



1. 数据基础

1.1 定义数据的概念

1.2 描述基本数据变量类型

- 布尔、数字、字符串

1.3 描述数据分析中使用的基本结构

- 表、行、列和列表

1.4 描述数据类别

- 定性、定量、元数据、大数据

2. 数据操作

2.1 导入、存储和导出数据

- ETL（提取、转换和加载）过程、数据操作工具（SQL、R、Python）、通用数据存储文件格式（分隔数据文件、XML、JSON）

2.2 清洁数据

- 目的和常见做法（处理空字符、特殊字符、修剪空格、格式不一致、删除重复项等）；验证数据

2.3 组织数据

- 目的和常见做法（排序、过滤、切片、转置、附加、截断等）

2.4 汇总数据

- 目的和常见做法（分组、合并、汇总、数据透视等）

3. 数据分析

3.1 描述和区分数据分析类型

- 描述性分析、诊断分析、假设检验、预测分析、规范分析

3.2 描述和区分数据聚合和解释指标

- 搜索、过滤、唯一值、聚合函数，如总和、最大值、最小值、计数、平均值/均值、模式、中值、标准偏差

3.3 描述和区分探索性数据分析方法

- 识别数据关系，描述数据钻探，描述数据挖掘概念（异常、相关性分析、模式、异常值等）

3.4 评估和解释数据分析结果

- 计算趋势，确定预期值，解释预测模型、p值、t检验、回归分析的结果

3.5 定义和描述人工智能在数据分析中的作用

- 定义人工智能、机器学习和算法；描述人工智能如何用于数据分析；描述如何在数据分析中使用机器学习算法（注意：特定算法不在知识点内）

4. 数据可视化和通信

4.1 报告数据

- 在表格和图表中有效显示信息；解释分解数据的时间和原因

4.2 从数据创建可视化

- 确定将误解可能性降至最低的数据可视化实践；确定表示底层数据结构和分析问题的可视化类型，包括比较、时间/趋势、部分到整体、关系、分布、相关图、方框图和箱型图、散点图（scatter chart、scatter plot）、条形图、桑基图、直方图、饼图、柱状图等

4.3 从数据可视化中得出结论

- 将数据的可视化表示转化为文字；根据分析及其图形表示claim之间的差异

4. 负责任的分析实践

5.1 描述数据隐私法和最佳实践

- GDPR、FERPA、HIPAA、IRB、PCI等。

5.2 描述负责任地处理数据的最佳实践

- 处理PII、保护数据安全和保护小数据集匿名性的方法；数据匿名化的重要性；平衡可解释性和准确性；利用有限样本数据进行总体级别概括的缺点

5.3 给定场景，描述影响数据收集和解释的偏差类型

- 确认偏差、人类认知偏差、动机偏差和抽样偏差；选择可视化/数据表示以避免偏差

