

试卷代号:24154

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 试题答案及评分标准

(供参考)

2023年7月

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)

- | | | | | |
|-------|-------|-------|-------|-------|
| 1. D | 2. D | 3. A | 4. D | 5. B |
| 6. D | 7. C | 8. D | 9. B | 10. B |
| 11. A | 12. D | 13. D | 14. C | 15. A |
| 16. D | 17. C | 18. D | 19. B | 20. D |

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. √ | 22. √ | 23. × | 24. × | 25. √ |
| 26. √ | 27. √ | 28. × | 29. √ | 30. √ |

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?

数据是对现象或事物进行测量和记录的结果,可用于制表、计算和分析等,也可以统指一切保存在电脑中的信息,能够进行电子化的记录,包括文本、图像、音频、视频等。

32. 大数据的成因是什么?

数据的存储和管理能力的增强;数据采集能力增强;大数据的挖掘和分析等技术的同步发展。

33. 数据整理的内容主要包括哪四个方面?

数据的提取;数据的连接;数据的聚合;去除冗余和重复。

34. 通过相关系数矩阵处理共线性问题的算法步骤是什么?

计算相关系数矩阵;确定最大的成对相关系数对应的预测变量 A 和预测变量 B;计算变量 A 与其他所有变量之间的平均绝对值相关系数,对变量 B 也做同样的计算;比较 A 与 B,谁的平均绝对值相关系数最大,删除谁;重复以上步骤,直到两两之间绝对值相关系数低于某一特定阈值。