

# CPDA 数据分析师专业技术考试大纲

(2021 年修订版)

## 考试介绍

### 一、考试目标

CPDA 数据分析师专业技术考试主要测试考生是否具备数据分析思维，是否具备数据分析基础知识，是否掌握数据分析工作流程及数据分析技术，是否具备利用数据分析技术解决实际业务问题的能力。

### 二、考试科目及考试形式

考试科目包括数据分析理论知识、数据分析算法与模型、数据分析应用，考试方式分为理论机考和实操机考，满分都为 100 分。

科目	考试方式	题型	总分	及格分	考试时长
数据分析理论知识		单选题、多选题、判断题	100 分	60 分	90 分钟
数据分析算法与模型	机考	操作计算题	100 分	60 分	120 分钟
数据分析应用		应用分析题 (操作+分析)	100 分	60 分	120 分钟

CPDA 数据分析师的认证考核采取全国统一时间，每年四次；

其他特殊情况以考前通知为准；

具体信息，请查询官方网站 [www.chinacpda.com](http://www.chinacpda.com)。

### 三、参考书籍

《数据分析基础》，中国商业联合会数据分析专业委员会编著，中国商业出版社，2021 年

《客户与产品数据分析》，中国商业联合会数据分析专业委员会编著，中国商业出版社，2021年

《供应链优化与投资分析》，中国商业联合会数据分析专业委员会编著，中国商业出版社，2021年

## 四、考试范围和要求

考试范围涉及四大模块：数据与数据分析、数据分析思维与数据分析流程、数据分析中的常用方法和数据分析应用。

考试大纲对四大模块专业知识的要求分掌握、熟悉和了解三个层次。

**掌握**即要求考生具备解决实际问题的能力；

**熟悉**即要求考生对该知识具有深刻的理解；

**了解**即要求考生对该知识有正确的认知。

以下是对考点的详细介绍，数据分析是一门综合学科，部分考点存在交叉和重叠为考试正常需要。

## 第一部分 数据与数据分析

数据与数据分析模块考察考生对数据分析基础知识的掌握情况，考试范围涉及数据分析概述和数据分析与计算机技术 2 方面。考点可分布在数据分析理论知识、数据分析算法与模型、数据分析应用 3 个科目。具体如下：

### 一、数据分析概述

#### 1、大数据：大数据的概念、典型特征、大数据分析思维。

考点分布：数据分析理论知识

考点要求：了解

#### 2、前沿技术：与大数据相关的前沿技术。

考点分布：数据分析理论知识

考点要求：了解

## 二、数据分析与计算机技术

**1、excel 的基础操作：设置数据类型、数据的排序、筛选、查找和替换、数据透视表、数据的合并与拆分、数据可视化。**

考点分布：数据分析理论知识、数据分析算法与模型、数据分析应用

考点要求：掌握

**2、excel 函数：excel 常用函数的使用和操作，如 SUM()、SUMIF()、IF()、AVERAGE()、MEDIAN()、MODE()、STDEV()、GEOMEAN()、COUNT()、COUNTA()、COUNTIF()、MAX()、MIN()、VLOOKUP()、CORREL()等。**

考点分布：数据分析理论知识、数据分析算法与模型、数据分析应用

考点要求：掌握

**3、python 编程基础：python 基础知识、数据类型（数字，字符串，列表，元组，字典，集合）、条件和循环。**

考点分布：数据分析理论知识

考点要求：掌握

## 第二部分 数据分析思维与数据分析流程

数据分析思维与数据分析流程考察考生对数据分析各流程中知识点的掌握情况，考试范围涉及确定业务问题、数据获取、数据探索和预处理和决策分析 4 个方面。考点可分布在数据分析理论知识、数据分析算法与模型、数据分析应用 3 个科目。具体如下：

