

**2020年兰州大学数学建模竞赛**

**中文标题**  混合物中特定成分的检测

队伍编号 1 1 1

学 院 信息科学与工程学院

专 业 数据科学

学生姓名 蔡琪琦

丁俊伟

曹延飞

**摘要**

这里写上摘要正文。

**关键字：**关键字1 关键字2

目录

[1. 问题重述 4](#_Toc45988533)

[2. 数据准备 4](#_Toc45988534)

[3. 模型假设 4](#_Toc45988535)

[4. 通用符号说明 4](#_Toc45988536)

[5. 问题一：给出判定特定成分存在的主要指标 4](#_Toc45988537)

[5.1 问题分析 4](#_Toc45988538)

[6. 问题二：成分存在与否的判定是否存在模糊区域 4](#_Toc45988539)

[6.1 问题分析 4](#_Toc45988540)

[7. 问题三：建立数学模型，判断测试数据 5](#_Toc45988541)

[7.1 问题分析 5](#_Toc45988542)

[8. 模型分析与评价 5](#_Toc45988543)

[8.1 模型优点 5](#_Toc45988544)

[8.2 模型改进 5](#_Toc45988545)

# 问题重述

为了检测某种混合物中是否含有特定成分，对该混合物的25000个样本进行了7项指标(记为V1,V2,…,V7)的检测。检测结果见附件“Data.xlsx”。数据文件包含两个表单。第一个表单是“训练数据”，包含20000个混合物样本，已知混合物中是否含有特定成分，其中1表示含有特定成分，0表示不含特定成分。第二个表单是“测试数据”，包含5000个混合物样本，未知混合物中是否含有特定成分。请完成以下三个任务：

1. 给出判定特定成分存在的主要指标；
2. 成分存在与否的判定是否存在模糊区域（对于7项指标），即在这些区域中很难判断成分是否存在。如果存在，试估计这一区域；
3. 建立数学模型，判断测试数据中每一个混合物是否含有特定成分。论文中给出前10个混合物的判定即可。

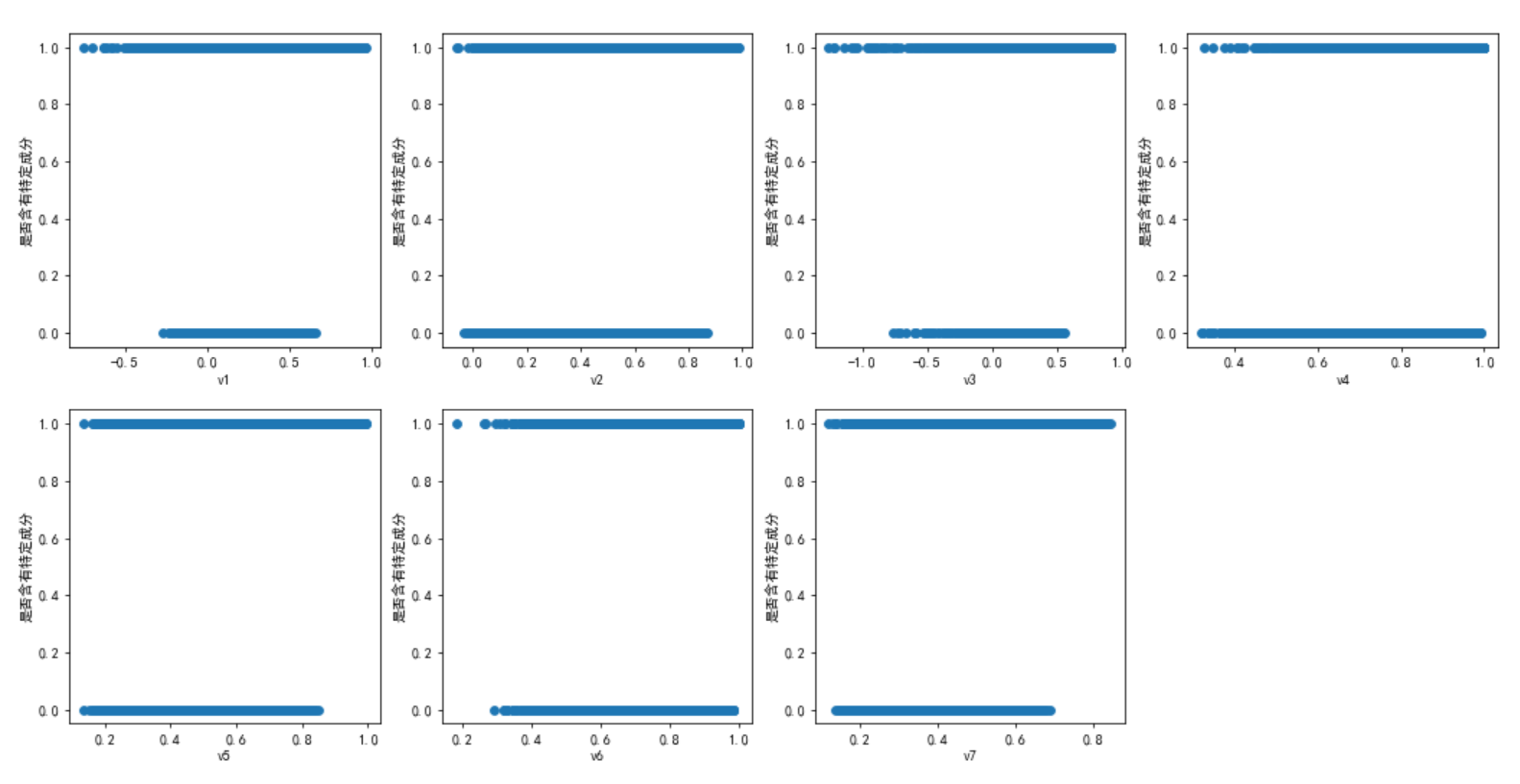
# 数据准备

### 数据来源

对某种混合物的25000个样本进行了7项指标(记为V1,V2,…,V7)的检测，判断是否含有特定成分。其中20000个样本作为训练数据，已知混合物中是否含有特定成分，1表示含有特定成分，0表示不含特定成分，5000个样本作为测试数据，未知混合物中是否含有特定成分。

### 数据清洗

在进行分析之前，需要先对数据进行清洗以及补值。因为数据集不存在缺失数据，而且通过分别绘制7项指标和是否含有特定成分的散点图可以看出，并不存在明显的异常离群值，故没有对原有数据进行任何操作。



# 模型假设

这里写模型假设。

# 通用符号说明

这里写符号说明。

# 问题一：给出判定特定成分存在的主要指标

## 问题分析

问题一要求给出判定特定成分存在的主要指标。首先选取合适的数据分析方法来对指标进行分析，本文采用的是数据挖掘方法；然后通过建立模型分析各项指标对判定结果的影响，本文采用逻辑回归模型；最后通过比较逻辑回归结果来得出判定特定成分存在的主要指标。

## 模型建立

## 模型求解

# 问题二：成分存在与否的判定是否存在模糊区域

## 问题分析

这里写问题2的分析。

# 问题三：建立数学模型，判断测试数据

## 问题分析

这里写问题3的分析。

# 模型分析与评价

## 模型优点

这里写模型优点。

## 模型改进

这里写模型改进。

**参考文献**

[1]迟畅. 基于主成分的回归和决策树对高血压数据的处理[D].吉林大学,2019.