

第一次作业要求

(布置时间: 9月18日, **提交时间: 9月25日上课前**)

一、 第一部分 (教材习题, 20 分)

完成教材《数据挖掘概念与技术》第二章习题 (P53) 中的如下习题:

- (1) 习题 2.2 (5 分)
- (2) 习题 2.3 (5 分)
- (3) 习题 2.4 (5 分)
- (4) 习题 2.6 (5 分)

二、 第二部分 (数据集 1 分析, 40 分)

1. 选用合适的统计描述指标和图形表达方式, 分析属性 Q1, Q2, Q3, Q4, Q7, Q11 的中心趋势和分散趋势特征。

- (1) 各个属性的类型 (二元、标称、序数、区间数值、比率数值等) (5 分)
- (2) 各个属性的中心趋势和分散趋势特征 (15 分)
- (3) 分析属性 Q2 (过去 2 年的违章次数) 与其他各个属性之间的关联性 (20 分)

三、 第三部分 (数据集 2 分析, 40 分)

2. 选用合适的统计描述指标和图形表达方式, 分析停车线周期流量、排队长度和实时饱和流率的总体统计特征、时变规律和周期变化规律。

- (1) 相位 4 的停车线周期流量、排队长度和实时饱和流率的中心趋势和分散趋势 (10 分)
- (2) 相位 8 的停车线周期流量、排队长度和实时饱和流率的总体统计特征 (一天)、时变规律和周期变化规律 (10 分)
- (3) 相位 4 和相位 8 的停车线周期流量、排队长度和实时饱和流率的关联性 (20 分)