**第2章 数据及数据分析基础**

**【练习题】**

1. **数据类型分类**  
   请列举以下数据的类型，并说明它们属于哪种数据维度（分类属性、有序属性或数值属性）。
2. 学生的性别  
   b) T恤的尺码（小、中、大）  
   c) 一年中的平均气温
3. **数据的基本统计特征**  
   假设一个数据集中有以下一组成绩数据：65,72,79,85,88,90,95。  
   a) 计算该数据集的均值、中位数、方差和标准差。  
   b) 请解释这些统计特征在数据分析中的意义。
4. **距离计算方法**  
   已知两个二进制字符串为："1010101" 和 "1110001"。  
   a) 使用Jaccard系数计算这两个字符串的距离。  
   b) 在数据元素为数值属性的情况下，如何计算两个元素之间的欧式距离？
5. **聚类分析**  
   a) 简述K-Means聚类算法的基本步骤。  
   b) 该算法的优缺点是什么？  
   c) 当使用K-Means算法进行聚类时，如何影响聚类结果的准确性？
6. **数据属性间的关联**  
   a) 请简述相关性分析与因果关系的区别。  
   b) 皮尔森相关系数的取值范围是什么？该系数为1和为0时分别表示什么？
7. **回归分析**  
   a) 线性回归的目标是什么？  
   b) 如何通过回归分析来预测一个变量的取值？请举一个实际生活中的应用例子。