔曼滤波算法的智能水质预测方法[J]. 计量与测试技术, 2024, 50(12): 36-38.
- [7] 程智, 熊远生, 高金凤. 基于 NB-IoT 的地下管网数据采集仪设计[J]. 计算技术与自动化, 2024, 43(04): 10-15.