

CDA LEVEL II 考试大纲

CERTIFIED DATA ANALYST LEVEL II EXAMINATION OUTLINE

一、总体目标

CDA (Certified Data Analyst), 即“CDA 数据分析师”, 是在数字经济大背景和人工智能时代趋势下, 面向全行业的专业技能认证, 旨在提升全球用户数字技能, 助力企业数字化转型, 推动行业数字化发展。「CDA 人才考核标准」是面向全行业数据相关岗位的一套科学化、专业化、国际化的人才技能准则, CDA 考试大纲规定并明确了数据分析师认证考试的具体范围、内容和知识点, 考生可按照大纲要求进行相关知识的学习, 获取技能, 成为专业人才。

二、考试形式与试卷结构

考试方式: 线下考试, 上机答题

考试题型: 客观选择题 (单选 85 题+多选 20 题+内容相关 10 题+案例分析 10 题)

考试时间: 120 分钟

考试成绩: 分为 A、B、C、D 四个层次, A、B、C 为通过考试, D 为不通过

考试要求: 闭卷上机答题, 请勿携带与考试无关的用品

三、知识要求

针对不同知识, 掌握程度的要求分为【领会】、【熟知】、【应用】三个级别, 考生应按照国家不同知识要求进行学习。

1. 领会: 考生能够了解规定的知识点, 并能够了解规定知识的内涵与外延, 了解其内容要点之间的区别与联系, 并能做出正确的阐述、解释和说明。
2. 熟知: 考生须掌握知识的要点, 并能够正确理解和记忆相关理论方法, 能够根据不同要求, 做出逻辑严密的解释、说明和阐述。此部分为考试的重点部分。
3. 应用: 考生须学会将知识点落地实践, 并能够结合相关工具(只考察 Python)进行商业应用, 能够根据具体要求, 给出问题的具体实施流程和策略。

四、考试科目

- ◆ **PART 1 数据分析思维（占比 2%）**
 1. 数字化时代的数据思维
 2. 数据思维在企业管理中的作用

- ◆ **PART 2 量化策略分析方法（占比 6%）**
 1. 量化策略分析框架
 2. 量化策略分析流程

- ◆ **PART 3 数据的采集与治理（占比 26%）**
 1. 数据采集方法
 2. 数据读取
 3. 数据整合
 4. 数据清洗
 5. 特征处理

- ◆ **PART 4 数据可视化（占比 12%）**
 1. 可视化的数据探索与统计分析
 2. 统计制图

- ◆ **PART 5 统计分析方法 I（占比 28%）**
 1. 假设检验
 2. 列联表分析与卡方检验
 3. 方差分析与 F 检验
 4. 相关系数
 5. 线性回归
 6. 逻辑回归

