**实验三 探索性数据分析方法与应用**

【实验目的】

1 了解频率和众数的含义与计算。

2 掌握百分位数的计算方法。

3 掌握常见集中趋势度量的方法与计算。

4 掌握常见分散程度的度量方法与计算。

5 熟悉常见数据分析图表的绘制方法。

【实验类型】

设计型

【实验学时】

2学时

【实验环境】

Windows 7以上操作系统

Python3.0以上版本

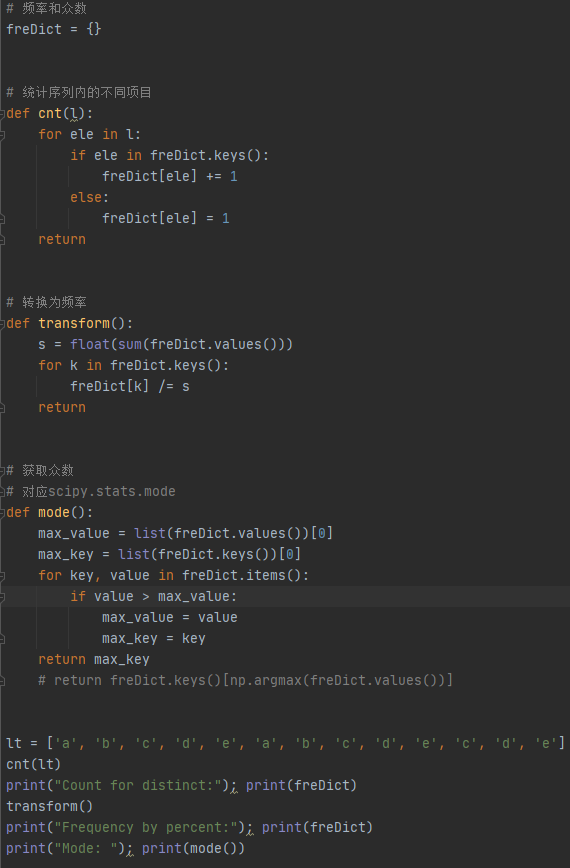
Pycharm开发环境

Spyder开发环境

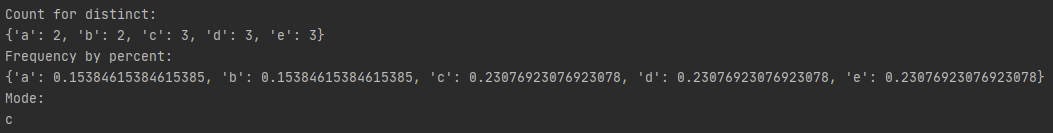
【实验要求】

一、频率和众数是简单描述数据分布状况的常见度量，请编写函数实现序列元素频率序列及其众数的计算，并自行构建数据验证方法。

**参考答案：**



**参考输出结果：**

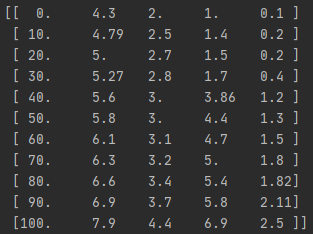


二、百分位数也是简单描述数据分布特征的常用度量，请编写函数实现数据序列百分位数的计算，并计算iris数据集中四个属性的百分位数。

**参考答案：**

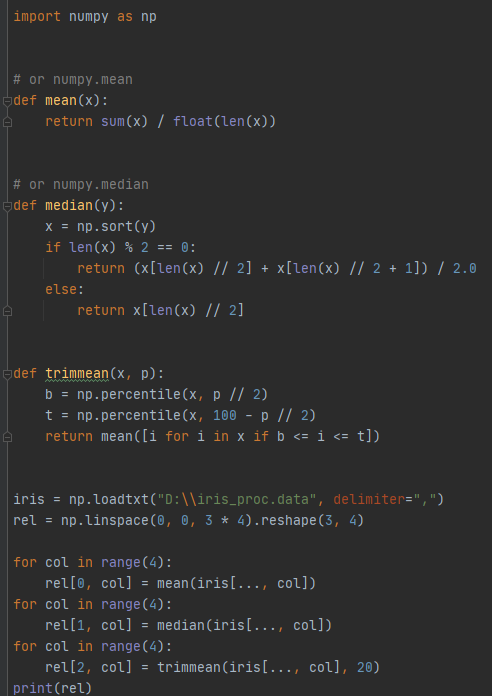


