|  |  |  |
| --- | --- | --- |
| 选题 | **2024年第十四届APMCM**  **亚太地区大学生数学建模竞赛（中文赛项）** | 参赛编号 |
|  | apmcm\*\*\*\*\*\* |

标题（此处换成论文的标题）

摘要

(说明：以下开始写摘要，正文从下一页开始。摘要及正文格式基本要求是宋体，小四号，单倍行距，没有要求的地方就自行处理。看完后删除该说明)

1. 问题重述

1.1问题背景

ChatGPT吧！

1.2问题要求

附件1给出了各地季风强度、地形排水、河流管理等基本信息，以及该地发生洪水的概率。为了能更好地根据已知信息预测洪水的发生概率，最小化灾害损失，现需结合实际情况与所给信息建立数学模型，分析以下问题。

问题1：依据附件1中给出的基本信息，可视化处理相关数据，并单独分析洪水发生概率与每个基本信息的相关性，判断其关联程度大小，给出相关结论和建议。

问题2：

问题3：

问题4：

1. 问题分析

2.1问题1的分析

针对问题一，首先对train.csv中所给数据进行探索性分析，该步骤包括基本统计量计算，数据预处理等基础性工作，为后续的数据分析奠定基础。然后，采用概率密度直方图以及核密度估计（Kernel Density Estimation, KDE）图可视化洪水概率的分布情况，用频数统计直方图可视化train.csv与test.csv中各个指标的分布情况，并计算上述所有指标分布的偏度与峰度，实现量化评估指标与洪水概率的分布情况，为量化联系提供帮助。随后，基于对数据分布情况的分析，进行数据相关性分析，以Spearman相关系数为桥梁，通过假设检验，量化20个等级化离散指标与洪水概率之间的关系。最后，综合以上分析，指出对洪水概率影响较大的部分指标，并据此提出相应的预防措施。

2.2问题2的分析

针对问题二，对聚类问题，首先标准化train.csv中洪水概率数据，并利用K-Means++聚类算法对其进行聚类分析；随后对聚类结果进行箱型图及频数分布直方图可视化，并计算类别分组基本统计量，用以评估聚类结果有效性；最后划分风险为高、中、低三类，将其作为分类标签，并对其分别进行统计分析可视化，以分析不同风险对应指标特征。对风险预警评价问题，其实际是一个基于连续数值到逻辑类别区间映射的回归分类问题，故首先对回归概率转类别的映射区间进行区间填充，训练逻辑回归模型从而建立符合数据分布情况的洪水概率到风险等级的连续映射，以保证最终回归模型的分类评价能力的鲁棒性；随后，基于所有指标构建线性回归模型，探索数据的线性相关性，并对所有指标之和与洪水概率进行可视化以发现其线性相关性，因此引入指标和作为关键指标，然后将已有20指标在样本维度上从小到大排序，构造排序指标，从而消除指标维度上的数据差异性；最后，将排序指标与指标和作为最终指标训练CatBoost模型实现了优于基准线性回归模型的效果，并且其在三分类任务上表现依旧优秀，训练模型的同时，利用K折交叉验证法验证了模型的低灵敏度与高泛化性。

2.3问题3的分析

针对问题三，对于基于问题1分析结果选取合适指标进行回归预测，首先基于各指标与洪水概率的Spearman相关系数选出较优指标数据作为基本指标，随后加入指标和作为关键指标，训练CatBoost模型，在洪水概率预测上的性能指标有较大提升。对于仅选择五个关键指，使用主成分分析（Principle Component Analysis，PCA）法进行数据降维，将20维数据降维至4维，

2.4问题4的分析

**文章总体思路如图1所示：**

**图1总体思路图**

1. 模型假设

四、符号说明

五、数据侧写

1. 问题1模型的建立与求解

6.1基于数据的数字特征进行描述

题目要求对数据进行可视化处理，为了更好地描述洪水发生概率分布、基本信息特征的分布规律，本文引入峰度系数、偏度系数来描述统计数据。

1. 峰度系数

峰度系数（Kurtosis）用于衡量数据分布的峰度程度。例如时，数据的分布为正态分布；时，数据分布的峰度较小，数据更分散；时，数据分布的峰度较大，数据更集中。并且越大，数据分布的图像更加“陡峭”。峰度系数的计算公式如下：



1. 偏度系数

偏度系数（Skewness）用于衡量数据分布的偏斜程度。例如时，数据的分布为正态分布；时，数据呈左偏分布；时，数据呈右偏分布。并且越大，数据分布的偏斜越明显。峰度系数的计算公式如下：