## 一、确定业务问题

### 1、确定和分解问题：如何确定问题，确定问题的结构化思维。

考点分布：数据分析理论知识

考点要求：熟悉

### 2、企业战略管理模型与应用：SWOT 分析、安索夫矩阵、BCG（波士顿）矩阵和 GE（通用电气）矩阵、PEST 分析和波特五力模型。

考点分布：数据分析理论知识

考点要求：熟悉

## 二、数据获取

### 1、数据获取的方法：内部数据获取和外部数据获取、传统数据和大数据获取方法与技术。

考点分布：数据分析理论知识

考点要求：熟悉

### 2、数据库与数据仓库：数据库与数据仓库的概念、特征和区别。

考点分布：数据分析理论知识

考点要求：熟悉

### 3、SQL 基础：SQL 的增、删、改、查命令语句。

考点分布：数据分析理论知识

考点要求：熟悉

### 三、 数据探索和预处理

#### 1、变量的类型：变量类型的划分、不同类型变量的特点。

考点分布：数据分析理论知识、数据分析算法与模型、数据分析应用

考点要求：掌握

#### 2、描述性统计中变量的均值、标准差、最大值、最小值、极差、中位数、众数、四分位数、四分位差、偏度和峰度、左偏和右偏、正态分布和相关操作。

考点分布：数据分析理论知识、数据分析算法与模型、数据分析应用

考点要求：掌握

#### 3、数据清洗中缺失值、异常值和数据不一致的处理方法和操作。

考点分布：数据分析理论知识、数据分析算法与模型、数据分析应用

考点要求：掌握

#### 4、相关性分析中皮尔逊相关系数、方差分析和卡方检验的原理、操作及结果解读。

考点分布：数据分析理论知识、数据分析算法与模型、数据分析应用

考点要求：掌握

#### 5、数据集成：数据集成的概念和意义。

考点分布：数据分析理论知识

考点要求：了解

#### 6、数据形态变换中的 Box-Cox 变换、作用和操作。

考点分布：数据分析理论知识、数据分析算法与模型、数据分析应用

考点要求：掌握

**7、标准化变换：最大最小值（Min-Max）标准化、零均值（Z-score）标准化的原理、作用和操作。**

考点分布：数据分析理论知识、数据分析算法与模型、数据分析应用

考点要求：掌握

**8、哑变量与独热编码：标签编码、哑变量和独热编码的原理、转换方法、区别和操作。**

考点分布：数据分析理论知识、数据分析算法与模型、数据分析应用

考点要求：掌握

**9、离散化：离散化的作用、等距离散和等频离散的原理和操作。**

考点分布：数据分析理论知识、数据分析算法与模型、数据分析应用

考点要求：掌握

**10、数据降维与因子分析：数据降维的作用和常用方法、因子分析与主成分分析的原理、区别、操作和结果解读。**

考点分布：数据分析理论知识、数据分析算法与模型、数据分析应用

考点要求：掌握

## 四、决策分析

**1、数据可视化技术：柱状图、条形图、直方图、折线图、散点图、气泡图、饼图、面积图、箱线图的操作。**

考点分布：数据分析理论知识、数据分析算法与模型、数据分析应用

考点要求：掌握

**2、数据可视化在案例中的应用：柱状图、条形图、直方图、折线图、散点图、气泡图、饼图、面积图、箱线图的应用与结果解读。**

考点分布：数据分析理论知识、数据分析算法与模型、数据分析应用

考点要求：掌握

### 第三部分 数据分析中的常用方法

数据分析中的常用方法考察考生对各类数据分析方法的掌握情况，考试范围涉及线性回归模型、分类算法、聚类算法、推荐算法和时间序列算法 5 个方面。考点可分布在数据分析理论知识、数据分析算法与模型、数据分析应用 3 个科目。具体如下：

#### 一、线性回归模型

**1、线性回归模型的原理与构造：一元和多元线性回归模型的原理、模型形式和操作。**

考点分布：数据分析理论知识、数据分析算法与模型、数据分析应用

考点要求：掌握

**2、线性回归模型的显著性检验：拟合优度检验、F 检验、t 检验原理和结果解读。**

考点分布：数据分析理论知识、数据分析算法与模型、数据分析应用

考点要求：掌握

**3、线性回归模型的评价和预测：R 方、回归方程的显著性、回归系数的显著性、对回归系数的解释、运用回归模型进行预测、预测区间和置信区间。**

考点分布：数据分析理论知识、数据分析算法与模型、数据分析应用

考点要求：掌握

## 二、 分类算法

**1、分类器的评价指标：训练集、测试集、正确率(Accuracy)、准确率(Precision)、召回率(Recall)、F-Score、AUC 值、ROC 曲线、样本均衡性、训练误差和泛化误差。**

考点分布：数据分析理论知识、数据分析算法与模型、数据分析应用

考点要求：掌握

**2、逻辑回归：逻辑回归的原理、模型形式、预处理、模型操作、结果解读和优缺点。**

考点分布：数据分析理论知识、数据分析算法与模型、数据分析应用

考点要求：掌握

**3、朴素贝叶斯：朴素贝叶斯的原理、计算公式、预处理、模型操作、结果解读和优缺点。**

考点分布：数据分析理论知识、数据分析算法与模型、数据分析应用

考点要求：掌握

**4、集成学习：集成学习的原理、预处理、模型操作、结果解读和优缺点。**

考点分布：数据分析理论知识、数据分析算法与模型、数据分析应用

考点要求：掌握

**5、决策树：决策树的原理、预处理、模型操作、结果解读和优缺点。**



考点分布：数据分析理论知识、数据分析算法与模型、数据分析应用

考点要求：掌握

**6、支持向量机（SVM）：支持向量机的原理、预处理、模型操作、结果解读和优缺点。**

考点分布：数据分析理论知识、数据分析算法与模型、数据分析应用

考点要求：掌握

**7、神经网络：神经网络的原理、预处理、模型操作、结果解读和优缺点。**

考点分布：数据分析理论知识、数据分析算法与模型、数据分析应用

考点要求：掌握

### 三、 聚类算法

**1、聚类算法的评价标准：聚类的原理、类别相似性、轮廓系数、平均轮廓系数、类中心点坐标。**

考点分布：数据分析理论知识、数据分析算法与模型、数据分析应用

考点要求：掌握

**2、K-means 聚类算法：K-means 聚类原理、对数据的要求、预处理、模型操作、结果解读和优缺点。**

考点分布：数据分析理论知识、数据分析算法与模型、数据分析应用

考点要求：掌握

**3、其他聚类算法：系统聚类和划分聚类的概念和原理、K-Medoids 聚类、基于密度的聚类（如：DBSCAN）等的原理和优缺点。**

考点分布：数据分析理论知识、数据分析算法与模型、数据分析应用

考点要求：掌握

## 四、推荐算法

1、关联规则 **Apriori** 算法: **Apriori** 算法原理 (支持度、置信度、项集、频繁项、有效的强关联规则等)、对数据的要求、预处理、模型操作、结果解读和优缺点。