**参考输出结果：**

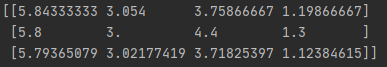


三、衡量数据序列集中程度的统计量通常由均值、中位数和截断均值，请编写函数分别实现均值、中位数和截断均值的计算，并分别针对iris数据集的四个属性进行计算。

**参考答案：**

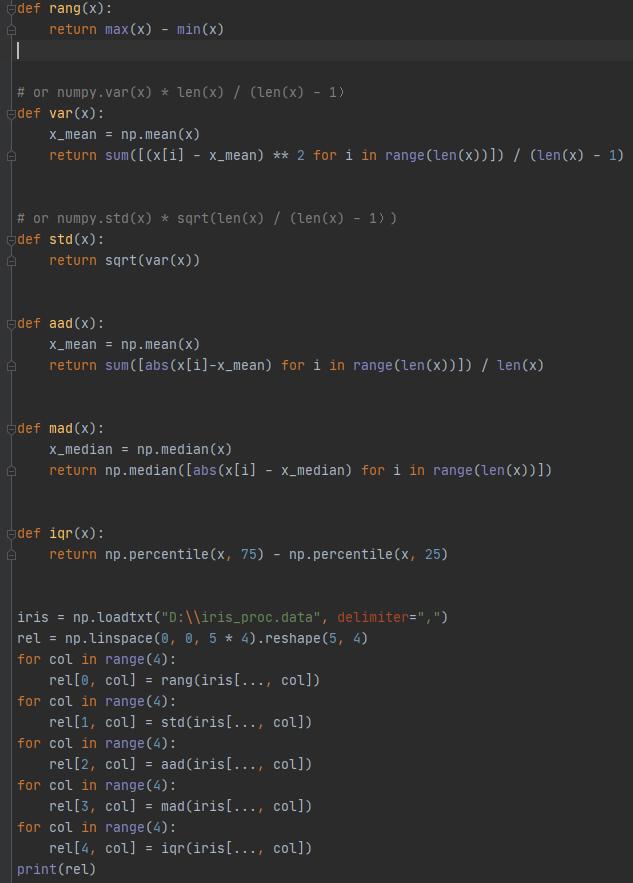


**参考输出结果：**

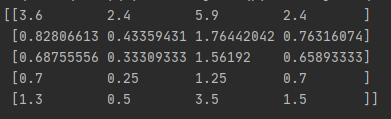


四、简单描述数据序列分散程度的统计量通常由极差、标准差、绝对平均偏差（AAD）、中位数绝对偏差（MAD）、四分位数极差（IQR）等，请编写函数实现这些统计量的计算，并针对iris数据集的四个属性进行计算。

**参考答案：**

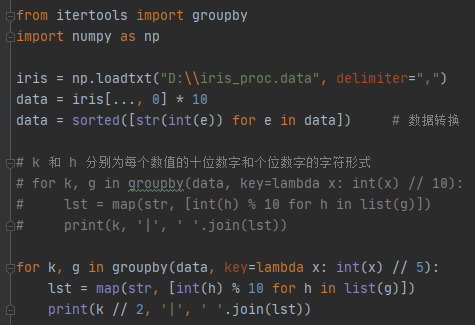


**参考输出结果:**



五、茎叶图是描述数据分布的一种简单可视化方法，请变成实现茎叶图的输出，完成iris数据集中的第一个属性萼片长度的茎叶图输出。

**参考答案：**



**参考输出结果：**



