

因此, 继续学习本章知识时应重视已知的综合应用。

习 题

一、选择题

1. () 主要代表的是数据的形态是否符合计算与算法要求。
A. 规整数据 B. 干净数据 C. 算法数据 D. 抽样数据
2. 规整数据应满足的基本原则有 ()。
A. 每个实体占且仅占一个关系 (表) B. 每个观察占且仅占一行
C. 每个变量占且仅占一列 D. 每一类观察单元构成一个关系 (表)
3. 数据科学中的 3C 精神指 ()。
A. 团队合作 B. 批判性思考 C. 原创设计 D. 好奇心提问
4. () 指对已有的数据在尽量少的先验假定下进行探索, 并通过作图、制表、方程拟合、计算特征量等手段探索数据的结构和规律的一种数据分析方法。
A. 预测性分析 B. 描述性分析 C. 探索性数据分析 D. 诊断性分析
5. () 指对于数据局部不良行为的非敏感性, 它是探索性数据分析追求的主要目标之一。
A. 鲁棒性 B. 稳定性 C. 可靠性 D. 耐抗性
6. 残差指 ()。
A. 观察值与均值之差 B. 实际观察值与估计值之差
C. 最大值和最小值之差 D. 预测值和实际值之差
7. 在探索性数据分析中, 通常用 () 描述数据分布是否为对称。
A. 正态分布 B. T 分布 C. 偏态 D. 峰态
8. () 的主要目的是提高数据质量, 使数据形态更加符合某一算法需求, 进而提高数据计算的效率并降低其复杂度。
A. 数据加工 B. 数据分析 C. 数据挖掘 D. 数据处理
9. 缺失数据的处理主要涉及的关键活动包括 ()。
A. 缺失数据的识别 B. 缺失数据的分析
C. 删除或插补缺失数据 D. 缺失数据的处理



10. 噪声数据的主要表现形式有 ()。
- A. 错误数据 B. 虚假数据
C. 异常数据 D. 缺失数据
11. 数据脱敏操作必须满足 ()。
- A. 单向性 B. 无残留
C. 易于实现 D. 不对称加密
12. 常用的数据规约方法有 ()。
- A. 目标向量规约
B. 特征矩阵规约
C. 维规约 D. 值规约
13. 预定义审计中可以依据的数据或方法有 ()。
- A. 数据字典
B. 用户自定义的完整性约束条件, 如字段“年龄”的取值范围为 20~40
C. 数据的自描述性信息, 如数字指纹(数字摘要)、校验码、XML Schema 定义
D. 属性的定义域与值域
14. 一个好的 Hash 函数应具备以下几个特征 ()。
- A. 容易计算 B. 便于加密 C. 单向性 D. 抗碰撞性

二、调研与分析题

1. 结合自己的专业领域, 调研该领域的数据预处理方法、技术与工具。
2. 调查研究典型的 2~3 个数据预处理工具(产品), 并探讨其关键技术和主要特征。
3. 调查分析关系数据库中常用的数据预处理方法。
4. 调查一项具体的数据科学项目, 分析其数据预处理活动, 并讨论预处理活动与数据计算活动之间的联系。

