

试卷代号:24154

座位号

国家开放大学2023年春季学期期末统一考试

### 大数据技术概论 试题

2023年7月

题号	一	二	三	总分
分数				

得分	评卷人

一、单项选择题(本题共 20 小题,每小题 3 分,共 60 分。请在给出的选项中,选出最符合题目要求的一项)

- 下列各项不属于数据的是( )。
  - A. 文本
  - B. 图像
  - C. 视频
  - D. 印象
- 下列各项不属于大数据特征的是( )。
  - A. 体量大
  - B. 种类多
  - C. 真实性
  - D. 数据生成慢
- 数据异常值的处理方法不包括( )。
  - A. 极小值替换
  - B. 删除
  - C. 忽略
  - D. 视为缺失值进行填补
- 下列各项不能用于描述数据集中趋势的是( )。
  - A. 方差
  - B. 平均数
  - C. 中位数
  - D. 峰值
- 下列各项不属于 Hadoop 的特点是( )。
  - A. 存储迅速
  - B. 成本高
  - C. 计算能力强
  - D. 灵活性强

○-○-○-  
 学号  
 姓名  
 分校(工作站)  
 ○-○-○-

题  
答  
要  
不  
内  
线  
封  
密

6. 在工业网络实时监控系统中,需要连续不断地采集和处理数据。以下( )不属于这种计算模式。

- A. 在线处理
- B. 实时处理
- C. 流式计算
- D. 批量计算

7. 下面不是研究数据方法的是( )。

- A. 统计学
- B. 机器学习
- C. 心理分析
- D. 数据挖掘

8. 下面不属于大数据的处理过程的是( )。

- A. 数据获取
- B. 数据清洗
- C. 数据分析
- D. 数据安全

9. 下面不属于大数据计算模式的类型的是( )。

- A. 批量计算
- B. 手动计算
- C. 流式计算
- D. 交互式计算

10. 下列各项属于合规数据的是( )。

- A. 非法收集隐私信息数据
- B. 取得使用者同意的个人资料数据
- C. 泄露的隐私信息数据
- D. 垄断数据

11. HDFS 架构中的主节点(名称节点)负责( )。

- A. 接收来自客户端的作业请求
- B. 接收来自用户的数据读取请求
- C. 创建数据块
- D. 删除数据块

12. 下列有关 SQL 数据库的局限性,描述错误的是( )。
- A. 结构化数据建模
  - B. 已有数据结构的变化成本高
  - C. 空间利用率低
  - D. 无法保证事务处理的可靠性
13. 为展示两个数值型变量之间的相关关系,常用的可视化方法是( )。
- A. 箱线图
  - B. 直方图
  - C. 饼图
  - D. 散点图
14. 无法用来反映高维数据之间关系的可视化方法是( )。
- A. 相关系数矩阵图
  - B. 散点图矩阵
  - C. 气泡图
  - D. 平行坐标图
15. 噪声是指测量数据中的( )。
- A. 随机误差
  - B. 均值
  - C. 方差
  - D. 极差
16. 下列各项不属于大数据在金融领域的应用的是( )。
- A. 风险管理
  - B. 客户分析
  - C. 运营优化
  - D. 社交网络舆情
17. 按照预先设定的搜索规则,编写程序自动浏览并下载目标网页信息的数据采集方法称作( )。
- A. 系统日志信息数据采集方法
  - B. ETL 工具数据采集方法
  - C. 网络爬虫数据采集方法
  - D. 传感器数据采集方法
18. 关于传统数据与大数据的获取与处理,下列描述中错误的是( )。
- A. 传统的数据获取来源单一
  - B. 采用关系型数据库和并行数据库可以处理传统数据
  - C. 获取大数据的数据来源广泛
  - D. 大数据的数据量相对较小

19. 常见的离散化方法中,( )的基本思想是将样本等分成若干份,各组所包含的样本个数相同,组限是相应分位数上的分位值。

- A. 组距分组
- B. 分位数分组
- C. 秩分组
- D. 均值-标准差分组

20. 下列各项不属于数据仓库的特点的是( )。

- A. 将分析处理与事务数据库分离
- B. 整合来自不同数据源、不同类型的数据
- C. 数据质量高
- D. 数据仓库是面向应用而设计的

得 分	评卷人

二、判断题(本题共 10 小题,每小题 2 分,共 20 分。以下叙述中,你认为正确的打“√”,错误的打“×”)

- 21. 根据数据在收集过程中是否控制有关因素,可以将数据分为观测数据和实验数据。( )
- 22. 时间序列分析中采用对数变换来消除异方差。( )
- 23. 关系型数据库不是用来存储和访问具有彼此相关性数据的数据库。( )
- 24. 气泡图中气泡的面积大小没有实际意义。( )
- 25. 数据科学是通过科学方法探索数据,以获得有价值的发现。( )
- 26. 数据科学的发展不仅可以推动学科的发展,而且能够助推相关产业的发展与进步。( )
- 27. 网页数据是一种半结构化数据。( )
- 28. 批量数据处理是一种在线计算模式,可以满足系统对数据进行实时处理的需要。( )
- 29. 银行业通过大数据技术可以有效分析经营过程中可能存在的风险因素。( )
- 30. 适当地使用数据脱敏技术,可以有效地减少敏感数据在采集、传输、使用等环节中的暴露,降低敏感数据泄露的风险。( )

密 封 线 内 不 要 答 题

得 分	评卷人

三、简答题(本题共 4 小题,每小题 5 分,共 20 分)

31. 数据的定义是什么?

32. 大数据的成因是什么?

33. 数据整理的内容主要包括哪四个方面？

34. 通过相关系数矩阵处理共线性问题的算法步骤是什么？