- ◆ **PART 6 统计分析方法 II（占比 20%）**
 1. 主成分分析
 2. 因子分析
 3. 聚类分析
 4. 决策树分析

◆ PART 7 数据管理与数据安全（占比 6%）

1. 企业数据管理方法论
2. 企业数据安全方法论

五、科目内容**PART 1 数据分析思维****◆ 1、数字化时代的数据思维****【领会】**

- 设计思维
- 敏捷思维
- 数据思维

◆ 2、数据思维在企业管理中的作用**【领会】**

- 运营管理和项目管理
- 在双钻模型中应用数据思维

PART 2 量化策略分析方法**◆ 1、量化策略分析框架****【领会】**

- 感知型企业的各个阶段的特征
- EDIT 数字化模型

【熟知】

- 企业分析的探索、诊断、指导阶段
- 常用的定量分析方法

◆ 2、量化策略分析流程**【领会】**

企业分析决策的常见模式

【熟知】

企业量化分析的流程

近因分析与根因分析方法

宏微观预测

制定与验证方案

PART 3 数据的采集与处理

◆ 1、数据采集方法

【领会】

一手数据与二手数据

【熟知】

概率抽样方法（简单随机抽样、分层抽样、系统抽样、分段抽样等）

非概率抽样方法（定额抽样、滚雪球抽样等）

各种抽样方法的区别与优缺点

问卷调查的基本流程与调研方法

问卷设计原理

问卷题型设置以及每类题型的数据编码

【应用】

根据给定条件选择最适合的抽样方式

计算概率抽样所需的样本量

实施问卷调查

◆ 2、数据读取

【领会】

常见数据源的特点及获取方法

【熟知】

结构化数据的特点

数据表的数据特点

【应用】

可以根据数据格式的不同，应用相关软件设置参数获取数据

◆ 3、数据整合

【领会】

数据整合的思路框架、意义

【熟知】

行、列的选择（切片）

数据的转换操作

横向拼接

纵向拼接

分组聚合

排序

【应用】

基于需求加工整合数据

◆ 4、数据清晰

【熟知】

重复值的识别、处理

缺失值的识别、处理

错误值的识别、处理

【应用】

能对存在重复、缺失、错误的数据做合理的加工处理

◆ 5、特征处理

【领会】

特征处理在算法建模中的重要意义和作用

【熟知】

分类型特征的编码处理

分类型特征的概化处理

连续型特征离群值的识别与处理

连续型特征的分布形态转换

连续型特征的中心标准化或归一化

特征降维（含特征筛选）

特征变换（非线性变换）

特征分箱

WoE 转换

【应用】

对数据和特征进行处理，以优化算法建模和数据分析的结果

PART 4 数据可视化

◆ 1、可视化的数据探索与统计分析

【领会】

描述性统计分析在数据可视化中的作用

【熟知】

分类型变量的常用统计量

连续型变量的分布、集中趋势、离散程度与相关统计量

数据分布的偏态与峰态

【应用】

借助可视化图形，基于相关统计指标探索数据

◆ 2、统计制图

【熟知】

常用的统计图表类型

统计制图的数据加工方法

【应用】

对数据进行处理，以达到制图的要求

根据分析意图对数据进行深度加工

确定分析类型

选择合适的图表类型，绘制并进行展示

PART 5 统计分析方法 I**◆ 1、假设检验****【领会】**

常用的随机变量分布：两点分布、二项分布、均匀分布、泊松分布、正态分布

常用的统计量分布： χ^2 分布、T 分布、F 分布

假设检验的基本概念、步骤

假设检验与区间估计的联系

假设检验中的两类错误

【熟知】

检验统计量、显著性水平及对应临界值（Critical Value）的基本定义

P 值的含义及计算

如何利用 P 值进行检验

t 检验统计量（含 z 检验统计量，两者不做区分）

F 检验统计量

其他常用检验统计量

【应用】

实现单样本检验

实现两独立样本、配对样本检验

各种检验应用的数据分析场景

◆ 2、列联表分析与卡方检验**【领会】**

列联表的基本概念

卡方检验的基本概念与步骤

【熟知】

列联表的构造与计算

卡方检验的原假设与统计量

应用 P 值进行卡方检验

【应用】

构造列联表，使用卡方检验进行分类数据分析

◆ 3、方差分析与 F 检验

【领会】

方差分析的基本概念
单因素与多因素方差分析的原理
方差分析的 F 统计量构造过程

【熟知】

单因素方差分析的基本步骤
总离差平方和（SST）、组间离差平方和（SSA）、组内离差平方和（SSE）的含义及计算
单因素方差分析的原假设及相应计算
多因素方差分析的种类及相应计算

【应用】

实现单因素方差分析的步骤
实现多因素方差分析的步骤
对方差分析表的解读

◆ 4、相关系数**【熟知】**

连续型变量间关系的描述：散点图
协方差的概念与计算
相关系数的计算与检验（含 Pearson、Spearman 等系数）

◆ 5、线性回归**【领会】**

相关分析散点图的绘制与作用
横截面和时间序列数据在回归建模上的差异

【熟知】

回归系数的最小二乘估计
线性回归的经典线性假设
经典线性假设的检验，及违反假设后出现的问题以及解决方法
模型评估、模型检验、系数检验
回归变量的筛选（调整 R 方，AIC，BIC 指标，逐步回归法）

