

# 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, 面向全行业的专业权威国际资格认证, 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 100 题+多选 20 题+内容相关 15 题+案例分析 15 题)

考试时间: 150 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, 并能够使用相关数据分析工具(如Excel, Python等)进行实操。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 数据采集与处理（占比 9%）

- a. 数据采集方法（占比 1%）
- b. 市场调研和数据录入（占比 3%）
  - 市场调研流程
  - 样本选取
  - 问卷设计及录入
- c. 数据探索与可视化（占比 2%）
- d. 数据预处理方法（占比 3%）

### ◆ PART 2 指标体系（占比 3%）

- a. 指标标准
- b. 指标体系的定义和意义
- c. 构建指标库

### ◆ PART 3 数据治理与数据模型管理（占比 6%）

- a. 数据治理（占比 3%）
  - 数据治理的驱动因素
  - 数据治理体系
  - 数据应用域
- b. 数据模型管理（占比 3%）
  - 数据分类
  - 关系模型
  - 数据仓库体系和 ETL

### ◆ PART 4 标签体系与用户画像（占比 5%）

- a. 标签体系设计原理（占比 3%）
- b. 标签的加工方式（占比 1%）
- c. 用户画像（占比 1%）

### ◆ PART 5 统计分析（占比 25%）

- a. 抽样估计（占比 5%）

- b. 假设检验（占比 5%）
- c. 方差分析（占比 5%）
- d. 一元线性回归分析（占比 10%）

◆ **PART 6 数据分析模型（占比 40%）**

- a. 主成分分析法、因子分析法（占比 10%）
- b. 多元回归分析法
  - 多元线性回归（占比 10%）
  - 逻辑回归（占比 10%）
- c. 聚类分析法
  - 系统聚类法（占比 2%）
  - K-Means 聚类法（占比 3%）
- d. 时间序列（占比 5%）

◆ **PART 7 数字化工作方法（占比 12%）**

- a. 业务探查与问题定位（占比 3%）
- b. 问题诊断
  - 近因分析（占比 3%）
  - 根本原因分析（占比 2%）
- c. 业务策略优化和指导
  - 业务目标设定原则（占比 1%）
  - 知识库，策略库，流程分析（占比 1%）
  - 线性和整数规划（占比 1%）
  - 二次优化（占比 1%）

## 五、科目内容

### PART 1 数据采集与处理

◆ **1、数据采集方法**

**【领会】**

一手数据与二手数据来源渠道

优劣势分析

使用注意事项

**【熟知】**

一手数据采集中的概率抽样与非概率抽样的区别与优缺点

**【运用】**

概率抽样方法，包括简单随机抽样、分层抽样、系统抽样、分段抽样明确每种抽样的优缺点

根据给定条件选择最可行的抽样方式

计算简单随机抽样所需的样本量

**◆ 2、市场调研和数据录入**

**【熟知】**

市场调研的基本步骤（提出问题、理论推演、收集材料、构建模型、归因分析）

样本选取方式的适应性及优缺点

问卷设计原理，问卷题型设置以及每类题型的数据编码及录入

**◆ 3、数据探索与可视化**

**【领会】**

数据探索的目的与意义

常用数据可视化工具软件（EXCEL BI、SPSS、PYTHON 等）

**【熟知】**

数据探索与数据预处理之间的关系

数据探索常用数据描述方法：集中趋势分析、离中趋势分析、数据分布关系、图分析

数据探索常用数理统计方法：假设检验、方差检验、相关分析、回归分析、因子分析

**【应用】**

能够通过使用数据可视化工具（EXCEL BI、SPSS、PYTHON 等）来完成相关数据分析项目的数据探索任务。（说明：考试中不会考核该部分工具和软件的使用方法）。

**◆ 4、数据预处理方法**

**【熟知】**

数据预处理的基本步骤，包括数据集成（不同数据源的整合）、数据探索、数据变换（标准化）、数据归约（维度归约技术、数值归约技术），这部分内容不涉及计算，只需要根据需求明确可选的处理技术即可。