考点分布: 数据分析理论知识、数据分析算法与模型、数据分析应用

考点要求: 掌握

2、协同过滤算法: 协同过滤算法的原理, 相似性度量、基于用户的相似性和基于产品的相似性推荐。

考点分布: 数据分析理论知识、数据分析算法与模型

考点要求: 熟悉

## 五、时间序列算法

1、时间序列原理: 时间序列的概念、平稳序列和非平稳序列、基本构成要素 (长期趋势、季节波动、循环变动和不规则变动)、加法模型和乘法模型、平滑指数、季节因子、周期长度和预测长度。

考点分布: 数据分析理论知识、数据分析算法与模型

考点要求: 掌握

2、时间序列算法: 移动平均、加权移动平均、一次指数平滑、**Holt** 指数平滑、**Winter** 指数平滑、季节分解法的原理、模型操作、结果解读和优缺点。

考点分布: 数据分析理论知识、数据分析算法与模型、数据分析应用

考点要求: 掌握

## 第四部分 数据分析应用

数据分析应用考察考生解决各类数据分析具体问题的能力，考试范围涉及客户分析、产品分析、供应链分析和投资分析 4 个方面。考点可分布在数据分析理论知识、数据分析算法与模型、数据分析应用 3 个科目。具体如下：

### 一、 客户分析

**1、客户分析概况：**客户分析相关概念，如客户满意度、客户忠诚度、客户价值等。

考点分布：数据分析理论知识

考点要求：了解

**2、客户画像：**用户画像的概念、标签、客户画像分析技术。

考点分布：数据分析理论知识、数据分析算法与模型、数据分析应用

考点要求：掌握

**3、RFM 客户价值分析：**RFM 模型原理、数据获取、模型操作、结果解读和优缺点。

考点分布：数据分析理论知识、数据分析算法与模型、数据分析应用

考点要求：掌握

**4、AARRR 模型：**AARRR 模型原理、漏斗模型可视化操作、各阶段分析指标和注意事项。

考点分布：数据分析理论知识、数据分析算法与模型

考点要求：掌握

## 二、 产品分析

**1、KANO 模型：**KANO 模型原理、产品属性划分、问卷设计、better/worse 系数计算、散点图可视化操作、结果解读和优缺点。

考点分布：数据分析理论知识、数据分析算法与模型

考点要求：掌握

**2、PSM 模型：**PSM 模型原理、问卷设计、可接受、可保留接受和不接受的人数百分比计算、可视化操作、结果解读和优缺点。

考点分布：数据分析理论知识、数据分析算法与模型

考点要求：掌握

**3、巴斯模型：**巴斯模型原理、创新系数  $p$  和模仿系数  $q$  等参数估计、模型操作、结果预测、可视化操作、结果解读和优缺点。

考点分布：数据分析理论知识、数据分析算法与模型

考点要求：掌握

**4、A/B test：**A/B test 原理和实现方法。

考点分布：数据分析理论知识

考点要求：熟悉

## 三、 供应链分析

**1、供应链数据分析概况：**供应链的概念、结构、数据分析在供应链中的应用。

考点分布：数据分析理论知识

考点要求：熟悉

**2、层次分析法：层次分析法原理、模型操作、结果解读和优缺点。**

考点分布：数据分析理论知识、数据分析算法与模型

考点要求：掌握

**3、规划求解：规划求解的原理、决策变量、目标函数、约束条件、模型操作、结果解读和优缺点。**

考点分布：数据分析理论知识、数据分析算法与模型

考点要求：掌握

**4、物流选址和物流配送：重心法和节约里程法的原理、操作、结果解读和优缺点。**

考点分布：数据分析理论知识、数据分析算法与模型

考点要求：掌握

**5、销售预测：销售预测常用定性和定量方法（如回归和时间序列法）。**

考点分布：数据分析理论知识、数据分析算法与模型、数据分析应用

考点要求：掌握

## **四、 投资分析**

**1、资金的时间价值：折现率、资金的现值、终值、年金的概念和计算。**

考点分布：数据分析理论知识

考点要求：掌握

**2、现金流量表：现金流量表的编制、现金流入项目、现金流出项目、净现金流量、所得税和固定资产回收调节税等。**

考点分布：数据分析理论知识

考点要求：掌握

**3、投资经济效益指标分析：静态指标、动态指标和指标间比较。**

考点分布：数据分析理论知识

考点要求：掌握

**4、风险投资决策分析：存在风险概率情况下不同方案的净现值比较。**

考点分布：数据分析理论知识

考点要求：掌握