【应用】

结合业务构建回归模型并解释回归系数

根据业务场景与变量分布情况进行函数转换

解释变量为分类变量时的处理方法

◆ 6、逻辑回归

【领会】

二分类逻辑回归的基本概念及参数估计

【熟知】

Logit 变换 (Sigmoid 函数) 及其他常用变换

二分类逻辑回归模型构建与变量筛选

分类模型评估的方法 (混淆矩阵、ROC 曲线、由混淆矩阵衍生出来的评估指标)

选择最优分类阈值的方法

【应用】

结合业务构建分类模型并且解释回归系数

根据模型结果进行新样本预测

PART 6 统计分析方法 II

◆ 1、主成分分析

【领会】

矩阵分析法的基本思想

主成分分析的原理

主成分分析的计算步骤

【熟知】

适用于主成分分析的变量类型

保留主成分个数的原则

特征值、特征向量、主成分三者的概念和关系

主成分分析前数据标准化的原因和计算方法

【应用】

明确主成分分析的适用场景

根据业务场景或建模需求, 决定是否使用主成分分析方法

◆ 2、因子分析

【领会】

因子分析模型的概念和计算（只考察主成分法的计算步骤）

【熟知】

适用于因子分析的变量类型

通过分析结果，选取合适的因子个数

常用因子旋转的方法

主成分分析与因子分析的联系与区别

【应用】

在遇到业务问题时，有能力决定使用因子分析还是主成分分析

根据原始变量在各因子上的权重明确每个因子的意义

在聚类前对数据进行描述，发现理想的聚类方式和数量

◆ 3、聚类分析

【领会】

用户细分与聚类的关系

聚类分析的基本概念

【熟知】

聚类方法的基本逻辑

聚类模型的评估方法

层次聚类的基本算法（包括两点距离、两类合并的计算方法等，着重Ward法的计算步骤）

K-Means 聚类的基本算法

聚类分析前可以进行的数据处理（数据标准化、主成分分析等）

距离的度量

【应用】

结合客户细分、商品聚类、离群值检验（反欺诈）等业务运用场景，选取合适的聚类方法与步骤。

K-Means聚类后的结果解释，根据聚类后变量分布情况归纳每类的特征和群特征画像。

◆ 4、决策树分析**【领会】**

决策树的基本概念

【熟知】

决策树的结构（根节点、内部节点、叶节点）

决策树节点的分裂准则

决策路径与分类规则

决策树的可视化

【应用】

根据需求，建立决策树模型并应用到业务场景中

应用决策树解读K-Means聚类后的分群特征

PART 7 数据管理与数据安全**◆ 1、企业数据管理方法论****【领会】**

数据管理基本概念和关键术语

数据管理的输入与资源需求

数据治理流程

输出和绩效评估

数据管理的成熟度模型

数据价值评估与数据资产入表

◆ 2、企业数据安全方法论**【领会】**

数据安全的需求、定义

数据安全的管理目标和原则

输入和资源需求

数据生命周期安全管理流程

输出和绩效评估

数据安全能力成熟度模型 (DSMM)

六、推荐学习书目

说明：推荐学习书目中，考生可根据自身需求选择性学习。参考书目不需全部学完，根据考试大纲知识点进行针对性学习即可。

- [1] CDA数据科学研究院. CDA二级认证教材:量化策略分析 (第4版) [M]. 电子工业出版社, 2025. (必读)
- [2] 龚德恩. 经济数学基础 (第3分册): 概率统计 (第五版) [M]. 四川人民出版社, 2016. (选读)
- [3] 贾俊平, 何晓群, 金勇进. 统计学 (第8版) [M]. 中国人民大学出版社, 2021. (选读)
- [4] 理查德等著, 陆璇等译. 实用多元统计分析 [M]. 清华大学出版社, 2008. (选读)
- [5] 李卫东. 应用多元统计分析 (第2版) [M]. 北京大学出版社, 2015. (选读)
- [6] 魏建国、曾珂、翟锬、常国珍. 金融数据分析和数据挖掘案例实战 (双色) [M]. 电子工业出版社, 2025. (选读)
- [7] 伍德里奇. 计量经济学导论 (第4版) [M]. 人民大学出版社, 2016. (选读)

CDA 数据分析认证考试委员会
CDA Institute

注：考试题库中约有5%的超纲题