**【应用】**

数据清洗，包括填补遗漏的数据值（根据业务场景使用常数、中位数、众数等方法，不

涉及多重查补的方法)、平滑有噪声数据(移动平均)、识别或除去异常值(单变量根据中心标准化值,多变量使用快速聚类),以及解决不一致问题(熟知概念即可),查重(只考核 SQL 的语句,不涉及其它语言)。

## **PART 2 指标体系**

### **【领会】**

指标标准

构建指标体系的意义

### **【熟知】**

指标体系的定义和概念

构建指标库

## **PART 3 数据治理与数据模型管理**

### **◆ 1、数据治理**

#### **【领会】**

数据治理的驱动因素

数据治理、数据管理、数据应用的概念

数据治理域、数据管理域、数据应用域各自包含的内容及工作

#### **【熟知】**

元数据、主数据、交易数据的概念

数据质量要求

### **◆ 2、数据模型管理**

#### **【领会】**

数据和信息的概念

数据库建模中概念、逻辑、物理模型之间的关系

数据库范式的概念、数据仓库和数据集市、ETL过程

#### **【熟知】**

关系模型与维度模型的使用场景

## PART 4 标签体系与用户画像

### ◆ 1、标签体系设计原理

#### 【领会】

区分标签和指标的概念

精准营销的概念

消费者决策进程

用户分群发展历程

#### 【熟知】

分层标签和分群标签

马斯洛需求层次理论

用户标签和用户画像的关系

### ◆ 2、标签的加工方式

#### 【领会】

标签的时态角度分类

#### 【熟知】

基础、统计、模型标签

### ◆ 3、用户画像

#### 【领会】

细分市场和STP模型

用户画像的概念

标签在数据科学的作用

#### 【熟知】

用户细分的方法

用户分群的精准营销

用户画像在EDIT模型中的作用和应用

## PART 5 统计分析

### ◆ 1、抽样估计

#### 【领会】

随机试验、随机事件、随机变量的概念

总体与样本的概念  
抽样估计的理论基础  
正态分布及三大分布的图像形式  
三大分布的构造原理

**【熟知】**

随机事件的概率  
点估计与区间估计方法的特点与优缺点  
总体与样本  
参数和统计量  
重复抽样与不重复抽样  
常见的抽样分布

**【应用】**

随机变量及其概率分布  
根据不同的场景，选择合适的统计量进行总体的参数估计

**◆ 2、假设检验****【领会】**

假设检验的基本概念  
其基本思想在数据分析中的作用  
假设检验的基本步骤  
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  
z 检验统计量  
t 检验统计量  
F 检验统计量  
 $\chi^2$  检验统计量