6.2数据可视化

利用python作出洪水发生概率密度图像、基本信息的分布柱状图如下：

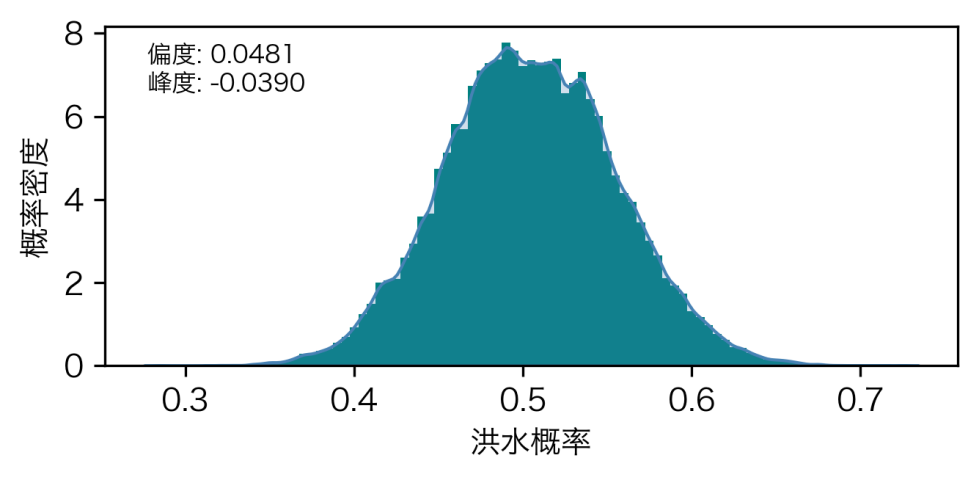
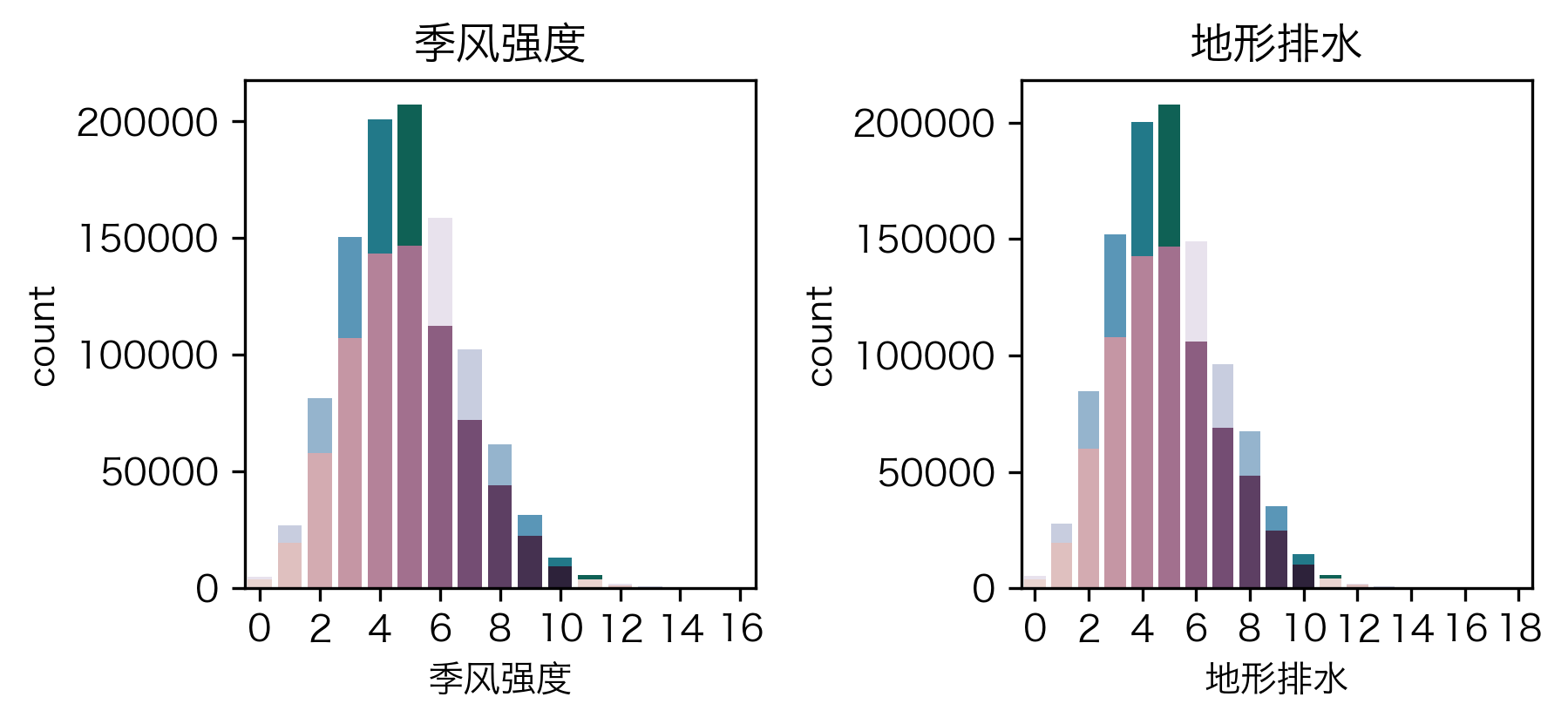


图2洪水概率密度的图像





七、问题2模型的建立与求解

八、问题3模型的建立与求解

九、问题4模型的建立与求解

十、模型的评价与推广

十一、参考文献

十二、附录

**参考文献 （可另起一页）**

参考文献的编号，如[1][3]等；引用书籍还必须指出页码。参考文献按正文中的引用次序列出，其中：**书籍的表述方式为**

[编号] 作者，书名，出版地：出版社，出版年。

**参考文献中期刊杂志论文的表述方式为**

[编号] 作者，论文名，杂志名，卷期号：起止页码，出版年。

**参考文献中网上资源的表述方式为**

[编号] 作者，资源标题，网址，访问时间（年月日）。

**附录（另起一页）**