**【应用】**

实现单样本  $t$  检验

实现配对样本  $t$  检验

两独立样本  $t$  检验的步骤和检验中使用的统计量与原假设

两种检验应用的数据分析场景

**◆ 3、方差分析****【领会】**

方差分析的相关概念

单因素方差分析的原理

多因素方差分析的原理

方差分析的统计量构造过程

**【熟知】**

单因素方差分析的基本步骤

总离差平方和（SST）的含义及计算

组间离差平方和（SSA）的含义及计算

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

多因素方差分析的种类

列联表检验的原理

**【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、一元线性回归分析****【领会】**

相关分析散点图的绘制与作用

相关系数的计算公式

**【熟知】**

相关分析的概念与特点  
相关关系与函数关系的区别与联系  
相关关系的种类及意义  
一元线性回归系数的最小二乘估计  
一元线性回归模型的检验

**【应用】**

计算相关系数与回归系数  
回归分析与相关分析的区别与联系

**PART 6 数据分析模型****总体要求**

领会模型基本原理，数值模型操作流程，懂得模型应用场景，能够完成数据建模分析报告。

**◆ 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  
主成分的个数的保留原则

**【应用】**

在深入理解主成分的意义的基础之上，在遇到业务问题时，有能力决定是否使用主成分分析方法  
明确主成分分析的适用场景

**◆ 2、因子分析****【领会】**

了解因子分析模型设置，只需要关注主成分法的计算步骤

**【熟知】**

适用于因子分析的变量度量类型，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

有能力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、回归分析

#### 【领会】

多元线性回归的原理

#### 【熟知】

使用最小二乘法计算多元回归模型的结果

明确线性回归的 6 个经典假设

明确违反上述假设后出现的问题以及解决方法

明晰横截面和时间序列数据在回归建模上的差异

模型评估、模型检验、系数检验

#### 【应用】

结合业务构建回归模型并且解释回归系数

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

### ◆ 4、分类分析

#### 【领会】

卡方检验计算原理与公式

二分类逻辑回归的计算公式

#### 【熟知】

分类变量是否存在相关关系的描述方法和检验方法，涉及列联表分析、卡方检验与 Logit

转换

二分类逻辑回归模型构建与变量筛选

模型评估的方法，涉及混淆矩阵、ROC 曲线、由混淆矩阵衍生出来的评估指标

#### 【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

进行客户流失预测、信用评级、精准营销等模型的基本步骤与注意事项

### ◆ 5、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】

聚类模型的评估方法

聚类方法的基本逻辑，涉及层次聚类、kmeans聚类

层次聚类和 K-Means 聚类的基本算法和优缺点

层次聚类的计算步骤，包括两点距离、两类合并的计算方法

K-Means 聚类的基本算法

聚类分析变量标准化的原因和计算方法

变量需要进行主成分分析的原因

距离的计算

#### 【应用】

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

聚类事后分析，根据聚类后变量分布情况获取每类的特征。

### ◆ 6、时间序列

#### 【领会】

时间序列分析原理

#### 【熟知】

时间序列分析的方法

明确趋势分解法、ARMA方法、ARIMA 方法差异和适用场景

ARIMA方法的建模流程

**【应用】**

结合业务（业绩预测、预警），选取合适的分析方法  
进行业务时间序列预测等模型的基本步骤与注意事项

**PART 7 数字化工作方法**

**◆ 1、业务探查与问题定位**

**【领会】**

异常事件严重度评估准则。  
业务流程等事件还原工具。

**【熟知】**

业务流程图的绘制。

**◆ 2、问题诊断**

**【领会】**

近因分析的头脑风暴法与量化分析工具的选取。  
根本原因分析中的 5WHY 分析法，原因型和对策型因果图。

**【熟知】**

通过帕累托分析识别要点。  
通过散点图、关联图、亲和图进行关联分析。  
通过漏斗分析、用户画像、留存分析、跟踪数字足迹进行探查。 绘制原因型因果图。

**◆ 3、业务策略优化和指导**

**【领会】**

业务目标设定原则  
线性规划的组成部分、标准形式。  
整数规划与去尾法线性规划的差异性。  
二次规划的组成部分、标准形式。  
知识库的类型和组成部分。  
策略库的类型和组成部分。

**【熟知】**

线性规划的建模步骤。

二次规划的建模步骤。

流程优化的分析方法和工具。

#### 【应用】

根据题目要求给出目标函数和约束条件。

## 六、推荐学习书目

说明：推荐学习书目中，部分书籍结合软件，但考试中不考查软件操作使用，考生可根据自身需求选择性学习。参考书目不需全部学完，根据考纲知识点进行针对性学习即可。

- [1] 常国珍等. 商业策略数据分析. 电子工业出版社, 2023. (必读)
- [2] 赵仁乾, 常国珍等. 金融商业算法建模: 基于python和SAS [M]. 机械工业出版社, 2021. (必读)
- [3] 比约恩等. 根原因分析-简化的工具和技术(第2版) [M]. 中国人民大学出版社, 2011. (必读)
- [4] 常国珍等. Python数据科学: 技术详解与商业实践 [M]. 机械工业出版社, 2018. (选读)
- [5] 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)
- [6] 金勇进, 杜子芳等. 抽样技术(第5版). 中国人民大学出版社, 2021. (选读)
- [7] 数据管理协会(DAMA国际). DAMA数据管理知识体系指南(原书第2版) [M]. 机械工业出版社, 2020. (选读)
- [8] 韩伯棠, 管理运筹学(第5版) [M]. 高等教育出版社, 2020. (选读)
- [9] 任寅姿, 季乐乐等. 标签类目体系 [M]. 机械工业出版社, 2021. (选读)

